



**UNIVERSIDADE ESTADUAL DA PARAÍBA**

**CENTRO DE CIÊNCIAS E TECNOLOGIA**

**DEPARTAMENTO DE ESTATÍSTICA**

**ALVARO AUGUSTO DO NASCIMENTO**

**Análise da estrutura de autocorrelação espacial dos casos  
notificados por COVID-19 nos municípios do estado da Paraíba**

**Campina Grande - PB**

**2022**

ALVARO AUGUSTO DO NASCIMENTO

**Análise da estrutura de autocorrelação espacial dos casos  
notificados por COVID-19 nos municípios do estado 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 em cumprimento às exigências legais para obtenção do título de Bacharel em Estatística.

Orientador: PROF. DR. RICARDO ALVES DE OLINDA

Campina Grande - PB

2022

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.

N244a Nascimento, Álvaro Augusto do.  
Análise da estrutura de autocorrelação espacial dos casos notificados por COVID-19 nos municípios do estado da Paraíba [manuscrito] / Alvaro Augusto do Nascimento. - 2022.  
62 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2022.

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

1. Análise Espacial. 2. Estatística Espacial. 3. COVID-19.  
4. Saúde pública. I. Título

21. ed. CDD 519.536

ALVARO AUGUSTO DO NASCIMENTO

## **Análise da estrutura de autocorrelação espacial dos casos notificados por COVID-19 nos municípios do estado 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 em cumprimento às exigências legais para obtenção do título de Bacharel em Estatística.

Aprovado em: 1º/07/2022.

### **BANCA EXAMINADORA**



---

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



---

Prof. Dr. João Domingos Scalon  
Universidade Federal de Lavras (UFLA)



---

Prof. Dr. Tiago Almeida de Oliveira  
Universidade Estadual da Paraíba (UEPB)

## **AGRADECIMENTOS**

Primeiramente agradeço a Deus pelo dom da vida.

Dedico esse trabalho aos meus pais Severino Marinho (in memoriam) e Alcinda Rodrigues (in memoriam), com todo amor do mundo e gratidão.

Um Agradecimento especial à minha esposa Tania Maria, aos meus filhos Lays, Laryssa e Alvaro Filho, aos meus genros Thales e Júlio César e principalmente a minhas netinhas Lívia e Maria Laura.

Agradeço ao meu orientador, Prof. Dr. Ricardo Alves de Olinda pela sua dedicação e pela orientação deste trabalho e a todos os professores do Departamento de Estatística da UEPB pelos conhecimentos transmitidos durante esta caminhada.

Agradeço, também, aos amigos que conquistei durante o curso de Estatística, em especial aos que iniciaram essa jornada comigo na turma 16.1 e permaneceram até o fim do curso: Eduardo Gomes, Rafaela e minha grande amiga Viviane Costa por quem tenho grande admiração. Também aos amigos Gilmar, Mateus Peixoto e Wylliam.

“A imaginação é a metade da doença; a tranquilidade é a metade do remédio; e a paciência é o primeiro passo para a cura. “

- Ibn Sina (980-1037), também conhecido como Avicena -

## RESUMO

O fim do ano de 2019 até o momento está marcado pelo desafio do enfrentamento da pandemia causada pelo novo coronavírus, SARS-CoV-2, agente etiológico responsável pela COVID-19. O primeiro caso de COVID-19 na Paraíba foi registrado em 18 de março de 2020 no município de João Pessoa, que chega a 396.442 casos no dia 30 de junho de 2021. O objetivo deste trabalho é mapear e analisar a evolução dos casos da COVID-19 no estado da Paraíba, nos meses de janeiro a junho de 2021. Trata-se de um estudo ecológico nos 223 municípios paraibanos utilizando-se de cálculo dos índices de Moran, mapas usados na Estatística Espacial e modelos espaciais para verificar a dependência espacial existente para os casos confirmados de COVID-19 nos municípios da Paraíba. As variáveis independentes analisadas foram o Índice de Desenvolvimento Humano Municipal (IDHM) e a taxa de urbanização, e a variável dependente foi a taxa de incidência de casos confirmados de COVID-19. Para este trabalho foi possível detectar uma estrutura de autocorrelação espacial significativa entre os municípios do estado da Paraíba formando vários *clusters*. O modelo de regressão espacial ajustado aos dados mostrou-se eficiente em descrever o relacionamento entre a taxa de incidência para casos de COVID-19 e as taxas de urbanização e o IDHM. Compreender a distribuição espacial da doença é fundamental para prever surtos locais e desenvolver estratégias de controle para COVID-19 e sendo assim, os resultados apresentados neste trabalho podem contribuir com o planejamento e ações do poder público principalmente os setores de saúde pública.

**Palavras-chave:** COVID-19. Análise Espacial. Estatística Espacial.

## ABSTRACT

The end of 2019 so far is marked by the challenge of facing the pandemic caused by the new coronavirus, SARS-CoV-2, the etiological agent responsible for COVID-19. The first case of COVID-19 in Paraíba was registered on March 18, 2020 in the municipality of João Pessoa, which reached 396,442 cases on June 30, 2021. The objective of this work is to map and analyze the evolution of cases in COVID-19 in the state of Paraíba, from January to June 2021. This is an ecological study in the 223 municipalities in Paraíba using the calculation of Moran indices, maps used in Spatial Statistics and spatial models to verify dependence existing space for confirmed cases of COVID-19 in the municipalities of Paraíba. The independent variables analyzed were the Municipal Human Development Index (IDHM) and the urbanization rate, and the dependent variable was the incidence rate of confirmed cases of COVID-19. For this work, it was possible to detect a significant spatial autocorrelation structure between the municipalities of the state of Paraíba, forming several clusters. The data-adjusted spatial regression model proved to be efficient in describing the relationship between the incidence rate for COVID-19 cases and urbanization rates and the MHDH. develop control strategies for COVID-19 and, therefore, the results presented in this work can contribute to the planning and actions of the public power, especially the public health sectors.

**Keywords:** COVID-19. Spatial Analysis. Spatial Statistics.



## LISTA DE FIGURAS

Figura 1 - Mapa de localização do Estado da Paraíba. ....	14
Figura 2 - Mapa da distribuição espacial do Índice de Desenvolvimento Humano dos municípios do estado da Paraíba.....	16
Figura 3 - Mapa da distribuição espacial da taxa de urbanização dos municípios do estado da Paraíba.....	17
Figura 4 - Roteiro da análise do trabalho .....	18
Figura 5 - Matriz de proximidade espacial usando o critério de fronteira entre as regiões .....	22
Figura 6 - Matriz de proximidade espacial com linhas normalizadas	<b>Error! Bookmark not defined.</b>
Figura 7 - Diagramas de espalhamento de Moran.....	27
Figura 8 - Distribuição espacial da taxa de incidência para os casos de COVID-19, de janeiro/2021 a junho/2021 no estado da Paraíba. ....	38
Figura 9 - Histograma da distribuição do número de vizinhos nos municípios paraibano para o critério contingência <i>queen</i> .....	39
Figura 10 - Mapa de conectividade dos municípios da Paraíba .....	39
Figura 11 - Diagrama de espalhamento para taxa de incidência nos casos de COVID-19 nos meses de janeiro a março/2021 .....	42
Figura 12 - Diagrama de espalhamento para taxa de incidência nos casos de COVID-19 nos meses de abril a junho/2021 .....	43
Figura 13 - Box map da taxa de incidência para casos de COVID-19, de janeiro a março/2021 no estado da Paraíba.....	44
Figura 14 - Box map da taxa de incidência para casos de COVID-19, de abril a junho/2021 no estado da Paraíba. ....	45
Figura 15 - Lisa map da taxa de incidência para casos de COVID-19, de janeiro a março/2021 no estado da Paraíba.....	47
Figura 16 - Lisa map da taxa de incidência para casos de COVID-19, de abril a junho/2021 no estado da Paraíba .....	48
Figura 17 - Moran map da taxa de incidência para casos de COVID-19, de janeiro a março/2021 no estado da Paraíba.....	50
Figura 18 - Moran map da taxa de incidência para casos de COVID-19, de abril a junho/2021 no estado da Paraíba.....	51

Figura 19 - Mapa dos resíduos para o Modelo Espacial de Durbin (SDM), de janeiro a março/2021 .....	58
Figura 20 - Mapa dos resíduos para o Modelo Espacial de Durbin (SDM), de abril a junho/2021. ....	59

### **LISTA DE TABELAS**

Tabela 1 - Análise descritiva da taxa de incidência para os casos de COVID-19, de janeiro/2021 a junho/2021 no estado da Paraíba. ....	37
Tabela 2 - Descrição do I de Moran e valor-p para da taxa de incidência para casos de .....	40
Tabela 3 - Avaliação das pressuposições no modelo clássico de regressão.....	52
Tabela 4 - Estimativa dos testes de multiplicadores de Lagrange .....	53
Tabela 5 - Tabela dos coeficientes estimados para os modelos de Regressão Espacial para janeiro/2021 .....	53
Tabela 6 - Tabela dos coeficientes estimados para os modelos de Regressão Espacial para fevereiro/2021.....	54
Tabela 7 - Tabela dos coeficientes estimados para os modelos de Regressão Espacial para março/2021 .....	54
Tabela 8 - Tabela dos coeficientes estimados para os modelos de Regressão Espacial para abril/2021.....	54
Tabela 9 - Tabela dos coeficientes estimados para os modelos de Regressão Espacial para maio/2021 .....	55
Tabela 10 - Tabela dos coeficientes estimados para os modelos de Regressão Espacial para junho/2021 .....	55
Tabela 11 - Tabela dos valores do AIC para os meses de janeiro/2020 a junho/2020. ....	56

## LISTA DE DA ABREVIATURAS E SIGLAS

IBGE	Instituto Brasileiro de Geografia e Estatística
OMS	Organização Mundial da Saúde
IDHM	Índice de Desenvolvimento Humano Municipal
IDH	Índice de Desenvolvimento Humano
TIC	Taxa de incidência de casos
TU	Taxa de urbanização
PNUD	Programa da Nações Unidas para o Desenvolvimento
SEDH	Secretaria de Estado do Desenvolvimento Humano
Q1	Primeiro quadrante
Q2	Segundo quadrante
Q3	Terceiro quadrante
Q4	Quarto quadrante
AA	Alto Alto
BB	Baixo Baixo
AB	Alto Baixo
BA	Baixo Alto
LISA	Local Indicators of Spatial Association (Indicadores Locais de Associação Espacial)
MQO	Mínimos Quadrados Ordinários
MRC	Modelo de Regressão Clássica
SAR	Spatial AutoRegressive Model (Modelo Espacial Autoregressivo Misto)
SEM	Spatial Error Model (Modelo de Erro Espacial)
SDM	Spatial Durbin Model (Modelo Espacial de Durbin)
AIC	Akaike Information Criterion (Critério de Informação de Akaike)

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>12</b>
<b>2</b>	<b>MATERIAL</b> .....	<b>14</b>
<b>2.1</b>	<b>Área de estudo</b> .....	<b>14</b>
<b>2.2</b>	<b>Banco de dados</b> .....	<b>15</b>
<b>2.2.1</b>	<i>Variáveis independentes</i> .....	<b>15</b>
<b>3</b>	<b>MÉTODO</b> .....	<b>18</b>
<b>4</b>	<b>ANÁLISE ESTATÍSTICA</b> .....	<b>20</b>
<b>4.1</b>	<b>Análise Espacial - Indicadores de dependência espacial</b> .....	<b>20</b>
<b>4.2</b>	<b>Matriz de proximidade espacial (W)</b> .....	<b>21</b>
<b>4.3</b>	<b>Média móvel espacial</b> .....	<b>23</b>
<b>4.4</b>	<b>Índice I de Moran (global)</b> .....	<b>23</b>
<b>4.5</b>	<b>Índice de C de Geary</b> .....	<b>25</b>
<b>4.6</b>	<b>Índice Local de Moran</b> .....	<b>25</b>
<b>4.7</b>	<b>Diagrama de ESPALHAMENTO de Moran</b> .....	<b>26</b>
<b>4.8</b>	<b>Box Map</b> .....	<b>28</b>
<b>4.9</b>	<b>LISA Map</b> .....	<b>28</b>
<b>4.10</b>	<b>Moran Map</b> .....	<b>29</b>
<b>4.11</b>	<b>Modelos de Regressão Espacial</b> .....	<b>29</b>
<b>4.11.1</b>	<i>Modelo de Regressão Clássica - MRC</i> .....	<b>30</b>
<b>4.11.2</b>	<i>Testes estatísticos</i> .....	<b>31</b>
<b>4.11.3</b>	<i>Modelos espaciais</i> .....	<b>32</b>
<b>4.11.3.1</b>	<i>Modelo de defasagem espacial – SAR</i> .....	<b>33</b>
<b>4.11.3.2</b>	<i>Modelo de erro autorregressivo espacial – SEM</i> . .....	<b>33</b>
<b>4.11.3.3</b>	<i>Modelo de defasagem espacial com erro autorregressivo espacial – SAC</i> . .....	<b>34</b>
<b>4.11.3.4</b>	<i>Modelo de Durbin espacial - SDM</i> . .....	<b>34</b>
<b>4.11.3.5</b>	<i>Modelo de Durbin espacial do erro - SDEM</i> .....	<b>35</b>
<b>4.11.3.6</b>	<i>Modelo Regressivo Cruzado Espacial - SLX</i> .....	<b>35</b>
<b>4.12</b>	<b>Escolha do modelo apropriado</b> .....	<b>36</b>
<b>5</b>	<b>RESULTADOS</b> .....	<b>37</b>

<b>5.1</b>	<b>Análise Espacial.....</b>	<b>38</b>
<b>5.1.1</b>	<b><i>Diagrama de espalhamento de Moran e Box map.....</i></b>	<b>40</b>
<b>5.1.2</b>	<b><i>LISA map .....</i></b>	<b>46</b>
<b>5.1.3</b>	<b><i>Moran map.....</i></b>	<b>49</b>
<b>5.2</b>	<b>REGRESSÃO ESPACIAL.....</b>	<b>52</b>
<b>5.2.1</b>	<b><i>Ajuste dos modelos .....</i></b>	<b>52</b>
<b>5.2.2</b>	<b><i>Análise de Regressão Espacial .....</i></b>	<b>55</b>
<b>6</b>	<b>CONCLUSÃO .....</b>	<b>60</b>
	<b>REFERÊNCIAS .....</b>	<b>61</b>

## 1 INTRODUÇÃO

O mundo passa pela mais grave pandemia de uma doença infecciosa causada por um novo vírus da família *Coronaviridae* (inicialmente chamado de Novo Coronavírus 2019-nCoV, e mais tarde renomeado para SARS-CoV-2). O nome COVID<sup>1</sup> é a junção de letras que se referem a *(co)rona (vi)rus (d)isease*, traduzindo para o português seria "doença do coronavírus", o número 19 está ligado ao ano que apareceram os primeiros casos. Em dezembro de 2019, na cidade de Wuhan (província de Hubei, na China), um surto de pneumonia de causa inicialmente desconhecida e posteriormente foi identificada a doença chamada COVID-19, representando um grande problema mundial de saúde pública nos últimos anos (NETTO & CORRÊA, 2020).

A SARS-CoV-2 é altamente transmissível por gotículas e contato, especialmente em locais fechados e hospitais. Locais fechados com pouca ventilação facilitam a transmissão do vírus (MEDEIROS, 2020). A transmissão do vírus ocorre de forma direta ou indireta. A transmissão direta ocorre de pessoa para pessoa, principalmente por meio de gotículas respiratórias e a transmissão indireta por meio de fômites que são vectores passivos (bancadas, corrimão, puxadores, interruptores da luz, telefones móveis, roupa etc), ou seja, objetos inanimados que podem levar e espalhar a doença e agentes infecciosos.

Em 30 de janeiro de 2020, a Organização Mundial da Saúde (OMS) declarou a epidemia do novo coronavírus como uma Emergência de Saúde Pública de Importância Internacional (ESPII) – o mais alto nível de alerta da Organização, conforme previsto no Regulamento Sanitário Internacional. Em 11 de março de 2020, a COVID-19 foi caracterizada pela OMS como uma pandemia.

O primeiro caso de COVID-19 no Brasil, e na América do Sul, foi registrado em 26 de fevereiro de 2020 em São Paulo. Até o dia 30 de junho de 2021, contabilizavam-se no Brasil 18.557.141 casos e 518.066 óbitos confirmados da doença. No estado da Paraíba, o primeiro caso de COVID-19 foi registrado em 18 de março de 2020 no município de João Pessoa. Até o dia 30 de junho de 2021, contabilizavam-se na Paraíba 396.442 casos e 8.606 óbitos confirmados da doença

---

<sup>1</sup> <https://portal.fiocruz.br/pergunta/por-que-doenca-causada-pelo-novo-coronavirus-recebeu-o-nome-de-covid-19>

Diante do contexto acima, usaremos os métodos de análise espacial para entender como uma determinada variável se distribui no espaço apresentando semelhanças entre si e diferenças com as demais. Conforme Cavalcante *et al.* (2020), na Saúde Coletiva essas análises vêm sendo utilizadas principalmente em situações de surtos e epidemias e é uma importante ferramenta para identificação de áreas de maior risco, conhecer como a doença está se distribuindo em uma região, compreender a sua propagação bem como a forma de dispersão geográfica de modo que possa servir como suporte para implementação de medidas de controle.

Dentre as técnicas utilizadas, temos as que mais se destacam: o índice de Moran local e o índice de Moran global, este avalia a relação de interdependência espacial entre todos os polígonos da área de estudo e é expressa por meio de um valor único para toda a região, aquele é um indicador utilizado quando se dispõe de grande número de áreas e identifica a relação existente entre um determinado polígono e a suas vizinhanças, verificando a existência de diferentes de correlação espacial em diferentes sub-regiões. Essas diferentes técnicas também podem ser visualizadas por meio de mapas *LISA* map, Moran map, Box map e diagrama de espalhamento (SANTOS *et al.*, 2007).

O objetivo deste trabalho é verificar a propagação da pandemia COVID-19, analisando as relações entre os casos confirmados de COVID-19 nos municípios paraibanos e indicadores sociodemográficos, como IDHM e taxa de urbanização durante o período de 1º de janeiro de 2021 a 30 de junho de 2021, mostrando municípios que podem influenciar os seus vizinhos de forma positiva, bem como de forma negativa na propagação do vírus. Este estudo tem como foco a primeira Lei da Geografia de Waldo Tobler (TOBLER, 1970), “Tudo está relacionado com tudo o mais. Mas as coisas próximas estão mais relacionadas do que as distantes”.

Compreender a distribuição espacial da doença é fundamental para prever surtos locais e desenvolver estratégias de controle para COVID-19. Assim, busca-se por meio da aplicação de métodos de Estatística Espacial, analisar a distribuição espacial da taxa de incidência da COVID-19 no estado da Paraíba. A Estatística Espacial é fundamental para representar e compreender a variação espacial existente nas regiões em estudo.

## 2 MATERIAL

### 2.1 Área de estudo

O estudo foi realizado na Paraíba, litoral oriental da região Nordeste do Brasil. O estado dispõe de área territorial de 56.468,435km<sup>2</sup>, divididos em 223 municípios. Conforme Figura 1 a Paraíba tem seus limites ao norte com o Rio Grande do Norte, ao sul com Pernambuco, ao leste com o Oceano Atlântico e ao oeste com o Ceará. Em 2010, a Paraíba contava com uma população de 3.766.528 pessoas. A população estimada para o ano de 2021 é de 4.059.905 pessoas. Em 2010 a Paraíba apresentava uma densidade demográfica de 66,70 hab/km<sup>2</sup> e um índice de desenvolvimento humano (IDH) de 0,658 (MENDES, 2019).

Figura 1 - Mapa de localização do Estado da Paraíba.



Fonte: Elaboração própria



## 2.2 Banco de dados

Para este estudo, os dados dos casos de COVID-19 nos municípios do estado da Paraíba foram levantados do site do Ministério da Saúde<sup>2</sup>, no período de janeiro/2021 a junho/2021. São números absolutos dos casos de COVID-19 nos municípios paraibanos e mostram claramente que os maiores números de casos ocorrem nas maiores cidades: João Pessoa, Campina Grande, Patos e Cajazeiras (acima de 5.000 casos registrados para o período citado acima). No entanto, é importante conhecer a distribuição relativa, ou seja, a distribuição de casos ponderada pelo número de habitantes (taxa de Incidência dos casos). Dessa forma, torna-se possível comparar, por exemplo, a incidência no mês de janeiro/2021 em João Pessoa (96,62 casos por 10.000 habitantes) com a da pequena cidade de Carrapateira (219,58 casos por 10.000 habitantes) sem o viés do peso populacional da metrópole.

A variável de interesse (variável dependente) é a Taxa de incidência de casos (TIC) de COVID-19 por 10.000 habitantes nos municípios do estado da Paraíba durante o período de janeiro de 2021 a junho de 2021, disponibilizada pelo Ministério da Saúde (<https://COVID.saude.gov.br/> - acesso em 1º/07/2021). Para o cálculo da taxa de incidência de cada município, o número de casos confirmados foi dividido pela população residente e multiplicado pela base populacional de 10.000 habitantes. Considerando que este trabalho envolve dados secundários e de livre acesso e sem identificação pessoal, o que, conforme a Resolução nº 510/2016, art. 1º parágrafo único, incisos I e III, do Conselho Nacional de Saúde, dispensa a necessidade prévia de aprovação do Comitê de Ética em Pesquisa com Seres Humanos

### 2.2.1 Variáveis independentes

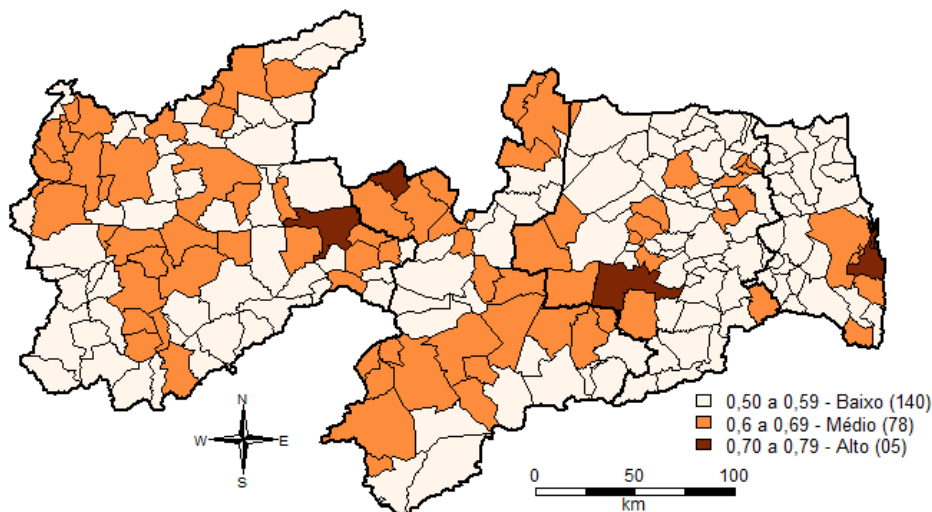
As variáveis independentes consideradas no trabalho foram o Índice de Desenvolvimento Humano Municipal (IDHM) e a taxa de urbanização (TU). Para o IDHM os dados foram coletados no Atlas de Desenvolvimento Humano do Brasil, as quais são

---

<sup>2</sup> <https://COVID.saude.gov.br/>

sintetizadas pelo do Programa da Nações Unidas para o Desenvolvimento (PNUD) estimado para cada município no ano de 2010 e podem ser classificado em muito baixo (0 a 0,499), baixo (0,500 a 0,599), médio (0,600 a 0,699), alto (0,700 a 0,799) e muito alto (0,800 a 1), sintetizando uma média de três dimensões básicas e universais da vida: o acesso ao conhecimento (educação), o direito a uma vida longa e saudável (longevidade) e o direito a um padrão de vida digno (renda) (Brasil, 2020a). A Figura 2 mostra a distribuição dos municípios paraibanos conforme suas respectivas faixas de IDHM, verifica-se que 140 municípios são classificados com o IDHM baixo, 78 municípios com IDHM médio e apenas 5 municípios paraibano são classificados com IDHM alto. Não há municípios classificados na faixas muito baixo e muito baixo.

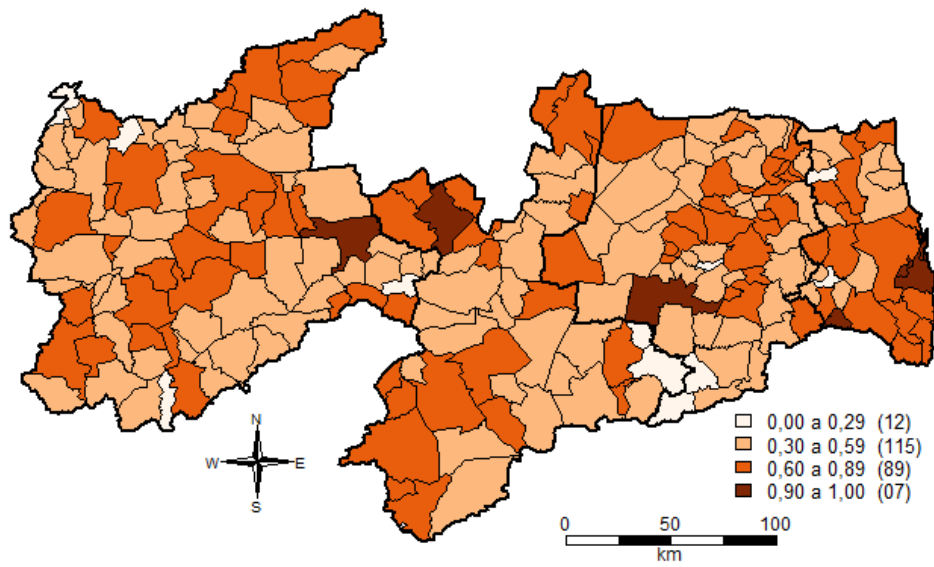
Figura 2 - Mapa da distribuição espacial do Índice de Desenvolvimento Humano dos municípios do estado da Paraíba.



Fonte: PNUD - Elaboração própria

A outra variável, Taxa de urbanização (TU), mostra a porcentagem da população residente constituída pelos moradores em domicílios urbanos em relação à população total, esses dados obtidos no Plano Estadual de Assistência Social da Paraíba (2020-2023) da Secretaria de Estado do Desenvolvimento Humano (SEDH) (Brasil, 2020a). Conforme Figura 3 tem-se que 7 municípios estão com a taxa de urbanização acima de 90%, sendo, em 2010, o município de Cabedelo o mais urbanizado (99,99%), 89 municípios pertencem a faixa de 60% a 90% de urbanização, 115 municípios estão na faixa de 30% a 60% e 12 municípios estão com a faixa de urbanização até 30%, sendo o município de Barra de Santana o menos urbanizado (8,91%), ou seja, 91,09% da população desse município reside na zona rural.

Figura 3 - Mapa da distribuição espacial da taxa de urbanização dos municípios do estado da Paraíba.

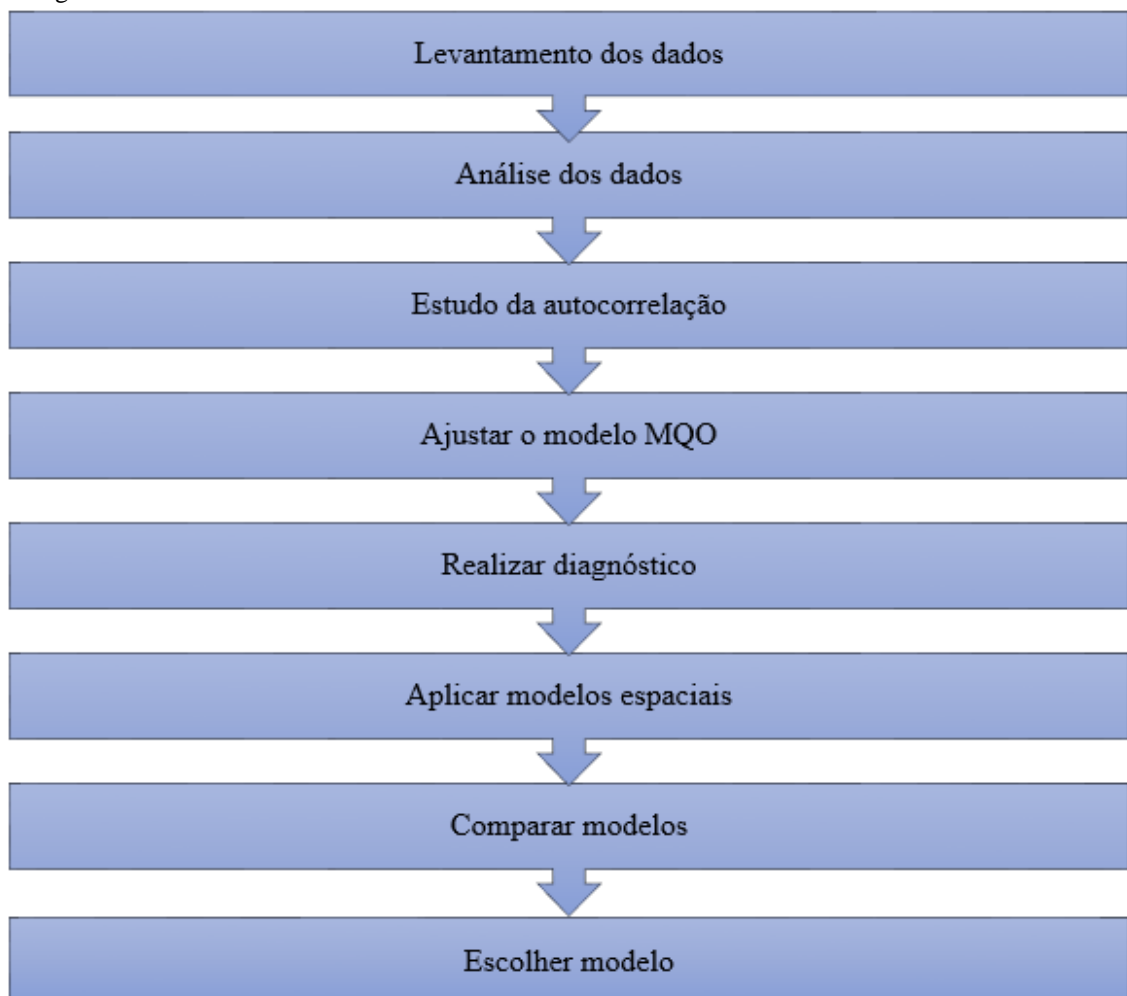


Fonte: SEDH – Elaboração própria

### 3 MÉTODO

O presente trabalho é um estudo ecológico, pois é um estudo em que os dados se referem a grupos de pessoas e não a indivíduos, a unidade de estudo é uma área geográfica delimitada por municípios do estado e os dados dessa área são comparados a outras, ou no espaço (séries temporais) (BONITA *et al*, 2010), cujas unidades de análise foram os 223 municípios do estado da Paraíba durante o período de janeiro de 2021 a junho de 2021. Para a confecção dos mapas temáticos, foi utilizada base cartográfica vetorial tipo *shapefile*, obtida no *site* do IBGE, com as divisões políticas da Paraíba por município. A análise espacial foi realizada com a linguagem de programação *Software R* versão 4.0.3 (2020-10-10), usando os pacotes: *spdep*, *sp*, *rgdal*, etc. A metodologia do trabalho está baseada nas fases que compõem a Figura 4

Figura 4 - Roteiro da análise do trabalho



Fonte: Elaboração própria

### Detalhamento de cada passo da análise

- **Levantamento dos dados** - Os dados utilizados neste trabalho foram retirados de sites oficiais conforme citado anteriormente.
- **Análise dos dados** - Faz-se necessário realizar uma análise descritiva dos dados
- **Estudo da autocorrelação** - De posse dos dados prontos, inicia-se o estudo da autocorrelação com o passo inicial que é a construção da matriz de vizinhança usando critérios de construção por polígonos contíguos ou polígonos dentro de uma certa distância e posteriormente aplicação da estatística global de Moran, estatística C de Geary e estatística local de Moran. A análise por meio de gráficos ou mapas são realizadas pelo diagrama de espalhamento de Moran, Box map, Lisa map e Moran map.
- **Ajuste do modelo MQO** - Após a análise da autocorrelação, inicia-se o estudo da modelagem com a finalidade identificar as variáveis que estão relacionadas globalmente com a variável dependente do estudo. O modelo de regressão linear foi estimado por MQO.
- **Realizar diagnóstico** - Para verificar se esta modelação está bem especificada é essencial que o modelo satisfaça os pressupostos quanto aos erros: que sua média deve ser zero, sua variância deve ser constante e ainda devem seguir a distribuição normal, bem como suas observações devem ser independentes entre si. A suposição de que as observações sejam independentes é pouco provável que aconteça, por causa da possibilidade de dependência espacial entre os termos do erro. Uma alternativa é utilizar o modelo espacial.
- **Aplicar modelos espaciais** - Após o diagnóstico, aplica-se os testes para confirmar a autocorrelação espacial. Caso haja autocorrelação espacial, o modelo gerado deve incorporar a estrutura espacial.
- **Comparar modelos** - A comparação entre os modelos espaciais é feita utilizando o  $R^2$  ajustado que indica a variabilidade da variável dependente ( $Y$ ) explicada pelo modelo, e utilizando o AIC que fornece informação sobre o desempenho do modelo.
- **Escolher modelo** - Para escolha do modelo mais apropriado, utiliza-se o Critério de Informação de Akaike (*AIC - Akaike Information Criterion*), quanto menor o valor, melhor o desempenho do modelo.

## 4 ANÁLISE ESTATÍSTICA

### 4.1 Análise Espacial - Indicadores de dependência espacial

Quando se pretende fazer algum tipo de análise com dados de área, que são espacialmente arranjados, uma das primeiras coisas que se faz é verificar e quantificar o grau de autocorrelação espacial (ou apenas correlação espacial) da variável de interesse existente entre as unidades amostrais vizinhas. Segundo Câmara *et al.* (2004), a ideia básica para essa análise exploratória é estimar a magnitude da autocorrelação espacial entre as áreas. Dessa forma, utiliza o índice global de Moran e o índice de Geary. Porém, quando o número de áreas for grande, a natureza dos processos envolvidos é tal que é muito provável a presença de diferentes regimes de correlação espacial em diferentes sub-regiões. Para particularizar estes regimes espaciais, podemos utilizar os indicadores locais de autocorrelação espacial e o mapa de espalhamento de Moran.

A correlação é uma medida que informa o quanto uma determinada variável varia em função de outra. No entanto, não se trata de analisar apenas como duas variáveis se correlacionam numericamente, mas sim de analisar o quanto o valor observado de uma única variável de área, numa área espacial é dependente dos valores dessa mesma variável nas áreas vizinhas fazendo-se necessário capturar a autocorrelação espacial entre todos os pares de localizações (LUZARDO, 2017)

Após determinar a estrutura espacial através da matriz de vizinhança, verifica ou não se há dependência espacial em um conjunto de dados. Em uma análise da existência ou não dessa dependência pode-se utilizar os índices de Moran (1950) e Geary (1954). O objetivo para o cálculo da autocorrelação espacial é de quanto o valor de uma variável em uma área é parecido com o do vizinho mais próximo e de quanto esse valor é diferente do vizinho mais distante. Esse pressuposto reflete a Lei Primeira da Geografia (TOBLER, 1970) onde “tudo está relacionado a tudo o mais, mas as coisas próximas estão mais relacionadas do que as coisas distantes”. Para Câmara *et al.* (2004), todas estas estatísticas dependem da definição de vizinhança adotada, de maneira que se entenda a matriz de proximidade espacial.

## 4.2 Matriz de proximidade espacial ( $W$ )

A representação formal da proximidade entre cada par de unidades de observação é realizada por meio de matrizes de proximidade espacial ( $W$ ), também denominadas como matriz de vizinhança e são indicadas para obter a relação espacial de cada área com as demais. Para análises da dependência espacial, o estudo primordial é a matriz de proximidade espacial ( $W$ ), tornando-se um procedimento necessário para a análise de dados de área. Essas matrizes compreendem a força de associação espacial de cada localização  $i$  com relação a todos os outros lugares  $j$  (ALMEIDA, 2012).

É uma matriz ( $W$ ) de ordem  $n \times n$ , em que  $n$  representa o número de áreas (observações) do estudo e cada elemento  $w_{ij}$ , com  $i = 1, 2, \dots, n$  e  $j = 1, 2, \dots, n$ , de  $W$  representa uma medida de proximidade (interdependência) entre uma área  $i$  ( $A_i$ ) e uma área  $j$  ( $A_j$ ) da região de estudo, de modo que os valores diferentes de zero na matriz de proximidade refletem o potencial de interação entre as observações. Elementos com valor igual a zero indicam inexistência de interação entre observações. De acordo com Câmara et al. (2004), a matriz  $W$  é definida como,

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \quad (4.1)$$

Há várias formas para construção da matriz de proximidade espacial ( $W$ ), tais como, proximidade física, compartilhamento de fronteiras, distâncias geográficas dos centroides, distâncias rodoviárias. As duas principais formas são por distâncias geográficas dos centroides ou por fronteiras.

- Matriz de k distância

$w_{ij} = 1$ , se o centroide de  $A_i$  está a uma determinada distância “x” do centroide de  $A_j$

$w_{ij} = 0$ , caso contrário

O valor de “x” é escolhido de acordo com a situação estudada

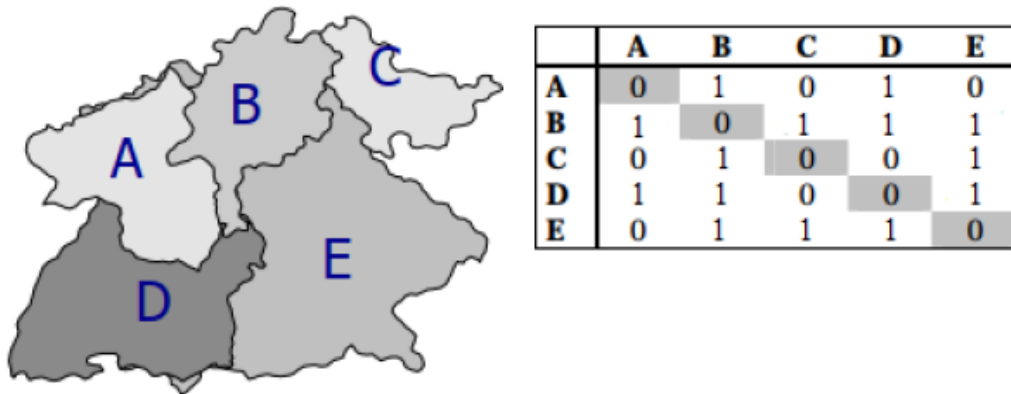
- Matriz de fronteira

$w_{ij} = 1$ , se  $A_i$  tem fronteira comum com  $A_j$ ;

$w_{ij} = 0$ , caso contrário

A Figura 5 apresenta uma matriz de proximidade espacial usando o critério de fronteira entre as regiões.

Figura 5 - Matriz de proximidade espacial usando o critério de fronteira entre as regiões

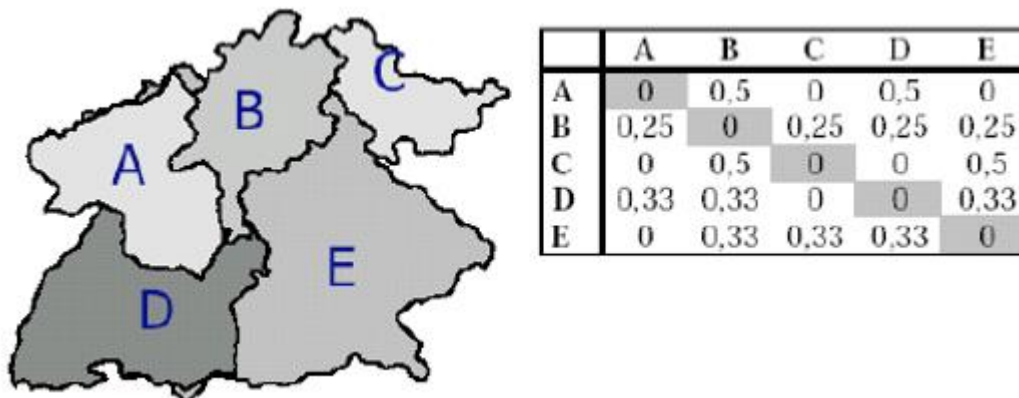


Fonte: Câmara et al., 2004

Segundo Araújo *et al.*, (2013), para simplificar os cálculos de índices de autocorrelação espacial na fase de análise exploratória, há necessidade de normalizar as linhas da matriz  $W$ . Esse procedimento garante que a soma dos valores de cada linha da matriz de proximidade espacial ( $W$ ) resulte no valor igual a 1.

A Figura 6 ilustra um exemplo anterior de matriz de proximidade espacial, em que os valores dos elementos da matriz refletem o critério de adjacência e foram normalizados.

Figura 6 - Matriz de proximidade espacial com linhas normalizadas



Fonte: Câmara et al., 2004



### 4.3 MÉDIA móvel espacial

A média móvel espacial é uma suavização dos dados espaciais na região de estudo para verificar a existência de padrões e detectar sub-regiões onde a tendência espacial é mais forte. Segundo Câmara et al. (2004), seu uso tende a reduzir a instabilidade espacial ao produzir uma superfície com menor oscilação que os dados originais. A média móvel é associada ao vetor dos desvios dados por  $z_i = y_i - \bar{y}$ , relativo à  $i$ -ésima área. Pode ser calculada a partir dos elementos  $w_{ij}$  da matriz normalizada de proximidade espacial  $W$ , tomando-se simplesmente a média dos vizinhos. Seu estimador é dado por:

$$\hat{\mu}_i = \sum_{j=1}^n w_{ij}(y_j - \bar{y}), \quad (4.2)$$

em que,  $w_{ij}$  é um elemento da matriz de proximidade espacial,  $y_i$  é o valor do atributo em cada região,  $\bar{y}$  é a média dos vizinhos e  $n$  é o número de regiões em estudos.

### 4.4 Índice I de Moran (global)

Segundo Anselin & Florax (1995), o índice I de Moran, proposto por Moran (1950) é uma das formas mais utilizadas para o cálculo de autocorrelação espacial, cujo valor tem uma variação de  $-1$  a  $+1$  e indica quanto cada área analisada é semelhante ou não à sua vizinhança imediata. Pode-se realizar um teste de hipótese para verificar a significância do índice, sendo que a hipótese nula ( $H_0$ ) é de independência espacial e a hipótese alternativa ( $H_1$ ) é de que existe dependência espacial.

O cálculo do índice I de Moran que representa o relacionamento existente entre os valores de uma variável e o seu valor médio pode ser expresso da seguinte forma,

$$I = \frac{n \sum_i \sum_j w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{(\sum_i (y_i - \bar{y})^2) (\sum_i \sum_j w_{ij})}, \quad (4.3)$$

onde,  $y_i$  é o valor da variável  $y$  na região  $i$ ,  $y_j$  é o valor da variável  $y$  na região  $j$ ,  $\bar{y}$  é a média da variável  $y$ ,  $w_{ij}$  é o elemento  $ij$  da matriz de proximidade espacial ( $W$ ) e  $n$  é o número de observações.

Para análise dos valores do índice de Moran, temos:

- Valores de I próximos de zero indicam a inexistência de autocorrelação espacial significativa entre os valores dos objetos e seus vizinhos, ou seja, o valor de uma variável em estudo tende a apresentar um comportamento aleatório;
- Valores próximos de 1 para o índice, indicam autocorrelação espacial positiva, ou seja, o valor do atributo de um objeto tende a ser semelhante aos valores dos seus vizinhos;
- Valores negativos para o índice indicam autocorrelação espacial negativa, ou seja, o valor do atributo de um objeto tende a ser diferente aos valores dos seus vizinhos.

O Índice I de Moran é uma das formas de detecção de similaridade entre as áreas. Portanto, o indicador é uma das maneiras de analisar se as regiões possuem algum padrão espacial de agrupamento. Conforme Câmara *et al.* (2004), os indicadores globais de autocorrelação espacial fornecem um único valor como medida da associação espacial para todo o conjunto de dados, o que é útil na caracterização da região de estudo como um todo. Dessa forma, verificamos que o I de Moran nos fornece três informações:

- Se os dados estão distribuídos aleatoriamente ou não
- O sinal da estatística de Moran, se positivo os dados estarão concentrados através das regiões e se for negativo, os dados estarão dispersos.
- A magnitude do I de Moran informa a força da autocorrelação espacial.

Este índice pode ocultar os padrões locais de autocorrelação espacial, por esse motivo que existem os indicadores locais pois há um melhor detalhamento para cada região, tornando possível, a identificação de diferentes padrões espaciais, como agrupamentos com valores parecidos (clusters) ou divergentes (outliers). Conforme Albuquerque (2009), ao estudar os padrões locais para autocorrelação é possível ter-se informações sobre a presença de heterogeneidade espacial, podendo então, descobrir associações espaciais do tipo alto-alto (AA) ou baixo-baixo (BB), e também do tipo alto-baixo (AB) ou baixo-alto (BA).

#### 4.5 Índice de C de Geary

Outro método usado para verificar a dependência espacial é o índice de C de Geary proposto por Geary (1954). A diferença deste método com o Moran Global, é que ele considera a diferença entre os pares, já o Moran utiliza a diferença entre cada ponto e a média global (CÂMARA *et al.* 2004). O índice de Geary pode ser explicado pela equação abaixo.

$$C = \frac{n-1}{2} \frac{n \sum_i \sum_j w_{ij} (y_i - \bar{y})^2}{(\sum_i (y_i - \bar{y})^2) (\sum_i \sum_j w_{ij})}, \quad (4.4)$$

em que,  $y_i$  é o valor da variável  $y$  na região  $i$ ,  $y_j$  é o valor da variável  $y$  na região  $j$ ,  $\bar{y}$  é a média da variável  $y$ ,  $w_{ij}$  é o elemento  $ij$  da matriz de proximidade espacial ( $W$ ) e  $n$  é o número de observações.

O resultado do índice C de Geary é semelhante ao I de Moran. Entretanto, seu valor varia entre 0 e 2, sendo:

- 0 para uma forte autocorrelação espacial positiva;
- 2 para uma forte autocorrelação espacial negativa;
- 1 representa a ausência de autocorrelação espacial.

#### 4.6 Índice Local de Moran

Outro indicador para detecção de autocorrelação espacial foi proposto na literatura por Anselin (1995) e refere-se a estes indicadores como estatística LISA (*Local Indicators of Spatial Association*). Segundo Câmara *et al.* (2004), os indicadores globais de autocorrelação espacial fornecem um único valor como medida da associação espacial para todo o conjunto de dados, o que é útil na caracterização da região de estudo como um todo. No entanto, muitas vezes é desejável examinar padrões numa escala de maior detalhe. Segundo Anselin (1995) os indicadores locais produzem um valor específico para cada área, identificando agrupamentos de valores extremos e a existência de vários regimes espaciais não identificados através do índice I global de Moran. O índice I local de Moran é representado por:

$$I_i = \frac{(y_i - \bar{y}) \sum_i \sum_j w_{ij} (y_j - \bar{y})}{\frac{\sum_i (y_i - \bar{y})^2}{n}}, \quad (4.5)$$

em que,  $y_i$  é o valor da variável  $y$  na região  $i$ ,  $y_j$  é o valor da variável  $y$  na região  $j$ ,  $\bar{y}$  é a média da variável  $y$ ,  $w_{ij}$  é o elemento  $ij$  da matriz de proximidade espacial ( $W$ ) e  $n$  é o número de observações.

De acordo com Moran (1950), o Índice de Moran local quando for valores positivos, significam que existem agrupamentos espaciais com valores similares (alto ou baixo) da variável em estudo. Para valores negativos significam que existem agrupamentos espaciais com valores dissimilares da variável entre as áreas e seus vizinhos.

#### 4.7 Diagrama de ESPALHAMENTO de Moran

Através dos gráficos podemos visualizar a dependência espacial de forma mais clara, tornando-se um facilitador para as análises. Dessa forma, o diagrama de espalhamento de Moran (Moran Scatterplot Map) é uma maneira alternativa de visualizar a dependência espacial. Construído com base nos valores normalizados (valores dos atributos subtraídos de sua média e divididos pelo desvio padrão), permite analisar o comportamento da variabilidade espacial (CÂMARA,2004).

A ideia é comparar os valores normalizados do atributo numa área com a média dos seus vizinhos, construindo um gráfico bidimensional de  $z$  (valores normalizados) por  $wz$  (média dos vizinhos, também normalizadas), que é dividido em quatro quadrantes, como mostrado na Figura 7. Após o cálculo do Índice de Moran global podemos visualizar graficamente o resultado calculado no indicador global da autocorrelação espacial por meio do diagrama de dispersão de Moran.

Segundo Câmara *et al.* (2004), o diagrama de Moran é um diagrama de dispersão de pontos representando as regiões de estudo, com a indicação da declividade da reta de regressão, onde o eixo  $x$  mostra a variável de interesse ( $z$ ) em função da defasagem espacial da variável de interesse no eixo  $y$  ( $Wz$ ), observando que os valores desses dois eixos deverão ser padronizados (valores normalizados: média = 0, variância = 1).

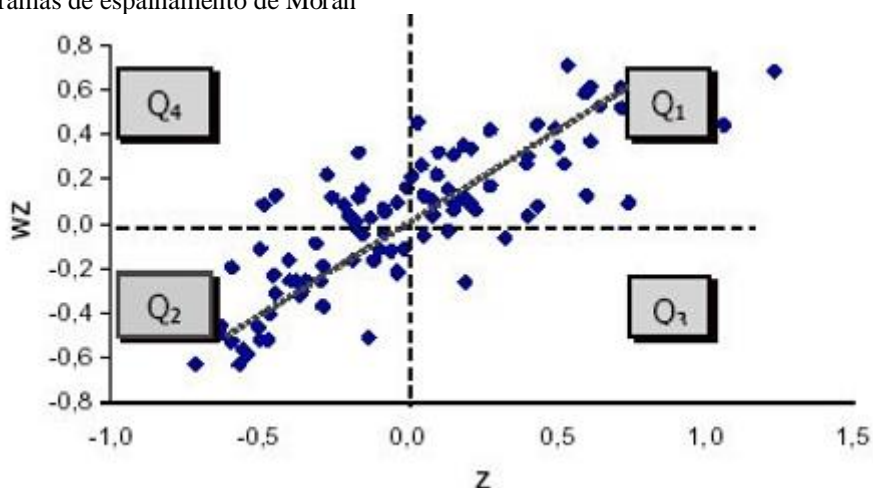
Para o cálculo da declividade dessa reta estima-se o coeficiente de regressão linear por Mínimos Quadrados Ordinários (CHEIN,2019), onde o coeficiente da inclinação da reta de regressão é dado pela estatística I de Moran. A inclinação da curva é obtida pela regressão de Wz contra z, e essa inclinação fornece o grau de ajustamento (Equação 4.6).

$$Wz = \alpha + \beta z + \varepsilon, \quad (4.6)$$

onde  $\alpha$  é a constante de regressão,  $\beta$  é o coeficiente angular (I de Moran) e  $\varepsilon$  é o erro aleatório.

Dessa forma temos que o I de Moran é o coeficiente angular da reta do gráfico de dispersão da defasagem espacial da variável de interesse (wz) contra a variável de interesse (z) estimado por mínimos quadrados ordinários (MQO) e representado pela figura abaixo (ALMEIDA, 2012).

Figura 7 - Diagramas de espalhamento de Moran



Fonte: Câmara *et al.* (2004)

O diagrama é dividido em quatro partes: Alto-Alto(Q<sub>1</sub>), Baixo-Baixo(Q<sub>2</sub>), Alto-Baixo(Q<sub>3</sub>) e Baixo-Alto (Q<sub>4</sub>) e cada ponto representa uma área (polígono). O quadrante superior direito (Q<sub>1</sub>), classificado como alto-alto (AA), são as regiões que apresentam altos valores para a variável em análise (valores acima da média) cercada por áreas vizinhas que também apresentam valores acima da média. O quadrante inferior esquerdo (Q<sub>2</sub>), classificado como baixo-baixo (BB), são as regiões que apresentam valores baixos para a variável em análise cercada por áreas vizinhas que também apresentam baixos valores. O quadrante inferior direito (Q<sub>3</sub>), classificado como alto-baixo (AB), é formado pelas áreas com altos valores para a variável em análise cercada por áreas com baixos valores. O quadrante superior esquerdo (Q<sub>4</sub>),

classificado como baixo-alto (BA), mostra as áreas com valores baixos cercadas por vizinhos que apresentam valores altos.

A maior concentração de pontos nos quadrantes  $Q_1$  (alto-alto) e  $Q_2$  (baixo-baixo): representam áreas que contribuem para autocorrelação positiva [ $I > 0$ ] e indicam clusters (aglomerações) de valores altos e baixos, respectivamente. Quando a maior concentração de pontos está nos quadrantes  $Q_3$  (alto-baixo) e  $Q_4$  (baixo-alto): representam áreas que contribuem para autocorrelação negativa [ $I < 0$ ] e indicam áreas de transição, com os valores altos cercados por valores baixos do atributo, e vice-versa. Caso os pontos estejam igualmente distribuídos pelos quatro quadrantes, há uma indicação de ausência de autocorrelação [ $I \approx 0$ ].

#### 4.8 Box Map

Outra forma de apresentar o Diagrama de Espalhamento de Moran é em forma de mapa, no qual cada polígono é apresentado indicando-se seu quadrante no diagrama de espalhamento e será classificado conforme sua posição em relação aos quadrantes do gráfico de espalhamento de Moran, onde os elementos de cada quadrante do gráfico são representados por uma cor específica com seus respectivos polígonos (LUZARDO, 2017).

#### 4.9 LISA Map

Com a significância estatística do índice local de Moran calculada, é muito útil gerar um mapa que indica as regiões que apresentam correlação local significativamente diferente do restante dos dados (CÂMARA *et al.*, 2004). Dessa forma geraremos os mapas para Indicadores Locais de Associação Espacial (LISA map) que são utilizados para identificar as regiões que apresentam correlação local significativamente diferente das demais (TEIXEIRA, 2015).

Na geração do LISA map, a avaliação da significância é feita comparando-se os valores dos índices de Moran local com uma série de valores obtidos por meio de permutações dos

valores dos atributos dos vizinhos (número de permutações definida pelo usuário), sob a hipótese nula de não existência de autocorrelação espacial (SILVA et al., 2011).

Os valores do índice local de Moran são classificados nos seguintes grupos:

- Não significantes
- Com significância de 95%
- Com significância de 99%
- Com significância de 99,9%

#### **4.10 Moran Map**

De maneira análoga ao LISA map, o Moran map também é utilizado para a análise dos indicadores locais de autocorrelação espacial. Neste mapa é avaliada a significância dos valores do índice de Moran local calculados para cada área, porém, são apresentadas apenas as áreas para as quais os valores dos índices locais de Moran foram considerados significativos. (valor- $p < 0,05$ ) e classificadas em quatro grupos, conforme sua localização no quadrante do gráfico de espalhamento.

Na construção deste mapa a classificação é realizada utilizando-se do diagrama de espalhamento de Moran. As áreas, em que o índice local apresentou significância, são classificadas em grupos de acordo com os quadrantes aos quais pertencem no gráfico de espalhamento de Moran e as demais áreas são classificadas como não significativas (SILVA et al., 2011).

#### **4.11 Modelos de Regressão Espacial**

O estudo da regressão espacial trata do problema de inferência em modelos (de regressão) com dados espaciais, tais modelos também necessitam dos três principais pressupostos do modelo de regressão clássica, ou seja, os erros são normais com média zero, têm variância constante (homocedasticidade) e são não correlacionados (ZILLI,2020). No

entanto, ao se incorporar na sua estrutura o fator “espaço”, na maioria das vezes, são eliminados os problemas de autocorrelação e heterocedasticidade. A inferência de modelos de regressão com dados espaciais pode ser suspeita. Em essência, isso ocorre porque as coisas próximas são semelhantes, conforme a 1ª lei da geografia (TOBLER, 1970) e pode não ser justo considerar casos individuais como independentes. Portanto, esses modelos precisam ser diagnosticados antes de relatá-los. Especificamente, é importante avaliar a existência de autocorrelação espacial nos resíduos (uma vez que estes são supostamente independentes, não correlacionados). Quando se faz uma análise de regressão, procura-se alcançar dois objetivos: Encontrar um bom ajuste entre os valores preditos pelo modelo e os valores observados da variável dependente e descobrir quais das variáveis explicativas contribuem de forma significativa para o relacionamento linear.

Segundo Anselin (1988), dados espaciais são caracterizados por estar presente autocorrelação espacial e pela heterogeneidade ou estrutura espacial. Entretanto, invalidam os resultados dos modelos tradicionais de regressão, pois violam alguns pressupostos como a independência e a homocedasticidade. Dessa forma, para introduzir efeitos espaciais em modelos de regressão, buscamos capturar a estrutura de correlação espacial através do adicionamento de um parâmetro no modelo de regressão clássica.

Segundo Golgher (2015) existem muitos tipos de modelos espaciais de forma que as regressões espaciais são modelos que possuem a matriz de vizinhança adicionados em algum dos seus parâmetros. Os modelos mais conhecidos são Spatial Autoregressive Model (SAR), Spatial Error Model (SEM), General Spatial Model (SAC), modelos Durbin Espacial (modelo SDM), e o Durbin Espacial do Erro (SDEM).

#### ***4.11.1 Modelo de Regressão Clássica - MRC***

Segundo Câmara et al. (2004), um modelo de regressão está baseado no relacionamento entre duas ou mais variáveis de modo que uma delas (variável dependente) pode ser explicada ou ter seu valor predito através de outras variáveis (variáveis independentes) e é descrito por:

$$Y = X\beta + \varepsilon, \quad (4.7)$$



onde,  $Y$  representa a variável dependente,  $X$  é uma matriz de variáveis explicativas (independentes),  $\beta$  é o coeficiente de regressão e  $\varepsilon$  é o erro aleatório com média zero e variância constante. Verificamos que nesse modelo não existe interação espacial entre as regiões, dessa forma, equação não incorpora nenhuma defasagem espacial.

A estimação dos parâmetros pode ser feita usando o método dos mínimos quadrados ordinários que consiste em minimizar a soma de quadrados dos resíduos. Um resíduo é a diferença entre o valor observado e o valor estimado em um certo ponto.

Uma das exigências do modelo descrito no modelo clássico é que os erros devem ser independentes. No entanto, para o caso de dados espaciais, onde está presente a dependência espacial, é muito pouco provável que a hipótese padrão de independência dos erros seja verdadeira. Câmara *et al.* (2004) afirmam que no caso mais comum os resíduos continuam apresentando a autocorrelação espacial presente nos dados, dessa forma, na situação dos dados espaciais, quando está presente a autocorrelação espacial as estimativas do modelo devem incorporar essa estrutura espacial, uma vez que a dependência entre as observações altera o poder explicativo do modelo.

Nos modelos de regressão linear, para torná-los válidos, são feitas várias pressuposições sobre o termo do erro, ou seja, os erros serão homocedásticos, independentes e seguir uma distribuição normal com média zero e variância constante. Se esses pressupostos forem violados não serão válidas as inferências com base no modelo.

#### **4.11.2 Testes estatísticos**

Para Almeida (2012), os testes mais utilizados para detectar a autocorrelação espacial são  $I$  de Moran e os testes do tipo multiplicador de Lagrange: LM Robusto (erro) e LM Robusto (defasagem). O teste  $I$  de Moran é o mais usado nos estudos de dados de corte transversal de unidades geográficas. Conforme Wooldridge,(2005), um conjunto de dados de corte transversal consiste em uma amostra de uma unidade de análise, tomada em um determinado ponto no tempo. Trata-se de um teste de fácil cálculo computacional, o problema deste teste é que ele não aponta qual tipo de autocorrelação espacial é predominante, ou seja, não identifica o tipo de efeito (erro ou defasagem espacial). Por esse motivo, serão utilizados testes mais específicos:

o LM Robusto (erro), para detectar efeitos de autocorrelação espacial no termo do erro e o LM Robusto (defasagem), para verificar a presença de efeitos de defasagem espacial na variável dependente. Com a confirmação da presença de autocorrelação espacial nos dados pelos testes citados acima, faz-se necessário introduzir extensões convenientes no modelo de Regressão Clássica, considerando-se os efeitos autocorrelação espacial nos erros ou os efeitos ocasionados pelas interações entre os valores da variável dependente.

#### ***4.11.3 Modelos espaciais***

Conforme Almeida (2012), os modelos espaciais podem ser divididos em modelos de alcance global, modelos de alcance local e modelos de alcance global-local da dependência espacial. Os modelos de alcance global capturam a estrutura de correlação espacial em um único parâmetro, apresenta a dependência espacial através do transbordamento do efeito espacial de forma global, ou seja, os efeitos ocorridos na variável dependente  $y$  em uma determinada região produz uma transmissão destes efeitos para o restante das regiões através do multiplicador espacial. Os modelos de alcance global existentes são: o Modelo de Defasagem Espacial (modelo SAR) e o Modelo de Erro Autorregressivo Espacial (modelo SEM) e o Modelo de Defasagem Espacial com Erro Autorregressivo Espacial (modelo SAC). Para este trabalho serão utilizados os modelos SAR e SEM.

O mesmo autor informa que os modelos de alcance local da dependência espacial são caracterizados por apresentar uma dependência espacial que gera impactos de maneira localizada, ou seja, o impacto da dependência espacial é apenas observado em algumas regiões em estudo, especialmente entre os vizinhos diretos e os vizinhos indiretos de segunda ordem (vizinhos dos vizinhos). Neste trabalho, será apresentado o modelo Regressivo Cruzado Espacial (modelo SLX). Para os modelos que apresentam o alcance da dependência espacial tanto global como local simultaneamente temos os modelos Durbin Espacial (modelo SDM) e o Durbin Espacial do Erro (SDEM). Neste trabalho, será detalhado o modelo Durbin Espacial (modelo SDM).

#### 4.11.3.1 Modelo de defasagem espacial – SAR

O modelo SAR expressa a interação existente entre a variável dependente em regiões vizinhas  $i$  e  $j$  (CÂMARA, 2004). Para este modelo, há de forma aditiva um termo ( $\rho WY$ ) em relação ao modelo clássico de regressão. Este modelo pode ser expresso em sua forma mista, da seguinte maneira:

$$Y = X\beta + \rho WY + \varepsilon, \quad (4.8)$$

em que,  $Y$  representa a variável dependente,  $WY$  é a matriz de vizinhança ou matriz de ponderação espacial com dependência espacial em  $Y$ ,  $\rho$  é o coeficiente de defasagem espacial, com valor pertencente ao intervalo aberto -1 e 1. Se  $\rho$  apresentar um sinal positivo, tem-se a indicação da presença de autocorrelação espacial global positiva, ou seja, um alto (baixo) valor de  $y$  nas regiões vizinhas aumenta (diminui) o valor de  $y$  na região  $i$ , se o coeficiente  $\rho$  for negativo indica a presença de autocorrelação espacial global negativa, ou seja, um alto (baixo) valor de  $y$  nas regiões vizinhas diminui (aumenta) o valor de  $y$  na região  $i$ , se o coeficiente  $\rho$  for igual a zero, indica que não existe dependência entre as variáveis,  $X$  é uma matriz de variáveis explicativas (independentes),  $\beta$  é o coeficientes de regressão e  $\varepsilon$  é o erros aleatórios com média zero e variância  $s^2$ . Neste modelo a variável dependente  $y$  é influenciada pela mesma variável nas regiões vizinhas (ALMEIDA, 2012).

#### 4.11.3.2 Modelo de erro autorregressivo espacial – SEM.

No modelo SEM a dependência espacial é encontrada nos resíduos do modelo (ALMEIDA, 2012), ou seja, a informação de vizinhança aparece no termo de erro, dessa forma, deixam de ser distribuídos aleatoriamente no espaço, estando autocorrelacionados espacialmente. O modelo SEM pode ser expresso por:

$$Y = X\beta + \xi \quad \text{onde} \quad \xi = \lambda W\xi + \varepsilon, \quad (4.9)$$

sendo,  $Y$  representando a variável dependente,  $W\xi$  uma matriz que representa uma média dos erros nas regiões vizinhas, acompanhados de um coeficiente  $\lambda$  que representa o parâmetro do

erro autorregressivo espacial, onde  $|\lambda| < 1$ ,  $X$  é uma matriz de variáveis explicativas (independentes),  $\beta$  é o coeficientes de regressão e  $\varepsilon$  é o termo de erro aleatório.

#### 4.11.3.3 Modelo de defasagem espacial com erro autorregressivo espacial – SAC.

O modelo SAC é uma generalização dos dois modelos acima. A estrutura de vizinhança  $W$  aparece tanto como variável explicativa como no erro aleatório, ou seja, utiliza duas matrizes de vizinhança  $W1$  e  $W2$ , uma para estimar o valor predito e outra para estimar o erro. O modelo SAC pode ser expresso por:

$$Y = \rho W y + X\beta + \xi \quad \text{onde} \quad \xi = \lambda W \xi + \varepsilon, \quad (4.10)$$

onde  $Y$  é a variável dependente,  $\rho$  é o coeficiente espacial autoregressivo,  $W$  é a matriz de proximidade espacial defasada espacialmente em  $y$ ,  $X$  são as variáveis independentes (explicativas e  $\varepsilon$  é o termo de erro aleatório (ALMEIDA, 2012).

#### 4.11.3.4 Modelo de Durbin espacial - SDM.

No modelo SDM as interações entre vizinhos são levadas em consideração tanto no nível de variável dependente defasada espacialmente  $WY$  quanto no nível de variáveis explicativas  $WX$ , esse modelo entra no rol dos modelos que apresentam um alcance global dado pelo multiplicador espacial decorrente da adição da defasagem espacial da variável dependente, e ainda um alcance local advindo da defasagem espacial das variáveis explicativas. O modelo SDM pode ser representado formalmente como:

$$Y = \rho W y + X\beta + WX\theta + \varepsilon, \quad (4.11)$$

onde  $Y$  é a variável dependente,  $\rho$  é o coeficiente espacial autoregressivo,  $W$  é a matriz de proximidade espacial defasada espacialmente em  $Y$ ,  $X$  são as variáveis independentes

(explicativas),  $\theta$  é o *lag* espacial das variáveis explicativas e  $\varepsilon$  é o termo de erro aleatório (ALMEIDA, 2012).

#### 4.11.3.5 Modelo de Durbin espacial do erro - SDEM.

O modelo SDEM apresenta elementos de transbordamento espacial de alcance local que difunde nas variáveis explicativas ( $X$ ), e elementos de transbordamento espacial global que interfere o termo de erro ( $\xi$ ) (ALMEIDA, 2012). O modelo SDEM pode ser representado com a seguinte estrutura:

$$Y = X\beta + WX\theta + \xi \text{ onde } \xi = \lambda W\xi + \varepsilon, \quad (4.12)$$

as notações são semelhantes às dos modelos anteriores, em que  $WX$  representa o transbordamento espacial local das variáveis explicativas, atingindo apenas os vizinhos diretos, enquanto  $W\xi$  representa o termo de erro defasado espacialmente que afeta todo os conjuntos de regiões.

#### 4.11.3.6 Modelo Regressivo Cruzado Espacial - SLX.

O modelo SLX é um modelo que apresenta a dependência espacial localizada, ou seja, os efeitos da autocorrelação espacial são observados em apenas algumas regiões vizinhas da área de estudo, especialmente entre vizinhos diretos. No modelo SLX, as variáveis explicativas  $X$  especificadas no modelo produzem um efeito de transbordamento localizado nas regiões vizinhas, ou seja, que as variáveis explicativas da região  $i$  influenciam a variável dependente  $Y$  na região vizinha  $j$ , e vice-versa. O modelo SLX pode ser representado por:

$$Y = X\beta + WX\tau + \varepsilon, \quad (4.13)$$

onde  $WX$  representa o transbordamento espacial das variáveis explicativas, e  $\tau$  o parâmetro que informa o grau de interação espacial localizado de  $WX$  nas regiões vizinhas, sem afetar todo o sistema (ALMEIDA, 2012).

#### 4.12 Escolha do modelo apropriado

Uma maneira de escolher o modelo pode ser feita pela comparação do valor testes LM (erro), LM (defasagem) e pelo índice  $I$  de Moran sobre os resíduos. Assim, quanto maior for o valor encontrado na estatística de teste, maior será o efeito espacial correspondente a esta estatística, conforme argumento de Anselin e Rey (1991). Outra forma de escolha de modelos é pela utilização dos critérios de informação de Akaike (AIC) e de Schwartz (SC).

Em termos gerais, estes critérios de informação assumem a forma

$$IC = -2L + f(k,n), \quad (4.14)$$

onde  $L$  é a máxima log verossimilhança e  $f(k,n)$  é uma função do número de variáveis independentes ( $k$ ) e do número de observações ( $n$ );  $f(k,n) = 2k$  para o Critério de Informação de Akaike (AIC) e  $f(k,n) = k \cdot \ln(n)$  para o Critério de Informação Schwartz (SC). Na escolha do modelo, o melhor será aquele com menores valores encontrados para AIC e SC (DANTAS *et al.*, 2010)

## 5 RESULTADOS

Realizou-se uma análise descritiva da taxa de incidência para os casos de COVID-19, por 10.000 habitantes, de janeiro/2021 a junho/2021. Pode-se verificar pela Tabela 1 que essa taxa, obteve seu valor mínimo em janeiro/2021 e maio/2021 com nenhum caso registrado e seu valor máximo em maio/2021 de 772,29 casos por 10.000 habitantes na cidade de Curral Velho (194 casos registrados para uma população de 2.512 habitantes), com a maior média obtida em junho/2021 (162,59 casos por 10.000 habitantes).

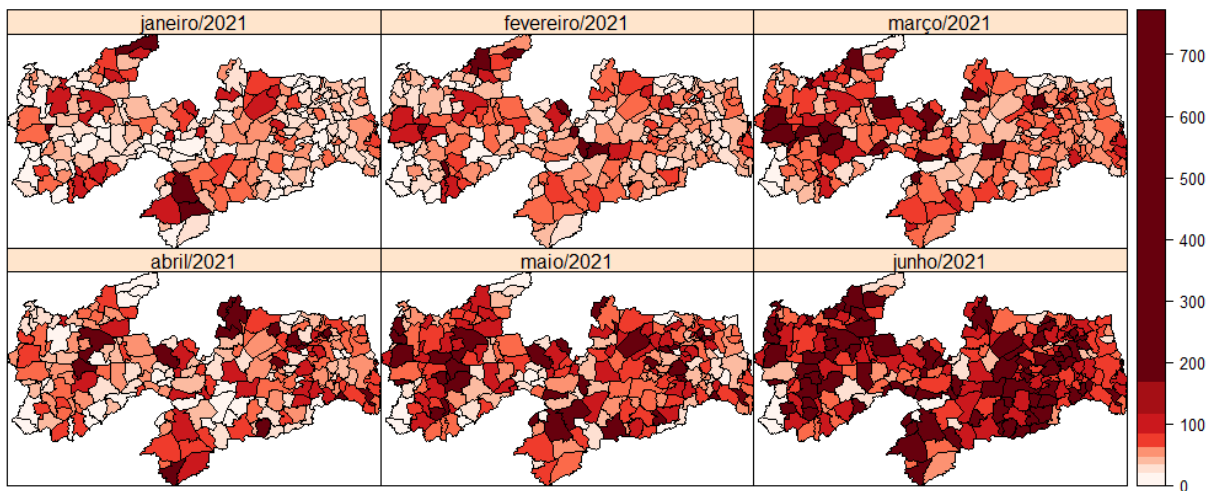
Tabela 1 - Análise descritiva da taxa de incidência para os casos de COVID-19, de janeiro/2021 a junho/2021 no estado da Paraíba.

	<b>Janeiro</b>	<b>Fevereiro</b>	<b>Março</b>	<b>Abril</b>	<b>Maió</b>	<b>Junho</b>
<b>Valor Mínimo</b>	0,00	2,21	1,05	2,78	0,00	8,35
<b>Valor Máximo</b>	227,24	714,55	546,08	418,76	772,29	597,35
<b>1º Quartil</b>	18,60	25,94	39,45	29,83	45,87	87,16
<b>3º Quartil</b>	61,09	70,90	95,62	95,56	132,85	212,86
<b>Média</b>	46,95	57,71	78,19	73,25	103,79	162,59
<b>Mediana</b>	32,28	44,40	62,51	56,40	82,94	141,16
<b>Variância</b>	1.847,02	3.728,67	4.582,38	3.929,64	8.222,95	11.511,16
<b>Desvio Padrão</b>	42,98	61,06	67,69	62,69	90,68	107,29
<b>Assimetria</b>	1,65	6,01	2,80	2,04	2,85	1,25
<b>Curtose</b>	2,87	58,14	12,56	6,32	14,26	1,81

Fonte: Elaboração própria com dados do Ministério da Saúde

A Figura 8 representa o mapa da distribuição espacial da taxa de incidência para os casos de COVID-19, por 10.000 habitantes dos municípios do estado Paraíba, mostrando a crescente taxa de casos no período de janeiro/2021 a junho/2021. Observa-se aglomerados de taxas elevadas nas mesorregiões do Sertão o do Agreste Paraibano.

Figura 8 - Distribuição espacial da taxa de incidência para os casos de COVID-19, de janeiro/2021 a junho/2021 no estado da Paraíba.



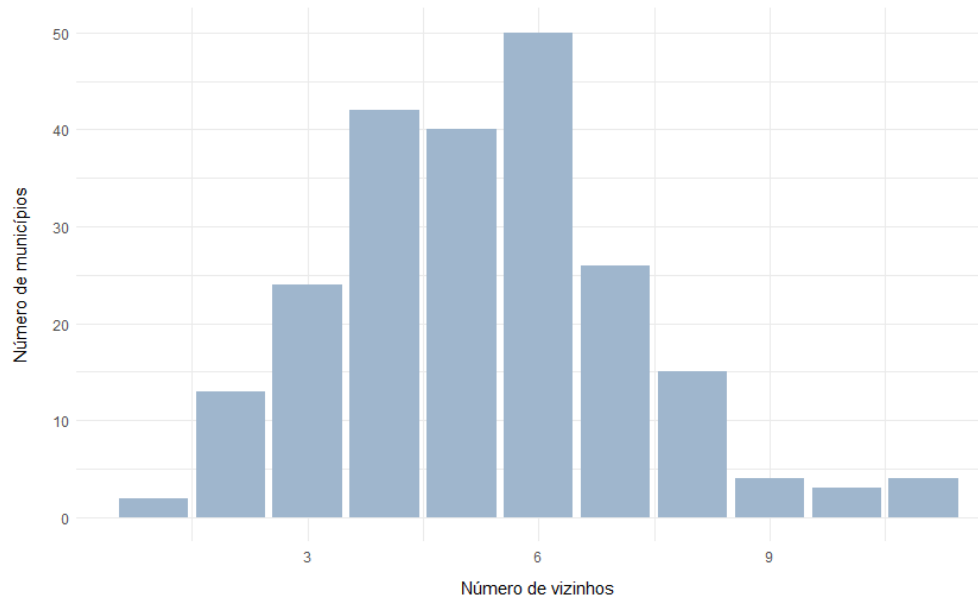
Fonte: Elaboração própria com dados do Ministério da Saúde

### 5.1 Análise Espacial

O passo inicial para análise espacial antes de aplicar o índice de Moran para calcular a autocorrelação espacial dos dados é construir a matriz de proximidade espacial para identificar as regras de vizinhança para cada município do estado do Paraíba. O critério de vizinhança utilizado neste trabalho foi o critério da contingência *queen* que atribuiu o valor de 1 para os polígonos que são vizinhos e 0 para os polígonos que não compartilham fronteiras, neste caso, os municípios do estado da Paraíba que são vizinhos de um determinado município X irão receber o valor 1, e o restante dos municípios o valor 0. A Figura 10 mostra o mapa de conectividade dos municípios da Paraíba para os 223 municípios utilizados neste trabalho. Percebe-se que neste critério a maioria dos municípios possuem entre 4 a 6 fronteiras em comum. A Figura 9, apresenta o histograma da distribuição do número de vizinhos das cidades paraibanas.



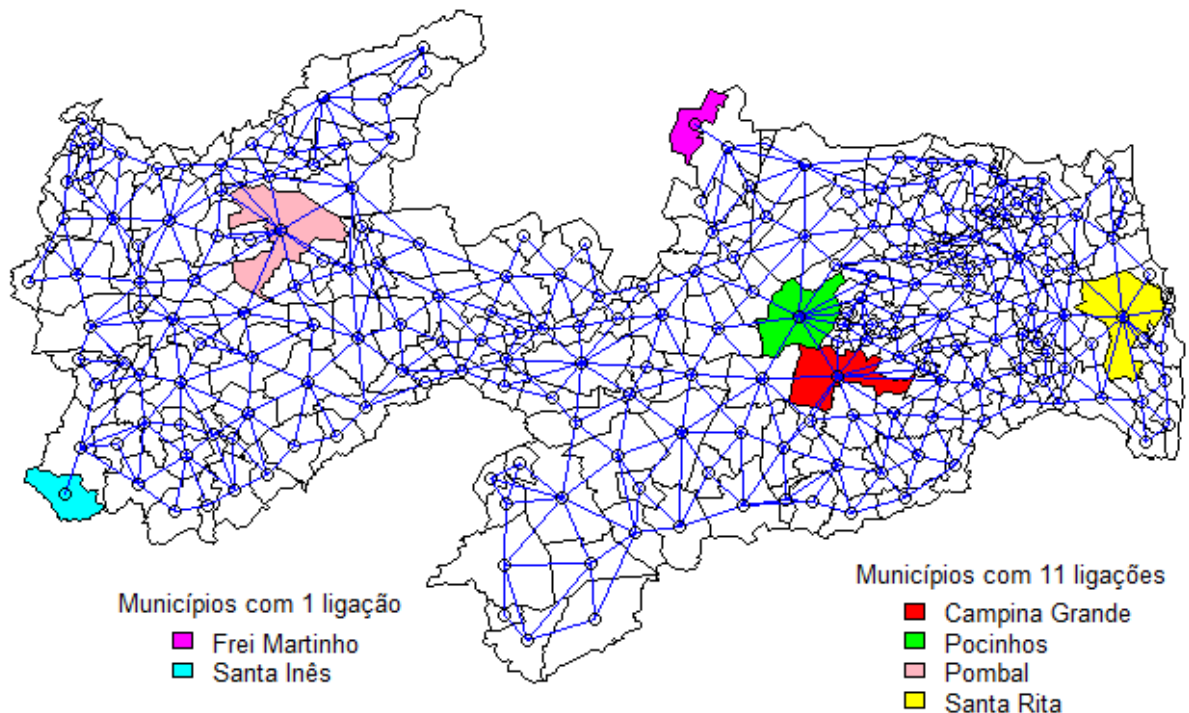
Figura 9 - Histograma da distribuição do número de vizinhos nos municípios paraibano para o critério contingência *queen*



Fonte: Elaboração própria

A Figura 10 mostra os extremos dos valores dos números de fronteiras, apresentando as cidade com mais vizinhos, ou seja, cidades que fazem fronteiras com 11 municípios (Campina Grande, Pocinhos, Pombal e Santa Rita) e as cidades com menos vizinhos, municípios que fazem fronteiras apenas com 1 município do estado da Paraíba (Frei Martinho e Santa Inês).

Figura 10 - Mapa de conectividade dos municípios da Paraíba



Fonte: Elaboração própria

Para a verificar a autocorrelação espacial global utilizou-se o índice de Moran global, que tem o objetivo de analisar se os dados estão distribuídos aleatoriamente no espaço, ou, se os dados estão autocorrelacionados, Na Tabela 2 estão dispostos os resultados da análise estatística espacial da taxa de casos dos períodos estudados. O coeficiente I de Moran de autocorrelação espacial para os meses de janeiro, fevereiro e abril/2021 apresentaram valores significativo ao nível de 5%, indicando ocorrência de dependência espacial

Tabela 2 - Descrição do I de Moran e valor-p para da taxa de incidência para casos de COVID-19, de janeiro/2021 a junho/2021 no estado da Paraíba, Brasil.

<b>Mês</b>	<b>Estatística I Moran</b>	<b>p-valor</b>
Janeiro	0.152111028	0.0001057
Fevereiro	0.146661304	1.635e-05
Março	0.019358072	0.2818
Abril	0.186342452	2.66e-06
Maiο	0.028789831	0.2092
Junho	0.024219818	0,25

Fonte: Elaboração própria com dados do Ministério da Saúde

### **5.1.1 Diagrama de espalhamento de Moran e Box map**

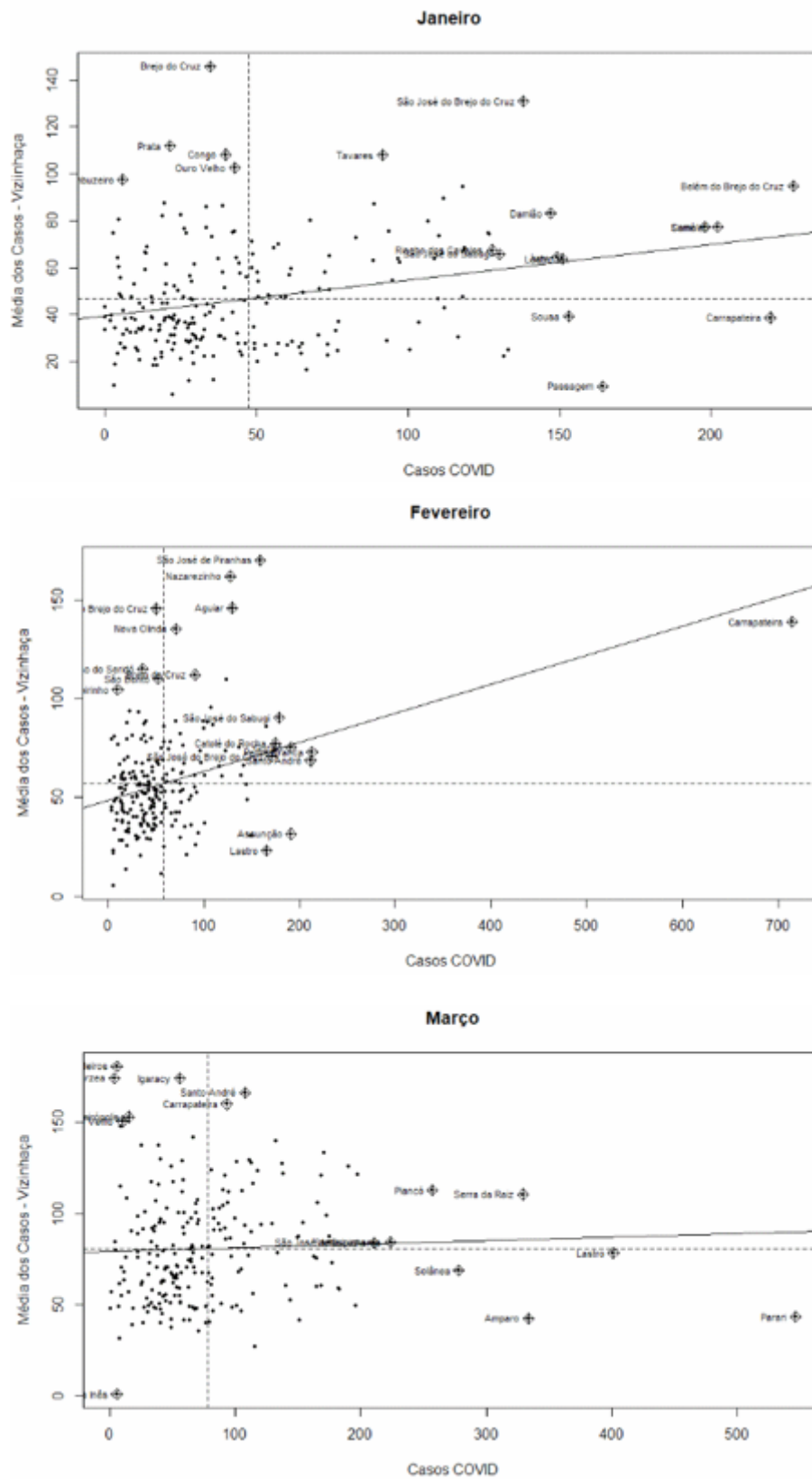
O diagrama de espalhamento de Moran, Figura 11 e Figura 12, é utilizado para visualizar a autocorrelação existente entre os municípios do estado da Paraíba. Para melhor visualização utiliza-se o Box map e para sua elaboração cada município paraibano será classificado conforme sua posição em relação aos quadrantes do gráfico de espalhamento de Moran, recebendo uma cor correspondente no mapa gerado conforme Figura 13 e Figura 14.

Por meio do Box map, verifica-se que alguns municípios tais como: Barra de São Miguel, Cajazeiras, Catolé do Rocha, Monteiro, Pombal, Santa Luzia e Solânea em alguns meses estão no primeiro quadrante (alto-alto), que são municípios que apresentam altos valores para a taxa de casos de COVID-19 (valores acima da média) cercada por vizinhos que também apresentam valores acima da média e em outros meses estão no terceiro quadrante (alto-baixo) com altos valores para taxa de casos de COVID-19 cercada por vizinhos com baixos valores,

ou seja, são municípios que estão sempre apresentando sua taxa de caso alta. Não foi encontrado nenhum município que permaneceu no primeiro quadrante durante todo período de estudo do trabalho (janeiro/2021 a junho/2021).

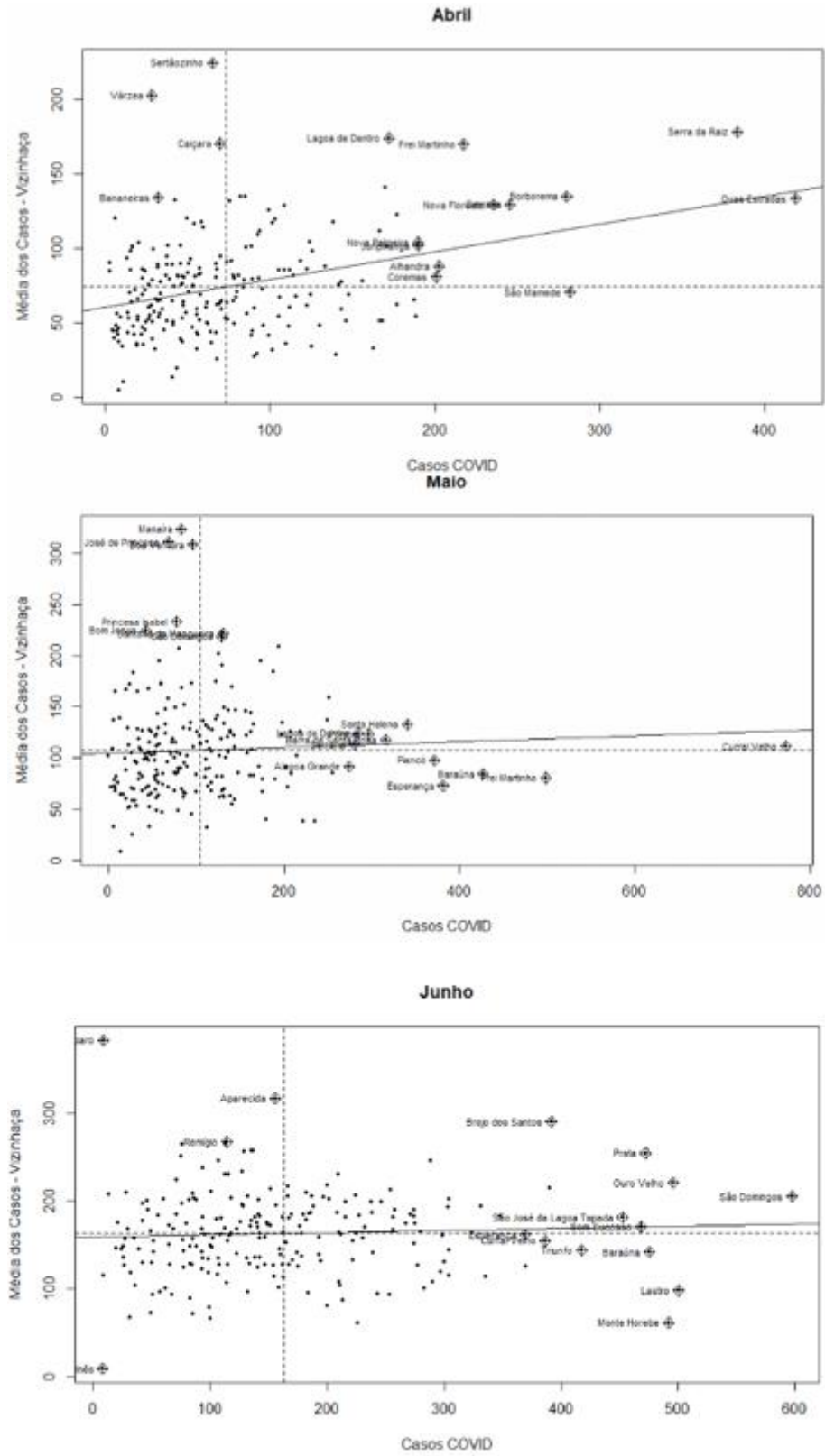
Verifica-se, também, que os municípios de Santa Inês, Uiraúna e Conceição pertencentes à Mesorregião do Sertão Paraibano, os municípios de Araruna e Campo de Santana pertencentes à Mesorregião do Agreste Paraibano e os municípios de Caldas Brandão, Mari, Sapé, Sobrado, Riacho de Santo Antônio, Capim e Mataraca pertencentes à Mesorregião da Mata Paraibana todos permaneceram no segundo quadrante (baixo-baixo) durante os todo período de estudo do trabalho, ou seja, formaram *clusters* em suas respectivas mesorregiões, influenciando dessa forma para diminuição dos casos, pois são regiões que apresentam valores baixos para a taxa de casos de COVID-19 cercada por áreas vizinhas que também apresentam baixas taxas.

Figura 11 - Diagrama de espalhamento para taxa de incidência nos casos de COVID-19 nos meses de janeiro a março/2021



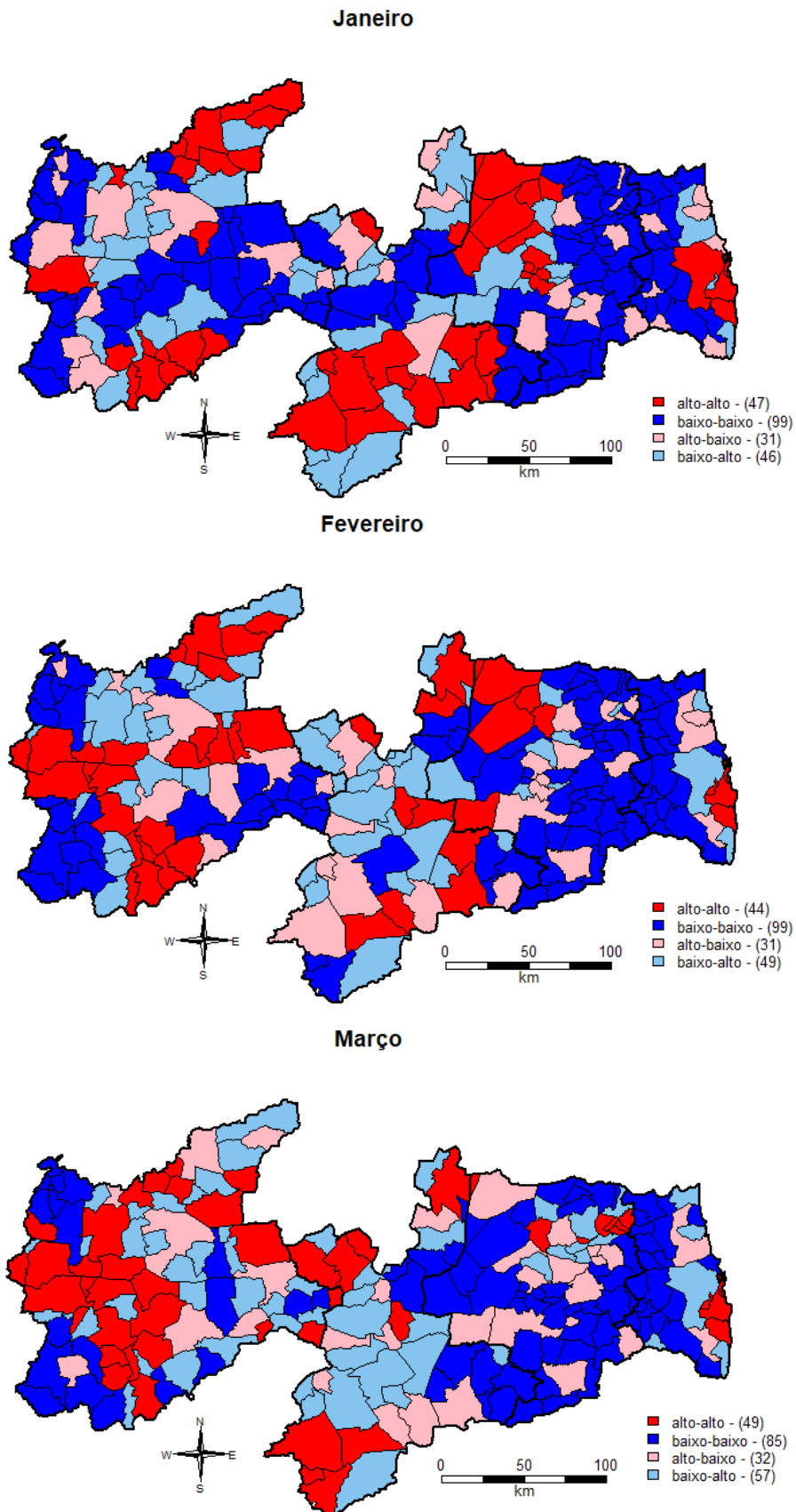
Fonte: Elaboração própria com dados do Ministério da Saúde

Figura 12 - Diagrama de espalhamento para taxa de incidência nos casos de COVID-19 nos meses de abril a junho/2021



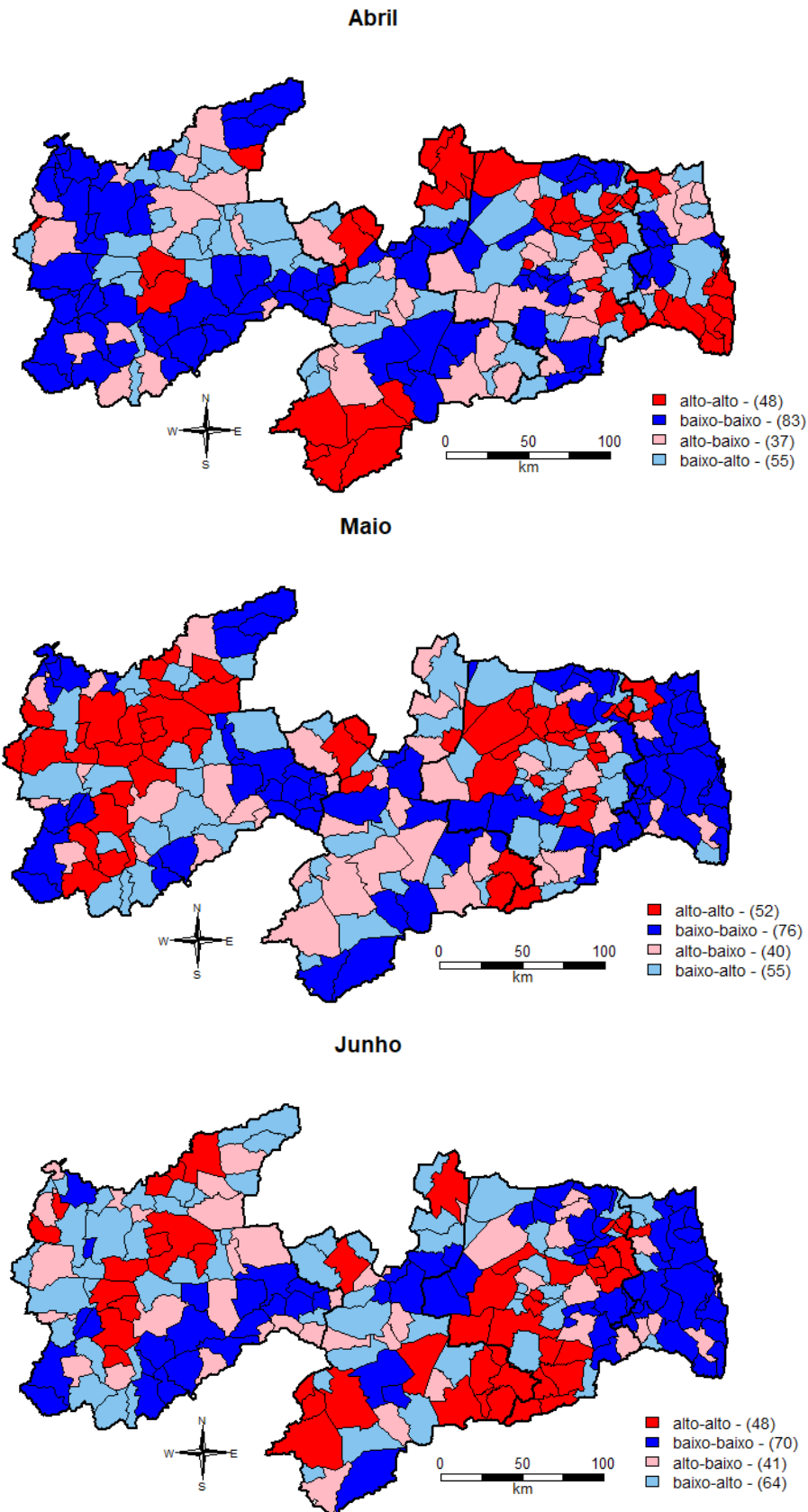
Fonte: Elaboração própria com dados do Ministério da Saúde

Figura 13 - Box map da taxa de incidência para casos de COVID-19, de janeiro a março/2021 no estado da Paraíba.



Fonte: Elaboração própria com dados do Ministério da Saúde

Figura 14 - Box map da taxa de incidência para casos de COVID-19, de abril a junho/2021 no estado da Paraíba.



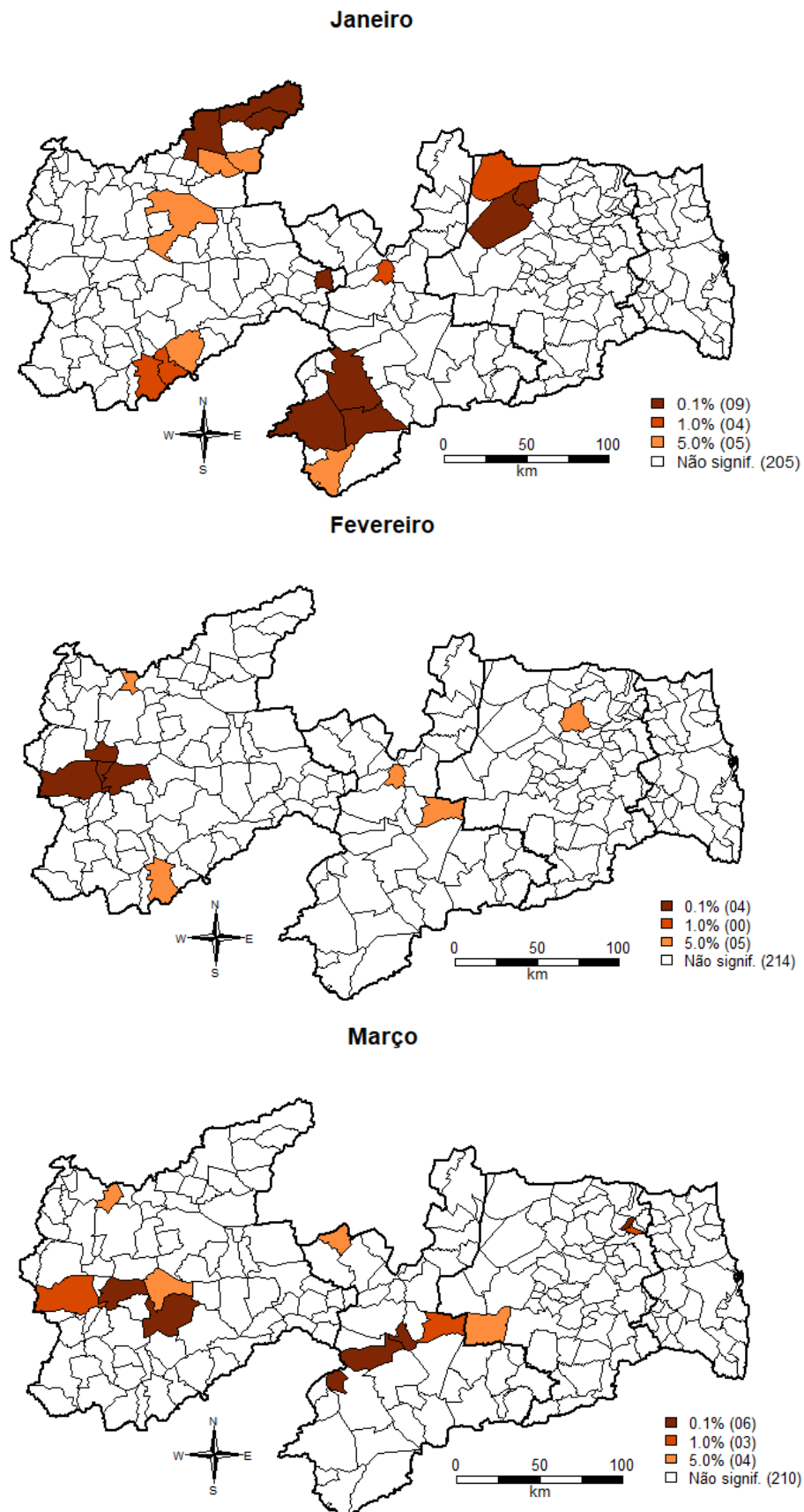
Fonte: Elaboração própria com dados do Ministério da Saúde

### 5.1.2 *LISA map*

Segundo Almeida (2012), para cada observação (município) é calculado um valor para o índice de Moran local ( $I_i$ ), assim, são calculadas 223 estatísticas  $I_i$  com seus respectivos níveis de significância. Nas Figura 15 e Figura 16 são apresentados os mapas de *clusters* LISA, que combina a informação do mapa de significância das medidas de associação local  $I_i$ . Observa-se na Figura 15 que o maior número de regiões com o Indicador de Moran Local significativo foi no mês de janeiro/2021, ou seja, 18 municípios, em contrapartida o mês de maio apresentou somente 7 municípios com esse índice significativo. Verifica-se que o mês de janeiro/2021 apresentou *clusters* significativo nas mesorregiões Sertão Paraibano, Borborema e Agreste Paraibano e posteriormente houve uma tendência de redução da quantidade *clusters* significativos ao longo dos meses. As zonas selecionadas no LISA Map serão utilizadas na próxima fase do trabalho para a identificação dos *clusters* espaciais

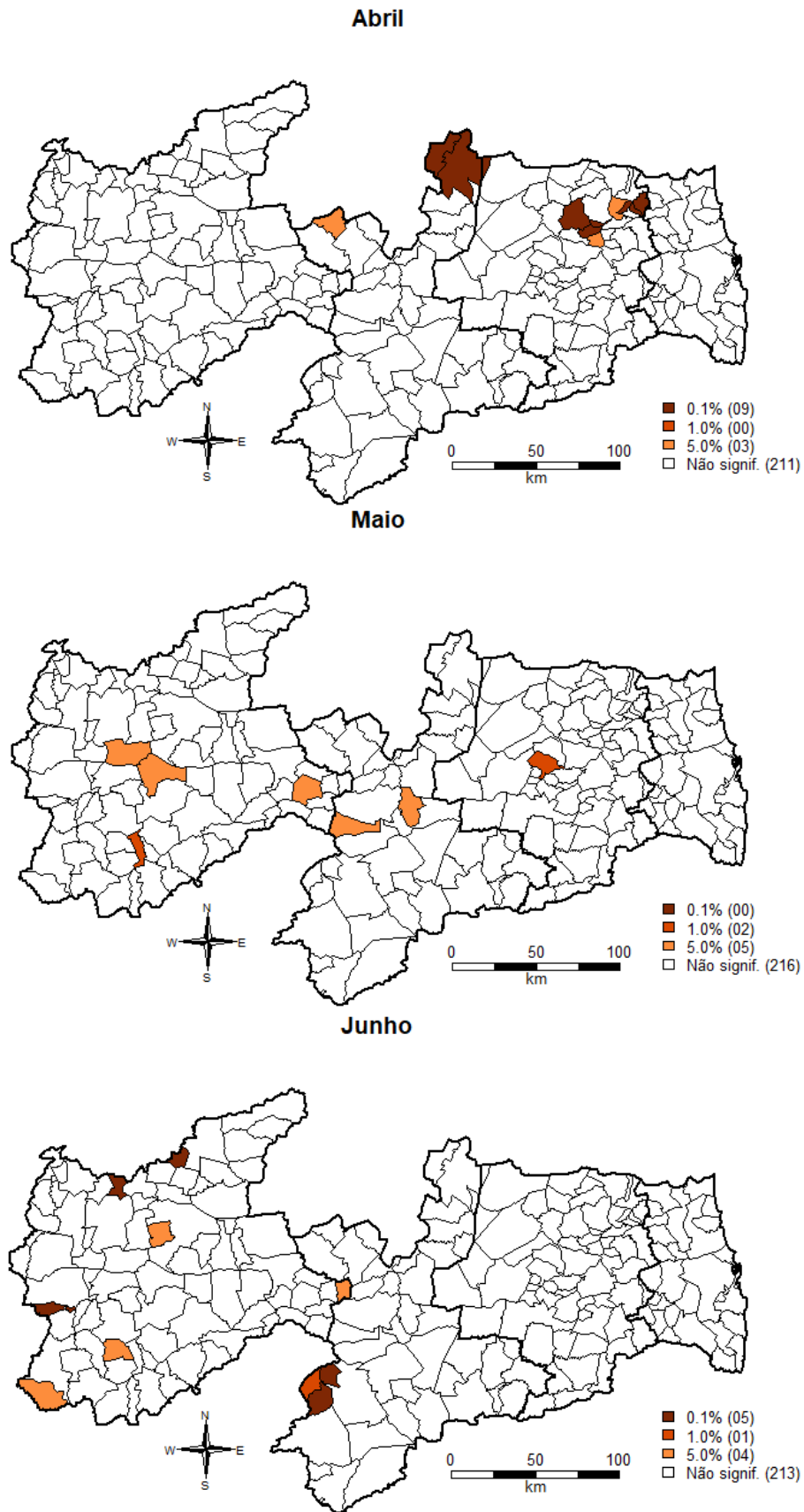


Figura 15 - Lisa map da taxa de incidência para casos de COVID-19, de janeiro a março/2021 no estado da Paraíba



Fonte: Elaboração própria com dados do Ministério da Saúde

Figura 16 - Lisa map da taxa de incidência para casos de COVID-19, de abril a junho/2021 no estado da Paraíba



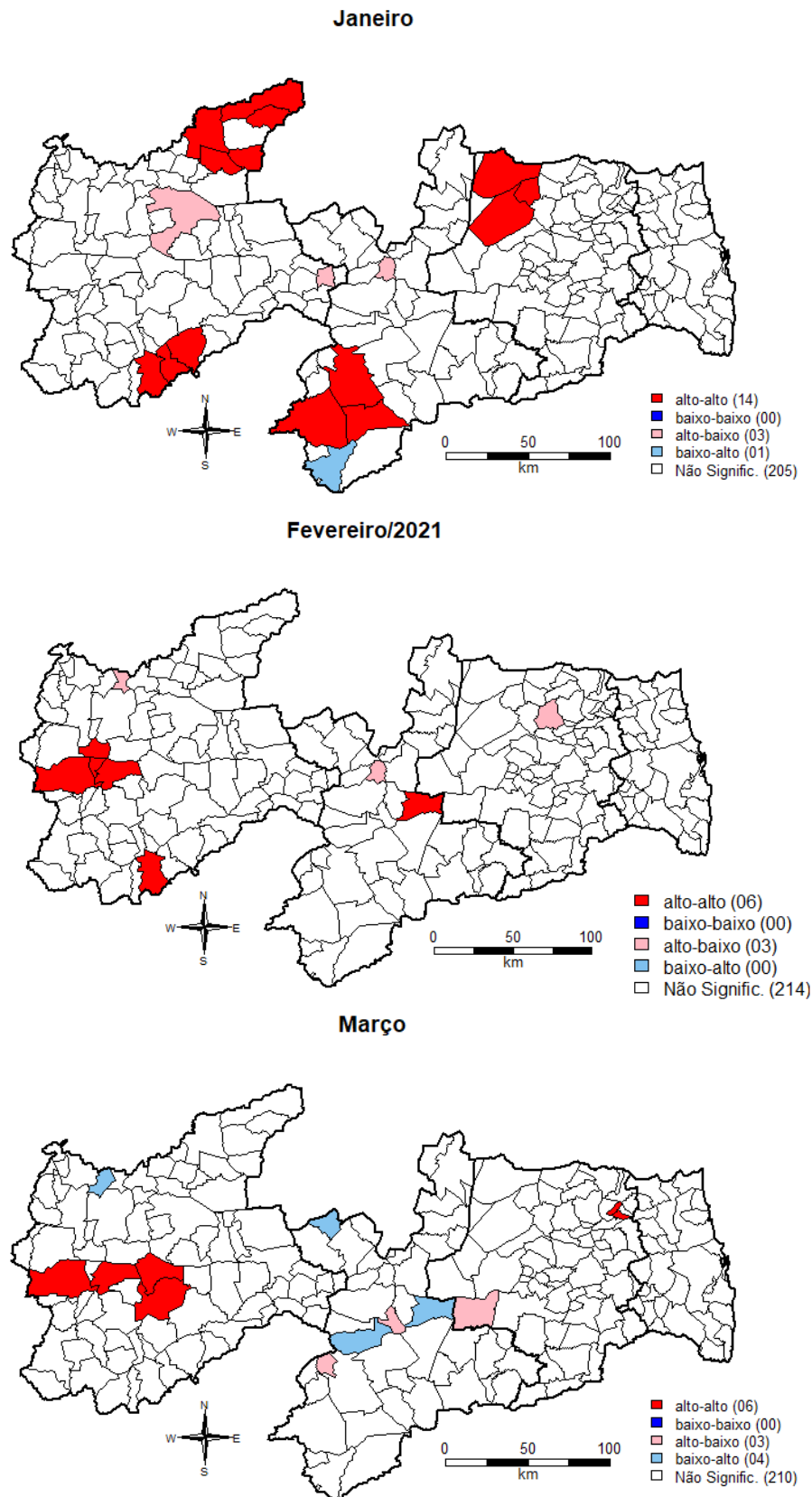
Fonte: Elaboração própria com dados do Ministério da Saúde

### 5.1.3 Moran map

As Figura 17 e Figura 18 apresentam os municípios que apresentaram valores significativos que foram obtidos através da aplicação de indicadores de autocorrelação espacial local conforme Figura 15 e Figura 16, mostrando como a taxa de incidência de casos em um município  $i$  altera com a variação dessa mesma taxa nos municípios vizinhos. Conforme Almeida (2012), as áreas em vermelho são municípios em que a taxa de incidência é elevada (Alto-Alto), tanto no município  $i$  como nos seus vizinhos. As áreas em azul escuro são os municípios em que esse fenômeno é reduzido (Baixo-Baixo). As áreas em azul claro (Baixo-Alto) e as em rosa (Alto-Baixo), indicam pontos extremos (outliers) em relação aos seus vizinhos. As áreas brancas (não significativas), são aquelas em que a taxa de casos de COVID-19 não difere, estatisticamente, da média do conjunto de municípios do estado.

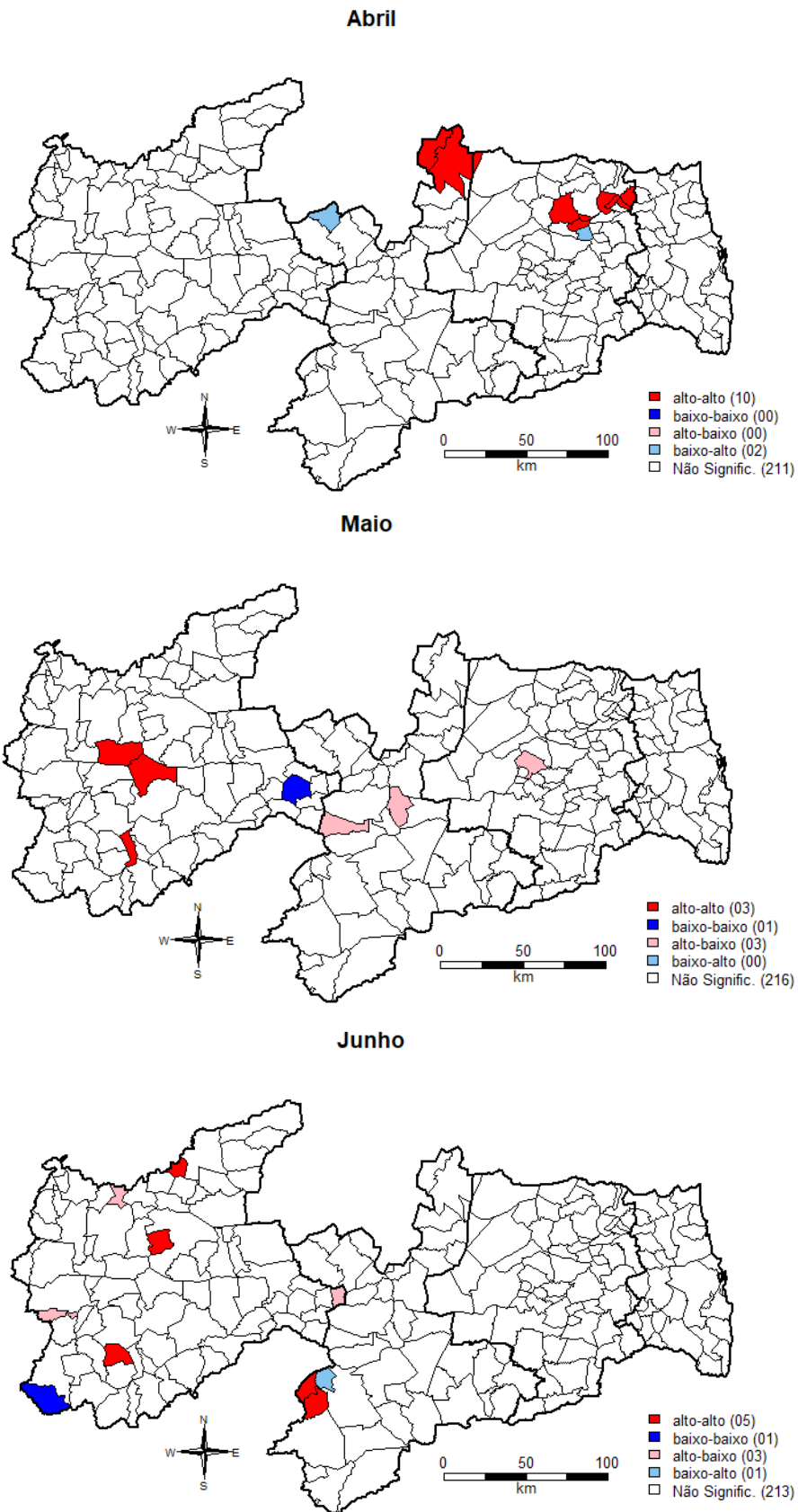
Nota-se que no mês de janeiro/2021 houve formação de 4 *clusters* (14 municípios) com elevada taxa de incidência, posteriormente houve diminuição do número dessas regiões, em abril/2021 apresentou 3 *clusters* (10) municípios, área que demandam nessas regiões uma maior intervenção das autoridades. Apenas os municípios de Cacimba de Areia e Santa Inês nos meses de maio/2021 e junho/2021, respectivamente, apresentaram-se na área azul escuro, ou seja, regiões que contribuem para redução da taxa de casos em sua vizinhança.

Figura 17 - Moran map da taxa de incidência para casos de COVID-19, de janeiro a março/2021 no estado da Paraíba.



Fonte: Elaboração própria com dados do Ministério da Saúde

Figura 18 - Moran map da taxa de incidência para casos de COVID-19, de abril a junho/2021 no estado da Paraíba.



Fonte: Elaboração própria com dados do Ministério da Saúde

## 5.2 REGRESSÃO ESPACIAL

### 5.2.1 Ajuste dos modelos

Numa primeira fase da modelagem, ajustou-se o modelo de regressão clássica usando como variável resposta o logaritmo dos casos registrados nos municípios do estado da Paraíba e a variáveis independentes consideradas foram o Índice de Desenvolvimento Humano Municipal (IDHM) e a taxa de urbanização. Na Tabela 3, têm-se as estimativas dos parâmetros do modelo clássico de regressão usando essas duas covariáveis.

Tabela 3 - Avaliação das pressuposições no modelo clássico de regressão

Meses	Intercepto	IDHM	Tx. Urb.	R <sup>2</sup> ajust.	AIC
<b>Janeiro</b>	-4,75 (0,00085)	12,07 (7,95e-06)	0,02 (0.00129)	0,23	733,62
<b>Fevereiro</b>	-3,42 (0,00804)	9,95 (4.92e-05)	0,02 (3.31e-05)	0,26	691,65
<b>Março</b>	-1,90 (0,15035)	7,50 (0.00276)	0,03 (2.03e-06)	0,24	705,46
<b>Abril</b>	-0,27 (0,840)	3,77 (0.135)	0,03 (8.25e-10)	0,26	711,19
<b>Mai</b>	-0,12 (0,9302)	5,01 (0.0536)	0,02 (1.35e-05)	0,17	724,16
<b>Junho</b>	-0,26 (0,83871)	6,67 (0.00512)	0,02 (0.00016)	0,18	683,67

Fonte: Elaboração própria

De acordo com a Tabela 3, faz-se necessário um modelo espacial adequado, pois o modelo de regressão clássica mostrou um R<sup>2</sup> muito baixo e esse valor implica que aproximadamente 75% dos casos COVID-19 nos municípios paraibanos são causados por variáveis desconhecidas que não são capturadas pelo modelo. Portanto, este estudo confirmou a necessidade de um modelo espacial para retratar os casos COVID-19, ou seja, para melhorar a qualidade de ajuste do modelo, deve-se incluir a autocorrelação espacial presente, usando um modelo de regressão com efeitos espaciais. Na Tabela 4 têm-se os resultados do diagnóstico da dependência espacial usando os testes de multiplicadores de Lagrange.

Tabela 4 - Estimativa dos testes de multiplicadores de Lagrange

<b>Meses</b>	<b>LM – error (valor p)</b>	<b>LM - lag (valor p)</b>
<b>Janeiro</b>	18,10 (2,095e-05)	10,40 (0,00126)
<b>Fevereiro</b>	16,06 (6,152e-05)	9,17 (0,002458)
<b>Março</b>	4,97 (0,02582)	3,04 (0,08101)
<b>Abril</b>	24,36 (7,976e-07)	17,36 (3,096e-05)
<b>Mai</b>	9,64 (0,001902)	6,35 (0,01174)
<b>Junho</b>	21,44 (3,65e-06)	12,03 (0,0005234)

Fonte: Elaboração própria

Através dos testes de multiplicadores de Lagrange acima, verifica-se que os resíduos do modelo de regressão clássica, apresentam dependência espacial ( $p < 0,05$ ). Ambos os testes, mostram significância para ajuste, tanto no modelo espacial autoregressivo, quanto do modelo de erro espacial. Conforme Câmara et al (2004), essa dependência espacial presente nos resíduos era esperada, pois tratando-se de dados espaciais, afirmam que no caso mais comum os resíduos continuam apresentando a autocorrelação espacial presente nos dados.

Uma vez detectada a presença de autocorrelação espacial nos dados, faz-se necessário a introdução de extensões convenientes no modelo de regressão clássica. Da tabela 5 a tabela 10 mostram as estatísticas resumidas para o modelo de defasagem espacial (SAR), modelo de erro espacial (SEM) e modelo de Durbin espacial (SDM) para os meses de janeiro/2021 a junho/2021, respectivamente. Esses modelos mostram uma melhoria de desempenho em relação ao modelo de regressão clássica na modelagem dos casos de COVID-19 na Paraíba.

Tabela 5 - Tabela dos coeficientes estimados para os modelos de Regressão Espacial para janeiro/2021

<b>Covariáveis</b>	<b>SAR (valor p)</b>	<b>SEM (valor p)</b>	<b>SDM (valor p)</b>
<b>Intercepto</b>	-5,30 (9,80e-05)	-6,82 (5,37e-06)	1,68 (0,4116)
<b>IDHM</b>	11,72 (7,68e-06)	15,95 (1,28e-08)	19,39 (6,77e-11)
<b>Taxa Urb.</b>	0,02 (0,0022)	0,01 (0,0178)	0,01 (0,0667)
<b>Rho</b>	0,24 (0,0037)	-	0,38 (2,95e-05)
<b>Lambda</b>	-	0,39 (3,72e-05)	-
<b>AIC</b>	727,21	718,61	712,61 (728,06)

Fonte: Elaboração própria

Tabela 6 - Tabela dos coeficientes estimados para os modelos de Regressão Espacial para fevereiro/2021

<b>Covariáveis</b>	<b>SAR (valor p)</b>	<b>SEM (valor p)</b>	<b>SDM (valor p)</b>
<b>Intercepto</b>	-3,95 (0,001476)	-4,83 (0,0003906)	1,50 (0,4269260)
<b>IDHM</b>	9,53 (6,353e-05)	12,56 (8,587e-07)	1,55 (1,361e-08)
<b>Taxa Urb.</b>	0,02 (4,814e-05)	0,018(0,0003653)	1,53 (0,0022516)
<b>Rho</b>	0,22 (0,00657)	-	0,36 (7,7152e-05)
<b>Lambda</b>	-	0,36 (0,0001471)	-
<b>AIC</b>	686,27	679,25	675,74 (689,37)

Fonte: Elaboração própria

Tabela 7 - Tabela dos coeficientes estimados para os modelos de Regressão Espacial para março/2021

<b>Covariáveis</b>	<b>SAR (valor p)</b>	<b>SEM (valor p)</b>	<b>SDM (valor p)</b>
<b>Intercepto</b>	-2,29 (0,080639)	-2,60 (0,0608153)	2,31 (0,2517777)
<b>IDHM</b>	7,32 (0,003096)	8,84 (0,0006691)	12,37 (2,277e-05)
<b>Taxa Urb.</b>	0,02 (2,311e-06)	0,023 (1,266e-05)	0,02 (0,0002557)
<b>Rho</b>	0,13 (0,1175)	-	0,21 (0,48813)
<b>Lambda</b>	-	0,21 (0,033962)	-
<b>AIC</b>	705,01	702,96	700,36 (703,3)

Fonte: Elaboração própria

Tabela 8 - Tabela dos coeficientes estimados para os modelos de Regressão Espacial para abril/2021

<b>Covariáveis</b>	<b>SAR (valor p)</b>	<b>SEM (valor p)</b>	<b>SDM (valor p)</b>
<b>Intercepto</b>	-1,26 (0,325)	-1,83 (0,19700)	3,55 (0,0692924)
<b>IDHM</b>	3,78 (0,118)	6,61 (0,01256)	9,46(0,0007961)
<b>Taxa Urb.</b>	0,03 (7,389e-10)	0,03 (6,048e-09)	0,03 (8,321e-08)
<b>Rho</b>	0,28 (0,0003)	-	0,38 (7,772e-06)
<b>Lambda</b>	-	0,39 (8,9514e-06)	-
<b>AIC</b>	700,26	693,46	690,19 (708,19)

Fonte: Elaboração própria



Tabela 9 - Tabela dos coeficientes estimados para os modelos de Regressão Espacial para maio/2021

<b>Covariáveis</b>	<b>SAR (valor p)</b>	<b>SEM (valor p)</b>	<b>SDM (valor p)</b>
<b>Intercepto</b>	-0,89 (0,51032)	-1,13 (0,43826)	3,44 (0,0986946)
<b>IDHM</b>	4,98 (0,05055)	6,81 (0,01271)	10,03 (0,0008598)
<b>Taxa Urb.</b>	0,02 (1,644e-05)	0,02 (4,952e-05)	0,02 (0,0003965)
<b>Rho</b>	0,20 (0,0229)	-	0,28 (0,00233)
<b>Lambda</b>	-	0,27 (0,0037445)	-
<b>AIC</b>	720,99	717,76	715,96 (723,22)

Fonte: Elaboração própria

Tabela 10 - Tabela dos coeficientes estimados para os modelos de Regressão Espacial para junho/2021

<b>Covariáveis</b>	<b>SAR (valor p)</b>	<b>SEM (valor p)</b>	<b>SDM (valor p)</b>
<b>Intercepto</b>	-1,20 (0,329248)	-1,30 (0,3290678)	2,57 (0,1611404)
<b>IDHM</b>	6,18 (0,007295)	8,27 (0,0009217)	10,71 (5,268e-05)
<b>Taxa Urb.</b>	0,02 (0,000122)	0,02 (8,964e-05)	0,02 (0,0002852)
<b>Rho</b>	0,26 (0,00218)	-	0,39 (8,75e-06)
<b>Lambda</b>	-	0,38 (2,47e-05)	-
<b>AIC</b>	676,28	667,89	663,11 (680,88)

Fonte: Elaboração própria

Verifica-se que em todos os modelos e em todos os meses existe uma relação direta entre o logaritmo do número de casos de COVID-19 com as variáveis independentes: Índice de Desenvolvimento Humano Municipal (IDHM) e taxa de urbanização, ou seja, as regiões (cidades) que apresentam alto IDHM e alta taxa de urbanização constituam os locais de maior número de casos.

### **5.2.2 Análise de Regressão Espacial**

Na análise de regressão espacial, temos a taxa da casos de COVID-19 como variável dependente e como variáveis independentes o Índice de Desenvolvimento Humano Municipal (IDHM) e a taxa de urbanização, esta mostra a porcentagem da população residente constituída pelos moradores em domicílios urbanos em relação à população total, aquela mede o

desenvolvimento humano em cada município sintetizando uma média de três dimensões básicas e universais da vida: o acesso ao conhecimento (educação), o direito a uma vida longa e saudável (longevidade) e o direito a um padrão de vida digno (renda) (Brasil, 2020a) e pode ser classificado em muito baixo (0 a 0,499), baixo (0,500 a 0,599), médio (0,600 a 0,699), alto (0,700 a 0,799) e muito alto (0,800 a 1).

Conforme os resultados apresentados na Tabela 11, observa-se que em todos os modelos ajustados pelo critério da informação de Akaike (AIC), os valores obtidos para Modelo Espacial de Durbin (SDM) são inferiores aos encontrados nos demais modelos, o que mostra a melhoria do ajustamento com modelos espaciais proposto neste trabalho.

verifica-se que o melhor modelo que se ajustou aos dados foi o Modelo Espacial de Durbin (SDM), pois o mesmo apresentou o menor valor de AIC em todos os seis meses de estudo.

Tabela 11 - Tabela dos valores do AIC para os meses de janeiro/2020 a junho/2020.

Meses	MRC	SAR	SEM	SDM
<b>Janeiro</b>	733,62	727,21	718,61	<b>712,61</b>
<b>Fevereiro</b>	691,65	686,27	679,25	<b>675,74</b>
<b>Março</b>	705,46	705,01	702,96	<b>700,36</b>
<b>Abril</b>	711,19	700,26	693,46	<b>690,19</b>
<b>Mai</b>	724,16	720,99	717,76	<b>715,96</b>
<b>Junho</b>	693,67	676,28	667,89	<b>663,11</b>

Fonte: Elaboração própria

Com a escolha do Modelo Espacial de Durbin (SDM), o ajuste para cada mês é mostrada nas equações abaixo.

Para o mês de janeiro/2020 a reta ajustada é dada pela equação 5.1

$$\widehat{TIC}(\hat{y}) = 19.3855 IDHM + 0.0099 TU + 1.6827, \quad (5.1)$$

dessa forma, o acréscimo de uma unidade no IDHM, há um aumento de 19,3855 na taxa da casos de COVID-19, bem como o aumento de uma unidade na taxa de urbanização, também há um acréscimo de 0,0099 na taxa de casos.

Para o mês de fevereiro/2020 a reta ajustada é dada pela equação 5.2

$$\widehat{TIC}(\hat{y}) = 1,5540 IDHM + 1,5319 TU + 1.5004, \quad (5.2)$$

Para o mês de março/2020 a reta ajustada é dada pela equação 5.3

$$\widehat{TIC}(\hat{y}) = 12,3672 IDHM + 0,0195 TU + 2,3067, \quad (5.3)$$

Para o mês de abril/2020 a reta ajustada é dada pela equação 5.4

$$\widehat{TIC}(\hat{y}) = 9,4588 IDHM + 0,0276 TU + 3,5500, \quad (5.4)$$

Para o mês de maio/2020 a reta ajustada é dada pela equação 5.5

$$\widehat{TIC}(\hat{y}) = 10,0312 IDHM + 0,0195 TU + 3,4416, \quad (5.5)$$

Para o mês de junho/2020 a reta ajustada é dada pela equação 5.6

$$\widehat{TIC}(\hat{y}) = 10,7095 IDHM + 0,0176 TU + 2,5672, \quad (5.6)$$

As conclusões dos meses anteriores (fevereiro/2020 a junho/2020) seguem a mesma interpretação do mês de janeiro/2020, ou seja, o acréscimo de uma unidade no IDHM há um aumento na taxa de incidência de casos de COVID-19 conforme o respectivo coeficiente na equação de cada mês. A mesma conclusão é dada, também, para a taxa de urbanização.

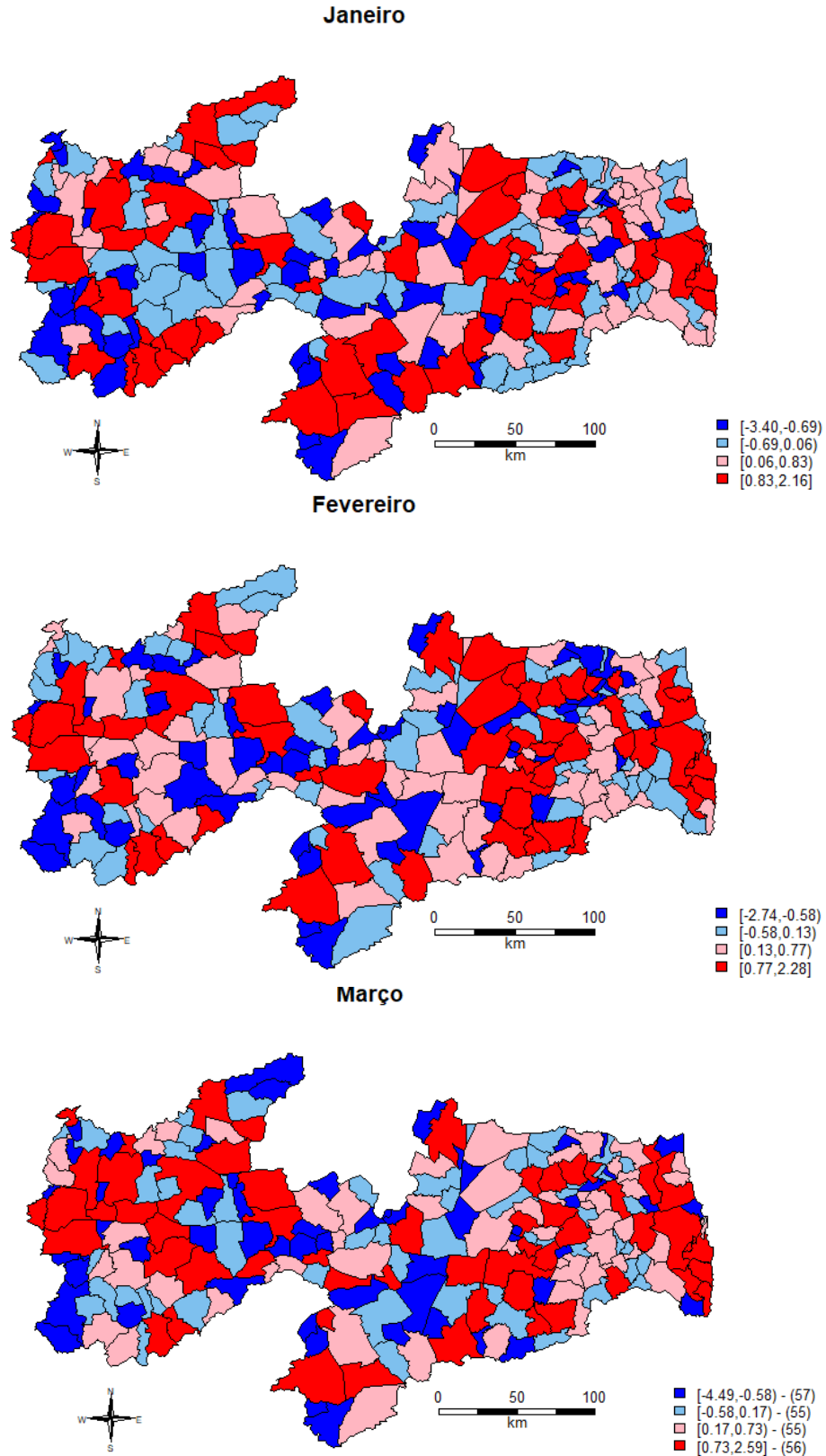
De posse das retas ajustadas, passamos a analisar o modelo através dos mapas de resíduos. O mapeamento dos resíduos é a forma de avaliar do ponto de vista espacial os resultados dos modelos (Anselin, 2005), indicando as regiões sub ou superestimadas que podem indicar padrões espaciais. Os resíduos são as distâncias entre nossos valores observados e os valores previstos. Então, em essência, eles são as diferenças da realidade observada do seu modelo e sua linha de regressão é otimizada para ser aquele que melhor representa seus dados. Portanto, os resíduos são muito úteis para diagnosticar se o seu modelo é uma boa representação da realidade ou não.

Os mapas dos resíduos do Modelo Espacial de Durbin (SDM) revelaram grandes agregados de resíduos positivos e negativos em todos os meses, o que mostra a permanência da dependência espacial comprovada pelo teste de Moran permitindo gerar resíduos distribuídos aleatoriamente pela área de estudo, como pode ser observado na Figura 19 e na Figura 20 . O mapa dos resíduos indica a existência de *clusters* de regiões onde o modelo de regressão apresentou erro, tanto para mais quanto para menos. A presença desses agrupamentos mostra a influência do espaço nos resultados, ou seja, que os resultados são espacialmente dependentes.

As áreas (municípios) vermelhas são locais onde os valores reais são maiores do que o modelo estimado. As áreas azuis são locais onde os valores reais são menores do que o modelo

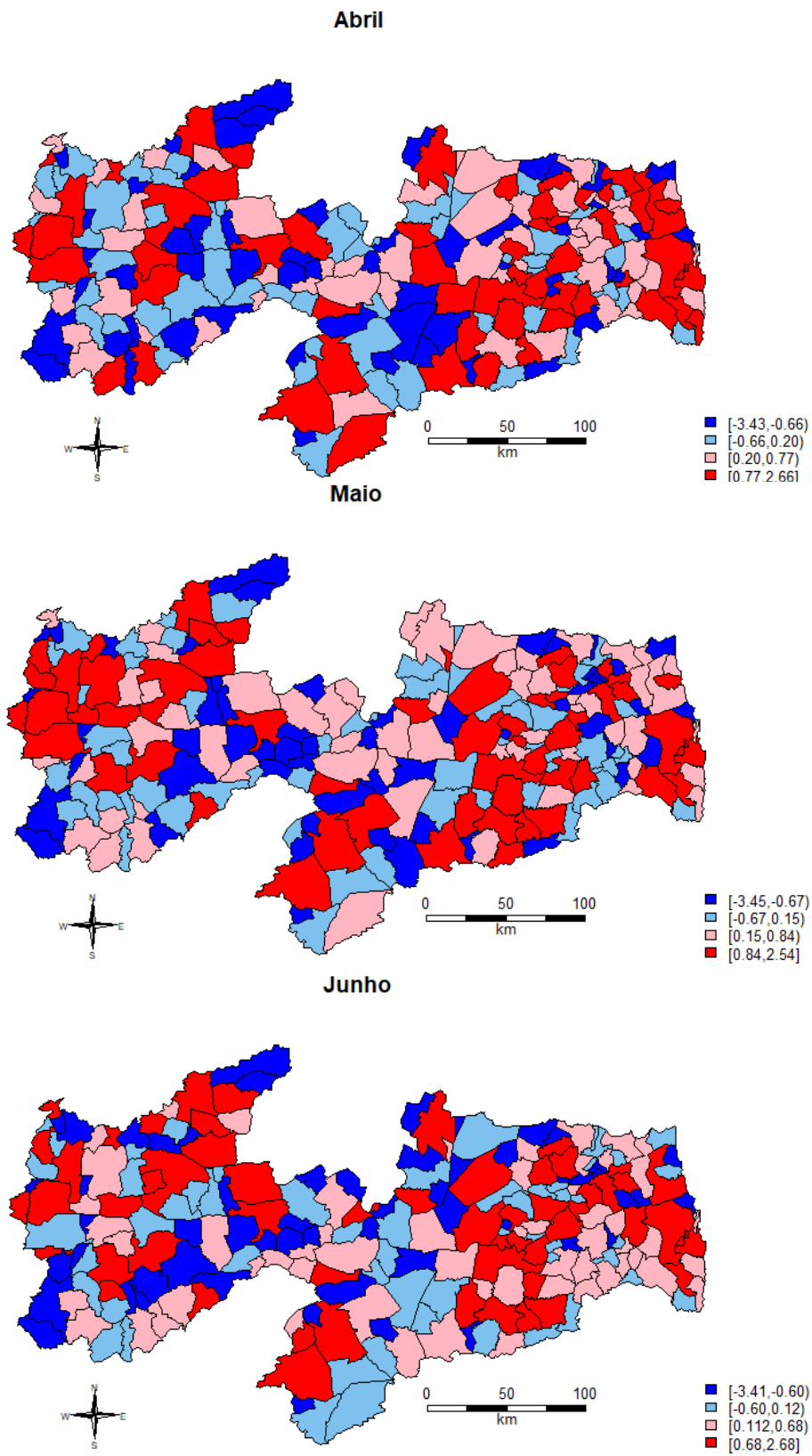
estimado. As áreas com as cores mais claras (vermelha e azul) indicam que o erro foi menor, dessa forma indica um melhor modelo.

Figura 19 - Mapa dos resíduos para o Modelo Espacial de Durbin (SDM), de janeiro a março/2021.



Fonte: Elaboração própria com dados do Ministério da Saúde

Figura 20 - Mapa dos resíduos para o Modelo Espacial de Durbin (SDM), de abril a junho/2021.



Fonte: Elaboração própria com dados do Ministério da Saúde

## 6 CONCLUSÃO

Ao lançar mais luz sobre a interação entre COVID-19 e características relacionadas à população do estado da Paraíba, este estudo pode analisar a distribuição espacial da taxa de incidência da COVID-19 no estado da Paraíba de modo a compreender a variação espacial existente entre seus municípios, bem como defender mais atenção para prever surtos locais e desenvolver estratégias de controle para COVID-19. Este trabalho mostra que foi possível identificar a presença de dependência espacial da taxa de incidência para casos de COVID-19, de janeiro/2021 a junho/2021 no estado da Paraíba e dessa forma ajustar um modelo espacial capaz de descrever o relacionamento entre a taxa de incidência para casos de COVID-19 e as taxas de urbanização e o IDHM.

Através das equações de predição desenvolvidas neste trabalho podemos identificar que as variáveis taxa de urbanização e IDHM influenciaram positivamente para o aumento da taxa de casos, ou seja, as áreas com maior taxa de urbanização e maior IDHM proporcionaram uma maior taxa de casos. Dessa forma esse estudo mostra a distribuição espacial da COVID-19 no Estado da Paraíba e sinaliza para um estudo de prevenção de surtos e no desenvolvimento de estratégias de controle pelos poderes públicos.

## REFERÊNCIAS

- ALBUQUERQUE, PHM. Curso de econometria espacial aplicada. Fortaleza: IPECE, 2009.
- ALMEIDA, Eduardo. Econometria espacial. Campinas–SP. Alínea, 2012.
- ANDRADE, Ana Lúcia et al. Introdução à estatística espacial para a saúde pública. 2007. Santos,2007
- ANSELIN, Luc. Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical analysis*, v. 20, n. 1, p. 1-17, 1988.
- ANSELIN, Luc; REY, Serge. Properties of tests for spatial dependence in linear regression models. *Geographical analysis*, v. 23, n. 2, p. 112-131, 1991.
- ANSELIN, Luc. Local indicators of spatial association—LISA. *Geographical analysis*, v. 27, n. 2, p. 93-115, 1995.
- ANSELIN, Luc; FLORAX, Raymond JGM. Small sample properties of tests for spatial dependence in regression models: Some further results. In: *New directions in spatial econometrics*. Springer, Berlin, Heidelberg, 1995. p. 21-74.
- ANSELIN, L. *Exploring Spatial Data with GeoDa: A Workbook*. Center for Spatially Integrated Social Science, 2005, 226p.
- ARAÚJO, Everton C. de; URIBE-OPAZO, Miguel A.; JOHANN, Jerry Adriani. Modelo de regressão espacial para estimativa da produtividade da soja associada a variáveis agrometeorológicas na região oeste do estado do Paraná. *Engenharia Agrícola*, v. 34, n. 2, p. 286-299, 2014.
- Bonita R, Beaglehole R, Kjellstrom T. *Epidemiologia Básica*. 2ª ed. São Paulo: Grupo Editorial Nacional; 2010.
- Brasil. (2020a). Atlas de Desenvolvimento Humano no Brasil. <<http://www.atlasbrasil.org.br/>>.
- BRASIL. Plano Estadual de Assistência Social da Paraíba (2020-2023). Secretaria de Estado do Desenvolvimento Humano (SEDH), Diretoria do Sistema Único de Assistência Social (DSUAS). Paraíba, 2020.
- CÂMARA, G.; CARVALHO, M. S.; CRUZ, O. G.; CORREA, V. Análise espacial de áreas. *Análise espacial de dados geográficos*, Empresa Brasileira de Pesquisa Agropecuária Brasília, v. 2, 2004
- CAVALCANTE, João Roberto; ABREU, Ariane de Jesus Lopes de. COVID-19 no município do Rio de Janeiro: análise espacial da ocorrência dos primeiros casos e óbitos confirmados. *Epidemiologia e Serviços de Saúde*, v. 29, p. e2020204, 2020.

CHEIN, Flávia. Introdução aos modelos de regressão linear: um passo inicial para compreensão da econometria como uma ferramenta de avaliação de políticas públicas. 2019.

DANTAS, Rubens Alves; MAGALHÃES, André Matos; VERGOLINO, José Raimundo de Oliveira. Um modelo espacial de demanda habitacional para a cidade do Recife. *Estudos Econômicos* (São Paulo), v. 40, p. 891-916, 2010.

GOLGHER, André Braz. Introdução à econometria espacial. Paco Editorial, 2015.

LUZARDO, Antonio José Rocha; CASTAÑEDA FILHO, Rafael March; RUBIM, Igor Brum. Análise espacial exploratória com o emprego do Índice de Moran. *GEOgraphia*, v. 19, n. 40, p. 161-179, 2017.

MEDEIROS, Eduardo Alexandrino Servolo. A luta dos profissionais de saúde no enfrentamento da COVID-19. *Acta Paulista de Enfermagem*, v. 33, 2020.

MENDES, Micheline da Silveira et al. Análise espacial da tuberculose em menores de 15 anos de idade e risco socioeconômico: um estudo ecológico na Paraíba, 2007-2016. *Epidemiologia e Serviços de Saúde*, v. 30, 2021.

MENDES, Micheline da Silveira. Tuberculose em menores de 15 anos no estado da Paraíba: uma análise espaço-temporal no período de 2007 a 2016. 2019. Dissertação (Mestrado Acadêmico em Saúde Pública) – Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2019.

MORAN, P. A. P. Notes on Continuous Stochastic Phenomena. *Biometrika* 37, p. 17-23, 1950.

NETTO, R. G. F.; DO NASCIMENTO CORRÊA, J. W. Epidemiologia do surto de doença por coronavírus (COVID-19). *Desafios-Revista Interdisciplinar da Universidade Federal do Tocantins*, v. 7, n. Especial-3, p. 18-25, 2020.

SILVA, N. C. N. et al. Análise de dados de área aplicada a dois indicadores econômicos de mesorregiões do estado de Minas Gerais. *Revista Brasileira de Biometria*, São Paulo, v. 29, n. 3, p. 369-395, 2011.

TEIXEIRA, F. J. A dependência espacial do valor do prêmio do seguro de automóvel. Dissertação (Mestrado) | UNIVERSIDADE FEDERAL DE LAVRAS, 2015. 75 p.

TOBLER, Waldo R. A computer movie simulating urban growth in the Detroit region. *Economic geography*, v. 46, n. sup1, p. 234-240, 1970.

WOOLDRIDGE, J. M. Introdução à econometria: Uma abordagem moderna. São Paulo: Thomson, 2005.

ZILLI, Carlos Augusto et al. Regressão geograficamente ponderada aplicada na avaliação em massa de imóveis urbanos. 2020.