

**ANÁLISE ESPACIAL DA MORTALIDADE
INFANTIL NO MUNICÍPIO DE ALFENAS, MG**

DENISMAR ALVES NOGUEIRA

2008

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

DENISMAR ALVES NOGUEIRA

**ANÁLISE ESPACIAL DA MORTALIDADE INFANTIL
NO MUNICÍPIO DE ALFENAS, MG**

Tese apresentada a Universidade Federal de Lavras
como parte das exigências do Programa de Pós-
graduação em Estatística e Experimentação
Agropecuária, para a obtenção do título de
“Doutor”.

Orientadora:

Profa. Dra. Thelma Sáfydi

Co-orientador:

Prof. Dr. João Domingos Scalon

LAVRAS

MINAS GERAIS – BRASIL

2008

DENISMAR ALVES NOGUEIRA

ANÁLISE ESPACIAL DA MORTALIDADE INFANTIL NO
MUNICÍPIO DE ALFENAS, MG

Tese apresentada a Universidade Federal de Lavras
como parte das exigências do Programa de Pós-
graduação em Estatística e Experimentação
Agropecuária, para a obtenção do título de
“Doutor”.

APROVADA em 4 de abril de 2008

Prof. Dr. João Domingos Scalon	UFLA
Prof. Dr. Renato Ribeiro de Lima	UFLA
Prof. Dr. Paulo Justiniano Ribeiro Júnior	UFPR
Profa. Dra. Miriam Monteiro de Castro Graciano	UNIFENAS

Profa. Dra. Thelma Sáfyadi
UFLA
(Orientadora)

LAVRAS
MINAS GERAIS – BRASIL

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Nogueira, Denismar Alves

Análise espacial da mortalidade infantil no município de Alfenas, MG
/ Denismar Alves Nogueira. -- Lavras : UFLA, 2008.

101 p. : il.

Tese (Doutorado) – Universidade Federal de Lavras, 2008.

Orientadora: Thelma Sáfadi.

Bibliografia.

1. Inferência Bayesiana. 2. Função *K*. 3. MLEG. I. Universidade Federal
de Lavras. II. Título.

CDD-519.542

*À mulher que eu amo, **Thaísa**. Companheira dedicada, soube suportar, mais uma vez, com abnegação e carinho os caminhos espinhosos por que passei nesta jornada. Obrigado.*

*Ao meu motivo de inspiração e alegria de viver, **Mateus***

*Aos meus pais, **Emar e Denaide**, com a humildade e o jeito simples de ser no qual eu sempre me inspirei.*

A todos os familiares que estiveram sempre presentes.

AGRADECIMENTOS

A Deus, por mais esta etapa vencida. A minha família, que esteve sempre ao meu lado e que acreditou em mim.

À Universidade Federal de Lavras, que sempre me acolheu como um filho e ao Programa de Pós-Graduação, que acreditou no meu trabalho.

A **Thelma Sáfadi**, pela orientação, confiança, dedicação, apoio, ensinamentos, amizade e carinho. Ao colega professor Scalon, pelos ensinamentos e dedicação.

Aos colegas **Eric (Barba)**, **Fabyano (Bia)**, **Janser**, **José Waldemar (Zé)**, **Luiz Beijo**, **Marcelo Cirilo**, **Washington e Aladir**, pela convivência que tivemos, pela amizade, carinho e apoio nas horas difíceis.

À Unifenas, pelo apoio e incentivo. Aos colegas professores e alunos desta instituição, pelo apoio e amizade.

A equipe Master Stat, agradeço a amizade e a compreensão, e desejo grande sucesso.

Aos professores e amigos não citados, mas que estão presentes no coração.

Obrigado à Capes, pela bolsa, permitindo dedicação a este trabalho.

SUMÁRIO

	Página
LISTA DE TABELAS	i
LISTA DE FIGURAS	ii
RESUMO.....	v
ABSTRACT	vi
1 INTRODUÇÃO.....	1
2 REFERENCIAL TEÓRICO.....	4
2.1 Mortalidade infantil.....	4
2.2 Análise espacial.....	9
2.3 Análise exploratória espacial.....	13
2.3.1 Aglomerados espaciais.....	13
2.3.1.1 Função K de Ripley	14
2.4 Modelagem	17
2.4.1 Análise espacial para o risco	17
2.4.2 Modelo linear espacial generalizado	21
2.4.3 Processo espacial.....	21
2.4.4 Teoria bayesiana	23
2.4.4.1 Considerações iniciais.....	23
2.4.4.2 Inferência bayesiana	24
2.4.5 Distribuição preditiva.....	26
2.4.6 Simulação estocástica via cadeias de Markov	26
2.4.6.1 Amostrador de Gibbs (“Gibbs Sampler”).....	27
2.4.6.2 Metropolis-Hastings	29
2.4.6.2.1 Langevin-Hastings.....	30
2.4.6.3 Verificação formal da convergência.....	31

2.4.6.3.1 Critério de Gelman & Rubin (1992).....	32
2.4.6.3.2 Critério multivariado de Brooks & Gelman (1998).....	35
2.4.6.3.3 Critério de Raftery & Lewis (1992).....	36
2.4.6.3.4 Critério de Geweke (1992)	37
2.4.6.3.5 Critério de Heidelberger & Welch (1983).....	39
2.4.7 Inferência bayesiana para modelo linear espacial generalizado.....	40
2.4.7.1 A definição do modelo.....	40
2.4.7.2 Distribuições a priori.....	43
2.4.7.2.1 Priori para o caso em que se considera ϕ conhecido	43
2.4.7.2.2 Priori para o caso em que se considera ϕ desconhecido.....	44
2.4.7.2.3 Distribuição a posteriori	44
2.4.7.3 Distribuição preditiva para o modelo linear espacial generalizado	45
3 METODOLOGIA.....	47
3.1 Área do estudo.....	47
3.2 Bases territoriais.....	50
3.3 Dados demográficos e sócio-econômicos.....	51
3.4 Análise exploratória.....	55
3.5 Modelo linear espacial generalizado	56
3.6 Inferência bayesiana para modelo linear espacial generalizado.....	57
4 RESULTADOS E DISCUSSÕES.....	60
4.1 Análise Exploratória.....	60
4.2 Ajuste dos modelos lineares espaciais generalizados via inferência bayesiana.....	72
4.3 Considerações finais	88
5 CONCLUSÕES.....	90
REFERÊNCIAS BIBLIOGRÁFICA	91

LISTA DE TABELAS

TABELA 1	Covariáveis selecionadas por setor censitário. IBGE censo 2000..	65
TABELA 2	Médias e I.C. HPD 95% dos parâmetros β_1 dos modelos de regressão logística de acordo com a covariável utilizada.....	73
TABELA 3	Resumo das estatísticas descritivas dos parâmetros, para o modelo que considera a covariável número de residentes.....	76
TABELA 4	Resultados dos testes de verificação da convergência para o modelo, considerando o número de mulheres em idade fértil.....	79
TABELA 5	Resumo das estatísticas descritivas dos parâmetros, para o modelo considerando a covariável densidade demográfica.....	82
TABELA 6	Resultados dos testes de verificação da convergência, para o modelo considerando a covariável densidade demográfica.....	86
TABELA 7	Comparação entre os <i>odds ratio</i> para 6 bairros no município de Alfenas, MG, segundo os dois modelos significativos	88

LISTA DE FIGURAS

FIGURA 1	Cartograma dos logradouros e bairros da cidade de Alfenas, MG...	48
FIGURA 2	Cartograma dos setores censitários do município de Alfenas, MG...	49
FIGURA 3	Cartograma dos setores censitários e os limites da área considerada no estudo, com respectivos centróides.....	50
FIGURA 4	Representação pontual dos casos de mortalidade infantil entre 2000 e 2004 no município de Alfenas, MG.....	53
FIGURA 5	Representação pontual dos casos de nascimentos entre 2000 e 2004 no município de Alfenas, MG	54
FIGURA 6	Número de mortes em cada setor censitário, localizado pelo centróide, entre 2000 e 2004, ocorridas no município de Alfenas.....	55
FIGURA 7	Representação gráfica da localização espacial de uma amostra dos dados de morte (+) e de nascimento (•), ocorridos entre 2000 e 2004 na cidade de Alfenas.....	60
FIGURA 8	Representação gráfica para a função K bivariada para os dados de mortalidade e de nascimento, estimada, a cada distância e os respectivos limites de significância.....	61
FIGURA 9	Representação gráfica para a função da diferença $D(h)$ para os dados de mortalidade e de nascimento, estimada a cada distância e os respectivos limites de significância.....	62
FIGURA 10	Representação pontual proporcional a ocorrência de mortes por setor, entre 2000 e 2004 no município de Alfenas, MG.....	63

FIGURA 11	Localização pontual das classes de dados (superior esquerdo), valores dos dados de acordo com as coordenadas (superior direito e inferior esquerdo) e histograma dos coeficientes de mortalidade.....	64
FIGURA 12	Análise exploratória para o número de mulheres em idade fértil, na cidade de Alfenas, de acordo com os setores censitários.....	66
FIGURA 13	Análise exploratória para o número de mulheres em idade de risco gestacional na cidade de Alfenas, de acordo com os setores censitários.....	67
FIGURA 14	Análise exploratória para o número de mulheres em idade fértil analfabetas na cidade de Alfenas, de acordo com os setores censitários.....	68
FIGURA 15	Análise exploratória para o número de residências com mais de seis moradores na cidade de Alfenas, de acordo com os setores censitários.....	69
FIGURA 16	Análise exploratória para a renda mensal da mulher, na cidade de Alfenas, de acordo com os setores censitários.....	70
FIGURA 17	Análise exploratória para a renda mensal do homem, na cidade de Alfenas, de acordo com os setores censitários.....	70
FIGURA 18	Análise exploratória para a densidade demográfica na cidade de Alfenas, de acordo com os setores censitários.....	71
FIGURA 19	Semivariograma para os dados de mortalidade, considerando uma tendência para número de residentes.....	72
FIGURA 20	Representação gráfica dos valores preditos para o risco, dados pelo modelo considerando o número de mulheres em idade fértil, no município de Alfenas, MG	74

FIGURA 21	Representação gráfica dos pontos amostrais e valores preditos, considerando a covariável número de mulheres em idade fértil...	75
FIGURA 22	Distribuições dos parâmetros β_1 e <i>odds ratio</i> , para o modelo, considerando a covariável número de mulheres em idade fértil...	78
FIGURA 23	Representação gráfica dos valores preditos, considerando a covariável densidade demográfica, para o risco de morte no município de Alfenas, MG	81
FIGURA 24	Distribuições dos parâmetros β_1 e <i>odds ratio</i> , para o modelo, considerando a covariável densidade demográfica do setor censitário.....	85

RESUMO

NOGUEIRA, Denismar Alves. **Análise espacial da mortalidade infantil no município de Alfenas, MG.** 2008. 101 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG*.

Os trabalhos epidemiológicos relacionados à saúde coletiva mostram que o processo de saúde-doença ocorre de maneira espacial e, em alguns casos, espacial-temporal. Procurou-se trabalhar com as bases teóricas para análise espacial de dados epidemiológicos, considerando um estudo ecológico de múltiplos grupos utilizando ferramenta analítico, no qual o número de óbitos em menores de um ano, durante os anos de 2000 a 2004, foram tomados como variável dependente e, como variáveis independentes, o número de mulheres em idade fértil, o número de mulheres em idade de risco gestacional, o número de mulheres em idade fértil analfabetas, o número de residências com mais de seis moradores, a renda mensal da mulher, a renda mensal do homem e a densidade demográfica do setor censitário. A análise exploratória de dados espaciais foi feita para detectar aglomerados espaciais, já a predição e a estimação dos parâmetros do modelo foram realizadas por inferência bayesiana. Utilizou-se a metodologia de geoestatística baseada em modelos para ajustar os dados com distribuição binomial, utilizando modelo linear generalizado com função de ligação logística e função exponencial para a covariância residual. Os dados analisados foram obtidos a partir do SIM e do SINASC do município de Alfenas, MG, bem como do censo de 2000 (IBGE). O estudo com a função K bivariada caracterizou um efeito de aproximação entre morte e nascimento e o uso da diferença entre as funções mostrou que ambas provêm de mesmo processo estocástico. O ajuste do modelo logístico utilizando inferência bayesiana possibilitou a realização de predição e a obtenção das distribuições dos parâmetros do modelo. A obtenção da distribuição do parâmetro *odds ratio* possibilitou caracterizar, assim, uma relação de causa entre o número de mulheres em idade fértil e densidade demográfica com a mortalidade infantil.

Palavras Chaves: inferência bayesiana, função K , MLEG.

* Comitê Orientador: Profa. Dra. Thelma Sáfydi – UFLA (orientadora) e Prof. Dr. João Domingos Scalon – UFLA.

ABSTRACT

NOGUEIRA, D. A. **Spatial analysis of infant mortality in the county of Alfenas, MG, Brazil**. 2008. 101 p. Thesis (Doctorate in Statistics and Agricultural Experimentation) – Federal University of Lavras, Lavras, MG*.

Epidemiological studies related to public health show that the process of health-disease occurs in a spatial way and, in some cases, in a spatial-temporal way. This paper is based on theoretical spatial analysis of epidemiological data considering an ecological study of analytical multiple groups. The data is taken from SIM and SINASC of the city of Alfenas, MG, Brazil as well as of the 2000 census provided by the Brazilian Institute of Geography and Statistics (IBGE). The deaths occurred in infants with inferior age to one year from 2000 to 2004 is taken as dependent variable. The covariates are: number of women in fertile age, number of women in age of gestational risk, number of illiterate women in fertile age, number of residences with more than six inhabitants, monthly income of the woman, monthly income of the man and demographic density of the census sector. It is performed exploratory spatial data analysis in order to detect spatial clusters. The parameters estimation of the model and predictions are carried out by using Bayesian inference. A model-based geostatistics is fitted to a binomial data using generalized linear model with a logistic link function and an exponential function for the residual covariance. The bivariate K function shows a significant evidence of spatial interaction between the spatial distributions of deaths and births. The difference between the univariate K functions for deaths and births shows that both spatial distributions come from the same random process. Through the fit of the logistic model by using Bayesian inference is possible to get both the distributions of the parameters of the model and the predictions. The distribution of the odds-ratio parameter allows characterizing a relationship between the dependent variable infant mortality and the covariates number of women in fertile age and demographic density of the census sector.

Keywords: Bayesian inference, K function, geostatistics, GLSM.

*Guidance Committee: Prof. DSc. Thelma Sáfaci – UFLA (Adviser), Prof. DSc. João Domingos Scalón – UFLA

1 INTRODUÇÃO

A análise estatística de dados, levando em consideração a localização espacial do evento, é, obviamente, muito importante para a identificação de áreas nas quais a saúde é precária, necessitando de atenção diferenciada. O uso de análises baseadas na distribuição espacial dos eventos pode trazer contribuição importante à gestão do sistema de saúde e à qualificação na atenção básica à população.

A informação de endereço e a eficiência do Sistema de Informação Geográfico (SIG), para localizar os eventos pontualmente (georreferenciamento), são fundamentais para possibilitar as análises dos padrões de distribuição das ocorrências de interesse. As maiores proporções de uma determinada moléstia, como a mortalidade no período pós-neonatal relacionadas a causas como doenças respiratórias e diarreias, ocorrem em situações de grande carência social e precário acesso a serviços de saúde. Este é um dos fatos de fundamental importância em um estudo epidemiológico, ou seja, trata-se de levar em consideração fatores externos, como os sócio-econômicos que estão ligados diretamente à localização espacial.

O surgimento de trabalhos científicos que incorporam na análise de dados à informação espacial é recente. Segundo Carvalho (1997), em levantamento feito em duas publicações nacionais dedicadas à saúde pública, de 1993 até março de 1995, menos de 20% tratavam de alguma forma a localização espacial do acontecimento,

Estudos de doenças em populações e diagnósticos de saúde de áreas determinadas são, freqüentemente, realizados a partir de dados relativos à situação sócio-econômica, meio-ambiente, perfil de morbi-mortalidade, disponibilidade de equipamentos urbanos e utilização de serviços de saúde (Castellanos, 1990). Um dos problemas usualmente encontrados em estudos utilizando setores é a extensão da base territorial de cada unidade geográfica, onde é freqüente que estejam agregados grupos sociais distintos (favelas e áreas nobres). Assim, o indicador calculado representa uma média que não reflete a verdadeira situação, pois considera todo o setor como sendo homogêneo. Evidentemente, os agravos à saúde e o acesso aos equipamentos urbanos não estão distribuídos homogeneamente na população.

É esperado que o uso de covariáveis que retratam informações sócio-econômicas possibilite introduzir, nos modelos propostos, a influência destes fatores externos, relacionados à mortalidade infantil. O tipo de relação e a existência desta poderão ser quantificados, permitindo melhoria na modelagem da mortalidade infantil.

A interação entre fatores individuais e ambientais exige, cada vez mais, a aplicação de novas técnicas de análise. O uso da modelagem geoestatística baseada em modelos para dados de mortalidade infantil é de cunho inovador, pois, permite a estimação de parâmetros que quantificam a relação de covariáveis e a predição do risco de morte para a região como um todo. Este estudo possibilitará determinar regiões bem distintas, que apresentem pouca acessibilidade aos serviços de saúde municipais, refletindo, assim, a qualidade da assistência à saúde. A definição de regiões de maior probabilidade de morte possibilitará a implantação de políticas de melhoria e prevenção da saúde no município de Alfenas, MG. O uso da metodologia não assume homogeneidade dentro de setores, por tratar a região de forma contínua, o que possibilita melhoria na qualidade das interpretações.

2 REFERENCIAL TEÓRICO

2.1 Mortalidade infantil

O Sistema de Informações sobre Nascidos Vivos (SINASC), ponto de partida da presente pesquisa, foi implantado no Brasil em 1990, com o objetivo de dar a conhecer o perfil epidemiológico dos nascimentos vivos, segundo as variáveis de peso ao nascer, duração da gestação, tipo de parto, idade da mãe e número de partos, dentre outras, constantes na Declaração de Nascido Vivo (DNV). O Sistema de Informação de Mortalidade (SIM), por sua vez, já havia sido implantado no Brasil em 1975, tendo seus dados informatizados e disponíveis desde 1976, quando se tornou elemento essencial para o Sistema de Vigilância Epidemiológica Brasileiro. Suas informações procedem da declaração de óbito (DO), (Rouquayrol & Almeida Filho, 2003).

É fundamental, para a Saúde Pública, que os indicadores calculados a partir do SINASC e do SIM reflitam de fato o perfil epidemiológico da população. Por isso, tem sido grande o interesse em avaliar suas informações, tanto sob o aspecto da cobertura do sistema quanto sob o da confiabilidade de suas informações (Theme Filha et al., 2004).

A mortalidade neonatal é um dos dois componentes da mortalidade infantil (óbitos de crianças menores de um ano, nascidas vivas) e diz respeito aos óbitos ocorridos até o 27º dia de vida. Esse componente da mortalidade infantil subdivide-se em dois outros componentes: (1) a mortalidade neonatal precoce, que abrange óbitos ocorridos nos seis primeiros dias de vida e (2) a mortalidade neonatal tardia, que são os óbitos ocorridos do 7º ao 27º dia de vida. O segundo componente da mortalidade infantil, a mortalidade pós-neonatal, compreende os óbitos ocorridos do 22º dia de nascimento ao 12º mês de vida (Brasil, 2004). Além destes dois componentes da mortalidade infantil, é necessário também o estudo da natimortalidade, ou mortalidade fetal tardia, que inclui os óbitos

ocorridos a partir da 22^a semana de gestação. Uma vez que se considera a natimortalidade como um componente invisível da mortalidade infantil, que possui os mesmos determinantes e soluções que os óbitos neonatais precoces, sugere-se, ainda, o cálculo da mortalidade perinatal, que contempla tanto os óbitos neonatais precoces quanto os natimortos (Leal et al., 2004).

É preciso analisar separadamente cada um desses componentes da mortalidade infantil porque, em cada um deles, se diferencia a rede de causalidade desencadeadora do óbito, bem como a natureza e a forma das intervenções possíveis e necessárias (Leal et al., 2004). O coeficiente de mortalidade infantil é calculado a partir da razão entre o número de óbitos ocorridos no período de um ano e o número de nascidos vivos naquele mesmo ano e lugar, multiplicado por mil (Rouquayrol & Almeida Filho, 2003).

A mortalidade infantil, no Brasil, vem apresentando queda desde a década de 1980, particularmente em função de uma queda do componente pós-neonatal. Esse declínio tem sido atribuído a vários fatores, tais como: intervenções ambientais, avanços da medicina, expansão do acesso a cuidados de saúde, diminuição da taxa de fecundidade, aumento do nível educacional da população, melhoria nutricional e das condições de vida, de modo geral (Oliveira & Mendes, 1995).

No mundo, ocorrem cerca de 7,6 milhões de mortes perinatais por ano, das quais 98% ocorrem nos países em desenvolvimento. Além disso, 57% dessas mortes consistem em óbitos fetais, cuja queda tem sido muito lenta (World Health Organization, 1996). Nesses países, a mortalidade neonatal tem se reduzido mais que a natimortalidade, mas, mesmo assim, a uma velocidade bem menos intensa do que a detectada nos países desenvolvidos, nos quais, por sinal, a mortalidade diminuiu igualmente em todas as faixas gestacionais (Joseph et al., 2000).

infantil, entre 1982 e 1993, não reduziu as diferenças entre os grupos sociais. No estudo de coorte de 1993, as crianças de famílias com renda igual ou inferior a um salário mínimo mensal apresentaram coeficientes de mortalidade seis vezes maiores do que as crianças de famílias com renda superior a 10 salários mínimos mensais (Victora et al., 2000).

Os dois componentes da mortalidade infantil (neonatal e pós-neonatal) têm importância variada, conforme a situação epidemiológica da área em análise. Maiores proporções de mortalidade no período pós-neonatal, relacionadas às causas como doenças respiratórias e diarreias, ocorrem em situações de grande carência social e precário acesso a serviços de saúde (Shimakura et al., 2001).

Andes & Davis (1995) estudaram a relação existente do contexto geográfico na mortalidade infantil no Alasca, sendo os resultados fundamentais para testar novas hipóteses a serem investigadas com o uso da análise de estatística espacial.

Segundo Santos et al. (2001), a qualidade da informação de endereço e a eficiência do sistema de informação geográfico (SIG) para localizar os eventos pontualmente são fundamentais para possibilitar as análises dos padrões de distribuição dessas ocorrências.

A disponibilidade de bancos de dados em saúde com a informação do endereço e a estruturação do SIG em diversos municípios brasileiros têm possibilitado o uso crescente da análise de padrões espaciais de morbimortalidade para diferenciar áreas dentro dos municípios (Lima & Ximenes, 1998; Santos, 1999; Ximenes et al., 1999).

Segundo Mendes (1996) e Teixeira et al. (1998), (*appud* Santos et al. 2001), o conhecimento de regiões de maior incidência de casos, ou de maior risco, independentemente da distribuição da população, é fundamental para a adequação de recursos e o direcionamento de ações de saúde. Além disso, o

ambiente de SIG dá margem à integração de informações oriundas de diversos setores, as quais poderão proporcionar visão mais abrangente da dinâmica dos processos sociais vigentes nas diferentes áreas delimitadas.

Shimakura et al. (2001) realizaram um estudo para o risco de mortalidade infantil na cidade de Porto Alegre, utilizando a metodologia de estatística espacial divulgada por Kelsall & Diggle (1998). Os autores objetivaram descrever a metodologia, por ser inovadora e identificaram a forma como se distribuem os óbitos e seus determinadores de risco no território.

Santos et al. (1999) realizaram um estudo utilizando estatística espacial por área, com o objetivo de detectar aglomerados espaciais de óbitos violentos na cidade de Porto Alegre. Estes autores utilizaram a técnica de suavização por kernel (Bailey & Gatrell, 1995), para a criação da superfície de risco. Os resultados possibilitaram a identificação de áreas, nas quais a concentração de residências foi incomum para as principais causas externas de morte, sendo de potencial interesse e de grande importância para a implantação de políticas de promoção e de prevenção da saúde da população.

Outro estudo, publicado no ano de 2001, foi o de Andrade & Szwarcwald, cujo objetivo foi estabelecer a distribuição espacial da mortalidade neonatal precoce no município do Rio de Janeiro, de 1995 a 1996. A metodologia utilizada incluiu as técnicas apresentadas por Cliff & Ord (1981), apropriadas para dados espaciais de área (*Lattice data*). A análise possibilitou a compreensão da distribuição geográfica da mortalidade, identificando os problemas relacionados à gravidez em adolescentes e à marginalidade social. Neste mesmo trabalho, comentou-se sobre estudos que têm sido realizados considerando estatística espacial, entre eles o de Sastry (1996), que estudou a mortalidade infantil no Brasil, que entre os seus objetivos, buscou associar o contexto ambiental e social em que a criança reside, afetando o risco de morrer. Estes efeitos foram, em geral, medidos pelos atributos da mãe, como o nível de

escolaridade dela. Outro trabalho foi o de Slogger & Joshi (1998), no qual se estudou a associação entre o grau de privação social e alguns agravos da saúde, com significância para associação com gravidez na adolescência.

2.2 Análise espacial

A análise estatística de dados, levando-se em consideração a posição espacial, tem ganhado cada vez mais adeptos, devido aos avanços na integração de ferramentas do sistema de informação geográfica e da análise espacial, em particular da análise estatística espacial. Paralelamente, os modelos inferenciais bayesianos ocupam espaço cada vez maior na análise de dados espaciais. Reis (2005) destaca que a ausência de softwares que fazem análises que integram o SIG e a inferência bayesiana. Outra maneira de definir a análise espacial, com a informação da localização dos dados, seria como uma técnica que busca descrever os padrões existentes nos dados espaciais e estabelecer, de forma quantitativa ou não, os relacionamentos das várias variáveis georreferenciadas. A ênfase, segundo Câmara & Monteiro (2004), deste tipo de análise, seria a de mensurar propriedades e relacionamentos, levando em conta a localização espacial do fenômeno em estudo de forma explícita, incorporando o espaço à análise.

Assumir que os dados são não correlacionados espacialmente é uma pressuposição forte, em alguns casos. Considerar, em uma análise, a estrutura espacial dos dados, de maneira correta, permite um estreitamento dos intervalos de confiança dos estimadores e da superestimação dos níveis de significância dos preditores (Gemperli et al., 2003).

A epidemiologia é definida como estudo dos fatores que determinam a ocorrência e a distribuição de doenças na população (Jekel et al., 2005). Assim, apresenta recursos metodológicos que buscam explicitar a desigualdade destas

distribuições da saúde nas sociedades, assim como subsidiar o planejamento das ações do setor saúde.

A utilização de mapas e a preocupação com a distribuição geográfica de dados epidemiológicos, como a de diversas doenças, são antigas. O exemplo que é utilizado por pesquisadores é o trabalho pioneiro de Jonh Snow, em 1854, o qual considerou a posição espacial da ocorrência de cólera para tentar explicar o fenômeno de epidemia ocorrida na cidade de Londres. Esse autor utilizou técnicas de mapeamento para localizar os casos dessa doença e os pontos de coleta de água, mostrando o papel das fontes na ocorrência da doença (Scholten & Lepper, 1991). A partir deste trabalho, vários outros foram escritos utilizando a distribuição espacial, descrevendo variações geográficas na distribuição de dados epidemiológicos.

Em estudos epidemiológicos, os modelos de localização são ferramentas de apoio à decisão para a distribuição espacial de dados de saúde. O crescimento do uso do SIG na área de saúde tem proporcionado melhores resultados quanto à indicação das áreas de risco ou de maior urgência de intervenção da saúde pública.

A análise estatística espacial, propriamente dita, está relacionada, de maneira sucinta, com a caracterização dos três tipos de dados encontrados: eventos ou padrão pontual, superfícies contínuas e áreas com contagens ou taxas (Bailey & Gatrell, 1995). Para a análise de padrão pontual, tem-se interesse apenas na localização espacial do evento em estudo. Para a análise de superfícies, conhecida como geoestatística, tem-se o objetivo de reconstruir uma superfície da área de interesse em estudo. Já a análise de área tem o objetivo de estudar o comportamento dos dados oriundos de levantamentos populacionais, tais como censos, estatísticas de saúde, etc. Neste caso, as áreas (setores censitários, bairros, municípios) são polígonos fechados, assumidos homogêneos (Câmara & Monteiro, 2004). O setor censitário é definido em função das

necessidades operacionais do censo, uma vez que corresponde à área que pode ser percorrida por um recenseador durante o período de coleta de dados (Fundação Instituto Brasileiro de Geografia e Estatística, FIBGE, 1993).

Outro fator que está sendo incorporado aos estudos é o uso do tempo, juntamente com o espaço, sendo, então, considerados estudos espaços-temporais (Kyriakidis & Journel, 1999). O interesse destes estudos está, principalmente, na interação entre tempo e espaço. Muitos modelos têm sido propostos para análise deste tipo de dados, sendo a abordagem bayesiana a que vem obtendo aumento substancial em popularidade, em relação à análise desses modelos. Essa popularidade se deve, principalmente, aos avanços computacionais e à adequada metodologia para inferência preditiva (Paez, 2004). Segundo Paez (2004), a interpolação e a previsão são técnicas “naturais”, sob o ponto de vista bayesiano, pois qualquer quantidade desconhecida pode ser caracterizada pela sua distribuição a posteriori.

A agregação espacial de doença, no espaço e no tempo, pode ser atribuída a fatores demográficos, genéticos, ambientais ou sócio-culturais superpostos geograficamente ao padrão de ocorrências observadas. A agregação temporal pode ser feita considerando-se certos padrões sazonais, entre outros, que podem contribuir para o esclarecimento dos mecanismos responsáveis pela geração de casos da doença (Werneck & Struchiner, 1997). De acordo com estes mesmos autores, a agregação espaço-temporal de doenças tem sido estudada principalmente em situações de doenças raras, ou em estudos como o de avaliação do impacto de programa de prevenção e para sugerir padrões de disseminação de doenças transmissíveis.

Segundo Knox (1988), a concentração de eventos pode ocorrer no espaço, no tempo ou em ambos. Um *cluster* ou um agrupamento de eventos pode ser definido como foco particular de alta incidência ou como grupo delimitado de ocorrências relacionadas entre si, mediante algum mecanismo

social ou biológico. Segundo alguns autores, como Paez (2004) e Silva & Dean (2006), existem trabalhos recentes utilizando a inferência bayesiana de modelos espaço-temporais. Entre estes se destacam: Le & Zidek (1992); Handcock & Wallis (1994); Wikle et al. (1998); Brown et al. (2001); Tonellato (2001); Shaddick & Wakefield (2002); Paez & Gamerman (2003); Barnerjee e al. (2004); Waller & Gotway (2004) e Silva et al. (2005).

Na epidemiologia, outro foco que se refere aos dados com o uso da referência espacial é a presença dos casos que refletem a distribuição espacial de risco da população de interesse. Os padrões atribuídos somente à população de casos não são de interesse, sendo de suma importância comparar os casos com uma amostra dos controles da população de interesse de risco. Estes estudos são os chamados casos-controle. O desenho caso-controle simples deve ser completamente aleatório, nos quais os controles são amostras aleatórias independentes da população de interesse. Delineamentos mais sofisticados de caso-controle envolvem estratificação ou marcação (Diggle, 2003).

Shimakura et al. (2001) estimaram e mapearam o perfil de risco de morte infantil em um desenho caso-controle na cidade de Porto Alegre, RS. Os autores utilizam a abordagem semi-paramétrica por meio de modelos aditivos generalizados (GAM). Outro trabalho interessante foi realizado por Seaman & Richardson, em 2001, usando inferência bayesiana em caso-controle, mas sem o uso de referência espacial. Os autores relatam um esforço computacional para o uso de modelos relativamente simples, mas de muitas aplicabilidades, com o uso de modelagem mais elaborada. Sugerem também que os modelos devam incluir efeitos aleatórios e prioris informativas para as razões de chances (*odds ratio*).

Kelsall & Diggle (1995) consideram, no seu estudo, um problema de estimação da variação espacial para dados de caso-controle, propondo, entre outros, um teste Monte Carlo para um risco constante sobre uma dada região.

2.3 Análise exploratória espacial

As técnicas de análise exploratória aplicadas a dados espaciais são essenciais no desenvolvimento de etapas da modelagem estatística espacial. Em geral, são sensíveis ao tipo de distribuição e à presença de valores extremos. Os dados explorados neste trabalho caracterizam um estudo ecológico de múltiplos grupos, no qual se utilizou os óbitos ocorridos em menores de um ano ocorridos durante os anos de 2000 a 2004. Os dados são valores pontuais no espaço, dessa forma, carregam a informação apenas da localização espacial. A análise exploratória foi utilizada para estudar o comportamento destes dados e, dessa forma, verificar se estes ocorrem de maneira aleatória no espaço.

Em uma segunda etapa, foi realizada uma transformação nos dados, agrupando-os de maneira a permitir o uso da geoestatística, ou seja, valores ocorridos em um mesmo setor censitário foram agrupados, possibilitando, assim, ter um conjunto com p casos de morte. O número de nascidos vivos foi retirado do IBGE (censo de 2000) da variável número de residentes de zero ano de vida. Dessa forma, referenciamento espacial foi feito pelo centróide do respectivo setor censitário, caracterizando, assim, uma distribuição binomial com n e p conhecidos. Com este novo conjunto de dados foi possível utilizar a modelagem que assume ocorrência contínua no espaço, os modelos lineares generalizados espaciais.

2.3.1 Aglomerados espaciais

A detecção de aglomerado espacial pode ser feita a partir da análise da distribuição de pontos (*point pattern*), em que se avalia se a distância entre os pontos de ocorrência de eventos é ou não aleatória. O aspecto central da análise de pontos, tal como desenvolvido por Knox & Bartlett (1964), é a detecção de um número acima do esperado de pares de casos excessivamente próximos ou de distâncias (normalmente a euclidiana) entre pares de casos. A distribuição base

CSR). Dessa forma, a ferramenta explora a comparação entre uma estimativa empírica ($\hat{K}(h)$) com o resultado de um processo padrão de pontos aleatórios, sendo $K(h) = \pi h^2$. Portanto, se existe aglomeração, então, se espera um excesso

$$\hat{K}_{12}(h) = \frac{A}{n_1 n_2} \sum_i^{n_1} \sum_{j, i \neq j}^{n_2} \frac{I_h(d_{ij})}{w_{ij}^*},$$

sendo: $w_{ij}^* = (n_1 w_{ij} + n_2 w_{ji}) / (n_1 + n_2)$.

Para testar a significância para a presença de atração ou de repulsão espacial, é realizado um teste utilizando técnicas de simulação de Monte Carlo. Assumindo um CSR, podem-se rodar m independentes simulações de n eventos da região em estudo. Para cada padrão pontual simulado, é estimado $K_{12}(h)$ e utilizam-se o máximo e o mínimo desta função para definir o limite superior e o limite inferior. Fazendo-se a representação gráfica dos valores estimados da função $K_{12}(h)$, limite inferior e limite superior, pelas distâncias é possível concluir que se, por acaso, a função $K_{12}(h)$ estiver acima do limite superior, caracteriza-se um efeito de atração. Se, por acaso, este estiver abaixo do limite inferior, caracteriza-se, portanto, repulsão entre as variáveis em uma distância específica.

No contexto de um estudo com duas variáveis, os nascimentos formam uma amostra espacial aleatória da mesma população das mortes. Portanto, sob a hipótese inicial, a função $D(h)$ é igual a zero, tal que a função $D(h)$ é dada por:

$$D(h) = K_{11}(h) - K_{22}(h),$$

em que $K_{11}(h)$ é a função univariada para a variável 1 e $K_{22}(h)$ é a função univariada para a variável 2.

Segundo Diggle (2003), $K_{22}(h)$ mede o grau de agregação espacial da população de risco e $K_{11}(h)$ mede o efeito acumulativo desta mesma agregação espacial. $D(h)$ mensura a diferença da aglomeração espacial entre mortes e

nascimentos do mesmo modo que $K(h) - \pi h^2$ mensura a agregação para um caso univariado.

Para testar a significância desta terceira situação, é também feito um teste utilizando-se técnicas de simulação de Monte Carlo. Como nos casos anteriores, para cada padrão pontual simulado é estimado $D(h)$. Fazendo a representação gráfica dos valores estimados da função $D(h)$, limite inferior e limite superior, pelas distâncias é possível concluir que se, por acaso, a estimação da função $D(h)$ estiver fora dos limites superior ou inferior, caracteriza-se que as variáveis vieram de processos diferentes.

2.4 Modelagem

2.4.1 Análise espacial para o risco

O interesse neste tipo de análise é descrever a distribuição espacial de doenças em uma população que reside em uma dada região A conhecida. A amostra que será feita da região é um conjunto de localizações $s_i \in A$; $i = (1, \dots, n)$. O modelo teórico para esta situação é assumir que se tem um par de processos pontuais independentes de Poisson, com intensidades dadas por $\lambda_1(s)$ e $\lambda_2(s)$. O objetivo é a obtenção da variação do risco da doença ou a função da Razão de Chances (*odds ratio*) espacial, $\rho(s) = \lambda_1(s) / \lambda_2(s)$. Em uma situação em que não existe variação espacial e a doença ocorre de maneira aleatória, não existirá influência da localização, portanto, o risco ($\rho(s)$) será constante em toda área A .

Para uma modelagem, é necessário conhecer os fatores de risco e identificar a variação espacial. O risco, $\rho(s)$, não assume valores negativos e, para isso, utiliza-se um modelo log-linear (Elliott et al., 2001):

$$\rho(s) = \exp \left\{ \alpha + \sum_{j=1}^p \beta_j x_j + U \right\},$$

em que α é o intercepto; x_j é a variável independente j representante dos p fatores de risco de variabilidade espacial conhecida, ou covariáveis; β_j o parâmetro do efeito da j -ésima variável, sendo p parâmetros dos efeitos sobre o logaritmo do risco e U a variação espacial não explicável ou o efeito de alguma covariável que carrega a informação espacial. Uma alternativa para explicar esta modelagem é considerar que se tem um único processo de Poisson com intensidade dada por $\lambda(s) = \lambda_1(s) + \lambda_2(s)$. Assumindo que estes dados estão condicionados ao conjunto completo de localizações espaciais s_i , então, y_i segue uma distribuição binomial, independente e identicamente distribuída, sendo a esperança dada por:

$$\mu_i = E[y_i | s_i] = P(y_i = 1 | s_i) = \frac{\lambda_1(s)}{\lambda_1(s) + \lambda_2(s)} = \frac{\rho(s_i)}{1 + \rho(s_i)}. \quad (2.1)$$

Fazendo $h[\mu] = \eta$, por alguma função de ligação $h[.]$, por exemplo, a logística, ter-se-á um modelo de regressão logística (Diggle & Rowlingson, 1994; Gemperli et al., 2003). Portanto, o modelo pode ser escrito da seguinte forma:

$$\text{Log} \left\{ \frac{\rho(s, x)}{1 + \rho(s, x)} \right\} = \alpha + \sum_{j=1}^p \beta_j x_j + U. \quad (2.2)$$

Segundo Diggle et al. (1998), deve-se inserir no modelo uma variação espacial não observada, U , que caracteriza-se por ser um processo espacial

gaussiano, isotrópico, estacionário latente. Sobre a região A , $U = (U_1, U_2, \dots, U_n) \sim N(0, \Sigma)$, sendo Σ_{ij} uma função paramétrica da distância d_{ij} entre os locais s_i e s_j .

A parametrização para covariância, Σ_{ij} , pode ser de várias maneiras. Ecker & Gelfand (1997) propõem variadas formas para representar a função de correlação espacial, entre elas gaussiana, cauchy e esférica. Uma representação utilizada é $\Sigma_{ij} = \sigma^2 \rho(d_{ij}; \phi)$, sendo σ^2 a variância do processo espacial e $\rho(d_{ij}; \phi)$ a função de correlação com um parâmetro escalar ϕ . Este parâmetro controla a taxa de correlação que decai com a distância, $d_{ij} = s_i - s_j$, que é a distância euclidiana entre as localizações s_i e s_j . Existem variadas maneiras de definir a forma da função de correlação. Uma família é dada por $\rho(d_{ij}; \phi) = \exp\left\{-d_{ij}/\phi\right\}$. Utilizando na equação $\kappa = 1$, tem-se uma função exponencial simples e se $\kappa = 2$, tem-se uma normal. Outros exemplos de funções com dois parâmetros é a classe de funções Matérn (Matérn, 1986). Em particular, se o risco é assumido constante na região, então, $U = 0$ e o modelo (2.2) é reduzido a um modelo de regressão logística usual (Hosmer & Lemeshow, 1989). O modelo (2.2) é um modelo de regressão logística estendido, com um componente aditivo U , que varia suavemente no espaço.

Para a situação em que U é uma função determinística de forma não especificada, Bithell (1990) propôs o uso de métodos padrão de estimação não paramétrica, sendo $\rho(s)$ estimado como a razão de duas estimativas de Kernel (Silverman, 1986). Kelsall & Diggle (1995) propõem algumas melhorias na metodologia, dando uma expressão teórica para o quadrado médio do erro de $\hat{\rho}(s)$.

Segundo Elliott et al. (2001), o limitante desta metodologia está no problema de se utilizar variáveis explicativas adicionais (x), sem informações

importantes. Uma solução frequentista é o uso de modelos aditivos generalizados (GAM) (Hastie & Tibshirani, 1990), sendo a estimação possível para α e β .

O uso do modelo de regressão não paramétrico está restrito ao parâmetro de largura de banda (*bandwidth*) da função de kernel. A escolha deste parâmetro pode ser feita de maneira intuitiva, pois varia conforme a distribuição espacial dos dados e do objetivo da análise. Alguns autores, como Fan et al. (1995) Kelsall (1996) e Kelsall & Diggle (1998), propõem métodos automáticos para a seleção deste parâmetro, chamados de métodos de validação cruzada.

Kelsall & Diggle (1998) utilizaram um modelo como dado na equação (2.2). No estudo destes autores, a transformação dos resíduos, $\exp\{U\}$, representa a *odds ratio* da doença após o ajuste das covariáveis. Os autores sugerem o uso de logaritmo na base 2, permitindo interpretar que o aumento de uma unidade na superfície implica em risco dobrado.

No artigo de Seaman & Richardson (2001), é proposta uma base para a modelagem mais complexa, utilizando a inferência bayesiana em estudos de caso-controle, com o uso de dois exemplos práticos.

Bailey (2001) publicou um artigo divulgando os métodos utilizados, até o momento, em estudos espaciais aplicados em saúde, mapeamento de doenças, estudos ecológicos, estudos de *cluster* de doenças e monitoramento, e acesso ambiental. Na área de mapeamento de risco de doenças, o modelo de regressão logística tem seu espaço garantido, apesar de os estudos para dados de padrões pontuais serem os de menor expressão numérica. Na finalização do trabalho, o autor deixa claro que há a necessidade de comentar que não apenas metodologias avançadas de estatística devem ser abordadas, mas que se deve preocupar com a qualidade da informação utilizada e com acesso a dados confiáveis de grau de detalhamento suficiente. Alguns autores (Santos et al., 2001; Shimakura et al., 2001) comentam a má qualidade da informação

disponível nos sistemas de informação brasileiro, pois, em alguns casos, podem chegar até a 50% de não aproveitamento da informação. Este problema ocorre por variados motivos, entre eles, a ausência de covariáveis ou a impossibilidade de georreferenciamento.

Kazembe et al. (2006) apresentam um estudo sobre o risco da malária na República do Maláui, com a utilização da metodologia proposta por Diggle et al. (1998), Diggle et al. (2003) e Diggle & Ribeiro Jr. (2007), *model-based geostatistical*, sendo realizada análise e predição do risco, com estatística espacial por área via inferência bayesiana.

2.4.2 Modelo linear espacial generalizado

Considerando o método clássico de modelagem linear generalizada, assume-se que as observações são independentes (y_1, \dots, y_n) e as esperanças $\mu_i = E[y_i]$ são especificadas por um preditor linear $h(\mu_i) = \sum_{j=1}^p \beta_j x_{ij}$, sendo $h(\cdot)$ a função de ligação (McCullagh & Nelder, 1989).

Quando se considera que as observações são independentes, mas, condicionais a valores de um conjunto de variáveis aleatórias U_1, \dots, U_n , o modelo passa a se chamar modelo linear misto generalizado (GLMM) (Breslow & Clayton, 1993).

O modelo linear espacial generalizado (GLSM) é um GLMM, pois assume que U_1, \dots, U_n são derivadas de um processo espacial $U(\cdot)$.

2.4.3 Processo espacial

De maneira geral, em modelos geoestatísticos, a seguinte notação é adotada. Primeiramente, os dados para a análise são definidos por (y_i, s_i) , para $i = 1, \dots, n$, em que s_1, s_2, \dots, s_n são localizações dentro da região em estudo A . Os

valores observados são dados por y_1, y_2, \dots, y_n e sua mensuração está associada com estas localizações, ou seja, $y_i = y(s_i)$, sendo $y(\cdot) = \{y(s) : s \in A\}$ chamado processo de mensuração. É postulada a existência de um processo estocástico não observado $U(\cdot) = \{U(s) : s \in A\}$, chamado de processo sinal, como sendo um processo estocástico gaussiano com $E[U(s_i)] = \sum_{j=1}^p \beta_j x_j(s_i)$, $VAR[U(s_i)] = \sigma^2$ e $\rho(d) = Corr[U(s_i), U(s_j)]$, sendo $d = \|s_i - s_j\|$. Portanto, é assumido que as distâncias são condicionais ao processo $U(\cdot)$, com esperança condicional μ_i e $h(\mu_i) = U(s_i)$, $i = 1, \dots, n$, para uma função de ligação conhecida. Neste modelo, o processo sinal é $\{h^{-1}(U(s)) : s \in A\}$.

Para um modelo geoestatístico, é assumida uma distribuição conjunta de um processo de mensuração e um processo sinal, ou seja:

$$P(U, Y | \theta) = P(Y | U, \theta)P(U, \theta), \quad (2.3)$$

em que $P(\cdot)$ denota a distribuição de probabilidades; o termo U representa o vetor de valores de $U(s_i)$, Y corresponde ao vetor de valores mensurados e θ é o representante de todos os parâmetros do modelo (Diggle & Ribeiro JR., 2007).

O GLSM abrange um caso especial do modelo linear gaussiano, no qual as variáveis respostas não seguem a distribuição padrão, normal, mas, certamente, descrevem o mecanismo amostral envolvido. Para dados de contagem, é comum adotar um modelo Poisson-log-linear e, para dados binomiais, é comum adotar o modelo logístico (função de ligação).

2.4.4 Teoria bayesiana

2.4.4.1 Considerações iniciais

O problema fundamental da estatística é fazer inferência. Alguns dados são observados e deseja-se fazer inferência sobre uma ou mais características desconhecidas do sistema que teria dado origem a estes dados.

A inferência estatística trabalha na presença de observações y , cujos valores inicialmente incertos são descritos por meio de uma distribuição de probabilidade $f(y|\theta)$. O parâmetro θ é o que interessa ao pesquisador e tem um sentido exato dentro da inferência bayesiana. O pesquisador deve ter alguma informação prévia sobre θ e esta informação deve ser incorporada à análise.

Para Gelman et al. (1997), a inferência bayesiana é o processo de ajustar um modelo de probabilidade para um grupo de dados e resumir o resultado por uma distribuição de probabilidade nos parâmetros do modelo e em quantidades não observáveis como predição para novas observações.

A inferência bayesiana é baseada no conceito de probabilidade subjetiva, que mede o grau de desconhecimento que alguém tem da ocorrência de um determinado evento do espaço amostral. Se existe uma incerteza, a descrição desta incerteza deve ser feita utilizando-se a probabilidade subjetiva. Então, a análise bayesiana descreve toda a quantidade desconhecida por meio da probabilidade. Ao descrever uma incerteza usando a interpretação subjetivista de probabilidade, deve-se levar em conta o axioma de coerência estabelecido por De Finetti (1937)* *appud* Henry & Kyburg Jr. (1964).

*DE FINETTI, B. La prévision: ses lois logiques, ses sources subjectives. Annales de l'Institut Henri Poincaré, Paris, v. 7, p. 1-68, 1937.

2.4.4.2 Inferência bayesiana

O pesquisador, ao incorporar à análise uma opinião sobre θ , por meio de uma densidade de probabilidade $P(\theta)$, está determinando a densidade a priori. A densidade a priori possui este nome por ser a distribuição de probabilidade de θ antes que se observem os dados. O parâmetro θ pode ser um escalar ou um vetor de parâmetros.

A densidade conjunta de um grupo de observações y_1, \dots, y_n , examinada como uma função do parâmetro, é denominada função de verossimilhança e é representada por $L(y_1, \dots, y_n | \theta)$, sendo n o número de observações.

Se $P(\theta)$ é a densidade a priori para θ , então, a densidade a posteriori de θ é dada pelo Teorema de Bayes, ou seja

$$P(\theta | Y) = \frac{L(Y | \theta)P(\theta)}{\int L(Y | \theta)P(\theta)d\theta},$$

sendo $Y = (y_1, y_2, \dots, y_n)$. Como o denominador não depende de θ , pode-se escrever:

$$P(\theta | Y) \propto L(Y | \theta)P(\theta),$$

em que \propto representa proporcionalidade.

Pode-se pensar no Teorema de Bayes como um mecanismo de atualização da opinião do estatístico sobre θ . O Teorema de Bayes é um elemento essencial para a análise bayesiana, pois, toda inferência é feita a partir da posteriori.

Por muitos anos, o uso de priori subjetiva gerou uma grande controvérsia para a inferência estatística baseada no Teorema de Bayes, pois,

dependendo da informação desta priori, a posteriori será muito influenciada, principalmente em um estudo com poucos dados. A escolha da distribuição a priori é um problema pertinente na abordagem bayesiana. Existem dois tipos de priori: a priori informativa e a priori não-informativa. Quando o pesquisador tem alguma informação prévia sobre o que está estudando, ele pode usar uma priori informativa. De todos os aspectos da inferência bayesiana, a priori informativa é a mais difícil e a que traz maior controvérsia. Pode acontecer que, em determinado estudo, o pesquisador tenha pouca ou nenhuma informação para incorporar à priori. Quando isso acontece, a distribuição considerada é a não-informativa. Existe igualmente controvérsia quanto à priori não-informativa, pois, muitas vezes, esta priori faz com que se tenha uma distribuição a posteriori imprópria. A priori não-informativa mais usada é a priori de Jeffreys, sendo invariante sobre reparametrização (Jeffreys, 1939).

A representação probabilística de todo o conhecimento incerto é a essência da inferência bayesiana, seja tal conhecimento relacionado ao futuro, quantidades observáveis ou parâmetros desconhecidos. Os conceitos de priori e posteriori são sempre relativos à observação considerada no momento.

Broemeling (1985) afirma que a base da inferência bayesiana é a distribuição a posteriori de θ , pois, qualquer conclusão é feita a partir desta distribuição.

A diferença formal entre a inferência bayesiana e a freqüentista é que, para a inferência bayesiana, a incerteza sobre o parâmetro θ é tratada utilizando-se uma distribuição de probabilidade, como uma variável aleatória. E, para a inferência freqüentista, os parâmetros são valores fixos, não sendo possível atribuir a eles uma distribuição de probabilidade.

O'Hagan (1994) comenta o porquê de usar a inferência bayesiana em vez da inferência clássica. Ele afirma que, em termos gerais, alguns dos argumentos a favor do enfoque bayesiano são os de que a análise bayesiana é

fundamentalmente completa, muito flexível, produz conclusões claras e diretas e faz uso de toda informação disponível. Em contraste, o enfoque clássico ignora a informação a priori.

2.4.5 Distribuição preditiva

Um outro elemento na inferência bayesiana é a *distribuição preditiva*, sendo usada quando se quer fazer uma inferência sobre um valor (\tilde{y}) não observado, utilizando-se as observações $Y = \{y_1, \dots, y_n\}$, o que seria a previsão de valores desconhecidos, porém, observáveis. Esse valor a ser predito pode ser considerado no modelo como um novo parâmetro a ser também estimado. A distribuição preditiva é baseada na distribuição a posteriori de \tilde{y} , $P(\tilde{y} | y_1, \dots, y_n)$, e pode ser obtida da posteriori de \tilde{y} e θ pela integração:

$$P(\tilde{y} | Y) = \int P(\tilde{y}, \theta | Y) d\theta = \int P(\tilde{y} | \theta, Y) P(\theta | Y) d\theta,$$

em que θ é o vetor de parâmetros envolvidos.

Na estatística espacial, o que se deseja realizar, com o uso da distribuição preditiva, é a interpolação espacial da variável resposta Y . As observações são realizadas em N locais (s_1, \dots, s_n) distintos no espaço, em que a correlação entre duas variáveis vai depender dos seus locais de observação s_i e s_j . O objetivo é prever o valor da variável resposta para locais $s_{n+1}, s_{n+2}, \dots, s_{n+q}$, em que o Y não foi observado.

2.4.6 Simulação estocástica via Cadeias de Markov

O avanço da informática propiciou métodos cada vez melhores para se obter soluções aproximadas dos processos analíticos.

O termo simulação refere-se ao tratamento de problemas reais a partir de reproduções em ambientes controlados pelo pesquisador, sendo, muitas vezes, esses ambientes os equipamentos computacionais. Alguns problemas apresentam componentes aleatórios, os quais não podem ser descritos de forma exata e são baseados em informações probabilísticas. O processo de simulação é estocástico, ou seja, baseado em distribuições de probabilidades.

Dos métodos de simulação que utilizam cadeias de Markov podem ser citados: o amostrador de Gibbs e o Metropolis-Hastings (Hastings, 1970). Recentemente, alguns autores (Christensen & Waagepetersen, 2002; Waagepetersen et al., 2008) têm trabalhado com uma variação do algoritmo do Metropolis-Hastings. Este algoritmo é conhecido como Langevin-Hastings e proporciona uma convergência mais rápida na maioria dos casos, segundo Christensen et al. (2001).

2.4.6.1 Amostrador de Gibbs (“Gibbs Sampler”)

O amostrador de Gibbs é, essencialmente, um esquema iterativo de amostragem de cadeias de Markov, cujo núcleo de transição é formado pelas distribuições condicionais completas. É uma técnica para gerar variáveis aleatórias de uma distribuição (marginal) sem que se conheça a sua densidade. A idéia do método de Monte Carlo via Cadeias de Markov (MCMC) é simular a um passeio aleatório no espaço do parâmetro θ , o qual converge para uma distribuição estacionária que, em estatística bayesiana, é a distribuição a *posteriori* $P(\theta|Y)$.

O amostrador de Gibbs tem sido extremamente útil na resolução de problemas multidimensionais e é definido em termos de subvetores de θ . Seja o vetor de parâmetros θ dividido em p subvetores $(\theta_1, \theta_2, \dots, \theta_p)$ e suponha que as distribuições condicionais de cada subvetor (que pode ser um parâmetro ou um vetor destes) θ_i , dados todos os outros, sejam conhecidas. Essas distribuições são

denotadas por $f(\theta_1|\theta_2, \dots, \theta_p, Y)$, $f(\theta_2|\theta_1, \theta_3, \dots, \theta_p, Y), \dots, f(\theta_n|\theta_1, \theta_2, \dots, \theta_{p-1}, Y)$, sendo Y o vetor de n observações. Estas distribuições são denominadas distribuições *condicionais completas*.

Para o amostrador de Gibbs é necessário que estas condicionais completas apresentem uma forma fechada, ou seja, conhecida de alguma distribuição. Caso contrário, há a necessidade de se utilizar outro amostrador, o Metropolis-Hastings.

Segundo Gamerman & Lopes (2006), o amostrador de Gibbs pode ser descrito da seguinte forma:

Passo 1 – assumir os valores iniciais $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$ para os parâmetros;

Passo 2 - gerando $m+n$ grupos de números aleatórios, retira-se iterativamente

$$\theta_1^{(1)} \text{ de } f_1(\theta_1|\theta_2^{(0)}, \theta_3^{(0)}, \dots, \theta_p^{(0)}, Y),$$

$$\theta_2^{(1)} \text{ de } f_2(\theta_2|\theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_p^{(0)}, Y),$$

↓

$$\theta_p^{(1)} \text{ de } f_p(\theta_p|\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{p-1}^{(1)}, Y),$$

obtendo-se, na primeira iteração, $\theta^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_p^{(1)})'$.

Passo 3 - descartando as primeiras m realizações no passo 2, as n restantes realizações são utilizadas para formar uma amostra aleatória $\{(\theta_1^{(i)}, \theta_2^{(i)}, \dots, \theta_p^{(i)})_{i=m+1}^{m+n}\}$ e estima-se a posteriori usando esta amostra aleatória.

À medida que o número de iterações aumenta $n \rightarrow \infty$, a seqüência se aproxima de sua condição de equilíbrio, ou seja, ela atinge a característica de estacionariedade com um comportamento assintótico. Assim, assume-se que a convergência é atingida em uma iteração cuja distribuição esteja arbitrariamente próxima da distribuição de equilíbrio, ou seja, a marginal desejada, e não no

sentido formal e inatingível do número de iterações tendendo para o infinito. Gamerman & Lopes (2006) comentam que, apesar de os resultados teóricos garantirem a convergência do amostrador de Gibbs, sua utilização, na prática, pode ser bastante complicada quando se têm modelos muito complexos. Essa complexidade faz com que a convergência do amostrador de Gibbs seja de difícil caracterização.

2.4.6.2 Metropolis-Hastings

Este procedimento é usado em situações em que a distribuição a posteriori condicional completa não apresenta uma forma conhecida, fechada, não identificável.

Seja $P(\theta | Y)$ uma distribuição não identificada. O algoritmo de Metropolis-Hastings gera seqüências de pontos aleatórios $(\theta_1, \theta_2, \theta_3, \dots)$ cuja distribuição converge para $P(\theta | Y)$.

O algoritmo é especificado por uma densidade $q(\theta, \beta)$ denominada de núcleo de transição ou densidade auxiliar. Esta densidade é selecionada de uma família de distribuições que requer a especificação de alguma característica do parâmetro, como escala ou posição. O núcleo de transição representa a probabilidade de movimento de θ para β . Geralmente o processo se move com mais freqüência de θ , para β do que de β para θ , contradizendo a condição de reversibilidade $P(\theta | Y) q(\theta, \beta) = P(\theta | Y) q(\beta, \theta)$. Este fato pode ser corrigido, se diminuirmos o número de movimentos de θ para β , e para isto introduzimos uma probabilidade de movimento p .

Os passos do algoritmo podem ser definidos da seguinte forma:

Passo 1 – inicialize com um valor arbitrário para θ ;

Passo 2 – gera-se um valor de β de $q(\theta, \cdot)$ e u de uma uniforme $(0,1)$;

Passo 3 – seja $p = \min \left\{ 1, \frac{P(\beta | Y) q(\theta^{(j)}, \beta)}{P(\theta^j | Y) q(\beta, \theta^{(j)})} \right\}$;

Passo 4 – se $u \leq p$, faça $\theta^{(j+1)} = \theta^{(j)}$;

Passo 5 – repita os passos 2 e 3 até a distribuição estacionária seja obtida.

O núcleo de transição q apenas define uma proposta de movimento, a qual deve (ou não) ser confirmada por p . Um problema crítico na eficiência do algoritmo é a escolha do fator escala ou dispersão da densidade q . O fator escala da densidade q afeta o comportamento da cadeia em pelo menos duas dimensões, sendo uma a taxa de aceitação e a outra é a região do espaço amostral que é coberta pela cadeia.

Um densidade auxiliar que pode ser escolhida é a normal multivariada (NM) com a média dada pelo valor correto de θ e com matriz de covariância hI , sendo h uma variância dada pelo usuário, $h > 0$, e I uma matriz identidade, isto é, $q(\theta, \beta)$ tem densidade $NM(\theta, hI)$. Em situações de alta dimensão o algoritmo pode convergir muito lentamente e produzir amostras altamente autocorrelacionadas (Christensen et al., 2001). As características de convergência do processo também são válidas para o algoritmo de Metropolis-Hastings, sendo na maioria dos casos mais lenta.

2.4.6.2.1 Langevin-Hastings

A proposta para algoritmo Langevin-Hastings é acelerar a velocidade de convergência do algoritmo Metropolis-Hastings. Este novo algoritmo utiliza a informação da derivada da função de verossimilhança. Dessa forma, uma densidade auxiliar que pode ser escolhida é a normal multivariada, como em Metropolis-Hastings, com média dada por $\theta + h\Delta \ln p(\theta | y)/2$, sendo θ o valor correto da média, h a variância estipulada pelo usuário e $\Delta \ln p(\theta | y)$ o gradiente da densidade a posteriori. A variância da densidade continua sendo hI , para $h > 0$. A idéia continua sendo a mesma do algoritmo Metropolis-Hastings, porém, a probabilidade de aceitação será:

$$p(\theta | y) \exp(-\frac{1}{2h} \|\theta - E(\theta)\|^2) / K q(\theta) \exp(-\frac{1}{2h} \|\theta - E(\theta)\|^2),$$

sendo $E(\cdot)$ a esperança matemática. Segundo Christensen et al. (2001), intuitivamente, a informação do gradiente ajuda o processo direcionando o algoritmo para a região de maior densidade da posteriori. Em aplicações de estatística espacial, o algoritmo de Langevin-Hastings apresenta um desempenho superior, se comparado ao algoritmo de Metropolis-Hastings (Christensen & Waagepetersen, 2002; Waagepetersen et al., 2008). As características de convergência deste processo também são válidas para o algoritmo de Langevin-Hastings, portanto, devem também ser investigadas.

Segundo Waagepetersen et al. (2008), a escolha da variância, h , deve ser feita para se obter uma taxa de aceitação de 25% para o passeio aleatório e de 60% para a atualização do algoritmo de Langevin-Hastings. Quanto à aplicação deste algoritmo no estudo de estatística espacial, Christensen & Ribeiro Jr. (2007) sugerem uma taxa de aceitação para o processo que deve ser de 60% e quando se utiliza o parâmetro ϕ com qualquer distribuição a priori, este deve ter uma taxa de aceitação, de preferência, entre 25% e 30%.

2.4.6.3 Verificação formal da convergência

A inferência bayesiana utilizando procedimento MCMC tem ganhado cada vez mais espaço, mas é necessário verificar a convergência do processo de simulação para se obter resultados confiáveis das distribuições de interesse. Para isso, é fundamental utilizar critérios para a verificação da convergência. Estes critérios permitem avaliar se há necessidade de se realizar um número de iterações maior dos processos iterativos das cadeias de Markov. A partir do trabalho de Nogueira et al. (2004), escolheu-se trabalhar, de maneira conjunta, com os testes univariados de Gelman & Rubin (1992), Raftery & Lewis (1992), Geweke (1992), Heidelberger & Welch (1983) e o multivariado de Brooks &

Gelman (1998). Estes métodos já estão presentes no pacote *Bayesian Output Analysis*, ou BOA, do software R (R Development Core Team, 2007). Para maiores informações a respeito do pacote BOA, ver Smith (2005). Neste pacote também estão presentes ferramentas de análise informal da convergência que são de grande importância e serão exploradas, como os estimadores do erro de Monte Carlo, autocorrelação e médias móveis.

2.4.6.3.1 Critério de Gelman & Rubin (1992)

O método envolve simulações provindas de uma complexa e, geralmente, multivariada distribuição de interesse, $P(\theta)$, indiretamente gerado de uma cadeia de Markov com a densidade de interesse, uma densidade estacionária. Geralmente, são simuladas $m \geq 1$ seqüências de simulações, cada uma de tamanho n , $(\theta_{j1}, \theta_{j2}, \dots, \theta_{jn})$, para $j = 1 \dots m$ (sendo cada θ_{jt} um vetor). Se $m > 1$, as m seqüências são, usualmente, mas nem sempre, simuladas independentemente.

Gelman & Rubin (1992) mostraram que, em muitos casos, a convergência pode ser facilmente determinada provinda de seqüências múltiplas independentes em paralelo, mas não pode ser diagnosticada utilizando-se o resultado da simulação provinda de qualquer seqüência simples. Estes autores propuseram um método que utiliza múltiplas repetições de cadeias para decidir se a estacionariedade foi alcançada dentro da segunda metade de cada amostra.

O método pressupõe que m seqüências tenham sido simuladas em paralelo, cada uma partindo de diferentes pontos iniciais. Tendo obtido os pontos iniciais satisfatórios, ou seja, pertencentes ao espaço paramétrico da posteriori, as seqüências são geradas para $2n$ iterações, das quais as primeiras n são descartadas para evitar o período de aquecimento (*burn-in*), ou seja, evitar o uso de valores de uma amostra não estacionária.

As m seqüências rendem m possíveis inferências. Se estas inferências são bastante similares, isto é um indicativo de que a convergência foi alcançada ou está próxima.

Considere uma variável aleatória θ , que tem média μ e variância σ^2 , sob a distribuição de interesse. Considere também que se tem um estimador $\hat{\mu}$ não viesado para μ . Dado θ_{ji} , denota o i -ésimo valor de um total de n de θ na seqüência j , toma-se $\hat{\mu} = \bar{\theta}_{..}$ e calcula-se a variação entre seqüência E/n e a variação dentro de seqüência D , definidas por:

$$E/n = \frac{1}{m-1} \sum_{j=1}^m (\bar{\theta}_j - \bar{\theta}_{..})^2$$

$$D = \frac{1}{m(n-1)} \sum_{j=1}^m \sum_{i=1}^n (\theta_{ji} - \bar{\theta}_j)^2.$$

Note-se que a análise de variância (ANAVA) assume um *pool* de variâncias total, que pode ser produzido porque, sobre convergência, a variação total das seqüências é igual, de fato, à variação dentro de seqüência, pois a variação entre seqüências se anula.

Depois de calculadas as estimativas anteriores, pode-se estimar $\hat{\sigma}^2$ pela média de E e D :

$$\hat{\sigma}^2 = \frac{n-1}{n} D + \frac{E}{n},$$

a qual pode ser uma estimativa não viesada da verdadeira variância se os valores iniciais das seqüências provierem da distribuição de interesse.

Considerando-se a variância do estimador um *pool* de variâncias:

$$\hat{V} = \hat{\sigma}^2 + \frac{E}{mn}.$$

A comparação entre as variâncias é expressa por:

$$R = \frac{\hat{V}}{\sigma^2},$$

o qual chama-se fator de redução de escala (da sigla inglesa SRF). O denominador de R não é conhecido, devendo ser estimado dos dados; pode-se, por isso, ganhar uma superestimativa de R , pois σ^2 será subestimada pelo D . Por isso, emprega-se uma superestimativa de R , utilizando:

$$\hat{R} = \frac{\hat{V}}{D} = \frac{m+1}{m} \left(\frac{\hat{\sigma}^2}{D} \right) - \frac{n-1}{mn},$$

o qual é conhecido como fator de redução de escala potencial (sigla em inglês PSRF) e pode ser interpretado como um fator de diagnóstico da convergência. Brooks & Gelman (1998) comentam que Gelman & Rubin (1992) adotaram incorretamente o fator de correção aplicado à variância para a variabilidade amostral, i.e. $d/(d-2)$, em que d representa os números de graus de liberdade. O uso deste fator incorreto conduziu a um número de problemas, em que o corrigido pode ser infinito ou, até mesmo, negativo. Então, seguindo o método de Fisher (1935), baseado em amostras independentes providas da distribuição normal, o parâmetro chave para este ajustamento são os graus de liberdade d . Para informar corretamente a variabilidade amostral, o fator correto é:

$$\hat{R} = \left(\frac{d+3}{d+1} \right) \frac{\hat{V}}{D}, \quad (2.4)$$

esta correção será importante para pequenas amostras. Quando o \hat{R} for grande, sugere-se que qualquer das estimativas das variâncias pode ser reduzida com a utilização de mais iterações, desde que as seqüências simuladas ainda não tenham feito um passeio por toda a distribuição de interesse. Se PSRF aproxima-se de 1, pode-se concluir que cada m conjunto de n iterações se aproxima da distribuição de interesse (Brooks et al., 1997). Gelman (1995) sugere aceitar, como garantia de convergência, valores de \hat{R} menores que 1,1 ou 1,2.

Em alguns casos, pode-se encontrar uma infinidade de parâmetros de interesse para serem monitorados. No entanto, para acompanhamento individual, acaba sendo inviável. Por isso, Brooks & Gelman (1998) estenderam o método para o caso de interesse não somente em funções escalares, mas algo mais geral, como o caso de função multivariada das seqüências.

2.4.6.3.2 Critério multivariado de Brooks & Gelman (1998)

O critério de Brooks & Gelman (1998) é a extensão multivariada para o critério de Gelman & Rubin (1992) (Equação 2.4). Por analogia, quando se está interessado na estimação de um vetor de parâmetros $\underline{\theta}$ com base nas observações $\theta_{ji}^{(p)}$, sendo $\theta_{ji}^{(p)}$ o p -ésimo elemento do vetor de parâmetros da seqüência j na iteração i . Para grandes dimensões, deve-se estimar a matriz de covariâncias a posteriori:

$$\hat{V} = \frac{n-1}{n}D + \left(1 + \frac{1}{m}\right)\frac{E}{n},$$

em que:

$$D = \frac{1}{m(n-1)} \sum_{j=1}^m \sum_{i=1}^n (\theta_{ji} - \bar{\theta}_j)(\theta_{ji} - \bar{\theta}_j)^T$$

e

$$\frac{E}{n} = \frac{1}{m-1} \sum_{j=1}^m (\bar{\theta}_j - \bar{\theta}_{..})(\bar{\theta}_j - \bar{\theta}_{..})^T$$

são matrizes p -dimensionais estimadas de p -variáveis. Pode-se, desta maneira, monitorar ambas as variâncias, \hat{V} e D , em que E e D são as matrizes de covariâncias entre e dentro de seqüências, respectivamente.

Resume-se a distância entre \hat{V} e D com uma medida escalar que deve se aproximar a 1 para verificar se a convergência foi alcançada. Uma maneira de se fazer isto é trabalhar com o máximo da raiz característica, que é o máximo do SRF de qualquer projeção linear de θ .

O máximo é dado a partir da derivada em relação a \underline{a} . Igualando-a a zero, tem-se:

$$\hat{R}^P = \max_{\underline{a}} \frac{\underline{a}^T \hat{V} \underline{a}}{\underline{a}^T D \underline{a}}, \quad (2.5)$$

que pode ser o multivariado PSRF (MPSRF). O lema pode ser rescrito por:

$$\hat{R}^P = \frac{n-1}{n} + \left(\frac{m+1}{m} \right) \lambda_1,$$

sendo λ_1 o maior autovalor da matriz positiva definida simétrica $D^{-1}E/n$. Sobre igualdade de médias entre seqüências, $\lambda_1 \rightarrow 0$ e $\hat{R}^P \rightarrow 1$, quando n for grande o suficiente.

2.4.6.3.3 Critério de Raftery & Lewis (1992)

O procedimento diagnóstico sugerido por Raftery & Lewis (1992) é um método que estima quantas iterações são necessárias para que o amostrador de Gibbs apresente convergência à distribuição estacionária, propondo

especificações de quantidades necessárias para isto. Deve-se ter uma seqüência do amostrador de Gibbs com um N mínimo, que é o número mínimo de iterações requeridas para se obter a precisão necessária de estimação. O método fornece as estimativas do *burn-in* (B), que é o número de iterações que devem ser descartadas, o número de iterações que devem ser computadas N (total) e o k , que é a distância mínima de uma iteração à outra, para se obter a subamostra aproximadamente independente (*thin*). Esses valores serão calculados mediante especificações para garantir que um quantil u de uma determinada função $f(\theta)$ seja estimado com precisão pré-definida.

A presença de inúmeras variáveis possibilita resultados dessa análise para cada uma delas; assim, devem-se tomar os maiores valores de B , N e k , pois a convergência só se dará quando todas as variáveis convergirem.

Outra saída importante fornecida pelo critério é o fator de dependência que é responsável pelo acréscimo multiplicativo ao número de iterações necessárias para se alcançar a convergência, devido à autocorrelação dentro de seqüência. Segundo Raftery & Lewis (1992), se este fator for maior que 5,0, pode-se dizer que a convergência ainda não foi obtida, sendo necessário reparametrizar o modelo em estudo.

2.4.6.3.4 Critério de Geweke (1992)

Os métodos de Monte Carlo Cadeias de Markov são especiais casos de séries temporais estacionárias, então, é natural usar métodos de séries temporais para testar a convergência. Estes dois próximos métodos existentes na literatura utilizam técnicas padrão de análise espectral para estimar a verdadeira variância via densidade espectral, $S(w)$.

A ausência de independência entre os dados da amostra é considerada um problema. Na literatura, tem-se tratado este problema de modo informal, denotando-o a um problema computacional. Usando técnicas padrão de análise

espectral, promove-se um erro padrão para a aproximação $E[g(\theta)]$ das médias amostrais correspondentes; dessa maneira, leva-se em conta qualquer autocorrelação em que $g(\theta)$ é uma função real. Com base nestes resultados, um diagnóstico para a ausência de convergência do amostrador de Gibbs é construído (GEWEKE, 1992).

Considerando a função real $g(\theta)$ e sua trajetória $g^{(1)}, g^{(2)}, \dots$, construída a partir de $g^{(i)} = g(\theta^{(i)})$, $i = 1, 2, \dots, n$, essa trajetória define uma série temporal. Pode-se, portanto, estimar a média desta série utilizando-se o seguinte estimador:

$$\bar{g}_i = \left(\frac{1}{n}\right) \sum_{i=1}^n g^{(i)}.$$

Ela fornece um estimador não viesado de $E[g(\theta)]$, cuja variância assintótica é dada por $S_g(0)/n$, sendo $S_g(w)$ a densidade espectral da série g de frequência 0. Em geral, S_g é desconhecido, sendo estimado por \hat{S}_g com base em análise espectral. Após um número suficiente grande, n , de iterações, determinam-se as médias \bar{g}_A e \bar{g}_B , sendo \bar{g}_A a média das primeiras n_A e \bar{g}_B , a média das n_B últimas iterações. Considerando, também, os estimadores independentes $\hat{S}_g^A(0)$ e $\hat{S}_g^B(0)$ das variâncias assintóticas de $\{g^{(i)} : i = 1, \dots, n_A\}$ e $\{g^{(i)} : i = n^*, \dots, n_B\}$, respectivamente, sendo $n^* = n - n_B + 1$, pode-se mostrar que, se n_A/n e n_B/n são fixos, com $(n_A + n_B)/n < 1$, quando $n \rightarrow \infty$, tem-se a diferença padronizada entre as médias:

$$\frac{\bar{g}_A - \bar{g}_B}{\sqrt{\left(\hat{S}_g^A(0)/n_A\right) + \left(\hat{S}_g^B(0)/n_B\right)}} \approx N(0,1), \quad (2.6)$$

se a seqüência $\{g^{(i)}\}$ for estacionária. Assim, um teste para uma seqüência pode ser construído. Se a diferença padronizada entre as médias for grande, existe indicação de ausência de convergência. As médias devem ser construídas após algumas iterações iniciais terem sido descartadas e que sejam usados os valores de $n_A = 0,1n$ e $n_B = 0,5n$ (Geweke, 1992).

Na prática, deve-se notar que esta tentativa de diagnóstico para verificar uma necessária, mas não suficiente, condição de convergência, informa ao pesquisador somente se a convergência não foi alcançada, e não se realmente convergiu (Nogueira, 2004).

2.4.6.3.5 Critério de Heidelberger & Welch (1983)

Heidelberger & Welch (1983) apresentaram um método para determinar o valor inicial de uma seqüência discreta de eventos, que é também apropriada para ser utilizada para o diagnóstico de convergência do amostrador de Gibbs. Para testar a hipótese nula de estacionariedade, a partir de uma seqüência de dados $\{g^{(i)} : i = 1, \dots, n\}$ oriundos de um processo não estacionário de desconhecida densidade espectral $S(w)$, utiliza-se a teoria browniana para fazer esta ligação e, assim, avaliar possível superdispersão na presença dos valores iniciais. Pode-se, portanto, utilizar um teste estatístico para testar a hipótese de nulidade, como Cramer-von Mises (Von Mises, 1931), que testa se os valores amostrados para cada variável provêm de uma distribuição estacionária. Se a hipótese nula é rejeitada para um dado valor, o teste é repetido depois de descartadas os primeiros 10% das iterações. Se a hipótese é novamente rejeitada, outros 10% são descartados após o descarte dos 10% primeiros. Este processo é repetido até se ter uma proporção de 50% (10%, 20%, ..., 50%). Se a hipótese for novamente rejeitada ou o teste não conseguir ser realizado, isto indica falha da estacionariedade, implicando que é necessário um número maior de iterações.

Se o teste for satisfatório, o número inicial de iterações descartadas é indicado como o tamanho do *burn-in* (Nogueira, 2004).

Com o teste da estacionariedade, existe o teste de *half-width*, que é utilizado para verificar se a média estimada está sendo calculada com uma acurácia pré-especificada, na porção da seqüência que passou no teste de estacionariedade, para cada variável em estudo. Calcula-se um intervalo de confiança com 95% para a média, utilizando-se o desvio padrão assintótico via métodos de séries temporais, ou seja, utilizando densidade espectral. Pelo teste, verifica-se que a razão entre o (limite superior - limite inferior)/2 e a estimativa da média é menor que a fração específica pré-definida (acurácia), o qual, segundo Heidelberg & Welch (1983), deverá ser de 0,1. O resultado passará no teste se isto for verdadeiro, indicando que a média amostral é estimada com um erro relativo aceitável, portanto, julgada ser a média a posteriori, sendo a amostra realizada da distribuição de interesse. Se o teste falhar, é necessário um número maior de iterações.

Heidelberg & Welch (1983) sugerem este processo iterativo como estimativa do tamanho do *burn-in*. O diagnóstico tem muito pouco poder para detectar a falta de convergência se a amostra for de tamanho muito pequeno, podendo ser problema também em obter estimativas fidedignas da densidade espectral, o que ocorre com o método de Geweke.

2.4.7 Inferência bayesiana para modelo linear espacial generalizado

2.4.7.1 A definição do modelo

Assume-se que Y_i segue uma distribuição binomial i.i.d. ($E[y_i|s_i] = \mu_i$) com o modelo de preditores dado por $h[\mu_i] = X^T \beta$, sendo $h[\cdot]$ a função de ligação logística para o risco de morte, X a matriz que inclui as covariáveis e β um vetor de parâmetros dos efeitos das covariáveis. Este modelo é dado por:

$$\text{Log} \left\{ \frac{p(s)}{1-p(s)} \right\} = \alpha + \beta_1 x_i + U(s_i), \quad (2.7)$$

sendo x_i representante da específica covariável em estudo e os β s os parâmetros dos efeitos da covariável sobre o logaritmo do risco relativo. O termo U_i caracteriza-se por ser um processo espacial gaussiano, isotrópico, estacionário latente, $U \sim N(0, \Sigma)$, sendo Σ_{ij} uma função paramétrica da distância d_{ij} . A matriz de covariância pode ser definida como $\Sigma_{ij} = \sigma^2 \rho(d_{ij}; \phi)$, sendo σ^2 a variância do processo espacial e $\rho(d_{ij}; \phi)$ a função de correlação com um parâmetro escalar ϕ , o qual controla a taxa de correlação, que decai com a distância, $d_{ij} = s_i - s_j$, em que d_{ij} é a distância euclidiana entre as localizações s_i e s_j .

Para um exemplo de situação, considere um modelo normal multivariado, no qual a média é explicada por um modelo de regressão linear:

$$Y \sim NM(X^T \beta, \sigma^2 R(\phi) + \tau^2 I) \quad (2.8)$$

sendo X a matriz ($n \times p$) de covariáveis, β o vetor ($p \times 1$) de parâmetros da regressão e $R(\phi)$ a matriz ij -ésimo elemento dado pela correlação $r_{ij} = \rho(d_{ij}; \phi)$.

Fazendo-se o conjunto de parâmetros que define a estrutura de covariância do modelo $\theta = (\tau^2, \sigma^2, \phi)$, então, $V(\theta) = \sigma^2 R(\phi) + \tau^2 I$. Portanto, $Y \sim NM(X^T \beta, V(\theta))$.

Dessa forma, a função de log-verossimilhança pode ser definida por:

$$l(Y | \beta, \theta) = -\frac{1}{2} \left\{ \log |V(\theta)| + (Y - X^T \beta)^T [V(\theta)]^{-1} (Y - X^T \beta) \right\}.$$

Para Diggle & Ribeiro Jr. (2007), usando uma estrutura de independência condicional existente entre os termos do modelo, pode-se definir

o modelo como os parâmetros da estrutura de covariância, θ , independente dos parâmetros do modelo de regressão, β , e os parâmetros da estrutura de covariância independentes dos dados mensurados. Esta estrutura permite uma melhor amostragem dos processos MCMC. Assim tem-se:

$$P(\theta|U) \propto P(U|\theta)P(\theta), \quad (2.9)$$

o termo U representa o vetor de valores de $U(s_i)$ e Y corresponde ao vetor de valores mensurados. Como U é uma variável latente e Y condicional a esta variável, pode-se utilizar a metodologia de dados aumentados para Y e U . Esta metodologia foi utilizada, para análise fatorial, por Sáfadi & Peña (2007), ao considerar que os fatores são variáveis latentes. A idéia essencial de se utilizar a metodologia de dados aumentados é determinar a distribuição condicional dos parâmetros desconhecidos condicional ao fator latente, posteriormente à distribuição condicional do fator latente, dados os demais parâmetros desconhecidos. Dessa forma, o processo MCMC é realizado em duas etapas. Para maiores detalhes a respeito do algoritmo, ver Diggle & Ribeiro Jr. (2007).

A estrutura de modelo linear generalizado implica que a verossimilhança pode ser dada por:

$$L(Y|\beta, U) = \prod_{j=1}^n L(Y_j|\beta, U_j). \quad (2.10)$$

Neste modelo, $L(Y_j|U_j, \beta) = P(y, \mu_j)$, para $\mu_j = h^{-1}[X_j^T \beta + U(s_j)]$, sendo $h[\cdot]$ a função de ligação do modelo linear generalizado.

2.4.7.2 Distribuições a priori

Os parâmetros que compõem o modelo são dados por: β , um vetor para os parâmetros da regressão, σ^2 e ϕ , os parâmetros da covariância. Considerando dois casos para o parâmetro ϕ , as prioris são definidas a seguir.

2.4.7.2.1 Priori para o caso em que se considera ϕ conhecido

Considere-se o modelo definido na equação (2.8), assumindo que o parâmetro da função de correlação é conhecido, ausência de efeito *nugget* ($\tau^2 = 0$) e que os parâmetros a serem considerados pelo modelo são β e σ^2 , a fórmula de priori conjugada a ser utilizada, sugerida por Diggle et al. (2003), é um qui-quadrado inverso escalado gaussiano, ou seja,

$$(\beta | \sigma^2, \phi) \sim N(m_b, \sigma^2 V_b), \quad (2.11)$$

$$(\sigma^2 | \phi) \sim \chi^2_{\text{Inverso escalado}}(n_\sigma, S_\sigma^2), \quad (2.12)$$

sendo a densidade do qui-quadrado inverso escalado dada por:

$$P(\sigma^2) \propto (\sigma^2)^{-(n_\sigma/2+1)} \exp\{-n_\sigma S_\sigma^2 / (2\sigma^2)\}.$$

Assim,

$$(\beta, \sigma^2 | \phi) \sim N\chi^2_{\text{Inverso escalado}}(m_b, V_b, n_\sigma, S_\sigma^2). \quad (2.13)$$

Na prática, ao assumir uma distribuição a priori, pode ser complicado e, por isso, prioris não informativas podem ser assumidas. Um exemplo é a priori de Jeffreys ($P(\beta, \sigma^2) \propto 1/\sigma^2$).

2.4.7.2.2 Priori para o caso em que se considera ϕ desconhecido

Diggle et al. (2003) sugerem o uso do produto de prioris,

$$P(\beta, \sigma^2, \phi) = P(\beta, \sigma^2 | \phi) P(\phi),$$

sendo $P(\beta, \sigma^2 | \phi)$ dado na equação (2.13) e $P(\phi)$, qualquer priori própria ou imprópria. Para a situação de priori própria é usual o uso da distribuição uniforme ou a exponencial. Prioris impróprias devem ser utilizadas com cautela, pois permitem o aparecimento de posteriores imprópria.

2.4.7.2.3 Distribuição a posteriori

A distribuição a posteriori condicional de β pode ser definida por:

$$P(\beta | U, Y) \propto L(Y | U, \beta) P(\beta).$$

Segundo Diggle & Ribeiro Jr. (2007), a posteriori completa para o modelo com distribuição normal multivariada (Equação 2.8), pode ser dada por:

$$(\beta, \sigma^2 | \phi, Y) \sim N\chi_{Inverso\ escalado}^2(\tilde{\beta}, V_{\tilde{\beta}}, n_{\sigma} + n, S^2),$$

sendo $\tilde{\beta} = V_{\tilde{\beta}}(V_b^{-1}m_b + X^T R^{-1}y)$, $V_{\tilde{\beta}} = (V_b^{-1} + X^T R^{-1}X)^{-1}$ e $S^2 = (n_{\sigma}S_{\sigma}^2 + m_b^T V_b^{-1}m_b + y^T R^{-1}y - \tilde{\beta}^T V_{\tilde{\beta}}^{-1}\tilde{\beta}) / (n_{\sigma} + n)$.

Para a distribuição de cada parâmetro a ser predito, a posteriori condicional pode ser dada por:

$$P(U | \beta, \phi, \sigma^2, Y) \propto P(Y | U, \beta) P(U | \phi, \sigma^2).$$

Para Diggle & Ribeiro Jr. (2007), $U(\cdot)$ é um processo normal e $P(U | \phi, \sigma^2)$ segue, portanto, uma distribuição normal multivariada.

2.4.7.3 Distribuição preditiva para o modelo linear espacial generalizado

A distribuição à posteriori de U é dada por:

$$P(U | Y) \propto \prod_{i=1}^n g(y_i; h^{-1}(s_i)) P(U),$$

sendo $g(y, \mu)$ a distribuição do erro parametrizada pela média e $P(U)$ a distribuição à priori de U .

A distribuição marginal de U pode ser dada por uma distribuição t multivariada, $t_{n_\sigma}(m_b, S_\sigma^2(R + XV_b X^T))$, sendo m_b o vetor de hiperparâmetros das médias de β (Equação 2.11); S_σ^2 o hiperparâmetro da variância da priori de σ^2 ; R a matriz de correlação espacial; X a matriz de covariáveis e V_b a matriz de hiperparâmetro de pesos das covariâncias, dada pela Equação (2.11) (Diggle et al. 2003).

Para obter as amostras, Diggle et al. (2003) utilizaram o algoritmo de Langevin-Hastings, com a seguinte reparametrização:

$$U = X^T m_b + S_\sigma (R + XV_b X^T) \Omega^{1/2} \Gamma,$$

sendo $\Omega^{1/2}$ a raiz quadrada de $\Omega = S_\sigma^2 (R + XV_b X^T)$ e uma priori para Γ , dada por distribuição t multivariada n -dimensional $t_{n+n_\sigma}(0, I_n)$. Segundo os autores, o

gradiente a ser utilizado, considerando que h é uma função de ligação canônica, é dado por:

$$\Delta \ln p(\theta | y) = -\theta(n - n_\sigma) / (n_\sigma + \|\theta\|^2) + (\Omega^{1/2})^T \{y - h^{-1}(s)\}.$$

3 METODOLOGIA

Os trabalhos epidemiológicos relacionados à saúde coletiva demonstram que o processo de saúde-doença ocorre de maneira espacial e temporal. Procura-se trabalhar aqui com as bases teóricas para análise espacial de dados epidemiológicos. O estudo é caracterizado como sendo um estudo ecológico de múltiplos grupos, tratando a linha do tempo de maneira transversal e explorando os seguintes aspectos: análise exploratória de dados espaciais com a verificação da ocorrência de aglomerados espaciais, estudo das covariáveis e, por fim, a modelagem baseada em geoestatística via inferência bayesiana.

3.1 Área do estudo

A área estudada correspondeu à região urbana do município de Alfenas, Minas Gerais. Alfenas está localizada na bacia hidrográfica do Rio São Tomé, que termina no lago de Furnas. Ocupa uma área de 849,2 km² de extensão e a altitude máxima é de 888 metros acima do nível do mar. A cidade situa-se segundo as coordenadas geográficas 21°25'45'' latitude Sul e 45°56'50'' W. Greenwich. Segundo o IBGE, é composta de 70 setores censitários, nos quais se localizam residências de todas as classes sociais. A base cartográfica completa, com ruas e lotes, foi digitalizada pela equipe de geoprocessamento do município de Alfenas (Figura 1) e a base de setores censitários obtida do programa *Estatcart*© do IBGE e convertida para o aplicativo *ArcMap*© (Figura 2).



FIGURA 1 Cartograma dos logradouros e bairros da cidade de Alfenas, MG.

A representação gráfica completa dos loteamentos da cidade de Alfenas e respectivos logradouros da região urbana e a localização dos bairros encontram-se na Figura 1.

Na Figura 2 estão apresentados os setores censitários da cidade de Alfenas. Neste cartograma, os limites considerados também são os limites do município, portanto, contêm a região urbana e a rural.



FIGURA 2 Cartograma dos setores censitários do município de Alfenas, MG.

As análises de estatística espacial foram realizadas excluindo-se os setores censitários do distrito industrial e regiões rurais, devido às grandes áreas e às pequenas populações destes setores censitários especiais. Os diferentes métodos de interpolação utilizados pressupõem continuidade espacial, existente no terreno, exceto nas regiões excluídas, como pode ser visto na Figura 3:

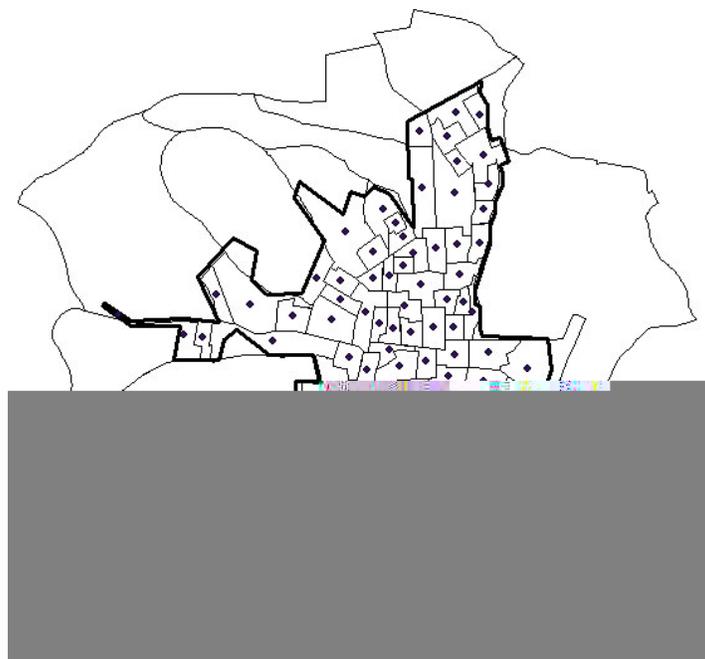


FIGURA 3 Cartograma dos setores censitários e os limites da área considerada no estudo, com respectivos centróides.

Na Figura 3 foi demarcado um novo limite, que foi a área considerada no estudo, utilizada como representação da área povoada. A representação espacial do setor censitário foi feita utilizando-se o centróide como representante das informações da região espacial específica e os limites dos setores foram ignorados, pois a geoestatística assume continuidade espacial.

3.2 Bases territoriais

A perspectiva deste trabalho foi a análise do perfil de risco de morte infantil da população urbana, considerando o coeficiente de mortalidade infantil. De um lado, está a base de dados dos logradouros da cidade de Alfenas. Do outro lado, está disponível a base cartográfica digital (a malha por setores censitários) do município em estudo. Na base de dados de logradouros, a partir

do endereço residencial, foi localizado o exato ponto em que residia cada indivíduo que veio a óbito em um ano e, assim, a base classificada como padrão pontual. Portanto, dessa forma, existem dois grupos: o das localizações das mortes e o dos nascimentos. Nas Figuras 4 e 5 estão apresentados os dados de mortes e de nascimentos, respectivamente. Na base de dados dos setores censitários, os dados pontuais de morte foram agrupados por setor censitário, computado o número de mortes ocorridas no respectivo setor, sendo o referenciamento espacial feito pelo centróide do setor (FIGURA 6). Para o cálculo do coeficiente de mortalidade, razão entre número de mortes por nascidos vivos, foi utilizado o número de mortes no setor pelo número de nascidos vivos no respectivo setor censitário.

3.3 Dados demográficos e sócio-econômicos

O problema, neste estudo, se refere a uma análise de dados com resposta binária, obtidos de uma amostra de localizações em uma região fixa A . Seja $\{s_i : s_i \in A \subseteq \mathbb{R}^d\}$ para $i = 1, \dots, n$, denotando um conjunto espacial de localizações.

Para o estudo, utilizou-se um Sistema de Informação Geográfica (SIG) que conta com bases cartográficas de logradouros, da malha de lotes e dos limites do município de Alfenas. Este conteúdo foi fornecido pela Equipe de Informação em Saúde, da Coordenadoria Geral de Vigilância em Saúde, da Secretaria Municipal de Saúde de Alfenas e da equipe de georreferenciamento.

Os registros de nascimentos e mortes foram coletados a partir dos bancos de dados provenientes do Sistema de Informação de Nascimentos (SINASC) e do Sistema de Informação de Mortes (SIM). O SINASC é um subsistema de informações de âmbito nacional, sob responsabilidade das secretarias estaduais e municipais de saúde, concebido para diminuir a subnumeração e melhorar a qualidade das informações sobre nascidos vivos.

Este sistema é alimentado pelas declarações de nascido vivo (DN), documento oficial emitido pelo hospital onde ocorre o nascimento (D'Orsi & Carvalho, 1998). O SIM é alimentado pelos atestados de óbito (DO). Os dados foram selecionados à medida que foi possível o seu georreferenciamento.

O período específico em estudo foi de 2000 a 2004. Do registro do SIM, foram selecionados apenas os natimortos e os óbitos até 1 ano de vida e, do SINASC, foram todos os nascidos vivos, neste período, na cidade de Alfenas, MG. Dessa forma foi caracterizado o conjunto de dados.

O georreferenciamento das residências foi realizado pela comparação entre o endereço que consta no sistema de informação com o cadastro de logradouros. As coordenadas geográficas de cada ponto foram calculadas a partir da interpolação da Inscrição Cadastral (IC) dos lotes de logradouros por meio do SIG, permitindo a visualização da distribuição dos eventos na base cartográfica, em que cada ponto corresponde à localização da residência de óbito ou de nascimento. Este referenciamento só permitiu localizar 46% dos locais de residência das vítimas pelo endereço e 24% nos dados de nascimento. Isso ocorreu devido à qualidade do preenchimento do banco de dados e de possíveis alterações no sistema de informação urbano. Alfenas é considerada uma cidade pólo regional, com a presença de hospital universitário. Por esta razão, é esperada uma migração, no intuito de se utilizar o sistema de saúde. Para tal utilização, é necessário um cadastro com endereço em Alfenas, o que, muitas vezes, é fictício.

As Figuras 4 e 5 são representações gráficas do conjunto de dados pontuais, sendo amostras das distribuições de morte e de nascimentos, respectivamente.

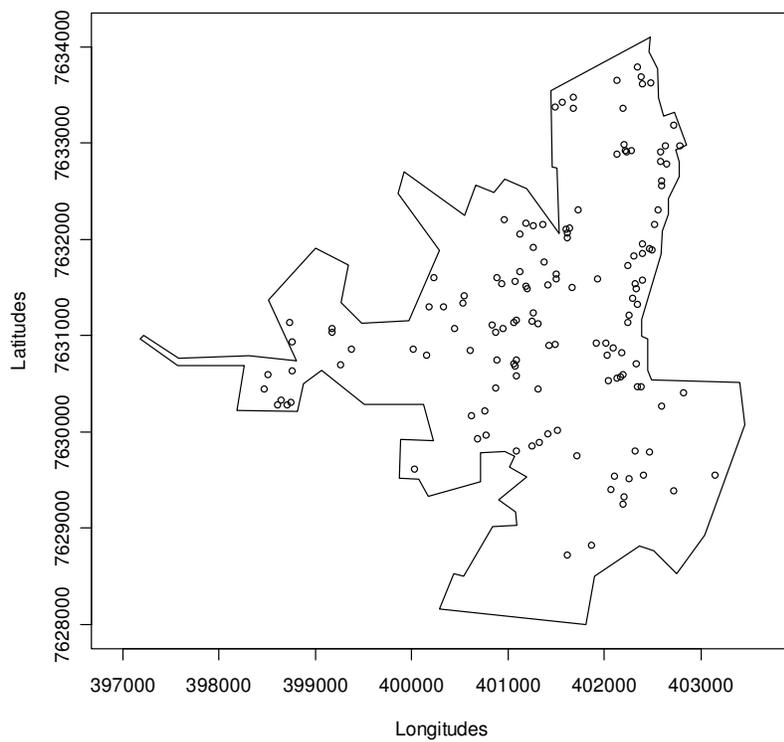
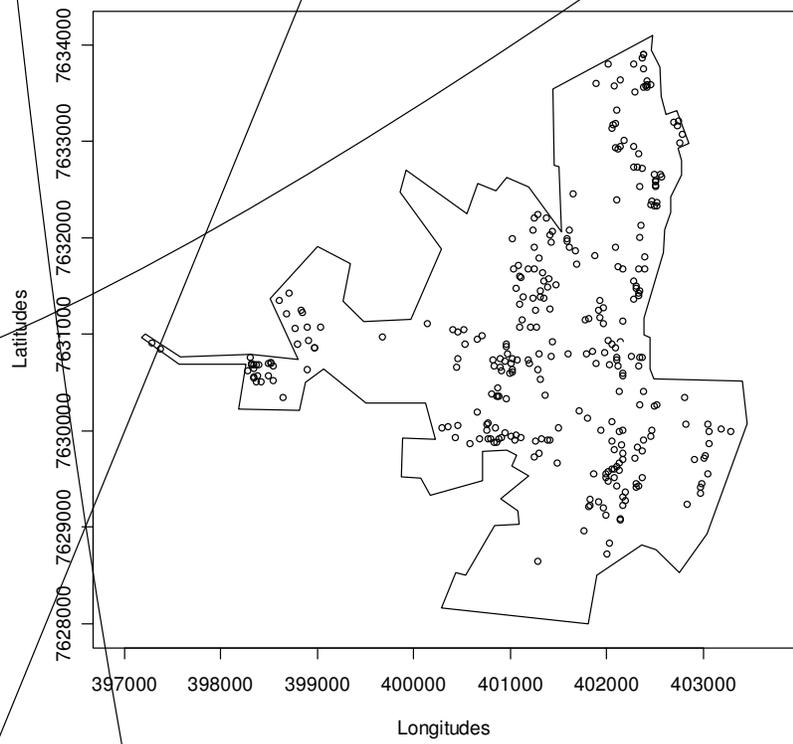


FIGURA 4 Representação pontual dos casos de mortalidade infantil, entre 2000 e 2004, no município de Alfenas, MG.

Para o ajuste do padrão pontual foram considerados 137 pontos de mortalidade e 1.229 pontos de nascimentos.



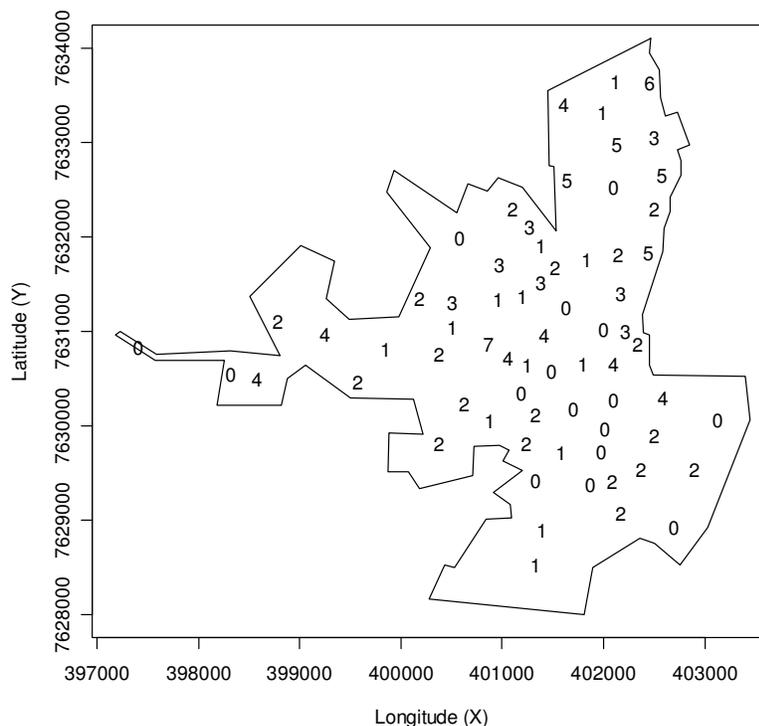


FIGURA 6 Número de mortes em cada setor censitário, localizado pelo centróide, entre 2000 e 2004, ocorridas no município de Alfenas.

A representação gráfica permite a visualização de alguns pontos com a ausência de ocorrência de casos e outros com até 7 casos.

3.4 Análise exploratória

Para a realização da análise descritiva, o uso da função K bivariada e da função K para a diferença foi considerado. Para a análise, foram utilizados os dados de mortalidade e de nascimento. A distância mínima utilizada no estudo foi de 0 (zero). A distância máxima escolhida neste trabalho foi determinada pela distância média entre as residências. Acredita-se que os aglomerados não

ocorrerem com menos do que 20 metros. A função K univariada não foi estimada, pois a comparação com um processo de Poisson homogêneo não foi de interesse uma vez que a população não se distribui de maneira homogênea na área em estudo. Devido a esta situação, para representar a distribuição espacial da população, foi utilizada a amostra de nascimento (Figura 5).

As simulações para a construção dos limites de significância do teste foram realizadas utilizando-se 1.000 iterações para cada distância. Para a realização das análises, o software utilizado foi o R (versão 2.5.1) (R Development Core Team, 2007) com o pacote *splanx* (versão 2.01-23) (Rowlingson & Diggle, 2007).

Na segunda etapa, a escolha das variáveis independentes para participar do modelo logístico como covariáveis foi feita de acordo com a importância de cada uma no contexto epidemiológico, com o intuito de investigar as variáveis de risco sócio-econômicas e as de risco biológico (Rouquayrol & Almeida Filho, 2003).

3.5 Modelo linear espacial generalizado

Com o perfil de óbitos e de nascimentos, é possível a obtenção da medida de risco, que é definida pela proporção de pessoas que estão afetadas no início de um período estudado, mas que desenvolvem o evento de risco durante este período, sendo o evento, no nosso caso, a morte (Jekel et al., 2005).

Seja Y_i a variável correspondente à resposta binária para o risco de morte na localização s_i . Assume-se que y_i , $i = (1, \dots, n)$, condicional ao processo espacial, são realizações de uma variável aleatória independente binomial $(n_i, p(s))$, com média $(E[y_i|s_i] = \mu_i)$ e o modelo de preditores dado por $h[\mu_i] = X^T \beta$, sendo $h[\cdot]$ a função de ligação logística para o risco de morte dada por:

$$\text{Log} \left\{ \frac{p(s, x)}{1 - p(s, x)} \right\} = \alpha + \beta_1 x_i + U_i, \quad (2.14)$$

sendo X_i representante da covariável em estudo e os β s os parâmetros dos efeitos da covariável sobre o logaritmo do risco relativo, no caso *odds ratio*. O termo U_i caracteriza-se por ser um processo espacial gaussiano, isotrópico, estacionário latente, $U \sim N(0, \Sigma)$, sendo $\Sigma_{ij} = \sigma^2 \exp(-d_{ij}/\phi)$, em que σ^2 é a variância do processo espacial e ϕ o parâmetro da função exponencial de correlação, o qual controla a taxa de correlação que decai com a distância, $d_{ij} = s_i - s_j$, que é a distância euclidiana entre as localizações s_i e s_j .

As variáveis definidas como covariáveis, obtidas do IBGE (censo de 2000), foram: número de mulheres em idade fértil, número de mulheres em idade de risco gestacional, número de mulheres em idade fértil analfabetas, número de residências com mais de seis moradores, renda mensal da mulher, renda mensal do homem e densidade demográfica do setor censitário.

3.6 Inferência bayesiana para modelo linear espacial generalizado

Para rodar a análise, foi assumido que $Y_i|U_i$ tem distribuição binomial($n_i, P_i(s_i)$). Para a verossimilhança, foi adotado o modelo da equação (2.10), portanto, um produto de binomiais. O modelo linear generalizado implica que $L(Y_j | U_j, \beta) = P(y, \mu_j)$ com $\mu_j = h^{-1} [X_j^T \beta + U(s_j)]$. Em que X é a matriz (68x2) da variável independente utilizada, β o vetor (2x1) de parâmetros da regressão. O processo espacial segue uma distribuição normal ($U_j \sim N(0, V(\theta))$). Fazendo $\theta = (\sigma^2, \phi)$, então, $V(\theta) = \sigma^2 R(\phi)$, sendo $R(\phi)$ a matriz ij -ésimo elemento, dado pela correlação $r_{ij} = \exp(-d_{ij}/\phi)$.

Foram definidas as prioris das seguintes formas:

$$P(\beta) \propto 1/\sigma^2; P(\phi) \propto \text{constante}$$

$$(\sigma^2 | \phi) \sim \chi_{\text{Inverso escalado}}^2(n_\sigma = 5, S_\sigma^2 = 3).$$

A distribuição a posteriori, considerando o produto entre a verossimilhança e as prioris, tem a seguinte equação:

$$P(\beta, \sigma^2 | \phi, Y) \propto L(Y | \beta, \phi)P(\beta)P(\sigma^2 | \phi)P(\phi)$$

em que não é conhecida sua distribuição. Portanto, utiliza-se de procedimentos de MCMC para obter amostras da distribuição a posteriori. Para obter as distribuições dos parâmetros e dos valores preditos, utilizou-se o software R (versão 2.5.1)(R Development Core Team, 2007) com o pacote geoRglm (versão 0.8-22) (Christensen & Ribeiro Jr., 2002).

A escolha do valor inicial de ϕ foi determinada por um pré-estudo, usando o semivariograma, pois, a partir deste, optou-se por utilizar o valor 1.500. Para o parâmetro da variância relativa *nugget*, foi assumido o valor nulo, ou seja, sem efeito pepita. No processo iterativo do MCMC, o número de iterações utilizadas foi de 2.000.000, sendo a amostra efetiva coletada após descarte de um *burn-in* de 500.000 e um pulo (*thin*) de 500, ou seja, caracterizando assim uma amostra de 4.000 observações, para cada cadeia de cada parâmetro. Para o estudo da convergência, foram simuladas 2 cadeias ($m = 2$) independentes e em paralelo, para, assim, possibilitar o uso do critério de Gelman & Rubin (1992) (Equação 2.4) e o de Brooks & Gelman (1998) (Equação 2.5). Os critérios de Geweke (1992) (Equação 2.6), Heidelberger & Welch (1983) também foram utilizados para determinar a convergência. A escolha do número de iterações, do *burn-in* e do pulo foi devido a um estudo

preliminar, com o método de Raftery & Lewis (1992), e posteriormente, com os demais critérios, do comportamento das cadeias geradas.

Para a predição, a distribuição condicional completa é dada por uma normal multivariada, dada por:

$$(U^* | Y, \beta, \sigma^2) \sim NM(X^{*T} \beta + r^T R^{-1}(Y - X^T \beta), \sigma^2(1 - r^T R^{-1} r)),$$

sendo U^* um vetor de q valores a serem preditos $U^* = (U_{(s_{n+1})}, \dots, U_{(s_{n+q})})$, e $X^{*T} \beta$ a média dos valores a serem preditos. Em que r é o vetor (68x1) de correlações entre $U_{(s_0)}$ e $U_{(s_i)}$, sendo s_0 uma arbitrária localização a ser predita. A distribuição preditiva $U^* | Y$ é obtida de:

$$P(U^* | Y) = \iint P(U^* | Y, \beta, \sigma^2) P(\beta, \sigma^2 | Y) d\beta d\sigma^2 .$$

4 RESULTADOS E DISCUSSÃO

4.1 Análise exploratória

O conjunto de observações está representado na Figura 7, em que estão apresentadas ambas as representações espaciais dos pontos. Os dados de mortalidade estão caracterizados com um “+” e os de nascimento com um ponto negro.

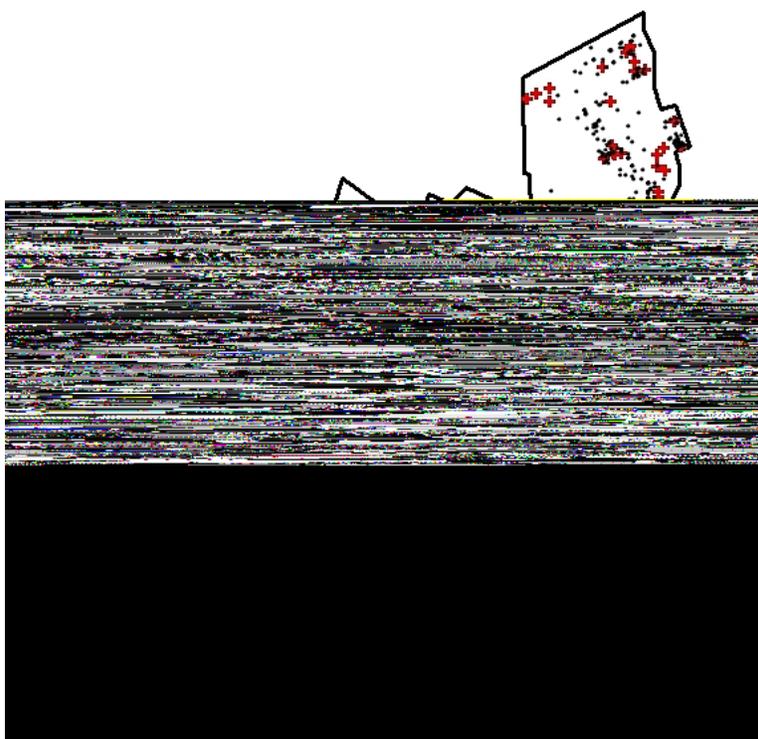


FIGURA 7 Representação gráfica da localização espacial de uma amostra dos dados de morte (+) e de nascimento (•), ocorridos entre 2000 e 2004, na cidade de Alfenas, MG.

Foram obtidas as funções K para efeitos bivariados e para a diferença entre a morte e o nascimento. Os limites superiores e inferiores simulados, para testar a hipótese de independência entre morte e nascimento e a hipótese de que

ambas vieram do mesmo processo estocástico, foram obtidos usando 1.000 simulações e as distâncias computadas em metros. Nas Figuras 8 e 9 estão representados os resultados para a função K bivariada e para a diferença entre as funções, respectivamente.

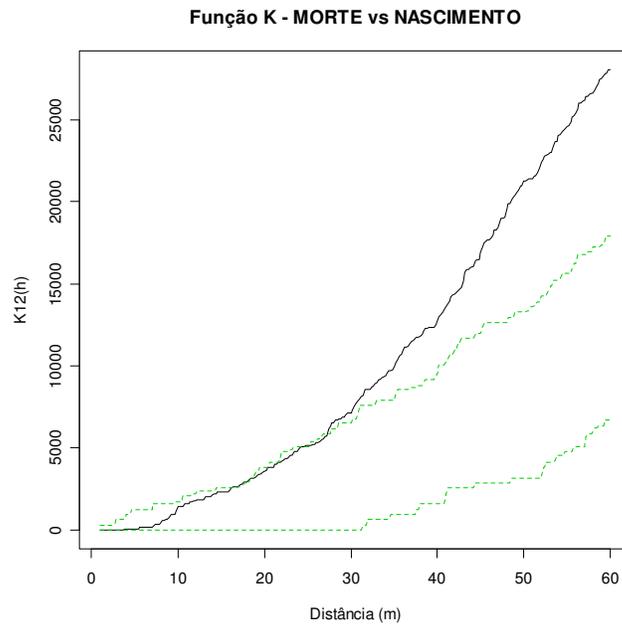


FIGURA 8 Representação gráfica para a função K bivariada, para os dados de mortalidade e de nascimento, estimada a cada distância e os respectivos limites de significância.

Com a função K bivariada (Figura 8) observou-se o interesse em avaliar a interação entre as mortes e os nascimentos. De acordo com o gráfico da Figura 8, pode-se afirmar que, a partir da distância de 27 metros, os dados de mortalidade e de nascimento apresentam a característica de atração, estatisticamente significativa, entre os eventos. Até uma distância de 27 metros, pode-se considerar a independência entre as variáveis. Como os dados foram referenciados pelos lotes, acredita-se que os 27 metros seria uma média mínima entre o centro de um lote e o seu vizinho. Dessa forma, é um resultado esperado,

pois, o local onde ocorrem as mortes está relacionado com a distribuição da população no espaço, que está sendo representada pelos dados de nascimentos.

Para o estudo da diferença entre a estimativa da função K da mortalidade e do nascimento, apresentado na Figura 9, é possível afirmar que a diferença é nula, em todas as distâncias, portanto, não existe diferença entre os processos estocásticos de ambas as variáveis.

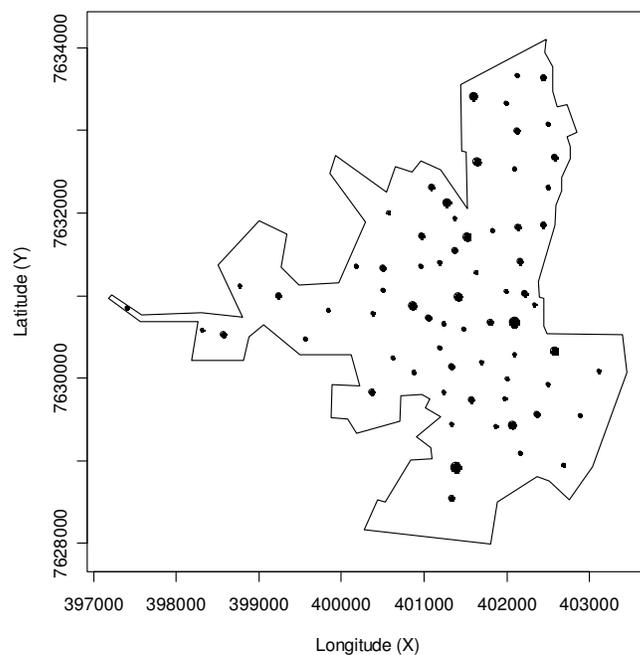
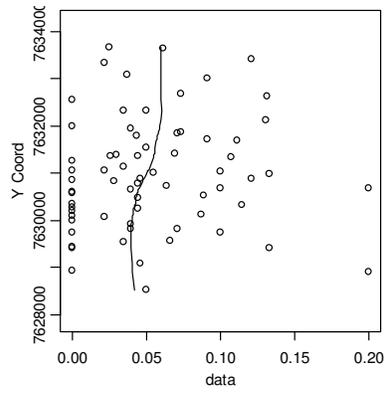
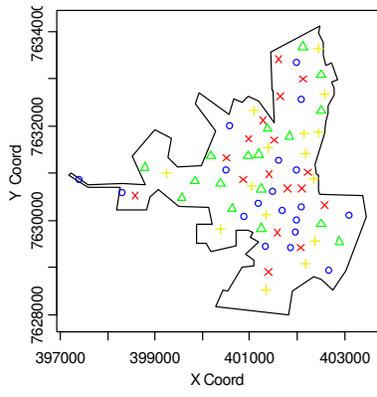


FIGURA 10 Representação pontual proporcional a ocorrência de mortes por setor, entre 2000 e 2004, no município de Alfenas, MG.

Na Figura 10, a localização dos dados é feita pelos círculos e o tamanho



presença significativa de superdispersão, que pode ser explicada por um processo espacial existente.

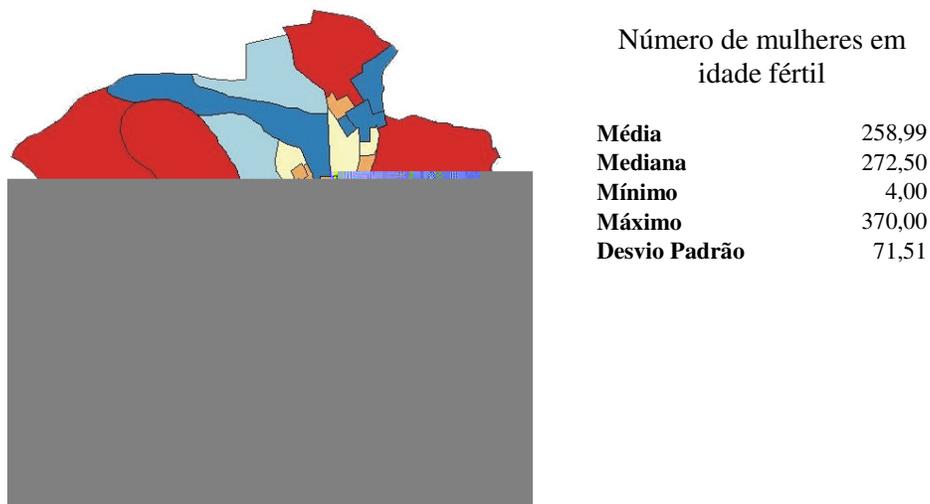
As covariáveis que foram escolhidas no banco de dados do IBGE (2000) para serem investigadas quanto à sua importância neste processo, se encontram na Tabela 1.

TABELA 1 Covariáveis selecionadas por setor censitário. IBGE, censo 2000.

-
1. Número de mulheres em idade fértil
 2. Número de mulheres em idade de risco gestacional
 3. Número de mulheres em idade fértil analfabetas
 4. Número de residências com mais de seis moradores
 5. Renda mensal da mulher
 6. Renda mensal do homem
 7. Densidade demográfica do setor censitário
-

Fonte: *Estatcart* – Base de Informações por Setor Censitário – Alfenas – IBGE 2000.

Nas Figura 12 a 18, estão apresentados a representação gráfica de cada covariável espacialmente distribuída e o resumo da estatística descritiva destas que serão utilizadas na modelagem, seguindo a seqüência apresentada na Tabela 1. Uma relação direta foi verificada com alguns setores censitários de Alfenas e as covariáveis, em que os maiores valores foram encontrados nos bairros de classe econômica desprivilegiada. Apenas as covariáveis rendimento de homens e de mulheres é que apresentaram um comportamento diferente.

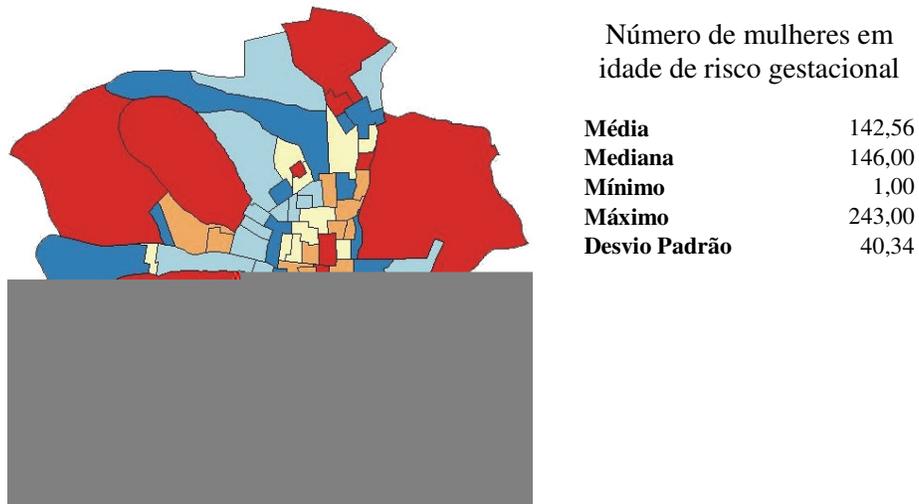


Fonte do Cartograma: *Estatcart*, IBGE 2000.

FIGURA 12 Análise exploratória para o número de mulheres em idade fértil, na cidade de Alfenas, de acordo com os setores censitários.

A variável do número de mulheres em idade fértil (Figura 12) apresenta um comportamento espacial, em que os maiores valores estão na periferia da cidade, ou seja, no centro da cidade os valores são baixos, podendo chegar a 200 mulheres, enquanto que, nos setores da periferia, esses valores podem chegar a mais de 310.

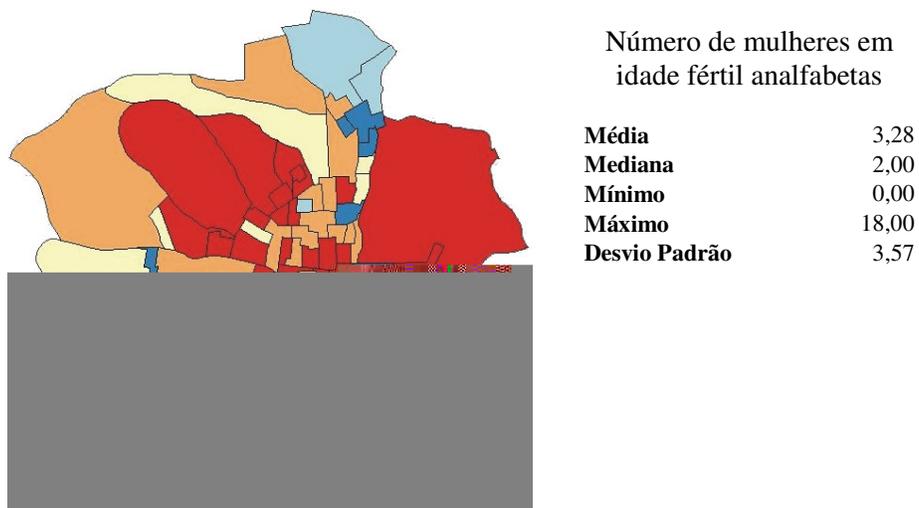
Na Figura 13, está apresentada a variável número de mulheres em idade de risco gestacional, ou seja, mulheres com idade entre 15 e 19 anos e as mulheres entre 35 e 39 anos de idade. Esta variável carrega a informação de um risco biológico. A distribuição espacial destes dados tem um comportamento parecido com a variável, mulheres em idade fértil. Existem setores ricos e setores mais desprivilegiados economicamente com uma contagem muito próxima para esta variável.



Fonte do Cartograma: *Estatcart*, IBGE 2000.

FIGURA 13 Análise exploratória para o número de mulheres em idade de risco gestacional na cidade de Alfenas, de acordo com os setores censitários.

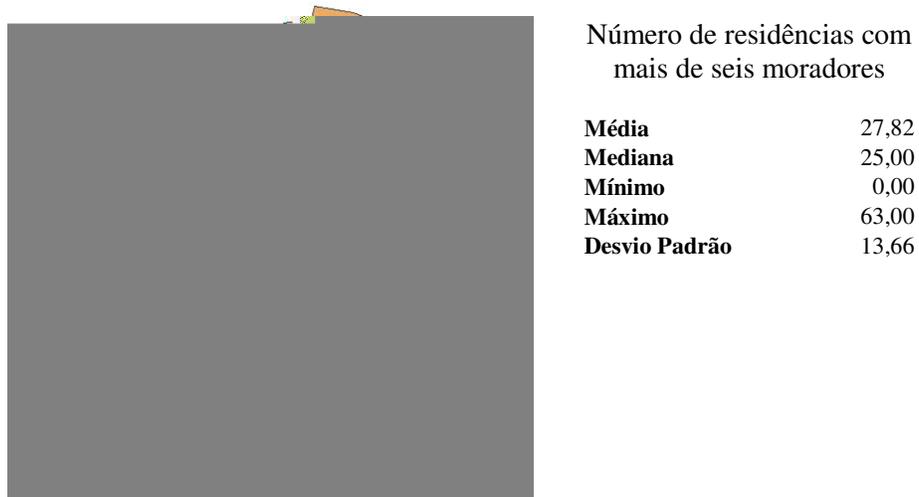
Na Figura 14, a representação gráfica da distribuição espacial do número de mulheres em idade fértil analfabetas é responsável por representar um risco social, com mulheres entre 20 e 34 anos de idade. Esta variável apresenta seus maiores valores também mais presentes na periferia da cidade, mas a diferença entre os setores é pequena. Dos 68 setores, 29 apresentam entre 0 a 2 domicílios com mais de 6 moradores. Percebe-se, no mapa, que os setores que apresentam os maiores valores são periféricos, nos bairros, como Chapada, Vila Santa Edwirges, Campo Elísio, Jardim Eunice, Vista Grande, Jardim Primavera e Jardim São Carlos.



Fonte do Cartograma: *Estatcart*, IBGE 2000.

FIGURA 14 Análise exploratória para o número de mulheres em idade fértil analfabetas na cidade de Alfenas, de acordo com os setores censitários.

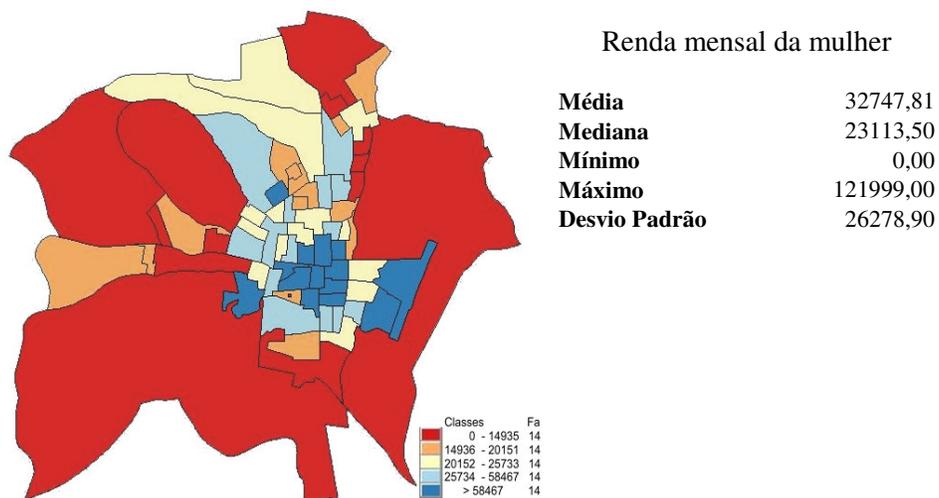
Para a variável número de residências com mais de seis moradores (Figura 15), seu uso representa um risco social, caracterizando a distribuição espacial onde se localizam o maior número de domicílios com maior número de moradores. De maneira geral, o comportamento desta variável também apresenta os seus maiores valores nas periferias, como era esperado.



Fonte do Cartograma: *Estatcart*, IBGE 2000.

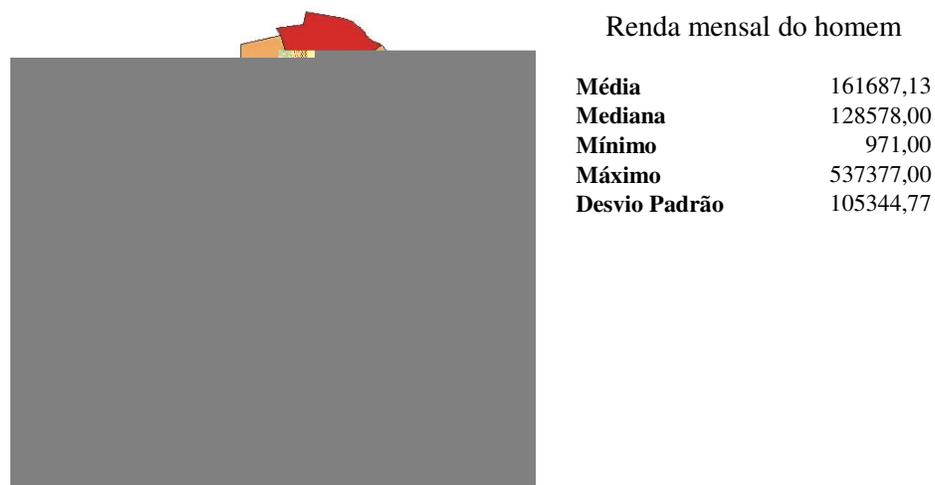
FIGURA 15 Análise exploratória para o número de residências com mais de seis moradores na cidade de Alfenas, de acordo com os setores censitários.

Nas Figuras 16 e 17 está graficamente representada a renda mensal das mulheres e dos homens, respectivamente. Os valores são a soma dos rendimentos mensais do setor censitário. A distribuição espacial da variável social, renda das mulheres, apresenta o comportamento de se concentrar no centro da cidade, onde se acreditam existir as famílias mais tradicionais da cidade e onde residem a grande maioria das viúvas e aposentadas de homens que eram bem inseridos socialmente nos melhores cargos, como, por exemplo, banqueiros, empresários e funcionários públicos e professores universitários. A variável renda mensal dos homens já tem um comportamento mais variado, mas apresenta valores mais baixos na periferia da cidade, como era esperado.



Fonte do Cartograma: *Estatcart*, IBGE 2000.

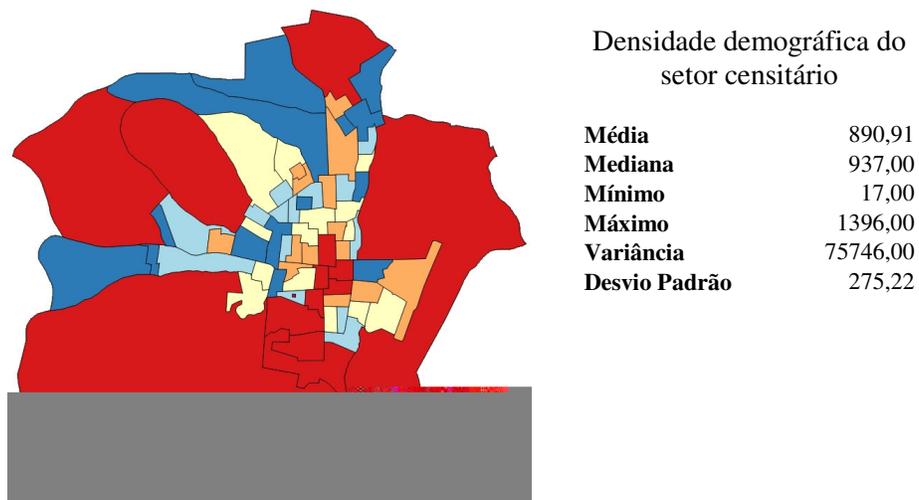
FIGURA 16 Análise exploratória para a renda mensal da mulher, na cidade de Alfenas, de acordo com os setores censitários.



Fonte do Cartograma: *Estatcart*, IBGE 2000.

FIGURA 17 Análise exploratória para a renda mensal do homem, na cidade de Alfenas, de acordo com os setores censitários.

Para representar a última variável a ser utilizada, tem-se a densidade demográfica do setor censitário (Figura 18). Esta também apresenta um comportamento de se concentrar na periferia da cidade, onde as residências ocupam menor espaço físico, devido às características sócio-econômicas.



Fonte do Cartograma: *Estatcart*, IBGE 2000.

FIGURA 18 Análise exploratória para a densidade demográfica na cidade de Alfenas, de acordo com os setores censitários.

Considera-se que, a partir do momento em que se levaram em conta as covariáveis, o efeito de tendência foi praticamente anulado, sendo estas variáveis importantes

Para ajudar na escolha dos valores dos parâmetros ϕ e σ^2 , utilizou-se o semivariograma apresentado na Figura 19.

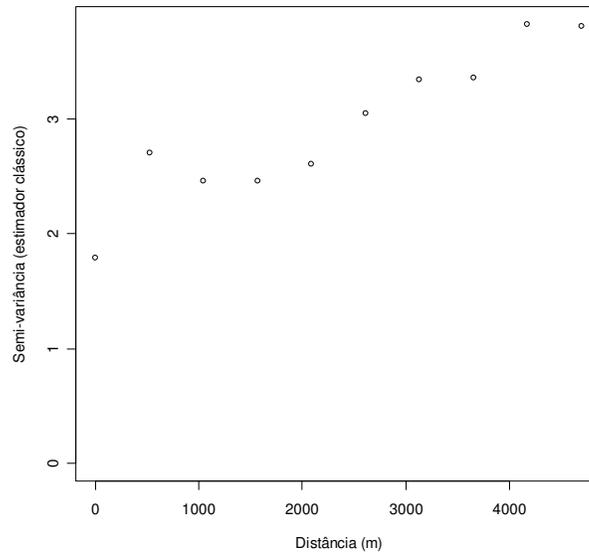


FIGURA 19 Semivariograma para os dados de mortalidade, considerando uma tendência para número de residentes.

Com a ferramenta *Eyefit* presente no software R, pacote *geoR* (Ribeiro Jr. & Diggle, 2001), é possível ter, de forma interativa, uma simulação do ajuste de modelos espaciais, com seus respectivos parâmetros. Portanto, utilizando o semivariograma e a ferramenta *Eyefit*, foi possível a escolha de valores iniciais: 0,05, para o σ^2 e 1.500, para o ϕ , em todos os modelos.

4.2 Ajuste dos modelos lineares espaciais generalizados via inferência bayesiana

A abordagem realizada neste trabalho é de uma modelagem espacial utilizando modelos lineares generalizados mistos via inferência bayesiana, utilizando um processo espacial gaussiano estacionário, assumindo

continuidade, ou seja, uma aplicação geoestatística. Os dados utilizados foram da razão entre o número de mortes sobre o número de nascidos vivos entre 2000 e 2004. Esta razão é chamada de coeficiente de mortalidade infantil. Para o ajuste dos sete modelos, considerou-se o número de nascidos como uma variável *offset*.

Os parâmetros (β_1) dos modelos de regressão logística estão apresentados na Tabela 2, na qual estão também apresentados os intervalos de credibilidade HPD com 95%. Analisando-se os resultados dos parâmetros, pode-se afirmar que os intervalos de credibilidade que contêm o valor 0, apresentam uma probabilidade de a incerteza ser nula, portanto não existindo efeito desta variável. De acordo com os valores apresentados, é possível afirmar que apenas as variáveis número de mulheres em idade fértil e densidade demográfica apresentaram significância estatística. Portanto, as distribuições dos parâmetros β_1 e *odds ratio* dos respectivos modelos significativos estão apresentados nas Figuras 22 e 24.

TABELA 2 Médias e I.C. HPD 95% dos parâmetros β_1 dos modelos de regressão logística de acordo com a covariável utilizada.

Covariáveis	Média (β_1)	I.C. HPD	
		LI 95%	LI 95%
Número de mulheres em idade fértil	0,0046	0,000227	0,00896
Número de mulheres em idade de risco gestacional	0,0056	-0,00108	0,01271
Número de mulheres em idade fértil analfabetas	0,0473	-0,01278	0,12024
Número de residências com mais de 6 moradores	0,0184	-0,00308	0,03867
Renda mensal da mulher	-0,000003	-0,00001	0,000008
Renda mensal do homem	-0,000001	-0,000004	0,000001
Densidade demográfica do setor censitário	0,0013	0,000108	0,002522

Um dos resultados obtidos para os modelos significativos foi a predição do risco em localizações arbitrárias, dentro da área em estudo. Para o modelo considerando o número de mulheres em idade fértil como variável independente, tem-se o gráfico do risco predito para a região em estudo (Figura 20).

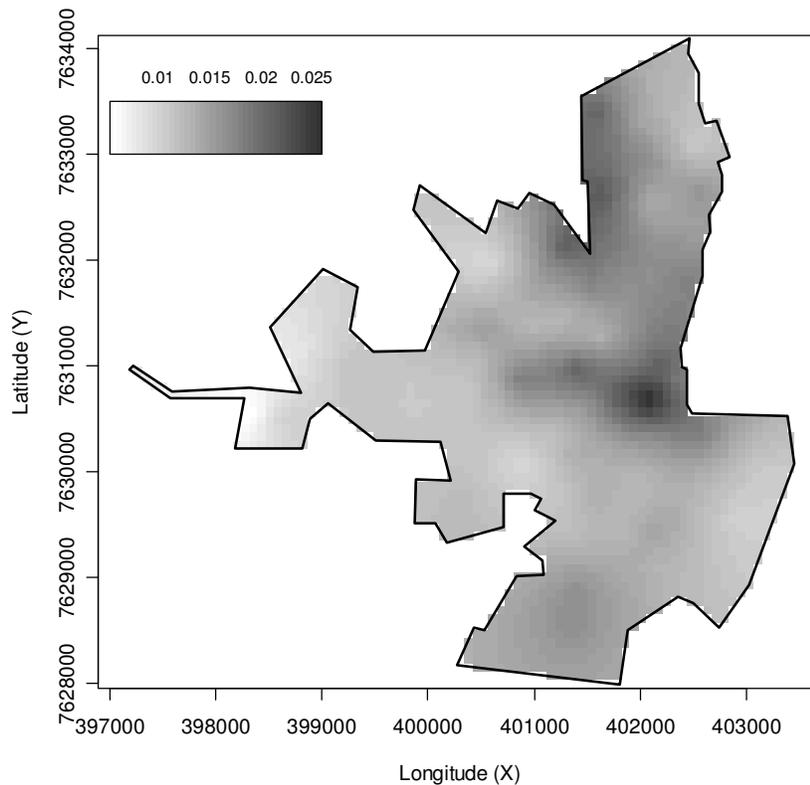


FIGURA 20 Representação gráfica dos valores preditos para o risco, dados pelo modelo considerando o número de mulheres em idade fértil, no município de Alfenas, MG.

O modelo ajustado permitiu realizar a predição para 2.342 pontos na área urbana de Alfenas. Na Figura 21, se tem a representação gráfica destes mesmos valores preditos, mas considerando a localização dos 68 pontos amostrais. Este tipo de gráfico possibilita a visualização de regiões com maior

probabilidade de morte infantil. Estes valores permitem referenciar os resultados da Tabela 3.

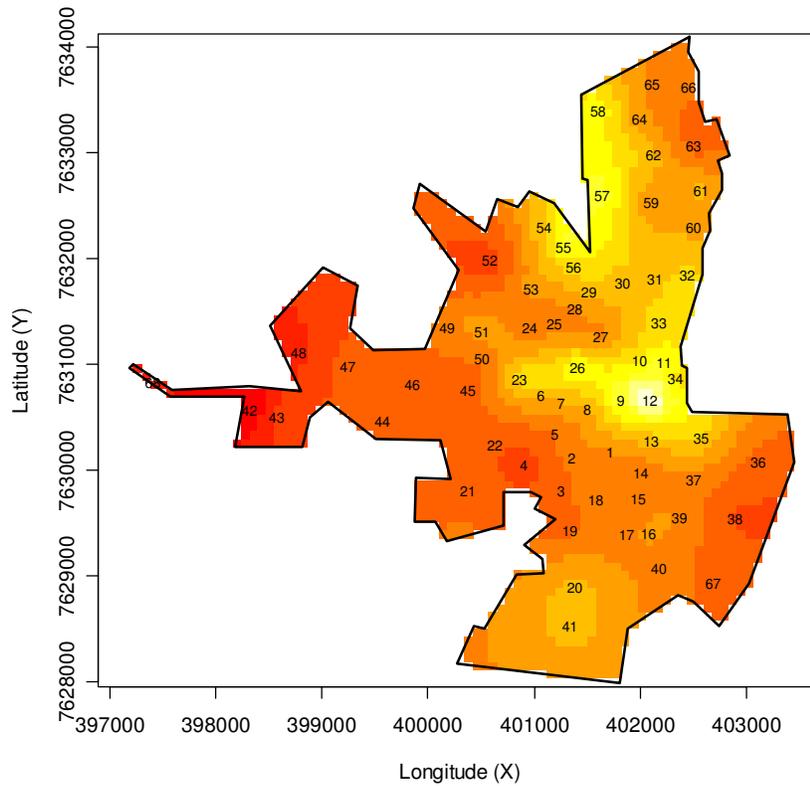


FIGURA 21 Representação gráfica dos pontos amostrais e valores preditos, considerando a covariável número de mulheres em idade fértil.

Segundo os valores preditos para a probabilidade de morte infantil, percebe-se que valores mais altos para a probabilidade estão associados a bairros como Vista Grande, Jardim São Carlos, Jardim América, Chapada e Jardim Panorama. Estes bairros são regiões de características sócio-econômicas desprivilegiadas, com relação à infra-estrutura e, como tais, são bairros das periferias.

Na Tabela 3 estão dispostos os 68 parâmetros preditos e os parâmetros do modelo de regressão logística. Cada parâmetro teve sua distribuição amostrada, portanto, as estatísticas descritivas de cada parâmetro também estão apresentadas. Também na Tabela 3 constam os intervalos de credibilidade de maior densidade de probabilidade (da sigla inglesa HPD) construídos com 95% .

TABELA 3 Resumo das estatísticas descritivas dos parâmetros, para o modelo que considera a covariável número de residentes.

Parâmetros	Média	Desvio padrão	Mediana	I.C. HPD	
				LI 95%	LS 95%
σ^2	0,7871	0,2661	0,7363	0,3574	1,3077
β_0	-4,4305	0,8228	-4,4371	-5,9989	-2,8004
β_1	0,0046	0,0023	0,0046	0,0002	0,0090
U[1]	0,0288	0,0160	0,0254	0,0049	0,0604
U[2]	0,0441	0,0188	0,0410	0,0130	0,0816
U[3]	0,0470	0,0186	0,0441	0,0152	0,0842
U[4]	0,0463	0,0185	0,0436	0,0153	0,0842
U[5]	0,0341	0,0145	0,0315	0,0101	0,0629
U[6]	0,0575	0,0186	0,0550	0,0251	0,0959
U[7]	0,0500	0,0194	0,0469	0,0170	0,0888
U[8]	0,0393	0,0180	0,0361	0,0115	0,0763
U[9]	0,0488	0,0245	0,0439	0,0110	0,0965
U[10]	0,0508	0,0220	0,0470	0,0150	0,0947
U[11]	0,0681	0,0251	0,0644	0,0262	0,1175
U[12]	0,0823	0,0338	0,0760	0,0262	0,1487
U[13]	0,0403	0,0201	0,0362	0,0093	0,0798
U[14]	0,0353	0,0177	0,0322	0,0069	0,0704
U[15]	0,0371	0,0184	0,0338	0,0081	0,0733
U[16]	0,0715	0,0315	0,0667	0,0212	0,1336
U[17]	0,0424	0,0202	0,0387	0,0117	0,0832
U[18]	0,0358	0,0187	0,0318	0,0082	0,0741
U[19]	0,0335	0,0181	0,0296	0,0061	0,0685
U[20]	0,0402	0,0266	0,0335	0,0051	0,0928
U[21]	0,0544	0,0257	0,0496	0,0111	0,1042

...continua...

TABELA 3 Continuação....

Parâmetros	Média	Desvio padrão	Mediana	I.C. HPD	
				LI 95%	LS 95%
U[22]	0,0447	0,0184	0,0416	0,0152	0,0812
U[23]	0,0911	0,0283	0,0878	0,0404	0,1475
U[24]	0,0508	0,0190	0,0483	0,0173	0,0878
U[25]	0,0520	0,0189	0,0496	0,0199	0,0900
U[26]	0,0669	0,0252	0,0628	0,0259	0,1184
U[27]	0,0463	0,0181	0,0438	0,0134	0,0805
U[28]	0,0576	0,0187	0,0554	0,0263	0,0971
U[29]	0,0671	0,0253	0,0630	0,0260	0,1201
U[30]	0,0486	0,0212	0,0451	0,0131	0,0900
U[31]	0,0782	0,0295	0,0738	0,0271	0,1347
U[32]	0,0658	0,0223	0,0629	0,0267	0,1098
U[33]	0,0650	0,0240	0,0614	0,0263	0,1152
U[34]	0,0738	0,0259	0,0705	0,0295	0,1271
U[35]	0,0879	0,0335	0,0836	0,0309	0,1552
U[36]	0,0396	0,0225	0,0347	0,0063	0,0849
U[37]	0,0462	0,0180	0,0434	0,0145	0,0811
U[38]	0,0455	0,0193	0,0423	0,0113	0,0820
U[39]	0,0503	0,0208	0,0473	0,0173	0,0939
U[40]	0,0511	0,0226	0,0471	0,0136	0,0968
U[41]	0,0361	0,0244	0,0300	0,0046	0,0866
U[42]	0,0327	0,0145	0,0302	0,0088	0,0605
U[43]	0,0430	0,0189	0,0398	0,0132	0,0818
U[44]	0,0457	0,0201	0,0421	0,0145	0,0857
U[45]	0,0562	0,0219	0,0528	0,0186	0,0996
U[46]	0,0336	0,0161	0,0308	0,0091	0,0648
U[47]	0,0452	0,0175	0,0428	0,0154	0,0795
U[48]	0,0451	0,0194	0,0416	0,0123	0,0831
U[49]	0,0343	0,0164	0,0312	0,0081	0,0669
U[50]	0,0467	0,0172	0,0443	0,0175	0,0805
U[51]	0,0622	0,0243	0,0584	0,0200	0,1096
U[52]	0,0392	0,0181	0,0361	0,0092	0,0747
U[53]	0,0733	0,0277	0,0693	0,0291	0,1322
U[54]	0,0605	0,0240	0,0565	0,0200	0,1100
U[55]	0,0594	0,0252	0,0551	0,0176	0,1085
U[56]	0,0537	0,0221	0,0498	0,0180	0,0965

...continua...

TABELA 3 Continuação....

Segundo o intervalo de credibilidade (HPD) para o parâmetro *odds ratio* (I.C. HPD_{95%}: [1,000227; 1,008998]), pode-se afirmar que existe relação entre o número de mulheres em idade fértil e a mortalidade infantil na cidade de Alfenas, MG e que esta relação é do tipo causal. Quanto maior o número de mulheres em idade fértil, maior a chance da ocorrência de mortalidade. Este fato permite afirmar que a exposição à regiões com mais mães em idade fértil pode causar maior coeficiente de mortalidade. Esta covariável coincide com regiões de periferia.

A convergência para este modelo foi verificada e os resultados para o teste de Gelman & Rubin (1992) e Geweke (1992) estão apresentados na Tabela 4. O teste multivariado de Brooks & Gelman (1998) apresentou como resultado a estimativa, $R^{71} = 1,0179$, que caracteriza a convergência. Os resultados para os testes de Raftery & Lewis (1992) e Heidelberger & Welch (1983) foram satisfatórios e ambos caracterizaram como obtida a convergência.

TABELA 4 Resultados dos testes de verificação da convergência para o modelo, considerando o número de mulheres em idade fértil.

Parâmetro	G&R	Geweke	Parâmetro	G&R	Geweke
	R	p-valor		R	p-valor
σ^2	1,0016	0,1280	U[10]	0,9999	0,6075
β_0	0,9998	0,4268	U[11]	0,9999	0,0836
β_1	0,9999	0,5994	U[12]	1,0001	0,2085
U[1]	1,0005	0,1704	U[13]	1,0013	0,7224
U[2]	1,0031	0,0059	U[14]	1,0009	0,4359
U[3]	1,0036	0,0976	U[15]	1,0013	0,4899
U[4]	0,9998	0,9771	U[16]	1,0025	0,9323
U[5]	1,0001	0,6581	U[17]	1,0010	0,9458
U[6]	0,9998	0,7563	U[18]	1,0003	0,6167
U[7]	1,0009	0,5013	U[19]	1,0015	0,7653
U[8]	1,0004	0,1932	U[20]	1,0041	0,1717
U[9]	0,9998	0,9275	U[21]	1,0011	0,4986

...continua...

TABELA 4 Continuação...

Parâmetro	G&R	Geweke	Parâmetro	G&R	Geweke
	R	p-valor		R	p-valor
U[22]	0,9999	0,3981	U[46]	1,0005	0,3112
U[23]	1,0007	0,2685	U[47]	0,9999	0,5695
U[24]	1,0017	0,9819	U[48]	1,0000	0,3862
U[25]	1,0008	0,4366	U[49]	1,0006	0,5381
U[26]	1,0017	0,4747	U[50]	1,0009	0,8378
U[27]	1,0005	0,5733	U[51]	0,9999	0,7453
U[28]	1,0008	0,7882	U[52]	1,0015	0,8938
U[29]	1,0023	0,3349	U[53]	0,9999	0,8241
U[30]	0,9999	0,5860	U[54]	0,9999	0,2630
U[31]	1,0000	0,9311	U[55]	1,0007	0,3160
U[32]	0,9999	0,7534	U[56]	1,0012	0,6915
U[33]	1,0007	0,4673	U[57]	0,9998	0,8539
U[34]	0,9999	0,1311	U[58]	0,9998	0,3943
U[35]	0,9998	0,2383	U[59]	1,0015	0,1551
U[36]	1,0010	0,8756	U[60]	0,9999	0,2845
U[37]	1,0012	0,1229	U[61]	1,0000	0,6359
U[38]	1,0001	0,6698	U[62]	1,0004	0,0842
U[39]	1,0012	0,3192	U[63]	0,9998	0,3453
U[40]	1,0025	0,0679	U[64]	1,0007	0,0376
U[41]	1,0033	0,5596	U[65]	1,0006	0,6770
U[42]	0,9999	0,4002	U[66]	0,9999	0,5669
U[43]	1,0005	0,4984	U[67]	1,0004	0,9115
U[44]	1,0009	0,9973	U[68]	0,9999	0,1842
U[45]	0,9999	0,7732			

Os resultados da Tabela 4 mostraram que a convergência pelo teste de Gelman & Rubin foi confirmada, pois todos os parâmetros apresentaram o Fator R próximo de 1. Pelo o teste de Geweke, a convergência foi caracterizada em praticamente todos os parâmetros; apenas 2 deles apresentaram significância, a 5%, para rejeição da estabilidade, mas esta quantidade de parâmetros que apresentou significância está dentro da quantidade aceita como erro tipo I, ou seja, pode-se afirmar que a convergência também foi verificada pelo teste de Geweke.

Este modelo permitiu identificar a relação do número de mulheres em idade fértil, como um fator de risco para a mortalidade infantil no município de Alfenas. Autores como Behm (1980), Monteiro (1982), Paim et al. (1987) e Yunes (1983) verificaram que é comum o fato de fatores sociais e econômicos estarem associados à mortalidade infantil. Este modelo, considerando a correlação espacial, possibilitou a verificação de tal associação, como um fator de risco. Para o modelo considerando a densidade demográfica do setor censitário como variável independente no modelo logístico, tem-se o gráfico do risco predito para a região em estudo (Figura 23).

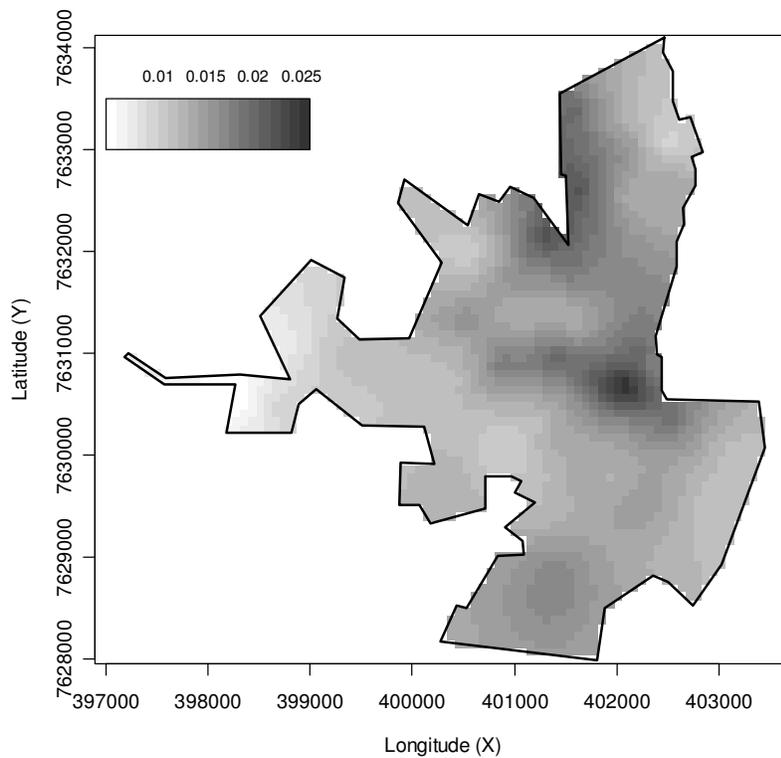


FIGURA 23 Representação gráfica dos valores preditos, considerando a covariável densidade demográfica, para o risco de morte no município de Alfenas, MG.

Realizando a análise com o modelo com a covariável densidade demográfica, foi possível realizar a predição para 2.342 pontos na área urbana de Alfenas. Na Figura 23, observa-se a representação gráfica desses valores preditos por este segundo modelo. Este gráfico caracteriza, com maior probabilidade de morte, as regiões da cidade de população de renda mais baixa. Comparando-se os resultados dos dois modelos, observa-se que ambos apresentaram resultados muito parecidos para os riscos e com praticamente as mesmas regiões críticas. Diferenças maiores foram encontradas nos bairros Vila Formosa e Jardim Tropical.

Na Tabela 5 estão apresentadas as estatísticas descritivas das distribuições dos parâmetros que compõem o modelo e os valores preditos nos mesmos pontos amostrais representados pela Figura 21. Estão presentes também, na Tabela 5, os intervalos de credibilidades de maior densidade de probabilidade (HPD) construídos com 95%.

TABELA 5 Resumo das estatísticas descritivas dos parâmetros, para o modelo considerando a covariável densidade demográfica.

Parâmetros	Média	Desvio padrão	Mediana	I.C. HPD	
				LI 95%	LS 95%
σ^2	0,7934	0,2690	0,7434	0,3489	1,3009
β_0	-4,4568	0,8080	-4,4626	-5,9823	-2,8552
β_1	0,0013	0,0006	0,0013	0,0001	0,0025
U[1]	0,0308	0,0164	0,0274	0,0065	0,0632
U[2]	0,0476	0,0201	0,0442	0,0149	0,0886
U[3]	0,0463	0,0183	0,0432	0,0145	0,0822
U[4]	0,0463	0,0185	0,0438	0,0143	0,0827
U[5]	0,0351	0,0147	0,0325	0,0111	0,0652
U[6]	0,0606	0,0194	0,0581	0,0271	0,1002
U[7]	0,0477	0,0188	0,0447	0,0157	0,0847
U[8]	0,0402	0,0182	0,0370	0,0118	0,0771
U[9]	0,0485	0,0243	0,0439	0,0134	0,0974

...continua...

TABELA 5 Continuação...

Parâmetros	Média	Desvio padrão	Mediana	I.C. HPD	
				LI 95%	LS 95%
U[10]	0,0489	0,0215	0,0452	0,0139	0,0910
U[11]	0,0722	0,0261	0,0685	0,0278	0,1240
U[12]	0,0795	0,0331	0,0733	0,0236	0,1438
U[13]	0,0399	0,0199	0,0359	0,0086	0,0779
U[14]	0,0342	0,0173	0,0312	0,0074	0,0693
U[15]	0,0405	0,0197	0,0370	0,0093	0,0788
U[16]	0,0666	0,0292	0,0620	0,0204	0,1249
U[17]	0,0444	0,0210	0,0405	0,0111	0,0857
U[18]	0,0370	0,0191	0,0328	0,0088	0,0761
U[19]	0,0315	0,0175	0,0278	0,0047	0,0642
U[20]	0,0383	0,0257	0,0316	0,0050	0,0892
U[21]	0,0496	0,0238	0,0452	0,0128	0,0983
U[22]	0,0440	0,0183	0,0409	0,0142	0,0793
U[23]	0,0913	0,0283	0,0882	0,0435	0,1511
U[24]	0,0515	0,0192	0,0489	0,0173	0,0887
U[25]	0,0499	0,0183	0,0475	0,0192	0,0867
U[26]	0,0693	0,0258	0,0652	0,0273	0,1224
U[27]	0,0435	0,0173	0,0409	0,0115	0,0752
U[28]	0,0625	0,0203	0,0605	0,0260	0,1024
U[29]	0,0668	0,0253	0,0627	0,0240	0,1176
U[30]	0,0479	0,0211	0,0444	0,0138	0,0901
U[31]	0,0693	0,0263	0,0652	0,0256	0,1219
U[32]	0,0697	0,0233	0,0667	0,0287	0,1163
U[33]	0,0648	0,0240	0,0611	0,0225	0,1115
U[34]	0,0753	0,0264	0,0721	0,0316	0,1310
U[35]	0,0883	0,0336	0,0841	0,0310	0,1560
U[36]	0,0365	0,0214	0,0317	0,0045	0,0786
U[37]	0,0449	0,0177	0,0420	0,0137	0,0790
U[38]	0,0422	0,0183	0,0390	0,0111	0,0779
U[39]	0,0496	0,0207	0,0467	0,0153	0,0916
U[40]	0,0540	0,0237	0,0499	0,0137	0,1010
U[41]	0,0353	0,0240	0,0295	0,0033	0,0821
U[42]	0,0332	0,0147	0,0308	0,0093	0,0621
U[43]	0,0456	0,0197	0,0422	0,0127	0,0845
U[44]	0,0472	0,0207	0,0437	0,0147	0,0884
U[45]	0,0563	0,0219	0,0530	0,0190	0,0990

...continua...

TABELA 5 Continuação...

Parâmetros	Média	Desvio padrão	Mediana	I.C. HPD	
				LI 95%	LS 95%
U[46]	0,0312	0,0154	0,0284	0,0076	0,0611
U[47]	0,0445	0,0173	0,0420	0,0148	0,0780
U[48]	0,0464	0,0199	0,0430	0,0133	0,0863
U[49]	0,0350	0,0164	0,0320	0,0081	0,0666
U[50]	0,0443	0,0166	0,0420	0,0160	0,0769
U[51]	0,0628	0,0245	0,0589	0,0203	0,1104
U[52]	0,0380	0,0178	0,0350	0,0089	0,0725
U[53]	0,0641	0,0245	0,0601	0,0252	0,1157
U[54]	0,0608	0,0243	0,0568	0,0196	0,1105
U[55]	0,0583	0,0250	0,0540	0,0187	0,1078
U[56]	0,0534	0,0221	0,0497	0,0181	0,0971
U[57]	0,1055	0,0369	0,1008	0,0428	0,1781
U[58]	0,0876	0,0356	0,0826	0,0299	0,1584
U[59]	0,0444	0,0197	0,0407	0,0124	0,0843
U[60]	0,0428	0,0168	0,0402	0,0137	0,0750
U[61]	0,0594	0,0202	0,0569	0,0235	0,0992
U[62]	0,0713	0,0235	0,0686	0,0313	0,1195
U[63]	0,0584	0,0198	0,0561	0,0227	0,0973
U[64]	0,0394	0,0171	0,0364	0,0122	0,0739
U[65]	0,0313	0,0158	0,0282	0,0065	0,0628
U[66]	0,0592	0,0195	0,0566	0,0244	0,0974
U[67]	0,0204	0,0173	0,0155	0,0009	0,0542
U[68]	0,0123	0,0158	0,0076	0,0002	0,0397

Todos os parâmetros apresentaram estimativas para todos os *lags* da autocorrelação inferiores a 0,1, caracterizando, assim, amostras independentes.

Para a segunda variável significativa, a distribuição dos parâmetros β_1 e *odds ratio* está apresentada na Figura 23.

TABELA 6 Resultados dos testes de verificação da convergência, para o modelo considerando a covariável densidade demográfica.

Parâmetro	G&R R	Geweke p-valor	Parâmetro	G&R R	Geweke p-valor
σ^2	1,0014	0,1679	U[34]	1,0004	0,1121
β_0	1,0002	0,4716	U[35]	1,0013	0,2229
β_1	1,0002	0,6985	U[36]	0,9999	0,8447
U[1]	1,0000	0,1802	U[37]	1,0008	0,1453
U[2]	1,0016	0,0079	U[38]	0,9998	0,6918
U[3]	1,0018	0,0830	U[39]	1,0000	0,3370
U[4]	1,0001	0,8879	U[40]	1,0000	0,0566
U[5]	1,0000	0,6099	U[41]	1,0005	0,5080
U[6]	0,9999	0,8197	U[42]	1,0011	0,4768
U[7]	1,0001	0,6452	U[43]	1,0003	0,5448
U[8]	1,0006	0,1914	U[44]	0,9998	0,9595
U[9]	1,0005	0,8813	U[45]	1,0005	0,8217
U[10]	1,0007	0,6112	U[46]	1,0000	0,3202
U[11]	1,0003	0,0914	U[47]	0,9998	0,5544
U[12]	1,0005	0,2254	U[48]	1,0010	0,2990
U[13]	1,0004	0,7228	U[49]	1,0008	0,5188
U[14]	0,9999	0,5681	U[50]	1,0011	0,9727
U[15]	1,0000	0,6525	U[51]	1,0000	0,7154
U[16]	0,9999	0,8906	U[52]	1,0034	0,9544
U[17]	0,9999	0,8847	U[53]	1,0005	0,7298
U[18]	1,0003	0,7898	U[54]	1,0016	0,2408
U[19]	0,9999	0,7142	U[55]	1,0007	0,3078
U[20]	1,0016	0,1642	U[56]	0,9998	0,7398
U[21]	1,0010	0,5334	U[57]	1,0000	0,8962
U[22]	1,0004	0,3239	U[58]	1,0001	0,4135
U[23]	1,0006	0,3356	U[59]	1,0011	0,1993
U[24]	0,9999	0,8744	U[60]	1,0005	0,2965
U[25]	1,0014	0,2720	U[61]	1,0004	0,7265
U[26]	1,0008	0,5911	U[62]	1,0001	0,0735
U[27]	0,9998	0,4879	U[63]	0,9999	0,2728
U[28]	1,0010	0,9136	U[64]	1,0001	0,0615
U[29]	0,9998	0,3246	U[65]	1,0001	0,6403
U[30]	1,0004	0,5826	U[66]	1,0000	0,5538
U[31]	1,0000	0,9165	U[67]	0,9999	0,8864
U[32]	1,0006	0,7405	U[68]	1,0003	0,2528
U[33]	0,9998	0,3463			

Os resultados apresentados na Tabela 6 possibilitaram caracterizar a convergência pelo teste de Gelman & Rubin, pois, todos os parâmetros apresentaram os fatores R próximos de 1. Para o teste de Geweke, como ocorrido com o modelo anterior, a convergência foi caracterizada; apenas uma cadeia apresentou significância, a 5%, para se rejeitar a estabilidade. Porém, essa quantidade está dentro do aceitável como erro tipo I, ou seja, pode-se afirmar que a convergência foi verificada.

A confirmação da relação entre o coeficiente de mortalidade e a densidade demográfica e também o conhecimento sobre a população urbana do município de Alfenas permitem afirmar que, juntamente com a variável densidade, existem outros fatores, pois esta variável caracteriza muito bem as regiões críticas da cidade por meio de seus valores, coincidindo estes com as periferias. Estas periferias apresentam uma série de problemas sociais e econômicos.

Costa et al. (2001) e Sastry (1996) encontraram relação de atributos medidos na mãe, mas, principalmente, ligados à escolaridade. A covariável utilizada que poderia representar melhor a escolaridade foi a de mulheres não alfabetizadas, mas tal covariável não apresentou significância.

As predições dos riscos de cada localidade podem ser comparadas verificando-se o *odds ratio* (OR) entre estas localidades. Separaram-se, para isso, três localidades consideradas regiões mais desprivilegiadas economicamente e três localidades de maior classe social. Para se calcular o OR, realizou-se uma comparação duas a duas. Os dados da Tabela 7 apresentam os valores de OR para as comparações, segundo os dois modelos significativos. Pode-se afirmar que os resultados dos modelos são muito parecidos e, como era de se esperar, os riscos foram realmente maiores nas regiões desprivilegiadas. Em alguns casos, o OR pode chegar a, aproximadamente, três vezes maior, ou

seja, no bairro de Vista Grande, o risco é quase três vezes maior de mortalidade infantil do que no centro.

TABELA 7 Comparação entre os *odds ratio* para 6 bairros no município de Alfenas, MG, segundo os dois modelos significativos.

Bairros (Csi vs Css ¹)	Número de mulheres em idade fértil	Densidade demográfica do setor censitário
	<i>odds ratio</i>	<i>odds ratio</i>
Vista Grande vs. Jd. Aeroporto	1,89	2,08
Vista Grande vs. Centro	2,98	2,84
Vista Grande vs. Jd. Colina	1,58	1,77
Vila Betânia vs. Jd. Aeroporto	1,47	1,58
Vila Betânia vs. Centro	2,33	2,17
Vila Betânia vs. Jd. Colina	1,23	1,35
Chapada vs. Jd. Aeroporto	1,81	1,88
Chapada vs. Centro	2,86	2,58
Chapada vs. Jd. Colina	1,51	1,60

¹ Csi – Classe sócio-econômica inferior; Css – Classe sócio-econômica superior.

Os bairros e as referidas localizações espaciais estão apresentados na Figura 1.

O OR para as comparações é muito parecido nos dois modelos, por isso, acredita-se que exista ainda algum fator social que esteja oculto nestas variáveis e que poderia explicar melhor o processo.

4.3 Considerações finais

A ausência de significâncias em algumas das covariáveis pode estar relacionada à existência de um efeito que os modelos não conseguiram explicar e a necessidade de utilização de outras covariáveis será de grande importância para melhorar os resultados e obter novas hipóteses.

A aplicação da metodologia apresentada por Diggle & Ribeiro Jr. (2007), de geoestatística baseada em modelos, foi muito eficiente no ajuste de

dados com distribuição binomial usando função ligação logística e a presença de covariáveis, sendo interessante para a predição de forma contínua no território em estudo.

O uso da distribuição de Poisson pode ser utilizada como uma distribuição para os dados, sendo uma outra possibilidade de análise. Este novo modelo assumiria que o número de nascimentos em cada ponto (centróide do setor) é desconhecido, portanto a análise seria da contagem de casos.

Uma opção que seria interessante é levar em consideração, na análise, o uso da localização dos Postos de Saúde da Família (PSF), pois estes atuam diretamente na população de maneira preventiva, podendo explicar algumas variações espaciais.

Em vez da utilização de dados agrupados em 5 anos, como ocorrido neste estudo, seria interessante o uso da presença do efeito do tempo, ou seja, uma análise espaço-tempo, que pode ser realizada na forma de geoestatística e, até mesmo, com a estatística espacial por área (setores censitários).

Existe a necessidade de estudar um banco de dados mais atualizado, pois os resultados mostraram que o bairro Chapada apresenta grande risco para a mortalidade infantil, em comparação com os demais. Sabe-se, hoje, que, nesta localidade, a partir de 2004, foi implantado um PSF mais eficiente e espera-se que os resultados tenham melhorado. Portanto, um conjunto atualizado poderia verificar a veracidade desta hipótese.

5 CONCLUSÕES

- O uso da análise de padrão pontual com a função K de forma bivariada confirmou uma atração entre as variáveis. O uso desta função pela diferença confirmou que a mortalidade infantil e o nascimento provêm de mesmo processo estocástico.
- O uso da modelagem usando geoestatística baseada em modelos possibilitou o ajuste de modelos logístico com distribuição binomial; tanto a inferência quanto a previsão foram feitas com precisão.
- O ajuste do modelo que considera as covariáveis número de mulheres em idade fértil e densidade demográfica do setor censitário caracterizou uma relação de causa destas com a variável mortalidade infantil, ou seja, maior número de mulheres em idade fértil e maior densidade demográfica causam maiores riscos de morte, permitindo também a construção das distribuições dos respectivos parâmetros de *odds ratio*.
- Os resultados mostraram que o risco de morte é influenciado pela localização espacial na região urbana do município de Alfenas, MG. Dessa forma, este tipo de modelagem pode ajudar na gestão destas localidades da cidade, permitindo avaliar quais fatores são preponderantes para prevenir a situação e, assim, melhorar o coeficiente municipal da mortalidade infantil

REFERÊNCIAS BIBLIOGRÁFICAS

- ANDES, N.; DAVIS, J. E. Linking public health data using geographic information system techniques: Alaskan community characteristics and infant mortality. **Statistics in Medicine**, v.14, p.481-490, 1995.
- ANDRADE, A. L. T.; SZWARCOWALD, C. L. Análise espacial da mortalidade neonatal. **Caderno de Saúde Pública**, Rio de Janeiro, v.17, n.5, p.1199-1210, set./out. 2001.
- BAILEY, T. Spatial statistics methods in health. **Caderno de Saúde Pública**, v.17, p.1083-1098, 2001.
- BAILEY, T.C.; GATRELL, A.C. **Interactive spatial data analysis**. London: Longman Scientific and Technical, 1995. 413 p.
- BANERJEE, S.; CARLIN, B. P.; GELFAND, A. E. **Hierarchical Modeling and Analysis for Spatial Data**. Boca Raton, FL: Chapman & Hall/CRC, 2004.
- VICTORA, C. G.; BARROS, F. C.; TOMASI, E.; MENEZES, A. M.; HORTA, B. L.; WEIDERPASS, E.; CESAR, A. J.; COSTA, J. S. D.; OLINTO, M. T.; HALPERN, R.; GARCIA, M. Del M.; VAUGHAN, J. P. Tendências e diferenciais na saúde materno-infantil: delineamento e metodologia das coortes de 1982 e 1993 de mães e crianças de Pelotas, RS. **Cadernos de Saúde Pública**, v.12, n.1, p.7-14, 1996
- BEHM, H. Determinantes económicas y sociales de la mortalidad en América Latina. **Revista Cubana de Salud**, v.6, p.1-30, 1980.
- BITHELL, J. F. Na application of density estimation to geographical epidemiology. **Statistics in Medicine**, v.9, p.691-701, 1990.
- BRASIL. Ministério da Saúde. **Manual dos comitês de prevenção de óbito infantil e fetal**. Brasília, 2004. (Normas e Manuais Técnicos).
- BRESLOW, N. E.; CLAYTON, D. G. Approximate inference in generalized linear mixed models. **Journal of the American Statistical Association**, v.88, p.9-25, 1993.
- BROEMELING, L. D. **Bayesian analysis of linear models**. New York: M. Dekker, 1985. 454 p.

- BROOKS, S. P.; DELLAPORTAS, P.; ROBERTS, G. O. An approach to diagnosing total variation convergence of MCMC algorithms. **Journal of Computational and Graphical Statistics**, Alexandria, v. 6, n. 3, p. 251-265, Sept. 1997.
- BROOKS, S. P.; GELMAN, A.; General methods for monitoring convergence of iterative simulations. **Journal of Computational and Graphical Statistics**, Alexandria, v.7, n.4, p.434-455, Dec. 1998.
- BROWN, P. E.; DIGGLE, P. J.; LORD, M. E.; YOUNG, P. C. Space-time calibration of radar-rainfall data. **Applied Statistics**, v.50, p.221-242, 2001.
- CÂMARA, G.; MONTEIRO, A. M. V. Conceitos básicos em ciência da geoinformação. **Introdução à ciência da geoinformação**. São José dos Campos, SP: INPE, 2004. Disponível em: <<http://www.dpi.inpe.br/gilberto/livro/introd/>>. Acesso em 22 fev. 2007.
- CARVALHO, M. S. **Aplicação de métodos de análise espacial na caracterização de áreas de risco a saúde**. 1997 p.179. Tese (Doutorado em Engenharia Biomédica) - Universidade Federal do Rio de Janeiro. Instituto Alberto Luiz Coimbra, Rio de Janeiro.
- CASTELLANOS, P. L. Sobre el concepto de salud-enfermedad. Descripción y explicación de la situación de salud. **Boletín Epidemiológico OPAS**, v.10, n.4, p.1-7, 1990.
- CHRISTENSEN, O. F.; MØLLER, J.; WAAGEPETERSEN, R. P. Geometric ergodicity of Metropolis-Hastings algorithms for conditional simulation in generalised linear mixed models. **Methodology and Computing in Applied Probability**, v.3, p.309-327, 2001.
- CHRISTENSEN, O. F.; RIBEIRO JR., P. J. GeoRglm: a package for generalised linear spatial models. **R-News**, v.2, n.2, p.26-28, 2002.
- CHRISTENSEN, O. F.; RIBEIRO JR., P. J. **GeoRglm**: a package for generalized linear spatial models – introductory session. Mar. 2007. Software. Disponível em: <<http://www.daimi.au.dk/~olefc/geoRglm/Intro/intro.html>>. Acesso em: 22 abr. 2007.
- CHRISTENSEN, O. F.; WAAGEPETERSEN, R. P. Bayesian prediction of spatial count data using generalised linear mixed models. **Biometrics**, v.58, p.280-286, 2002.

CLIFF, A. D.; ORD, J. K. **Spatial process models and applications**. London: Pion, 1981.

COSTA, M. C. N.; AZI, P. A.; PAIM J. S.; SILVA, L. M. V. Mortalidade infantil e condições de vida: a reprodução das desigualdades sociais em saúde na década de 90. **Caderno de Saúde Pública**, v.7, p.555-567, 2001.

HENRY E.; KYBURG Jr., Foresight: its logical laws, its subjective sources. In: KYBURG JR., H. E.; SMOKLER, H. E. (Ed.). **Studies in subjective probability**. New York: J. Wiley, 1964. p. 53-118.

DE LORENZI D.R.S.; TANAKA, A.C.A. BOZZETTI, M.C.; RIBAS, F.E.; WEISHEIMER, L. A natimortalidade como indicador de saúde perinatal. **Caderno de Saúde Pública**, v.17, p.141-146, 2001.

DIGGLE, P. J.; ROWLINGSON, B. S. A conditional approach to point process modelling of elevated risk. **Journal of the Royal Statistical Society, Series A**, v.157, p.433-440,1994.

DIGGLE, P. J. HEAGERTY, P. LIANG, K. Y. ZEGER, S. L. **Analysis of Longitudinal data**. Oxford. Science, 2002. 396 p.

DIGGLE, P. J.; TAWN, J. A.; MOYEED, R. A. Model-based Geostatistics. **Applied Statistics**, v.47, p.299-326, 1998.

DIGGLE, P. J. **Statistics analysis of spatial point patterns**. London: Arnold, 2003.

DIGGLE, P. J.; RIBEIRO JR., P. J.; CHRISTENSEN, O. An introduction to model-based geostatistics. In MOLLER, J. (Ed.). **Spatial statistics and computational methods**. New York: Springer, 2003.

DIGGLE, P. J.; RIBEIRO JR., P. J. **Model-based geostatistics**. Hardcover: Springer, 2007. 230 p. (Series in Statistics).

D'ORSI, E.; CARVALHO, M. S. Perfil de Nascimentos no município do Rio de Janeiro: uma análise espacial. **Caderno de Saúde Pública**, v.14, p.367-379, 1998.

ECKER, M.; GELFAND, A.E. **Spatial modeling and prediction under range anisotropy**. Connecticut: University of Connecticut, 1997. (Technical Report 97 – 26).

- ELLIOTT, P.; WAKEFIELD, J.; BEST, N.; BRIGGS, D. **Spatio epidemiology: methods and applications**. London: Oxford University, 2001.
- FAN, J. Q.; HECKMAN, N. E.; WAND, M. P. Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. **Journal of the American Statistical Association**, v.90, p.141-150, 1995.
- FISHER, R. A. **The desing of experiments**. Edinburgh: Oliver and Boyd, 1935.
- FUNDAÇÃO INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo Demográfico 1991** - resultado do universo relativo às características da população e dos domicílios. Rio de Janeiro, 1993.
- GAMERMAN, D.; LOPES, H. F. **Markov Chain Monte Carlo: stochastic simulation for bayesian inference**. Illinois: Chapman and Hall, 2006. 344 p.
- GELMAN, A. Inference and monitoring convergence. In: GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. J. (Ed.). **Markov Chain Monte Carlo in practice**. New York: Chapman and Hall, 1995. cap. 8.
- GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian data analysis**. London: Chapman and Hall, 1997. 526 p.
- GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical Science**, Hayward, v. 7, n. 4, p. 457-511, 1992.
- GEMPERLI, A., VOUNATSOU, P., KLEINSCHMIDT, I., BAGAYOKO, M., LENGELER, C., SMITH T. Spatial patters of infant mortality in mali: the effect of malaria endemicity. **American Journal of Epidemology**, v.159, n.1, p.64-72, 2003.
- GEWEKE, J. Evaluating the accurary of sampling-based approaches to the calculation of posterior moments. *N*: BERNARDO, J. M.; BERGER, J. O.; DAWID, A. P.; SMITH, A. F. M. (Ed.). **Bayesian statistics 4**. New York: Oxford University, 1992. p.625-631.
- HANDCOCK, M. S.; WALLIS, J. R.; An approach to statistical spatio-temporal modeling of meteorological fields (with discussion). **Journal of the American Statistical Association**, v.89, p.368-378, 1994.

HARTZ, Z. M. A. Mortalidade infantil "evitável" em duas cidades do nordeste do Brasil: Indicador de qualidade do sistema local de saúde. **Revista de Saúde Pública**, v.30, p.310-318, 1996.

HASTIE, T. J.; TIBSHIRANI, R. J. Generalized additive models. **Statistical Science**, v.1, p.297-318, 1990. (With Discussion).

HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, London, v.57, n.1, p.97-109, Apr. 1970.

HEIDELBERGER, P.; WELCH, P. Simulation run length control in the presence of an initial transient. **Operations Research**, Landing, v. 31, n. 6, p. 1109-1144, 1983.

HOLLSTEIN, R. D.; VEJA, J.; CARVAJAL, Y. Desigualdades sociales e salud. Nivel socioeconomico y mortalidad infantil Chile, 1985-1995. **Revista Med. Chile**, v.126, p.333-340, 1998.

HOSMER, D. W.; LEMESHOW, S. **Applied logistic regression**. New York: J. Wiley, 1989

JEFFREYS, H. **Theory of probability**. Oxford: Clarendon, 1939. 380 p.

JEKEL, J. F.; KATZ, D. L.; ELMORE, J. G. **Epidemiologia, bioestatística e medicina preventiva**. 2.ed. Porto Alegre: Artmed, 2005.

JOSEPH, K. S.; KRAMER, M. S.; ALLEN, A. C.; CYR, M.; FAIR, M.; OHLSSON, A.; WEN, S. W. Gestational age- and birthweight-specific declines in infant mortality in Canada, 1985-94. For the Fetal and Infant Health Study Group of the Canadian Perinatal Surveillance System. **Paediatr Perinat Epidemiol**, v.14, p. 332-339, 2000.

KAHHALE, S. Quando vamos melhorar a qualidade da assistência perinatal no Brasil? **Revista Ginecologia Obstetrícia**, v.11, n.1, p.1, jan./mar. 2000

KAZEMBE, L. N.; KLEINSCHMIDT, I.; HOLTZ, T. H.; SHARP, B. L. Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. **Journal of Health Geographics**, v.5, n.41, 2006. Disponível em: < <http://www.ij-healthgeographics.com/content/5/1/41>>. Acesso em: 10 nov. 2007.

KELSALL, J. E.; DIGGLE, P. J. Non parametric estimation of spatial variation in relative risk. **Statistics in Medicine**, v.14, p.2335-2342, 1995.

KELSALL, J. E.; DIGGLE, P. J. Spatial variation in risk of disease: A nonparametric binary regression approach. **Applied Statistics**, v.47, p.559-573., 1998.

KELSALL, J. E. **Kernel smoothing for application in environmental epidemiology**. 1996. Thesis (Ph.D.) - Lancaster University. Department of Mathematics and Statistics, Lancaster.

KNOX, E. G.; BARTLETT, M. S. The Detection of Space-Time Interactions. **Applied Statistics**, v.13, n.1. p.25-30, 1964.

KNOX, E. G. Detection of clusters. In: ELLIOTT, P. (Ed.). **Methodology of enquiries into disease clustering**. London: London School of Hygiene and Tropical Medicine, 1988. p. 17-22.

KYRIAKIDIS, P. C.; JOURNEL, A. G. Geoostatistical space-time models. **Mathematical Geology**, v.31, p.651-684, 1999.

LANSKY, S.; FRANCA, E.; LEAL, M. do C. Mortalidade perinatal e evitabilidade: revisão da literatura. **Revista de Saúde Pública**, v.36, n.6, p.759-772, dez. 2002.

LEAL, M. C.; GAMA, S. G. N.; CAMPOS, M. R. *et al.* Fatores associados à morbi-mortalidade perinatal em uma amostra de maternidades públicas e privadas do Município do Rio de Janeiro, 1999-2001. **Caderno de Saúde Pública**, v.20, n.1, p.20-33, 2004.

LE, N. D.; ZIDEK, J. V. Interpolation with uncertain spatial covariances: A Bayesian alternative to Kriging. **Journal of Multivariate Analysis**, v.43, p.351-374, 1992.

LIMA, M. L. C de.; XIMENES, R. Violência e morte: diferenciais da mortalidade por causas externas no espaço urbano do Recife. **Caderno de Saúde Pública**, v.14, n.4, p.829-840, 1998.

MATÉRN, B. **Spatial variation**. 2.ed. Berlin: Springer-Verlag, 1986.

MCCULLAGH, P.; NELDER, J. A. **Generalized linear models**. 2.ed. London: Chapman and Hall, 1989.

- MENDES, E. V. **Uma agenda para a saúde**. São Paulo: Hucitec, 1996.
- MONTEIRO, C. A. Contribuição para o estudo do significado do coeficiente da mortalidade infantil no Município de São Paulo, SP (Brasil), nas três últimas décadas (1950-1979). **Revista de Saúde Pública**, v.16, p.7-18, 1982.
- NOGUEIRA, D. A. **Proposta e avaliação de critérios de convergência para o método de Monte Carlo via Cadeias de Markov**: casos uni e multivariados. 2004. 121p. Dissertação (Mestrado em Agronomia, Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, MG.
- NOGUEIRA, D. A.; SAFADI, T.; FERREIRA, D. F. Avaliação de critérios de convergência univariados para o método de Monte Carlo via cadeias de Markov. **Revista Brasileira de Estatística**, Rio de Janeiro, v.65, n. 224, p.59-88, jul./dez. 2004.
- O'HAGAN, A. **Kendall's advanced theory of statistics**: Bayesian inference [S.l]: E. Arnold, 1994. v.2b.
- OLIVEIRA, L. A. P.; MENDES, M. M. S. Mortalidade Infantil no Brasil: Uma avaliação de tendências recentes. In: MINAYO, M. C. S. (Org.). **Os muitos Brasis**: saúde e população na década de 80. São Paulo: Hucitec, 1995. p.291-303.
- PAEZ, M. S. **Análise de modelos para a estimação e previsão de processos espaço-temporais**. 2004. 161p. Tese (Doutorado em Estatística) - Universidade Federal do Rio de Janeiro. Instituto de Matemática, Rio de Janeiro.
- PAEZ, M. S.; GAMERMAN, D. Study of the space-time effects in the concentration of airborne pollutants in the metropolitan region of Rio de Janeiro. **Environmentrics**, v.14, p.387-408, 2003.
- PAIM, J. S.; COSTA, M. C. N.; CABRAL, V.; MOTA, I. A.; BATISTA-NEVES, R. B. Spatial distribution of proportional infant mortality and certain socioeconomic variables in Salvador, Bahia, Brazil. **Bulletin of the Pan-American Health Organization**, v.21, p.225-239, 1987.
- R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2007. Disponível em: <<http://www.R-project.org>>. Acesso em: 10 nov. 2007.

RAFTERY, A. E.; LEWIS, S. Comment: one long run with diagnostics: implementation strategies for Markov chain Monte Carlo. **Statistical Science**, Hayward, v. 7, n. 4, p. 493-497, 1992.

REIS, I. A. **O Estado da arte da integração entre sistemas de informação geográfica e modelos inferenciais bayesianos**. Pós-graduação em Sensoriamento Remoto. São José dos Campos, SP: INPE, 2005. Disponível em: <<http://www.dpi.inpe.br/~ilka>>. Acesso em: 08 mar. 2007.

RIBEIRO JR., P. J.; DIGGLE, P. J. geoR: a package for geostatistical analysis **R-News**, v.1, n.2, p.15-18, June 2001. Software.

RIPLEY, B. D. **Spatial statistics**. Chichester: J. Wiley, 1981. Chap. 6/8.

ROUQUAYROL, M. Z.; ALMEIDA FILHO, N. **Epidemiologia e saúde**. Rio de Janeiro: MEDSI, 2003.

ROWLINGSON, B.; DIGGLE, P. J. R. **Splancs**: spatial and space-time point pattern analysis. R package version 2.01-23. Disponível em: <<http://www.maths.lancs.ac.uk/~rowlings/Splancs/>>. Acesso em: 10 nov. 2007.

SÁFADI, T.; PEÑA, D. Bayesian analysis of dynamic factor models: an application to air pollution and mortality in São Paulo, Brazil. **Environmetrics**, (DOI: 10.1002/env.899), 2007. Disponível em: <<http://dx.doi.org/10.1002/env.899>>. Acesso em: 13 dez 2007

SANTOS, S. **Análise da distribuição espacial dos homicídios no ano de 1996 em Porto Alegre/RS**. 1999. Dissertação (Mestrado em Epidemiologia) - Escola Nacional de Saúde Pública, Rio de Janeiro.

SANTOS, S. M.; BARCELLOS, C.; CARVALHO, M. S.; FLÔRES, R. Detecção de aglomerados espaciais de óbitos por causas violentas em Porto Alegre, Rio Grande do Sul, Brasil, 1996. **Caderno de Saúde Pública**, v.17, n.5, p.1141-1151, 2001.

SANTOS, S. M.; NORONHA, C.P. Padrões espaciais de mortalidade e diferenciais socioeconômicos na cidade do Rio de Janeiro. **Caderno de Saúde Pública**, v.17, p.1099-110, 2001.

SASTRY, N. Community characteristics, individual and household attributes, and child survival in Brazil. **Demography**, v.33, p.211-229, 1996.

SCHOLTEN, H. J.; LEPPER, M. J. C. The benefits of the application of geographic information systems in public and environmental health. **World Health Statistical Quarterly**, v.44, p.160-170, 1991.

SEAMAN S. R.; RICHARDSON S. Bayesian analysis of case-control studies with categorical covariates. **Biometrika**, v.88, n.4, p.1073-1088, 2001.

SHADDICK, G.; WAKEFIELD, J. Modeling multivariate pollutant data at multiple sites. **Applied Statistics**, v.51, p.351-372, 2002.

SHIMAKURA, S. E.; CARVALHO, M. S.; AERTS, D. R. G. C.; FLORES, R. Distribuição espacial do risco: modelagem da mortalidade infantil em Porto Alegre, Rio Grande do Sul, Brasil. **Cadernos de Saúde Pública**, Rio de Janeiro, v.17, n.5, p.1251-1266, 2001.

SILVA, G. L. Da.; DEAN, C. B. Uma introdução à análise de modelos espaço-temporais para taxas, proporções e processos de multi-estados. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA 17., 2006, Caxambu. **Anais...** Caxambu, MG: ABE, 2006. 112p. (Mini-Curso).

SILVA, G. L. da; DEAN, C. B.; NIYONSENGA, T.; VANASSE, A. **Hierarchical Bayesian spatiotemporal analysis of revascularization odds using smothing splines**. Lisboa: Technical University of Lisboa. IST, 2005. (Technical Report).

SILVERMAN, B. W. **Density estimation for statistics and data analysis**. London: Chapman and Hall, 1986.

SLOGGER, A.; JOSHI, H. Deprivation indicators as predictors of life events 1981-1992 based on the UK ONS longitudinal study. **Journal of Epidemiology and Community Health**, v.52, p.228-233, 1998.

SMITH, B. J. **Bayesian Output Analysis Program (BOA)**. Version 1.1.5. The University of Iowa, 2005. Disponível em: <<http://www.public-health.uiowa.edu/boa>>. Acesso em: 23 mar. 2007.

TEIXEIRA, C. F.; PAIM, J. S.; VILAS BÔAS, A. L. SUS, modelos assistenciais e vigilância da saúde. 1998. **Informe Epidemiológico do SUS**, 7: 7-28.

THEME FILHA, M. M.; GAMA, S. G. N.; CUNHA, C. B.; LEAL, M. do C. Confiabilidade do Sistema de Informações sobre Nascidos Vivos Hospitalares no Município do Rio de Janeiro, 1999-2001. **Caderno de Saúde Pública**, v.20, n.1, p.83-91, 2004.

TONELLATO, S. F. A multivariate time series model for the analysis e prediction of carbon monoxide atmospheric concentrations. **Applied Statistics**, v.50, p.187-200, 2001.

VICTORA, C. G.; VAUGHAN, J. P; BARROS, F. C; SILVA, A. C; TOMASI, E. Explaining trends in inequities: evidence from Brazilian child health studies. **The Lancet**, v.356, p.1093-1098, Sept. 2000.

VON MISES, R. **Wahrscheinlichkeitsrechnung**. Vienna: Deutiche, 1931.

WAAGEPETERSEN, R.; IBANEZ, N.; SORENSEN, D. A comparison of strategies for Markov chain Monte Carlo computation in quantitative genetics **Genetics, Selection, Evolution**, v.40, p.161-176, 2008.

WALLER, L. A.; GOTWAY, C. A. **Applied Spatial Statistics for Public Health Data**. New York: J. Wiley, 2004.

WERNECK, G.L.; STRUCHINER, C.J. Estudos de agregados de doença no espaço-tempo: conceitos, técnicas e desafios. **Caderno de Saúde Pública**, Rio de Janeiro, v.13, n.4, p.611-624, out./dez. 1997.

WIGGLESWORTH, J. S. Monitoring perinatal mortality - A pathophysiological approach. **The Lancet**, 1980, 27:684-686

WIKLE, C. K.; BERLINER, L. M.; CRESSIE, N. Hierarchical Bayesian space-time models. **Environmental and Ecological Statistics**, 5, 117-154. 1998

WORLD HEALTH ORGANIZATION. **Perinatal mortality. A listing of available information**. Geneva: World Health Organization; 1996.

XIMENES, R. A. A.; MARTELLI, C. M. T.; SOUZA, W. V.; LAPA, T. M.; ALBUQUERQUE, M. F. M; ANDRADE, A. L. S.; MORAIS NETO, O. L.; SILVA, S. A.; LIMA, M. L. C.; PORTUGAL, J. L. Vigilância das doenças endêmicas em áreas urbanas: A interface entre mapas de setores censitários e indicadores de morbidade. **Cadernos de Saúde Pública**, 1999, 15:53-61, 1999.

YUNES, J. Características sócio-econômicas da mortalidade infantil em São Paulo. **Pediatria**, v.5, p.162-168, 1983.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)