



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE BACHARELADO EM ESTATÍSTICA**

GISLÂNIA CAUANNY SILVA SOUZA

**AJUSTE DE MODELOS LINEARES GENERALIZADOS APLICADO AO
IMPACTO DE MATERIAL PARTICULADO EM MUNICÍPIOS DA
PARAÍBA**

CAMPINA GRANDE - PB

2023

GISLÂNIA CAUANNY SILVA SOUZA

**AJUSTE DE MODELOS LINEARES GENERALIZADOS APLICADO AO
IMPACTO DE MATERIAL PARTICULADO EM MUNICÍPIOS DA
PARAÍBA**

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Bacharel em Estatística.

Orientador: Prof. Dr. Ricardo Alves de Olinda

**CAMPINA GRANDE - PB
2023**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

S729a Souza, Gislania Cauanny Silva.

Ajuste de modelos lineares generalizados aplicado ao impacto de material particulado em municípios da Paraíba [manuscrito] / Gislania Cauanny Silva Souza. - 2023.

39 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2023.

"Orientação : Prof. Dr. Ricardo Alves de Olinda, Coordenação do Curso de Estatística - CCT. "

1. Carbono negro. 2. Dados assimétricos. 3. Poluição do ar. 4. Saúde pública. I. Título

21. ed. CDD 519.5

GISLÂNIA CAUANNY SILVA SOUZA

AJUSTE DE MODELOS LINEARES GENERALIZADOS APLICADO AO IMPACTO
DE MATERIAL PARTICULADO EM MUNICÍPIOS DA PARAÍBA

Trabalho de Conclusão de Curso apresentado
ao curso de Bacharelado em Estatística do
Departamento de Estatística do Centro de Ci-
ências e Tecnologia da Universidade Estadual
da Paraíba, como requisito parcial à obtenção
do título de Bacharel em Estatística.

Trabalho aprovado em 24/11/2023.

BANCA EXAMINADORA



Prof. Dr. Ricardo Alves de Olinda Primeiro
Membro (Orientador)
Universidade Estadual da Paraíba (UEPB)



Profa. Dr. Divanilda Maia Esteves Segundo
Membro
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Silvio Xavier Ferreira
Júnior Terceiro Membro
Universidade Estadual da Paraíba (UEPB)

Dedico este trabalho em primeiro lugar a Deus; sem Ele eu não teria capacidade para chegar até aqui.

AGRADECIMENTOS

Em primeiro lugar, expresso minha profunda gratidão a Deus, o autor da minha vida, que permitiu que eu vivenciasse todas essas experiências. Que darei eu ao Senhor, por todos os benefícios que me tem feito? (Salmos 116:12). À minha amada família, meu pai Gilvan, minha mãe Carmita e minhas irmãs Nathália e Lara, quero expressar minha eterna gratidão pelo apoio incondicional que me deram durante esta jornada. Palavras não são suficientes para descrever o quanto sou grata. Ao meu companheiro, Marcilio, agradeço por todo o amor, cumplicidade, suporte e compreensão que demonstrou em todos os momentos, ele sempre esteve pronto para me ajudar e ouvir. Ao meu cunhado Jonas por todo apoio. Aos meus amigos de turma, Daví, Elyda, Gabriel, Hellen e Marcela, quero agradecer por tornarem os dias mais leves e sorridentes e por sua pronta colaboração. Não consigo imaginar como teria chegado até aqui sem vocês. Às minhas primas, Elibe e Ewellyn, que são exemplos inspiradores para mim, minha gratidão é imensa. Ao meu orientador, Ricardo, agradeço por todo o conhecimento e dedicação que compartilhou comigo. Sua disponibilidade e orientação foram fundamentais para o meu sucesso. Também desejo agradecer ao professor Domingos por sua calma, paciência e colaboração, que desempenharam um papel crucial. Sem cada um de vocês, não posso imaginar como seria esta jornada. Expresso meu carinho e profundo agradecimento do fundo do meu coração a cada um de vocês.

"O produto mais valioso que conheço é a informação."
(Gordon Gekko)

RESUMO

O material particulado consiste em uma mistura de partículas de diversos materiais que pode ser encontrado em suspensão na atmosfera, seja na forma líquida ou sólida, permanecendo em suspensão no ar. De acordo com a Agência de Proteção Ambiental dos Estados Unidos (EPA), o material particulado pode ser classificado em duas categorias: Pm2.5, que são partículas com um diâmetro de até 2,5 micrômetros, frequentemente encontradas em nevoeiros e fumaça; e Pm10, composto por partículas com diâmetro de 2,5 a 10 micrômetros, geralmente presentes em áreas próximas a indústrias. Um impacto ambiental significativo é o carbono negro (fuligem), proveniente da queima incompleta de diesel e incêndios, sendo um dos principais contribuintes para o aquecimento global. Quanto à saúde, a exposição a essas partículas está associada a ataques cardíacos, arritmias cardíacas e desenvolvimento de asma em crianças. Para conscientizar a população sobre a importância desse estudo, serão aplicados métodos e técnicas dos Modelos Lineares Generalizados (MLGs). O objetivo principal é avaliar a relação entre a variável resposta (Pm2.5) e as variáveis explicativas (precipitação, temperatura, umidade e vento) nos municípios de Campina Grande, João Pessoa e Patos, localizados no estado da Paraíba. Os resultados obtidos demonstraram uma relação significativa entre o Pm2.5 e as variáveis climáticas, indicando que o aumento do Pm2.5 está associado a temperaturas mais elevadas, baixa umidade e oscilações na precipitação. A integração de dados climáticos neste estudo fornece informações valiosas para o desenvolvimento de regulamentações ambientais mais eficazes, tais como padrões de qualidade do ar, restrições de emissões e práticas de controle da poluição. Além disso, não podemos ignorar as graves consequências para a saúde da população que tem contato com esse poluente, sendo as crianças e os idosos os principais afetados, com o desenvolvimento de doenças respiratórias e outros danos. Isso destaca a importância de compreender os riscos à saúde associados à poluição do ar e fornece uma base sólida para a adoção de medidas de controle. Essas informações são de extrema relevância para a sociedade e são fundamentais na formulação de políticas públicas voltadas para a conscientização sobre a gravidade do material particulado no meio ambiente.

Palavras-chave: Carbono negro; dados assimétrico; poluição do ar; saúde pública.

ABSTRACT

Particulate matter consists of a mixture of particles of various materials that can be found suspended in the atmosphere, either in liquid or solid form, remaining suspended in the air. According to the US Environmental Protection Agency (EPA), particulate matter can be classified into two categories: Pm2.5, which are particles with a diameter of up to 2.5 micrometers, often found in fog and smoke; and Pm10, made up of particles with a diameter of 2.5 to 10 micrometers, usually present in areas close to industries. A significant environmental impact is black carbon (soot), which comes from the incomplete burning of diesel and fires, and is one of the main contributors to global warming. In terms of health, exposure to these particles is associated with heart attacks, cardiac arrhythmias and the development of asthma in children. To raise awareness of the importance of this study, methods and techniques from Generalized Linear Models (GLMs) will be applied. The main objective is to evaluate the relationship between the response variable (Pm2.5) and the explanatory variables (precipitation, temperature, humidity and wind) in the municipalities of Campina Grande, João Pessoa and Patos, located in the state of Paraíba. The results obtained showed a significant relationship between Pm2.5 and the climatic variables, indicating that an increase in Pm2.5 is associated with higher temperatures, low humidity and oscillations in precipitation. The integration of climate data in this study provides valuable information for the development of more effective environmental regulations, such as air quality standards, emission restrictions and pollution control practices. In addition, we cannot ignore the serious consequences for the health of the population that comes into contact with this pollutant, with children and the elderly being the main ones affected, with the development of respiratory diseases and other damages. This highlights the importance of understanding the health risks associated with air pollution and provides a solid basis for adopting control measures. This information is extremely relevant to society and is fundamental in formulating public policies aimed at raising awareness of the seriousness of particulate matter in the environment.

Keywords: Black carbon; asymmetric data; air pollution; public health.

LISTA DE ILUSTRAÇÕES

Figura 1 – Pontos de queimadas ao longo dos anos de 2015 à 2019 no estado da Paraíba	28
Figura 2 – Correlação entre as variáveis explicativas e variável resposta dos municípios de Campina Grande, Patos e João Pessoa	30
Figura 3 – Análise de diagnóstico referente ao ajuste do modelo gama para o município de Campina Grande	32
Figura 4 – Análise de diagnóstico referente ao ajuste do modelo gama para o município de Patos	34
Figura 5 – Análise de diagnóstico referente ao ajuste do modelo gama para o município de João Pessoa	36

LISTA DE TABELAS

Tabela 1 – Algumas distribuições importantes da família exponencial	17
Tabela 2 – Algumas distribuições importantes da função geradora de momentos .	19
Tabela 3 – Funções de ligações canônicas	22
Tabela 4 – Análise descritiva das variáveis explicativas e da variável resposta para o município de Campina Grande	29
Tabela 5 – Análise descritiva das variáveis explicativas e da variável resposta para o município de Patos	29
Tabela 6 – Análise descritiva das variáveis explicativas e da variável resposta para o município de João Pessoa	29
Tabela 7 – Estimativas dos parâmetros referente ao modelo gama ajustado para o município de Campina Grande	31
Tabela 8 – Estimativas dos parâmetros referente ao modelo gama ajustado para o município de Patos	33
Tabela 9 – Estimativas dos parâmetros referente ao modelo gaussiano ajustado para o município de João Pessoa	35

LISTA DE ABREVIATURAS E SIGLAS

MLGs	Modelos Lineares Generalizados
SISAM	Sistema de Informações Ambientais Integrado a Saúde
Pm2.5	Material Particulado
Pe	Precipitação
Te	Temperatura
Um	Umidade Relativa do Ar
Ve	Vento
AIC	Critério de Informação de Akaike
BIC	Critério de Informação de Bayes
CETESB	Companhia de Tecnologia de Saneamento Ambiental

LISTA DE SÍMBOLOS

Γ	Gama
Λ	Lambda
μ	Mi
ϕ	Fi
θ	Theta
σ	Sigma
π	Pi
ν	Nu
∂	Partial
η	Eta
β	Beta
α	Alfa
l	ell

SUMÁRIO

1	INTRODUÇÃO	13
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	Métodos Estatísticos	15
2.1.1	<i>Família Exponencial Uniparamétrica</i>	15
2.1.2	<i>Componente Aleatório</i>	15
2.1.3	<i>Função Geradora de Momentos</i>	18
2.1.4	<i>Estatística Suficiente</i>	19
2.2	Modelo Linear Generalizado	20
2.2.1	<i>Ligação Canônica</i>	21
2.2.2	<i>Estimação dos Parâmetros</i>	21
2.3	Seleção e Validação de Modelos	24
2.3.1	<i>Qualidade do Ajuste</i>	24
2.3.2	<i>Seleção de modelos</i>	25
2.4	Técnicas de Diagnóstico	26
2.5	Materiais	26
3	RESULTADOS E DISCUSSÃO	28
3.1	Análise exploratória	28
3.2	Ajuste de modelos	30
4	CONCLUSÃO	37
	REFERÊNCIAS	38

1 INTRODUÇÃO

Os danos causados pelos aerossóis provenientes da queima de biomassa, presentes na poluição do ar, têm sido amplamente estudados em todo o mundo. Estudos anteriores, como o de (ANDRADE PAULO ARTAXO, 2012), (ARAUJO, 2020), (ROCHA, 2016) e (RIBEIRO; ASSUNÇÃO, 2002), demonstram os impactos que as queimadas têm na saúde humana. As queimadas são responsáveis por liberar altas concentrações de partículas de aerossóis na atmosfera, o que acarreta efeitos adversos diretos e indiretos no clima, incluindo a redução da precipitação e a alteração do funcionamento dos ecossistemas, conforme descrito por (FILHO et al., 2013).

Pesquisas realizadas por (Radojevic; Hassan, 1999) em Brunei Darussalam, nas ilhas Bornéu, indicam alguns dos efeitos que as queimadas de florestas desencadeiam na região: drástica redução da visibilidade, fechamento de aeroportos e escolas, aumento de acidentes de tráfego, destruição da biota pelo fogo, aumento na incidência de doenças, diminuição da produtividade, restrição das atividades de lazer e de trabalho, efeitos psicológicos e custos econômicos. Dentre os sintomas de doenças é relatado infecções do sistema respiratório superior, asma, conjuntivite, bronquite, irritação dos olhos e garganta, tosse, falta de ar, nariz entupido, vermelhidão e alergia na pele, e desordens cardiovasculares.

Outro efeito significativo em termos climáticos é que as partículas em suspensão acabam por interferir no processo de formação de núcleos de condensação, alterando os mecanismos de formação de nuvens e o albedo. Pesquisas realizadas dentro do projeto SCAR-B indicaram que o raio dos aerossóis emitidos por queimadas teve um aumento de 60% na atmosfera nos três primeiros dias devido à condensação e à coagulação, indicando uma correlação positiva entre os núcleos de condensação e as concentrações de fumaça que afetam as propriedades das nuvens (KAUFMAN et al., 1999).

A Análise de Regressão é hoje uma das técnicas estatísticas mais usadas em diversas áreas do conhecimento. Na Análise de Regressão pretende-se encontrar uma relação estocástica entre duas ou mais variáveis com o objetivo de explicar determinado fenômeno em estudo e prever a evolução desse fenômeno. Mais precisamente, analisar a influência que uma ou mais variáveis (variáveis independentes ou explicativas) têm sobre uma variável de interesse (variável dependente ou resposta). Entretanto, no ajuste dos modelos de regressão linear, para relacionar a variável resposta com as variáveis explicativas, é muito frequente a violação de pressupostos como as hipóteses de linearidade da relação e homoscedasticidade dos componentes da variável resposta.

Para situações em que a variável resposta não segue uma distribuição normal, foi proposto por (NELDER; WEDDERBURN, 1972) os Modelos Lineares Generalizados (MLGs) como uma extensão do modelo linear normal, são modelos para análise de dados em que a suposição de normalidade não é plausível. É possível modelar variáveis de interesse que assumem a forma de contagem, contínuas simétricas e assimétricas, binárias

e categóricas, assume-se que a densidade da função de distribuição da variável resposta pertence à família exponencial.

Com o objetivo de destacar a relevância deste estudo para a população e políticas públicas, foi conduzida uma análise abrangente dos efeitos do material particulado e seu impacto na saúde e no meio ambiente para o estado da Paraíba nos municípios de Campina Grande, João Pessoa e Patos. Essa pesquisa visa informar ações e diretrizes voltadas para a proteção da saúde da população e preservação do ambiente. Além disso, contribuirá para o desenvolvimento de estratégias de mitigação dos impactos negativos do material particulado, promovendo assim uma melhor qualidade de vida e um ambiente mais saudável para as comunidades dessas regiões.

2 FUNDAMENTAÇÃO TEÓRICA

Uma análise teórica foi conduzida sobre os MLGs com o propósito de aplicá-los em situações práticas envolvendo dados contínuos com assimetria positiva para distribuição gama, e dados simétricos positivos para distribuição gaussiana. Esta pesquisa abarca conceitos essenciais, tais como componente aleatório, componente sistemático, função de ligação, avaliação da qualidade do ajuste e diagnóstico de modelo.

2.1 Métodos Estatísticos

Conforme descrito na literatura, o conceito de família exponencial foi introduzido na estatística por Fisher, mas os modelos da família exponencial apareceram na mecânica estatística no final do século XIX e foram desenvolvidos por Maxwell, Boltzmann e Gibbs. A importância da família exponencial de distribuições teve maior destaque, na área dos modelos de regressão, a partir do trabalho pioneiro de (NELDER; WEDDERBURN, 1972) com os modelos lineares generalizados. Os modelos lineares generalizados são uma síntese e extensão notáveis de modelos de regressão familiar.

2.1.1 Família Exponencial Uniparamétrica

A família exponencial uniparamétrica é caracterizada por uma função (de probabilidade ou densidade), conforme (DEMÉTRIO, 2008), na forma

$$f(y; \theta) = h(y) \exp [\eta(\theta) t(y) - b(\theta)],$$

em que $\eta(\theta)$, $b(\theta)$, $t(y)$ e $h(y)$ assumem valores em subconjuntos dos reais. As funções $\eta(\theta)$, $b(\theta)$ e $t(y)$ não são únicas.

Segundo (DEMÉTRIO, 2008), a família exponencial na forma canônica é definida a partir da uniparamétrica considerando que as funções $\eta(\theta)$ e $t(y)$ são iguais à função identidade, de forma que

$$f(y; \theta) = h(y) \exp[\theta y - b(\theta)],$$

onde, θ é chamado de parâmetro canônico.

2.1.2 Componente Aleatório

O componente aleatório de um modelo linear generalizado é definido a partir da família exponencial uniparamétrica na forma canônica com a introdução de um parâmetro $\phi > 0$ denominado de parâmetro de perturbação. (NELDER; WEDDERBURN, 1972) ao fazerem isso, conseguiram incorporar distribuições biparamétricas no componente aleatório do modelo. Definido por,

$$f(y; \theta, \phi) = \exp\{\phi^{-1}[y\theta - b(\theta)] + c(y, \phi)\}, \quad (2.1)$$

em que $b(\cdot)$ e $c(\cdot)$ são funções conhecidas.

A esperança e variância de uma distribuição pertencente a família exponencial são dadas por,

$$E(Y) = \mu = b'(\theta) \quad \text{e} \quad \text{Var}(Y) = \phi b''(\theta), \quad (2.2)$$

em que, $b''(\theta) = \frac{d\mu}{d\theta}$ é uma função de μ e pode ser representada por $V(\mu)$, função da variância e ϕ é o parâmetro de dispersão do modelo e seu inverso ϕ^{-1} , uma medida de precisão. O logaritmo da função de verossimilhança é definida da forma,

$$\ln f(y|\theta) = \ell(\theta|y),$$

portanto, a função escore

$$U = U(\theta) = \frac{\partial \ln(Y|\theta)}{\partial \theta}$$

por propriedade da função escore,

$$E(U) = 0 \quad \text{e} \quad E[U^2(\theta)] = -E\left[\frac{\partial^2 \ln(Y|\theta)}{\partial \theta^2}\right] = I(\theta),$$

onde $I(\theta)$ é a informação esperada de Fisher. Então:

$$\begin{aligned} U(\theta) &= \frac{\partial U(\theta)}{\partial \theta} = \phi^{-1}[y - b'(\theta)] \\ &\text{e} \\ U'(\theta) &= \frac{\partial^2 U(\theta)}{\partial \theta^2} = -\phi^{-1}b''(\theta). \end{aligned}$$

Portanto,

$$E(U) = \phi^{-1}[y - b'(\theta)] = 0 \quad \text{implica em} \quad E(Y) = b'(\theta)$$

assim,

$$\text{Var}(U) = E(U') = -\phi^{-1}b''(\theta) \quad \text{então,} \quad \text{Var}(U) = E(U^2) = -\phi^{-2}\text{Var}(Y).$$

Logo,

$$\text{Var}(Y) = \phi b''(\theta),$$

em que, $b'' = V(\mu)$. A função gama $\Gamma(\cdot)$ é dada por $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$, $\alpha > 0$.

A Tabela 1 apresenta algumas distribuições importantes na família exponencial, caracterizando as funções $b(\theta)$, $c(y, \phi)$ e a função da média e variância.

Tabela 1 – Algumas distribuições importantes da família exponencial

Distribuições	ϕ	θ	$b(\theta)$	$c(y, \phi)$	$\mu(\theta)$	$V(\mu)$
Normal: $N(\mu, \sigma^2)$	σ^2	μ	$\frac{\theta^2}{2}$	$-\frac{1}{2} \left[\frac{y^2}{\sigma^2} + \log(2\pi\sigma^2) \right]$	θ	1
Poisson: $P(\mu)$	1	$\log(\mu)$	e^θ	$-\log y!$	e^θ	μ
Binomial: $B(m, \pi)$	1	$\log\left(\frac{\mu}{m-\mu}\right)$	$m \log(1 + e^\theta)$	$\log\left(\frac{m}{y}\right)$	$\frac{me^\theta}{1+e^\theta}$	$\frac{\mu}{m}(m - \mu)$
Binomial Negativa: $BN(\mu, k)$	1	$\log\left(\frac{\mu}{\mu+k}\right)$	$-k \log(1 - e^\theta)$	$\log\left[\frac{\Gamma(k+y)}{\Gamma(k)y^i}\right]$	$\frac{ke^\theta}{1-e^\theta}$	$\mu\left(\frac{\mu}{k} - 1\right)$
Gama: $G(\mu, \nu)$	ν^{-1}	$-\frac{1}{\mu}$	$-\log(-\theta)$	$\nu \log(\nu y) - \log y - \log \Gamma(\nu)$	$-\frac{1}{\theta}$	μ^2
Normal Inversa: $IG(\mu, \sigma^2)$	σ^2	$-\frac{1}{2\mu^2}$	$-\left(-2\theta\right)^{1/2}$	$-\frac{1}{2} \left[\log(2\pi\sigma^2 y^3) + \frac{1}{\sigma^2 y} \right]$	$\left(-2\theta\right)^{-1/2}$	μ^3

Fonte: (Cordeiro e Demétrio, 2013)

2.1.3 Função Geradora de Momentos

Conforme (DEMÉTRIO, 2008), a função geradora de momentos da família exponencial é dada por,

$$M(t; \theta, \phi) = E(e^{tY}) = \exp\{\phi^{-1}[b(\phi t + \theta) - b(\theta)]\}, \quad (2.3)$$

Prova: Será feita apenas para o caso de variáveis aleatória contínuas. No caso discreto, basta substituir a integral pelo somatório.

Lembrando-se que

$$\int f(y; \theta, \phi) dy = 1,$$

chega-se a

$$\int \exp\{\phi^{-1}[\theta y - b(\theta)] + c(y, \phi)\} dy = 1,$$

obtendo -se

$$\int \exp[\phi^{-1}\theta y + c(y, \phi)] dy = \exp[\phi^{-1}b(\theta)].$$

Logo,

$$\begin{aligned} M(t; \theta, \phi) &= E(e^{tY}) = \int \exp(ty) f(y) dy \\ &= \int \exp\{\phi^{-1}[(\phi t + \theta)y - b(\theta)] + c(y, \phi)\} dy \\ &= \frac{1}{\exp[\phi^{-1}b(\theta)]} \int \exp[\phi^{-1}(\phi t + \theta)y + c(y, \phi)] dy \end{aligned}$$

conforme a equação (2.3) tem-se

$$M(t; \theta, \phi) = \frac{\exp[\phi^{-1}b(\phi t + \theta)]}{\exp[\phi^{-1}b(\theta)]}$$

ou ainda, demonstrando,

$$M(t; \theta, \phi) = \exp\{\phi^{-1}[b(\phi t + \theta) - b(\theta)]\}.$$

A função geradora de cumulantes (f.g.c.) correspondente é,então,

$$\varphi(t; \theta, \phi) = \log[M(t; \theta, \phi)] = \phi^{-1}[b(\phi t + \theta) - b(\theta)],$$

derivando-se, sucessivamente, em relação a t, tem-se

$$\begin{aligned}
\varphi'(t; \theta, \phi) &= \phi^{-1} b'(\phi t + \theta) \phi = b'(\phi t + \theta) \\
\varphi''(t; \theta, \phi) &= \phi b''(\phi t + \theta) \\
\varphi'''(t; \theta, \phi) &= \phi^2 b'''(\phi t + \theta) \\
&\vdots \\
\varphi^{(r)}(t; \theta, \phi) &= \phi^{r-1} b^{(r)}(\phi t + \theta).
\end{aligned}$$

Em que, $b^{(r)}$ indica a derivada de r-ésima ordem de $b(\cdot)$ em relação a t . Para $t = 0$, obtém-se o r-ésimo cumulante da família exponencial como

$$\begin{aligned}
\varphi^{(1)} &= b'(\theta) = E(Y) \\
\varphi^{(2)} &= \phi b''(\theta) = Var(Y).
\end{aligned}$$

Abaixo, tem-se a Tabela 2, onde, representa a função geradora de momentos para algumas distribuições de grande importância.

Tabela 2 – Algumas distribuições importantes da função geradora de momentos

Distribuição	Função geradora de momentos $M(t; \theta, \phi)$
Normal $N(\mu, \sigma^2)$	$\exp\left(t\mu + \frac{\sigma^2 t^2}{2}\right)$
Poisson $P(\mu)$	$\exp[\mu(e^t - 1)]$
Binomial $B(m, \pi)$	$\left(\frac{m-\mu}{m} + \frac{\mu}{m} e^t\right)^m$
Binomial Negativa $BN(\mu, k)$	$\left[1 + \frac{\mu}{k}(1 - e^t)\right]^{-k}$
Gama $G(\mu, \nu)$	$\left(1 - \frac{t\mu}{\nu}\right)^{-\nu}, t < \frac{\nu}{\mu}$
Normal Inversa $IG(\mu, \sigma^2)$	$\exp\left\{\frac{1}{\sigma^2} \left[\frac{1}{\mu} - \left(\frac{1}{\mu^2} - 2t\sigma^2\right)^{\frac{1}{2}}\right]\right\}, t < \frac{1}{2\sigma^2\mu^2}$

Fonte: (Cordeiro e Demétrio, 2013)

2.1.4 Estatística Suficiente

Segundo Cordeiro e Demétrio (2008), seja Y_1, \dots, Y_n uma amostra aleatória (a.a.) de uma distribuição que pertence à família exponencial, então, a distribuição conjunta de Y_1, \dots, Y_n é dada por

$$\begin{aligned}
f(y; \theta, \phi) &= \sum_{i=1}^n f(y_i; \theta, \phi) = \sum_{i=1}^n \exp\{\phi^{-1}[y_i\theta - b(\theta)] + c(y_i, \phi)\} \\
&= \exp\left\{\phi^{-1} \left[\theta \sum_{i=1}^n y_i - nb(\theta)\right]\right\} \exp\left[\sum_{i=1}^n c(y_i, \phi)\right].
\end{aligned}$$

Pelo teorema da fatoração de Neyman-Fisher, conforme Silvey (1975), tem-se que $T = \sum_{i=1}^n Y_i$ é uma estatística suficiente para θ , pois

$$f(y; \theta, \phi) = g(t, \theta)h(y_1, \dots, y_n), \quad (2.4)$$

sendo que $g(t, \theta)$ depende de θ e dos y 's apenas através de t e $h(y_1, \dots, y_n)$ independente de θ .

Isto mostra, que sob amostragem aleatória, se uma densidade pertence à família exponencial com um parâmetro, então, existe uma estatística suficiente. Na realidade, usando - se o Teorema de Lehmann-Scheffé. Mendenhall (1981) mostra-se que $T = \sum_{i=1}^n Y_i$ é uma estatística suficiente minimal.

2.2 Modelo Linear Generalizado

Os modelos lineares generalizados são uma extensão do modelo linear clássico, (SILVA, 2000)

$$\mathbf{Y} = X\beta + \epsilon,$$

onde, X é uma matriz de dimensão $n \times p$ de especificação do modelo associada a um vetor $\beta = (\beta_1, \dots, \beta_n)^T$ de parâmetros e ϵ é um vetor de erros aleatórios com distribuição que se supõe $N(\mathbf{0}, \sigma^2 \mathbf{I})$.

Estas hipóteses implicam obviamente que $E(\mathbf{Y}|Z) = \mu$ com $\mu = \mathbf{Z}\beta$, ou seja, o valor esperado da variável resposta é uma função linear das covariáveis.

(MCCULLAGH; NELDER, 1989) definem que MLGs são caracterizados pela seguinte estrutura:

1. Componente aleatório

Dados o vetor de covariáveis \mathbf{x}_i as variáveis Y_i são independentes com distribuição pertencente à família exponencial (2.1), com $E(Y_i|\mathbf{x}_i) = \mu_i = b'(\theta_i)$ para $i = 1, \dots, n$ e, possivelmente, um parâmetro de dispersão ϕ não dependente de i .

2. Componente sistemático ou estrutural

As variáveis explicativas entram na forma de uma soma linear de seus efeitos

$$\eta_i = \sum_{j=1}^p x_{ij}\beta_j = \mathbf{x}_i^T \beta, \quad \text{ou} \quad \boldsymbol{\eta} = \mathbf{X}\beta, \quad (2.5)$$

onde, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ representa a matriz do modelo, $\beta = (\beta_1, \dots, \beta_p)^T$ o vetor de parâmetros desconhecidos e $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$ o preditor linear.

3. Função de ligação

A função de ligação relaciona a média ao preditor linear, estabelecendo uma relação entre o componente aleatório ao componente sistemático. Então:

$$\eta_i = g(\mu_i), \quad (2.6)$$

sendo $g(\cdot)$ uma função monótona, derivável.

Assim, para a especificação do modelo, os parâmetros θ_i da família exponencial não são de interesse direto mas sim um conjunto menor de parâmetros $\beta_1, \beta_2, \dots, \beta_p$ tais que uma combinação linear dos β' s seja igual a alguma função do valor esperado Y_i .

2.2.1 Ligação Canônica

Se a função de ligação é escolhida da forma $\eta_i = g(\mu_i) = \theta_i$ o preditor linear modela diretamente o parâmetro canônico e tal função de ligação é chamada função de ligação canônica. Conforme (PAULA, 2013), supondo ϕ conhecido, o logaritmo da função de verossimilhança com respostas independentes pode ser expresso na forma

$$\ell(\beta) = \sum_{i=1}^n \phi\{y_i\theta_i - b(\theta_i)\} + \sum_{i=1}^n c(y_i; \phi).$$

Um caso importante ocorre quando o parâmetro θ coincide com o preditor linear, ou seja, $\theta_i = \eta_i = \sum_{j=1}^p x_{ij}\beta_j$. Então, $\ell(\beta)$ fica dado por

$$\ell(\beta) = \sum_{i=1}^n \phi\{y_i \sum_{j=1}^p x_{ij}\beta_j\} + \sum_{i=1}^n c(y_i; \phi).$$

Definindo a estatística $S_j = \phi \sum_{i=1}^n Y_i x_{ij}$, $\mathbf{L}(\beta)$

$$\ell(\beta) = \sum_{j=1}^p s_j \beta_j - \phi \sum_{i=1}^n b\left(\sum_{j=1}^p x_{ij}\beta_j\right) + \sum_{i=1}^n c(y_i; \phi).$$

Pelo teorema da fatorização a estatística $\mathbf{S} = (S_1, \dots, S_p)^T$ é suficiente minimal para o vetor $\beta = (\beta_1, \dots, \beta_p)^T$. Na tabela segue algumas ligações canônicas

2.2.2 Estimação dos Parâmetros

A estimação dos MLGs é baseada na máxima verossimilhança

Estimação de β

Segundo (DEMÉTRIO, 2008), se um MLG segue sua estrutura com expressões (2.1), (2.4) e (2.5) pressupondo que os dados é representado por um vetor $\mathbf{y} = (y_1, \dots, y_n)^T$. Dada a função de verossimilhança com o vetor \mathbf{y} é definida $\ell(\beta) = \ell(\beta; \mathbf{y})$ tem-se

Tabela 3 – Funções de ligações canônicas

Distribuição	Ligação canônica
Normal	Identidade: $\eta = \mu$
Poisson	Logarítmica: $\eta = \ln(\mu)$
Binomial	Logística: $\eta = \ln\left(\frac{\pi}{1-\pi}\right) = \ln\left(\frac{\mu}{m-\mu}\right)$
Gama	Recíproca: $\eta = \frac{1}{\mu}$
Normal Inversa	Recíproca ² : $\frac{1}{\mu^2}$

Fonte: (Cordeiro e Demétrio, 2013)

$$\ell(\beta) = \phi^{-1} \sum_{i=1}^n [y_i \theta_i - b(\theta_i)] + \sum_{i=1}^n c(y_i, \phi),$$

em que, $\theta_i = q(\mu_i)$, $\mu_i = g^{-1}(\eta_i)$ e $\eta_i = \sum_{r=1}^p x_{ir} \beta_r$.

Conforme Demétrio (2002), por propriedade a família exponencial de distribuições seus elementos satisfazem as condições de regularidades suficientes para garantir que o máximo global do logaritmo da função de verossimilhança $\ell(\theta; \mathbf{y})$ é dado unicamente pela solução da equação $\mathbf{U}_\theta = \frac{d\ell}{d\theta} = \mathbf{0}$, equivalente a $\mathbf{U}_\beta = \frac{d\ell}{d\beta} = \mathbf{0}$ (Cox & Hinkley, 1986). Então,

$$U_j = \sum_{i=1}^n \frac{\partial \ell(\theta_i; y_i, \phi)}{\partial \beta_j} = \sum_{i=1}^n \frac{\partial \ell_i}{\partial \beta_j}.$$

Pela regra da cadeia, tem-se

$$\begin{aligned} U_j &= \sum_{i=1}^n \frac{d\ell_i}{d\theta_i} \frac{d\theta_i}{d\mu_i} \frac{d\mu_i}{d\eta_i} \frac{\partial \eta_i}{\partial \beta_j} \\ &= \sum_{i=1}^n \frac{1}{a_i(\phi)} [y_i - b'(\theta_i)] \frac{1}{\mu_i / d\theta_i} \frac{d\mu_i}{d\eta_i} x_{ij}. \end{aligned}$$

Como já visto em (2.2), temos que:

$$U_j = \frac{\partial \ell}{\partial \beta} = \sum_{i=1}^n \frac{1}{a_i(\phi)} (y_i - \mu_i) \frac{1}{V(\mu_i)} \frac{d\mu_i}{d\eta_i} x_{ij}. \quad (2.7)$$

As equações $U_j = 0$, $j = 1, 2, \dots, p$ são não lineares e têm que ser resolvidas numericamente por processos iterativos de Newton-Raphson, em que, para solução de uma equação $f(x) = 0$ é baseado na aproximação de Taylor para a função $f(x)$ nas vizinhanças no ponto x_0 . Conforme Demétrio (2002), para obter a solução de U_β pelo método de Newton-Raphson, tem-se:

$$\hat{\beta}^{(\mathbf{m}+1)} = \hat{\beta}^{(\mathbf{m})} + (\mathbf{I}_0^{(-1)})^{(\mathbf{m})} \mathbf{U}^{(\mathbf{m})},$$

sendo, $\hat{\beta}^{(m)}$ e $\hat{\beta}^{(m+1)}$ os vetores de parâmetros estimados nos passos m e $m + 1$, $\mathbf{U}^{(m)}$ o vetor escore e $I(\cdot)$ é a matriz de informação de Fisher, dada por:

$$\mathbf{I}(\beta) = E \left[\frac{-\partial \ell(\beta; \mathbf{y})}{\partial \beta_j \partial \beta_k} \right] = E \left[\frac{\partial \ell(\beta; \mathbf{y})}{\partial \beta_j} \frac{\partial \ell(\beta; \mathbf{y})}{\partial \beta_k} \right]$$

então, tem-se

$$\begin{aligned} I(\beta) &= E \left[\left(\frac{(Y_i - \mu_i)x_{ij}}{a_i(\phi)V(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right) \left(\frac{(Y_i - \mu_i)x_{ik}}{a_i(\phi)V(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) \right) \right] \\ &= E \left[\frac{(Y_i - \mu_i)^2 x_{ij} x_{ik}}{a_i^2(\phi)[V(\mu_i)^2]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \right] \\ &= E \left[\frac{a_i(\phi)V(\mu_i)}{a_i^2(\phi)[V(\mu_i)^2]} x_{ij} x_{ik} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \right] \\ &= \frac{x_{ij} x_{ik}}{a_i(\phi)V(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2. \end{aligned}$$

Estimação de ϕ

Conforme (DEMÉTRIO, 2008), o método do desvio é baseado na aproximação χ_{n-p}^2 para o desvio escalonado. Para um modelo bem ajustado às observações é esperado que o desvio escalonado S_p tenha um valor esperado igual a $n - p$. Logo, as estimativa de ϕ é

$$\hat{\phi}_d = \frac{D_p}{n - p},$$

em que, o desvio D_p é obtido pela expressão (2.6) como função das observações y e dos valores ajustados $\hat{\mu}$

O método de Pearson é baseado na aproximação da distribuição de estatística de Pearson X_p^2 generalizada, dividida por ϕ , pela distribuição χ_{n-p}^2 . Então

$$\phi_p = \frac{1}{n - p} \sum_{i=1}^2 \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}.$$

Valida para o modelo normal $\hat{\phi}_d = \hat{\phi}_p$. Para os demais modelos contínuos, esses estimadores diferem em valor.

Segundo (DEMÉTRIO, 2008), é possível para estimar o parâmetro ϕ pelo método da máxima verossimilhança, porém, quando não existir o estimador de máxima verossimilhança torna - se desafiador computacionalmente e pode ser expresso da seguinte maneira

$$\ell(\beta, \phi) = \phi^{-1} \sum_{i=1}^n [y_i \theta_i - b(\theta_i)] + \sum_{i=1}^n c(y_i, \phi).$$

A função escore relativa ao parâmetro ϕ é dada por

$$U_\phi = \frac{\partial(\beta, \phi)}{\partial\phi} = \phi^{-2} \sum_{i=1}^n [y_i\theta_i - b(\theta_i)] + \sum_{i=1}^n \frac{dc(y_i, \phi)}{d\phi}.$$

Note que, U_ϕ é função de β através θ e de ϕ , supondo y conhecido.

2.3 Seleção e Validação de Modelos

De acordo com (SILVA, 2000), quando lidamos com muitos parâmetros, há um interesse em identificar qual modelo é mais parcimonioso. Isto é, aquele que possui o menor número de variáveis explicativas, mas que ainda proporciona uma interpretação eficaz do problema em questão e, ao mesmo tempo, se adequar de maneira precisa aos dados. O modelo ideal deve ser bem ajustado, parcimonioso e oferecer uma interpretação sólida dos dados. Como escrito na literatura segundo (DEMÉTRIO, 2008), considerando n observações, podendo ser ajustado até n parâmetros.

- **Modelo Nulo**

Considerado um modelo simples, com apenas um parâmetro, presume que todas as variáveis Y_i compartilham o mesmo valor médio μ . Embora seja um modelo de fácil interpretação, ele raramente consegue capturar a estrutura presente nos dados.

- **Modelo Completo**

Também chamado de modelo saturado, este modelo atribui toda variação dos dados à componente sistemática. Como as estimativas da máxima verossimilhança dos μ_i são as próprias observações $\hat{\mu}_i = y_i$, se ajusta de forma exata reproduzindo os próprios dados. No entanto, ele não é explicativo, uma vez que replica os dados sem fornecer uma interpretação significativa.

- **Modelo Minimal**

É o modelo mais simples, com menor número de parâmetros que se ajusta adequadamente aos dados.

- **Modelo Maximal**

É o modelo que contém o maior número de parâmetros que podem ser considerados, logo, é o modelo mais complexo.

- **Modelo Corrente**

O modelo corrente, é qualquer modelo com q parâmetros linearmente independentes situado entre o modelo maximal e minimal, e que está a ser sujeito a investigação.

2.3.1 Qualidade do Ajuste

Função Desvio

Conforme (PAULA, 2013), a qualidade do ajuste de um MLG é avaliada através da função desvio, consiste em uma distância entre o logaritmo da função de verossimilhança do modelo saturado (com n parâmetros) e do modelo corrente (com p parâmetros) avaliando a estimativa de máxima verossimilhança $\hat{\beta}$,

$$D^*(\mathbf{y}; \hat{\mu}) = \phi D(\mathbf{y}; \hat{\mu}) = 2L(\mathbf{y}; \mathbf{y}) - L(\hat{\mu}; y),$$

onde, $L(\mathbf{y}; \mathbf{y})$ é o logaritmo da função de verossimilhança do modelo saturado e $L(\hat{\mu}; y)$ é a estimativa de máxima verossimilhança do modelo corrente de μ_i , dada por, $\tilde{\mu}_i = y_i$. O modelo $D^*(\mathbf{y}; \hat{\mu})$ é denominado de desvio escalonado, já $D(\mathbf{y}; \hat{\mu})$ é chamada de desvio. Descrita na forma

$$D(\mathbf{y}; \hat{\mu}) = 2 \sum_{i=1}^n y_i (\tilde{\theta}_i - \hat{\theta}_i) + (b(\hat{\theta}_i) - b(\tilde{\theta}_i)), \quad (2.8)$$

em que, $\hat{\theta}_i = \theta(\hat{\mu}_i)$ e $\tilde{\theta}_i = \theta_i(\tilde{\mu}_i)$ as estimativas de máxima verossimilhança de θ para os modelos com p parâmetros ($p < n$) e saturado ($p = n$).

Estatística de Pearson Generalizada

Segundo (SILVA, 2000), outra medida de adequabilidade é a estatística de Pearson generalizada X_p^2 , cuja expressão é dada por

$$X_p^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)},$$

sendo $V(\hat{\mu}_i)$ a função de variância estimada sob o modelo ajustado aos dados.

2.3.2 Seleção de modelos

Nas análises estatísticas por meio de Modelos Lineares Generalizados, frequentemente nos deparamos com um grande número de parâmetros que podem ser potencialmente significativos na explicação da variabilidade e influência dos dados. Com o intuito de selecionar o modelo mais apropriado, de acordo com (SILVA, 2000), duas estatísticas frequentemente utilizadas para a seleção de modelos são o AIC e BIC.

Critério de Informação Akaike

O critério de informação de Akaike (AIC), criado por Akaike (1974), como descrito por (PAULA, 2013), tem como objetivo selecionar um modelo que seja parcimonioso. Uma vez que o logaritmo da função de verossimilhança $L(\beta)$ aumenta com o acréscimo de parâmetros no modelo, a essência do teste reside em identificar o modelo com menor valor para essa função

$$AIC = -2L(\hat{\beta}) + 2p,$$

onde, p é o número de parâmetros.

Critério de Informação de Bayes

O método de Schwartz (1978), conhecida por Critério de Informação de Bayes (BIC), conforme (PAULA, 2013), consiste em maximizar $L(\beta)$ também minimizando o número de coeficientes da regressão, com penalização diferente. O método equivalente a minimizar a função

$$BIC = -2L(\hat{\beta}) + p \log(n).$$

Para regressão múltipla tem - se que

$$BIC = n \log D(\mathbf{y}; \hat{\mu}/n) + p \log(n).$$

2.4 Técnicas de Diagnóstico

Resíduo de Pearson

Conforme (DEMÉTRIO, 2008), o resíduo mais simples é o de Pearson, definido por

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\hat{V}_i^{1/2}}.$$

É uma quantidade do componente da estatística de Pearson generalizada $X_p^2 = \sum_{i=1}^2 r_i^{P^2}$. A desvantagem do resíduo de Pearson é que sua distribuição é assimétrica para modelos não - normais.

Desvio Residual

De acordo com (SILVA, 2000), o desvio residual corresponde à i-ésima observação definida por

$$R_i^D = \delta_i \sqrt{d_i},$$

em que, $\delta_i = \text{sin}(\arcsin((y_i - \hat{\mu}_i) / \sqrt{\hat{\phi}(1 - h_{ii})}))$. O desvio residual padronizado é obtido dividindo o desvio residual R_i^D por $\sqrt{\hat{\phi}(1 - h_{ii})}$, logo

$$R_i^{*D} = \frac{R_i^D}{\sqrt{\hat{\phi}(1 - h_{ii})}}.$$

2.5 Materiais

Os dados foram obtidos no site do Sistema de Informações Ambientais Integrado a Saúde (SISAM)¹, trata-se de uma ferramenta de análise de dados pontuais e espaciais que combina informações de concentrações de poluentes oriundas e estimativas de emissões de queimadas e emissões urbanas/indústrias, dados de monitoramento de focos de queimadas e dados meteorológicos decorrido. Os dados eram coletados 4 vezes ao dia nos horários de

¹ <https://www2.cprh.pe.gov.br/sistema-integrado-de-servicos-ambientais-sisam/>

(00:00; 06:00; 12:00; 18:00), onde, para este estudo foi obtido a média mensal dos anos de 2015 à 2019 para os municípios de Campina Grande, João Pessoa e Patos localizados na Paraíba.

As variáveis em análise foram Material Particulado (Pm2.5), constituído por partículas de até 2,5 micrômetros, que representa uma complexa mistura de partículas sólidas e líquidas suspensas no ar, apresentando desafios significativos para a qualidade do ar e a saúde (JOHNSON; GRAHAM, 2005); Precipitação (Pe), desempenha um papel fundamental nos ciclos naturais da água e no funcionamento dos ecossistemas; Temperatura (Te), um dos principais indicadores das condições climáticas, sendo fundamental para compreensão das mudanças climáticas e previsão do tempo; Umidade Reativa do Ar (Um), desempenha um papel crucial na influência das condições atmosféricas sobre o conforto humano e saúde, afeta os processos meteorológicos, como a formação de nuvens e precipitação (GUARNIERI BIANCA OLIVIERI, 2023) e Vento (Ve), tem papel vital na redistribuição de calor e umidade na atmosfera, influenciando diretamente os padrões climáticos.

Será feito ajustes utilizando a distribuição gama com função de ligação log para Campina Grande e Patos, já para o município de João Pessoa será utilizado a distribuição gaussiana com função de ligação identidade. A variável resposta para todos os modelos será Pm2.5 e as demais variáveis independentes.

As análises foram realizadas no *software R* os pacotes utilizados foram *dplyr* (Wickham, 2014), *lubridate* (Grolemund; Wickham, 2014), *GGally* (Schloerke; Cook; Larmarange; Briatte; Wickham, 2012), *car* (Fox; Weisberg, 2011), e para visualizar os pontos de queimadas na Paraíba foi utilizado os pacotes *maptools* (Bivand; Lewin-Koh, 2022) e *spatstat* (Baddeley; Turner, 2000).

3 RESULTADOS E DISCUSSÃO

Conforme (SPOSITO et al., 2007), na análise de processos pontuais estuda-se a ocorrência de eventos no espaço, considerando a localização dos eventos como aleatória. Obtivemos o shape do estado da Paraíba no site do Instituto Brasileiro de Geografia e Estatística (IBGE) com o propósito de monitorar os pontos de queimadas no estado. É importante notar que as queimadas também resultam na emissão de material particulado, a análise foi realizada para os cinco anos em estudo.

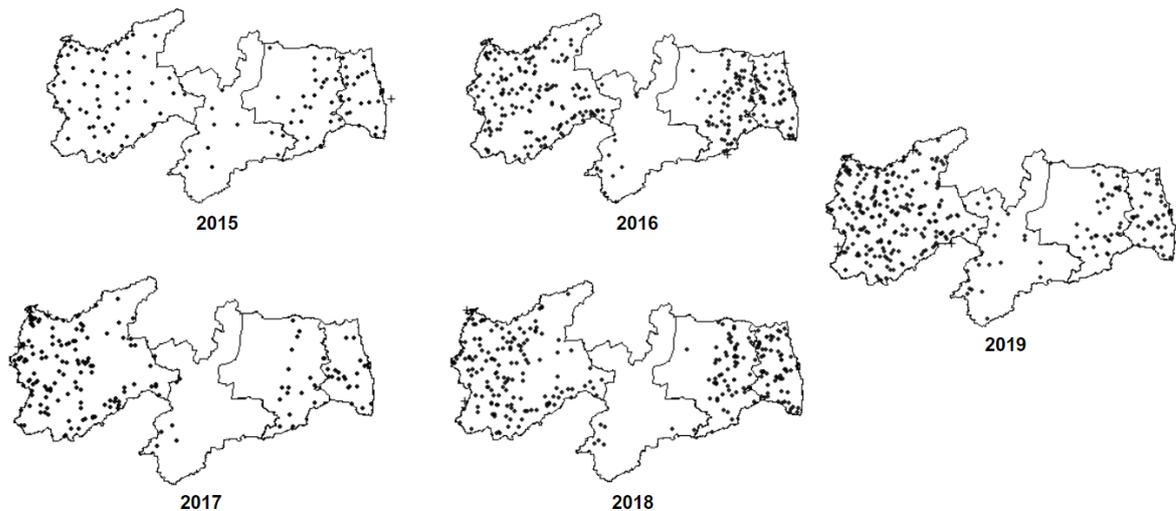


Figura 1 – Pontos de queimadas ao longo dos anos de 2015 à 2019 no estado da Paraíba

Fonte: Elaborada pelo autor, 2023.

Conforme demonstrado na Figura 5, podemos observar que as mesorregiões da Paraíba que apresentaram a maior incidência de queimadas foram o Sertão Paraibano, o Agreste e a Mata Paraibana. Estas são as áreas em que estão localizados os três municípios que foram objeto de análise.

O Sertão Paraibano se destaca, pois teve as maiores incidências de queimadas ao longo dos cinco anos analisados. Essas observações são importantes para entender a distribuição geográfica das queimadas e as potenciais fontes de material particulado gerada pelas queimadas no estado da Paraíba.

3.1 Análise exploratória

Nas Tabelas 4, 5 e 6, apresentam-se as análises descritivas de todas as variáveis em estudo. Notavelmente, o município de Patos registrou o maior índice de Pm2.5, com uma concentração de $19,9\mu\text{g}/\text{m}^3$, enquanto João Pessoa se destacou por apresentar a menor concentração, com $11,7\mu\text{g}/\text{m}^3$. Este dado é de grande relevância, uma vez que as

partículas Pm2.5 possuem um tamanho diminuto, o que as torna capazes de penetrar profundamente no sistema respiratório, chegando até os alvéolos pulmonares (CETESB, 2020).

Quanto às demais variáveis, como precipitação, umidade e temperatura, observamos que, em média, elas apresentam valores próximos entre os municípios, conforme estudos realizado por (RODRIGUES SAMYA DE LARA PINHEIRO, 2017). Entretanto, devido à localização em diferentes mesorregiões da Paraíba, cada município possui características e variações distintas.

Tabela 4 – Análise descritiva das variáveis explicativas e da variável resposta para o município de Campina Grande

Variável	Mínimo	Média	Variância	Desvio Padrão	Máximo	Assimetria	Curtose
Pm2.5	4,95	7,93	2,94	1,71	12,69	0,67	0,59
Precipitação	0	1,73	2,69	1,64	7,46	1,34	1,42
Temperatura	20,92	24,48	1,80	1,34	26,45	-0,68	-0,25
Umidade	71,28	84,32	71,99	8,48	96,10	-0,08	-1,61
Vento	92,94	123,58	170,72	13,07	149,23	-0,23	-0,86

Fonte: Elaborada pelo autor, 2023.

Tabela 5 – Análise descritiva das variáveis explicativas e da variável resposta para o município de Patos

Variável	Mínimo	Média	Variância	Desvio Padrão	Máximo	Assimetria	Curtose
Pm2.5	3,09	8,53	19,30	4,39	19,92	0,73	-0,35
Precipitação	0	1,25	2,71	1,65	6,54	1,44	1,22
Temperatura	20,07	26,95	1,21	1,10	28,60	-0,50	-0,89
Umidade	53,14	72,31	193,36	13,91	92,68	0,25	-1,76
Vento	79,19	114,94	184,39	13,58	141,64	-0,41	-0,56

Fonte: Elaborada pelo autor, 2023.

Tabela 6 – Análise descritiva das variáveis explicativas e da variável resposta para o município de João Pessoa

Variável	Mínimo	Média	Variância	Desvio Padrão	Máximo	Assimetria	Curtose
Pm2.5	4,92	8,07	1,76	1,33	11,67	-0,01	0,19
Precipitação	0,03	3,30	11,97	3,46	14,33	1,54	1,61
Temperatura	23,47	26,04	1,11	1,05	27,45	-0,60	-0,71
Umidade	73,09	84,56	69,11	8,31	96,51	0,20	-0,11
Vento	93,63	130,87	244,77	15,65	162,05	-0,11	-0,59

Fonte: Elaborada pelo autor, 2023.

A análise de correlação desempenha um papel crucial em nossa pesquisa, permitindo-nos explorar as relações entre diferentes variáveis e compreender suas interações. No caso

dos três municípios que examinamos, observamos uma correlação positiva entre a variável resposta Pm2.5 e as variáveis explicativas precipitação, temperatura e umidade. No entanto, notou-se uma correlação negativa com a variável vento. Em outras palavras, nossos resultados sugerem que um aumento na concentração de Pm2.5 na atmosfera está associado a um impacto mais significativo nos ciclos ecológicos.

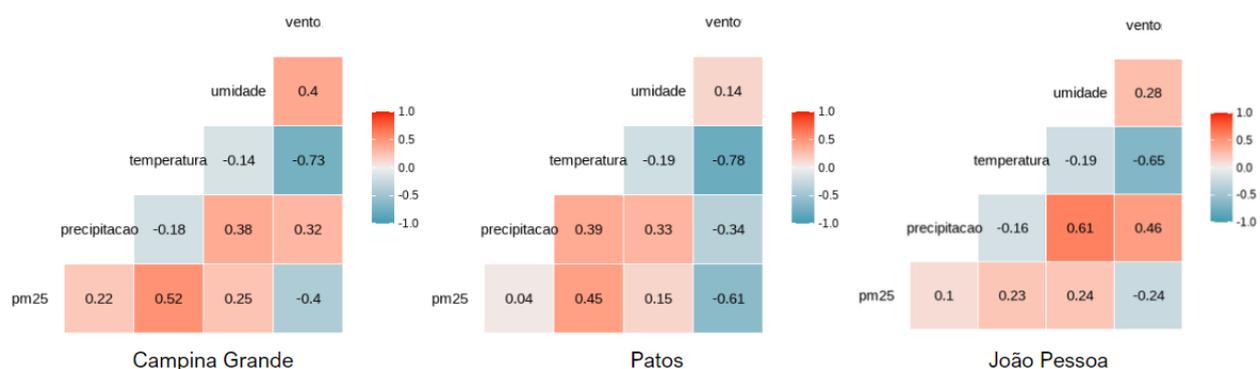


Figura 2 – Correlação entre as variáveis explicativas e variável resposta dos municípios de Campina Grande, Patos e João Pessoa

Fonte: Elaborada pelo autor, 2023.

A estatística descritiva desempenha um papel fundamental na seleção de candidatos apropriados para descrever a distribuição de um conjunto de distribuições paramétricas sob análise. Além disso, o uso de medidas como a assimetria e curtose, relacionadas ao terceiro e quarto momentos, oferece uma abordagem para a tomada de decisões sobre qual distribuição seguir (DELIGNETTE-MULLER; DUTANG, 2015). Quando a assimetria possui um valor diferente de zero, isso indica uma falta de simetria na distribuição empírica, enquanto o valor da curtose fornece uma medida do peso das caudas em comparação com a distribuição normal, na qual a curtose é igual a 3.

3.2 Ajuste de modelos

Para o município de Campina Grande, observou-se que apenas a variável vento não apresentou relevância para o modelo. Além disso, os dados exibiram um ajuste satisfatório à distribuição gama. A escolha da função de ligação foi determinada com base no critério de seleção AIC (Critério de Informação de Akaike), e a função de ligação log foi selecionada devido ao fato de resultar no menor valor de AIC. Na Tabela 7 são apresentadas as estimativas dos parâmetros para o modelo gama em Campina Grande.

Tabela 7 – Estimativas dos parâmetros referente ao modelo gama ajustado para o município de Campina Grande

Coeficientes		Estimativas	Erro Padrão	P-Valor
Constante	β_0	-0,900	0,432	0,041608
Precipitação	β_1	0,047	0,012	0,000459
Temperatura	β_2	0,093	0,015	5.79e-08
Umidade	β_3	0,007	0,002	0,006968

Fonte: Elaborada pelo autor, 2023.

A estimativa do parâmetro de dispersão é de $\hat{\phi} = 0,0234$, para saber se um modelo é bem ajustado aos dados é feito a análise do desvio, que consiste em avaliar a qualidade do ajuste, testando com $p < 0,05$ de significância obteve-se que $D^*(y; \hat{\mu}) = 0,842$, ou seja, tem-se evidências que o modelo é bem ajustado aos dados.

Conforme observado por (PéREZ, 2022), devido ao uso da função de ligação logarítmica no modelo, podemos interpretar os coeficientes exponenciados, ou seja, e^{β_j} , como efeitos multiplicativos no valor esperado. Portanto, para a variável de precipitação, temos $e^{0,047} = 1,048$; para a variável de temperatura, $e^{0,093} = 1,097$; e para a variável de umidade, $e^{0,007} = 1,001$.

A estimação do modelo é dada por:

$$\hat{y} = e^{-0,900+0,047Pr+0,093Te+0,007Um},$$

- Para Precipitação, a cada milímetro que aumenta corresponde a um acréscimo de 5% em Pm2.5, mantendo todas as outras variáveis constantes.
- Para Temperatura, indica que um aumento em $1^\circ C$ resulta em um acréscimo de cerca de 10% em Pm2.5, mantendo todas as outras variáveis constantes.
- Para Umidade relativa do ar, indica que a cada acréscimo de 1 unidade, haverá um aumento de 1% em Pm2.5, mantendo todas as outras variáveis constantes.

Esses resultados são preocupantes, uma vez que altas concentrações desse material podem estar associadas a temperaturas mais elevadas, precipitações descontroladas (como eventos de chuva intensa) e baixa umidade na atmosfera. Isso reflete desafios significativos no município, especialmente à luz das condições atuais.

Com intuito de observar se o modelo foi bem ajustado aos dados, é feito a análise de resíduos e apresentada por meio do gráfico diagnostico conforme a Figura 2, é observado se os dados são homogêneos (A), se seguem normalidade (B), homocedásticos (C) e se apresenta outlier (D). Os elementos [22],[39] e [42] graficamente aparentam ser pontos discrepantes, porém, foi analisado e todos estão dentro da métrica dos dados não apresentando discrepância.

A análise será baseada nos resíduos e resíduos de Pearson padronizados é essencial para garantir que o modelo seja apropriado para os dados e que não haja valores atípicos ou problemas substanciais que possam afetar a validade das inferências feitas a partir do modelo. É uma prática comum na modelagem estatística para verificar a qualidade do ajuste e a robustez das conclusões tiradas com base nos resultados do modelo.

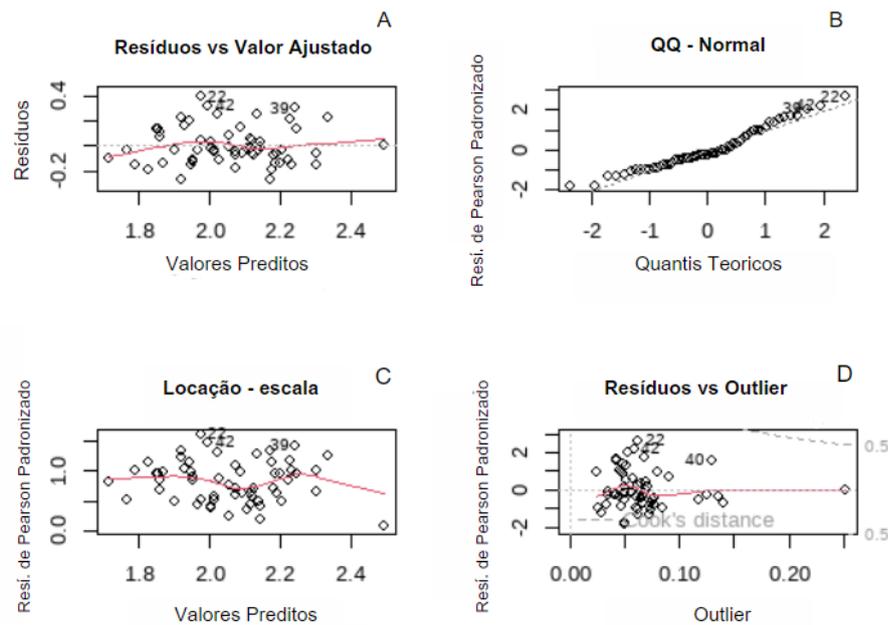


Figura 3 – Análise de diagnóstico referente ao ajuste do modelo gama para o município de Campina Grande

Fonte: Elaborada pelo autor, 2023.

Ao analisar os gráficos, observamos que os dados são homogêneos Figura 2A e 2C, em que, no gráfico 2A os pontos estão dispersos de forma aleatória e a linha em vermelho segue de forma horizontal, sendo este analisado pelos resíduos; o gráfico 2C é analisado pelo resíduo de Pearson padronizado, indicando também que são homocedásticos. A Figura 2B sugere que os resíduos seguem uma distribuição normal. Isso é evidenciado pela linha de teste no gráfico. Além disso, aplicamos o teste de Shapiro-Wilk (1965) com um nível de significância de 5%, e o resultado foi um $valor - p = 0,197$, o que confirma que os resíduos seguem uma distribuição normal. Em relação à presença de outliers (valores atípicos ou aberrantes), analisamos graficamente os resíduos de Pearson padronizados e não identificamos a presença de pontos atípicos.

Em Patos, da mesma forma que em Campina Grande, os dados se caracterizaram e se ajustaram bem a uma distribuição gama com função de ligação log, que foi escolhida com base no critério AIC. A diferença notável é que a variável temperatura não demonstrou relevância para o modelo em Patos, ao passo que a variável vento desempenhou um papel

significativo nesse município.

Vale ressaltar que Patos está situada na mesorregião do sertão paraibano. De acordo com (FRANCISCO, 2017), a distribuição espacial da temperatura média mensal do ar, de janeiro a dezembro, no estado da Paraíba, apresenta maiores valores de temperatura média ao longo do ano em áreas de baixa altitude. Esse fenômeno ocorre principalmente no Litoral e na mesorregião do sertão da Paraíba, que também se caracterizam por apresentar baixos índices de precipitação. Na Tabela 8 são apresentadas as estimativas dos parâmetros para o município de Patos.

Tabela 8 – Estimativas dos parâmetros referente ao modelo gama ajustado para o município de Patos

	Coefficientes	Estimativas	Erro Padrão	P-Valor
Constante	β_0	4,392	0,545	6.25e-11
Precipitação	β_1	-0,082	0,035	0,02215
Umidade	β_2	0,011	0,004	0,00482
Vento	β_3	-0,026	0,004	3.80e-08

Fonte: Elaborada pelo autor, 2023.

A estimativa do parâmetro de dispersão do modelo é $\hat{\phi} = 0,1626$. Avaliando a análise do desvio ao nível $p < 0,05$, obteve-se que, $D^*(y; \hat{\mu}) = 0,289$ indicando ser um modelo bem ajustado aos dados.

Da mesma forma que no município de Campina Grande, devido à utilização da função logarítmica no modelo, podemos interpretar os coeficientes exponenciais da seguinte maneira: para a precipitação, temos $e^{-0,082} = 0,921$; para a umidade, temos $e^{0,011} = 1,011$; e para o vento, temos $e^{-0,026} = 0,974$.

As estimativas do modelo é dada por:

$$\hat{y} = e^{4,39-0,082Pr+0,011Um-0,026Ve},$$

- Para Precipitação, a cada acréscimo de 1 milímetro corresponde a uma diminuição de 8% na concentração de material particulado, mantendo os demais parâmetros constantes.
- Para Umidade relativa do ar, a cada aumento em 1 unidade é observado um aumento de 1% de Pm2.5, mantendo os demais parâmetros constantes.
- Para Vento, a cada acréscimo de 1 km/h resulta numa redução de 3% de material particulado, mantendo os demais parâmetros constantes.

Com intuito de avaliar se o modelo é bem ajustado para os dados no município de Patos, foi feita a análise dos resíduos Figura 3, assim como em Campina Grande

graficamente foi identificado alguns pontos aberrantes [21], [22] e [30], porém, quando analisados todos estão na métrica de elementos em análise não apresentando discrepância.

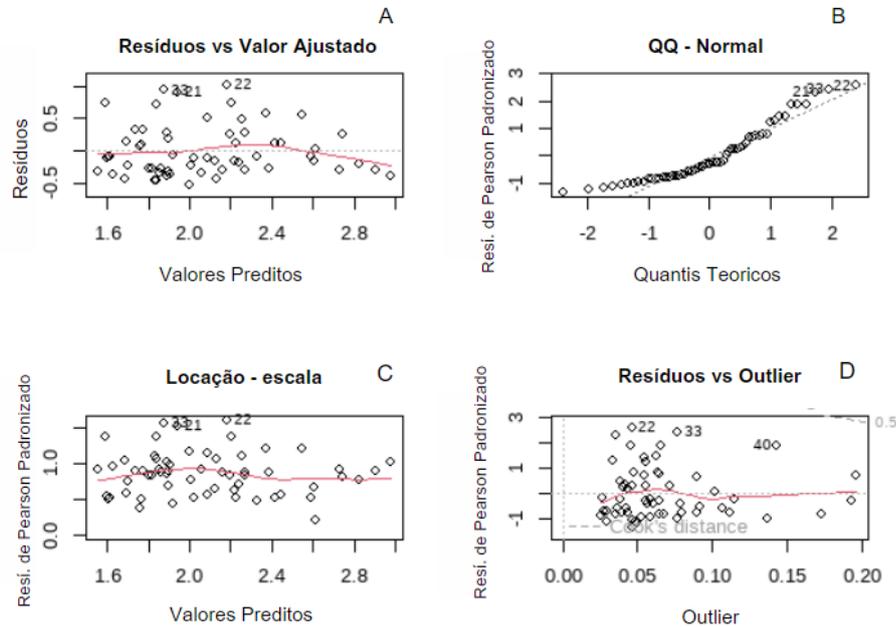


Figura 4 – Análise de diagnóstico referente ao ajuste do modelo gama para o município de Patos

Fonte: Elaborada pelo autor, 2023.

Com base na análise dos gráficos da Figura 3, Os dados exibem homocedasticidade, o que significa que a variabilidade dos erros é constante em todos os níveis das variáveis independentes (conforme ilustrado nos gráficos 3A e 3C).

Entretanto, ao avaliar se os resíduos seguem uma distribuição normal, como demonstrado no gráfico 3B, notamos que os dados seguem parcialmente a linha de teste, com a presença de alguns pontos fora da linha. Para aprofundar a análise quanto à suposição de normalidade, aplicamos o teste de Shapiro-Wilk (1965) com um nível de significância de 5%, resultando em um $valor - p = 0,01$. Isso indica que os dados não seguem uma distribuição normal.

Além disso, ao verificar a presença de outliers nos dados, conforme mostrado na Figura 3D, não identificamos nenhum ponto atípico. É importante observar que os pontos destacados no gráfico estão dentro dos critérios da métrica de estudo e não afetam o modelo.

João Pessoa, a capital da Paraíba situada na mesorregião da mata paraibana, é conhecida por seu clima tropical quente e úmido. Os elevados índices pluviométricos na região são principalmente atribuídos à umidade vinda do Oceano Atlântico. A distribuição de dados que se mostrou mais apropriada para este município foi a distribuição gaussiana, com função de ligação identidade que se assemelha a uma regressão clássica.

No contexto de João Pessoa, as variáveis precipitação e temperatura não demonstraram relevância para o ajuste do modelo estatístico. Por outro lado, as variáveis umidade e vento desempenharam um papel significativo na modelagem deste município. A Tabela 9 apresenta as estimativas obtidas para o município de João Pessoa.

Tabela 9 – Estimativas dos parâmetros referente ao modelo gaussiano ajustado para o município de João Pessoa

Coefficientes		Estimativas	Erro Padrão	P-Valor
Constante	β_0	6,449	0,029	6.36e-06
Umidade	β_1	0,060	0,0002	0,00196
Vento	β_2	-0,026	0,0001	0,00604

Fonte: Elaborada pelo autor, 2023.

A estimativa do parâmetro de dispersão do modelo é $\hat{\phi} = 1,4262$. Pela análise dos resíduos e as pressuposições de uma análise de regressão clássica o modelo foi bem ajustado aos dados.

Como a distribuição gaussiana com função de ligação identidade se assemelha a uma análise de regressão clássica a interpretação para os parâmetros é de forma semelhante.

A estimativa do modelo é dada por

$$\hat{y} = 6,45 + 0,06Um - 0,03Ve,$$

- A cada aumento de uma unidade na umidade relativa do ar, prevê-se um acréscimo de 0,06 unidades no valor esperado do material particulado, mantendo os outros parâmetros constantes.
- Ao considerar a influência do Vento, a cada quilômetro adicional, espera-se uma redução de 0,03 unidades no valor médio previsto de Pm2.5, mantendo todas as outras variáveis constantes.

Para avaliar a adequação do modelo ajustado, realizamos uma análise de diagnóstico usando os resíduos e os resíduos de Pearson padronizados, conforme ilustrado na Figura 4. Similarmente aos ajustes anteriores para os municípios, dois pontos foram destacados, são eles [22] e [38]. Assim como nos ajustes anteriores, esses pontos foram analisados separadamente e não indicam problemas significativos nos ajustes do modelo.

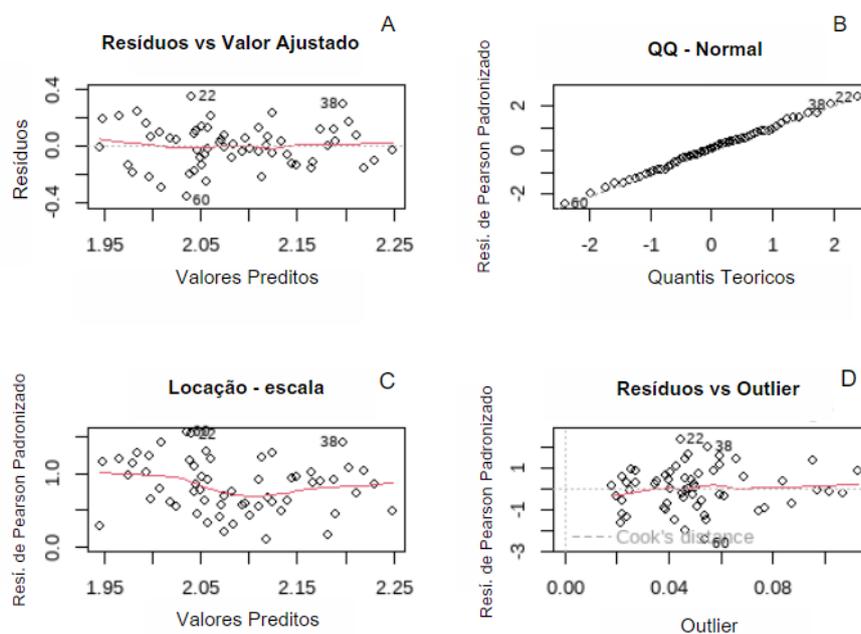


Figura 5 – Análise de diagnóstico referente ao ajuste do modelo gama para o município de João Pessoa

Fonte: Elaborada pelo autor, 2023.

Com base nas análises dos gráficos na Figura 4, os resíduos exibem homogeneidade, o que significa que os pontos estão dispersos de forma aleatória, e a linha vermelha segue horizontalmente (conforme ilustrado nos gráficos 4A e 4C). Em relação à normalidade dos dados, ao analisar graficamente, observamos que eles seguem uma distribuição normal. Além disso, o teste de Shapiro-Wilk realizado a um nível de significância de 5% resultou em um $valor - p = 0,942$, indicando fortemente que os dados seguem uma distribuição normal (conforme ilustrado no gráfico 4B). Por fim, ao verificar a presença de pontos discrepantes, identificamos três elementos destacados [22], [33] e [60]. No entanto, após análise minuciosa, ficou evidente que esses pontos estão dentro da mesma faixa métrica dos demais elementos do modelo, indicando que não prejudicam o ajuste do modelo (conforme demonstrado na Figura 4D).

4 CONCLUSÃO

Os modelos lineares generalizados (MLGs) representam uma extensão significativa dos modelos lineares tradicionais, permitindo a modelagem de uma ampla gama de distribuições e lidando com dados que não seguem a distribuição normal e os pressupostos tradicionais. Com a flexibilidade oferecida pelos MLGs, é possível explorar as relações entre variáveis de maneira mais abrangente e realizar análises mais precisas e adaptadas às características dos dados em estudo. Isso torna os MLGs uma ferramenta poderosa na análise estatística e na compreensão das complexidades dos fenômenos.

As variáveis em análise, de natureza contínua com comportamento assimétrico positivo e dados simétricos positivos, mostraram-se adequadas para uma abordagem por meio do modelo gama com função de ligação log. Essa escolha se revelou de grande importância em um estudo sobre material particulado e seu impacto nos ciclos naturais do meio ambiente, bem como nas consequências que isso acarreta para a saúde da população.

Grandes concentrações de material particulado nos municípios podem causar alterações nos ciclos naturais do meio ambiente, também conhecidos como ciclos biogeoquímicos. Esses ciclos representam os processos naturais pelos quais ocorre a reciclagem dos elementos, ou seja, a transferência de componentes físico-químicos dos organismos vivos de volta para o meio ambiente. Esse fenômeno pode desencadear condições climáticas desfavoráveis, como temperaturas elevadas, precipitação irregular e umidade relativa do ar descontrolada.

Conforme evidenciado em um estudo conduzido por (RODRIGUES SAMYA DE LARA PINHEIRO, 2017), foi identificado que essas variáveis climáticas cooperam significativamente para a compreensão do impacto do material particulado no meio ambiente. Isso destaca a importância de considerar não apenas os efeitos diretos do material particulado, mas também como ele pode influenciar os processos naturais e o clima local.

A saúde da população tem grande impacto com a exposição ao material particulado e as alterações no meio ambiente, os principais públicos afetados são as crianças e idosos, causando doenças respiratórias conforme (SILVA et al., 2013), estudos epidemiológicos no Brasil têm evidenciado associações de material particulado com a incidência de doenças. (CARABALLO, 2012) ressaltam que os gases e partículas ultrafinas provenientes da queima de combustíveis, principalmente do diesel, são fatores responsáveis pela maior incidência de asma brônquica e outras doenças alérgicas em indivíduos.

Pesquisas, como as de (CARABALLO, 2012), enfatizam que os gases e partículas ultrafinas originados da queima de combustíveis, especialmente do diesel, desempenham um papel significativo no aumento da incidência de condições como asma brônquica e outras doenças alérgicas em indivíduos. Essas descobertas destacam a importância de entender os riscos à saúde associados à poluição do ar e fornecem embasamento para a adoção de medidas de controle e políticas públicas visando à proteção da saúde da população, especialmente daqueles mais suscetíveis aos efeitos adversos.

REFERÊNCIAS

- ANDRADE PAULO ARTAXO, S. H. C. N. C. V. S. Influência do material particulado (pm2.5) de queimadas e variáveis meteorológicas na morbidade respiratória de crianças em manaus, . *REVISTA GEONORTE*, 2012. Citado na página 13.
- ARAUJO, N. M. d. R. Julia Manfredini de. Poluição atmosférica associada ao material particulado no estado de São Paulo: Análise baseada em dados de satélite. *Brazilian Journal of Environmental Sciences*, 2020. Citado na página 13.
- CARABALLO, J. M. S. Características físico-químicas de los gases y partículas contaminantes del aire. su impacto en el asma. *Iatreia*, v. 25, n. 4, oct. 2012. Disponível em: <<https://revistas.udea.edu.co/index.php/iatreia/article/view/13133>>. Citado na página 37.
- CETESB. *Qualidade do ar no estado de São Paulo 2019/CETESB*. [S.l.]: <<http://cetesb.sp.org/ar/publicacoes-relatorio>>, 2020. 228 p. Citado na página 29.
- DELIGNETTE-MULLER, M. L.; DUTANG, C. fitdistrplus: An r package for fitting distributions. *Journal of Statistical Software*, 2015. Disponível em: <<https://www.jstatsoft.org/index.php/jss/article/view/v064i04>>. Citado na página 30.
- DEMÉTRIO, G. M. C. e C. G. B. *Modelos Lineares Generalizados e Extensão*. 1. ed. Piracicaba: Centro de Documentação e Disseminação de Informações. Fundação Instituto Brasileiro de Geografia e Estatística, 2008. Citado 6 vezes nas páginas 15, 18, 21, 23, 24 e 26.
- FILHO, V. S. d. A. et al. Aerosols from biomass burning and respiratory diseases in children, manaus, northern brazil. *REVISTA DE SAÚDE PÚBLICA*, 2013. Citado na página 13.
- FRANCISCO, D. S. P. R. M. *Climatologia do Estado da Paraíba*. [S.l.]: EDUFPG/UFPG, 2017. Citado na página 33.
- GUARNIERI BIANCA OLIVIERI, G. S. A. V. G. Relative humidity and its impact on the immune system and infections. *International Journal of Molecular Sciences*, Department of Medicine, University of Verona, 37134 Verona, Italy, 2023. Citado na página 27.
- JOHNSON, P. R. S.; GRAHAM, J. J. Fine particulate matter national ambient air quality standards: Public health impact on populations in the northeastern united states. *Environmental health perspectives*, National Institute of Environmental Health Sciences. National Institutes of Health. Department of Health, Education and Welfare, v. 113, n. 9, p. 1140–1147, 2005. Citado na página 27.
- KAUFMAN, Y. et al. Smoke, clouds, and radiation-brazil (scar-b) experiment. *Journal of Geophysical Research: Atmospheres*, v. 103, 1999. Citado na página 13.
- MCCULLAGH, P.; NELDER, J. A. *Generalized linear models (Second edition)*. [S.l.]: London: Chapman & Hall, 1989. 500 p. Citado na página 20.
- NELDER, J. A.; WEDDERBURN, R. W. M. Smoke, clouds, and radiation-brazil (scar-b) experiment. *Journal of the Royal Statistical Society, Series A, General*, v. 135, 1972. Citado 2 vezes nas páginas 13 e 15.

- PAULA, G. A. *Modelos de regressão: com apoio computacional*. [S.l.]: IME-USP, 2013. Citado 3 vezes nas páginas 21, 25 e 26.
- PÉREZ, F. L. *Modelos Lineares Generalizados*. [S.l.]: Catálogo USP, 2022. Citado na página 31.
- Radojevic, M.; Hassan, H. Air quality in Brunei Darussalam during the 1998 haze episode. *Atmospheric Environment*, v. 33, n. 22, p. 3651–3658, 1999. Citado na página 13.
- RIBEIRO, H.; ASSUNÇÃO, J. V. de. Efeitos das queimadas na saúde humana. *Estudos Avançados*, 2002. Citado na página 13.
- ROCHA, L. R. L. A correlação entre doenças respiratórias e o incremento das queimadas em alta floresta e peixoto de azevedo norte do mato grosso - amazônia legal. *Revista Brasileira de Políticas Públicas*, 2016. Citado na página 13.
- RODRIGUES SAMYA DE LARA PINHEIRO, W. J. E. I. S. d. S. H. Poliany Cristiny de O. Variabilidade climática aumenta a morbimortalidade associada ao material particulado. *Revista de Saúde Pública*, 2017. ISSN 0034-8910. Disponível em: <<https://www.redalyc.org/articulo.oa?id=67249591090>>. Citado 2 vezes nas páginas 29 e 37.
- SILVA, A. M. C. d. et al. Material particulado originário de queimadas e doenças respiratórias. *Revista de Saúde Pública*, Faculdade de Saúde Pública da Universidade de São Paulo, v. 47, n. 2, p. 345–352, Apr 2013. ISSN 0034-8910. Disponível em: <<https://doi.org/10.1590/S0034-8910.2013047004410>>. Citado na página 37.
- SILVA, M. A. A. T. e G. L. *Modelos Lineares Generalizados - da Teoria à Prática*. [S.l.]: Edições SPE, 2000. Citado 4 vezes nas páginas 20, 24, 25 e 26.
- SPOSITO, M. B. et al. Spatial pattern of trees affected by black spot in citrus groves in brazil. *Plant Disease*, 2007. Citado na página 28.