

Isabelle de Oliveira Pinto

**Associação entre poluição atmosférica de
queimadas e internação por doenças
cardiovasculares**

Niterói - RJ, Brasil

2 de Fevereiro de 2022

Isabelle de Oliveira Pinto

**Associação entre poluição
atmosférica de queimadas e
internação por doenças
cardiovasculares**

Trabalho de Conclusão de Curso

Monografia apresentada para obtenção do grau de Bacharel em
Estatística pela Universidade Federal Fluminense.

Orientador(a): Profa. Dra. Ludmilla Viana Jacobson

Niterói - RJ, Brasil

2 de Fevereiro de 2022

Isabelle de Oliveira Pinto

**Associação entre poluição atmosférica de
queimadas e internação por doenças
cardiovasculares**

Monografia de Projeto Final de Graduação sob o título “*Associação entre poluição atmosférica de queimadas e internação por doenças cardiovasculares*”, defendida por Isabelle de Oliveira Pinto e aprovada em 2 de Fevereiro de 2022, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Profa. Dra. Ludmilla Viana Jacobson
Departamento de Estatística – UFF

Profa. Dra. Ana Beatriz Monteiro Fonseca
Departamento de Estatística – UFF

Profa. Dr. Rafael Santos Erbisti
Departamento de Estatística – UFF

Niterói, 2 de Fevereiro de 2022

Ficha catalográfica automática - SDC/BIME
Gerada com informações fornecidas pelo autor

P659a Pinto, Isabelle de Oliveira
Associação entre poluição atmosférica de queimadas e
internação por doenças cardiovasculares / Isabelle de
Oliveira Pinto ; Ludmilla Viana Jacobson, orientadora.
Niterói, 2022.
44 f. : il.

Trabalho de Conclusão de Curso (Graduação em
Estatística)-Universidade Federal Fluminense, Instituto de
Matemática e Estatística, Niterói, 2022.

1. Queimada. 2. Doenças cardiovasculares. 3. Amazônia
legal. 4. Poluição atmosférica. 5. Produção intelectual.
I. Jacobson, Ludmilla Viana, orientadora. II. Universidade
Federal Fluminense. Instituto de Matemática e Estatística.
III. Título.

CDD -

Resumo

O objetivo principal deste projeto é avaliar o efeito da poluição atmosférica no período de seca sobre as internações por doenças cardiovasculares nos municípios da Amazônia Legal. A área de estudo é definida pelos municípios que compõem a Amazônia Legal e o período limitado ao ano de 2019, nos meses de Maio, Junho, Julho, Agosto, Setembro e Outubro. A escolha desses meses é devido a ocorrência de queimadas, quando geralmente se observa um aumento nas internações hospitalares. Foram utilizados os dados do SISAM, sobre as concentrações de material particulado fino $PM_{2.5}$, temperatura e umidade, por mês e município. Para o desfecho de saúde foram utilizados os dados disponíveis no DATASUS, por meio do Sistema de Internações Hospitalares do SUS (SIH/SUS). Na análise dos dados foi estimado um Modelo de Regressão Linear Múltipla. A variável resposta do modelo é o log da taxa de internação e as variáveis explicativas são as variáveis ambientais, que são temperatura, umidade e concentração de $PM_{2.5}$, além dos meses e interações das variáveis ambientais. Como objetivos específicos pretende-se: (i) Avaliar se existe diferença entre as concentrações médias das variáveis ambientais entre os meses de Maio à Outubro nos municípios da Amazônia Legal; (ii) Estimar o efeito das variáveis ambientais na taxa de internação a partir de Modelo de Regressão Linear Múltipla; (iii) Avaliar o efeito de interação entre $PM_{2.5}$ e umidade na taxa de internação; (iv) Avaliar o efeito de interação entre $PM_{2.5}$ e temperatura na taxa de internação; (v) Avaliar o efeito de interação entre umidade e temperatura na taxa de internação. Após todas as análises e estimação do modelo ficou evidente a necessidade de estudar as 3 variáveis ambientais conjuntamente. Além de ser observado que quando a temperatura alta é esperado um aumento na taxa de internação, o mesmo ocorre quando a umidade é baixa e concentração de $PM_{2.5}$ alta, quando avaliados separadamente. Já em relação ao efeito de interação tripla observou-se que o efeito esperado do material particulado fino na taxa de internação é um aumento de 34,23%, quando a umidade baixa e temperatura alta. Todas as análises foram realizadas no programa estatístico R.

Palavras-chave: Poluição atmosférica. Queimadas. Doenças cardiovasculares. Modelo de Regressão Linear Múltipla.

Agradecimentos

Agradeço primeiramente à Deus pela oportunidade de cursar e finalizar Estatística na Universidade Federal Fluminense.

Agradeço aos meus pais e meu padrasto pelo apoio e incentivo aos meus estudos. Também aos meus irmãos por mesmo me estressando trazerem muitas alegrias na minha vida.

Agradeço ao meu marido por sempre estar ao meu lado, me consolar nos momentos difíceis e ter muita paciência nos meus momentos de desespero. Também à Masha, que chegou durante a pandemia para somar muito na minha vida, e ao Pupilo que esteve presente em toda essa jornada sendo sempre muito companheiro.

Agradeço aos meus amigos que conheci na UFF, como a Larissa Estudante, Maqueise, Gabriel, Pedro e Felipe por tantos momentos de alegria que passamos juntos, como as tardes conversando, os almoços e as idas ao cinema. Especialmente agradeço à Larissa Pires Machado, pela amizade que com certeza levaremos para toda a vida e por toda caminhada que vivenciamos juntas até aqui. Também aos meus amigos da vida que estiveram presente nesta etapa.

Agradeço à professora Ludmilla pela orientação deste trabalho, pelos incentivos, as conversas e também por fazer com que essa etapa difícil fosse leve e agradável. Também agradeço à todos os professores e profissionais que estiveram presente na minha caminhada pela Universidade Federal Fluminense.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 10
1.1	Objetivos	p. 12
1.1.1	Objetivo Geral	p. 12
1.1.2	Objetivos Específicos	p. 12
1.2	Organização	p. 12
2	Materiais e Métodos	p. 13
2.1	Material	p. 13
2.1.1	Local e Período de Estudo	p. 13
2.1.2	Dados de Saúde	p. 14
2.1.3	Dados do SISAM	p. 14
2.1.4	Banco de Dados	p. 14
2.2	Métodos	p. 14
2.2.1	Análise Exploratória	p. 14
2.2.2	Modelo de Regressão Linear Múltiplo	p. 16
2.2.2.1	Estimação do Modelo	p. 18
2.2.2.2	Resíduos e valores ajustados	p. 19
2.2.2.3	Teste de Hipótese	p. 20
2.2.2.4	Adequação do modelo	p. 21

2.2.2.5	Pressupostos do Modelo	p. 22
2.2.2.6	Variáveis dummy	p. 23
2.2.2.7	Interações	p. 24
3	Análise dos Resultados	p. 25
3.1	Análise Exploratória	p. 25
3.2	Modelo de Regressão Linear	p. 31
3.2.1	Transformação das variáveis	p. 31
3.2.2	Interação	p. 33
3.2.3	Escolha do melhor modelo	p. 34
4	Conclusões	p. 37
	Referências	p. 38
	Apêndice 1 – Boxplots variáveis ambientais	p. 40
	Apêndice 2 – Primeiro modelo	p. 42
	Apêndice 3 – Modelos	p. 43

Lista de Figuras

1	Mapa da Amazônia Legal, Brasil.	p. 13
2	Dendrograma exemplo	p. 16
3	Boxplot da média de $PM_{2.5}$ em cada mês. Amazônia Legal, 2019.	p. 25
4	Dendrograma de todas as variáveis. Amazônia Legal, 2019.	p. 26
5	Mapas da Concentração Média de $PM_{2.5}$ em cada mês. Amazônia Legal, 2019.	p. 27
6	Mapas da Temperatura Máxima em cada mês. Amazônia Legal, 2019.	p. 28
7	Mapas da Umidade Mínima relativa em cada mês. Amazônia Legal, 2019.	p. 29
8	Mapas da Taxa de Internação em cada mês. Amazônia Legal, 2019.	p. 30
9	Histograma do Log da Taxa de Internação. Amazônia Legal, 2019.	p. 32
10	QQ-plot dos modelos. Amazônia Legal, 2019.	p. 32
11	Interação entre as variáveis ambientais. Amazônia Legal, 2019.	p. 33
12	QQ-plot e histograma dos resíduos do Modelo 1. Amazônia Legal, 2019.	p. 34
13	Efeito do Log da Taxa de Internação em cada mês e interações. Amazônia Legal, 2019. ²	p. 36
14	Boxplot da média de $PM_{2.5}$ em cada mês por Estado. Amazônia Legal, 2019.	p. 40
15	Boxplot da média da temperatura em cada mês por Estado. Amazônia Legal, 2019.	p. 41
16	Boxplot da média da umidade em cada mês por Estado. Amazônia Legal, 2019.	p. 41

Lista de Tabelas

1	Descrição das variáveis	p. 15
2	Correlação de Spearman da Taxa de Internação com as variáveis ambientais. Amazônia Legal, 2019.	p. 26
3	Efeito de interação das variáveis ambientais. Amazônia Legal, 2019. . .	p. 33
4	Modelos estimados. Amazônia Legal, 2019.	p. 34
5	Estimativas dos parâmetros do Modelo Final. Amazônia Legal, 2019. ¹	p. 35
6	Estimativas dos parâmetros do modelo. Amazônia Legal, 2019.	p. 42
7	Estimativas dos parâmetros do Modelo 2. Amazônia Legal, 2019. . . .	p. 43
8	Estimativas dos parâmetros do Modelo 3. Amazônia Legal, 2019. . . .	p. 43
9	Estimativas dos parâmetros do Modelo 4. Amazônia Legal, 2019. . . .	p. 44
10	Estimativas dos parâmetros do Modelo 5. Amazônia Legal, 2019. . . .	p. 44

1 Introdução

Neste trabalho foi abordado a possível associação entre poluição atmosférica, causada pelas queimadas, e internação por doenças cardiovasculares no ano de 2019 na Amazônia Legal. A Amazônia Legal é composta pelos Estados Acre, Amapá, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia, Roraima e Tocantins. Foram escolhidos os meses de Maio à Outubro pois são os meses com menor índice de umidade e o ano abordado foi selecionado devido ao aumento de aproximadamente 45,1% de focos de queimadas em relação ao ano anterior, 2018. No ano de 2019 os 6 Estados brasileiros com maior porcentagem de focos de queimadas pertencem à Amazônia Legal, sendo o primeiro lugar ocupado pelo Mato Grosso (INPE, 2019). As doenças do aparelho circulatório foram escolhidas para este trabalho por serem a principal causa de morte em todo mundo nos últimos 20 anos (OPAS, 2020).

Entre as principais doenças cardiovasculares estão doença coronariana (doença dos vasos sanguíneos que irrigam o músculo cardíaco), doença cerebrovascular (doença dos vasos sanguíneos que irrigam o cérebro), doença arterial periférica (doença dos vasos sanguíneos que irrigam os membros superiores e inferiores), doença cardíaca reumática (danos nos músculos do coração e válvulas cardíacas devido à febre reumática, causada por bactérias estreptocócicas), cardiopatia congênita (malformações na estrutura do coração) e trombose venosa profunda e embolia pulmonar (coágulos sanguíneos nas veias das pernas) (OMS, 2021).¹

Foi lançado no ano de 2004 a primeira declaração científica, pela *American Heart Association*, sobre poluição atmosférica e doenças cardiovasculares, onde foi concluído que a exposição ao material particulado contribui para morbidade e mortalidade por doenças cardiovasculares. Uma revisão sistemática mostrou que muitos estudos concordam com a ideia que a exposição prolongada ao $PM_{2.5}$ aumenta o risco de morte por doenças cardiovasculares, reduzindo a expectativa de vida da população (BROOK et al., 2010).

¹Tradução livre do texto da OMS.

O $PM_{2.5}$, também conhecido como material particulado fino, tem se mostrado um indicador robusto de risco à saúde. Ele descreve partículas finas inaláveis, com diâmetro de até $2,5\mu m$ (EPA, 2021). O limite diário de exposição ao material particulado fino, em áreas urbanas, é de $25\mu g/m^3$ e o limite para a média anual é de $10\mu g/m^3$, de acordo com a OMS (LIM et al., 2013). Um estudo sobre a exposição à longo prazo ao $PM_{2.5}$ concluiu que essa exposição, durando mais de 16 anos, é um fator de risco para mortalidade por doenças cardiopulmonares e câncer de pulmão (III et al., 2002).

Um estudo feito no ano de 2005, sobre a mortalidade por doenças circulatórias nos idosos devido à exposição de $PM_{2.5}$ na Amazônia brasileira, mostrou que a taxa de mortalidade média era 1,58 por 100 habitantes. Os resultados desse estudo mostraram que a mortalidade por doenças cardiovasculares aumenta devido à exposição, acima do limite diário recomendado, ao material particulado fino. A qualidade do ar na Amazônia é altamente comprometida durante o período de seca que é quando tem um aumento no número de queimadas. Mais de 65% do material particulado, gerado pela queima de biomassa, é feito de partículas finas e ultrafinas, que podem ser prejudiciais à saúde humana (REAL; IGNOTTI; HACON, 2013).

De acordo com um estudo sobre variabilidade do clima, morbidade e mortalidade associados com material particulado foi dito que o clima pode influenciar na associação do $PM_{2.5}$ com morbidade e mortalidade. Notou-se que mesmo a temperatura sendo um parâmetro muito estudado alguns autores observaram que umidade, pressão atmosférica e as estações do ano também podem ser um fator que influencia no efeito do $PM_{2.5}$ na saúde. Foi concluído, nesse estudo, que a temperatura média e máxima tem correlação direta com $PM_{2.5}$ e que a umidade possui alta correlação inversa com a temperatura máxima (RODRIGUES et al., 2017).

O impacto causado pelas queimadas no ecossistema amazônico durante o período de seca foi abordado em um estudo feito nos anos de 2001 à 2012, onde o foco foi a população acima de 60 anos, no Estado de Rondônia. Nesse estudo os autores observaram que entre as causas de óbito a que teve maior prevalência foram as doenças cardiovasculares, com 35% do total de óbitos. Os dois tipos mais frequentes de doenças cardiovasculares observados foram acidente vascular cerebral e infarto agudo do miocárdio, com 22 e 21% dos diagnósticos, respectivamente (ANDRADE FILHO et al., 2017).

As discussões neste trabalho visam entender o efeito do $PM_{2.5}$ na taxa de internação por doenças cardiovasculares, através de um Modelo de Regressão Linear Múltipla. Sua importância é devido ao fato de ter poucos estudos sobre o tema no ano de 2019, onde

houve um aumento expressivo no número de focos de queimadas na Amazônia Legal.

1.1 **Objetivos**

1.1.1 **Objetivo Geral**

Verificar associação entre poluição atmosférica de queimadas ($PM_{2.5}$) e internações por doenças cardiovasculares na Amazônia Legal, no ano de 2019 e suas interações com as variáveis temperatura e umidade.

1.1.2 **Objetivos Específicos**

- Avaliar se existe diferença entre as concentrações médias das variáveis ambientais entre os meses de Maio à Outubro nos municípios da Amazônia Legal.
- Estimar o efeito das variáveis ambientais na taxa de internação a partir de Modelo de Regressão Linear Múltipla.
- Avaliar o efeito de interação entre $PM_{2.5}$ e umidade na taxa de internação.
- Avaliar o efeito de interação entre $PM_{2.5}$ e temperatura na taxa de internação.
- Avaliar o efeito de interação entre umidade e temperatura na taxa de internação.

1.2 **Organização**

Este trabalho está organizado da seguinte forma, Capítulo 1 contém a motivação para o desenvolvimento do trabalho, revisão bibliográfica e objetivos. O Capítulo 2 contém a descrição dos materiais e métodos que foram utilizados no estudo. Por fim, nos Capítulos 3 e 4, respectivamente, estão presentes os resultados das análises e conclusões.

2 Materiais e Métodos

O capítulo de Materiais e Métodos está dividido em duas partes, na qual a primeira contém características sobre os dados, local de estudo e o banco de dados. A segunda seção é composta pela metodologia que foi utilizada.

2.1 Material

2.1.1 Local e Período de Estudo

A área de estudo é a Amazônia Legal, que ocupa cerca de 58,9 % do território brasileiro e é composta pelos Estados do Acre, Amapá, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia, Roraima e Tocantins. Foram selecionados para este estudo os indivíduos internados por doenças cardiovasculares residentes nos municípios da Amazônia Legal no ano de 2019, nos meses Maio, Junho, Julho, Agosto, Setembro e Outubro, que fazem parte do período de seca (IBGE, 2021). Na Figura 1 contém o mapa da Amazônia Legal.

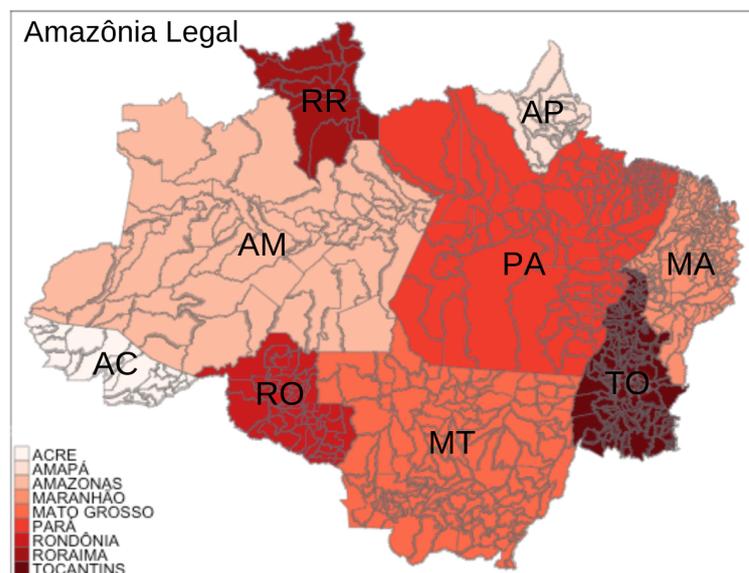


Figura 1: Mapa da Amazônia Legal, Brasil.

2.1.2 Dados de Saúde

Os dados sobre as internações por doenças cardiovasculares foram obtidos através do site do Departamento de Informática do SUS (DATASUS), por meio do Sistema de Internações Hospitalares (SIH/SUS) e código CID-10 I00-I99, capítulo IX, Doenças do aparelho circulatório. A variável internação é o número de internações em cada município, por mês (DATASUS, 2019).

2.1.3 Dados do SISAM

Os dados sobre umidade, temperatura e concentração de material particulado fino ($PM_{2.5}$) foram obtidos no site do Sistema de Informações Ambientais Integrado a Saúde (SISAM), de cada município e mês deste trabalho. Os dados estavam na base constando dia e horário e a partir disso foram criadas novas variáveis como a média, mínimo, máximo e amplitude mensais de cada variável ambiental por município, também a variável quantil 90 que foi criada a partir dos dados diários sobre temperatura, com o objetivo de verificar os valores observados em 90% das observações (SISAM, 2019).

2.1.4 Banco de Dados

Todas as análises e modificações no banco de dados foram feitas utilizando o programa RStudio (R Core Team, 2014). A partir da variável internação e da população estimada no ano de 2019 foi criada uma nova variável denominada Taxa de Internação, os valores da população estimada foram obtidos através do site do IBGE, onde foram utilizados os dados sobre projeção da população no ano de 2019. A Amazônia Legal possui 9 Estados contendo um total de 772 municípios e a base de dados possui um total de 4632 observações.

2.2 Métodos

2.2.1 Análise Exploratória

Para avaliar a relação entre as variáveis x e y utilizamos a correlação de Spearman. O valor do coeficiente deve estar no intervalo $[-1, 1]$, onde -1 significa maior correlação negativa enquanto o valor 1 significa maior correlação positiva, baseando-se então na proximidade ao zero onde indica que as variáveis não possuem correlação. O coeficiente

Tabela 1: Descrição das variáveis

Variável	Descrição
Internação	número de internações em cada município, por mês
Taxa de Internação	$\frac{\text{número de internações}}{\text{população estimada}} \times 100000$
Média da Temperatura	Temperatura média por mês
Mínimo da Temperatura	Temperatura mínima por mês
Máximo da Temperatura	Temperatura máxima por mês
Amplitude da Temperatura	Temperatura máxima – Temperatura mínima, por mês
Quantil 90 da Temperatura	Quantil 90 da Temperatura por mês
Média da Umidade	Umidade média por mês
Mínimo da Umidade	Umidade mínima por mês
Máximo da Umidade	Umidade máxima por mês
Amplitude da Umidade	Umidade máxima – Umidade mínima, por mês
Média de $PM_{2.5}$	Concentração média de $PM_{2.5}$ por mês
Mínimo de $PM_{2.5}$	Concentração mínima de $PM_{2.5}$ por mês
Máximo de $PM_{2.5}$	Concentração máxima de $PM_{2.5}$ por mês
Amplitude de $PM_{2.5}$	Máximo de $PM_{2.5}$ – Mínimo de $PM_{2.5}$, por mês

de correlação de Spearman equivale ao coeficiente de correlação de Pearson, que é escrito da seguinte forma:

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}} = r_s \quad (2.1)$$

Onde os dados de x e y são convertidos em postos (HOLLANDER; CHICKEN, 2014).

Com o intuito de verificar a significância da estimativa utilizamos o seguinte Teste de Hipóteses:

$$H_0 : \rho = 0 \quad H_1 : \rho \neq 0 \quad (2.2)$$

A estatística de teste definida como

$$t_r = r_s \sqrt{\frac{n-2}{1-r_s^2}} \quad , \quad (2.3)$$

tem uma distribuição T-Student, com $n - 2$ graus de liberdade. E sua região crítica é dada por:

$$|t_r| > t_{1-\frac{\alpha}{2}; n-2}, \quad (2.4)$$

se o valor de t_r pertencer à região crítica, então H_0 é rejeitado.

A função de densidade da distribuição T-Student, com v graus de liberdade, está definida no teorema abaixo.

Teorema 2.2.1. *Seja Z uma v.a. $N(0, 1)$ e Y uma v.a. $\chi^2(v)$, com Z e Y independentes. Então, a v.a.*

$$t = \frac{Z}{\sqrt{\frac{Y}{v}}}, \quad (2.5)$$

tem densidade dada por

$$f(t; v) = \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{\pi v}} \left(1 + \frac{t^2}{v}\right)^{-\frac{v+1}{2}}, \quad -\infty < t < \infty \quad (2.6)$$

Em seguida foi utilizado um diagrama chamado dendrograma, que é criado tratando cada objeto como cluster e unindo os dois itens com menor distância para verificar proximidade entre as variáveis. A Figura 2 apresenta um exemplo do dendrograma.

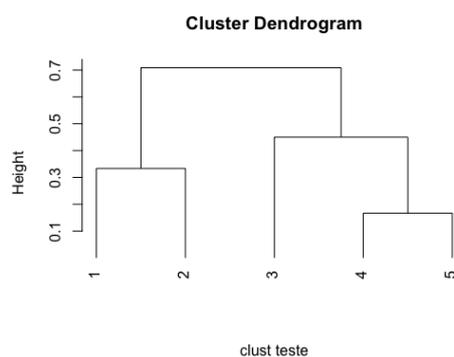


Figura 2: Dendrograma exemplo

O método para construção do dendrograma foi o método de Ward, que é um procedimento de agrupamento hierárquico, nele a medida de distância entre dois clusters é dada pela soma das distâncias ao quadrado entre os dois clusters. Este método consiste em minimizar a perda de informação, juntando dois grupos (JOHNSON; WICHERN, 2007).

Além disso, também foram produzidos diversos mapas coropléticos, boxplots e gráficos de dispersão.

2.2.2 Modelo de Regressão Linear Múltiplo

Para estudar a relação das variáveis foi utilizado o Modelo de Regressão Linear Múltiplo, onde sua definição é apresentada em Definição 2.2.1.

Definição 2.2.1. *O Modelo de Regressão Linear Múltiplo Normal é o modelo que define uma relação estatística linear ou não linear entre a variável resposta y e $p - 1$ variáveis independentes: x_1, x_2, \dots, x_{p-1} . A suposição básica desse modelo é que a média da distribuição de y varia de forma linear com as variáveis x_1, x_2, \dots, x_{p-1} . Essa relação pode ser estabelecida por:*

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_{p-1} x_{i,p-1} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2) \text{ são v.a. independentes} \quad (2.7)$$

onde,

- $x_{i,j}$ é o valor da j -ésima variável independente na i -ésima observação
- y_i é o valor da variável resposta na i -ésima observação
- ϵ_i o erro aleatório para a i -ésima observação
- $\beta_0, \beta_1, \beta_2, \dots, \beta_{p-1}$ e σ^2 são parâmetros do modelo

Costuma-se utilizar no Modelo de Regressão Linear Múltiplo a notação matricial, que é escrita na forma:

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{1,1} + \beta_2 x_{1,2} + \dots + \beta_{p-1} x_{1,p-1} + \epsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{2,1} + \beta_2 x_{2,2} + \dots + \beta_{p-1} x_{2,p-1} + \epsilon_2 \\ y_3 = \beta_0 + \beta_1 x_{3,1} + \beta_2 x_{3,2} + \dots + \beta_{p-1} x_{3,p-1} + \epsilon_3 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n,1} + \beta_2 x_{n,2} + \dots + \beta_{p-1} x_{n,p-1} + \epsilon_n \end{cases} \quad (2.8)$$

Podendo ser escrito da maneira abaixo (KUTNER et al., 2004).

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,p-1} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,p-1} \\ 1 & x_{3,1} & x_{3,2} & \cdots & x_{3,p-1} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,p-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{p-1} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{pmatrix} \quad (2.9)$$

O modelo geral também pode ser definido como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2 I) \quad (2.10)$$

Onde

- n é o número total de observações da amostra;
- p é o número de parâmetros do tipo β ;
- \mathbf{Y} é o vetor aleatório de dimensão n ;
- \mathbf{X} é a matriz $n \times p$ de valores constantes;
- β é o vetor de dimensão p de parâmetros desconhecidos;
- ϵ é o vetor aleatório de dimensão n .

2.2.2.1 Estimação do Modelo

A fim de encontrar os estimadores para os parâmetros e a função de regressão estimada pode-se utilizar o método dos Estimadores por Mínimos Quadrados ou por Máxima Verossimilhança. O estimador encontrado a partir dos Mínimos Quadrados coincide com o estimador encontrado a partir da Máxima Verossimilhança (KUTNER et al., 2004).

- Estimadores por Mínimos Quadrados

O estimador para β pelo método dos Mínimos Quadrados minimiza a soma de quadrados dos erros (KUTNER et al., 2004). A equação dos estimadores por Mínimos Quadrados para o modelo de regressão linear é:

$$\mathbf{X}^T \mathbf{X} \beta = \mathbf{X}^T \mathbf{Y} \quad (2.11)$$

E temos o estimador $\hat{\beta}$ pelo método dos Mínimos Quadrados na Equação 2.12.

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (2.12)$$

O desvio de \mathbf{Y} deve ser calculado em torno de sua própria média estimada. Sendo assim:

$$\epsilon = \mathbf{Y} - \mathbf{X} \hat{\beta} \quad (2.13)$$

e a soma dos quadrados denotada por SQE.

$$SQE = \mathbf{Y}^T \mathbf{Y} - \hat{\beta} \mathbf{X}^T \mathbf{Y} \quad (2.14)$$

A soma dos quadrados possui $n - p$ graus de liberdade. Então, a média dos quadrados apropriada é denotada por MQE .

$$MQE = \frac{SQE}{n - p} \quad (2.15)$$

Temos então que MQE é o estimador não-viesado de σ^2 para o modelo de regressão (KUTNER et al., 2004).

$$\sigma^2 = E(MQE) \quad (2.16)$$

2.2.2.2 Resíduos e valores ajustados

Na regressão múltipla o valor ajustado para a i -ésima observação é definido na equação 2.17 e o i -ésimo resíduo é dado na equação 2.18.

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i,1} + \hat{\beta}_2 x_{i,2} + \cdots + \hat{\beta}_{p-1} x_{i,p-1} \quad (2.17)$$

$$e_i = y_i - \hat{y}_i \quad (2.18)$$

Temos que

$$\hat{\mathbf{Y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \\ \vdots \\ \hat{y}_n \end{pmatrix} \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{pmatrix} \quad (2.19)$$

O vetor com os n valores ajustados e o vetor com os n resíduos estão representados nas equações 2.20 e 2.21, respectivamente.

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \quad (2.20)$$

$$\boldsymbol{\epsilon} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}} \quad (2.21)$$

2.2.2.3 Teste de Hipótese

Para analisar a significância de β foi realizado o Teste de Hipóteses. Foi construído um Teste de Hipóteses separadamente para cada β_k . As hipóteses do teste são dadas a seguir:

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

A estatística de teste é :

$$t_{0k} = \frac{\hat{\beta}_k}{\sqrt{MQE \times C_{kk}}} \quad (2.22)$$

- t segue uma distribuição T-Student com $n - p$ graus de liberdade
- n é o total de observações da amostra
- $k = 1, \dots, p - 1$
- $SQE = \mathbf{Y}^T \mathbf{Y} - \hat{\beta}^T \mathbf{X}^T \mathbf{Y}$
- $MQE = \frac{SQE}{n-p}$
- C_{kk} é o elemento da diagonal principal da matriz $(\mathbf{X}^T \mathbf{X})^{-1}$

A região crítica é dada por:

$$|t_{0k}| > t_{1-\frac{\alpha}{2}; n-p} \quad (2.23)$$

Se o valor de t_{0k} pertencer à região crítica, então H_0 é rejeitado.

Temos que o intervalo de confiança de $100(1 - \alpha)\%$, para cada β_k é dado na equação 2.24 (MONTGOMERY; PECK; VINING, 2012).

$$IC_{100(1-\alpha)\%}(\beta_k) = [\hat{\beta}_k \pm t_{1-\frac{\alpha}{2}; n-p} \sqrt{MQE \times C_{kk}}] \quad (2.24)$$

2.2.2.4 Adequação do modelo

Para verificar a adequação do modelo utilizou-se algumas medidas como o teste ANOVA, R^2 ajustado e o AIC, que é uma medida de comparação de modelos.

- ANOVA

Com a finalidade de testar a significância da regressão foi utilizado o teste para determinar se é existente relação entre a variável resposta e as covariáveis do modelo. As hipóteses deste teste seguem abaixo.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$$

$$H_1 : \beta_k \neq 0, \text{ para pelo menos um } k$$

Caso a hipótese nula seja rejeitada implicará que pelo menos uma das covariáveis contribui significativamente para o modelo. O procedimento que é realizado neste teste é uma generalização da Análise de Variância utilizada no modelo de regressão linear simples. Temos então que

$$SQT = SQR + SQE \quad (2.25)$$

$$F_0 = \frac{\frac{SQR}{p-1}}{\frac{SQE}{n-p}} = \frac{MQR}{MQE} \sim F_{p-1, n-p} \quad (2.26)$$

onde H_0 é rejeitado caso $F_0 > F_{1-\alpha, p-1, n-p}$ (MONTGOMERY; PECK; VINING, 2012).

- R^2 e R^2 ajustado

O R^2 é conhecido como coeficiente de Determinação e ele pode ser utilizado também para verificar a adequabilidade do modelo. Ele é utilizado como medida descritiva da qualidade do ajuste, e é definido como:

$$R^2 = 1 - \frac{SQE}{SQT} \quad (2.27)$$

De maneira geral o R^2 aumenta quando é adicionada um variável no modelo, entretanto isso não significa necessariamente que houve uma melhora no modelo (MONTGOMERY; PECK; VINING, 2012). A partir da tentativa de corrigir os problemas mencionados anteriormente foi definido o R^2 ajustado, onde:

$$R_a^2 = 1 - \frac{\frac{SQE}{n-p}}{\frac{SQT}{n-1}}. \quad (2.28)$$

- AIC

A equação do AIC é dada por:

$$AIC = 2p - 2 \log L(\hat{\beta}), \quad (2.29)$$

onde $L(\hat{\beta})$ é a função de verossimilhança e p é a quantidade de parâmetros que o modelo possui. O AIC é uma ferramenta que é utilizada na comparação de modelos, sendo melhor menores valores de AIC para o modelo (MONTGOMERY; PECK; VINING, 2012).

2.2.2.5 Pressupostos do Modelo

O modelo deve seguir alguns pressupostos, os erros devem:

- ter média zero e variância constante;
- ser não correlacionados;
- seguir uma distribuição normal;
- ser independentes.

Para verificar os pressupostos do modelo foram feitos testes e análises gráficas. Com o objetivo de verificar a normalidade dos erros utilizou-se o Teste de Normalidade onde as hipóteses testadas foram:

H_0 : os erros seguem uma distribuição normal

H_1 : os erros não seguem uma distribuição normal.

O teste utilizado para avaliar a normalidade foi o Teste Shapiro. No programa R(R Core Team, 2014) foi utilizada a função *shapiro.test* para realizar o teste e *qqnorm* para a construção do gráfico QQ-plot.

A fim de verificar a Homocedasticidade dos erros, foi realizado uma análise gráfica e utilizado o teste de Breush-Pagan, onde as hipóteses testadas seguem abaixo. A função utilizada nesse teste foi *bptest*.

H_0 : a variância dos erros é constante
 H_1 : a variância dos erros não é constante

Além dos testes citados anteriormente, também foi executado o teste de Durbin-Watson para avaliar se os erros são autocorrelacionados, a função utilizada no software foi *dwttest*. As hipóteses testadas seguem abaixo:

$$\begin{aligned} H_0: \rho &= 0 \\ H_1: \rho &\neq 0 \end{aligned}$$

Quando a hipótese nula do teste é rejeitada, há evidências de que os erros são autocorrelacionados.

Como neste trabalho não foi possível satisfazer os pressupostos do Modelo de Regressão Linear Múltiplo de variância dos erros constante, normalidade e autocorrelação dos erros foi necessário fazer uma transformação nos dados. O procedimento apropriado para realizar a transformação dos dados é:

$$y^\lambda = \begin{cases} \frac{y^\lambda - 1}{\lambda y^{\lambda-1}}, & \lambda \neq 0 \\ y' \ln y, & \lambda = 0 \end{cases} \quad (2.30)$$

Onde λ é o ponto que maximiza a função de Verossimilhança e $y' = \ln[\frac{1}{n} \sum_{i=1}^n \ln y_i]$ é a média geométrica das observações. É possível achar um intervalo de confiança aproximado para a transformação do parâmetro λ . A transformação converte a variável resposta de Y para Y^λ (MONTGOMERY; PECK; VINING, 2012).

$$\mathbf{Y}^\lambda = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2 I) \quad (2.31)$$

Quando λ próximo de 0, a equação utilizada no modelo passa a ser:

$$\ln \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma^2 I). \quad (2.32)$$

2.2.2.6 Variáveis dummy

As variáveis dummy são variáveis indicadoras que representam uma variável categórica. Segue um exemplo:

$$Z = \begin{cases} 1, & \text{se a característica estiver presente} \\ 0, & \text{caso contrário.} \end{cases} \quad (2.33)$$

As demais categorias da variável categórica são comparadas com a categoria de referência, que é representada por 0 (MONTGOMERY; PECK; VINING, 2012).

2.2.2.7 Interações

O modelo de regressão linear múltipla com interações, utilizado neste trabalho, contém 3 interações duplas e 1 interação tripla.

Para verificar a significância das interações foi utilizado o Teste F, através da função *anova* do programa R (R Core Team, 2014). Se rejeitada a hipótese nula do teste então há evidências que o efeito de interação é significativo para o modelo, mantendo então esse efeito no modelo (MONTGOMERY; PECK; VINING, 2012).

Para todas as análises feitas neste trabalho foi adotado nível de significância de 5%.

3 Análise dos Resultados

3.1 Análise Exploratória

Inicialmente foi feita a análise exploratória, com o objetivo de conhecer melhor os dados. Através da Figura 3 notou-se um valor observado muito discrepante dos demais na concentração média de $PM_{2.5}$ no mês de maio, atingindo aproximadamente $250\mu g/m^3$.

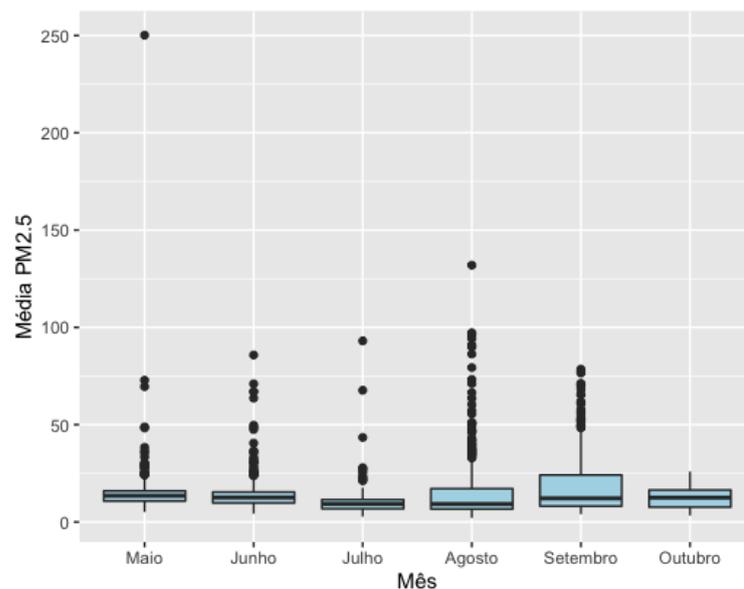


Figura 3: Boxplot da média de $PM_{2.5}$ em cada mês. Amazônia Legal, 2019.

Para verificar a proximidade entre as variáveis foi executado um dendrograma, que está na Figura 4. Podemos perceber que, no dendrograma, existe uma proximidade entre as variáveis Taxa de Internação, Média da Temperatura, Média de $PM_{2.5}$ e Máximo de $PM_{2.5}$. Também foi observado a proximidade entre as variáveis Máximo da Umidade e Mínimo de $PM_{2.5}$, indicando uma correlação inversa.

Posteriormente foi feita a correlação de Spearman para verificar a correlação entre as variáveis ambientais e a Taxa de Internação. As variáveis que apresentaram a maior correlação com a Taxa de Internação, por cada variável ambiental, foram Mínimo da

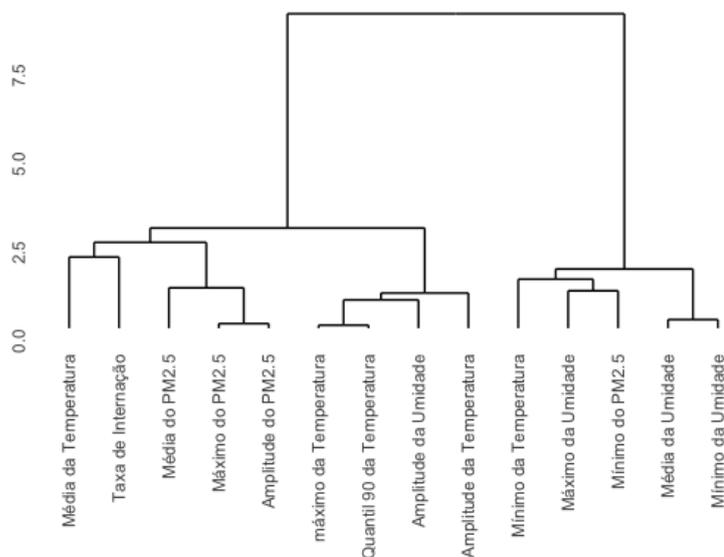


Figura 4: Dendrograma de todas as variáveis. Amazônia Legal, 2019.

Temperatura, Média da Umidade e Amplitude do $PM_{2.5}$. Os resultados constam na Tabela 2.

Tabela 2: Correlação de Spearman da Taxa de Internação com as variáveis ambientais. Amazônia Legal, 2019.

Variável	ρ	p-valor
Média da Temperatura	-0,1447	< 0,0001
Mínimo da Temperatura	-0,2847	< 0,0001
Máximo da Temperatura	0,0188	0,2017
Amplitude da Temperatura	0,1996	< 0,0001
Quantil 90 da Temperatura	0,0503	0,0007
Média da Umidade	-0,1719	< 0,001
Mínimo da Umidade	-0,1519	< 0,0001
Máximo da Umidade	-0,0886	< 0,0001
Amplitude da Umidade	0,1016	< 0,0001
Média de $PM_{2.5}$	0,0041	0,7798
Mínimo de $PM_{2.5}$	-0,0396	0,0071
Máximo de $PM_{2.5}$	0,0821	< 0,0001
Amplitude de $PM_{2.5}$	0,0840	< 0,0001

Além da Correlação de Spearman e do Dendrograma foram executados diversos mapas coropléticos para observar as variáveis ambientais em cada mês deste estudo. Para esta análise foram escolhidas as variáveis Média da Concentração de $PM_{2.5}$, Temperatura Máxima e Umidade Mínima devido a forte correlação entre essas variáveis. Além da forte correlação entre as variáveis, um estudo do INPE comprovou que "Temperatura máxima acima de 30°C e Umidade Relativa mínima abaixo de 40 % para o ar aumentam linear-

mente o Risco de Fogo no dia de interesse; abaixo e acima destes limites, respectivamente, diminuem o Risco de Fogo.” (SETZER; SISMANOGLU; SANTOS, 2019).

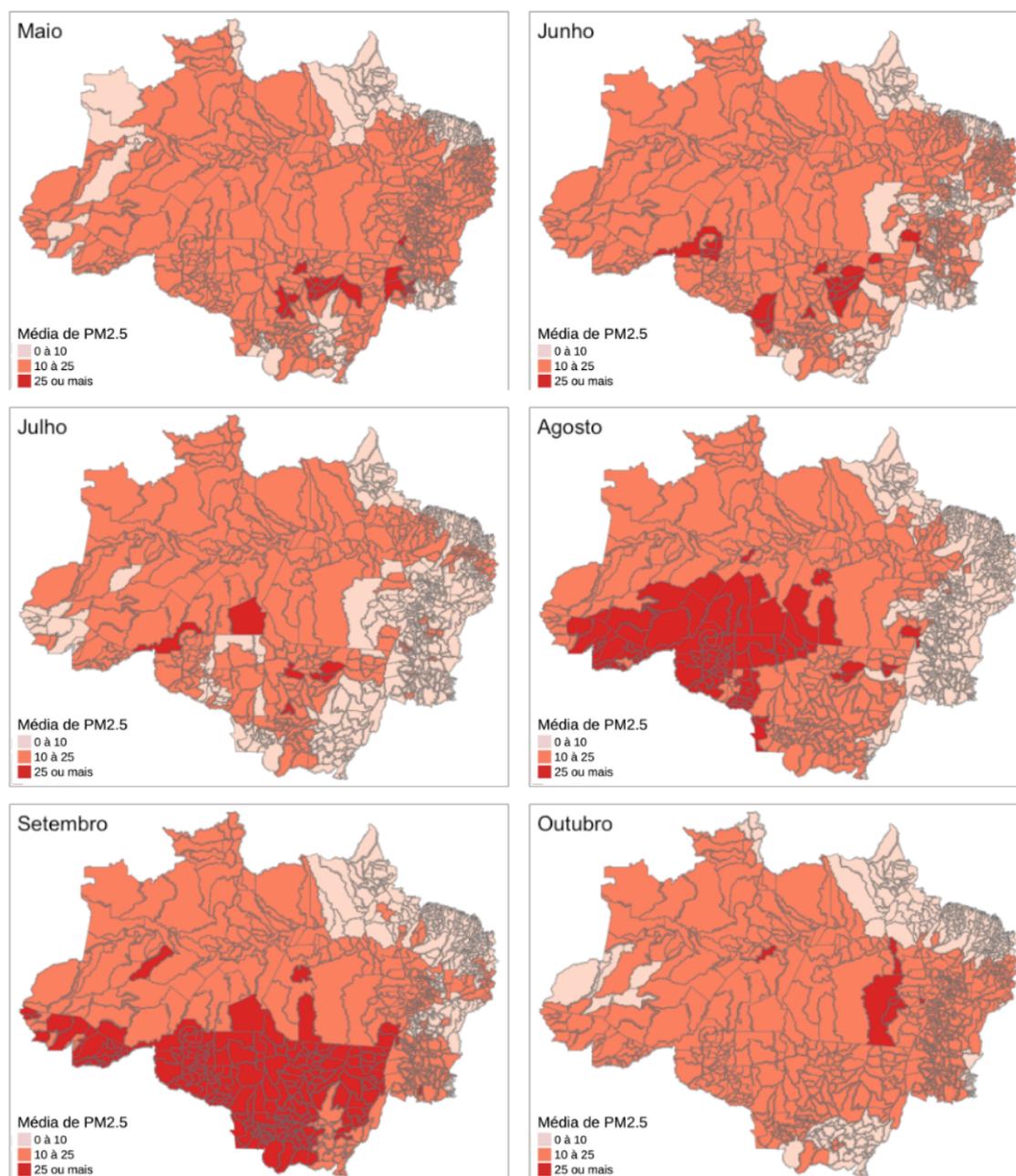


Figura 5: Mapas da Concentração Média de $PM_{2.5}$ em cada mês. Amazônia Legal, 2019.

A Figura 5 mostra os mapas da média da concentração de $PM_{2.5}$ em cada mês, o intervalo foi escolhido com base nas informações da OMS, onde até $10\mu g/m^3$ é o limite para a média anual de exposição ao material particulado fino, $25\mu g/m^3$ é o limite para a média diária de exposição e acima desse valor é altamente prejudicial à saúde. Podemos observar que nos meses de Maio à Outubro a maior parte da Amazônia Legal tem uma concentração média de material particulado acima de $10\mu g/m^3$, e nos meses de Agosto e

Setembro existe uma região onde a concentração média é acima de $25\mu\text{g}/\text{m}^3$.

Os mapas da temperatura máxima em cada mês estão na Figura 6, essa variável foi selecionada pois temperaturas altas aumentam o risco de queimadas. Podemos observar que nos meses de Maio à Outubro a maior parte da Amazônia Legal apresentou temperatura máxima acima de 30°C e nos meses de Agosto à Outubro mais de 50% dos municípios apresentaram temperatura máxima entre 35°C e 43°C .

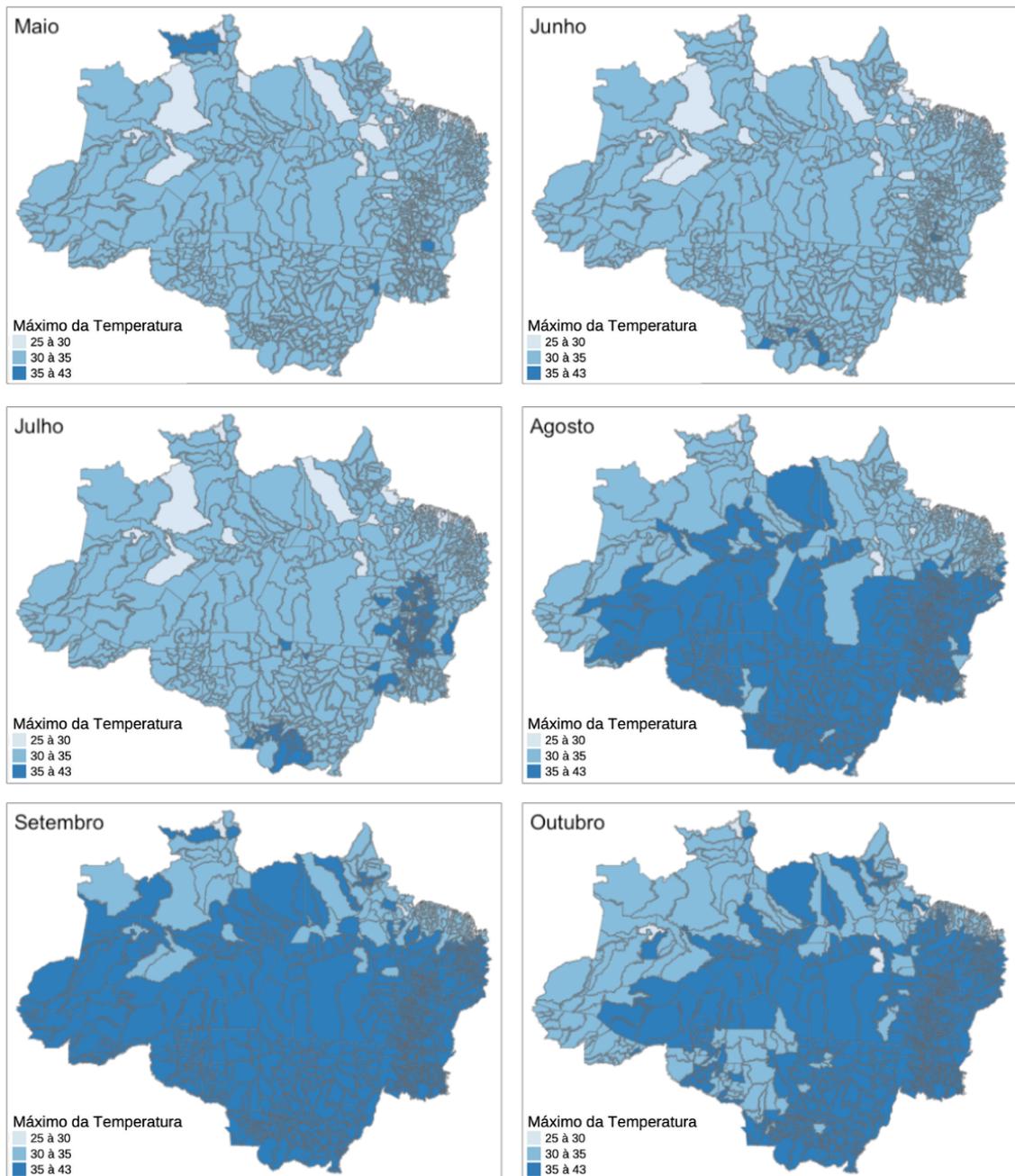


Figura 6: Mapas da Temperatura Máxima em cada mês. Amazônia Legal, 2019.

Além das variáveis ambientais Temperatura e Concentração de $PM_{2.5}$ também foi analisado o mínimo da Umidade em cada mês. Os mapas do mínimo da umidade da Amazônia Legal no ano de 2019 estão na Figura 7, onde os intervalos foram selecionados com base na recomendação da OMS, que a umidade mínima deva ser acima de 30%, mas que a umidade ideal é entre 50 e 80%. Podemos observar que nos meses de julho à setembro houve uma grande parte da Amazônia Legal onde atingiu a umidade mínima abaixo de 30%.

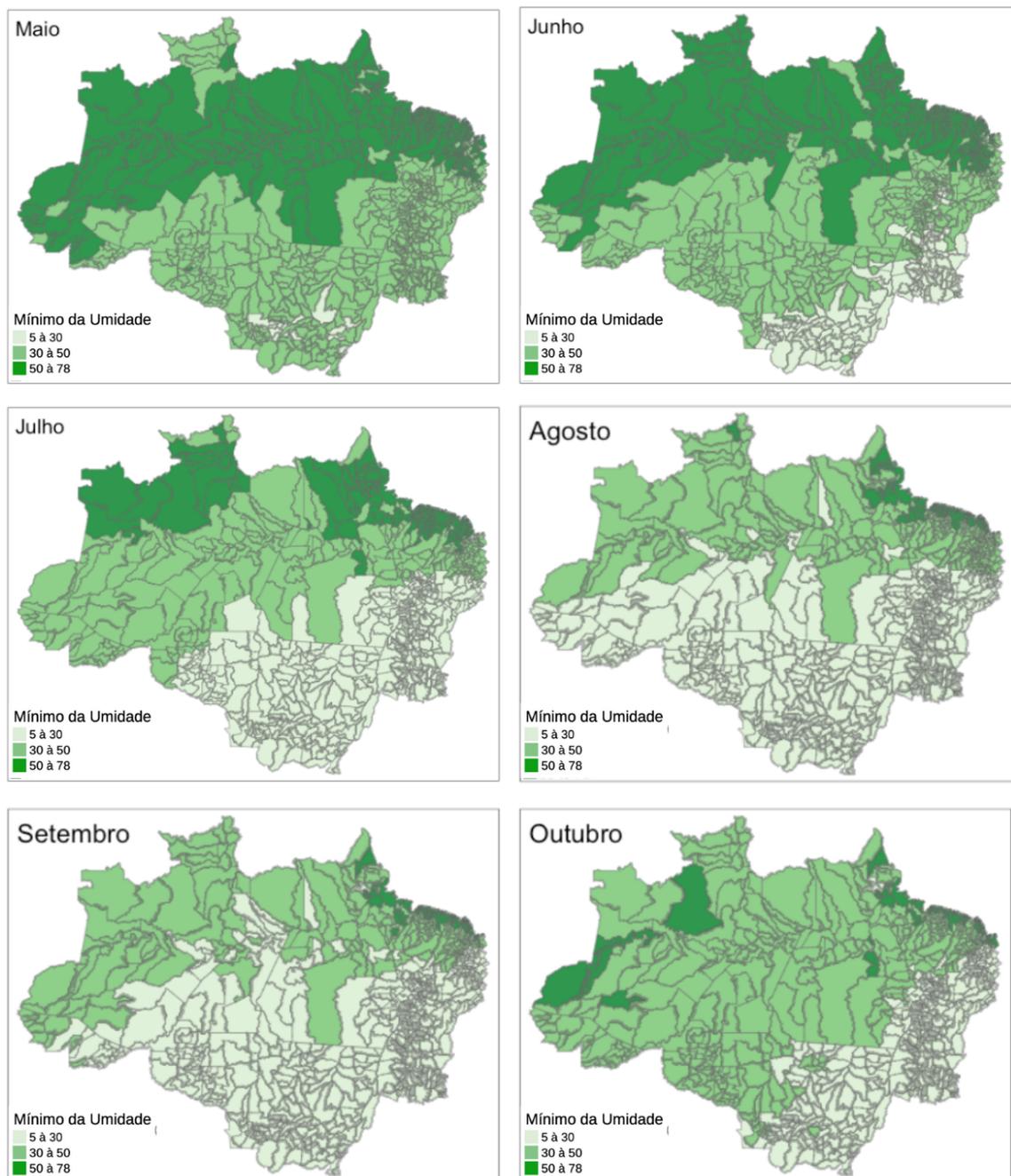


Figura 7: Mapas da Umidade Mínima relativa em cada mês. Amazônia Legal, 2019.

Nos mapas da variável Taxa de Internação, que se encontram na Figura 8, percebe-se que nos meses de maio à agosto houve mais municípios com a Taxa de Internação acima do valor médio que é de 28,5 por 100.000 habitantes.

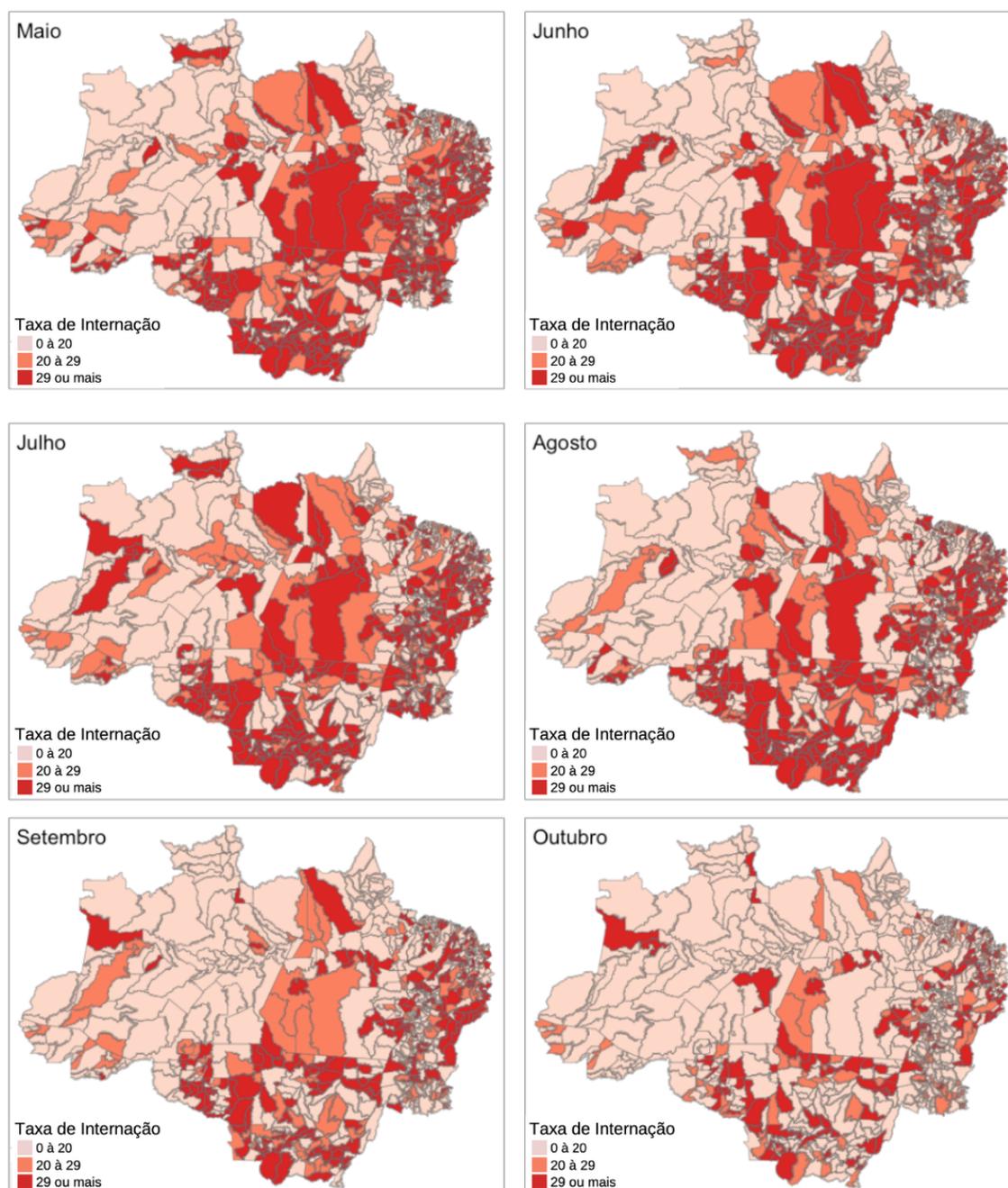


Figura 8: Mapas da Taxa de Internação em cada mês. Amazônia Legal, 2019.

Diante da diferença de comportamento das variáveis ambientais em cada mês, notou-se a necessidade em considerar os meses nas análises.

3.2 Modelo de Regressão Linear

3.2.1 Transformação das variáveis

Diante dos resultados obtidos através da análise exploratória, foram selecionadas as variáveis ambientais Máximo de $PM_{2.5}$, Máximo da Temperatura e Mínimo da Umidade. A variável Máximo de $PM_{2.5}$ foi escolhida devido à sua proximidade com a Taxa de Internação, como foi observado no dendrograma na Figura 4. O primeiro modelo testado, contendo todas as variáveis selecionadas e suas interações, encontra-se no Apêndice 2, onde pode-se observar através do p-valor que as variáveis não foram significativas. Com o objetivo de melhorar o ajuste do modelo foi necessário dividir as variáveis ambientais em categorias, que constam nas equações 3.1, 3.2 e 3.3. A variável dummy Categoria Temperatura foi criada com base no valor médio observado no Máximo da temperatura, que era de $34,23^{\circ}\text{C}$. A dummy Categoria $PM_{2.5}$ foi criada levando em consideração o limite máximo para a média diária, com base nas informações da OMS, que é de $25\mu\text{g}/\text{m}^3$. E a variável dummy Categoria Umidade, foi criada também com base nas informações da OMS que a umidade relativa deve estar acima de 30%.

$$\text{Categoria } PM_{2.5} = \begin{cases} 1, & \text{se concentração acima de } 25\mu\text{g}/\text{m}^3 \\ 0, & \text{caso contrário.} \end{cases} \quad (3.1)$$

$$\text{Categoria Temperatura} = \begin{cases} 1, & \text{se temperatura acima de } 34,23^{\circ}\text{C} \\ 0, & \text{caso contrário.} \end{cases} \quad (3.2)$$

$$\text{Categoria Umidade} = \begin{cases} 1, & \text{se umidade abaixo de } 30\% \\ 0, & \text{caso contrário.} \end{cases} \quad (3.3)$$

A variável Mês também é uma dummy, onde Maio é a categoria de referência. A partir da seleção das variáveis, foi estimado o primeiro modelo contendo todas as variáveis e interações entre elas. Em seguida verificou-se que a variável resposta não seguia uma distribuição normal, então foi feita utilizada a função *boxcox* do programa R(R Core Team, 2014). O intuito foi verificar uma possível transformação na variável resposta. A partir do Boxcox, onde indicava um λ próximo ao 0, 1, foi testado a transformação Log, por λ se aproximar de 0. Então, considerando $\lambda = 0$, a variável resposta do modelo foi transformada em Log da Taxa de Internação, onde o gráfico consta na Figura 9.

Após a transformação da variável foi feito a análise dos resíduos dos dois modelos,

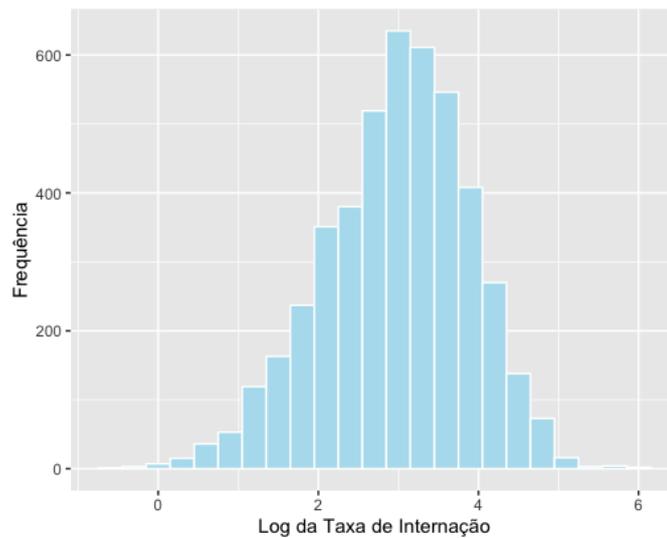


Figura 9: Histograma do Log da Taxa de Internação. Amazônia Legal, 2019.

onde o Modelo 1 é o modelo completo e o Modelo 2 é o modelo completo com a variável resposta transformada. A partir dessa análise foram gerados 2 gráficos quantil-quantil que estão na Figura 10, onde podemos observar uma melhora significativa do Modelo 2 em relação ao Modelo 1. Diante da análise gráfica os dados do Modelo 2 os resíduos aparentam se aproximar de uma distribuição normal.

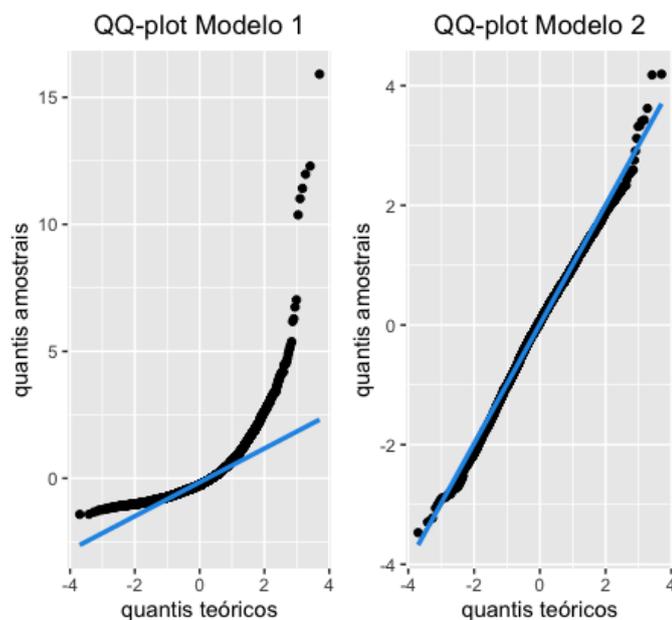


Figura 10: QQ-plot dos modelos. Amazônia Legal, 2019.

3.2.2 Interação

Em seguida foi iniciada a análise da interação entre as variáveis ambientais. Para essa análise foi utilizado o teste F, onde os resultados do p-valor de cada teste encontram-se abaixo na Tabela 3.

Tabela 3: Efeito de interação das variáveis ambientais. Amazônia Legal, 2019.

	Variáveis	Teste F	P-valor
Interação	$PM_{2.5}$ vs Umidade	13,85	0,0002
	Temperatura vs Umidade	2,67	0,1022
	$PM_{2.5}$ vs Temperatura	12,24	0,0005
	$PM_{2.5}$ vs Temperatura vs Umidade	10,923	<0,0001

Diante dos resultados obtidos através do Teste F observou-se que o efeito de interação entre $PM_{2.5}$ e Umidade é significativo, bem como o efeito de interação entre $PM_{2.5}$ e Temperatura e também a interação entre todas as variáveis ambientais. Em relação à interação entre Temperatura e Umidade, apresentou um p-valor próximo de 0,10, e em conjunto com a análise gráfica apresentada na Figura 11 observamos que aparenta existir uma influência mútua entre essas variáveis. Quando a temperatura é mais alta predomina os casos onde a umidade mínima é abaixo de 30 %, enquanto nos casos onde a temperatura é abaixo de 30°C houveram poucas observações que constem umidade abaixo de 30 %.

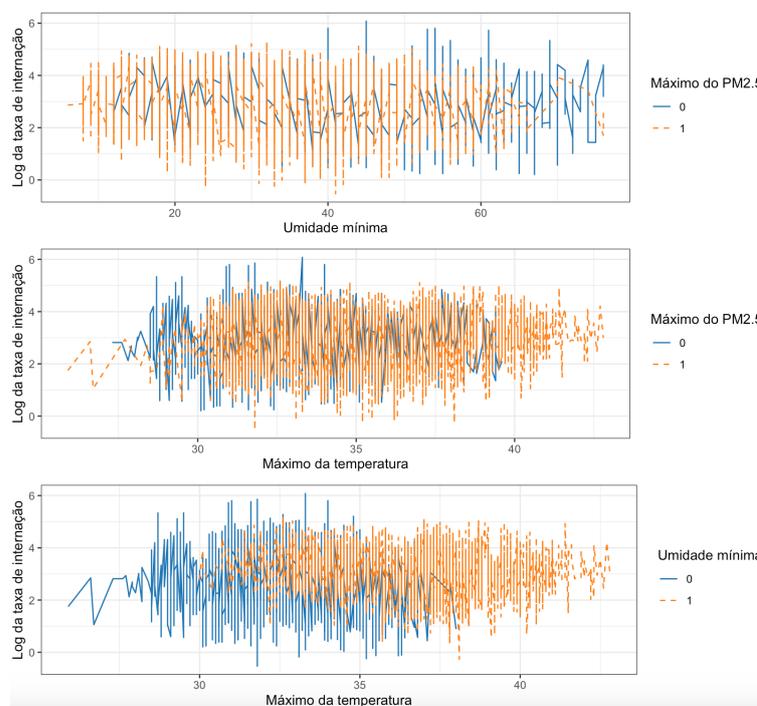


Figura 11: Interação entre as variáveis ambientais. Amazônia Legal, 2019.

3.2.3 Escolha do melhor modelo

A fim de escolher o melhor modelo foi feita a análise dos resíduos e utilizado medidas de comparação como AIC e R^2 ajustado de diversos modelos diferentes, que encontram-se na Tabela 4. No Apêndice 3 encontram-se os modelos presente na Tabela 4, citados anteriormente. Após a análise dos resíduos não foi observada uma diferença entre os modelos, então a escolha do melhor modelo foi com base no menor AIC e maior R^2 ajustado. Apesar dos valores serem parecidos o modelo escolhido foi o Modelo 1.

Tabela 4: Modelos estimados. Amazônia Legal, 2019.

Modelo	R^2 ajustado	AIC
Modelo 1: completo com todas interações	0,105	11826
Modelo 2: completo sem interações	0,101	11842
Modelo 3: com interação entre Umidade e Temperatura	0,101	11844
Modelo 4: com interação entre Umidade e $PM_{2.5}$	0,101	11844
Modelo 5: com interação entre Temperatura e $PM_{2.5}$	0,102	11839.

Após a escolha do Modelo 1, foram executados os testes de Normalidade Lilliefors (p-valor $< 0,0001$) e Shapiro (p-valor $< 0,0001$). Além desses também foi utilizado o teste de Breusch-Pagan para verificar se variância dos erros era constante, porém todos os testes rejeitaram a hipótese nula, com p-valor menor que 0,0001 retornando um resultado de que os erros não seguiam uma distribuição normal e que a variância dos erros não eram constantes. Entretanto, diante de uma análise gráfica que consta na Figura 12 é razoável considerar que os resíduos seguem uma distribuição aproximadamente normal.

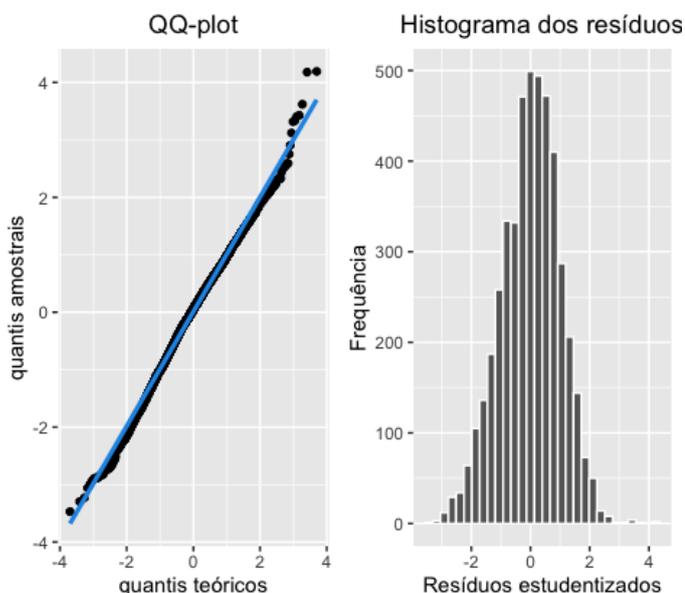


Figura 12: QQ-plot e histograma dos resíduos do Modelo 1. Amazônia Legal, 2019.

Apesar de o modelo final não ser o que melhor se ajusta aos dados, as estimativas dos parâmetros e p-valores seguem na tabela 5.

Tabela 5: Estimativas dos parâmetros do Modelo Final. Amazônia Legal, 2019. ¹

Variáveis	Estimativa	p-valor
Intercepto	3,0324	<0,0001
mesJunho	-0,1057	0,01852
mesJulho	-0,2069	<0,0001
mesAgosto	-0,29953	<0,0001
mesSetembro	-0,5605	<0,0001
mesOutubro	-0,8452	<0,0001
UmidadeMin	0,55240	<0,0001
TemperaturaMax	0,38635	<0,0001
PM2.5Max	0,12682	0,0018
UmidadeMin:TemperaturaMax	-0,48000	0,0003
UmidadeMin:PM _{2,5} Max	-0,25800	0,01804
TemperaturaMax:PM _{2,5} Max	-0,45258	<0,0001
UmidadeMin:TemperaturaMax:PM _{2,5} Max	0,66958	<0,0001

1

De acordo com as estimativas do Modelo final, descrito na tabela 5, os efeitos do material particulado fino na taxa de internação são interpretados a seguir:

- quando a concentração máxima do material particulado está acima de $25\mu\text{g}/\text{m}^3$ espera-se um aumento médio de aproximadamente 13,5% na taxa de internação;
- caso a temperatura máxima seja acima de $34,23^\circ\text{C}$, o efeito do $PM_{2,5}$, acima de $25\mu\text{g}/\text{m}^3$, na taxa de internação é de um decréscimo de 6,41% [$IC_{95\%} : -6,53; -6,29$];
- o efeito esperado do $PM_{2,5}$ na taxa de internação é um aumento de 34,23% [$IC_{95\%} : 34,11; 34,35$], quando a umidade relativa está abaixo de 30%;
- já o efeito esperado do material particulado fino, quando a temperatura alta e a umidade baixa, na taxa de internação é um aumento de 121,75% [$IC_{95\%} : 121,42; 122,09$].

Podemos observar que é esperado um aumento 47,16% na taxa de internação quando a temperatura está acima de $34,23^\circ\text{C}$. Para a umidade relativa mínima abaixo de 30% é esperado um aumento médio de 73,74% na taxa de internação. Analisando os meses observou-se que nos meses de setembro e outubro é esperado uma redução de 42,91% e 57,05%, respectivamente, na taxa de internação em relação ao mês de maio. Podemos

¹TemperaturaMax, UmidadeMin e $PM_{2,5}$ Max são as variáveis ambientais categorizadas

observar também, na figura 13 que o mês de maio foi o mês com maior valor do log da Taxa de Internação.

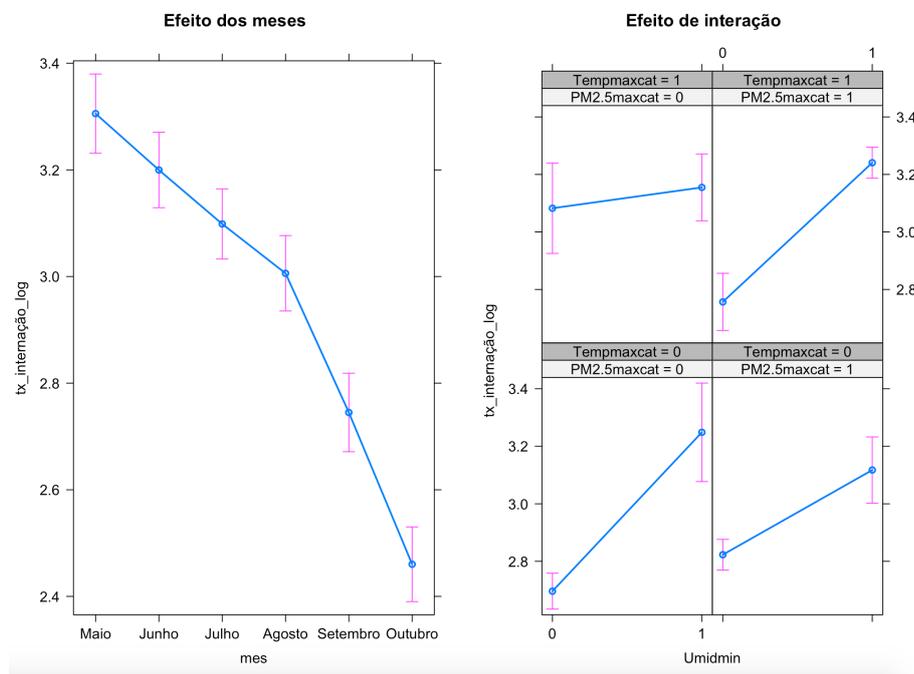


Figura 13: Efeito do Log da Taxa de Internação em cada mês e interações. Amazônia Legal, 2019. ²

2

Se a umidade mínima relativa estiver abaixo de 30%, o efeito da temperatura na taxa de interação é de um decréscimo de 8,94% [$IC_{95\%} : -9,15; -8,73$]. O efeito esperado da Temperatura, quando a umidade está baixa e o $PM_{2.5}$ alto, na taxa de internação é um aumento de 187,46% [$IC_{95\%} : 187,05; 187,88$] na taxa de internação. Na Figura 13 podemos observar o efeito da umidade mínima, temperatura máxima e $PM_{2.5}$ máximo na taxa de internação. Notou-se que nos casos onde a umidade mínima é baixa, concentração máxima de $PM_{2.5}$ alta e temperatura máxima alta teve um maior efeito na taxa de internação, além de se observar um intervalo de confiança menor nesse caso. Após todos os resultados obtidos ficou evidente a importância de estudar as 3 variáveis ambientais conjuntamente.

²Tempmaxcat, Umidmin e $PM_{2.5}$ maxcat são as variáveis ambientais categorizadas, tx_internação_log é o Log da Taxa de Internação

4 Conclusões

No modelo final observamos que era esperado um aumento de 47,16% na taxa de internação em relação à temperatura quando abaixo da média. Além disso, foi observado que para a umidade relativa mínima abaixo de 30% é esperado um aumento médio de 73,74% na taxa de internação. Também notou-se que no mês de outubro é esperado uma redução de 57,05% na taxa de internação quando comparado ao mês de referência, que é agosto. O efeito do material particulado fino na taxa de internação, quando a temperatura alta e a umidade baixa, é um aumento de 121,75%. Além disso, o efeito esperado do $PM_{2.5}$ na taxa de internação é um aumento de 34,23%, quando a umidade relativa está abaixo de 30%. Com isso, o objetivo principal deste trabalho foi alcançado, que visava entender o efeito da concentração do material particulado fino na taxa de internação, junto com as variáveis umidade e temperatura. Além de avaliar as interações entre as variáveis ambientais.

Os resultados obtidos corroboram com um estudo sobre variabilidade do clima, onde foi observado que umidade e as estações do ano também podem ser um fator que influencia no efeito do $PM_{2.5}$. No mesmo estudo foi concluído que a temperatura média e máxima tem correlação direta com $PM_{2.5}$ e que a umidade possui alta correlação inversa com a temperatura máxima, corroborando com os resultados obtidos no presente estudo e evidenciando a necessidade de estudar as 3 variáveis ambientais (RODRIGUES et al., 2017).

Quanto às limitações deste estudo observamos que mesmo após a transformação da variável resposta em Log, o modelo não satisfazia todos os pressupostos necessários, a variância dos erros não era constante e os erros eram correlacionados. Com isso, os estimadores obtidos não são tão eficientes. Existem então outros modelos que podem se ajustar melhor aos dados, como por exemplo um Modelo Linear Hierárquico com 2 níveis, sendo o primeiro nível os meses e o segundo nível os municípios.

Referências

- ANDRADE FILHO, V. S. de et al. Distribuição espacial de queimadas e mortalidade em idosos em região da amazônia brasileira, 2001 – 2012. *Ciência Saúde Coletiva*, v. 22, n. 1, p. 245–253, 2017.
- BROOK, R. D. et al. Particulate matter air pollution and cardiovascular disease. v. 121, n. 21, p. 2331–2378, 2010.
- DATASUS. *Sistema de Informações Hospitalares do SUS (SIH/SUS)*, Dezembro 2019. Disponível em: <http://tabnet.datasus.gov.br/cgi/defthtm.exe?sih/cnv/nrbr.def>. Acesso em: 7 de outubro 2021.
- EPA. Particulate matter (pm2.5) trends. *United States Environmental Protection Agency*, Maio 2021. Disponível em: <https://www.epa.gov/air-trends/particulate-matter-pm25-trends>. Acesso em: 14 de Fevereiro 2022.
- HOLLANDER, D. A. W. M.; CHICKEN, E. *Nonparametric Statistical Methods*. 4. ed. [S.l.]: John Wiley Sons, Inc., 2014.
- IBGE. Amazônia legal. *IBGE*, Julho 2021.
- III, C. A. P. et al. Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution. *JAMA*, v. 287, n. 9, p. 1132–1141, 03 2002. ISSN 0098-7484. Disponível em: <https://doi.org/10.1001/jama.287.9.1132>.
- INPE. Focos por estado. *INPE*, Outubro 2019. Disponível em: <https://queimadas.dgi.inpe.br/queimadas/bdqueimadas#graficos>. Acesso em: 20 de agosto 2021.
- JOHNSON, R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. 6. ed. [S.l.]: Pearson Education, Inc., 2007.
- KUTNER, M. H. et al. *Applied Linear Statistical Models*. 5. ed. [S.l.]: The McGraw-Hill Companies, Inc., 2004.
- LIM, S. S. et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the global burden of disease study 2010. *Lancet*, p. 2224–2260, 2013.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. *Introduction to Linear Regression Analysis*. 5. ed. [S.l.]: John Wiley and Sons, Inc., 2012.
- OMS. Cardiovascular diseases (cvds). *OMS*, Junho 2021. Disponível em: [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)). Acesso em: 30 de agosto 2021.

OPAS. Oms revela principais causas de morte e incapacidade em todo o mundo entre 2000 e 2019. *OPAS*, Dezembro 2020. Disponível em: <https://www.paho.org/pt/noticias/9-12-2020-oms-revela-principais-causas-morte-e-incapacidade-em-todo-mundo-entre-2000-e>). Acesso em: 20 de agosto 2021.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2014. Disponível em: <http://www.R-project.org/>.

REAL, K. V.; IGNOTTI, N. E.; HACON, S. de S. Circulatory disease mortality rates in the elderly and exposure to pm2.5 generated by biomass burning in the brazilian amazon in 2005. *Cadernos Saúde Pública*, v. 29, n. 3, p. 589–598, 2013.

RODRIGUES, P. C. de O. et al. Climatic variability and morbidity and mortality associated with particulate matter. *Revista de Saúde Pública*, v. 51, n. 91, 2017.

SETZER, A. W.; SISMANOGLU, R. A.; SANTOS, J. G. M. dos. Método do cálculo do risco de fogo do programa do inpe. *INPE*, Junho 2019. Disponível em: https://queimadas.dgi.inpe.br/~rqueimadas/documentos/RiscoFogo_Sucinto.pdf. Acesso em: 23 de Janeiro 2022.

SISAM. *Sistema de Informações Ambientais Integrado a Saúde*, Dezembro 2019. Disponível em: <https://queimadas.dgi.inpe.br/queimadas/sisam/v2/dados/download/>. Acesso em: 7 de outubro 2021.

APÊNDICE 1 – Boxplots variáveis ambientais

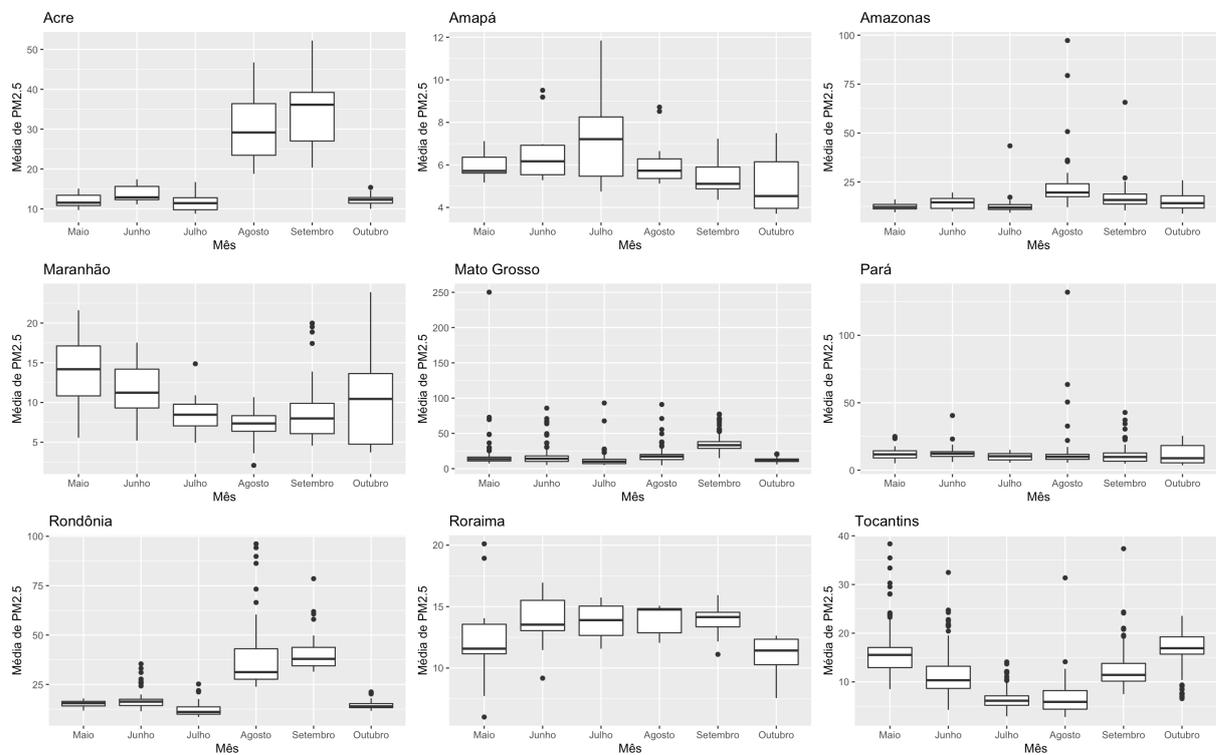


Figura 14: Boxplot da média de $PM_{2.5}$ em cada mês por Estado. Amazônia Legal, 2019.

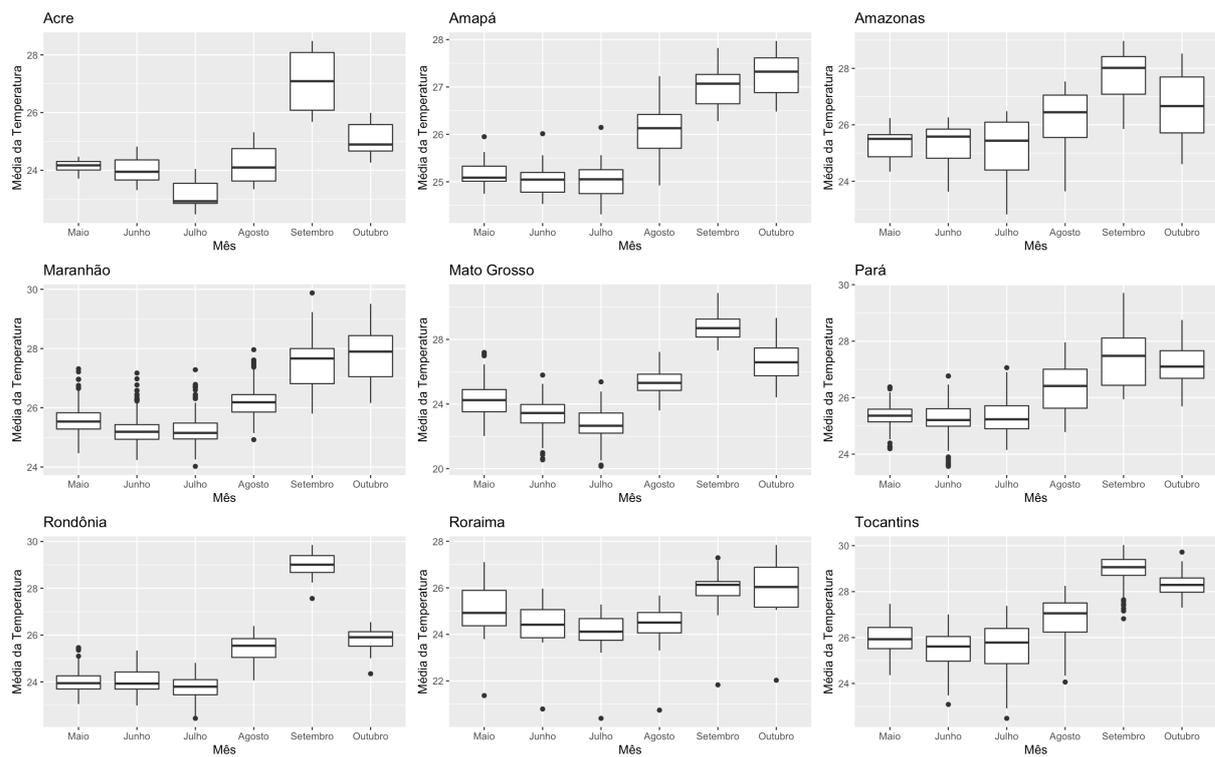


Figura 15: Boxplot da média da temperatura em cada mês por Estado. Amazônia Legal,2019.

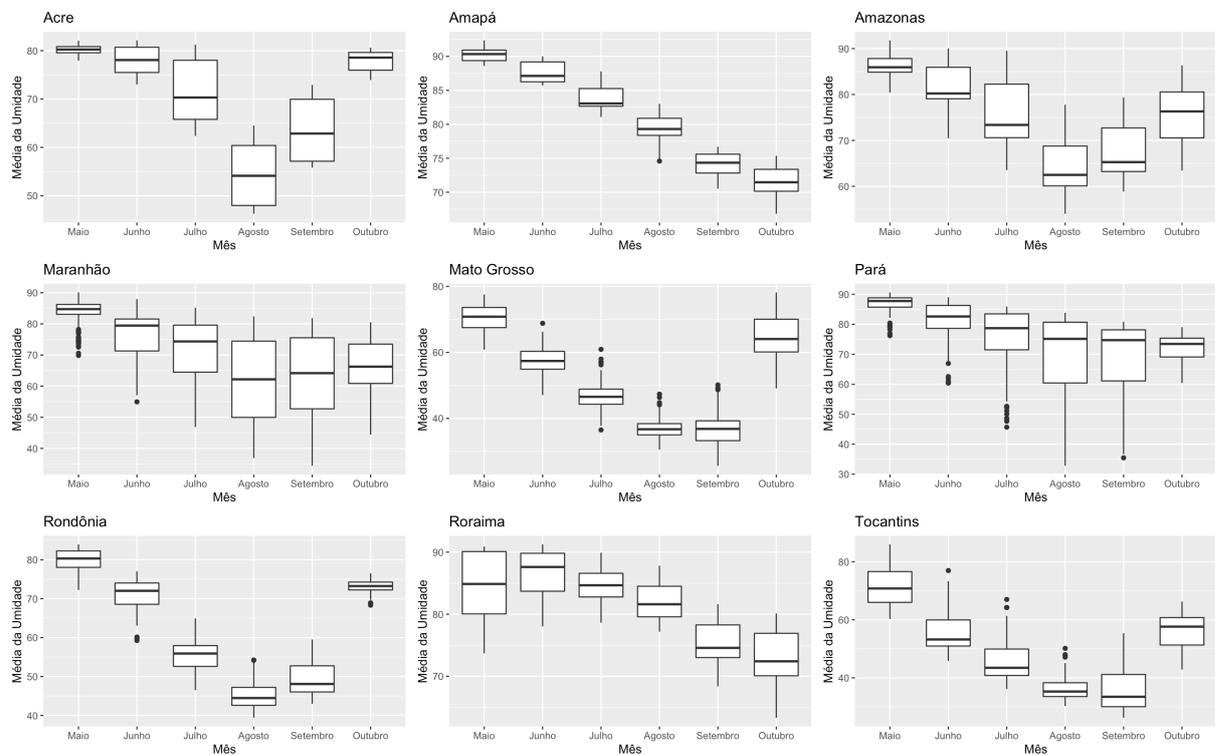


Figura 16: Boxplot da média da umidade em cada mês por Estado. Amazônia Legal,2019.

APÊNDICE 2 – Primeiro modelo

Tabela 6: Estimativas dos parâmetros do modelo. Amazônia Legal, 2019.

Variáveis	Estimativa	p-valor
Intercepto	4.686	<0,0001
mesJunho	-0,1367	0,0024
mesJulho	-0,2615	<0,0001
mesAgosto	-0,3572	<0,0001
mesSetembro	-0,5671	<0,0001
mesOutubro	-0,7769	<0,0001
UmidadeMinima	0,0098	0,4950
TemperaturaMaxima	-0,0171	0,3230
$PM_{2.5}$ Maximo	-0,005142	0,4243
TemperaturaMax: UmidadeMinima	-0,0010	0,0307
UmidadeMinima: $PM_{2.5}$ Max	0,0002	0,4248
TemperaturaMax: $PM_{2.5}$ Max	0,0002	0,3648
TemperaturaMax: $PM_{2.5}$ Max:UmidadeMinima	-0,000006	0,3890

APÊNDICE 3 – Modelos

Tabela 7: Estimativas dos parâmetros do Modelo 2. Amazônia Legal, 2019.

Variáveis	Estimativa	p-valor
Intercepto	3,0908	<0,0001
mesJunho	-0,1036	0,0204
mesJulho	-0,2137	<0,0001
mesAgosto	-0,3182	<0,0001
mesSetembro	-0,5793	<0,0001
mesOutubro	-0,8598	<0,0001
UmidadeMin	0,3761	<0,0001
TemperaturaMax	0,0751	0,0857
$PM_{2,5}Max$	0,0448	0,1386

Tabela 8: Estimativas dos parâmetros do Modelo 3. Amazônia Legal, 2019.

Variáveis	Estimativa	p-valor
Intercepto	3,0909	<0,0001
mesJunho	-0,1034	0,0215
mesJulho	-0,2132	<0,0001
mesAgosto	-0,3182	<0,0001
mesSetembro	-0,5792	<0,0001
mesOutubro	-0,8595	<0,0001
UmidadeMin	0,3740	<0,0001
TemperaturaMax	0,0735	0,1608
$PM_{2,5}Max$	0,044738	0,1395
UmidadeMin:TemperaturaMax	0,0037	0,9571

Tabela 9: Estimativas dos parâmetros do Modelo 4. Amazônia Legal, 2019.

Variáveis	Estimativa	p-valor
Intercepto	3,0846	<0,0001
mesJunho	-0,1042	0,0198
mesJulho	-0,2147	<0,0001
mesAgosto	-0,3150	<0,0001
mesSetembro	-0,5756	<0,0001
mesOutubro	-0,8571	<0,0001
UmidadeMin	0,3964	<0,0001
TemperaturaMax	0,0533	0,1531
$PM_{2,5}Max$	0,0732	0,0962
UmidadeMin: $PM_{2,5}Max$	-0,0260	0,6965

Tabela 10: Estimativas dos parâmetros do Modelo 5. Amazônia Legal, 2019.

Variáveis	Estimativa	p-valor
Intercepto	3,0554	<0,0001
mesJunho	-0,1068	0,01680
mesJulho	-0,2174	<0,0001
mesAgosto	-0,3017	<0,0001
mesSetembro	-0,5599	<0,0001
mesOutubro	-0,8455	<0,0001
UmidadeMin	0,3848	<0,0001
TemperaturaMax	0,1663	0,00614
$PM_{2,5}Max$	0,0938	0,01301
TemperaturaMax: $PM_{2,5}Max$	-0,1411	0,03035