



UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CAMPUS DE CURITIBA
DEPARTAMENTO DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA
E DE MATERIAIS - PPGEM

YARA DE SOUZA TADANO

ANÁLISE DO IMPACTO DE MP₁₀ NA SAÚDE
POPULACIONAL: ESTUDO DE CASO EM
ARAUCÁRIA, PR

CURITIBA

FEVEREIRO - 2007

Livros Grátis

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

Milhares de livros grátis para download.

YARA DE SOUZA TADANO

**ANÁLISE DO IMPACTO DE MP₁₀ NA SAÚDE
POPULACIONAL: ESTUDO DE CASO EM
ARAUCÁRIA, PR**

Dissertação apresentada como requisito parcial à obtenção do título de Mestre em Engenharia, do Programa de Pós-Graduação em Engenharia Mecânica e de Materiais, Área de Concentração em Engenharia Térmica, do Departamento de Pesquisa e Pós-Graduação, do Campus de Curitiba, da UTFPR.

Orientador: Prof. Cássia Maria Lie Ugaya, Dr. Eng.

Co-orientador: Prof. Admilson Teixeira Franco, Dr. Eng.

CURITIBA

FEVEREIRO - 2007

TERMO DE APROVAÇÃO

YARA DE SOUZA TADANO

ANÁLISE DO IMPACTO DE MP₁₀ NA SAÚDE POPULACIONAL: ESTUDO DE CASO EM ARAUCÁRIA, PR

Esta Dissertação foi julgada para a obtenção do título de mestre em engenharia, área de concentração em engenharia térmica, e aprovada em sua forma final pelo Programa de Pós-graduação em Engenharia Mecânica e de Materiais.

Prof. Neri Volpato, Ph.D.
Coordenador de Curso

Banca Examinadora

Prof. Cássia Maria Lie Ugaya, Dr. Eng.
(UTFPR)

Prof. Lourdes Conceição Martins, Dr.
(Faculdade de Medicina do ABC)

Prof. Cezar Otaviano Ribeiro Negrão,
Ph.D. (UTFPR)

Prof. Márcia Silva de Araújo, Ph.D.
(UTFPR)

Curitiba, 15 de Fevereiro de 2007

AGRADECIMENTOS

A Deus por me dar forças para enfrentar as dificuldades.

Aos meus pais Maria das Graças e Mario pelo incentivo e compreensão.

Aos meus orientadores Cássia e Admilson pela orientação, dedicação e paciência.

À Agência Nacional do Petróleo (ANP) e à Financiadora de Estudos e Projetos (FINEP) por meio do Programa de Recursos Humanos da ANP para o Setor de Petróleo e Gás – PRH – ANP/MCT (PRH10 – UTFPR) pelo apoio financeiro.

Ao Instituto Ambiental do Paraná (IAP), Secretaria Municipal do Meio Ambiente de Araucária, Lactec (Instituto de Tecnologia para o Desenvolvimento) e Secretaria Municipal da Saúde de Araucária pelo fornecimento dos dados necessários para a realização do trabalho.

Aos professores Lourdes Conceição Martins e Alfésio Luís Ferreira Braga por ajudarem na realização das análises estatísticas.

Aos meus amigos Eveline, Paula, Carla, Augusto e Felipe.

Ao Marcelo por ter acreditado na minha capacidade e me dar forças durante estes dois anos difíceis.

“No fim tudo dá certo, se não deu certo é porque ainda não chegou ao fim.”

Fernando Sabino

TADANO, Yara de Souza, **Análise do Impacto de MP₁₀ na Saúde Populacional: Estudo de Caso em Araucária, PR**, 2007, Dissertação (Mestrado em Engenharia) - Programa de Pós-graduação em Engenharia Mecânica e de Materiais, Universidade Tecnológica Federal do Paraná, Curitiba, 119p.

RESUMO

As atividades de refino de petróleo são responsáveis por 42% das emissões de poluentes atmosféricos industriais na Região Metropolitana do Rio de Janeiro, apesar do número reduzido de indústrias de refino de petróleo nessa região. Os modelos estatísticos mais utilizados para avaliar o impacto da poluição atmosférica na saúde populacional são as análises de regressão, pois são ferramentas capazes de relacionar uma ou mais variáveis explicativas (variáveis independentes) com uma única variável resposta (variável dependente). O objetivo deste estudo foi encontrar um modelo estatístico capaz de avaliar, bem como prever o impacto do MP₁₀ na saúde humana, para então aplicá-lo para a população do município de Araucária. O modelo estatístico escolhido foi o modelo linear generalizado para regressão de Poisson, em que o número de pessoas com doenças respiratórias por faixa etária, bem como o número de crianças com pneumonia e gripe foram considerados como variáveis respostas. Utilizando apenas a concentração de MP₁₀ como variável explicativa pode-se superestimar o coeficiente de regressão, pois não estão sendo considerados outros fatores que afetam a saúde humana como a temperatura e umidade, assim como a grande diferença no número de atendimentos hospitalares entre os dias da semana e nos feriados. Devido a estes fatores foram consideradas como variáveis explicativas a concentração de MP₁₀, a temperatura, a umidade, os dias da semana e feriados. O modelo se ajustou melhor aos dados de doenças respiratórias em maiores de 60 anos e de pneumonia e gripe em crianças entre 5 e 14 anos. Apesar da população do município de Araucária ser predominantemente adulta, os gráficos de risco relativo mostraram que crianças e idosos foram os mais afetados pela concentração de MP₁₀. Baseado na análise realizada pode-se concluir que o modelo de regressão de Poisson é adequado para avaliar, bem como prever impactos do MP₁₀ na saúde populacional.

Palavras-chave: análise de regressão, material particulado, impacto ambiental.

TADANO, Yara de Souza, **Analysis of PM₁₀ impact on population's health: Case Study in Araucaria, PR**, 2007, Dissertation (Master of Engineering) – Post-graduate program in Mechanical and Material Engineering, Federal Technological University of Parana, Curitiba, 120p.

ABSTRACT

Despite the reduced number of oil refining industries in the metropolitan region of Rio de Janeiro, the oil refine activities are responsible for 42% of the industrial air pollutants emissions in this region. The most used statistical model to evaluate the relation between air pollution and population's health is regression analysis, as it is able to relate one or more explanatory variables with one response variable. This research aims to find a statistical model to evaluate and predict the impact of PM₁₀ concentration on human health and apply it to Araucaria's city. After studying several statistical models, the generalized linear model with Poisson regression was chosen due to the explanatory variable characteristics. The number of people with respiratory diseases or the number of children with pneumonia and influenza were considered as response variables. Using only PM₁₀ concentration as explanatory variable can over estimate the regression coefficient because it does not consider some important factors that affect human health as temperature and humidity; as well as the huge difference between hospital admission on days of week and holidays. Therefore, temperature, humidity, days of week and holidays were also considered as explanatory variables. The model for respiratory diseases in elderly people and pneumonia and influenza in 5 to 14 years old children had the best fits. The relative risk plots showed that children and elderly people were the most affected by PM₁₀ concentration, despite the predominance of adults in Araucaria's city. The results showed that it is possible to use the Poisson regression model to evaluate and predict PM₁₀ concentration on population's health.

Keywords: regression analysis, particulate matter, environmental impact.

SUMÁRIO

RESUMO.....	v
ABSTRACT	vi
LISTA DE FIGURAS	ix
LISTA DE TABELAS	xiii
LISTA DE QUADROS	xv
LISTA DE ABREVIATURAS E SIGLAS	xvi
LISTA DE SÍMBOLOS.....	xix
1 INTRODUÇÃO.....	1
1.1 Material Particulado e a Saúde Humana.....	3
1.2 Fontes Emissoras de Material Particulado.....	7
1.3 Objetivos.....	8
1.4 Estrutura da Dissertação	9
2 REVISÃO BIBLIOGRÁFICA	10
2.1 Impacto da Poluição Atmosférica na Saúde	10
2.2 Análise dos Modelos Estatísticos para Estudos Epidemiológicos	11
2.3 Análise de Regressão	16
2.3.1 Modelos de regressão	18
2.3.2 Modelos lineares generalizados.....	19
2.3.3 Modelos de regressão de Poisson	21
3 METODOLOGIA	24
3.1 Coleta e Análise dos Dados	24
3.2 Escolha do Modelo Estatístico	25
3.2.1 Estimativa dos parâmetros.....	27
3.2.2 Ajuste de tendências temporais	28
3.2.3 Avaliação do ajuste	32
3.2.4 Análise gráfica.....	33
3.3 Análise dos Resultados.....	35
3.4 Risco Relativo.....	36
4 ESTUDO DE CASO.....	39
4.1 Características do Município de Araucária	39
4.2 Dados de Saúde da Região Sul do Brasil.....	42
4.3 Coleta e Análise dos Dados	43
4.4 Resultados Preliminares	47
4.5 Escolha do Modelo Estatístico	52
4.6 Aplicação do Modelo de Regressão de Poisson.....	57
4.7 Análise por Faixa Etária	61

4.7.1	Análise para avaliação	61
4.7.2	Análise para previsão	75
5	CONCLUSÕES.....	84
5.1	Sugestões.....	86
	PRODUÇÃO CIENTÍFICA NO PERÍODO (Fev. 2005 – Fev. 2007)	87
	REFERÊNCIAS.....	88
	Anexo A – Informações sobre as Estações de Monitoramento da Qualidade do Ar da RMC	95
	Anexo B – Distribuição t de <i>Student</i>	96
	Anexo C – Estimativa dos Parâmetros de Regressão (Paula, 2004)	97

LISTA DE FIGURAS

Figura 1.1 – Porcentagem de aumento de doenças respiratórias e intervalo de confiança de 95% quando os poluentes foram incluídos em modelos simples (símbolos brancos) e em modelos com todos os poluentes (símbolos pretos) (Braga et al., 2001)	4
Figura 1.2 - Regiões do aparelho respiratório de acordo com a deposição de material particulado em função do diâmetro médio e da fração de deposição	6
Figura 2.1 – Gráfico da distribuição de Poisson para quatro valores de μ (Schmidt, 2003)	22
Figura 3.1 - Representação gráfica de um histograma com distribuição aproximadamente normal (Hair Jr. et al., 2005)	26
Figura 3.2 – Matriz de dispersão (Adaptado de Hair Jr. et al., 2005)	27
Figura 3.3 – Exemplo de gráfico da função de autocorrelação (FAC) parcial em relação a dias de defasagem (<i>lag</i>)	31
Figura 3.4 – Exemplo de gráfico dos desvios residuais em função dos valores ajustados pelo modelo	34
Figura 3.5 – Exemplo de gráfico dos valores observados em função dos valores ajustados pelo modelo	35
Figura 3.6 – Fluxograma de escolha do modelo estatístico apropriado aos dados coletados.....	38
Figura 4.1 – Mapa da Região Metropolitana de Curitiba (Adaptado de IPPUC, 2006)	40
Figura 4.2 - População do município de Araucária desde a década de 70 até o ano 2000 (IBGE, 2000 apud Prefeitura de Araucária, 2005).....	41
Figura 4.3 – Localização das estações de monitoramento da qualidade do ar na RMC (Adaptado de IAP, 2005).....	42
Figura 4.4 – Porcentagem de dados faltantes para o ano de 2004.....	45
Figura 4.5 – Porcentagem de dados faltantes para o ano de 2005.....	45

Figura 4.6 – Gráfico da percentagem mensal de pessoas com problemas respiratórios por faixa etária em Araucária para o período de 2004 a 2005.....	47
Figura 4.7 – Gráficos do número de pessoas com DR total diário (A), concentração de MP ₁₀ (B), temperatura (C) e umidade (D) médias diárias no município de Araucária para 2004 e 2005.....	49
Figura 4.8 – Gráfico da percentagem de pessoas com doenças respiratórias por faixa etária em relação à temperatura média mensal	50
Figura 4.9 – Gráfico da percentagem de pessoas com doenças respiratórias por faixa etária em relação à média mensal da umidade relativa do ar	51
Figura 4.10 – Gráfico da percentagem de pessoas com doenças respiratórias por faixa etária em relação à concentração média mensal de MP ₁₀	51
Figura 4.11 – Histogramas das variáveis respostas.....	53
Figura 4.12 – Matrizes de dispersão para pessoas menores de 4 anos (A), entre 5 e 14 anos (B), entre 15 e 19 anos (C) e entre 20 e 59 anos (D)	56
Figura 4.13 – Matriz de dispersão para pessoas maiores de 60 anos (E)	57
Figura 4.14 – Gráficos de desvios residuais e valores observados em função dos valores ajustados para o modelo com ligação logarítmica (A) e o modelo com ligação raiz quadrada (B)	60
Figura 4.15 – Função de autocorrelação parcial em relação aos dias de defasagem nos modelos sem resíduo para pessoas menores de 4 anos (A), entre 5 e 14 anos (B), entre 15 e 19 anos (C), entre 20 e 59 anos (D) e maiores de 60 anos (E)	62
Figura 4.16 - Função de autocorrelação parcial em relação aos dias de defasagem nos modelos com resíduo para pessoas menores de 4 anos (A), entre 5 e 14 anos (B), entre 15 e 19 anos (C), entre 20 e 59 anos (D)	63
Figura 4.17 – Gráficos dos desvios residuais em relação aos valores ajustados para pessoas menores de 4 anos (A) e entre 5 e 14 anos (B), para modelo sem resíduo (1) e com resíduo (2).....	64

Figura 4.18 – Gráficos dos desvios residuais em relação aos valores ajustados para pessoas entre 15 e 19 anos (C), entre 20 e 59 anos (D) e maiores de 60 anos (E) para modelo sem resíduo (1) e com resíduo (2)	65
Figura 4.19 - Gráficos dos valores observados em relação aos valores ajustados para pessoas menores de 4 anos (A), entre 5 e 14 anos (B) e entre 15 e 19 anos (C), para modelo sem correção (1) e com correção (2).....	66
Figura 4.20 - Gráficos dos valores observados em relação aos valores ajustados para pessoas entre 20 e 59 anos (D) e maiores de 60 anos (E) para modelo sem correção (1) e com correção (2)	67
Figura 4.21 – Gráficos dos desvios residuais e dos valores observados em relação aos valores ajustados para pneumonia e gripe em crianças menores de 4 anos (A) e entre 5 e 14 anos (B) para modelo com resíduo	70
Figura 4.22 – Risco relativo com intervalo de confiança de 95% de doenças respiratórias para cada faixa etária associado ao aumento na concentração de MP_{10}	72
Figura 4.23 - Risco relativo com intervalo de confiança de 95% de doenças respiratórias para cada faixa etária associado a um aumento de $100 \mu g/m^3$ na concentração de MP_{10}	74
Figura 4.24 – Risco relativo com intervalo de confiança de 95% para pneumonia e gripe em crianças associado a um aumento de $100 \mu g/m^3$ na concentração de MP_{10}	75
Figura 4.25 – Gráficos dos desvios residuais em relação aos valores ajustados para pessoa menores de 4 anos (A), entre 5 e 14 anos (B), entre 15 e 19 anos (C) e entre 20 e 59 anos (D)	76
Figura 4.26 – Gráfico dos desvios residuais em relação aos valores ajustados para pessoas maiores de 60 anos (E).....	77
Figura 4.27 - Gráficos dos valores observados em relação aos valores ajustados para pessoas menores de 4 anos (A), entre 5 e 14 anos (B), entre 15 e 19 anos (C) e entre 20 e 59 anos (D)	77

Figura 4.28 - Gráfico dos valores observados em relação aos valores ajustados para pessoas maiores de 60 anos (E).....78

Figura 4.29 – Gráficos dos desvios residuais e dos valores observados em relação aos valores ajustados para pneumonia e gripe em crianças menores de 4 anos (A) e entre 5 e 14 anos (B).....80

LISTA DE TABELAS

Tabela 2.1 – Funções de ligação canônica de algumas distribuições da família exponencial (McCullagh e Nelder, 1989)	20
Tabela 4.1 – População do município de Araucária nos anos de 2004 e 2005 por faixa etária (Ministério da Saúde e IBGE, 2006)	41
Tabela 4.2 – Dados epidemiológicos e demográficos por região brasileira, bem como para os estados da região Sul (Ministério da Saúde e IBGE, 2006)	43
Tabela 4.3 – Porcentagem de dados de cada variável fornecida pelo IAP e SMMA/Lactec.....	44
Tabela 4.4 – Estatística descritiva de DR para cada faixa etária	54
Tabela 4.5 – Estatística descritiva para variáveis explicativas	55
Tabela 4.6 – Pseudo R^2 , estatística de Pearson (χ^2) e razão entre χ^2 e graus de liberdade para modelos com funções de ligação diferentes.....	59
Tabela 4.7 – Pseudo R^2 , estatística de Pearson (χ^2) e razão entre χ^2 e graus de liberdade dos modelos com e sem acréscimo de resíduo por faixa etária	64
Tabela 4.8 – Coeficientes de regressão, erro padrão e valores de t para cada faixa etária para modelo com resíduo**	69
Tabela 4.9 – Pseudo R^2 , estatística de Pearson (χ^2) e razão entre χ^2 e graus de liberdade para modelo com resíduo em crianças com pneumonia e gripe	69
Tabela 4.10 – Coeficientes de regressão, erro padrão e valores de t para pneumonia e gripe em crianças para modelo com resíduo	71
Tabela 4.11 – Pseudo R^2 , estatística de Pearson (χ^2) e razão entre χ^2 e graus de liberdade dos modelos com variável mês sem resíduo por faixa etária	76
Tabela 4.12 – Coeficientes de regressão, erro padrão e valores de t para cada faixa etária	79

Tabela 4.13 – Pseudo R^2 , estatística de Pearson (χ^2) e razão entre χ^2 e graus de liberdade para o modelo com variável mês de crianças com pneumonia e gripe	79
Tabela 4.14 – Coeficientes de regressão, erro padrão e valores de t para pneumonia e gripe em crianças	81
Tabela 4.15 – Coeficientes de regressão com respectivos erros padrões para todas as variáveis explicativas de todas as análises realizadas	82
Tabela B.1.1 – Distribuição t de <i>Student</i> (Fonseca e Martins, 1996)	96

LISTA DE QUADROS

Quadro 1.1 - Eventos de poluição atmosférica excessiva (Derisio, 2000)	2
Quadro 1.2 – Comparação entre partículas finas e espessas (EPA, 1998)	5
Quadro 2.1 - Exemplos de estudos sobre o impacto da poluição do ar na saúde humana	12
Quadro 2.2 – Estudos com enfoque na comparação entre modelos estatísticos.....	13
Quadro 2.3 - Histórico dos estudos que levaram à modelagem estatística denominada MLG (McCullagh e Nelder, 1989).....	19
Quadro A.1 - Estações de monitoramento da qualidade do ar do município de Araucária no ano de 2005 (IAP, 2005).....	95

LISTA DE ABREVIATURAS E SIGLAS

Al	- Alumínio
CaCO ₃	- Carbonato de Cálcio
Cd	- Cádmiio
CETESB	- Companhia de Tecnologia de Saneamento Ambiental
CIAR	- Centro Industrial de Araucária
CIC	- Cidade Industrial de Curitiba
CO	- Carbono
CONAMA	- Conselho Nacional do Meio Ambiente
COV	- Compostos orgânicos voláteis
Cu	- Cobre
DR	- Doenças Respiratórias
EPA	- <i>Environmental Protection Agency</i>
FAC	- Função de autocorrelação
Fe	- Ferro
gl	- Graus de liberdade
GNV	- Gás natural veicular
H ⁺	- Íon de Hidrogênio
HPAs	- Hidrocarbonetos Policíclicos Aromáticos
IAP	- Instituto Ambiental do Paraná
IBAMA	- Instituto Brasileiro de Meio Ambiente e dos Recursos Naturais Renováveis
IBGE	- Instituto Brasileiro de Geografia e Estatística
IC	- Intervalo de Confiança
IM	- Infarto do Miocárdio

Kg	- Kilogramas
Lactec	- Instituto de Tecnologia para o Desenvolvimento
MAG	- Modelos Aditivos Generalizados
mbar	- Milibar
ml	- Mililitros
MLG	- Modelos Lineares Generalizados
Mn	- Manganês
MP	- Material Particulado
MP ₁₀	- Partículas com diâmetro aerodinâmico menor ou igual a 10µm
MP _{2,5}	- Partículas com diâmetro aerodinâmico menor ou igual a 2,5µm
MP _{2,5-10}	- Partículas com diâmetro aerodinâmico entre 2,5 e 10µm
NaCl	- Cloreto de Sódio
NH ₃	- Amônia
NH ₄ ⁺	- Amônio
NHPAs	- Hidrocarbonetos Policíclicos Aromáticos Nitrogenados
Ni	- Níquel
NO ₂	- Dióxido de Nitrogênio
NO ₃ ⁻	- Nitrato
NO _x	- Óxidos de Nitrogênio
O ₃	- Ozônio
°C	- Graus Celsius
Pb	- Chumbo
PI	- Partículas Inaláveis
PTS	- Partículas Totais em Suspensão

REPAR	- Refinaria Presidente Getúlio Vargas
RMC	- Região Metropolitana de Curitiba
Si	- Silício
SIMEPAR	- Tecnologias e Informações Ambientais
SMMA	- Secretaria Municipal do Meio Ambiente de Araucária
SO ₂	- Dióxido de Enxofre
SO ₄ ⁻²	- Sulfato
Ti	- Titânio
UEG	Usina Elétrica a Gás de Araucária
V	- Vanádio
W/m ²	- Watts por metro quadrado
Zn	- Zinco
μm	- Micrômetro

LISTA DE SÍMBOLOS

β_i	- Coeficientes de regressão
ε	- Erro de previsão (erro padrão)
σ^2	- Variância
η	- Preditor linear
μ	- Média
θ	- Parâmetro canônico
p	- Probabilidade
ϕ	- Parâmetro de dispersão
χ^2	- Estatística de Pearson ou Qui-Quadrado
$\hat{\mu}$	- Média estimada

1 INTRODUÇÃO

Ao contrário do que se possa imaginar, a poluição do ar não é uma exclusividade da industrialização. Desde a descoberta do fogo, aproximadamente 800 mil anos antes de Cristo, o homem já contribuía de forma atuante para a deterioração da qualidade do ar, passando a sofrer as conseqüências de seus próprios atos (Braga et al., 2005).

Com o surgimento das áreas urbanas, as pessoas tiveram acesso a muitas facilidades, como transporte, energia, educação, abastecimento, porém, como conseqüência, a qualidade do ar começou a piorar. Desde então, a humanidade focou em extrair do planeta o que lhe fosse útil, sem se preocupar com os danos causados ao meio ambiente, até ocorrerem diversos eventos de poluição atmosférica excessiva.

Um dos registros mais antigos de poluição do ar urbano aconteceu no ano de 1283, na cidade inglesa de Nottingham, quando se formaram nuvens de fumaça preta, provenientes da queima de carvão mineral utilizado no aquecimento das residências durante o ano (Danni-Oliveira, 1999).

Com o crescimento da população e conseqüente aumento no consumo de carvão mineral, e ainda com o início do processo maciço de industrialização, notadamente a produção de ferro (século XVIII) e de aço (século XIX), houve uma redução na qualidade do ar devido à intensificação e diversificação dos poluentes emitidos para a atmosfera das cidades, em particular as inglesas, precursoras do processo de concentração industrial (Danni-Oliveira, 1999).

Não foi por outra razão que o primeiro manual sobre climatologia urbana tenha surgido em Londres no ano de 1818, de autoria de Luke Howard (Danni-Oliveira, 1999).

No século XX, a presença cada vez maior de veículos a combustão, associada à industrialização, e condições meteorológicas desfavoráveis à dispersão de poluentes provocou em diversas regiões do planeta a ocorrência de vários fenômenos críticos e alarmantes de poluição atmosférica excessiva, registrando perdas de vidas ou danos à saúde.

Alguns exemplos de eventos de poluição atmosférica excessiva são apresentados no Quadro 1.1.

Quadro 1.1 - Eventos de poluição atmosférica excessiva (Derisio, 2000)

Ano	Local	Poluente	Causa	Conseqüências
1930	Vale do Rio Meusa, Bélgica	Diversos poluentes	Inversão térmica	63 mortes por DR
1948	Donnora, Pensilvânia, EUA	MP e SO₂	Fabricas de ácido sulfúrico e beneficiamento de zinco	17 mortes e 6000 pessoas com DR
1952	Londres, Inglaterra	MP e SO ₂	---	Aumento de 4000 mortes

Nota: DR – Doenças Respiratórias.

O fenômeno ocorrido na região industrial do vale do rio Meusa ocasionou mortes por problemas respiratórios principalmente em crianças e idosos (Derisio, 2000).

Na cidade de Donnora na Pensilvânia, Estados Unidos, em 1948, ao invés de duas mortes (quantidade comum para uma região com 14.000 habitantes), ocorreram 17 mortes e 6.000 habitantes apresentaram problemas respiratórios e forte irritação nos olhos (Branco e Murgel, 2001).

O mais clássico e grave dos fenômenos de poluição atmosférica excessiva, aconteceu em Londres, no período de inverno de 1952. Nesta ocasião a cidade ficou coberta por partículas de SO₂ e material particulado (MP) em suspensão (em concentrações até nove vezes maiores do que a média para ambos os poluentes). Este fenômeno ficou conhecido como “Névoa Negra”, provocando um aumento de 4.000 mortes, na semana seguinte ao evento (acima da média no mesmo período) e mais 4.000 nos dois meses seguintes ao fenômeno (Braga et al., 2005).

Após tantos acontecimentos preocupantes, os pesquisadores perceberam a necessidade de controlar a emissão de poluentes atmosféricos. A Agência de Proteção Ambiental Americana (EPA) estabeleceu então padrões de qualidade do ar para seis poluentes específicos: partículas totais em suspensão (PTS), dióxido de enxofre (SO₂), monóxido de carbono (CO), dióxido de nitrogênio (NO₂), ozônio (O₃) e chumbo (Pb) (Braga et al., 2005).

Na Europa, em 1956 e 1968, o Parlamento Inglês, influenciado pelo fenômeno ocorrido em 1952 na cidade de Londres, desenvolveu ações controladoras como os “*Clean Air Acts*”, ampliando os controles de emissão de poluentes atmosféricos industriais (Braga et al., 2005).

O crescimento da questão ambiental nos anos seguintes resultou na criação da Agência de Proteção Ambiental norte-americana (EPA) em 1990, que tem desenvolvido metodologias de monitoramento e controle de qualidade do ar (Danni-Oliveira, 1999).

Os países desenvolvidos aperfeiçoaram formas rígidas de controle ambiental, exigindo muito das indústrias, que começaram então a migrar para países onde a legislação e o controle eram menos rígidos e até mesmo inexistentes, provocando um aumento das concentrações de poluentes atmosféricos nos países em desenvolvimento.

No Brasil, foi criado o Conselho Nacional do Meio Ambiente – CONAMA, em 1981, vinculado ao Instituto Brasileiro de Meio Ambiente e dos Recursos Naturais Renováveis – IBAMA, sendo atualmente o órgão federal que estabelece as normas de monitoramento e controle da poluição atmosférica, bem como os padrões de qualidade do ar que devem ser seguidos (Derisio, 2000).

Os órgãos acima mencionados estabeleceram, na Resolução nº 3 de 28/6/90 – CONAMA, os padrões de qualidade do ar para: partículas totais em suspensão (PTS), fumaça, partículas inaláveis (PI), dióxido de enxofre (SO₂), monóxido de carbono (CO), ozônio (O₃) e dióxido de nitrogênio (NO₂), bem como os níveis de concentração considerados críticos para estes poluentes, com o objetivo de prevenir fenômenos agudos de poluição (Derisio, 2000).

1.1 Material Particulado e a Saúde Humana

O termo material particulado (MP) é usado para uma mistura de partículas sólidas e líquidas suspensas no ar. Estas partículas originam-se de uma variedade de fontes, e são formadas na atmosfera por transformações de emissões gasosas. Sua composição química e física depende da localização, época do ano, e clima.

Alguns autores como Bates (1996) apud Freitas et al. (2004), dizem que o material particulado (MP) deve ser colocado em primeiro lugar como estratégia de controle, por ser um poluente capaz de adsorver e carregar poluentes tóxicos para o aparelho respiratório, além de interagir com outras substâncias poluidoras, como NO_x , SO_2 , NH_3 e COV's (Compostos Orgânicos Voláteis) (Pires, 2005).

De acordo com o estudo de Braga et al. (2001), o MP não sofreu efeitos de outros poluentes, ou seja, os efeitos causados pela poluição atmosférica na saúde da população de São Paulo foram causados principalmente pelo material particulado. O autor realizou uma análise para cada poluente separadamente e uma análise considerando todos os poluentes juntos. Os resultados mostraram que os efeitos encontrados nas análises separadas para concentração de O_3 , SO_2 , CO e NO_2 foram superestimadas, pois na análise conjunta verificou-se que os efeitos destes poluentes na saúde da população diminuíram, enquanto para o MP_{10} permaneceu aproximadamente igual (Figura 1.1).

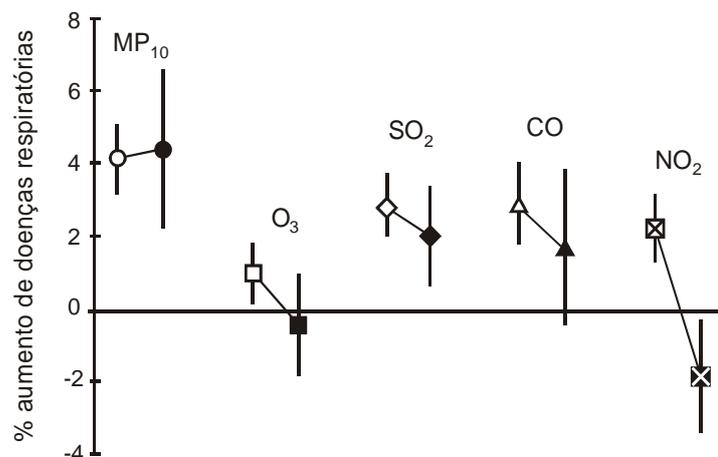


Figura 1.1 – Porcentagem de aumento de doenças respiratórias e intervalo de confiança de 95% quando os poluentes foram incluídos em modelos simples (símbolos brancos) e em modelos com todos os poluentes (símbolos pretos) (Braga et al., 2001)

Em processos de refino de petróleo e geração de energia, a emissão de MP é menor que de outros poluentes, porém entre todos os processos industriais que emitem MP, os maiores emissores são os processos de refino de petróleo e geração de energia, além de o MP ser emitido pela maioria das fontes poluidoras do ar, como em processos de combustão, indústrias, queima de resíduos sólidos e por veículos automotores (Pires, 2005).

De uma forma geral, o MP é dividido em partículas finas e espessas. Suas características como, formação, composição, fonte de emissão, tempo de vida e distância percorrida estão relacionadas no Quadro 1.2.

Dentre as várias diferenças entre as partículas finas e espessas mostradas no Quadro 1.2, cabe salientar a grande diferença entre a distância percorrida e o tempo de vida destas partículas, pois as partículas que percorrem a maior distância e têm o maior tempo de vida são também as mais prejudiciais à saúde.

Quadro 1.2 – Comparação entre partículas finas e espessas (EPA, 1998)

	Partículas Finas	Partículas Espessas
Formação	Gases	Sólidos/líquidos
Formado através de	Reações químicas, nucleação, condensação, coagulação; evaporação de névoa e gotas de nuvens nas quais gases se dissolveram ou reagiram.	Rompimento mecânico (ex. trituração, moagem, desgaste de superfície, evaporação de <i>sprays</i> , suspensão de poeira).
Composição	Sulfato (SO_4^{-2}); Nitrato (NO_3^-); amônia (NH_4^+); íon de hidrogênio (H^+); carbono elementar, compostos orgânicos (ex. HPAs, NHPAs); metais (Pb, Cd, V, Ni, Cu, Zn, Mn, Fe); partículas de água.	Poeira resuspensa (ex. poeira do solo, poeira das ruas), cinzas de carvão e óleo suspensas; óxidos metálicos de elementos geológicos (Si, Al, Ti, Fe); CaCO_3 , NaCl, sal marinho, pólen, esporos; fragmentos de plantas/animais; restos de pneus.
Solubilidade	Largamente solúvel	Largamente insolúvel
Fontes	Combustão de óleo, carvão, gasolina, diesel; floresta; produtos de transformações atmosféricas de NO_x , SO_2 ; compostos orgânicos incluindo espécies biogênicas (ex. “ <i>terpenes</i> ”); processos a alta temperatura; fundidor; fábricas de aço e etc.	Resuspensão de poeira industrial e terra nas rodovias; suspensões de solos perturbados (ex. agricultura, exploração de minas, rodovias não pavimentadas); fontes biológicas; construção e demolição; combustão de óleo e carvão; pulverizações oceânicas.
Tempo de vida	Dias a semanas	Minutos a horas
Distância alcançada	Centenas e milhares de quilômetros	unidades a dezenas de quilômetros

A capacidade do material particulado em produzir efeitos adversos à saúde humana depende de sua deposição no aparelho respiratório, que é dependente principalmente do tamanho e propriedades aerodinâmicas das partículas.

Existem diversas divisões para o MP, porém a mais utilizada é dividir em partículas com diâmetro aerodinâmico entre 2,5 e 10 μm (MP_{2,5-10}) correspondentes às partículas inaláveis (PI) e partículas com diâmetros aerodinâmicos menores ou

iguais a $2,5\mu\text{m}$ ($\text{MP}_{2,5}$), chamadas partículas finas, correspondentes à fração de partículas respiráveis.

As partículas inaláveis são capazes de serem depositadas na região superior do aparelho respiratório e as partículas respiráveis são capazes de penetrar na região alveolar do pulmão, causando maiores danos à saúde por não ser possível expeli-las por tosse ou espirro.

Outra divisão, utilizada pelos órgãos que monitoram a qualidade do ar no Brasil é em MP_{10} , correspondendo às partículas com diâmetro aerodinâmico menor ou igual a $10\mu\text{m}$, e partículas totais em suspensão (PTS) correspondendo a todas as partículas presentes no ar com diâmetro aerodinâmico até cerca de $50\mu\text{m}$ (Almeida, 1999).

A Figura 1.2 mostra de forma detalhada, as regiões do aparelho respiratório onde o material particulado é depositado de acordo com o diâmetro da partícula e a fração de deposição, verificando-se que as partículas menores se depositam em regiões mais profundas do pulmão, ou seja, as regiões bronquiais.

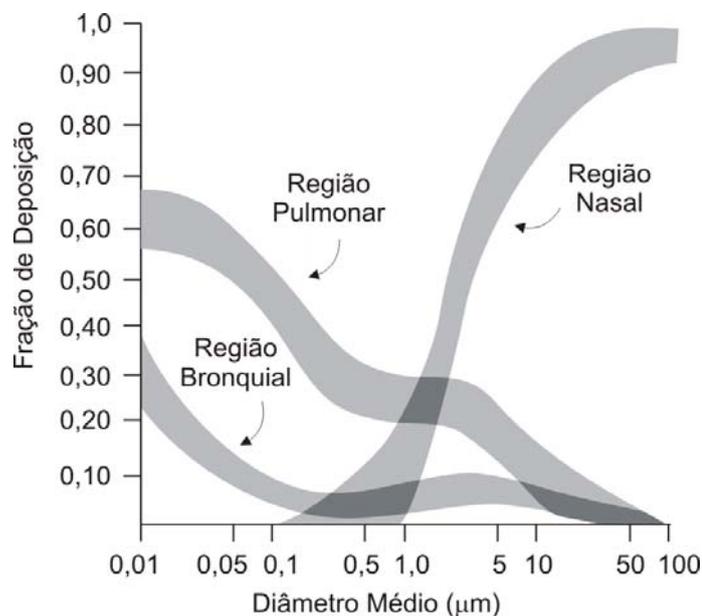


Figura 1.2 - Regiões do aparelho respiratório de acordo com a deposição de material particulado em função do diâmetro médio e da fração de deposição

1.2 Fontes Emissoras de Material Particulado

As indústrias, principalmente de processamento de petróleo, são as principais fontes fixas emissoras de poluentes atmosféricos. Sendo os COV, compostos sulfurados, nitrogenados e material particulado, os poluentes mais comumente emitidos por estas fontes poluidoras (Szwarcfiter, 2004).

Conforme Pires (2005), as atividades de refino de petróleo são responsáveis por 42% das emissões de poluentes atmosféricos industriais na Região Metropolitana do Rio de Janeiro, apesar das indústrias de refino de petróleo estarem em número reduzido na Região Metropolitana do Rio de Janeiro. Ele ressalta ainda a contribuição das atividades geradoras de energia com 31% do total de poluentes emitidos pelas indústrias da Região Metropolitana do Rio de Janeiro.

Deve-se destacar, entretanto, que a poluição decorrente das refinarias de petróleo ocorre tanto no processo de refino de petróleo, quanto no uso dos produtos derivados do petróleo. No Brasil, o refino de petróleo desempenha um importante papel na economia, sendo responsável por cerca de 30% da energia total consumida no país, e praticamente toda a energia consumida no setor de transporte. No entanto, os poluentes gerados nas unidades de processo das refinarias, são modestos frente à poluição total gerada pelo consumo dos produtos derivados do petróleo, seja pelo consumo para o transporte, para a manufatura de produtos químicos, para a geração de energia elétrica ou para usos comerciais (Mariano, 2001).

De acordo com Szwarcfiter (2004), as emissões de poluentes veiculares representam parte importante das emissões totais em grandes centros urbanos, sendo que, muitas vezes, são os maiores responsáveis pela poluição atmosférica.

Em um estudo realizado por Castanho (1999), foi identificada a contribuição de algumas fontes emissoras de material particulado para a poluição atmosférica da cidade de São Paulo. O autor verificou que as emissões veiculares são responsáveis por aproximadamente 28% do $MP_{2,5}$ suspenso no ar no período de inverno e 24% no período de verão. Entretanto, o transporte de veículos também está associado ao processo de ressuspensão de solo, que é responsável por aproximadamente 20%

do $MP_{2,5}$ suspenso no período de inverno e 30% no período de verão. Apenas em torno de 5% no inverno e 6% no verão estão relacionados a fontes industriais.

Os veículos movidos a gasolina, diesel, álcool e GNV (gás natural veicular) produzem gases, vapores e material particulado. A diferença está nas quantidades que cada um emite para a atmosfera. Os combustíveis de origem fóssil (gasolina e diesel) liberam, por exemplo, na queima incompleta do combustível, CO, hidrocarbonetos, aldeídos, NO_x , SO_x e material particulado (Derisio, 2000; Branco e Murgel, 2001).

A substituição do óleo diesel pelo GNV nos veículos pesados (caminhões e ônibus), que está ocorrendo constantemente, é importante para a redução de emissões de poluentes locais, em especial o material particulado, um dos grandes problemas do uso do óleo diesel (Szwarcfiter, 2004).

1.3 Objetivos

Devido à grande preocupação com os efeitos da poluição atmosférica na saúde humana e à carência de informações no que se refere à relação do material particulado e doenças respiratórias para a Região Metropolitana de Curitiba (RMC), o objetivo geral do trabalho será determinar um modelo estatístico capaz de avaliar e prever o impacto da poluição atmosférica na saúde humana. A partir deste objetivo geral serão desdobrados os seguintes objetivos específicos:

- Obter dados de concentração de MP_{10} e meteorológicos junto aos órgãos responsáveis de Araucária, bem como dados de saúde, mais precisamente do número de atendimentos por doenças respiratórias;
- Comparar os modelos estatísticos mais utilizados para escolher o mais apropriado para o problema em questão;
- Aplicar o modelo escolhido para avaliar o impacto do MP_{10} no número de pessoas com doenças respiratórias em Araucária a fim de testar o modelo;
- Avaliar o ajuste produzido pelo modelo e analisar os resultados obtidos;

- Obter a equação que rege a influência do MP_{10} na saúde da população de Araucária.

1.4 Estrutura da Dissertação

Assim, quanto à estrutura da Dissertação, a parte introdutória enuncia os conceitos pertinentes ao problema em si, a importância do problema junto com alguns dados da literatura, os objetivos geral e específicos da pesquisa e a forma de desenvolvimento do texto.

O Capítulo 2 apresenta uma revisão das pesquisas já realizadas nas áreas envolvidas no trabalho, ou seja, estudos sobre o impacto da poluição atmosférica na saúde populacional, bem como estudos de caráter estatístico.

O Capítulo 3 descreve a metodologia utilizada no desenvolvimento da pesquisa, ressaltando as principais atitudes a serem tomadas na escolha do modelo que melhor se ajuste aos dados observados.

O Capítulo 4 se refere ao estudo de caso, mostrando alguns resultados preliminares sobre a relação entre as variáveis utilizadas e resultados mais precisos obtidos com a aplicação da família de Poisson para o modelo linear generalizado (MLG) e os respectivos ajustes ao conjunto de dados para o período entre 01 de Janeiro de 2004 e 31 de Dezembro de 2005.

As conclusões e sugestões encontram-se no Capítulo 5.

2 REVISÃO BIBLIOGRÁFICA

Muitos estudos epidemiológicos¹ têm demonstrado a importância da poluição do ar como um possível fator de influência na saúde humana, principalmente em doenças por causas respiratórias e cardiovasculares, até mesmo em concentrações abaixo dos limites regulamentares. Para tanto, é preciso avaliar os impactos dos poluentes na saúde.

2.1 Impacto da Poluição Atmosférica na Saúde

A resposta a uma quantidade de poluente é diferente para cada pessoa, dependendo de sua faixa etária, da genética e das condições sociais (Bakonyi, 2003).

Idosos e crianças, por exemplo, são mais suscetíveis à poluição atmosférica. As crianças são mais suscetíveis por consumirem o dobro da quantidade de ar consumida pelos adultos. Dichiade (1992) apud Bakonyi (2003) relata que “em condições de repouso e equilíbrio térmico, um lactente consome 7 ml/kg de oxigênio por minuto, contra 3,5 ml/kg por minuto para um adulto nas mesmas condições, ou seja, o volume de ar que passa pelos pulmões de um lactente é duas vezes maior que o de um adulto em repouso”. Já os idosos são mais afetados por possuírem um sistema imunológico e respiratório frágil, além de já terem sido expostos a uma grande quantidade de poluição atmosférica ao longo da vida.

Quanto à genética, pessoas que possuem doenças crônicas ou alergias, como bronquite e asma são mais sensíveis à poluição do ar.

Já em relação às influências da condição social na saúde populacional, as regiões com infra-estrutura precária, como por exemplo, falta de saneamento básico e asfalto, acarretam maiores riscos à saúde dos moradores devido à maior concentração de poluentes presentes nestas regiões.

¹ Estudos epidemiológicos – Estudos de processos saúde-doença em coletividades humanas que analisam a distribuição e os fatores determinantes das enfermidades, danos à saúde e eventos associados à saúde coletiva, propondo medidas específicas de prevenção, controle ou erradicação da doença (Rouquayrol, 1999).

Apesar de existirem diversos fatores que influenciam na saúde populacional em relação à concentração dos poluentes, este trabalho limitar-se-á a avaliar os efeitos respiratórios decorrentes da concentração de MP_{10} no ar subdividindo por faixa etária, pois não é possível considerar fatores como genética e condição social em uma análise para toda a população de uma região e conseqüentemente não se pode utilizar o modelo para realizar previsões.

2.2 Análise dos Modelos Estatísticos para Estudos Epidemiológicos

Em estudos epidemiológicos, os modelos freqüentemente utilizados são os estatísticos e analíticos que segundo Conceição et al. (2001) “constituem ferramentas extremamente úteis para resumir e interpretar dados. Em particular, estes modelos podem facilitar a avaliação da forma e da intensidade de associações de interesse em estudos epidemiológicos”.

Os principais modelos de regressão utilizados são o de Poisson, proporcional de Cox e logística condicional. Exemplos de estudos que utilizaram estes modelos de regressão estão relacionados no Quadro 2.1. Todos os estudos mencionados no Quadro 2.1 mostraram que a poluição atmosférica está relacionada com um agravo na saúde populacional. De acordo com estudos que avaliaram a diferença entre as faixas etárias, as crianças foram mais afetadas pela poluição do ar.

Com base nos resultados encontrados no estudo realizado em Curitiba, Bakonyi et al. (2004) relataram que “os níveis de poluição do ar em Curitiba, apesar de não serem tão altos ou mesmo quando não ultrapassaram o padrão de qualidade do ar, interferem no perfil da morbidade² respiratória da população infantil da cidade”.

² Morbidade – Conjunto de indivíduos que adquiriram doenças num intervalo de tempo.

Quadro 2.1 - Exemplos de estudos sobre o impacto da poluição do ar na saúde humana

	Poluentes considerados	Causa estudada	Cidade	Período	Citação
Regressão de Poisson	MP ₁₀ , CO, O ₃	Morbidade respiratória em menores de 15 anos e mortalidade em idosos	São Paulo, Brasil	1993-1997	Freitas et al. (2004)
	MP ₁₀ , O ₃ , SO ₂ , CO, NO ₂	DR em crianças	São Paulo, Brasil	1993-1997	Braga et al. (2001)
	MP ₁₀ , O ₃ , SO ₂ , CO, NO ₂	Morbidade por pneumonia e gripe em idosos	São Paulo, Brasil	1996-1998	Martins et al. (2002)
	MP, fumaça, NO ₂ , O ₃	Morbidade por DR em crianças	Curitiba, Brasil	1999-2000	Bakonyi et al. (2004)
	MP ₁₀ , NO ₂ , SO ₂ , CO, O ₃	Mortalidade por faixa etária	Seul, Coréia	1995-1999	Ha et al. (2005)
Regressão Proporcional de Cox	MP ₁₀ , O ₃ , SO ₂ , NO ₂ , SO ₄	Mortalidade por causas específicas em adultos separados por sexo	Califórnia, EUA	1977-1992	Abbey et al. (1999)
	MP, fumaça, O ₃ , SO ₂ , NO ₂ , CO	Mortalidade	EUA	---	Dockery et al. (1993)
	MP _{2.5} , MP ₁₀ , sulfatos, SO ₂ , NO ₂ , CO, O ₃	Mortalidade por causas respiratórias em voluntários da sociedade americana de câncer	América	1982-1998	Pope et al. (2002)
Regressão Condicional Logística	MP _{2.5} , carbono elementar e poluentes gasosos	Infarto do Miocárdio (IM)	Boston, EUA	1995-1996	Peters et al. (2001)
	PTS	Mortalidade	Filadélfia, EUA	1973-1980	Neas et al. (1999)
	PTS, SO ₂ , O ₃	Mortalidade	Seul, Coréia	1991-1995	Lee e Schwartz (1999)
Regressão Linear	MP, CO, NO ₂ , SO ₂	Pressão sanguínea e batimentos cardíacos em pacientes com doenças coronárias do coração	Finlândia, Alemanha, Países Baixos	1998-1999	Ibald-Mulli et al. (2004)

Existem alguns estudos que compararam os resultados obtidos através de modelos estatísticos diferentes, bem como estudos que propuseram novos modelos capazes de avaliar o impacto da poluição atmosférica na saúde. Alguns destes estudos estão relacionados no Quadro 2.2.

Quadro 2.2 – Estudos com enfoque na comparação entre modelos estatísticos

Modelos comparados	Caso estudado	Conclusão	Autor
Análise caso-controle ³ para regressão condicional logística ⁴ e Regressão de Poisson ⁵	Associação entre concentração de PTS e mortalidade entre 1973 e 1980 na Filadélfia, EUA	Os resultados obtidos com a análise caso-controle confirmaram os resultados obtidos com uma regressão de Poisson	Neas et al. (1999)
Regressão linear múltipla, regressão de Poisson e regressão logística	Associação entre concentração de O ₃ e MP ₁₀ na mortalidade diária para vários estudos.	A metodologia proposta para comparação entre estudos funciona perfeitamente	Baxter et al. (1997)
Método desenvolvido pelo autor	Associação entre concentração de CO, NO ₂ e MP ₁₀ e riscos à saúde para algumas cidades da Austrália	O método desenvolvido é prático, transparente e satisfaz os aspectos probabilísticos da avaliação de risco na saúde	Beer e Ricci (1999)
Modelo de regressão de Poisson	Dados de um laboratório de análises clínicas de São Luiz Gonzaga/RS, entre Agosto e Setembro de 2002	O uso da estatística na área de saúde está evoluindo	Schmidt (2003)
MLG e MAG	Associação entre concentração de MP ₁₀ , SO ₂ , CO e O ₃ e mortalidade em idosos de São Paulo, entre 1994 e 1997	As duas classes de modelos apresentadas produziram resultados coerentes	Conceição et al. (2001)

O estudo realizado por Neas et al. (1999) tinha como objetivo demonstrar a metodologia de análise caso-controle, com o uso da análise de regressão condicional logística, bem como comparar com o modelo de regressão de Poisson. Os autores concluíram que a primeira análise confirmou os resultados obtidos em um estudo anterior em que o modelo de regressão de Poisson foi utilizado (um aumento na mortalidade diária esteve associado ao material particulado).

Conforme Neas et al. (1999), a desvantagem da aproximação caso-controle é que o modelo para mortalidade total tem um erro padrão maior que da regressão de

³ Análise caso-controle – “Este tipo de estudo inicia-se pelos doentes identificados (casos), estabelece controles para eles (sujeitos comparáveis aos casos, porém reconhecidamente não-doentes) e, retrospectivamente, procura conhecer os níveis de exposição ao suposto fator de risco” (Rouquayrol, 1999).

⁴ Regressão Condicional Logística – Modelo que utiliza dados binomiais como variável resposta (mais detalhes na página 15)

⁵ Regressão de Poisson – Utilizada quando a variável resposta é não-negativa em forma de contagem (mais detalhes em 2.3.3 Modelos de regressão de Poisson)

Poisson, entretanto, a aproximação caso-controle em princípio permite utilizar dados adicionais de cada morte individual, tais como, idade exata do paciente, contribuições para a causa de morte, o que não é possível através da regressão de Poisson.

Baxter et al. (1997) mostraram como estudos que usam diferentes modelos estatísticos podem ser comparados. Para realizar a comparação, os autores propuseram primeiro selecionar um valor comum para x e y , por exemplo, x_0 e y_0 , com os quais seria possível avaliar os resultados de cada estudo. Escolhas prováveis para y_0 incluem a taxa de mortalidade média observada para todos os estudos considerados, a taxa de mortalidade média para as regiões geográficas nas quais os dados foram coletados ou a taxa de mortalidade média para todo o país. Similarmente, possíveis escolhas para x_0 incluem o valor médio de MP_{10} para todos os estudos considerados, o valor médio de MP_{10} para as áreas geográficas nas quais os dados foram coletados ou para todo o país. Assim, avaliam-se os resumos estatísticos de interesse para cada estudo em (x_0, y_0) . Os autores concluíram que a metodologia proposta para comparação entre estudos funciona perfeitamente.

O estudo realizado por Conceição et al. (2001) descreveu e comparou duas classes de modelos que podem ser utilizadas para avaliar a associação entre poluição atmosférica e morbidade ou mortalidade por causas específicas: os modelos lineares generalizados (MLG) e os modelos aditivos generalizados (MAG). Nos MLG, foram utilizados, como variáveis explicativas⁶, 12 indicadores para meses do ano, 4 para anos de estudo, 7 para dias da semana, temperatura, umidade e o número de óbitos por outras causas exceto respiratórias.

Já nos MAG, “os indicadores para meses e anos foram substituídos por uma curva de alisamento⁷ do número de óbitos em função do número de dias transcorridos. Os polinômios envolvendo as medidas de temperatura e umidade foram substituídos por curvas de alisamento do número de óbitos em função dessas variáveis” (Conceição et al., 2001)

⁶ Variáveis explicativas são comumente conhecidas como variáveis independentes.

⁷ Curva de alisamento: São curvas ajustadas através de uma função que por definição deve ser mais “suave” do que os valores de y , ou seja, devem ter menor variabilidade do que os valores de y . Como exemplos de curvas de alisamento podem ser citados: média móvel, *splines*, *locally weighted running line smoother* (loess), entre outros (Conceição et al., 2001).

Conceição et al. (2001) concluíram que as duas classes de modelos apresentadas produziram resultados coerentes, sendo que a abordagem baseada em curvas de alisamento utilizada nos MAG permite que o padrão sazonal seja definido pelos próprios dados, sem a imposição de uma estrutura rígida e, talvez, menos fidedigna. Existem, entretanto, desvantagens em utilizar curvas de alisamento, pois os coeficientes estimados correspondentes nos modelos de regressão não são interpretáveis.

Nos MLG, foram consideradas 18 variáveis explicativas, enquanto apenas 4 foram requeridas pelos MAG. Assim, aparentemente, os MAG são modelos mais parcimoniosos, ou seja, necessitam de um número menor de variáveis explicativas, o que justificaria o fato de detectarem um número maior de associações (Conceição et al., 2001).

Dentre os modelos estatísticos utilizados em estudos epidemiológicos, verificou-se pelos estudos citados acima, que os mais utilizados são os modelos de regressão logística, proporcional de Cox e de Poisson. A principal diferença entre estes modelos é o tipo de variável resposta (comumente chamada de variável dependente) considerada.

O modelo de regressão logística é utilizado para problemas em que a variável resposta é do tipo binomial, onde se atribui apenas valores 0 ou 1, representando fracasso ou sucesso na análise (Everitt, 2003).

Este tipo de observação é conveniente para experiências médicas, onde no final do período de estudo, ou o paciente se recuperou ($Y=1$) ou não se recuperou ($Y=0$).

Com este modelo é possível considerar fatores como idade exata, sexo, uso ou não de medicamentos, entre outros. Desta forma, espera-se que os resultados expliquem melhor o problema, porém a extensão do resultado para outros grupos implicará em erros provavelmente maiores que com a aplicação da regressão de Poisson, pois ele não fornece uma visão do impacto geral da poluição em uma determinada região.

Outro tipo de variável que não fornece uma visão geral é a chamada de variável de sobrevivência, correspondendo ao tempo de vida de um componente, ou tempo

de sobrevivência de um paciente, muito utilizada para experiências médicas ou para experimentos de segurança industrial onde a ênfase é dada para o tempo de falha de um equipamento ao invés do tempo de sobrevivência. No caso de experiências médicas, é necessário ter um acompanhamento do paciente durante o período em estudo, o que exige tempo e dedicação maiores. O modelo de regressão proporcional de Cox tem como variável resposta dados de sobrevivência (Everitt, 2003).

Já o modelo de regressão de Poisson, muito comum em estudos epidemiológicos, tem por característica a análise de dados contados na forma de proporções ou razões de contagem, ou seja, leva em consideração o total de pessoas que receberam um tipo de diagnóstico, dados facilmente adquiridos através da secretaria de saúde da região estudada (McCullagh e Nelder, 1989). Os resultados deste modelo têm um aspecto generalizado. Assim, prevêem o impacto geral de um tipo de variável na saúde da população de uma determinada região.

Os modelos estatísticos utilizados na análise da relação entre a poluição atmosférica e o impacto na saúde são análises de regressão, pois “é uma ferramenta útil para avaliar a relação entre uma ou mais variáveis explicativas (variáveis independentes, preditoras ou covariáveis) (x_1, x_2, \dots, x_n) e uma única variável resposta (variável dependente, prevista) (y) ” (Martins, 2000).

2.3 Análise de Regressão

A análise de regressão que envolve apenas uma variável explicativa é chamada de regressão simples, enquanto a análise envolvendo duas ou mais variáveis explicativas é denominada regressão múltipla (Hair Jr. et al., 2005).

A regressão linear múltipla é dada por:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \varepsilon \quad \text{Eq. 2.1}$$

onde y é a variável resposta e x_i com $(i = 1, 2, \dots, n)$ são as variáveis explicativas. β_0 representa o valor de y quando as variáveis explicativas são nulas, os termos β_i são chamados de coeficientes de regressão e o resíduo (ε) é o erro de previsão, ou

seja, a diferença entre os valores reais e os previstos da variável resposta, que é assumido normalmente distribuído com média zero e variância σ^2 (Hair Jr. et al., 2005).

A relação linear existente na regressão linear é entre y e os parâmetros $\beta_1, \beta_2, \dots, \beta_n$, e não entre y e os x 's. Por exemplo, uma expressão da forma $y = \beta_0 + \beta_1 x_1^2 + \beta_2 \log x_2 + \varepsilon$ é um modelo de regressão linear múltipla, já $y = \beta_0 + \beta_1 x_1^{\beta_2} + \beta_3 x_2^3 + \varepsilon$ é não-linear (Werkema e Aguiar, 1996).

O objetivo da análise de regressão linear múltipla, assim como de todos os tipos de regressão é encontrar uma equação (chamada de equação de regressão, variável estatística de regressão ou modelo de regressão) que melhor prevê a variável resposta a partir de uma combinação das variáveis explicativas, ou seja, deseja-se encontrar os valores dos β 's que melhor se ajustem aos dados do problema (Hair Jr. et al., 2005).

Encontrados os β 's, é necessário validar o modelo de regressão, ou seja, verificar se os sinais e magnitude dos coeficientes fazem sentido no contexto do fenômeno estudado (Werkema e Aguiar, 1996).

Durante muitos anos os modelos lineares foram utilizados para descrever a maioria dos fenômenos aleatórios. Mesmo quando o fenômeno estudado não apresentava uma resposta em que a suposição de normalidade fosse razoável, tentava-se algum tipo de transformação no sentido de alcançar a normalidade procurada. Estes modelos são baseados na idéia relativamente simples de empregar uma equação para expressar o relacionamento entre as variáveis de interesse (Paula, 2004).

Nem sempre é possível aplicar um modelo de regressão linear em análises do impacto da poluição na saúde devido ao caráter não linear da variável resposta. Nestes casos, geralmente utiliza-se as classes de modelos que oferecem uma poderosa alternativa para a transformação de dados, chamadas de MLG e MAG (Santos, 2000 apud Schmidt, 2003).

Neste trabalho busca-se avaliar o impacto da poluição na população total de uma região, além de almejar-se um modelo capaz de realizar previsões. Para tanto é

necessário utilizar a família de Poisson para o MLG, pois como mostram Conceição et al. (2001), os coeficientes estimados utilizando o MAG não são possíveis de ser interpretados.

2.3.1 Modelos de regressão

A partir do desenvolvimento computacional, na década de 70, algumas análises de regressão que exigiam a utilização de processos iterativos na estimativa dos parâmetros começaram a ser utilizados e aperfeiçoados.

Nesta época, Nelder e Wedderburn (1972) propuseram uma teoria unificadora da modelagem estatística a qual denominaram modelos lineares generalizados (MLG), como sendo uma extensão dos modelos lineares clássicos. Os autores mostraram que uma série de técnicas, comumente estudadas separadamente havia mais de um século, podia ser reunida.

Os estudos que levaram a essa visão geral da modelagem estatística denominada MLG baseado em McCullagh e Nelder (1989) estão relacionados no Quadro 2.3.

A regressão linear múltipla, envolvendo distribuição normal, iniciou-se no século XIX com os matemáticos Legendre e Gauss (McCullagh e Nelder, 1989).

Fisher (1935) apud McCullagh e Nelder (1989) desenvolveu, em 1934, a família exponencial de distribuições, como sendo uma classe de distribuições com propriedades “boas” para a estimativa dos parâmetros.

O trabalho de Dike e Patterson (1952) apud McCullagh e Nelder (1989) foi realizado através de uma análise de dados cruzados relacionando a proporção de pessoas que possuem um bom conhecimento sobre câncer.

O trabalho de Birch (1963) apud McCullagh e Nelder (1989) é usado freqüentemente nas ciências sociais e o de Nelder (1966) apud McCullagh e Nelder (1989), em trabalhos biológicos para expressar a forma das curvas de resposta, o crescimento das curvas entre outros. Por sua vez, os trabalhos de Feigl e Selen (1965); Zippin e Armitage (1966) e Glasser (1967) apud McCullagh e Nelder (1989),

são usados geralmente para tratamentos clínicos e cirúrgicos, por serem modelos para análise de dados de sobrevivência.

Quadro 2.3 - Histórico dos estudos que levaram à modelagem estatística denominada MLG (McCullagh e Nelder, 1989)

Ano	Autor	Estudo realizado	Tipo de distribuição envolvida
1920 a 1935	Fisher	Análise de variância para experimentos planejados	Distribuição binomial com função de ligação logit
1922	Fisher	Procedimento geral para inferência estatística	Qualquer distribuição
1922	Fisher	Modelo complemento log-log para ensaios de diluição	Distribuição binomial
1935	Bliss	Modelo probit para proporções	Distribuição binomial
1952	Dike e Patterson	Modelos logísticos para proporções	Distribuição binomial com função de ligação logit
1960	Rasch	Modelo logístico para análise de itens	Distribuição Bernoulli
1963	Birch	Modelos log-lineares para contagens	Distribuição de Poisson e multinomial
1966	Nelder	Polinômios inversos	Distribuição gama
1965	Feigl e Selen	Modelos para dados de sobrevivência	Distribuição exponencial com função de ligação recíproca ou log
1966	Zippin e Armitage		
1967	Glasser		

2.3.2 Modelos lineares generalizados

Os modelos lineares generalizados (MLG) são a união de modelos lineares e não-lineares com uma distribuição da família exponencial, que é formada pela distribuição normal, Poisson, binomial, gama, normal inversa e incluem modelos lineares tradicionais (erros com distribuição normal), bem como modelos logísticos (Schmidt, 2003).

Desde 1972, inúmeros trabalhos relacionados com modelos lineares generalizados foram publicados. Por exemplo, as ferramentas computacionais GLIM (*Generalized Linear Interactive Models*), S-Plus, R, SAS, STATA e SUDAAN, bem como extensões destes modelos (Paula, 2004).

Os MLG são definidos por uma distribuição de probabilidade, membro da família exponencial de distribuições e são formados pelas seguintes componentes (McCullagh e Nelder, 1989):

- Componente Aleatória: n valores explicativas y_1, \dots, y_n , de uma variável resposta que segue uma distribuição da família exponencial, com valor esperado $E(y_i) = \mu$;
- Componente Sistemática: compõe uma estrutura linear para o modelo de regressão $\eta = \beta \mathbf{X}^T$, chamado de preditor linear, onde $\mathbf{X}^T = (x_{i1}, x_{i2}, \dots, x_{ip})^T$, $i = 1, \dots, n$, são as chamadas variáveis explicativas;
- Função de ligação: Uma função monótona e diferenciável g , chamada de função de ligação, capaz de ligar as componentes aleatória e sistemática, ou seja, liga a média da variável resposta (μ) e a estrutura linear, definida nos MLG por $g(\mu) = \eta$, onde:

$$\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \tag{Eq. 2.2}$$

ou em forma matricial:

$$\eta = \beta \mathbf{X}^T \tag{Eq. 2.3}$$

Com o coeficiente de regressão $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ representando o vetor de parâmetros a ser estimado (McCullagh e Nelder, 1989).

Na Tabela 2.1, cada distribuição tem uma função de ligação especial, chamada de função de ligação canônica que ocorre quando $\eta_i = \theta_i$, onde θ é o chamado parâmetro de localização ou parâmetro canônico (McCullagh e Nelder, 1989).

Tabela 2.1 – Funções de ligação canônica de algumas distribuições da família exponencial (McCullagh e Nelder, 1989)

Distribuição	Normal	Poisson	Binomial	Gamma	Gaussiana Inversa
Função de ligação canônica (η)	μ	$\log(\mu)$	$\log\{\mu/(1-\mu)\}$	μ^{-1}	μ^{-2}

De acordo com Myers e Montgomery (2002), a utilização da função de ligação canônica implica em algumas interessantes propriedades, porém não quer dizer que deva ser utilizada sempre. Essa escolha é conveniente porque além de simplificar as estimativas dos parâmetros do modelo, também facilita o cálculo do intervalo de confiança⁸ para a média da variável resposta. Contudo, a conveniência não implica necessariamente em qualidade de ajuste do modelo.

Conforme o exposto acima, se η é a função logarítmica⁹ e y_i possui distribuição de Poisson, o modelo resultante é o modelo de regressão de Poisson com função de ligação canônica, utilizado para avaliar dados não-negativos em forma de contagens, frequentemente encontrados em estudos epidemiológicos.

2.3.3 Modelos de regressão de Poisson

O modelo de regressão de Poisson é um tipo específico dos MLG e MAG, que teve origem por volta de 1970, quando Wedderburn (1974) desenvolveu a teoria da quasi-verossimilhança, analisada com mais detalhes por McCullagh (1983).

A variável resposta de uma regressão de Poisson deve seguir uma distribuição de Poisson e os dados devem possuir igual dispersão, ou seja, a média da variável resposta deve ser igual à variância. Entretanto, conforme Ribeiro (2006), “quando se trabalha com dados reais, esta propriedade é frequentemente violada”. Assim, pode-se ter uma superdispersão, quando a variância é maior que a média ou uma subdispersão, quando a variância é menor que a média (Schmidt, 2003). Nestes casos, ainda é possível aplicar o modelo de Poisson realizando-se alguns ajustes.

⁸ O intervalo de confiança (IC) é um conjunto de valores calculados com base nos dados. Pressupõe-se que cubra o parâmetro de interesse com um ‘certo’ grau (nível) de confiança. O grau de confiança mais comumente utilizado é o de 95% (Bergamaschi e Souza, 2005).

⁹ Os termos logarítmica e o símbolo “log” que aparecerão no decorrer deste trabalho se referem ao logarítmico neperiano (ln). Esta nomenclatura será utilizado, pois é a forma que aparece nos livros sobre MLG e nos trabalhos sobre o impacto da poluição atmosférica na saúde populacional.

Distribuição de Poisson

A distribuição de Poisson pode ser usada para determinar a probabilidade da ocorrência de sucessos quando os eventos ocorrem continuamente no tempo ou espaço. Esta distribuição é largamente empregada quando se deseja contar o número de eventos de certo tipo que ocorrem num intervalo de tempo, superfície ou volume. Por exemplo: o número de chamadas telefônicas recebidas durante cinco minutos; número de falhas de uma máquina em um dia de operação, bem como, o número de pessoas com problemas respiratórios durante um ano (Schmidt, 2003).

Na distribuição de Poisson o valor da probabilidade (p) é muito pequeno, enquanto o número de observações (n) é muito grande, assim o produto np tenderá à média μ (Schmidt, 2003). Neste caso a função densidade de probabilidade é dada por:

$$P(Y = y_i) = \frac{e^{-\mu} \mu^{y_i}}{y_i!}; \quad i = 0, 1, 2, \dots \quad \text{Eq. 2.4}$$

Uma propriedade peculiar de uma variável aleatória de Poisson é que sua média é igual a sua variância ($\mu = \sigma^2$) (Meyer, 1983).

A representação das probabilidades da distribuição de Poisson para $\mu = 1, 2, 4, 6$ encontra-se na Figura 2.1 (Schmidt, 2003).

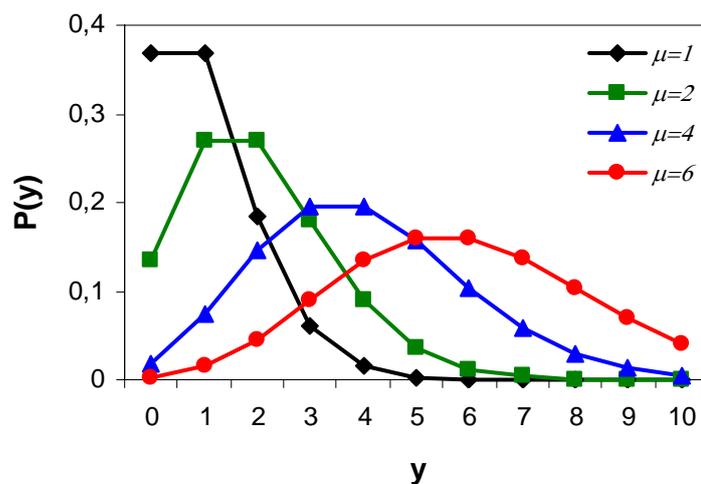


Figura 2.1 – Gráfico da distribuição de Poisson para quatro valores de μ (Schmidt, 2003)

Na Figura 2.1, observa-se um achatamento da curva e o seu deslocamento para a direita quando μ aumenta. Para valores grandes de μ , a distribuição é aproximadamente normal (Schmidt, 2003).

Os procedimentos necessários para verificar a confiabilidade dos dados e estimar os parâmetros de regressão β nos MLG serão apresentados no capítulo seguinte.

3 METODOLOGIA

Toda pesquisa requer uma busca pelo modelo mais simples, razoável, que descreva de melhor maneira possível os dados observados. Para tanto, o primeiro passo consiste na coleta e análise dos dados, em seguida escolhe-se o modelo adequado, estimam-se os parâmetros do modelo e finalmente avalia-se o ajuste do modelo escolhido.

3.1 Coleta e Análise dos Dados

Antes de escolher um modelo estatístico, é necessário especificar o escopo do estudo para realizar a coleta dos dados necessários. Em estudos epidemiológicos é importante coletar dados separados por faixa etária devido à grande diferença existente na reação de cada faixa etária a um determinado fator.

Compreender os processos que conduziram a dados faltantes na amostra é essencial na seleção do curso de ação apropriado para a melhoria do banco de dados, pois dados faltantes são uma realidade em qualquer tipo de análise (Hair Jr. et al., 2005).

Para alguns problemas em particular, existem na literatura aproximações padrões. Utilizar estas aproximações acarretará um aumento na confiabilidade dos dados. Por exemplo, os órgãos responsáveis pelo monitoramento da qualidade do ar seguem um padrão para o cálculo de médias diárias, mensais e anuais de variáveis meteorológicas e de poluentes.

O exame dos dados pode parecer uma tarefa comum e sem importância, porém é fundamental em qualquer análise de regressão, pois o banco de dados é a parte fundamental da análise, assim a confiabilidade do banco de dados é essencial (Hair Jr. et al., 2005).

Com os devidos ajustes no banco de dados, parte-se para a escolha da modelagem estatística que descreva melhor os dados.

3.2 Escolha do Modelo Estatístico

De acordo com a revisão bibliográfica realizada, o modelo adequado para o problema em questão é o linear generalizado com família de Poisson, devido às características da variável resposta. Para confirmar que o modelo escolhido é adequado, é preciso verificar primeiro se existe possibilidade de se aplicar a regressão estatística mais simples (regressão linear). Uma característica importante da regressão linear é a normalidade dos dados, ou seja, os dados devem possuir uma distribuição aproximadamente normal.

O teste de diagnóstico mais simples para verificar a normalidade é o histograma¹⁰, que compara os valores dos dados observados com uma distribuição aproximadamente normal, como mostra a Figura 3.1. Este teste é atraente devido a sua simplicidade, porém é um método problemático para amostras com menos de 100 observações. Para estes casos, existem outros métodos mais confiáveis (Hair Jr. et al., 2005).

Em estudos sobre o impacto da poluição atmosférica na saúde frequentemente tem-se mais de 100 observações, pois são necessários dados diários de pelo menos dois anos para a realização de uma boa análise. Portanto, neste trabalho foi utilizado apenas o histograma para verificar a normalidade dos dados.

¹⁰ Histograma – É uma representação gráfica de uma única variável que representa a frequência de ocorrências (valores dos dados) dentro de categorias de dados.

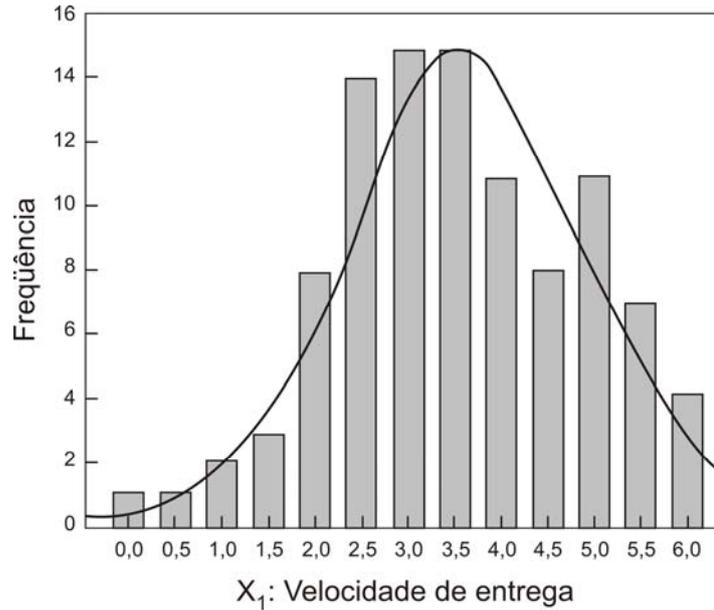


Figura 3.1 - Representação gráfica de um histograma com distribuição aproximadamente normal (Hair Jr. et al., 2005)

Uma outra característica importante dos dados é a linearidade, ou seja, a magnitude da relação linear entre duas variáveis. A linearidade pode ser observada através de um diagrama de dispersão¹¹ (Hair Jr. et al., 2005).

Em uma matriz de dispersão, como no exemplo da Figura 3.2 são apresentados o diagrama de dispersão abaixo da diagonal principal, o histograma na diagonal principal e os coeficientes de correlação¹² entre as variáveis acima da diagonal principal.

¹¹ Diagrama de dispersão: “Gráfico de pontos baseado em duas variáveis, onde uma variável define o eixo horizontal e a outra define o eixo vertical. As variáveis podem ser observações, valores esperados ou mesmo resíduos. Uma forte organização dos pontos ao longo de uma linha reta caracteriza uma relação linear ou correlação” (Hair Jr. et al. 2005).

¹² Coeficiente de correlação (r) – Coeficiente que indica a força da associação entre quaisquer duas variáveis métricas. O sinal (+ ou -) indica a direção da relação. O valor pode variar de -1 a +1, em que +1 indica uma perfeita relação positiva, 0 indica relação nenhuma e -1, uma perfeita relação negativa ou reversa (quando uma variável se torna maior, a outra fica menor) (Hair Jr. et al., 2005).

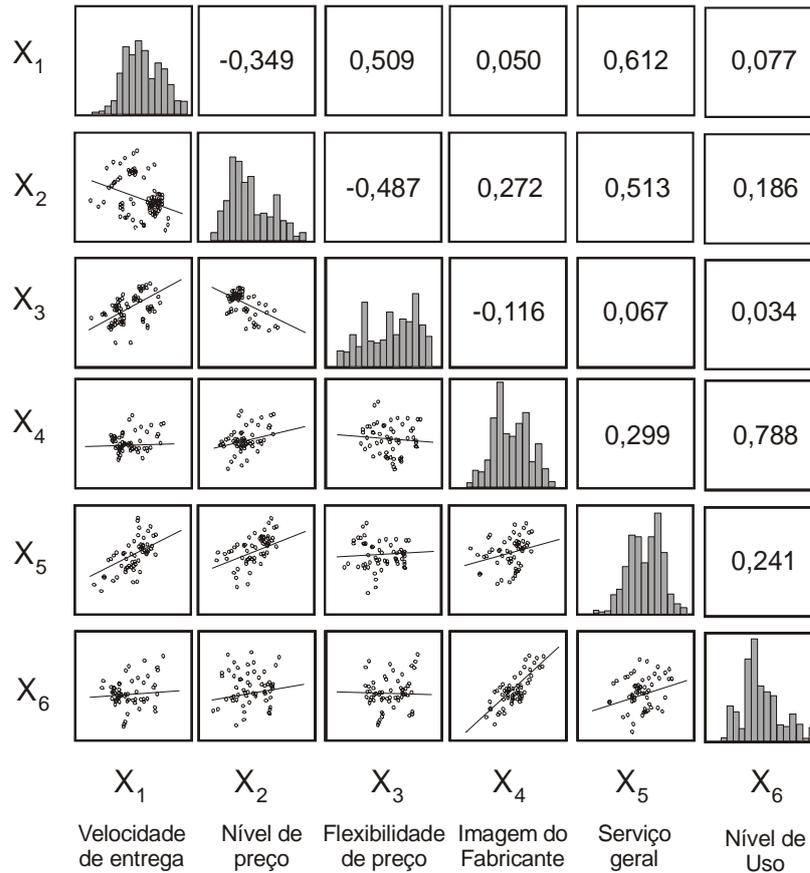


Figura 3.2 – Matriz de dispersão (Adaptado de Hair Jr. et al., 2005)

Dados que não seguem uma distribuição normal e são não negativos e em forma de contagem indicam a possibilidade de aplicação da família de Poisson dos MLG. Para se aplicar o modelo de Poisson, os dados devem possuir distribuição de Poisson, ou seja, o histograma deve apresentar uma tendência de acordo com as linhas da Figura 2.1, exceto para $\mu = 6$ que indica uma distribuição normal (Schmidt, 2003).

Se os dados seguem uma distribuição de Poisson, inicia-se a aplicação da análise de regressão de Poisson.

3.2.1 Estimativa dos parâmetros

Com todas as suposições satisfeitas, inicia-se a estimativa do modelo que pode ser feita com o auxílio de programas estatísticos capazes de encontrar os coeficientes β_i da Equação 2.2.

A estimativa dos parâmetros pode ser uma tarefa relativamente simples se houver somente um parâmetro a ser estimado, contudo não é o que ocorre normalmente. Nestes casos, processos iterativos devem ser empregados para resolver o sistema de equações.

Nos MLG, o método utilizado para estimar os valores dos parâmetros de regressão β_i é conhecido como método score de Fisher¹³ para maximização da função de verossimilhança¹⁴, que coincide com o método de Newton-Raphson quando a função de ligação é a canônica (Paula, 2004).

Desta forma, para o modelo de regressão de Poisson, a função densidade de probabilidade¹⁵ é:

$$f_y(y; \theta; \phi) = \exp\{y \log(\mu) - \mu - \log(y!)\} \quad \text{Eq. 3.1}$$

com $\theta = \log(\mu)$, $b(\theta) = e^\theta$, $\phi = 1$, $c(y, \phi) = -\log y!$, $V(\mu) = \mu$ e $w = \mu$ quando a função de ligação considerada é a ligação canônica ($\log \mu = \eta$), sendo que y é a variável resposta, θ é o parâmetro de dispersão, ϕ é o parâmetro canônico, V é a variância, w é a função peso, c e b são funções específicas, e η é a função de ligação.

Um aspecto importante de destacar em relação à regressão de Poisson é a respeito dos casos em que ocorre superdispersão ou subdispersão¹⁶. Uma maneira de acomodá-las nos modelos é assumir que a variância é um múltiplo da média, isto é, $V(\mu) = \phi\mu$ e estimar o parâmetro ϕ , ao invés de assumi-lo igual a 1 (Ribeiro, 2006).

3.2.2 Ajuste de tendências temporais

a) Dias da semana

Em estudos sobre o impacto da poluição atmosférica na saúde, é necessário levar em consideração a tendência temporal. Por exemplo, nos finais de semana o

¹³ Mais detalhes do método score de Fisher são apresentados no Anexo C.

¹⁴ A função de verossimilhança é baseada na função densidade de probabilidade, mais detalhes no Anexo C.

¹⁵ A função densidade de probabilidade é a probabilidade de que uma variável aleatória contínua X assumira um particular valor x (Fonseca e Martins, 1996).

número de atendimentos hospitalares é menor do que nos dias de semana. Uma forma de ajustar esta tendência é acrescentar uma variável qualitativa para dia da semana, que varia de 1 a 7, começando a contagem no domingo.

O número de atendimentos hospitalares nos feriados também é menor do que nos dias em que não é feriado. Esta tendência é ajustada através do acréscimo de uma variável binomial, ou seja, os dias de feriado recebem valor 1 e os dias em que não é feriado recebem valor 0.

b) Sazonalidade

Outra tendência temporal importante é a sazonalidade, pois as variáveis meteorológicas e concentração de poluentes variam no decorrer do ano. Para ajustar estas tendências frequentemente utiliza-se uma função chamada *spline*, pois esta função fornece uma aproximação melhor que as tendências polinomais (Chapra e Canale, 1987).

Spline é um tipo de função que fornece uma aproximação do comportamento das funções que têm mudanças locais e abruptas. Nos *splines*, ao invés de utilizar apenas um polinômio para todo o conjunto de dados, definem-se alguns intervalos e estima-se uma função polinomial para cada um dos intervalos (Chapra e Canale, 1987).

O *spline* mais utilizado para suavização de curvas é o *spline* cúbico. O objetivo dos *splines* cúbicos é derivar um polinômio de terceira ordem para cada intervalo entre 2 nós (ponto onde dois *splines* se encontram) que tem a seguinte forma:

$$f_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i \quad \text{Eq. 3.2}$$

Para $n+1$ pontos ($i = 1, 2, \dots, n$) existem n intervalos e, conseqüentemente, $4n$ constantes desconhecidas (a_i, b_i, c_i, d_i) a determinar. Assim, são necessárias $4n$ equações ou condições.

1 – Os valores das funções devem ser constantes nos nós, exceto para o primeiro e último nó ($2n - 2$ condições). Esta condição leva a:

¹⁶ Definições para superdispersão e subdispersão estão apresentadas em 2.3.3 Modelos de regressão de Poisson.

$$a_{i-1}x_{i-1}^3 + b_{i-1}x_{i-1}^2 + c_{i-1}x_{i-1} + d_{i-1} = f(x_{i-1}) \quad \text{Eq. 3.3}$$

$$a_i x_{i-1}^3 + b_i x_{i-1}^2 + c_i x_{i-1} + d_i = f(x_{i-1}) \quad \text{Eq. 3.4}$$

para $i = 2, 3, \dots, n$;

2 – A primeira e a última função devem passar pelos pontos finais. (2 condições), ou seja:

$$a_1 x_0^3 + b_1 x_0^2 + c_1 x_0 + d_1 = f(x_0) \quad \text{Eq. 3.5}$$

$$a_n x_n^3 + b_n x_n^2 + c_n x_n + d_n = f(x_n) \quad \text{Eq. 3.6}$$

3 – As primeiras derivadas nos nós (exceto o primeiro e o último) devem ser iguais para garantir a continuidade (n-1 condições);

4 – As derivadas segundas nos nós (exceto o primeiro e o último) devem ser iguais (n-1 condições).

Neste caso, tem-se 4n-2 condições e 4n constantes a serem determinadas. Portanto é preciso fazer alguma suposição para obter mais duas condições.

Se os valores das derivadas segundas nos nós finais não são nulas (ou seja, existe uma curvatura), esta informação pode ser usada para suprir as 2 condições necessárias. Porém, existe um caso particular de *spline* cúbico chamado de *spline* cúbico natural, onde a função *spline* torna-se uma linha reta nos nós finais, neste caso a condição necessária para completar o conjunto de equações é:

5 – As derivadas segundas nos pontos finais devem ser iguais a zero (2 condições).

Assim, utilizando as condições especificadas acima, obtêm-se os polinômios de terceiro grau para cada intervalo considerado. (Chapra e Canale, 1987). Em estudos como o apresentado neste trabalho, utilizam-se 4 a 5 nós por ano, pois a sazonalidade ocorre devido às diferenças existentes entre as estações do ano.

Apesar dos *splines* fornecerem uma boa aproximação para tendências temporais, não possuem interpretação física, portanto não podem ser utilizados para realizar previsões. Nestes casos, opta-se por utilizar uma variável qualitativa para os

meses do ano que possui valores entre 1 (referente à janeiro) e 12 (referente à dezembro) (Conceição et al., 2001).

c) Correlação de dados com o tempo

As tendências temporais podem provocar autocorrelação entre os dados, ou seja, os dados de um dia podem estar correlacionados com os dados do dia anterior. Assim, para verificar a existência de correlação entre os dados e decidir quais as providências a serem tomadas, são construídos gráficos da função de autocorrelação parcial *versus* tempo de defasagem (*lag*).

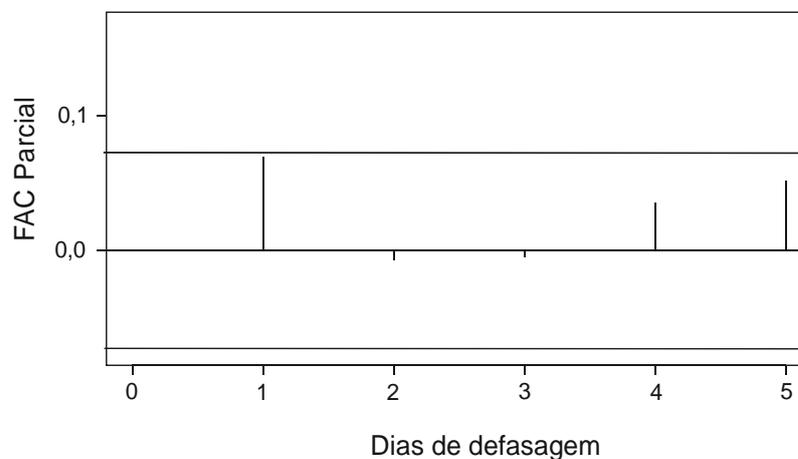


Figura 3.3 – Exemplo de gráfico da função de autocorrelação (FAC) parcial em relação a dias de defasagem (*lag*)

A função de autocorrelação do resíduo é dada por:

$$FAC = \frac{C_k}{C_0} \quad \text{Eq. 3.7}$$

onde $C_k = \frac{1}{n} \sum_{i=1}^{n-k} (y_i - \mu)(y_{i+k} - \mu)$ sendo que n é o número de observações e k representa os dias de defasagem (Box et al., 1994).

No gráfico da função de autocorrelação, os resíduos devem ser o menor possível, encontrando-se numa faixa entre $[-2/n^{1/2}, 2/n^{1/2}]$, onde n é o número de observações presentes no problema, como mostra a Figura 3.3 (Ferraz et al., 1999).

No caso de estudos epidemiológicos, as autocorrelações importantes e possíveis de serem interpretadas ocorrem nos primeiros 4 dias. Estas

autocorrelações geralmente são devido ao reduzido número de atendimentos hospitalares nos finais de semana.

Se os dados estão correlacionados, deve-se ajustar o modelo levando em consideração essas autocorrelações. Esta correção é feita através da inserção do resíduo no modelo.

Quando o modelo é ajustado para os resíduos (autocorrelação) é preciso comparar o modelo original com o modelo ajustado para autocorrelações, com o intuito de verificar se houve uma melhora significativa no modelo, caso contrário, opta-se pela utilização do modelo original, pois o modelo ajustado não pode ser utilizado para realizar previsões.

Uma vez escolhido o modelo que melhor se ajuste aos dados, é necessário avaliar este ajuste.

3.2.3 Avaliação do ajuste

Um teste interessante e fácil de ser aplicado para avaliar o ajuste de um modelo é chamado de estatística pseudo R^2 que é similar ao coeficiente de determinação obtido nos modelos lineares clássicos, e é definida por:

$$R_D^2 = 1 - \frac{D_c}{D_0} \quad \text{Eq. 3.8}$$

onde D_c é o desvio do modelo ajustado e D_0 é o desvio do modelo nulo, ou seja, desvio antes da aplicação do modelo. Esta estatística mede a redução no desvio devido à inclusão de variáveis explicativas (Ribeiro, 2006).

A principal ferramenta capaz de avaliar o ajuste do modelo, porém, é a análise de resíduos. Para os MLG, são definidos quatro tipos diferentes de resíduos capazes de avaliar o ajuste do modelo, o desvio residual, resíduo *working*, resíduo de Pearson e resíduo resposta. Dentre eles, será utilizado o desvio residual (*deviance residual*), que é capaz de detectar observações atípicas que influenciam o processo de ajuste do modelo e o resíduo de Pearson por ser uma versão aprimorada (*rescaled*) do resíduo *working* (S-Plus 7, 2005).

O resíduo de Pearson está presente em uma estatística bastante utilizada para avaliar o ajuste do modelo, chamada estatística de Pearson ou Qui-quadrado χ^2 capaz de comparar a distribuição observada com a determinada pelo modelo através da seguinte expressão (Schmidt, 2003):

$$\chi^2 = \sum (y - \hat{\mu})^2 / V(\hat{\mu}) \quad \text{Eq. 3.9}$$

onde $V(\hat{\mu})$ é a função de variância estimada para a distribuição em questão (Schmidt, 2003).

Pode-se dizer que a estatística de Pearson é a soma dos resíduos de Pearson para cada observação.

Um modelo que se ajuste bem aos dados possui estatística de Pearson χ^2 aproximadamente igual ao seu grau de liberdade (gl), ou seja, $\chi^2 / gl \sim 1$, caso contrário, pode se dizer que o modelo é inadequado, podendo tratar-se de um problema de superdispersão (Wang et al., 1996 apud Russo, 2002).

3.2.4 Análise gráfica

Entre os tipos de análises gráficas comumente utilizadas para avaliar o ajuste de um MLG utilizou-se (S-Plus 7, 2005):

1 – Gráfico dos desvios residuais de cada observação *versus* valores ajustados pelo modelo.

Um modelo bem ajustado possui o gráfico com os pontos o mais próximo possível de zero no intervalo entre -2 e 2, como mostra a Figura 3.4 (Everitt, 2003).

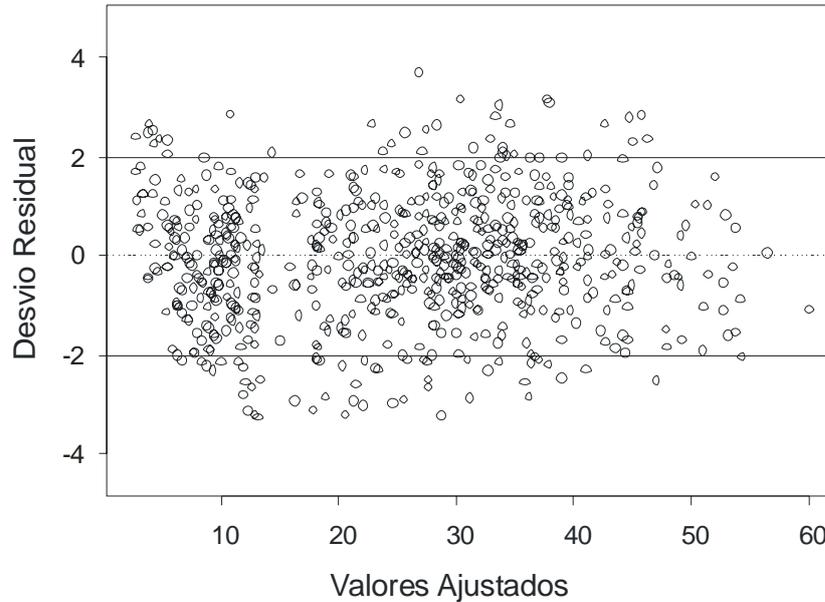


Figura 3.4 – Exemplo de gráfico dos desvios residuais em função dos valores ajustados pelo modelo

O desvio residual, que em alguns trabalhos aparece como *deviance residual* ou resíduo *deviance*, é definido por:

$$r_i^D = |y_i - \hat{\mu}_i| \sqrt{d_i} \quad \text{Eq. 3.10}$$

onde d_i é a contribuição da i -ésima observação para o resíduo (S-Plus 7, 2005).

2 – Gráfico dos valores observados da variável resposta *versus* valores ajustados pelo modelo.

Um exemplo deste gráfico está na Figura 3.5. Os pontos deste gráfico devem estar próximos da linha em que $y = x$, indicando que os valores ajustados estão próximos dos valores observados.

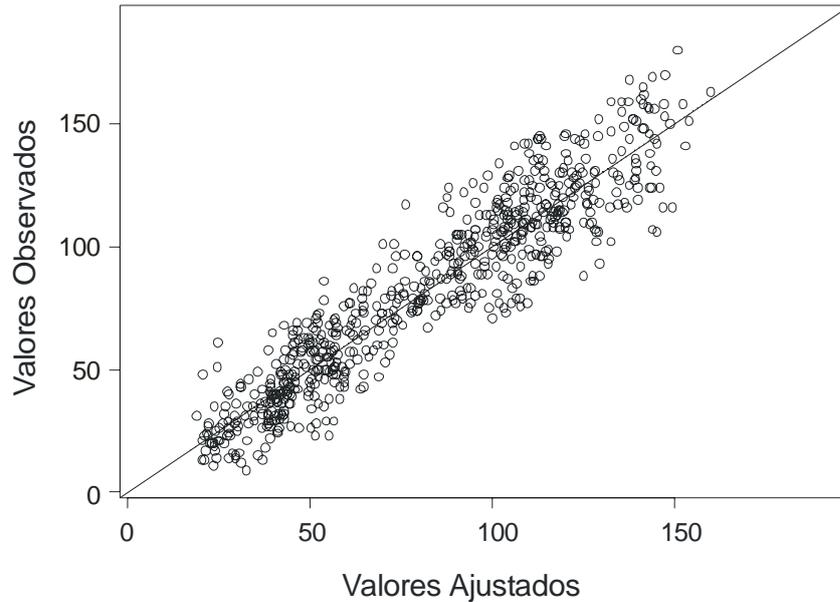


Figura 3.5 – Exemplo de gráfico dos valores observados em função dos valores ajustados pelo modelo

3.3 Análise dos Resultados

Para verificar a significância estatística dos coeficientes de regressão ajustados utiliza-se um teste de hipótese¹⁷. A hipótese estatística a ser testada é designada por H_0 , chamada hipótese nula, expressa por uma igualdade. A hipótese alternativa é dada por uma desigualdade (Fonseca e Martins, 1996).

O teste de hipótese possui várias finalidades, uma delas é verificar se o coeficiente de regressão estimado pode ser desprezado. Neste caso, consideram-se as hipóteses $H_0 : \beta = 0$ e $H_1 : \beta \neq 0$.

A estatística teste utilizada para verificar esta hipótese é:

$$t_0 = \beta / \varepsilon \quad \text{Eq. 3.11}$$

onde ε é o erro padrão do coeficiente de regressão (β) estimado.

A rejeição da hipótese nula ocorre quando $|t_0| > t_{\alpha/2, n-k-1}$ (n = número de observações, k = número de variáveis explicativas), indicando que o valor

¹⁷ Teste de hipótese – É uma regra de decisão utilizada para verificar a validade ou não de uma hipótese formulada a partir dos dados observados. (Fonseca e Martins, 1996).

encontrado para o coeficiente de regressão é estatisticamente significativo, ou seja, a variável explicativa considerada influencia nos resultados da variável resposta.

Os valores $t_{\alpha/2, n-k-1}$ são apresentados na Tabela A.2 de distribuição t de *student*¹⁸ do Anexo A, onde gl é o grau de liberdade, dado por $n-k-1$ e α é o nível de significância considerado.

Para a análise dos resultados em estudos epidemiológicos frequentemente utiliza-se uma medida chamada risco relativo, calculada através dos parâmetros estimados no modelo (Baxter et al., 1997).

3.4 Risco Relativo

O risco relativo é uma medida da associação entre um fator particular (por exemplo, a concentração de poluentes atmosféricos) e o risco de um certo resultado (por exemplo, o número de pessoas com problemas respiratórias em uma região) (Everitt, 2003).

De forma mais específica, a função risco relativo para um nível x de um poluente (Y) é definida por (Baxter et al., 1997):

$$RR(x) = \frac{E(Y|x)}{E(Y|x=0)} \quad \text{Eq. 3.12}$$

Para o modelo de regressão de Poisson, o risco relativo é dado por:

$$RR(x) = e^{\beta x} \quad \text{Eq. 3.13}$$

Isto indica que, o risco de uma pessoa exposta a uma concentração de MP_{10} (x) adquirir doenças respiratórias é $RR(x)$ vezes maior que uma pessoa que não foi exposta a esta concentração.

Um risco relativo igual a cinco para uma concentração de MP_{10} de $100 \mu\text{g}/\text{m}^3$, por exemplo, indica que uma pessoa exposta a uma concentração de MP_{10} de 100

¹⁸ Distribuição t de *student* – É uma distribuição contínua semelhante à distribuição normal utilizada para realizar testes de hipótese (Fonseca e Martins, 1996).

$\mu\text{g}/\text{m}^3$ possui cinco vezes mais chance de adquirir uma doença respiratória que uma pessoa que não foi exposta a esta concentração.

O processo de busca pelo modelo que melhor se ajuste a um conjunto de dados, apresentado neste capítulo, pode ser utilizado para qualquer problema que envolva uma variável resposta não negativa em forma de contagem.

Na Figura 3.6, é apresentado um fluxograma com as principais etapas no processo de busca pelo modelo mais adequado aos dados coletados.

A fim de verificar se a metodologia proposta pode ser aplicada para avaliar o impacto da poluição atmosférica na saúde, realizou-se um estudo de caso para o município de Araucária.

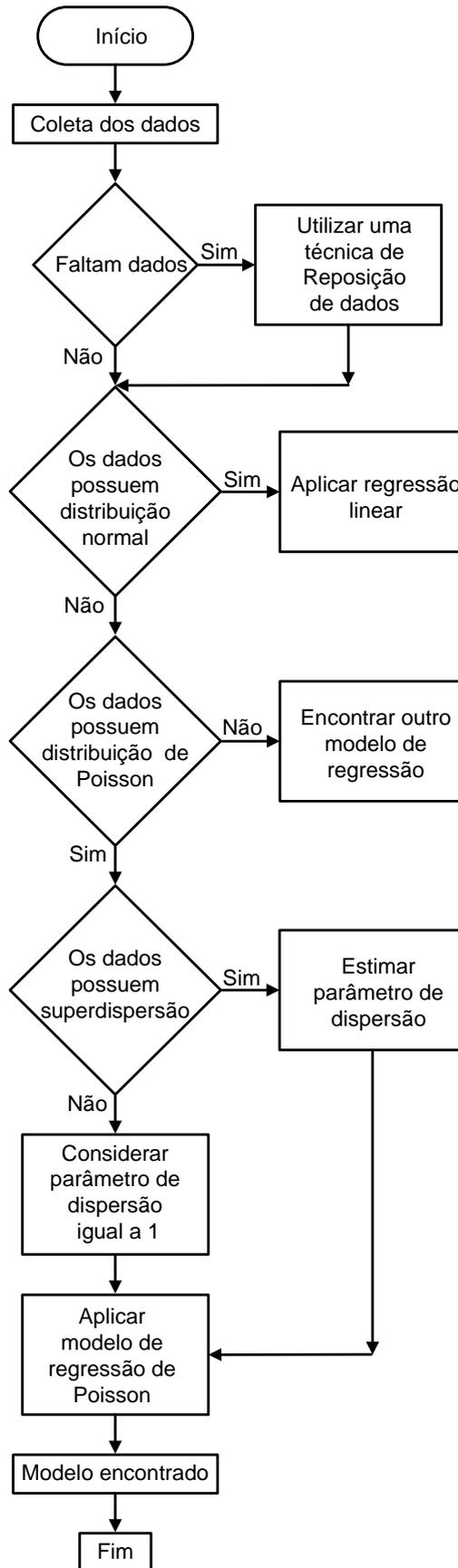


Figura 3.6 – Fluxograma de escolha do modelo estatístico apropriado aos dados coletados

4 ESTUDO DE CASO

No intuito de verificar se o modelo de regressão de Poisson é adequado para avaliar e prever o impacto da poluição atmosférica na saúde foram realizadas análises considerando a relação do número de pessoas com doenças respiratórias (DR) no município de Araucária com a concentração de MP_{10} , bem como com algumas variáveis meteorológicas.

Para verificar a existência de uma relação explícita entre as variáveis explicativas e o número de pessoas com doenças respiratórias no município de Araucária, foi realizada uma análise inicial através de gráficos bidimensionais entre estas variáveis. Em seguida, através das etapas propostas na metodologia, aplicou-se o modelo de regressão de Poisson dos MLG para os dados de Araucária, com o objetivo de quantificar a relação entre as variáveis consideradas.

4.1 Características do Município de Araucária

O município de Araucária faz parte da Região Metropolitana de Curitiba (RMC) que atualmente abrange 26 municípios, como mostra a Figura 4.1.



Figura 4.1 – Mapa da Região Metropolitana de Curitiba (Adaptado de IPPUC, 2006)

Na década de 70, o município de Araucária apresentou mudanças significativas em seu perfil econômico e populacional, especialmente com a implantação da Refinaria Presidente Getúlio Vargas (REPAR), a criação da Cidade Industrial de Curitiba (CIC) e do Centro Industrial de Araucária (CIAR). A partir de então, houve um rápido crescimento populacional na região, maior que o crescimento do Brasil na mesma época, como mostra a Figura 4.2. O crescimento populacional do Brasil da década de 80 para 90 foi de 21% enquanto o crescimento populacional do município de Araucária foi de 70% na mesma época.

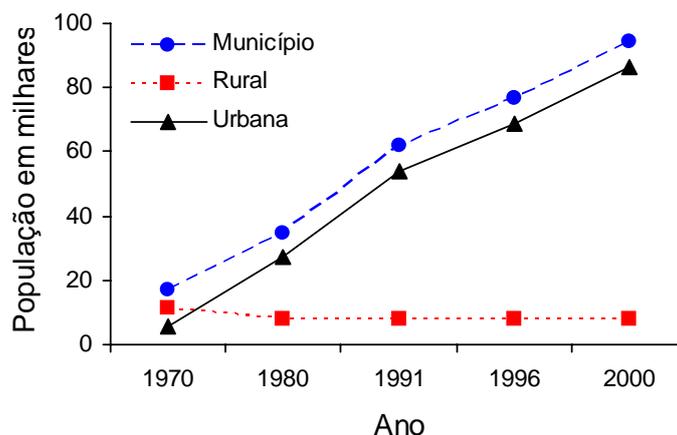


Figura 4.2 - População do município de Araucária desde a década de 70 até o ano 2000 (IBGE, 2000 apud Prefeitura de Araucária, 2005)

De acordo com dados do Ministério da Saúde e IBGE (Instituto Brasileiro de Geografia e Estatística) (2006), a população do município de Araucária continua aumentando, como mostra a Tabela 4.1, onde é possível verificar uma predominância de adultos na região.

Tabela 4.1 – População do município de Araucária nos anos de 2004 e 2005 por faixa etária (Ministério da Saúde e IBGE, 2006)

Faixa etária	2004	2005
Menores de 1 ano	2.275	2.428
1 a 4 anos	9.212	9.829
5 a 9 anos	11.368	12.129
10 a 14 anos	10.865	11.594
15 a 19 anos	10.760	11.482
20 a 29 anos	20.824	22.218
30 a 39 anos	17.615	18.794
40 a 49 anos	12.230	13.050
50 a 59 anos	6.406	6.836
60 a 69 anos	3.621	3.863
70 a 79 anos	1.705	1.820
Maiores de 80 anos	569	607
Total	107.450	114.650

Devido ao rápido crescimento populacional e à grande quantidade de indústrias presentes no município de Araucária, implantou-se a rede de monitoramento da qualidade do ar na RMC. Atualmente, a RMC conta com 12 estações de monitoramento da qualidade do ar, como mostra a Figura 4.3. A localização e a

categoria das estações e os parâmetros (poluentes e variáveis meteorológicas) monitorados podem ser observados no Quadro A.1 do Anexo A.

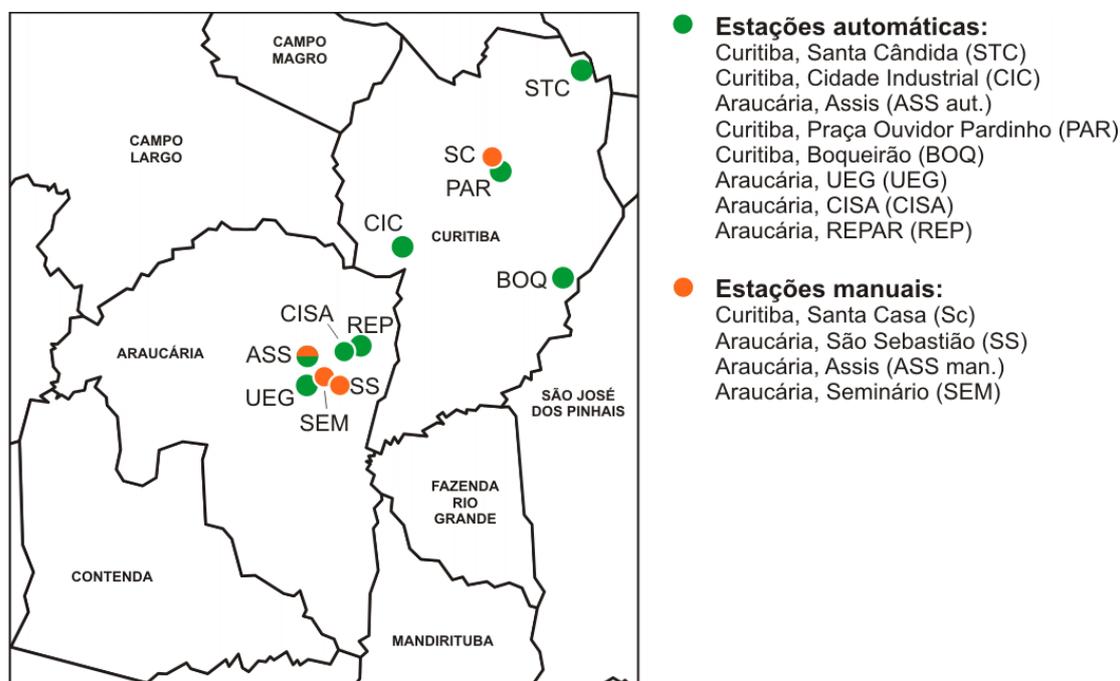


Figura 4.3 – Localização das estações de monitoramento da qualidade do ar na RMC (Adaptado de IAP, 2005)

4.2 Dados de Saúde da Região Sul do Brasil

De acordo com os dados do Ministério da Saúde e IBGE (Instituto Brasileiro de Geografia e Estatística) (2006), no período de 2004 e 2005, a Região Sul era a região brasileira com a maior porcentagem de pessoas com problemas respiratórios, sendo que o estado da região Sul com a maior porcentagem foi o Paraná (Tabela 4.2).

Apesar da cidade de Curitiba possuir uma população aproximadamente 15 vezes maior que a população do município de Araucária, a incidência de doenças respiratórias na população da cidade de Curitiba para o período de 2004 e 2005 foi de 0,46% e 0,48% respectivamente, enquanto no município de Araucária foi de 0,91% e 0,90%; ou seja, a porcentagem de pessoas com problemas respiratórias no

município de Araucária é aproximadamente duas vezes maior que na cidade de Curitiba.

Tabela 4.2 – Dados epidemiológicos e demográficos por região brasileira, bem como para os estados da região Sul (Ministério da Saúde e IBGE, 2006)

Ano	Região	Estados	Pessoas com DR	População Total	% DR/Pop. Total
2004	Norte		139.835	14.373.260	0,97
	Nordeste		513.559	50.427.274	1,02
	Sudeste		556.941	77.374.720	0,72
	Sul	PR	126.313	10.135.388	1,25
		SC	57.534	5.774.178	0,99
		RS	129.036	10.726.063	1,20
		Total	343.649	26.635.629	1,29
	Centro-Oeste		160.012	12.770.141	1,25
Brasil		1.713.996	181.581.024	0,94	
2005	Norte		136.042	14.698.878	0,93
	Nordestes		457.231	51.019.091	0,89
	Sudeste		516.554	78.472.017	0,66
	Sul	PR	312.883	26.973.511	1,16
		SC	137.917	10.261.856	1,34
		RS	61.420	5.866.568	1,05
		Total	144.312	10.845.087	1,33
	Centro-Oeste		143.033	13.020.767	1,10
Brasil		1.565.743	184.184.264	0,853	

4.3 Coleta e Análise dos Dados

Para verificar a influência da concentração de MP_{10} no número de pessoas com problemas respiratórios em Araucária realizou-se uma análise para o período entre 01 de janeiro de 2004 e 31 de dezembro de 2005. Para tanto, a Secretaria Municipal da Saúde de Araucária forneceu dados de atendimentos hospitalares por DR nos hospitais públicos do município de Araucária para o período de estudo de acordo com a Classificação Internacional das Doenças (Organização Mundial da Saúde, 1997).

De acordo com a Classificação Internacional das Doenças, são consideradas doenças respiratórias: infecções agudas das vias aéreas superiores; influenza (gripe) e pneumonia; outras infecções agudas das vias aéreas inferiores; outras doenças das vias aéreas superiores; doenças crônicas das vias aéreas inferiores; doenças pulmonares devido a agentes externos; outras doenças respiratórias que

afetam principalmente o interstício; afecções necróticas e supurativas das vias aéreas inferiores; outras doenças da pleura e outras doenças do aparelho respiratório.¹⁹

Os dados horários da concentração de MP₁₀ e variáveis meteorológicas foram fornecidos pelo IAP (Instituto Ambiental do Paraná). A SMMA (Secretaria Municipal do Meio Ambiente de Araucária) junto ao Lactec (Instituto de Tecnologia para o Desenvolvimento) também forneceram dados horários de variáveis meteorológicas. A porcentagem de dados fornecidos tanto pelo IAP quanto pela SMMA/Lactec está na Tabela 4.3.

Tabela 4.3 – Porcentagem de dados de cada variável fornecida pelo IAP e SMMA/Lactec

	Dados do IAP		Dados da SMMA/Lactec	
	2004 (%)	2005 (%)	2004 (%)	2005 (%)
Concentração de MP₁₀ (µg/m³)	75,25	93,18	-	-
Temperatura (°C)	38,41	93,08	91,95	96,44
Umidade relativa do ar (%)	-	7,01	91,95	96,43
Radiação Total (W/m²)	80,85	-	90,00	97,03
Pressão Atmosférica (mbar)	-	93,56	89,96	96,07
Velocidade do Vento (m/s)	96,48	93,64	89,67	96,42
Direção do Vento (graus)	96,48	93,55	89,71	33,14

Nota: Os dados fornecidos pelo IAP foram da estação de monitoramento de qualidade do ar da UEG e os dados da SMMA/Lactec foram da estação de monitoramento de qualidade do ar Assis automática (ASS).

Devido à grande quantidade de dados faltantes não foi possível realizar uma média entre os dados fornecidos pelos diferentes órgãos, assim optou-se por utilizar os dados de temperatura e umidade fornecidos pela SMMA/Lactec e os dados de concentração de MP₁₀ fornecidos pelo IAP devido à maior quantidade de dados disponíveis.

Dentre as variáveis disponíveis foram utilizadas como variáveis explicativas a temperatura, a umidade e a concentração de MP₁₀, como nos estudos de Martins (2000), Bakonyi et al. (2004), Bakonyi (2003), Martins et al. (2002), Braga et al. (2001) e Abbey et al. (1999).

¹⁹ Mais detalhes em Organização Mundial da Saúde (1997).

As Figuras 4.4 e 4.5 mostram a porcentagem de dados faltantes das variáveis explicativas escolhidas para 2004 e 2005 respectivamente. Os dados de concentração de MP₁₀ para os meses de setembro e novembro de 2004 não foram fornecidos e a porcentagem de dados disponíveis para 2005 foi maior que para 2004.

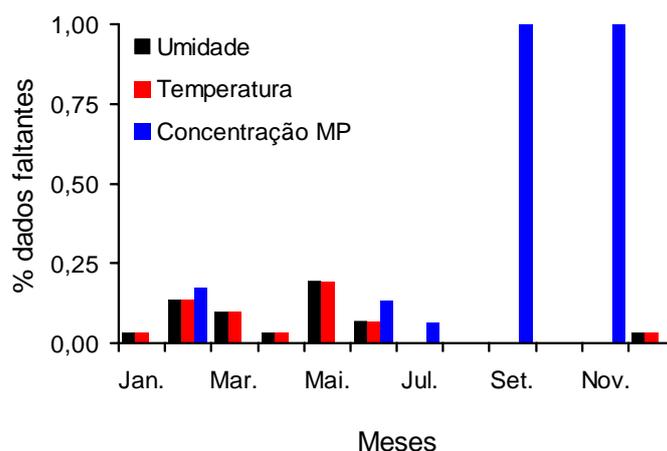


Figura 4.4 – Porcentagem de dados faltantes para o ano de 2004

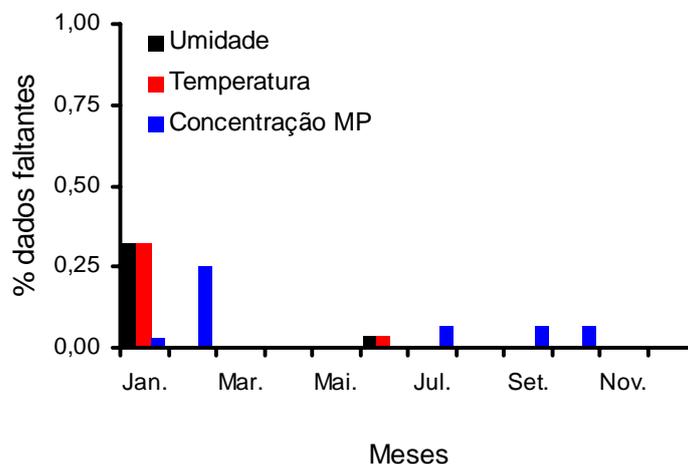


Figura 4.5 – Porcentagem de dados faltantes para o ano de 2005

A partir dos dados horários das variáveis meteorológicas e da concentração de MP₁₀ foram calculadas as médias diárias seguindo o critério utilizado pelo IAP (2004) e outros órgãos responsáveis pelo monitoramento da qualidade do ar como a

CETESB (2005) (Companhia de Tecnologia de Saneamento Ambiental). Estes órgãos consideram representativa uma média diária baseada em pelo menos 16 médias horárias válidas.

Após o cálculo das médias diárias, faltaram dados de temperatura e umidade para 18 dias em 2004 e 11 dias em 2005. Para a concentração de MP_{10} faltaram dados de 71 dias em 2004, (meses de setembro e novembro inteiros) e 14 dias em 2005.

A fim de obter um banco de dados completo, foram utilizados dados de regiões próximas ao município de Araucária, por possuírem um clima semelhante.

No site do SIMEPAR (Tecnologias e Informações Ambientais) foram obtidos os dados de temperatura da cidade de Curitiba. Através da comparação dos dados de temperatura de Curitiba (SIMEPAR) e Araucária (IAP) verificou-se uma correlação linear de 0,92 entre os dados. Desta forma, os dados de temperatura de Curitiba foram utilizados para reposição dos dados faltantes de Araucária. Em dois dias de 2004 coincidiu de não possuir dados tanto para Curitiba quanto Araucária, sendo necessário calcular a média móvel entre os dias anterior e posterior.

Os dados de umidade de Curitiba não tiveram uma boa correlação linear com os dados de Araucária (0,51), porém a correlação linear entre os dados de Araucária e de Pinhais foi de 0,83. Neste caso, os dados de umidade de Pinhais foram utilizados para repor os dados faltantes de Araucária.

O maior desafio foi completar o banco de dados para concentração de MP_{10} , pois a SMMA/Lactec não monitora MP_{10} e apesar de terem sido solicitados ao IAP dados para o período entre 2000 e 2005, foram fornecidos apenas dados de 2004 e 2005.

Como o objetivo principal do trabalho é estudar e validar o modelo estatístico utilizado para avaliar o impacto da poluição atmosférica na saúde populacional e não simplesmente quantificar a influência da concentração de MP_{10} na saúde populacional do município de Araucária, os dados de 2004 foram utilizados para repor os dados faltantes de 2005 e vice-versa, apesar da baixa correlação linear entres os dados de 2004 e 2005 (0,29).

Devido à baixa correlação entre os dados de concentração de MP_{10} de 2004 e 2005 quando faltavam apenas 2 dias consecutivos de dados foi utilizada a média móvel dos dias anterior e posterior.

Dos 71 dias com dados faltantes de 2004 foi possível repor apenas 4 dias com a média móvel, sendo que os 67 dias restantes foram repostos com dados de 2005, incluindo os meses de setembro e novembro inteiros. Para o ano de 2005, os dados de apenas cinco dias foram repostos por dados de 2004.

Com o banco de dados completo foram realizadas algumas análises preliminares antes de aplicar o modelo de regressão de Poisson.

4.4 Resultados Preliminares

A fim de verificar a relação existente entre as variáveis meteorológicas, a concentração de MP_{10} e o número de pessoas com doenças respiratórias no município de Araucária foram realizadas as análises gráficas apresentadas abaixo.

Apesar dos adultos serem maioria no município de Araucária, o número de crianças com doenças respiratórias (DR) para o período entre 2004 e 2005 foi maior que as demais faixas etárias (Figura 4.6).

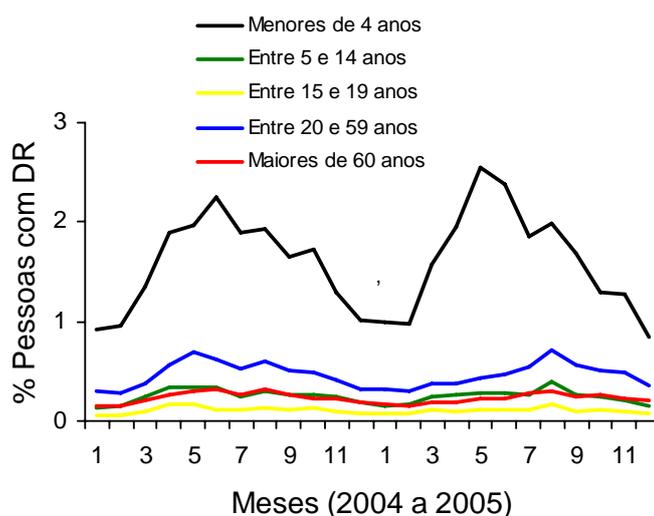


Figura 4.6 – Gráfico da porcentagem mensal de pessoas com problemas respiratórios por faixa etária em Araucária para o período de 2004 a 2005

De acordo com a resolução nº 3 de 28/6/90 – CONAMA, a concentração de MP_{10} não pode exceder o limite diário de $150 \mu\text{g}/\text{m}^3$ mais de uma vez ao ano. Para o período entre 2004 e 2005, a concentração de MP_{10} em Araucária ultrapassou o limite diário em 13 dias de 2004 e 4 dias de 2005, sendo que a concentração diária máxima em 2004 foi de $230,66 \mu\text{g}/\text{m}^3$ no dia 18 de agosto e em 2005 foi de $158,13 \mu\text{g}/\text{m}^3$ para o dia 04 de agosto.

O gráfico da Figura 4.7 A apresenta o número diário de atendimentos hospitalares de pessoas com doenças respiratórias nos hospitais públicos do município de Araucária para o período de estudo, mostrando que o número de pessoas com DR foi maior nos períodos de inverno. Nesta época, as temperaturas são mais baixas, como mostra a Figura 4.7 C.

Um fator peculiar da região estudada é a presença de alta umidade relativa do ar durante todo o ano devido ao clima atípico da região. Por esta razão, a Figura 4.7 D não mostra uma diferença na umidade relativa do ar entre as estações do ano (Araucária, 2003 apud Souza, 2006).

A partir da Figura 4.7 B nota-se que as concentrações de MP_{10} foram maiores no inverno, sendo que a concentração de MP_{10} no inverno de 2005 estava mais alta que em 2004.

Comparando os gráficos da Figura 4.7, pode-se concluir que o número de pessoas com DR nos períodos de inverno é maior devido às condições meteorológicas desfavoráveis (baixa temperatura) à dispersão dos poluentes.

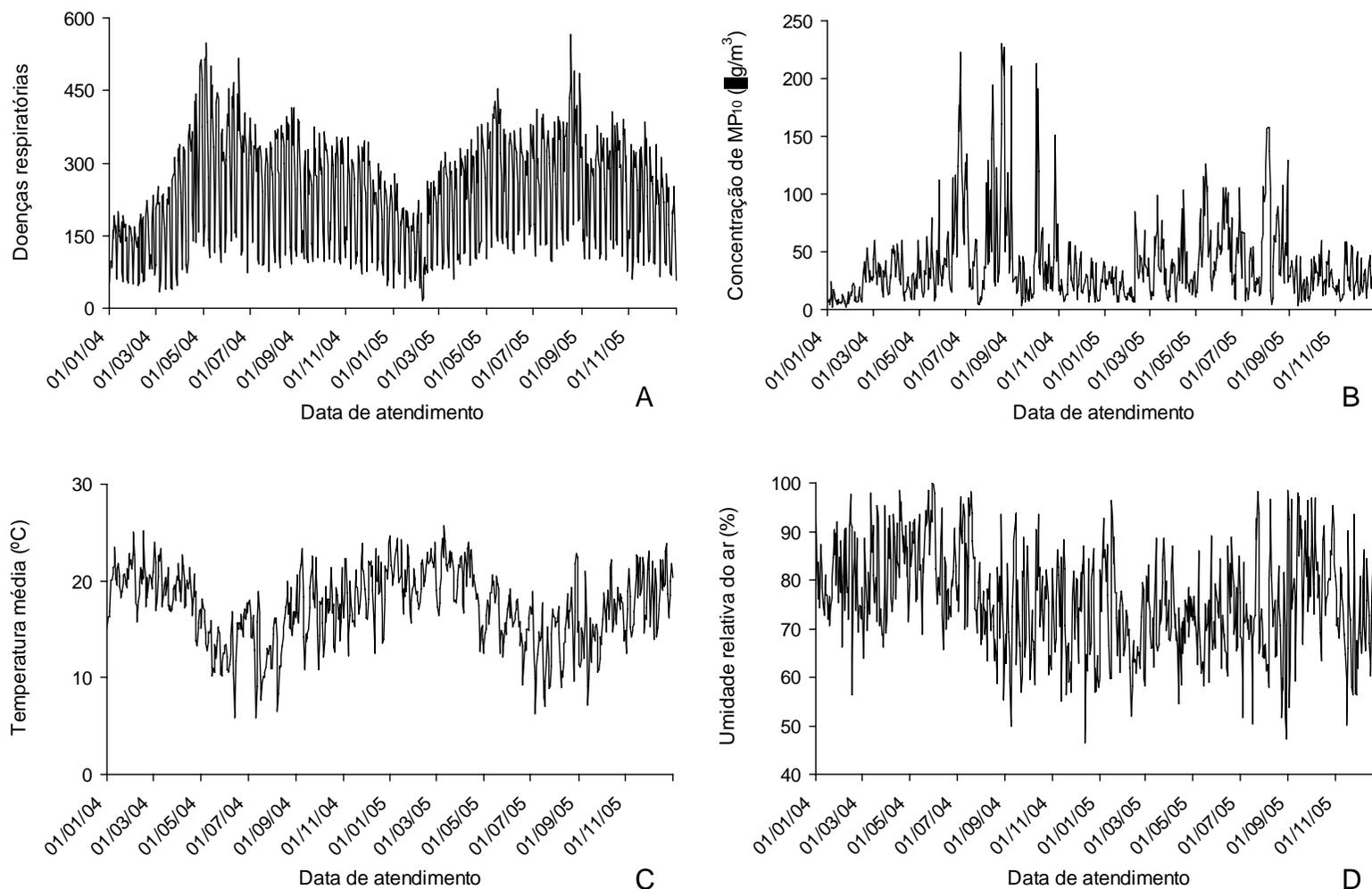


Figura 4.7 – Gráficos do número de pessoas com DR total diário (A), concentração de MP₁₀ (B), temperatura (C) e umidade (D) médias diárias no município de Araucária para 2004 e 2005

No intuito de verificar a existência de uma relação clara entre o número de pessoas com DR e as variáveis meteorológicas foram construídos os gráficos das Figuras 4.8 e 4.9.

Pela Figura 4.8 nota-se que a porcentagem de pessoas com problemas respiratórios diminuiu conforme a temperatura aumentou, confirmando os resultados apresentados na Figura 4.7.

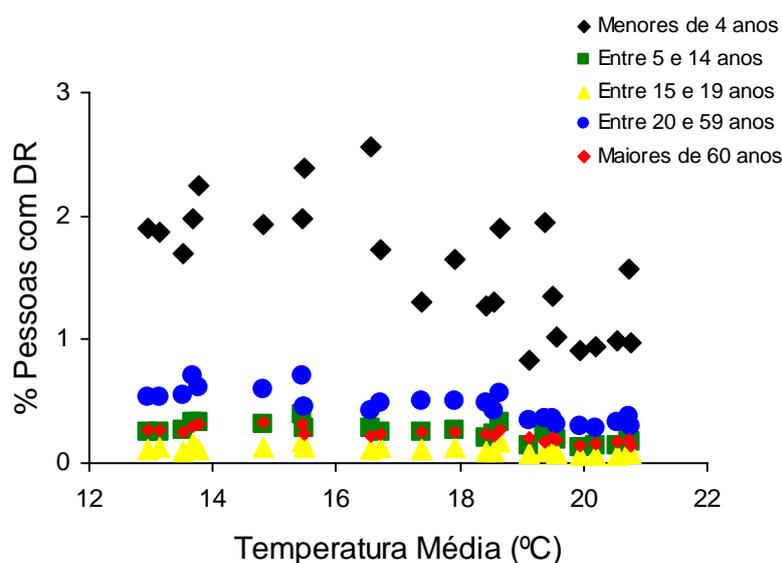


Figura 4.8 – Gráfico da porcentagem de pessoas com doenças respiratórias por faixa etária em relação à temperatura média mensal

O material particulado por definição é uma mistura de partículas sólidas e líquidas suspensas no ar que possui solubilidade maior que os poluentes gasosos, portanto, espera-se que a concentração de MP_{10} esteja fortemente relacionada com a umidade relativa do ar (Seinfeld, 1986). Neste caso, a umidade relativa do ar deve estar inversamente relacionada com o número de pessoas com DR. A Figura 4.9 mostra que a porcentagem de pessoas com DR não está relacionada com a umidade relativa do ar, provavelmente devido à alta umidade existente durante todo o ano na região em estudo.

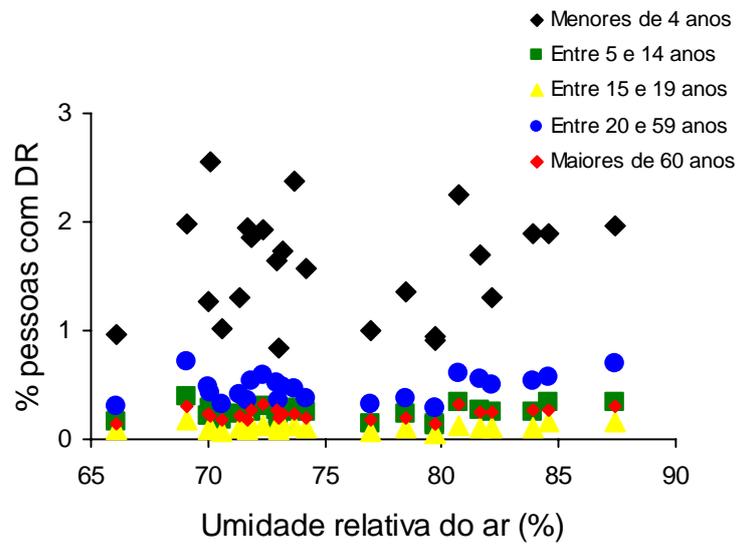


Figura 4.9 – Gráfico da porcentagem de pessoas com doenças respiratórias por faixa etária em relação à média mensal da umidade relativa do ar

A partir da Figura 4.10 verifica-se que a porcentagem de pessoas com DR possui uma relação razoavelmente linear com a concentração de MP_{10} (0,46), ou seja, conforme a concentração de MP_{10} aumenta o número de pessoas com problemas respiratórios também aumenta.

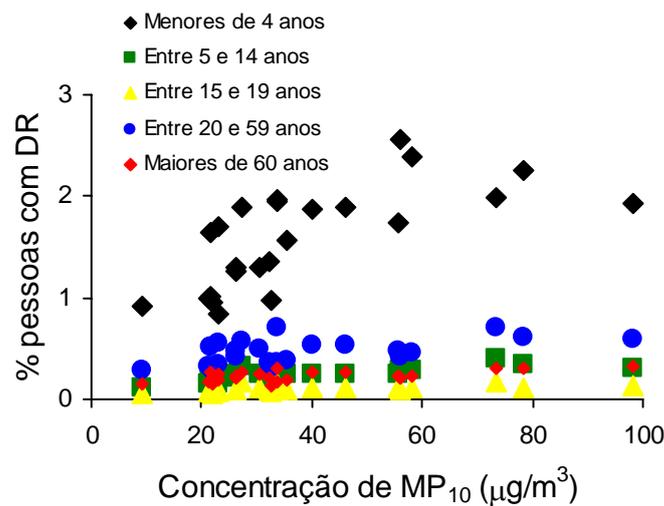


Figura 4.10 – Gráfico da porcentagem de pessoas com doenças respiratórias por faixa etária em relação à concentração média mensal de MP_{10} .

A fim de dimensionar a relação entre o número de pessoas com doenças respiratórias, a concentração de MP_{10} e as variáveis meteorológicas com maior precisão foram realizadas análises estatísticas com base nas características da variável resposta (pessoas com doenças respiratórias) e das variáveis explicativas (concentração de MP_{10} , temperatura e umidade).

4.5 Escolha do Modelo Estatístico

Para confirmar a possibilidade de se aplicar o modelo de regressão de Poisson aos dados coletados realizou-se a análise dos histogramas das variáveis respostas, ou seja, comparou-se o gráfico que representa a frequência de ocorrência dos valores da variável considerada apresentado no eixo das abscissas com a distribuição de Poisson.

Uma distribuição de Poisson tende para uma distribuição normal conforme a média aumenta, portanto a distribuição normal é um caso específico da distribuição de Poisson (Schmidt, 2003). Para verificar se os histogramas da Figura 4.11 seguem uma distribuição de Poisson, realizou-se uma comparação com as curvas apresentadas na Figura 2.1.

A Figura 4.11 mostra que todas as variáveis seguem uma distribuição aproximadamente de Poisson, sendo que para algumas faixas etárias a distribuição está mais próxima de uma distribuição normal, indicando um caso de superdispersão.

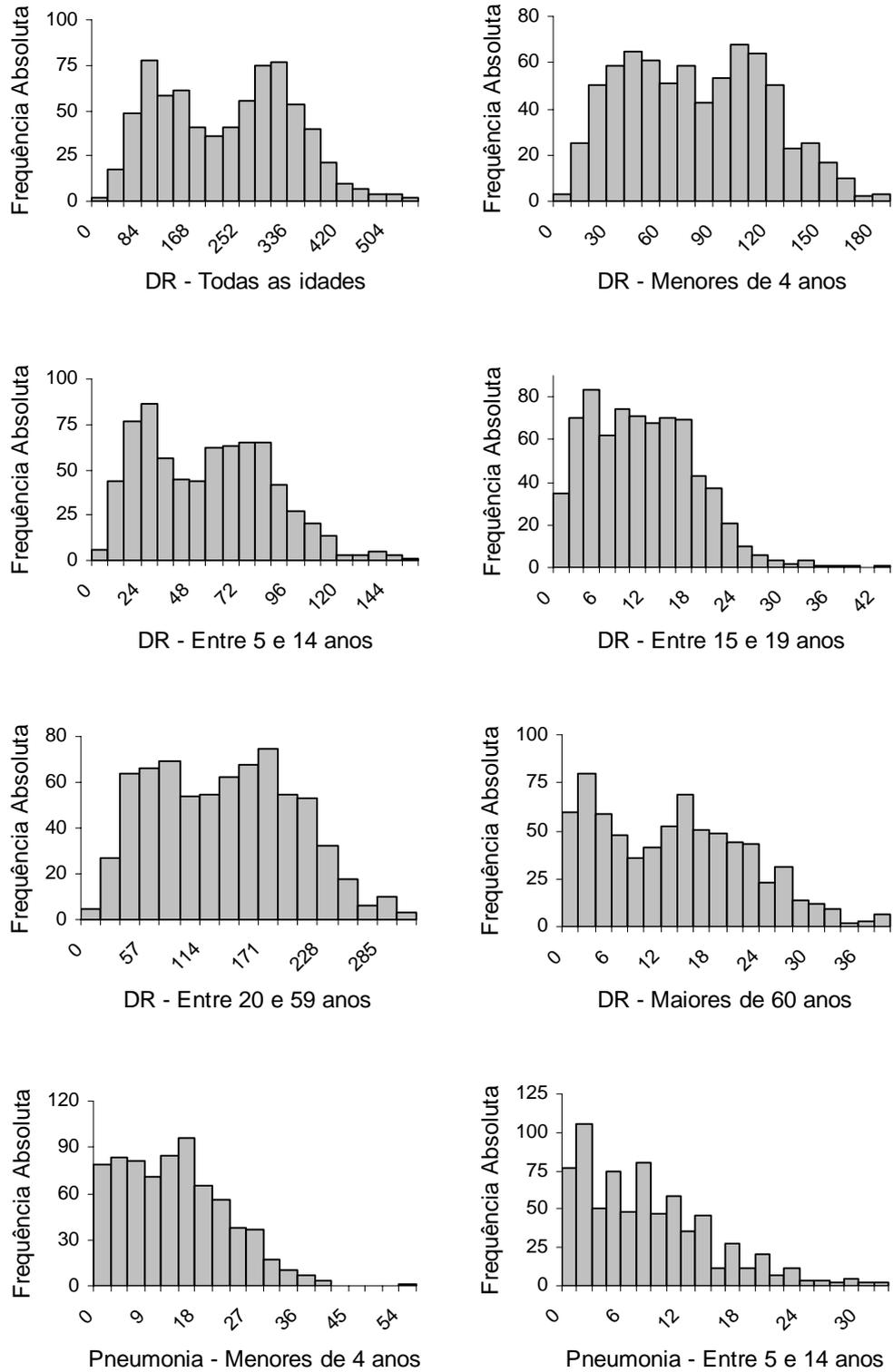


Figura 4.11 – Histogramas das variáveis respostas

No intuito de identificar a presença de superdispersão dos dados, a Tabela 4.4 mostra a estatística descritiva para as variáveis respostas consideradas neste estudo. A média diária com seu respectivo desvio padrão, o número diário de pessoas com doenças respiratórias mínimo e máximo referem-se a todo o período de estudo, ou seja, dois anos. Em todos os casos existe superdispersão, ou seja, a variância é maior que a média. Os dados com maior dispersão foram de DR para todas as idades (variância aproximadamente 55 vezes maior que a média) e os dados com menor dispersão foram de doenças respiratórias para pessoas entre 15 e 19 anos (variância aproximadamente 4 vezes maior que a média).

O modelo de regressão de Poisson pode ser aplicado mesmo em casos com superdispersão, porém é necessário estimar o parâmetro de dispersão ao invés de considerá-lo igual a 1.

Tabela 4.4 – Estatística descritiva de DR para cada faixa etária

	Todas as idades	Menores de 4 anos	Entre 5 e 14 anos	Entre 15 e 19 anos	Entre 20 e 59 anos	Maiores de 60 anos	Pneumonia - Menores de 4 anos	Pneumonia - Entre 5 e 14 anos
Média	233,693	81,617	57,551	12,211	141,801	14,190	14,823	8,844
Desvio padrão	113,896	39,949	31,196	7,063	71,011	9,101	9,374	6,614
Variância	12.972,385	1.595,914	973,171	49,885	5.042,578	82,823	87,87	43,741
Mínimo	17	5	5	0	8	0	0	0
Máximo	565	191	156	45	390	40	58	33

Os resultados da análise descritiva para as variáveis explicativas adotadas no estudo estão na Tabela 4.5. Os valores médios também se referem à média para o período de estudo, ou seja, média de dois anos. O valor do desvio padrão para a concentração de MP_{10} está próximo ao valor da média, indicando um fraco ajuste dos dados para MP_{10} . Como o objetivo do trabalho é verificar a validade do modelo estatístico apresentado, estes dados foram utilizados apesar de não serem tão fidedignos.

Tabela 4.5 – Estatística descritiva para variáveis explicativas

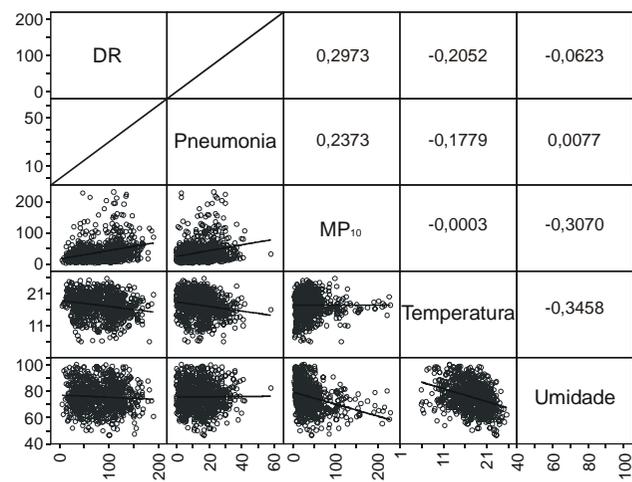
	MP₁₀	Temperatura	Umidade
Média	38,652	17,350	75,674
Desvio padrão	35,561	3,721	10,483
Mínimo	2,278	5,904	46,407
Máximo	230,665	25,679	99,920

A fim de verificar a existência de linearidade entre os dados, foram construídas as matrizes de dispersão das Figuras 4.12 e 4.13. O eixo das abscissas dos diagramas de dispersão (abaixo da diagonal principal) equivale à variável apresentada na mesma coluna e o eixo das coordenadas equivale à variável apresentada na mesma linha.

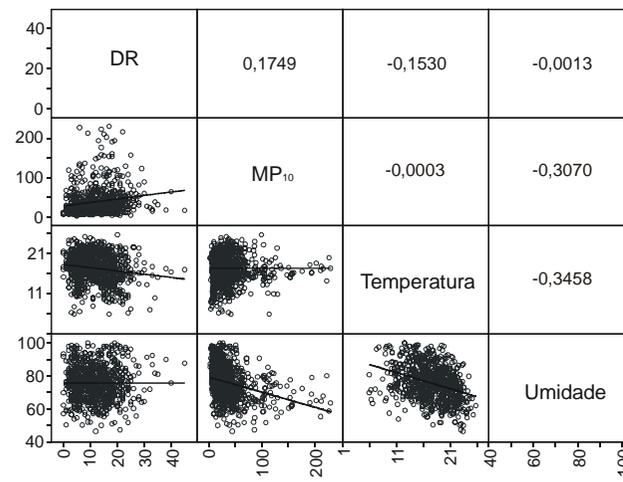
Os diagramas de dispersão mostram uma forte correlação entre duas variáveis quando os pontos seguem uma linha reta. Observando os diagramas de dispersão apresentados abaixo da diagonal principal das matrizes de dispersão das Figuras 4.12 e 4.13 verificou-se uma baixa correlação entre as variáveis. As correlações lineares entre as variáveis são apresentadas acima da diagonal principal. O valor 0,2973 da Figura 4.12 A, por exemplo, equivale à correlação entre o número de crianças menores de 4 anos com doenças respiratórias e a concentração de MP₁₀.

Os valores das correlações entre as variáveis explicativas e o número de pessoas com DR são baixos, porém é possível verificar que o número de pessoas com doenças respiratórias para todas as faixas etárias possui maior correlação com a concentração de MP₁₀ do que com a temperatura e a umidade. Em todas as faixas etárias a correlação entre DR e temperatura possui sinal negativo, indicando uma correlação inversamente proporcional entre estas variáveis, ou seja, o número de pessoas com DR aumenta quando a temperatura diminui (Figuras 4.12 e 4.13).

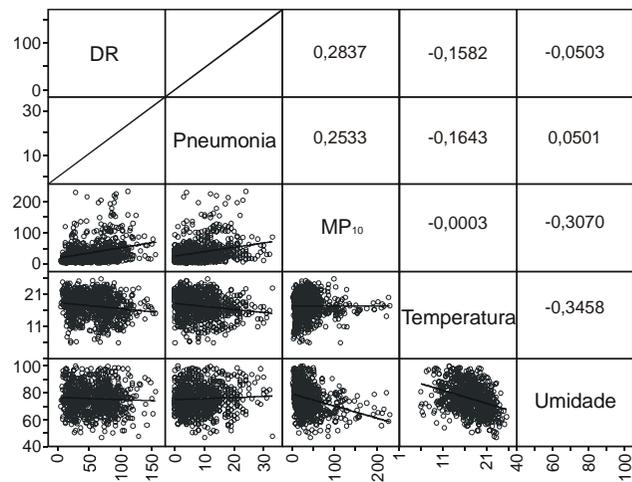
As baixas correlações lineares entre DR para todas as faixas etárias e a umidade relativa do ar apresentadas nas Figuras 4.12 e 4.13 estão de acordo com os resultados observados na Figura 4.9 (relação entre o número de pessoas com DR e umidade).



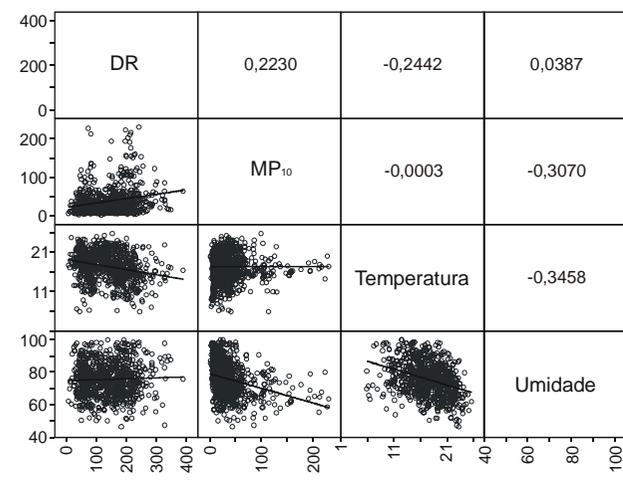
A



C



B



D

Figura 4.12 – Matrizes de dispersão para pessoas menores de 4 anos (A), entre 5 e 14 anos (B), entre 15 e 19 anos (C) e entre 20 e 59 anos (D)

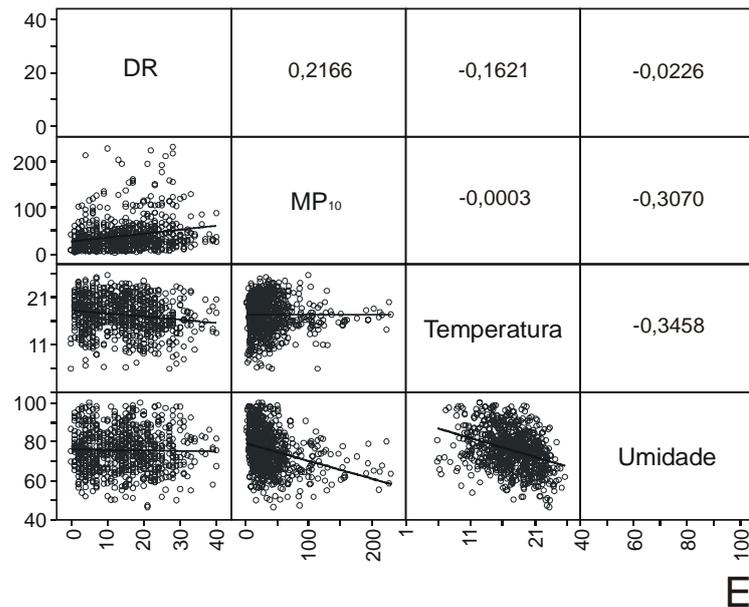


Figura 4.13 – Matriz de dispersão para pessoas maiores de 60 anos (E)

De acordo com os resultados apresentados pelos histogramas (Figura 4.11), pelas estatísticas descritivas (Tabela 4.4) e pelas matrizes de dispersão (Figuras 4.12 e 4.13), verificou-se que o MLG (modelo linear generalizado) com família de Poisson é apropriado para o problema em questão, pois as variáveis respostas possuem uma baixa correlação linear com as variáveis explicativas e os histogramas apresentaram distribuição de Poisson para todas as variáveis respostas.

4.6 Aplicação do Modelo de Regressão de Poisson

Finalmente, com todas as suposições satisfeitas, inicia-se a estimativa do modelo de regressão de Poisson com o auxílio do programa estatístico S-Plus 7.0. O critério de convergência utilizado nas análises foi o critério padrão do S-Plus 7.0 de $\varepsilon = 10^{-7}$ (Schmidt, 2003 e S-Plus 7.0).

Em estudos sobre o impacto da poluição atmosférica na saúde, é necessário levar em consideração a variação dos fatores meteorológicos e da concentração de poluentes durante o ano. Uma forma de suavizar esta tendência chamada sazonalidade é utilizar a função *spline* cúbica natural²⁰, que não possui interpretação

²⁰ *Spline cúbica natural* – Definição em 3.2.2 Ajuste de tendências temporais - Sazonalidade

física, conseqüentemente um modelo com função *spline* não pode ser utilizado para realizar previsões, neste caso utiliza-se uma variável qualitativa para os meses do ano (Conceição et al., 2001).

Uma análise utilizando os diferentes tipos de função de ligação existentes para o modelo de regressão de Poisson foi realizada no intuito de verificar qual função de ligação se ajusta melhor aos dados.

No intuito de avaliar o impacto da concentração de MP_{10} na saúde populacional do município de Araucária, serão realizadas primeiramente análises utilizando a função *spline* cúbica natural para suavização da sazonalidade; em seguida, a fim de prever o impacto da poluição atmosférica na saúde serão realizadas análises utilizando uma variável qualitativa para os meses do ano para suavizar a sazonalidade.

Escolha da função de ligação

Para o modelo de regressão de Poisson existem três tipos de funções de ligação, a função identidade ($\eta = \mu$) utilizada para dados com caráter linear, a função raiz quadrada e a função logarítmica (ligação canônica) (S-Plus 7.0).

Como os dados para o problema em questão não possuem boa correlação linear, foram realizadas comparações apenas entre a função de ligação raiz quadrada e logarítmica. Nos dois modelos foi acrescentada uma variável explicativa para dias da semana e outra para feriados, bem como a suavização para sazonalidade utilizando a função *spline* cúbica natural.

Para o modelo com função de ligação raiz quadrada, deseja-se estimar os coeficientes da seguinte equação:

$$\sqrt{DR} = \beta_0 + \beta_1 temp + \beta_2 umid + \beta_3 MP_{10} + \beta_4 ns + \beta_5 dds + \beta_6 fer \quad \text{Eq. 4.1}$$

onde DR = doenças respiratórias; temp = temperatura; umid = umidade; MP_{10} = concentração de MP_{10} , ns = *spline* cúbico natural, dds = dia da semana e fer = feriado.

Para o modelo com função de ligação logarítmica tem-se:

$$\log(DR) = \beta_0 + \beta_1 temp + \beta_2 umid + \beta_3 MP_{10} + \beta_4 ns + \beta_5 dds + \beta_6 fer \quad \text{Eq. 4.2}$$

Conforme apresentado na metodologia, existem alguns métodos estatísticos utilizados para avaliar e comparar o ajuste de modelos de regressão. Entre estes métodos, tem-se o pseudo R^2 e a razão entre a estatística de Pearson (χ^2) e graus de liberdade que devem possuir valores aproximadamente iguais a um (Ribeiro, 2006 e Russo, 2002). Dentre as análises gráficas capazes de avaliar o ajuste de um modelo de regressão tem-se o gráfico dos desvios residuais em relação aos valores ajustados pelo modelo (em que a maioria dos pontos devem estar entre -2 e 2), bem como o gráfico dos valores observados em relação aos valores ajustados (onde os valores ajustados devem estar próximos dos valores observados).

Para o problema em questão, o método de ajuste mais importante é a razão entre a estatística de Pearson e graus de liberdade, pois leva em consideração a função de variância estimada pelo modelo de regressão Poisson que está relacionada com o parâmetro de dispersão estimado (Schmidt, 2003).

Os ajustes dos modelos referentes às Equações 4.1 e 4.2 apresentados na Tabela 4.6 mostram que os valores de R^2 e χ^2/gl não estiveram próximos de um para ambos os modelos, porém o modelo com ligação logarítmica se ajustou melhor.

Tabela 4.6 – Pseudo R^2 , estatística de Pearson (χ^2) e razão entre χ^2 e graus de liberdade para modelos com funções de ligação diferentes

Função de ligação	R^2	χ^2	χ^2/gl^*
Logarítmica	0,884	1.348,299	1,902
Raiz quadrada	0,866	1.390,521	1,961

Nota: *gl – graus de liberdade do resíduo (número de observações menos número de parâmetros do modelo)

Nos gráficos dos desvios residuais (Figura 4.14), existem diversos pontos fora do intervalo entre -2 e 2 para ambos os modelos, porém para o modelo com ligação raiz quadrada os pontos estão mais dispersos, com valores próximos de -10.

Os valores ajustados estão próximos dos valores observados para ambos os modelos como apresenta a Figura 4.14, sendo que o modelo para ligação raiz quadrada possui mais pontos afastados que o modelo com ligação logarítmica.

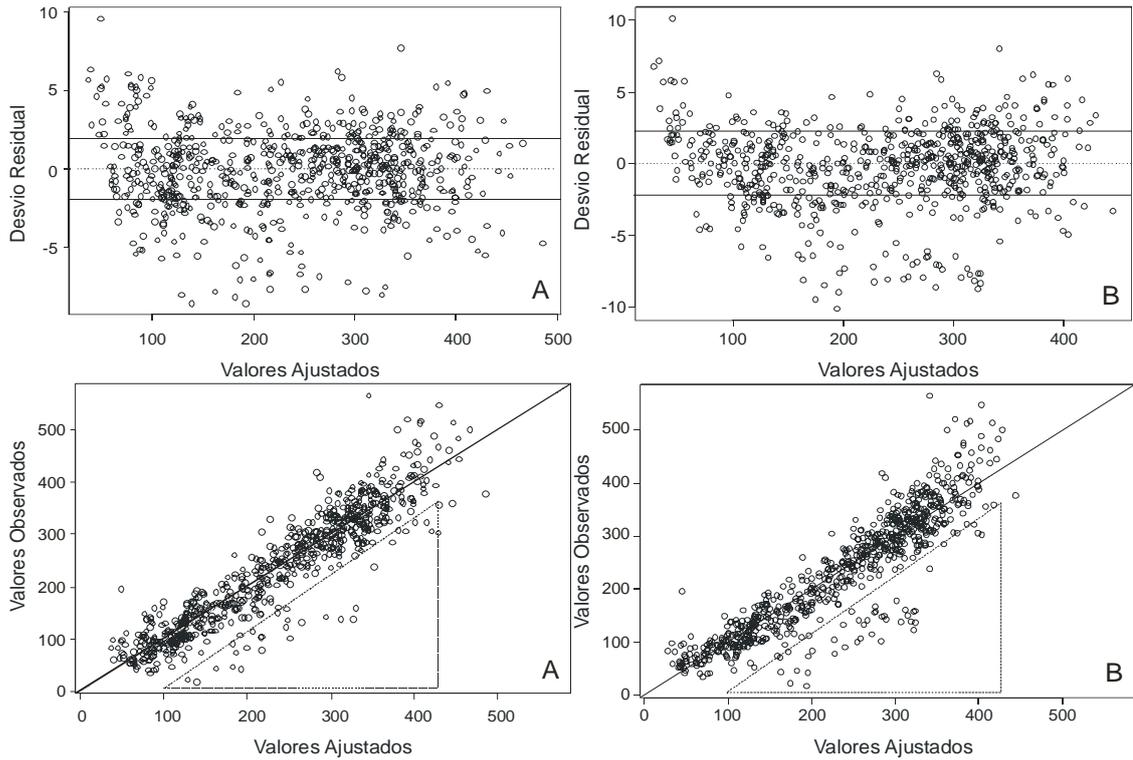


Figura 4.14 – Gráficos de desvios residuais e valores observados em função dos valores ajustados para o modelo com ligação logarítmica (A) e o modelo com ligação raiz quadrada (B)

As análises de ajuste dos modelos mostraram uma pequena diferença entre as diferentes funções de ligação, porém optou-se por utilizar a função de ligação logarítmica, pois além de ter se ajustado razoavelmente melhor, conforme Myers e Montgomery (2002), utilizar a função de ligação canônica (neste caso, ligação logarítmica) implica em simplificações nas estimativas dos parâmetros de regressão.

A análise considerando todas as pessoas com doenças respiratórias não se ajustou de forma satisfatória aos dados como mostra a Tabela 4.6, portanto, foram realizadas análises separadas por faixa etária, no intuito de obter melhores resultados, bem como verificar a diferença entre as faixas etárias.

4.7 Análise por Faixa Etária

As faixas etárias consideradas nas análises estão de acordo com a classificação do Ministério da Saúde (2006)²¹ exceto para crianças devido à forma como a Secretaria Municipal da Saúde de Araucária disponibilizou os dados. As faixas etárias consideradas neste estudo foram então crianças menores de 4 anos e entre 5 e 14 anos, pessoas entre 15 e 19 anos (adolescentes), entre 20 e 59 anos (adultos) e maiores de 60 anos (idosos). Foram considerados todos os adultos em uma única faixa etária, devido à baixa incidência de doenças respiratórias nesta faixa etária, como pode ser observado na Figura 4.6 que mostra a porcentagem mensal de pessoas com problemas respiratórios por faixa etária para o período estudado.

4.7.1 Análise para avaliação

Para avaliar o impacto da concentração de MP_{10} no número de pessoas com doenças respiratórias no município de Araucária foram realizadas análises utilizando a função de ligação logarítmica por ter se ajustado melhor aos dados e a suavização da sazonalidade através da função *spline* cúbica natural.

Conforme apresentado na metodologia, as tendências temporais podem provocar autocorrelação entre os dados devido à diferença entre os atendimentos hospitalares nos finais de semana e feriados; mesmo após serem incluídas variáveis explicativas para dias da semana e feriado. Para o problema em questão, as linhas que estiverem fora do intervalo entre -0,07 e 0,07 indicam uma forte autocorrelação entre os dados de um dia com dados de dias anteriores. Uma forma de ajustar autocorrelações é inserir o resíduo no modelo (Ferraz et al., 1999).

A Figura 4.15 mostra os gráficos da função de autocorrelação parcial em relação aos dias de defasagem para todas as faixas etárias consideradas. Pode-se verificar que todos os modelos, exceto para pessoas maiores de 60 anos, apresentaram resíduo para um dia de defasagem, sendo que para alguns, observou-se a presença de resíduo para mais dias de defasagem.

²¹ Considera-se criança, a pessoa até doze anos de idade incompletos; adolescente aquela entre doze e dezoito anos de idade e idosos, as pessoas com idade igual ou superior a 60 anos (Ministério da Saúde, 2006a e b).

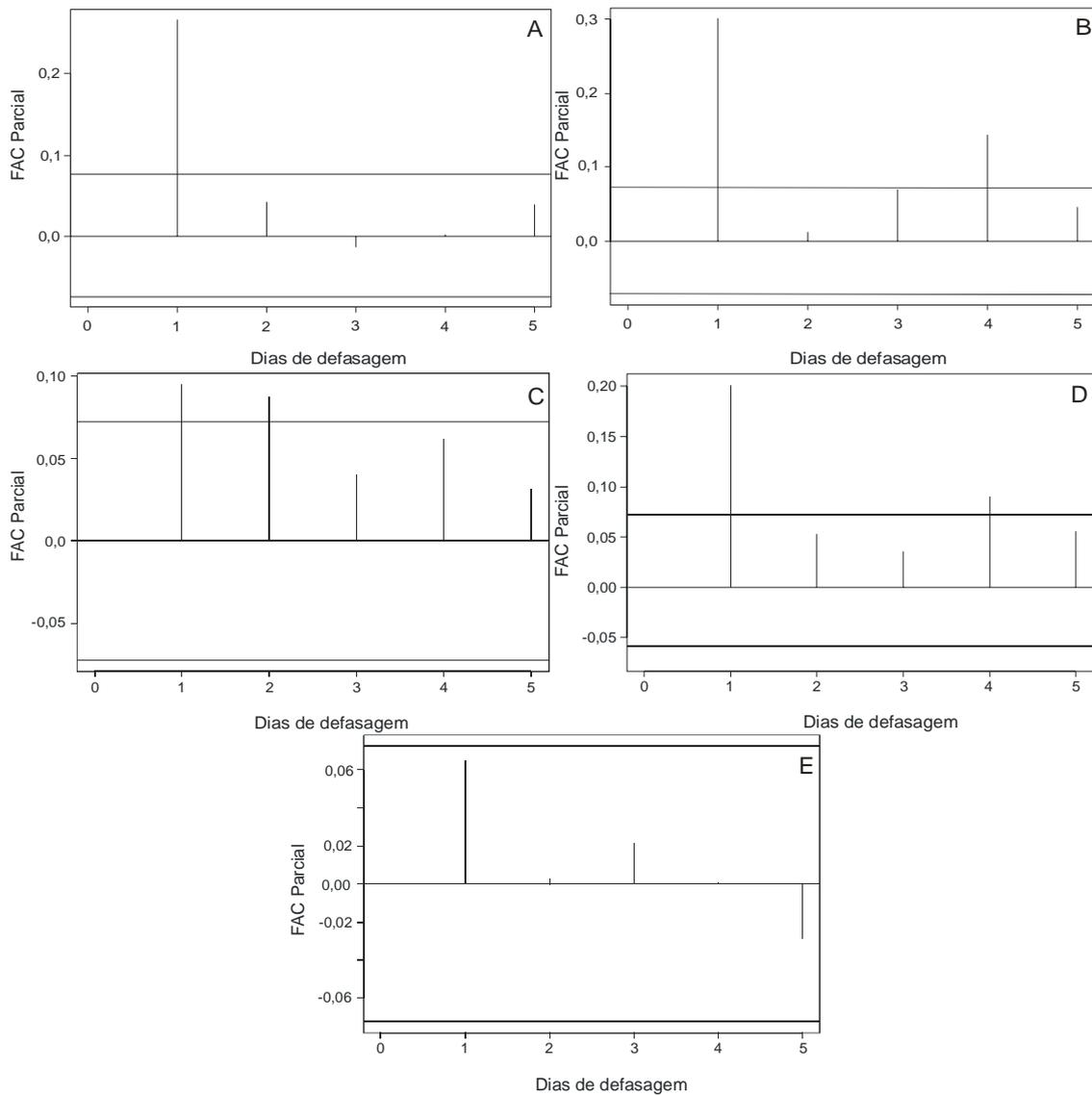


Figura 4.15 – Função de autocorrelação parcial em relação aos dias de defasagem nos modelos sem resíduo para pessoas menores de 4 anos (A), entre 5 e 14 anos (B), entre 15 e 19 anos (C), entre 20 e 59 anos (D) e maiores de 60 anos (E)

Os modelos com acréscimo de resíduo para todas as faixas etárias não apresentaram autocorrelação, como mostra a Figura 4.16. Cabe lembrar que a ausência de autocorrelações não implica em um ajuste satisfatório do modelo, sendo necessário comparar os modelos com e sem acréscimo de resíduo para verificar qual obteve melhor ajuste.

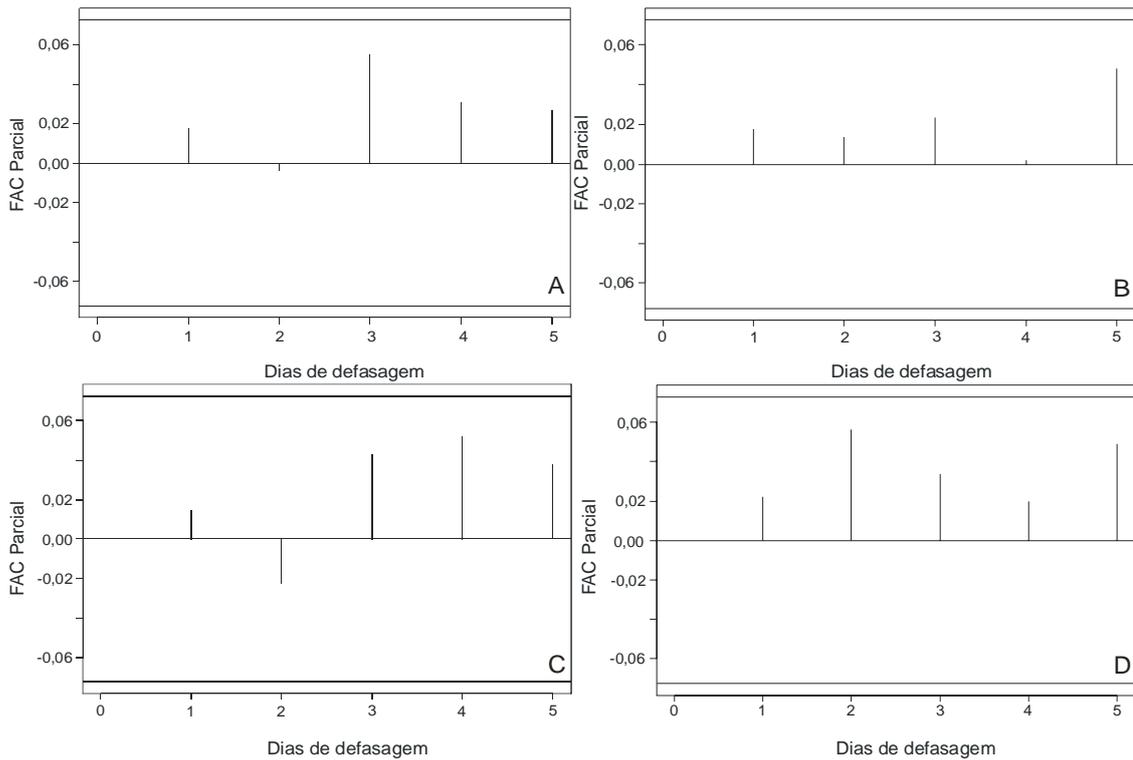


Figura 4.16 - Função de autocorrelação parcial em relação aos dias de defasagem nos modelos com resíduo para pessoas menores de 4 anos (A), entre 5 e 14 anos (B), entre 15 e 19 anos (C), entre 20 e 59 anos (D)

Através das estatísticas pseudo R^2 e χ^2/gI apresentadas na Tabela 4.7, verifica-se que os modelos com acréscimo de resíduo mostraram melhor ajuste que os modelos sem acréscimo de resíduo para todas as faixas etárias, exceto para pessoas maiores de 60 anos em que não foi preciso acrescentar o resíduo, pois não foram observadas autocorrelações entre os dados de um dia com dados de dias anteriores.

Comparando os ajustes para as diferentes faixas etárias, apresentados na Tabela 4.7, verifica-se que o modelo para pessoas maiores de 60 anos apresentou melhor ajuste, com χ^2/gI aproximadamente igual a 1.

Tabela 4.7 – Pseudo R², estatística de Pearson (χ^2) e razão entre χ^2 e graus de liberdade dos modelos com e sem acréscimo de resíduo por faixa etária

Faixa etária	Resíduo	R ²	χ^2	χ^2/gf
Menores de 4 anos	não	0,836	997,645	1,407
	sim	0,844	977,095	1,378
Entre 5 e 14 anos	não	0,828	960,516	1,355
	sim	0,846	890,428	1,256
Entre 15 e 19 anos	não	0,692	610,155	0,861
	sim	0,701	598,118	0,843
Entre 20 e 59 anos	não	0,839	1.290,313	1,820
	sim	0,846	1.238,813	1,747
Maiores de 60 anos	não	0,745	679,214	0,955
	sim	-	-	-

Observando os gráficos dos desvios residuais para as diferentes faixas etárias, apresentados nas Figuras 4.17 e 4.18, verifica-se que os modelos para pessoas entre 15 e 19 anos e para maiores de 60 anos se ajustaram melhor.

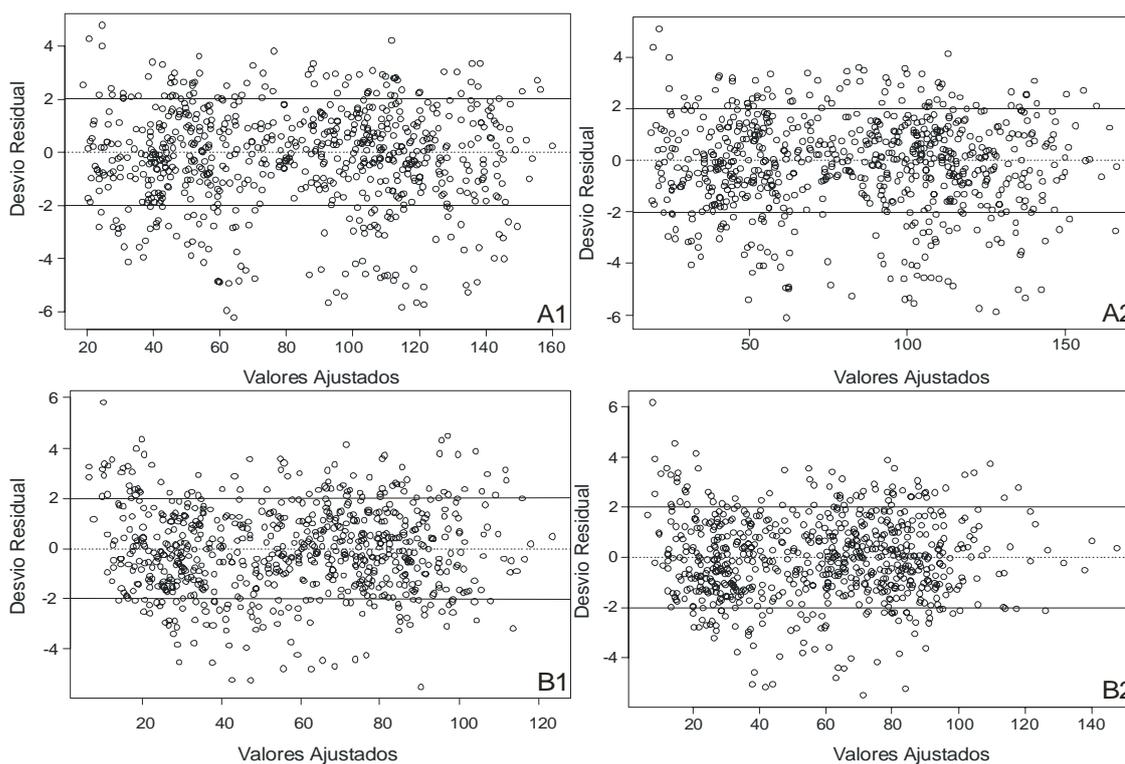


Figura 4.17 – Gráficos dos desvios residuais em relação aos valores ajustados para pessoas menores de 4 anos (A) e entre 5 e 14 anos (B), para modelo sem resíduo (1) e com resíduo (2)

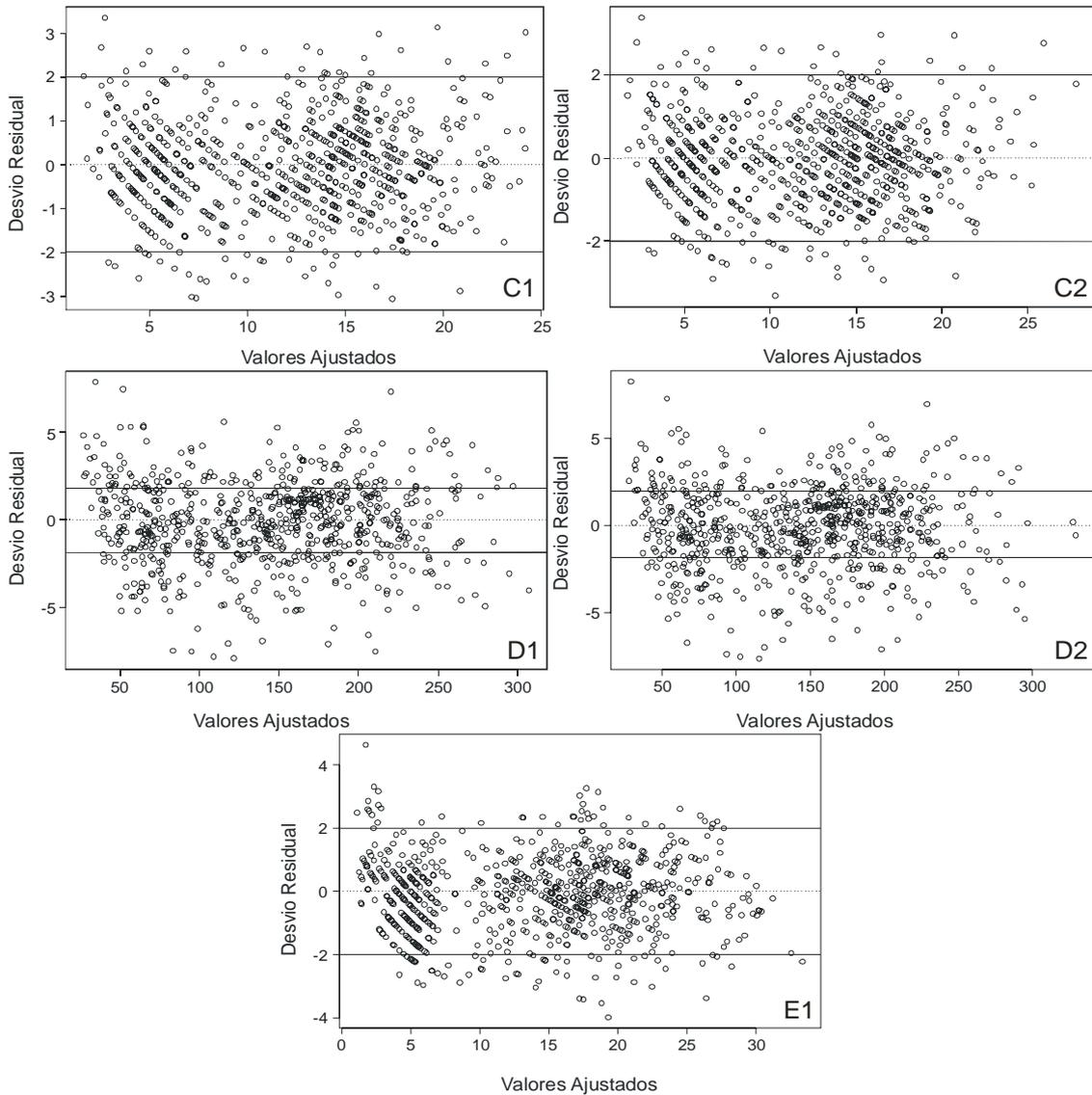


Figura 4.18 – Gráficos dos desvios residuais em relação aos valores ajustados para pessoas entre 15 e 19 anos (C), entre 20 e 59 anos (D) e maiores de 60 anos (E) para modelo sem resíduo (1) e com resíduo (2)

Os gráficos dos valores observados em relação aos valores ajustados apresentados nas Figuras 4.19 e 4.20 mostram que os valores ajustados pelo modelo estiveram próximos dos valores observados para todas as faixas etárias.

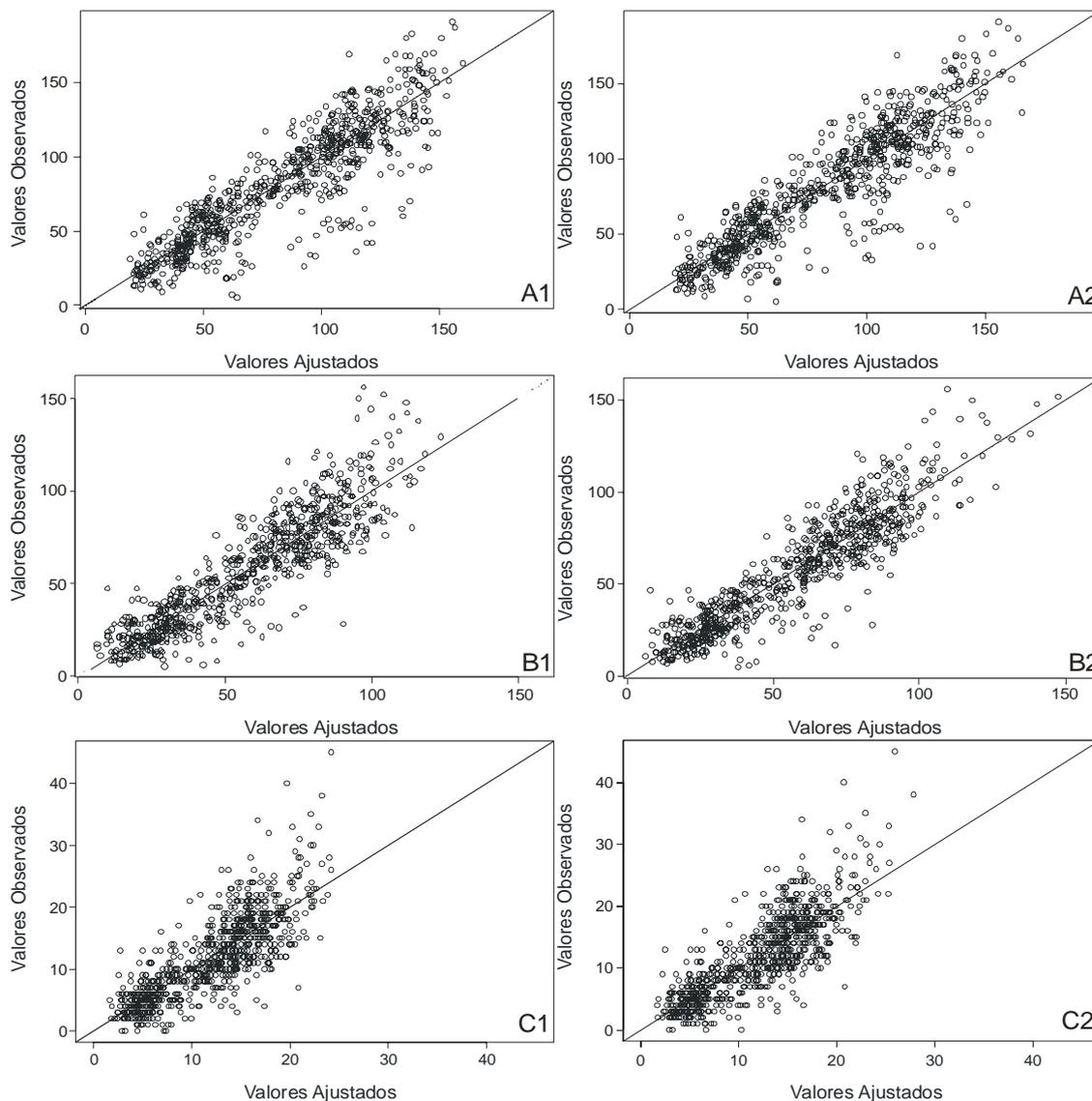


Figura 4.19 - Gráficos dos valores observados em relação aos valores ajustados para pessoas menores de 4 anos (A), entre 5 e 14 anos (B) e entre 15 e 19 anos (C), para modelo sem correção (1) e com correção (2)

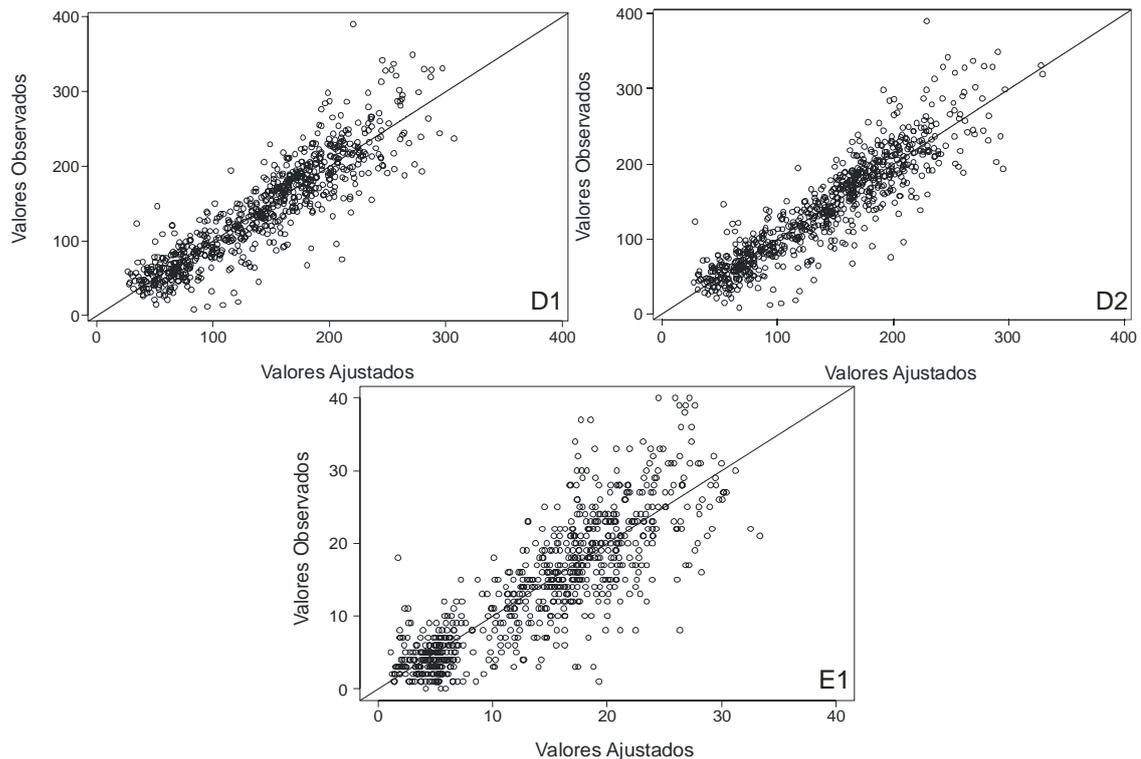


Figura 4.20 - Gráficos dos valores observados em relação aos valores ajustados para pessoas entre 20 e 59 anos (D) e maiores de 60 anos (E) para modelo sem correção (1) e com correção (2)

Através de todas as análises de ajuste dos modelos apresentadas acima, concluiu-se que o modelo para pessoas maiores de 60 anos se ajustou melhor.

Após avaliar o ajuste de cada modelo, cabe verificar se os coeficientes de regressão estimados são estatisticamente significativos, ou seja, verificar se as variáveis explicativas contribuíram de forma significativa para o aumento de pessoas com DR no município de Araucária.

Os resultados dos modelos com acréscimo de resíduo para as diferentes faixas etárias estão relacionados na Tabela 4.8, na qual se encontram os coeficientes de regressão estimados, os erros padrões e os valores de t para concentração de MP_{10} , temperatura e umidade. Pode se observar que a umidade relativa do ar tem um coeficiente de regressão negativo, indicando uma relação inversamente proporcional entre a umidade e o número de pessoas com problemas respiratórios, ou seja, o número de pessoas com doenças respiratórias aumenta quando a umidade diminui. Este fator está diretamente ligado à alta solubilidade do material particulado (Seinfeld, 1986).

Para verificar se os coeficientes de regressão estimados são estatisticamente significativos é necessário comparar os valores de t encontrados com o valor de t tabelado. Para o problema em questão, o nível de significância (α) considerado foi de 5%, o número de observações foi igual a 731 e foram consideradas 19 ou 21 variáveis explicativas no modelo (9 ou 11 equivalentes à quantidade de nós considerados na função *spline* cúbica natural para cada modelo e 6 à variável dia da semana, por serem variáveis qualitativas²²). O grau de liberdade foi de 711 ou 709 (=731-20 ou =731-22). Conforme a Tabela B.1 do Anexo B, para problemas com grau de liberdade maior que 120, o valor de t é 1,96. Isto indica que um coeficiente de regressão estatisticamente significativo deve possuir um valor de t maior que 1,96.

Apesar do modelo para pessoas maiores de 60 anos ter se ajustado melhor, os coeficientes de regressão apresentados na Tabela 4.8 não se mostraram estatisticamente significativos nesta faixa etária, indicando que não existe uma forte relação entre a concentração de MP_{10} , temperatura e umidade e a incidência de doenças respiratórias em pessoas maiores de 60 anos. Este resultado pode ter sido ocasionado por erros na coleta dos dados experimentais de concentração de MP_{10} , por ter sido considerado todos os tipos de doenças respiratórias em um único modelo ou porque os dados de doenças respiratórias são referentes aos atendimentos hospitalares apenas de hospitais públicos, não sendo considerados os atendimentos em hospitais e consultórios particulares.

De acordo com os resultados da Tabela 4.8, verifica-se que os coeficientes de regressão dos modelos para pessoas entre 15 e 19 anos e entre 20 e 59 anos também não são estatisticamente significativos. Para crianças menores de 4 anos, apenas o coeficiente para temperatura não é estatisticamente significativo e para crianças entre 5 e 14 anos, todos os coeficientes são estatisticamente significativos.

²² Variáveis qualitativas – Se referem a uma característica e não a uma quantidade, sendo que uma das características é considerada como característica de referência e as demais são contadas como diferentes variáveis explicativas.

Tabela 4.8 – Coeficientes de regressão, erro padrão e valores de t para cada faixa etária para modelo com resíduo**

Faixa etária	Variável	Coeficiente	Erro padrão	Valor de t
Menores de 4 anos	Temperatura	0,0048	0,0028	1,7056
	Umidade	-0,0043	0,0009	-4,9506*
	MP ₁₀	0,0005	0,0002	2,1443*
Entre 5 e 14 anos	Temperatura	0,0084	0,0032	2,6346*
	Umidade	-0,0029	0,0010	-3,0048*
	MP ₁₀	0,0010	0,0002	3,8699*
Entre 15 e 19 anos	Temperatura	0,0076	0,0046	1,6480
	Umidade	-0,0014	0,0014	-0,9991
	MP ₁₀	0,0001	0,0004	0,2090
Entre 20 e 59 anos	Temperatura	0,0031	0,0028	1,1046
	Umidade	-0,0011	0,0008	-1,3257
	MP ₁₀	0,0004	0,0002	1,8780
Maiores de 60 anos	Temperatura	0,0057	0,0049	1,1736
	Umidade	-0,0028	0,0015	-1,9334
	MP ₁₀	0,0005	0,0004	1,4483

Nota: * Coeficientes estatisticamente significativos

** Os resultados para pessoas maiores de 60 anos é referente ao modelo sem acréscimo de resíduo.

Devido à grande incidência de pneumonia e gripe em crianças foram realizadas análises considerando como variável resposta o número de atendimentos hospitalares por pneumonia e gripe de acordo com a CID-10 em crianças menores de 4 anos e entre 5 e 14 anos.

Como nas análises anteriores (considerando todas as doenças respiratórias), os modelos com acréscimo de resíduo mostraram melhores resultados, para pneumonia e gripe em crianças foram considerados então apenas os resultados dos modelos com acréscimo de resíduo.

Através dos resultados do ajuste dos modelos com acréscimo de resíduo apresentados na Tabela 4.9, verifica-se que a razão χ^2/gl esteve próxima de 1 para ambas faixas etárias, sendo que o modelo para crianças entre 5 e 14 anos se ajustou melhor.

Tabela 4.9 – Pseudo R², estatística de Pearson (χ^2) e razão entre χ^2 e graus de liberdade para modelo com resíduo em crianças com pneumonia e gripe

Faixa etária	R ²	χ^2	χ^2/gl
Menores de 4 anos	0,635	836,558	1,180
Entre 5 e 14 anos	0,650	727,696	1,026

Os gráficos dos desvios residuais (Figura 4.21) não mostraram uma diferença visível entre os modelos para as diferentes faixas etárias, sendo que os dois modelos são satisfatórios, ou seja, a maioria dos pontos para os dois modelos está dentro do intervalo entre -2 e 2.

Os valores ajustados estiveram próximos dos valores observados para as duas faixas etárias consideradas, como mostra a Figura 4.21.

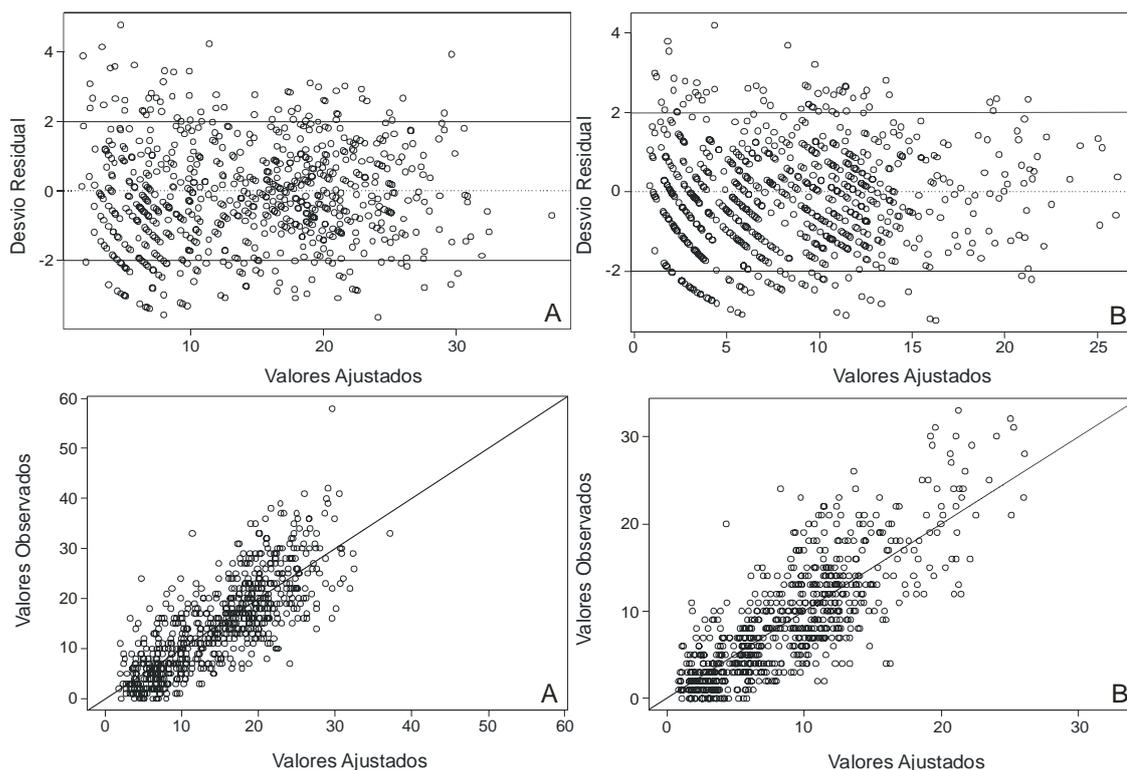


Figura 4.21 – Gráficos dos desvios residuais e dos valores observados em relação aos valores ajustados para pneumonia e gripe em crianças menores de 4 anos (A) e entre 5 e 14 anos (B) para modelo com resíduo

A partir dos resultados das análises para crianças, verificou-se que o modelo se ajustou melhor para pneumonia e gripe ($\chi^2/\text{gl} = 1,180$ para crianças menores de 4 anos e $\chi^2/\text{gl} = 1,026$ para crianças entre 5 e 14 anos) do que para todas as doenças respiratórias ($\chi^2/\text{gl} = 1,378$ para crianças menores de 4 anos e $\chi^2/\text{gl} = 1,256$ para crianças entre 5 e 14 anos). Esta diferença ocorreu possivelmente porque no modelo para todas as doenças respiratórias, estão sendo considerados alguns tipos de doenças respiratórias que não estão relacionadas com a poluição atmosférica.

Os coeficientes de regressão, os erros padrões e os valores de t para os modelos considerando apenas pneumonia e gripe estão na Tabela 4.10. No modelo para crianças menores de 4 anos, apenas o coeficiente de regressão para umidade é estatisticamente significativo e para crianças entre 5 e 14 anos apenas o coeficiente de regressão para umidade não é estatisticamente significativo.

Tabela 4.10 – Coeficientes de regressão, erro padrão e valores de t para pneumonia e gripe em crianças para modelo com resíduo

Faixa etária		Coeficiente	Erro padrão	Valor de t
Menores de 4 anos	Temperatura	0,0075	0,0058	1,2879
	Umidade	-0,0043	0,0018	-2,4188*
	MP ₁₀	0,0002	0,0005	0,4155
Entre 5 e 14 anos	Temperatura	0,0194	0,0066	2,9427*
	Umidade	-0,0007	0,0020	-0,3313
	MP ₁₀	0,0015	0,0005	2,9405*

Nota: * Coeficientes estatisticamente significativos

Através de todas as análises estatísticas realizadas acima, concluiu-se que o coeficiente de regressão para a concentração de MP₁₀ se mostrou estatisticamente significativo apenas para doenças respiratórias nas crianças e para pneumonia e gripe em crianças entre 5 e 14 anos, possivelmente devido à falta de fidedignidade dos dados de concentração de MP₁₀ disponíveis.

Como foram consideradas apenas crianças para os modelos com pneumonia e gripe, os resultados das análises para todas as doenças respiratórias nas crianças também foram utilizados na construção dos gráficos dos riscos relativos.

Risco relativo

Através dos gráficos de risco relativo é possível comparar o impacto do MP₁₀ no número de pessoas com doenças respiratórias nas diferentes faixas etárias. Na maioria dos casos, o risco relativo não aumentou de forma significativa para um aumento na concentração de MP₁₀ como mostram os gráficos da Figura 4.22 em que o risco de uma pessoa exposta a uma concentração de MP₁₀ de aproximadamente 220 µg/m³ adquirir uma DR foi aproximadamente 10% maior que de uma pessoa exposta a uma concentração de aproximadamente 10 µg/m³, exceto para crianças entre 5 e 14 anos, em que o risco relativo para uma pessoa exposta a

uma concentração de aproximadamente 220 $\mu\text{g}/\text{m}^3$ foi aproximadamente 30 % maior que de uma pessoa exposta a uma concentração de aproximadamente 10 $\mu\text{g}/\text{m}^3$.

As crianças (pessoas menores de 4 anos e entre 5 e 14 anos) e os idosos (pessoas maiores de 60 anos) foram mais afetadas pela poluição atmosférica no período estudado (Figura 4.22).

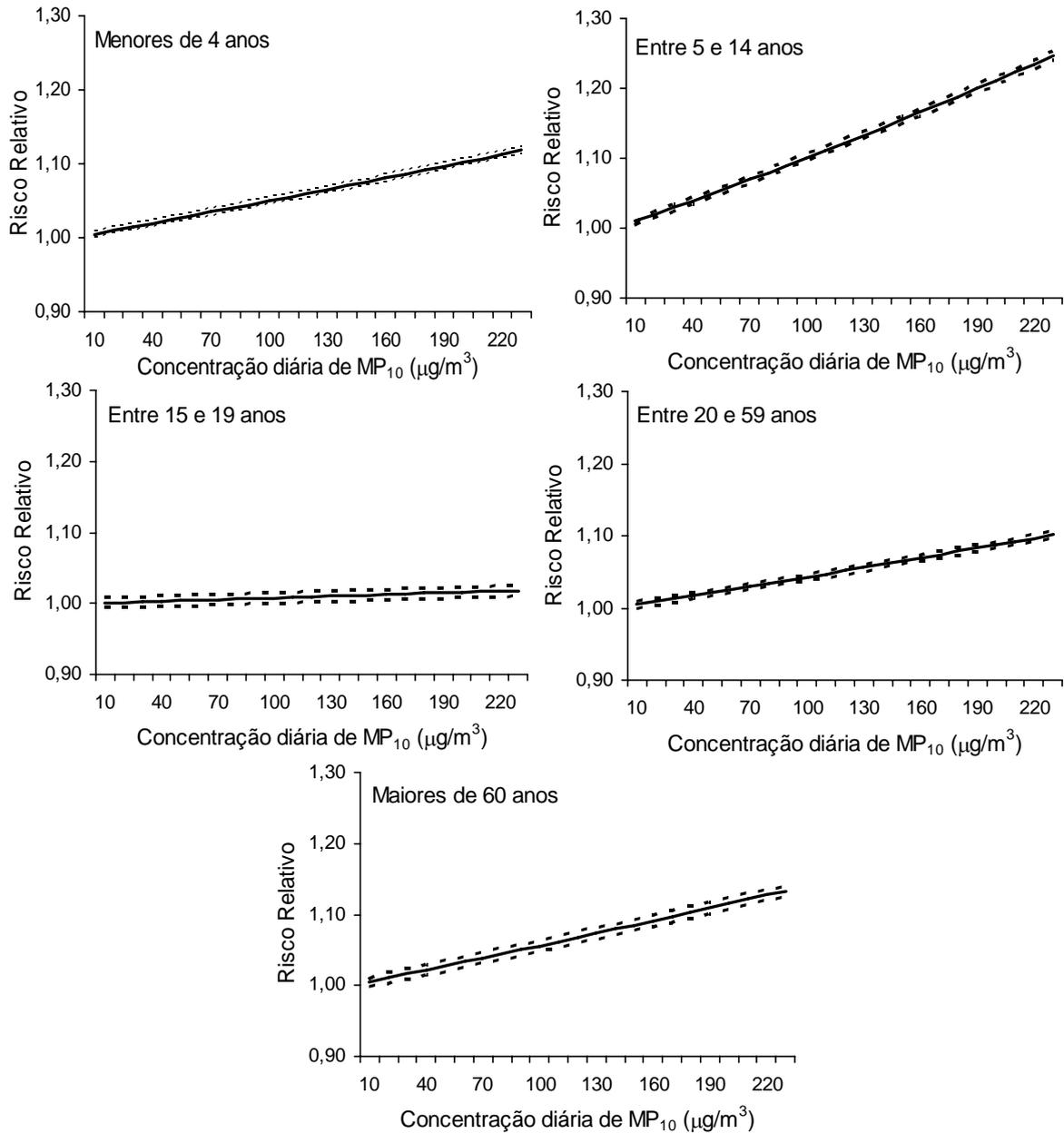


Figura 4.22 – Risco relativo com intervalo de confiança de 95% de doenças respiratórias para cada faixa etária associado ao aumento na concentração de MP₁₀

Os atendimentos hospitalares efetuados em um determinado dia podem estar relacionados tanto à poluição do dia quanto de dias anteriores, portanto avaliou-se o risco relativo entre os casos de doenças respiratórias de um dia com a poluição atmosférica de dias anteriores (Braga, 1998 apud Bakonyi, 2003).

O resultado encontrado na análise do risco relativo de doenças respiratórias em relação à concentração de MP_{10} de dias anteriores não esteve de acordo com o encontrado em outros estudos (em que o risco relativo aumenta quando relacionado com a poluição de dias anteriores) conforme mostra a Figura 4.23. Para crianças entre 5 e 14 anos e pessoas entre 20 e 59 anos, o risco relativo relacionado com a concentração de MP_{10} de quatro dias anteriores diminuiu aproximadamente 5% em relação à concentração do mesmo dia, já para pessoas maiores de 60 anos, este valor foi de aproximadamente 10%.

Não foi possível verificar uma alteração no risco relativo para crianças menores de 4 anos. Já para as pessoas entre 15 e 19 anos, o risco relativo relacionado com a concentração de MP_{10} de dois e quatro dias anteriores esteve aproximadamente igual ao risco relativo para a concentração do mesmo dia e o risco relativo relacionado com a concentração de um e três dias anteriores aumentou 5% em relação à concentração do mesmo dia.

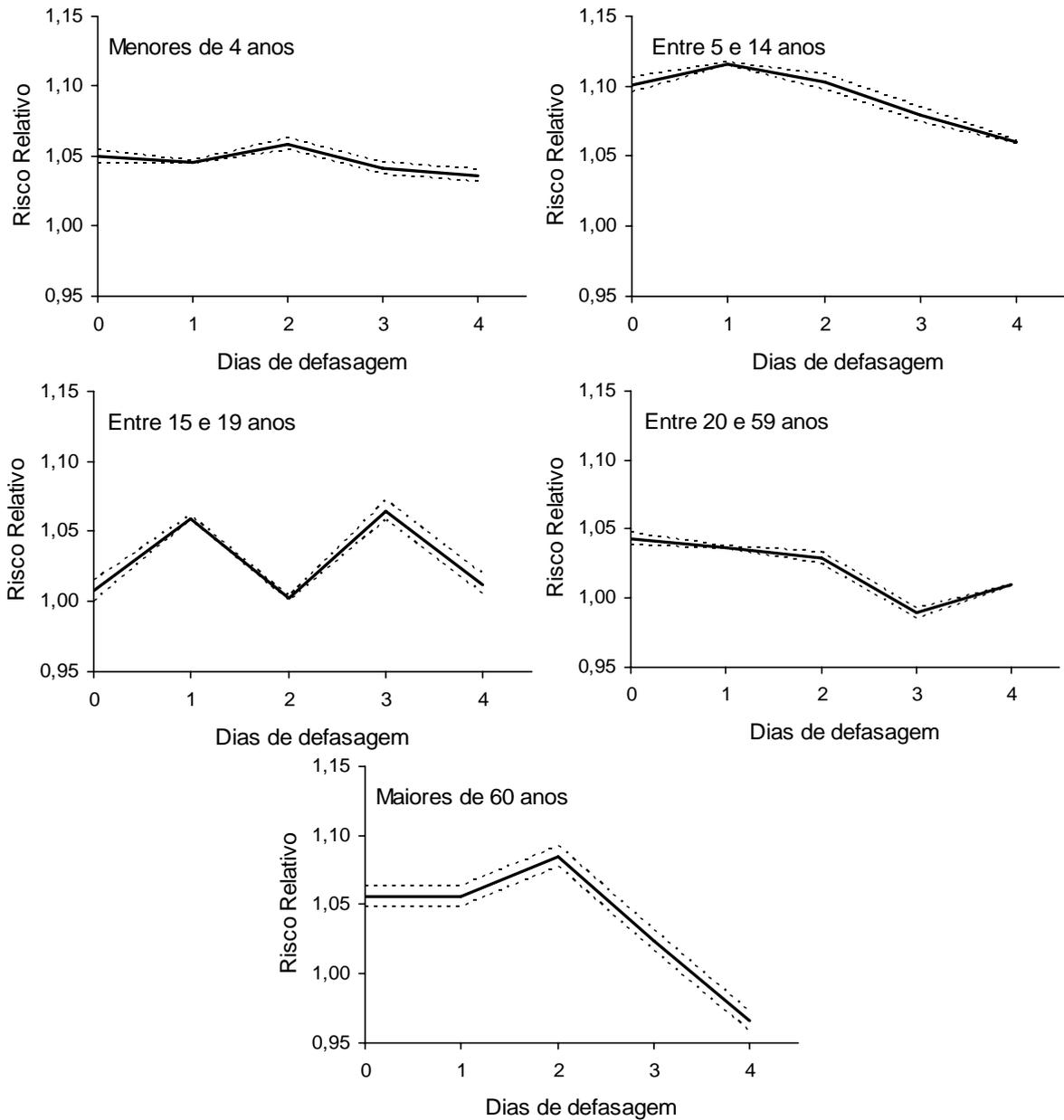


Figura 4.23 - Risco relativo com intervalo de confiança de 95% de doenças respiratórias para cada faixa etária associado a um aumento de $100 \mu\text{g}/\text{m}^3$ na concentração de MP_{10}

O risco relativo de pneumonia e gripe nas crianças não mostrou uma alteração razoável de acordo com a concentração de MP_{10} de dias anteriores como mostra a Figura 4.24. A diferença entre o risco de pneumonia e gripe em crianças relacionado com a concentração de MP_{10} de dias anteriores e a concentração do mesmo dia alcançou valores abaixo de 5%.

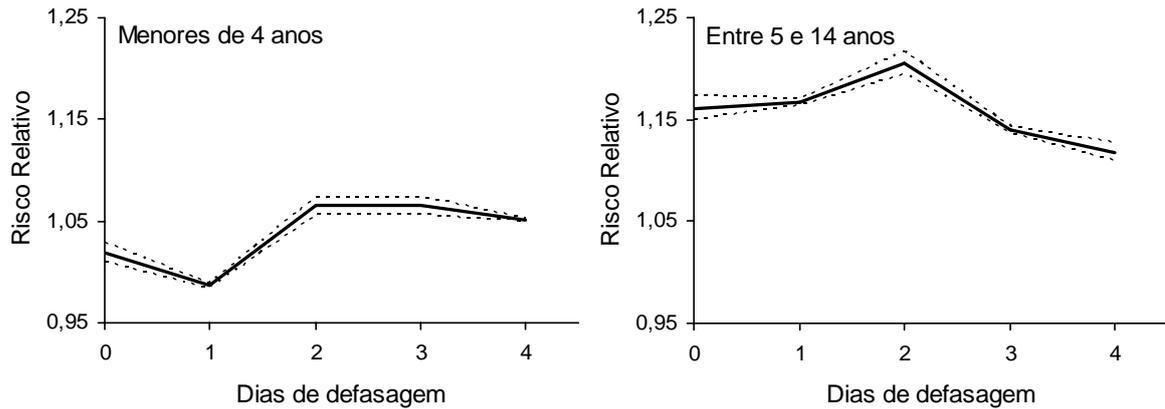


Figura 4.24 – Risco relativo com intervalo de confiança de 95% para pneumonia e gripe em crianças associado a um aumento de $100 \mu\text{g}/\text{m}^3$ na concentração de MP_{10} .

A incoerência nos resultados para o risco relativo de doenças respiratórias e pneumonia e gripe podem ter sido ocasionados pela falta de qualidade nos dados de concentração de MP_{10} .

4.7.2 Análise para previsão

Conforme apresentado na metodologia, os modelos com acréscimo de resíduo e que utilizam a função *spline* cúbica natural para suavização da sazonalidade não podem ser utilizados para realizar previsões.

Devido ao objetivo principal do trabalho de prever o impacto da poluição atmosférica na saúde populacional, foram realizadas, então, análises utilizando uma variável para os meses do ano a fim de suavizar a sazonalidade e não foi acrescentado o resíduo nestes modelos.

Os coeficientes de regressão que serão estimados para o modelo com variável mês são apresentados na Equação 4.3.

$$\log(DR) = \beta_0 + \beta_1 temp + \beta_2 umid + \beta_3 MP_{10} + \beta_4 mes + \beta_5 dds + \beta_6 fer \quad \text{Eq. 4.3}$$

Comparando os ajustes para as diferentes faixas etárias, presentes na Tabela 4.11, verifica-se que o modelo para pessoas maiores de 60 anos apresentou melhor ajuste, com χ^2/gI aproximadamente igual a 1.

Tabela 4.11 – Pseudo R^2 , estatística de Pearson (χ^2) e razão entre χ^2 e graus de liberdade dos modelos com variável mês sem resíduo por faixa etária

Faixa etária	R^2	χ^2	χ^2 / gl
Menores de 4 anos	0,821	1.048,901	1,479
Entre 5 e 14 anos	0,810	1.010,137	1,425
Entre 15 e 19 anos	0,667	632,678	0,892
Entre 20 e 59 anos	0,795	1.446,149	2,040
Maiores de 60 anos	0,735	692,561	0,977

Observando os gráficos dos desvios residuais para as diferentes faixas etárias, apresentados nas Figuras 4.25 e 4.26 verifica-se que o modelo para pessoas entre 15 e 19 anos e maiores de 60 anos se ajustaram melhor.

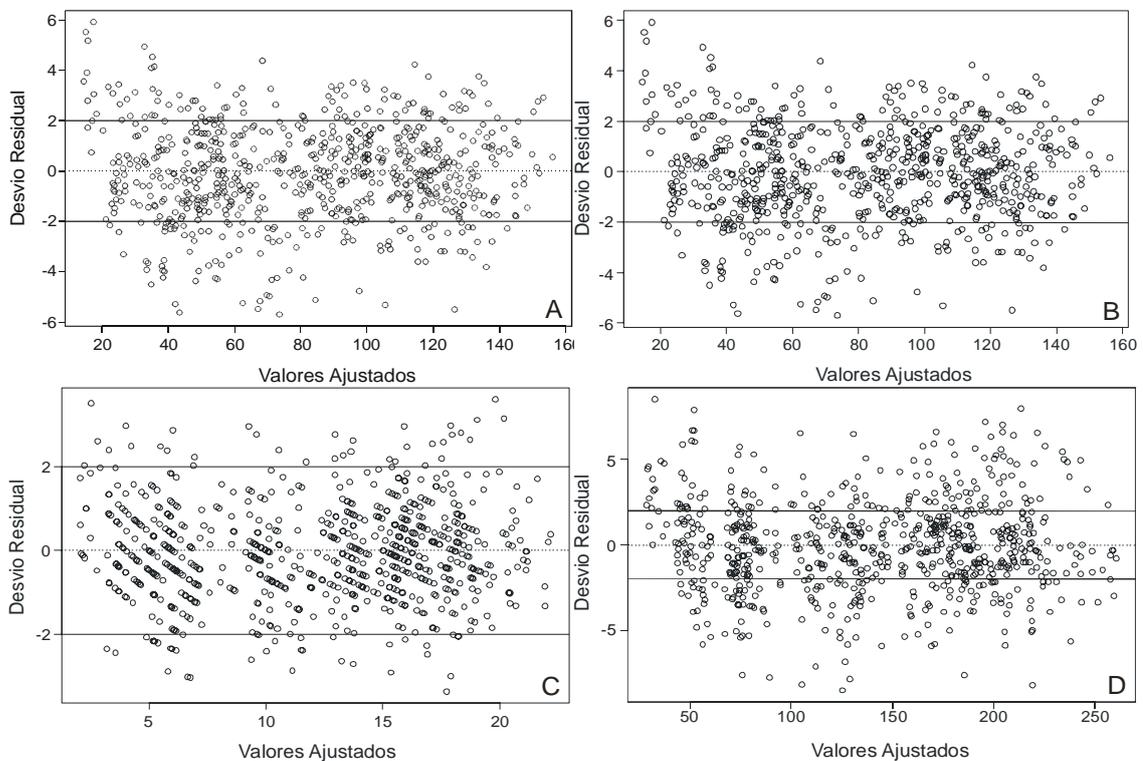


Figura 4.25 – Gráficos dos desvios residuais em relação aos valores ajustados para pessoa menores de 4 anos (A), entre 5 e 14 anos (B), entre 15 e 19 anos (C) e entre 20 e 59 anos (D)

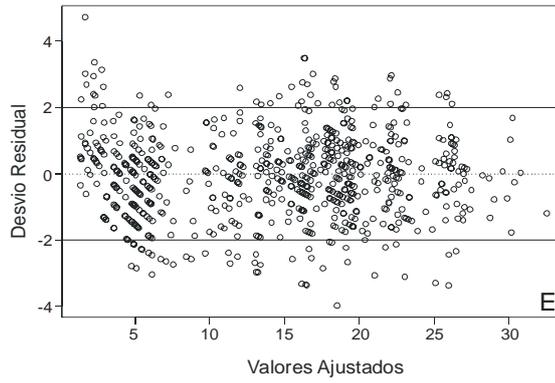


Figura 4.26 – Gráfico dos desvios residuais em relação aos valores ajustados para pessoas maiores de 60 anos (E)

Os gráficos dos valores observados em relação aos valores ajustados apresentados nas Figuras 4.27 e 4.28 mostram que os valores ajustados estão próximos dos valores observados para todas as faixas etárias.

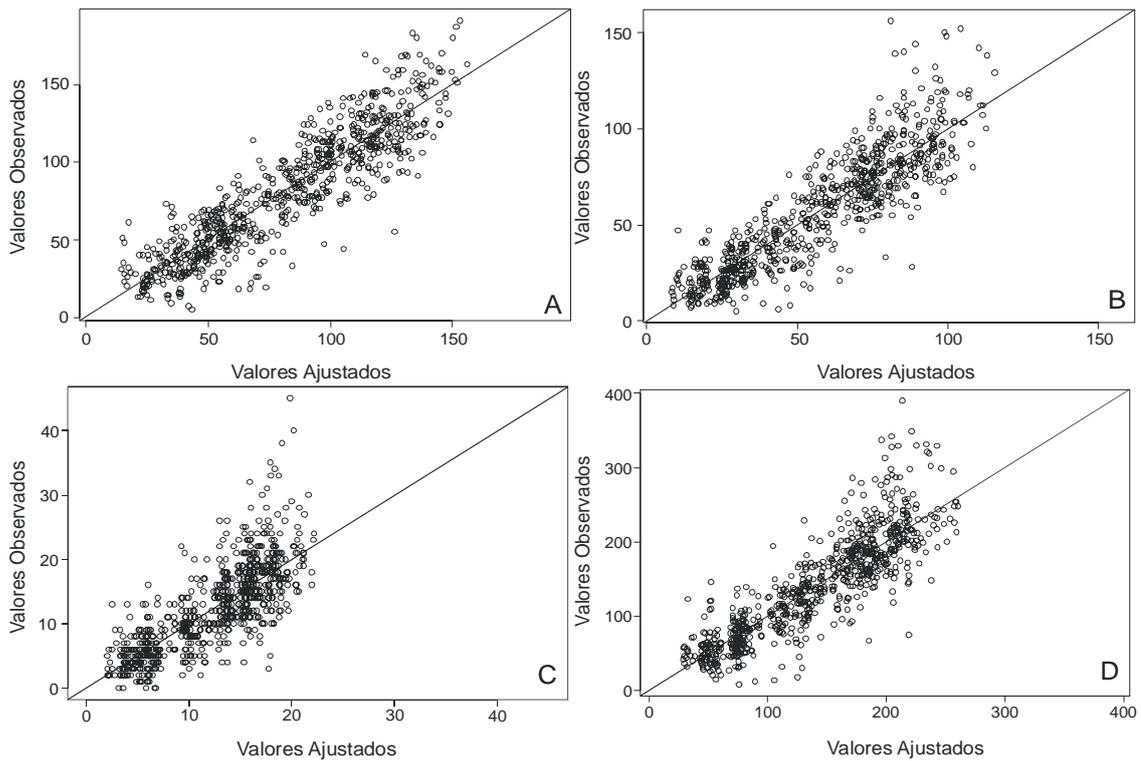


Figura 4.27 - Gráficos dos valores observados em relação aos valores ajustados para pessoas menores de 4 anos (A), entre 5 e 14 anos (B), entre 15 e 19 anos (C) e entre 20 e 59 anos (D)

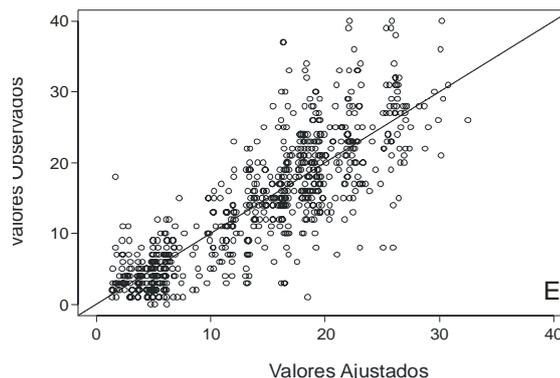


Figura 4.28 - Gráfico dos valores observados em relação aos valores ajustados para pessoas maiores de 60 anos (E)

Através de todas as análises de ajuste dos modelos apresentadas acima, concluiu-se que o modelo para pessoas maiores de 60 anos se ajustou melhor.

Novamente, após avaliar o ajuste de cada modelo, cabe verificar se os coeficientes de regressão estimados são estatisticamente significativos.

Os resultados das análises por faixa etária estão relacionados na Tabela 4.12, na qual se encontram os coeficientes de regressão estimados, os erros padrões e os valores de t para concentração de MP_{10} , temperatura e umidade. Os coeficientes de regressão das demais variáveis estão apresentados na Tabela 4.15.

Como foi apresentado na análise para avaliação, para verificar se os coeficientes de regressão estimados são estatisticamente significativos é necessário comparar os valores de t encontrados com o valor de t tabelado. Para o problema em questão, o nível de significância (α) considerado foi de 5%, o número de observações foi igual a 731 e foram consideradas 22 variáveis explicativas no modelo (11 equivalentes à variável mês e 6 à variável dia da semana, por serem variáveis qualitativas). Assim, o grau de liberdade foi de 709 (= 731-22). Como nas análises anteriores, se o valor de t for maior que 1,96; o coeficiente de regressão é considerado estatisticamente significativo.

Apesar do modelo para pessoas maiores de 60 anos ter se ajustado melhor, os coeficientes de regressão apresentados na Tabela 4.12 não se mostraram estatisticamente significativos, indicando que não existe uma forte relação entre a concentração de MP_{10} , temperatura e umidade e a incidência de doenças respiratórias em pessoas maiores de 60 anos.

De acordo com os resultados da Tabela 4.12, verifica-se que os coeficientes de regressão dos modelos para pessoas entre 15 e 19 anos e entre 20 e 59 anos também não são estatisticamente significativos. Para crianças menores de 4 anos apenas o coeficiente para umidade é estatisticamente significativo e para crianças entre 5 e 14 anos, apenas o coeficiente para concentração de MP₁₀ não é estatisticamente significativo.

Tabela 4.12 – Coeficientes de regressão, erro padrão e valores de t para cada faixa etária

Faixa etária	Variável	Coefficiente	Erro padrão	Valor de t
Menores de 4 anos	Temperatura	0,0035	0,0030	1,1852
	Umidade	-0,0047	0,0008	-5,5222*
	MP ₁₀	0,0001	0,0003	0,4496
Entre 5 e 14 anos	Temperatura	0,0083	0,0035	2,3943*
	Umidade	-0,0022	0,0010	-2,2754*
	MP ₁₀	-0,0000	0,0003	-0,0773
Entre 15 e 19 anos	Temperatura	0,0027	0,0049	0,5649
	Umidade	-0,0007	0,0014	-0,4981
	MP ₁₀	-0,0007	0,0004	-1,7741
Entre 20 e 59 anos	Temperatura	-0,0023	0,0032	-0,7233
	Umidade	0,0017	0,0009	1,8108
	MP ₁₀	0,0003	0,0003	1,1238
Maiores de 60 anos	Temperatura	0,0011	0,0049	0,2344
	Umidade	0,0003	0,0014	0,1899
	MP ₁₀	0,0006	0,0004	1,5609

Nota: * Coeficientes estatisticamente significativos

Análises para pneumonia e gripe em crianças menores de 4 anos e entre 5 e 14 anos também foram realizadas, devido à grande incidência de pneumonia e gripe nas crianças.

Através dos resultados do ajuste dos modelos apresentados na Tabela 4.13, verifica-se que os valores do pseudo R² para as duas faixas etárias não estiveram próximos de 1, porém a razão χ^2/g para crianças entre 5 e 14 anos mostrou um ajuste melhor.

Tabela 4.13 – Pseudo R², estatística de Pearson (χ^2) e razão entre χ^2 e graus de liberdade para o modelo com variável mês de crianças com pneumonia e gripe

Faixa etária	R ²	χ^2	χ^2/g
Menores de 4 anos	0,620	861,979	1,216
Entre 5 e 14 anos	0,597	799,381	1,127

Através da Figura 4.29 não é possível verificar uma diferença visível entre os modelos para as diferentes faixas etárias, sendo que a maioria dos pontos para os dois modelos está dentro do intervalo entre -2 e 2.

Os valores ajustados estiveram próximos dos valores observados, como mostra a Figura 4.29, não podendo ser verificada uma diferença visível entre os modelos para as duas faixas etárias consideradas.

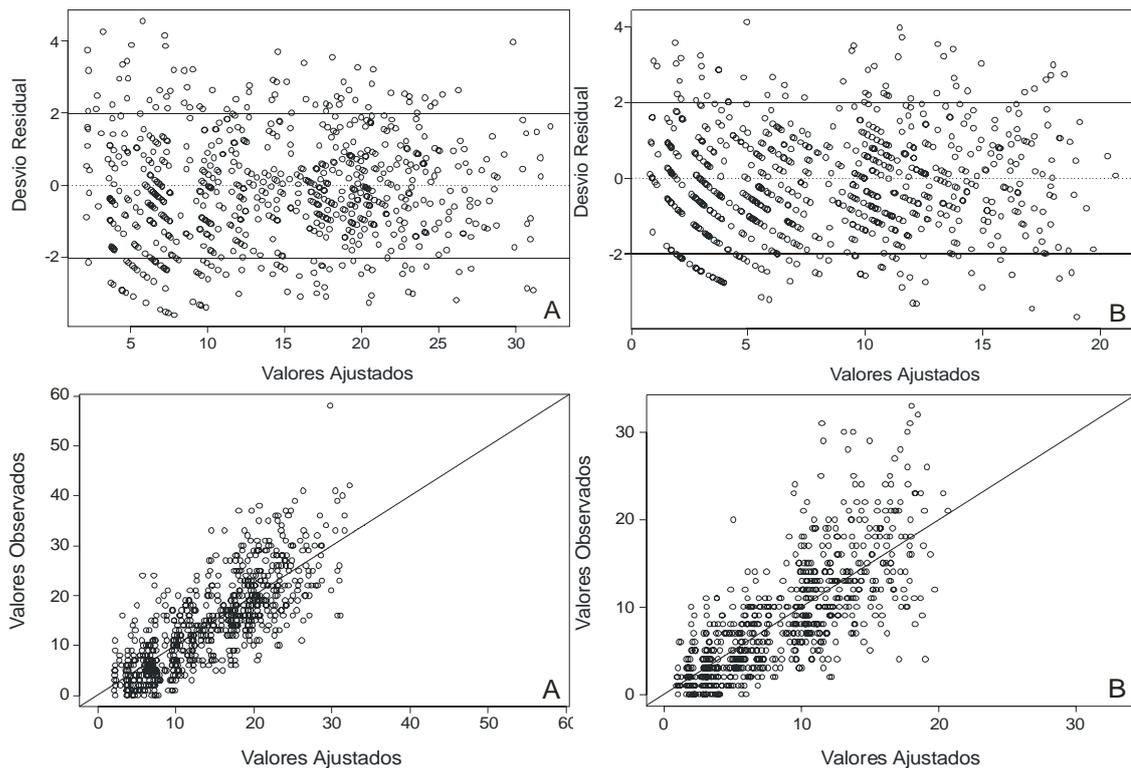


Figura 4.29 – Gráficos dos desvios residuais e dos valores observados em relação aos valores ajustados para pneumonia e gripe em crianças menores de 4 anos (A) e entre 5 e 14 anos (B)

A partir dos resultados das análises para crianças, verificou-se que o modelo se ajustou melhor para pneumonia e gripe ($\chi^2/gI = 1,216$ para crianças menores de 4 anos e $\chi^2/gI = 1,127$ para crianças entre 5 e 14 anos) do que para todas as doenças respiratórias ($\chi^2/gI = 1,479$ para crianças menores de 4 anos e $\chi^2/gI = 1,425$ para crianças entre 5 e 14 anos). Esta diferença ocorreu possivelmente porque no modelo para todas as doenças respiratórias, estão sendo considerados alguns tipos de doenças respiratórias que não estão relacionados com a poluição atmosférica.

Os coeficientes de regressão, os erros padrões e os valores de t da concentração de MP_{10} , temperatura e umidade para os modelos considerando apenas pneumonia e gripe são apresentados na Tabela 4.14, os demais coeficientes de regressão são apresentados na Tabela 4.15. O coeficiente de regressão para concentração de MP_{10} não é estatisticamente significativo em nenhum dos casos. Para crianças entre 5 e 14 anos, os coeficientes para temperatura e umidade são estatisticamente significativos e para crianças menores de 4 anos nenhum coeficiente se mostrou estatisticamente significativo.

Tabela 4.14 – Coeficientes de regressão, erro padrão e valores de t para pneumonia e gripe em crianças

Faixa etária		Coeficiente	Erro padrão	Valor de t
Menores de 4 anos	Temperatura	0,0075	0,0058	1,2797
	Umidade	-0,0006	0,0016	-0,3480
	MP_{10}	0,0002	0,0005	0,4840
Entre 5 e 14 anos	Temperatura	0,0162	0,0071	2,2820*
	Umidade	0,0051	0,0020	2,5638*
	MP_{10}	0,0005	0,0006	0,7697

Nota: * Coeficientes estatisticamente significativos

Através de todas as análises estatísticas realizadas, concluiu-se que o coeficiente de regressão para concentração de MP_{10} não se mostrou estatisticamente significativo em nenhum dos casos, possivelmente devido à falta de fidedignidade dos dados de concentração de MP_{10} disponíveis.

Através das análises apresentadas neste trabalho é possível realizar previsões sobre o número de pessoas que podem adquirir uma doença respiratória em relação à concentração de MP_{10} presente no ar e às condições meteorológicas diárias.

Os coeficientes de regressão e os erros padrões de todas as análises realizadas a fim de prever o impacto da poluição atmosférica na saúde populacional são apresentados na Tabela 4.15. Como a variável mês, dias da semana e feriado são variáveis qualitativas, tem-se um coeficiente de regressão para cada uma das qualidades exceto àquela considerada como referência, que para a variável mês foi dezembro, para dia da semana foi sábado e para feriado, o dia em que não é feriado.

Tabela 4.15 – Coeficientes de regressão com respectivos erros padrões para todas as variáveis explicativas de todas as análises realizadas

	Variáveis explicativas	Menores de 4 anos	Pneumonia – menores de 4 anos	Entre 5 e 14 anos	Pneumonia – entre 5 e 14 anos	Entre 15 e 19 anos	Entre 20 e 59 anos	Maiores de 60 anos
β_0	Intercepto	4,6174 (0,0977)*	2,4807 (0,1897)*	3,9799 (0,1130)*	1,3296 (0,2303)*	2,4376 (0,1570)*	4,7733 (0,1054)*	2,4463 (0,1609)*
β_1	Temperatura	0,0035 (0,0030)	0,0075 (0,0058)	0,0083 (0,0035)*	0,0162 (0,0071)*	0,0027 (0,0049)	-0,0023 (0,0032)	0,0011 (0,0049)
β_2	Umidade	-0,0047 (0,0008)*	-0,0006 (0,0016)	-0,0022 (0,0010)*	0,0051 (0,0020)*	-0,0007 (0,0014)	0,0017 (0,0009)	0,0003 (0,0014)
β_3	Concentração de MP ₁₀	0,0001 (0,0002)	0,0002 (0,0005)	-0,0000 (0,0003)	0,0004 (0,0006)	-0,0007 (0,0004)	0,0003 (0,0003)	0,0006 (0,0004)
β_4 – Meses	Janeiro	0,0369 (0,0247)	-0,0165 (0,0480)	0,0868 (0,0281)*	0,0492 (0,0599)	0,0433 (0,0345)	0,0159 (0,0247)	-0,0260 (0,0380)
	Fevereiro	0,1371 (0,0122)*	0,1191 (0,0241)*	0,1388 (0,0137)*	0,1728 (0,0289)*	0,1021 (0,0192)*	0,0505 (0,0131)*	0,0593 (0,0198)*
	Março	0,1471 (0,0077)*	0,1357 (0,0153)*	0,1327 (0,0090)*	0,1468 (0,0187)*	0,1225 (0,0125)*	0,0918 (0,0088)*	0,0766 (0,0134)*
	Abril	0,1147 (0,0062)*	0,1461 (0,0119)*	0,0994 (0,0074)*	0,1602 (0,0142)*	0,0838 (0,0104)*	0,0803 (0,0072)*	0,0704 (0,0108)*
	Mai	0,0690 (0,0053)*	0,0783 (0,0103)*	0,0624 (0,0062)*	0,0804 (0,0124)*	0,0409 (0,0089)*	0,0379 (0,0060)*	0,0467 (0,0091)*
	Junho	0,0281 (0,0045)*	0,0215 (0,0088)*	0,0191 (0,0053)*	0,0112 (0,0110)	0,0194 (0,0073)*	0,0312 (0,0048)*	0,0394 (0,0072)*
	Julho	0,0197 (0,0038)*	0,0227 (0,0075)*	0,0475 (0,0043)*	0,0660 (0,0084)*	0,0494 (0,0059)*	0,0426 (0,0041)*	0,0391 (0,0061)*
	Agosto	0,0045 (0,0031)	0,0056 (0,0061)	0,0080 (0,0035)*	0,0070 (0,0073)	0,0031 (0,0050)	0,0138 (0,0032)*	0,0174 (0,0050)*
	Setembro	0,0145 (0,0026)*	0,0197 (0,0050)*	0,0088 (0,0031)*	0,0148 (0,0061)*	0,0161 (0,0041)*	0,0129 (0,0028)*	0,0151 (0,0043)*
	Outubro	-0,0100 (0,0027)*	-0,0004 (0,0051)	-0,0016 (0,0030)	-0,0016 (0,0062)	-0,0008 (0,0042)	0,0086 (0,0027)*	0,0097 (0,0042)*
	Novembro	-0,0225 (0,0026)*	-0,0265 (0,0052)*	-0,0248 (0,0030)*	-0,0349 (0,0065)*	-0,0157 (0,0040)*	-0,0128 (0,0027)*	-0,0080 (0,0040)*
		Dezembro**						
β_5 – Dias da semana	Domingo	0,4765 (0,0171)*	0,5862 (0,0349)*	0,5331 (0,0202)*	0,6521 (0,0440)*	0,5401 (0,0289)*	0,5466 (0,0187)*	0,8281 (0,0331)*
	Segunda	0,1362 (0,0082)*	0,1351 (0,0166)*	0,1591 (0,0096)*	0,1742 (0,0202)*	0,1828 (0,0135)*	0,1337 (0,0090)*	0,1744 (0,0148)*
	Terça	0,0778 (0,0054)*	0,0603 (0,0111)*	0,1060 (0,0061)*	0,0773 (0,0133)*	0,1011 (0,0087)*	0,0879 (0,0059)*	0,1242 (0,0091)*
	Quarta	0,0420 (0,0041)*	0,0504 (0,0082)*	0,0457 (0,0048)*	0,0567 (0,0099)*	0,0459 (0,0067)*	0,0284 (0,0046)*	0,0354 (0,0072)*
	Quinta	0,0045 (0,0035)	0,0498 (0,0064)*	-0,0063 (0,0042)	0,0515 (0,0077)*	0,0151 (0,0056)*	0,0082 (0,0038)*	0,0060 (0,0060)
	Sexta	-0,0982 (0,0041)*	-0,1225 (0,0089)*	-0,1179 (0,0051)*	-0,1440 (0,0114)*	-0,1093 (0,0069)*	-0,1033 (0,0045)*	-0,1383 (0,0080)*
		Sábado**						
β_6	Feriado	-0,4277 (0,0381)*	-0,5071 (0,0746)*	-0,5165 (0,0459)*	-0,5578 (0,0930)*	-0,4605 (0,0611)*	-0,4108 (0,0411)*	-0,7689 (0,0747)*

Nota: * Coeficientes estatisticamente significativos

** Mês e dia da semana considerados como referência.

$$\log(DR) = \beta_0 + \beta_1 temp + \beta_2 umid + \beta_3 MP_{10} + \beta_4 mes + \beta_5 dds + \beta_6 fer - \text{Eq. 4.3.}$$

Para realizar uma previsão, é necessário ter os dados de temperatura, umidade e concentração de MP_{10} para um dia específico do ano e substituir os valores dos coeficientes de regressão apresentado na Tabela 4.15 na Equação 4.3.

Considerando, por exemplo, que para o dia 04 de Março de 2007 (domingo – não feriado) a temperatura é de 20°C, a umidade é de 80% e deseja-se avaliar qual o número de pessoas maiores de 60 anos que adquirirão uma doença respiratória devido a uma concentração de 100 $\mu\text{g}/\text{m}^3$ de MP_{10} , devem ser considerados os valores da última coluna apresentada na Tabela 4.15. Calculando-se, pode-se concluir que aproximadamente 31 pessoas adquiririam uma doença respiratória.

Cabe lembrar que os modelos para previsão estimados neste estudo não apresentaram um bom ajuste devido a erros na coleta de dados experimentais de concentração de MP_{10} e fatores relacionados à não utilização de funções de suavização para as variáveis explicativas consideradas.

Através das análises realizadas neste estudo, notou-se então que é necessário investigar detalhadamente os fatores que levaram aos erros encontrados nos modelos para previsão, bem como verificar quais fatores podem ser considerados a fim de melhorar o ajuste do modelo, para então ser possível utilizar o modelo no intuito de realizar previsões.

De acordo com os resultados apresentados neste capítulo, segue no Capítulo 5 as conclusões encontradas, bem como sugestões para pesquisas futuras.

5 CONCLUSÕES

Os modelos estatísticos mais utilizados na análise da relação entre a poluição atmosférica e a saúde populacional são as análises de regressão, pois são ferramentas úteis para avaliar a relação de uma ou mais variáveis explicativas com uma única variável resposta.

A maior dificuldade neste trabalho foi obter um banco de dados completo e fidedigno, devido à falta de dados de concentração de MP_{10} para o período de estudo.

A falta de dados provocou um equívoco entre os resultados presentes nos relatórios de qualidade do ar do IAP de 2004 e 2005 e os resultados obtidos neste trabalho. De acordo com os relatórios de qualidade do ar do IAP, a concentração média diária máxima de MP_{10} em 2004 foi de $126,3 \mu\text{g}/\text{m}^3$ no dia 22 de setembro e em 2005 foi de $110,6 \mu\text{g}/\text{m}^3$ no dia 2 de agosto. No presente trabalho, verificou-se que a concentração média diária máxima de MP_{10} para 2004 foi de $230,66 \mu\text{g}/\text{m}^3$ no dia 18 de agosto e para 2005 foi de $158,13 \mu\text{g}/\text{m}^3$ no dia 4 de agosto.

Cabe mencionar que os dados horários para setembro de 2004 não foram fornecidos e a concentração média diária de MP_{10} calculada para o dia 2 de agosto de 2005 foi de $153,64 \mu\text{g}/\text{m}^3$.

Os modelos estatísticos mais encontrados em estudos sobre o impacto da poluição atmosférica na saúde populacional são os modelos de regressão de Poisson para MLG e MAG, de regressão condicional logística e proporcional de Cox. Comparando estes modelos, optou-se pelo modelo linear generalizado (MLG) para regressão de Poisson, pois a variável resposta presente no estudo é não negativa e em forma de contagens e possui distribuição de Poisson, além de almejar-se realizar previsões com o modelo escolhido.

Os resultados foram apresentados em duas etapas. A primeira etapa foi realizada para avaliar o impacto da concentração de MP_{10} na saúde populacional do município de Araucária para o período de 2004 a 2005. O modelo utilizado nesta etapa foi o modelo de regressão de Poisson com função de ligação canônica, ajuste

de sazonalidade através da função *spline* cúbica natural e acréscimo de resíduos no modelo, por ser o modelo que apresentou melhor ajuste.

Na segunda etapa, foi proposto um modelo capaz de prever o impacto da concentração de MP₁₀ na saúde populacional do município de Araucária em que foi utilizado o modelo de regressão de Poisson com uma variável qualitativa para meses do ano a fim de suavizar a sazonalidade e não foram acrescentados resíduos no modelo.

A variável resposta escolhida em todas as análises foi o número de pessoas com DR e pneumonia e gripe no município de Araucária e como variáveis explicativas foram consideradas a concentração de MP₁₀, a temperatura, a umidade, dias da semana e feriado bem como a suavização para sazonalidade escolhida.

Em relação às análises para avaliação, concluiu-se que o modelo de regressão de Poisson se ajustou melhor aos dados para doenças respiratórias em maiores de 60 anos ($\chi^2/\text{gl} = 0,955$), sendo que esta foi a única faixa etária em que não foi preciso acrescentar resíduo para melhorar o ajuste do modelo. O modelo para pneumonia e gripe em crianças entre 5 e 14 anos também se ajustou de forma satisfatória ($\chi^2/\text{gl} = 1,026$).

Em relação às análises para previsão, concluiu-se que o modelo para pessoas maiores de 60 anos se ajustou melhor ($\chi^2/\text{gl} = 0,977$).

Tanto nas análises para avaliação quanto para previsão os modelos para pneumonia e gripe (crianças menores de 4 anos: avaliação $\chi^2/\text{gl} = 1,180$ e previsão $\chi^2/\text{gl} = 1,216$ e para crianças entre 5 e 14 anos: avaliação $\chi^2/\text{gl} = 1,026$ e previsão $\chi^2/\text{gl} = 1,127$) se ajustou melhor que o modelo para todas as doenças respiratórias (crianças menores de 4 anos: avaliação $\chi^2/\text{gl} = 1,378$ e previsão $\chi^2/\text{gl} = 1,479$ e para crianças entre 5 e 14 anos: avaliação $\chi^2/\text{gl} = 1,256$ e previsão $\chi^2/\text{gl} = 1,425$).

Os coeficientes de regressão para a concentração de MP₁₀ se mostraram estatisticamente significativos nos modelos para avaliação de doenças respiratórias em crianças menores de 4 anos, de doenças respiratórias e de pneumonia e gripe em crianças entre 5 e 14 anos.

Nos modelos para previsão, o coeficiente de regressão para a concentração de MP_{10} não se mostrou estatisticamente significativo em nenhuma das faixas etárias, provavelmente devido à falta de confiabilidade dos dados de concentração de MP_{10} e por ter sido utilizada a variável mês para suavização da sazonalidade, ao invés da função *spline* cúbica natural.

Apesar da população de Araucária ser predominantemente adulta, os gráficos de risco relativo mostraram que as crianças e idosos foram os mais afetados.

O modelo de regressão de Poisson dos MLG é adequado para avaliar e prever o impacto da poluição atmosférica na saúde populacional, porém é necessário investigar possíveis fatores que possam contribuir para a melhoria no ajuste do modelo.

Com os resultados encontrados nesta pesquisa espera-se contribuir com as questões ambientais da região estudada, bem como apresentar uma modelagem estatística que pode ser utilizada em estudos epidemiológicos e em diversos outros problemas de engenharia.

5.1 Sugestões

Uma forma de melhorar os resultados apresentados neste estudo seria considerar separadamente os tipos de doenças respiratórias que possuem como agente causador principal a poluição atmosférica, como foi feito para as crianças. Outra opção seria incluir no modelo outros fatores que possuem influência sobre a variável resposta, como por exemplo, considerar condições sociais diferentes.

No intuito de obter dados de concentração de poluentes mais fidedignos, bem como avaliar o impacto de uma única fonte poluidora, sugere-se realizar a simulação numérica da dispersão dos poluentes para a região em estudo.

PRODUÇÃO CIENTÍFICA NO PERÍODO (Fev. 2005 – Fev. 2007)

TADANO, Y. S.; UGAYA, C. M. L.; FRANCO, A. T. Life Cycle Impact Assessment: PM₁₀ effects on Araucaria's population health. CILCA 2007; São Paulo, SP, Fevereiro 2007.

PEGORARO, L. A.; GOEDKOOP, M.;UGAYA, C. M. L.; TADANO, Y. S. The Life Cycle Impact Assessment (LCIA) and some important considerations to Brazil. CILCA 2007; São Paulo, SP, Fevereiro 2007.

TADANO, Y. S.; BOGEA, L. C.; UGAYA, C. M. L.; FRANCO, A. T. LCIA for respiratory diseases caused by particulate matter in Araucaria, Parana, Brazil. SETAC North American Annual Meeting, Montreal, Quebec, Canada, Novembro 2006.

TADANO, Y. S.; UGAYA, C. M. L.; FRANCO, A. T. Avaliação do impacto do ciclo de vida: Efeitos dos poluentes e das condições atmosféricas na saúde da população de Araucária. Encontro de PRHs Região Sul, Curitiba, PR, Novembro 2006.

TADANO, Y. S.; UGAYA, C. M. L.; FRANCO, A. T. Análise estatística do impacto da poluição atmosférica na saúde populacional. Rio Oil & Gas 2006 Expo and Conference, Setembro 2006.

REFERÊNCIAS

ABBEY, D. E.; NISHINO, N.; MCDONNELL, W. F.; BURCHETTE, R. J.; KNUTSEN, S. F.; BEESON, W. L.; YANG, J. X. Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. **Am. J. Respir. Crit. Care Med.**, v. 159, p. 373-382, 1999.

ALMEIDA, I. T. **A poluição atmosférica por material particulado na mineração a céu aberto**. 1999. 186 f. Dissertação (Mestrando em Engenharia Mineral) - Escola Politécnica da Universidade de São Paulo, São Paulo, 1999.

BAKONYI, S. M. C. **Poluição do ar e doenças respiratórias em Curitiba/PR. 2003**. 198 f. Dissertação (Mestrado em Geografia) – Setor de Ciências da Terra, Universidade Federal do Paraná, Curitiba, 2003.

BAKONYI, S. M. C.; DANNI-OLIVEIRA, I. M.; MARTINS, L. C.; BRAGA, A. L. F. Poluição atmosférica e doenças respiratórias em crianças na cidade de Curitiba, PR. **Rev. Saúde Pública**, v. 38, n. 5, p. 695-700, 2004.

BAXTER, L. A.; FINCH, S. J.; LIPFERT, F. W.; YU, Q. Comparing estimates of the effects of air pollution on human mortality obtained using different regression methodologies. **Risk Analysis**, v. 17, n. 3, p. 273-278, 1997.

BEER, T.; RICCI, P. F. A quantitative risk assessment method based on population and exposure distributions using Australian air quality data. **Environ. Intern.**, v. 25, n. 6/7, p. 887-898, 1999.

BERGAMASCHI D. P.; SOUZA, J. M. P. **Bioestatística aplicada a Nutrição, Aula 11**. 2005. Universidade de São Paulo. Disponível em: <<http://www.fsp.usp.br/hep103/Aula11.pdf>>. Acesso em 27 de Set. 2006.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis: Forecasting and control**. Terceira Edição. EUA: Prentice-Hall, Inc., 1994. 598 p.

BRAGA, A. L. F.; PEREIRA, L. A. A. e SALDIVA, P. H. N. **Poluição atmosférica e seus efeitos na saúde humana.** Disponível em: <http://www.cgu.unicamp.br/energia2020/papers/paper_Saldiva.pdf>. Acesso em 20 Nov. 2005.

BRAGA, A. L. F.; SALDIVA, P. H. N.; PEREIRA, L.A.A.; MENEZES, J.J.C.; CONCEIÇÃO, G.M.S.; LIN, C.A., ZANOBETTI, A.; SCHWARTZ, J.; DOCKERY, D.W. Health effects of air pollution exposure on children and adolescents in São Paulo, Brazil. **Pediatric Pulmonology**, v. 31, p. 106-113, 2001.

BRANCO, S. M.; MURGEL, E. **Poluição do ar.** São Paulo, Editora Moderna, 2001, 87 p.

CASTANHO, A. D. A. **A determinação quantitativa de fontes de material particulado na atmosfera da cidade de São Paulo.** 1999. 140 f. Dissertação (Mestrado em Física) – Universidade de São Paulo. São Paulo. 1999.

CETESB, **Relatório de qualidade do ar no estado de São Paulo, 2005.** Disponível em <http://www.cetesb.sp.gov.br/Ar/ar_geral.asp>. Acesso em 03 Out. 2006.

CHAPRA, S. C.; CANALE, R. P. **Numerical methods for engineers with personal computer applications.** Segunda edição. EUA: McGraw-Hill International Editions, 1987. 570 p.

CONCEIÇÃO, G. M. S.; SALDIVA, P. H. N.; SINGER, J. M. Modelos MLG e MAG para análise da associação entre poluição atmosférica e marcadores de morbimortalidade: uma introdução baseada em dados da cidade de São Paulo. **Rev. Bras. Epidemiol.**, v. 4, n. 3, p. 206-219, 2001.

DANNI-OLIVEIRA, I. M., **A cidade de Curitiba/PR e a poluição do ar.** 1999. 320 f. Tese (Doutorado em Geografia) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 1999.

DERISIO, J. C. **Introdução ao controle de poluição ambiental.** Segunda edição. São Paulo: Signus Editora, 2000.

DOCKERY, D. W.; POPE III, C. A.; XU, X.; SPENGLER, J. D.; WARE, J. H.; FAY, M. E.; FERRIS, B. G. Jr.; SPEIZER, F. E. An association between air pollution and mortality in six U.S. cities. **New England Journal of Medicine**, v. 29, n. 24, p. 1753-1759, 1993.

EPA (Environmental Protection Agency). **A comparison of CALPUFF modeling results to two tracer field experiments**. 1998. Disponível em: <<http://www.epa.gov/ttn/scram/7thconf/calpuff/tracer.pdf>>. Acesso em 16 Nov. 2005.

EVERITT, B. S. **Modern medical statistics**. Nova Iorque: Oxford University Press Inc., 2003. 235 p.

FERRAZ, M. I. F.; SÁFADI, T.; LAGE, G. Uso de modelos de séries temporais na previsão de séries de precipitação pluviométricas mensais no município de Lavras – MG. **Revista Brasileira de Agrometeorologia**, Santa Maria, v. 7, n.2, p. 259-267, 1999.

FONSECA J. S.; MARTINS, G. A. **Curso de Estatística**. Sexta edição. São Paulo: Atlas, 1996. 320 p.

FREITAS, C.; BREMNER S. A.; GOUVEIA, N.; PEREIRA, L. A. A.; SALDIVA, P. H. N. Internações e óbitos e sua relação com a poluição atmosférica em São Paulo, 1993-1997. **Rev. Saúde Pública**, v. 38, n. 6, p. 751-757, 2004.

HA, E-H.; LEE, J-T. KIM, H.; HONG, Y-C.; LEE, B-E.; PARK H-S. Infant susceptibility of mortality to air pollution in Seoul, South Korea. **Pediatrics**, v. 111, n. 2, p. 283-290, 2005.

HAIR JR. J. F., ANDERSON R. E., TATHAM R. L., BLACK W. C. **Análise Multivariada de dados**. Quinta Edição. São Paulo: Bookman, 2005. 593 p.

IAP, **Relatório de qualidade do ar na Região Metropolitana de Curitiba 2004**. Disponível em: <http://www.pr.gov.br/meioambiente/pdf/relatorio_ar_2004.pdf> Acesso em Setembro de 2005.

IAP, **Relatório de qualidade do ar na Região Metropolitana de Curitiba 2005**. Disponível em: <http://www.pr.gov.br/meioambiente/pdf/relatorio_ar_2005.pdf> Acesso em 19 Jul. de 2006.

IBALD-MULLI, A.; TIMONEN, K. L.; PETERS, A.; HEINRICH, J.; WÖLKE, G.; LANKI, T.; BUZORIUS, G.; KREYLING, W. G.; HARTOG, J.; HOEK, G.; BRINK, H. M.; PEKKANEN, J. Effects of particulate air pollution on blood pressure and heart rate in subjects with cardiovascular disease: A multicenter approach. **Environmental Health Perspectives**, v. 112, n. 3, p. 369-377, 2004.

IPPUC – **Instituto de Pesquisa e Planejamento Urbano de Curitiba**. Disponível em: <<http://www.ippuc.org.br>>. Acesso em 13 Nov. 2006.

LEE, J. T.; SCHWARTZ, J., Reanalysis of the effects of air pollution on daily mortality in Seoul, Korea: A case-crossover design. **Environmental Health Perspectives**, v. 107, n. 8, p. 633-636, 1999.

MARIANO, J. B. **Impactos ambientais do refino de petróleo**. 2001. 289 f. Tese (Doutorado em Ciências em planejamento energético) – Universidade Federal do Rio de Janeiro. Rio de Janeiro. 2001.

MARTINS, L. C. **Relação entre poluição atmosférica e algumas doenças respiratórias em idosos**: Avaliação do rodízio de veículos no município de São Paulo. 2000. (Dissertação de Mestrado) – Faculdade de Medicina, Universidade de São Paulo, São Paulo, 2000.

MARTINS, L. C.; LATORRE, M. R. D. O.; CARDOSO, M. R.; GONÇALVES, F. L. T.; SALDIVA, P. H. N.; BRAGA, A. L. F. Poluição atmosférica e atendimentos por pneumonia e gripe em São Paulo, Brasil. **Rev. Saúde Pública**, v. 36, n. 1, p. 88-94, 2002.

McCULLAGH, P., Quasi-likelihood functions, **Annals of Statistics**, v. 11, p. 59-67, 1983.

McCULLAGH, P.; NELDER, J. A., **Generalized Linear Models**, Segunda edição. Chapman & Hall, 1989. 511 p.

MEYER, P. L. **Probabilidade**: Aplicações à Estatística. Segunda Edição. Rio de Janeiro: Livros Técnicos e Científicos Editora, 1983.

Ministério da Saúde e IBGE. Disponível em <<http://tabnet.datasus.gov.br>> Acesso em 05 de Outubro de 2006.

Ministério da Saúde. **Estatuto da criança e do adolescente**. Terceira edição. Brasília: Editora do Ministério da Saúde, 2006a. 96 p.

Ministério da Saúde. **Estatuto do Idoso**. Segunda edição revisada. Brasília: Editora do Ministério da Saúde, 2006b. 70 p.

MYERS, R. H.; MONTGOMERY, D. C. **Response Surface Methodology: Process and Product Optimization Using Designed Experiments**. Segunda edição. Nova Iorque: John Wiley & Sons, 2002. 798 p.

NEAS, L. M.; SCHWARTZ, J.; DOCKERY, D. A. Case-crossover analysis of air pollution and mortality in Philadelphia. **Environmental Health Perspectives**, v. 107, n. 8, p. 629-631, 1999.

NELDER, J. A.; WEDDERBURN, R. W. M. Generalized Linear Models. **Journal of the Royal Statistical Society A**, v. 135, n. 2, p. 370-384, 1972.

Organização Mundial da Saúde. **CID-10 Classificação estatística internacional das doenças e problemas relacionados à saúde**, v.1. Décima edição revisada. São Paulo: Universidade de São Paulo; 1997.

PAULA, G. A. **Modelos de regressão com apoio computacional**. 2004. Instituto de Matemática e Estatística. Universidade de São Paulo. 2004. Disponível em: <<http://www.ime.usp.br/~giapaula/livro.pdf>>. Acesso em 25 Jan. 2006.

PETERS, A.; DOCKERY, D. W.; MULLER, J. E.; MITTLEMAN, M.A. Increased particulate air pollution and the triggering of myocardial infarction. **Circulation**, v. 103, p. 2810-2815, Jun. 2001.

PIRES, D. O. **Inventário de emissões atmosféricas de fontes estacionárias e sua contribuição para a poluição do ar na região metropolitana do Rio de Janeiro**. 2005. 194 f. Dissertação (Mestrado em Engenharia de planejamento energético) – Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, 2005.

POPE III, C. A.; BURNETT, R. T.; THUN, M. J.; CALLE, E. E.; KREWSKI, D.; ITO, K.; THURSTON, G. D. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution, **Journal of the American Medical Association**, v. 287, n. i9, p. 1132-1142, 2002.

Prefeitura de Araucária. Disponível em: <<http://www.atontecnologia.com.br/clientes/araucaria>>. Acesso em 16 Nov. 2005.

RIBEIRO A. J. F. **Um estudo sobre mortalidade dos aposentados por invalidez do regime geral da previdência social (RGPS)**. 2006. 191 f. Tese (Doutorado do Centro de Desenvolvimento e Planejamento Regional) - Faculdade de Ciências Econômicas, Universidade Federal de Minas Gerais, Belo Horizonte, 2006.

ROUQUAYROL, M. Z.; ALMEIDA FILHO, N. **Epidemiologia e Saúde**, Quinta edição. Rio de Janeiro: Ed. Médica Científica, 1999. 667 p.

RUSSO, S. L. **Gráficos de controle para variáveis não-conformes autocorrelacionadas**. 2002. 120 f. Tese (Doutorado em Engenharia de Produção) - Universidade Federal de Santa Catarina. Florianópolis, 2002.

SCHMIDT, C. M. C. **Modelo de regressão de Poisson aplicado à área da saúde**. 2003. 98 f. Dissertação (Mestrado em Modelagem Matemática) - Universidade Regional do Noroeste do Estado do Rio Grande do Sul, Ijuí, 2003.

SEINFELD J. H. **Atmospheric chemistry and physics of air pollution**. Nova Iorque: John Wiley & Sons, Inc., 1986. 738 p.

SIMEPAR (Tecnologia e Informações Ambientais). Disponível em: <<http://www.simepar.br>>. Acesso em 18 Ago. 2006.

SOUZA, L. S. **Poluição do ar e doenças respiratórias, Araucária – PR**. 206 f. Dissertação (Mestrado em Geografia) – Setor de Ciências da Terra, Universidade Federal do Paraná, Curitiba, 2006.

S-PLUS 7, **Guide to Statistics**. Volume I,. Insightful Corporation, Seattle, Washington. Abr. 2005.

S-PLUS version 7.0 for Windows (computer software), **Insightful Corporation** (1988-2005), Seattle, Washington: Insightful Corporation. 2005.

SZWARCFITER, L. **Opções para o aprimoramento do controle de emissões de poluentes atmosféricos por veículos leves no Brasil: uma avaliação do potencial de programas de inspeção e manutenção e de renovação acelerada da frota**. 2004. 275 f. Tese (Doutorado em Ciências em planejamento energético) – Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2004.

WEDDERBURN, R.W.M. Quasilikelihood functions, generalized linear models and the Gauss-Newton method, **Biometrika**, v. 61, p. 439-447, 1974.

WERKEMA, M. C. C.; AGUIAR, S. **Análise de regressão: Como entender o relacionamento entre as variáveis de um processo**. Série Ferramentas da qualidade - 7. Belo Horizonte, Fundação Christiano Ottoni da Escola de Engenharia da UFMG, 1996. 311 p.

ANEXO A – INFORMAÇÕES SOBRE AS ESTAÇÕES DE MONITORAMENTO DA QUALIDADE DO AR DA RMC

Quadro A.1 - Estações de monitoramento da qualidade do ar do município de Araucária no ano de 2005 (IAP, 2005)

	Estação de Monitoramento	Localização	Parâmetros medidos no ano de 2005		Período de funcionamento/ Responsável pelo custo operacional
			Poluentes	Meteorologia	
Automática	Assis Automática (ASS aut.)	Centro/ Norte de Araucária	SO ₂ , NO, NO ₂ , O ₃ , PTS	Todas as estações: temperatura, umidade relativa, radiação global, pressão, velocidade e direção do vento.	Desde Abril de 2000/ SMMA Araucária
	UEG ⁽²⁾ (UEG)	Região Central de Araucária	SO ₂ , CO, O ₃ , PI		Desde Maio de 2003/ IAP
	CSN – CISA ⁽³⁾ (CISA)	Centro/ Nordeste de Araucária	SO ₂ , NO, NO ₂ , CO, O ₃ , PI, PTS, HCT ⁽¹⁾	Exceções: ASS: sem radiação UVA, UVB, velocidade e direção do vento. CISA: sem radiação global, UVA, UVB REP: sem radiação UVA, UVB, umidade relativa, pressão	Desde Agosto de 2002/ CISA
	REPAR	Centro/ Nordeste de Araucária	SO ₂ , NO, NO ₂ , CO, O ₃ , PTS, PI, benzeno, tolueno, etilbenzeno		Desde Julho de 2003/ REPAR
Manual	São Sebastião (SS)	Centro/ Leste de Araucária	Fumaça, SO ₂ , NH ₃		Desde 1985/ IAP
	Assis (ASS man.)	Centro/ Norte de Araucária	Fumaça, SO ₂ , NH ₃		Desde 1985/ IAP
	Seminário (SEM)	Região central de Araucária	Fumaça, SO ₂ , NH ₃		Desde 1985/ IAP

Notas: 1) Hidrocarbonetos Totais

2) UEG – Usina Elétrica a Gás de Araucária

3) CISA/CSN – Aços Revestidos S/A – Araucária, PR.

ANEXO B – DISTRIBUIÇÃO T DE STUDENT

Tabela B.1.1 – Distribuição t de Student (Fonseca e Martins, 1996)

$\alpha \backslash \varphi$	0,50	0,25	0,10	0,05	0,025	0,01	0,005
1	1,00000	2,4142	6,3138	12,7060	25,5420	63,6570	127,3200
2	0,81650	1,6036	2,9200	4,3127	6,2053	9,9248	14,0890
3	0,76489	1,4226	2,3534	3,1825	4,1765	5,8409	7,4533
4	0,74070	1,3444	2,1318	2,7764	3,4954	4,6041	5,5976
5	0,72669	1,3009	2,0150	2,5706	3,1634	4,0321	4,7733
6	0,71756	1,2733	1,9432	2,4469	2,9687	3,7074	4,3168
7	0,71114	1,2543	1,8946	2,3646	2,8412	3,4995	4,0293
8	0,70639	1,2403	1,8595	2,3060	2,7515	3,3554	3,8325
9	0,70272	1,2297	1,8331	2,2622	2,6850	3,2498	3,6897
10	0,69981	1,2213	1,8125	2,2281	2,6338	3,1693	3,5814
11	0,69745	1,2145	1,7959	2,2010	2,5931	3,1058	3,4966
12	0,69548	1,2089	1,7823	2,1788	2,5600	3,9545	3,4284
13	0,69384	1,2041	1,7709	2,1604	2,5326	3,0123	3,3725
14	0,69200	1,2001	1,7613	2,1448	2,5096	2,9768	3,3257
15	0,69120	1,1967	1,7530	2,1315	2,4899	2,9467	3,2860
16	0,69013	1,1937	1,7459	2,1199	2,4729	2,9208	3,2520
17	0,68919	1,1910	1,7396	2,1098	2,4581	2,8982	3,2225
18	0,68837	1,1887	1,7341	2,1009	2,4450	2,8784	3,1966
19	0,68763	1,1866	1,7291	2,0930	2,4334	2,8609	3,1737
20	0,68696	1,1848	1,7247	2,0860	2,4231	2,8453	3,1534
21	0,68635	1,1831	1,7207	2,0796	2,4138	2,8314	3,1352
22	0,68580	1,1816	1,7171	2,0739	2,4055	2,8188	3,1188
23	0,68531	1,1802	1,7139	2,0687	2,3979	2,8073	3,1040
24	0,68485	1,1789	1,7109	2,0639	2,3910	2,7969	3,0905
25	0,68443	1,1777	1,7081	2,0595	2,3846	2,7874	3,0782
26	0,68405	1,1766	1,7056	2,0555	2,3788	2,7787	3,0669
27	0,68370	1,1757	1,7033	2,0518	2,3734	2,7707	3,0565
28	0,68335	1,1748	1,7011	2,0484	2,3685	2,7633	3,0469
29	0,68304	1,1739	1,6991	2,0452	2,3638	2,7564	3,0380
30	0,68276	1,1731	1,6973	2,0423	2,3596	2,7500	3,0298
40	0,68066	1,1673	1,6839	2,0211	2,3289	2,7045	2,9712
60	0,67862	1,1616	1,6707	2,0003	2,2991	2,6603	2,9146
120	0,67656	1,1559	1,6577	1,9799	2,2699	2,6174	2,8599
∞	0,67449	1,1503	1,6449	1,9600	2,2414	2,5758	2,8070

ANEXO C – ESTIMATIVA DOS PARÂMETROS DE REGRESSÃO (PAULA, 2004)

O método escore de Fisher é baseado na chamada função de verossimilhança.

A função de verossimilhança é baseada na densidade de probabilidade de uma distribuição e de forma geral é dada por:

$$f(y; \theta; \phi) = \exp\{\phi[y\theta_i - b(\theta_i)] + c(y, \phi)\} \quad \text{Eq. C.1}$$

onde y é a variável resposta, θ é o parâmetro canônico, ϕ é o parâmetro de dispersão e $b(\theta_i)$ e $c(y, \phi)$ são funções específicas de cada distribuição (Paula, 2004).

Matematicamente é mais conveniente utilizar o logaritmo da função de verossimilhança, chamada de função de log-verossimilhança:

$$\ell(\beta; y) = \sum_{i=1}^n \phi \{y_i \theta_i - b(\theta_i)\} + \sum_{i=1}^n c(y_i, \phi) \quad \text{Eq. C.2}$$

No caso particular em que o parâmetro canônico (θ) coincide com o preditor linear, isto é, quando $\theta_i = \eta_i = \sum_{j=1}^p x_{ij} \beta_j$ a log-verossimilhança é:

$$\ell(\beta; y) = \sum_{i=1}^n \phi \left\{ y_i \sum_{j=1}^p x_{ij} \beta_j - b \left(\sum_{j=1}^p x_{ij} \beta_j \right) \right\} + \sum_{i=1}^n c(y_i, \phi) \quad \text{Eq. C.3}$$

Os valores dos parâmetros que maximizam a função de log-verossimilhança são obtidos através do cálculo das derivadas parciais da equação log-linear para cada parâmetro, ou seja, resolver o sistema igualando a Equação C.4 a zero:

$$U(\beta) = \partial \ell(\beta, y) / \partial \beta \quad \text{Eq. C.4}$$

A equação acima é conhecida como função escore e é obtida pela regra da cadeia:

$$U(\beta) = \sum_{i=1}^n \frac{\partial \ell(\beta, y)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \ell}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} \quad \text{Eq. C.5}$$

A derivada parcial da log-verossimilhança em relação ao parâmetro canônico é dada por:

$$\frac{\partial \ell}{\partial \theta_i} = \phi \left(y_i - \frac{db(\theta_i)}{d\theta_i} \right) \quad \text{Eq. C.6}$$

E a derivada de $\mu_i = \frac{db(\theta_i)}{d\theta_i}$ em relação a θ_i é:

$$\frac{\partial \mu_i}{\partial \theta_i} = \frac{d}{d\theta_i} \left(\frac{db(\theta_i)}{d\theta_i} \right) = \frac{\partial^2 b(\theta_i)}{\partial \theta_i^2} = V(\mu) \quad \text{Eq. C.7}$$

onde $V(\mu)$ é denominada função de variância.

A derivada em relação a β_i do preditor linear $\eta = \mathbf{X}^T \boldsymbol{\beta}$ é:

$$\frac{\partial \eta_i}{\partial \beta_j} = x_{ij} \quad \text{Eq. C.8}$$

Assim, substituindo-se as Equações C.6, C.7 e C.8 na Equação C.5 obtém-se:

$$\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial \ell(\boldsymbol{\beta}; \mathbf{y})}{\partial \boldsymbol{\beta}} = \phi \mathbf{X}^T \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu}) \quad \text{Eq. C.9}$$

onde $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$ é a matriz de pesos cujos valores são dados por $w_i = (d\mu_i/d\eta_i)^2 / V_i$; \mathbf{X} é uma matriz $n \times p$ cujas linhas serão denotadas por x_i^T , $i = 1, \dots, n$, $\mathbf{V} = \text{diag}\{V_1, \dots, V_n\}$, $\mathbf{y} = (y_1, \dots, y_n)^T$ e $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$.

Outra definição útil na estimação dos parâmetros de regressão é a matriz de informação de Fisher $\mathbf{K}(\boldsymbol{\beta})$, que corresponde ao valor esperado da segunda derivada parcial da log-verossimilhança em relação a $\boldsymbol{\beta}$, ou seja:

$$\mathbf{K}(\boldsymbol{\beta}) = E \left\{ -\frac{\partial^2 \ell(\boldsymbol{\beta}; \mathbf{y})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right\} = \phi \mathbf{X}^T \mathbf{W} \mathbf{X} \quad \text{Eq. C.10}$$

Para o caso especial em que se utiliza uma ligação canônica, os valores da função score e da matriz de informação de Fisher, tomam formas simplificadas:

$$\mathbf{U}(\boldsymbol{\beta}) = \phi \mathbf{X}^T (\mathbf{y} - \boldsymbol{\mu}) \quad \text{e} \quad \mathbf{K}(\boldsymbol{\beta}) = \phi \mathbf{X}^T \mathbf{V} \mathbf{X} \quad \text{Eq. C.11}$$

ou seja, a matriz de pesos é a própria função de variância \mathbf{V} .

Definidas todas as funções envolvidas na estimativa dos parâmetros, aplica-se o processo iterativo de Newton-Raphson para a obtenção da estimativa de máxima verossimilhança de β . Este método é definido através da expansão da função escore $U(\beta)$ em torno de um valor inicial $\beta^{(0)}$, tal que:

$$U(\beta) \cong U(\beta^{(0)}) + U'(\beta^{(0)})(\beta - \beta^{(0)}) \quad \text{Eq. C.12}$$

em que $U'(\beta^{(0)})$ é a primeira derivada de $U(\beta)$ em relação a β . Assim, considerando $\beta^{(m+1)}$ o valor de β na iteração $m+1$ e, substituindo $\beta^{(0)}$ da Equação C.12 por $\beta^{(m)}$, obtém-se:

$$\beta^{(m+1)} \cong \beta^{(m)} + \{-U'(\beta^{(m)})\}^{-1} U(\beta^{(m)}) \quad \text{Eq. C.13}$$

com $m = 0, 1, \dots$. A matriz $-U'(\beta)$ pode não ser positiva, assim, pode ser mais conveniente substituir a matriz $-U'(\beta)$ pelo valor esperado correspondente. Ou seja, a Equação C.13 torna-se:

$$\beta^{(m+1)} \cong \beta^{(m)} + K^{-1}(\beta^{(m)}) U(\beta^{(m)}) \quad \text{Eq. C.14}$$

Finalmente, substituindo os valores da função escore e da matriz de informação de Fisher (Equações C.9 e C.10) na Equação C.14 tem-se:

$$\beta^{(m+1)} \cong \left(X^T W^{(m)} X \right)^{-1} X^T W^{(m)} z^{(m)} \quad \text{Eq. C.15}$$

onde $z = \eta + W^{-1/2} V^{-1/2} (y - \mu)$. Nos MLG z é uma variável resposta modificada, ou seja, é uma forma linearizada da função de ligação aplicada para y . A convergência da Equação C.15 ocorre quando as mudanças forem suficientemente pequenas, ou seja, quando $|\beta^{(m+1)} - \beta^{(m)}| < \varepsilon$ (McCullagh e Nelder, 1989).

A convergência incide em um número finito de passos, independente dos valores iniciais utilizados, porém é usual iniciar com $\eta^{(0)} = g(y)$.

Livros Grátis

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

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

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

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

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