

UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

ESTELA CRISTINA CARNESECA

**Problemas respiratórios e fatores ambientais: uma análise  
Bayesiana para dados de Ribeirão Preto**

Ribeirão Preto - SP  
2011



ESTELA CRISTINA CARNESECA

**Problemas respiratórios e fatores ambientais: uma análise  
Bayesiana para dados de Ribeirão Preto**

Dissertação apresentada ao Departamento de  
Medicina Social da Faculdade de Medicina de  
Ribeirão Preto da Universidade de São Paulo  
para a obtenção do título de Mestre em Saúde  
na Comunidade.

Área de concentração: Saúde na Comunidade

Orientador: Prof. Dr. Jorge Alberto Achcar

Ribeirão Preto - SP  
2011



## **DEDICATÓRIA**

À minha família pelo incentivo e apoio incondicional neste e em todos os projetos de minha vida.

## **AGRADECIMENTOS**

À Deus, que esteve ao meu lado ao longo desses 25 anos de vida.

Aos meus pais, tios, avós e irmão, que por acreditarem em mim, me ensinaram a acreditar em mim mesma.

À todos os meus amigos que de alguma forma contribuíram para a concretização deste trabalho. Em especial, agradeço a Mayara, Juliana e Ana pelos momentos de alegrias, conselhos e por compartilharem angústias e dúvidas em momentos difíceis.

À Profa. Dra. Doris Lieth Nunes Peçanha, pelo incentivo e motivação para que eu investisse em minha carreira acadêmica.

Ao meu orientador, Prof. Dr. Jorge Alberto Achcar, pela transmissão do conhecimento e pela confiança depositada em mim para a realização deste trabalho.

Ao Prof. Dr. Edson Zangiacomi Martinez por transmitir seu conhecimento e experiência ao longo de meu mestrado e de meu trabalho no CEMEQ. Sua participação e apoio foram fundamentais para que esta dissertação fosse realizada.

À todos os colegas, professores e funcionários do Departamento de Medicina Social e do programa de pós-graduação em Saúde na Comunidade pelo convívio e aprendizado.

Enfim, a todos aqueles que de alguma forma contribuíram para que este trabalho pudesse ser concretizado.

“Escolhe um trabalho de que gostes,  
e não terás que trabalhar nem um  
dia na tua vida.”

Confúcio

## RESUMO

CARNESECA, E. C. **Problemas respiratórios e fatores ambientais: uma análise Bayesiana para dados de Ribeirão Preto**. Dissertação (mestrado). Ribeirão Preto: Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto, 2011. 84 f.

Estudos envolvendo o meio ambiente estão sendo cada vez mais desenvolvidos devido ao fato dos níveis de poluição e das mudanças climáticas estarem causando a degradação da qualidade do ar e dos reservatórios de água de maneira alarmante nos últimos anos, comprometendo sobretudo, a qualidade de vida do ser humano. Dado que estes fatores são preponderantes nos agravos e complicações respiratórias dos indivíduos, buscou-se compreender com este estudo a relação entre as condições atmosféricas e os problemas respiratórios nos residentes do município de Ribeirão Preto, interior de São Paulo, onde há um elevado número de focos de queimadas nos períodos de estiagem e, conseqüentemente, altas concentrações de poluentes, como o material particulado. Considerando os dados mensais de contagem de inalações/nebulizações, foram assumidos diferentes modelos de regressão de Poisson na presença de um fator aleatório que captura a variabilidade extra-Poisson entre as contagens. A análise dos dados foi feita sob enfoque Bayesiano, utilizando métodos de simulação MCMC (Monte Carlo em Cadeias de Markov) para obter os sumários a posteriori de interesse.

**Palavras-chave:** *Modelo de regressão de Poisson, Análise Bayesiana, Problemas respiratórios, Queimadas, Material particulado.*



## ABSTRACT

CARNESECA, E. C. **Respiratory problems and environmental factors: a Bayesian analysis for data from Ribeirão Preto City**. Dissertation (master degree). Ribeirão Preto: University of São Paulo, Faculty of Medicine of Ribeirão Preto, 2011. 84 p.

Many studies involving the environment are being developed in the last years due to the fact that the levels of pollution and climate changes are causing the degradation of air quality and water reservoirs at an alarming rate in recent years, with great consequences for the quality of life of the population. Since these factors are prevalent in respiratory disorders and complications of the health for the individuals, we intended to understand from this study the relationship between weather conditions and respiratory problems for the residents of the municipality of Ribeirão Preto, São Paulo, which has a high number of outbreaks of fires in drought periods and, consequently, high concentrations of pollutants such as particulate matter. Considering the monthly count of inhalations / nebulizations, we assumed different Poisson regression models in the presence of a random factor that captures the extra-Poisson variability between the counts. The data analysis was performed under a Bayesian approach using MCMC simulation methods (Markov Chain Monte Carlo) to get the posterior summaries of interest.

**Keywords:** *Poisson regression model, Bayesian Analysis, Respiratory problems, Fires, Particulate matter.*

## LISTA DE FIGURAS

1 Localização do município .....	19
2 Área total plantada de cana-de-açúcar na região de Ribeirão Preto, no período de 2002 a 2008 .....	20
3 Mapa de Ribeirão Preto com a indicação do local (A) onde se situa a estação de monitoramento da qualidade do ar, no bairro Ipiranga.....	23
4 Série do total de inalações mensais no período .....	25
5 Descrições das variáveis climáticas e da inalação ao longo do tempo .....	26
6 Gráficos de dispersão envolvendo as inalações e os valores de precipitação, temperatura mínima e material particulado .....	27
7 Boxplot das inalações segundo categorias do material particulado.....	42
8 Gráfico dos resíduos <i>versus</i> observação, considerando os valores contínuos do material particulado .....	56
9 Gráfico dos resíduos <i>versus</i> observação, considerando os valores do material particulado categorizados .....	57
10 Gráfico de CPO <i>versus</i> o número da observação para os modelos 2 e 3, considerando $MP_{10}$ contínuo e categorizado, separadamente .....	57
11 Riscos relativos e respectivos intervalos de credibilidade (95%) para inalações com concentrações crescentes do $MP_{10}$ .....	58
12 Dados imputados na série das quantidades mensais de inalação/nebulização.....	61
13 Riscos relativos e respectivos intervalos de credibilidade (95%) para inalações/nebulizações com concentrações crescentes do $MP_{10}$ , considerando as estimativas dos modelos com e sem imputação dos dados.....	63
14 Gráfico de CPO <i>versus</i> o número da observação, considerando $MP_{10}$ as estimativas do modelo 2 com e sem imputação dos dados.....	64

## LISTA DE TABELAS

1 Estatísticas descritivas segundo estação do ano. Ribeirão Preto, 2004 a 2010.....	26
2 Criação das variáveis indicadoras através dos valores de $MP_{10}$ categorizado.....	43
3 Estimativas dos parâmetros do modelo clássico sem $w_i$ , com seus respectivos intervalos de confiança (IC) 95%. (material particulado contínuo) .	51
4 Estimativas dos parâmetros do modelo clássico com a presença de $w_i$ , com seus respectivos intervalos de confiança (IC) 95%. (material particulado contínuo).....	51
5 Estimativas dos parâmetros do modelo 1, com seus respectivos intervalos de credibilidade (IC) 95%. (material particulado contínuo).....	51
6 Estimativas dos parâmetros do modelo 2, com seus respectivos intervalos de credibilidade (IC) 95%. (material particulado contínuo).....	52
7 Estimativas dos parâmetros do modelo 3, com seus respectivos intervalos de credibilidade (IC) 95%. (material particulado contínuo).....	52
8 Soma dos resíduos absolutos de cada modelo. (material particulado contínuo).....	53
9 Estimativas dos parâmetros do modelo clássico sem $w_i$ , com seus respectivos intervalos de confiança (IC) 95%. (material particulado categorizado) .	54
10 Estimativas dos parâmetros do modelo clássico com a presença de $w_i$ , com seus respectivos intervalos de confiança (IC) 95%. (material particulado categorizado). .	54
11 Estimativas dos parâmetros do modelo 1, com seus respectivos intervalos de credibilidade (IC) 95%. (material particulado categorizado)..	54
12 Estimativas dos parâmetros do modelo 2, com seus respectivos intervalos de credibilidade (IC) 95%. (material particulado categorizado). .	55
13 Estimativas dos parâmetros do modelo 3, com seus respectivos intervalos de credibilidade (IC) 95%. (material particulado categorizado). .	55
14 Soma dos resíduos absolutos de cada modelo. (material particulado categorizado). .	55

15 Descrição dos valores de inalação/nebulização gerados por cada uma das cinco imputações realizadas.....	61
16 Estimativas combinadas pontuais e intervalares (95%) dos parâmetros do modelo 2. (material particulado contínuo).....	61
17 Estimativas combinadas pontuais e intervalares (95%) dos parâmetros do modelo 2. (material particulado categorizado).....	62

# SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	13
1.1 Interações entre o clima e a saúde humana.....	13
1.2 O uso do fogo e a saúde pública.....	14
1.2.1 O processo de combustão da biomassa.....	16
1.3 O município de Ribeirão Preto.....	18
1.3.1 Aspectos geográficos e econômicos.....	18
1.3.2 Ar seco e problemas respiratórios.....	20
<b>2 OBJETIVO</b> .....	22
<b>3 COLETA E DESCRIÇÃO DOS DADOS</b> .....	23
3.1 Obtenção dos dados.....	23
3.2 Análise exploratória.....	24
<b>4 MODELAGEM ESTATÍSTICA</b> .....	28
4.1 Metodologia Bayesiana.....	28
4.1.1 Fórmula de Bayes e inferência Bayesiana.....	29
4.1.2 Distribuições a priori não-informativas.....	31
4.1.3 Métodos de simulação para obtenção dos sumários a posteriori.....	32
4.1.3.1 Método de Monte Carlo.....	32
4.1.3.2 Método de Monte Carlo em cadeias de Markov.....	34
4.1.3.2.1 Amostrador de Gibbs.....	36
4.1.3.2.2 Algoritmo de Metropolis-Hastings.....	37
4.1.4 Formulação de modelos.....	38
4.1.5 Convergência do algoritmo.....	44
4.1.6 Seleção de modelos.....	45
4.2 Imputação de dados faltantes.....	46
4.2.1 Imputação múltipla.....	47

4.2.1.1 Método de imputação baseado na Regressão Linear Bayesiana	48
<b>5 RESULTADOS</b>	50
5.1 Análise estatística dos dados sem imputação	50
5.1.1 Ajustes considerando os valores contínuos do material particulado	50
5.1.2 Resultados considerando os valores categorizados do material particulado	53
5.1.3 Seleção do melhor modelo	56
5.2 Análise estatística dos dados com imputação	59
<b>6 DISCUSSÃO DOS RESULTADOS OBTIDOS</b>	65
<b>7 CONCLUSÕES E PERSPECTIVAS FUTURAS</b>	70
<b>REFERÊNCIAS</b>	71
<b>A APÊNDICE - Programas</b>	79
A.1 Seção 4.1.4	79
A.1.1 Modelo 1 (material particulado contínuo)	79
A.1.2 Modelo 1 (material particulado categorizado)	80
A.1.3 Modelo 2 (material particulado contínuo)	81
A.1.4 Modelo 2 (material particulado categorizado)	82
A.1.5 Modelo 3 (material particulado contínuo)	83
A.1.6 Modelo 3 (material particulado categorizado)	84

## 1 INTRODUÇÃO

Muitos são os fatores que influenciam a saúde de uma população. Entre eles, pode-se destacar o clima e/ou as condições atmosféricas. Juntos, estes fatores se tornam de grande valia, cuja importância varia com a doença em questão e com as características de cada indivíduo (PITTON & DOMINGOS, 2004).

As relações entre as condições atmosféricas e a saúde humana tem sido objeto de diversos estudos desenvolvidos nas últimas décadas. Nesse contexto, o presente trabalho buscou compreender a relação entre as condições atmosféricas e complicações respiratórias nos residentes em Ribeirão Preto – SP, escolhida por apresentar um clima muito seco em quase todo o ano e por ter um elevado número de focos de queimadas urbanas e rurais em períodos de estiagem, que são características que propiciam o surgimento ou o agravamento de doenças respiratórias.

Esta dissertação está estruturada da seguinte forma: No Capítulo 1, é apresentada uma introdução sobre o tema abordado. No Capítulo 2, o objetivo do estudo. No Capítulo 3, uma descrição da forma de coleta dos dados e análise exploratória dos mesmos. No Capítulo 4, apresenta-se a abordagem Bayesiana para ajuste e seleção dos modelos. No Capítulo 5, resultados. No Capítulo 6, discussão. No Capítulo 7, conclusões e perspectivas futuras.

### 1.1 Interações entre o clima e a saúde humana

Ainda que muitas vezes seus efeitos sejam ignorados, as condições climáticas sempre influenciam os seres vivos e sua capacidade de resposta a elas próprias. Nesse sentido, a climatologia médica considera o clima como o fator determinante de efeitos favoráveis ou desfavoráveis sobre os seres humanos (SOUZA E NETO, 2008). No passado, Hipócrates, na tentativa de eliminar as causas sobrenaturais sobre as doenças, já destacava que o sol, a água, os ventos e os elementos climáticos eram essenciais para a manutenção e recuperação da saúde; ao mesmo tempo, é comum a afirmativa de que o ar puro e a adequada exposição ao sol são fatores favoráveis à saúde, bem como a busca por um local com clima ameno foi, durante muitos anos, um importante procedimento terapêutico para o tratamento da tuberculose (ver por exemplo, SOUZA E NETO, 2008).

As condições climáticas são resultantes de alterações da paisagem natural provocadas pelas intensas atividades humanas, que a substitui por um ambiente construído (AYOADE, 1991). O clima urbano, por sua vez, é capaz de comprometer, dentre outras, a qualidade da saúde humana. Apesar do corpo humano possuir um sistema que regula e mantém o equilíbrio térmico (homeotermia), se pessoas predispostas, como idosos e crianças, e as portadoras de doenças crônicas forem expostas a situações extremas de calor ou de frio, ficarão mais propensas às enfermidades cardiovasculares, cerebrovasculares e respiratórias (PITTON e DOMINGOS, 2004).

A partir do acelerado processo de urbanização, característico do século XX, e com o surgimento de novas áreas urbanas nos países desenvolvidos e em desenvolvimento, a degradação ambiental e a queda na qualidade de vida nas cidades foram agravadas (DREW, 1998).

Segundo Ayoade (1991), devido à direta influência das atividades praticadas pelo homem sobre o clima, o seu maior impacto ocorre nas áreas urbanas, ocasionando um clima distinto das áreas rurais circundantes.

Monteiro (1976) afirma que a cidade é a própria responsável por seu clima. Sendo assim, o clima seria resultado da ação dos fatores que atuam sobre o contorno urbano, alterando o clima em escala local. Segundo o autor, a população sentiria os efeitos através do conforto térmico, qualidade do ar, entre outras manifestações que fossem capazes de prejudicar a qualidade de vida dos habitantes, bem como desorganizar a vida da própria cidade. Para Ayoade (1991), dois aspectos de maiores consequências devem ser considerados: o aumento da temperatura nessas áreas urbanas, responsável pela formação de ilhas de calor, e a poluição do ar, altamente prejudiciais à saúde.

## 1.2 O uso do fogo e a saúde pública

No período de estiagem prolongada, é comum a prática do uso do fogo em todo o Brasil. Porém, esta é uma prática que há séculos acompanha o homem, e é através do uso do fogo que a história da humanidade é registrada (JACOBI, 2007). Os ancestrais da espécie humana somente deram início a seu domínio sobre a vida na terra, ao adquirirem o controle sobre o fogo, quando a partir de então, passaram a alterar os ecossistemas (JACOBI, 2007).



Segundo Maia, et al. (2001), desde a ocupação portuguesa no país, as queimadas se tornaram uma das principais ferramentas utilizadas para a derrubada da vegetação nativa e abertura de novos campos para a prática da agricultura, pecuária e mineração.

Particularmente, a prática da queima da biomassa tem sido intensamente difundida entre os municípios brasileiros, principalmente em épocas de inverno, as quais estão associadas aos períodos de estiagem. Um dos principais propósitos dessa prática é a “limpeza” de áreas com folhas e galhos secos. Algumas vezes o objetivo da queima é apenas a redução do volume de lixo doméstico (TRESMONDI et al., 2008).

O Brasil é o maior produtor de cana-de-açúcar do mundo. Segundo dados da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), atualmente 4.5 milhões de hectares do território nacional são utilizados para o plantio de cana, representando 1% das terras agricultáveis. A produção nacional atingiu 436.8 milhões de toneladas na safra 2005/2006. Da safra canavieira, 55% são reservados à produção do álcool e seus derivados, enquanto o restante (45%) são destinados à produção de açúcar e subprodutos. O estado de São Paulo, com exceção do litoral, Serra do Mar e Vale do Ribeira, é responsável por cerca de 60% da cana brasileira (ÚNICA – União da Indústria de cana-de-açúcar).

Em cerca de 25% da produção brasileira o corte da cana é feito de forma mecanizada, já em relação a produção paulista, este percentual sobe para 40%; o restante sofre queima pré-corte e o corte é feito manualmente (RIBEIRO, 2008).

A queima de resíduos da cana-de-açúcar é uma prática generalizada no mundo, sobretudo em países em desenvolvimento e tem por principais objetivos limpar o terreno antes e após a colheita e controlar pragas e plantas daninhas que aparecem em processos de monocultura, além de propiciar uma maior facilidade e rendimento do corte e melhor qualidade tecnológica da matéria-prima, redução de custos de corte, carregamento, transporte e processo industrial. (OLIVEIRA E BARROCAS, 2004). A queimada consiste em atear fogo no canavial de forma que grande parte da matéria-prima considerada descartável, como folhas secas e verdes, sejam destruídas (LANGOWSKI, 2007).

Outra prática comum entre os brasileiros são as chamadas queimadas urbanas, prática-crime frequente em grande parte dos municípios, e que representam um importante componente na deterioração da qualidade do ar. Nas cidades, geralmente o fogo é empregado no fundo de quintais, terrenos e áreas abandonadas que servem de

depósitos clandestinos de lixo e lotes com construção apresentando vegetação de pequeno porte.

Apesar de ser comum, há séculos o uso do fogo nas matas e terrenos é condenado pelos manuais de conservação do solo, pelas consequências nocivas que provoca na terra, uma vez que queimadas são grandes responsáveis pelas alterações climáticas e consequente destruição da cobertura vegetal nativa, o que deixam desprotegidos os mananciais e nascentes e provoca uma alteração irreversível no ciclo das chuvas (FERREIRA, 2006).

### 1.2.1 O processo da combustão

A combustão é um processo químico pelo qual um material reage rapidamente com o oxigênio do ar produzindo luz e calor intenso e, no caso da biomassa, se faz em três estágios: ignição, combustão com chama e combustão sem chama (ARBEX et al., 2004).

Como citado por Arbex et al. (2004), a prolongada exposição à fumaça produzida pela queima de biomassa tem sido associada a efeitos deletérios à saúde, como infecções respiratórias agudas em crianças (KOSSOVE, 1982; PANDEY et al., 1989), doença pulmonar obstrutiva crônica (DOSSING, 1994; DENNIS et al., 1996) e tuberculose pulmonar (MISHRA et al., 1999; PEREZ-PADILLA et al., 2001), entre outros.

A incineração da biomassa é a maior fonte de produção de gases tóxicos, material particulado e gases do efeito estufa no planeta (CRUTZEN E ANDREAE, 1999). Em geral, pessoas idosas e crianças são as mais afetadas pelos níveis elevados de poluentes no ar.

Entre os elementos provenientes da combustão, o material particulado é o que tem gerado maior quantidade de estudos, devido, principalmente, a sua elevada toxicidade (SEATON et al., 1995). É composto basicamente por partículas sólidas ou por gotículas dispersas no ar que penetram com facilidade o sistema respiratório, ocasionando sérios problemas inflamatórios (SEATON et al., 1995).

As partículas inaláveis, denominação dada às partículas com diâmetro aerodinâmico menor que  $10\mu\text{m}$  ( $\text{MP}_{10}$ ), resultam de emissões de uma grande variedade de fontes e são emitidas diretamente (através das queimadas) ou são formadas na atmosfera ao reagirem com outros poluentes (TRESMONDI, 2003).

Apesar de existirem diferentes classificações quanto ao tamanho das partículas, de modo geral, tanto partículas finas quanto as mais grossas podem penetrar no organismo e se acumular no sistema respiratório, porém, as partículas mais finas tendem a agravar problemas respiratórios já existentes, aumentando o risco de internações emergenciais e de mortes prematuras (ARBEX et al., 2004).

De acordo com a explicação de livros didáticos de medicina (ver por exemplo, FATTINI E DANGELO, 2002) sobre o sistema respiratório do corpo humano, um conjunto de pequenos pêlos que existem no interior do nariz é capaz de filtrar o ar e reter as partículas maiores de poeira. Contudo, partículas menores conseguem passar e, se chegarem aos alvéolos (local onde ocorrem as trocas gasosas entre o ar e o sangue), desencadearão problemas pulmonares.

A superfície dos bronquíolos mais largos, dos brônquios e da traquéia é revestida por uma série de cílios e por uma camada de muco e as pequenas partículas que escapam da filtração realizada pelo nariz fixam-se nesse muco. Os cílios, que estão em constante movimento, empurram continuamente o muco para cima em direção à entrada da laringe. Chegando lá, o muco é engolido com a saliva, sem que a pessoa perceba. Assim, as partículas que se fixam no muco deixam de oferecer riscos à saúde pulmonar (FATTINI E DANGELO, 2002).

Quando o ar inspirado está extremamente seco, parte considerável da água presente no muco que reveste as vias respiratórias evapora, o que as deixa ressecadas. Isso provoca tosse, dificuldade para respirar e favorece a entrada de poeira nos pulmões. Assim, a presença de vapor de água no ar é muito importante, pois auxilia o organismo a respirar adequadamente. Nas épocas do ano em que o ar fica mais seco ocorrem, com maior frequência, problemas respiratórios na população, sobretudo em crianças e idosos (FATTINI E DANGELO, 2002).

Nos últimos anos, tem se percebido um aumento no número de estudos que tentam compreender especificamente a influência da poluição atmosférica, em especial do material particulado, sobre o aumento de internações por doenças respiratórias, como asma e pneumonia, ou cardiovasculares (MARTINS et al., 2002; FREITAS et al., 2004; SALES e MARTINS, 2005; NASCIMENTO et al., 2006; BRAGA et al., 2007; CARMO et al., 2010). Em todos eles, os autores sugerem que a poluição atmosférica promove efeitos adversos para a saúde tanto de idosos quanto de crianças.

A literatura científica que trata de efeitos da queima de biomassa à saúde, porém, ainda é bastante restrita. No Brasil, os estudos centralizam-se no interior do estado de

São Paulo, onde a população de áreas canavieiras fica exposta aos poluentes atmosféricos oriundos da queima. Araraquara, município da região central do estado, tem sediado várias pesquisas, pois em seus arredores há extensas plantações de cana. Em um desses estudos, Arbex et al. (2000) concluiu que a queima de cana pode ter efeitos adversos à saúde da população exposta. Semelhantemente, um estudo realizado em Piracicaba (SP) indicou que as internações hospitalares por doenças respiratórias em crianças e adolescentes estavam relacionadas às variações do material particulado de forma que um aumento de  $10\mu\text{g}/\text{m}^3$  do material particulado associava-se a um aumento de 21% nas internações (CANÇADO, 2003). Lopes e Ribeiro (2006), por sua vez, verificaram uma maior incidência de internações por doenças respiratórias em áreas onde há queimadas de cana.

Embora seja nítido o crescente número de produções científicas nessa área, a idéia de se associar a poluição do ar aos grandes centros urbanos, com a imagem de poluentes sendo eliminados exclusivamente por automóveis ou pelas indústrias, permanece. Um exemplo disso é a grande quantidade de estudos que analisam a relação entre poluição atmosférica e doenças respiratórias sendo desenvolvidos em cidades como São Paulo (MARTINS et al., 2002; FREITAS et al., 2004; GOUVEIA et al., 2006; BRAGA et al., 1999; BRAGA et al., 2001; PEREIRA et al., 1998), onde a poluição é predominantemente gerada pela atividade industrial e pela grande frota de veículos. É imprescindível considerar que atualmente grande parte da população do planeta convive com outra fonte de poluição: a queima da biomassa. Nesse sentido, este trabalho busca contribuir com a literatura da área, estudando os possíveis efeitos da queimada sobre a saúde humana numa região que possui uma grande extensão de monocultura canavieira e elevado índice de queimadas urbanas, como é o caso do município de Ribeirão Preto, apresentado a seguir.

### 1.3 O Município de Ribeirão Preto

#### 1.3.1 Aspectos geográficos e econômicos

Segundo dados da Prefeitura Municipal, a cidade de Ribeirão Preto (Figura 1), com quase 600 mil habitantes, situa-se em um planalto na região Norte/Nordeste do Estado de São Paulo. Seu relevo caracteriza-se por ser plano com pequenos declives e ondulações naturais, o que favorece a mecanização agrícola.

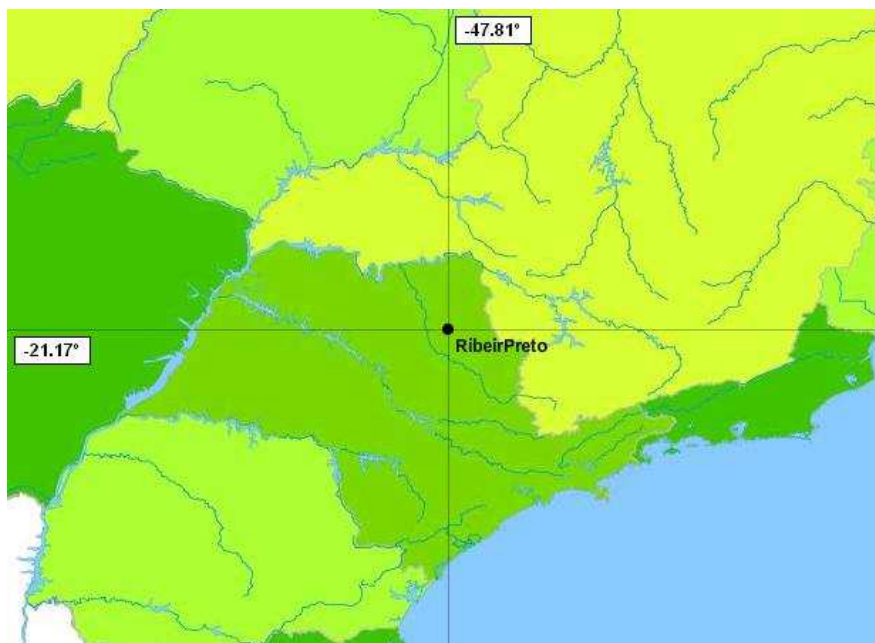


Figura 1. Localização do município.

Fonte: Página online da prefeitura municipal de Ribeirão Preto – [www.ribeiraopreto.sp.gov.br](http://www.ribeiraopreto.sp.gov.br). Acessado em agosto de 2010.

O clima típico do município é o tropical úmido, caracterizado pelo verão chuvoso e pelo inverno seco. Contudo, Ribeirão Preto é considerada por seus habitantes uma cidade que possui um calor excessivo, sendo inclusive apontada por jornalistas especialistas em matérias de meteorologia como a cidade mais quente do estado em algumas épocas do ano. Segundo dados da Prefeitura, no verão a temperatura máxima média é de cerca de 30 °C com um índice pluviométrico superior a 200 mm de chuva/mês e umidade relativa do ar em torno de 80%. O inverno apresenta uma temperatura mínima com média mensal em torno de 13 °C com precipitação média oscilando de 20 a 30 mm e umidade relativa do ar em cerca de 60%.

Os bons índices econômicos e sociais de Ribeirão Preto advêm da estrutura econômica forte e diversificada, tanto do município quanto da região, graças, principalmente, à agricultura. A qualidade do solo (terra roxa) e do clima faz com que esta seja uma das principais regiões agrícolas do Estado de São Paulo e do País, caracterizando-se por uma grande produção e por elevados níveis de rendimento das culturas, com destaque para a cana de açúcar, laranja, entre outras.

De acordo com a prefeitura municipal (página eletrônica: <http://www.ribeiraopreto.sp.gov.br/crp/dados/i01principal.htm>), a região é a maior produtora mundial de açúcar e álcool. As usinas representam uma das principais

atividades econômicas da região. São 21 usinas que empregam em torno de 8000 empregados. Segundo os resultados das pesquisas agrícolas municipais realizadas anualmente pelo IBGE (Figura 2), houve, na região, um aumento de quase 400 mil hectares de lavoura de cana de açúcar em 2008 sobre os 869 mil hectares registrados na pesquisa de 2002, o que representa um crescimento de 45,1% no período de 6 anos.

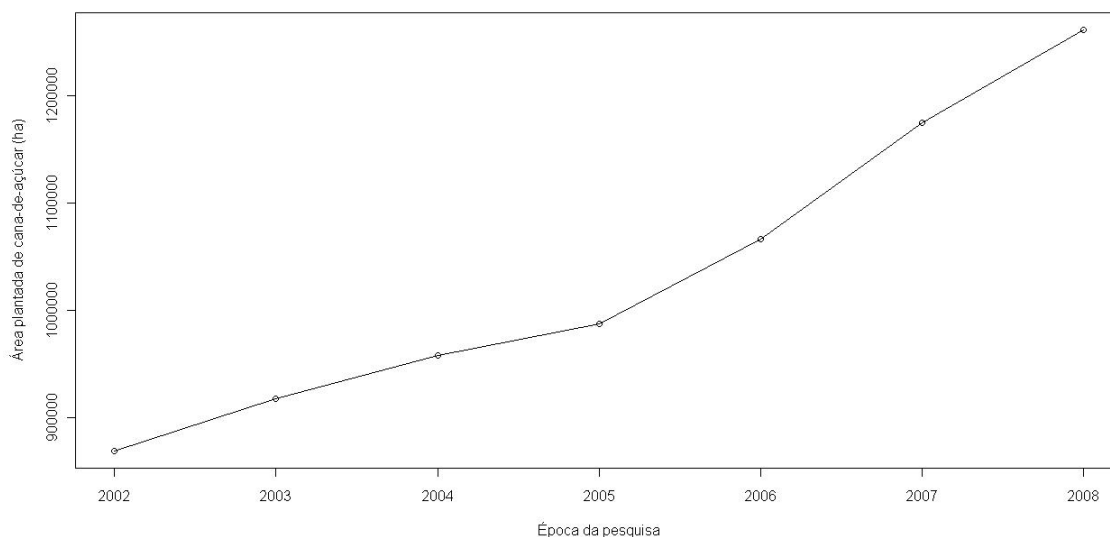


Figura 2. Área total plantada de cana-de-açúcar na região de Ribeirão Preto, no período de 2002 a 2008.

Fonte: Pesquisa Agrícola Municipal, IBGE.

### 1.3.2 Ar seco e problemas respiratórios

Um agravante ao clima seco de Ribeirão Preto é o aumento de queimadas nestes períodos de estiagem. A degradação ambiental nesta região do estado de São Paulo, causada pelo setor sucroalcooleiro, é muito grande, sendo as queimadas altamente prejudiciais à saúde, e apontadas por Ferreira (2006) como a principal fonte de poluição atmosférica da região. Porém, além das queimadas provocadas por este setor, existem ainda as chamadas queimadas urbanas, muito comuns na cidade e visíveis em calçadas ou quintais, onde qualquer pessoa pode facilmente entrar em contato com a fumaça. Tal prática acontece sem qualquer controle ou critério que minimize seus efeitos no meio ambiente ou na saúde da população.

Na região, até a década de 1970 havia 22% de cobertura florestal nativa. Com o advento do Proálcool (que tinha por objetivo estimular a produção do álcool, visando o atendimento das necessidades do mercado interno e externo e da política de combustíveis automotivos), essa área foi reduzida para menos de 3% nos dias atuais.

Mesmo com a quase inexistente cobertura vegetal para manter o equilíbrio ecológico da região, o fogo continua a ser utilizado na flora nativa remanescente (FERREIRA, 2006).

Como os canaviais não são plantados em áreas distantes, e como a queimada urbana é realizada próxima de áreas residenciais (normalmente em épocas de estiagem), os ventos facilmente dispersam a fumaça, fazendo com que qualquer pessoa possa entrar em contato com ela.

## **2 OBJETIVO**

Este estudo foi desenvolvido com a finalidade de analisar os efeitos da exposição ao material particulado, neste caso, ao  $MP_{10}$ , sobre o número de procedimentos de inalação/nebulização, uma vez que o fato de uma pessoa ser encaminhada a um ambulatório necessitando destes procedimentos seria indicativo de uma complicação respiratória, que, por sua vez, poderia ter sido causada pela piora na qualidade do ar.



### 3 COLETA E DESCRIÇÃO DOS DADOS

#### 3.1 Obtenção dos dados

As medidas de material particulado inalável ( $MP_{10}$ ) foram obtidas da Companhia de Tecnologia de Saneamento Ambiental (CETESB) no período de janeiro de 2004 a dezembro de 2010. Com apenas uma estação de monitoramento da qualidade do ar no Município de Ribeirão Preto, localizada no bairro Ipiranga (região norte da cidade) (Figura 3), não houve coleta diária dos dados – sua obtenção era realizada de duas a três vezes por semana; assim, preferiu-se trabalhar com a média mensal de  $MP_{10}$ , tomando os valores disponíveis como sendo representativos do mês a que se referiam.



Figura 3. Mapa de Ribeirão Preto com a indicação do local (A) onde se situa a estação de monitoramento da qualidade do ar, no bairro Ipiranga. (fonte:Google maps)

A partir das mensurações diárias de temperatura mínima e precipitação, obtidas pelo Centro Integrado de Informações Agro-Meteorológicas (CIAGRO) do Estado de São Paulo, obtiveram-se as médias mensais de temperatura mínima ( $^{\circ}C$ ) e o volume total mensal de precipitação (mm).

Os números de procedimentos de Inalação/Nebulização foram obtidos através da página eletrônica do Departamento de Informações e Informática do Sistema Único de Saúde (DATASUS), órgão responsável por coletar, processar e disseminar informações sobre saúde, referentes exclusivamente ao SUS. Provavelmente por problemas de subnotificação, as informações dos meses de julho de 2006 e janeiro, abril e maio de 2008 ficaram muito abaixo das restantes em todo o período 2004-2010, por isso preferiu-se excluí-las da análise, evitando assim um possível viés.

Optou-se por trabalhar com os dados de inalação/nebulização porque os dados de internação hospitalar por doença respiratória, como pneumonia, não trazem qualquer evidência de que a doença tenha se desencadeado na data da internação, uma vez que, as internações por tais causas no município são agendadas e não feitas em caráter de emergência, como ocorre com as doenças cardiovasculares. Ou seja, entre o paciente ser exposto às variações meteorológicas ou poluição atmosférica, ter a doença diagnosticada e ser internado existe um período de tempo difícil de se mensurar, que pode variar de semanas a meses. Assim, procurou-se trabalhar com as informações no nível da Atenção Primária em Saúde, ou seja, dados ambulatoriais (do SUS) que trariam uma informação mais imediata dos efeitos da qualidade do ar na saúde da população, uma vez que a inalação/nebulização de vapor de água ajuda a umidificar as vias aéreas e a reduzir os problemas respiratórios característicos das épocas de ar seco, sendo uma das primeiras medidas terapêuticas aplicadas quando ocorrem complicações respiratórias, seja por doenças pré-existentes como a asma, ou em problemas respiratórios eventuais, como gripes, resfriados e pneumonias.

### 3.2 Análise exploratória

O número mensal de inalações/nebulizações apresentadas em Ribeirão Preto no período de janeiro de 2004 a dezembro de 2010 é descrito na Figura 4 e, as séries de tempo para o volume total de precipitação por mês e médias mensais de temperatura mínima e  $MP_{10}$  são exibidas na Figura 5. Considerando todo o período estudado, observou-se uma média mensal de 3.777 inalações/ nebulizações, sendo a menor quantidade registrada em novembro de 2010 (936 procedimentos) e a maior em julho de 2004 (11.900 procedimentos). Nota-se, no gráfico, que o número de inalações/nebulizações tem decrescido nos últimos anos do período. Porém, percebe-se

uma tendência sazonal, com maiores ocorrências de procedimentos entre os meses de maio e outubro.

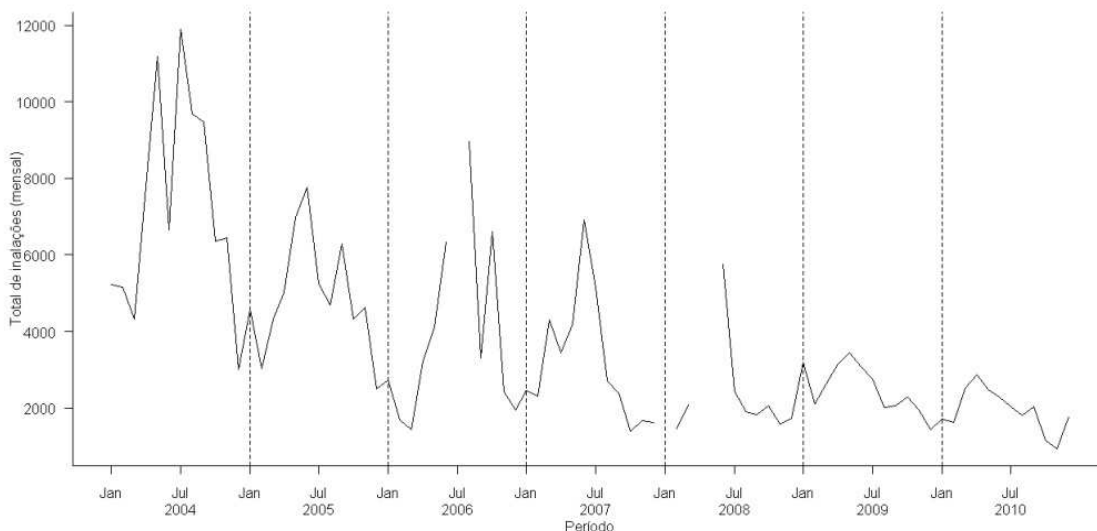


Figura 4. Série do total de inalações/nebulizações mensais no período.

As séries temporais exibidas na Figura 5 evidenciam o comportamento sazonal do volume total de precipitação por mês e médias mensais de temperatura mínima e  $MP_{10}$  durante o período estudado e sugerem uma associação com o número de inalações/nebulizações. O material particulado é liberado mais frequentemente nos meses intermediários do ano, quando também há baixos volumes de precipitação e menores temperaturas, períodos coincidentes aos de maiores quantidades de inalação. Da mesma forma, percebe-se um maior número de inalações nos períodos de outono e inverno (Tabela 1), épocas do ano em que as mudanças bruscas de temperatura e os baixos índices de umidade do ar por si só são fatores de risco para problemas respiratórios, agravados pela inalação de fumaça e poluentes liberados pelas queimas.

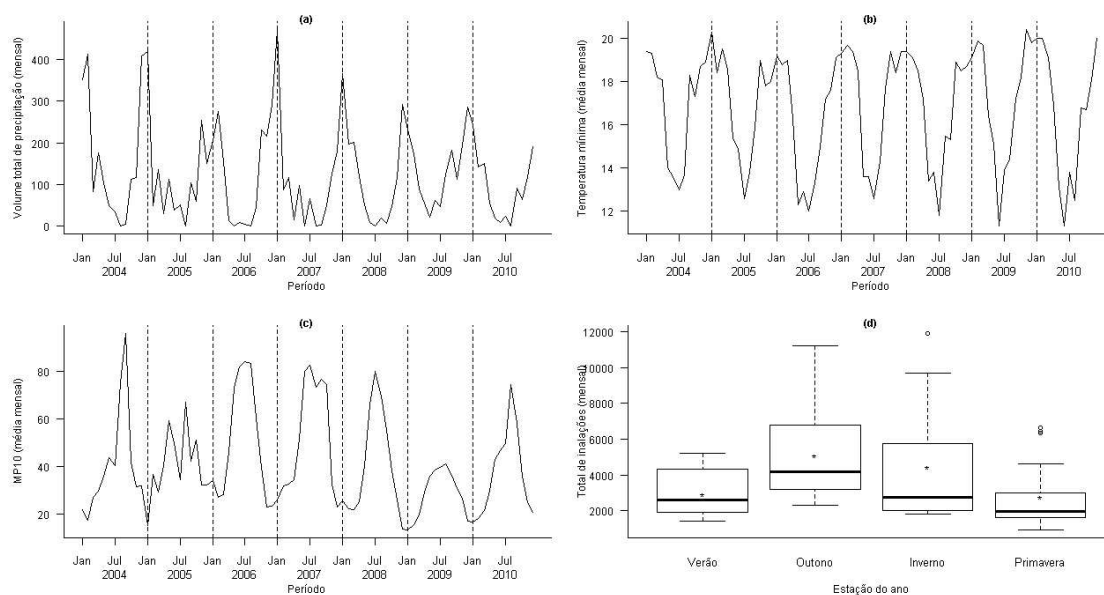


Figura 5. Descrições das variáveis climáticas e da inalação ao longo do tempo.

Nota: O “\*” no boxplot indica a média da quantidade de inalações/nebulizações correspondente à categoria.

Tabela 1. Estatísticas descritivas segundo estação do ano. Ribeirão Preto, 2004 a 2010.

Variável	Estação	Média	Desvio padrão	Mínimo	Mediana	Máximo
Número de inalações	Outono e inverno	4753,6	2798,7	1807	3453	11900
	Primavera e verão	2847,6	1548,2	936	2323	6621
MP <sub>10</sub>	Outono e inverno	54,7	19,2	25,4	49,7	95,8
	Primavera e verão	28	11	13,5	26,5	74,4
Temperatura mínima média	Outono e inverno	14,6	2,1	11,3	13,9	18,6
	Primavera e verão	18,9	0,8	16,7	19,1	20,4
Volume total de precipitação	Outono e inverno	44,4	48,3	0	27,4	181,2
	Primavera e verão	194,4	110,4	47,6	176,6	469,9

A média mensal de MP<sub>10</sub> observada no período em estudo, 41,3 µg/m<sup>3</sup>, está abaixo do valor referencial da Organização Mundial de Saúde (OMS) para médias de 24 horas, de 50 µg/m<sup>3</sup>. Porém, em vários meses do período estudado, a concentração do poluente ultrapassou essa medida de referência, atingindo seu maior valor médio (95,8 µg/m<sup>3</sup>) no mês de setembro de 2004. Observa-se na Tabela 1 que, no período, as médias mensais de MP<sub>10</sub> são maiores no outono e inverno (54,7 µg/m<sup>3</sup>), estações com menores registros mensais de precipitação.

Para descrever melhor a relação entre as variáveis e as inalações estão apresentados os gráficos de dispersão correspondentes, em que se percebem leves tendências de decréscimo do número de inalações/nebulizações conforme a temperatura

mínima e o volume de precipitação aumentam. Da mesma forma, parece que um aumento nos níveis de  $MP_{10}$  aumentam, por sua vez, os procedimentos inalatórios/de nebulização. Estas relações serão estudadas nos próximos capítulos com o ajuste de modelos apropriados para a análise.

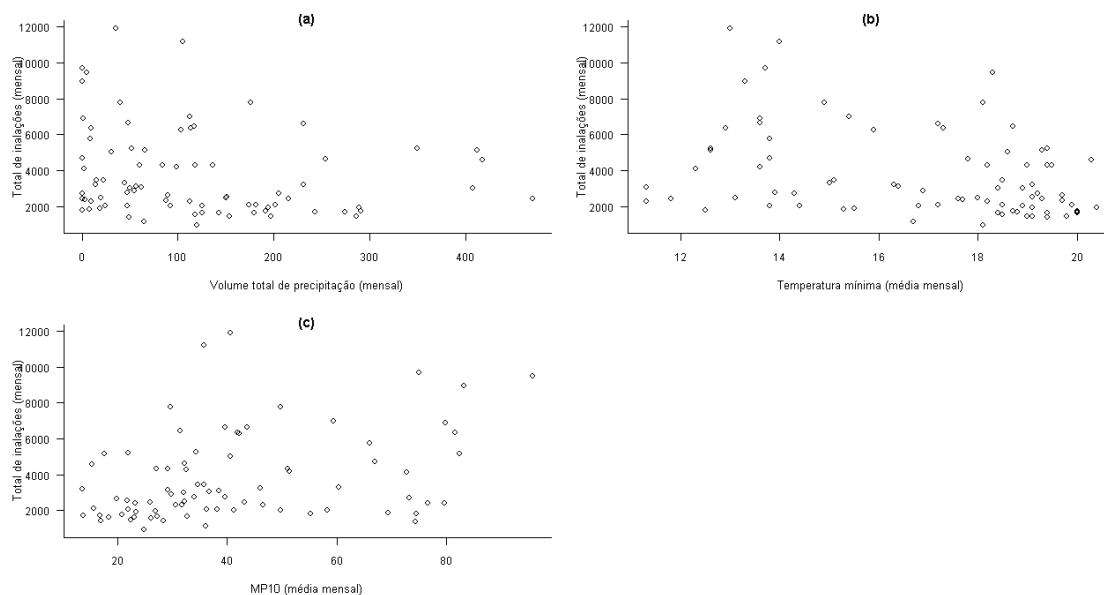


Figura 6. Gráficos de dispersão envolvendo as inalações e os valores de precipitação (a), temperatura mínima (b) e material particulado (c).

## 4 MODELAGEM ESTATÍSTICA

### 4.1 Metodologia Bayesiana

Na estatística, a informação sobre uma determinada quantidade desconhecida da população em estudo, por exemplo, um parâmetro  $\theta$ , é essencial. Assim, pode-se dizer que a idéia fundamental da estatística é tentar reduzir esse desconhecimento. Além disso, a magnitude da incerteza sobre  $\theta$  pode assumir graus variados. Sob o enfoque Bayesiano, estes diferentes graus de incerteza são representados por distribuições de probabilidade nos dados e nos parâmetros. Segundo Paulino (2002), o paradigma Bayesiano expõe que o que é desconhecido, no caso o parâmetro, é incerto e toda incerteza deve ser quantificada em termos probabilísticos. Sendo assim, é natural que os graus de conhecimento e incerteza a respeito de  $\theta$  variem de pesquisador para pesquisador. Este fato faz com que além das quantidades observáveis, também os parâmetros de um modelo estatístico sejam considerados quantidades aleatórias.

Diferentemente da Inferência Clássica, a Inferência Bayesiana não se utiliza apenas da amostra para obter informação sobre o parâmetro. Apesar de parte de essa informação provir dela (representada pela função de verossimilhança), outra parte se deve ao conhecimento prévio do pesquisador, que se dá através de uma distribuição de probabilidade para o parâmetro, conhecida como distribuição *a priori*.

A estimação do parâmetro é realizada com a utilização conjunta destas duas informações, através da fórmula de Bayes, e tem como resultado a distribuição *a posteriori* do parâmetro, que representa todo o conhecimento existente sobre o problema investigado. A distribuição *a posteriori* é o principal elemento de toda análise bayesiana de dados, pois é ela que possui toda a informação necessária para a realização de inferências, permitindo a análise de qualquer aspecto do parâmetro em estudo, como média, mediana, percentis, etc (ver por exemplo, CARLIN & LOUIS, 1996).

O surgimento e disponibilidade de vários algoritmos computacionais, além dos variados softwares estatísticos livres capazes de ajustarem modelos bayesianos, como o R (R Development Core Team, 2010) e o WinBugs (SPIEGELHALTER et al., 2000), facilitaram o grande crescimento do uso da metodologia Bayesiana em pesquisas nas últimas décadas.

Além da possibilidade de incorporação de informações prévias sobre o tema estudado (distribuição *a priori*), a Inferência Bayesiana possui a vantagem de ser mais

flexível quanto ao ajuste dos modelos, ao mesmo tempo em que proporciona maior facilidade de interpretação através do intervalo de credibilidade obtido diretamente dos quantis da distribuição *a posteriori* (PAULINO et al., 2002). Além disso, sob o enfoque bayesiano usualmente não é necessário o uso de testes de hipóteses (p-valores) para as inferências de interesse o que é comum usando a inferência clássica. Outro ponto importante: a metodologia bayesiana fornece melhores possibilidades de comparação do ajuste dos modelos propostos usando densidades preditivas.

#### 4.1.1 Fórmula de Bayes e inferência Bayesiana

A fundamentação da teoria de Inferência Bayesiana é baseada na fórmula de Bayes dada a seguir.

Sejam os eventos  $A_1, A_2, \dots, A_k$  formando uma sequência de eventos mutuamente exclusivos e exaustivos, criando uma partição do espaço amostral  $\Omega$ , isto é,  $\bigcup_{j=1}^k A_j = \Omega$

e  $A_i \cap A_j = \emptyset$  (conjunto vazio) para  $i \neq j$  tal que  $P\left(\bigcup_{j=1}^k A_j\right) = \sum_{j=1}^k P(A_j) = 1$ .

Assim, para qualquer outro evento  $B$  ( $B \subset \Omega$ ), temos,

$$P(A_i | B) = \frac{P(B | A_i)P(A_i)}{\sum_{j=1}^k P(B | A_j)P(A_j)} \quad (2.1)$$

para todo  $i$  variando de 1 até  $k$ .

A fórmula de Bayes (2.1) pode ser interpretada da seguinte forma: antes do conhecimento de qualquer informação sobre o evento  $A_i$ , atribui-se uma probabilidade a priori dada para  $A_i$  ( $P(A_i)$ ), atualizada a partir da ocorrência do evento  $B$ . Essa probabilidade atualizada, ou probabilidade condicional do evento  $A_i$  dado a ocorrência do evento  $B$  (probabilidade a posteriori), ou seja,  $P(A_i | B)$ , é dada pela fórmula de Bayes (2.1).

Assumiremos agora que temos um vetor de dados  $\mathbf{y} = (y_1, \dots, y_n)'$  e quantidades desconhecidas  $\theta$  representando os parâmetros de uma distribuição de probabilidade associada com a variável aleatória  $Y_i$  com valores observados  $y_i$ ,  $i = 1, \dots, n$ .

Suponha que o parâmetro  $\theta$  assume valores contínuos num dado intervalo. Considerando uma amostra aleatória  $\mathbf{y} = (y_1, \dots, y_n)'$  (os dados são independentes e identicamente distribuídos com uma distribuição conjunta dada pela densidade  $f(\mathbf{y} | \theta)$ , também definida como função de verossimilhança para  $\theta$ ), e uma distribuição a priori para  $\theta$  (dada por  $\pi(\theta)$ ), temos de (2.1), a distribuição a posteriori para  $\theta$  dado  $\mathbf{y}$ ,

$$\pi(\theta | \mathbf{y}) = \frac{f(\mathbf{y} | \theta)\pi(\theta)}{\int f(\mathbf{y} | \theta)\pi(\theta)d\theta} \quad (2.2)$$

em que a integral no denominador de (2.2) é definida no intervalo de variação de  $\theta$ .

Em geral, não é necessário o cálculo da integral no denominador e, assim, a *posteriori* é proporcional à priori multiplicada pela verossimilhança. No caso discreto, como da distribuição Poisson, substitui-se a integral do denominador por  $\sum_{\theta}$ .

Ao invés de se utilizar apenas o valor estimado do parâmetro, estimar um intervalo de valores no qual se espera que ele esteja, com um certo grau de confiança, é de interesse tanto da inferência clássica, quanto da Bayesiana. Na inferência clássica esse intervalo é chamado de intervalo de confiança. Considerando um nível de significância ( $\alpha$ ), por exemplo igual a 5%, a interpretação para um parâmetro  $\theta$  é a seguinte: se retirássemos da população um número grande de amostras de tamanho  $n$ , 95% destas amostras iriam gerar intervalos de confiança que contém o verdadeiro valor de  $\theta$ . Assim, em termos frequentistas, o parâmetro é considerado fixo e o intervalo de confiança é aleatório.

Diferente da inferência clássica, a construção e interpretação de um intervalo para um parâmetro de interesse sob enfoque Bayesiano é feita de maneira mais intuitiva, levando muitos cientistas e não-estatísticos a interpretarem intervalos de confiança no sentido dos intervalos de credibilidade bayesianos. Seja  $\theta$  um parâmetro unidimensional e sua distribuição *a posteriori*, unimodal. Uma forma de estimador por intervalo Bayesiano com probabilidade  $(1 - \alpha)$  será dado por  $(\theta_*, \theta^*)$ , para

$$\int_{-\infty}^{\theta_*} \pi(\theta | \mathbf{y})d\theta = \frac{\alpha}{2} \quad (2.3)$$

e

$$\int_{\theta^*}^{\infty} \pi(\theta | \mathbf{y})d\theta = \frac{\alpha}{2}$$



O intervalo  $(\theta_*, \theta^*)$  é chamado intervalo de credibilidade para  $\theta$  com probabilidade  $(1 - \alpha)$ , ou seja, a probabilidade de que  $\theta$  esteja no intervalo  $(\theta_*, \theta^*)$  é de  $1 - \alpha$ .

Nos casos em que a distribuição a posteriori é assimétrica, uma outra forma de se encontrar intervalos de credibilidade para o parâmetro, é encontrar intervalos que minimizem  $\theta_* - \theta^*$ , mas com uma cobertura  $1 - \alpha$ .

#### 4.1.2 Distribuições a priori não-informativas

A escolha da distribuição a priori é um importante tópico da Inferência Bayesiana. É através dela que a opinião de um especialista pode ser incorporada na forma de uma distribuição de probabilidade, já que permite construir, antes da realização da pesquisa, uma distribuição que represente uma proposição sobre o conhecimento probabilístico que se tem sobre o parâmetro a ser estimado.

A informação a priori sobre o parâmetro é, então, estabelecida por uma densidade de probabilidade, de acordo com a experiência e o conhecimento do pesquisador sobre os parâmetros de interesse. Dessa forma, as distribuições a priori podem ser classificadas como sendo informativas e não informativas, cuja principal diferença está na existência de algum conhecimento a respeito do parâmetro.

Quando não se tem um conhecimento prévio sobre o tema, é necessário adotar alguma forma de análise que consiga captar esta noção de priori e que possua um efeito mínimo sobre os dados.

Vários são os métodos existentes para se escolher uma distribuição a priori, porém nas situações em que não existem informações a priori substanciais, o uso de distribuições a priori não informativas é o mais indicado, como por exemplo, a priori de Jeffreys (JEFFREYS, 1998).

### 4.1.3 Métodos de simulação para obtenção dos sumários *a posteriori*

Na obtenção de sumários *a posteriori* de interesse, geralmente não é possível encontrar uma solução analítica, fato comum quando se tem um vetor  $\theta$  de parâmetros. Como solução deste problema, várias alternativas foram introduzidas na literatura. Um exemplo é o uso da aproximação de Laplace que utiliza uma aproximação normal para a distribuição *a posteriori* de interesse mas esse método muitas vezes pode não ser o método mais indicado, já que dependendo da complexidade das integrais, o tempo gasto nessas resoluções pode ser muito grande, especialmente quando a dimensão do vetor de parâmetros for grande (em geral maior ou igual à 5) pois envolve a determinação de máximos e derivadas até segunda ordem de expressões matemáticas complicadas. Uma alternativa que tornou o uso de métodos bayesianos muito popular foi o aparecimento de várias técnicas de simulação de Monte Carlo. De forma geral, o desenvolvimento computacional propiciou que os métodos de simulação pudessem ser utilizados de forma intensiva para a obtenção dos sumários *a posteriori* de interesse, sendo o método de Monte Carlo via função por importância e o de Monte Carlo via Cadeias de Markov (métodos MCMC) muito utilizados pelos pesquisadores das mais diversas áreas.

#### 4.1.3.1 Método de Monte Carlo

Uma distribuição *a posteriori* pode ser resumida como sendo esperanças de funções particulares do parâmetro  $\theta$ , isto é

$$E[g(\theta) | \mathbf{y}] = \int g(\theta) \pi(\theta | \mathbf{y}) d\theta \quad (2.4)$$

ou distribuições *a posteriori* marginais quando  $\theta$  for multidimensional, ou seja

$$p(\theta_1 | x) = \int p(\theta | x) d\theta_2 \quad (2.5)$$

onde  $\theta = (\theta_1, \theta_2)$ .

Assim, o problema geral consistirá em calcular tais valores esperados segundo a distribuição *a posteriori* de  $\theta$ . Alguns exemplos são,

1. Se  $g(\theta) = \theta$ , então a média *a posteriori* será  $\mu = E(\theta | x)$ .
2. Quando  $g(\theta) = (\theta - \mu)^2$ , então a variância *a posteriori* será  $\sigma^2 = \text{var}(\theta) = E((\theta - \mu)^2 | x)$ .

Pelo método de Monte Carlo ordinário, simular uma amostra  $\theta_1, \dots, \theta_k$  da distribuição a posteriori  $\pi(\theta | \mathbf{y})$ . Assim, (2.4) pode ser aproximado por,

$$\hat{E}[g(\theta) | \mathbf{y}] = \frac{1}{n} \sum_{i=1}^n g(\theta_i) \quad (2.6)$$

Uma vez que as gerações são independentes, pela Lei Forte dos Grandes Números segue que  $\hat{E}[g(\theta) | \mathbf{y}]$  converge quase certamente para  $E[g(\theta) | \mathbf{y}]$ .

A precisão dessa aproximação pode ser medida pelo erro-padrão de Monte Carlo, dado por,

$$\frac{1}{\sqrt{n(n-1)}} \left\{ \sum_{i=1}^n \left[ g(\theta_i) - \frac{1}{n} \sum_{i=1}^n g(\theta_i) \right]^2 \right\}^{\frac{1}{2}} \quad (2.7)$$

Intervalos de credibilidade para  $\theta$  podem ser obtidos usando o método de Monte Carlo ordinário. Ordenando a amostra simulada de  $\pi(\theta | \mathbf{y})$ :  $\theta_{(1)} < \theta_{(2)} < \dots < \theta_{(n)}$ , um intervalo de credibilidade  $100(1-\alpha)\%$  para  $\theta$  será dado por,

$$R_c(\alpha) = \left[ \theta_{\frac{(1-\alpha)}{2}}; \theta_{\frac{(1+\alpha)}{2}} \right] \quad (2.8)$$

cujos extremos definem quantis de probabilidade *a posteriori*  $\frac{(1-\alpha)}{2}$  e  $\frac{(1+\alpha)}{2}$  de  $\theta$ .

$$\text{Ou seja, } P \left[ \theta \leq \theta_{\frac{(1-\alpha)}{2}} | \mathbf{y} \right] = \frac{(1-\alpha)}{2} \text{ e } P \left[ \theta \leq \theta_{\frac{(1+\alpha)}{2}} | \mathbf{y} \right] = 1 - \frac{(1-\alpha)}{2} = \frac{(1+\alpha)}{2}.$$

Quando não é possível simular uma amostra diretamente da distribuição *a posteriori*  $\pi(\theta | \mathbf{y})$ , uma amostragem via função de importância, simulando uma amostra de uma distribuição semelhante à *posteriori*, é uma alternativa muito utilizada.

Seja  $p(\theta)$  uma densidade da qual seja fácil simular amostras e que aproxime a distribuição  $\pi(\theta | \mathbf{y})$ . Podemos escrever (2.4) na forma,

$$\begin{aligned}
\int g(\theta)\pi(\theta|\mathbf{y})d\theta &= \frac{\int g(\theta)f(\mathbf{y}|\theta)\pi(\theta)d\theta}{\int f(\mathbf{y}|\theta)\pi(\theta)d\theta} \\
&= \frac{\int g(\theta)\frac{f(\mathbf{y}|\theta)\pi(\theta)}{p(\theta)}p(\theta)d\theta}{\int \frac{f(\mathbf{y}|\theta)\pi(\theta)}{p(\theta)}p(\theta)d\theta} \\
&= \frac{\int g(\theta)w(\theta)p(\theta)d\theta}{\int w(\theta)p(\theta)d\theta}
\end{aligned} \tag{2.9}$$

em que  $w(\theta) = \frac{f(\mathbf{y}|\theta)\pi(\theta)}{p(\theta)}$ ,  $f(\mathbf{y}|\theta)$  é a função de verossimilhança para  $\theta$  e  $\pi(\theta)$  é a distribuição a priori para  $\theta$ .

Obtendo uma amostra  $\theta_1, \dots, \theta_n$  de  $p(\theta)$ , encontramos a aproximação de Monte Carlo para  $E[g(\theta)|\mathbf{y}]$  dada por,

$$\hat{E}[g(\theta)|\mathbf{y}] = \frac{1}{\sum_{i=1}^n w_i} \sum_{i=1}^n w_i g(\theta_i) \tag{2.10}$$

em que  $w_i = \frac{f(\mathbf{y}|\theta_i)\pi(\theta_i)}{p(\theta_i)}$ .

#### 4.1.3.2 Método de Monte Carlo em cadeias de Markov

Encontrar uma densidade de importância que seja uma boa aproximação da posteriori e ao mesmo tempo, fácil de ser amostrada, pode ser difícil em muitos casos (ver por exemplo, VISMARA, 2007). Nesse contexto, os métodos de Monte Carlo via cadeias de Markov são uma alternativa aos métodos não iterativos quando se trabalha com problemas de maior complexidade. A idéia ainda é obter uma amostra da distribuição a posteriori e calcular estimativas de características desta distribuição, porém aqui serão utilizadas técnicas de simulação iterativa, baseadas em cadeias de Markov (dado o estado presente, o estado futuro é independente do passado), e assim os valores gerados não serão mais independentes.

Assim, supor que não podemos gerar diretamente uma amostra de uma distribuição a posteriori  $\pi(\boldsymbol{\theta}|\mathbf{y})$ ,  $\boldsymbol{\theta} \in R^k$ . Entretanto, supor que podemos construir uma cadeia de Markov com espaço de estados no espaço paramétrico  $\Theta$  (conjunto de todos os valores possíveis de  $\boldsymbol{\theta}$ ) que é simples para simular e cuja distribuição de equilíbrio

seja dada por  $\pi(\boldsymbol{\theta} | \mathbf{y})$ . Se temos muitas simulações dessa cadeia, os valores simulados podem ser usados para sumarizar características da posteriori  $\pi(\boldsymbol{\theta} | \mathbf{y})$ .

Segundo Besag (1994), se a distribuição conjunta a posteriori  $\pi(\boldsymbol{\theta} | \mathbf{y})$  for positiva em  $\Theta_1 \times \Theta_2 \times \dots \times \Theta_k$ , com  $\Theta_i$  sendo suporte para a distribuição de  $\theta_i$ ,  $i = 1, \dots, k$ , então a distribuição a posteriori  $\pi(\boldsymbol{\theta} | \mathbf{y})$  é unicamente determinada pelas distribuições condicionais completas  $\pi(\theta_i | \mathbf{y}, \boldsymbol{\theta}_{(i)})$  para  $i = 1, \dots, k$  em que  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$  e  $\boldsymbol{\theta}_{(i)}$  é o vetor de todos os componentes de  $\boldsymbol{\theta}$  exceto  $\theta_i$ , isto é,  $\boldsymbol{\theta}_{(i)} = \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k$ .

Sob algumas condições de regularidade é fácil ver que os resultados simulados da cadeia com distribuição de equilíbrio  $\pi(\boldsymbol{\theta} | \mathbf{y})$  podem ser supostos com uma amostra aleatória de  $\pi(\boldsymbol{\theta} | \mathbf{y})$ .

Se  $\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(t)}, \dots$  é uma realização de uma cadeia, temos,

$$\boldsymbol{\theta}^{(t)} \xrightarrow{D} \boldsymbol{\theta} \sim \pi(\boldsymbol{\theta} | \mathbf{y}) \quad (2.11)$$

em que o símbolo “ $D$ ” significa convergência em distribuição.

Da mesma forma para estimar o valor esperado de  $g(\boldsymbol{\theta})$  com respeito à  $\pi(\boldsymbol{\theta} | \mathbf{y})$ , isto é,

$$E[g(\boldsymbol{\theta}) | \mathbf{y}] = \int g(\boldsymbol{\theta}) \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta} \quad (2.12)$$

observamos que

$$\frac{1}{t} \sum_{i=1}^t g[\boldsymbol{\theta}^{(i)}] \xrightarrow{q.c.} E[g(\boldsymbol{\theta}) | \mathbf{y}] \quad (2.13)$$

(“ $q.c.$ ”: convergência quase certa).

Na prática,  $\boldsymbol{\theta}^{(i)}$  pode estar correlacionado entre si, mas poderíamos considerar espaços adequados entre os  $\boldsymbol{\theta}^{(i)}$  gerados para garantir uma amostra aleatória de  $\pi(\boldsymbol{\theta} | \mathbf{y})$ .

É necessário gerarmos cadeias de Markov cuja distribuição de equilíbrio convirja para  $\pi(\boldsymbol{\theta} | \mathbf{y})$ , para isso podemos utilizar os algoritmos Metropolis-Hastings e o amostrador de Gibbs (ver por exemplo, GELFAND e SMITH, 1990; CHIB e GREENBERG, 1995).

## 4.1.3.2.1 Amostrador de Gibbs

Supor que estamos interessados em obter inferências da distribuição a posteriori conjunta,  $\pi(\boldsymbol{\theta} | \mathbf{y})$ ,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ . Para isso simulamos quantidades aleatórias de distribuições condicionais completas  $\pi(\theta_i | \mathbf{y}, \boldsymbol{\theta}_{(i)})$ ,  $i = 1, \dots, k$ , que produzem uma cadeia de Markov.

Observar que  $\pi(\theta_i | \mathbf{y}, \boldsymbol{\theta}_{(i)})$  são facilmente identificadas como funções de  $\theta_i$  por inspeção da forma de  $\pi(\boldsymbol{\theta} | \mathbf{y})$ , a distribuição a posteriori para  $\boldsymbol{\theta}$  dado  $\mathbf{y}$ . (GAMERMAN, 1997).

Atribuindo um conjunto arbitrário de valores iniciais  $\boldsymbol{\theta}_1^{(0)}, \boldsymbol{\theta}_2^{(0)}, \dots, \boldsymbol{\theta}_k^{(0)}$  para o vetor de parâmetros  $\boldsymbol{\theta}$ , podemos escrever o algoritmo da forma:

$$(i) \text{ Gerar } \theta_1^{(1)} \text{ de } \pi(\theta_1 | \mathbf{y}, \theta_2^{(0)}, \dots, \theta_k^{(0)}); \quad (2.14)$$

$$(ii) \text{ Gerar } \theta_2^{(1)} \text{ de } \pi(\theta_2 | \mathbf{y}, \theta_1^{(1)}, \theta_3^{(0)}, \dots, \theta_k^{(0)});$$

$$(iii) \text{ Gerar } \theta_3^{(1)} \text{ de } \pi(\theta_3 | \mathbf{y}, \theta_1^{(1)}, \theta_2^{(1)}, \theta_4^{(0)}, \dots, \theta_k^{(0)});$$

.

.

.

$$(k) \text{ Gerar } \theta_k^{(1)} \text{ de } \pi(\theta_k | \mathbf{y}, \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_{k-1}^{(1)}).$$

Então, o próximo passo será substituir os valores iniciais com uma nova realização  $\boldsymbol{\theta}^{(1)} = \theta_1^{(1)}, \theta_2^{(1)}, \dots, \theta_k^{(1)}$  de  $\boldsymbol{\theta}$  e repetir o processo acima.

Para  $t$  suficientemente grande, o valor  $\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_k^{(t)}$  converge para um valor da quantidade aleatória com distribuição  $\pi(\boldsymbol{\theta} | \mathbf{y})$  (GEMAN e GEMAN, 1984). Além disso,  $\theta_j^{(t)}$  pode ser considerado como uma observação simulada da distribuição a posteriori marginal  $\pi(\theta_j | \mathbf{y})$ ,  $j = 1, 2, \dots, k$ .

Replicando o processo acima  $B$  vezes obtemos  $B$  vetores  $\boldsymbol{\theta}_{1g}^{(t)}, \boldsymbol{\theta}_{2g}^{(t)}, \dots, \boldsymbol{\theta}_{kg}^{(t)}$ ;  $g = 1, 2, \dots, B$ .

Da convergência do amostrador de Gibbs, qualquer característica da densidade a posteriori marginal  $\pi(\theta_j | \mathbf{y})$  pode ser obtida, como esperança, variância, entre outras.

Em particular, se  $\pi(\theta_j | \boldsymbol{\theta}_{(j)}, \mathbf{y})$  é dada em forma fechada, então

$$\hat{\pi}[g_j | \mathbf{y}] = \frac{1}{B} \sum_{g=1}^B \pi(\theta_{(j)}^g, y) \quad (2.15)$$

em que  $j = 1, \dots, k$ .

#### 4.1.3.2.2 Algoritmo de Metropolis-Hastings

O amostrador de Gibbs seleciona amostras de todas as densidades a posteriori condicionais de  $\pi(\boldsymbol{\theta} | \mathbf{y})$ . Quando usamos uma priori conjugada, as distribuições condicionais em geral são de forma conhecida, como, por exemplo, as distribuições normal, gama, Poisson, Beta, etc, e a simulação de amostras dessas distribuições são disponíveis em diversos softwares, como o R e o WinBugs.

Quando as distribuições condicionais não são facilmente identificadas, devemos usar o algoritmo de Metropolis-Hastings ou métodos de amostragem por importância.

Supor que desejamos simular amostras de uma densidade não-regular  $\pi(\theta_i | \boldsymbol{\theta}_{(i)}, \mathbf{y})$ , ou simplesmente  $\pi(\theta_i | \boldsymbol{\theta}_{(i)})$ , em que  $\boldsymbol{\theta}_{(i)} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$ .

Definir o núcleo de transição  $q(\theta, \beta)$  da distribuição  $p(\theta)$  que representa  $\pi(\theta_i | \boldsymbol{\theta}_{(i)})$  e que transforma  $\theta$  em  $\beta$ . Se  $\theta$  é uma variável real com amplitude em toda reta  $R$ , podemos construir  $q$  tal que  $\beta \leftarrow \theta + \sigma z$ , com  $Z \sim N(0, \sigma^2)$ , em que  $\sigma^2$  reflete a variância condicional de  $\theta$  em  $p(\theta)$ .

Se  $\theta$  é limitado com amplitude  $(a, b)$  usar uma transformação que leva  $(a, b)$  em  $(-\infty, \infty)$  e daí usar o núcleo de transição  $q$  e aplicar o algoritmo de Metropolis para a densidade da variável transformada. O algoritmo de Metropolis é dado por:

- (i) Iniciar com um valor  $\theta^{(0)}$  e indicador de estágio,  $j = 0$ ;
- (ii) Gerar um ponto  $\beta$  do núcleo de transição  $q(\theta^{(j)}, \beta)$ ;
- (iii) Atualizar  $\theta^{(j)}$  por  $\theta^{(j+1)} = \beta$  com probabilidade,

$$p = \min \left\{ 1, \frac{p(\beta) q[\theta^{(j)}, \beta]}{p[\theta^{(j)}] q[\beta, \theta^{(j)}]} \right\} \quad (2.16)$$

Ficar com  $\theta^{(j)}$  com probabilidade  $1 - p$ , caso contrário,

- (iv) Repetir os estágios (ii) e (iii) até conseguir uma distribuição estacionária.

Observar que:

(a) O algoritmo de Metropolis Hastings é especificado pela densidade candidata para geração  $q(x,y)$ ;

(b) Se um valor candidato é rejeitado, o valor atual é considerado na próxima etapa;

(c) O cálculo de  $p$  em (2.15) não depende da constante normalizadora;

(d) Se a densidade candidata para geração das amostras é simétrica, isto é,  $q(x,y)=q(y,x)$ , a probabilidade de movimento se reduz à  $\frac{p(\beta)}{p[\theta^{(j)}]}$ . Assim, se  $p(\beta) > p[\theta^{(j)}]$ , a cadeia se move para  $\beta$ ; em caso contrário, ela se move para  $\beta$  com probabilidade  $\frac{p(\beta)}{p[\theta^{(j)}]}$ . Em outras palavras: um salto na direção "ascendente" é sempre aceito; um salto na direção "descendente" é aceito com uma dada probabilidade.

#### 4.1.4 Formulação de modelos

Na análise estatística dos dados, o número de procedimentos mensais de inalação/nebulização a cada mês ( $N_i$ , sendo o índice  $i$  associado ao mês respectivo) será considerado uma variável dependente e as variáveis de controle serão definidas como a estação do ano (codificada como variável indicadora – SUITS, 1957), temperatura mínima mensal e volume mensal de precipitação.

Para estimar a associação existente entre o número mensal de procedimentos de inalação/nebulização e material particulado será utilizado um modelo de regressão de Poisson (MCCULLAGH & NELDER, 1989) sob um enfoque bayesiano (SPIEGELHALTER, 2000). O uso da distribuição de Poisson justifica-se pela variável dependente ( $N_i$ ) ser um evento de contagem. O modelo é então definido por

$$P(N_i = n_i | \lambda_i) = \frac{e^{-\lambda_i} \lambda_i^{n_i}}{n_i!}, n_i = 0,1,2,\dots, i = 1,2,\dots,k, \quad (2.17)$$

onde  $n_i$  denota o número de inalações/nebulizações no  $i$ -ésimo mês e  $k$  é o número de meses contido no período em estudo.

Associado a cada mês, será considerada a presença das seguintes covariáveis: material particulado ( $X_{1i}$ ), temperatura mínima ( $X_{2i}$ ), precipitação ( $X_{3i}$ ) e as variáveis binárias indicadoras (SUITS, 1957) para outono, inverno e primavera, respectivamente denotadas por  $X_{4i}$ ,  $X_{5i}$ , e  $X_{6i}$  (verão é então considerada uma classe de referência). Para a



ligação das variáveis com o parâmetro de locação  $\lambda_i$  da distribuição de Poisson, será inicialmente considerado o modelo de regressão,

$$\lambda_i = \exp\{\beta_0 + \beta_1(X_{1i} - \bar{X}_1) + \beta_2(X_{2i} - \bar{X}_2) + \beta_3(X_{3i} - \bar{X}_3) + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i}\} \quad (2.18)$$

em que  $\bar{X}_j$  é a média da covariável (contínua)  $X_j$ ,  $j = 1, 2, 3$  e  $i = 1, 2, \dots, k$ .

O modelo pode ainda ser reescrito na forma

$$\log(\lambda_i) = \beta_0 + \beta_1(X_{1i} - \bar{X}_1) + \beta_2(X_{2i} - \bar{X}_2) + \beta_3(X_{3i} - \bar{X}_3) + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i}. \quad (2.19)$$

Este modelo é o que denotaremos por Modelo 1. Para uma análise Bayesiana, assumiremos que  $\beta_l, l = 1, \dots, 6$  são independentes e consideraremos para eles distribuições a priori normais:

$$\begin{aligned} \beta_0 &\sim N(a, b); a, b \text{ conhecidos} \\ \beta_1, \dots, \beta_6 &\sim N(c, d); c, d \text{ conhecidos} \end{aligned} \quad (2.20)$$

Em cada mês existe uma taxa própria de procedimentos de inalação/nebulização que depende de fatores externos, difíceis de serem mensurados. Por este motivo, incluiremos um efeito aleatório, responsável por capturar a variabilidade extra-Poisson (overdispersion) e, conseqüentemente, proporcionar uma melhora no ajuste do modelo.

Assim, serão ajustados outros dois modelos, incorporando um efeito aleatório, que denotaremos por  $w_i$ .

O modelo que denotaremos por Modelo 2 modificará a regressão dada em (2.18) para

$$\lambda_i = \eta_i e^{w_i} \quad (2.21)$$

em que  $\eta_i = \exp\{\beta_0 + \beta_1(X_{1i} - \bar{X}_1) + \beta_2(X_{2i} - \bar{X}_2) + \beta_3(X_{3i} - \bar{X}_3) + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i}\}$  e no qual  $w_i$  segue uma distribuição normal com variância desconhecida,

$$w_i \stackrel{iid}{\sim} N(0, \sigma^2). \quad (2.22)$$

Assumindo o modelo definido por (2.19), a função de verossimilhança para  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1)^T$ , dado  $N_i, w_i$  e as covariáveis  $X_{1i}, X_{2i}, \dots, X_{6i}$ ,  $i = 1, 2, \dots, k$ , é dada por,

$$L(\boldsymbol{\theta}) = \prod_{i=1}^n \frac{e^{-\lambda_i} \lambda_i^{n_i}}{n_i!} \quad (2.23)$$

em que os efeitos aleatórios  $w_i$  tem a distribuição especificada em (2.22).

Como  $N_i$  segue uma distribuição de Poisson, tem-se que  $E(N_i | \lambda_i) = \text{Var}(N_i | \lambda_i) = \lambda_i$ . Isto é,

$$E(N_i | \beta_l, w_i, x_i) = \text{Var}(N_i | \beta_l, w_i, x_i) = \exp(\beta_l x_i + w_i). \quad (2.24)$$

De  $E(N_i | \beta_l, x_i) = E[E(N_i | \beta_l, w_i, x_i)]$ , tem-se por (2.24) que  $E(N_i | \beta_l, x_i) = e^{\beta_l x_i} E(e^{w_i})$ .

Como  $w_i$  são normalmente distribuídos (2.22),  $e^{w_i}$  segue uma distribuição log-normal com média  $E(e^{w_i}) = e^{\sigma^2/2}$  e variância  $\text{Var}(e^{w_i}) = (e^{\sigma^2} - 1)e^{\sigma^2}$ . Ou seja,

$$E(N_i | \beta_l, x_i) = e^{\beta_l x_i} e^{\sigma^2/2}. \quad (2.25)$$

De  $\text{Var}(N_i | \beta_l, x_i) = \text{Var}[E(N_i | \beta_l, w_i, x_i)] + E[\text{Var}(N_i | \beta_l, w_i, x_i)]$ , tem-se por (2.24) que  $\text{Var}(N_i | \beta_l, x_i) = e^{2\beta_l x_i} \text{Var}(e^{w_i}) + e^{\beta_l x_i} E(e^{w_i})$ , isto é,

$$\text{Var}(N_i | \beta_l, x_i) = e^{2\beta_l x_i} (e^{\sigma^2} - 1)e^{\sigma^2} + e^{\beta_l x_i} e^{\sigma^2/2}. \quad (2.26)$$

De (2.25) e (2.26), é possível observar que a média e a variância de  $N_i$  dado  $\beta_l$  e  $x_i$ ,  $l=1, \dots, 6$ ;  $i=1, \dots, k$ , são diferentes, ou seja, tem-se a presença da variabilidade extra-Poisson dada pelo termo  $e^{2\beta_l x_i} (e^{\sigma^2} - 1)e^{\sigma^2}$ , incorporada ao modelo (2.21).

Para uma análise Bayesiana hierárquica, vamos considerar dois estágios: para o primeiro estágio, assumiremos distribuições a priori normais para os parâmetros  $\beta_0, \beta_1, \dots, \beta_5$  e  $\beta_6$ , já definidas em (2.20) para o Modelo 1.

Para o segundo estágio da análise Bayesiana hierárquica, no qual  $w_i \stackrel{iid}{\sim} N(0, \sigma^2)$ , assumiremos,

$$\sigma^2 \sim \text{Gama}(f, f); f \text{ conhecido}. \quad (2.27)$$

No Modelo 3, o modelo de regressão será definido por

$$\lambda_i = w_i \eta_i. \quad (2.28)$$

Neste caso,  $\beta_0, \beta_1, \dots, \beta_5$  e  $\beta_6$  assumirão as mesma prioris dos modelos anteriores e  $w_i$  assumirá uma distribuição gama com hiperparâmetros conhecidos:

$$w_i \stackrel{iid}{\sim} \text{Gama}(f, f) \quad (2.29)$$

Observa-se que  $E(w_i)=1$  e  $\text{Var}(w_i)=\frac{1}{f}$ .

Essa estrutura de fragilidade “ $w_i$ ” está relacionada aos modelos aditivos gama, introduzidos na literatura por Korsgaard e Andersen (1998) e Li (2002).

De  $E(N_i | \beta_l, x_i) = E[E(N_i | \beta_l, w_i, x_i)]$ , tem-se que  $E(N_i | \beta_l, x_i) = e^{\beta_l x_i} E(w_i) = e^{\beta_l x_i}$ , sendo  $E(w_i)=1$  para  $i=1, \dots, k$ .

Do resultado,  $\text{Var}(N_i | \beta_l, x_i) = \text{Var}[E(N_i | \beta_l, w_i, x_i)] + E[\text{Var}(N_i | \beta_l, w_i, x_i)]$ , tem-se que  $\text{Var}(N_i | \beta_l, x_i) = \frac{1}{f} e^{2\beta_l x_i} \text{Var}(e^{w_i}) + e^{\beta_l x_i} E(w_i)$ , isto é,  $\text{Var}(N_i | \beta_l, x_i) = \frac{1}{f} e^{2\beta_l x_i} + e^{\beta_l x_i}$ .

Para melhor interpretação das estimativas obtidas pelos modelos, será estimado o risco relativo através da função exponencial de cada parâmetro do vetor  $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$  e  $\beta_6)$  obtidos do modelo de regressão, com seus respectivos intervalos de credibilidade 95% (IC95%).

Arbex (2001) analisou a relação do peso do sedimento do material particulado sobre o número de inalações no município de Araraquara (SP). Em uma de suas análises, Arbex utilizou categorias do peso, como uma tentativa de verificar se a relação entre o número de inalações e o peso do sedimento era dose-dependente. Por se tratar de uma forma interessante de análise e pouco vista na literatura, aqui reproduziremos sua metodologia. Assim, ajustaremos os mesmos modelos descritos anteriormente, considerando quatro categorias das concentrações de material particulado em lugar da variável contínua:

1ª categoria (MP<sub>10\_a</sub>): inclui os meses em que a concentração do poluente assumiu os valores 25% mais baixos de todo o período;

2ª categoria (MP<sub>10\_b</sub>): inclui os meses em que a concentração do poluente assumiu os valores entre 26 e 50%;

3ª categoria (MP<sub>10\_c</sub>): inclui os meses em que a concentração do poluente assumiu os valores entre 51 e 75%;

4ª categoria (MP<sub>10\_d</sub>): inclui os meses em que a concentração do poluente assumiu os valores 25% mais altos de todo o período.

Dessa forma, as concentrações do material particulado serão categorizadas como

$$\begin{cases} \text{MP}_{10\_a} : \leq 26,0 \\ \text{MP}_{10\_b} : \text{entre } 26,1 \text{ e } 35,8 \\ \text{MP}_{10\_c} : \text{entre } 35,9 \text{ e } 51,1 \\ \text{MP}_{10\_d} : > 51,1 \end{cases}$$

Pela figura a seguir (Figura 7), vemos que a relação entre o poluente e as inalações/nebulizações fica um pouco mais clara daquela que o gráfico de dispersão entre os dois tinha sugerido anteriormente, já que percebemos que o aumento dos valores do poluente, indo da primeira categoria para a terceira, faz com que as quantidades de inalação/nebulização também aumentem.

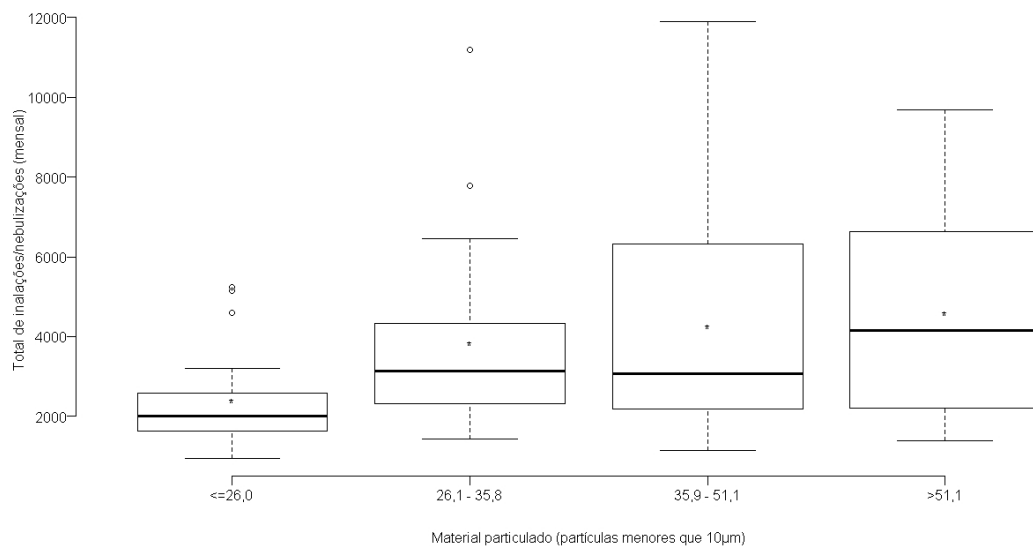


Figura 7. Boxplots das inalações segundo categorias do material particulado.

Nota: O “\*” indica a média da quantidade de inalações/nebulizações correspondente à categoria.

Para o ajuste dos modelos, será necessária a criação de variáveis indicadoras (SUITS, 1957). Neste caso, devido a um problema de convergência dos modelos optamos por criar duas novas variáveis, da seguinte forma:

Tabela 2. Criação das variáveis indicadoras através dos valores de MP<sub>10</sub> categorizados.

Categoria do material particulado	Variáveis	
	X <sub>1i</sub>	X <sub>2i</sub>
MP <sub>10_a</sub>	0	0
MP <sub>10_b</sub>	1	0
MP <sub>10_c</sub>	0	1
MP <sub>10_d</sub>	1	1

Com a mudança, o Modelo 1, e conseqüentemente os demais, com o acréscimo do  $w_i$ , será denotado por

$$\log(\lambda_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{3i} - \bar{X}_3) + \beta_4 (X_{4i} - \bar{X}_4) + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i}, \quad (2.30)$$

onde as variáveis temperatura mínima, precipitação, e as variáveis indicadoras para outono, inverno e primavera serão denotadas por  $X_{3i}, X_{4i}, X_{5i}, X_{6i}$  e  $X_{7i}$ , respectivamente.

Neste modelo,

$X_{1i}$  = indicadora de ( $X_{1i} = 1$  se a concentração de material particulado do  $i$ -ésimo mês corresponder às categorias MP<sub>10\_b</sub> ou MP<sub>10\_d</sub>;  $X_{1i} = 0$ , se corresponder às categorias MP<sub>10\_a</sub> ou MP<sub>10\_c</sub>);

$X_{2i}$  = indicadora de ( $X_{2i} = 1$  se a concentração de material particulado do  $i$ -ésimo mês corresponder às categorias MP<sub>10\_c</sub> ou MP<sub>10\_d</sub>;  $X_{2i} = 0$ , se corresponder às categorias MP<sub>10\_a</sub> ou MP<sub>10\_b</sub>).

Suponha que as médias mensais de inalação/nebulização sejam dadas por  $\mu_0$  quando as concentrações do poluente estiverem na categoria MP<sub>10\_a</sub>;  $\mu_1$  quando estiverem na categoria MP<sub>10\_b</sub>;  $\mu_2$  quando estiverem na categoria MP<sub>10\_c</sub>; e  $\mu_3$  quando estiverem na categoria MP<sub>10\_d</sub>. Do modelo dado em (2.30) temos que

$$\mu_0 = E(N_i | MP_{10_a}) = \exp(\beta_0);$$

$$\mu_1 = E(N_i | MP_{10_b}) = \exp(\beta_0 + \beta_1);$$

$$\mu_2 = E(N_i | MP_{10_c}) = \exp(\beta_0 + \beta_2);$$

$$\mu_3 = E(N_i | MP_{10\_d}) = \exp(\beta_0 + \beta_1 + \beta_2).$$

Note que a partir dessas estimativas, as diferenças entre as médias de inalações/nebulizações na ocorrência de cada categoria do poluente são fáceis de serem obtidas.

Como para os modelos anteriores, serão obtidos os riscos relativos. Neste caso, relacionaremos as três últimas categorias com a primeira, que será tomada como sendo de referência. Dessa forma, teremos que os riscos relativos de  $MP_{10\_b}$ ,  $MP_{10\_c}$  e  $MP_{10\_d}$  com relação ao  $MP_{10\_a}$  serão dados pelo  $\exp(\beta_1)$ ,  $\exp(\beta_2)$  e  $\exp(\beta_1 + \beta_2)$ , respectivamente.

Para efeito de comparação com a metodologia Bayesiana, serão ajustados quatro modelos clássicos, ora considerando a presença de um efeito aleatório, ora sem esse efeito, ajustando-se o  $MP_{10}$  como uma variável contínua, num primeiro momento, e categorizada, num segundo.

Quando se analisa os dados sob o enfoque clássico, na maioria dos problemas com modelos lineares generalizados, como é o caso do modelo de regressão de Poisson (MCCULLAGH & NELDER, 1989), é mais prático encontrar as estimativas dos parâmetros a partir de processos iterativos, como o de Newton-Raphson, por exemplo. Porém, tais métodos normalmente se iniciam dando-se valores iniciais aos parâmetros. Sendo assim, as estruturas dos modelos ajustados sob paradigma frequentista serão as mesmas daquelas ajustadas sob enfoque Bayesiano, considerando os modelos 1 e 2, com exceção do ajuste pelo efeito aleatório que seguirá uma distribuição dada por  $w_i \sim N(0, \sigma^2)$ , sendo que utilizaremos  $\sigma^2$  igual a 1 como valor inicial. Os parâmetros  $\beta_l$  também receberão valores iniciais, iguais a 0,01.

#### 4.1.5 Convergência do algoritmo

Para verificar a convergência dos algoritmos, existem na literatura diversas técnicas que podem ser consideradas. Entre elas, estão técnicas gráficas propostas por Gelfand e Smith (1990) e Geweke (1992). Na técnica proposta por Gelfand e Smith, várias cadeias paralelas são geradas a partir de valores iniciais diferentes. Após um grande número de iterações em cada cadeia, comparam-se os histogramas para cada componente  $\theta_j$  de  $\boldsymbol{\theta}$ . Histogramas similares indicam convergência da cadeia. Geweke

(1992) sugere métodos gráficos baseados em séries temporais das amostras selecionadas. Ainda outra técnica é a proposta por Gelman e Rubin (1992), baseada na análise de variância.

É importante observar que, na geração de amostras de Gibbs, devemos considerar as  $q$  primeiras iterações como período de aquecimento ("*burn-in-samples*"), que devem ser descartadas para eliminar o efeito de valores iniciais.

#### 4.1.6 Seleção de modelos

A seleção do melhor modelo é um problema fundamental da estatística. Sob paradigma Bayesiano, a literatura apresenta diferentes abordagens, sendo o fator de Bayes, a mais popular. Várias modificações deste fator são apresentadas na literatura, sendo que Geisser e Eddy (1979) e Gelfand e Gosh (1998) sugerem uma abordagem preditiva, na qual uma observação futura é descrita por uma distribuição condicional aos dados passados, denominada distribuição preditiva.

Considere uma escolha entre dois modelos paramétricos, cada um sendo denotado por sua função de verossimilhança  $L(\boldsymbol{\theta}_i; \mathbf{y}, M_i)$  ou por sua densidade conjunta  $f(\mathbf{y}; \boldsymbol{\theta}_i, M_i)$ ,  $i = 1, 2$ . Suponha que  $p_i$  é a probabilidade de selecionar o modelo  $M_i$  e  $f(\mathbf{y} | M_i)$  é a distribuição preditiva para o modelo  $M_i$ , isto é,

$$f(\mathbf{y} | M_i) = \int f(\mathbf{y} | \boldsymbol{\theta}_i, M_i) \pi(\boldsymbol{\theta}_i | M_i) d\boldsymbol{\theta}_i \quad (2.31)$$

onde  $\pi(\boldsymbol{\theta}_i | M_i)$  é a função de densidade a priori para o modelo  $M_i$ . Se  $y_0$  denota o conjunto de dados amostrais, então o modelo que apresentar o maior valor de  $p_i f(y_0 | M_i)$  será o escolhido. Geralmente, considera-se  $p_i = 0,5$  e  $i = 1, 2$ , e o fator de Bayes de  $M_1$  com relação a  $M_2$  é definido como

$$B_{12} = \frac{f(y_0 | M_1)}{f(y_0 | M_2)} \quad (2.32)$$

A distribuição preditiva (2.31) pode ser aproximada por Monte Carlo usando  $S$  amostras geradas da densidade a priori  $\pi(\boldsymbol{\theta}_i | M_i)$ , ou seja,

$$\hat{f}(\mathbf{y} | M_i) = \frac{1}{S} \sum_{s=1}^S f(\mathbf{y} | \boldsymbol{\theta}_i^{(s)}, M_i) \quad (2.33)$$

em que  $\boldsymbol{\theta}_i^{(s)}$  é a  $i$ -ésima amostra do vetor  $\boldsymbol{\theta}_i$ . Algumas modificações da estimativa são propostas na literatura, porém para a seleção de modelos podemos utilizar a técnica da densidade preditiva condicional ordenada (CPO) (GELFAND & DEY, 1994), dada por

$$CPO_r = f(y_r | \mathbf{y}, M_i) = \int f(y_r | \boldsymbol{\theta}_i, \mathbf{y}_{(r)}, M_i) \pi(\boldsymbol{\theta}_i | \mathbf{y}_{(r)}, M_i) d\boldsymbol{\theta}_i \quad (2.34)$$

em que  $\mathbf{y}_{(r)} = (y_1, \dots, y_{r-1}, y_{r+1}, \dots, y_n)$  e  $\pi(\boldsymbol{\theta}_i | \mathbf{y}_{(r)}, M_i)$  é a densidade a posteriori de  $\boldsymbol{\theta}_i$ , baseada em  $\mathbf{t}(r)$ .

Usando o amostrador de Gibbs, as densidades (2.34) podem ser aproximadas por

$$\hat{f}(y_r | \mathbf{t}_{(r)}, M_i) = \frac{1}{S} \sum_{s=1}^S f(y_r | \mathbf{t}_{(r)}, \boldsymbol{\theta}_i^{(s)}, M_i) \quad (2.35)$$

em que  $\boldsymbol{\theta}_i^{(s)}$  denota a  $s$ -ésima amostra de Gibbs do vetor  $\boldsymbol{\theta}_i$ . Podemos utilizar as estimativas de  $C\hat{P}O_r = \hat{f}(t_r | \mathbf{t}_{(r)}, M_i)$  na seleção dos modelos, considerando um gráfico dos valores de  $C\hat{P}O$  versus o número da observação  $r$ ,  $r = 1, \dots, n$  para diferentes modelos. Neste caso, valores maiores (em média), indicam o melhor modelo. Além disso, também podemos considerar para a escolha do melhor modelo, aquele que

$$\text{maximiza } MCPO = \frac{\sum_{r=1}^n \log(C\hat{P}O)}{n} \quad (\text{ACHCAR, 2001}).$$

Neste trabalho, utilizaremos tanto o gráfico, quanto o MCPO para a escolha do modelo.

## 4.2 Imputação de dados faltantes

Um comum fator complicador de pesquisas científicas é a ocorrência de dados faltantes (RUBIN, 1996). Determinar a melhor forma de se trabalhar com um banco de dados com observações incompletas é uma questão complexa, por isso, o desenvolvimento de métodos estatísticos direcionados a solucionar tais problemas tem sido a base de muitas investigações nas últimas décadas (RUBIN, 1987; LITTLE, 1992; SCHAFER, 1999; ZHANG, 2003).

Em situações como esta, é comum a prática de restringir a análise aos sujeitos com observações completas, porém, as estimativas obtidas podem ser viesadas, especialmente se os indivíduos incluídos na análise forem diferentes quanto a uma ou mais variáveis daqueles que foram excluídos. Como solução a esses problemas, desde a



década de 80 tem surgido técnicas estatísticas envolvendo a substituição dos dados faltantes por estimativas de valores que sejam plausíveis. Tais técnicas são chamadas de imputação de dados faltantes e seu uso vem gradativamente aumentando nas áreas de pesquisa (RUBIN, 1996; SCHAFER, 2002). Seu objetivo é “completar” os bancos de dados e possibilitar a análise com todos os indivíduos do estudo.

As primeiras técnicas de imputação desenvolvidas envolviam métodos simples, como substituição dos dados faltantes pela média ou mediana da variável, por interpolação ou por meio de uma regressão linear (NUNES et al., 2009). Entretanto, o fato de estas técnicas preencherem os dados uma única vez (imputação única) faz com que a incerteza quanto à imputação realizada deva ser considerada, já que os resultados obtidos com estes valores irreais poderiam não ser válidos. De incertezas podemos considerar segundo Zhang (2003), citado por Nunes et al. (2009), por exemplo, aquelas devidas à amostragem, que prevê a seleção aleatória dos valores a serem imputados e as incertezas quanto aos parâmetros do modelo. Esta idéia levou ao desenvolvimento da técnica de Imputação Múltipla proposta por D. B. Rubin, em meados da década de 70.

As análises estatísticas foram e ainda são desenvolvidas, em sua maioria, para dados completos. Apesar de vários métodos de inferência na presença de dados incompletos estarem sendo desenvolvidos, tais como os modelos de efeitos aleatórios (MCCULLAGH & NELDER, 1989), podem ainda surgir complicações computacionais que comprometam a precisão das inferências.

Mesmo uma pequena quantidade de dados faltantes pode ocasionar estimativas viesadas e ineficientes, por isso, devido às quantidades faltantes de inalação/nebulização para os meses de julho de 2006 e janeiro, abril e maio de 2008, consideramos a necessidade de analisar não apenas os dados excluindo as observações perdidas, mas analisar os dados utilizando um método adequado de imputação, no caso, o de imputação múltipla, melhor descrito no tópico a seguir.

#### 4.2.1 Imputação múltipla

O fato de ser possível trabalhar tanto com dados quantitativos quanto qualitativos faz com que a imputação múltipla seja vista como um método flexível, se tornando uma técnica cada vez mais utilizada para tratar dados faltantes (WHITE, 2007). Ademais, essa técnica é capaz de corrigir o maior problema associado à

imputação única através da inclusão da incerteza da imputação dos dados na variância das estimativas (NUNES, 2007).

Basicamente, o método consiste de três passos (SCHAFER, 1997):

- (i) Obter  $m$  bancos de dados completos através de técnicas adequadas de imputação;
- (ii) Analisar separadamente, pelo método estatístico apropriado, os  $m$  bancos de dados completos;
- (iii) Combinar os  $m$  resultados encontrados de forma a se obter a chamada inferência da imputação repetida.

De forma geral, o modelo de imputação múltipla é ajustado sob o enfoque Bayesiano, uma vez que é através do resultado da distribuição a posteriori que um conjunto de extrações aleatórias para os dados faltantes a partir dos dados observados, é feito. Esse processo é repetido  $m$  vezes, originando  $m$  bancos de dados completos. Independente do método de imputação múltipla utilizado, o princípio é o mesmo: as imputações são feitas a partir de uma regressão linear em que a variável resposta será a variável a ser imputada (RUBIN, 1987).

Um método adequado para a imputação de variáveis quantitativas, como é o caso das quantidades de inalação/nebulização, e talvez o método mais comum de se prever  $Y_i$  a partir de um conjunto de variáveis  $X_i$  é o modelo da Regressão Linear Bayesiana (NUNES, 2007).

#### 4.2.1.1 Método de imputação baseado na Regressão Linear Bayesiana

Seja o modelo de regressão linear múltiplo dado por  $Y = \alpha + \beta X$ ,  $Y \sim N(X\beta; I\sigma^2)$ , sendo  $\theta = (\beta, \sigma^2)$ ,  $\beta$  um vetor de  $q$  componentes, em que  $q$  é o número de variáveis preditoras, e  $\sigma$  um escalar. Assume-se uma distribuição a priori não informativa para  $\theta$ ,  $P(\theta)$ .

**Resultado:** (RUBIN, 1987). A distribuição a posteriori de  $\theta$  envolve somente as unidades com  $Y_i$  observados. A estimação dos parâmetros será feita a partir da distribuição a posteriori de  $\theta$ .

Basicamente, serão estimados  $\beta$  e  $\sigma$  para cada um dos  $m$  modelos. A partir daí, os valores faltantes de  $y$  serão preenchidos com os próprios valores preditos pelos  $m$  modelos independentes.

A combinação dos resultados obtidos nas  $m$  diferentes análises se dá através de regras criadas por Rubin, que podem ser usadas independentemente do método utilizado para se fazer a imputação múltipla (SCHAFER & GRAHAM, 2002 citado por NUNES, 2007).

O princípio é que se obtenham as  $m$  estimativas para o parâmetro de interesse  $Q$ , sendo ele por exemplo, a média  $Y$  na população. A estimativa combinada será a média das estimativas individuais:

$$\bar{Q} = \frac{1}{m} \sum_{j=1}^m \hat{Q}_j \quad (2.32)$$

Para se obter a variância combinada, inicialmente calcula-se a variância média dentro das imputações,  $\bar{U} = \frac{1}{m} \sum_{j=1}^m U_j$ , em que  $U_j$  são as variâncias dos estimadores dentro de cada imputação, para  $j=1,2,\dots,m$ , e em seguida, calcula-se a variância entre as imputações,  $B = \frac{1}{m-1} \sum_{j=1}^m (\hat{Q}_j - \bar{Q})^2$ .

Assim, a variância combinada, será  $T = \bar{U} + \left(1 + \frac{1}{m}\right)B$ .

Estimativas intervalares para  $Q$  são formadas usando uma distribuição  $t$  com  $v = (m-1) \left(1 + \frac{1}{r}\right)^2$  graus de liberdade, em que  $r$  é o aumento relativo da variância devido a não-resposta:

$$r = \frac{\left(1 + \frac{1}{m}\right)B}{\bar{U}} \quad (2.33)$$

Então, da maneira padrão, uma estimativa intervalar  $100(1-\alpha)\%$  de  $Q$  é dada por

$$Q \pm t_{v, \alpha/2} \sqrt{T} \quad (2.34)$$

**Observação:** A implementação computacional dos diferentes métodos de imputação podem ser feitos em diversos softwares, como SAS, NORM e R. Porém, por este último se tratar de um software livre e gratuito, foi o utilizado para o cálculo das imputações.

Através do pacote Multivariate Imputation by Chained Equations (MICE - VAN BUUREN & OUDSHOORN, 2000), o software R permite que a imputação múltipla seja feita de forma rápida e eficiente.

## 5 RESULTADOS

### 5.1 Análise estatística dos dados sem imputação

#### 5.1.1 Ajustes considerando os valores contínuos do material particulado

Para a análise Bayesiana dos dados, foi assumido para os modelos 1, 2 e 3 os seguintes valores para os hiperparâmetros das distribuições a priori (2.20), (2.28) e (2.29):  $a=c=0$ ;  $b=10$ ;  $d=1$  e  $f=1$ .

Esses hiperparâmetros foram escolhidos para que as distribuições a priori fossem aproximadamente não-informativas e para que a convergência dos algoritmos MCMC usados para a simulação de Gibbs sample para a distribuição a posteriori de interesse fosse observada. As análises utilizaram o programa computacional WinBugs. Os códigos dos programas podem ser vistos no Apêndice.

Para o modelo 1 foram simuladas 51.000 amostras, sendo descartadas as 1000 primeiras (“*burn-in-samples*”) para evitar algum efeito dos valores iniciais e, posteriormente foi simulada uma amostra final para cada parâmetro de interesse de tamanho 5000, escolhida de 10 em 10. Para os modelos 2 e 3 foram simuladas um total de 9.020.000 amostras, sendo as primeiras 20.000 amostras descartadas. Para se ter uma amostra de Gibbs aproximadamente não correlacionada, foi considerada a 3000<sup>a</sup>, 6000<sup>a</sup>, 9000<sup>a</sup>,..., que resultou em uma amostra final de tamanho 3000 para cada parâmetro. Para todos os modelos, através da leitura de valores iniciais, foram ajustadas duas cadeias paralelas visando uma melhor convergência do algoritmo e resultados mais precisos.

As convergências dos algoritmos foram monitoradas utilizando o método de Gelman e Rubin (1992), que utiliza a técnica de análise de variância para determinar se mais iterações são necessárias, além de gráficos usuais de séries temporais das amostras simuladas.

Os ajustes feitos sob enfoque frequentista foram realizados utilizando o software SAS, através do procedimento *NLMIXED* (WOLFINGER, 2000).

Os resultados dos ajustes sob paradigma clássico podem ser vistos nas Tabelas 3 e 4. Já, sob enfoque bayesiano, os resultados a posteriori para os parâmetros dos modelos 1, 2 e 3 podem ser vistos nas Tabelas 5 a 7.

Tabela 3. Estimativas dos parâmetros do modelo clássico sem  $w_i$ , com seus respectivos intervalos de confiança (IC) 95%. (material particulado contínuo)

<b>Parâmetro</b>	<b>Estimativa</b>	<b>IC 95%</b>
Intercepto ( $\beta_0$ )	7,659	(7,645; 7,673)
Medidas mensais de MP <sub>10</sub> , $\mu\text{g}/\text{m}^3$ ( $\beta_1$ )	0,0094	(0,0092; 0,0097)
Temperatura mínima, médias mensais, em °C ( $\beta_2$ )	-0,015	(-0,017; -0,013)
Volume total de precipitação, mm ( $\beta_3$ )	0,0014	(0,0014; 0,0015)
Estação do ano – Outono x Verão ( $\beta_4$ )	0,485	(0,469; 0,501)
Estação do ano – Inverno x Verão ( $\beta_5$ )	0,229	(0,212; 0,246)
Estação do ano – Primavera x Verão ( $\beta_6$ )	-0,097	(-0,108; -0,085)

Tabela 4. Estimativas dos parâmetros do modelo clássico com a presença de  $w_i$ , com seus respectivos intervalos de confiança (IC) 95%. (material particulado contínuo)

<b>Parâmetro</b>	<b>Estimativa</b>	<b>IC 95%</b>
Intercepto ( $\beta_0$ )	7,998	(7,688; 8,309)
Medidas mensais de MP <sub>10</sub> , $\mu\text{g}/\text{m}^3$ ( $\beta_1$ )	0,011	(0,002; 0,019)
Temperatura mínima, médias mensais, em °C ( $\beta_2$ )	-0,017	(-0,093; 0,059)
Volume total de precipitação, mm ( $\beta_3$ )	0,001	(0,000; 0,003)
Estação do ano – Outono x Verão ( $\beta_4$ )	0,439	(-0,055; 0,933)
Estação do ano – Inverno x Verão ( $\beta_5$ )	0,048	(-0,480; 0,576)
Estação do ano – Primavera x Verão ( $\beta_6$ )	-0,195	(-0,514; 0,125)

Tabela 5. Estimativas dos parâmetros do modelo 1, com seus respectivos intervalos de credibilidade (IC) 95%. (material particulado contínuo)

<b>Parâmetro</b>	<b>Estimativa</b>	<b>IC 95%</b>
Intercepto ( $\beta_0$ )	8,045	(8,035; 8,056)
Medidas mensais de MP <sub>10</sub> , $\mu\text{g}/\text{m}^3$ ( $\beta_1$ )	0,0094	(0,0092; 0,010)
Temperatura mínima, médias mensais, em °C ( $\beta_2$ )	-0,015	(-0,017; -0,013)
Volume total de precipitação, mm ( $\beta_3$ )	0,00144	(0,0014; 0,0015)
Estação do ano – Outono x Verão ( $\beta_4$ )	0,485	(0,469; 0,501)
Estação do ano – Inverno x Verão ( $\beta_5$ )	0,229	(0,212; 0,246)
Estação do ano – Primavera x Verão ( $\beta_6$ )	-0,097	(-0,108; -0,085)

Tabela 6. Estimativas dos parâmetros do modelo 2, com seus respectivos intervalos de credibilidade (IC) 95%. (material particulado contínuo)

<b>Parâmetro</b>	<b>Estimativa</b>	<b>IC 95%</b>
Intercepto ( $\beta_0$ )	7,983	(7,672; 8,299)
Medidas mensais de MP <sub>10</sub> , $\mu\text{g}/\text{m}^3$ ( $\beta_1$ )	0,011	(0,001; 0,020)
Temperatura mínima, médias mensais, em °C ( $\beta_2$ )	-0,016	(-0,094; 0,062)
Volume total de precipitação, mm ( $\beta_3$ )	0,002	(0,000; 0,003)
Estação do ano – Outono x Verão ( $\beta_4$ )	0,451	(-0,055; 0,958)
Estação do ano – Inverno x Verão ( $\beta_5$ )	0,057	(-0,484; 0,591)
Estação do ano – Primavera x Verão ( $\beta_6$ )	-0,180	(-0,522; 0,162)

Tabela 7. Estimativas dos parâmetros do modelo 3, com seus respectivos intervalos de credibilidade (IC) 95%. (material particulado contínuo)

<b>Parâmetro</b>	<b>Estimativa</b>	<b>IC 95%</b>
Intercepto ( $\beta_0$ )	8,143	(7,627; 8,688)
Medidas mensais de MP <sub>10</sub> , $\mu\text{g}/\text{m}^3$ ( $\beta_1$ )	0,010	(-0,007; 0,027)
Temperatura mínima, médias mensais, em °C ( $\beta_2$ )	-0,036	(-0,174; 0,104)
Volume total de precipitação, mm ( $\beta_3$ )	0,001	(-0,002; 0,004)
Estação do ano – Outono x Verão ( $\beta_4$ )	0,357	(-0,430; 1,156)
Estação do ano – Inverno x Verão ( $\beta_5$ )	0,114	(-0,784; 1,018)
Estação do ano – Primavera x Verão ( $\beta_6$ )	-0,130	(-0,720; 0,458)

Os modelos ajustados forneceram resultados semelhantes, uma vez que em todos eles, os intervalos para o parâmetro  $\beta_1$ , associado às médias mensais de MP<sub>10</sub>, não contiveram o valor 0, com exceção do Modelo 3, bayesiano. Apesar da metodologia clássica ter apresentado intervalos de confiança parecidos com os intervalos de credibilidade obtidos sob enfoque bayesiano, inclusive com os valores das somas de resíduos absolutos (Tabela 8) entre ambos os modelos sendo também muito próximos, os intervalos de credibilidade bayesianos tem a vantagem de serem mais facilmente interpretados.

Os valores das somas dos resíduos absolutos de cada modelo ajustado (Tabela 8) sugerem que o Modelo 3 fornece o melhor ajuste dentre os demais, uma vez que ofereceu o menor valor entre as medidas calculadas, porém, dada a magnitude dos dados estudados, podemos dizer que este valor não se mostrou muito diferente do observado para o Modelo 2.

Tabela 8. Soma dos resíduos absolutos de cada modelo. (material particulado contínuo)

Modelos		Soma dos resíduos absolutos
Clássico	Sem $w_i$	130.626,2
	Com $w_i$	132,0
Bayesiano	Modelo 1	130.627,0
	Modelo 2	130,4
	Modelo 3	61,2

### 5.1.2 Resultados considerando os valores categorizados do material particulado

Os modelos de regressão considerando o material particulado de maneira categorizada foi construído de forma idêntica aos modelos considerando os valores contínuos do MP<sub>10</sub>. A única diferença está na inclusão de um parâmetro, como já exposto na Tabela 2 e no modelo (2.30), sendo assim, o vetor de parâmetros  $\beta$ , será composto agora por  $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$  e  $\beta_7$ .

Analogamente aos modelos anteriores,  $\beta_0 \sim N(a,b)$  e  $\beta_1, \dots, \beta_7 \sim N(c,d)$ , com os hiperparâmetros conhecidos,  $a=c=0$ ;  $b=10$ ;  $d=1$ , e escolhidos de forma que as distribuições a priori fossem aproximadamente não-informativas e para que a convergência dos algoritmos fosse observada.

Para o modelo 1, foram simuladas duas cadeias com 505.000 amostras cada, sendo descartadas as 1000 primeiras para evitar algum efeito dos valores iniciais e, posteriormente foi simulada uma amostra final para cada parâmetro de interesse de tamanho 5000, escolhida de 100 em 100. Para os modelos 2<sup>i</sup> (considerando  $w_i \sim N(0, \sigma^2)$  e  $\sigma^2 \sim Gama(1,1)$ ) e 3<sup>ii</sup> (considerando  $w_i \sim Gama(1,1)$ ) foram geradas duas cadeias de tamanho 9.020.000 para cada parâmetro. As primeiras 20.000 amostras foram descartadas e as restantes selecionadas de 3000 em 3000, totalizando uma amostra de tamanho 3000.

<sup>i</sup>  $\log(\lambda_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{3i} - \bar{X}_3) + \beta_4 (X_{4i} - \bar{X}_4) + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i} + w_i$

<sup>ii</sup>  $\lambda_i = w_i \exp[\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{3i} - \bar{X}_3) + \beta_4 (X_{4i} - \bar{X}_4) + \beta_5 X_{5i} + \beta_6 X_{6i} + \beta_7 X_{7i}]$

Os resultados dos ajustes dos modelos clássicos podem ser vistos nas Tabelas 9 e 10 e os resultados dos modelos ajustados sob enfoque bayesiano podem ser vistos nas Tabelas 11 a 13.

Tabela 9. Estimativas dos parâmetros do modelo clássico sem  $w_i$ , com seus respectivos intervalos de confiança (IC) 95%. (material particulado categorizado)

<b>Parâmetro</b>	<b>Estimativa</b>	<b>IC 95%</b>
Intercepto ( $\beta_0$ )	8,080	(8,034; 8,126)
X1 ( $\beta_1$ )	0,196	(0,188; 0,204)
X2 ( $\beta_2$ )	0,219	(0,207; 0,230)
Temperatura mínima, médias mensais, em °C ( $\beta_3$ )	-0,023	(-0,025; -0,021)
Volume total de precipitação, mm ( $\beta_4$ )	0,00127	(0,00121; 0,0032)
Estação do ano – Outono x Verão ( $\beta_5$ )	0,454	(0,438; 0,470)
Estação do ano – Inverno x Verão ( $\beta_6$ )	0,255	(0,238; 0,272)
Estação do ano – Primavera x Verão ( $\beta_7$ )	-0,095	(-0,107; -0,083)

Tabela 10. Estimativas dos parâmetros do modelo clássico na presença do  $w_i$ , com seus respectivos intervalos de confiança (IC) 95%. (material particulado categorizado)

<b>Parâmetro</b>	<b>Estimativa</b>	<b>IC 95%</b>
Intercepto ( $\beta_0$ )	7,729	(7,377; 8,081)
X1 ( $\beta_1$ )	0,249	(-0,009; 0,508)
X2 ( $\beta_2$ )	0,289	(-0,079; 0,658)
Temperatura mínima, médias mensais, em °C ( $\beta_3$ )	-0,022	(-0,100; -0,055)
Volume total de precipitação, mm ( $\beta_4$ )	0,001	(0,000; 0,003)
Estação do ano – Outono x Verão ( $\beta_5$ )	0,411	(-0,093; 0,915)
Estação do ano – Inverno x Verão ( $\beta_6$ )	0,069	(-0,469; 0,607)
Estação do ano – Primavera x Verão ( $\beta_7$ )	-0,190	(-0,514; 0,134)

Tabela 11. Estimativas dos parâmetros do modelo 1, com seus respectivos intervalos de credibilidade (IC) 95%. (material particulado categorizado)

<b>Parâmetro</b>	<b>Estimativa</b>	<b>IC 95%</b>
Intercepto ( $\beta_0$ )	7,842	(7,830; 7,854)
X1 ( $\beta_1$ )	0,196	(0,188; 0,204)
X2 ( $\beta_2$ )	0,218	(0,207; 0,230)
Temperatura mínima, médias mensais, em °C ( $\beta_3$ )	-0,023	(-0,025; -0,021)
Volume total de precipitação, mm ( $\beta_4$ )	0,00127	(0,00121; 0,0032)
Estação do ano – Outono x Verão ( $\beta_5$ )	0,454	(0,439; 0,470)
Estação do ano – Inverno x Verão ( $\beta_6$ )	0,255	(0,238; 0,272)
Estação do ano – Primavera x Verão ( $\beta_7$ )	-0,095	(-0,107; -0,083)



Tabela 12. Estimativas dos parâmetros do modelo 2, com seus respectivos intervalos de credibilidade (IC) 95%. (material particulado categorizado)

<b>Parâmetro</b>	<b>Estimativa</b>	<b>IC 95%</b>
Intercepto ( $\beta_0$ )	7,721	(7,372; 8,058)
X1 ( $\beta_1$ )	0,249	(-0,026; 0,521)
X2 ( $\beta_2$ )	0,286	(-0,103; 0,685)
Temperatura mínima, médias mensais, em °C ( $\beta_3$ )	-0,024	(-0,109; 0,060)
Volume total de precipitação, mm ( $\beta_4$ )	0,001	(0,000; 0,003)
Estação do ano – Outono x Verão ( $\beta_5$ )	0,406	(-0,097; 0,904)
Estação do ano – Inverno x Verão ( $\beta_6$ )	0,074	(-0,472; 0,602)
Estação do ano – Primavera x Verão ( $\beta_7$ )	-0,181	(-0,514; 0,154)

Tabela 13. Estimativas dos parâmetros do modelo 3, com seus respectivos intervalos de credibilidade (IC) 95%. (material particulado categorizado)

<b>Parâmetro</b>	<b>Estimativa</b>	<b>IC 95%</b>
Intercepto ( $\beta_0$ )	7,815	(7,153; 8,449)
X1 ( $\beta_1$ )	0,263	(-0,257; 0,775)
X2 ( $\beta_2$ )	0,333	(-0,335; 1,020)
Temperatura mínima, médias mensais, em °C ( $\beta_3$ )	-0,024	(-0,164; 0,117)
Volume total de precipitação, mm ( $\beta_4$ )	0,001	(-0,001; 0,005)
Estação do ano – Outono x Verão ( $\beta_5$ )	0,394	(-0,393; 1,178)
Estação do ano – Inverno x Verão ( $\beta_6$ )	0,195	(-0,708; 1,078)
Estação do ano – Primavera x Verão ( $\beta_7$ )	-0,109	(-0,682; 0,506)

Assim como ocorreu ajustando o MP<sub>10</sub> contínuo, os modelos aqui ajustados também forneceram resultados semelhantes. Em todos os modelos ajustados na presença de  $w_i$  não se observou nenhum intervalo que não contivesse o valor 0 para os parâmetros. Porém, mais uma vez comparando os valores obtidos com a soma dos resíduos absolutos de cada modelo (Tabela 14), observamos que o modelo Bayesiano obtém melhores resultados.

Tabela 14. Soma dos resíduos absolutos de cada modelo. (material particulado contínuo)

<b>Modelos</b>		<b>Soma dos resíduos absolutos</b>
<b>Clássico</b>	<b>Sem <math>w_i</math></b>	133.806,3
	<b>Com <math>w_i</math></b>	132,6
<b>Bayesiano</b>	<b>Modelo 1</b>	133.805,0
	<b>Modelo 2</b>	119,9
	<b>Modelo 3</b>	50,8

Apesar dos resultados entre os enfoques clássico e Bayesiano terem sido similares, a inferência Bayesiana oferece resultados de fácil interpretação e não depende de testes de hipóteses, ao contrário do método clássico. Por isso, nas próximas seções estudaremos mais detalhadamente apenas o ajuste dos modelos Bayesianos.

### 5.1.3 Seleção do melhor modelo

Além dos valores obtidos pelo cálculo da soma dos resíduos absolutos (Tabelas 8 e 14), é possível notar a sensível melhora que a inclusão do efeito aleatório ofereceu aos ajustes através dos gráficos contendo os resíduos *versus* observação obtidos de cada modelo (Figuras 8 e 9). Percebe-se que, ao contrário do modelo sem efeito aleatório, os modelos com a presença de  $w_i$  predizeram quase que perfeitamente os valores observados de inalação/nebulização durante todo o período do estudo, resultado que ocorreu incluindo o MP<sub>10</sub> contínuo e categorizado. Enquanto o ajuste dos dois modelos sem  $w_i$  (considerando o MP<sub>10</sub> contínuo e em categorias) obtiveram uma diferença entre os valores de inalação/nebulização observados e os preditos entre -2000 e 8000, os modelos com a presença do efeito aleatório obtiveram resíduos insignificantes dada a magnitude dos dados, variando de -2 a 10.

Na Figura 8 tem-se o gráfico dos resíduos para cada observação, considerando os diferentes modelos ajustados.

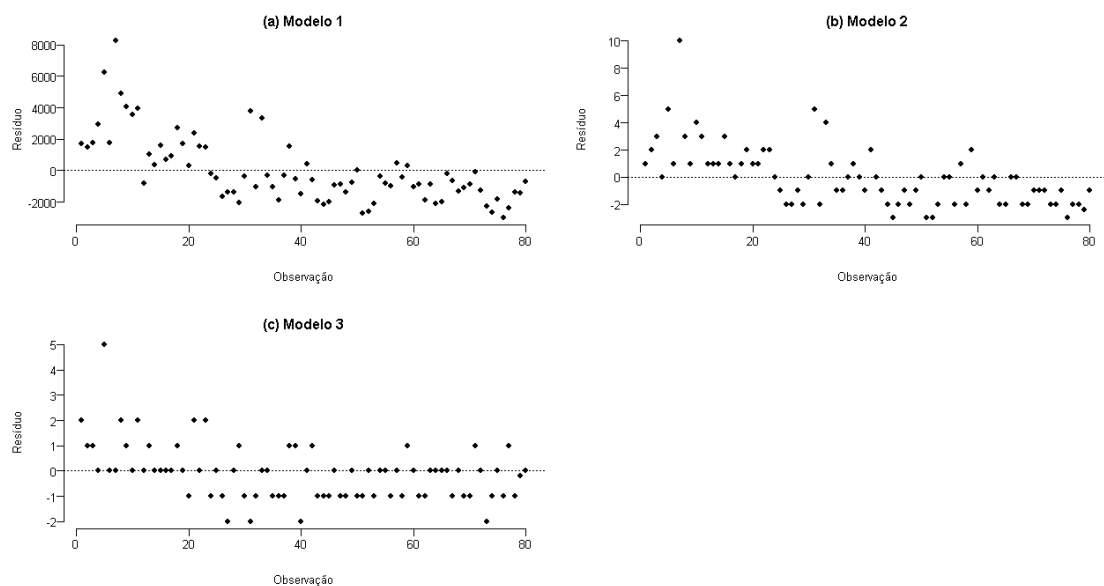


Figura 8. Gráfico dos resíduos *versus* observação, considerando os valores contínuos do material particulado.

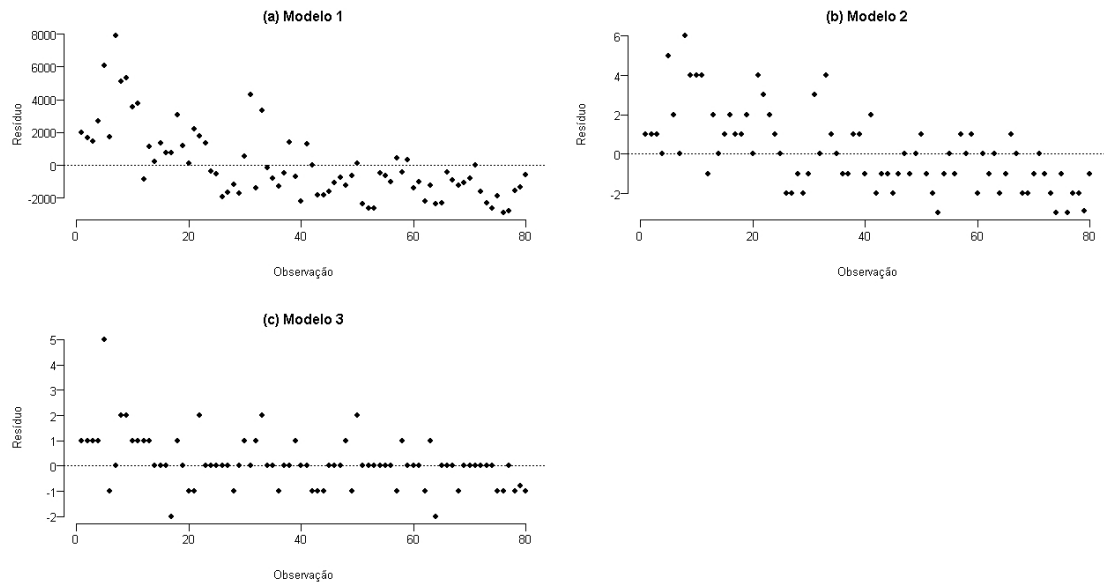


Figura 9. Gráfico dos resíduos *versus* observação, considerando os valores do material particulado categorizados.

O ganho mudando a formulação do modelo e consequentemente, a priori do efeito aleatório, parece ser pequeno quando comparamos os ajustes entre os dois modelos, tanto quando consideramos o  $MP_{10}$  contínuo, quanto categorizado.

Utilizando as estimativas de CPO, calculadas para cada modelo, como critério para escolha do melhor modelo, foi possível construir os gráficos, a seguir, dos valores de  $\hat{CPO}$  versus o número das observações e calcular o MCPO correspondente.

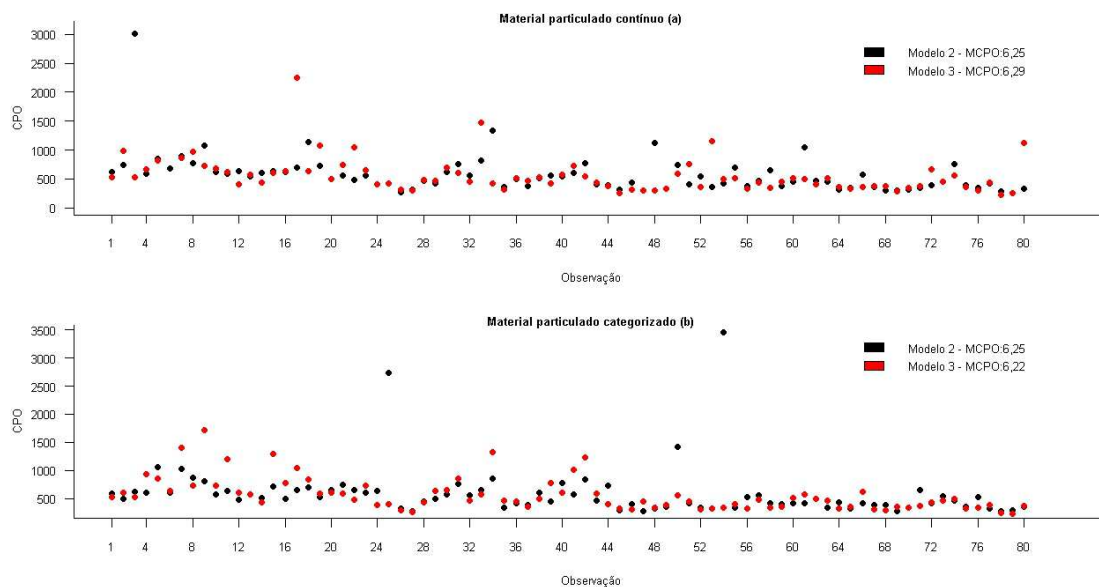


Figura 10. Gráfico de CPO *versus* o número da observação para os modelos 2 e 3, considerando  $MP_{10}$  contínuo e categorizado, separadamente.

Pelo MCPO calculado para cada modelo, vemos que as quatro medidas foram muito próximas, sendo que o Modelo 3 obteve um valor ligeiramente maior no ajuste pelas medidas contínuas do poluente, e o Modelo 2 um valor maior no ajuste do material particulado categorizado.

Analisando todos os modelos ajustados e ainda o conjunto de medidas utilizadas para seleção de modelos (MCPO, gráfico de resíduos e soma dos resíduos absolutos), percebe-se que o Modelo 2 produz resultados muito semelhantes ao Modelo 3, porém com intervalos de credibilidade bem mais “fechados” que este último. Assim, para fins de interpretação, detalharemos as estimativas obtidas pelo modelo considerando  $w_i$  como uma distribuição normal (Modelo 2) para as duas formas de ajuste pelo  $MP_{10}$ .

O coeficiente  $\beta_1$ , associado às médias mensais de  $MP_{10}$  em uma escala contínua, foi estimado em 0,011, com um IC95% de (0,001; 0,020). Por outro lado, não foram evidenciados efeitos das médias mensais de temperatura mínima ou do volume total de precipitação sobre os números mensais de procedimentos de inalação/nebulização (os respectivos IC95% cobrem o valor 0).

A partir do ajuste do modelo pelas categorias criadas com os valores das concentrações mensais do poluente foi possível calcular o risco relativo e o intervalo de credibilidade (95%) correspondente a cada faixa de valores. O resultado é mostrado na figura a seguir.

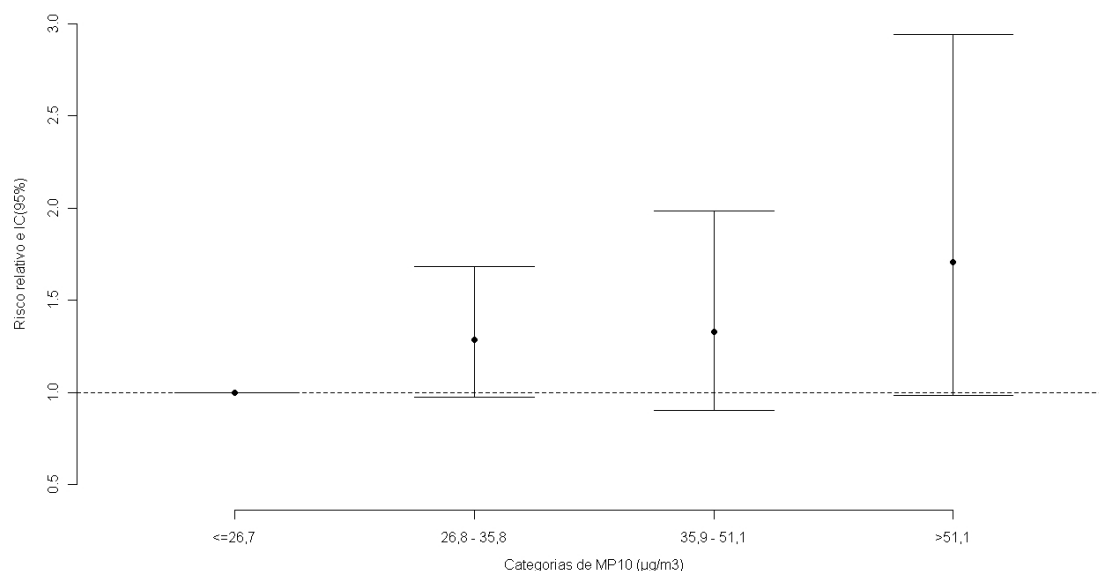


Figura 11. Riscos relativos e respectivos intervalos de credibilidade (95%) para inalações com concentrações crescentes do  $MP_{10}$ .

O gráfico sugere que as concentrações do poluente, quando categorizado, não interferem no número mensal de inalações/nebulizações. Apesar de os intervalos referentes à segunda e quarta faixas do poluente apresentarem um limite superior bem elevado, próximos a 1.7 e 3.0 respectivamente, os limites inferiores de ambos ficaram abaixo de 1, o que indica a não evidência de associação entre as categorias do poluente e as quantidades mensais de procedimentos.

## 5.2 Análise estatística dos dados com imputação

No item anterior (5.1), desconsideramos os meses em que as quantidades de inalações/nebulizações estavam sub-notificadas e ajustamos os modelos considerando apenas os meses em que as informações de todas as variáveis envolvidas estavam completas. Entretanto, essa exclusão pode gerar inferências inválidas, principalmente se as informações referentes aos meses que permaneceram na análise forem diferentes daquelas faltantes ou que foram excluídas. Por esse motivo preferiu-se ajustar novos modelos utilizando um método adequado de imputação de dados, já detalhado no fim do capítulo 4.

Em pesquisas, a ausência de dados é praticamente inevitável, porém, é aconselhável que ainda na etapa de planejamento do estudo, o pesquisador investigue potenciais variáveis preditoras dos valores faltantes e utilize-as na coleta de dados. Pensando nisso, Rubin (1996) destaca em seu trabalho três tipos de mecanismos de não-resposta, que são:

- a) Não-resposta ou dado faltante completamente aleatório: quando os motivos para a perda do dado independem das respostas dos sujeitos;
- b) Não-resposta ou dado faltante ao acaso: quando a perda do dado pode ser prevista de acordo com os outros dados coletados;
- c) Não-resposta ou dado faltante não aleatório: quando a não-resposta está relacionada aos valores não observados. Quando o dado faltante se trata de outliers, ou seja, quando possui um valor diferente do padrão da amostra coletada, sua chance de resposta à variável de interesse (por exemplo, renda familiar mensal) diminui.

Na prática, o processo que levou à ocorrência da não-resposta é difícil de ser determinado. Segundo Kenward e Carpenter (2007), essa afirmação não pode ser testada, somente suposta.

No presente trabalho, julgou-se aceitável considerar o dado faltante aleatório como mecanismo de não-resposta. Como a porcentagem de dados faltantes é muito pequena em vista de outros trabalhos da literatura, de 4.7% (4 casos), a escolha do mecanismo de não resposta tornou-se ainda mais difícil. Devido à ausência da informação da quantidade de inalação/nebulização ter ocorrido em maior proporção nos períodos de estiagem (meses de outono e inverno), quando as temperaturas mínimas e o volume de precipitação são mais baixos e a concentração do material particulado se eleva, consideramos que os dados faltantes poderiam estar relacionados a essas variáveis e optamos por não fazer ajustes especiais na imputação que aumentaria sua complexidade, talvez desnecessariamente. Assim, todas as variáveis estudadas, material particulado, temperatura mínima, precipitação e estação do ano foram incluídas na imputação pelo método da Regressão Linear Bayesiana.

Foram realizadas um total de 5 imputações –  $m$  igual a 5 é a quantidade mais frequentemente utilizada na literatura –, e em seguida, os bancos de dados completados foram analisados segundo o modelo 2, considerando primeiramente os valores contínuos das concentrações de material particulado e posteriormente, as concentrações divididas em categorias.

A Tabela 15 descreve as quantidades de inalação/nebulização geradas por cada imputação para cada mês com ausência da informação. É possível observar que os valores gerados para os meses de janeiro e abril de 2008 não tiveram grande variabilidade entre as imputações, diferentemente do que ocorreu com o mês de julho de 2006, quando este número variou de 2494 a 6344. É o fato de considerar essa variabilidade na estimativa final do parâmetro de interesse que faz com que a imputação múltipla seja preferível em relação às técnicas mais simples de imputação. A Figura 12, contendo as séries observadas dos procedimentos de inalação/nebulização ao longo do período com o acréscimo dos dados imputados, permite uma melhor visualização dos valores gerados pelas cinco imputações realizadas.

Tabela 15. Descrição dos valores de inalação/nebulização gerados por cada uma das cinco imputações realizadas.

Mês com dado faltante	Valores gerados em cada imputação					Média	Desvio Padrão
	1 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	5 <sup>a</sup>		
Jul/06	6344	3013	6282	5157	2494	4658,0	1810,9
Jan/08	1568	1807	2065	1730	2054	1844,8	214,2
Abr/08	2387	4315	3080	2759	4183	3344,8	862,3
Mai/08	2709	2494	2735	3139	3033	2822,0	261,2

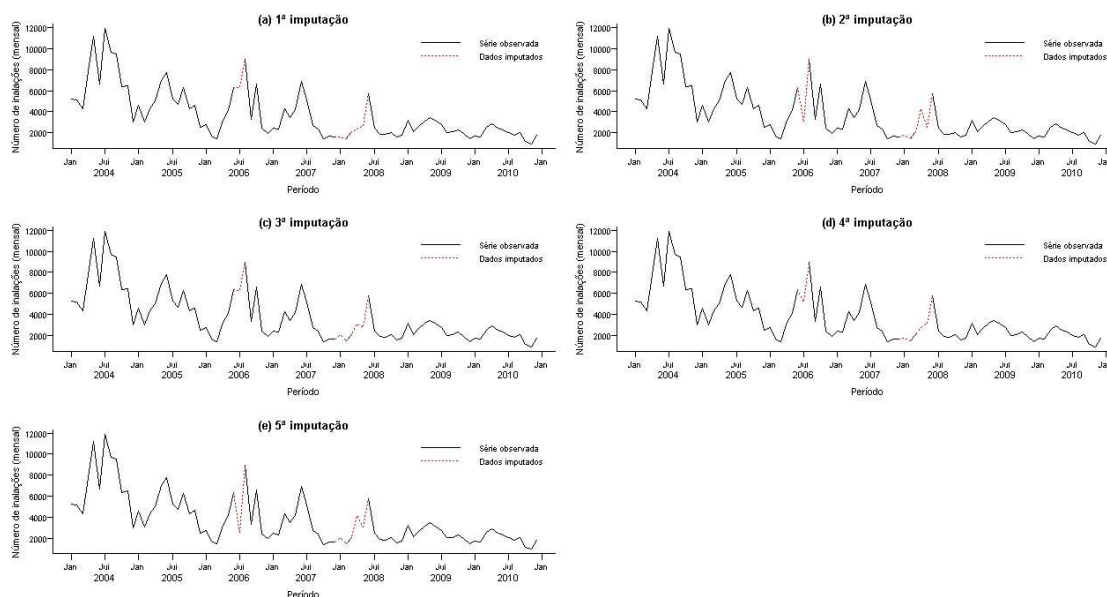


Figura 12. Dados imputados na série das quantidades mensais de inalação/nebulização.

A partir das regras propostas por Rubin, as estimativas combinadas pontuais e intervalares com um nível de significância de 5% para os parâmetros de interesse puderam ser obtidas para cada modelo. As Tabelas 16 e 17 descrevem, respectivamente, as estimativas obtidas para o modelo com os valores contínuos do poluente e com seus valores categorizados.

Tabela 16. Estimativas combinadas pontuais e intervalares (95%) dos parâmetros do modelo 2. (material particulado contínuo)

Parâmetro	Estimativa	IC 95%
Intercepto ( $\beta_0$ )	7,982	(7,675; 8,286)
Medidas mensais de MP <sub>10</sub> , $\mu\text{g}/\text{m}^3$ ( $\beta_1$ )	0,011	(0,002; 0,020)
Temperatura mínima, médias mensais, em °C ( $\beta_2$ )	-0,016	(-0,094; 0,062)
Volume total de precipitação, mm ( $\beta_3$ )	0,001	(0,000; 0,003)
Estação do ano – Outono x Verão ( $\beta_4$ )	0,422	(-0,050; 0,903)
Estação do ano – Inverno x Verão ( $\beta_5$ )	0,062	(-0,469; 0,592)
Estação do ano – Primavera x Verão ( $\beta_6$ )	-0,169	(-0,503; 0,162)

Tabela 17. Estimativas combinadas pontuais e intervalares (95%) dos parâmetros do modelo 2. (material particulado categorizado)

<b>Parâmetro</b>	<b>Estimativa</b>	<b>IC 95%</b>
Intercepto ( $\beta_0$ )	7,718	(7,359; 8,081)
X1 ( $\beta_1$ )	0,258	(-0,010; 0,522)
X2 ( $\beta_2$ )	0,267	(-0,107; 0,640)
Temperatura mínima, médias mensais, em °C ( $\beta_3$ )	-0,024	(-0,103; 0,056)
Volume total de precipitação, mm ( $\beta_4$ )	0,001	(0,000; 0,003)
Estação do ano – Outono x Verão ( $\beta_5$ )	0,405	(-0,086; 0,895)
Estação do ano – Inverno x Verão ( $\beta_6$ )	0,096	(-0,444; 0,627)
Estação do ano – Primavera x Verão ( $\beta_7$ )	-0,165	(-0,500; 0,169)

O coeficiente  $\beta_1$ , associado às médias mensais de  $MP_{10}$  em uma escala contínua, foi estimado em 0,011 pelos modelos com e sem imputação de dados. As estimativas intervalares também foram quase idênticas, sendo de 0,002 a 0,020 no primeiro caso, contra 0,001 a 0,020 no segundo. Estima-se, portanto, um risco relativo de 1,11 de procedimentos para um aumento de  $10 \mu\text{g}/\text{m}^3$  na concentração de material particulado, ou seja, interpreta-se que um aumento na concentração mensal do poluente em  $10 \mu\text{g}/\text{m}^3$  provocaria um acréscimo de aproximadamente 11% nas quantidades mensais de inalação. Como anteriormente, não foram evidenciados efeitos das médias mensais de temperatura mínima ou do volume total de precipitação sobre os números mensais de procedimentos de inalação/nebulização (os respectivos IC95% cobrem o valor 0). O mesmo resultado pôde ser observado na análise com o poluente categorizado em quatro faixas. Pela Figura 13 é possível observar que assim como o ajuste com o banco de dados incompleto, o ajuste dos dados com imputação não evidenciou associação entre as categorias do poluente e as quantidades mensais de inalação/nebulização.



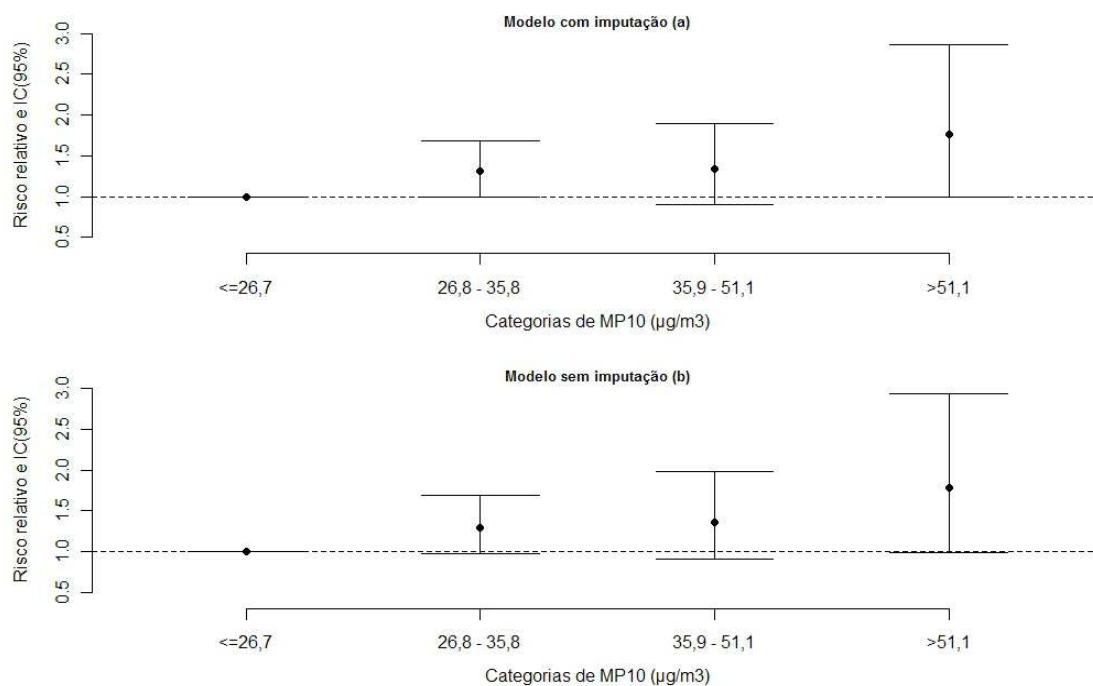


Figura 13. Riscos relativos e respectivos intervalos de credibilidade (95%) para inalações/nebulizações com concentrações crescentes do  $MP_{10}$ , considerando as estimativas dos modelos com e sem imputação dos dados.

Complementando a análise comparativa entre os modelos com dados completos e incompletos, foram calculados os valores de CPO e MCPO também para o ajuste dos dados com imputação. As estimativas combinadas obtidas podem ser observadas na Figura 14. Pelo gráfico podemos observar que os ajustes com os dados completos apresentaram maiores valores de CPO para a maioria das observações, sendo o modelo que também maximizou o MCPO. Tal fato seria um indicativo de que o modelo com os dados imputados ofereceria um melhor ajuste.

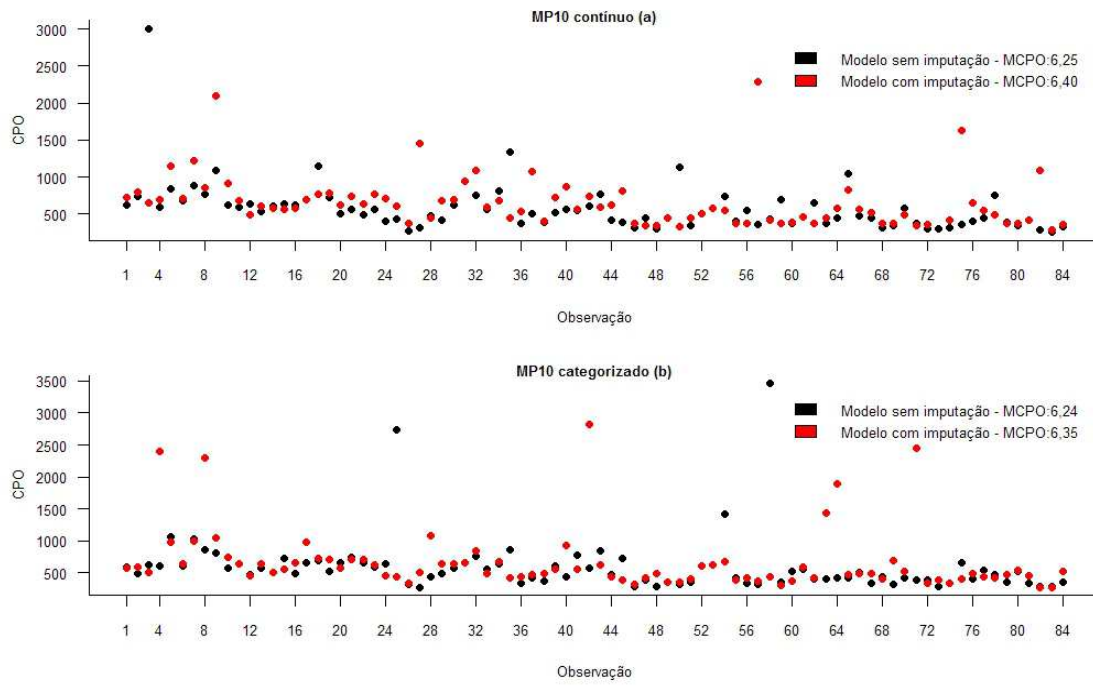


Figura 14. Gráfico de CPO *versus* o número da observação, considerando as estimativas do modelo 2 com e sem imputação dos dados.

## 6 DISCUSSÃO DOS RESULTADOS OBTIDOS

Uma das limitações deste estudo é o fato de se tratar de um estudo ecológico, em que a unidade de observação não é o indivíduo, mas um grupo de pessoas que representam uma parcela da população de Ribeirão Preto, formada pelos usuários do Sistema Único de Saúde (SUS) do município. Entretanto, estudos como estes têm se mostrado muito relevantes ao abordarem os efeitos da poluição sobre a saúde, já que estudos experimentais em seres humanos não podem ser realizados (BRAGA et al., 1999; FREITAS et al., 2004).

Outra limitação do estudo diz respeito ao período de análise relativamente curto (2004 a 2010), uma vez que se trata de uma série mensal de dados. Sabe-se que as variáveis sazonais que podem confundir a associação entre exposição e efeito podem ser mais bem controladas quanto maior for o período de tempo estudado. As informações quanto à demanda de procedimentos de inalação/nebulização foram, neste caso, um importante fator limitador, já que estes valores apenas são disponibilizados pelo SUS de forma mensal. No entanto, os dados de material particulado foram os principais causadores dessa limitação, uma vez que, no município de Ribeirão Preto, a medição do poluente não é realizada de maneira regular e, por isso, só foi possível utilizar os dados coletados a partir de janeiro de 2004.

A utilização de dados diários possibilitaria analisar se o efeito da poluição ocorreria com uma determinada defasagem (lag) de tempo. Além disso, sabe-se que grupos etários formados por crianças e idosos são mais suscetíveis a agravos e desenvolvimento de problemas respiratórios, devido a seus frágeis sistemas imunológicos. Sendo assim, o fato de não ser disponível pelo DATASUS as informações das inalações/nebulizações divididas por grupos etários foi mais um fator limitador do estudo.

A despeito das limitações e atendo-se aos resultados obtidos, o ajuste do modelo de regressão utilizando os dados imputados gerou resultados análogos ao modelo com o banco de dados incompleto. Talvez a baixa proporção de informações faltantes (cerca de 4,7%) tenha feito com que isso acontecesse. Neste caso, seria inevitável pensar que o tempo despendido para ajustar os cinco modelos utilizando o conjunto de dados completo poderia ter sido evitado, sendo o ajuste com os dados incompletos suficiente para se ter bons resultados. Esse pensamento vem ao encontro de autores como Harrell (2001), que define a escolha do método de imputação de acordo com a proporção de dados faltantes em determinada variável. Conforme o autor, para uma proporção de até

5% de não respostas, utilizar métodos de imputação única ou considerar na análise apenas os dados que tiverem todas as informações seria aceitável.

Entretanto, alguns estudos já observaram que as estimativas obtidas através de análises de dados com imputações são mais confiáveis do que na análise de um conjunto de dados com informações incompletas, mesmo que a porcentagem de dados faltantes seja baixa. Entre eles podemos citar o estudo de Nunes et al. (2009), que ao avaliar fatores de risco para a ocorrência do óbito de pacientes num período de 30 dias a partir da realização de uma laparotomia exploradora, gerou, a partir de um banco de dados completo, dois conjuntos de dados incompletos, um com 5% e outro com 20% de dados faltantes para a variável albumina (uma das variáveis preditoras do modelo de regressão). As pesquisadoras ajustaram três diferentes modelos de regressão logística: um considerando o conjunto de dados completo, outro para o conjunto com dados faltantes e outro com o conjunto de dados completado por imputação múltipla. As estimativas utilizando o banco de dados com os valores imputados ficaram bem mais próximas aos valores estimados pelo banco de dados completo do que as estimativas obtidas com o banco de dados sem 5% das informações de albumina. Tal discrepância foi ainda maior quando da análise do conjunto de dados com 20% de informações faltantes. Assim, as autoras chegaram à conclusão que a imputação múltipla foi superior à análise desconsiderando os casos incompletos, independente do percentual de não-respostas na base de dados.

Atualmente, os avanços computacionais têm permitido que o pesquisador faça uso de métodos de imputação de forma mais fácil e ágil. Sendo assim, em muitos casos é preferível não comprometer os resultados de uma pesquisa simplesmente desconsiderando os casos com informações faltantes e analisando o conjunto de dados incompletos, uma vez que dessa forma, a chance de se obter resultados enviesados é grande.

De qualquer forma, independente do modelo ajustado (com ou sem os dados imputados), as estimativas obtidas neste estudo evidenciaram que o material particulado produzido pelas queimadas tem influência sobre a saúde dos habitantes de Ribeirão Preto, dado que as variações mensais do poluente mostraram-se associadas às variações mensais de procedimentos de inalações e nebulizações.

Como este estudo, vários outros tem mostrado que a poluição do ar é a causa de agravos de saúde em milhares de pessoas em todo o mundo (SEATON et al., 1995; SCHWELA, 2000), especialmente aquelas portadoras de doenças respiratórias e

cardiovasculares (BROOK et al., 2010). Estima-se que a poluição do ar seja responsável por cerca de 5% de um total de mais de cinquenta milhões de mortes que ocorrem anualmente no planeta (WHO, 2000).

Mesmo que este estudo não tivesse encontrado qualquer indício de associação entre o poluente e as complicações respiratórias dos residentes de Ribeirão que os levaram a realizar procedimentos de inalação/nebulização, somente os transtornos causados pela queimada, como sujidade nas casas e em áreas públicas, danos nas linhas de transmissão, levando à interrupção do fornecimento de energia elétrica e os vários acidentes ocorridos nas rodovias próximas às áreas queimadas, devido à baixa visibilidade provocada pela fumaça, já seriam motivos suficientes para que a eliminação das queimadas fosse colocada em prática. Além disso, a degradação ambiental provocada pela queima de biomassa não deve ser esquecida, já que a queima diminui a fertilidade do solo, destrói a biodiversidade e ainda altera ecossistemas (FERREIRA, 2006).

No estado de São Paulo, a Lei Estadual nº 11.241 de setembro de 2002, regulamentada em março de 2003, dispõe sobre a gradativa eliminação da queima da cana-de-açúcar. Para as áreas mecanizáveis, o prazo se estende até o ano de 2021 e para as áreas não mecanizáveis, o prazo vai até 2031. De acordo com a lei, fica proibida a queima da cana a um quilômetro do perímetro urbano e de áreas indígenas; a cem metros de subestações de energia elétrica; a cinquenta metros de estações ecológicas, reservas e demais áreas de conservação; a vinte e cinco metros de estações de telecomunicações; a quinze metros de linhas de transmissão de energia; a quinze metros de áreas de domínio de ferrovias e rodovias estaduais e federais; e em áreas de plantio superiores a quinhentos hectares de terra. Está ainda previsto que as queimadas devem ser feitas preferencialmente no período noturno (período em que o fogo é gerado com menor intensidade, já que a temperatura é mais baixa e a umidade do ar é mais alta), além da obrigação por parte das usinas de sinalizarem as estradas próximas aos canaviais e manterem vigilância constante sobre o controle do fogo.

Entre as dificuldades para a eliminação total da queima que a lei paulista prevê, estão algumas restrições apresentadas pelos produtores e pelos próprios cortadores de cana. No primeiro caso, a restrição se dá ao alto preço das colheitadeiras mecânicas e ao aumento dos custos envolvidos com o transporte da cana, além das deficiências que existiriam na própria máquina. No segundo caso, as restrições se baseiam no fato de os cortadores manuais ganharem por produção. Uma vez que a queima da cana-de-açúcar

possibilita uma maior produtividade, além de ser a responsável por diminuir o risco de animais peçonhentos na área, sua eliminação prejudicaria os ganhos dos cortadores (BRAUNBECK & MAGALHÃES, 2004).

São nos canaviais em que se concentram a maior parte da força de trabalho agrícola no estado de São Paulo, trabalhadores esses, em sua maioria, com baixo nível de escolaridade (BRAUNBECK & MAGALHÃES, 2004). Por isso, talvez o maior questionamento sobre as vantagens e desvantagens da eliminação da queima da cana seja a de que essa ação provocaria a troca da mão-de-obra trabalhadora pelas colheitadeiras, deixando milhares de cortadores de cana desempregados.

Por outro lado, nos últimos anos, o número de mortes de trabalhadores em canaviais tem aumentado. Alguns pesquisadores, como Alves (2006), atribuem essas mortes ao extremo esforço físico a que esses trabalhadores se impõem na busca constante por uma maior produtividade no corte da cana. Por outro lado, Langowski (2007) acredita que essas mortes sejam causadas pelas más condições ambientais de trabalho às quais o cortador de cana queimada fica exposto. Isso porque a temperatura ambiente no canavial queimado se elevaria a até 45 °C. Além disso, o autor acredita que a substância particulada inalada por estes trabalhadores, pode estar associada aos casos de mortes por problemas cardíacos. Nesse sentido, muitos estudos tem sido desenvolvidos para analisar os agravos à saúde dos cortadores de cana provocados pela queima. Bosso (2000), por exemplo, identificou substâncias cancerígenas que estão presentes na fuligem da cana, na urina desses trabalhadores. Sendo assim, substituir a queima da cana por uma máquina colheitadeira seguramente não beneficiaria somente as condições de saúde das pessoas que residem próximas aos canaviais, mas também melhoraria a saúde do próprio trabalhador rural, exposto a condições inadequadas de trabalho.

Contudo, apesar dos benefícios, para que a colheita mecanizada da cana-de-açúcar gere uma real melhora na saúde dos cortadores, é preciso que programas de requalificação desses trabalhadores para que haja uma efetiva absorção destes pelo mercado de trabalho sejam desenvolvidos e efetivamente aplicados.

A proibição das queimadas, no entanto, não se restringe apenas às áreas onde há a colheita da cana. A legislação também proíbe a prática de queimadas urbanas e cabe aos municípios sua fiscalização.

Em Ribeirão Preto, de acordo com a prefeitura municipal, a fiscalização tem sido intensificada pelo Departamento de Fiscalização Geral, principalmente nas épocas

de estiagem, e, quando confirmada a infração, a Secretaria Municipal da Fazenda fica responsável por emitir uma multa ao proprietário do terreno que varia de R\$501,00 a R\$2.000,00, de acordo com o tamanho do terreno.

A prática desse tipo de queimada é antiga. Sendo assim, o combate a ela requer além da fiscalização do município, uma participação ativa da mídia local e ações educativas nas próprias escolas no intuito de que os donos de terrenos urbanos sejam conscientizados de que a queima de lixo e mato não é a melhor maneira de se manter o terreno limpo e que com essa prática, além de causar danos aos vizinhos, sua própria saúde é prejudicada.

Ainda que a lei que proíbe a queima da cana no estado de São Paulo seja cumprida efetivamente pelas usinas canavieiras, a cana continuará a ser queimada no restante do País. Da mesma forma ocorre com as queimadas urbanas. Portanto, a eficiência das campanhas de prevenção quanto à saúde do brasileiro, dependerá também da eficiência de todos os municípios do país no controle das queimadas (RIBEIRO, 2008).

A melhora na saúde pública só será efetiva se houver a união entre sociedade e governo na criação de políticas públicas, que ao mesmo tempo em que conscientizem as pessoas sobre a gravidade dos problemas gerados pelas queimadas, exijam dos órgãos competentes, mudanças imediatas na qualidade de vida e de educação da população. Quando isso ocorrer, o problema da poluição do ar, mesmo que não tiver sido solucionado, ao menos terá melhorado.

## 7 CONCLUSÕES E PERSPECTIVAS FUTURAS

O estudo evidenciou uma associação entre o material particulado emitido pela queima de biomassa e um indicador de complicações respiratórias, representado pelas quantidades de procedimentos de inalação/nebulização realizados em ambulatórios do SUS da cidade de Ribeirão Preto.

Adicionalmente, futuros estudos sobre a problemática poderiam enfocar, além de doenças e sintomas de doenças respiratórias, o surgimento de neoplasias, impactos às atividades diárias das pessoas diretamente afetadas e episódios de re-hospitalizações em decorrência de agravos à saúde.

Como bem citado por Arbex (2001), “estudos futuros deveriam ainda ampliar o tempo de avaliação utilizando estudos de coorte em populações de risco, como pacientes portadores de asma brônquica, portadores de doença pulmonar obstrutiva crônica e portadores de doenças cardiovasculares”.

Considerando que a saúde pública visa continuamente a promoção da saúde e a garantia da qualidade de vida da sociedade, este trabalho buscou colaborar para que novas suposições sobre o tema pudessem ser formuladas e assim, novos estudos pudessem ser realizados. Talvez o aumento de estudos na área seja uma forma de agilizar o processo de eliminação de todas as formas de queima, seja ela urbana ou rural, conscientizando a população da gravidade de seus efeitos contra a própria saúde.



## REFERÊNCIAS

ACHCAR, J. A. **Bayesian analysis for software reliability data**. In: BALAKRISHNAN, N.; RAO, C. R. (ed.), Handbook of statistics. Holland: Elsevier Science, 2001. v.20, p.733-748. (Advances in Reliability)

ALVES, F. Por que morrem os cortadores de cana? **Saúde e Sociedade**, v.15, n.3, p.90-98, 2006.

ARBEX, M. A.; BOHM, G. M.; SALDIVA, P. H.; CONCEIÇÃO, G. Assessment of the effects of sugar cane plantation burning on daily counts of inhalation therapy. **Journal of the Air & Waste Management Association**, v.50, n.10, p.1745-1749, 2000.

ARBEX, M. A. **Avaliação dos efeitos do material particulado proveniente da queima da plantação de cana-de-açúcar sobre a morbidade respiratória na população de Araraquara – SP**. São Paulo. 204p. Dissertação (Doutorado) – Faculdade de Medicina, Universidade de São Paulo, 2001.

ARBEX, M. A.; CANÇADO, J. E. D.; PEREIRA, L. A. A.; BRAGA, A. L. F.; SALDIVA, P. H. N. Queima de biomassa e efeitos sobre a saúde. **Jornal Brasileiro de Pneumologia**, v. 30, p.158-175, 2004.

AYOADE, J. O. **Introdução à climatologia para os trópicos**. Editora Bertrand Brasil. p.300-305, 1991.

BESAG, J. E. Discussion on the paper by Grenander and Miller. **Journal of the Royal Statistical Society**, v.56, p.591-592, 1994.

BOSSO, R. M. V. **Avaliação da atividade mutagênica da fuligem sedimentável proveniente da queima da cana-de-açúcar e da urina dos cortadores de cana através de ensaios com mutação gênica reversa em Salmonella Typhimurim**. 146 f. Dissertação (Mestrado) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2000.

BRAGA, A. L. F. et al. Air pollution and pediatric respiratory hospital admissions in São Paulo, Brazil. **Journal of Environmental Medicine**, vol.2, n.1, p.95-102, 1999.

BRAUNBECK, A. O.; MAGALHÃES, P. S. G. Colheita sustentável, com aproveitamento integral da cana. **Visão Agrícola**, v.1, n.1, p.72-78, 2004.

BROOK, R. D.; RAJAGOPALAN, S.; POPE, C. A. et al. Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. **Circulation**, v.121, n.21, p.2331-2378, 2010.

CANÇADO, J. E. D. **A poluição atmosférica e sua relação com a saúde humana na região canavieira de Piracicaba – SP**. Dissertação (Doutorado). São Paulo: Faculdade de Medicina da USP, 2003.

CARLIN, B. P.; LOUIS, T. A. **Bayes and empirical bayes methods for data analysis**. New York: Chapman and Hall, 1996.

CHIB, S.; GREENBERG, E. Understanding the metropolis-hastings algorithm. **The American Statistician**, v.49, n.4, p.327-335, 1995.

CRUTZEN, P. J.; ANDREAE, M. O. Biomass burning in the tropics: Impacts on atmospheric chemistry and biogeochemical cycles. **Science**, v.250, p.1669-1678, 1990.

DENNIS, R. J.; MALDONADO, D.; NORMAN, S.; BAENA, E.; MARTINEZ, G. Woodsmoke exposure and risk for obstructive airways disease among women. **Chest**, v.109, p.115-119, 1996.

DOSSING, M. Risk factors for chronic obstructive lung disease in Saudi Arabia. **Respiratory Medicine**, v.88, p.519-522, 1994.

DREW, D. **Processos interativos homem-meio ambiente**. Tradução de João Alves dos Santos. 4ª ed. Rio de Janeiro: Bertrand Brasil, 1998.

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária. **Agroecologia da cana-de-açúcar**. Disponível em < <http://www.cana.cnpm.embrapa.br/agroeco.html> > Acessado em setembro de 2010.

FATTINI, C. A.; DANGELO, J. G. **Anatomia humana sistêmica e segmentar para o estudante de medicina**. 2ª ed. São Paulo: Editora Atheneu, 2002.

FERREIRA, M. E. T. **A queimada da cana e seu impacto socioambiental**. 2006. Disponível em < <http://www.sucre-ethique.org/A-queimada-da-cana-e-seu-impacto.html> >. Acessado em setembro de 2010.

FREITAS, C. et al. Internações e óbitos e sua relação com a poluição atmosférica em São Paulo, 1993 a 1997. **Revista de Saúde Pública**, vol.38, n.6, p.751-757, 2004.

GEISSER, S.; EDDY, W. A predictive approach to model selection. **Journal of the American Statistical Association**, New York, v.74, p.153-160, 1979.

GELFAND, A. E.; SMITH, A. F. M. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, v.85, p.398-409, 1990.

GELFAND, A. E.; DEY, D. K. Bayesian model choice: asymptotics and exact calculations. **Journal of the Royal Statistical Society: Series B**, London, v.56, p.501-514, 1994.

GELFAND, A. E.; GOSH, S. K. Model choice. A minimum posterior predictive loss approach. **Biometrika**, London, v.85, p.398-409, 1998.

GELMAN, A.; RUBIN, B. D. Inference from iterative simulation using multiple sequences. **Statistical Science**, v.7, p.457-511, 1992.

GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v.6, p.721-741, 1984.

GEWEKE, J. **Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments.** In: J.O. Berger, J.M. Bernardo, A.P. Dawid, e A.F.M. Smith. (Eds.). *Bayesian Statistics 4*, Oxford: Oxford University Press, p.169-194, 1992.

GOUVEIA, N.; FREITAS, C. U. de; MARTINS, L. C.; MARCILIO, I. O. Hospitalizações por causas respiratórias e cardiovasculares associadas à contaminação atmosférica no Município de São Paulo, Brasil. **Cadernos de Saúde Pública**, vol.22, n.12, p. 2669-2677, 2006.

JACOBI, L. F. **Queimadas acidentais em campo em Santa Maria – RS.** Dissertação (Doutorado). Universidade Federal de Santa Maria, 2007.

JEFFREYS, H. **Theory of probability.** Oxford University Press, USA, 3a ed, 1998.

KENWARD, M. G.; CARPENTER, J. Multiple imputation: current perspectives. **Statistical Methods in Medical Research**, v.16, n.3, p.199-218, 2007.

KORSGAARD, I. R.; ANDERSEN, A. H. The additive genetic gamma frailty model. **Scandinavian Journal of Statistics**, v.25, p.255-269, 1998.

KOSSOVE, D. Smoke-filled rooms and lower respiratory disease in infants. **South African Medical Journal**, v.61, p.622-624, 1982.

LANGOWSKI, E. **Queima da cana: uma prática usada e abusada.** Cianorte. 2007. Disponível em: <<http://www.apromac.org.br/QUEIMA%20DA%20CANA.pdf>>. Acessado em setembro de 2010.

LI, H. Z. An additive genetic gamma frailty model for linkage analysis of diseases with variable age of onset using nuclear families. **Lifetime Data Analysis**, v.8, p.315-334, 2002.

LITTLE, R. J. A. Regression with missing Xs – a review. **Journal of the American Statistical Association**, v.87, p.227-37, 1992.

LOPES, F. S.; RIBEIRO, H. Mapeamento de internações hospitalares por problemas respiratórios e possíveis associações à exposição humana aos produtos da queima da palha de cana-de-açúcar no estado de São Paulo. **Revista Brasileira de Epidemiologia**, v.9, n.2, p.215-25, 2006.

MAIA, N. B.; MARTOS H. L.; BARRELLA, W. **Indicadores ambientais: conceitos e aplicações**. Pontifícia Universidade Católica de São Paulo, 2001.

MARTINS, L. C. et al. Poluição atmosférica e atendimentos por pneumonia e gripe em São Paulo, Brasil. **Revista de Saúde Pública**, vol.36, n.1, p. 88-94, 2002.

MCCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**. Chapman and Hall: London, 1989.

MISHRA, V. K.; RETHEFORD, R. D.; SMITH, K. R. Biomass cooking fuels and prevalence of tuberculosis in India. **International Journal of Infectious Diseases**, v.3, p. 119-129, 1999.

MONTEIRO, C. A. F. **Teoria e clima urbano**. São Paulo, Universidade de São Paulo: 181 p, 1976.

NUNES, L. **Métodos de imputação de dados aplicados na área da saúde**. Porto Alegre. 120p. Dissertação (Doutorado) – Faculdade de Medicina, Universidade Federal do Rio Grande do Sul, 2007.

NUNES, L. N.; KLÜCK, M. M.; FACHEL, J. M. G. Uso da imputação múltipla de dados faltantes: uma simulação utilizando dados epidemiológicos. **Cadernos de saúde pública**, v.25, n.2, p.268-278, 2009.

OLIVEIRA, L.; BARROCAS, R. **Vantagens e desvantagens na queimada da cana no estado de São Paulo, Brasil**. 2004. Disponível em <

<http://observatoriogeograficoamericalatina.org.mx/egal8/Geografiasocioeconomica/Geografiaagricola/04.pdf> > Acessado em setembro de 2010.

PANDEY, M. R.; BOLEIJ, J. S.; SMITH, K. R.; WAFULA, E. M. Indoor air pollution in developing countries and acute respiratory infection in children. **Lancet**, v.1, p.427-429, 1989.

PAULINO, C.; TURKMAN, M.; MURTEIRA, B. **Estatística Bayesiana**. Lisboa, Portugal: Fundação Calouste Gulbenkian, 446 p., 2002.

PEREIRA, L. A. A. et al. Association between air pollution and intrauterine mortality in São Paulo, Brazil. **Environmental Health Perspectives**, l.106, n.6, p.325-329, 1998.

PEREZ-PADILLA R.; PEREZ-GUZMAN C.; BAEZ-SALDANA R.; TORRES-CRUZ A. Cooking with biomass stoves and tuberculosis: a case control study. **The International Journal of Tuberculosis and Lung Disease**, v.5, p.441-7, 2001.

PITTON, S. E. e DOMINGOS, A. E. Tempos e doenças: efeitos dos parâmetros climáticos nas crises hipertensivas nos moradores de Santa Gertrudes – SP. **Estudos geográficos**. Rio Claro, v.2, n.1, p.75-86, 2004.

R Development Core Team **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. 2010. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

RIBEIRO, H. Queimadas de cana-de-açúcar no Brasil: efeitos à saúde respiratória. **Revista de Saúde Pública**, v.42, p.370-376, 2008.

RUBIN, D. B. **Multiple imputation for nonresponse in surveys**. New York: Wiley, 1987.

RUBIN, D. B. Multiple imputation after 18+ years. **Journal of the American Statistical Association**, v.91, p.473-89, 1996.

SÃO PAULO. Lei n. 11.241, de 19 de setembro de 2002. Dispõe sobre a eliminação gradativa da queima da palha da cana-de-açúcar e dá providências correlatas.

SCHAFER, J. L. **Analysis of incomplete multivariate data**. London: Chapman & Hall, 1997.

SCHAFER, J. L. Multiple imputation: a primer. **Statistical Methods in Medical Research**, v.8, p.3-15, 1999.

SCHAFER, J. L.; GRAHAM, J. W. Missing data: our view of the state of the art. **Psychological Methods**, v.7, p.147-77, 2002.

SCHWELA, D. Air pollution and health in urban areas. **Reviews on Environmental Health**, v.15, n.1-2, p.13-42, 2000.

SEATON, A.; MACNEE, W.; DONALDSON, K.; GODDEN, D. Particulate air pollution and acute health effects. **Lancet**, v.345, n.8943, p.176-178, 1995.

SOUZA, C. G., NETO, J. L. S. Geografia da saúde e climatologia médica: ensaios sobre a relação clima e vulnerabilidade. **HYGEIA, Revista Brasileira de Geografia Médica e da Saúde**, v.3, n.6, p.116-126, 2008.

SPIEGELHALTER, D. J.; MYLES, J. P.; JONES, D. R.; ABRAMS K. R. Bayesian methods in health technology assessment: a review. **Health Technology Assessment**, v.4, n.38, p.1-130, 2000.

SPIEGELHALTER, D.; LUNN, D.J.; THOMAS, A.; BEST, N. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. **Statistics and Computing**, v.10, p.325-337, 2000.

SUITS, D. B. Use of dummy variables in regression equations. **Journal of the American Statistical Association**, v.52, n.280, p.548-551, 1957.

TRESMONDI, A. C. C. L. **Qualidade do ar na area de influência do pólo industrial de Paulínia-SP: 2000-2002**. Dissertação (Doutorado). UNICAMP. Campinas. 303 p, 2003.

ÚNICA – União da Agroindústria Canavieira de São Paulo. **Agroindústria da cana-de-açúcar: Alta competitividade canavieira**. Disponível em < [http://www.unica.com.br/pages/agroindustria\\_alta.asp](http://www.unica.com.br/pages/agroindustria_alta.asp) > Acessado em setembro de 2010.

VAN BUUREN, S.; OUDSHOORN, C. G. M. **Multivariate imputation by chained equations**. MICE V1.0 user's manual. Leiden: TNO Preventie en Gezondheid, 2000.

VISMARA, L.S.; KARAM, D.; MORITA, L.H.M. Aplicação das inferências clássica e bayesiana na estimação dos parâmetros do modelo de densidade populacional de plantas daninhas. **Planta Daninha**. Viçosa-MG, v.25, n.4, p.661-669, 2007.

WHITE, I. A.; WOOD, A.; ROYSTON, P. Editorial: multiple imputation in practice. **Statistical Methods in Medical Research**, v.16, p.195-197, 2007.

WHO – World Health Organization. **Air Pollution**. Fact Sheet, n.187, 2000.

WHO – World Health Organization. **Air Quality Guidelines: Global Update 2005**. Copenhagen: World Health Organization, 2006. Disponível em: [http://www.euro.who.int/\\_\\_data/assets/pdf\\_file/0005/78638/E90038.pdf](http://www.euro.who.int/__data/assets/pdf_file/0005/78638/E90038.pdf).

WOLFINGER, R. D. **Fitting nonlinear mixed models with the new NLMIXED procedure**. Cary: SAS Institute Inc, 2000.

ZHANG, P. Multiple imputation: theory and method. **International Statistical Review**, v.71, p.581-592, 2003.



## A APÊNDICE - Programas

Este apêndice apresenta alguns programas computacionais utilizados para que os ajustes de regressão apresentados na dissertação fossem realizados.

### A.1 Seção 4.1.4

Os programas desenvolvidos no software WinBugs para os ajuste dos Modelos 1, 2 e 3 estão apresentado a seguir.

#### A.1.1 Modelo 1 (material particulado contínuo)

---

#### Listagem 1: Código do Software WinBugs.

---

```

1 model
2 {
3 for( i in 1 : n ) {
4   inalacao[i] ~ dpois(lambda[i])
5   log(lambda[i]) <- beta0+beta1*(MP10[i]-mean(MP10[]))+ beta2*(temperatura[i]-
6   mean(temperatura[]))+ beta3*(precipitacao[i]- mean(precipitacao[]))+
7   beta4*outono[i]+beta5*inverno[i]+beta6*primavera[i]
8     }
9     beta0 ~ dnorm(0,0.1)
10    beta1 ~ dnorm(0,1)
11    beta2 ~ dnorm(0,1)
12    beta3 ~ dnorm(0,1)
13    beta4 ~ dnorm(0,1)
14    beta5 ~ dnorm(0,1)
15    beta6 ~ dnorm(0,1)
16 }

```

## A.1.2 Modelo 1 (material particulado categorizado)

Listagem 2: Código do Software WinBugs.

---

```
1 model
2 {
3 for( i in 1 : n ) {
4     inalacao[i] ~ dpois(lambda[i])
5     log(lambda[i]) <- beta0+beta1*x1[i]+beta2*x2[i]+beta3*(temperatura[i]-
6     mean(temperatura[])) + beta4*(precipitacao[i]-mean(precipitacao[])) +
7     beta5*outono[i]+beta6*inverno[i]+beta7*primavera[i]
8     }
9     beta0 ~ dnorm(0,0.1)
10    beta1 ~ dnorm(0,1)
11    beta2 ~ dnorm(0,1)
12    beta3 ~ dnorm(0,1)
13    beta4 ~ dnorm(0,1)
14    beta5 ~ dnorm(0,1)
15    beta6 ~ dnorm(0,1)
16    beta7 ~ dnorm(0,1)
17 }
```

## A.1.3 Modelo 2 (material particulado contínuo)

Listagem 3: Código do Software WinBugs.

---

```
1 model
2 {
3 for( i in 1 : n ) {
4     w[i] ~ dnorm(0,tau)
5     inalacao[i] ~ dpois(lambda[i])
6     log(lambda[i])<- beta0+beta1*(MP10[i]-mean(MP10[]))+beta2*(temperatura[i]-
7     mean(temperatura[])) + beta3*(precipitacao[i]-mean(precipitacao[])) +
8     beta4*outono[i]+beta5*inverno[i]+beta6*primavera[i]+w[i]
9     }
10     tau ~ dgamma(1,1)
11     beta0 ~ dnorm(0,0.1)
12     beta1 ~ dnorm(0,1)
13     beta2 ~ dnorm(0,1)
14     beta3 ~ dnorm(0,1)
15     beta4 ~ dnorm(0,1)
16     beta5 ~ dnorm(0,1)
17     beta6 ~ dnorm(0,1)
18 }
```

## A.1.4 Modelo 2 (material particulado categorizado)

Listagem 4: Código do Software WinBugs.

---

```
1 model
2 {
3   for( i in 1 : n ) {
4     w[i] ~ dnorm(0,tau)
5     inalacao[i] ~ dpois(lambda[i])
6     log(lambda[i]) <- beta0+beta1*x1[i]+beta2*x2[i]+beta3*(temperatura[i]-
7     mean(temperatura[])) + beta4*(precipitacao[i]-mean(precipitacao[])) +
8     beta5*outono[i]+beta6*inverno[i]+beta7*primavera[i]+w[i]
9     }
10    tau ~ dgamma(1,1)
11    beta0 ~ dnorm(0,0.1)
12    beta1 ~ dnorm(0,1)
13    beta2 ~ dnorm(0,1)
14    beta3 ~ dnorm(0,1)
15    beta4 ~ dnorm(0,1)
16    beta5 ~ dnorm(0,1)
17    beta6 ~ dnorm(0,1)
18    beta7 ~ dnorm(0,1)
19 }
```

## A.1.5 Modelo 3 (material particulado contínuo)

Listagem 5: Código do Software WinBugs.

---

```
1 model
2 {
3   for( i in 1 : n ) {
4     w[i] ~ dgamma(1,1)
5     inalacao[i] ~ dpois(lambda[i])
6     lambda[i] <- exp(beta0+beta1*(MP10[i]-mean(MP10[]))+beta2*(temperatura[i]-
7     mean(temperatura[])) + beta3*(precipitacao[i]-mean(precipitacao[])) +
8     beta4*outono[i]+beta5*inverno[i]+beta6*primavera[i])*w[i]
9     }
10    beta0 ~ dnorm(0,0.1)
11    beta1 ~ dnorm(0,1)
12    beta2 ~ dnorm(0,1)
13    beta3 ~ dnorm(0,1)
14    beta4 ~ dnorm(0,1)
15    beta5 ~ dnorm(0,1)
16    beta6 ~ dnorm(0,1)
17 }
```

## A.1.6 Modelo 3 (material particulado categorizado)

Listagem 6: Código do Software WinBugs.

---

```
1 model
2 {
3 for( i in 1 : n ) {
4     w[i] ~ dgamma(1,1)
5     inalacao[i] ~ dpois(lambda[i])
6     lambda[i] <- exp(beta0+beta1*x1[i]+beta2*x2[i]+beta3*(temperatura[i]-
7     mean(temperatura[])) + beta4*(precipitacao[i]-mean(precipitacao[])) +
8     beta5*outono[i]+beta6*inverno[i]+beta7*primavera[i])*w[i]
9     }
10     beta0 ~ dnorm(0,0.1)
11     beta1 ~ dnorm(0,1)
12     beta2 ~ dnorm(0,1)
13     beta3 ~ dnorm(0,1)
14     beta4 ~ dnorm(0,1)
15     beta5 ~ dnorm(0,1)
16     beta6 ~ dnorm(0,1)
17     beta7 ~ dnorm(0,1)
18 }
```