



UNIVERSIDADE ESTADUAL DA PARAÍBA  
CENTRO DE CIÊNCIAS E TECNOLOGIA  
DEPARTAMENTO DE ESTATÍSTICA

Arthur Oliveira Costa

# **Modelagem Espacial no Mapeamento de Doenças Transmitidas pelo Aedes Aegypti no Estado da Paraíba**

Campina Grande - PB

01 de dezembro de 2017

Arthur Oliveira Costa

# **Modelagem Espacial no Mapeamento de Doenças Transmitidas pelo *Aedes Aegypti* no 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: Ricardo Alves de Olinda

Campina Grande - PB  
01 de dezembro de 2017

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.

C837m Costa, Arthur Oliveira.

Modelagem espacial no mapeamento de doenças transmitidas pelo *Aedes Aegypti* no estado da Paraíba [manuscrito] : / Arthur Oliveira Costa. - 2017.

39 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2017.

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

1. Métodos estatísticos. 2. Índice I de Moran. 3. Regressão espacial. 4. Dados de área.

21. ed. CDD 519.5

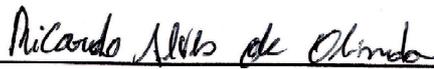
Arthur Oliveira Costa

## **Modelagem Espacial no Mapeamento de Doenças Transmitidas pelo Aedes Aegypti no 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.

Trabalho aprovado em 01 de dezembro de 2017.

### **BANCA EXAMINADORA**



Dr. Ricardo Alves de Olinda  
Universidade Estadual da Paraíba



Dr. Tiago Almeida de Oliveira  
Universidade Estadual da Paraíba



Dra. Divanilda Maia Esteves  
Universidade Estadual da Paraíba



*Dedico este trabalho a minha mãe, Maria do Socorro, ao meu pai Josenildo Costa, por me encorajar a conquistar os meus sonhos e que sempre me apoiou em todos os momentos difíceis.*

# Agradecimentos

A Deus, por tudo.

A minha família, principalmente ao meu pai, pelo apoio, incentivo, ajuda financeira e oportunidade de alcançar esse objetivo, a minha mãe, pela presença, conforto e carinho todos esses anos.

Ao professor Ricardo Alves de Olinda, pelas horas de dedicação, comprometimento e ensinamentos, por acreditar e tornar possível esse trabalho, por ser meu orientador e amigo, muito obrigado.

A todos os professores do Departamento de Estatística da UEPB, especialmente a Tiago Almeida de Oliveira, Divanilda Maia Esteves, João Gil de Luna, Ana Patrícia Bastos Peixoto, Gustavo Henrique Esteves pela dedicação e principalmente a amizade.

Enfim, a todos que fazem parte da minha vida e torcem pelo meu sucesso.

*“Tudo o que um sonho precisa para ser  
realizado é alguém que acredite que  
ele possa ser realizado.”  
(Roberto Shinyashiki)*

# Resumo

Uma das principais causas que vem preocupando diversos países, inclusive o Brasil, atualmente é a disseminação do mosquito *Aedes aegypti*, vetor responsável por transmitir os vírus, da *Dengue*, *Zika vírus* e *Chikungunya*. A dinâmica da difusão da infecção dos vírus é um processo complexo. O uso de metodologias que melhor destacam a influência dos processos ambientais e sociais nos padrões de transmissão espacial dessas doenças são de extrema valia para adoção de medidas mitigadoras para prevenção e controle do *Aedes aegypti*. Na análise espacial sob o enfoque de dados de área, as variáveis são provenientes de levantamentos populacionais, sendo as áreas definidas por polígonos fechados com dados internos homogêneos. A análise espacial compõe um conjunto de procedimentos cujo foco é encontrar um modelo inferencial que incorpore explicitamente as relações espaciais constituintes de um fenômeno. Para isso, objetivou-se neste trabalho detectar a existência de dependência espacial dos indivíduos contaminados com os vírus transmitidos pelo *Aedes aegypti* no Estado da Paraíba, levando-se em consideração novas metodologias para análise de dados de área. Portanto, utilizou-se das técnicas de dados de área, realizando primeiramente uma análise descritiva que em uma tabela apresentou-se o município que tem a maior incidência no ano, depois se utilizou os métodos descritivos de visualização da variável e observou-se o comportamento da variável resposta, em seguida apresentou-se a existência de autocorrelação espacial global e local no Estado da Paraíba, apresentou-se também os gráficos de espalhamento de *Moran*, *Box Map*, *Lisa Map* e *Moran Map*. Utilizou-se os modelos de regressão espacial para explicar o relacionamento entre a proporção dos casos notificados por dengue com as variáveis socioeconômicas e ambientais e observou-se que o MEAR foi o que melhor ajustou-se aos dados. Por fim, realizou-se uma análise de resíduos para validar os modelos que foram ajustados aos dados.

**Palavras-chave:** *Dados de área. Índice I de Moran. Regressão Espacial.*

# Abstract

Nowadays one of the major causes which is alerting various countries including Brazil is the *Aedes Aegypti* mosquito dissemination, the responsible vector for transmitting Dengue, Zika and Chikungunya virus. The diffusion dynamics of the virus infection involves a complex process. The usage of methodologies which best points out the influence from the environmental and social processes on the spatial transmission patterns from those illness are of extreme importance in the adoption of mitigating measures to the *Aedes aegypti* prevention and control process. On the spatial analysis under the focus of the area data, the variables are provenient from populational researches, where the areas are defined by closed polygons with homogeneous internal data. The spatial analysis composes a set of procedures in which the focus is to find an inference model that explicitly fills the spatial relationships, part of a phenomenon. To this, the objective of this research was to detect the existence of a spatial dependence from the contaminated individuals with the virus transmitted by *Aedes Aegypti* in Paraiba's State, taking in consideration new methodologies to the area data analysis. Therefore, were used area data techniques, carrying out first a descriptive analysis which showed up in a table the city that has the highest annual incidence, after that were used the descriptive methods of visualization from the variable response, then was presented the existence of local global spatial autocorrelation in Paraiba's State, it was also showed the spread graphs from Moran, Box Map, Lisa Map and Moran Map. Using the spatial regression models to explain the relationship between the proportion from the notified cases of Dengue with the socioeconomics and environment variables, it was also observed that SAR was the one which best adjusted to the data. Lastly, it was carried out a residues analysis to validate the models which were adjusted according to the data.

**Key-words:** Area data. Moran Index I. Spatial Regression.

# Lista de ilustrações

Figura 1 – Municípios da Paraíba com casos de notificações por dengue. . . . .	28
Figura 2 – Mapa de intervalos iguais para proporção de casos notificados por dengue no Estado da Paraíba. . . . .	29
Figura 3 – Mapa de quintis para proporção de casos notificados por dengue no Estado da Paraíba. . . . .	29
Figura 4 – Mapa de desvios padrão para proporção de casos notificados por dengue no Estado da Paraíba. . . . .	30
Figura 5 – Índice $I_i$ de <i>Moran</i> Local da proporção de casos notificados por dengue no Estado da Paraíba. . . . .	31
Figura 6 – Mapa da média móvel para proporção de casos notificados por dengue no Estado da Paraíba. . . . .	31
Figura 7 – Mapa do espalhamento de <i>Moran</i> para proporção de casos notificados por dengue no Estado da Paraíba. . . . .	32
Figura 8 – <i>Box Map</i> da proporção de casos notificados por dengue no Estado da Paraíba. . . . .	33
Figura 9 – <i>LISA Map</i> da proporção de casos notificados por dengue no Estado da Paraíba. . . . .	33
Figura 10 – <i>Moran Map</i> da proporção de casos notificados por dengue no Estado da Paraíba. . . . .	34
Figura 11 – Mapas de resíduos para os modelos de regressão clássica, espacial autorregressivo misto, erro espacial e espacial de <i>Durbin</i> para a proporção de casos notificados por dengue no Estado da Paraíba. . . . .	36

# Lista de tabelas

Tabela 1 – Tabela dos municípios com maior número de casos de notificações por dengue no Estado da Paraíba. . . . .	28
Tabela 2 – Tabela das estimativas do Índice de <i>Moran</i> e a Estatística <i>c</i> de <i>Geary</i> .	30

# Lista de abreviaturas e siglas

OMS	Organização Mundial da Saúde
SINAN	Sistema de Informação de Agravos de Notificação
IBGE	Instituto Brasileiro de Geografia e Estatística
PROPOR	Proporção dos casos notificados por dengue em relação a população do município da Paraíba
IDH	Índice de Desenvolvimento Humano
GINI	Índice de Gini
PRECIP	Precipitação média anual
TEMP	Temperatura média anual em graus Celsius
PDR8	Proporção de domicílios com oito residentes
PDSREND	Proporção de domicílios sem renda mensal
PDEOD	Proporção de domicílios com eletricidade de origem desconhecida.
PSB	Proporção de Saneamento Básico
AIC	Critério de Informação de Akaike
RC	Regressão Clássica
MEAR	Modelo Espacial Autorregressivo Misto
MEE	Modelo de Erro Espacial
MED	Modelo Espacial de Durbin
Q1	Primeiro quadrante
Q2	Segundo quadrante
Q3	Terceiro quadrante
Q4	Quarto quadrante
AA	Alto Alto
BB	Baixo Baixo

AB Alto Baixo

BA Baixo Alto

# Lista de símbolos

$\mu$	Letra grega minúscula mu
$\alpha$	Letra grega minúscula alpha
$\beta$	Letra grega minúscula beta
$\varepsilon$	Letra grega minúscula epsilon
$\sigma$	Letra grega minúscula sigma
$\rho$	Letra grega minúscula rho
$\theta$	Letra grega minúscula theta
$\Theta$	Letra grega maiúscula Theta
$\tau$	Letra grega minúscula tau
$\in$	Pertence
$\ell$	El
$\pi$	Letra grega minúscula pi
$\delta$	Letra grega minúscula delta

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>15</b>
<b>2</b>	<b>REVISÃO DE LITERATURA</b>	<b>17</b>
<b>2.1</b>	<b>Material</b>	<b>17</b>
<b>2.2</b>	<b>Métodos Estatísticos</b>	<b>18</b>
2.2.1	Métodos Descritivos de Visualização da Variável	19
2.2.2	Matriz $W$	19
2.2.3	Autocorrelação Espacial	19
2.2.4	Média Móvel	21
2.2.5	Gráfico de Espalhamento de <i>Moran</i>	21
2.2.6	<i>Box Map</i> , <i>LISA Map</i> e <i>Moran Map</i>	22
2.2.7	Regressão Clássica	22
2.2.8	Modelo Espacial Autorregressivo Misto	23
2.2.9	Modelo de Erro Espacial	23
2.2.10	Modelo Espacial de <i>Durbin</i>	24
2.2.11	Estimação dos Parâmetros em Modelos de Regressão Espacial	24
2.2.12	Seleção de Modelos	26
<b>3</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>28</b>
<b>4</b>	<b>CONCLUSÃO</b>	<b>37</b>
	<b>REFERÊNCIAS</b>	<b>38</b>

# 1 Introdução

Atualmente, grande parte do planeta está sob o risco epidêmico devido à rápida propagação do mosquito *Aedes aegypti* e à difusão de vários vírus associados a ele, principalmente a *dengue*, o *Zika vírus* e a *Chikungunya*. O Ministério da Saúde<sup>1</sup> e a Organização Mundial da Saúde (OMS)<sup>2</sup> já confirmaram que a infecção causada pelo *Zika vírus* está mesmo relacionada aos casos de microcefalia e também a outras síndromes neurológicas como a de *Guillain-Barrém*. O problema é que ainda se sabe muito pouco sobre como combater essa epidemia, pois é a primeira vez que pesquisadores associam o *Zika vírus* à microcefalia, portanto não existem referências na literatura - nem para o diagnóstico, nem para o tratamento.

A infestação do *Aedes aegypti* é elevada em áreas com serviços deficientes de coleta de lixo e água potável. Fatores ambientais de períodos de chuva com temperaturas e umidade elevadas também favorecem o ciclo de vida e disseminação do mosquito (MARINO et al., 2016). O uso de metodologias que melhor destacam a influência dos processos ambientais e sociais nos padrões de transmissão espacial dessas doenças são de extrema valia para adoção de medidas mitigadoras para prevenção e controle do *Aedes aegypti*.

A dinâmica da difusão da infecção dos vírus é um processo complexo. Modelos estatísticos, que levam em consideração a dependência espacial, podem relacionar a incidência da doença com fatores ambientais e socioeconômicos mostrando que o risco aumenta em áreas com elevada infestação de mosquitos *Aedes aegypti*, densidade demográfica e sobretudo mobilidade urbana.

Em um novo avanço da dengue no país, 135 cidades brasileiras terminaram 2015 com epidemia da doença. Isso significa que esses municípios tinham mais de 300 casos a cada 100 mil habitantes, patamar utilizado pela OMS para definir o quadro de epidemia. Dessas 135 cidades que começaram o ano de 2016 enfrentando quadro epidêmico, 29 estão na região Nordeste, região com maior surto do *Zika vírus* no país. Segundo o site G1/PB<sup>3</sup>, o Estado da Paraíba segue em quarto lugar no *ranking* dos casos confirmados de microcefalia, uma infecção que provoca má-formação do cérebro de bebês e que o governo brasileiro confirmou a relação com o *Zika vírus*.

A principal questão que motivou esta pesquisa baseia-se na hipótese de que pode haver dependência espacial na propagação do mosquito *Aedes aegypti*, vetor de doenças, que transmite os vírus da *dengue*, o *Zika vírus* e a *Chikungunya*. A dependência espacial

---

<sup>1</sup> <http://portalsaude.saude.gov.br/>

<sup>2</sup> <http://www.who.int/en/>

<sup>3</sup> <http://g1.globo.com/pb/paraiba/noticia/2016/02/apenas-tres-cidades-tem-atendimento-para-criancas-com-microcefalia-na-pb.html>

pode ser um forte indicador de que as ações de controle não estejam produzindo o resultado desejado de minimizar as chances de transmissão em regiões nas quais são detectados focos do problema. Além disso, é possível relacionar essas regiões a fatores ambientais e socioeconômicos.

A análise espacial compõe um conjunto de procedimentos cujo foco é encontrar um modelo inferencial que incorpore explicitamente as relações espaciais constituintes de um fenômeno. Normalmente, a modelagem é iniciada pela análise exploratória associada à visualização dos dados por meio de gráficos e mapas e, posteriormente, identificam-se padrões de dependência espacial das variáveis em estudo. É importante ressaltar que o problema a ser abordado nesta pesquisa leva em consideração o número de indivíduos infectados por unidade de área, isto é, número de indivíduos infectados em cada um dos 223 municípios do Estado Paraíba.

Na análise espacial sob o enfoque de dados de área, as variáveis são provenientes de levantamentos populacionais, sendo as áreas definidas por polígonos fechados com dados internos homogêneos. No entanto, por questões operacionais ou políticas, na coleta de informações, não há nenhuma garantia da homogeneidade dos dados, ou seja, pode haver, por exemplo, grupos sociais distintos agregados em uma mesma região de coleta, como favela e área nobre. Assim, indicadores globais podem apresentar distorções, havendo necessidade do uso de técnicas para ajustes locais (CÂMARA; MONTEIRO, 2001). Sendo assim, este trabalho tem por objetivos: detectar a existência de dependência espacial dos indivíduos contaminados com os vírus transmitidos pelo *Aedes aegypti* no Estado da Paraíba, levando-se em consideração novas metodologias para análise de dados de área. Ajustar índices de autocorrelação espacial sobre a proporção de pessoas infectadas com os vírus que são transmitidos pelo mosquito *Aedes aegypti* no Estado da Paraíba. Quantificar através de mapas temáticos, valores preditos da variabilidade espacial, em escala global e local, a propagação do mosquito *Aedes aegypti* no Estado da Paraíba. Implementar um modelo que seja capaz de identificar espacialmente áreas de risco de contaminação pelo mosquito *Aedes aegypti* no Estado da Paraíba em tempo real, dando subsídio aos órgãos responsáveis quanto a avaliação de suas ações.

## 2 Revisão de Literatura

Considerando as etapas do processo de um tratamento estatístico de dados espaciais, de acordo com Bailey e Gatrell (1995), os processos de análise espacial incluem métodos de visualização, métodos que investiguem o padrão espacial dos dados e sua correlação espacial, métodos que auxiliem na escolha de um modelo estatístico e a estimação dos parâmetros desse modelo.

Para LANDIM (1998), quando se utilizam modelos da Estatística Clássica para representar as propriedades dos valores amostrais distribuídos espacialmente, supeita-se que esses sejam realizações de uma variável aleatória, que as posições relativas da amostra são ignoradas e presume-se que todos os valores amostrais tenham a mesma probabilidade de serem escolhidos. Assim, muitos modelos da Estatística Clássica exigem a independência das variáveis.

De acordo com LANDIM (1998), Estatística Espacial é um outro conjunto de modelos que traz resultados diferentes daqueles obtidos pela Estatística Clássica, sendo os primeiros geralmente mais informativos por incorporarem a dimensão espacial. Para sua análise, são necessárias, pelo menos, as informações sobre a localização e os atributos, que são os valores associados aos dados. Parte-se do pressuposto de que os dados podem ser espacialmente dependentes.

Para mostrar a importância da localização amostral, cita-se Câmara e Monteiro (2001), que apresentam o caso estudado pelo médico inglês *John Snow*, na cidade de Londres, em 1854, onde consta o caso de uma grave epidemia de cólera, doença desconhecida para a época quanto à forma de contaminação. Como uma das estratégias para lidar com o problema, foi identificada no mapa da cidade a localização exata das residências com óbitos por cólera e dos poços de água que abasteciam a cidade. Com a representação espacial dos dados, percebeu-se que a maioria dos casos estava concentrada em torno do poço de *Broad Street*, como consequência, este foi lacrado, contribuindo para acabar com a epidemia.

### 2.1 Material

Os dados utilizados nesta pesquisa são de fonte secundária do período de 2007 a 2016, relativos às séries de dados das doenças de notificação compulsória do SINAN. Dados relativos a indicadores socioeconômicos foram obtidos através do site do IBGE.<sup>1</sup>

---

<sup>1</sup> <http://www.ibge.gov.br/home/>

Dados relativos a variáveis ambientais foram obtidos por meio dos sites CLIMATE<sup>2</sup> e AESA<sup>3</sup>. Dados relativos aos casos notificados por dengue foram cedidos por meio da secretária de saúde do Estado da Paraíba, onde a variável dependente é PROPOR. As 8 variáveis independentes estão listadas abaixo, tem-se 223 observações, onde cada observação corresponde a um município da Paraíba.

- PROPOR: Proporção dos casos notificados por dengue em relação a população do município da Paraíba;
- IDH: Índice de Desenvolvimento Humano;
- GINI: Índice de Gini;
- PRECIP: Precipitação média anual;
- TEMP: Temperatura média anual em graus Celsius;
- PDR8: Proporção de domicílios com oito residentes;
- PDSREND: Proporção de domicílios sem renda mensal;
- PDEOD: Proporção de domicílios com eletricidade de origem desconhecida;
- PSB: Proporção de Saneamento Básico.

A Paraíba é uma das 27 unidades federativas do Brasil. Está situada a leste da região Nordeste e tem como limites o estado do Rio Grande do Norte ao norte, o Oceano Atlântico, a leste, Pernambuco ao sul e o Ceará a oeste. Ocupa uma área de  $56.439 \text{ km}^2$  (pouco menor que a Croácia). A população do estado nas estimativas do IBGE em 2015 foi de 3.943.885 milhões de habitantes, sendo o décimo terceiro estado mais populoso do Brasil.

## 2.2 Métodos Estatísticos

Existem vários métodos e técnicas na estatística espacial, um deles é a autocorrelação espacial que apresenta o nível de similaridade, proximidade, correlação e aleatoriedade dos valores quantitativos amostrais representados. Uma técnica muito utilizada para esse tipo de análise são os mapas de *cluster* que representam as informações geográficas agrupadas em valores próximos (BIVAND; HAUKE; KOSSOWSKI, 2013).

Os dados de área referem-se a um mapa de um região  $R$ , particionado em  $n$  áreas  $A_i$ ,  $i = 1, 2, \dots, n$ , onde em cada área é feita uma observação aleatória  $Y_i$ , em que  $Y_i$  é

---

<sup>2</sup> <http://www.climate-data.org>

<sup>3</sup> <http://www.aesa.pb.gov.br>

uma agregação sobre a área  $A_i$ , onde essa medição refere-se a toda área  $A_i$ , não a um ponto particular.

### 2.2.1 Métodos Descritivos de Visualização da Variável

Antes de iniciar qualquer análise estatística de um conjunto de dados é necessário fazer uma análise descritiva. Três métodos de visualização da variável resposta são realizados nesse trabalho: método de intervalos iguais, que é a diferença entre o valor máximo e o valor mínimo da variável, onde é dividida pelo número de classes, obtendo-se o valor do intervalo de cada classe. Método dos quintis, que é o intervalo de valores de cada classe, onde é calculado por meio do ordenamento dos valores da variável, é uma forma de separar as classes em percentual de valores da variável. E o método de número de desvios padrão, onde são geradas classes separadas pelo valor do desvio padrão.

### 2.2.2 Matriz W

Formalmente, a associação de observações na vizinhança definido para cada localização pode ser expressa por meio de uma contiguidade ou pesos espaciais numa matriz  $\mathbf{W}$  de ordem  $n$ , ou seja,

$$\mathbf{W} = \begin{bmatrix} W_{11} & W_{12} & \cdots & W_{1n} \\ W_{21} & W_{22} & \cdots & W_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n2} & \cdots & W_{nn} \end{bmatrix}$$

onde  $n$  representa o número de locais (observações). A entrada na  $i$ -ésima linha e  $j$ -ésima coluna, denotado como  $W_{ij}$ , corresponde ao par  $(i, j)$  de localizações. Os elementos da matriz  $W_{ij}$  assumem valores diferentes de zero quando as áreas (observações)  $i$  e  $j$  são considerados vizinhas, e zero caso contrário (DUNCAN et al., 2014).

### 2.2.3 Autocorrelação Espacial

A autocorrelação espacial (associação) é a correlação entre as observações de uma única variável estritamente atribuível à proximidade destas observações no espaço geográfico. Esta definição é resumida pela primeira lei da geografia descrita por Tobler, que afirma que "*todas as coisas são parecidas, mas coisas mais próximas se parecem mais que coisas mais distantes*" (TOBLER, 1970). Atualmente, uma série de medidas de autocorrelação espacial estão disponíveis (BIVAND; PIRAS, 2015).

As medidas de autocorrelação espacial lidam com a covariância ou a correlação entre as observações vizinhas de uma determinada variável, por exemplo a autocorrelação espacial do número de indivíduos contaminados com os vírus transmitidos pelo mosquito

*Aedes aegypti* nos municípios do Estado da Paraíba. Sendo assim, compara-se dois tipos de informação: semelhança de observações (valor de similaridade) e similaridade entre locais (FISCHER; WANG, 2011).

As medidas e testes de autocorrelação espacial (associação) podem ser diferenciadas pelo alcance ou escala de análise. Geralmente distingue-se entre as medidas globais e locais (CHEN, 2013). Global implica que todos os elementos na matriz  $\mathbf{W}$  estão incluídas no cálculo da correlação espacial, produzindo um valor de autocorrelação espacial para qualquer matriz de pesos espaciais. Em contrapartida, as medidas locais estão concentradas. Isto é, avaliam a autocorrelação espacial associada com uma ou algumas unidades de área particulares. Ambas indicam o grau de associação espacial do conjunto de dados. O índice  $I$  de *Moran* calcula a autocorrelação espacial como uma covariância, a partir do produto dos desvios em relação à média. Esse índice indica a magnitude da associação espacial presente no conjunto de dados com  $n$  localizações, e a estatística  $c$  de *Geary* utiliza o quadrado dessas diferenças (FISCHER; WANG, 2011).

Duas medidas que têm sido amplamente utilizadas para dados de área e variáveis em escala intervalar são o índice  $I$  de *Moran* e a Estatística  $c$  de *Geary*. O índice  $I$  de *Moran* é dado pela seguinte expressão

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

A estatística  $c$  de *Geary* é calculada como

$$c = \frac{(n-1)}{2 \sum_{i=1}^n \sum_{j \neq i}^n w_{ij}} \times \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^n (z_i - \bar{z})^2}.$$

O Índice  $I$  de *Moran* varia em um intervalo de  $(-1,1)$ , onde  $-1$  significa dispersão perfeita,  $0$  representa um comportamento aleatório e  $1$  significa associação perfeita. Já a estatística  $c$  de *Geary* varia em intervalo de  $(0,2)$ , onde  $0$  significa associação perfeita,  $1$  representa um comportamento aleatório e  $2$  significa dispersão perfeita.

Assumindo-se que  $z_i$  são observações de variáveis aleatórias  $Z_i$  cuja distribuição é normal, então,  $I$  possui distribuição apropriadamente normal com os momentos

$$E(I) = -\frac{1}{(n-1)}$$

$$Var(I) = \frac{n^2(n-1)W_1 - n(n-1)W_2 - 2W_0^2}{(n+1)(n-1)^2W_0^2}$$

Ainda que essas medidas globais sejam muito úteis para fornecer uma indicação de agrupamento global dos dados, esta medida precisa ser complementada por estatísticas locais. A fórmula para o cálculo do índice de *Moran* local para cada área  $A_i$  é dada por:

$$I_i = \frac{(y_i - \bar{y}) \sum_{i=1}^n \sum_{j=1}^n W_{ij} (y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2 / n}.$$

A estatística pode ser interpretada da seguinte forma: valores positivos de  $I_i$  significam que existem *clusters* espaciais com valores similares (alto ou baixo) da variável em estudo, valores negativos significam que existem *clusters* espaciais com valores dissimilares da variável entre as áreas e seus vizinhos.

### 2.2.4 Média Móvel

Para analisar a variação de tendência espacial da taxa de pessoas infectadas pelo *Aedes aegypti* no Estado da Paraíba, pode-se utilizar a média móvel local para cada município em estudo da seguinte forma

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

onde  $i = 1, \dots, n$  e  $j = 1, \dots, n$  representam, respectivamente, os elementos da matriz de proximidade espacial padronizada de primeira ordem e o valor observado da variável taxa, por município, de pessoas infectadas pelo *Aedes aegypti* no Estado da Paraíba.

### 2.2.5 Gráfico de Espalhamento de Moran

Para observar graficamente as relações entre os desvios  $\hat{\alpha}_i = (Y_i - \bar{Y})$  do atributo numa área  $A_i$  com média móvel  $\hat{\mu}_i$  podem ser construídos gráficos, denominados de espalhamento de Moran, o que permite uma comparação visual entre os valores do atributo e de seus vizinhos de maneira quase imediata. Dessa forma, dividem-se o diagrama em quatro quadrantes (Q1, Q2, Q3 e Q4). Os quadrantes são interpretados dessa maneira:

- Q1: Primeiro quadrante (Localizado na parte superior direita) mostra as áreas que apresentam altos valores para variável em análise cercada por áreas vizinhas que também apresentam valores acima da média para a variável em análise. Esse quadrante é classificado como alto-alto (AA,+ +).
- Q2: Segundo quadrante (Localizado na parte inferior esquerda) mostra as áreas que apresentam baixos valores para variável em análise cercada por áreas vizinhas que também apresentam valores abaixo da média para variável em análise. Esse quadrante é classificado como baixo-baixo (BB,- -).
- Q3: Terceiro quadrante (Localizado na parte inferior direita) mostra as áreas que apresentam altos valores para variável em análise cercada por áreas vizinhas que apresentam valores abaixo da média para variável em análise. Esse quadrante é classificado como alto-baixo (AB,+ -).
- Q4: Quarto quadrante (Localizado na parte superior esquerda) mostra as áreas que apresentam baixos valores para variável em análise cercada por áreas que apresentam

valores acima da média para a variável em análise. Esse quadrante é classificado como baixo-alto (BA,- +).

Os municípios localizados nos quadrantes Q1 e Q2 apresentam autocorrelação local positiva, ou seja, os municípios formam *clusters*, valores similares. Já os municípios localizados nos quadrantes Q3 e Q4 apresentam autocorrelação local negativa, ou seja, há dissimilaridade entre os municípios vizinhos.

### 2.2.6 Box Map, LISA Map e Moran Map

- *Box Map*: este mapa é uma extensão do gráfico de espalhamento de *Moran*, onde os elementos de cada quadrante do gráfico são representados por uma cor específica com os seus respectivos polígonos;
- *LISA Map*: indica as regiões que apresentam correlação local significativamente diferente das demais, onde são classificados nos seguintes grupos, não significativos, significativo a 5%, 1% e 0,1% ;
- *Moran Map*: de forma semelhante ao *LISA Map*, somente os valores significativos são apresentados, porém classificados em quatro grupos, conforme os quadrantes a que pertencem no gráfico, além dos não significativos.

### 2.2.7 Regressão Clássica

Um modelo de regressão baseia-se no relacionamento entre duas ou mais variáveis de forma que uma delas possa ser explicada ou ter seu valor predito por meio de outras variáveis (AZEVEDO et al., 2015). No caso de dados espaciais, havendo autocorrelação espacial, o modelo gerado deve incorporar a estrutura espacial, já que a dependência entre as observações afeta a capacidade de explicação do modelo (CÂMARA; MONTEIRO, 2001).

Para cada observação  $i$  (área: municípios do Estado da Paraíba), com  $i = 1, \dots, n$ , mantém-se a seguinte relação

$$Y_i = \sum_{q=1}^Q X_{iq}\beta_q + \varepsilon_i, \quad (2.1)$$

em que  $Y_i$  é uma observação da variável dependente,  $X_{iq}$  é uma observação sobre uma variável explicativa (variável independente), com  $q = 1, \dots, Q$ ,  $\beta_q$  os coeficientes de regressão correspondentes, e  $\varepsilon_i$  é o erro associado as observações da variável dependente. Na especificação da regressão clássica, os termos do erro possuem média zero, isto é,  $E[\varepsilon_i] = 0$  para todo  $i$ , e são independentes e identicamente distribuídos(*i.i.d.*). Por isso a sua variância é constante,  $Var[\varepsilon_i] = \sigma^2$ , para todo  $i$ , e eles são não correlacionados,  $E[\varepsilon_i\varepsilon_j] = E[\varepsilon_i]E[\varepsilon_j]$  para  $i \neq j$  (RIBEIRO; SOUSA; PEREIRA, 2016).

A suposição de que as observações sejam independentes simplifica o modelo, mas no contexto de dados de área esta simplificação é pouco provável que aconteça, por causa da possibilidade de dependência espacial entre os termos do erro. Uma alternativa é utilizar o modelo espacial autoregressivo misto (*Spatial AutoRegressivo* - SAR ou *Spatial Lag Model*), que atribui à variável resposta  $y_i$  a autocorrelação espacial ignorada. Outra alternativa é pelo modelo do erro espacial (*Spatial Error Model* ou *Conditional AutoRegressive* - CAR), que considera os efeitos espaciais como um ruído, isto é, como um fator a ser removido (FISCHER; LESAGE, 2010).

### 2.2.8 Modelo Espacial Autorregressivo Misto

O modelo espacial autorregressivo misto, permite que as observações da variável dependente  $y_i$  na área  $i$  ( $i = 1, \dots, n$ ) dependam de observações em áreas vizinhas com  $j \neq i$ , assumindo a forma

$$y_i = \rho \sum_{j=1}^n W_{ij} y_j + \sum_{q=1}^Q X_{iq} \beta_q + \varepsilon_i, \quad (2.2)$$

em que  $\varepsilon_i$  é o erro, sendo *i.i.d.*,  $W_{ij}$  é o  $(i, j)$ -ésimo elemento da matriz espacial de ordem  $n$  (isto é,  $n$  por  $n$ ). O escalar  $\rho$  definido na Equação (2.2) é um parâmetro (a ser estimado) que vai determinar a intensidade da relação espacial autorregressiva entre  $y_i$  e  $\sum_j W_{ij} y_j$ , esse parâmetro possui como interpretação o efeito médio da variável dependente relativo à vizinhança espacial na região em questão. O vetor  $\mathbf{W}\mathbf{y}$  é conhecido como *lag* espacial, a matriz  $\mathbf{X}$  contém as observações das variáveis independentes e o vetor  $\boldsymbol{\beta}$  possui coeficientes para as variáveis independentes.

### 2.2.9 Modelo de Erro Espacial

Outra forma de dependência espacial ocorre quando a dependência é obtida através do processo de erros, em que os erros das diferentes áreas podem apresentar covariância espacial (BIVAND; PIRAS, 2015), determinado da seguinte forma

$$\varepsilon_i = \rho \sum_{j=1}^n W_{ij} \varepsilon_j + u_i, \quad (2.3)$$

em que  $\rho$  é o parâmetro auto-regressivo que indica a intensidade da autocorrelação espacial entre os resíduos da equação observada, esse parâmetro mensura o efeito médio dos erros dos vizinhos em relação ao resíduo da região em questão e  $u_i$  é o termo do erro aleatório, tipicamente assumindo ser *i.i.d.* A autocorrelação espacial nesses modelos de erro espacial aparece nos termos de erro. Em notação matricial a Equação (2.3) pode ser escrita como

$$\boldsymbol{\varepsilon} = \rho \mathbf{W}\boldsymbol{\varepsilon} + \mathbf{u}. \quad (2.4)$$

Assumindo  $|\rho| < 1$  e resolvendo a Equação (2.4) para  $\boldsymbol{\varepsilon}$  chega-se ao seguinte resultado

$$\boldsymbol{\varepsilon} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{u}. \quad (2.5)$$

Inserindo a Equação (2.5) no modelo de regressão (2.1) chega-se ao seguinte modelo de erro espacial

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{u} \quad (2.6)$$

com  $E[\mathbf{u}\mathbf{u}'] = \sigma^2\mathbf{I}$  sendo que a matriz de variâncias e covariâncias é definida da seguinte forma

$$E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \sigma^2(\mathbf{I} - \rho\mathbf{W})^{-1}(\mathbf{I} - \rho\mathbf{W}')^{-1} \quad (2.7)$$

O Modelo de Erro Espacial (MEE) pode ser visto como uma combinação do modelo de regressão padrão com o modelo autorregressivo espacial em termos do erro  $\boldsymbol{\varepsilon}$ , e, portanto, seu valor esperado é igual ao modelo de regressão padrão. Para grandes amostras, as estimativas pontuais do vetor de parâmetros  $\boldsymbol{\beta}$ , a partir do MEE é semelhante ao modelo de regressão padrão, mas em pequenas amostras pode haver um ganho de eficiência ao modelar corretamente a dependência espacial em termos do erro (BIVAND; PIRAS, 2015).

### 2.2.10 Modelo Espacial de *Durbin*

Uma extensão dos modelos SAR e SEM é o modelo espacial de *Durbin* onde é incluído o “*spatial lag*” na variável dependente  $Y_i$  e também na variável independente.

$$Y_i = \rho \sum_{j=1}^n W_{ij}y_j + \sum_{q=1}^Q X_{iq}\beta_q + \sum_{j=1}^n W_{ij}X_{iq}\theta_q + \varepsilon_i, \quad (2.8)$$

sendo que  $\mathbf{y}$  contém um vetor de variáveis dependentes e  $\mathbf{X}$  representa a matriz de dados contendo as variáveis independentes com um vetor de parâmetro associado  $\boldsymbol{\beta}$ .  $\mathbf{W}$  é a matriz de peso espacial e o parâmetro  $\rho$  é um coeficiente de “*spatial lag*” da variável resposta. Um conjunto adicional de variáveis explicativas é adicionado ao modelo construindo um intervalo espacial das variáveis explicativas usando o produto da matriz  $\mathbf{W}\mathbf{X}$ , com os parâmetros associados  $\boldsymbol{\theta}$ . Este conjunto de variáveis representa variáveis explicativas construídas como médias de observações vizinhas.

### 2.2.11 Estimação dos Parâmetros em Modelos de Regressão Espacial

A estimação dos modelos de regressão espacial é tipicamente realizada por meio do método da máxima verossimilhança, na qual a probabilidade conjunta da distribuição de todas as observações é maximizada no que diz respeito a uma série de parâmetros relevantes.

A máxima verossimilhança tem propriedades teóricas assintóticas desejáveis, tais como consistência, eficiência e normalidade assintótica, e também é considerada robusta para pequenos desvios da suposição de normalidade (KELEJIAN; PRUCHA, 2010).

Dada uma distribuição de probabilidade cujos parâmetros são desconhecidos, tem-se como objetivo, a partir de uma amostra de uma determinada população, estimar o valor do parâmetro da distribuição de probabilidade que representa toda a população (OLINDA et al., 2014). Assim, considerando  $\mathbf{Y} = (Y_1, \dots, Y_n)$  uma amostra aleatória *i.i.d.* com distribuição de probabilidade  $p(\mathbf{y}; \boldsymbol{\theta})$  em que  $\boldsymbol{\theta}$  é o vetor de parâmetros a serem estimados, deseja-se obter uma estimativa do valor de  $\boldsymbol{\theta}$  ou de uma função de  $\tau(\boldsymbol{\theta})$  a partir dos valores observados  $y_1, \dots, y_n$ . O método da máxima verossimilhança, por ser um método de estimação pontual, nos fornecerá um valor numérico de uma estatística  $T(Y)$  para estimar  $\tau(\boldsymbol{\theta})$ .

Dada uma amostra  $\mathbf{y} = (y_1, \dots, y_n)$ , o método de máxima verossimilhança consiste em maximizar a função de verossimilhança definida por  $L(\boldsymbol{\theta}; \mathbf{y}) = p(\mathbf{y}; \boldsymbol{\theta}) = p(y_1; \boldsymbol{\theta}) \times p(y_2; \boldsymbol{\theta}) \times \dots \times p(y_n; \boldsymbol{\theta})$ ,  $\boldsymbol{\theta} \in \Theta$  em que  $\Theta$  representa o espaço paramétrico de  $\boldsymbol{\theta}$ , isto é, o método consiste em encontrar  $\hat{\boldsymbol{\theta}}$  (estimador do vetor de parâmetros  $\boldsymbol{\theta}$ ) tal que  $\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}; \mathbf{y})$  em que  $\hat{\boldsymbol{\theta}}$  é o vetor das estimativas de  $\boldsymbol{\theta}$  e  $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(Y)$  é o vetor dos estimadores de máxima verossimilhança de  $\boldsymbol{\theta}$ .

Como o valor de  $\boldsymbol{\theta}$  que maximiza  $L(\boldsymbol{\theta}; \mathbf{y})$  é o mesmo que maximiza a função logarítmica, ou seja,  $\ell_p(\boldsymbol{\theta}; \mathbf{y}) = \ln(L(\boldsymbol{\theta}; \mathbf{y}))$  e  $\ell_p(\boldsymbol{\theta}; \mathbf{y})$  sempre está definida pois  $L(\boldsymbol{\theta}; \mathbf{y})$  é positiva, então, por simplicidade, maximiza-se  $\ell_p(\boldsymbol{\theta}; \mathbf{y})$ , também chamada de *função suporte* (OLINDA et al., 2013).

O modelo espacial autorregressivo misto contém, como parâmetro desconhecido, o coeficiente  $\rho$  e  $\sigma^2$ , para estimação não utiliza-se o estimador de mínimos quadrados ordinários, pois o vetor de *lags* espaciais  $\mathbf{W}\mathbf{y}$  é correlacionado com o vetor de resíduos  $\boldsymbol{\varepsilon}$ , portanto o estimador é inconsistente. Utiliza-se então a máxima verossimilhança, a partir da hipótese de que o vetor de resíduos  $\boldsymbol{\varepsilon}$  possui distribuição multivariada com média zero e covariância  $\sigma^2\mathbf{I}$ . Têm-se então

$$\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}, \quad (2.9)$$

o vetor de variáveis observadas  $\mathbf{y}$  possui distribuição normal multivariada, com média condicional

$$E[\mathbf{y}|\mathbf{X}] = (\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}, \quad (2.10)$$

e matriz de variância e covariância condicional

$$\Sigma_{\mathbf{y}|\mathbf{X}} = \sigma^2(\mathbf{I} - \rho\mathbf{W})^{-1}[(\mathbf{I} - \rho\mathbf{W})^{-1}]^T. \quad (2.11)$$

Obtém-se a função de log-verossimilhança condicional  $\log L(\rho, \boldsymbol{\beta}, \sigma^2)$  da distribuição  $\mathbf{y}$ . A função de log-verossimilhança é maximizada em relação aos parâmetros do modelo, encontram-se as estimativas para os coeficientes e para a variância dos resíduos.

Igualmente aos modelos espacial autorregressivo misto a estimação dos parâmetros desconhecidos  $\rho$  e  $\sigma^2$  do MEE utiliza-se o método de máxima verossimilhança, têm-se

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\varepsilon}, \quad (2.12)$$

onde  $\boldsymbol{\varepsilon}$  possui distribuição multivariada com média nula e covariância  $\sigma^2\mathbf{I}$ . O vetor de variável dependente  $\mathbf{y}$  possui distribuição normal multivariada com média condicional

$$E[\mathbf{y}|\mathbf{X}] = \mathbf{X}\boldsymbol{\beta}, \quad (2.13)$$

e matriz de variância e covariância condicional

$$\Sigma_{\mathbf{y}|\mathbf{X}} = \sigma^2(\mathbf{I} - \rho\mathbf{W})^{-1}[(\mathbf{I} - \rho\mathbf{W})^{-1}]^T. \quad (2.14)$$

Obtém-se a função de log-verossimilhança condicional  $\log L(\rho, \boldsymbol{\beta}, \sigma^2)$  da distribuição  $\mathbf{y}$ . A função de log-verossimilhança é maximizada em relação aos parâmetros do modelo, encontram-se as estimativas para os coeficientes e para a variância dos resíduos.

O modelo espacial de *Durbin* a estimação também foi feita pelo método de máxima verossimilhança, onde log-verossimilhança tem a forma:

$$\ln L = -\frac{n}{2}\ln(\pi\sigma^2) + \ln|I_n - \rho\mathbf{W}| - \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{2\sigma^2} \quad (2.15)$$

$$\boldsymbol{\varepsilon} = \mathbf{y} - \rho\mathbf{W}\mathbf{y} - \mathbf{Z}\boldsymbol{\delta} \quad (2.16)$$

onde  $\boldsymbol{\varepsilon}$  possui distribuição multivariada com média nula e covariância  $\sigma^2\mathbf{I}$ , sendo  $\mathbf{Z} = [\mathbf{X} \quad \mathbf{W}\mathbf{X}]$ ,  $\boldsymbol{\delta} = [\boldsymbol{\beta} \quad \boldsymbol{\theta}]$ .

### 2.2.12 Seleção de Modelos

Em geral, o algoritmo de ajuste deve ser aplicado não a um modelo espacial isolado, mas a vários modelos de um conjunto bem amplo que deve ser, realmente, relevante para a natureza das observações que se pretende analisar (DOBSON; BARNETT, 2011). Se o processo é aplicado a um único modelo, não levando em consideração possíveis modelos alternativos, existe o risco de não se obter um dos modelos mais adequados aos dados.

Similarmente aos modelos de regressão usual, pelo critério de Akaike, deve-se encontrar um submodelo para o qual a quantidade abaixo seja minimizada

$$AIC = D^*(\mathbf{y}; \hat{\boldsymbol{\mu}}) + 2p, \quad (2.17)$$

em que  $D^*(\mathbf{y}; \hat{\boldsymbol{\mu}})$  é a distância entre o logaritmo da função de verossimilhança do modelo saturado (com  $q$  parâmetros) e do modelo sob investigação (com  $p$  parâmetros) avaliado na estimativa de máxima verossimilhança  $\hat{\boldsymbol{\beta}}$ . Um valor pequeno para a função desvio indica que, para um número menor de parâmetros, obtêm-se um ajuste tão bom quanto o ajuste com o modelo saturado. O procedimento *stepwise* é utilizado para a seleção de um subconjunto de variáveis independentes para serem incorporados ao modelo.

As análises foram realizadas por meio do software estatístico *R*, versão 3.3.2 (<http://www.r-project.org/>), com auxílio dos pacotes *geoR*, *POT*, *spatialExtremes*, *spatial* e *spdep*.

### 3 Resultados e Discussão

Inicialmente elaborou-se uma tabela dos municípios com maior número de casos notificados por dengue na última década, por meio da Tabela 1 observou-se que a capital João Pessoa é o município que tem a maior incidência, com exceção no ano de 2008 onde a cidade de Patos apresentou a maior incidência no ano, a alta densidade populacional da capital pode ser um dos motivos para a disseminação da dengue.

Tabela 1 – Tabela dos municípios com maior número de casos de notificações por dengue no Estado da Paraíba.

Ano	Notificação no Estado	Maior Incidência	Notificação no Município
2007	12,868	João Pessoa	3,584
2008	11,262	Patos	1,032
2009	1,596	João Pessoa	267
2010	8,712	João Pessoa	1,348
2011	16,264	João Pessoa	4,538
2012	11,482	João Pessoa	4,620
2013	18,441	João Pessoa	3,323
2014	7,552	João Pessoa	2,377
2015	29,880	João Pessoa	4,880
2016	44,464	João Pessoa	6,935

Na Figura 1 optou-se por dicotomizar a variável dependente em ausência ou presença de casos notificados por dengue por ano. Observou-se que nos últimos 4 anos que quase todos os municípios tiveram a presença de pelo menos uma notificação de dengue.

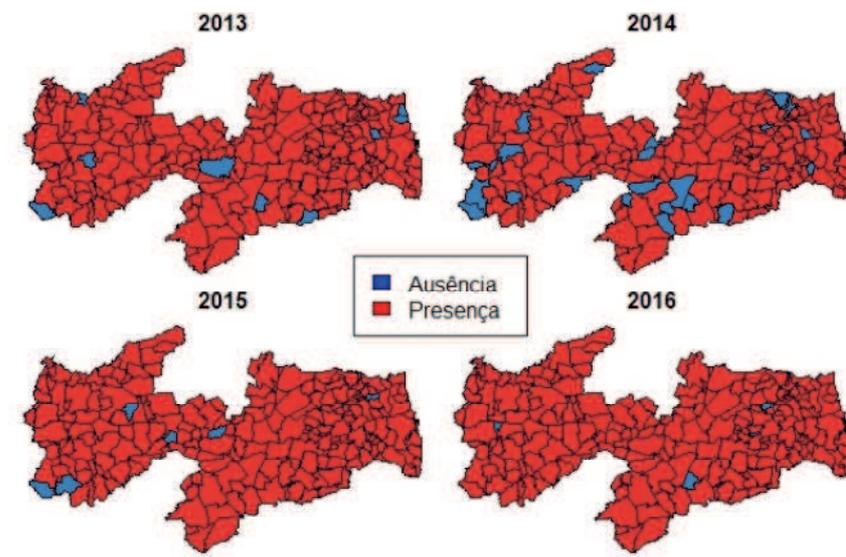


Figura 1 – Municípios da Paraíba com casos de notificações por dengue.

Na análise com métodos descritivos de visualização da variável dependente, como primeira análise exploratória foi construído o mapa de intervalos iguais (Figura 2) e observou-se inicialmente que cinco cidades (Carrapateira, Monte Horebe, Monteiro, Princesa Isabel, Zabelê), representadas pela cor vermelha no mapa, apresentaram alta proporção de casos notificados por dengue entre  $[0,17096; 0,21360]$ .

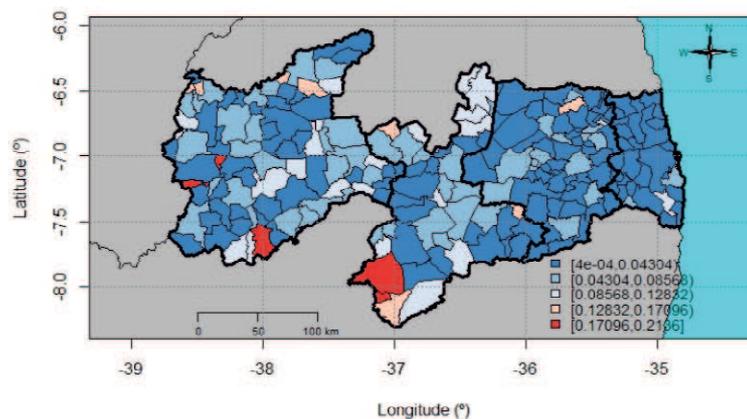


Figura 2 – Mapa de intervalos iguais para proporção de casos notificados por dengue no Estado da Paraíba.

Na Figura 3 construímos o mapa de quintis onde as classes foram separadas em percentuais, observou-se que 80% dos municípios, representados pelas cores azuis e bege, possuem baixa proporções de casos notificados por dengue entre  $[0;0,07558]$ , enquanto que os municípios representados pela cor vermelha apresenta alta proporção de casos notificados por dengue entre  $[0,07558;0,21360]$ .

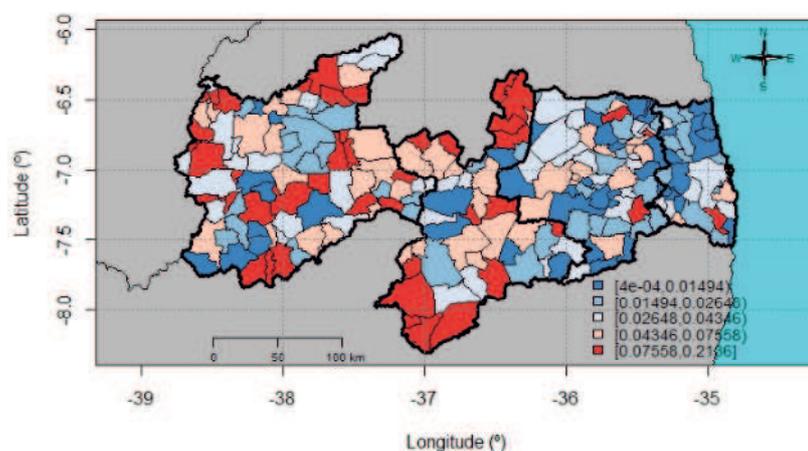


Figura 3 – Mapa de quintis para proporção de casos notificados por dengue no Estado da Paraíba.

Por meio da Figura 4 observou-se que no mapa foram geradas seis classes, onde cada classe foi obtida através do valor do desvio para cada município, observamos que os municípios que obtiveram os desvios padrão mais altos, representados pela cor vermelha no mapa, foram os mesmos que apresentaram alta proporção de casos notificados por dengue na Figura 2.

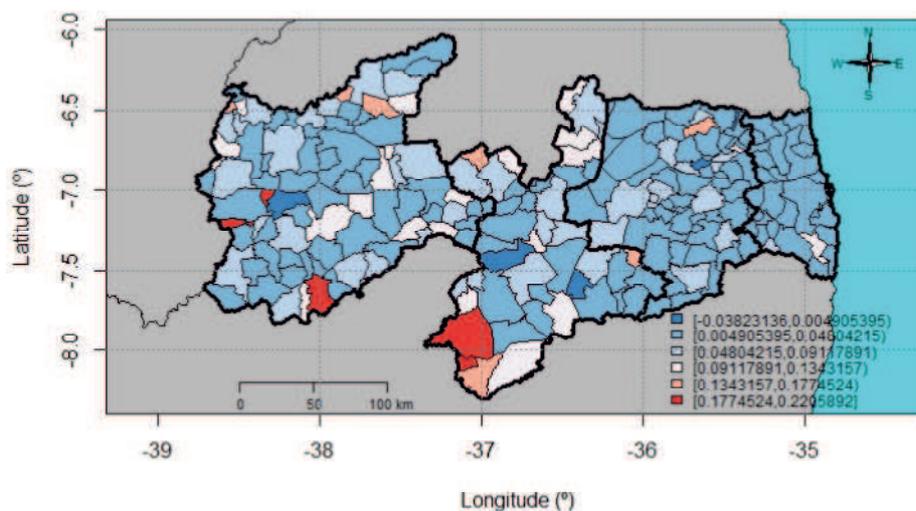


Figura 4 – Mapa de desvios padrão para proporção de casos notificados por dengue no Estado da Paraíba.

Em um estudo realizado no estado do Pará para verificar a existência de dependência espacial entre notificações por dengue nos municípios do estado, utilizando a metodologia de dados de área Santos et al. (2013) constataram que as notificações por dengue no estado do Pará é espacialmente dependente com um valor de Índice de *Moran* significativo ao nível de 5%. Em nossos resultados obtivemos um valor do Índice de *Moran* significativo à 0,1% para a dependência espacial entre os municípios do Estado da Paraíba, indicando que há uma dependência espacial na proporção dos casos notificados por dengue nos municípios do estado. Pode-se observar, por meio da Tabela 2, a existência de autocorrelação espacial positiva ( $p < 0,01$ ), ou seja, existe uma dependência na propagação do mosquito *Aedes Aegypti* entre municípios vizinhos de Estado da Paraíba.

Tabela 2 – Tabela das estimativas do Índice de *Moran* e a Estatística *c* de *Geary*.

Estatísticas	Estimativas	Valor de p
<i>Moran</i>	0,20510	< 0,001
<i>Geary</i>	0.73434	< 0,001

Sabe-se que existe autocorrelação espacial pela Tabela 2, entretanto quando se têm muitas áreas em estudo o ideal é verificar o quanto essas áreas vizinhas são similares,

por isso geramos o mapa do índice de *Moran* local (Figura 5 ). Valores positivos de  $I_i$  representado pela cor vermelha significa que existe *clusters* espaciais com valores similares, valores negativos de  $I_i$  representado pela cor azul significa que existe *clusters* espaciais dissimilares entre os municípios vizinhos. Observa-se que no mapa há uma prevalência da cor vermelha no litoral, indicando que grande parte dos municípios do litoral têm valores similares de casos notificados por dengue.

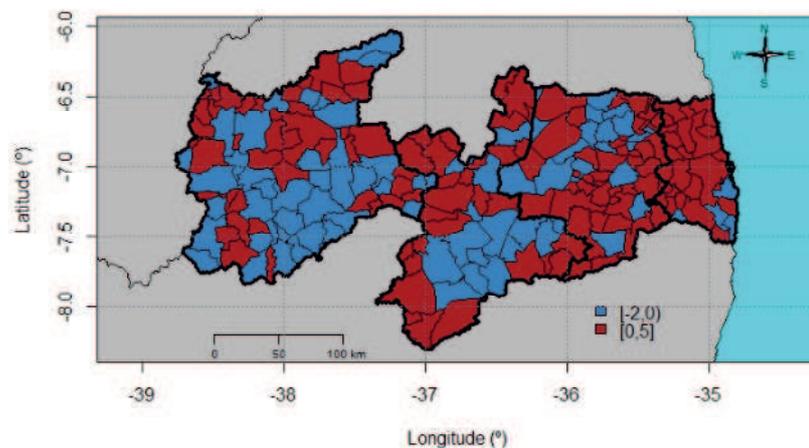


Figura 5 – Índice  $I_i$  de *Moran* Local da proporção de casos notificados por dengue no Estado da Paraíba.

Na Figura 6 construímos um mapa da média móvel para analisar a tendência espacial da taxa de pessoas infectadas pelo *Aedes aegypti*. Verifica-se que a medida que aproxima-se do interior do estado a média móvel aumenta, indicando existir uma tendência espacial mais fortemente observada nesta localidade.

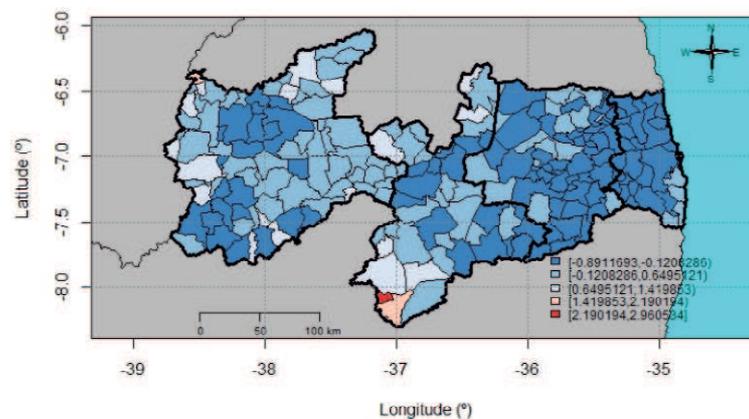


Figura 6 – Mapa da média móvel para proporção de casos notificados por dengue no Estado da Paraíba.

Observou-se no gráfico de espalhamento de *Moran* (Figura 7), que municípios localizados na parte superior a direita (primeiro quadrante) e na parte inferior a esquerda (segundo quadrante) apresentaram autocorrelação espacial local positiva, ou seja, os municípios formam *clusters* de valores similares entre os municípios vizinhos. Esses municípios (Bernardinho Batista, Brejo dos Santos, Riacho dos Cavalos, Monte Horebe e Várzea) do primeiro quadrante que se destacam no gráfico são preocupantes, pois além de apresentarem alta proporção de casos notificados por dengue eles também estão cercados por municípios que apresentam valores acima da média de proporção de casos notificados por dengue. A reta que passa pelo o primeiro e segundo quadrante corresponde ao coeficiente angular da regressão simples, esse valor é igual ao índice de *Moran* global.

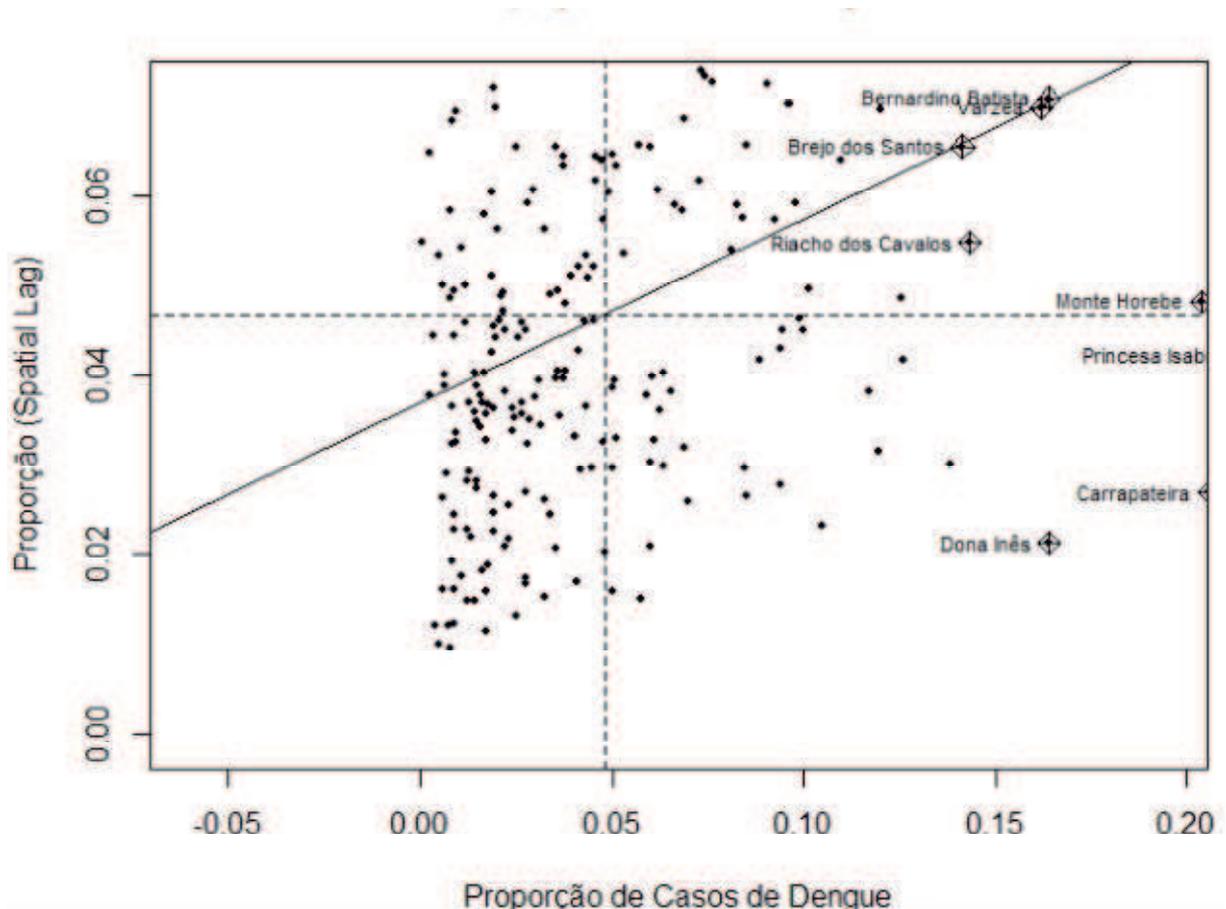


Figura 7 – Mapa do espalhamento de *Moran* para proporção de casos notificados por dengue no Estado da Paraíba.

No Box Map (Figura 8), que é uma extensão do espalhamento de *Moran*, municípios representados por cores vermelha e azul apresentam autocorrelação espacial positiva. Por outro lado, os municípios representados por cores verde e amarela apresentam dissimilaridade entre os municípios vizinhos com relação a proporção dos casos notificados por dengue no Estado da Paraíba. Enxerga-se que, há uma prevalência de municípios com similaridade entre os valores observados para os casos notificados por dengue no Estado.

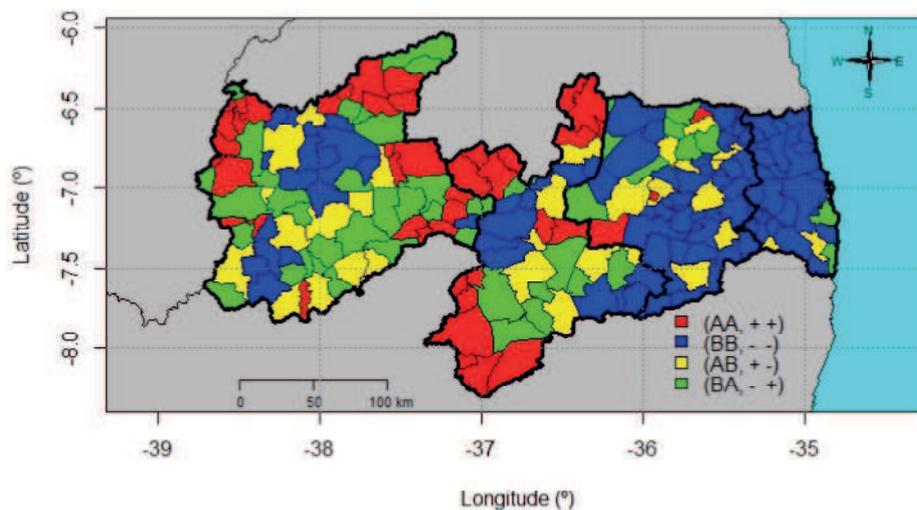


Figura 8 – *Box Map* da proporção de casos notificados por dengue no Estado da Paraíba.

Na Figura 9 tem-se o *LISA Map* onde é possível observar que os municípios que apresentam correlação local significativa ao nível de 0,1% estão representados pela cor vermelha. Nesses municípios a dependência entre a proporção de casos notificados por dengue observada em municípios vizinhos é ainda mais forte.

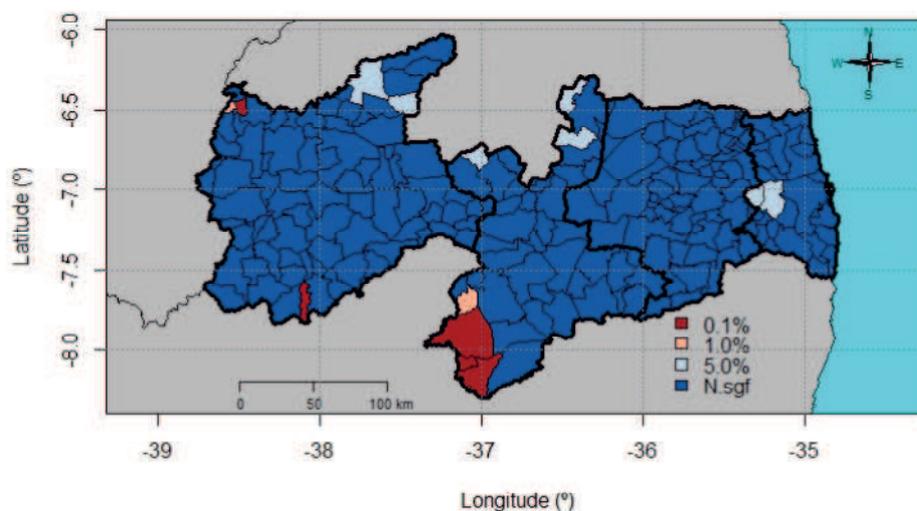


Figura 9 – *LISA Map* da proporção de casos notificados por dengue no Estado da Paraíba.

De maneira semelhante ao *LISA Map*, o *Moran Map* (Figura 10) apresenta os p-valores de cada município do índice de *Moran* local, diferenciando-se por apresentar os municípios que não foram significativos ao nível 10%. Nota-se que os municípios com

uma autocorrelação local mais significativo são encontrados na mesma região observada no *LISA Map*.

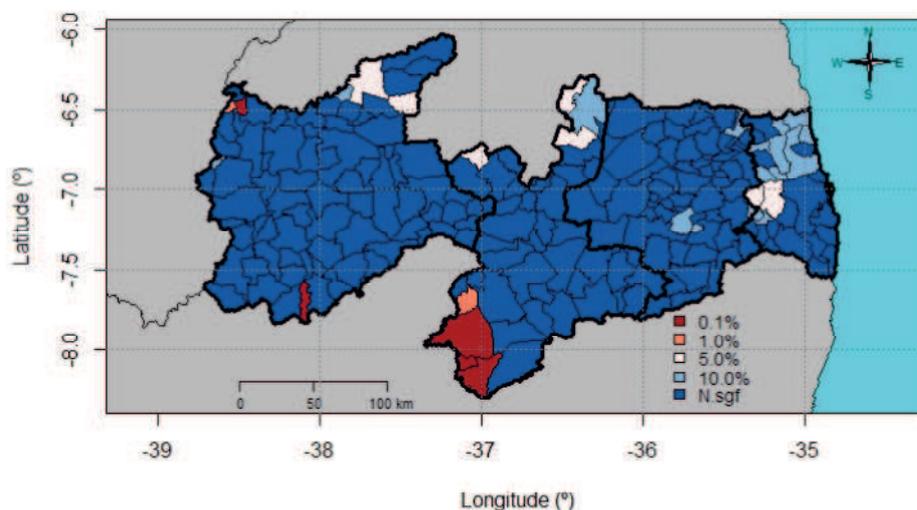


Figura 10 – *Moran Map* da proporção de casos notificados por dengue no Estado da Paraíba.

Outro artigo relacionado com o tema dengue analisou áreas representativas de bairros (MACHADO; OLIVEIRA; SOUZA-SANTOS, 2009), onde o objetivo do estudo era analisar a relação entre as condições de vida e a ocorrência dos casos notificados por dengue e investigar a possível relação entre as desigualdades sócio econômicas e variações temporais da incidência da doença por bairros do Município de Nova Iguaçu, Estado do Rio de Janeiro, Brasil, no período de 1996 a 2004, embora um modelo estatístico para explicar essa relação não tenha sido apresentado, os autores observaram um maior acometimento da dengue em áreas com desigualdades sociais localizadas próximas às vias de acesso. Analisando-se os resultados apresentados na Tabela 3, pode-se observar que o modelo que melhor se ajustou aos dados foi o modelo espacial auto-regressivo (MEAR), pois o mesmo apresentou o menor valor de AIC. Pode-se observar também a significância das covariáveis Índice de Desenvolvimento Humano (IDH) e o Índice de GINI na representação da proporção de casos notificados por dengue nos municípios do Estado da Paraíba. Sendo assim, observa-se a relação entre o acometimento por dengue e as covariáveis IDH e índice de GINI, demonstrando que os municípios com baixos valores desses índices são mais vulneráveis ao acometimento da dengue, direcionando os gestores para decisões que respeitem as particularidades de cada local.

Tabela 3: Tabela de coeficientes estimados e dos índices de *Moran* Residuais dos modelos de Regressão.

Covariáveis	RC(p-valor)	MEAR(p-valor)	MEE(p-valor)	MED(p-valor)
Intercepto	-0,096(0,032)	-0,091(0,034)	-0,090(0,048)	-0,125(0,120)
IDH	0,466(< 0, 01)	0,403(< 0, 01)	0,431(< 0, 01)	0,336(< 0, 01)
GINI	-0,297(< 0, 01)	-0,259(< 0, 01)	-0,256(< 0, 01)	-0,259(< 0, 01)
PRECIP	-0,00002(0,026)	-0,00001(0,099)	-0,00002(0,053)	-0,00004(0,081)
TEMP	-	-	-	-0,007(0,050)
AIC	-798,5	<b>-805,4</b>	-803,4	-804,8
$\rho$	-	0,278(0,002)	0,260(0,008)	0,262(0,005)
Resíduo teste	0,87(< 0, 001)	2,11(0,146)	-	0,30(0,583)

Em outro trabalho (SILVA; SILVA; CHAVES, 2014) o objetivo foi analisar a influência das variações das estações do ano na distribuição geográfica dos casos de dengue no município de João Pessoa - PB, os autores obtiveram em seus resultados que, a dengue teve um maior acometimento em regiões e meses cujo os resultados para a variável pluviométrica teve maior representatividade. No entanto, não foi exposto neste trabalho um modelo estatístico que considere a dependência espacial entre as áreas de notificação para verificar essa influência. Em nossos resultados trabalhamos com variáveis termo-pluviométricas e socioeconômicas, ajustando modelos estatísticos que consideram essa dependência em busca de uma possível influência dessas variáveis na proporção de casos notificados por dengue. Nos resultados encontrados as variáveis termo-pluviométricas não foram significativas ao nível de 5% de significância na estimação dos modelos, e apenas as variáveis socioeconômicas foram significativas. Indicando que essas variáveis estão diretamente ligadas a proporção de casos notificados por dengue.

No Rio Grande do Sul um estudo que concentrou-se na identificação dos elementos socioeconômicos e demográficos que explicar-se as taxas de incidência de dengue nos municípios, entre os anos de 2009 a 2015 obtiveram-se um índice de *Moran* global para cada ano (FAHL; ADAMCZYK, 2016). Um mapeamento de *clusters* por meio do índice local de *Moran* foi apresentado para cada ano também e por meio desse mapeamento os autores confirmaram a existência de dependência espacial. Nesse estudo no Rio Grande do Sul, modelos de regressão espacial foi estimado o MEAR e o MEE, mas por não apresentarem normalidade o modelo logit espacial foi o modelo que se ajustou aos dados, assim as variáveis IDH e GINI foram significativas a 1% e densidade populacional foi significativa a 5%. Portanto, assim como o Estado do Rio Grande do Sul, o Estado da Paraíba tem as variáveis explicativas socioeconômicas influenciando na propagação do mosquito *Aedes Aegypti*.

Abaixo são exibidas figuras com os mapas residuais para os modelos de regressão utilizados neste trabalho. A cor branca indica os municípios com o menor erro na estimativa

da proporção dos casos notificados por dengue no Estado da Paraíba. Enquanto que, as cores vermelho escuro e verde escuro apresentam os municípios com os maiores erros na estimativa desses valores.

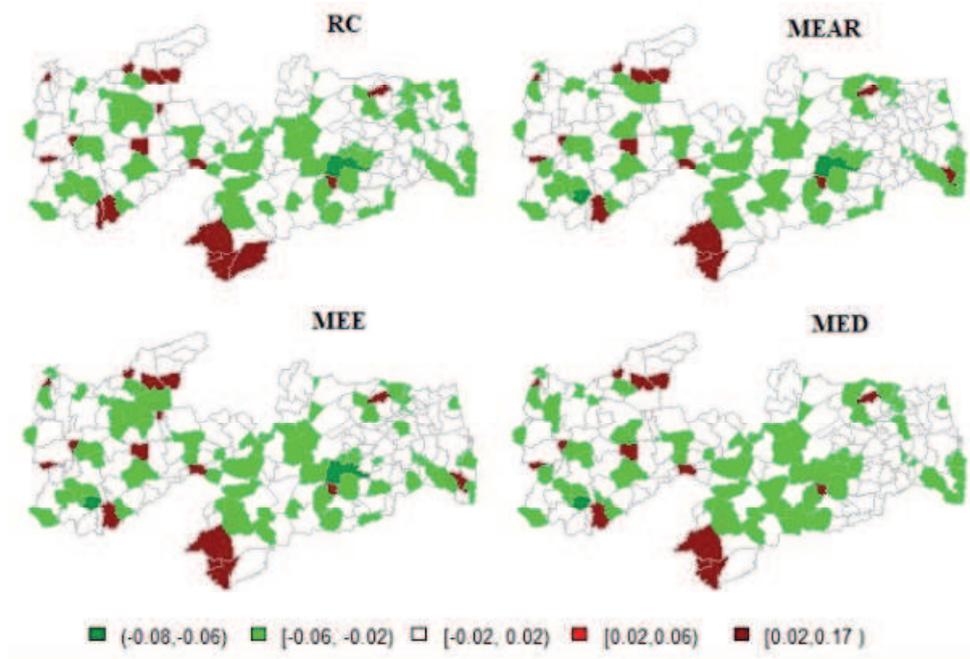


Figura 11 – Mapas de resíduos para os modelos de regressão clássica, espacial autorregressivo misto, erro espacial e espacial de *Durbin* para a proporção de casos notificados por dengue no Estado da Paraíba.

## 4 Conclusão

Diante do exposto, uma autocorrelação espacial positiva foi encontrada, indicando a dependência no acometimento dos casos de dengue nos municípios do Estado da Paraíba. Observamos também um padrão espacial em que se mostrou presente os mais altos números em relação a esta variável na região oeste do mapa da Paraíba, mostrando que a proporção de casos notificados por dengue é maior no sertão paraibano. Nos modelos de regressão estimados, verificamos que apenas as variáveis socioeconômicas foram significativas, e que o modelo que melhor se ajustou foi modelo autorregressivo misto, em que é levado em consideração a dependência espacial entre as variáveis.

Portanto, o estudo da variabilidade espacial e temporal da propagação do mosquito *Aedes aegypti*, em pequena e grande escala, torna-se de suma importância para os órgãos competentes, na gestão adequada de riscos minimizando as chances de transmissão da doença e para a sociedade em geral, pois será possível quantificar por meio de mapas os fatores que estão associados a propagação do mosquito, de modo que se possa avaliar o quanto esses fenômenos surtem impactos na Saúde Coletiva do Estado da Paraíba.

## Referências

- AZEVEDO, J. V. V. d.; SANTOS, C. A. C. d.; ALVES, T. L. B.; AZEVEDO, P. V. d.; OLINDA, R. A. d. Influence of climate on the incidence of acute respiratory infections in children at the districts of campina grande and monteiro, paraíba, brazil. *Revista Brasileira de Meteorologia*, SciELO Brasil, v. 30, n. 4, p. 467–477, 2015. Nenhuma citação no texto.
- BAILEY, T. C.; GATRELL, A. C. *Interactive spatial data analysis*. [S.l.]: Longman Scientific & Technical Essex, 1995. v. 413. Nenhuma citação no texto.
- BIVAND, R.; HAUKE, J.; KOSSOWSKI, T. Computing the jacobian in gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geographical Analysis*, v. 45, n. 2, p. 150–179, 2013. Disponível em: <<http://www.jstatsoft.org/v63/i18/>>. Nenhuma citação no texto.
- BIVAND, R.; PIRAS, G. Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, v. 63, n. 18, p. 1–36, 2015. Disponível em: <<http://www.jstatsoft.org/v63/i18/>>. Nenhuma citação no texto.
- CÂMARA, G.; MONTEIRO, A. M. V. Conceitos básicos em ciência da geoinformação. *Câmara G, Davis C, Monteiro AMV, organizadores. Introdução à ciência da geoinformação*. [acessado 2016 Fev 01]. Disponível em: [http://www.dpi.inpe.br/gilberto/livro/introd\[Links\]](http://www.dpi.inpe.br/gilberto/livro/introd[Links]), 2001. Nenhuma citação no texto.
- CHEN, Y. New approaches for calculating morans index of spatial autocorrelation. *PloS one*, Public Library of Science, v. 8, n. 7, p. e68336, 2013. Nenhuma citação no texto.
- DOBSON, A. J.; BARNETT, A. *An introduction to generalized linear models*. [S.l.]: CRC press, 2011. Nenhuma citação no texto.
- DUNCAN, D. T.; KAWACHI, I.; MELLY, S. J.; BLOSSOM, J.; SORENSEN, G.; WILLIAMS, D. R. Demographic disparities in the tobacco retail environment in boston: A citywide spatial analysis. *Public Health Reports*, Association of Schools of Public Health, v. 129, n. 2, p. 209, 2014. Nenhuma citação no texto.
- FAHL, G. B.; ADAMCZYK, W. B. Incidência de dengue no rio grande do sul: uma análise econométrica espacial. 2016. Nenhuma citação no texto.
- FISCHER, M. M.; LESAGE, J. Spatial econometric methods for modeling origin destination flows. *Handbook of applied spatial analysis*. Springer, Berlin/Heidelberg, p. 409–432, 2010. Nenhuma citação no texto.
- FISCHER, M. M.; WANG, J. *Spatial data analysis: models, methods and techniques*. [S.l.]: Springer Science & Business Media, 2011. 91 p. Nenhuma citação no texto.
- KELEJIAN, H. H.; PRUCHA, I. R. Spatial models with spatially lagged dependent variables and incomplete data. *Journal of geographical systems*, Springer, v. 12, n. 3, p. 241–257, 2010. Nenhuma citação no texto.

LANDIM, P. Análise de dados geológicos. *São Paulo: EDUNESP*, 1998. Nenhuma citação no texto.

MACHADO, J. P.; OLIVEIRA, R.; SOUZA-SANTOS, R. Análise espacial da ocorrência de dengue e condições de vida na cidade de nova iguaçu, estado do rio de janeiro, brasil spatial analysis of dengue occurrence and living conditions in nova iguaçu, rio de janeiro state, brazil. *Cadernos de Saúde Pública, Rio de Janeiro*, v. 25, n. 5, p. 1025–1034, 2009. Nenhuma citação no texto.

MARINO, R. A.; BESERRA, E. B.; BEZERRA-GUSMAO, M. A.; PORTO, V. S.; OLINDA, R. A.; SANTOS, C. A. C. Effects of temperature on the life cycle, expansion, and dispersion of aedes aegypti (diptera: Culicidae) in three cities in paraíba, brazil. In: . University of Idaho: *Journal of Vector Ecology*, 2016. (1, 1), p. 1–20. Nenhuma citação no texto.

OLINDA, R. A.; BLANCHET, J.; OZAKI, V. A.; JR, P. J. R. Spatial modeling applied to extreme precipitation data in the state of paraná. *Journal of Hydrology. No Prelo.*, 2014. Nenhuma citação no texto.

OLINDA, R. A.; OZAKI, V. A.; JULIETTE, B.; JUNIOR, P. J. R. Estudo de simulação de extremos espaciais com base em processos max-stable. *Revista Brasileira de Biometria, São Paulo*, v. 31, n. 1, p. 132–156, 2013. Nenhuma citação no texto.

RIBEIRO, M. C.; SOUSA, A. J.; PEREIRA, M. J. A coregionalization model can assist specification of geographically weighted poisson regression: application to an ecological study. *Spatial and Spatio-temporal Epidemiology*, Elsevier, 2016. Nenhuma citação no texto.

SANTOS, C. A. dos; MACEDO, M. R. A.; ROCHA, J. F. da; MIRANDA, R. S. Epidemiologia espacial para casos de dengue no estado do pará. 2013. Nenhuma citação no texto.

SILVA, R. M. da; SILVA, A. M.; CHAVES, J. J. S. Vulnerabilidade espacial da dengue e sua relação com a variabilidade termopluviométrica em João Pessoa-PB. *Hygeia: Revista Brasileira de Geografia Médica e da Saúde*, Associação Nacional de Pesquisa e Pós-Graduação em Geografia, Grupo de Trabalho de Geografia da Saúde, v. 10, n. 18, p. 177, 2014. Nenhuma citação no texto.

TOBLER, W. R. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, JSTOR, v. 46, p. 234–240, 1970. Nenhuma citação no texto.