
Terceiro Relatório Parcial - Fiotec

Emergências de Doenças em Contextos de Iniquidades Sociais: Abordagem Epidemiológica

5 de abril de 2024

Isabel S Diniz ¹
Louisiana R Quinino ²

1 INTRODUÇÃO

Para análise inicial, uma das doenças em estudo é a dengue que faz parte do grupo de doenças chamada *arboviroses*. No Brasil, o vetor da dengue é a fêmea do mosquito *Aedes aegypti*.

"Aspectos como a urbanização, o crescimento desordenado da população, o saneamento básico deficitário e os fatores climáticos mantêm as condições favoráveis para a presença do vetor, com reflexos na dinâmica de transmissão desses arbovírus. A dengue possui padrão sazonal, com aumento do número de casos e o risco para epidemias, principalmente entre os meses de outubro de um ano a maio do ano seguinte."(MINISTÉRIO DA SAÚDE, 2024)

Em Pernambuco, segundo o governo, houve um aumento de 542% novos casos prováveis, em relação ao mesmo período do ano passado e 18 mortes estão em investigação. (DIÁRIO DE PERNAMBUCO, 2024)

2 MATERIAL E MÉTODO

2.1 Coleta dos Dados de Dengue

Os dados de dengue foram coletados no site do DATASUS a partir do SINAN (DATASUS TABNET, 2024). Foram coletados dados desde 2001 até 2024, por mês de primeiro sintoma e município de residência. Os dados brutos estruturados da seguinte forma: para cada ano baixado, as linhas correspondem aos municípios e seus respectivos identificadores, pre-estabelecido pelo IBGE, e as colunas são os meses de primeiro sintoma, em que os dados correspondem aos casos de dengue acumulados dentro do mês, e aqueles que não apresentam números, o mês ou o ano não aparecem, são preenchidos por '0', pois significa que não houve casos de dengue naquela observação.

Após o tratamento dos dados, o conjunto apresenta as seguintes colunas:

¹UFPE, belsdiniz@gmail.com

²Universidade, gmail

- **GEOCODIGO**: código identificador do município;
- **MUNICIPIO**: nome do município;
- **TIPO_REGIAO**: Interior ou Região Metropolitana (RMR);
- **GERES**: Gerências Regional de Saúde;
- **MACRO_RS**: macrorregião de saúde;
- **CENSO**: censo em que a quantidade populacional se refere (2000, 2010 ou 2022);
- **QUANT_POP**: população do município;
- **N_CASOS**: casos de dengue no mês;
- **DATA**: mês e ano (ANO-MÊS) que ocorreu;
- **TX_DENGUE**: taxa de incidência de dengue;

2.1.1 Variável de interesse

Para esse estudo, a taxa de incidência de dengue por 100mil habitantes foi analisada, e calculada da seguinte forma:

$$\text{Tx de Incidência} = \left(\frac{\text{número de casos de dengue no mês}}{\text{população}} \right) \times c \quad (1)$$

em que:

- **número de casos de dengue no mês**: número dos casos de dengue por município de residência e mês no período de 2001 a 2024;
- **população**: população do município baseada na população do censo de 2000, 2010 ou 2022 (IBGE, 2024);
- **c**: constante = 100.000. São utilizados números múltiplos de 10 com objetivo de se evitar muitas casas decimais e facilitar a apresentação dos resultados.

Em relação à população para o cálculo da taxa de dengue dos casos ocorridos entre o ano de 2000 a 2006, foi considerado o censo de 2000, entre o ano de 2007 a 2016 foi considerado o censo de 2010, e para o ano de 2017 a 2024 foi considerado o censo de 2022.

2.2 Ferramentas Utilizadas

Para obter os resultados, utilizou-se a linguagem R (R Core Team, 2024a) e os respectivos pacotes foram: **dplyR** (Wickham *et al.*, 2023), para manipulação dos dados; **ggplot2** (Wickham, 2016), para análise gráfica; **sf** (Pebesma & Bivand, 2023a), **ggspatial** (Dunington, 2023) e **tmap** (Tennekes, 2018), para leitura do shapefile e construção do mapa; **spdep** (Pebesma & Bivand, 2023b) para realizar o teste de moran global e local e obter o diagrama de Moran. Para a decomposição da série e modelagem foram utilizados os pacotes **stats** (R Core Team, 2024b) e **forecast** (Hyndman & Khandakar, 2008); entre outros pacotes para a manipulação e tratamento de dados.

2.3 Análise descritiva

A análise descritiva é a primeira fase da análise estatística após a coleta de dados. Usa-se tal técnica para organizar, descrever e resumir as informações relevantes e importantes de um conjunto de características vinda de um conjunto de observações ou mesmo comparar dois conjuntos ou mais de dados analisando suas características. O objetivo vai desde organizar a identificar anomalias, como erro de digitação ou dispersão alta dos dados causada pela presença de um valor muito diferente dos demais, conhecido como *outlier*?. Na análise exploratória espacial, pode-se descrever padrões espaciais, inter-relações ou observações atípicas através da análise de mapas ou gráficos (MINISTÉRIO DA SAÚDE, 2007a).

2.4 Análise Exploratória Espacial

2.4.1 Dados de Área

São dados agregados em unidades de análise frequentemente delimitadas por polígonos fechados. Segundo Waller & Gotway (1965), na análise espacial os polígonos podem ser especificados pela localização do mesmo, e este é formado por um conjunto de pontos $u_i, v_i, i = 0, \dots, n$ conectados por segmentos de linha, definindo os vértices do polígono, em que o último ponto e o primeiro são equivalentes ($u_0 = u_n, v_0 = v_n$). Nesse caso, esses pontos são as coordenadas, e referenciam os polígonos espacialmente. O limites dos polígonos são usualmente divisões políticos-administrativas e, em alguns campos de estudo, estes limites são definidos pelo próprio pesquisador.

Municípios ou estados tem sido unidades básicas de referência geográfica para dados epidemiológicos. Mas, através da agregação de dados, o ambiente de sistema de informação geográfica (SIG) permite a utilização de diferentes unidades espaciais que melhor transpareça a informação ou distribuição de um determinado evento. Dentro destas unidades estabelecidas tem uma população com um determinado perfil epidemiológico, e nestes há condições ambientais e sociais associados. (MINISTÉRIO DA SAÚDE, 2006).

2.4.2 Sistema de informações geográficas

Os SIGs são utilizados para a captura, armazenamento, georrenciamento e análise de informação geográfica, em que permite uma visão mais detalhada de como o objeto de estudo se comporta no espaço.

Dados tratados em SIG são georreferenciados, que possuem localização podendo ser expressa por meio de coordenadas, e assim como é importante localizar estes no espaço, também é encontrar relações do mesmo com outros tipos de dados (Hino *et al.*, 2006; MINISTÉRIO DA SAÚDE, 2006).

2.4.3 Critério de vizinhança

Geralmente deseja-se identificar uma vizinhança, para cada grupo de estudo (regiões), o conjunto de regiões. Uma maneira de identificar esse conjunto é através de uma medida de proximidade espacial, fornecendo pesos adequados para os valores que estão sendo calculados. A vizinhança pode ser obtida de diferentes maneiras, e uma delas é por adjacência, ou contiguidade (Waller & Gotway, 1965).

2.4.4 Vizinhos espaciais

São considerados vizinhos espaciais polígonos que compartilham pontos de limite, em que a lista destes vizinhos é criada com base em regiões que compartilham pelo menos um ponto. O modo esses vizinhos como são obtidos é do tipo *Queen*, em que são considerados vizinhos todos os polígonos que compartilham a mesma fronteira ou um mesmo vértice, e estes são definidos como vizinhos de primeira ordem (Bivand *et al.*, 2008).

2.4.5 Matriz de Vinhança

A matriz de vinhança pode ser composta por apenas uma lista de vizinhos de cada município, e é utilizada para estimar a variabilidade espacial de dados de área.

Seja um conjunto de n áreas A_1, A_2, \dots, A_n , contruiu-se uma matriz $W^{(1)}$ ($n \times n$) onde cada um dos elementos w_{ij} é uma medida de proximidade entre A_i e A_j , e mesmo é definido por:

$$w_{ij} = \begin{cases} 1 & \text{se } A_i \text{ compartilha algum ponto em comum da fronteira com } A_j; \\ 0 & \text{caso contrário.} \end{cases} \quad (2)$$

A matriz de vizinhos $W^{(1)}$ é binária e simétrica. Uma possibilidade é normalizar esta matriz para que a soma dos pesos das linhas seja sempre igual a 1 (Câmara *et al.*, 2004).

2.4.6 Dependência espacial e associação espacial

Um dos passos fundamentais na análise exploratória de dados espaciais é mostrar como os valores estão relacionados no espaço, isto é, medir o quanto o valor observado da variável de estudo em um determinado local é dependente dos valores desta mesma variável em localizações vizinhas.

A associação espacial mede a correlação entre a variável de estudo e o espaço e varia de -1 a 1 , se for 0 não há correlação, e quanto mais próximo de 1 for essa medida, mais semelhantes são os vizinhos (MINISTÉRIO DA SAÚDE, 2006).

2.4.7 Teste global: Índice de Moran global

O Índice de Moran Global é uma das medidas para identificar se há associação espacial e ele mede o quanto um valor em uma unidade de área é semelhante aos outros que o cercam, dado por:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad \forall i \neq j. \quad (3)$$

em que:

y_i : i -ésima observação;

\bar{y} : média da variável de interesse;

w_{ij} : peso espacial da matriz de vizinhança W correspondente à ligação entre i e j .

Ele está relacionado a um teste onde a hipótese nula é de independência espacial.

2.4.8 Teste local: Índice de Moran local

O indicador local de associação espacial (Local Indicators of Spatial Association - LISA), ou Índice de Moran local, mede a autocorrelação espacial, que explica o quanto o valor observado da variável em um determinado local depende dos valores da mesma em locais vizinhos, possibilitando identificar padrões de dependência espacial. Esta estatística para uma observação i é dada por ?:

$$I_i = (y_i - \bar{y}) \sum w_{ij} y_j - \bar{y}.$$

Para a padronização do índice é definida como:

$$I_i = \frac{(y_i - \bar{y})}{\sum_{k=1}^n (y_k - \bar{y})^2 / n} \sum_{j=1}^n w_{ij} (y_j - \bar{y}), \quad (4)$$

em que o denominador $(y_k - \bar{y})^2 / (n)$ é o segundo momento central, ou variância.

E o teste de Moran local é tal que a hipótese nula é também de independência espacial. Detalhes sobre a estatística de teste para o teste de Moran Local e Global podem ser vistos em (Bivand *et al.*, 2008, p. 275-277).

2.4.9 Diagrama de espalhamento de Moran e *Moran map*

O diagrama de espalhamento de Moran objetiva mostrar graficamente o comportamento da variabilidade espacial dos dados de estudo. Esta variabilidade é analisada a partir da dispersão entre o desvio da variável de estudo \mathbf{Z} (desvio da variável: diferença da variável com sua média) e \mathbf{WZ} (média dos desvios), em que \mathbf{W} é a matriz de vizinhança, como já visto. A partir disto, são obtidos quatro quadrantes cartesianos no gráfico centrado nos valores zero nos eixos horizontal e vertical, e estes são interpretados da seguinte forma:

- Alto-Alto (desvios positivos e médias positivas) e Baixo-Baixo (desvios negativos e médias negativas): associação espacial positiva, em que a localização possui vizinhos com valores semelhantes;
- Alto-Baixo (desvios positivos e médias negativas) e Baixo-Alto (desvios negativos e médias positivas): associação espacial negativa, em que a localização possui vizinhos com valores distintos.

Os valores correspondentes aos vistos no diagrama de espalhamento de Moran, e aqueles que deram significativos no teste de Moran Local são apresentados espacialmente no *Moran map* (Ver MINISTÉRIO DA SAÚDE (2007a)) aparecendo de forma classificada em quatro grupos, de acordo com sua localização no quadrante do gráfico de espalhamento (NEVES *et al.*, 2000).

2.5 Análise da série temporal

2.5.1 Decomposição Sazonal - Método LOESS

A componente sazonal é determinada por suavização de LOESS (local polynomial regression - regressão polinomial local) da subsérie sazonal, com a janela sendo periódica, a suavização é

efetivamente substituída pela média. Os valores sazonais são removidos e o restante suavizado para encontrar a tendência. O nível geral é removido do componente sazonal e adicionado ao componente de tendência. Este processo é iterado algumas vezes. O componente restante são os resíduos do ajuste sazonal mais a tendência ([DATA SCIENCE INSTITUTE, 2024](#)).

2.5.2 Modelo Autorregressivo de Média Móvel (ARIMA)

Um modelo ARIMA pode ser entendido descrevendo cada um de seus componentes da seguinte forma ([Hyndman & Khandakar, 2008](#)):

- Autorregressão (AR): refere-se a um modelo que mostra uma variável que regride em seus próprios valores defasados, ou anteriores;
- Média móvel (MA): incorpora a dependência entre uma observação e um erro residual de um modelo de média móvel aplicado a observações defasadas.

Os parâmetros dos modelos são os seguintes:

- p: o número de observações de defasagem no modelo, também conhecido como ordem de defasagem;
- d: o número de vezes que as observações brutas são diferenciadas; também conhecido como grau de diferenciação;
- q: o tamanho da janela da média móvel, também conhecida como ordem da média móvel;

2.5.3 Algoritmo de Holt-Winters

O algoritmo Holt-Winters é um algoritmo de previsão mais popular para séries temporais, voltados para fins de detecção de anomalias e, especialmente, na previsão de tempo, e de simples implementação. o objetivo é introduzir tendência e/ou sazonalidade de forma aditiva ou multiplicativa no algoritmo de alisamento exponencial. Mais detalhes sobre o algoritmo pode ser visto em [Hyndman *et al.* \(2008\)](#).

3 Resultados

3.1 Análise descritiva

Uma análise exploratória dos casos de dengue e da taxa de incidência de dengue de 2001 a 2024 (TX_DENGUE) em Pernambuco foi realizada com objetivo de avaliar o seu comportamento, e valores discrepantes.

Observou-se que o maior número de casos de dengue foi 14.107 casos correspondente ao município de Recife no mês de Fevereiro de 2002, com sua taxa também sendo a maior observada, com aproximadamente 77.759 a cada 100 mil habitantes. A média de casos foi de 12 casos no mês e por município e média de 45 casos a cada 100 mil habitantes. O desvio padrão (DP) foi de

122 casos por mês e município e 607 casos a cada 100mil habitantes. Ao total, Pernambuco teve 609.370 casos de dengue entre 2001 e 2024. Os dados apresentaram 7744 outliers, que devem ser investigados por meio da análise temporal e espacial.

3.2 Análise Espacial

Nos anos 2023, 2024 e dentro de todo o período de estudo, o município com maior taxa média de incidência foi Riacho das Almas com 611, 2579 e 3158 casos a cada 100mil habitantes, respectivamente.

Ao fazer o mapa da distribuição da taxa de dengue na escala contínua (Figura 2 A,B e C) do período de estudo, dos anos de 2023 e 2024, foi visto a presença de outliers em diferentes locais de Pernambuco, cujos os municípios e suas respectivas taxas estão nas tabelas 1, 2 e 3. Já na Figura 2 D, E e F foi possível observar aleatoriedade espacial, ou seja, não há presença de agrupamentos ou tendência espacial da dengue.

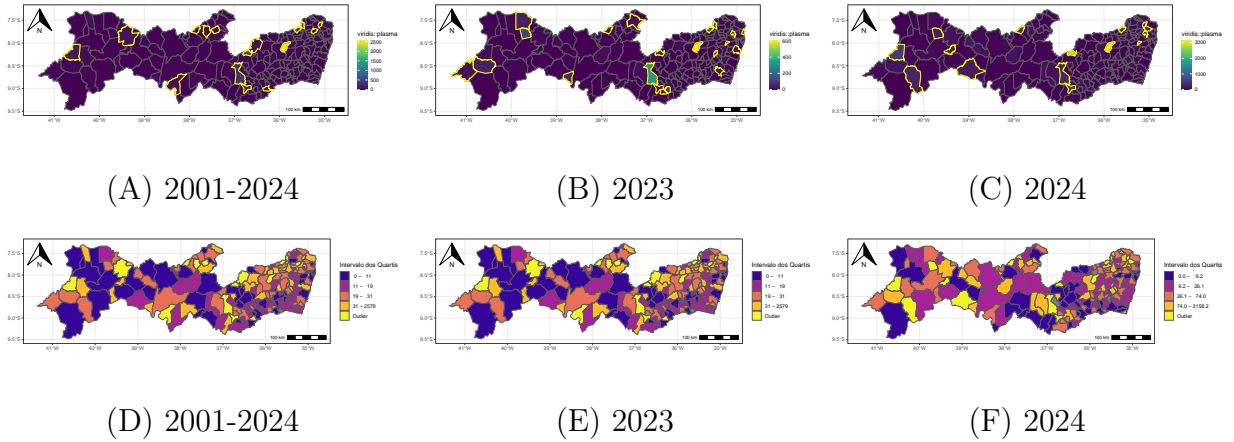


Figura 1: Distribuição da taxa de incidência da dengue por 100mil habitantes por município em uma escala contínua (A,B e C) e por quartis (D,E e F). Contorno e forma em amarelo são os municípios em que as taxas são outliers. Em cinza são os municípios que não apresentaram casos de dengue - Pernambuco (2001-2024)

Tabela 1: Taxa de incidência de dengue dos municípios detectados como outliers no período de estudo com as respectivas mesorregiões e casos acumulados - Pernambuco (2001 a 2024)

	NM_MUN	NM_MESO	CUM_CASOS	TX_DENGUE
1	Afogados da Ingazeira	Sertão Pernambucano	6789	67.53
2	Condado	Mata Pernambucana	4008	60.66
3	Ingazeira	Sertão Pernambucano	842	66.59
4	Jataúba	Agreste Pernambucano	3844	88.82
5	Nazaré da Mata	Mata Pernambucana	5363	62.81
6	Orobó	Agreste Pernambucano	23690	379.56
7	Palmeirina	Agreste Pernambucano	1860	85.59
8	Pedra	Agreste Pernambucano	19510	324.83
9	Petrolândia	São Francisco Pernambucano	7419	85.99
10	Poção	Agreste Pernambucano	14091	463.33
11	Quixaba	Sertão Pernambucano	1314	71.34
12	Riacho das Almas	Agreste Pernambucano	136825	2579.41
13	Saloá	Agreste Pernambucano	9451	236.05
14	Santa Filomena	Sertão Pernambucano	4525	129.84
15	São Vicente Férrer	Agreste Pernambucano	5121	111.15
16	Serrita	Sertão Pernambucano	4089	80.59

Tabela 2: Taxa de incidência de dengue dos municípios detectados como outliers em 2023 com as respectivas mesorregiões e casos acumulados - Pernambuco (2023)

	NM_MUN	NM_MESO	CUM_CASOS	TX_DENGUE
1	Afrânio	São Francisco Pernambucano	42	21.23
2	Araçoiaba	Metropolitana de Recife	209	90.18
3	Brejão	Agreste Pernambucano	86	78.89
4	Chã de Alegria	Mata Pernambucana	39	26.24
5	Chã Grande	Mata Pernambucana	92	37.41
6	Dormentes	São Francisco Pernambucano	92	47.45
7	Exu	Sertão Pernambucano	167	43.76
8	Gameleira	Mata Pernambucana	49	22.67
9	Granito	Sertão Pernambucano	100	119.55
10	Itacuruba	São Francisco Pernambucano	33	62.59
11	Ilha de Itamaracá	Metropolitana de Recife	169	61.56
12	Jaboatão dos Guararapes	Metropolitana de Recife	1582	20.16
13	Orobó	Agreste Pernambucano	391	153.89
14	Pedra	Agreste Pernambucano	871	317.18
15	Poção	Agreste Pernambucano	67	53.18
16	Quixaba	Sertão Pernambucano	30	38.41
17	Riacho das Almas	Agreste Pernambucano	1489	611.04
18	Saloá	Agreste Pernambucano	32	19.71
19	São José do Egito	Sertão Pernambucano	71	18.99
20	São Vicente Férrer	Agreste Pernambucano	39	19.90
21	Terezinha	Agreste Pernambucano	29	36.36

Tabela 3: Taxa de incidência de dengue dos municípios detectados como outliers em 2024 com as respectivas mesorregiões e casos acumulados - Pernambuco (2024)

	NM_MUN	NM_MESO	CUM_CASOS	TX_DENGUE
1	Araçoiaba	Metropolitana de Recife	164	283.04
2	Belém do São Francisco	São Francisco Pernambucano	104	189.22
3	Chã de Alegria	Mata Pernambucana	126	339.06
4	Condado	Mata Pernambucana	129	174.89
5	Granito	Sertão Pernambucano	39	186.49
6	Itaquitinga	Mata Pernambucana	157	315.98
7	Lagoa Grande	São Francisco Pernambucano	191	270.00
8	Orobó	Agreste Pernambucano	194	305.42
9	Pedra	Agreste Pernambucano	187	272.39
10	Poção	Agreste Pernambucano	907	2879.64
11	Quixaba	Sertão Pernambucano	89	455.85
12	Riacho das Almas	Agreste Pernambucano	1924	3158.19
13	Saloá	Agreste Pernambucano	225	554.32
14	Santa Filomena	Sertão Pernambucano	177	508.05
15	Terra Nova	São Francisco Pernambucano	47	177.38

Ao realizar o teste de Moran Global dentro do período, foi visto que a taxa de incidência média dentro do período não apresentou associação espacial significativa com valor-p de 0.42 e Índice de Moran Global de aproximadamente -0.002 , ou seja, bem próximo de 0. Os dados também foram analisados para os últimos anos de 2023 e 2024, e também não apresentou correlação global significativa, apresentando aproximadamente o valor-p 0.67 e 0.64, respectivamente e índice de Moran próximos de 0. O diagrama de Moran ilustra essa associação espacial, em que uma reta mais inclinada indica maior associação espacial. Na Figura , podemos observar que a reta está pouco inclinada e há um ponto de alavanca, que é o outlier máximo correspondente à taxa de incidência do município Riacho das Almas.

Portanto, podemos concluir que não há dependência espacial significativa, comprovando o que foi observado na Figura 2 D, E e F.

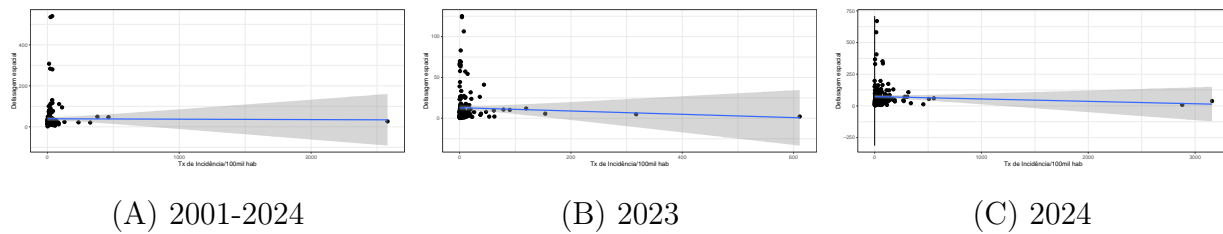


Figura 2: Digrama de Moran dos anos (A) 2001-2024 (B) 2023 (C) 2024 - Pernambuco (2001-2024)

3.3 Análise Temporal

Na 4 observou-se a presença de sazonalidade nos casos de dengue durante o período de 2001 a 2023, com maior número de casos registrados em 2002.

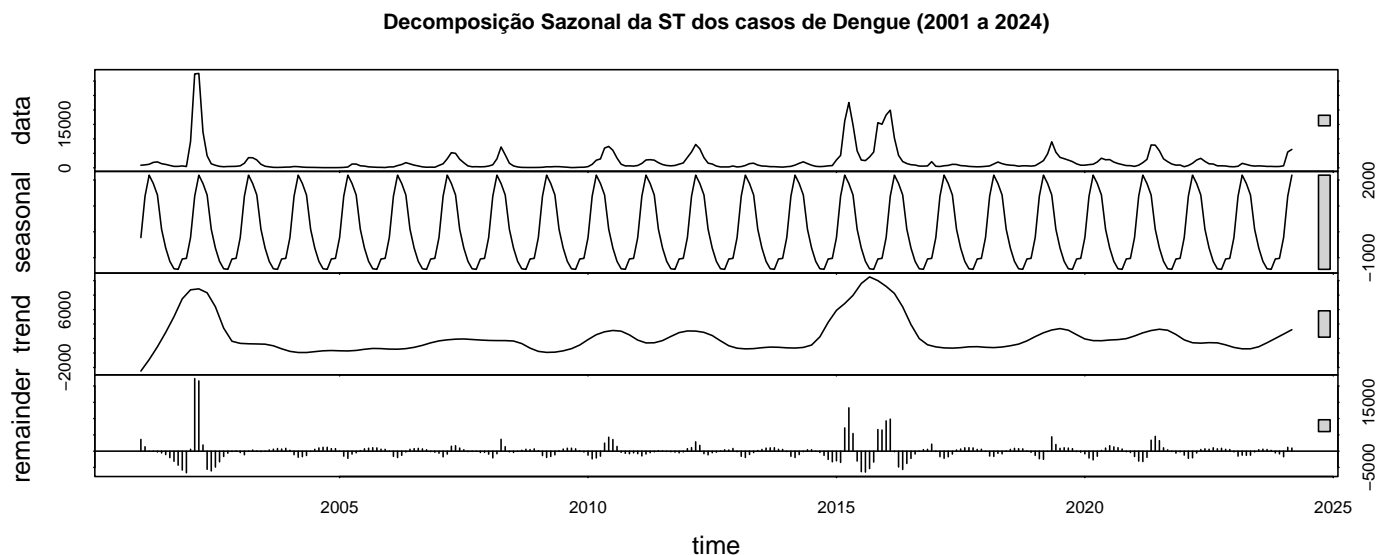


Figura 3: Decomposição Sazonal LOESS da Série Temporal dos casos de Dengue - Pernambuco (2001-2024)

Pra a modelagem da série temporal, os dados foram divididos em treino e teste, em que os dados de treino foram os casos registrados até o fim de 2023, e os dados de teste os casos entre Janeiro e Março de 2024. Dessa forma, os dados de treino foram ajustados usando o algoritmo multiplicativo Holt Winters, considerando o componente de sazonalidade, e o modelo ARIMA.

O primeiro passo antes de construir o Modelo Arima foi verificar se os dados já são estacionários ou não, pois Arima só pode ser usado quando os dados forem estacionários. Logo, aplicou-se o teste Dickey-Fuller que apresentou o valor-p de 0.01, ou seja, rejeitou-se a hipótese de não estacionaridade ao nível de 5% de significância. Ao fazer o ajuste do ARIMA o modelo ótimo encontrado foi o $ARIMA(5,0,0)(1,0,0)$.

Ao fazer as previsões para os três meses subsequentes aos dados de treino, correspondentes aos meses de Janeiro, Fevereiro e Março, foi possível ver que o modelo $ARIMA(5,0,0)(1,0,0)$ apresentou menor erro de previsão em relação ao Algoritmo Multiplicativo de Holt-Winters (faixas vermelha e verde, respectivamente).

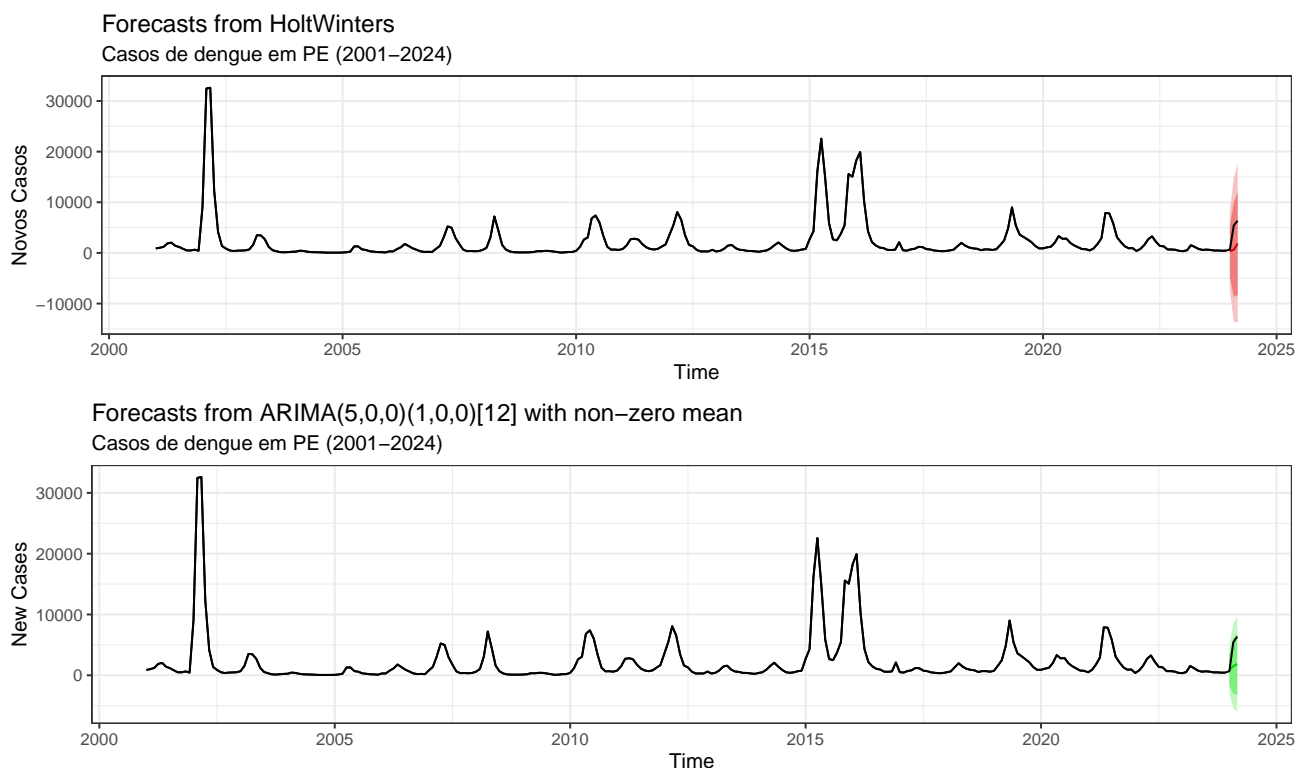


Figura 4: Previsão dos meses Janeiro a Março de 2024 da Série Temporal dos casos de Dengue dos dados de treino. Faixas vermelha e verde correspondem aos erros de previsão do algoritmo Holt-Winters multiplicativo e o modelo ajustado ARIMA(5,0,0)(1,0,0), respectivamente - Pernambuco (2001-2024)

É possível ver que o modelo ajustado ARIMA apresentou menores erros em comparação ao algoritmo de Holt-Winters tanto para os dados de treino quanto para o de teste. Ou seja, ele foi o melhor ajustado.

Ajuste	conjunto	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1	Theil's U
Holt-Winters	Treino	-391.33	3985.98	1322.18	-40.25	70.05	0.51	0.64	
	Teste	3222.79	3810.74	3222.79	70.32	70.32	1.25	-0.21	1.02
ARIMA	Treino	2.57	2019.50	922.99	-88.96	108.31	0.36	0.01	
	Teste	2741.73	3466.11	2902.16	36.03	59.62	1.12	-0.11	0.84

Na análise de resíduos, foi realizado o teste de normalidade de Shapiro-Wilk, que apresentou significância estatística ao nível 5% com um valor-p menor que 0.0001. Portanto os resíduos não apresentam normalidade significativa, indicando a necessidade de normalização dos dados antes do ajuste do modelo.

Ao realizar o teste Ljung Box, para testar se os dados possuem autocorrelação nos erros de previsão, foi visto que o ARIMA(5,0,0)(1,0,0) não possuíram autocorrelação significativa ao nível 5% nos erros de previsão (valor-p igual a 0.9). Mas, o algoritmo multiplicativo Holt-Winters possui autocorrelação nos erros de previsão com valor-p menor que 0.0001 ao nível de significância 5%. Ou seja, a presença de autocorrelação nos resíduos de um modelo pode indicar que o modelo não capturou completamente a estrutura de dependência temporal dos dados, logo o ARIMA(5,0,0)(1,0,0) se sobressai ao Holt-Winters.

A tabela abaixo mostra as previsões do modelo $ARIMA(5,0,0)(1,0,0)$ e os valores verdadeiros correspondentes aos dados de teste:

Previsões	921	1496	1850
Teste	680	5456	6356

Ou seja, o modelo previu que o número de casos iria crescer muito menos do que realmente cresceu.

4 Conclusão

Ao fazer a análise descritiva, foi visto que em Fevereiro de 2002 houve um número alto de casos de dengue no município de Recife, que refletiu na série temporal estudada, exigindo um estudo mais detalhado sobre essa ocorrência. Na análise espacial foi visto também a presença de outliers e pontos de alavanca no Estado, e foi observado que a taxa de dengue se distribui de maneira aleatória na área de estudo, o que exige um estudo espacial mais detalhado dentro das mesorregiões e inserção de fatores ambientais e pontos de alagamento, para um estudo mais detalhado da distribuição de dengue no espaço. Já na análise temporal, observou-se que os casos de dengue apresentam sazonalidade e o modelo que melhor se ajustou aos dados foi o $ARIMA$, trazendo melhores previsões. Entretanto, a rejeição de normalidade nos resíduos no teste de shapiro-wilk, indica a necessidade de normalização dos dados antes do ajuste, para maior precisão.

5 Trabalhos Futuros

Como trabalho futuro, indicamos que os dados sejam analisados de forma espacial considerando a estratificação por períodos chuvosos e secos, e por mesorregião. Dessa forma indica-se as seguintes análises:

- Análise da série temporal com os dados normalizados;
- Gráfico de controle, para estudo dos pontos discrepantes temporais;
- Estudo espaço-temporal da dengue;
- Estudo de fatores espaciais externos que podem influenciar no aumento de casos de dengue, tais como pluviosidade e temperatura;
 - teste de correlação dos dados de dengue com esses fatores no período de estudo e estado;
 - teste de correlação dos dados de dengue com esses fatores por estratificação mencionada anteriormente;
- Análise espaço-temporal da dengue relacionada aos dados de pluviosidade e temperatura

A concentração de valores baixos e presença de 0 nos dados pode indicar o ajuste da distribuições inflacionadas no zero, e para a modelagem da taxa de incidência por habitante, indica-se a distribuição Beta Inflacionada no Zero ou outras distribuições apropriadas para valores entre 0e1, como a Simplex, por exemplo, no modelo que será aplicado.

Referências

- Bivand, Roger S, Pebesma, Edzer J, Gómez-Rubio, Virgilio, & Pebesma, Edzer Jan. 2008. *Applied spatial data analysis with R*. Vol. 747248717. Springer.
- Câmara, Gilberto, Carvalho, Marília S, Cruz, Oswaldo G, & Correa, Virginia. 2004. Análise espacial de áreas. *Análise espacial de dados geográficos. Brasília: EMBRAPA*, 157–82.
- DATA SCIENCE INSTITUTE. 2024. *Time Series Decomposition in R*.
- DATASUS TABNET. 2024. *Ministério da Saúde/SVS - Sistema de Informação de Agravos de Notificação (SINAN)*.
- DIÁRIO DE PERNAMBUCO. 2024. *Estado atinge marca de 15,1 mil casos prováveis de dengue; aumento é de mais de 500*
- Dunnington, Dewey. 2023. ggspatial: Spatial Data Framework for ggplot2. *R package version 1.1.8*.
- Hino, Paula, Scatena Villa, Tereza Cristina, Midori Sasaki, Cinthia, de Almeida Nogueira, Jordana, & Benedita dos Santos, Claudia. 2006. Geoprocessamento aplicado à área da saúde. *Revista Latino-Americana de Enfermagem*, **14**(6).
- Hyndman, Rob, Koehler, Anne B, Ord, J Keith, & Snyder, Ralph D. 2008. Forecasting with exponential smoothing: the state space approach. *Springer Science & Business Media*.
- Hyndman, Rob J, & Khandakar, Yeasmin. 2008. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, **26**(3), 1–22.
- IBGE. 2024. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - Censo Demográfico 2000, 2010 e 2022.
- MINISTÉRIO DA SAÚDE. 2006. Abordagens Espaciais na Saúde Pública. *I edn. Vol. 1. MS*.
- MINISTÉRIO DA SAÚDE. 2007a. Introdução a Estatística Espacial para Saúde Pública. *I edn. Vol. 3. MS*.
- MINISTÉRIO DA SAÚDE. 2024. DENGUE - MINISTÉRIO DA SAÚDE.
- NEVES, Marcos C, Ramos, Frederico Roman, Camargo, Eduardo Celso Gerbi, Câmara, GILBERTO, & Monteiro, ANTÔNIO MIGUEL. 2000. *Análise exploratória espacial de dados sócio-econômicos de São Paulo*. Salvador: GIS Brasil2000.
- Pebesma, Edzer, & Bivand, Roger. 2023a. Spatial Data Science: With applications in R. *Chapman and Hall/CRC*.
- Pebesma, Edzer, & Bivand, Roger S. 2023b. Spatial Data Science With Applications in R. *Chapman Hall*.
- R Core Team. 2024a. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria*.

- R Core Team. 2024b. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.*
- Tennekes, Martijn. 2018. tmap: Thematic Maps in R. Journal of Statistical Software, 84(6), 1–39.*
- Waller, Lance A., & Gotway, Carol A. 1965. Applied Spatial Statistics for Public Health Data. JOHN WILEY SONS, INC.*
- Wickham, Hadley. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.*
- Wickham, Hadley, François, Romain, Henry, Lionel, Müller, Kirill, & Vaughan, Davis. 2023. dplyr: A Grammar of Data Manipulation. R package version 1.1.2.*