
Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Estatística

Gráficos de Controle para Monitorar a Taxa de
Letalidade da COVID-19 nos Estados da Região
Nordeste e no Brasil.

Natanael Fernandes Sarmiento

Dezembro/2021

Natanael Fernandes Sarmiento

Gráficos de Controle para Monitorar a Taxa de Letalidade da COVID-19 nos Estados da Região Nordeste e no Brasil.

Monografia apresentada ao Curso de Bacharelado em Estatística da Universidade Federal da Paraíba como requisito parcial para obtenção do Grau de Bacharel. Área de Concentração: Séries Temporais e Controle Estatístico de Processos.

João Pessoa
Dezembro de 2020

Catálogo na publicação
Seção de Catalogação e Classificação

S246g Sarmiento, Natanael Fernandes.

Gráficos de controle para monitorar a taxa de letalidade da COVID-19 nos estados da região nordeste e no Brasil / Natanael Fernandes Sarmiento. - João Pessoa, 2021.

46 p. : il.

Orientação: Tatiene Correia de Souza.

TCC (Graduação/Bacharelado em Estatística) - UFPB/CCEN.

1. Gráficos de controle. 2. Análise de resíduos. 3. Letalidade da Covid-19. 4. Séries temporais. I. Souza, Tatiene Correia de. II. Título.

UFPB/CCEN

CDU 519.171(043.2)



ATA DE DEFESA DO TRABALHO DE CONCLUSÃO DE CURSO

“Gráficos de Controle para Monitorar a Taxa de Letalidade da COVID-19 nos Estados da Região Nordeste e no Brasil.”

Natanael Fernandes Sarmiento

No nono dia do mês de Dezembro de 2021 às 14:00 horas, de modo remoto, via plataforma Google Meet, através de videoconferência no endereço eletrônico: <https://meet.google.com/cyz-wgcg-ojk>, constituiu-se a Banca Examinadora do Trabalho de Conclusão de Curso do discente Natanael Fernandes Sarmiento, matrícula nº 2016040029, composta pelos professores: Dra. Tatiene Correia de Souza, Presidente/Orientadora (Departamento de Estatística - UFPB), Dr. Luiz Medeiros de Araujo Lima Filho, Examinador (Departamento de Estatística - UFPB), Dra. Tarciana Liberal Pereira de Araujo, Examinadora (Departamento de Estatística - UFPB) e Dra. Maria Lidia Coco Terra, Examinadora Suplente (Departamento de Estatística - UFPB). Iniciando-se os trabalhos, a presidente da Banca Examinadora cumprimentou os presentes, comunicou aos mesmos a finalidade da reunião e passou a palavra ao candidato para que se fizesse, oralmente, a exposição do Trabalho de Conclusão de Curso intitulado “**Gráficos de Controle para Monitorar a Taxa de Letalidade da COVID-19 nos Estados da Região Nordeste e no Brasil.**”. Concluída a apresentação, o candidato foi arguido pela Banca Examinadora que sugeriu que o discente fizesse algumas alterações até o dia 15 de Dezembro de 2021. Uma vez entregue a versão final do Trabalho de Conclusão de Curso à Coordenação do Bacharelado em Estatística com as alterações solicitadas pela Banca Examinadora, dentro do prazo do prazo estabelecido, o discente será aprovado com a nota 8,7 (oito inteiros e sete décimos), resultante da média aritmética das notas atribuídas pelos membros da Banca Examinadora.

Dra. Tatiene Correia de Souza
(Professora Orientadora)

Dra. Tarciana Liberal Pereira de Araujo
(Professora Examinadora)

Dr. Luiz Medeiros de Araujo Lima Filho
(Professor Examinador)

Natanael Fernandes Sarmiento
(Discente)

João Pessoa, 09 de Dezembro de 2021.

**Gráficos de Controle para Monitorar a Taxa de Letalidade da COVID-19 nos
Estados da Região Nordeste e no Brasil.**

Natanael Fernandes Sarmiento

Banca Examinadora

Prof^o Dr. Tatiene Correia de Souza
Orientadora - UFPB

Prof^o Dr. Luiz Medeiros de Araujo Lima Filho
Examinador - UFPB

Prof^o Dr. Tarciana Liberal Pereira
Examinadora - UFPB

CONCEITO FINAL: _____

*Dedico este trabalho a minha família e aos amigos
que fiz durante minha trajetória acadêmica.*

AGRADECIMENTOS

- Em primeiro lugar a minha mãe, Francisca Anúbia Fernandes, que é meu pilar na vida, é meu anjo e minha guerreira, sempre me deu apoio incondicional em todas as minhas decisões, aos meus irmãos, Caio Fernandes Sarmiento e Miguel Fernandes Perreira, pela união e parceria de sempre, ao meu padastro Radier Fernandes Pereira, a toda minha família, em especial a minha tia Maria Andréia Fernandes.
- Ao minha orientadora, Prof. Dr. Tatiene Correia de Souza, que esteve presente na minha trajetória da graduação me orientando no PIBIC, TCC I e TCC II, prestando uma espetacular orientação, depositando confiança e me incentivando sempre a ir mais longe, muito obrigado.
- Ao Departamento de Estatística que conta com excelentes docentes.
- Aos examinadores deste Trabalho.
- Aos amigos que fiz durante o curso, em especial aos amigos Eduardo, Matheus, João Paulo, Lukas, Natasha, Mardonio, Daniele, Francielle, Kleber, Marina, Milleny e Natan.
- Aos amigos que fiz na vida Lucas, Vitor, Gabriel, Romário, Ivo, Halysen e Pedro Yuri.

*Você não controla tudo o que acontece na sua vida
mas a maneira como reage é sua responsabilidade,
você é apenas um instante no tempo, um breve
instante, apenas pó, veio do pó e logo voltará ao pó
mas até lá seja, mas seja tudo o que puder ser.*

- LetoDie (Memento Mori)

O objetivo deste trabalho é utilizar gráficos de controle para monitorar a taxa de letalidade da Covid-19 nos estados da região Nordeste e no Brasil. Diante disso foram calculadas a taxa de letalidade da covid-19 por semana epidemiológica nos estados da região Nordeste e no Brasil. O período estudado foi de 3 de maio de 2020 até 18 de setembro de 2021, este período representa as semanas epidemiológicas 19 de 2020 até a semana 37 de 2021, totalizando 72 semanas. Modelos de séries temporais foram ajustados a cada taxa de letalidade calculada, obedecendo ao critério de seleção adotado BIC e a análise de resíduos. Os gráficos de controle foram aplicados aos resíduos dos modelos selecionados, em que contribuiu para monitorar a letalidade da covid-19 e detectar semanas incomuns representando pontos fora de controle. Os gráficos foram importantes para detectar grandes mudanças na letalidade e o gráfico de médias móveis exponencialmente ponderadas (MMEP) apresentou resultados que detectaram pequenas mudanças no fenômeno representando uma combinação bastante útil no monitoramento de variáveis epidemiológicas. Os resultados descritivos foram preocupantes uma vez que o valor máximo obtido para taxa de letalidade foi de 10,559, representando que a cada 100 pessoas contaminadas pelo vírus, aproximadamente 11 pessoas morreram e a maior taxa de letalidade média foi de 3,396, mostrando ser necessário o monitoramento da doença. Os resultados dos gráficos de controle se mostraram muito importantes para o monitoramento da covid-19, uma vez que os gráficos de controle aplicados nos estados da região Nordeste e Brasil sinalizaram que a taxa de letalidade não está sob controle. Assim mostrando ser necessário aplicar medidas em determinados períodos evitando um aumento na taxa de letalidade.

Palavras-chaves: Gráficos de Controle, Análise de Resíduos, Letalidade da Covid-19, Séries Temporais.

ABSTRACT

The objective of this work is to use control charts to monitor the Covid-19 fatality rate in the states of the Northeast region and in Brazil. Therefore, the covid-19 fatality rate per epidemiological week was calculated in the states of the Northeast region and in Brazil. The period studied was from May 3, 2020 to September 18, 2021, this period represents the epidemiological weeks 19 of 2020 to week 37 of 2021, totaling 72 weeks. Time series models were adjusted to each calculated lethality rate, following the selection criteria adopted by BIC and the residual analysis. Control charts were applied to the residuals of the selected models, which helped to monitor covid-19 lethality and detect unusual weeks representing out-of-control points. The graphs were important to detect large changes in lethality and the graph of exponentially weighted moving averages (EWMA) showed results that detected small changes in the phenomenon, representing a very useful combination in monitoring epidemiological variables. The descriptive results were worrying since the maximum value obtained for the lethality rate was 10.559, representing that for every 100 people infected by the virus, approximately 11 people died and the highest average lethality rate was 3.395, showing that monitoring is necessary of the disease. The results of the control charts proved to be very important for the monitoring of covid-19, since the control charts applied in the states of the Northeast region and Brazil indicated that the lethality rate is not under control. Thus showing that it is necessary to apply measures in certain periods, avoiding an increase in the lethality rate.

Keywords: Control Charts, Residual Analysis, Covid-19 Lethality, Time Series.

Lista de Figuras	vii
Lista de Tabelas	viii
Lista de abreviaturas e siglas	ix
1 Introdução	1
2 Metodologia	4
2.1 Séries Temporais	4
2.2 Gráficos de Controle	10
3 Resultados	17
4 Conclusão	30
Referências bibliográficas	32
A	36
A.1 Resultados τ sdiag dos estados da região nordeste e Brasil.	36
A.2 Script R.	39

LISTA DE FIGURAS

2.1	Exemplo Resíduo modelo AR(1).	9
2.2	Exemplo FAC modelo AR(1).	9
2.3	Exemplo de Gráfico de Controle.	11
3.1	Gráfico da taxa de letalidade por semana epidemiológica nos estados, AL, BA, CE, MA, PB, PE, PI, RN e SE e na região nordeste.	20
3.2	Gráfico da taxa de letalidade por semana epidemiológica no Brasil.	21
3.3	Gráfico dos Resíduos por semana epidemiológica obtidos a partir dos modelos selecionados considerando a taxa de letalidade por estados do Nordeste e para a região Nordeste.	23
3.4	Gráfico dos Resíduos por semana epidemiológica obtidos a partir do modelo selecionado considerando a taxa de letalidade no Brasil.	24
3.5	Gráficos de controle do tipo Shewhart e MMEP aplicado aos resíduos obtidos a partir dos modelos estimado para a taxa de letalidade dos estado de Alagoas, Bahia, Ceará, Maranhão e Paraíba, respectivamente.	26
3.6	Gráficos de controle do tipo Shewhart e MMEP aplicado aos resíduos obtidos a partir dos modelos estimado para a taxa de letalidade dos estado do Pernambuco, Piauí, Rio Grande do Norte, Sergipe e da região Nordeste, respectivamente.	27
3.7	Gráficos de controle do tipo Shewhart e MMEP aplicado aos resíduos obtidos a partir do modelo estimado para a taxa de letalidade no Brasil.	28
A.1	Gráficos dos resultados da função <code>tsdiag</code> nos estados de Alagoas e Bahia.	36
A.2	Gráficos dos resultados da função <code>tsdiag</code> nos estados de Ceará e Maranhão.	37
A.3	Gráficos dos resultados da função <code>tsdiag</code> nos estados de Paraíba e Pernambuco.	37
A.4	Gráficos dos resultados da função <code>tsdiag</code> nos estados de Piauí e Rio Grande do Norte.	38

A.5	Gráficos dos resultados da função $tsdiag$ no estado de Sergipe e na região Nordeste.	38
A.6	Gráficos dos resultados da função $tsdiag$ no Brasil.	39

LISTA DE TABELAS

2.1	ARLs em controle para os gráficos MMEP e o de medidas individuais para várias distribuições gama e distribuição normal.	14
3.1	Resultados descritivos referente a letalidade nos estados estudados, no Nordeste e no Brasil.	21
3.2	Resultados da ordem do modelo ARIMA selecionado referente a taxa letalidade nos estados estudados, no Nordeste e no Brasil.	22
3.3	Resultados dos teste de normalidade (KS) e teste de Ljung-Box com $k = 6$ nos resíduos dos modelos.	24

LISTA DE ABREVIATURAS E SIGLAS

AIC Critério de Informação de Akaike

AICC Critério de Informação de Akaike Corrigido

AR Autorregressivo

ARIMA Autorregressivo Integrado e Médias Móveis

ARMA Autorregressivo e Médias Móveis

BIC Critério de Informação Bayesiano

CEP Controle Estatístico de Processo

FAC Função de Autocorrelação

LC Linha Central

LIC Limite Inferior de Controle

LSC Limite Superior de Controle

MA Médias Móveis

MMEP Médias Móveis Exponencialmente Ponderadas

SUS Sistema Único de Saúde

CAPÍTULO 1

INTRODUÇÃO

Na China em dezembro de 2019 foi identificado um surto de síndrome respiratória aguda causada pelo novo coronavírus (SARS-CoV-2) em trabalhadores de um mercado de alimentos e animais vivos em Wuhan (GUILLIER *et al.*, 2020). Esse vírus é pertencente à família Coronaviridae e provoca a doença respiratória denominada COVID-19. No ano de 2020, no dia 11 de março a OMS (Organização Mundial da Saúde) declarou o novo coronavírus (Covid-19) como uma pandemia. No Brasil, o primeiro caso de Covid-19 foi confirmado pelo MS (Ministério da Saúde) em 20 de fevereiro de 2020 (Ministério da Saúde, 2021a). E em 20 de março, por meio do decreto (BRASIL, 2020), o Brasil declarou calamidade pública, visando criar campanhas e medidas para combater a Covid-19.

A Covid-19 é uma infecção respiratória aguda causada pelo coronavírus SARS-CoV-2, potencialmente grave, de elevada transmissibilidade e de distribuição global (Ministério da Saúde, 2021b). O SARS-CoV-2, da família Coronaviridae, é um vírus envelope de RNA (ácido ribonucleico) de fita simples, senso positivo, que está intimamente relacionado com o SARS (síndrome respiratório agudo grave) como coronavírus e com base na análise filogenética, esses coronavírus têm um ancestral comum que se assemelha ao coronavírus de morcego HKU9-1 (HE *et al.*, 2021). Os impactos da Covid-19 começaram a ser percebidos no setor da saúde e na economia mundial no início de 2020. No Brasil, o Ministério da Saúde procurou criar medidas para o monitoramento da doença após o anúncio da OMS. Entre as medidas indicadas pelo MS, estão as não farmacológicas, como distanciamento social, etiqueta respiratória e de higienização das mãos, uso de máscaras, limpeza e desinfecção de ambientes, isolamento de casos suspeitos, confirmados e quarentena dos contatos dos casos de covid-19, conforme orientações médicas.

Diante deste problema, é de suma importância entender o comportamento da doença

e encontrar possíveis causas para criar medidas de combate do coronavírus. Ferramentas estatísticas são muito importantes na análise de dados e podem contribuir muito nestas situações. Na literatura, há vários estudos que utilizam essas técnicas para entender a Covid-19 como, por exemplo, Chan J. S. Malik Peiris (2011), Bayyurt & Bayyurt (2020), Benvenuto *et al.* (2020), Biryukov *et al.* (2020), Chen *et al.* (2020), Ferraz *et al.* (2020), Guillier *et al.* (2020), Sarkodie & Owusu (2020), Tzampoglou & Loukidis (2020), Agarwal *et al.* (2021), He *et al.* (2021), Loh *et al.* (2021), Noreen & Saqib (2021), Yang, Li & Cao (2021), Zhang *et al.* (2021), Zheng *et al.* (2021) dentre outros. Então é importante aplicar ferramentas estatísticas que ajudem a monitorar, detectar problemas e entender o comportamento da doença. Variáveis epidemiológicas são observações obtidas ao longo do tempo que necessitam de estudos e monitoramento, essas observações são autocorrelacionadas e assim limitando ferramentas que não suporte autocorrelação. As séries temporais são observações obtidas ao longo do tempo, a partir de fenômenos correlacionados que precisam ser estudados, monitorados e previstos. Para descrever um fenômeno autocorrelacionado avaliado ao longo do tempo, a aplicação de modelo de série temporal, é necessária para entender o comportamento e fazer previsões (MONTGOMERY; JOHNSON, 1976).

Os gráficos de controle são ferramentas amplamente utilizadas e importantes no campo do CEP pois, ajudam a monitorar o comportamento do fenômeno e de verificar visualmente observação fora de controle. Os gráficos de controle podem ser utilizados em variáveis ou atributos a serem controlados, visando entender se o processo está em um estado de controle ou não (MONTGOMERY, 2009). Uma combinação bastante útil em detectar pequenas mudanças no processo e grandes mudanças é aplicar gráfico de controle do tipo Shewhart e gráfico de controle MMEP (Médias Móveis Exponencialmente Ponderadas) para medidas individuais no fenômeno estudado. Uma combinação bastante útil em avaliar fenômenos autocorrelacionados é utilizar modelos de séries temporais e aplicar gráficos de controle nos resíduos extraído do modelo. Montgomery (2009) afirma que gráficos de controle convencionais podem ser aplicados à sequência de resíduos obtidos de modelos de séries temporais, para descrever fenômenos autocorrelacionados. Estas ferramentas podem ser usadas para monitorar a covid-19 buscando detectar causas especiais, apresentando resultados que podem ajudar os governos a criar medidas para controlar a covid-19, organizar campanhas junto ao SUS em determinados períodos, entre outras formas de controle que podem ser usadas.

Assim, este trabalho propõe o uso de Séries Temporais com Gráficos de Controle, apresentando os passos para aplicação, análise e construção dos gráficos, visando criar um procedimento que ajude de forma rápida a monitorar e detectar problemas no fenômeno. Este estudo está dividido da seguinte forma: na Seção de metodologia são apresentados

alguns conceitos de séries temporais, análise dos modelos e gráficos de controle. Na Seção dos resultados são apresentados a análise da taxa de letalidade da covid-19 nos estados da região Nordeste e no Brasil e na Seção conclusão são apresentadas as principais conclusões.

Atualmente vemos um grande fluxo de informações e o aumento da necessidade de análise dessas informações, em grande maioria esses dados são correlacionados e assim necessitam de ferramentas que ajudem na análise e interpretação. Uma combinação bastante útil para monitorar observações autocorrelacionadas é usar modelos de séries temporais e aplicar gráficos de controle aos resíduos do modelo, (MONTGOMERY, 2009). Vários estudos usam essas duas ferramentas para entender melhor o fenômeno em particular, por exemplo, em Veloso (2018) são aplicados gráficos de controle com séries temporais em indicadores epidemiológicos. Também em Boaventura (2020) são explicados os benefícios da combinação de séries temporais e gráfico de controle e outros como Noreen & Saqib (2021) e outros trabalhos. A seguir é apresentado mais detalhes sobre séries temporais e gráficos de controle.

2.1 Séries Temporais

Séries Temporais são observações obtidas ao longo do tempo, a partir de fenômenos que precisam ser estudados, monitorados e previstos e muitas vezes esses fenômenos são correlacionados. Segundo Montgomery & Johnson (1976) e Ehlers (2005), para descrever o processo ou fenômeno aplica-se técnicas de Séries Temporais, visando entender o comportamento da série e realizar previsões. Na estatística há algumas ferramentas que não suportam dados autocorrelacionados, como por exemplo gráficos de controle e modelos de regressão onde ambos necessitam que as observações sejam independentes. Os modelos de séries temporais são ferramentas que suportam dados autocorrelacionados, como por

exemplo modelos autorregressivos, apresentando ser uma ferramenta boa para entender esse tipo de observações. Assim, há ferramentas estatísticas bastantes úteis para analisar processos temporais onde através de modelos reduzidos como AR e MA ou gerais como ARIMA e ARMA, podemos analisar o comportamento da série e realizar previsões de observações futuras.

Os modelos ARMA (modelo autorregressivo de médias móveis) são bastante utilizados para estudos de fenômenos e conseguem se adequar a categoria de séries correlacionadas. Sendo assim os modelos ARMA é uma generalização dos processos do tipo autorregressivos $AR(p)$ e de médias móveis $MA(q)$. Então seja Y_t , uma série temporal em que $t = 1, 2, 3, \dots, n$, o modelo $ARMA(p, q)$ é apresentado da seguinte forma,

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q},$$

em que ϵ_t é ruído branco com média zero e variância σ^2 . O modelo ARMA pode ser reescrito usando o operador defasagem B

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) Y_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) \epsilon_t,$$

Então o modelo pode ser escrito como,

$$\phi(B) Y_t = \theta(B) \epsilon_t$$

em que ,

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

$$\theta(B) = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q).$$

No modelo ARMA há uma limitação quando as observações não são estacionárias. Sendo assim os modelos ARIMA (modelo autorregressivo integrado de médias móveis) são obtidos considerando o comportamento não estacionário da série definindo d como o número de diferenças necessárias para tornar a série estacionária e combinando-se processos autorregressivos de médias móveis. Então considere um modelo do tipo $ARMA(p, q)$, e considere agora a série temporal é não estacionária. Um modelo do tipo $ARIMA(p, d, q)$ pode ser utilizado em que d é um número inteiro que representa o número de diferenças realizadas, para a série ficar estacionária. Usando o operador de defasagem B , o processo ARIMA pode ser representado como $\Delta^d Y_t = (1 - B)^d Y_t$.

O modelo ARIMA pode ser representado da seguinte forma:

$$\phi(B)(1 - B)^d Y_t = \theta(B)\epsilon_t.$$

Ha uma particularidade entre os modelos AR e MA, onde os mesmos podem ser obtidos a partir de um modelo ARMA. Um processo autorregressivo de ordem p , $AR(p)$, pode ser obtido como casos especiais a partir de um modelo $ARMA(p,q)$ quando $q = 0$, ou seja,

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \epsilon_t,$$

em que, ϕ_0, \dots, ϕ_p são os parâmetros do modelo e ϵ_t representa o ruído branco com média zero e variância σ^2 .

O modelo AR pode ser apresentado em função de um operador de defasagem que é escrito da seguinte forma:

$$(1 - \phi_1 B - \dots - \phi_p B^p) Y_t = \epsilon_t,$$

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

Já em um processo de médias móveis de ordem q , $MA(q)$, pode ser obtido como casos especiais a partir de um modelo $ARMA(p,q)$ quando $p = 0$, por definição, o processo pode ser escrito em função das defasagens do erro aleatório, isto é

$$Y_t = \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q},$$

em que, $\theta_1, \dots, \theta_q$ são os parâmetros do modelo e ϵ_t representa o ruído branco com média zero e variância σ^2 . Este modelo também pode ser reescrito em função do operador de defasagem.

$$Y_t = (1 + \theta_1 B + \dots + \theta_q B^q) \epsilon_t,$$

em que,

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

Apresentados os modelos AR, MA, ARMA e ARIMA é importante que os parâmetros desses modelos sejam estimados e a ordem seja escolhida, sendo preferível selecionar modelos com o menor número de parâmetros. Em Box & Jenkins (1970) é apresentado que os modelos ideais seriam aqueles com máxima simplicidade e o menor número de parâmetros possível seguindo uma representação adequada da série temporal trabalhada, um modelo parcimonioso.

Os parâmetros do modelo precisam ser estimados e representar com eficiência a série a ser estudada. Os estimadores de máxima verossimilhança (EMV) são obtidos a partir do método de máxima verossimilhança que produz estimativas eficientes, suficientes e com ótimas características. Outro método bastante utilizado é o método de mínimos quadrados que apresenta ótima qualidade, em certas condições os estimadores de mínimos quadrados serão, aproximadamente EMVs. Na literatura há outros métodos que podem ser adotados, para mais detalhes ver Morettin & Tolo (2018), Cryer & Chan (2008), Brockwell & Davis (2016) e outro.

Os critérios de seleção são usados para penalizar a quantidade de parâmetros no modelo, assim escolhendo modelos com o menor número possível de parâmetros, os principais critérios são: Critério de Informação de Akaike (AIC), Critério de Akaike Corrigido (AICC) e o Critério Bayesiano de Informação (BIC) (AKAIKE, 1973). O critério AIC, se baseia no valor negativo do logaritmo natural da função de verossimilhança, o critério AICC é uma versão com o viés corrigido do critério de informação AIC e o BIC também é calculado a partir do logaritmo natural da função de verossimilhança. Estes Critério de Informação, consiste em encontrar a ordem do modelo ao minimizar o valor da medida calculada, assim, modelo considerado mais adequado seria aquele com menor valor correspondente. Para uma revisão geral destes e outros critérios de informação no contexto de séries temporais, ver AKAIKE (1973), Ehlers (2005), Morettin & Tolo (2018).

O processo de escolha da ordem dos parâmetros em um modelo de séries temporais é fundamental para escolha de um bom modelo para representar o fenômeno em estudo, em que esse processo pode ser demorado e repetitivos até escolher o modelo adequado. Tendo em vista este problema Hyndman & Khandakar (2008), apresentaram o algoritmo chamado `auto.arima` que ajuda na seleção do modelo. Este algoritmo determina o valor de d através de testes e por critérios de informações determina os valores de p e q , para mais detalhes ver Cryer & Chan (2008) e Hyndman & Khandakar (2008).

Após a escolha do modelo obtido através da função `auto.arima` pelo critério escolhido é preciso fazer uma análise de diagnóstico para verificar se o modelo representa adequadamente a séries a ser estudada. Assim Ehlers (2005), aponta que após um modelo ter sido ajustado a uma série temporal deve-se verificar se ele fornece uma representação adequada dos dados.

Para verificar se o modelo é adequado aos dados da série temporal, a análise de resíduos é necessária. Os resíduos (ϵ_t) , são obtidos fazendo a diferença entre o valor observado na série de dados (Y_t) e o valor predito pelo modelo (\widehat{Y}_t) , tal que:

$$\epsilon_t = Y_t - \widehat{Y}_t$$

Nos modelos ARIMA, ARMA, AR e MA apresentados o resíduo pode ser reescrito em função dos parâmetros do modelo, onde o mesmo fica definido como o erro de previsão. Por exemplo, em um modelo AR(1), se $\hat{\phi}$ é a estimativa do coeficiente autorregressivo então o valor ajustado no tempo t é $\hat{\phi}y_{t-1}$ e o resíduo correspondente $\epsilon_t = y_t - \hat{\phi}y_{t-1}$.

Se o modelo tiver um bom ajuste, espera-se que os resíduos se distribuam aleatoriamente em torno de zero, com variância constante e sejam não correlacionados (ruído branco). O termo Ruído Branco se refere a uma sequência de variáveis aleatórias não correlacionadas que possuem média zero e variância constante. Então o aspecto básico para a construção de modelos de séries temporais é este tipo de sequência. Para isto, a necessidade de testar se os resíduos estão ou não correlacionados. Em Ehlers (2005) é apresentado que correlação serial nos resíduos é uma indicação de que uma ou mais características da série não foram adequadamente descritas pelo modelo.

Uma maneira de verificar se os resíduos apresentam correlação é através do gráfico da função de autocorrelação (FAC). A FAC é uma medida de correlação entre as observações de uma série temporal que são separadas por t -s unidades representando os lags presente no gráfico. Esta medida permite uma análise mais detalhada da estrutura de autocorrelação, indicando possíveis dependências dos termos no modelo e o gráfico dos resíduos poderá revelar a presença de dados discrepantes, efeitos de correlação ou padrões cíclicos. Para mais detalhes sobre FAC e o gráfico dos resíduos ver Ehlers (2005), Cryer & Chan (2008), Brockwell & Davis (2016), Morettin & Toloí (2018).

O exemplo a seguir ajuda a entender melhor como utilizar o gráfico da FAC para verificar se os resíduos são correlacionados e o gráfico dos resíduos. Nesse exemplo, é considerado um modelo AR de ordem $p = 1$, com o parâmetro estimado $\hat{\phi} = 0,5$ e este modelo foi gerado a partir das observações considerando $n = 100$. Na Figura 2.1 é apresentado o gráfico dos resíduos, note que não há presença de padrões e os resíduos estão aleatoriamente distribuídos em torno de uma constante. Na Figura 2.2 é apresentado o gráfico da FAC dos resíduos, note que as linhas pontilhadas representam os limites que indica que os resíduos não apresentar correlação. Como todos os lags estão dentro dos limites, concluímos que os resíduos desse exemplo não apresentam autocorrelação. Assim como mostrado na Figura 2.1 e 2.2 o modelo é adequado para representar o fenômeno em estudo.

Além de observar individualmente as correlações dos resíduos através do FAC, é útil ter um teste que considere suas magnitudes como um grupo. Para isto, uma alternativa, é o teste proposto por Ljung & Box (1978). Por exemplo, pode ser que ao verificar a FAC dos resíduos obtidos do modelo de series temporais as autocorrelações residuais sejam moderadas entre as defasagens, ou seja os lags estejam dentro dos limites, mas quando

Figura 2.1: Exemplo Resíduo modelo AR(1).

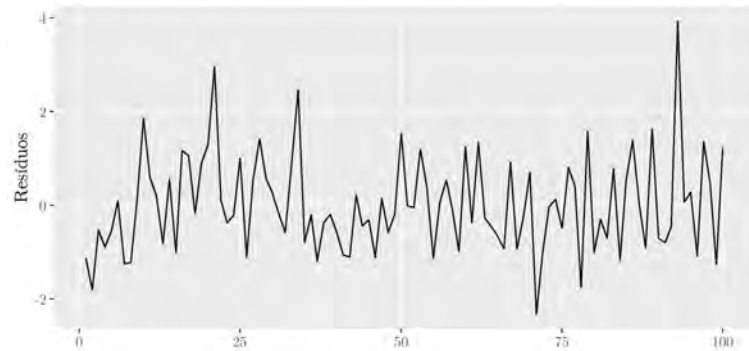
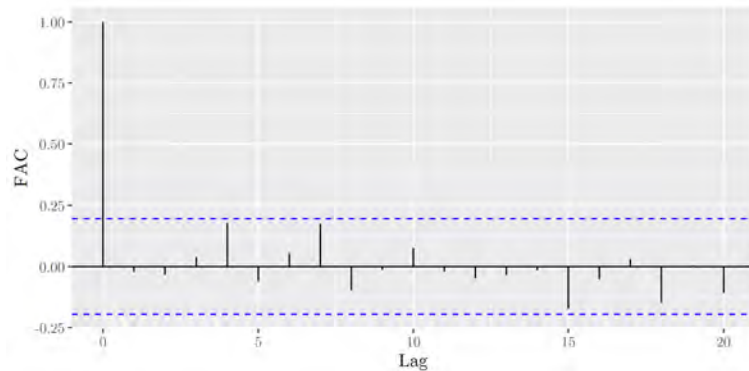


Figura 2.2: Exemplo FAC modelo AR(1).



verificado em conjunto podem apresentar correlação nos resíduos, em que k representa a unidade de tempo entre as observações para determinar conjunto que será verificado. As hipóteses do teste de Ljung-Box são:

$$\begin{cases} H_0 : & \text{Os resíduos são não correlacionados;} \\ H_1 : & \text{Os resíduos são correlacionados} \end{cases}$$

No exemplo visto anteriormente com os resíduos obtidos a partir do modelo AR(1), que mostramos nas Figuras 2.1 e 2.2, o teste de Ljung-Box apresenta um p -valor de 0,4857. Então considerando um nível de significância de 5%, não há evidências estatísticas para rejeitar a hipótese que os resíduos são não correlacionados. Para finalizar a análise de modelo de series temporais, é preciso testar se os resíduos seguem distribuição normal. Na literatura, há vários testes de normalidade que podem ser aplicados, um exemplo é o teste de Kolmogorov Smirnov. Para uma revisão detalhada sobre a análise dos resíduos e modelos de series temporais ver Sousa (2021), Ehlers (2005), Morettin & Toloí (2018), Cryer & Chan (2008) e Brockwell & Davis (2016).

2.2 Gráficos de Controle

O controle estatístico de processo (CEP) é uma ferramenta bastante importante para entender o comportamento e tomar decisão sobre o fenômeno. CEP é um sistema de inspeção por amostragem, operando ao longo do processo, visando verificar a presença de causas especiais, ou seja, causas que não são naturais ao processo e podem prejudicar a qualidade do produto, para mais detalhes ver Louzada *et al.* (2013).

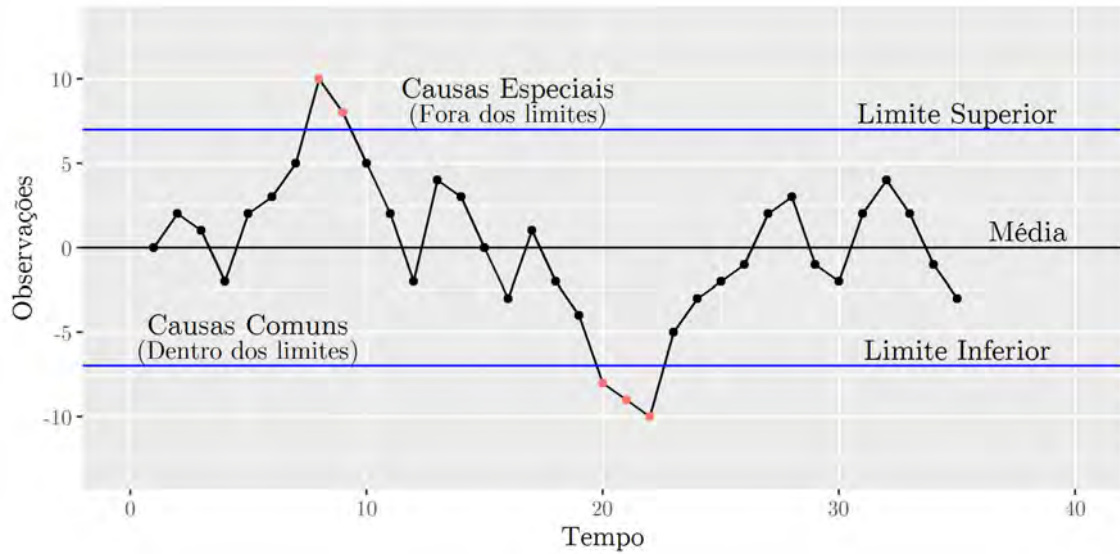
O objetivo desse sistema é a identificação de causas comuns e causas especiais no processo e atuar sobre elas, melhorando os processos e a qualidade final do fenômeno estudado. Para isso, o CEP fornece uma espécie de varredura no processo, identificando sua variabilidade e possibilitando o controle dessa variabilidade ao longo do tempo. De acordo com a ISO 21747 (2006), causas comuns são aquelas fontes de variação aleatórias e intrínsecas ao processo que não são controladas e representam pouca influência no processo. Enquanto que as causas especiais são aquelas fontes de variação oriundas de falhas ocasionais que ocorrem durante o processo, e por sua vez, apresentam grande influência no processo, devendo assim ser controladas. Vale ressaltar que algumas regras são estabelecidas na literatura, onde é possível verificar alguns padrões no comportamento do processo antes de acontecer as causas especiais. Essas regras são: tendência crescente e tendência decrescente, mudança de média para cima e mudança de média para baixo, ciclo, corrida, estratificação e a última regra aberração padrão e aberração pontual. Em Boaventura (2020) é explicado cada uma dessas regras e também é mostrado exemplos.

Gráficos de controle pode ser utilizada em processos que possuem variáveis ou atributos a serem controlados e possui uma conexão com um teste de hipótese que busca saber se o processo está em estado de controle estatístico ou não. Esta ferramenta é bastante utilizada e importante na área de CEP, pois ela facilita visualmente verificar o comportamento do processo (MONTGOMERY, 2009).

O Gráfico de Controle consiste em três linhas paralelas: uma linha média que reflete o nível de operação do processo, ou seja, que representa o valor médio (LC) da característica da qualidade, e duas linhas externas denominadas limite superior de controle (LSC) e limite inferior de controle (LIC), obtidos em função da dispersão da variável sob a média do processo ver Shewhart (1931). Para ajudar a entender foi criado uma ilustração apresentada na Figura 2.3.

De forma geral, seja $X_t, t \in T$ uma variável que será monitorada, seja um valor fixo μ_0 o valor médio desejado para o processo, o gráfico de controle tem seus limites e linha

Figura 2.3: Exemplo de Gráfico de Controle.



central são definidos por:

$$LSC = \mu_0 + L\sigma_x,$$

$$LC = \mu_0,$$

$$LIC = \mu_0 - L\sigma_x,$$

em que, L está associado à largura do limite de controle em relação à variação do processo e σ_x é o desvio padrão da variável observada no processo.

Vale lembrar que nem sempre o valor de μ_0 e a variação do processo σ_x são conhecidos e, por isso, devem ser substituídos por seus valores estimados. Os limites devem ser calculados, de forma geral, tomando como base três desvios padrões de distância em relação à Linha Central ($L = 3$) Shewhart (1931).

Existem diversos tipos de gráficos de controle, porém eles podem ser classificados em dois grandes grupos: gráficos de controle por atributos, em que a característica de qualidade sob estudo é mensurada em uma escala qualitativa, por exemplo, julgar se o processo está conforme ou não conforme; e gráficos de controle para variáveis, os quais a característica de qualidade de interesse é mensurada em uma escala quantitativa (por exemplo: comprimento, diâmetro, etc). Para mais detalhes Samohyl (2009), Gouveia (2018), Ramos, Almeida & Araújo (2012) dentre outros.

Os principais gráficos por atributos estudado na literatura são: o gráfico P , onde a estatística de interesse é a proporção de produtos defeituosos em um subgrupo ou amostra; o gráfico NP que monitora o número total de unidades defeituosas na amostra; o gráfico C que controla o número de defeitos por unidade em um subgrupo e o gráfico U , é usado

para controlar a taxa de defeitos por unidade.

Já os principais gráficos de variáveis apresentado na literatura são: o gráfico \bar{X} , cuja estatística de interesse é o valor médio das características de qualidade; os gráficos R e S , que monitoram a variabilidade do processo com base na amplitude e no desvio padrão, respectivamente; e os gráficos para valores individuais e amplitudes móveis, são utilizados em situações que o tamanho amostral usado para o controle do processo é $n = 1$. Para mais detalhes sobre os gráficos de controle para atributos e gráficos de controle para variáveis ver Montgomery (2009), Samohyl (2009), Gouveia (2018), Ramos, Almeida & Araújo (2012), (LOUZADA *et al.*, 2013).

Como apresentado anteriormente sobre séries temporais, uma combinação bastante útil em avaliar variáveis correlacionada é utilizar modelos de séries temporais e aplicar gráfico de controle nos resíduos do modelo. Como os resíduos obtidos de modelos adequadamente estimados são aproximadamente normal e independentemente distribuídos com média zero e variância constante. Montgomery (2009) apresenta que os gráficos de controle convencionais podem ser aplicados à sequência de resíduos. Por exemplo, o gráfico de controle do tipo Shewhart e o gráfico de Médias Móveis Exponencialmente Ponderadas (MMEP) para medidas individuais.

Para a construção do gráfico do tipo Shewhart para valores individuais utilizamos a média amostral \bar{X} e para a dispersão dos dados utiliza a média das amplitudes móveis \overline{MR} , que é definida pela distância absoluta entre dois valores consecutivos, ou seja,

$$MR = |x_{t+1} - x_t|, t = 1, 2, 3, \dots, m.$$

Sendo assim, temos que os limites de controle do gráfico para valores individuais são dados da seguinte forma,

$$LSC = \bar{X} + \frac{L * \overline{MR}}{d_2},$$

$$LC = \bar{X},$$

e

$$LIC = \bar{X} - \frac{L * \overline{MR}}{d_2},$$

em que d_2 é um valor tabelado que pode ser obtido no apêndice VI em Montgomery (2009). Para mais detalhes sobre gráfico de Shewhart para medidas individuais ver Samohyl (2009), Montgomery (2009), Gouveia (2018), Loh *et al.* (2021).

Quando há necessidade em detectar pequenas mudanças no processo, o gráfico de

média móvel exponencialmente ponderada (MMEP) aparecerá como um substituto para o gráfico de controle do tipo Shewhart. A construção do gráfico MMEP consiste em uma média ponderada de todas as observações anteriores. Essa ponderação ocorre de tal forma que conforme o tempo passa das observações os pesos atribuídos decrescem geometricamente, ou seja, é dado menor peso para observações mais antigas (MONTGOMERY, 2004). Segundo Montgomery (2004), o gráfico do MMEP é ideal para observações individuais e não requer suposições sobre a distribuição das observações, pois este método é não paramétrico. No caso típico de uma amostra de tamanho n igual a 1 a suavização dos dados é realizada da seguinte forma:

$$z_t = \lambda x_t + (1 - \lambda)z_{t-1},$$

em que, x_t representa o valor observado, λ é uma constante que está definida entre $0 < \lambda \leq 1$ e o valor inicial, exigido quando $t = 1$, é o alvo do processo ($z_0 = \mu_0$), ou ainda, quando dados preliminares são utilizados, a média do processo ($z_0 = \bar{X}$).

Essa suavização (z_t) pode ser reescrita de forma a evidenciar o peso dado a cada observação, ela é escrita como:

$$z_t = \sum_{j=0}^{t-1} (1 - \lambda)^j x_{t-j} + (1 - \lambda)^t z_0,$$

E a variável a ser representada no gráfico de controle é a variável suavizada z_t , assim é necessário calcular a variância desses valores suavizados, $\text{Var}(z_i)$. Supondo que as observações (x_t) sejam variáveis aleatórias independentes com variância igual a σ^2 esse cálculo é feito segundo a expressão a seguir:

$$\text{Var}(z_t) = \sigma_{z_t}^2 = \sigma^2 \frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2t}].$$

Dessa forma, o Gráfico de controle MMEP, segundo Montgomery (2009), pode ser construído ao plotar cada um dos valores suavizados (z_t) pelo tempo t em que foi observado. A linha central e os limites de controle são definidos por:

$$LSC = \mu_0 + L\sigma \sqrt{\frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2t}]},$$

$$LC = \mu_0,$$

$$LIC = \mu_0 - L\sigma \sqrt{\frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2t}]},$$

em que, L representa a largura do limite de controle e μ_0 e σ são a média e desvio padrão

das observações. É interessante perceber que após o gráfico MMEP está sendo utilizado por longo período de tempo, o termo nos limites dado anteriormente $[1 - (1 - \lambda)^{2i}]$ se aproxima de 1, (MONTGOMERY, 2009). Então quando isto acontecer os limites de controle se aproximarão dos valores dados por:

$$LSC = \mu_0 + L\sigma\sqrt{\frac{\lambda}{2 - \lambda}},$$

$$LIC = \mu_0 - L\sigma\sqrt{\frac{\lambda}{2 - \lambda}},$$

Para construir o gráfico de controle do MMEP, os valores de λ e L precisam ser definidos para calcular os limites LIC e LSC . A escolha da distância entre os limites de controle e a linha central do gráfico é determinada diretamente pelo valor de λ . Além disso, Montgomery (2009) apontou que valores de λ entre 0,05 e 0,25 apresenta bons resultados e que $\lambda = 0,05$, $\lambda = 0,10$ e $\lambda = 0,20$ são as escolhas mais populares. Uma regra é que um pequeno valor de λ ajuda a detectar pequenas mudanças no valor médio exigido pelo processo. Também afirma que geralmente é eficaz escolher $L = 3$ para valores maiores de λ , mas para valores menores que $\lambda = 0,1$ é vantajoso usar um valor de L definido entre 2,6 e 2,8 para reduzir a largura do limite.

Montgomery (2009), apresenta os resultados dos ARL considerando distribuição normal e não normal das observações, onde é definido alguns valores para λ e L para o gráfico MMEP. Os resultados estão expostos na Tabela 2.1:

Tabela 2.1: ARLs em controle para os gráficos MMEP e o de medidas individuais para várias distribuições gama e distribuição normal.

λ	MMEP			Shewhart
	0,05	0,1	0,2	1
L	2,492	2,703	2,86	3
Normal	370,4	270,8	370,5	370,4
Gama(4,1)	372	341	259	97
Gama(3,1)	372	332	238	85
Gama(2,1)	372	315	208	71
Gama(1,1)	369	274	163	55
Gama(0.1,1)	357	229	131	45

Fonte:(MONTGOMERY, 2009), página 424.

O ARL (comprimento médio de execução) é o número médio de pontos que devem ser plotados antes que um ponto indique uma condição fora de controle. Em Montgomery (2009) é apresentado gráfico de controle em condições que o processo está sob controle e com limites de três sigma $L = 3$, apresentam ARL igual a 370, assim representando um bom valor de ARL. Note que mesmo distribuições moderadamente não normais têm o

efeito de reduzir significativamente o ARL sob controle do gráfico de Shewhart, aumentando drasticamente a taxa de alarmes falsos. O gráfico MMEP apresentam bons valores de ARL para distribuições normais e não normais e os valores de $\lambda = 0,2$ e $L = 2,86$ apresentam bons resultados de ARL. Um bom uso desses dois gráficos de controle em conjunto é em observações com distribuição normal, pois ambos apresentam bons resultados de ARL. Para mais detalhes sobre gráfico MMEP ver Loh *et al.* (2021), Montgomery (2009) dentre outros. Em Montgomery (2009) é mostrado que gráfico do tipo Shewhart para medidas individuais não são sensíveis em detectar pequenas mudanças no processo e recomenda que para melhorar essa sensibilidade, usar gráficos de controle MMEP para medidas individuais. Assim utilizar ambos os gráficos para detectar grande mudanças no processo e usar o gráfico MMEP para medidas individuais para detectar pequenas mudanças no processo.

Quando é utilizado gráficos de controle em dados reais, muitas vezes essas observações dependem de outras variáveis ou são autocorrelacionados. Uma alternativa é aplicar os gráficos de controle aos resíduos de modelos de regressão, muitos estudos fazem uso dessa técnica, para uma revisão mais detalhada listamos alguns estudos a seguir, Boaventura (2020), Lai *et al.* (2021), Pedrini (2009), Park, Jung & Kim (2020), García-Bustos *et al.* (2021) dentre outros. Em Boaventura (2020) é mostrado que quando é usado gráficos de controle em resíduos de modelo de séries temporais o desempenho desse método mostra-se superior quando comparado ao gráfico de controle de regressão. A suposição mais importante com relação aos gráficos de controle é a da independência das observações, pois os gráficos de controle convencionais não funcionam bem se o fenômeno estudado apresentar correlação ao longo do tempo. Esses gráficos de controle irão fornecer resultados enganosos na forma de muitos alarmes falsos se os dados estiverem positivamente correlacionados, para mais detalhes ver (MONTGOMERY, 2009).

Conforme apresentados, a seção referente a séries temporais e a seção sobre gráficos de controle, usar modelos de séries temporais em observações autocorrelacionadas e aplicar gráfico de controle nos resíduos obtidos a partir do modelo estimado, são uma alternativa em monitorar observações correlacionadas. Assim, para estudar o comportamento e propor decisões sobre o fenômeno em observações autocorrelacionadas recomendamos aplicar gráfico de controle do tipo Shewhart e gráfico de controle MMEP para medidas individuais em resíduos de modelos de séries temporais estimados para as observações. Então para aplicar o que foi aprestando nas duas seções os seguintes passos devem ser seguidos:

- 1º** : Fazer uma análise descritiva no processo e gráfico das observações.
- 2º** : Selecionar o modelo de series temporais e estimar os parâmetros do modelo.
- 3º** : Extrair os resíduos do modelo e representar graficamente.
- 4º** : Fazer uma análise de diagnostico verificando presença de correlação dos resíduos, aplicar teste de Ljung-Box e teste de normalidade nos resíduos.
- 5º** : Caso o modelo não passe na análise de resíduos voltar para seleção de modelo, escolher outro modelo e refazer os passos 3 e 4.
- 6º** : Apos passar pelos itens anteriores prosseguir para a construção dos gráficos de controle a partir dos resíduos obtidos do modelo de série temporal selecionado nos passos anteriores. Os gráficos de controle do tipo Shewhart e gráfico MMEP para medidas individuais são aplicados.

Os resultados foram processados com uso do software R **Cran** versão 4.0.2 (R Core Team, 2020). E as funções necessárias foram obtidas através dos pacotes: **qcc** (SCRUCCA, 2004), **forecast** (HYNDMAN; KHANDAKAR, 2008) e **ggplot2** (WICKHAM, 2016).

CAPÍTULO 3

RESULTADOS

Os dados da covid-19 utilizados nesse trabalho foram obtidos a partir de boletins epidemiológicos emitidos pelas Secretarias Estaduais de Saúde (SES). Estes boletins são apresentados para cada estado do Brasil, porém precisam ser tratados e organizados essas informações em um único conjunto de dados. A equipe do Brasil.IO coleta os boletins de cada estado, verifica as informações, organiza em um banco de dados e disponibilizam para pesquisadores realizarem estudos (Brasil.IO, 2021) disponível no site <<https://brasil.io/dataset/covid19/>>. Este conjunto é disponibilizado por dia, a descrição das variáveis usadas nesse estudo estão apresentadas a seguir.

1. state: sigla da unidade federativa.
2. new_confirmed: número de novos casos confirmados.
3. new_deaths: número de novos óbitos.
4. date: data de coleta dos dados no formato YYYY-MM-DD.
5. epidemiological_week: número da semana epidemiológica no formato YYYYWW.

Após obter o conjunto de dados foi feito um tratamento criando outros banco de dados filtrado por estado, por região e para todo o Brasil. A variável de interesse para monitorar e para tomada de decisão é a taxa de letalidade e foi calculada para os estados, regiões e para o Brasil. Esta variável é obtida a partir da divisão do número de óbitos pelo número de casos da doença multiplicado por 100, sendo assim a variável está definida entre 0 e 100. A importância de analisar a letalidade se dá porque esta medida representa a

gravidade da doença e nela está presente o número de casos e o número de mortes, onde uma redução ou aumento nessas variáveis impactam diretamente na letalidade.

Inicialmente a taxa de letalidade foi calculada por dia e alguns problemas apareceram, como, por exemplo, valores negativos, acima de 100, zero, “NaN” e “Inf”. alguns desses problemas pode ser causado por pontos faltantes, já os valores negativos as informações passadas é que naquela localidade houve remanejamento de paciente para outro local. Então a taxa de letalidade foi calculada por semana epidemiológica ao invés de ser por dia, com isto reduzindo ou em alguns casos evitando esses problemas na variável. As semanas epidemiológicas começam no domingo, indo até o sábado, representando os 7 dias da semana. A divisão dos 365 dias do ano em semanas epidemiológicas fica de 52 semanas ou 53 semanas. Assim constituindo o calendário epidemiológico que o ano inicia na semana epidemiológica 1 até a semana 52 ou 53. No início da pandemia ainda apareceram os problemas na letalidade por semana, como valores negativos, acima de 100, zero, “NaN” e “Inf”. Diante disso o início da coleta dos dados foi definido para 03/05/2020 representando a semana epidemiológica 19 do ano de 2020. Os dados obtidos referente a Covid-19 no Brasil, apresentando última atualização na data de 18/09/2021 representando a semana 37 de 2021. Então os dados estão definidos no intervalo da semana epidemiológica 19 de 2020 até a semana epidemiológica 37 de 2021, assim constituindo 72 observações.

Neste trabalho apresentaremos os resultados da taxa de letalidade por semana epidemiológica nos estados da região Nordeste (NE) que é constituído pelos seguintes estados: Alagoas (AL), Bahia (BA), Ceará (CE), Maranhão (MA), Paraíba (PB), Pernambuco (PE), Piauí (PI), Rio Grande do Norte (RN) e Sergipe (SE). A letalidade também foi calculada para a região Nordeste e para o Brasil (BR) que é representado pelos 26 estados da Federação mais o Distrito Federal, assim foram criados 11 conjuntos de dados. As Figuras 3.1 e 3.2 apresentam a taxa de letalidade ao decorrer das semanas epidemiológicas em cada um dos 11 conjuntos de dados.

Note que a taxa de letalidade em cada estados, Nordeste e Brasil apresentam comportamento diferente. Para ilustrar considere os estados da Paraíba e Pernambuco. Onde no ano de 2021 em Pernambuco apresenta um comportamento mais controlado do que no ano de 2020. A Paraíba no final do ano de 2020 apresentou pouca variação e no início de 2021 houve um pico significativo quando comparado com as semanas anteriores. Isto pode ser explicado pelo fato que no início da pandemia cada estado publicou decretos referentes as medidas de controle da Covid, visando suspender atividades não essenciais presencialmente. Para mais detalhes ver Costin & Coutinho (2021).

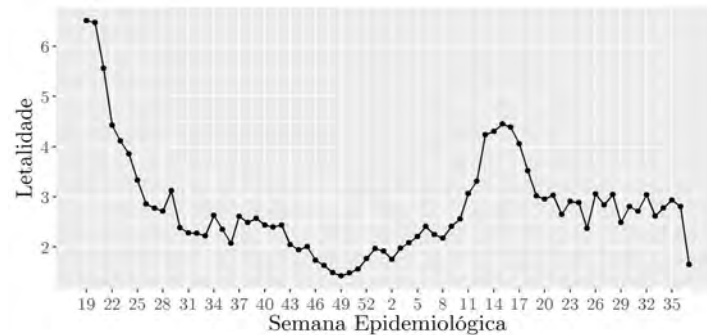
É possível notar que há um padrão que se repete em alguns estados, nas semanas entre 11 e 17 de 2021, há uma elevação na letalidade na Bahia, Maranhão, Paraíba, Rio

Grande do Norte, no Nordeste e no Brasil. Note que nas semanas entre 11 e 17 de 2021, aconteceram os feriados da semana santa e o dia de Tiradentes, estes feriados podem ter contribuído para este aumento na letalidade, devido ao aumento de viagens e aglomeração. Esta situação precisa ser detectada antecipadamente para que os governos locais possam determinar medidas que evitem a elevação da taxa de letalidade por covid-19.

Figura 3.1: Gráfico da taxa de letalidade por semana epidemiológica nos estados, AL, BA, CE, MA, PB, PE, PI, RN e SE e na região nordeste.



Figura 3.2: Gráfico da taxa de letalidade por semana epidemiológica no Brasil.



A Tabela 3.1 apresenta os resultados descritivos da taxa de letalidade nos nove estados da região Nordeste, na região Nordeste e no Brasil, em que mínimo (Min), máximo (Max), a média das observações e o desvio padrão (DP). Note que em média o Brasil apresenta uma taxa de letalidade de 2,801, isto quer dizer que a cada 100 pessoas com covid-19 aproximadamente 3 pessoas morrem pelo vírus, o maior valor na taxa de letalidade foi de 10,559 em Sergipe e o menor valor na taxa de letalidade foi de 0,285 no Rio Grande do Norte. Dos nove estados, região nordeste e Brasil o que apresentou maior valor na média da taxa de letalidade foi em Pernambuco com 3,395, Piauí foi o estado que apresentou menor valor na média que foi de 2,138. E o desvio padrão, em Pernambuco foi o estado com o maior valor que foi de 2,076. Note que a diferença no maior e menor valor da média e do mínimo não é alta, já o máximo apresenta uma diferença muito alta de 4,523 para 10,559.

Tabela 3.1: Resultados descritivos referente a letalidade nos estados estudados, no Nordeste e no Brasil.

	Min.	Max.	Média	DP.
Alagoas	1,008	7,009	2,993	1,411
Bahia	0,766	4,864	2,309	0,848
Ceará	0,484	6,973	2,550	1,444
Maranhão	1,079	8,197	3,036	1,214
Paraíba	1,048	4,863	2,145	0,773
Pernambuco	1,361	9,680	3,395	2,076
Piauí	1,081	4,523	2,138	0,714
Rio Grande do Norte	0,285	6,907	2,495	1,498
Sergipe	0,484	10,559	2,486	1,548
Nordeste	1,034	5,419	2,456	0,894
Brasil	1,424	6,511	2,801	1,025

Fonte: autor

Seguindo os passos descritos na metodologia, o gráfico da série já foi apresentado e a análise descritiva dos mesmos já foram apresentadas. Como essas observações apresentaram comportamentos únicos como períodos com variabilidade baixa e períodos com

variabilidade alta. Estes variáveis epidemiológicas apresentam comportamento de observações autocorrelacionado, então será aplicado os próximos descritos na metodologia que é, aplicar modelos de séries temporais na taxa de letalidade dos nove estados, na taxa de letalidade da região nordeste e na taxa de letalidade do Brasil. Na Tabela 3.2 é apresentado os modelos selecionados com a ordem da parte autorregressiva p , médias móveis q e se necessário d é o número de diferença para a série ficar estacionária. Vale ressaltar que os modelos foram selecionados considerando o critério de informação BIC, pois este critério apresenta modelos mais reduzidos e a seguir seguiremos para o próximo passo que é a extração dos resíduos e análise dos mesmos.

Tabela 3.2: Resultados da ordem do modelo ARIMA selecionado referente a taxa letalidade nos estados estudados, no Nordeste e no Brasil.

ARIMA	AL	BA	CE	MA	PB	PE	PI	RN	SE	NE	BR
(p,d,q)	(1,0,0)	(0,1,1)	(0,1,1)	(0,1,1)	(1,0,0)	(1,1,0)	(1,0,0)	(0,1,1)	(2,0,2)	(1,2,1)	(1,0,0)

Fonte: autor

Figura 3.3: Gráfico dos Resíduos por semana epidemiológica obtidos a partir dos modelos selecionados considerando a taxa de letalidade por estados do Nordeste e para a região Nordeste.

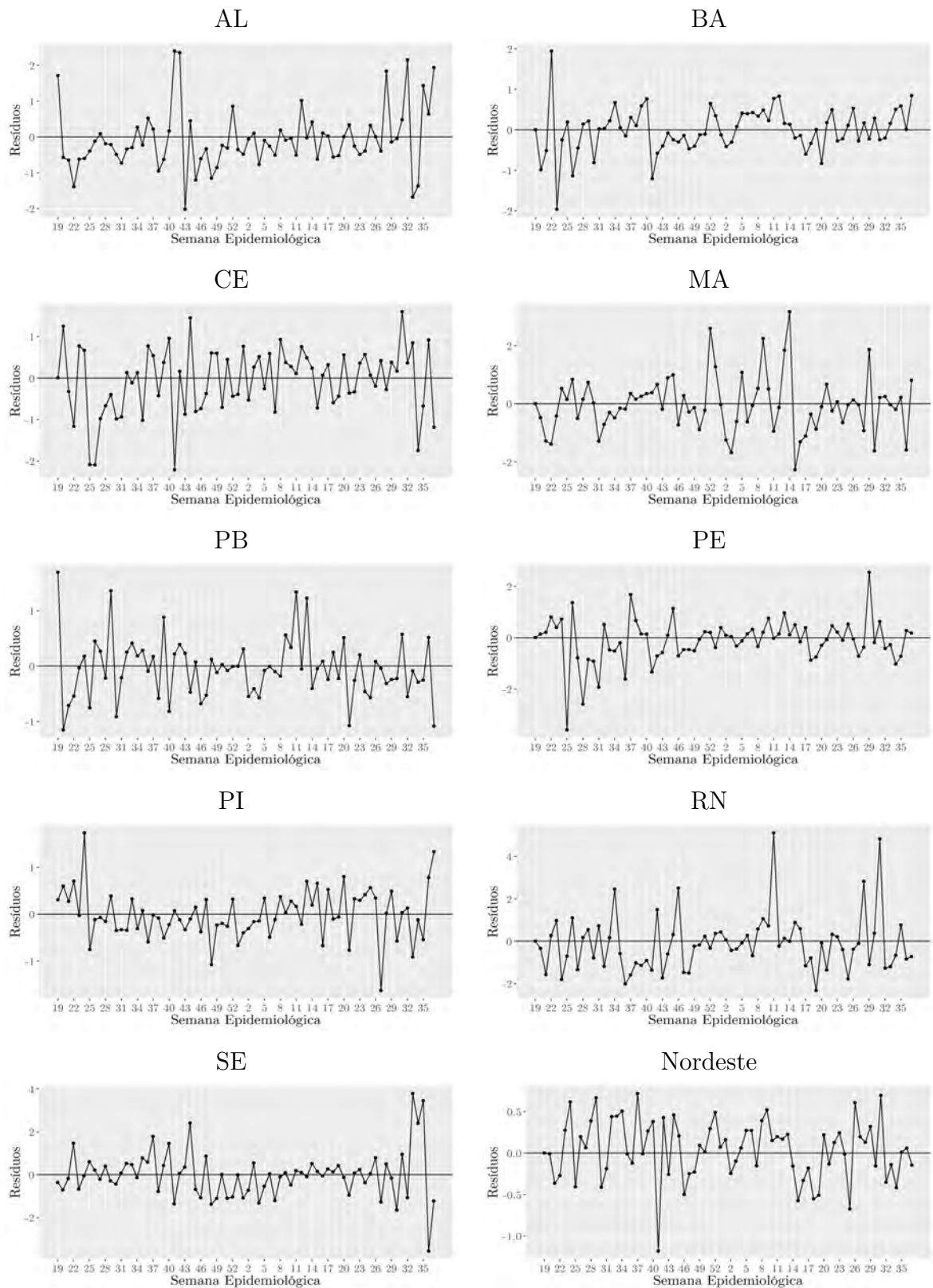
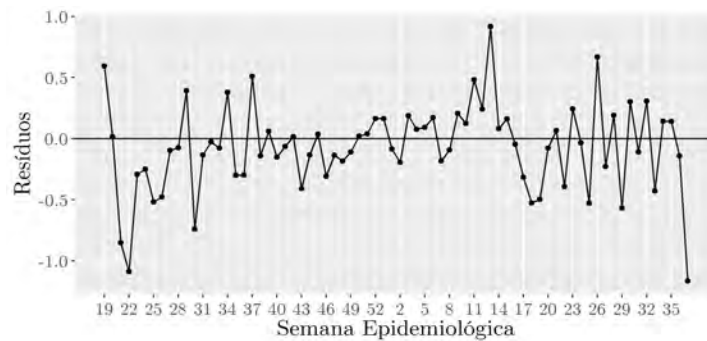


Figura 3.4: Gráfico dos Resíduos por semana epidemiológica obtidos a partir do modelo selecionado considerando a taxa de letalidade no Brasil.



As Figuras 3.3 e 3.4 estão apresentados os gráficos dos resíduos obtidos a partir dos modelos estimados para a taxa de letalidade da Covid-19 nos nove estados da região Nordeste, na região Nordeste e no Brasil. De forma geral os resíduos apresentam variação em torno de uma constante, não apresenta padrões e não tem muitos pontos discrepantes, assim demonstrando um comportamento de ruído branco. Note que os pontos que aparecem nos resíduos e no gráfico da série são os picos podendo ser o máximo ou o mínimo, como, por exemplo, em RN na semana 11 de 2021 foi o máximo no gráfico da série e no resíduo este ponto aparece com o mesmo comportamento. Na Tabela 3.3 os resultados do teste de normalidade usando `ks.test` (Kolmogorov-Smirnov) e do teste de Ljung-Box com $k = 6$ são apresentados para os resíduos dos modelos analisados, em que k representa a unidade de tempo entre as observações para determinar o grupo, por exemplo para $k = 6$ representa grupo de 6 defasagens. Com relação à normalidade ao nível de significância de 5% não há evidências estatísticas de rejeitar a hipótese que os resíduos estudados tenham distribuição normal. Já para verificar a correlação residual em grupo, o teste de Ljung-Box usando $k = 6$ apresentou um p -valor maior que ao nível de significância de 5%, os resíduos não apresentam correlação, quando calculado para um conjunto de 6 lags.

Tabela 3.3: Resultados dos teste de normalidade (KS) e teste de Ljung-Box com $k = 6$ nos resíduos dos modelos.

	AL	BA	CE	MA	PB	PE	PI	RN	SE	NE	BR
Normalidade	0,094	0,610	0,478	0,437	0,677	0,459	0,903	0,213	0,108	0,895	0,510
Ljung-Box	0,869	0,323	0,499	0,183	0,287	0,650	0,333	0,536	0,688	0,138	0,653

Sobre a verificação da correlação individual dos resíduos a FAC função de autocorrelação é apresentada no Apêndice por meio da `tsdiag`, esta função também apresenta o gráfico dos resíduos padronizados e o gráfico com os p -valores do teste de Ljung-Box para diferentes valores de k . De forma geral os resíduos passaram pelos testes, concluindo que eles apresentam comportamento do ruído branco indicando que os modelos estão bem ajustados as observações.

Feito isto, já que os modelos foram bem ajustados as observações o próximo passo é aplicar gráfico de controle nos resíduos dos modelos estimados. A seguir são apresentados os gráficos de controle do tipo Shewhart e MMEP com $L = 2,86$ e $\lambda = 0,2$ nos resíduos obtidos a partir dos modelos estimados e considerando a taxa de letalidade dos nove estados da região Nordeste separadamente, um modelo para a região Nordeste e um modelo para o Brasil considerando os 26 estados da federação mais o distrito federal. Estes gráficos de controle podem ser vistos nas Figuras 3.5 a 3.7.

Figura 3.5: Gráficos de controle do tipo Shewhart e MMEP aplicado aos resíduos obtidos a partir dos modelos estimado para a taxa de letalidade dos estado de Alagoas, Bahia, Ceará, Maranhão e Paraíba, respectivamente.

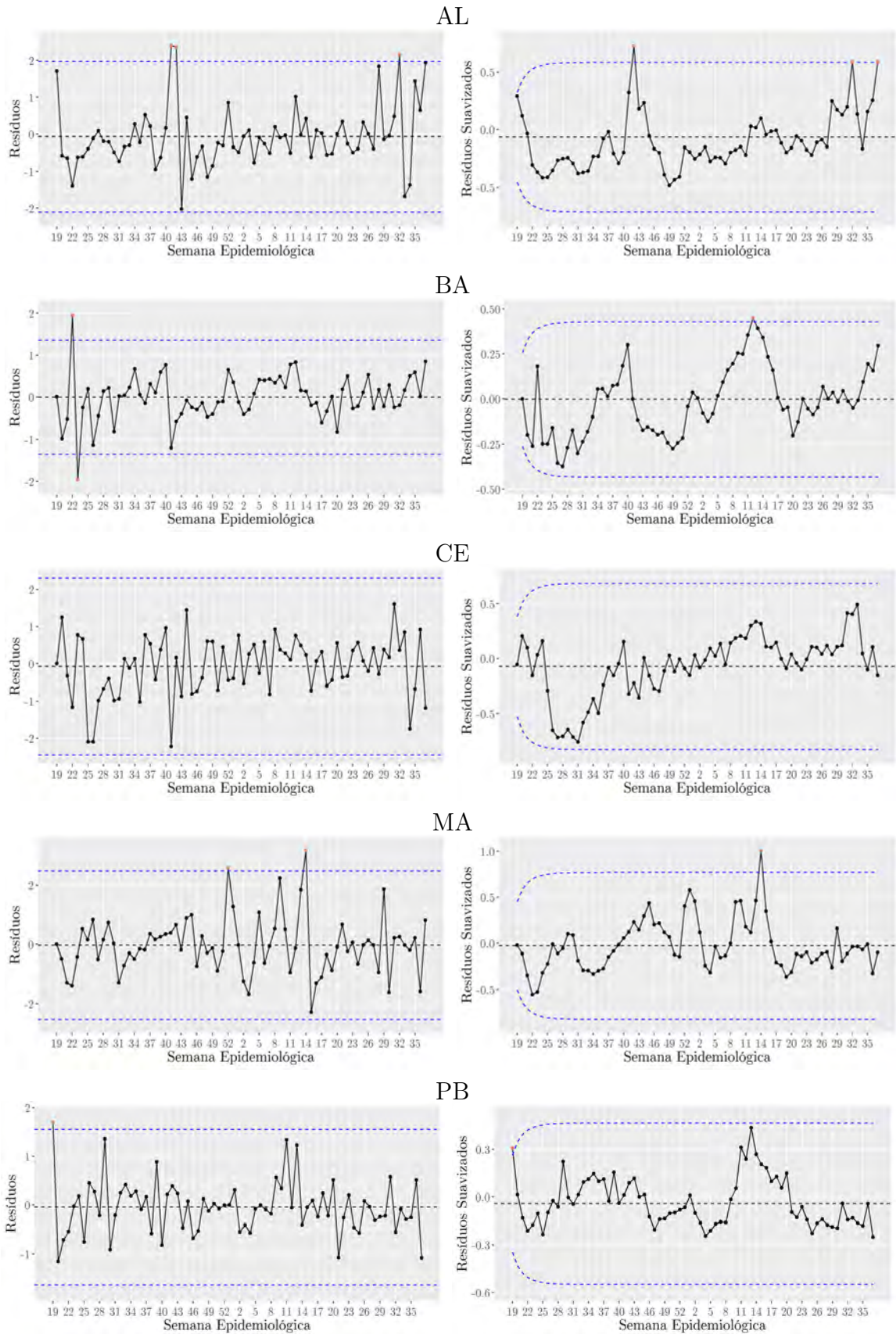


Figura 3.6: Gráficos de controle do tipo Shewhart e MMEP aplicado aos resíduos obtidos a partir dos modelos estimado para a taxa de letalidade dos estado do Pernambuco, Piauí, Rio Grande do Norte, Sergipe e da região Nordeste, respectivamente.

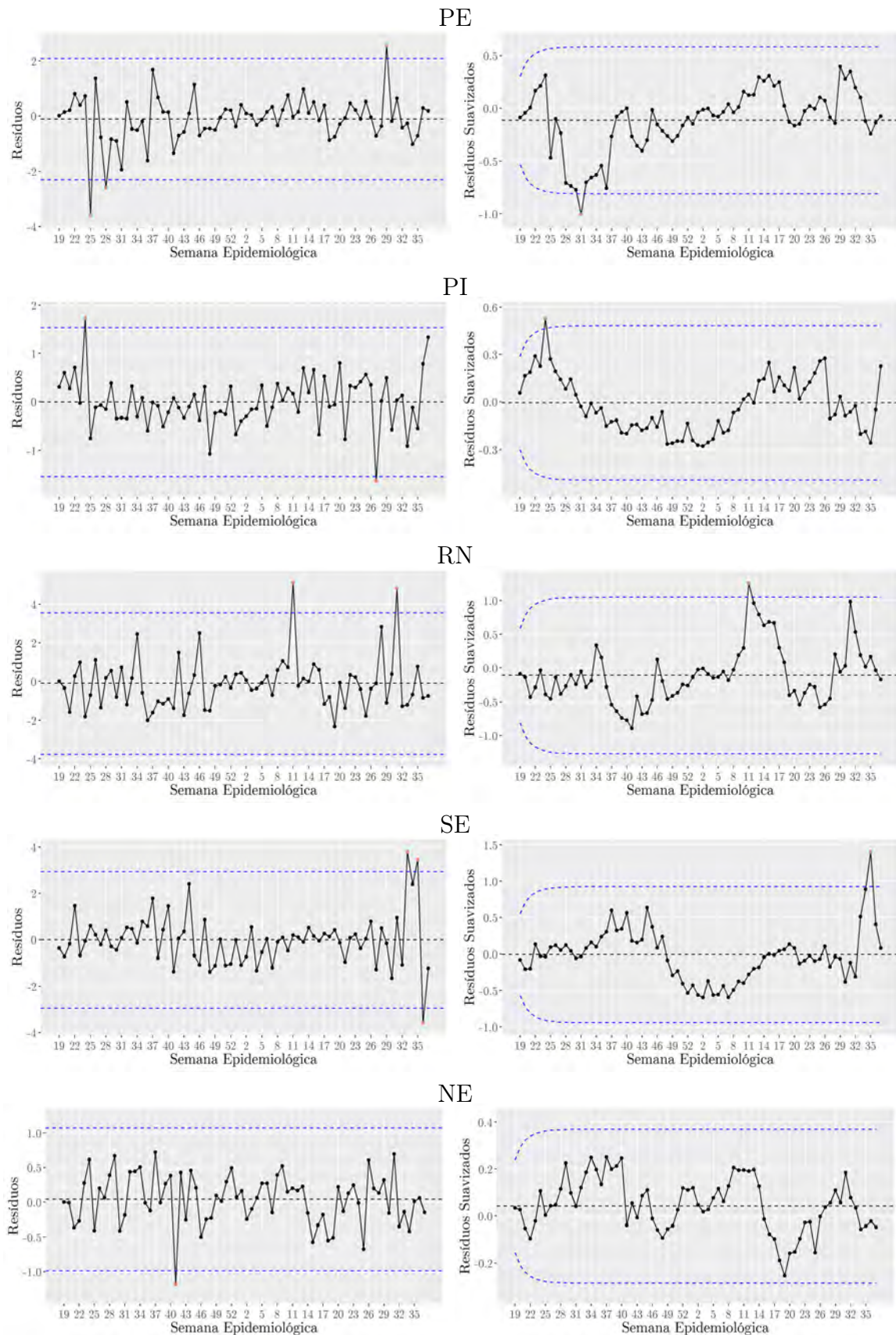
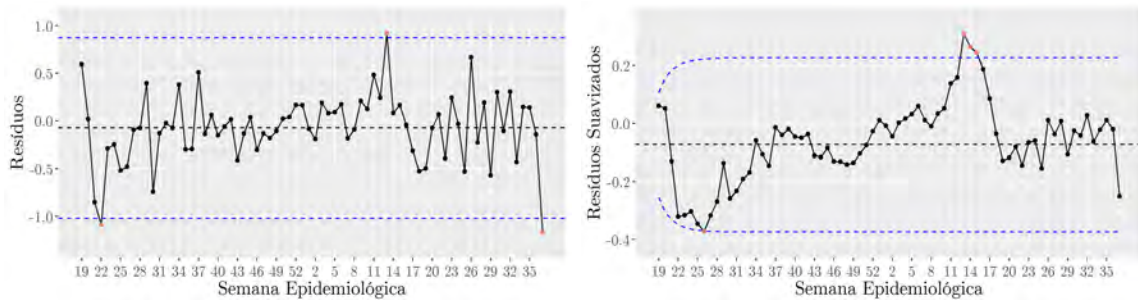


Figura 3.7: Gráficos de controle do tipo Shewhart e MMEP aplicado aos resíduos obtidos a partir do modelo estimado para a taxa de letalidade no Brasil.



Note que a aplicação dos gráficos de controle no resíduo obtidos a partir dos modelos estimados, apresentam bons resultados, pois ao contrário do gráfico das observações, como nas Figuras 3.1 e 3.4, os resíduos não apresentam variação alta e padrões no comportamento que prejudique a construção dos limites. Por exemplo, em Pernambuco no início do período dos dados é apresentado o máximo da letalidade e final do período a letalidade fica controlada e próximo da média, neste caso a construção dos limites de controle nas observações não ia apresentar nenhum ponto fora de controle no final do período. Quando é analisado o gráfico de controle aplicados aos resíduos obtidos do modelo estimado para o estado do Pernambuco isto não acontece, pois, no gráfico de Shewhart a semana epidemiológica 29 de 2021 fica fora de controle como mostrado na Figura 3.6.

O estado do Ceará foi o único dentre os estados estudados, Nordeste e Brasil, que não apresentou semanas epidemiologias fora de controle nos gráficos de controle usados no período estudado. O Brasil foi o que apresentou mais semanas fora de controle, pois com a aplicação dos dois gráficos de controle foi apresentado 7 semanas fora de controle em que no gráfico de Shewhart ficaram 3 semanas fora de controle e no gráfico MMEP ficaram 4 semanas fora de controle, totalizando 7 semanas. Com a aplicação do gráfico de Shewhart os estados de Alagoas, Pernambuco e Sergipe e no Brasil, foram os únicos que apresentaram mais semanas fora de controle para esse gráfico de controle, totalizando 3 semanas. E no gráfico MMEP o Brasil foi o que apresentou mais semanas fora de controle, no total foram 4 semanas fora de controle.

Uma limitação nesta etapa foi a falta de informação para poder descobrir o motivo daquele ponto está fora de controle, precisando que seja especulada o motivo como, por exemplo, na semana santa e o feriado de Tiradentes que aconteceram entre as semanas 11 e 17 de 2021. Neste período os gráficos detectaram pontos fora de controle em alguns estados, como na Bahia a semana 12 ficou fora de controle no gráfico MMEP, no Maranhão a semana 14 ficou fora de controle em ambos os gráficos, no Rio grande do Norte semana 11 ficou fora de controle em ambos os gráficos e no Brasil a semana 13 ficou fora de

controle em ambos os gráficos e as semanas 14 e 15 no gráfico MMEP. Nesses casos o que pode ser explicado é o aumento de viagens e aglomeração podem ter contribuído para estas semanas fora de controle.

Outra coisa bastante interessante, é sobre o uso do gráfico Shewhart e o gráfico MMEP para detectar grande mudanças nas observações e usar o gráfico MMEP para detectar pequenas mudanças no processo. Considerando a taxa de letalidade da Covid-19 no estado da Bahia, note que no gráfico MMEP a semana 11 fica fora de controle e no gráfico de Shewhart isto não acontece como visto na Figura 3.5, assim esta semana apresentou uma pequena mudança nas observações. E em Pernambuco acontece o inverso, o gráfico de Shewhart detecta uma grande mudança no processo que é a semana 29 que sai de controle e no gráfico MMEP esta semana apresenta comportamento em controle, isto pode ser visto na Figura 3.6. Também pode acontecer que os dois gráficos apresentem os mesmos pontos fora de controle como, por exemplo, no Rio grande do Norte na Figura 3.6, a semana 11 fica fora de controle nos dois gráficos. Assim como mostrado o uso dos dois gráficos em paralelo ajudam a detectar casos especiais no processo.

CAPÍTULO 4

CONCLUSÃO

Entender o comportamento de uma doença em uma pandemia é muito importante, pois precisa criar medidas, orientar profissionais de saúde permitindo diagnosticar a situação da doença e a manter em controle. Neste propósito foi apresentada a utilização de Séries Temporais e Gráficos de Controle Estatístico aplicados a taxa de letalidade da Covid-19 em nove Estados da Região Nordeste, na região Nordeste e no Brasil.

Então, foram obtidos os dados referentes a covid-19 nos estados do Brasil através de boletins epidemiológicos divulgados pelas secretarias estaduais de cada estado. Com os dados coletados foram filtradas as variáveis por semana epidemiológicas no período da semana de 19 de 2020 até a semana 37 de 2021 e foi calculado a taxa de letalidade para cada estado, para cada região e para o Brasil. Neste trabalho foram estudados a taxa de letalidade em cada um dos nove estados da região Nordeste, a taxa de letalidade do Nordeste e do Brasil, representando 11 conjunto de dados. Após apresentar o gráfico da taxa de letalidade dos 11 conjunto de dados foi feito uma análise descritiva para cada um deles. E foi selecionado modelos de séries temporais por meio do critério BIC para cada um dos conjuntos, foi estimado os parâmetros dos modelos e foi feito uma análise de resíduos em cada modelo. Após a análise de resíduos nos modelos, pois os mesmos representam medidas individuais, foram aplicados gráfico de controle do tipo Shewhart e gráfico MMEP aos resíduos dos modelos estimados. Assim apresentando resultados dos dois gráficos de controle para cada um dos nove estados da região Nordeste, para o Nordeste e para o Brasil.

A análise descritiva resultados preocupantes, pois semanas com letalidade de 10,559 que foi no estado de Sergipe, precisam ser evitadas. Com relação à média da letalidade, a

média mais alta dos 11 conjuntos foi de 3,395 em Pernambuco, em que essa média é alta e precisa ser monitorada. A variação apresentada dos Estados também é preocupante como em Pernambuco o desvio padrão foi de 2,076. A recomendação é de evitar esses tipos de situações principalmente período de alta, onde se acontecer, descobrir o que causou e evitar. A aplicação dos gráficos de controle nos resíduos se mostraram ser muito úteis, pois apresentaram pontos fora de controle que no gráfico da série não era perceptível. A taxa de letalidade calculada para o Brasil no período estudado foi o que apresentou mais pontos fora de controle, que foram 7 semanas fora de controle, onde o gráfico de Shewhart apresentou 3 semanas fora de controle e o gráfico MMEP apresentou 4 semanas fora de controle. Já o estado do Ceará foi o único que não apresentou pontos fora de controle nos dois gráficos de controle aplicados aos resíduos obtidos dos modelos estimados para a taxa de letalidade desse estado. Outro resultado importante foi que os gráficos de controle detectaram pontos fora de controle no mesmo intervalo entre as semanas 11 e 17 de 2021, nos estados da Bahia, Maranhão, Rio Grande do Norte e no Brasil. Nesse período aconteceram a semana santa e o feriado de Tiradentes, assim com o aumento de viagens e aglomeração podem ter contribuído para estas semanas ficarem fora de controle. Uma combinação importante foi que os dois gráficos contribuirão para detectar grandes mudanças na taxa de letalidade e o gráfico MMEP foi muito útil em detectar os pontos afetados por pequenas mudanças na taxa de letalidade.

A combinação do uso do gráfico de controle nos resíduos de modelos de séries temporais se mostrou muito úteis, pois variáveis epidemiológicas apresentam comportamento de autocorrelação e isto é uma limitação na aplicação de gráficos de controle direto nas observações. Então essas duas ferramentas são uma alternativa em analisar variáveis epidemiológicas, os resultados apresentados nesse trabalho se mostraram muito importante, pois com o estudo do motivo que aconteceu para aquela semana ficar fora de controle pode ser evitado para que não apareça outras semanas fora de controle. Assim essas duas ferramentas ajudando na criação de medidas como, por exemplo, foi detectado semanas fora de controle nos feriados, estudado o motivo e aplicar medidas antecipadamente para assim manter a taxa de letalidade da covid-19 sob controle. Dessa maneira, gráfico de controle com séries temporais, contribuirá para o fortalecimento das ações na taxa de letalidade do coronavírus, tendo em vista a possibilidade de, rapidamente, fornecer resultados para governos e profissionais de saúde tomarem as devidas decisões com relação às ações a serem adotadas.

Por fim, como proposta de trabalhos futuros, iremos aplicar estas ferramentas para outros estados e outros países, fazer um estudo de simulação gerando dados autocorrelacionados e comparando esta técnica com outras como, utilizam modelos de regressão, ou gráficos aplicados diretamente nos dados e estudar as vantagens.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGARWAL, N. *et al.* Indoor air quality improvement in covid-19 pandemic: Review. **Sustainable Cities and Society**, v. 70, p. 102942, 2021. ISSN 2210-6707. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2210670721002274>>.
- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. **2nd International Symposium on Information Theory**, 1973.
- BAYYURT, L.; BAYYURT, B. Forecasting of covid-19 cases and deaths using arima models. **medRxiv**, Cold Spring Harbor Laboratory Press, 2020. Disponível em: <<https://www.medrxiv.org/content/early/2020/04/22/2020.04.17.20069237>>.
- BENVENUTO, D. *et al.* Application of the arima model on the covid-2019 epidemic dataset. **Data in Brief**, v. 29, p. 105340, 2020. ISSN 2352-3409. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S2352340920302341>>.
- BIRYUKOV, J. *et al.* Increasing temperature and relative humidity accelerates inactivation of sars-cov-2 on surfaces. **mSphere**, v. 5, n. 4, p. e00441–20, 2020.
- BOAVENTURA, L. L. **Gráficos de controle de aprendizagem dos processos: uma flexibilização de CEP com IA**. Dissertação (Mestrado) — Universidade Federal da Bahia, IME - UFBA, Fevereiro 2020. Disponível em: <<http://repositorio.ufba.br/ri/handle/ri/33499>>.
- BOX, G. E. P.; JENKINS, G. M. Time series analysis: forecasting and control. **Holden-day**, 1970.
- BRASIL. Decreto legislativo, nº 6, de 2020. **Diário Oficial da União**, Brasília, DF, 2020. ISSN 1677-7042. Disponível em: <<https://www.in.gov.br/en/web/dou/-/decreto-legislativo-249090982>>.
- Brasil.IO. **Boletins epidemiológicos da COVID-19 por município por dia**. 2021. Disponível em: <<https://brasil.io/dataset/covid19/>> (última atualização: 28 de 10 de 2021, acesso em 28 de 10 de 2021).

- BROCKWELL, P.; DAVIS, R. **Introduction to Time Series and Forecasting**. [S.l.]: Springer International Publishing, 2016. (Springer Texts in Statistics). ISBN 9783319298542.
- CHAN J. S. MALIK PEIRIS, S. Y. L. L. L. M. P. K. Y. Y. W. H. S. K. H. The effects of temperature and relative humidity on the viability of the sars coronavirus. **Advances in Virology**, v. 2011, p. 7, 2011.
- CHEN, N. *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study. **The Lancet**, v. 395, n. 10223, p. 507–513, 2020. ISSN 0140-6736. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0140673620302117>>.
- COSTIN, C.; COUTINHO, A. Experiences with risk-management and remote learning during the covid-19 pandemic in brazil: Crises, destitutions, and (possible) resolutions. In: _____. **Primary and Secondary Education During Covid-19: Disruptions to Educational Opportunity During a Pandemic**. Cham: Springer International Publishing, 2021. p. 39–78. ISBN 978-3-030-81500-4. Disponível em: <https://doi.org/10.1007/978-3-030-81500-4_2>.
- CRYER, J.; CHAN, K. **Time Series Analysis: With Applications in R**. [S.l.]: Springer, 2008. (Springer Texts in Statistics). ISBN 9780387759586.
- EHLERS, R. **Análise de Séries Temporais**. Dissertação (Mestrado) — Departamento de Estatística, UFPR, Fevereiro 2005. Disponível em: <<http://leg.est.ufpr.br/~ehlers/notas>>.
- FERRAZ, C. *et al.* Covid-19: Monitoramento por gráficos de shewhart. **Revista Brasileira de Estatística**, v. 78, n. 245, p. 23–41, 2020. ISSN 2675-3243. Disponível em: <<https://rbes.ibge.gov.br/>>.
- GARCÍA-BUSTOS, S. *et al.* A control chart based on pearson residuals for a negative binomial regression: application to infant mortality data. **International Journal of Quality & Reliability Management**, 2021.
- GOUVEIA, M. A. da C. **Controle estatístico da qualidade**. [S.l.]: Editora e Distribuidora Educacional S.A, 2018. ISBN 978-85-522-0585-2.
- GUILLIER, L. *et al.* Modeling the inactivation of viruses from the <i>coronaviridae</i> family in response to temperature and relative humidity in suspensions or on surfaces. **Applied and Environmental Microbiology**, v. 86, n. 18, p. e01244–20, 2020.
- HE, Z. *et al.* The influence of average temperature and relative humidity on new cases of covid-19: Time-series analysis. **JMIR Public Health Surveill**, v. 7, n. 1, p. e20495, Jan 2021. ISSN 2369-2960.
- HYNDMAN, R. J.; KHANDAKAR, Y. Automatic time series forecasting: The forecast package for r. **Journal of Statistical Software**, v. 27, n. 3, p. 1–22, 2008. Disponível em: <<https://www.jstatsoft.org/index.php/jss/article/view/v027i03>>.
- ISO 21747. **Process performance and capability statistics for measured quality characteristics**. [S.l.], 2006. Disponível em: <<https://www.iso.org/standard/34685.html>>.

LAI, X. *et al.* Residuals based ewma control charts with risk adjustments for zero-inflated poisson models. **Quality and Reliability Engineering International**, 2021.

LJUNG, G.; BOX, G. On a measure of lack of fit in time series models. **Biometrika**, v. 65, 08 1978.

LOH, H. C. *et al.* Positive global environmental impacts of the covid-19 pandemic lockdown: a review. **GeoJournal**, Springer, p. 1–13, 2021.

LOUZADA, F. *et al.* **Controle Estatístico de Processos-Uma Abordagem Prática para Cursos de Engenharia e Administração**. [S.l.]: LTC, 2013. ISBN 9788521622048.

Ministério da Saúde. **Brasil confirma primeiro caso da doença**. 2021. Disponível em: <<https://www.gov.br/saude/pt-br/assuntos/noticias/brasil-confirma-primeiro-caso-de-novo-coronavirus>>. Acesso em: 21 novembro 2021.

Ministério da Saúde. **O que é a Covid-19?** 2021. Disponível em: <<https://www.gov.br/saude/pt-br/coronavirus/o-que-e-o-coronavirus>>. Acesso em: 21 novembro 2021.

MONTGOMERY, D. **Introduction to Statistical Quality Control**. [S.l.]: Wiley, 2009. ISBN 9780470233979.

MONTGOMERY, D. C. Design and analysis of experiments. **John Wiley and Sons Inc**, 2001.

MONTGOMERY, D. C. **Introdução ao controle estatístico da qualidade**. [S.l.]: Grupo Gen - LTC, 2004.

MONTGOMERY, D. C.; JOHNSON, L. A. **Forecasting and Time Series Analysis**. [S.l.]: McGraw-Hill, 1976. ISBN 9780070428577.

MORETTIN, P.; TOLOI, C. **Análise de séries temporais: modelos lineares univariados**. [S.l.]: Blucher, 2018. ISBN 9788521213529.

NOREEN, S.; SAQIB, S. Covid-19 and equity trading decisions: A control chart analysis. **Asian Finance Research Journal**, 2021.

PARK, K.; JUNG, D.; KIM, J.-M. Control charts based on randomized quantile residuals. **Applied Stochastic Models in Business and Industry**, v. 36, n. 4, p. 716–729, 2020.

PEDRINI, D. C. **Proposta de um método para aplicação de gráficos de controle de regressão no monitoramento de processos**. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, 2009. Disponível em: <<http://hdl.handle.net/10183/17154>>.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2020. Disponível em: <<https://www.R-project.org/>>.

RAMOS, E. M. L. S.; ALMEIDA, S. dos Santos de; ARAÚFO, A. dos R. **Controle Estatístico da Qualidade**. [S.l.]: bookman, 2012. ISBN 978-85-65837-45-3.

- SAMOHYL, R. W. **Controle estatístico da qualidade**. [S.l.]: Elsevier, 2009. ISBN 978-85-352-3220-2.
- SARKODIE, S. A.; OWUSU, P. A. Impact of meteorological factors on covid-19 pandemic: Evidence from top 20 countries with confirmed cases. **Environmental Research**, v. 191, p. 110101, 2020. ISSN 0013-9351. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0013935120309981>>.
- SCRUCCA, L. qcc: an r package for quality control charting and statistical process control. **R News**, v. 4/1, p. 11–17, 2004. Disponível em: <<https://cran.r-project.org/doc/Rnews/>>.
- SHEWHART, W. Economic control of quality of manufactured product. **ASQ Quality Press.**, 1931.
- SOUSA, A. R. dos S. **Análise de séries temporais**. [S.l.]: Grupo A, 2021. ISBN 9786556902876.
- TZAMPOGLOU, P.; LOUKIDIS, D. Investigation of the importance of climatic factors in covid-19 worldwide intensity. **International Journal of Environmental Research and Public Health**, v. 17, n. 21, 2020. ISSN 1660-4601. Disponível em: <<https://www.mdpi.com/1660-4601/17/21/7730>>.
- VELOSO, L. T. **Modelos de Séries Temporais e Gráficos de Controle Estatístico aplicados a Indicadores de Vigilância Epidemiológica do Ministério da Saúde**. 2018. Disponível em: <https://bdm.unb.br/bitstream/10483/24566/1/2018_LuizaTulerVeloso_tcc.pdf>.
- WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. Disponível em: <<https://ggplot2.tidyverse.org>>.
- WOODALL, W. H.; MONTGOMERY, D. C. Research issues and ideas in statistical process control. **Journal of Quality Technology, Taylor & Francis**, v. 31, n. 4, p. 376–386, 1999. Disponível em: <<https://doi.org/10.1080/00224065.1999.11979944>>.
- YANG, X.-D.; LI, H.-L.; CAO, Y.-E. Influence of meteorological factors on the covid-19 transmission with season and geographic location. **International Journal of Environmental Research and Public Health**, v. 18, n. 2, 2021. ISSN 1660-4601. Disponível em: <<https://www.mdpi.com/1660-4601/18/2/484>>.
- ZHANG, X. *et al.* The impact of weather condition and social activity on covid-19 transmission in the united states. **Journal of Environmental Management**, v. 302, p. 114085, 2021. ISSN 0301-4797. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0301479721021472>>.
- ZHENG, H.-L. *et al.* Effects of climate variables on the transmission of covid-19: a systematic review of 62 ecological studies. **Environmental Science and Pollution Research**, Springer, v. 28, n. 39, p. 54299–54316, 2021.

APÊNDICE A

A.1 Resultados `tsdiag` dos estados da região nordeste e Brasil.

Figura A.1: Gráficos dos resultados da função `tsdiag` nos estados de Alagoas e Bahia.

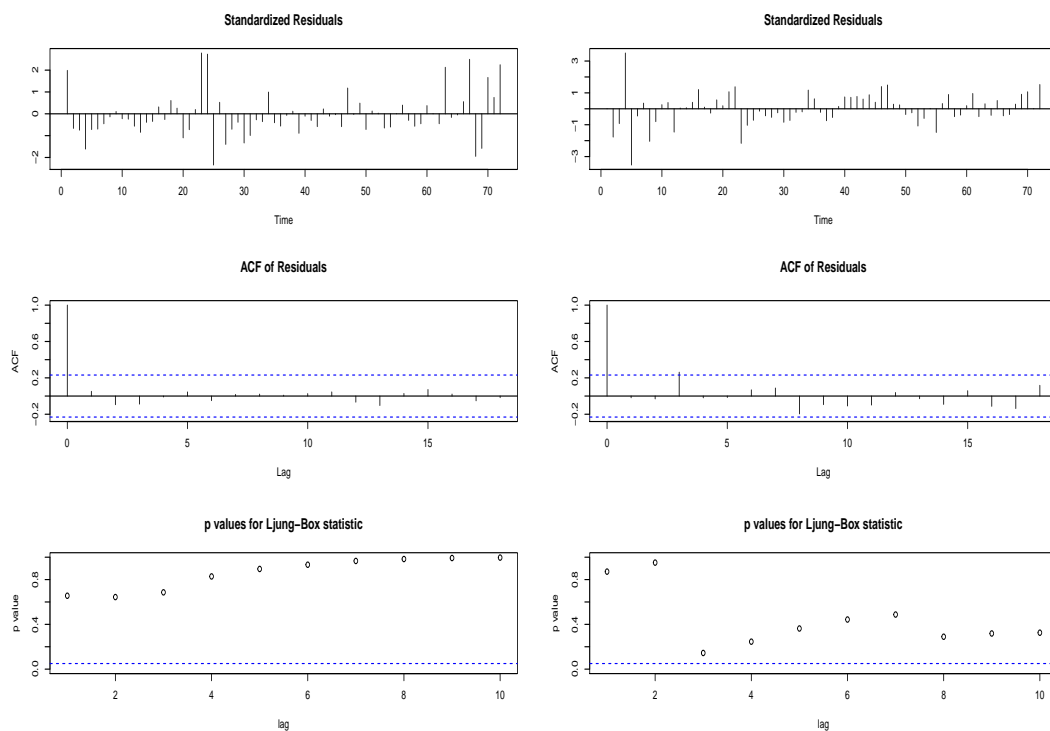


Figura A.2: Gráficos dos resultados da função `tsdiag` nos estados de Ceará e Maranhão.

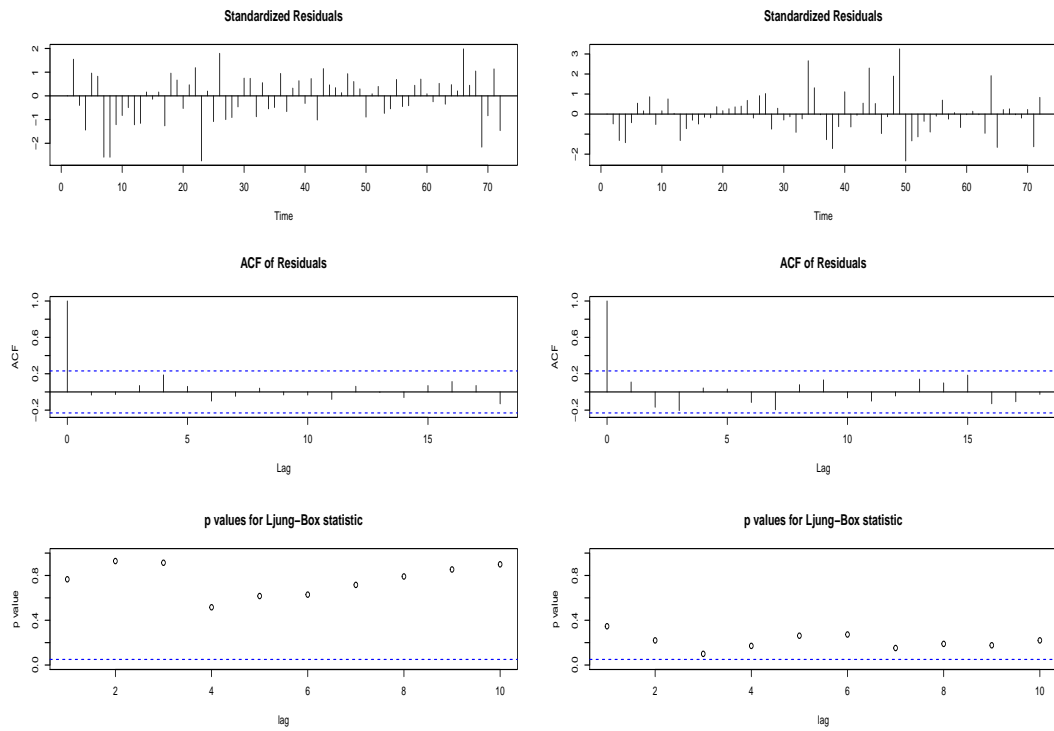


Figura A.3: Gráficos dos resultados da função `tsdiag` nos estados de Paraíba e Pernambuco.

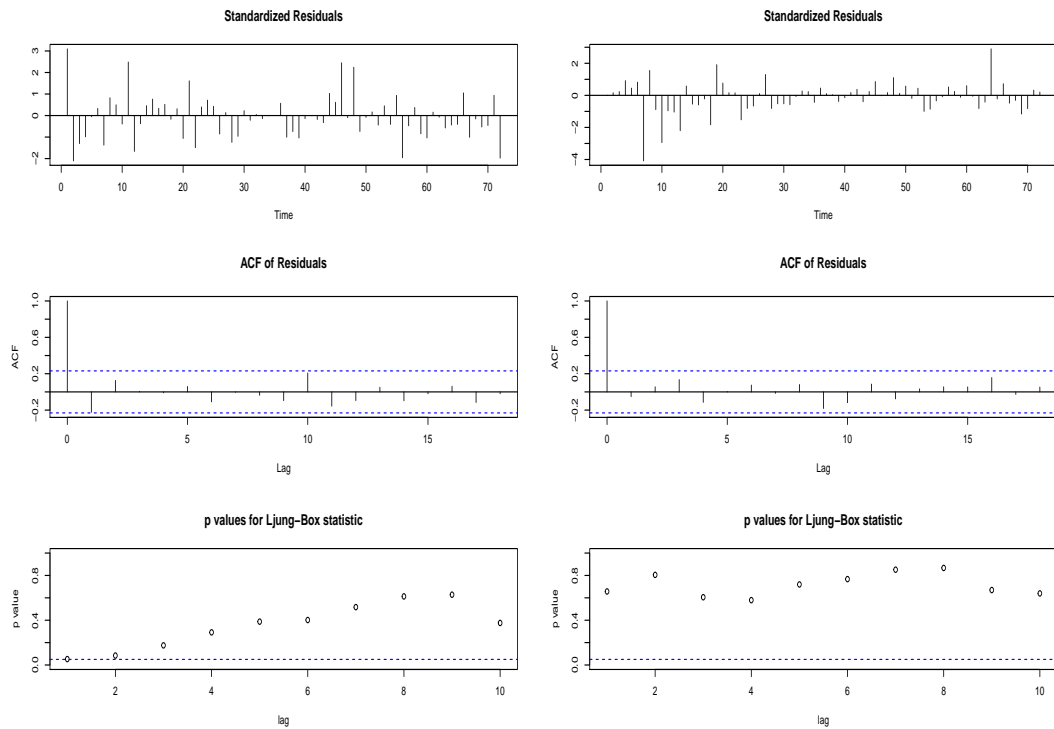


Figura A.4: Gráficos dos resultados da função `tsdiag` nos estados de Piauí e Rio Grande do Norte.

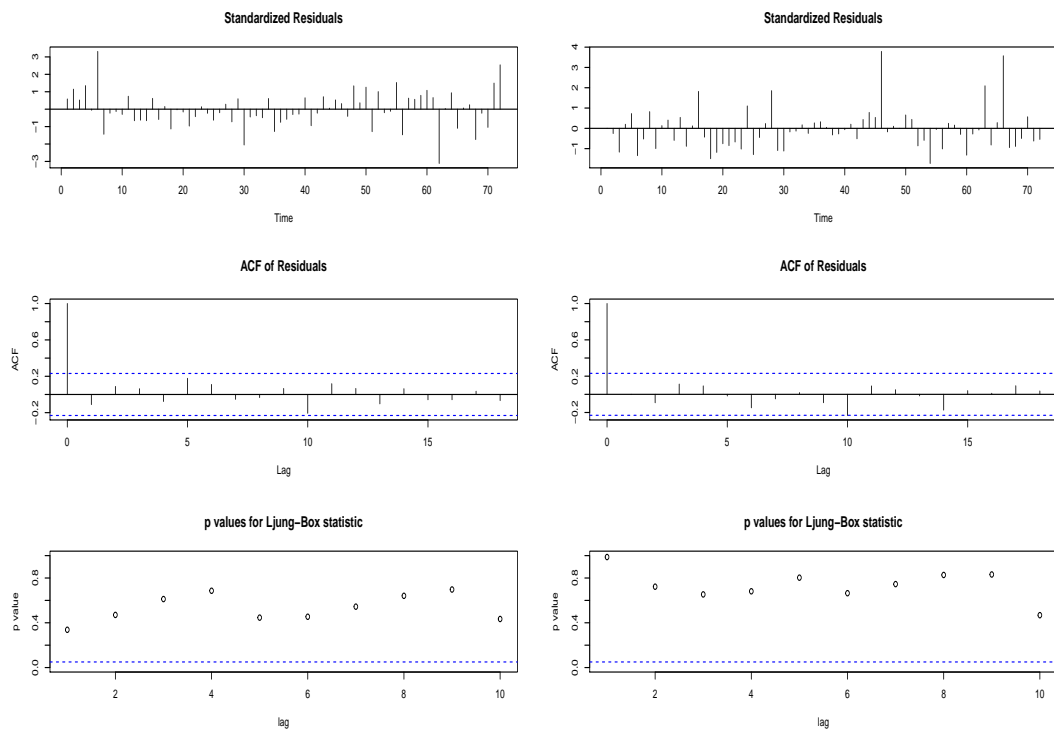


Figura A.5: Gráficos dos resultados da função `tsdiag` no estado de Sergipe e na região Nordeste.

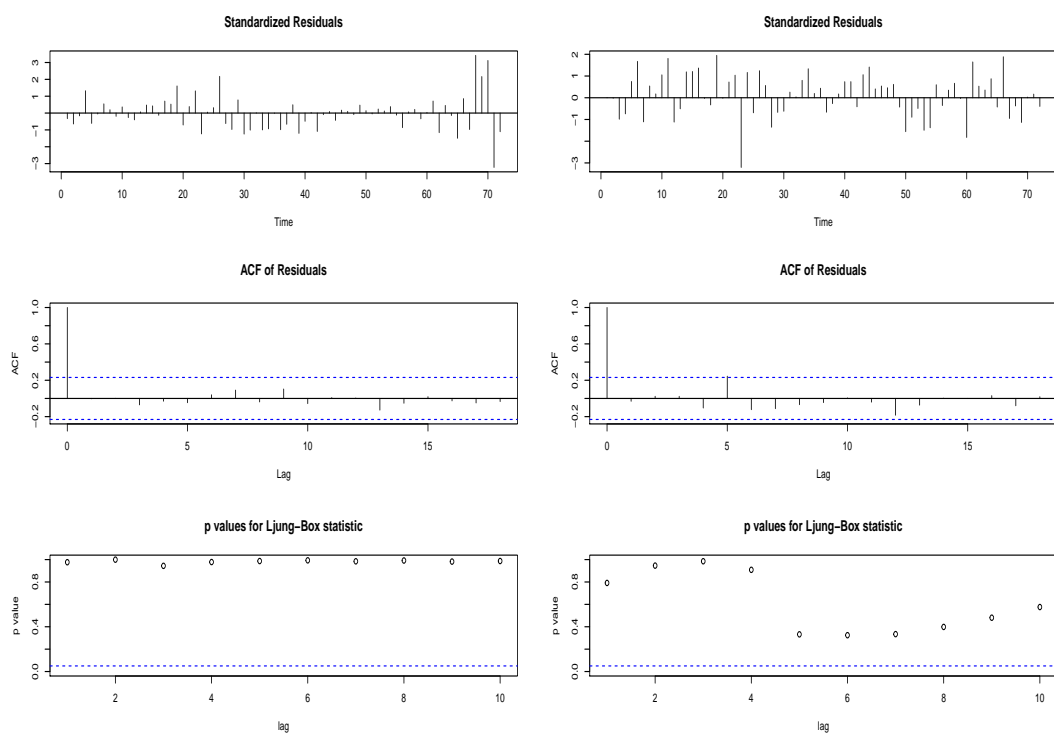
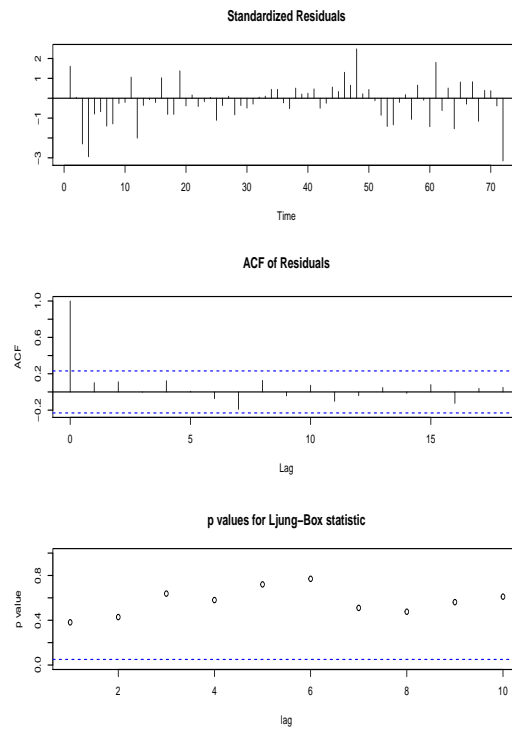


Figura A.6: Gráficos dos resultados da função `tsdiag` no Brasil.

A.2 Script R.

```
##### Script Usado para Obter os Resultados do Trabalho #####
```

```
#### Pacotes necessários
```

```
library(dplyr)
library(data.table)
library(tidyverse)
library(lubridate)
library(forecast)
library(timeSeries)
library(tseries)
library(lubridate)
library(ggplot2)
library(nortest)
library(TSA)
library(qcc)
```

```

##### Carregando banco obtido do Brasil.IO

Dados_Completos<-read.csv2("covid19-CasosFull-todos os estados.csv", header = T, se

##### Filtrando o banco principal em cada estado

### Dados Diários ###
### Por Estados, Regiões e Brasil.

Estados<-unique(Dados_Completos$state) # Coletando os nomes dos estados
Dados_diario<-list() # Lista que conterar todos os estados, região e Brasil
# Processo de filtragem
Dados_Completos<-data.table(Dados_Completos)
Dados_Completos <- Dados_Completos %>% group_by(date)
for(i in Estados){
print(i)
D<-Dados_Completos[which(Dados_Completos$state==i),c(1,2,8,11,10,12)]
Dados_diario[[i]]<-D %>% summarise(
new_confirmed = sum(new_confirmed),
new_deaths = sum(new_deaths),
last_available_deaths=sum(last_available_deaths),
last_available_confirmed=sum(last_available_confirmed),
epidemiological_week=max(epidemiological_week))
};rm(D)

### Carregando as informações para o Brasil
Dados_diario[["Brasil"]]<-Dados_Completos %>% summarise(
new_confirmed = sum(new_confirmed),
new_deaths = sum(new_deaths),
last_available_deaths=sum(last_available_deaths),
last_available_confirmed=sum(last_available_confirmed),
epidemiological_week=max(epidemiological_week)
)

#### Criando lista com o nome dos estados de cada Região
Regiao<-list()
Regiao[["Norte"]]=c('AC', 'AP', 'AM', 'PA', 'RO', 'RR', 'TO')
Regiao[["Nordeste"]]=c('MA', 'PI', 'CE', 'RN', 'PB', 'PE', 'AL', 'SE', 'BA')
Regiao[["Centro-Oeste"]]=c('GO', 'MT', 'MS', 'DF')

```

```

Regiao[["Sudeste"]]=c('ES', 'MG', 'RJ', 'SP')
Regiao[["Sul"]]=c('PR', 'SC', 'RS')
### Processo que carrega as informações de cada Região
for( i in names(Regiao)){
  dados_R<-NULL
  print(i)
  for( j in Regiao[[i]]){
    dados_R_a<-Dados_Completos[which(Dados_Completos$state==j),c(1,2,4,8,11,10,12)]
    dados_R<-rbind(dados_R,dados_R_a)
  }
  Dados_diario[[i]]<-dados_R %>% summarise(
    new_confirmed = sum(new_confirmed),
    new_deaths = sum(new_deaths),
    last_available_deaths=sum(last_available_deaths),
    last_available_confirmed=sum(last_available_confirmed),
    epidemiological_week=max(epidemiological_week))
}; rm(dados_R,dados_R_a)

### O banco original apresenta os dados por dia
### Os processo acima salvam as informações de cada estado, região e Brasil
# em bancos novos na lista Dados_diario

#### Abaixo o processo usa a lista Dados_diario para obter as informações
# por semana endemiológica

Dados_semanal<-list() # Lista que será salvo as informações por semana ep.
for(i in names(Dados_diario)){

  print(i)
  Dados_Comp <- Dados_diario[[i]] %>% group_by(epidemiological_week)

  Dados_semanal[[i]]<- Dados_Comp %>% summarise(
    new_confirmed = sum(new_confirmed),
    #c(max(last_available_confirmed)-min(last_available_confirmed)),
    new_deaths = sum(new_deaths),
    #c(max(last_available_deaths)-min(last_available_deaths)),
    last_available_deaths=last_available_deaths[ length(last_available_deaths)],
    last_available_confirmed=last_available_confirmed[ length(last_available_confirmed)
    date=max(date))
}

```



```

}; rm(D)

#### Os bancos estão filtrados e separados por dia e por semana
### Abaixo procedimentos para calcular a Letalidade

#### Calculando a Variável Var onde var=new_deaths/new_confirmed * 100
# Letalidade Nos Dados por dia
Var_Nova_Diario<-list()
for(i in names(Dados_diario)){
  Var_Nova_Diario[[i]]<-data.table(date=Dados_diario[[i]]$date,
  Var = (Dados_diario[[i]]$new_deaths/
  Dados_diario[[i]]$new_confirmed)*100)
}
# Letalidade Nos Dados por seman
Var_Nova_semanal<-list()
for(i in names(Dados_semanal)){
  Var_Nova_semanal[[i]]<-data.table(date = Dados_semanal[[i]]$date,
  sem = Dados_semanal[[i]]$epidemiological_week,
  Var = (Dados_semanal[[i]]$new_deaths/
  Dados_semanal[[i]]$new_confirmed)*100)
}

### Retirando as primeiras observações em todos os estados e regiões
#### o período que foi setado for 01/04/2020 e a semana 19 de 2020.
for(i in names(Var_Nova_Diario)){
  Var_Nova_Diario[[i]]<-Var_Nova_Diario[[i]][
  -which(Var_Nova_Diario[[i]]$date<"2020-05-01")]
  Var_Nova_semanal[[i]]<-Var_Nova_semanal[[i]][
  -which(Var_Nova_semanal[[i]]$sem<"202019")]
}

#### Criando um banco so para a letalidade de cada estados, regiões e Brasil
# Coluna 1 e 2 sera data e semana as demais colunas serão a letalidade
Dados_Var_Semanal<-cbind(Var_Nova_semanal[["Brasil"]]$date,
  Var_Nova_semanal[["Brasil"]]$sem)
for(i in names(Var_Nova_semanal)){
  Dados_Var_Semanal<-cbind(Dados_Var_Semanal,Var_Nova_semanal[[i]][,3])
}
# colocando os nomes

```

```

colnames(Dados_Var_Semanal)<-c("date", "sem", names(Var_Nova_semanal))
# Salvando o Banco de Letalidade para Analise
write.csv2(Dados_Var_Semanal,file="Dados_Var_Semanal.csv",row.names = F)

#### Proximos pessós é os Resultados
### Funções criadas para os resultados.

# plote da Série
plot.gg<-function(Dados,est,brek=3,...){
Data<-data.frame("x"=epiweek(as.Date(Dados$date)), "y"=Dados[,est])
break_x<-seq(1,nrow(Data),brek)
g<-ggplot(Data,aes(x=1:nrow(Data), y=Data$y),...) +
labs(#title = Estados.nomes[,est],
x='Semana Epidemiológica',y='Letalidade') +
geom_line() + geom_point() +
scale_x_continuous( breaks = break_x,labels = Data$x[break_x]) +
theme(axis.text.x = element_text(angle = 0, hjust = 0.5, vjust = 0.5),
text=element_text(size=12, family="CMU Serif"))

return(g)
}

### Plote para os Resíduo do modelo
plot.gg.resid<-function(Resid,X,brek=3,...){
Data<-data.frame("x"=epiweek(X), "y"=Resid)
break_x<-seq(1,nrow(Data),brek)
g<-ggplot(Data,aes(x=1:nrow(Data), y=Data$y),...) +
labs(x='Semana Epidemiológica',y='Resíduos') +
geom_line() + geom_point() +
geom_hline(yintercept = 0, linetype="solid") +
scale_x_continuous( breaks = break_x,labels = Data$x[break_x]) +
theme(axis.text.x = element_text(angle = 0, hjust = 0.5, vjust = 0.5),
text=element_text(size=12, family="CMU Serif"))
return(g)
}

### Plote Carta Xbar one
plot.xbar.one<-function(Chat,Nomes,ylimt=0.15, X, brek=3,ang=0){
LIC<-Chat$limits[,1]; LSC<-Chat$limits[,2]
Data<-data.frame("x"=epiweek(X), "y"=Chat$data[,1])
break_x<-seq(1,nrow(Data),brek)

```

```

data.col<-rbind(rbind( subset(Data, c(y>LSC)),subset(Data,c(y<LIC))))
p<-ggplot(Data,aes(x=1:nrow(Data), y=Data$y,ymin =min(c( min(Data$y),min(LIC)))-ylim)
labs(#title = paste(Nomes,"- Xbar One"),
x='Semana Epidemiológica',y ='Resíduos') +
geom_line() + geom_point() +
geom_hline(yintercept = Chat$center, linetype="dashed") +
geom_hline(yintercept = c(LIC,LSC), linetype="dashed", colour = "blue") +
scale_x_continuous( breaks = break_x,labels = Data$x[break_x]) +
theme(axis.title.y = element_text(size = 15), axis.title.x = element_text(size = 15),
plot.title = element_text(size = 25), text=element_text(size=12, family="CMU Serif"),
axis.text.x = element_text(angle = ang, hjust = 0.5, vjust = 0.5))
if(nrow(data.col)>=1){
return(p+geom_point(data =data.col, aes(x=as.numeric(row.names(data.col)), y=y, col
}else{
return(p)
}
}

#### Plote para a carta EWMA
plot.ewma<-function(Chat,Nomes, ylimt=0.05, X, brek=3, ang=0){
data.ewma<-data.frame("x"=epiweek(X), "y"=Chat$y)
break_x<-seq(1,nrow(data.ewma),brek)
data.col<-rbind( subset(data.ewma, c(y > Chat$limits[,2] )), subset(data.ewma,c(y <
p<-ggplot(data.ewma,aes(x=1:nrow(data.ewma), y=Chat$y ,ymin =min(c( min(Chat$y),min
labs(#title = paste(Nomes,"- EWMA"),
x='Semana Epidemiológica',y='Resíduos Suavizados') +
geom_line() + geom_point() +
scale_x_continuous( breaks = break_x,labels = data.ewma$x[break_x]) +
geom_hline(yintercept = Chat$center, linetype="dashed") +
geom_line(aes(x=1:nrow(data.ewma), y = Chat$limits[,1]), linetype = "dashed") +
geom_line(aes(x=1:nrow(data.ewma), y = Chat$limits[,2]), linetype = "dashed") +
theme(axis.title.y = element_text(size = 15), plot.title = element_text(size = 25),
axis.text = element_text(size = 15), text=element_text(size=12, family="CMU Serif"),
axis.text.x = element_text(angle = ang, hjust = 0.5, vjust = 0.5))
if(nrow(data.col)>=1){
return(p+geom_point(data =data.col, aes(x=as.numeric(row.names(data.col)),
y=y, colour = "RED"), show.legend = FALSE))
}else{
return(p)
}
}

```

```
}

```

```
#### Carregar o banco de dados salvo dos processos anteriores

```

```
Dados_Var<-read.csv2("Dados_Var_Semanal.csv", header = T, sep=";")

```

```
Nomes<-names(Dados_Var)[c(-1,-2)] # Carregando nomos salvos nos bancos

```

```
### Aqui ja foi definido quais estados serão estudados

```

```
# essas variáveis são necessárias para os plotes

```

```
Estados<-c('AL', 'BA', 'CE', 'MA', 'PB', 'PE', 'PI', 'RN', 'SE', "Nordeste", "Brasil")

```

```
Estados.nomes<-matrix(c("Alagoas", "Bahia", "Ceará", "Maranhão", "Paraíba", "Pernambuco"

```

```
colnames(Estados.nomes)<-Estados

```

```
#### Serie Temporais

```

```
# Plote da Serie, exemplo Brasil

```

```
plot.gg(Dados_Var, "Brasil", X=Dados_Var$date)

```

```
# Selecionando a ordem do modelo, criterio "bic", exemplo para o Brasil

```

```
auto.arima(Dados_Var$Brasil, ic="bic")

```

```
# Estimando os Parâmetros do modelo

```

```
fit.Brasil<-arima(Dados_Var$Brasil, order = c(1,0,0))

```

```
# Extraindo os Resíduos do modelo

```

```
residuo.Brasil<-residuals(fit.Brasil)

```

```
# Plot dos residuos

```

```
plot.gg.resid(residuo.Brasil, Dados_Var$date)

```

```
# Aplicando teste de Normalidade nos resíduos

```

```
ks.test(residuo.Brasil, "pnorm", mean(residuo.Brasil), sd(residuo.Brasil))

```

```
### Aplicando teste de Ljung Box nos resíduos

```

```
# k=6 como a ordem do Brasil é AR(1) fitdf=1

```

```
Box.test(residuo.Brasil, lag = 6 ,type=c("Ljung-Box"), fitdf = 1)

```

```
### Aplicando função tsdiag nos resíduos que apresenta ACF,

```

```
# plot dos residuos e plot Ljung Box para diferentes valores de k

```

```
tsdiag(residuo.Brasil)

```

```
### Aplicando as Cartas de Controle

```

```
## Carta Xbar.one obtendo resultados

```

```
Chat.Brasil.Xbar.one<-qcc(residuo.Brasil, type = "xbar.one", plot = F)

```

```
# Fazendo o plote

```

```
plot.xbar.one(Chat.Brasil.Xbar.one, X=Dados_Var$date)

```

```
## Carta EWMA obtendo os resultados para L-2.86 e lambda=0.2
Chat.Brasil.EWMA<-ewma(residuo.Brasil , lambda = 0.2, nsigmas=2.86, plot = F)
# Fazendo o plote
plot.ewma(Chat.Brasil.EWMA,"Brasil",X=Dados_Var$date)
```