

Anthony Iury Moura Soares

**MODELAGEM E MONITORAMENTO DAS
TAXAS DE INCIDÊNCIA DE DENGUE DA
CIDADE DE NATAL/RN**

Natal/RN

11 de março de 2019

Anthony Iury Moura Soares

MODELAGEM E MONITORAMENTO DAS TAXAS DE INCIDÊNCIA DE DENGUE DA CIDADE DE NATAL/RN

Monografia de Graduação apresentada ao Departamento de Estatística do Centro de Ciências Exatas e da Terra da Universidade Federal do Rio Grande do Norte como requisito parcial para a obtenção do grau de Bacharel em Estatística.

Universidade Federal do Rio Grande do Norte

Centro de Ciências Exatas e da Terra

Departamento de Estatística

Orientador: Prof.^a Dr.^a Carla Almeida Vivacqua

Natal/RN

11 de março de 2019

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI
Catalogação de Publicação na Fonte. UFRN - Biblioteca Setorial Prof. Ronaldo Xavier de Arruda - CCET

Soares, Anthony Iury Moura.

Modelagem e monitoramento das taxas de incidência de dengue da cidade de Natal/RN / Anthony Iury Moura Soares. - 2019. 42f.: il.

Monografia (Bacharelado em Estatística) - Universidade Federal do Rio Grande do Norte, Centro de Ciências Exatas e da Terra, Departamento de Estatística. Natal, 2019.

Orientadora: Carla Almeida Vivacqua.

1. Análise de séries temporais - Monografia. 2. ARIMA - Monografia. 3. Gráfico de controle - Monografia. 4. EWMA - Monografia. 5. Resíduos - Monografia. 6. Epidemiologia - Monografia. I. Vivacqua, Carla Almeida. II. Título.

RN/UF/CCET

CDU 519.246.8

Anthony Iury Moura Soares

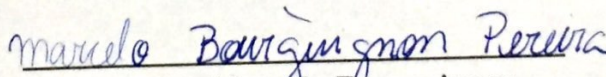
MODELAGEM E MONITORAMENTO DAS TAXAS DE INCIDÊNCIA DE DENGUE DA CIDADE DE NATAL/RN

Monografia de Graduação apresentada ao Departamento de Estatística do Centro de Ciências Exatas e da Terra da Universidade Federal do Rio Grande do Norte como requisito parcial para a obtenção do grau de Bacharel em Estatística.

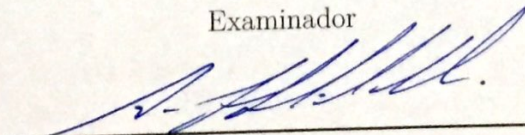
Trabalho aprovado. Natal/RN, 01 de Março de 2019 .



Prof.^a Dr.^a Carla Almeida Vivacqua
Orientadora



Prof. Dr. Marcelo Bourguignon
Pereira
Examinador



Prof. Dr. André Luís Santos de Pinho
Examinador

Natal/RN
28 de fevereiro de 2019

Soli Deo gloria.

Agradecimentos

A Deus, sem Ele nada poderia ser feito.

À minha família pelo apoio e incentivo, mostrando-me que o estudo é sempre recompensatório.

À minha namorada Suzana por todo incentivo, companhia e paciência em todos os momentos.

Aos meus grandes amigos Kaio Breno, Diêgo da Costa, Igor Antônio, Daniel Nóbrega, Luana Pontes, Félix Barbosa, Rodrigo Matheus, Érika Rayanne e Lucas Sales por todo o companheirismo e trocas de conhecimento durante o curso.

À Professora Carla Almeida Vivacqua, que participou diretamente da elaboração deste trabalho, pela orientação, paciência, dedicação, respeito, apoio e ensinamentos sempre direcionando e facilitando o aprendizado.

Aos meus colegas do estágio na Coordenadoria Estadual da Infância e Juventude, Tribunal de Justiça, por todo apoio e ensinamentos.

Aos Professores Marcelo Bourguignon e André Pinho por aceitarem fazer parte da banca.

Ao Departamento de Estatística, em especial aos professores, que tanto me ensinaram.

À Universidade Federal do Rio Grande do Norte e a CAPES por todo suporte, principalmente financeiro.

“Portanto, quer comais quer bebais, ou façais outra qualquer coisa, fazei tudo para glória de Deus.” 1 Coríntios 10:31

Resumo

No ano de 2016 houve uma epidemia do vírus da zika em todo o Brasil ocasionando diversos problemas clínicos e sociais. A zika se tornou um dos mais graves problemas de saúde pública nos últimos anos nos países em desenvolvimento e, conseqüentemente, o mosquito *Aedes aegypti* voltou a ganhar notoriedade pelo fato de ser o vetor de doenças virais, tais como: dengue, febre amarela, chikungunya e zika. Este trabalho tem o propósito de modelar a taxa dos casos de dengue coletados junto ao Centro de Controle de Zoonoses (CCZ) da cidade de Natal/RN entre os anos de 2000 e 2016, fazer previsões e monitorar as taxas de incidência de dengue, com o uso de técnicas de séries de temporais e gráficos de controle. O modelo ARMA(3,3) apresenta-se adequado aos dados analisados e é considerado um bom preditor para as primeiras 23 semanas de 2017. Um gráfico de controle construído a partir dos resíduos do modelo possibilita o monitoramento das taxas de incidência. Conclui-se que houve um decréscimo notório nas taxas de casos, proveniente, possivelmente, de campanhas de combate e prevenção à dengue.

Palavras-chave: Análise de Séries Temporais; ARIMA; Gráfico de Controle; EWMA; Resíduos; Epidemiologia; Dengue.

Abstract

In the year 2016 there was an epidemic of the zika virus throughout Brazil causing several clinical and social problems. Zika has become one of the most serious public health problems in recent years in developing countries, and consequently the *Aedes aegypti* mosquito has once again gained notoriety due to its being the vector of viral diseases such as: dengue, yellow fever, chikungunya and zika. This work aims to model the rates of dengue cases collected at the Centro de Controle de Zoonoses (CCZ) in the city of Natal/RN between the years 2000 and 2016, make predictions and monitor dengue incidence rates, with the use of time series techniques and control charts. The ARMA (3,3) model is adequate for the data analyzed and is considered a good predictor for the first 23 weeks of 2017. A control chart constructed from the residuals of the model allows the monitoring of incidence rates. It was concluded that there was a notable decrease in the rates of cases, possibly coming from campaigns to combat and prevent dengue.

Keywords: Time Series Analysis; ARIMA; Control Chart; EWMA; Residuals; Epidemiology; Dengue.

Lista de ilustrações

Figura 1 – Boxplot das taxas de casos da cidade de Natal/RN para cada semana.	31
Figura 2 – Série temporal das taxas de casos da cidade de Natal/RN para cada ano.	31
Figura 3 – Série temporal das taxas de casos da cidade de Natal/RN entre 2000 e 2016.	32
Figura 4 – Gráficos dos resíduos padronizados, ACF dos resíduos e resultados do teste Ljung-Box.	34
Figura 5 – Série temporal das taxas de casos observadas e previstas para o ano de 2017.	35
Figura 6 – Gráfico de Controle EWMA para os resíduos ($\lambda = 0.2, k = 2$).	37
Figura 7 – Gráfico de Controle EWMA para os resíduos ($\lambda = 0.5, k = 2$).	37
Figura 8 – Gráfico de Controle EWMA para os resíduos ($\lambda = 0.8, k = 2$).	38

Lista de tabelas

Tabela 1 – Estatísticas descritivas das taxas de dengue da cidade de Natal/RN, por semana, entre os anos 2000 e 2016.	30
Tabela 2 – Teste de Dickey-Fuller Aumentado.	33
Tabela 3 – Teste Ljung-Box.	33
Tabela 4 – Critério de informação AIC para os respectivos modelos da família ARIMA (p, d, q) ajustados.	33
Tabela 5 – Modelo ARIMA, coeficientes e Critério de Informação Akaike para Natal, Brasil.	33
Tabela 6 – Teste Ljung-Box.	34
Tabela 7 – Predições das 37 semanas de 2017 (intervalo de predição 95%) para a cidade de Natal.	35

Sumário

1	INTRODUÇÃO	12
1.1	Objetivos	13
2	CONCEITOS BÁSICOS	14
2.1	Séries Temporais	14
2.1.1	Estacionariedade	15
2.1.2	Modelos ARIMA	15
2.1.2.1	Modelos Autorregressivos	17
2.1.2.2	Modelos de Médias Móveis	18
2.1.2.3	Modelos Autorregressivos e de Médias Móveis	18
2.1.2.4	Modelos Autorregressivos Integrados de Médias Móveis	19
2.2	Controle Estatístico de Processos	20
2.2.1	Variabilidade e Causa	20
2.2.2	Gráficos de Controle	20
2.2.2.1	Gráficos de Controle para Variáveis	21
2.2.2.1.1	Gráfico \bar{X} e R	21
2.2.2.1.2	Gráfico de Controle EWMA	24
2.2.2.1.3	Gráficos de Controle para Resíduos de Modelos	25
3	MATERIAIS E MÉTODOS	28
3.1	Descrição dos Dados	28
3.2	Análise dos Dados	29
4	RESULTADOS	30
4.1	Análise Descritivas	30
4.2	Ajuste do Modelo ARIMA	32
4.3	Aplicação do Gráfico de Controle EWMA	36
5	CONCLUSÃO	39
	REFERÊNCIAS	40

1 Introdução

Recentemente, mais especificamente no ano 2016, houve uma grande epidemia do vírus da zika em todo o Brasil, ocasionando diversos problemas clínicos e sociais, como a microcefalia (G1, 2016) e a queda da natalidade no Brasil (GLOBO, 2017), respectivamente. Devido a todos esses problemas, a zika se tornou um dos mais graves problemas de saúde pública nos últimos anos nos países em desenvolvimento, sendo esta uma doença transmitida por hospedeiros por meio de seus vetores. O mosquito *Aedes aegypti* voltou a ganhar notoriedade pelo fato de ser o vetor de doenças virais como dengue, febre amarela e chikungunya, além de transmitir a zika. Este vetor é um dos mais preocupantes pelo grande impacto que causa em quase metade da população mundial.

A epidemiologia da dengue é modulada pela população suscetível dos seres humanos, a densidade do mosquito, o perfil dos sorotipos circulantes e as condições ambientais (BRADY et al., 2015; STANAWAY et al., 2016; CORTES et al., 2018). Não há tratamento antiviral específico para a dengue. O controle depende principalmente de intervenções de vigilância e controle integrados em áreas urbanas (GUZMAN; HARRIS, 2015; WILDER-SMITH et al., 2017).

Segundo Porta (2014), a vigilância de doenças infecciosas é definida como a notificação sistemática, monitoramento e análise de dados de casos, visando a prevenção e controle da saúde pública. O Brasil possui um dos mais abrangentes sistemas de vigilância da dengue (BRADY et al., 2015) e tem sido amplamente utilizado para descrever a epidemiologia (JUNIOR et al., 2005; TEIXEIRA et al., 2013; RUBERTO et al., 2015; COELHO et al., 2016), definir surtos (RUNGE-RANZINGER et al., 2008; BRADY et al., 2015) e a dinâmica da infecção em diferentes áreas urbanas do país (AMAKU et al., 2015; AMAKU et al., 2016). Frequentemente, os modelos matemáticos e estatísticos são usados para descrever a dinâmica da transmissão da dengue (SILAWAN et al., 2008; AMAKU et al., 2015; CHEN; COOK; LIM, 2015; AMAKU et al., 2016).

Diversas análises em séries temporais descreveram padrões de dengue em algumas cidades brasileiras, tais como: Rio de Janeiro (LUZ et al., 2008), Campinas (MARTINEZ; SILVA, 2011), Ribeirão Preto (MARTINEZ; SILVA; FABBRO, 2011), além de Recife e Goiania (CORTES et al., 2018). Já Johansson et al. (2016) avaliaram o desempenho das previsões de doenças infecciosas para o México, mostrando que os dados climáticos não melhoraram significativamente o modelo autorregressivo sazonal.

Na epidemiologia, há a necessidade não só de prever o futuro mas também intervir nos processos do presente e buscar uma solução para a melhora do problema em questão. Existe essa necessidade por se tratar, muitas das vezes, de um assunto de vida ou morte.

Porém, antes de se antever ou até mesmo de reconhecer o que pode interferir no processo é necessário um passo anterior o qual chamamos de modelagem. Portanto, a necessidade da modelagem no contexto de séries temporais se torna imprescindível para, por exemplo, ter o controle e prever casos de doenças transmitidas pelo vetor *Aedes aegypti*. Por meio dos modelos de séries temporais, pode-se avaliar os padrões de tendência, sazonalidade e são úteis para a previsão (CORTES et al., 2018).

Os gráficos estatísticos de controle de processo foram inicialmente desenvolvidos por Walter A. Shewhart, em 1924, para serem usados em processos industriais. E, desde então, diversos tipos de gráficos continuam a ser desenvolvidos. São alguns destes: o Gráfico das Médias Móveis Exponencialmente Ponderadas (EWMA) e o Gráfico de Controle das Somas Acumuladas (CUSUM), além dos gráficos multivariados. Além disso, ao passar dos anos, esses gráficos têm sido aplicados em diversas situações, sendo ainda escasso na área da saúde. Desse modo, o uso dos gráficos de controle podem ser úteis para o controle de doenças infecciosas.

1.1 Objetivos

Este trabalho, portanto, pretende modelar os dados de ocorrência de casos de dengue da cidade de Natal/RN entre os anos 2000 a 2016 por meio de um modelo ARIMA — que servirá para levar em conta a autocorrelação presente nos dados —, prever a taxa de incidência de dengue nas 37 primeiras semanas de 2017 e, posteriormente, aplicar o gráfico de controle EWMA aos resíduos do modelo para, assim, se ter o controle das taxas de incidências de dengue. Uma vez que o modelo é adequadamente ajustado, os resíduos se comportam de forma independente e identicamente distribuídos (PSARAKIS; PAPALEONIDA, 2007).

2 Conceitos Básicos

Para um melhor entendimento acerca das metodologias adotadas neste trabalho, este capítulo apresenta conceitos básicos sobre Séries Temporais e Gráficos de Controle.

2.1 Séries Temporais

Pode-se definir uma série temporal por uma coleção de observações de uma variável aleatória, obtidas sequencialmente, em instantes de tempo. Esta possui como principal característica a presença de dependência entre as observações (BOX; JENKINS; REINSEL, 1994). Enquanto em modelos de regressão, por exemplo, a ordem das observações é irrelevante para a análise, em séries temporais a ordem dos dados é crucial. Ao realizar uma análise de uma série temporal estamos interessados em descrever, interpretar e realizar previsões acerca de fenômenos de interesse.

A construção dos modelos para séries temporais pode ser feito a partir de dois diferentes enfoques. Quando a análise é feita no domínio temporal, utilizam-se modelos paramétricos, tendo um número finito de parâmetros. Caso a análise ocorra no domínio de frequências são propostos modelos não-paramétricos. O modelo Autorregressivo e de Médias Móveis, definido e detalhado na Seção 2.1.2.3, é um exemplo de modelo paramétrico e será utilizado neste trabalho.

Segundo Morettin e Tolo (2006), dada a série temporal Y_1, \dots, Y_t observada nos instantes $1, \dots, t$, consideram-se os seguintes objetivos na análise de séries temporais:

- i. Investigar o mecanismo gerador da série temporal;
- ii. Fazer previsões de valores futuros da série;
- iii. Descrever apenas o comportamento da série, o que envolve a construção do gráfico, verificação da existência de tendências, ciclos e variações sazonais, construção de histogramas e diagramas de dispersão;
- iv. Procurar periodicidades relevantes nos dados.

Segundo os mesmos autores, tais modelos devem ser simples e parcimoniosos (o número de parâmetros deve ser o menor possível sem o comprometimento da qualidade do modelo), e sua manipulação não deve apresentar dificuldades.

2.1.1 Estacionariedade

Morettin e Tolo (2006) dizem que uma das suposições mais frequentes que se faz a respeito de uma série temporal é a respeito da sua estacionariedade, ou seja, ela se desenvolve aleatoriamente no tempo, ao redor de uma média constante, refletindo alguma forma de equilíbrio estável. Na maior parte das séries encontradas na prática apresenta alguma forma de não-estacionariedade, na maioria dos procedimentos de análise estatística de séries temporais, supõe-se, porém, que estas sejam estacionárias.

Uma série temporal Y_t diz-se fracamente estacionária se e somente se:

- i. $E(Y_t) = \mu, \quad t \in T;$
- ii. $\text{Var}(Y_t) = \sigma_y^2 < \infty, \quad t \in T;$
- iii. $\gamma_{t,s} = \gamma_{t-s}, \quad t = s,$

em que $E(Y_t)$ é a média da série, $\text{Var}(Y_t)$ é a variância e $\gamma_{t,s}$ é a autocovariância entre Y_t e Y_s .

É comum encontrar, na prática, séries que não são estacionárias. Desta forma, é necessário transformar os dados originais. Segundo Morettin e Tolo (2006), a transformação mais usual consiste em tomar diferenças sucessivas da série original, até se obter uma série estacionária, sendo mais comum o uso apenas de uma diferença, isto é,

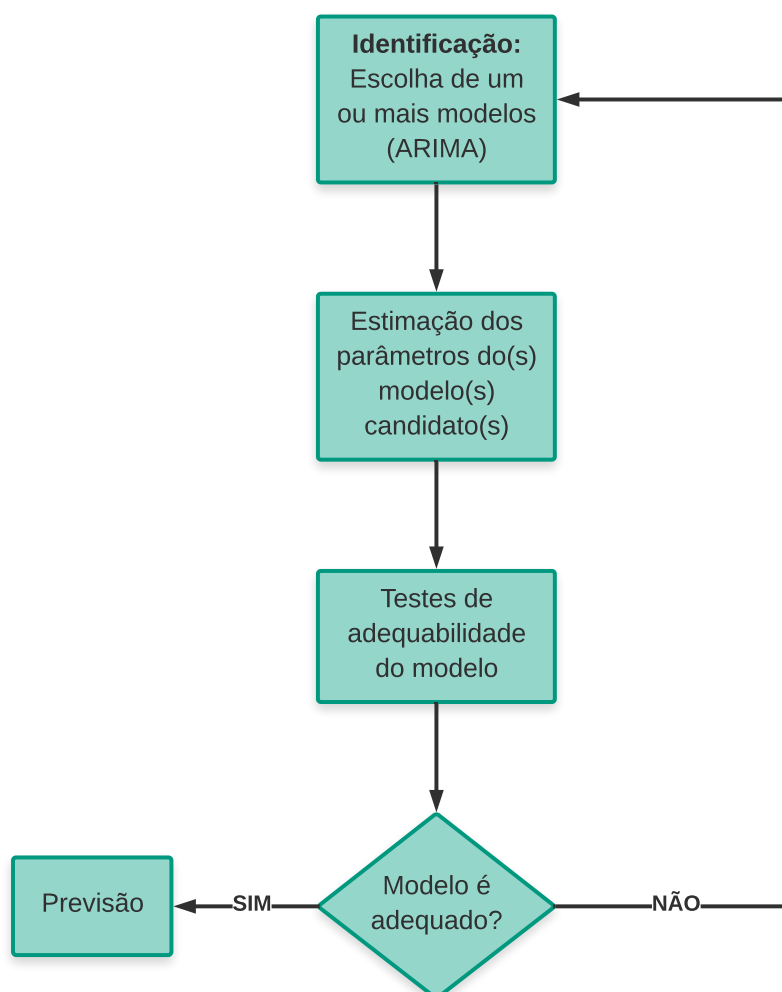
$$\Delta Y_t = Y_t - Y_{t-1}. \quad (2.1)$$

2.1.2 Modelos ARIMA

Morettin e Tolo (2006) afirmam que uma metodologia conhecida como abordagem de Box e Jenkins (1970) é amplamente utilizada na análise de modelos paramétricos. Ela consiste no ajuste de modelos autorregressivos integrados de médias móveis, $\text{ARIMA}(p, d, q)$, a um conjunto de dados.

Os modelos $\text{ARIMA}(p, d, q)$ são resultados da combinação de três filtros: o componente Autorregressivos (AR), de integração (I) e de Médias Móveis (MA). Para a modelagem de determinada série temporal, tanto se pode utilizar os três, quanto apenas um subconjunto destes (FAVA, 2000). A ordem de integração que é representada pela letra d do modelo ARIMA corresponde à possível diferenciação tomada na série para torná-la estacionária. Caso o modelo não apresente esta componente, então poderá ser um modelo Autorregressivo (AR), um modelo de Médias Móveis (MA), ou então uma junção destes dois que é representado pela sigla ARMA.

A construção deste modelo tem como base um ciclo iterativo, cujos estágios são dados pelo fluxograma abaixo.



Na etapa 1 há a identificação do modelo. Neste estágio encontra-se os valores apropriados das ordens p , d e q do(s) modelo(s). Há identificação de alguns modelos, com base na análise de autocorrelação, autocorrelação parcial e outros critérios como o AIC (Critério de Informação de Akaike), AICc (Critério de informação de Akaike corrigido) e o BIC (Critério de Informação Bayesiano).

Na etapa 2 há estimação. Depois de identificados os valores apropriados de p , d e q , neste estágio deve-se estimar os parâmetros dos termos autorregressivos (AR) e dos termos de média móvel (MA) incluídos no modelo. Com a utilização dos pacotes estatísticos, a estimação desses parâmetros torna-se sistemática.

Na etapa 3 há diagnóstico do modelo ajustado através da análise de resíduos. Uma vez escolhido o(s) modelo(s) ARIMA e tendo seus parâmetros conhecidos, deve-se verificar se o modelo selecionado ajusta-se bem aos dados. Para verificar, é necessário analisar se os resíduos são ruídos brancos com média zero e variância constante, se forem, poderemos

aceitar o modelo bem ajustado aos dados; caso contrário, é necessário recomeçar as etapas. Uma ferramenta para diagnóstico comumente usada para testar a falha do ajuste de um modelo de série temporal é o teste de Ljung e Box (1978). Este teste é aplicado aos resíduos de uma série temporal após o ajustamento de um modelo ARIMA (p, d, q) aos dados.

Na etapa 4 há previsão. Uma das razões da popularidade da modelagem por meio da abordagem Box e Jenkins é o grande sucesso na previsão, podendo ser pontual ou intervalar.

Pode-se, então, dizer que a metodologia Box-Jenkins é um processo iterativo. Caso o modelo não seja adequado, o ciclo é repetido, voltando-se à fase de identificação. Um procedimento que é utilizado muitas vezes é identificar não só um único modelo, mas alguns modelos que serão então estimados e verificados.

2.1.2.1 Modelos Autorregressivos

De acordo com esse modelo, o processo Y_t é descrito apenas por seus valores passados e pelo ruído branco a_t , a ser expresso por AR(p)

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t,$$

sendo μ o termo independente do processo, $\phi = (\phi_1, \phi_2, \dots, \phi_p)$ é o vetor de coeficientes autorregressivos e a_t é o ruído branco no instante t .

Neste caso, o valor de Y_t depende dos p valores anteriores da série e do ruído branco no instante t .

Se definirmos o operador autorregressivo estacionário de ordem p :

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p,$$

então, sendo $\mu = 0$ e B o operador defasagem, o processo pode ser representado por

$$\phi(B)Y_t = a_t.$$

O caso mais simples AR(1) ocorre quando $p = 1$ e o termo independente é nulo

$$Y_t = \phi_1 Y_{t-1} + a_t.$$

A condição de estacionariedade de um processo AR(p) é equivalente à equação característica $\phi(B) = 0$ tenha raízes fora do círculo unitário.

A função de autocorrelação de um processo autorregressivo tem o comportamento de acordo com uma mistura de polinômios, exponenciais e senóides amortecidas.

2.1.2.2 Modelos de Médias Móveis

Um processo de médias móveis de ordem q , a ser denotado por $MA(q)$, pode ser escrito, de acordo com Morettin e Tolo (2006), da seguinte maneira:

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q},$$

sendo μ o termo independente do processo, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_q)$ é o vetor de coeficientes de médias móveis e a_t é o ruído branco no instante t .

Se definirmos o operador médias móveis de ordem q :

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q,$$

então, o processo pode ser representado por

$$Y_t = \theta(B)a_t.$$

Considerando $\mu = 0$ e $q = 1$, tem-se o modelo mais simples de um processo de médias móveis $MA(1)$:

$$Y_t = a_t - \theta_1 a_{t-1}.$$

Para os parâmetros θ 's não há restrição para que o processo seja estacionário.

De acordo com Box, Jenkins e Reinsel (1994), um processo $MA(q)$ tem a função de autocorrelação parcial com o comportamento similar à função de autocorrelação de um processo $AR(p)$, isto é, apresenta um comportamento determinado por exponenciais e/ou senóides amortecidas.

2.1.2.3 Modelos Autorregressivos e de Médias Móveis

O modelo ARMA é uma combinação dos dois modelos anteriores: apresenta termos autorregressivos e termos de médias móveis, tendo a seguinte forma:

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q},$$

em que μ é o termo independente, o vetor ϕ representa o vetor de coeficientes autorregressivos, θ é o vetor de coeficientes de médias móveis e a_t é o ruído branco no instante t .

Substituindo, com $\mu = 0$, pelo operador autorregressivo e de médias móveis, tem-se

$$\phi(B)Y_t = \theta(B)a_t.$$

O modelo mais simples é o ARMA(1, 1), com $\mu = 0$, que tem a seguinte expressão:

$$Y_t = \phi_1 Y_{t-1} - \theta_1 a_{t-1} + a_t.$$

O processo ARMA consiste na combinação de processos AR e MA, desta forma o processo é estacionário quando todas as raízes de $\phi(B) = 0$ estão fora do círculo unitário.

Ainda pode-se verificar que se $q > p$, a função de autocorrelação (f.a.c) consiste numa mistura de exponenciais e/ou de senóides amortecidas. No entanto, se $q = p$, os primeiros $q - p - 1$ valores não seguirão este padrão (BOX; JENKINS; REINSEL, 1994; MORETTIN; TOLOI, 2006).

2.1.2.4 Modelos Autorregressivos Integrados de Médias Móveis

Os modelos de séries temporais estudados até aqui são apropriados para descrever séries (fracamente) estacionárias, isto é, apresentam de forma geral média e variância constantes e covariância invariante no tempo. Na prática, no entanto, muitas séries temporais são não estacionárias, necessitando, portanto, serem diferenciadas d vezes.

Sendo Y_t não estacionária, mas $\nabla Y_t = Y_t - Y_{t-1}$ é estacionária, então, diz-se que Y_t é integrada de ordem 1. Caso a mesma série Y_t precise de duas diferenças para se tornar estacionária, então é representada

$$\nabla^2(Y_t) = (\nabla^2 Y_t) = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}),$$

diz-se, então, que Y_t é integrada de ordem 2.

Diferenciando Y_t d vezes e aplicando o modelo ARMA(p, q), diz-se que a série temporal original é ARIMA(p, d, q), ou seja, é uma série temporal autorregressiva integrada de médias móveis, sendo p o número de termos autorregressivos, d o número de vezes que a série deve ser diferenciada antes de tornar-se estacionária e que q o número de termos de média móvel. Esse modelo pode ser representado na forma

$$\phi(B) \nabla^d Y_t = \theta(B)a_t,$$

em que $\phi(B)$ é o operador autorregressivo de ordem p , $\theta(B)$ é o operador médias móveis de ordem q , a_t ruído branco, d número de diferenças, $1 - B$ é operador diferença.

2.2 Controle Estatístico de Processos

O Controle Estatístico de Processos (CEP) é um método quantitativo para monitorar um processo repetitivo a fim de observar se este está operando adequadamente. Este método utiliza coletas de dados de processos e compara as medições atuais com os determinados medidores de desempenho do processo. De acordo com Montgomery (2009), o CEP é uma poderosa gama de ferramentas que tem como intuito a solução de problemas que são úteis para obter a estabilidade do processo e melhorar sua capacidade através da redução de sua variabilidade.

Além do mais, de acordo com o mesmo autor, o principal objetivo do CEP é detectar de forma rápida a ocorrência de causas especiais, que fazem o processo sair de controle, de forma que a investigação e a correção sejam executadas o mais cedo possível, antes que muitas unidades sejam produzidas, no contexto industrial, e antes que mais pessoas fiquem doentes, no contexto epidemiológico.

2.2.1 Variabilidade e Causa

Os gráficos de controle foram inicialmente produzidos por Shewhart, em 1924, com o objetivo de que fossem usados em processos industriais. De acordo com ele, todo e qualquer processo, por mais bem projetado e bem controlado que seja, possui em sua variabilidade um componente intrínseco ao processo e impossível de ser eliminado, que é fruto de uma série de causas aleatórias. Quando o processo apresenta apenas essa variabilidade intrínseca, diz-se, então, que ele está sob controle.

Todos os processos, segundo Costa (2008), também estão sujeitos à ocorrência ocasional de perturbações maiores, chamadas causas especiais, fazendo com que os pontos plotados tendam a sair dos limites de controle. Portanto, uma causa especial significa um modo de operação anormal do processo e pode ser corrigido ou até mesmo ser eliminado. Segue-se que se um processo que opera com a presença de causas especiais, segundo Montgomery (2009), é denominado fora de controle. Exemplos de causas especiais são: máquinas mal ajustadas ou controladas, erros do operador, matéria-prima defeituosa.

2.2.2 Gráficos de Controle

A principal ferramenta do controle estatístico de processo é o gráfico de controle, que é baseado no monitoramento periódico de amostras de uma determinada característica de qualidade de interesse.

Esta ferramenta pode ser utilizada para monitorar um processo através de uma estatística pré-definida, como por exemplo a média amostral ou a amplitude amostral. Os principais elementos de um gráfico de controle são: Limite Superior de Controle (LSC), Limite Inferior de Controle (LIC), Linha média (LM). O valor médio para o processo é calculado com base em desempenhos passados, já os limites de controle dependem do tamanho amostral que foi utilizado. Em geral, esses limites de controle são estabelecidos em três desvios-padrão (3σ) acima e abaixo da média. De acordo com Montgomery (2009), os limites são escolhidos de tal forma que, caso o processo esteja sob controle, praticamente todos os pontos amostrais estarão entre eles.

Os gráficos de controle podem ser classificados em dois tipos. Caso a característica de qualidade é mensurável, ela é chamada de variável. Nestes casos, é comum medir a tendência central e a variabilidade desta variável, construindo os respectivos gráficos de controle para variáveis. Os gráficos \bar{X} são comuns para estudar a tendência central, já os gráficos baseados na amplitude ou no desvio padrão da amostra são utilizados para estudar a variabilidade. Contudo, diversas características de qualidade não são passíveis de medição, pois envolvem certo grau de subjetividade. Neste caso, utiliza-se gráficos de controle para atributos.

2.2.2.1 Gráficos de Controle para Variáveis

Segundo Montgomery (2009), uma variável é uma característica de qualidade digna de mensuração numérica. Para lidar com esse tipo de característica é comum o monitoramento de sua média e variabilidade. Para o controle da média do processo é normalmente realizado o gráfico de controle \bar{X} , enquanto o gráfico de controle R é utilizado para monitorar a variabilidade.

2.2.2.1.1 Gráfico \bar{X} e R

O gráfico de controle \bar{X} , conhecido também por gráfico de Shewhart, foi proposto pelo engenheiro e estatístico Walter A. Shewhart na década de 30. Este gráfico se baseia em duas suposições:

- i São retiradas m amostras de tamanho n , em que cada uma das m amostras são independentes entre si e dentro da amostra também há independência.
- ii Assume-se normalidade para a variável monitorada. A suposição de normalidade é um pouco mais flexível, tendo em mente que se a variável de interesse não seguir distribuição normal, mas a amostra for suficientemente grande, utiliza-se, pelo teorema central do limite, a aproximação normal.

Supondo uma variável com distribuição normal de média μ_0 e desvio padrão σ_0 conhecidos. Se x_1, x_2, \dots, x_n é uma amostra de tamanho n , então a média desta amostra é:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}.$$

Sabe-se que \bar{X} tem distribuição normal de média $\mu_{\bar{X}}$ e desvio padrão $\sigma_{\bar{X}} = \sigma / \sqrt{n}$. Assim, pode-se usar os seguintes limites de controle:

$$\begin{aligned} LSC &= \mu_{\bar{X}} + k\sigma_{\bar{X}}, \\ LM &= \mu_{\bar{X}}, \\ LIC &= \mu_{\bar{X}} - k\sigma_{\bar{X}}. \end{aligned} \tag{2.2}$$

Shewhart propôs a utilização de $k = 3$, pois afirmava que isso iria prevenir intervenções desnecessárias na produção, tornando o gráfico útil e eficiente para as condições de uma indústria de produção. Para essa situação em que foi utilizado 3σ , a probabilidade de um ponto qualquer se situar entre os limites de controle é de 99,73% e, conseqüentemente, existe um risco igual a 0,27% de um ponto pertencer à distribuição normal estar fora dos limites de controle.

Na prática, contudo, em geral é comum não conhecermos a média (μ) e o desvio padrão (σ) da variável. Neste caso, esses parâmetros devem ser estimados a partir das amostras retiradas quando o processo está sob controle. Supondo m amostras, cada uma contendo n observações na característica a ser mensurada. Sejam $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ as médias de cada amostra. Então, o melhor estimador de μ , isto é, a média do processo, pode ser estimada por

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m}{m}.$$

Para construir os limites de controle é necessário estimar também o parâmetro σ (desvio padrão). O estimador do desvio padrão pode ser obtido tanto a partir das amplitudes quanto dos desvios padrões das m amostras. Considerando, então, o primeiro caso, tem-se que se x_1, x_2, \dots, x_m é uma amostra de tamanho n , então a amplitude R é obtida a partir da diferença entre a maior e a menor observação (MONTGOMERY, 2009).

$$R = x_{max} - x_{min}.$$

Sejam R_1, R_2, \dots, R_m as amplitudes das m amostras. Tem-se, então, que a amplitude média é dada por

$$\bar{R} = \frac{R_1 + R_2 + \dots + R_m}{m}.$$

Desta forma, tem-se os limites de controle para o gráfico \bar{x}

$$LSC = \bar{\bar{x}} + A_2 \bar{R},$$

$$LM = \bar{\bar{x}},$$

$$LIC = \bar{\bar{x}} - A_2 \bar{R},$$

em que os valores de A_2 são tabelados em Montgomery (2009).

É possível também estimar σ a partir da equação

$$\hat{\sigma} = \frac{\bar{R}}{d_2},$$

sendo d_2 é um valor tabelado em Montgomery (2009).

Com isso, tem-se os limites de controle para o gráfico

$$LSC = \bar{\bar{x}} + \frac{3}{d_2} \frac{\bar{R}}{n},$$

$$LM = \bar{\bar{x}},$$

$$LIC = \bar{\bar{x}} - \frac{3}{d_2} \frac{\bar{R}}{n}.$$

Pode-se observar a seguinte relação entre os valores tabulados A_2 e d_2 .

$$A_2 = \frac{3}{d_2} \frac{1}{n}.$$

De acordo com Montgomery (2009), os limites de controle para o gráfico R são por

$$LSC = D_4 \bar{R},$$

$$LM = \bar{R},$$

$$LIC = D_3 \bar{R},$$

em que D_3 e D_4 são constantes e tabelados para diversos valores de n em Montgomery (2009).

Montgomery (2009) sugere a determinação analítica dos limites de controle para monitoramento da média, conforme a Equação (2.2). No entanto, sabe-se que em alguns casos a distribuição da variável monitorada não é conhecida. Nestas situações, são utilizados algoritmos de simulação para determinar os limites de controle.

2.2.2.1.2 Gráfico de Controle EWMA

O gráfico de controle da Média Móvel Ponderada Exponencialmente, *Exponentially Weighted Moving Average* (EWMA), em inglês, é muito sensível em detectar pequenas e moderadas alterações dos parâmetros do processo e, portanto, é outra alternativa além do gráfico de controle Shewhart. O gráfico EWMA apresenta desempenho bastante similar ao gráfico CUSUM e é geralmente utilizado com observações individuais.

Neste gráfico, são plotados valores da estatística z_i :

$$z_i = \lambda x_i + (1 - \lambda)z_{i-1},$$

em que, $0 < \lambda < 1$ é a constante de suavização, x_i é o valor da série temporal e o valor inicial z_0 é o valor alvo da média:

$$z_0 = \mu_0.$$

À medida que λ se aproxima de 1, mais peso é dado à média mais recente e à medida que λ diminui, mais peso é dado para os dados mais antigos. Quando $\lambda = 0$, então

$$z_i = z_0.$$

Note que z_i é uma média ponderada de todas as observações anteriores, substituindo, tem-se:

$$\begin{aligned} z_i &= \lambda x_i + (1 - \lambda)z_{i-1} \\ &= \lambda x_i + (1 - \lambda)[\lambda x_{i-1} + (1 - \lambda)z_{i-2}] \\ &= \lambda x_i + \lambda(1 - \lambda)x_{i-1} + (1 - \lambda)^2 z_{i-2}. \end{aligned}$$

Substituindo de forma recursiva, obtém-se:

$$z_i = \lambda \sum_{j=0}^{i-1} (1 - \lambda)^j x_{i-j} + (1 - \lambda)^i z_0.$$

Portanto, tendo em mente que o gráfico EWMA pode ser considerado como uma média móvel ponderada de todas as observações passadas e corrente, o gráfico de EWMA é insensível à hipótese de normalidade, e é o ideal para ser usado em observações individuais.

A variância de z_i é dada por

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2i}].$$

Portanto, o gráfico de controle EWMA é construído com base nos pares ordenados (i, z_i) . Uma observação é considerada fora de controle, isto é, quando há alteração na média do processo, quando z_i não pertence ao seguinte limites de controle

$$\begin{aligned} LSC &= \mu_0 + k\sigma \sqrt{\left(\frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2i}]}, \\ LM &= \mu_0, \\ LIC &= \mu_0 - k\sigma \sqrt{\left(\frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2i}]}, \end{aligned} \tag{2.3}$$

em que o fator k é a largura dos limites de controle, λ é o fator de ponderação, μ_0 é um valor objetivo que pode ser substituído pela média do processo e σ é o desvio padrão do processo.

Note que à medida que o valor de i aumenta, a quantidade $[1 - (1 - \lambda)^{2i}]$ tende para 1, desta forma, os limites de controle tendem assintoticamente para

$$\begin{aligned} LSC &= \mu_0 + k\sigma \sqrt{\frac{\lambda}{2 - \lambda}}, \\ LM &= \mu_0, \\ LIC &= \mu_0 - k\sigma \sqrt{\frac{\lambda}{2 - \lambda}}. \end{aligned}$$

Para valores pequenos de i , Montgomery (2009) recomenda o uso dos limites de controle exato em (2.3), haja vista que o gráfico de controle em questão terá um melhor desempenho.

2.2.2.1.3 Gráficos de Controle para Resíduos de Modelos

Como ressaltado por Montgomery (2009), é comum nos processos industriais a suposição de que as observações são independentes não ser satisfeita. Uma das explicações

se dá pelo fato de que em sistemas de monitoramento do processos as observações são coletadas em intervalos de tempos pequenos para essa suposição, contribuindo, assim, para que as observações sejam autocorrelacionadas ao longo do tempo.

Devido a isso, um das abordagens usuais é ajustar um modelo de séries temporais às observações. Com isso, levará em conta a autocorrelação presente nos dados e, em seguida e, posteriormente, aplicar gráficos de controles tradicionais ou modificados aos resíduos (MONTGOMERY, 2009). É plausível o uso de gráficos controles residuais, pois, uma vez que o modelo seja bem ajustado, possuem resíduos que se comportam como uma variável aleatória independente e identicamente distribuídas (PSARAKIS; PAPALEONIDA, 2007) e conseqüentemente, os pressupostos necessários para a construção dos gráficos de controle serão atendidas. Diversos trabalhos, como de Henning et al. (2014), Pedrini e Caten (2008) e Montgomery e Mastrangelo (1991), são encontradas algumas aplicações de modelo de séries temporais com o intuito de levar em conta a dependência nas observações.

Um dos gráficos bastante usuais para o monitoramento dos resíduos dos modelos é o gráficos EWMA. Lu e Reynolds (1999) estudaram o desempenho dos gráficos EWMA para o monitoramento da média e da variância dos processos autocorrelacionados. English et al. (2000) e Koehler, Marks e O'connell (2001) estudaram o comportamento dos gráficos EWMA para resíduos e para observações autocorrelacionadas. Outra alternativa para o monitoramento dos resíduos padronizados dos modelos ARIMA é apresentada por Alwan e Roberts (1988).

Conforme a Equação (2.4), o resíduo de um modelo é definido como sendo a diferença entre o valor observado da variável resposta e o valor ajustado pelo modelo,

$$e_i = y_i - \hat{y}_i. \quad (2.4)$$

O desvio-padrão dos resíduos do modelo é estimado pela raiz quadrada do quadrado médio dos erros (QMR), conforme a Equação (2.5).

$$\hat{\sigma}_e = \sqrt{QMR} = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n - k}}, \quad (2.5)$$

em que k é o número de parâmetros do modelo.

O resíduo padronizado do modelo é obtido através da Equação abaixo:

$$z_i = \frac{e_i}{\hat{\sigma}_e}.$$

Desta forma, será possível haver a aplicação de diferentes tipos de gráficos de controle desde que nas fórmulas apresentadas dos limites substituam os valores das observações pelos respectivos resíduos.

3 Materiais e Métodos

Neste capítulo serão apresentados a descrição dos dados e os procedimentos metodológicos deste trabalho com uso de técnicas estatísticas.

3.1 Descrição dos Dados

A Secretaria Municipal de Saúde (SMS) de Natal atua no Centro de Informações Estratégicas de Vigilância em Saúde (CIEVS/Natal), como uma estratégia para fortalecimento da capacidade de alerta e resposta do Sistema Único de Saúde (SUS) frente às emergências em Saúde Pública.

Em sintonia com a SMS de Natal, a CIEVS/Natal identifica as emergências epidemiológicas de modo contínuo e sistemático, por meio da notificação telefônica, eletrônica e informações provenientes ou captadas de outras fontes. Os dados referentes ao número de casos de doenças de dengue ocorridos entre os anos de 2000 a 2017 (até a 37^a semana) foram obtidos no Centro de Controle de Zoonoses da cidade (CCZ) de Natal/RN. No entanto, a partir do ano de 2015, esses dados passaram a ter a junção dos números de casos de dengue e zika. A sua série histórica, é distribuída espacialmente pelos bairros da cidade com representação semanal de acordo com o calendário epidemiológico referente aos casos ocorridos em cada um deles.

As informações epidemiológicas são verificadas pelos agentes de forma minuciosa, realizando, caso haja dúvidas, visitas nas residências dos pacientes para confirmar as informações da localização do bairro na qual houve a ocorrência da notificação. O mesmo procedimento ocorre com relação às informações que chegam em forma de documentos, quanto os que precisam ser averiguados em campo.

Em relação aos dados, foi constatado que havia dados faltantes em dois bairros de Natal, a saber, os bairros Salinas e o bairro do Planalto. O primeiro apresentou dados faltantes entre o período de 2000 a 2006 e que foi constatado que havia bastante zeros em sua distribuição entre os anos de 2007 e 2016, por isso foi considerado como zero os dados faltantes. O mesmo aconteceu para o bairro Planalto.

Tendo em mente que a população de Natal ao passar dos anos aumentou, foi-se necessário lidar não com a quantidade de casos e sim com a taxa destes por 100.000 habitantes. A estimativa do total da população foi retirado do site Datasus¹.

¹ Disponível em: <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?ibge/cnv/poprn.def>. Acessado em: 10 set. 2018.

3.2 Análise dos Dados

Após a obtenção dos dados, a elaboração e revisão do banco de dados, uma análise exploratória foi realizada sobre os dados de dengue, entre o período de 2000 a 2016 para identificar as características da distribuição dos casos. Verificou-se a distribuição semanal das taxas de casos de dengue na cidade de Natal, no Rio Grande do Norte.

Foram ajustados modelos de séries temporais usando a metodologia de Box e Jenkins e, posteriormente, a construção do gráfico de controle EWMA nos resíduos da série.

Para a identificação do modelo, foi usada a função *auto.arima()* do pacote *forecast* (HYNDMAN; KHANDAKAR, 2008) do *Software R Core Team* (2017) versão 3.4.4. Esta função pesquisa por meio de combinações de ordens dos parâmetros e seleciona o conjunto que otimiza os critérios de ajuste do modelo. Além disso, permite que o usuário especifique a ordem máxima para p, d, q , que é definida como cinco por padrão.

Para eleger o melhor dentre os possíveis modelos identificados, foi utilizado o critério do menor valor de AIC (AKAIKE, 1974), realizou-se a validação diagnóstica considerando a distribuição dos resíduos padronizados. Aplicou-se o teste de Ljung-Box aos resíduos para o diagnóstico do modelo; os resíduos devem ser equivalentes ao ruído branco (BOX; PIERCE, 1970; LJUNG; BOX, 1978); (LJUNG; BOX, 1978).

Utilizou-se o modelo final ARIMA para prever as taxas de casos de dengue para as primeiras 37 semanas de 2017, com intervalo de previsão de 95% (IC 95%). E, com base nesse modelo procedeu-se com a aplicação do Gráfico das Médias Móveis Exponencialmente Ponderadas (EWMA) para os resíduos para, assim, ter o controle das taxas de incidência de dengue.

4 Resultados

Neste Capítulo é mostrado de forma sucinta os resultados obtidos. Primeiramente, realizou-se análises descritivas, posteriormente, ajustou-se o modelo ARIMA aos dados e aplicou gráficos de controle aos resíduos do modelo.

4.1 Análise Descritivas

Foi realizada uma análise descritiva dos dados. Na Tabela 1 constam as taxas de dengue por semana, entre os anos 2000 e 2016, acompanhados da média, desvio padrão e coeficiente de variação.

Pode-se, então, observar, por meio da Figura 1 e Tabela 1, que as maiores médias de incidência da doença e os maiores coeficientes de variação ocorrem entre as semanas 10 e 24. Enquanto os menores valores para essa medida ocorrem entre as semanas 40 e 52, isto é, à medida que se encaminha para o fim do ano as taxas de incidência tende a diminuir.

Tabela 1 – Estatísticas descritivas das taxas de dengue da cidade de Natal/RN, por semana, entre os anos 2000 e 2016.

Semana	Média	Desvio Padrão	Coefficiente de Variação	Semana	Média	Desvio Padrão	Coefficiente de Variação
1	7,28	6,26	0,86	27	22,81	16,68	0,73
2	8,43	7,52	0,89	28	20,59	16,79	0,82
3	10,11	10,34	1,02	29	20,24	14,76	0,73
4	12,01	12,74	1,06	30	17,52	15,53	0,89
5	15,69	17,30	1,10	31	16,70	13,28	0,80
6	18,25	21,54	1,18	32	15,09	10,98	0,73
7	22,44	30,70	1,37	33	13,33	10,09	0,76
8	29,12	42,31	1,45	34	10,74	7,73	0,72
9	29,69	37,57	1,27	35	9,06	5,90	0,65
10	35,61	44,37	1,25	36	7,96	5,38	0,68
11	36,26	35,28	0,97	37	7,52	5,22	0,69
12	35,50	31,89	0,90	38	6,86	4,49	0,65
13	34,89	31,66	0,91	39	5,94	3,86	0,65
14	40,57	38,42	0,95	40	5,58	4,01	0,72
15	40,61	41,17	1,01	41	5,14	3,60	0,70
16	39,59	39,59	1,00	42	5,42	3,37	0,62
17	39,73	37,80	0,95	43	4,79	3,65	0,76
18	39,65	40,52	1,02	44	4,75	3,28	0,69
19	38,98	38,42	0,99	45	4,69	3,17	0,68
20	39,65	41,48	1,05	46	4,42	3,08	0,70
21	36,58	38,92	1,06	47	4,82	3,30	0,69
22	36,27	34,32	0,95	48	5,06	3,21	0,64
23	33,07	29,90	0,90	49	4,60	3,28	0,71
24	30,91	27,08	0,88	50	5,13	3,56	0,69
25	27,19	22,95	0,84	51	5,10	3,75	0,73
26	23,50	17,29	0,74	52	4,82	3,91	0,81

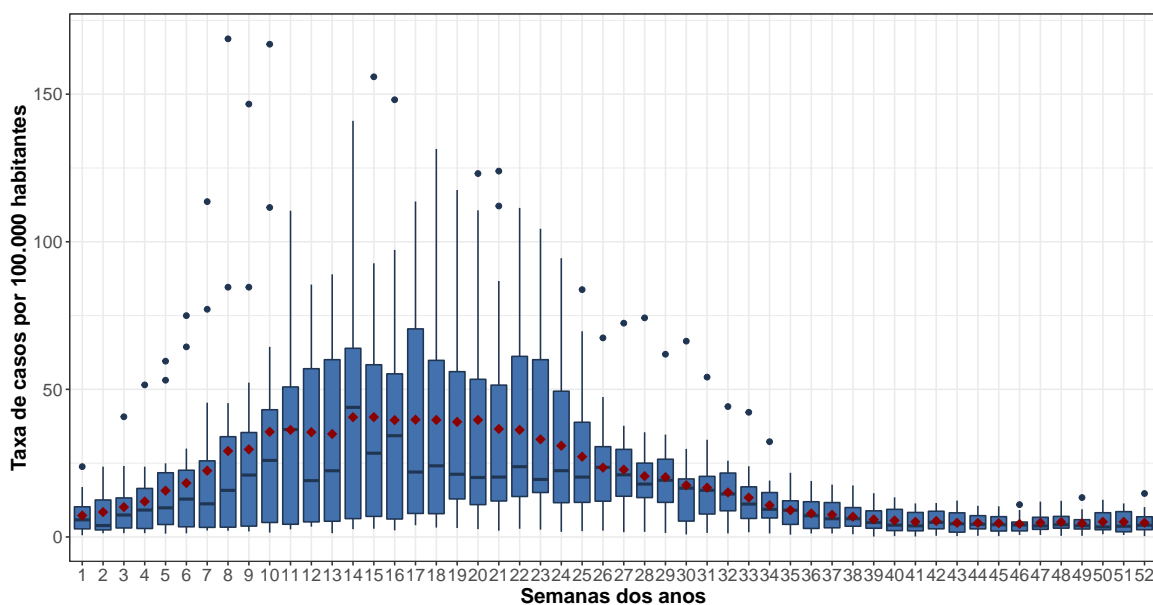


Figura 1 – Boxplot das taxas de casos da cidade de Natal/RN para cada semana.

A Figura 2 possibilita a visualização das série temporal por ano. Percebemos, claramente, que no ano de 2016 houve a maior epidemia dentre os anos em questão, conforme comentado na Seção 1.

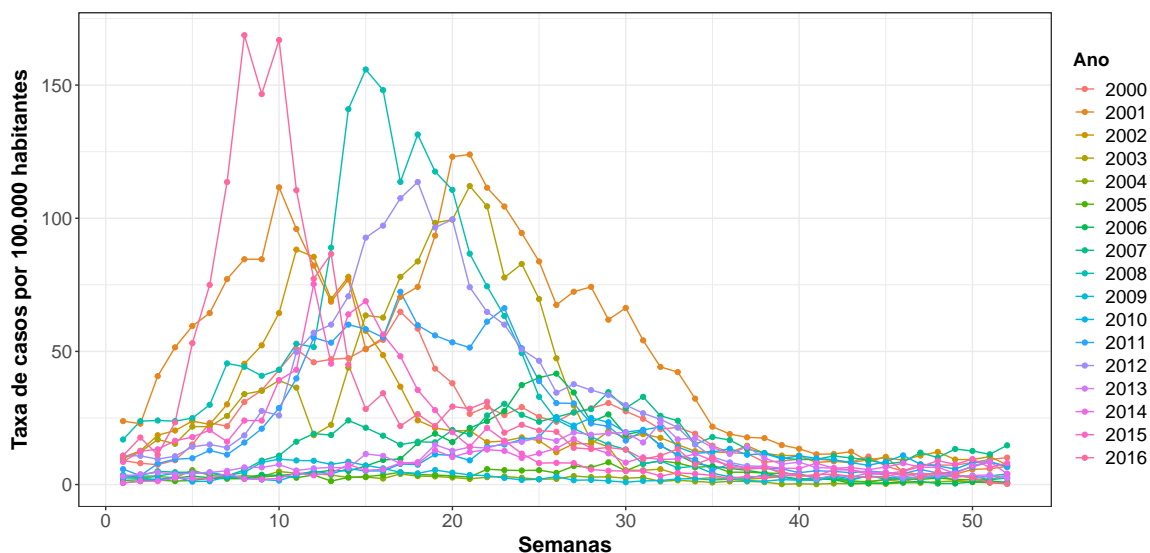


Figura 2 – Série temporal das taxas de casos da cidade de Natal/RN para cada ano.

Foi-se necessário verificar se os dados utilizados atendem os pressupostos como independência e o ajuste da distribuição empírica das taxas de dengue se ajusta à distribuição teórica normal para que, assim, os gráficos de controle estatístico tradicionais possam ser utilizados e, consequentemente, serem interpretado corretamente.

Desta forma, aplicou-se o teste de Shapiro-Wilks, no qual se constatou que a distribuição das taxas de incidência de dengue não se ajusta a uma distribuição Normal, isto é, $p - \text{valor} < 0,01$.

Sendo assim, usou-se na próxima seção uma abordagem apropriada para trabalhar com dados autocorrelacionados e não Normais.

4.2 Ajuste do Modelo ARIMA

A Figura 3 representa a série temporal das taxas de incidência de dengue na Cidade de Natal/RN (C_t). Os dados foram coletados junto ao Centro de Controle de Zoonoses (CCZ) da cidade de Natal/RN. As taxas foram analisadas entre os anos 2000 e 2016. Nota-se, de forma visual, que o maior pico ocorreu no ano 2016 em que houve a epidemia da zika.

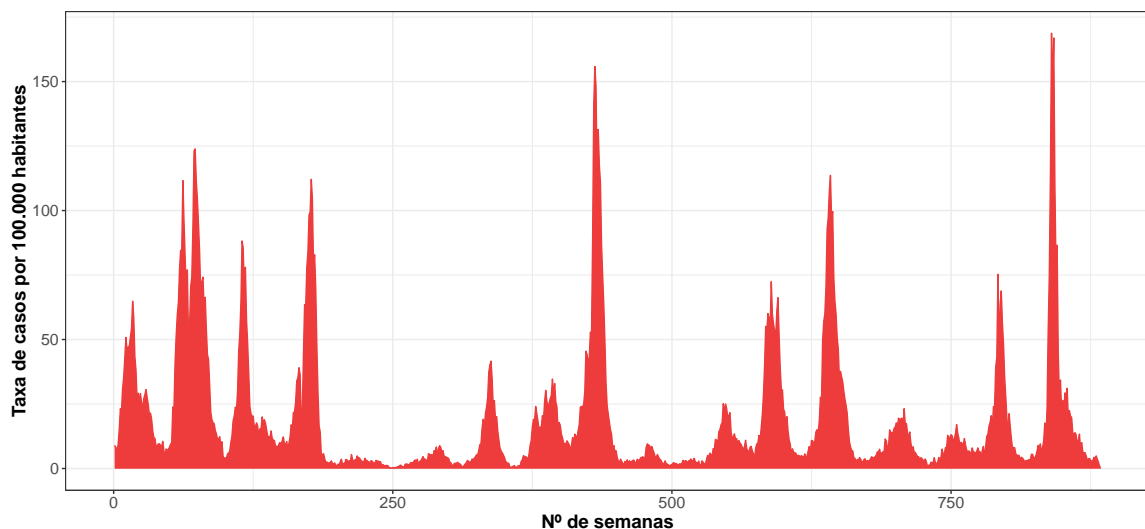


Figura 3 – Série temporal das taxas de casos da cidade de Natal/RN entre 2000 e 2016.

Com o intuito de avaliar se a série C_t é estacionária, é feito o teste Aumentado de Dickey-Fuller (*Augmented Dickey Fuller Test*, em inglês) — uma versão aumentada do Teste de Dickey-Fuller. Este é um teste de raiz unitária em séries temporais. A estatística ADF é um número negativo, e quanto mais negativo, mais indicativo o teste se torna de rejeitar a hipótese nula de que existe raiz unitária na série. A hipótese do teste é dada por

H_0 : tem raiz unitária (não é estacionária).

H_1 : não tem raiz unitária (é estacionária).

A Tabela 2 mostra a estatística do teste juntamente com seu respectivo $p - \text{valor}$.

Tabela 2 – Teste de Dickey-Fuller Aumentado.

$$\text{Dickey-Fuller} = -6 \quad \text{lag} = 10 \quad p - \text{valor} = 0,01$$

Portanto, com $\alpha = 0,05$, há evidências que a série é estacionária, ou seja, $E(C_t)$ é constante, $\text{Var}(C_t) < \infty$, $\gamma_{t,s} = \gamma_{t-s}$ para todo $t \in T$, conforme está na Seção 2.1.1.

Além disso, foi verificado se a série apresentava autocorrelação. Inicialmente, portanto, calculou-se o teste Ljung-Box da série C_t para os 10 primeiros lags (Tabela 3).

Tabela 3 – Teste Ljung-Box.

$$X^2 = 3976,6 \quad \text{lag} = 10 \quad p - \text{valor} = 0$$

O valor do teste indica indícios de presença de autocorrelação nos dados. Para levar em conta o problema de dependência da série foi ajustado um modelo ARIMA. Para seleção do modelo ARIMA, utilizou-se a função *auto.arima()*. Tem-se, então, na Tabela 4 a listagem de possíveis modelos aspirantes à série C_t . Realizamos a construção de um modelo para o período 2000 – 2016 e, posteriormente, previmos para as 37 primeiras semanas de 2017.

Tabela 4 – Critério de informação AIC para os respectivos modelos da família ARIMA (p, d, q) ajustados.

Modelo	AIC	Modelo	AIC
ARIMA(2,0,2) com média diferente de zero	5863,90	ARIMA(2,0,3) com média diferente de zero	5863,17
ARIMA(0,0,0) com média diferente de zero	8291,90	ARIMA(3,0,4) com média diferente de zero	5864,15
ARIMA(1,0,0) com média diferente de zero	5987,01	ARIMA(2,0,3) com média zero	5876,7
ARIMA(0,0,1) com média diferente de zero	7351,67	ARIMA(1,0,3) com média diferente de zero	5865,43
ARIMA(0,0,0) com média zero	8669,16	ARIMA(3,0,3) com média diferente de zero	5862,85
ARIMA(1,0,2) com média diferente de zero	5907,45	ARIMA(4,0,4) com média diferente de zero	Inf
ARIMA(3,0,2) com média diferente de zero	5865,51	ARIMA(3,0,3) com média zero	5877,36
ARIMA(2,0,1) com média diferente de zero	5885,93	ARIMA(4,0,3) com média diferente de zero	5864,67

Observa-se na Tabela 4 que o melhor modelo em questão foi o ARIMA $(3, 0, 3)$ — conhecido também como ARMA $(3, 3)$. A escolha do modelo foi com base no menor valor de AIC, de acordo com a mesma Tabela.

Dado que se chegou a um modelo por meio do critério AIC, é necessário a estimação dos parâmetros do modelo. Conforme dito na Seção 3.2, usou-se a função *auto.arima()* para a estimação dos parâmetros das respectivas ordens. O modelo ARMA $(3, 3)$ é representado na Tabela 6 abaixo.

Tabela 5 – Modelo ARIMA, coeficientes e Critério de Informação Akaike para Natal, Brasil.

ARIMA (p, d, q)	ar ₁	ar ₂	ar ₃	ma ₁	ma ₂	ma ₃	média	AIC
ARIMA $(3,0,3)$	0,9263	0,312	-0,3192	0,1773	-0,0371	0,2084	19,00	5862,85

Conforme está na Seção 2.1.2, é necessário fazer o diagnóstico do modelo ajustado através da análise de resíduos, com o intuito de investigar se o modelo acima foi corretamente ajustado.

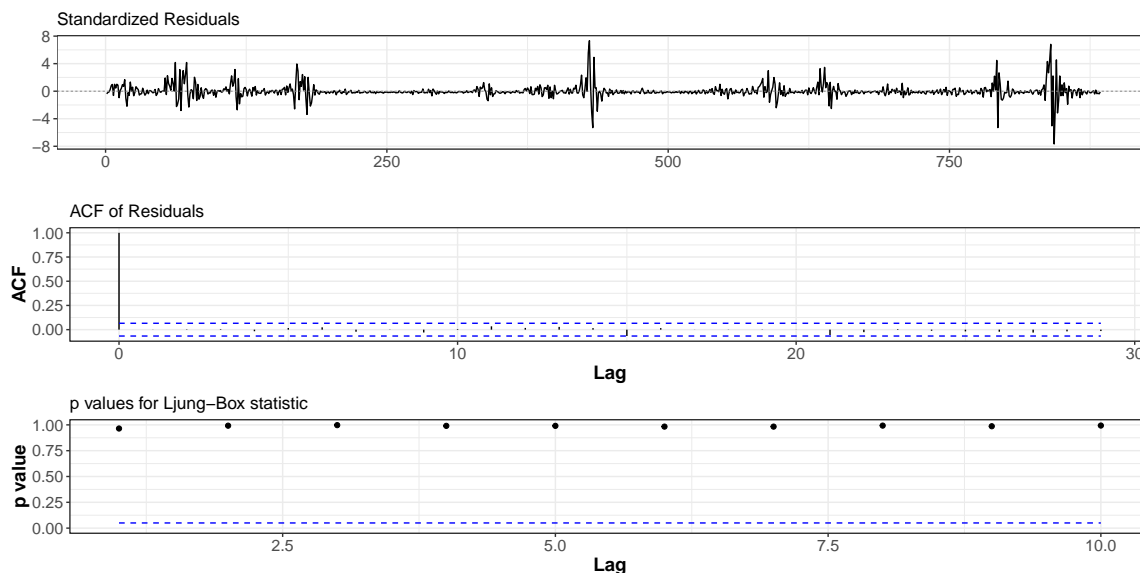


Figura 4 – Gráficos dos resíduos padronizados, ACF dos resíduos e resultados do teste Ljung-Box.

E a estatística do teste é dada pela Tabela 4 abaixo.

Tabela 6 – Teste Ljung-Box.

$$\underline{\underline{X^2 = 2,3 \quad lag = 10 \quad p - valor = 0,9935}}$$

Portanto, por meio do teste de Ljung-Box, com $\alpha = 0,05$, há evidências de que os resíduos são não correlacionados para os 10 primeiros *lags* da série. Ou seja, o modelo acima é adequado e há a constatação de que os resíduos se comportam como um ruído branco (a_t).

O intuito deste trabalho é encontrar um melhor modelo que se adeque à série C_t em questão. Usou-se da estratégia de modelar a série até o ano de 2016, ou seja, até 884ª semana para, assim, prever as taxas. A Tabela 7 mostra as taxas observadas, o modelo ajustado a essas taxas e a previsão para as primeiras 37 semanas de 2017.

Tabela 7 – Predições das 37 semanas de 2017 (intervalo de predição 95%) para a cidade de Natal.

Semana	Predição taxa dos casos	Inferior 95%	Superior 95%	Semana	Predição taxa dos casos	Inferior 95%	Superior 95%
1	0.87	–	13.85	20	17.67	–	69.02
2	1.69	–	21.02	21	17.86	–	69.25
3	3.11	–	28.75	22	18.03	–	69.45
4	4.67	–	36.36	23	18.17	–	69.61
5	6.30	–	42.71	24	18.29	–	69.75
6	7.83	–	48.04	25	18.40	–	69.86
7	9.27	–	52.33	26	18.48	–	69.96
8	10.55	–	55.81	27	18.56	–	70.04
9	11.71	–	58.59	28	18.62	–	70.11
10	12.71	–	60.81	29	18.68	–	70.17
11	13.60	–	62.60	30	18.73	–	70.22
12	14.36	–	64.03	31	18.77	–	70.26
13	15.03	–	65.18	32	18.80	–	70.29
14	15.60	–	66.11	33	18.83	–	70.32
15	16.09	–	66.87	34	18.85	–	70.35
16	16.51	–	67.48	35	18.88	–	70.37
17	16.87	–	67.98	36	18.89	–	70.39
18	17.18	–	68.40	37	18.91	–	70.41
19	17.44	–	68.74				

Outra forma de visualização e comparação dos valores observados e preditos é por meio da Figura 5.

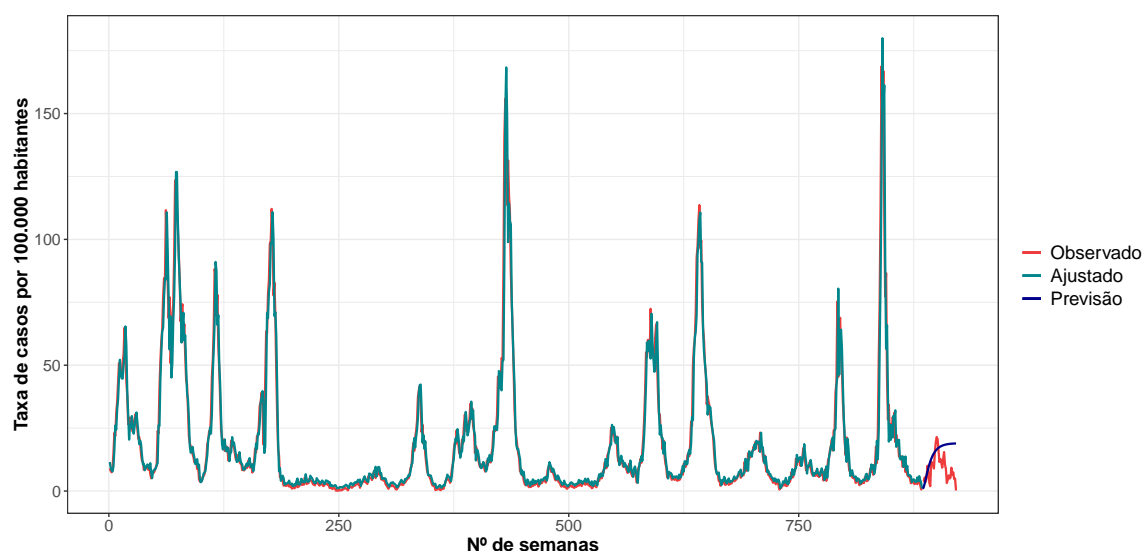


Figura 5 – Série temporal das taxas de casos observadas e preditas para o ano de 2017.

A Figura 5 possibilita, de forma visual, a comparação da série temporal observada com os valores preditos para o ano de 2017. Com isso, observa-se que o modelo ARMA(3, 3) é um bom modelo para predições para as primeiras 23 semanas de 2017.

4.3 Aplicação do Gráfico de Controle EWMA

Nos processos industriais, os limites dos gráficos de controle utilizam, na maioria dos casos a média como medida a ser estudada, e mais ou menos três desvios-padrão.

Já em relação aos processos de vigilância em saúde há a necessidade de identificar o quanto antes uma alteração no comportamento da doença. Alguns trabalhos na área de saúde, portanto, recomendam adotar dois desvios-padrão como limites controle ao invés de três (ALVES, 2004).

Por este motivo, optou-se pelo uso de dois desvios-padrão na plotagem dos gráficos com os limites de controle, tendo em vista que é de extrema importância a detecção precoce de possível epidemias de dengue.

Adotou-se o uso do Gráfico das Médias Móveis Exponencialmente Ponderadas (EWMA) para os resíduos do modelo ARMA(3,3). Adotou-se, também, três valores para a constante ponderação (λ), sendo este: 0,20, 0,50 e 0,80. A escolha desses valores foram escolhidos com o intuito de avaliar este gráfico em diferentes cenários.

Para construir o gráfico EWMA foram utilizadas, inicialmente, 884 observações dos resíduos, que correspondem ao período de janeiro de 2000 a dezembro de 2016, para ajustar os limites do gráfico. As 37 semanas de 2017 foram plotadas nos limites estabelecidos. Pretende-se com essa abordagem investigar se as taxas de incidência média permaneceu sobre controle estatístico, isto é, se as taxas estão dentro do esperado, nesse período.

Para estimar os limites de controle para os diversos cenários é necessária uma situação de controle das variáveis. Desse modo, foram retiradas algumas observações para construir os limites de controle, que foram utilizados para monitoramento das taxas de incidências de dengue. Uma vez estabelecido os limites do GC foram plotadas o restante de observações nos limites ajustados.

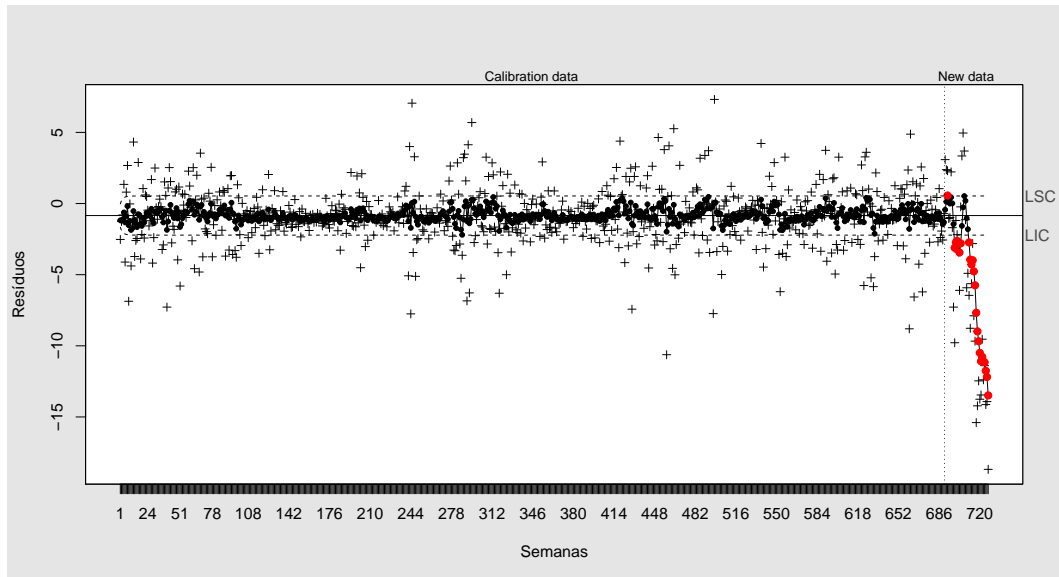


Figura 6 – Gráfico de Controle EWMA para os resíduos ($\lambda = 0.2, k = 2$).

Conforme observado na Figura 6, ao plotar o restante dos resíduos nos limites calibrados foi observado 22 pontos fora de controle, indicando que nessas semanas as taxas de incidências não estavam sobre controle estatístico.

Ainda neste gráfico, percebe-se uma observação acima do limite superior de controle, além de uma sucessão de pontos consecutivos e em declínio entre as últimas observações, culminando com 21 observações fora do limite inferior caracterizando a ação de uma causa especial.

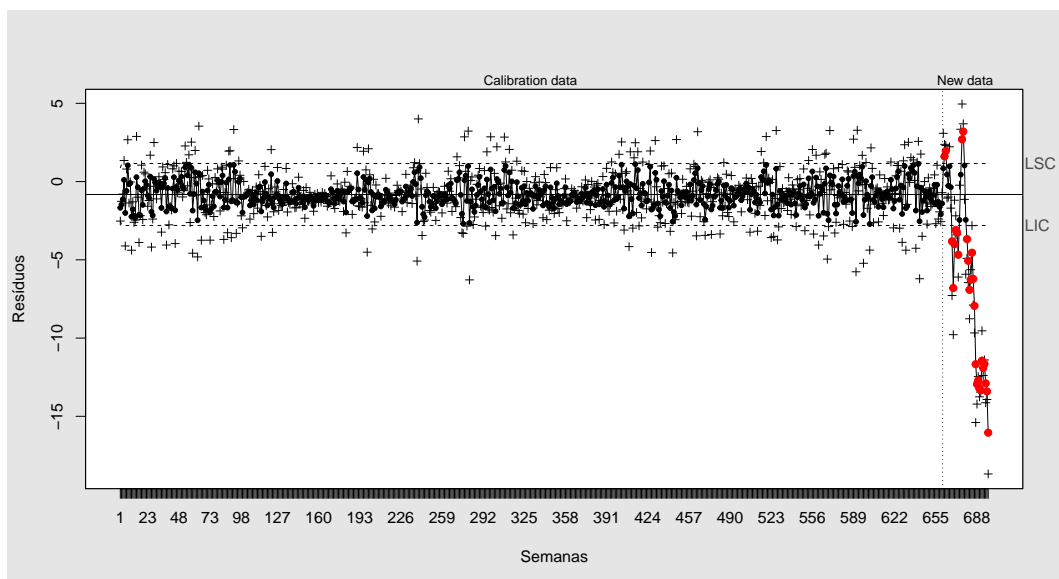


Figura 7 – Gráfico de Controle EWMA para os resíduos ($\lambda = 0.5, k = 2$).

Em relação à Figura 7, ao plotar o restante dos resíduos nos limites calibrados foi

observado 28 pontos fora de controle, indicando que nessas semanas as taxas de incidências não estavam sob controle estatístico.

Neste gráfico, percebe-se também quatro observações fora do limite superior, além disso, há uma sucessão de pontos consecutivos e em declínio entre as últimas observações, culminando com 24 observações fora do limite inferior.

Na Figura 8, ao plotar o restante dos resíduos nos limites calibrados foram observado 17 pontos fora de controle, indicando, assim que nessas semanas as taxas de incidências não estavam sob controle estatístico.

Ainda neste gráfico, percebe-se também quatro observações fora do limite superior, além disso, há uma sucessão de pontos consecutivos e em declínio entre as últimas observações, culminando com 24 observações fora do limite inferior.

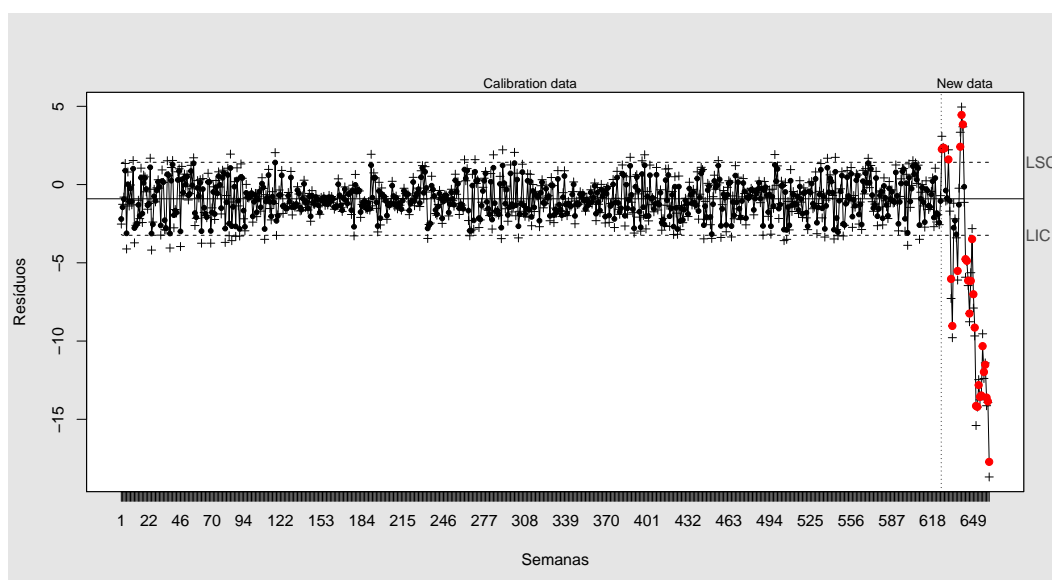


Figura 8 – Gráfico de Controle EWMA para os resíduos ($\lambda = 0.8, k = 2$).

Esta causa especial mencionada acima “pode ser atribuída a um conjunto de fatores, com a mobilização nacional contra as doenças e a maior proteção pessoal da população, a escassez de chuvas em determinadas regiões do país, o que desfavorece a proliferação do mosquito, e a proteção natural que as pessoas adquirem ao ter alguma das doenças em anos anteriores”¹. Já os pontos acima do Limite Superior de Controle pode ser explicado pelo fato de ter havido um crescimento nas taxas nas primeiras semanas de 2017, o que poderia já sugerir uma atenção nessas primeiras semanas.

¹ Disponível em: <http://www.combateaedes.saude.gov.br/pt/noticias/908-casos-de-dengue-no-brasil-caem-90-em-2017>. Acessado em: 15 out. 2018.

5 Conclusão

Este trabalho tem como intuito modelar, predizer e propor um gráfico de controle para monitorar as taxas de incidências de dengue da cidade de Natal/RN.

As taxas de incidência de dengue na cidade de Natal/RN não seguiram uma distribuição Normal e apresentaram autocorrelação.

O modelo ARIMA(3,0,3) adequou-se à série temporal das taxas de casos de dengue para a cidade de Natal. Este modelo foi identificado por meio da função *auto.arima()*. O modelo resultante foi considerado um bom preditor para as taxas de incidência, o que pode ser um importante suporte de orientação e apoio para a predição de taxas futuras.

Além do mais, o ajuste do modelo foi necessário para levar em conta a autocorrelação presente nas observações e, posteriormente, haver aplicação do gráficos de controle EWMA aos resíduos.

Sugere-se, nesse trabalho, o gráfico de controle como uma ferramenta adicional para identificar tendências e períodos de altas taxas incidência de dengue. Com a utilização do GC seria possível monitorar o comportamento das incidências e assim fornecer informações adicionais para a tomada de decisão.

A aplicação dos modelos ARIMA juntamente com a construção dos gráficos de controle pode ser um importante suporte de orientação e apoio para a predição e monitoramento de dados de vigilância epidemiológica, em geral.

Os gráficos de controle podem facilitar a análise de dados em saúde pública orientando gestores a tomar ações efetivas de combate, prevenção e controle com base nos resultados observados, embora existam diversas dificuldades em se trabalhar com dados de vigilância epidemiológica — demora no preenchimento dos bancos de dados, presença de autocorrelação ou dados não normais.

É, também, necessário ressaltar quando se trata de vigilância de saúde, o tempo é essencial para a detecção precoce de possíveis epidemias ou surtos para que as medidas de controle sejam adotadas oportunamente, de tal forma que grande número de casos e óbitos não venha a ser grande.

Referências

- AKAIKE, H. A new look at the statistical model identification. *IEEE transactions on automatic control*, Ieee, v. 19, n. 6, p. 716–723, 1974. 29
- ALVES, R. G. *Utilidade do Diagrama de Controle de Doenças na Previsão e Controle de Epidemias*. Tese (Doutorado) — Universidade Federal do Rio de Janeiro, 2004. 36
- ALWAN, L. C.; ROBERTS, H. V. Time-series modeling for statistical process control. *Journal of Business & Economic Statistics*, Taylor & Francis Group, v. 6, n. 1, p. 87–95, 1988. 26
- AMAKU, M. et al. Interpretations and pitfalls in modelling vector-transmitted infections. *Epidemiology & Infection*, Cambridge University Press, v. 143, n. 9, p. 1803–1815, 2015. 12
- AMAKU, M. et al. Magnitude and frequency variations of vector-borne infection outbreaks using the ross–macdonald model: explaining and predicting outbreaks of dengue fever. *Epidemiology & Infection*, Cambridge University Press, v. 144, n. 16, p. 3435–3450, 2016. 12
- BOX, G.; JENKINS, G.; REINSEL, G. *Tim Series Analysis, Foseactingand Control*. [S.l.]: Prentice-Hall, 1994. 14, 18, 19
- BOX, G. E.; PIERCE, D. A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, Taylor & Francis Group, v. 65, n. 332, p. 1509–1526, 1970. 29
- BRADY, O. J. et al. Dengue disease outbreak definitions are implicitly variable. *Epidemics*, Elsevier, v. 11, p. 92–102, 2015. 12
- CHEN, Y.; COOK, A. R.; LIM, A. X. Randomness of dengue outbreaks on the equator. *Emerging infectious diseases*, Centers for Disease Control and Prevention, v. 21, n. 9, p. 1651, 2015. 12
- COELHO, G. E. et al. Sensibilidade do sistema de vigilância da dengue no brasil para detecção de casos hospitalizados. *PLoS negligenciadas doenças tropicais*, Public Library of Science, v. 10, n. 5, 2016. 12
- CORTES, F. et al. Time series analysis of dengue surveillance data in two brazilian cities. *Acta tropica*, Elsevier, v. 182, 2018. 12, 13
- COSTA, A. F. B. *Controle Estatístico de Qualidade*. 2. ed. São Paulo: Atlas, 2008. 336 p. 20
- ENGLISH, J. R. et al. Detecting changes in autoregressive processes with x and ewma charts. *Iie Transactions*, Springer, v. 32, n. 12, p. 1103–1113, 2000. 26
- FAVA, V. L. *Análise de séries de tempo*. In: Vasconcellos, MAS; Alves, D. Manual de econometria: nível intermediário. São Paulo: Ed. Atlas, 2000. 15

- G1. *Microcefalia é causada por vírus da zika mutante, mostra estudo brasileiro*. 2016. Acessado em: 20 jun. 2018. Disponível em: <<http://glo.bo/1Sw8x61>>. 12
- GLOBO, O. *Natalidade brasileira caiu em 2016, auge da epidemia da Zika*. 2017. Acessado em: 20 jun. 2018. Disponível em: <<https://glo.bo/2krDLnn>>. 12
- GUZMAN, M. G.; HARRIS, E. Dengue. *The Lancet*, Elsevier, v. 385, n. 9966, p. 453–465, 2015. 12
- HENNING, E. et al. Um estudo para a aplicação de gráficos de controle estatístico de processo em indicadores de qualidade da água potável. *Sistemas & Gestão*, v. 9, n. 1, p. 2–13, 2014. 26
- HYNDMAN, R.; KHANDAKAR, Y. Automatic time series forecasting: the forecast package for r. *Journal of Statistical Software*, v. 26, n. 3, p. 1–22, 2008. 29
- JOHANSSON, M. A. et al. Evaluating the performance of infectious disease forecasts: A comparison of climate-driven and seasonal dengue forecasts for mexico. *Scientific reports*, Nature Publishing Group, v. 6, p. 33707, 2016. 12
- JUNIOR, J. B. S. et al. Dengue and dengue hemorrhagic fever, brazil, 1981–2002. *Emerging infectious diseases*, Centers for Disease Control and Prevention, v. 11, n. 1, p. 48, 2005. 12
- KOEHLER, A.; MARKS, N.; O'CONNELL, R. Ewma control charts for autoregressive processes. *Journal of the Operational Research Society*, Taylor & Francis, v. 52, n. 6, p. 699–707, 2001. 26
- LJUNG, G. M.; BOX, G. E. On a measure of lack of fit in time series models. *Biometrika*, Oxford University Press, v. 65, n. 2, p. 297–303, 1978. 17, 29
- LU, C.-W.; REYNOLDS, M. R. J. Ewma control charts for monitoring the mean of autocorrelated processes. *Journal of Quality technology*, Taylor & Francis, v. 31, n. 2, p. 166–188, 1999. 26
- LUZ, P. M. et al. Time series analysis of dengue incidence in rio de janeiro, brazil. *The American journal of tropical medicine and hygiene*, ASTMH, v. 79, n. 6, p. 933–939, 2008. 12
- MARTINEZ, E. Z.; SILVA, E. A. S. d. Predicting the number of cases of dengue infection in ribeirão preto, são paulo state, brazil, using a sarima model. *Cadernos de saude publica*, SciELO Public Health, v. 27, p. 1809–1818, 2011. 12
- MARTINEZ, E. Z.; SILVA, E. A. S. d.; FABBRO, A. L. D. A sarima forecasting model to predict the number of cases of dengue in campinas, state of são paulo, brazil. *Revista da Sociedade Brasileira de Medicina Tropical*, SciELO Brasil, v. 44, n. 4, p. 436–440, 2011. 12
- MONTGOMERY, D. *Introdução ao controle estatístico da qualidade*. Rio de Janeiro: LTC, 2009. v. 4. 20, 21, 22, 23, 24, 25, 26
- MONTGOMERY, D. C.; MASTRANGELO, C. M. Some statistical process control methods for autocorrelated data. *Journal of Quality Technology*, Taylor & Francis, v. 23, n. 3, p. 179–193, 1991. 26

- MORETTIN, P. A.; TOLOI, C. *Análise de séries temporais*. 2. ed. São Paulo: Blucher, 2006. 14, 15, 18, 19
- PEDRINI, D. C.; CATEN, C. S. ten. Comparação entre gráficos de controle para resíduos de modelos. *Revista GEPROS*, v. 4, n. 4, p. 123, 2008. 26
- PORTA, M. *A dictionary of epidemiology*. [S.l.]: Oxford University Press, 2014. 12
- PSARAKIS, S.; PAPALEONIDA, G. Spc procedures for monitoring autocorrelated processes. *Quality Technology & Quantitative Management*, Taylor & Francis, v. 4, n. 4, p. 501–540, 2007. 13, 26
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2017. Disponível em: <<https://www.R-project.org/>>. 29
- RUBERTO, I. et al. The availability and consistency of dengue surveillance data provided online by the world health organization. *PLoS neglected tropical diseases*, Public Library of Science, v. 9, n. 4, p. e0003511, 2015. 12
- RUNGE-RANZINGER, S. et al. What does dengue disease surveillance contribute to predicting and detecting outbreaks and describing trends? *Tropical Medicine & International Health*, Wiley Online Library, v. 13, n. 8, p. 1022–1041, 2008. 12
- SILAWAN, T. et al. Temporal patterns and forecast of dengue infection in northeastern thailand. *Southeast Asian Journal of Tropical Medicine and Public Health*, v. 39, n. 1, p. 90, 2008. 12
- STANAWAY, J. D. et al. The global burden of dengue: an analysis from the global burden of disease study 2013. *The Lancet doenças infecciosas*, Elsevier, v. 16, n. 6, 2016. 12
- TEIXEIRA, M. G. et al. Tendências epidemiológicas da dengue no brasil (2000 - 2010): uma busca e análise sistemática da literatura. *PLoS negligenciadas doenças tropicais*, Public Library of Science, v. 7, n. 12, 2013. 12
- WILDER-SMITH, A. et al. Epidemic arboviral diseases: priorities for research and public health. *The Lancet doenças infecciosas*, Elsevier, v. 17, n. 3, p. e101 – e106, 2017. 12