

**Amanda Ribeiro de Souza**

**Análise espacial dos casos de dengue nos  
municípios do estado do Rio de Janeiro**

Niterói - RJ, Brasil

19 de dezembro de 2023

**Amanda Ribeiro de Souza**

**Análise espacial dos casos de dengue  
nos municípios do estado do Rio de  
Janeiro**

**Trabalho de Conclusão de Curso**

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

Orientador(a): Prof. Dr. Jony Arrais Pinto Junior

Niterói - RJ, Brasil

19 de dezembro de 2023

**Amanda Ribeiro de Souza**

**Análise espacial dos casos de dengue nos  
municípios do estado do Rio de Janeiro**

Monografia de Projeto Final de Graduação sob o título “*Análise espacial dos casos de dengue nos municípios do estado do Rio de Janeiro*”, defendida por Amanda Ribeiro de Souza e aprovada em 19 de dezembro de 2023, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

---

**Prof. Dr. Jony Arrais Pinto Junior**  
Departamento de Estatística – UFF

---

**Prof. Dr. Victor Eduardo Leite de Almeida Duca**  
Departamento de Estatística – UFF

---

**Prof. Dra. Márcia Marques de Carvalho**  
Departamento de Estatística – UFF

Niterói, 19 de dezembro de 2023

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

S719a Souza, Amanda Ribeiro de  
Análise espacial dos casos de dengue nos municípios do  
estado do Rio de Janeiro / Amanda Ribeiro de Souza. - 2023.  
41 f.: il.

Orientador: Jony Arrais Pinto Junior.  
Trabalho de Conclusão de Curso (graduação)-Universidade  
Federal Fluminense, Instituto de Matemática e Estatística,  
Niterói, 2023.

1. Dengue. 2. Inferência bayesiana. 3. Produção  
intelectual. I. Pinto Junior, Jony Arrais, orientador. II.  
Universidade Federal Fluminense. Instituto de Matemática e  
Estatística. III. Título.

CDD - XXX

# Resumo

O estudo de dados agregados por divisões espaciais fazem parte do dia a dia de pesquisadores das mais diversas áreas. Este trabalho tem como foco principal fazer uma análise espacial dos casos de dengue, doença viral transmitida aos seres humanos principalmente por mosquitos do gênero *Aedes*, sobretudo o *Aedes aegypti*. Esse estudo foi realizado nos municípios do estado do Rio de Janeiro com objetivo de avaliar o comportamento da doença, identificar as regiões que são mais acometidas e detectar fatores que estejam associados ao número de casos dessa doença. Para isso, utilizou-se um modelo de Poisson com efeitos espaciais estruturados e não estruturados, e a inferência foi realizada sob o enfoque Bayesiano. Foram testados diferentes distribuições *a priori* para os efeitos espaciais, tais como, distribuições normais independentes, o *CAR* Intrínseco e o Leroux, com diferentes covariáveis. O modelo mais adequado para a representação dos dados foi o *CAR* Intrínseco, considerando as covariáveis Índice de Vulnerabilidade Social, Taxa da população que vive em domicílios urbanos sem o serviço de coleta de lixo e Taxa de analfabetismo - 18 anos ou mais, escolhido através do critério Deviance Information Criterion (DIC). Após a análise deste modelo, foi concluído que os municípios mais afetados pela falta de coleta de lixo possuíam um maior número de casos de dengue.

Palavras-chave: Estatística espacial. Inferência Bayesiana. modelo Independente. *CAR* Intrínseco. Leroux. Deviance Information Criterion. Dengue.

# Agradecimentos

Gostaria de expressar minha imensa gratidão a todos que tornaram possível a realização deste trabalho. Em primeiro lugar, agradeço à Deus, sem Ele nada disso seria possível. À minha família, que sempre esteve ao meu lado, dando apoio incondicional e compreensão nos momentos desafiadores deste percurso acadêmico. Aos amigos e colegas de curso, obrigada por compartilharem experiências, trocarem ideias e proporcionarem um ambiente de aprendizado colaborativo.

Agradeço também ao meu orientador, Jony Arrais, pela orientação, paciência e dicas que foram fundamentais para o desenvolvimento deste TCC. Aos membros da banca, meu apreço pelo tempo disponibilizado e contribuições enriquecedoras durante a defesa.

Expresso meu reconhecimento a Universidade Federal Fluminense (UFF), pelos recursos disponibilizados e pelo ambiente de aprendizagem. Aos professores que, ao longo dessa jornada, compartilharam conhecimentos que foram fundamentais para a minha formação como pessoa e profissional e, que sem dúvidas, contribuíram também para a construção deste trabalho.

Por fim, agradeço a todos que, de alguma forma, contribuíram para o sucesso deste projeto. Este TCC representa não apenas meu esforço individual, mas o resultado de um esforço coletivo. Cada palavra escrita é um reflexo do incentivo e colaboração recebidos ao longo desta jornada. Obrigada pela confiança, apoio e inspiração.

# Sumário

Lista de Figuras

Lista de Tabelas

<b>1</b>	<b>Introdução</b>	p. 9
1.1	Objetivos . . . . .	p. 13
1.2	Organização . . . . .	p. 13
<b>2</b>	<b>Materiais e Métodos</b>	p. 14
2.1	Materiais . . . . .	p. 14
2.2	Estatística Espacial . . . . .	p. 14
2.3	Análise exploratória de dados de área . . . . .	p. 16
2.3.1	Matriz de Vizinhança . . . . .	p. 17
2.3.2	Autocorrelação Espacial . . . . .	p. 18
2.4	Modelo de Poisson . . . . .	p. 19
2.5	Inferência no Modelo de Poisson . . . . .	p. 21
2.5.1	Distribuição <i>a priori</i> dos efeitos espaciais . . . . .	p. 21
2.5.2	Modelos Especificados . . . . .	p. 23
2.6	Comparação de modelos . . . . .	p. 25
<b>3</b>	<b>Análise dos resultados</b>	p. 26
<b>4</b>	<b>Conclusão</b>	p. 33
	<b>Referências</b>	p. 36

Apêndice A - Cadeias do modelo Independente	p. 38
Apêndice A - Cadeias do modelo CAR Intrínseco	p. 39
Apêndice A - Cadeias do modelo Leroux	p. 40



# Lista de Figuras

1	Exemplo de mapa cloroplético . . . . .	p. 17
2	Mapa cloroplético dos casos de dengue no Estado do RJ . . . . .	p. 26
3	Mapa cloroplético das covariáveis selecionadas . . . . .	p. 27
4	Mapa municípios do estado do Rio de Janeiro . . . . .	p. 28
5	Gráfico de efeitos espaciais no modelo Independente . . . . .	p. 30
6	Gráfico de efeitos espaciais no modelo CAR Intrínseco . . . . .	p. 31
7	Gráfico de efeitos espaciais no modelo Leroux . . . . .	p. 32
8	Gráfico das cadeias - Modelo Independente . . . . .	p. 38
9	Gráfico das cadeias - Modelo CAR Intrínseco . . . . .	p. 39
10	Gráfico das cadeias - Modelo CAR Leroux . . . . .	p. 40

# Lista de Tabelas

1	Tabela das covariáveis. . . . .	p. 15
2	Análise do Índice de Moran para o número de casos de dengue no ano de 2021. . . . .	p. 28
3	Estimativa pontual e intervalar para os parâmetros do Modelo Independente. . . . .	p. 29
4	Estimativa pontual e intervalar para os parâmetros do modelo CAR Intrínseco. . . . .	p. 29
5	Estimativa pontual e intervalar para os parâmetros do modelo Leroux. . . . .	p. 29
6	Tabela DIC dos modelos ajustados. . . . .	p. 30

# 1 Introdução

A **dengue** é uma doença viral, transmitida principalmente pela picada da fêmea do mosquito *Aedes aegypti*, além disso, possui quatro sorotipos diferentes denominados por DENV-1, DENV-2, DENV-3 e DENV-4. Sua transmissão ocorre quando os ovos dos *Aedes aegypti* são postos em recipientes que possam haver acúmulo de água da chuva, tais como pneus, garrafas, caixas d'água, entre outros. Poucos minutos após entrar em contato com a água esses ovos eclodem dando origem a novos mosquitos.

A infecção por dengue pode ser assintomática, apresentar sinais leves ou até mesmo evoluir para um quadro mais grave. Em geral, a doença começa a se manifestar com uma febre alta ( $> 38^{\circ}\text{C}$ ) e de início abrupto acompanhada de dor de cabeça, dores no corpo e articulações, fraqueza, falta de apetite, vômitos, dor atrás dos olhos e manchas vermelhas na pele. Em sua forma mais grave, conhecida como dengue hemorrágica, os pacientes podem apresentar também dores abdominais intensas e contínuas, sangramento de mucosas, convulsões e irritabilidade. A doença tem em média uma duração de 5 a 7 dias, mas em casos mais graves pode-se estender por várias semanas ou até mesmo levar o paciente ao óbito. (FIOCRUZ, 2023)

Atualmente existem diversos métodos diretos e indiretos para diagnosticar a dengue. Dentre eles, podem ser citados a pesquisa de genoma do vírus feita por transcrição reversa seguida de reação em cadeia da polimerase (RT-PCR), o teste de neutralização por redução de placas (PRNT) e a inibição da hemaglutinação (IH). Embora existam alguns estudos avançados sobre a prevenção da dengue, ainda não há nenhuma vacina disponível para aplicação de toda a população, que imunizará contra os quatro tipos de vírus. Dessa forma, para prevenir essa doença a principal medida é reduzir a infestação de mosquitos por meio da eliminação de criadouros, mantendo cobertos e limpos locais que possam acumular água.

Segundo Fiocruz (2023), o *Aedes aegypti* surgiu na África e foi se espalhando para a Ásia e Américas, principalmente através do tráfego marinho. No Brasil, a doença chegou

por volta do século XVIII pelas embarcações que transportavam escravos e, de lá para cá foram notificadas diversas epidemias, causadas pelo diferentes sorotipos. Nos últimos 10 anos foi observado um aumento no número de casos e de hospitalizações. Em 1998, registrou-se uma média de 4 por 100.000 habitantes, já no período de 2000 a 2010 passaram a 49.7 por 100.000 habitantes. Outro fator importante e que vem mudando nos últimos cinco anos é a distribuição de casos por faixa etária, antes a doença era predominante em adultos e, após 2006 é possível verificar maior incidência em crianças.

Desde 1846 há relatos de epidemias de dengue no estado do Rio de Janeiro. Em 1982, os sorotipo DENV-1 e DENV-4 foram isolados em Boa Vista (RO). Porém, no ano de 1986, o sorotipo DENV-1 foi reintroduzido no Brasil, na cidade de Nova Iguaçu (RJ) e, a partir daí, a doença começou a se espalhar rapidamente para as cidades vizinhas. Desta forma, somente no primeiro ano, foram notificados mais de 33 mil casos, já atingindo outros estados do país. Nos anos seguintes, a introdução dos sorotipos DENV-2 e DENV-3 também ocorrera no estado do Rio de Janeiro (BARRETO; TEIXEIRA, 2008).

Os dados de dengue são usualmente disponibilizados de forma agregada por alguma divisão espacial. Por exemplo, suponha que o espaço de interesse seja o Estado do Rio de Janeiro, que está subdividido em função de seus municípios, facilmente pode-se obter o número de casos de dengue para cada município do Estado. Esses dados são conhecidos na literatura como dados de área.

A análise exploratória e a modelagem de dados de área é discutido de forma ampla na literatura, incluindo diversos trabalhos sobre a análise de casos de dengue. Essas análises buscam entender a dinâmica da superfície de risco, identificando áreas de maior risco e quais fatores afetam essa superfície. A análise espacial de casos de dengue no Estado do Rio de Janeiro já foi abordada em diversos trabalhos. Um exemplo pode ser visto no estudo de Alves (2022), em que o objetivo era identificar as áreas de maior risco de transmissão de dengue, utilizando indicadores ambientais e socioeconômicos no município do Rio de Janeiro. Para isso, foi considerado o modelo Poisson ICAR e sua inferência foi realizada por meio de métodos MCMC. Concluiu-se que as regiões da Zona Oeste e as regiões com maior vulnerabilidade social possuem maior risco de transmissão do que o esperado na cidade.

Em 2017, Carvalho, Magalhães e Medronho (2017) fizeram uma análise da distribuição espacial de casos da dengue no município do Rio de Janeiro nos anos de 2011 e 2012. Para a realização deste estudo eles aplicaram a técnica de georreferenciamento dos casos notificados no Sistema de Informação de Agravos de Notificação utilizando os campos

logradouro e número. Foi realizado o processo automático da ferramenta Geocoding do programa ArcGis10 e foi estimada a densidade dos casos de dengue clássicos e de dengue grave por meio de uma função Kernel apropriada. Cada densidade pode ser vista como um mapa de calor. Nesse contexto não foi utilizado um modelo estatístico e sim técnicas de análise exploratória de padrão de pontos. Os resultados apontaram que as áreas quentes para dengue clássica não coincidem geograficamente com as áreas quentes da dengue grave, estando localizadas dentro ou próximas de favelas. Além disso, o processo de georreferenciamento mostrou uma perda de 17% dos registros de dengue grave e 41% de dengue clássica.

SILVA et al. (2020) compararam o modelo de regressão clássica, o modelo de regressão do erro espacial, o modelo de regressão espacial autorregressivo e o modelo de regressão espacial de Durbin para analisar a distribuição dos casos de dengue e sua relação com fatores socioambientais no estado da Paraíba, nos anos de 2007 a 2016. Na escolha do melhor modelo para ajuste das variáveis levou-se em consideração aquele que apresentou menor valor da log-verossimilhança e menores valores do critério AIC.

Segundo Gomes et al. (2011), as condições climáticas, tais como, o aumento das chuvas, a diminuição da temperatura e a duração da luz do dia influenciam no aumento da incidência de dengue no município de João Pessoa, na Paraíba. Esse resultado foi obtido por meio do ajuste do modelo de decisão fuzzy, que identificou os períodos com risco elevados de epidemia.

No trabalho de Ferreira (2004) sobre a distribuição dos casos de dengue na cidade do Rio de Janeiro, no período de 1986 a 2002, foram utilizadas três formas distintas de análise dos casos de dengue. A primeira constituiu uma análise puramente temporal, seguida de uma análise puramente espacial e uma análise espaço-temporal. Na primeira ele utilizou os Modelos Lineares Generalizados e analisou a série histórica de casos de dengue mensalmente. No segundo, o autor combinou várias covariáveis relacionadas à fatores sócio-econômicos e de infraestrutura, e adotou modelos que incorporam uma estrutura de dependência espacial para realizar o mapeamento da doenças. Os modelos aplicados na análise espaço-temporal combinaram o que foi mencionado anteriormente com os Modelos Lineares Generalizados Dinâmicos. Na primeira abordagem foi possível detectar um padrão sazonal nas notificações mensais de dengue. Já na análise espacial identificou-se uma influência significativa de diversos fatores sócioeconômicos e de infra-estrutura. Por fim, o autor concluiu através do critério DIC que a abordagem espaço-temporal apresentou melhores resultados.

Já no estudo realizado por Cabral e Freitas (2012) sobre a distribuição espacial e determinantes socioeconômicos e demográficos da dengue nos municípios brasileiros, no ano de 2010, a metodologia utilizada foram as Regressões Ponderadas Geograficamente. Essa técnica consiste em atribuir pesos para cada observação a partir de um ponto de calibragem, onde a influência destas observações diminui à medida que se tornam mais distantes desse ponto. Os resultados mostraram que a densidade demográfica é essencial para explicar o padrão da distribuição dos casos de dengue e que, os municípios observados independentemente do nível de desenvolvimento estão suscetíveis a propagação dos casos de dengue.

Os trabalhos apresentados anteriormente se propuseram a estimar a superfície de risco, de forma exploratória com o uso de Kernel ou por meio de modelos de regressão. Alguns modelos incluíram uma estrutura de dependência espacial como as apresentadas por Alves (2022) e Ferreira (2004), utilizando o modelo de Poisson ICAR e usando o método MCMC para estimação dos parâmetros. Entretanto, o uso do ICAR em modelos espaciais pode suavizar excessivamente a superfície de risco. Sendo assim, é importante avaliar o uso de opções que não supõem uma forte dependência espacial como a proposta de (LEROUX; LEI; BRESLOW, 2000).

Na revisão de literatura feita não encontrou-se trabalhos que avaliavam os casos de dengue, considerando todo o estado do Rio de Janeiro no período de 2021. Além disso, os modelos utilizados pelos trabalhos descritos acima ajustaram modelos que assumem a existência de correlação espacial global forte. Diante disso, seria interessante estudar o número de casos de dengue no Estado do Rio de Janeiro no período em questão, considerando modelos mais flexíveis, isto é, que permitem correlações espaciais menores e comparar com a abordagem usada nos trabalhos levantados. Neste trabalho será utilizada a estatística espacial para analisar o comportamento e fazer o mapeamento da doença, além da identificação de aglomerados espaciais. Esta ferramenta serve para contribuir na tomada de decisão e por meio da aplicação dos modelos de regressão espacial, auxiliar na interpretação da dinâmica espacial dos casos dessa infecção. Para realizar a modelagem desses dados de área serão utilizados modelos com efeitos espaciais, como o *CAR* Intrínseco e o Leroux.

## 1.1 Objetivos

O principal objetivo deste trabalho é realizar uma análise espacial, incluindo o ajuste de modelos com e sem efeitos espaciais para o número de casos de dengue nos municípios do Estado do Rio de Janeiro, no ano de 2021, a fim de estimar uma superfície de risco da doença e verificar o impacto de fatores sócio-ambientais e econômicos. A análise incluirá o ajuste dos modelos Independente, *CAR* Intrínseco e Leroux.

## 1.2 Organização

Este trabalho está disposto em 4 capítulos. O Capítulo 2 se refere a metodologia usada, este capítulo apresenta os dados utilizados e o modelo utilizado no estudo do mapeamento dos casos de dengue, além de técnicas e conceitos estatísticos. Já no Capítulo 3 poderá ser encontrada a aplicação dos métodos vistos no capítulo anterior, onde poderão ser vistos os resultados obtidos neste estudo. Por fim, no Capítulo 4, será feito um breve resumo com as principais conclusões tiradas a partir da análise desses dados.

## 2 Materiais e Métodos

No capítulo de Materiais e Métodos serão abordadas as metodologias que serão utilizadas para a análise dos casos de dengue. Este capítulo está dividido em seções, nas quais serão apresentadas conceitos e modelos utilizados para obtenção dos resultados.

### 2.1 Materiais

A base de dados utilizada neste trabalho é composta por 92 linhas, referentes aos municípios que compõem o estado do Rio de Janeiro e 17 colunas, referentes as variáveis investigadas nas análises. A variável resposta trata-se do número de casos de dengue em cada município, no ano de 2021, e foi retirada do Infodengue<sup>1</sup> consultada na data 14/06/2023. Já as covariáveis utilizadas neste trabalho serão apresentadas na Tabela 1:

Estas variáveis foram retiradas do IPEA<sup>2</sup>, com base na pesquisa do Censo Demográfico do IBGE no ano de 2010 e consultada na data 02/08/2023. A seguir serão apresentados os conceitos e as ferramentas que contribuirão para este estudo.

### 2.2 Estatística Espacial

A Estatística Espacial é uma subárea da Estatística responsável pela análise de dados que possuem uma referência espacial. As ferramentas de Estatística Espacial usam a referência espacial de forma explícita nas análises. Estas ferramentas englobam uma análise exploratória de dados espaciais, incluindo a confecção de mapas, a verificação da existência de um padrão espacial para o fenômeno de interesse, bem como avaliar a existência de possíveis fatores associados ao desfecho de interesse por meio de modelos estatísticos.

Pode-se citar como exemplo de dados que possuem referências espaciais: a precipitação

---

<sup>1</sup><https://info.dengue.mat.br/>

<sup>2</sup><http://ivs.ipea.gov.br/index.php/pt/planilha>



Tabela 1: Tabela das covariáveis.

<b>Covariáveis utilizadas</b>
População total
Índice de Vulnerabilidade Social (IVS)
Índice da dimensão Infraestrutura Urbana
Índice da dimensão Capital Humano
Índice da dimensão Renda e Trabalho
Índice de Desenvolvimento Humano Municipal (IDHM)
Índice da dimensão Longevidade
Índice sintético da dimensão Educação
Índice da dimensão Renda
Índice de Gini
Taxa de analfabetismo - 18 anos ou mais
Taxa de pessoas em domicílios com abastecimento de água e esgotamento sanitário inadequados
Taxa da população que vive em domicílios urbanos sem o serviço de coleta de lixo
Taxa de pessoas que vivem em domicílios com renda per capita inferior a meio salário mínimo (de 2010) e que gastam mais de uma hora até o trabalho

pluviométrica na cidade de São Paulo, o número de casos de Covid-19 nos bairros da cidade de Niterói ou ainda a localização exata dos sinistros de carros segurados na cidade do Rio de Janeiro em Novembro de 2022.

Para realizar as análises espaciais é preciso inicialmente realizar uma análise exploratória espacial que envolve a criação de mapas. Estes mapas podem auxiliar na visualização e, conseqüentemente, na compreensão de padrões espaciais e formulação de hipóteses, como por exemplo, mapas. Os dados espaciais podem ser classificados de algumas formas, dentre as mais comuns, tem-se: superfícies contínuas, padrão de pontos e dados de área.

Nos dados de superfície contínua, as variáveis de interesse são contínuas no espaço, ou seja, são observados valores em um conjunto de localidades que podem ser distribuídas de forma regular ou irregular, chamadas de estações de monitoramento. Usualmente, esse tipo de dado é aplicado em levantamentos de recursos naturais.

Os dados chamados de padrão de pontos representam um conjunto de localizações exatas, usualmente um par de coordenadas, do local onde o fenômeno de interesse ocorreu. Neste contexto, o fenômeno é conhecido, o que é aleatório é o local onde ele ocorre. Pode-se citar como exemplos de aplicação, a localização da residência dos indivíduos

diagnosticados com uma determinada doença ou a localização da ocorrência de um crime.

Por fim, existem os dados de área. Nesse tipo de dado, tem-se uma região de interesse dividida em sub-regiões (chamadas de áreas) e observado uma quantidade agregada para cada área, pois acredita-se que o evento estudado se comporte de forma homogênea nas divisões adotadas. Dessa forma, é possível afirmar que esses valores associados não estão relacionados a localização específica do evento estudado. Em geral, esses dados têm proveito nas áreas de economia, segurança pública e epidemiologia.

Os dados de interesse deste trabalho se encontram exatamente no último tipo de dado espacial apresentado. As análises descritas a seguir são voltadas para este contexto.

## **2.3 Análise exploratória de dados de área**

A análise exploratória de dados de área se inicia com a construção de um mapa coroplético. Os mapas coropléticos são representações visuais de dados agregados para áreas de uma região de interesse, onde diferentes áreas são coloridas ou sombreadas de acordo com os valores de uma variável específica.

Os mapas coropléticos são construídos dividindo-se uma área geográfica em regiões menores, como estados, províncias ou municípios, e atribuindo a cada região uma cor ou tonalidade que representa a intensidade ou valor da variável em questão. Geralmente, cores mais escuras indicam valores mais altos, enquanto cores mais claras indicam valores mais baixos.

Esses mapas são úteis para comunicar informações complexas de maneira visualmente acessível e podem ser empregados em diversas áreas, como demografia, economia, saúde pública e ciências sociais. No entanto, é importante interpretar esses mapas com cautela, pois a escolha da paleta de cores e a divisão das categorias podem influenciar a percepção dos dados. Além disso, o tamanho das áreas geográficas pode distorcer a representação visual, sendo necessário considerar a densidade populacional ou a área das regiões ao interpretar os resultados. Na Figura 1 é apresentado um exemplo de mapa coroplético.

Essa técnica é frequentemente usada para destacar padrões regionais e variações em dados, permitindo uma rápida percepção das tendências geográficas. A ideia de investigar padrões é verificar se o fenômeno avaliado se comporta de forma semelhante em áreas vizinhas. A seguir, é discutido como especificar uma estrutura de vizinhança por meio da matriz de vizinhança.

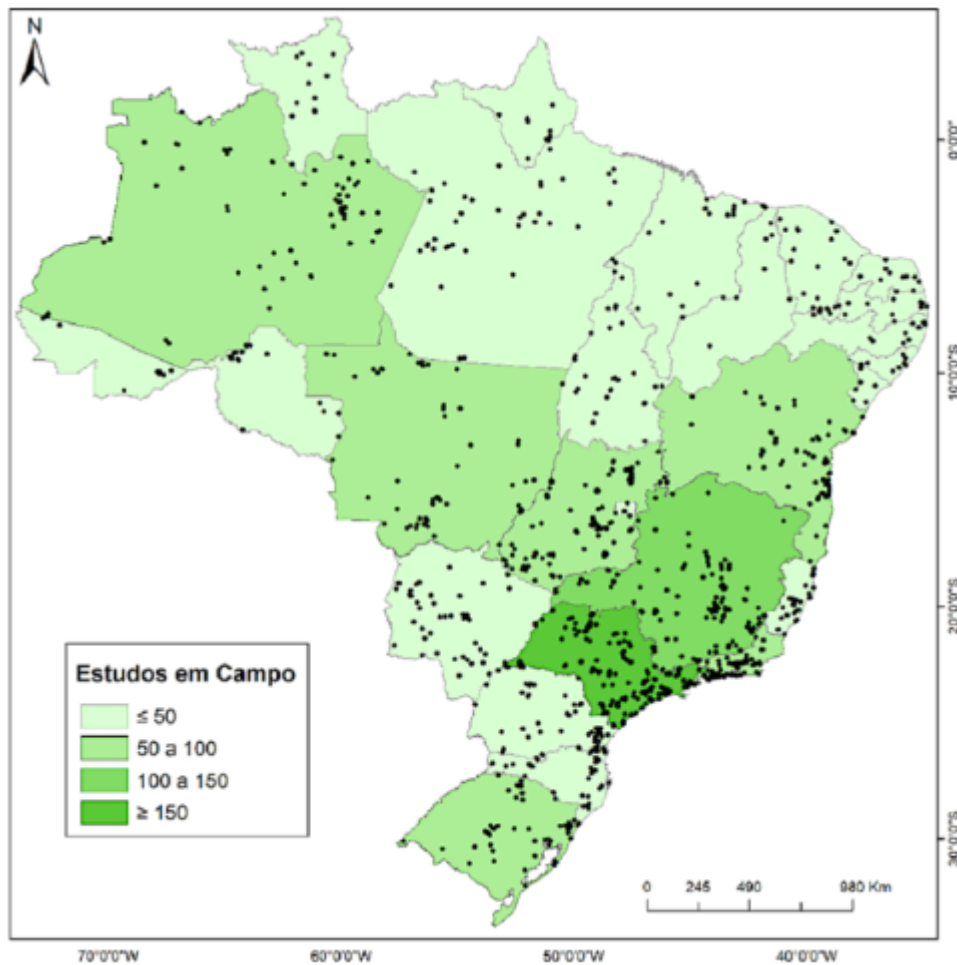


Figura 1: Exemplo de mapa coroplético

### 2.3.1 Matriz de Vizinhaça

A Matriz de Proximidade Espacial, também conhecida como Matriz de Vizinhaça, é utilizada para especificar o formato da relação espacial de dados de área. A construção dessa matriz pode ser feita baseando-se em diversos critérios, tais como, contiguidade ou distância.

Para construir a Matriz de Vizinhaça com base no critério de distância, é preciso calcular a distância entre as áreas definidas na região de interesse, com o objetivo de identificar o quão próxima uma área está das demais. Essa distância é calculada com base nos centroides das áreas. Já para construir a matriz com base no critério de contiguidade serão avaliadas as fronteiras das áreas envolvidas.

Para a definição da matriz de vizinhaça  $W_{n \times n}$  ( que será chamada de  $W$  a partir daqui), considere que o espaço de interesse  $S$  está subdividido em  $n$  áreas,  $A_1, A_2, \dots, A_n$ ,

isto é,

$$S = \bigcup_{i=1}^n A_i \text{ e } A_i \cap A_j = \emptyset.$$

A matriz  $W$  especifica duas características das relações entre as áreas. A primeira é a indicação de quem é vizinha de quem e qual o peso dessa relação. Cada elemento  $w_{i,j}$  da matriz  $W$  representa uma medida de proximidade entre as áreas  $A_i$  e  $A_j$ , tendo,  $i, j = 1, \dots, n$ , com  $i \neq j$  e  $w_{i,j} = 0$  se  $i = j$ . A especificação de  $w_{i,j}$  deve depender do problema analisado, da região de interesse, do conhecimento sobre a dependência espacial do fenômeno, entre outros fatores. Abaixo serão apresentados os principais critérios utilizados para a definição de  $w_{i,j}$ :

$$w_{i,j} = \begin{cases} 1, & \text{se o centróide de } A_i \text{ está a uma determinada distância de } A_j. \\ 0, & \text{caso contrário.} \end{cases}$$

$$w_{i,j} = \begin{cases} 1, & \text{se o centróide de } A_i \text{ é uma das áreas mais próximas do centróide de } A_j. \\ 0, & \text{caso contrário.} \end{cases}$$

$w_{i,j} = l_{i,j}/l_i$ , onde  $l_{i,j}$  é o comprimento da fronteira entre  $A_i$  e  $A_j$  e  $l_i$  é perímetro de  $A_i$ .

$$w_{i,j} = \begin{cases} 1, & \text{se } A_i \text{ tem fronteira comum com } A_j. \\ 0, & \text{caso contrário.} \end{cases}$$

Os critérios para a especificação dos pesos mais utilizados são o binário e o padronizado por linha. No binário é assumindo peso 1 para todos os elementos que são vizinhos e no padronizado cada  $w_{i,j}$  da matriz é dividido pela soma da linha a qual o elemento pertence, assim a soma dos pesos de cada linha será igual a 1, ou seja,  $\sum_{j=1}^n w_{i,j} = 1$ .

### 2.3.2 Autocorrelação Espacial

Uma medida de interesse na análise de dados de área é o que será denominado como autocorrelação espacial.

A autocorrelação espacial é um conceito utilizado em estatística espacial e análise espacial para avaliar a relação entre o fenômeno observado em diferentes locais geográficos, neste caso, nas diferentes áreas. Em outras palavras, refere-se à medida de quão semelhantes são os valores de uma variável em locais próximos uns dos outros, ou seja, em locais vizinhos.

Uma das formas utilizadas para calcular a dependência espacial na análise de dados de área é usando o índice I de Moran (BENITES, 2022), calculado da seguinte forma:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{i,j} (y_i - \bar{y})(y_j - \bar{y})}{(\sum_{i=1}^n (y_i - \bar{y})^2) (\sum \sum_{i \neq j} w_{i,j})} \quad (2.1)$$

onde  $n$  é número de observações,  $w_{i,j}$  é elemento da matriz de vizinhança,  $y_i$ ,  $y_j$  e  $\bar{y}$  indicam o valor da variável na localização  $i$  e  $j$ , respectivamente e  $\bar{y}$  é a média da variável,

Este coeficiente admite valores no intervalo de  $[-1,1]$ . Os números próximos de  $-1$  indicam a existência de uma autocorrelação negativa, o que significa que locais vizinhos apresentam diferenças em relação à variável de estudo. Caso o índice seja próximo de  $0$ , pode-se afirmar que existem evidências de que os dados sejam independentes. Por fim, a autocorrelação positiva, que corresponde a valores positivos, quanto mais próximo de  $1$ , maior será a dependência no espaço, isto é, indica que as regiões vizinhas tendem a apresentar valores similares da variável de interesse.

Além disso, foi realizado um teste estatístico do índice de Moran cuja hipótese nula representa a existência de independência espacial, condição esta dada por  $I=0$ . Porém, esta hipótese nula somente será rejeitada se o índice resultar em um valor diferente de zero, para um nível de significância predeterminado.

A seguir serão apresentados alguns conceitos importantes para a especificação dos modelos que serão utilizados neste trabalho.

## 2.4 Modelo de Poisson

A distribuição de Poisson é constantemente utilizada para modelar dados de contagem, como será feito neste trabalho. Um dos principais objetivos dessa pesquisa é avaliar o impacto das variáveis explicativas, mencionadas anteriormente, no número de casos de dengue nos municípios do Estado do Rio de Janeiro. Para tal, será realizada uma modelagem desses dados por meio da distribuição de Poisson.

Existem situações em que a suposição de normalidade e linearidade dos dados não é satisfeita, impossibilitando modelar o comportamento médio da variável por meio da combinação de covariáveis explicativas e um erro normalmente distribuído, processo conhecido como regressão linear. Para resolver essa questão, utiliza-se a classe de modelos lineares generalizados (MLGs), introduzidos por Nelder e Wedderburn (1972), que pres-

supõem que a densidade da função de distribuição da variável resposta pertence a família exponencial.

Uma variável aleatória  $Y$  tem distribuição pertencente à família exponencial se a sua função de probabilidade pode ser escrita na forma:

$$f(y; \theta) = \exp\{a(y)b(\theta) + c(y) + d(\theta)\}, \quad (2.2)$$

em que  $\theta$  é parâmetro escalar e  $a(\cdot)$ ,  $b(\cdot)$  e  $c(\cdot)$  são funções conhecidas.

Por meio dos modelos lineares generalizados é possível modelar variáveis de interesse de natureza contínua, de natureza dicotômica ou na forma de proporções, além da forma de contagem. Assim como os modelos Normal e Gama, o modelo de regressão de Poisson, representa um caso particular dos MLGs e, como se sabe, este é muito usado na análise de dados na forma de contagem (SILVA, 2000).

Para a definição do modelo considere que o espaço de interesse  $S$  está dividido em  $n$  áreas e  $Y_i$  representa o número de casos de dengue no município  $i$ ,  $i = 1, 2, \dots, n$  (variável resposta) e que foram observadas  $p$  variáveis explicativas  $\mathbf{x}_i = (1, x_{1,i}, x_{2,i}, \dots, x_{p,i})^T$ . O modelo de regressão de Poisson será definido como:

$$Y_i \sim Poi(\lambda_i) \quad (2.3)$$

$$\lambda_i = r_i \theta_i \quad (2.4)$$

$$\log(\theta_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \gamma_i. \quad (2.5)$$

O modelo descrito acima é um modelo hierárquico. No primeiro nível é dito que a variável resposta na área  $i$  tem média  $\lambda_i$ . No segundo é suposta que a média tem uma decomposição multiplicativa em uma quantidade conhecida  $r_i$ , representando um offset (a população do município  $i$ ), e a segunda é estocástica  $\theta_i$ . O terceiro nível propõe uma estrutura log-linear para  $\theta_i$ , na qual  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$  é o vetor com os efeitos das variáveis explicativas e  $\gamma_i$  é o efeito espacial do município  $i$ . A inclusão destes efeitos espaciais é no intuito de acomodar o problema de autocorrelação existente nos dados.

Os parâmetros desconhecidos do modelo são  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$  e  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)^T$ . Para realizar inferência sobre os modelos será especificada a sua função de verossimilhança dada por:

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}; y) = \prod_{i=1}^n \frac{\exp\{-r_i \exp\{x_i^T \boldsymbol{\beta}\}\} (r_i \exp\{x_i^T \boldsymbol{\beta}\})^{y_i}}{y_i!}. \quad (2.6)$$

Para a completa especificação do modelo, sob a perspectiva Bayesiana, é preciso definir as distribuições *a priori* dos parâmetros envolvidos. A seguir serão especificadas inicialmente as distribuições *a priori* dos efeitos espaciais e, em seguida, dos demais parâmetros.

## 2.5 Inferência no Modelo de Poisson

Para realizar inferência sob a perspectiva Bayesiana no modelo definido na seção anterior é preciso inicialmente especificar a distribuição *a priori* conjunta dos parâmetros desconhecidos,  $\beta$  e  $\gamma$ .

A distribuição *a priori* conjunta para  $\beta$ , isto é,  $\pi(\beta)$  será dado por:

$$\pi(\beta) = \pi(\beta_0, \beta_1, \dots, \beta_p) = \pi(\beta_0)\pi(\beta_1) \dots \pi(\beta_p), \quad (2.7)$$

no qual é assumido independência *a priori* entre todos os elementos de  $\beta$  e os mesmos são distribuídos normalmente.

A seguir, serão apresentadas algumas possibilidades para a especificação da distribuição *a priori* dos efeitos espaciais  $\gamma$ .

### 2.5.1 Distribuição *a priori* dos efeitos espaciais

O efeito espacial  $\gamma_i$  incluído na Equação (2.5) pode ser modelado por diversas formas distintas. É possível permitir que os mesmos sejam independentes ou incluir alguma estrutura de dependência.

Quando não inclui nenhuma estrutura de dependência no modelo, assume-se que:

$$\gamma_i \sim N(a_\gamma, \sigma^2), \quad i = 1, \dots, n. \quad (2.8)$$

$$\sigma^2 \sim \text{Gama} - \text{Inversa}(a, b), \quad (2.9)$$

Quando é incluída alguma estrutura de dependência espacial, em geral, ela é feita por meio de modelos autoregressivos condicionais (conhecidos como *CAR*). São utilizadas como distribuição *a priori* para representar esses efeitos em modelos hierárquicos (LEE, 2013).

O modelo *CAR Intrínseco*, proposto por (BESAG; YORK; MOLLÍÉ, 1991) e o modelo *CAR Leroux*, desenvolvido por (LEROUX; LEI; BRESLOW, 2000) são alguns dos diversos modelos autoregressivos condicionais propostos na literatura. Esses modelos

são empregados em um cenário bayesiano, no qual a inferência é baseada na simulação da cadeia de Markov.

A distribuição *a priori* do efeito espacial  $\gamma$  é definido de forma condicional, isto é, é definida a distribuição de  $\gamma_i$  condicionada a  $\gamma_{-i}$ , definida como:

$$\gamma_i | \gamma_{-i}, W, \sigma^2 \sim N \left( \frac{\sum_{i=1}^n w_{ki} \gamma_i}{\sum_{i=1}^n w_{ki}}, \frac{\sigma^2}{\sum_{i=1}^n w_{ki}} \right), \quad (2.10)$$

$$\sigma^2 \sim \text{Gama} - \text{Inversa}(a, b), \quad (2.11)$$

A esperança condicional do efeito aleatório  $\gamma_i$  é a média dos efeitos de seus vizinhos. Além disso, a variância do efeito espacial é função do número de vizinhos, isto é, quanto maior for o número de vizinhos menor é a variabilidade de  $\gamma_i$ . Este tipo de modelo é particularmente útil quando há autocorrelação espacial forte nos dados. Ele tende a suavizar excessivamente as estimativas.  $\sigma^2$  é um hiperparâmetro de variância, na qual é assumida uma distribuição *Gama - Inversa*( $a, b$ ).

Como o modelo Intrínseco representava apenas dados que possuíam autocorrelação espacial forte, (LEROUX; LEI; BRESLOW, 2000) propuseram um *CAR* alternativo para modelar forças variadas de autocorrelação espacial, usando apenas um único conjunto de efeitos aleatórios. Seu modelo é dado por:

$$\gamma_i | \gamma_{-i}, W, \sigma^2, \rho \sim N \left( \frac{\rho \sum_{i=1}^n w_{ki} \gamma_i}{\rho \sum_{i=1}^n w_{ki} + 1 - \rho}, \frac{\sigma^2}{\rho \sum_{i=1}^n w_{ki} + 1 - \rho} \right), \quad (2.12)$$

$$\sigma^2 \sim \text{Gama} - \text{Inversa}(a, b), \quad (2.13)$$

$$\rho \sim \text{Uniforme}(0, 1). \quad (2.14)$$

Esse efeito possui uma distribuição Normal com uma média que depende os efeitos das sub-regiões vizinhas, onde  $\rho$  é uma medida de autocorrelação espacial global que varia no intervalo  $[0, 1]$  de modo que, se  $\rho = 0$ , configura-se um cenário de independência espacial. Se  $\rho = 1$ , recai-se no modelo *CAR* Intrínseco, visto anteriormente.

A inferência no modelo será feita sob a ótica bayesiana usando métodos de Monte Carlo via Cadeia de Markov por meio dos algoritmos de Gibbs e Metropolis-Hasting. Para maiores detalhes ver (MIGON; GAMERMAN; LOUZADA, 2014).

Diversos softwares permitem realizar o ajuste deste modelo, como o R, o OPENBUGS, entre outros. Neste trabalho será utilizado o pacote *CARBayes* (LEE, 2013) do R, pois além da facilidade de uso no ajuste de modelos *CAR*, ele pode implementar uma classe muito mais ampla de modelos de unidades espaciais do que os listados anteriormente.



## 2.5.2 Modelos Especificados

Neste trabalho, o espaço de interesse é o estado do Rio de Janeiro particionado em seus municípios e a variável desfecho é o número de casos de dengue nos municípios do Rio de Janeiro no ano de 2021.

A seguir são apresentados os modelos investigados: o Modelo 1 será chamado de modelo Independente, o Modelo 2 será chamado de modelo CAR Intrínseco e Modelo 3 será chamado de modelo Leroux, especificados abaixo:

Modelo 1: Independente

$$Y_i \sim Poi(\lambda_i), i = 1, \dots, 92 \quad (2.15)$$

$$\lambda_i = r_i \theta_i \quad (2.16)$$

$$\log(\theta_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \gamma_i \quad (2.17)$$

$$\gamma_i \sim N(0, \sigma^2) \quad (2.18)$$

$$\sigma^2 \sim Gama - Inversa(0.1, 0.1) \quad (2.19)$$

$$\beta_j \sim N(0, 1000), j = 0, 1, \dots, 3 \quad (2.20)$$

$$(2.21)$$

Modelo 2: Car Intrínseco

$$Y_i \sim Poi(\lambda_i), i = 1, \dots, 92 \quad (2.22)$$

$$\lambda_i = r_i \theta_i \quad (2.23)$$

$$\log(\theta_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \gamma_i \quad (2.24)$$

$$\gamma_i | \boldsymbol{\gamma}_{-i}, W, \sigma^2, \rho \sim N \left( \frac{\sum_{i=1}^n w_{ki} \gamma_i}{\sum_{i=1}^n w_{ki}}, \frac{\sigma^2}{\sum_{i=1}^n w_{ki}} \right) \quad (2.25)$$

$$\sigma^2 \sim Gama - Inversa(0.1, 0.1) \quad (2.26)$$

$$\beta_j \sim N(0, 1000), j = 0, 1, \dots, 3 \quad (2.27)$$

$$(2.28)$$

Modelo 3: Leroux

$$Y_i \sim Poi(\lambda_i), i = 1, \dots, 92 \quad (2.29)$$

$$\lambda_i = r_i \theta_i \quad (2.30)$$

$$\log(\theta_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \gamma_i \quad (2.31)$$

$$\gamma_i | \boldsymbol{\gamma}_{-i}, W, \sigma^2, \rho \sim N \left( \frac{\rho \sum_{i=1}^n w_{ki} \gamma_i}{\rho \sum_{i=1}^n w_{ki} + 1 - \rho}, \frac{\sigma^2}{\rho \sum_{i=1}^n w_{ki} + 1 - \rho} \right) \quad (2.32)$$

$$\sigma^2 \sim Gama - Inversa(0.1, 0.1) \quad (2.33)$$

$$\beta_j \sim N(0, 1000), j = 0, 1, \dots, 3 \quad (2.34)$$

$$\rho \sim Uniforme(0, 1). \quad (2.35)$$

onde:

- $Y_i$  é o número de casos de dengue no ano de 2021 no município  $i$ .
- $\mathbf{x}_i^T$  é o vetor com  $p$  variáveis explicativas.
- $\boldsymbol{\beta}$  é o vetor com  $p + 1$  parâmetros de regressão relacionados ao vetor das variáveis explicativas.
- $\gamma_i$  é o efeito espacial no município  $i$ .
- $w_{ki}$  são os elementos da matriz de vizinhança.
- $\sigma^2$  é a variância dos efeitos aleatórios.
- $\rho$  é a autocorrelação espacial.

Como visto anteriormente, o primeiro modelo, Independente, assume independência dos efeitos espaciais. Já os últimos dois modelos levam em consideração efeitos espaciais estruturados, ou seja, consideram que os municípios do Rio de Janeiro possuem uma autocorrelação espacial. No modelo Leroux, a distribuição condicional autoregressiva depende dos hiperparâmetros  $\sigma^2$  e  $\rho$ , que correspondem, respectivamente, a variância e a força da autorrelação espacial. Já o Modelo CAR Intrínseco, é um caso particular do Modelo Leroux quando  $\rho = 1$ , o que significa que o modelo é mais adequado para cenários com autocorrelação global forte. Por outro lado, o modelo Independente, é um caso particular do Modelo Leroux quando  $\rho = 0$ .

A inferência de cada modelo foi realizada usando Métodos de Monte Carlo via Cadeia de Markov, com base em uma amostra de tamanho 1.000. Para obter tal amostra, foi

considerado um período de aquecimento de tamanho 100.000 e para evitar autocorrelações foi considerado um espaçamento de 10.000. As cadeias foram obtidas por meio do pacote CARBAYES (LEE, 2013). A convergência das cadeias foi analisada de forma visual e podem ser observadas através das Figuras 8, 9 e 10 dispostas no Apêndice.

Para análise dos modelos foram consideradas as estimativas pontuais das covariáveis e o intervalo de credibilidade dos efeitos dessas covariáveis, obtidos na saída do R.

Os resultados obtidos a partir dos ajustes destes modelos aos casos de dengue nos municípios do estado do Rio de Janeiro poderão ser vistos a seguir.

## 2.6 Comparação de modelos

Quando existem diversos modelos com diferentes níveis de complexidade para realizar os ajustes dos dados, faz-se necessária a escolha do melhor modelo para ser utilizado. Um dos critérios mais utilizados para comparação e seleção de modelos é o Deviance Information Criterion (DIC), proposto por (SPIEGELHALTER et al., 2002). Este método trata-se de uma generalização dos critérios Akaike Information Criterion (AIC) e Bayesian Information Criterion (BIC), e é dado por:

$$DIC(M_k) = \bar{D}_k + p_k = 2\bar{D}_k - D(\bar{\theta}_k) \quad (2.36)$$

Em que  $\theta_k$  é o vetor de parâmetros do modelo  $k$ ,  $\bar{\theta}_k$  é a média à posteriori de  $\theta_k$ ;  $D$  é a função deviance, que mede o ajuste do modelo;  $\bar{D}_k = E_{\theta_k|y}(D(\theta_k))$ , representando a média à posteriori da deviance;  $D(\bar{\theta}_k)$  é uma estimativa pontual da deviance obtida através da utilização da média à *posteriori* de  $\theta$  na expressão do desvio;  $p_k = E_{\theta_k|y}(D(\theta_k)) - D(E_{\theta_k|y}(\theta_k)) = \bar{D}_k - D(\bar{\theta}_k)$  é um termo penalizador que mede a complexidade do modelo e  $D(\theta_k) = -2 \log l(y|\theta_k, k) + 2 \log f(y)$  é a *posteriori* da deviance, sendo  $l(y|\theta_k, k)$  a função de verossimilhança e  $f(y)$  uma função de padronização dos dados.

Assim como os demais critérios, o melhor modelo será o que apresentar o menor valor de DIC.

### 3 Análise dos resultados

Este capítulo do trabalho apresenta os resultados obtidos a partir da modelagem dos casos de dengue nos municípios do estado do Rio de Janeiro. Foram utilizados métodos gráficos e numéricos para obterem os resultados. Além disso, foram comparados três modelos e a decisão do melhor modelo se deu através do critério DIC.

Como proposto na Seção 2.3, primeiramente foram plotados mapas coropléticos da variável número de casos de dengue em 2021 e das covariáveis definidas para modelagem desses dados para o estado do Rio de Janeiro. Nestes mapas, as áreas geográficas foram coloridas ou sombreadas em tons vermelho de acordo com a variação da variável de interesse.

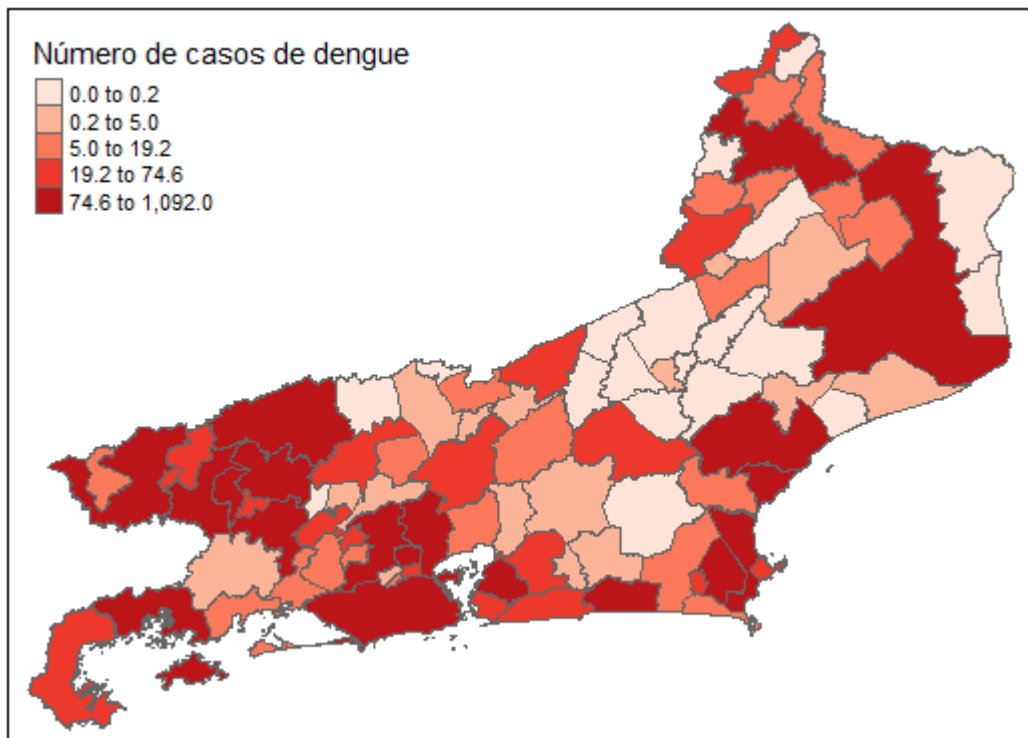


Figura 2: Mapa coroplético dos casos de dengue no Estado do RJ

Os modelos Independente, CAR Intrínseco e Leroux foram ajustados com os diferentes

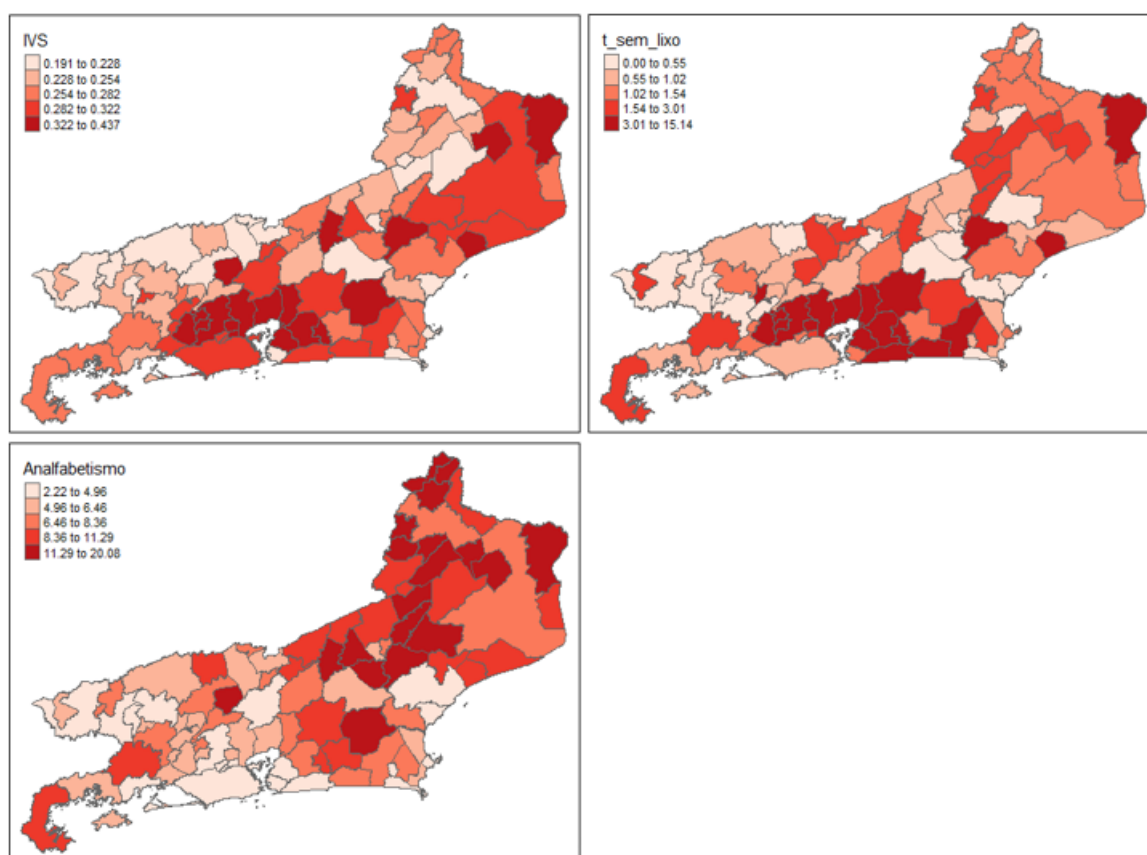


Figura 3: Mapa cloroplético das covariáveis selecionadas

subconjuntos de covariáveis, apresentadas na Seção 2.1, porém os melhores resultados foram obtidos considerando as variáveis explicativas: Índice de Vulnerabilidade Social (IVS), Taxa da população que vive em domicílios urbanos sem o serviço de coleta de lixo (tsemlixo) e Taxa de analfabetismo - 18 anos ou mais (Analfabetismo).

Na Figura2 nota-se que o número de casos de dengue está bem disposto em toda a região do mapa, porém é possível perceber sutilmente uma maior concentração de casos nas regiões do Médio Paraíba, Metropolitana e Noroeste Fluminense, que podem ser prontamente identificadas através do Mapa 4. Além disso, percebe-se também que as regiões vizinhas possuem um comportamento similar, ou seja, as observações em locais próximos são mais semelhantes entre si do que as observações em locais mais distantes.

Na Figura3, percebe-se que os indicadores estão menos acentuados na região do Médio Paraíba, uma das regiões que apresenta maior número de casos de dengue, o que sugere que nessa região há uma relação contrária dos indicadores IVS, Taxa da população que vive em domicílios urbanos sem o serviço de coleta de lixo e Taxa de analfabetismo com os casos da doença. Já na região Metropolitana, apesar de um número alto de casos de dengue, a taxa de analfabetismo é menor, indicando também uma relação contrária,

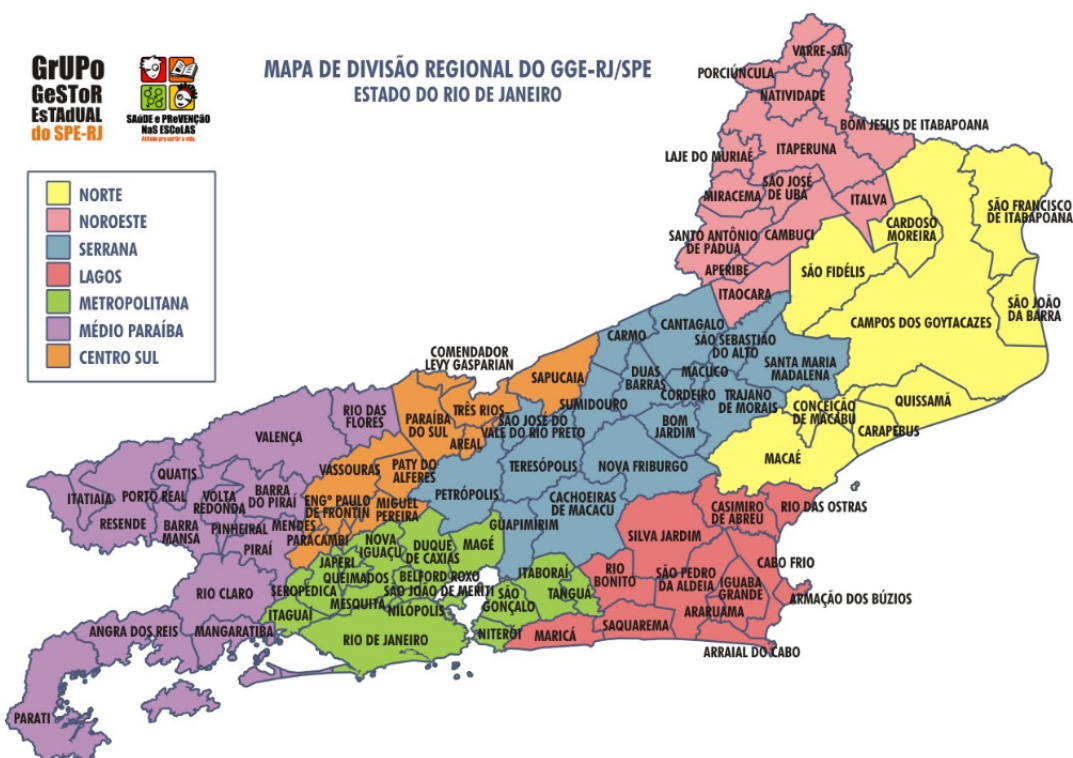


Figura 4: Mapa municípios do estado do Rio de Janeiro

diferentemente dos outros dois indicadores. Por fim, no Noroeste Fluminense, parece haver uma relação direta da concentração de casos nesta região com os indicadores Taxa da população que vive em domicílios urbanos sem o serviço de coleta de lixo e Taxa de analfabetismo.

Em seguida, calculou-se a matriz de vizinhança, apresentada na Seção 2.3.1, para representar a proximidade espacial entre os municípios do Rio de Janeiro. Para o cálculo, foi considerado o critério de borda comum, para considerar dois municípios vizinhos, e utilizou-se o critério binário para a especificação dos pesos, no qual é assumindo peso 1 para os elementos que são vizinhos.

A autocorrelação global dos casos de dengue foi estimada pelo Índice de Moran, método usado para avaliar a presença de padrões espaciais em dados de área, apresentada na Tabela 2.

Tabela 2: Análise do Índice de Moran para o número de casos de dengue no ano de 2021.

Índice de Moran	P-valor
0,1585	0,0009

A partir dos cálculos do Índice de Moran é possível concluir que, apesar do índice ter resultado em um valor baixo, o que indica pouca correlação espacial, há evidências da

existência de dependência espacial significativa, com base no p-valor = 0,0009.

Após isso, foram ajustados alguns modelos com diferentes subconjuntos de covariáveis até chegar no então modelo utilizado, com as variáveis Índice de Vulnerabilidade Social (IVS), Taxa da população que vive em domicílios urbanos sem o serviço de coleta de lixo (tsemlixo) e Taxa de analfabetismo - 18 anos ou mais (Analfabetismo).

As Tabelas 3, 4 e 5 apresentam as estimativas pontuais, o intervalo de credibilidade para os efeitos das covariáveis observadas, além dos hiperparâmetros dos modelos ajustados. Abaixo também poderão ser vistos os gráficos dos efeitos espaciais dos modelos.

Tabela 3: Estimativa pontual e intervalar para os parâmetros do Modelo Independente.

<b>Coefficiente</b>	<b>Média</b>	<b>IC (2,5%)</b>	<b>IC (97,5%)</b>	<b>Probab.</b>
Intercepto	-3,9550	-8,3442	0,4249	-
IVS	-14,0599	-30,6085	3,5806	0,95
tsemlixo	0,1318	-0,2601	0,5025	0,77
Analfabetismo	-0,1573	-0,4500	0,0684	0,91
$\sigma^2$	2,8914	1,8339	4,9787	-

Tabela 4: Estimativa pontual e intervalar para os parâmetros do modelo CAR Intrínseco.

<b>Coefficiente</b>	<b>Média</b>	<b>IC (2,5%)</b>	<b>IC (97,5%)</b>	<b>Probab.</b>
Intercepto	-6,1744	-10,6236	-1,5136	-
IVS	-4,6711	-23,3815	12,7387	0,72
tsemlixo	0,0933	-0,2366	0,4058	0,75
Analfabetismo	-0,1955	-0,5272	0,1038	0,90
$\sigma^2$	8,7709	5,3490	15,5020	-

Tabela 5: Estimativa pontual e intervalar para os parâmetros do modelo Leroux.

<b>Coefficiente</b>	<b>Média</b>	<b>IC (2,5%)</b>	<b>IC (97,5%)</b>	<b>Probab.</b>
Intercepto	-5,5002	-10,4195	-0,0911	-
IVS	-7,1775	-25,4924	11,7807	0,73
tsemlixo	0,0925	-0,2800	0,4313	0,71
Analfabetismo	-0,1942	-0,5794	0,1120	0,89
$\sigma^2$	6,5391	3,3317	13,1078	-
$\rho$	0,5378	0,1807	0,9097	-

Observando as Figuras 5, 6 e 7 percebe-se que os gráficos possuem padrões semelhantes, mas não são iguais. Percebe-se que há uma variedade espacial significativa estimada pelos efeitos espaciais. Vale ressaltar que os gráficos não possuem a mesma escala no eixo y e que os pontos no gráfico dos efeitos espaciais no modelo Independente estão mais dispersos em relação à linha de tendência, o que pode indicar maior variabilidade espacial.

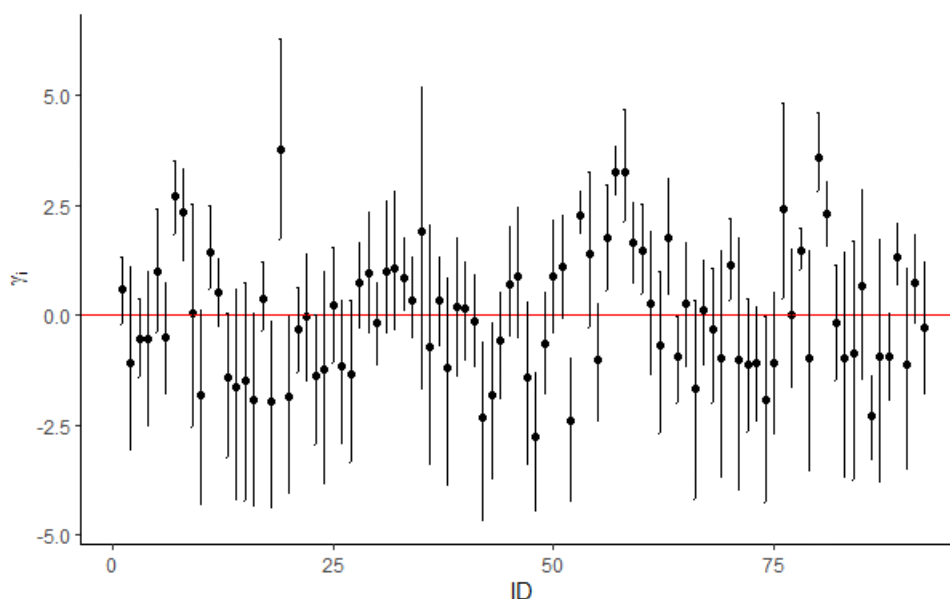


Figura 5: Gráfico de efeitos espaciais no modelo Independente

Tabela 6: Tabela DIC dos modelos ajustados.

Modelo	DIC
Independente	622,8323
CAR Intrínseco	<b>612,5283</b>
Leroux	618,1976

Além disso, no modelo Independente, a variância dos efeitos aleatórios, representada pelo  $\sigma^2$ , é igual a 2,8914, um valor menor do que nos modelos com efeitos espaciais.

Os resultados obtidos na Tabela 6 por meio do critério DIC para comparação de modelos, sugerem que o melhor modelo de ajuste dos dados é o CAR Intrínseco. Fazendo a comparação de modelos com a inclusão do efeito espacial de forma independente e estrutural, notou-se que os modelos espaciais performaram melhor do que o modelo Independente. Sendo assim, serão apresentadas somente as interpretações do modelo CAR Intrínseco.

Veja que no modelo CAR Intrínseco, escolhido através do critério DIC, a estimativa do efeito do Índice de Vulnerabilidade Social apresenta uma grande amplitude no intervalo de credibilidade, o que significa que o modelo tem dificuldade de avaliar o efeito desta covariável, indicando grande incerteza. Este intervalo de credibilidade contém, inclusive o zero, porém calculando a probabilidade desse efeito ser negativo, já que sua estimativa pontual deu -4,6711, concluiu-se que possui 72% de chance do IVS estar relacionado ao número de casos de dengue de forma inversa, isto é, regiões com mais vulneráveis possuem menores quantidade de casos de dengue, pois quanto maior o valor do IVS, mais vulnerável



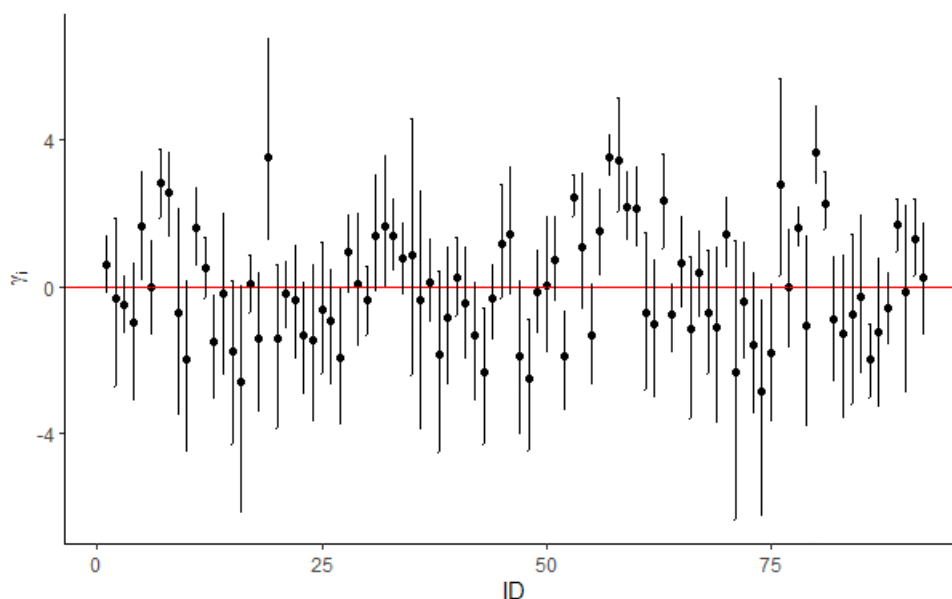


Figura 6: Gráfico de efeitos espaciais no modelo CAR Intrínseco

é o município.

No indicador Taxa da população que vive em domicílios urbanos sem o serviço de coleta de lixo, nota-se que a amplitude do intervalo de credibilidade é menor, indicando uma maior precisão na estimação deste efeito. Porém o IC também contém 0, no entanto, há 75% de probabilidade da taxa da população sem serviço de coleta de lixo estar relacionado de forma direta ao número de casos de dengue, já que sua estimativa pontual foi maior que 0, ou seja, os locais mais afetados pela falta de coleta de lixo possuem mais casos da doença.

Por último, o efeito do indicador Taxa de Analfabetismo, também possui amplitude pequena do intervalo de credibilidade e o 0 pertence ao IC. Entretanto, como sua estimativa pontual foi de -0,1955, calculando a probabilidade desse efeito no modelo ser negativo, chegou-se a conclusão de que existe uma probabilidade de 90% da Taxa de Analfabetismo ter uma relação inversa com o número de casos de dengue nos municípios do RJ. O que leva a crer que as regiões que apresentam maiores taxas de analfabetismo em indivíduos com 18 ou mais anos têm menos casos de dengue.

Como visto anteriormente, as covariáveis Índice de Vulnerabilidade Social e Taxa de Analfabetismo indicaram uma relação inversa com o número de casos de dengue, o que não era bem o esperado neste estudo. Entretanto, pode ser que existam outras variáveis que não estão contempladas no modelo e podem estar relacionadas com estas, ou ainda, a dificuldade ou uma pior qualidade de registro desses dados nas regiões observadas, impossibilitando a identificação de correlação dessas variáveis.

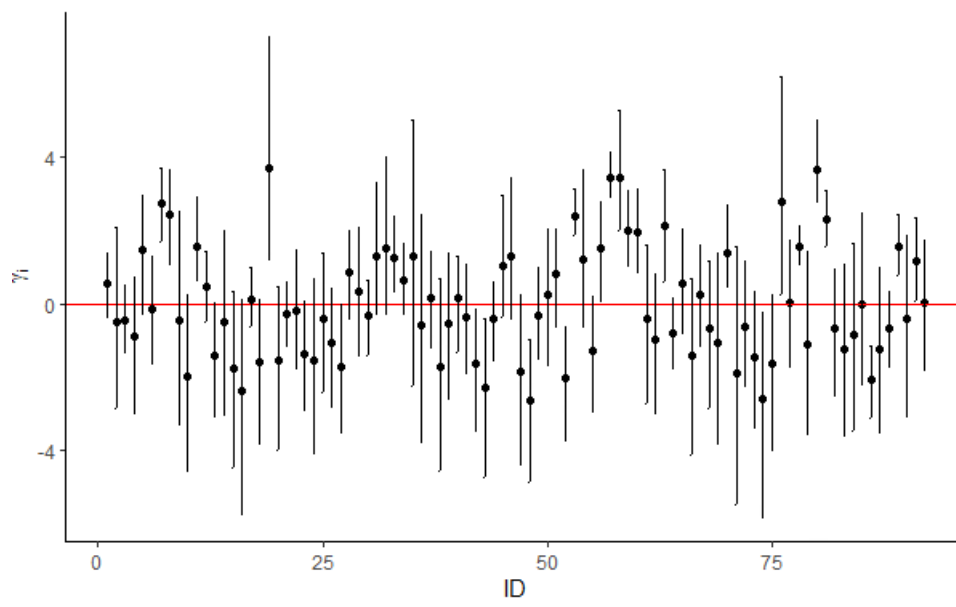


Figura 7: Gráfico de efeitos espaciais no modelo Leroux

## 4 Conclusão

Como visto no Capítulo 1, algumas cidades do Rio de Janeiro foram cenários de grandes marcos na história do surgimento da dengue no país, como é o caso de Nova Iguaçu. Diante deste fato e considerando que não foram encontrados estudos que avaliassem os casos de dengue nos municípios do estado do Rio de Janeiro, foi proposto neste trabalho uma análise dos casos de dengue nesses municípios, no ano de 2021, utilizando modelos que não consideram efeitos espaciais estruturados e modelos que consideram correlações espaciais menores, comparados aos modelos utilizados nos trabalhos levantados na revisão de literatura.

Para analisar o comportamento da doença, utilizou-se a estatística espacial e ferramentas como mapas para verificação da existência de padrões espaciais e fatores associados a estes dados de área. Para tal feito, foram ajustados três modelos considerando as covariáveis Índice de Vulnerabilidade Social (IVS), Taxa da população que vive em domicílios urbanos sem o serviço de coleta de lixo (tsemlixo) e Taxa de analfabetismo - 18 anos ou mais (Analfabetismo).

Através do mapa cloroplético do estado do Rio de Janeiro identificou-se que as regiões do Médio Paraíba, Metropolitana e Noroeste Fluminense tinham uma maior concentração de casos da doença e que as regiões vizinhas possuíam um comportamento similar. Já nos mapas cloropléticos das variáveis que foram abordadas nesta análise, notou-se que os indicadores estavam menos acentuados na região do Médio Paraíba, o que sugeria que estes indicadores apresentavam uma relação contrária à quantidade de casos de dengue nesta região. Na região Metropolitana o mesmo ocorreu para o indicador Taxa de analfabetismo. Já no Noroeste Fluminense, as covariáveis Taxa da população que vive em domicílios urbanos sem o serviço de coleta de lixo e Taxa de analfabetismo demonstraram terem uma relação direta à concentração de casos nesta região.

Após o cálculo do Índice de Moran e com base nas tabelas das estimativas pontuais e intervalar das covariâncias e hiperparâmetros dos modelos Independente, CAR Intrínseco

e Leroux, concluiu-se que os dados possuíam dependência espacial, logo os modelos que levavam em consideração efeitos espaciais estruturados deveriam performar melhor do que modelos que não consideram estruturas espaciais, o que de fato foi observado.

Para definir o melhor modelo a ser usado no ajuste dos dados de dengue, foi utilizado o critério Deviance Information Criterion (DIC) e o escolhido foi o modelo CAR Intrínseco, que apresentou menor valor DIC.

Nos resultados obtidos no ajuste do CAR Intrínseco percebeu-se que todos os intervalos de credibilidade das covariáveis observadas continham o 0, porém calculando as probabilidades dos efeitos destas covariáveis no modelo, constatou-se que havia alta probabilidade dos indicadores IVS, Taxa da população que vive em domicílios urbanos sem o serviço de coleta de lixo e Taxa de analfabetismo estarem relacionados ao número de casos de dengue nas regiões observadas. Esta análise só foi possível, pois a inferência do modelo foi feita sob uma perspectiva Bayesiana.

Para as covariáveis IVS e Taxa de analfabetismo foi possível inferir que os locais com maiores vulnerabilidades sociais e com maiores percentuais de analfabetos, possuíam menos casos registrados de dengue. Já para o indicador, Taxa da população que vive em domicílios urbanos sem o serviço de coleta de lixo, a estimativa pontual foi maior que 0, indicando que os municípios mais afetados pela falta de coleta de lixo possuíam mais casos de dengue.

Diferentemente do que era esperado neste estudo, o Índice de Vulnerabilidade Social e a Taxa de analfabetismo apontaram uma relação contrária ao número de casos de dengue nos municípios do estado do Rio de Janeiro, porém este resultado pode estar relacionado a qualidade dos registros nos municípios mais vulneráveis ou ainda por não terem sido consideradas outras variáveis importantes na modelagem.

O problema de registro de dados em alguns municípios pode envolver uma série de desafios relacionados à coleta, registro, comunicação e disponibilidade de informações. Alguns dos problemas podem ser: notificação incompleta, infraestrutura de saúde insuficiente, falta de padronização nos registros, entre outros. Municípios que apresentam desigualdades socioeconômicas significativas podem enfrentar dificuldades adicionais na coleta de dados. A falta de acesso a educação, serviços de saúde e outros recursos podem contribuir para registro desses dados. Além disso, a falta de recursos humanos, financeiros e tecnológicos pode impactar negativamente na qualidade desses registros.

Diante deste fato, para uma análise mais assertiva seria preciso uma pesquisa mais

aprofundada desses indicadores analisados ou até mesmo empregar no modelo outras covariáveis que pudessem estar relacionados ao casos de dengue.

# Referências

- ALVES, C. M. Análise espacial dos casos de dengue na cidade do rio de janeiro. 2022.
- BARRETO, M. L.; TEIXEIRA, M. G. Dengue no brasil: situação epidemiológica e contribuições para uma agenda de pesquisa. *Estudos avançados*, SciELO Brasil, v. 22, p. 53–72, 2008.
- BENITES, L. *Moran's I: definição, exemplos*. 2022. Disponível em: [https://statologos.com/moran-yo/#google\\_vignette](https://statologos.com/moran-yo/#google_vignette). Acesso em: 25 de novembro de 2023.
- BESAG, J.; YORK, J.; MOLLIE, A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, v. 43, p. 1–59, 1991. Disponível em: <https://api.semanticscholar.org/CorpusID:13799826>.
- CABRAL, J. A.; FREITAS, M. V. Distribuição espacial e determinantes socioeconômicos e demográficos da dengue nos municípios brasileiros. *Revista brasileira de estudos regionais e urbanos*, v. 6, n. 1, p. 81–95, 2012.
- CARVALHO, S.; MAGALHÃES, M. d. A. F. M.; MEDRONHO, R. d. A. Análise da distribuição espacial de casos da dengue no município do rio de janeiro, 2011 e 2012. *Revista de Saúde Pública*, SciELO Brasil, v. 51, 2017.
- FERREIRA, G. d. S. Análise espaço-temporal da distribuição dos casos de dengue na cidade do rio de janeiro no período de 1986 a 2002. Universidade Federal do Rio de Janeiro, 2004.
- FIOCRUZ. *Dengue*. 2023. Disponível em: <https://www.cpqrr.fiocruz.br/pg/dengue/>. Acesso em: 21 de julho de 2023.
- GOMES, R. G. S. et al. Modelagem da incidência do dengue através de aspectos climáticos, entomológicos e sócio-demográficos no município de João Pessoa, Paraíba. Universidade Federal da Paraíba, 2011.
- LEE, D. Carbayes: An R package for spatial areal unit modelling with conditional autoregressive priors. *Journal of Statistical Software*, v. 55, n. 13, 2013.
- LEROUX, B. G.; LEI, X.; BRESLOW, N. E. Estimation of disease rates in small areas: A new mixed model for spatial dependence. In: . [s.n.], 2000. Disponível em: <https://api.semanticscholar.org/CorpusID:118331591>.
- MIGON, H. S.; GAMERMAN, D.; LOUZADA, F. Statistical inference: An integrated approach, second edition. In: . [s.n.], 2014. Disponível em: <https://api.semanticscholar.org/CorpusID:86672689>.
- NELDER, J.; WEDDERBURN, R. W. M. Generalized linear models. *Journal of the Royal Statistical Society*, v. 135, n. 3, p. 370–384, 1972.

SILVA, A. T. e G. *Modelos Lineares Generalizados - da teoria à prática -*. Paraná: ufpr.br, 2000.

SILVA, E. T. C. d. et al. Análise espacial da distribuição dos casos de dengue e sua relação com fatores socioambientais no estado da paraíba, brasil, 2007-2016. *Saúde em Debate*, SciELO Public Health, v. 44, p. 465–477, 2020.

SPIEGELHALTER, D. J. et al. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, v. 64, 2002. Disponível em: <https://api.semanticscholar.org/CorpusID:70115>.

## APÊNDICE A – Cadeias do modelo Independente

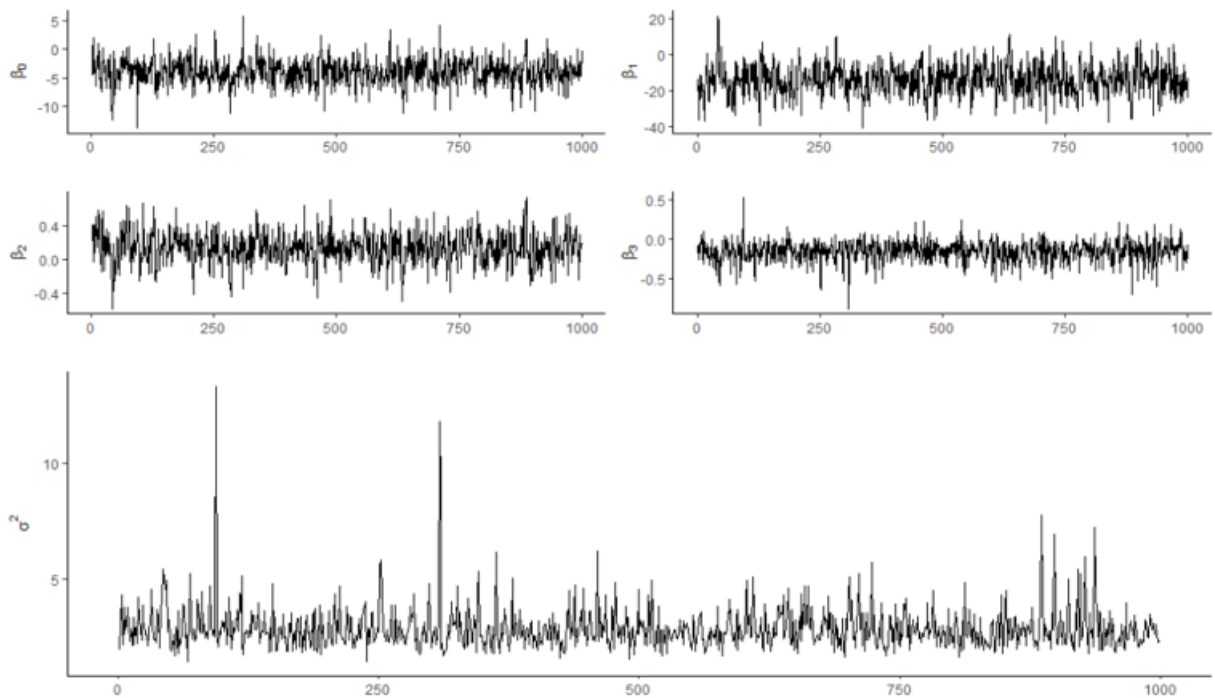


Figura 8: Gráfico das cadeias - Modelo Independente



## APÊNDICE A – Cadeias do modelo CAR Intrínseco

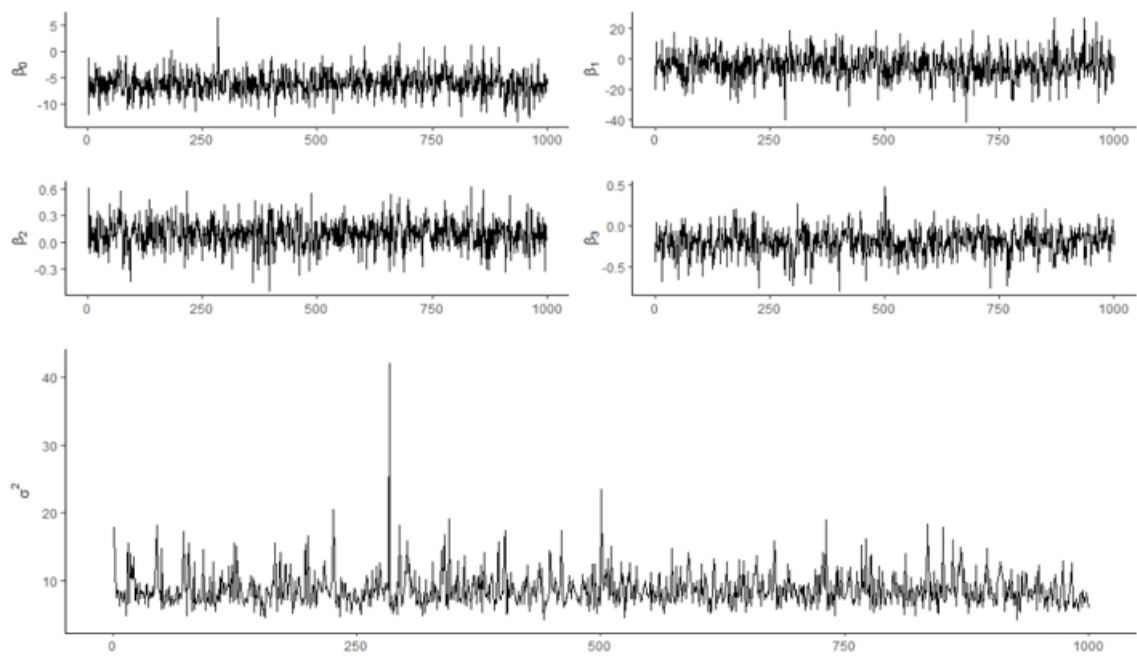


Figura 9: Gráfico das cadeias - Modelo CAR Intrínseco

## APÊNDICE A – Cadeias do modelo Leroux

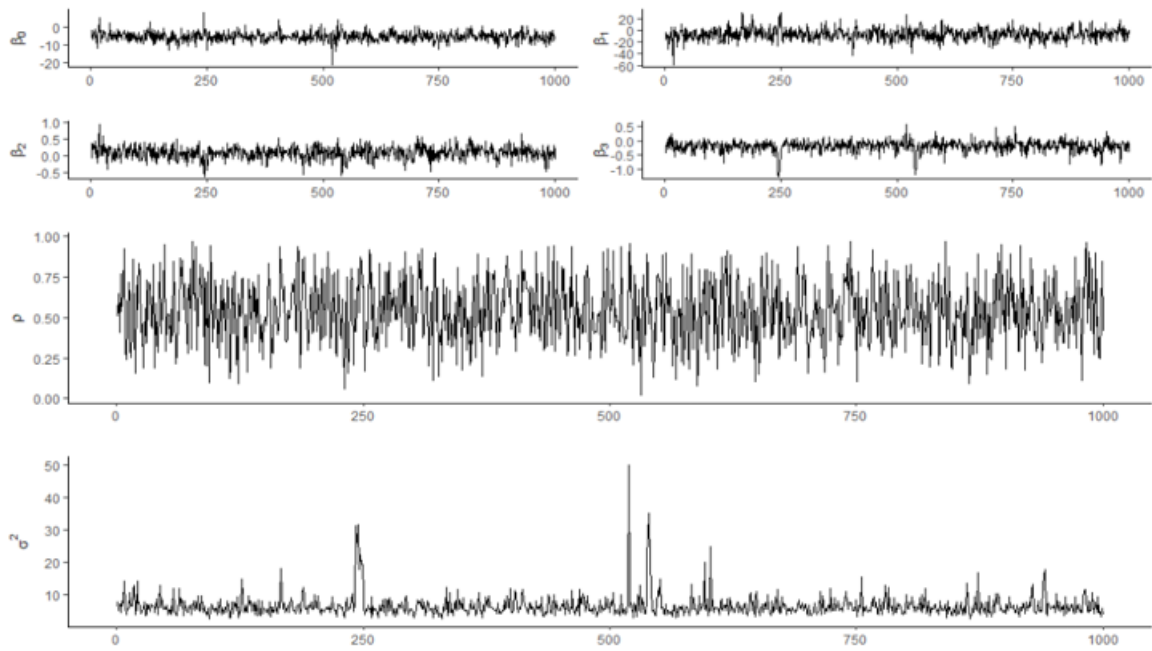


Figura 10: Gráfico das cadeias - Modelo CAR Leroux