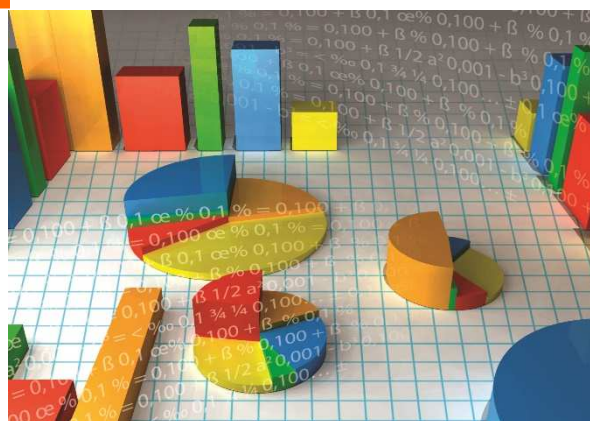


EDIÇÃO ESPECIAL - O uso da estatística para a compreensão das dinâmicas sociais e econômicas da Covid-19

**REVISTA BRASILEIRA
DE ESTATÍSTICA**

ISSN 2675-3243



volume 78

número 245

julho/dezembro 2020

Instituto Brasileiro de Geografia e Estatística – IBGE

REVISTA BRASILEIRA DE ESTATÍSTICA

Volume 78 - número 245 – jul/dez 2020

ISSN 2675-3243

R. Bras. Estat., Rio de Janeiro, v.78, n.245, p. 1-201, jul/dez 2020

@IBGE. 2020

Revista Brasileira de Estatística, ISSN 2675-3243

Órgão oficial do IBGE e da Associação Brasileira de Estatística – ABE
Publicação semestral que se destina a promover e ampliar o uso de métodos estatísticos através de divulgação de artigos inéditos tratando de aplicações da Estatística nas mais diversas áreas do conhecimento. Temas abordando aspectos do desenvolvimento metodológico serão aceitos, desde que relevantes para a produção e uso de estatísticas públicas.

Os originais para publicação deverão ser submetidos para o site <http://rbes.net.br>.

Os artigos submetidos à RBE não devem ter sido publicados em outros periódicos.

A Revista não se responsabiliza pelos conceitos emitidos em matéria assinada.

Editor Responsável

José André de Moura Brito - ENCE/IBGE

Editores Executivos

Marcel de Toledo Vieira - UFJF

Paulo Justiniano Ribeiro Junior – UFPR

Editores Associados

Alinne de Carvalho Veiga, ENCE/IBGE

Cristiano Ferraz, UFPE

Denise Britz do Nascimento Silva, ENCE/IBGE

Fernando Antônio Basile Colugnati, UFJF

Francisco Louzada-Neto, USP

Gustavo da Silva Ferreira, ENCE/IBGE

Ismenia Blavatsky, UFRN

Juvencio Santos Nobre, UFC

Maysa Sacramento de Magalhães, ENCE/IBGE

Paulo de Martino Jannuzzi, ENCE/IBGE

Pedro Luis do Nascimento Silva, ENCE/IBGE

Taiane Schaedler Prass, UFRGS

Vera Lucia Damasceno Tomazella, UFSCAR

Viviana Giampaoli, USP

Editores Convidados

César Augusto Marques

Daniel Takata Gomes

Editoração

José André de Moura Brito

Capa

Ilustração da Capa

Informações Adicionais

Revista Brasileira de Estatística/IBGE - v.1, n.1 (jan/mar. 1940) – Rio de Janeiro: IBGE, 1940.v. trimestral (1940-1986), semestral (1987-).

Continuação de: Revista de economia e estatística. Índices acumulados de autor e assunto publicados no v.43 (1940-1979) e v.50 (1980-1989). Co-edição com a Associação Brasileira de Estatística a partir do v.58.

ISSN publicação online 2675-3243, a partir de 2019.

I. Estatística – Periódicos. II. IBGE. III. Associação Brasileira de Estatística.

Gerência de Biblioteca e Acervos Especiais CDU 31(05)
RJ-IBGE/88-05 (ver. 2009) PERIÓDICO

Sumário

Nota do Editor 5

Artigos

Validade, Subnotificação e Acesso - Uma Análise Computacional de Efeitos sobre a Estimção da Proporção de Infectados.....7

João Gabriel Malaguti

Julia Machado Vieira

CCOVID-19: Monitoramento por Gráficos de Shewhart23

Cristiano Ferraz, Ademir J. Petenate

Andre Leite Wanderley

Raydonal Ospina

João Torres

Anna Peruzzi Moreira

Modelagem e Predição das quantidades de casos confirmados da COVID-19 no estado do Mato Grosso do Sul42

Erlandson Ferreira Saraiva

Leandro Sauer

Modelando Óbitos por Covid-19 em Cenários de Heterogeneidade Espacial: Uma Abordagem Utilizando Modelos Hierárquicos Bayesianos69

Ricardo Junqueira de Souza

Patrícia Viana da Silva

Jony Arrais Pinto Junior

Estimação do Número de Reprodutibilidade da Pandemia por COVID-19 em São Paulo91

Chang Chiann

Michel H. Montoril

Pedro A. Morettin

Evolução do Número de Testes para Covid-19 e Recomendações para o Número Mínimo de Testes em Diferentes Cenários. Estudo de Caso: Sergipe – Brasil110

Thiago de Jesus dos Santos

Jesy Karolayne Sales dos Santos

Carlos Raphael Araújo Daniel

Rita de Cássia de Lima Idalino

Distribuição Geográfica e Difusão Espacial dos Casos de COVID-19 no Estado de Mato Grosso142

Luiz Geraldo Mendes

Gizelle Prado da Silva Fonseca

José Manuel Cañizal Salceda

Adriana Queiroz Nascimento Pinhorati

CovidMetrika: Aplicativos em Shiny para Monitoramento da COVID-19166

Gustavo Machado Utpott

Juliana Sena de Souza

Gabriel Holmer Saul

Márcia Helena Barbian

Rodrigo Citton Padilha dos Reis

Contextualização do Cenário de Covid-19 na Região do Baixo
Tocantins-Pará188

Carla Braga Pereira

Luisa Dias Barros

Estevão José da Silva Barbosa

Nota do Editor

É com grande satisfação que anunciamos a publicação do volume 78 (número 245) da Revista Brasileira de Estatística. Este número, bem como o número 244, fazem parte da edição temática cujo título é “O uso da estatística para a compreensão das dinâmicas sociais e econômicas da Covid-19”. A edição traz artigos que propõem reflexões das dimensões sociais e econômicas do novo coronavírus a partir de estatísticas públicas e modelos estatísticos.

De modo geral, o estabelecimento e continuidade da pandemia de Covid-19 alterou profundamente as dinâmicas sociais, econômicas e científicas durante o ano de 2020. Em diferentes graus, países em todas as partes do globo implementaram medidas de mitigação da pandemia, alcançando graus diversos de sucesso na propagação da doença e na diminuição dos óbitos a ela associados.

Para tal, o campo científico foi desafiado a incorporar reflexões e pesquisas sobre a Covid-19 de diversas maneiras, muitas vezes de modo bastante imediato. Isso incluiu pesquisas não só para o desenvolvimento de vacinas e tratamentos médicos, mas também para a compreensão dos efeitos socioeconômicos da pandemia ao longo dos diversos tempos-espacos que ela afetou.

Mesmo considerando que as realidades da pandemia sejam relativamente novas e inéditas para as atuais gerações, é interessante reiterar que, no campo científico, a pesquisa sobre as diversas questões relacionadas ao Covid-19 ocorre a partir de um terreno já rico e diverso, tanto teórica como metodologicamente, considerando as diversas vantagens comparativas de cada campo do conhecimento.

É nesse cenário que as ciências estatísticas e os institutos de estatística atuaram na pandemia, acelerando tendências já construídas e desenvolvendo conhecimento que permita a compreensão da pandemia. Nesse número temático da Revista Brasileira de Estatística temos uma parcela de tais contribuições, com foco específico em modelagem estatística.

No primeiro artigo, João Malaguti e Julia Vieira exploram subnotificação e estimação de proporções de infectados através de efeitos investigados de maneira computacional a partir de simulações de Monte Carlo.

Já Cristiano Ferraz, Ademir Petenate, André Wanderley, Raydonal Ospina, João Torres e Anna Moreira utilizam gráficos de Shewhart, ferramentas tradicionais de Controle Estatístico de Processos, para monitoramento de dados de óbitos da Covid-19 no Brasil.

No terceiro artigo, Erlandson Saraiva e Leandro Sauer fazem uma modelagem das quantidades de casos confirmados da Covid-19 no estado do Mato Grosso do Sul, com previsões para a máxima quantidade de notificações e a data em que as taxas de notificações começarão a decrescer.

O quarto artigo, de autoria de Ricardo de Souza, Patrícia da Silva e Jony Arrais, propõe a utilização do modelo CAR dissimilaridade para levar em consideração a

possível existência de variações locais na autocorrelação espacial no estado do Rio de Janeiro no contexto da pandemia da Covid-19.

Chang Chiann, Michel Montoril e Pedro Morettin, por sua vez, estudam a taxa de reprodutibilidade, que mede a velocidade de contágio do vírus, no município e no estado de São Paulo, através de modelo autorregressivo com coeficiente variável no tempo e de uma decomposição das séries em componentes de tendência e ruído.

No sexto artigo, Thiago dos Santos, Jesy dos Santos, Carlos Daniel e Rita Idalino propõem um número mínimo de testes a serem realizados de modo a conseguir monitorar adequadamente a situação da pandemia da Covid-19 em Sergipe, através de informações como número de óbitos, taxa de letalidade e proporção de indivíduos sintomáticos.

Luiz Geraldo Mendes, Gizelle Fonseca, José Manuel Salceda e Adriana Pinhorati analisam, no sétimo artigo, a difusão espacial da pandemia de Covid-19 através dos mapas de vetores de mobilidade espacial do centro geográfico ponderado, nas Regiões Geográficas Intermediárias do Estado de Mato Grosso e módulos de fluxos rodoviários.

O oitavo artigo, de autoria de Gustavo Utpott, Juliana de Souza, Gabriel Saul, Márcia Barbian e Rodrigo dos Reis, descreve o CovidMetrika, grupo de estudos com o objetivo de desenvolver e implementar um sistema de vigilância espacial de casos de Covid-19, assim como monitorar a situação dos leitos hospitalares no estado do Rio Grande do Sul.

Por fim, no último artigo deste volume, Carla Pereira, Luisa Barros e Estevão Barbosa trazem uma análise estatística e cartográfica dos dados do novo coronavírus norteado em função de casos confirmados, óbitos, letalidade, números de leitos e médicos para esboçar o panorama espacial da pandemia na região do Baixo Tocantins, no Pará.

É uma satisfação trazer a público esta edição especial de uma revista de longa história como a RBE, com o irrestrito apoio do IBGE e Associação Brasileira de Estatística, ABE.

Junto-me ao colega Editor Associado deste número temático, César Augusto Marques da Silva, bem como aos colegas Editores Marcel de Toledo Vieira, José André de Moura Brito e Paulo Justiniano Ribeiro Junior, em agradecimento a autores e revisores, que, anonimamente, contribuíram para mais este número da Revista Brasileira de Estatística.

Desejo a todos uma excelente leitura.

Daniel Takata Gomes.

Editor Convidado.

VALIDADE, SUBNOTIFICAÇÃO E ACESSO - UMA ANÁLISE COMPUTACIONAL DE EFEITOS SOBRE A ESTIMAÇÃO DA PROPORÇÃO DE INFECTADOS

João Gabriel Malaguti

joagmalaguti@gmail.com

Escola Nacional de Ciências Estatísticas

Julia Machado Vieira

juliamachado.vieira12@gmail.com

Faculdade de Ciências Médicas e da Saúde de Juiz de Fora - SUPREMA

Resumo: Os dados relacionados à estimação da proporção de infectados por COVID-19 estão sujeitos a efeitos de diversos elementos como a subnotificação, mas como muitas variáveis são desconhecidas e as pesquisas com esse foco não podem ir a campo da mesma maneira que antes da pandemia são poucos os estudos sobre os limites dos dados. No entanto, alguns efeitos podem ser investigados de maneira computacional a partir de simulações Monte Carlo. Para este artigo, criamos estudos de simulação para observar os efeitos da validade do teste (sensitividade e especificidade), de diferentes níveis de subnotificação e de desigualdade no acesso aos testes. Resultados apontam aumentos de cerca de 90%, 160% e 215% no erro padrão quando se compara uma subnotificação 1:1 com 1:5, 1:10 e 1:15, respectivamente. Temos também que a desigualdade no acesso aos testes envia os resultados quando diferentes setores da população possuem diferentes proporções de infecção.

Palavras-chave: Subnotificação; Simulação Monte Carlo; Sensitividade; Especificidade

Abstract: The data related to the estimation of the proportion of those infected by COVID-19 are subject to effects from diverse elements such as underreporting, but since many variables are unknown and surveys with that focus aren't able to go on the field in the same manner as before the pandemic, there is little research on the limits of that data. However, some effects can be investigated computationally using Monte Carlo simulations. For this paper, we created simulation studies in order to observe the effects of test validity (sensitivity and specificity), of different levels of underreporting and of inequality in access to the tests. Results show increases of 90%, 160% and 215% in the standard error when we compare an underreporting of 1:1 to 1:5, 1:10 and 1:15, respectively. We also find that inequality of access to the tests biases the results when different sections of the population have differing infection proportions.

Keywords: Underreporting; Monte Carlo simulation, Sensitivity; Specificity

1. INTRODUÇÃO

A pandemia do coronavírus (COVID-19) afetou e afeta o mundo inteiro nas mais diferentes escalas. Várias pesquisas se encontram em hiato, e pesquisadores focam em conseguir informações relacionadas à essa crise. (FRASER *et al.*, 2020). No entanto, os dados relacionados à proporção de infectados são sujeitos a efeitos de diversos elementos, como a subnotificação, a validade dos testes e o acesso da população aos testes.

A subnotificação é a falha em notificar ou contabilizar todos os casos da doença (Porta, 2014). É uma das fraquezas conhecidas quando se refere a proporção de infectados, com pesquisadores estimando sua dimensão (embora não seu efeito) desde o início da pandemia (Ribeiro & Bernardes, 2020).

A validade de um teste diz respeito às métricas de sensibilidade e especificidade (Neta *et al.*, 2014), usadas para determinar se o resultado do teste corresponde à realidade. A sensibilidade é a habilidade de identificar corretamente os indivíduos que tem de fato a doença e a especificidade é a habilidade de identificar corretamente aqueles indivíduos que não tem a doença. No Brasil, Castro *et al.* (2020) identificaram que a especificidade dos testes registrados se encontra na faixa de 94 a 100 e a sensibilidade varia de 55,6 a 100.

Desigualdade de acesso é quando parcelas da população têm acesso diferenciado devido à classe socioeconômica a qual pertencem (Roemer & Trannoy, 2015). Sabemos que populações de menor renda são mais vulneráveis (Fiocruz, 2020a) tanto por sua maior exposição quanto por seu menor acesso à serviços de saúde de qualidade. Decorre, portanto, que a desigualdade afeta a estimação da proporção de infectados em duas frentes: na diferença da proporção em diferentes classes sociais e no acesso de cada classe à testagem.

Utilizando o método de simulações Monte Carlo, podemos investigar tais efeitos em um campo controlado, melhorando nosso entendimento de como esses aspectos reais impactam na qualidade dos dados.

Isolando os fatores que queremos estudar da fragilidade associada às atuais estimativas e dados relativos à COVID-19 podemos visualizar padrões de relação entre as variáveis, informando pesquisas futuras sobre essa pandemia específica e até outras.

Para este artigo, criamos estudos de simulação para observar o efeito da validade do teste (sensibilidade e especificidade), de diferentes níveis de subnotificação e de desigualdade no acesso aos testes.

Como principais resultados, temos que a subnotificação gera aumentos no erro padrão da estimação e temos que a desigualdade no acesso aos testes envia os resultados quando diferentes setores da população possuem diferentes proporções de infecção.

Este artigo se organiza da seguinte maneira: primeiramente, definimos alguns termos médicos e sociais (Seção 1), seguido pela manipulação necessária da fórmula de Bayes que permite a simulação (Seção 2). Na Seção 3 descrevemos as simulações, e analisamos seus resultados na Seção 4. A Seção 5 contém as considerações finais.

2.DESENVOLVIMENTO

2.1 VALIDADE DO TESTE

Podemos definir validade como a eficiência relativa de diferentes técnicas para identificar indivíduos em um grupo que apresentam evidências sugestivas da doença de interesse (Yerushalmy, 1947).

A validade de um teste pode ser determinada por duas métricas básicas (sensibilidade e especificidade) (Neta *et al.*, 2014), que são usadas para determinar se o resultado do teste corresponde à realidade, sendo comparado a um teste mais acurado.

A sensibilidade é a habilidade de identificar corretamente os indivíduos que tem de fato a doença ou que foram expostos ao objeto de estudo. Já a especificidade de um teste é a habilidade de identificar corretamente aqueles indivíduos que não tem a doença ou que não foram expostos ao objeto de estudo. Existe uma limitação inerente ao uso das métricas, pelo fato de que poucos testes são inerentemente dicotômicos.

Os valores da sensibilidade e especificidade podem variar de acordo com o ponto de corte utilizado no estudo para separar os indivíduos considerados doentes dos não doentes, por exemplo, carga viral acima de um certo valor. O quanto mais próximo um indivíduo estiver desse ponto, maiores são as chances de que ele seja classificado erroneamente. A escolha do ponto de corte, então, é extremamente importante pois determina os valores da sensibilidade e especificidade.

Outra dificuldade inerente ao teste é a capacidade de diferenciar quais resultados são verdadeiros ou falsos: se um teste tem resultado positivo e a pessoa não é infectada, temos um falso positivo (erro do tipo I). Por outro lado, se o teste é negativo e a pessoa de fato se encontra infectada, temos um falso

negativo (erro do tipo II), dessa forma podemos ter uma classificação errada, o que prejudica a validade do teste em questão (Newman, 2001).

Existem três tipos de teste para SARS-CoV-2: RT-PCR (*Real-time reverse transcription PCR*), teste por detecção de antígenos e teste por detecção de anticorpos (Rafiq, Batool, & Bazaz, 2020). O teste RT-PCR em tempo real (qRT-PCR) é o considerado o "padrão ouro", com um número de falsos positivos e falsos negativos muito menor que os outros (Fiocruz, 2020b). Este tipo de teste necessita de mais tempo para apresentar o resultado e amostras de má qualidade podem tornar necessária a coleta de novo material.

Os outros testes, apesar de mais rápidos, têm validade muito menor, por exemplo a sensibilidade do teste por detecção de antígenos varia entre 34% e 80%. O teste por detecção de anticorpos também é sujeito à reação cruzada com outros patógenos, o que significa que na presença de anticorpos de outros vírus similares (como a influenza) o teste retorna um falso positivo para COVID-19 (Rafiq, Batool, & Bazaz, 2020).

Na meta-análise sobre a validade de testes para COVID-19 feita por Bastos *et al.* (2020) levanta-se que a grande maioria das pesquisas (48 de 49) apresentava viés de seleção, escolhendo pacientes de alto risco, e 36 de 49 apresentavam viés de performance ou interpretação dos testes.

O caso brasileiro é analisado por Castro *et al.* (2020), que compara informações de todos os testes registrados pela Agência Nacional de Vigilância Sanitária (ANVISA) até março. A especificidade se encontra na faixa de 94 a 100 e a sensibilidade varia de 55,6 a 100. É importante notar que esses valores foram providenciados pelos fabricantes dos testes e nenhuma publicação revisada por pares foi encontrada. Além disso, alguns testes foram validados com 20 amostras ou menos e apenas metade deles incluiu mais de 150 amostras.

2.2 SUBNOTIFICAÇÃO

De maneira simples, subnotificação é a falha ao identificar ou contabilizar todos os casos *de facto* da doença (Porta, 2014). A teoria epidemiológica, no entanto, nem sempre corresponde à realidade; em muitas doenças os indivíduos contaminados podem não ser classificados como tal. Temos dois principais fatores responsáveis: epidemiológicos e os socioeconômicos.

Dentre os fatores epidemiológicos, podemos citar a severidade da doença, que está relacionada com o número de notificações; estes sendo influenciados pela morbidade e mortalidade associada a tal. O tamanho do surto da doença também é digno de nota. Os fatores socioeconômicos serão expandidos na subseção seguinte (2.3).

A subnotificação pode acontecer em qualquer nível do sistema de saúde, seja ele paciente, clínica, laboratorial ou no sistema como um todo, sendo os motivos os mais diversos possíveis (Noufaily, 2020).

Um exemplo de falha é a preferência por reportar casos severos durante o período de notificação ou durante estudos de coorte, negligenciando casos leves ou assintomáticos, enviesando certas medidas como a taxa de letalidade (Atkins *et al.*, 2015).

O número de notificações é de grande importância para determinar as necessidades do sistema de saúde, que está entrando em colapso em muitos países atingidos pela pandemia de COVID-19. Decorre então que a subnotificação é um grande problema pois as autoridades não sabem ao certo o real tamanho do problema que enfrentam, agindo no escuro (Ribeiro & Bernardes, 2020).

No início da pandemia brasileira (março 2020), pesquisadores estimaram o valor de 1:7,7 para a taxa de subnotificação, significando que o número real de casos deveria ser pelo menos 7 vezes maior do que o número de casos notificados.

As principais dificuldades são a ausência de infraestrutura laboratorial e de pessoal qualificado; a dificuldade na obtenção de testes diagnósticos devido a elevada demanda mundial e a baixa quantidade de fornecedores e a logística de distribuição nacional em um país com a dimensão continental como o Brasil apresenta (Ribeiro & Bernardes, 2020). Entretanto, essa não é uma realidade somente brasileira, já que diversos outros países enfrentam grandes dificuldades em determinar o real número de pessoas infectadas, sendo um empecilho a falta de testes diagnósticos para COVID-19.

2.2 ACESSO

O Brasil é um país muito desigual, sobretudo em sua distribuição de renda (Hoffmann *et al.*, 2019). Não é, então, uma suposição forte que diferentes eventos afetam classes sociais de maneira diferente. Uma das maneiras de caracterizar desigualdade é como desigualdade de oportunidades ou desigualdade de acesso (Roemer & Trannoy, 2015) na qual parcelas da população têm acesso diferenciado devido à classe socioeconômica a qual pertencem.

Sabemos que populações de menor renda são mais vulneráveis (Fiocruz, 2020a) tanto por sua maior exposição quanto por seu menor acesso à serviços de saúde de qualidade. Decorre, portanto, que a desigualdade afeta a

estimação da proporção de infectados em duas frentes: na diferença da proporção em diferentes classes sociais e no acesso de cada classe à testagem.

Ramos (2020) coloca que as características patogênicas da pandemia aliadas aos os fatores sociais de vulnerabilidade presentes no Brasil, como, por exemplo, as moradias urbanas precárias em locais de alta densidade populacional e de pobreza, sem saneamento e infraestrutura básica, podem se tornar condições perfeitas para uma “carnificina dos mais pobres”, usando as palavras da pesquisadora. O adensamento domiciliar excessivo (quando se tem mais de três moradores utilizando o mesmo cômodo como dormitório num domicílio), outro fator que auxilia a contaminação, compreende 14,5% dos pobres e 5,6% da população em geral.

Ao mesmo tempo, os setores da classe trabalhadora vinculados aos serviços essenciais estão sendo submetidos a jornadas mais intensas de trabalho e riscos de contágio aumentados, devido à ausência de equipamentos de segurança (Ramos, 2020).

A rede social a qual a pessoa pertence e a conectividade da população influenciam também a infecção e a notificação. Por exemplo, caso uma pessoa seja diagnosticada com a doença, assim que os sintomas começarem a aparecer em vizinhos as chances de notificação tornam-se maiores. (Gamado *et al.*, 2013).

Enquanto que o número de testes depende principalmente da máquina pública, o acesso aos mesmos se limita, dentre outros fatores, pela distância. De acordo com a nota técnica da Pesquisa Regiões de Influência das Cidades sobre informações de deslocamentos para serviços de saúde (IBGE, 2020), a média de deslocamento para serviços de saúde de baixa e média complexidade é de 72 quilômetros e a média de deslocamento para serviços de saúde de alta complexidade é de 155 km. Essa média mascara as diferenças regionais, visto que para Sudeste e Sul o valor é próximo de 100 km enquanto que para o Norte chega a ser 276 km e 256 km para a região Centro-Oeste. A população mais pobre não tem nem a disponibilidade, nem o capital necessário para fazer tais viagens ameio a pandemia.

2.4 CÁLCULO DE PROPORÇÃO

Suponha uma população de tamanho N com dois eventos de interesse: a pessoa possui a doença de interesse (D) e a pessoa possui sintomas (S). $(D \cap S)$ é o indivíduo infectado com sintomas, $(D \cap S^c)$ é o indivíduo infectado assintomático, $(D^c \cap S^c)$ é o indivíduo saudável, e $(D^c \cap S)$ é o indivíduo sem a infecção de interesse, mas com sintomas similares. Por exemplo, se o interesse

é a detecção de COVID-19 e a pessoa está infectada com outro tipo de coronavírus ou doenças como tuberculose.

Um teste médico é capaz de detectar a doença, embora sujeito a sua validade, como discutido na seção 2.1 deste artigo. Assumindo que os testes são realizados apenas em pessoas sintomáticas, ele seria capaz de fornecer a estimativa para o parâmetro $P(D|S)$.

Para que se possa estimar, a partir dos resultados do teste, a proporção dos infectados na população total, utilizamos o teorema de Bayes (Ross, 2010):

$$P(D|S) = \frac{P(S|D) \cdot P(D)}{P(S)} \quad (1)$$

O alvo de nossa estimação é $P(D)$, e podemos estimar $P(D|S)$ e $P(S)$. Como a estimação de $P(S|D)$ está fora do escopo do presente trabalho, assumiremos que os pesquisadores conseguem determinar o valor exato deste, retirando assim os efeitos e incertezas relacionados com estimação deste valor na estimação de $P(D)$.

Reescrevendo a equação acima, temos:

$$P(D) = \frac{P(S) \cdot P(D|S)}{P(S|D)} \quad (2)$$

Para a estimação, temos:

$$p(D) = \frac{\left(\frac{\text{Número de Pessoas com Sintomas}}{N}\right) \cdot \left(\frac{\text{Número de Resultados Positivos}}{\text{Número de Testes Realizados}}\right)}{P(S|D)} \quad (3)$$

Onde $p(D)$ é a estimativa para $P(D)$ e N é o tamanho da população.

3.SIMULAÇÃO MONTE CARLO

3.1 BASE TEÓRICA

Um estudo de simulação, como uma simulação Monte Carlo (MC) é um método computacional para se estimar respostas para problemas probabilísticos a partir de amostras independentes (Brandimarte, 2017; Lewis & Orav, 1989). Sua precisão é relacionada ao número de amostras (iterações ou rodadas do estudo), seguindo a lei dos grandes números que descreve que com um grande número de ensaios, o valor esperado dos resultados observados converge para a média verdadeira.

Os erros padrão deste artigo são calculados a partir da simulação MC, ou seja, calcula-se o erro padrão utilizando os A valores estimados para a proporção. Quanto maior for o valor de A, melhor é a estimativa do erro padrão da medida.

3.2 FRAGILIDADE DOS DADOS

Os mecanismos relacionados à COVID-19 ainda não são bem entendidos (Bartoszko *et al.*, 2020; Rafiq, Batool, & Bazaz, 2020). O período de incubação, por exemplo, é aproximadamente 5-6 dias, mas pode variar de 0 a 24 dias (Aquino *et al.*, 2020). A própria transmissão continua a ser investigada (Bielecki *et al.*, 2020).

Ainda existem poucos estudos sobre COVID-19 (Keeley *et al.*, 2020) e muitos desses são feitos com amostras pequenas, levando à uma precisão menor. Duas grandes pesquisas sobre transmissão assintomática, Huang *et al.* (2020) e Kimball *et al.* (2020) acompanharam 22 e 82 indivíduos, respectivamente. Na revisão sistemática de Keeley *et al.* (2020), apenas quatro estudos tinham informações sobre os sintomas dos pacientes que foram à óbito; suas amostras foram de 6801, 109, 101 e 20. Na mesma revisão, 10 estudos sobre os sintomas dos sobreviventes foram estudados, 7 deles com amostras menores que 200 e destes, 4 com tamanho de amostra menor que 100.

Os dados de secretarias e ministérios de saúde, além de sujeitos à subnotificação como descrito anteriormente, têm de lidar com diferentes definições de “caso” de país a país e também a classificação de COVID-19 como outro diagnóstico, como Síndrome Respiratória Aguda Grave (SRAG) (Fiocruz, 2020b). O COVID-19 não é o único agente etiológico causador de SRAG, apesar de ser o que mais contribui para os aumentos dos casos brasileiros em 2020.

O aumento expressivo de óbitos ocorridos em domicílios, sem assistência médica e por causas mal definidas influenciam também na fragilidade dos dados sobre a infecção (Fiocruz, 2020c).

Outro grande fator que aumenta a imprecisão dos dados é a diferença entre assintomáticos e pré-sintomáticos (que testam positivos e depois desenvolvem sintomas). Esta diferença muitas vezes não é reportada (Korownyk *et al.*, 2020) e isso enviesam a estimativa de quantos portadores assintomáticos de COVID-19 existem.

Na pesquisa de Kimball *et al.* mencionada anteriormente, dos 13 indivíduos positivos e assintomáticos no momento do teste, 10 desenvolveram sintomas durante a semana seguinte. Dois modelos matemáticos tentaram estimar o número de infecções causadas por pessoas assintomáticas e pré-sintomáticas, mas variaram bastante entre si, um estimando cerca de metade das infecções e o outro quatro quintos (Furukawa, Brooks & Sobel, 2020).

3.3 ESTUDOS DE SIMULAÇÃO

Utilizando a linguagem de programação R (R Core Team, 2020), geramos uma população de tamanho $N=100.000$, com as proporções sendo apresentadas na Tabela 1 a seguir. Os valores escolhidos foram arbitrários, para separar este estudo da fragilidade dos dados mencionada na seção anterior. As outras proporções podem ser calculadas a partir das apresentadas.

Tabela 1 – Parâmetros da Simulação

Parâmetro	Valor
D	0,10
S	0,12
S D	0,75
S D ^c	0,05

Definimos o número de testes de modo a que estes funcionem como quatro cenários de subnotificação: 1:1, 1:5, 1:10 e 1:15; ou seja, para cada 1 caso testado existem 1, 5, 10 ou 15 pessoas com a doença que não foram testadas. Por exemplo, se existem apenas testes para metade da proporção de infectados teríamos subnotificação 1:1.

Para o primeiro estudo, definimos três valores para sensibilidade e especificidade (85, 90 e 95). Com esses três valores, temos 9 combinações a serem estudadas.

Para o segundo estudo, que foca na desigualdade de acesso, fixamos a validade do teste em 90 para ambas as medidas por configurar o cenário mediano.

Para que possamos analisar essa desigualdade, precisamos que a população contenha também informações de classe socioeconômica. Essa população foi gerada com as divisões de classe de 20% de população na classe alta, 30% de população na classe baixa e 50% de população entre os dois extremos (classe média).

Diferentes cenários foram imaginados em relação à distribuição da infecção entre as classes, sempre mantendo a proporção da doença na população como um todo. No primeiro cenário, que serve de *benchmark*, todas as classes tem a mesma proporção. No segundo e terceiro cenários, uma das classes que não a alta tem o dobro de proporção das outras. No quarto cenário, as duas classes que não a alta têm o dobro da proporção de infectados da alta. No último cenário a classe média têm proporção 3 vezes maior do que a classe

alta e a classe baixa, cinco vezes maior. Estes cenários receberam os nomes de "1:1:1", "1:1:2", "1:2:1", "1:2:2" e "1:3:5".

O acesso aos testes será representado definindo as parcelas do total de testes reservadas a cada classe, sendo o benchmark a divisão igualitária de 33% para cada ("33-33-33"). Outra divisão é "50-30-20", que garante à classe mais alta 50% de todos os testes, 30% dos testes para a classe média e 20% dos testes para a classe baixa. As outras duas divisões estudadas continuam a concentrar a maior quantidade de testes na classe mais alta, com as porcentagens sendo "60-30-10" e "70-20-10".

Todos os estudos foram realizados com $A=100.000$ (cem mil) iterações Monte Carlo para aumentar a precisão das estimativas.

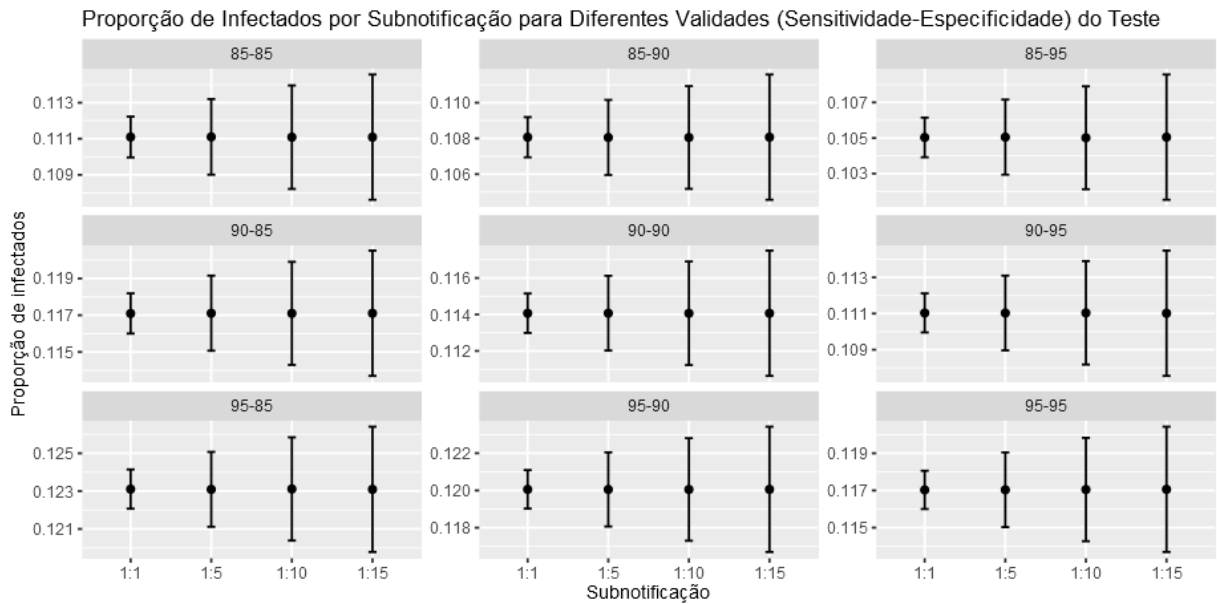
4.RESULTADOS

Como a subnotificação é central a este trabalho, convém começar a análise dos resultados por ela. Como podemos notar na Figura 1 a seguir, existe um padrão entre a subnotificação e a amplitude do intervalo de confiança da proporção, mais especificamente, quanto maior a taxa de subnotificação maior o erro padrão da proporção.

Este padrão é o mesmo para todas as combinações de sensibilidade e especificidade estudadas. Em média, o erro padrão da subnotificação "1:5" (0,00205) é 1,90 vezes maior do que a notificação *benchmark* "1:1" (0,00108), com os erros padrão das subnotificações "1:10" (0,00282) e "1:15" (0,00343) sendo respectivamente 2,62 e 3,18 vezes maior. É importante frisar que os eixos verticais são diferentes para cada validade.

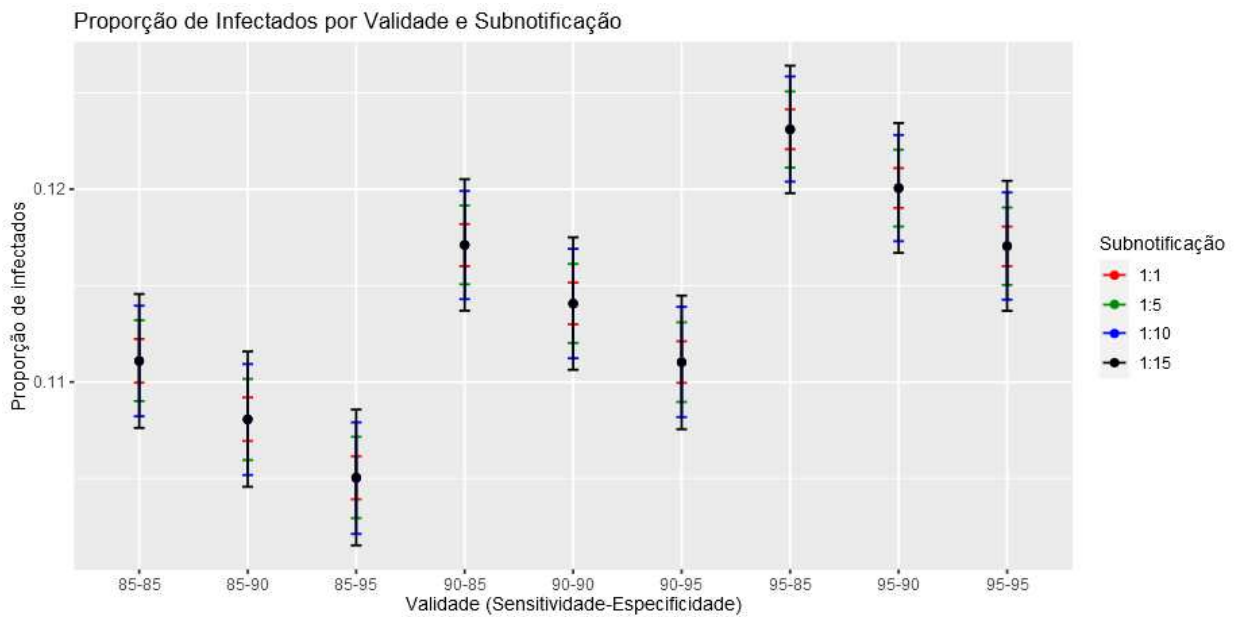
Utilizando os mesmos dados, mas alterando a disposição deles, podemos ver os efeitos da sensibilidade e da especificidade (Figura 2).

Figura 1 – Proporção de Infectados por Subnotificação para Diferentes Validades (Sensitividade-Especificidade) do Teste



Fonte – Autoria própria

Figura 2 – Proporção de Infectados por Validade e Subnotificação



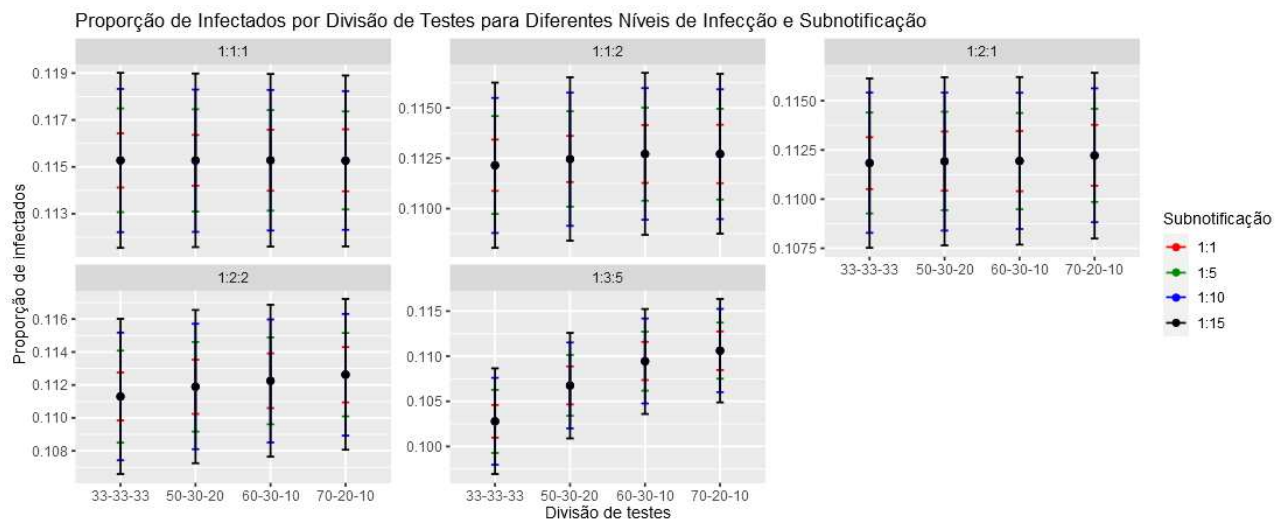
Fonte – Autoria própria

Fixada uma mesma sensibilidade, podemos notar o efeito da especificidade, que quanto maior leva a uma estimativa menor da proporção de infectados. Isso é pelo fato da quantidade menor de falsos positivos nos resultados dos testes.

Em compensação, quanto maior a sensibilidade, maior é a estimativa para a estimação devido a quantidade menor de falsos negativos dentre os resultados.

Para o estudo sobre acesso, o caso médio (sensibilidade e especificidade de 90%) foi selecionado para todas as combinações de proporção de infectados por classe e acesso aos testes. Os resultados estão ilustrados na Figura 3.

Figura 3 – Proporção de Infectados por Divisão de Testes para Diferentes Níveis de Infecção e Subnotificação

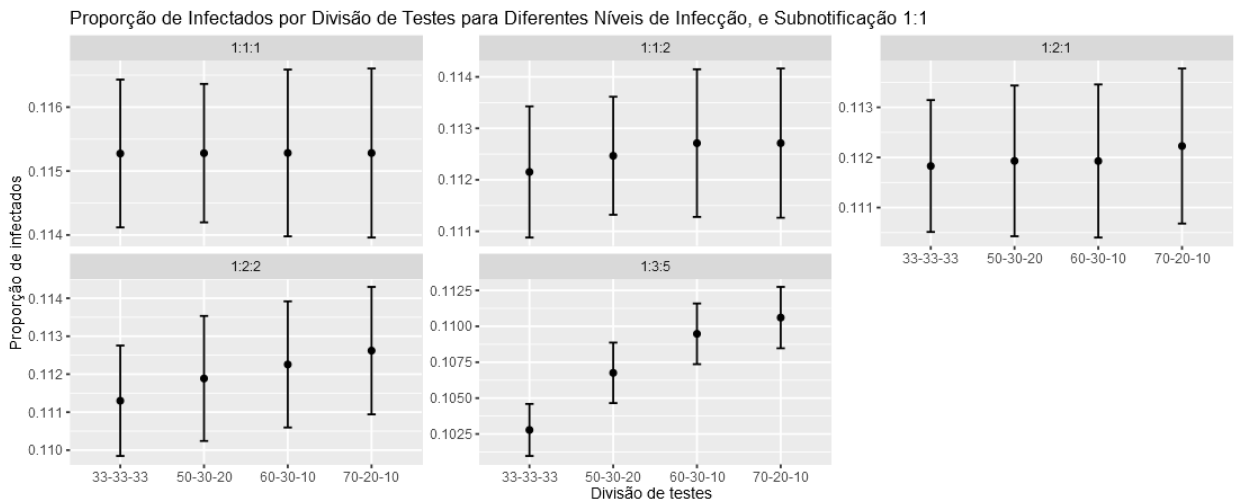


Fonte – Autoria própria

Em nenhum dos cinco gráficos que compõem a Figura 3 há muita diferença no tamanho das barras de erro, e nos três primeiros gráficos (1:1:1, 1:1:2 e 1:2:1), não há muita diferença no posicionamento da estimativa pontual para a proporção.

Para o cenário no qual a classe média e a classe baixa têm o dobro do nível de infecção da classe alta (1:2:2) existe uma tendência pequena (e não significativa) de aumento na estimativa pontual conforme a concentração no número de testes se desloca de uma divisão igualitária para uma concentrada na classe alta. Para o cenário mais provável, dado as vulnerabilidades as quais as populações mais pobres estão sujeitas (1:3:5) temos a mesma tendência que é significativa para subnotificação 1:1, mas não para as outras que têm a amplitude do intervalo muito maior. Isso é melhor visualizado na Figura 4, que mostra apenas a menor subnotificação.

Figura 4 – Proporção de Infectados por Divisão de Testes para Diferentes Níveis de Infecção, e Subnotificação 1:1



Fonte – Autoria própria

Temos então que a divisão dos testes aliada com a desigualdade da proporção de infectados por classe enviesará os resultados.

5. CONSIDERAÇÕES FINAIS

Apesar das simulações feitas neste trabalho serem bastante simples, elas são capazes de isolar os efeitos que certos aspectos têm sobre a estimação da proporção de infectados, o que seria difícil, se não impossível, com os dados reais, sujeitos a inúmeros diferentes efeitos.

E isso sem entrar na fragilidade dos dados e informações atuais, que é influenciada, por exemplo, pelo pequeno tamanho das amostras, pelas diferenças de definição de “caso” nos diversos países, pela mistura de casos com outros agentes epidemiológicos como influenza, entre outros fatores.

Os valores escolhidos para este artigo em si são arbitrários, mas nos permitem visualizar os padrões de relação entre as variáveis, nos informando para as pesquisas futuras sobre essa pandemia específica e outras sem a fragilidade associada às tentativas de estimação de tais parâmetros.

Quando estimativas mais confiáveis possam ser acessadas, é possível retornar à essas simulações e substituir os valores iniciais, assim como também é possível sua alteração visando incluir outros fatores de análise.

De um modo geral, temos que a subnotificação gera aumentos de cerca de 90%, 160% e 215% no erro padrão (quando se compara uma subnotificação 1:1 com 1:5, 1:10 e 1:15) e temos também que a desigualdade no acesso aos

testes envia os resultados quando diferentes setores da população possuem diferentes proporções de infecção.

6.REFERÊNCIAS BIBLIOGRÁFICAS

Atkins, K. E.; Wenzel, N. S.; Ndeffo-Mbah, M.; Altice, F. L.; Townsend, J. P. & Galvani, A. P. (2015). Under-reporting and case fatality estimates for emerging epidemics. *British Medical Journal*, 350:h1115.

Aquino, E. M., Silveira, I. H., Pescarini, J. M., Aquino, R., & Souza-Filho, J. A. D. (2020). Medidas de distanciamento social no controle da pandemia de COVID-19: potenciais impactos e desafios no Brasil. *Ciência & Saúde Coletiva*, 25, 2423-2446.

Bartoszko, J. J., Farooqi, M. A. M., Alhazzani, W., & Loeb, M. (2020). Medical masks vs N95 respirators for preventing COVID-19 in healthcare workers: A systematic review and meta-analysis of randomized trials. *Influenza and other respiratory viruses*.

Bastos, M. L., Tavaziva, G., Abidi, S. K., Campbell, J. R., Haraoui, L. P., Johnston, J. C., ... & Menzies, D. (2020). Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis. *British Medical Journal*, 370.

Bielecki, M., Züst, R., Siegrist, D., Meyerhofer, D., Cramer, G. A. G., Stanga, Z. G., ... & Deuel, J. W. (2020). Social distancing alters the clinical course of COVID-19 in young adults: A comparative cohort study. *Clinical Infectious Diseases*.

Brandimarte, P. (2014). *Handbook in Monte Carlo Simulation – Applications in Financial Engineering, Risk Management, and Economics* (1a ed.). Hoboken: John Wiley & Sons.

Castro, R., Luz, P. M., Wakimoto, M. D., Veloso, V. G., Grinsztejn, B., & Perazzo, H. (2020). COVID-19: a meta-analysis of diagnostic test accuracy of commercial assays registered in Brazil. *The Brazilian Journal of Infectious Diseases*. 24(2):180–187.

Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálffy, M., & Coates, J. A. (2020). Preprinting a pandemic: the role of preprints in the COVID-19 pandemic. *bioRxiv*.

Fundação Oswaldo Cruz – Fiocruz. (2020a) Estimativa de risco de espalhamento da COVID-19 no Brasil e avaliação da vulnerabilidade socioeconômica nas microrregiões brasileiras. Nota Técnica 3.

Fundação Oswaldo Cruz – Fiocruz. (2020b) Situando Síndrome Respiratória Aguda Grave (SRAG) e Covid-19. Nota Técnica 5.

Fundação Oswaldo Cruz – Fiocruz. (2020c) Óbitos desassistidos no Rio de Janeiro. Análise do excesso de mortalidade e impacto da Covid-19. Nota Técnica 11.

Furukawa, N. W., Brooks, J. T., & Sobel, J. (2020). Evidence supporting transmission of severe acute respiratory syndrome coronavirus 2 while presymptomatic or asymptomatic. *Emerging infectious diseases*, 26(7).

Gamado, K.M., Streftaris, G. & Zachary, S. (2013). Modelling under-reporting in epidemics. *Journal of Mathematical Biology*, 69, 737-765.

Hoffmann, R.; Botassio, D. C. & de Jesus, J. G. (2019). *Distribuição de Renda – Medidas de Desigualdade, Pobreza, Concentração, Segregação e Polarização*. (2a ed.). São Paulo: Editora da Universidade de São Paulo - Edusp.

Huang, L., Zhang, X., Zhang, X., Wei, Z., Zhang, L., Xu, J., ... & Xu, A. (2020). Rapid asymptomatic transmission of COVID-19 during the incubation period demonstrating strong infectivity in a cluster of youngsters aged 16-23 years outside Wuhan and characteristics of young patients with COVID-19: a prospective contact-tracing study. *Journal of Infection*. 80: e1–e13.

Instituto Brasileiro de Geografia e Estatística – IBGE. (2020) Pesquisa Regiões de Influência das Cidades – informações de deslocamentos para serviços de saúde – Notas Técnicas.

Keeley, P., Buchanan, D., Carolan, C., Pivodic, L., Tavabie, S., & Noble, S. (2020). Symptom burden and clinical profile of COVID-19 deaths: a rapid systematic review and evidence summary. *BMJ Supportive & Palliative Care*. 0:1–4.

Kimball, A., Hatfield, K. M., Arons, M., James, A., Taylor, J., Spicer, K., ... & Bell, J. M. (2020). Asymptomatic and presymptomatic SARS-CoV-2 infections in residents of a long-term care skilled nursing facility—King County, Washington, March 2020. *Morbidity and Mortality Weekly Report*, 69(13), 377.

Korownyk, C., Allan, G. M., Dugré, N., Lindblad, A. J., McCormack, J., & Kolber, M. R. (2020). Rapid review of COVID-19. *Canadian Family Physician*, 66(6), 429-429.

Lewis, P. A.W. & Orav, E. J. (1989) *Simulation Methodology for Statisticians, Operations Analysts and Engineers, Volume 1*. (1a ed.) Belmont: Wadsworth.

Neta, G.; Samet, J. M. & Rajaraman, P. (2014). Quality Control and Good Epidemiological Practice. In: Ahrens, W. & Pigeot, I. (dir.) *Handbook of Epidemiology* (2a ed., páginas 525-576). New York: Springer.

Newman, S. C. (2001). *Biostatistical Methods in Epidemiology* (1a ed.). New York: John Wiley & Sons.

Noufaily, A. (2020). Underreporting and Reporting Delays. In: Held, L.; Hens, N.; O'Neill, P. & Wallinga, J. (dir.) *Handbook of Infectious Disease Data Analysis* (1a ed., páginas 437-454). Boca Raton: CRC Press.

Porta, M. (dir.) (2014). *A Dictionary of Epidemiology* (6a ed.). Oxford: Oxford University Press.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Rafiq, D., Batool, A., & Bazaz, M. A. (2020). Three months of COVID-19: A systematic review and meta-analysis. *Reviews in Medical Virology*, e2113.

Ramos, V. R. G. (2020). Pandemia do coronavírus (COVID-19) e a classe trabalhadora brasileira em xeque mate. *Revista Estudos Libertários*, 2(4), 6-22.

Ribeiro, L. C. & Bernardes, A. T. (2020) Estimate of underreporting of COVID-19 in Brazil by Acute Respiratory Syndrome hospitalization reports. *Notas Técnicas Cedeplar – UFMG* 010.

Roemer, J. E. & Trannoy, A. (2015). Equality of Opportunity. In: Atkinson, A. B. & Bourguignon, F. (dir.) *Handbook of Income Distribution* (1a ed., páginas 217-300). Amsterdã: Elsevier.

Ross, S. (2010). *A First Course in Probability* (8a ed.). Upper Saddle River: Prentice Hall.

Yerushalmy, J. (1947). Statistical Problems in Assessing Methods of Medical Diagnosis, with Special Reference to X-Ray Techniques. *Public Health Reports*, 62(40), 1432-1449.

COVID-19: MONITORAMENTO POR GRÁFICOS DE SHEWHART

Cristiano Ferraz

cferraz@castlab.org

Universidade Federal de Pernambuco
Computational Agricultural Statistics Laboratory

Ademir J. Petenate

ademir@unicamp.br

Universidade Estadual de Campinas

Departamento de Estatística

André Leite Wanderley

leite@castlab.org

Universidade Federal de Pernambuco
Computational Agricultural Statistics Laboratory

Raydonal Ospina

raydonal@castlab.org

Universidade Federal de Pernambuco
Computational Agricultural Statistics Laboratory

João Torres

miqueiasjoaoeudes@gmail.com

Universidade Federal de Pernambuco
Computational Agricultural Statistics Laboratory

Anna Peruzzi Moreira

annperuzzi@hotmail.com

Universidade Federal do Rio de Janeiro, Macaé
Computational Agricultural Statistics Laboratory

Resumo: A pandemia de COVID-19 chegou oficialmente ao Brasil em 25 de fevereiro de 2020, no estado de São Paulo. Desde então, o seu histórico tem sido acompanhado em todas as Unidades da Federação, com divulgação de diversos indicadores diários. Num primeiro momento, modelos epidemiológicos para essas variáveis tiveram o importante papel de examinar cenários futuros frente a possibilidades de ações governamentais. Parâmetros foram ajustados de acordo com informações de outros países que já tinham passado pelo início da epidemia e vivenciavam situações críticas. Passada essa fase, dados locais começaram a ser utilizados, gerando um rico conjunto de análises. Dentre os desafios desde então enfrentados, destaca-se a necessidade de monitoramento da evolução dos dados da doença, levando-se em consideração as muitas limitações relacionadas ao registro de informações, sendo a subnotificação talvez a principal delas. Este artigo apresenta uma proposta de uso de gráficos de Shewhart, ferramentas tradicionais de Controle Estatístico de Processos, para monitoramento de dados de óbitos da COVID-19. Apesar de, como todo método, estar sujeito a qualidade dos dados que são disponibilizados, seu histórico de sucesso no uso para monitoramento de dados em diversas áreas respaldam seu potencial para aplicação no combate a pandemia.

Palavras-chave: SARS-CoV-2; Monitoramento; Novo Coronavírus; Controle Estatístico de Processos.

Abstract: The COVID-19 pandemic has officially reached Brazil in February 25th, 2020. Since the first case registered in Sao Paulo, all the Brazilian states have been registering disease numbers, making several related indicators available on a daily basis. In the beginning, epidemiological models had the important role of simulating different future scenarios to support governmental decision makers. At that time, model parameters were adjusted based on other countries' information. After local data has been sufficiently collected, several other analyses started to emerge to describe Brazilian reality. Now, the need for monitoring the evolution of COVID-19 raises as one of the challenges that need to be faced, given several data limitations, including non-documentation. This paper proposes the use of traditional Shewhart control charts to monitor COVID-19 deaths data in Brazil. Although also subjected to data limitations, as any other method, its history of success monitoring data in several other areas, support their potential for use as a tool to fight against the pandemics.

Keywords: SARS-CoV-2; Surveillance; New Coronavirus; Statistical Process Control.

1. INTRODUÇÃO: CORONAVÍRUS E COVID- 19

O novo Coronavírus SARS-CoV-2, denominado provisoriamente de 2019-nCoV, é responsável por causar a doença COVID-19 (Coronavírus disease 2019), considerada como pandemia em 2020 (Bedford, et al., 2020). Trata-se de um vírus encapsulado de RNA fita simples (Rothan & Byrareddy, 2020). É o sétimo Coronavírus humano (HCoVs) conhecido, recebendo esse nome pelo fato de seu envoltório ser semelhante a uma coroa. As cepas HCoV-OC43, HCoV-HKU1, HCoV-NL63 e HCoV-229E são responsáveis por um percentual de 10% a 35% dos resfriados comuns, dependendo da estação do ano (Longo et al., 2013). Já a SARS-CoV-2 é capaz de gerar quadros de maior gravidade, assim como a SARS-CoV e a MERS-CoV, responsáveis por epidemias passadas (Lu, et al., 2020; Holmes, 2020).

A COVID-19 possui um quadro clínico diversificado. As manifestações preponderantes assemelham-se as de uma síndrome gripal (febre, tosse seca, dispneia, mialgia). Entretanto, em sua forma mais grave, acometendo principalmente idosos e portadores de comorbidades, pode produzir um quadro de síndrome respiratória aguda grave, denominada de SARS – Severe Acute Respiratory Syndrome. A gravidade da doença vai além da sua manifestação em estado agudo, apresentando velocidade de contaminação entre indivíduos muito superiores aos casos de experiências passadas.

Enquanto o SARS-CoV apresenta uma taxa de letalidade estimada em 9,6% e o MERS-CoV em 34,4%, o SARS-CoV-2 apresenta taxa de letalidade relativamente baixa, de 3,8% (Park, 2020). Não obstante, em dois meses de seu início ser anunciado na cidade de Wuhan-China, já havia superado o SARS-CoV, epidemia ocorrida em 2003 com 8 meses de duração, em mais de 70.000 casos (Wilder-Smith, et al., 2020; Baud et al., 2020; WHO, 2020).

Relatos de uma pneumonia de causas até então desconhecidas, atingindo a cidade de Wuhan, chegaram a Organização Mundial de Saúde - OMS em 31 de dezembro de 2019. Em 11 de março de 2020, a OMS já declarava estado de pandemia por causa do novo Coronavírus SARS-CoV-2. No Brasil, o primeiro caso, confirmado oficialmente pelo Ministério da Saúde, ocorreu em São Paulo, em 25 de fevereiro de 2020, quando um homem de 61 anos de idade, foi atendido pelo Hospital Israelita Albert Einstein, com histórico de viagem a Itália. Em maio, o país já registrava mais de 135 mil casos confirmados (7 de maio, 135.106 casos).

São vários os fatores que contribuem para a velocidade de espalhamento da doença. Dentre aqueles intrínsecos ao vírus estão a ocorrência de transmissão direta e indireta (ou seja, por meio de indivíduos infectados e objetos contaminados, respectivamente) associada a alta estabilidade do mesmo no meio, podendo persistir em horas, e até dias, variando de acordo com o material (Santarpia, et al., 2020; Park 2020; Van Doremalen, et al., 2020; Chin et al., 2020). O vírus SARS-CoV-2 realiza uma interação importante com a enzima conversora da angiotensina 2 (ACE-2), presente sistemicamente em células humanas (Park, 2020; Yan et al., 2020). Essa ligação ocorre de forma ainda mais competente que seu antecessor SARS-CoV por apresentar um aprimoramento no local de clivagem de proteínas da fusão viral, aumentando seu tropismo tecidual e patogenicidade (Walls, et al., 2020). O SARS-Cov-2 tem um período longo de transmissibilidade, ocorrendo desde o início do desenvolvimento dos sintomas, e decrescendo gradativamente, havendo registros ainda de ocorrer em casos de pacientes assintomáticos ou com

sintomas leves. Já o SARS-CoV possibilita transmissão no período mais sintomático, sem registros de transmissão por assintomáticos, facilitando, assim, seu diagnóstico e isolamento adequado (Yuen, et al., 2020; Santarpia, et al., 2020; Park, 2020; Wilder-Smith, et al., 2020). Por fim, a baixa taxa de letalidade do SARS-Cov-2, aliada a alta preponderância de manifestações menos graves, aumenta o número de infectados capazes de transmitir o vírus de forma comunitária.

As características intrínsecas ao SARS-Cov-2, as informações de situações críticas enfrentadas por países da Europa que demoraram a tomar providências de isolamento social, bem como os resultados de estudos técnicos internacionais, utilizando modelos de predição para examinar cenários futuros de curto e médio prazos, frente a possibilidades de ações governamentais, alertaram para a gravidade da situação a ser enfrentada no Brasil. A adoção de medidas de distanciamento social, os esforços para ampliação de leitos de UTI, a mobilização de profissionais de saúde, os esforços para aquisição de material de EPI (Equipamento de proteção individual) e respiradores, foram provavelmente as ações de maior visibilidade no combate a doença por parte dos estados, nos primeiros momentos da crise.

Ainda que os governos estaduais tenham adotado medidas importantes, de forma tempestiva, a gravidade dos efeitos da epidemia tem sido verificada em diversos estados. São Paulo, Rio de Janeiro, Ceará, Pernambuco, Amazonas e Pará figuram entre os de maior incidência de número de casos, quando este artigo estava sendo finalizado. O momento atual aponta para a importância do monitoramento adequado da evolução da doença, apesar das muitas limitações relacionadas ao registro de informações, dentre as quais, destaca-se a subnotificação.

Este artigo apresenta uma proposta de uso de gráficos de Shewhart, tradicionais ferramentas do Controle Estatístico de Processos, para monitoramento de dados diários de óbitos por COVID-19. Apesar de, como todo método, estar sujeito à qualidade dos dados que são disponibilizados, seu histórico de sucesso no uso para monitoramento de dados em diversas áreas respaldam seu potencial para aplicação no monitoramento de dados da pandemia. Exemplos de gráficos com dados oficiais são apresentados, com uma breve reflexão sobre o momento a que eles se referem. Foi desenvolvida uma plataforma web que permite ao usuário a construção de gráficos de Shewhart adaptados para analisar dados de óbitos divulgados pelo Ministério da Saúde para o país, estados e municípios. A metodologia pode ser adaptada para uso com outros indicadores.

O artigo está organizado em cinco seções, incluindo esta introdução. A seção 2 traz informações sobre o SARS-Cov-2 e uma análise descritiva introdutória da situação da epidemia, desde seu início. A seção 3 apresenta uma descrição conceitual das principais estratégias de monitoramento epidemiológico, destacando o uso de gráficos de monitoramento de Shewhart. A seção 4 explica como gráficos de Shewhart podem ser adaptados para analisar dados relacionados a COVID-19. Ela apresenta ainda exemplos de análises de monitoramento do número de óbitos diários para os casos do Brasil, Pernambuco e São Paulo, bem como descreve a plataforma web utilizada para construir os gráficos. A seção 5 encerra o artigo com uma breve reflexão sobre o que ainda pode ser feito no combate a epidemia no Brasil.

2. A EPIDEMIA NO PAÍS E ESTADOS

Dados relativos à epidemia de COVID-19 têm sido disponibilizados diariamente pelo Ministério da Saúde, através do site covid.saude.gov.br. É possível acompanhar uma série de informações, incluindo o número de casos confirmados e de óbitos da doença no país e por estado da federação. A Figura 1 destaca a evolução do número de óbitos acumulados por estado, numa escala logarítmica, tendo como ponto de referência o dia de registro do primeiro óbito de cada estado. Os dados ilustram a situação até 18 de junho de 2020. Várias leituras do comportamento das curvas são possíveis. A primeira delas é a evidência de que os estados com maior número de mortes acumuladas são São Paulo, Rio de Janeiro, Ceará, Pará, Pernambuco e Amazonas. São Paulo e Rio de Janeiro destacam-se por terem adotado medida de distanciamento social após o início de registros de óbitos. Os dois são também os estados que apresentam maior gravidade da epidemia. São Paulo, no entanto, destaca-se acentuadamente no gráfico devido a velocidade com que as mortes devido a COVID-19 foram observadas.

Percebe-se que nos primeiros dias após o primeiro óbito registrado, o acúmulo do número de mortes de diversos estados era consistente com um crescimento exponencial em que o número de óbitos dobraria a cada dia (São Paulo, Pernambuco, Espírito Santo, Pará e Maranhão). Os casos de Rio de Janeiro, Bahia e Ceará, apresentaram padrão compatível com mortes que dobram a cada dois dias.

A Figura 2 apresenta a série temporal do número de óbitos diários registrados no Brasil até a finalização deste artigo. Observando-se o seu comportamento, percebe-se a variação naturalmente intrínseca a ação da doença. Em determinados dias o número de óbitos se apresenta menor que os do dia anterior, sem necessariamente implicar que esteja havendo uma

diminuição no padrão de mortes. Da mesma forma, em muitos casos, o registro de um número maior de óbitos pode não significar aumento no padrão de mortalidade da doença. É preciso considerar a série temporal como um todo para observar que o padrão de tendência observado no gráfico mostra uma nítida ascensão. A capacidade de determinar pontos de mudança no comportamento do processo, em meio a tais variações estocásticas, próprias do fenômeno, torna-se requisito de grande relevância para uma estratégia de monitoramento.

Figura 1- Evolução do número acumulado de mortes por Covid-19 nos estados brasileiros

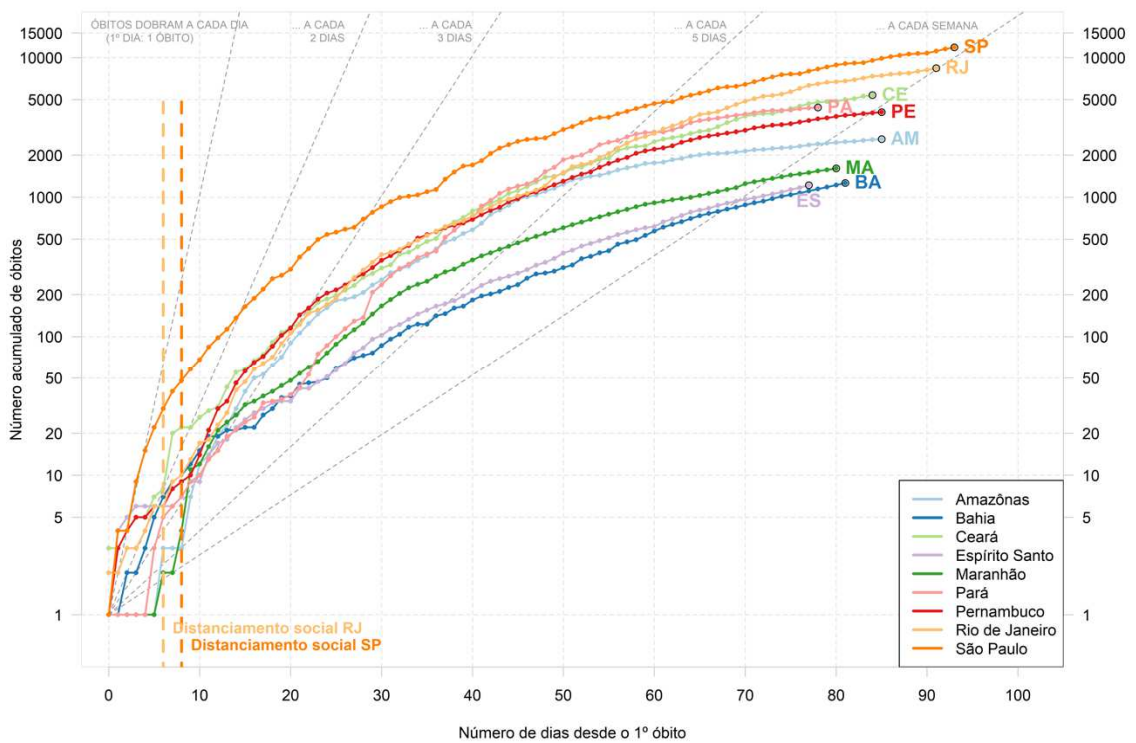
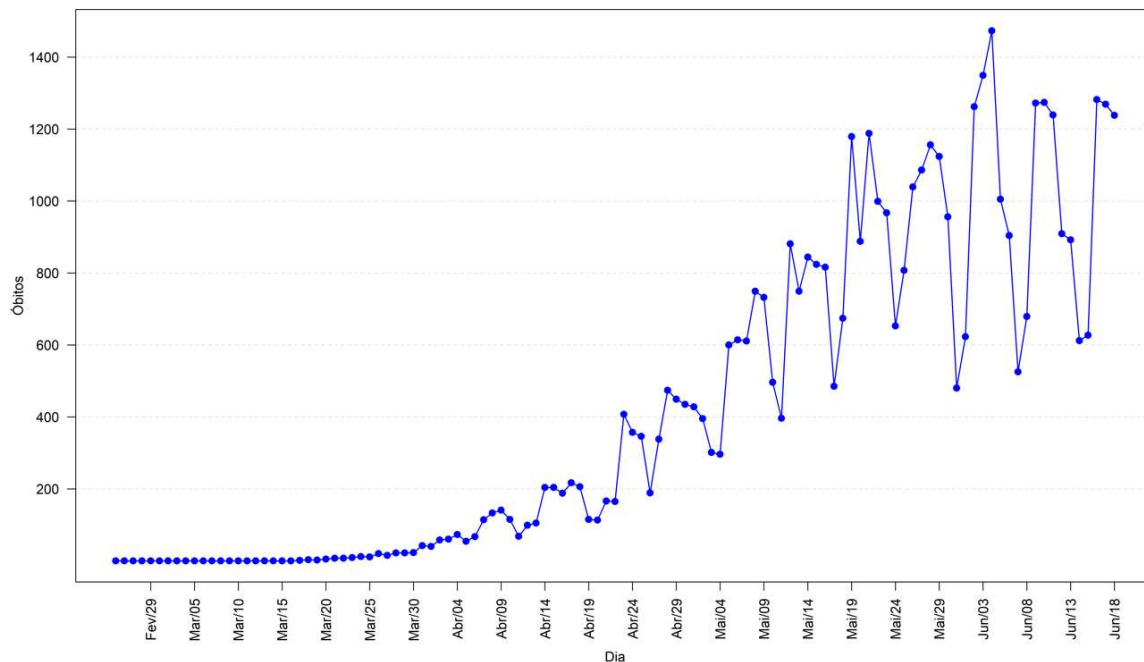


Figura 2 - Evolução do número diário de novas mortes por Covid-19 no país



3. ESTRATÉGIAS DE MONITORAMENTO

Monitorar a evolução da epidemia é importante por diversas razões. Ao acompanhar indicadores como o número de casos confirmados de COVID-19, ou o número de óbitos atribuídos a doença diariamente, tem-se ideia da frequência com que casos novos são registrados, demandando assistência por parte dos estados. Registros acumulados desses indicadores permitem visualizar a velocidade dos efeitos da epidemia, bem como avaliar a gravidade do impacto do problema sobre o sistema de saúde como um todo. Observando o comportamento dos números referentes à doença e indicadores de distanciamento social é possível avaliar o impacto dessas medidas e embasar, com critérios científicos, tomadas de decisão a respeito da melhor época para que elas sejam atenuadas. Estudos de monitoramento têm adotado diferentes objetivos, dentre os quais, destacam-se:

- a) Identificar (predizer) quando ocorrerá o pico dos principais indicadores da epidemia, tendo como referência o número de dias desde o primeiro caso oficial registrado;
- b) Prever o patamar de valores de indicadores a serem atingidos em datas de referência futura; e
- c) Identificar mudanças na evolução de indicadores e o estado atual da epidemia.

Vários modelos podem ser utilizados para dar suporte a análises visando os objetivos acima. Talvez o modelo SIR (Susceptible, Infected, Recovered), apresentado originalmente por Kermack and McKendrick (1927), e suas variantes, incluindo o SEIR (Susceptible, Exposed, Infected, Recovered) sejam os mais populares. Modelos Generalizados de Crescimento também têm sido considerados para avaliar as curvas epidêmicas (Vasconcelos, et al.; 2020). Menos comentados, mas com grande potencial para monitoramento do processo epidemiológico, são os gráficos de controle de Shewhart (Shewhart & Deming, 1986). Através deles, é possível atingir os objetivos descritos na parte c).

Desenvolvidos por Walter Shewhart entre o final da década de 20 e início de 30, os chamados gráficos de controle, que levam seu nome, são ferramentas tradicionais utilizadas no monitoramento estatístico de processos de produção e de serviços. Leitores interessados em explorar e estudar o assunto no contexto de Qualidade, podem consultar por exemplo Montgomery, D.C. (2016), ou Vardeman e Jobe (2016).

A estrutura de um gráfico de controle considera que, em cada tempo $t=1,2,\dots,m$, uma amostra $S(t)$, de n_t itens produzidos por um processo é obtida, e em cada item, uma característica é medida. Uma estatística (indicador) $Q(t)$ de interesse (média, desvio padrão, porcentagem, etc.), cuja escolha depende da característica medida em cada item, é calculada para resumir os valores medidos na amostra $S(t)$, representando o comportamento do processo naquele tempo t . Registram-se os valores de $Q(t)$ em um gráfico indexado no tempo, na ordem em que os valores são gerados. A média dos m valores deste indicador, $\hat{\mu}_Q$, define a linha central a ser traçada no gráfico. Calcula-se ainda uma medida de variação da estatística $Q(t)$, aqui denotada por σ_Q , que pode por exemplo ser um desvio-padrão ou amplitude. A escolha vai depender do tipo de característica medida em cada item (valor contínuo, contagem ou classificação) e do tamanho da amostra, se $n_t=1$, ou $n_t>1$. Quando $n_t=1$, o gráfico é chamado usualmente de Gráfico de Medidas Individuais, ou simplesmente Gráfico de Individuais. A medida de variação $\hat{\sigma}_Q$, nesse caso, pode ser calculada usando amplitudes móveis de tamanho 2:

$$\overline{MR} = \frac{\sum_{t=2}^m Q(t) - Q(t-1)}{m-1}; \tag{1}$$

$$\hat{\sigma}_{Q(\text{Individuais})} = 1,128 \overline{MR}; \tag{2}$$

A medida de variação $\hat{\sigma}_Q$ é multiplicada por uma constante k , tipicamente igual a 3, para gerar valores correspondentes aos limites inferior, $LI = \hat{\mu}_Q - k \hat{\sigma}_Q$, e superior, $LS = \hat{\mu}_Q + k \hat{\sigma}_Q$, de controle. Em seguida, inicia-se a análise do gráfico para identificar se o indicador apresenta comportamento estável ou se há indicação de causas especiais atuando sobre ele.

No contexto de uma epidemia como a de COVID-19, é possível analisar a evolução de indicadores de maneira análoga a de um processo de produção. O princípio básico por trás do método assume que observações são realizadas periodicamente, o que equivale a ter amostras de tamanho $n_t=1$, analisadas por meio de $Q(t)$, que nesse caso corresponde a própria observação. Tais estatísticas são confrontadas com los limites de monitoramento (linha central e limites de controle inferior e superior) traçados no gráfico ao longo do tempo. A Figura 3 ilustra o aspecto de um gráfico genérico de Shewart.

Os limites superior e inferior do gráfico na Figura 3 delimitam a área sombreada de referência para monitoramento. Tradicionalmente, esses limites são representados por linhas tracejadas, ou de cores distintas. Observando o comportamento do processo, representado pela estatística $Q(t)$, é possível identificar períodos em que haja estabilidade. Os parâmetros que definem valores para linha central e limites de controle podem ser estimados com base em experiências passadas, ou dados que estão sendo observados ao longo do tempo. Após um período de observação inicial é possível obter estimativas desses valores estabelecendo um primeiro gráfico de Shewhart. Os limites e linha central são então projetados para o futuro, para servirem de referência no monitoramento de dados subseqüentes. Esses dois passos iniciais são exemplificados na Figura 4, utilizando um conjunto inicial de sete observações (pontos). Dados continuam sendo coletados, e os valores do indicador, registrados no gráfico. A Figura 5 ilustra mudança de comportamento do processo, compatível com alteração de média do indicador. Uma nova fase é então detectada, e novos limites e linha central, estabelecidos para continuidade de monitoramento. Tanto na Figura 4 quanto na 5, os gráficos usam notação típica de softwares estatísticos, na qual $\mu_Q = \bar{X}$, $LI=LCL$ ("lower control limit"), e $LS=UCL$ ("upper control limit").

Figura 3 - Exemplo genérico de um gráfico de Shewhart para uma estatística $Q(t)$.

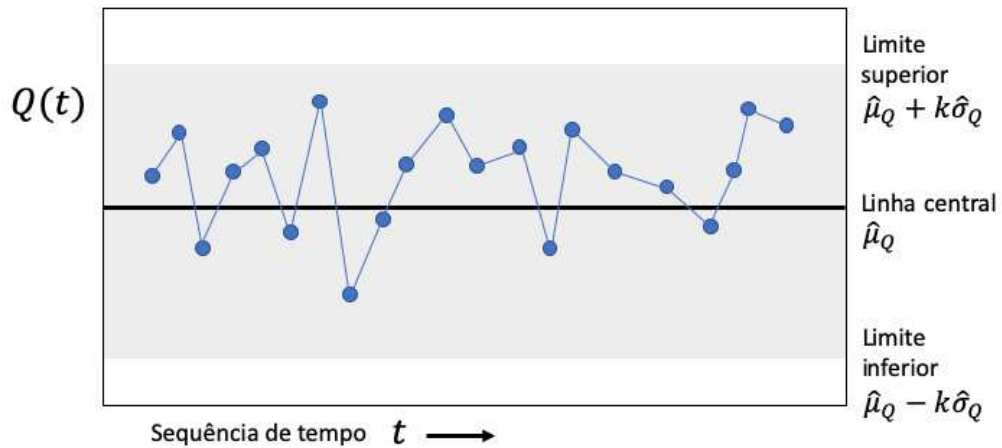


Figura 4 - Passos iniciais de monitoramento de processo

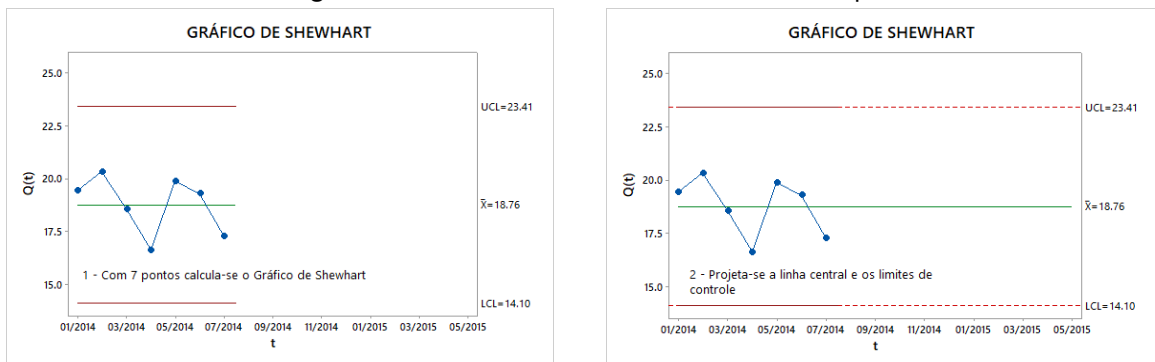
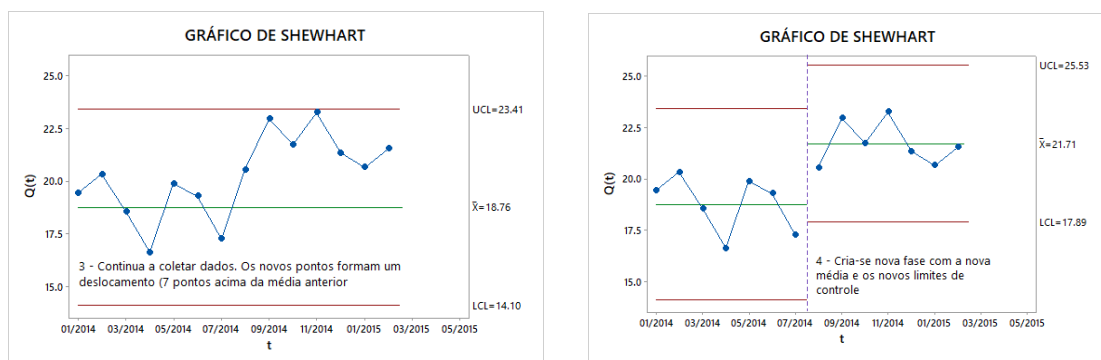


Figura 5 - Detecção de nova fase de processo.



Indicadores são usualmente considerados estáveis quando apenas causas comuns de variação estão atuando sobre ele. Para Shewhart, variações observadas em indicadores resultam da composição de variações de causa comum e de causa especial.

Variações de causa comum são variações aleatórias intrínsecas ao fenômeno em observação. A análise de dados de COVID-19 usa a própria observação de um indicador como estatística. Nesse artigo, o indicador considerado será o número diário de óbitos pela doença. A natureza aleatória dessa variável está ligada ao próprio fenômeno. Assim variações no número de mortes diárias são esperadas, como consequência da ação do novo Coronavírus (causa comum). Variações introduzidas por causas especiais estão relacionadas a mudanças no comportamento do indicador que não refletem o comportamento natural do fenômeno. No caso da epidemia, variações introduzidas por falhas nos registros dos dados, mudanças nas estratégias de isolamento social ou evolução do processo de tratamento dos contaminados são exemplos de ações que podem gerar variações de causas especiais agindo sobre o processo. Saiba-se, por exemplo, que em algumas situações no Brasil, o número registrado de mortes num determinado dia não coincide com o número de óbitos ocorridos de fato, pois parte desse número refere-se a data em que a causa mortis por COVID-19 foi identificada. Em um cenário usual de controle de processos, esforços são feitos para identificar, analisar e agir em conformidade com os efeitos produzidos pela ação de causas especiais nos indicadores. Mesmo na presença delas, é possível realizar monitoramento, tendo consciência de seus efeitos sobre o gráfico que está sendo utilizado.

4. GRÁFICOS DE SHEWHART PARA DADOS DE COVID-19

O uso mais frequente dos gráficos de Shewhart é em situações em que, nos períodos de estabilidade, os pontos oscilam aleatoriamente com uma mesma distribuição, mesma média e mesmo desvio-padrão. É possível fazer uma analogia e afirmar que o sinal é uma constante (média) e o ruído é a variação dos dados em torno do sinal. No caso do indicador de número de mortes por dia, o sinal não é constante. Em geral, seu crescimento é exponencial. Nesses casos, é possível modelar o sinal, e o que deve ser verificado, em termos de estabilidade, é o comportamento dos resíduos do modelo ajustado. Se o crescimento pode ser modelado por uma função de crescimento exponencial, descrita por

$$Q(t)=y_t=ae^{bt}, \quad (3)$$

na qual y_t representa o número de óbitos registrados no dia t ($t=1,2,\dots,m$) e, a e b são constantes a serem estimadas, então o $\log(y_t)$ é uma função linear do tempo t :

$$\log(y_t) = \log(a) + bt. \quad (4)$$

Assim, o algoritmo para adaptar o gráfico de Shewhart pode ser descrito como:

i) Ajusta-se $\log(y_t)$ em função de t através de um modelo de regressão linear simples:

$$\log(y_t) = \beta_0 + \beta_1 t + \varepsilon_t, \quad (5)$$

onde β_0 e β_1 são coeficientes, e ε_t um erro aleatório de valor esperado nulo e variância constante;

ii) Calculam-se os resíduos do modelo (5) ajustado:

$$\hat{\varepsilon}_t = \log(y_t) - \widehat{\log(y_t)} \quad ; \quad (6)$$

iii) Elabora-se o gráfico de Individuais de Shewhart para os resíduos definidos em (6) e obtém-se o limite superior do gráfico:

$$LS_{\text{resíduos}} = \hat{\varepsilon} + 3\hat{\sigma}_{Q(\text{Individuais})} \quad (7)$$

iv) Soma-se e subtrai-se o valor do limite superior $LS_{\text{resíduos}}$, calculado em (iii), aos valores ajustados pela regressão:

$$LI_{\text{regressão}} = \log(y_t) - LS_{\text{resíduos}}; LS_{\text{regressão}} = \log(y_t) + LS_{\text{resíduos}} \quad (8)$$

v) Calcula-se a linha central e os respectivos limites de controle do gráfico de Shewhart adaptado, exponenciando os valores ajustados, e os limites calculados no passo (iv).

Na implementação do algoritmo, decidiu-se utilizar o logaritmo na base 10, uma vez que:

$$\ln(y_t)\log_e(y_t) = \log_{10}(y_t)/\log_e(10) = 2.3026\log_{10}(y_t) \quad (9)$$

Por outro lado, é plausível que em algumas situações sejam observados valores nulos, isto é, $y_t = 0$, para algum $t=1,2 \dots m$. Por isso, para a implementação do Gráfico, foi usado o valor $\log_{10}(1+y_t)$ como variável resposta.

A adaptação descrita acima foi originalmente proposta por Lloyd Provost, e pode ser encontrada no artigo de Perla et al. (2020). Nele, os autores propõem o uso de um gráfico híbrido no qual o algoritmo de adaptação é usado para monitorar parte do processo, destacando-se a importância de identificação de ao menos três fases: pré-crescimento, crescimento e pós-crescimento (decaimento) dos números do indicador.

No presente artigo, propõe-se usar o gráfico adaptado de Shewhart, gerado pelo algoritmo acima, para monitorar todo o processo, com uma flexibilidade maior para a determinação de fases, definindo assim, uma proposta diferenciada em relação a Perla et al. (2020).

Uma vez estabelecido o Gráfico de Shewhart (adaptado) para monitorar indicadores como o número de óbitos por COVID-19, cabe adotar ainda uma ou mais regras que indiquem mudanças no comportamento do indicador ao longo do tempo, para que haja flexibilidade na definição de fases. Um vasto conjunto de regras para identificação de causas especiais atuando sobre indicadores estão disponíveis na literatura de Controle Estatístico de Processos. Lloyd S. Nelson (Nelson, 1984) propõe uma série dessas, dentre as quais, as três seguintes são as mais comumente utilizadas:

- *Regra 1:*

(Ponto astronômico) – Um ponto acima ou abaixo dos limites de controle

- *Regra 2:*

(Deslocamento) - Uma sequência de 7 ou mais pontos consecutivos abaixo ou acima da linha central do gráfico de monitoramento.

- *Regra 3:*

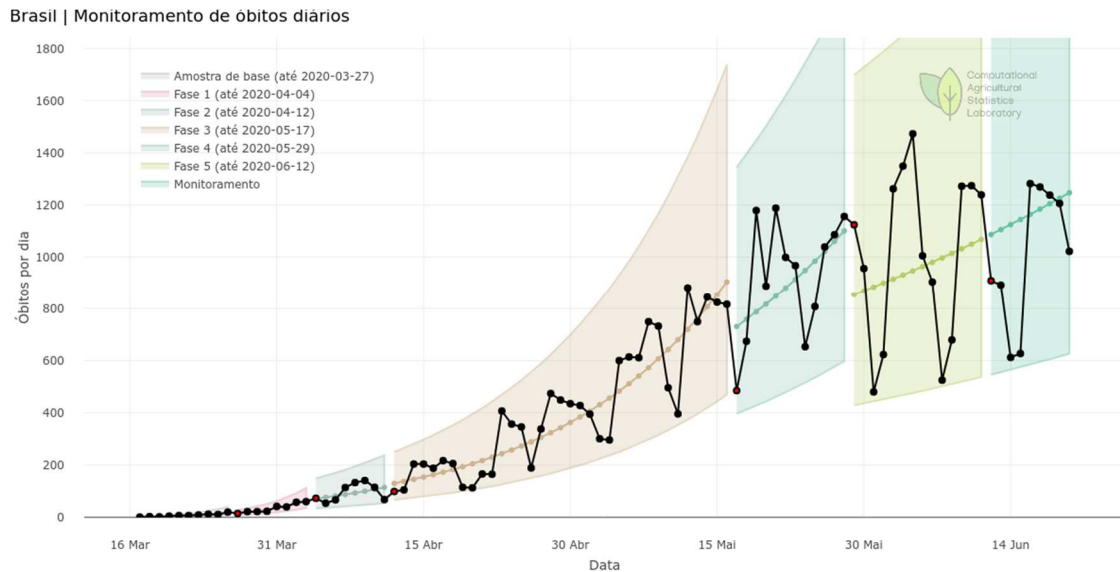
(Tendência) - Uma sequência de 6 ou mais pontos consecutivos crescentes, ou decrescentes.

Na literatura há variações em relação ao ponto de corte das Regras 2 e 3, podendo ser maiores ou menores dependendo dos valores dos riscos associados aos erros Tipo I (identificar como causa especial uma causa comum) e Tipo II (identificar como causa comum uma causa especial).

Nesse artigo propõe-se que seja usada apenas a regra do deslocamento para identificar possíveis mudanças na taxa de aumento (ou redução) do indicador. Propõe-se também uma abordagem adaptativa em que, ao analisar o histórico de um indicador, sejam escolhidos pontos de forma a definir fases, simulando uma estratégia de monitoramento prospectiva.

Exemplos de Gráficos de Shewhart (adaptados), utilizando dados do Ministério da Saúde (<https://covid.saude.gov.br/>) até a data de 20 de junho de 2020, e construídos com base na estratégia proposta, são apresentados a seguir. A Figura 6 mostra o Gráfico para o Brasil, até a data de finalização deste artigo. Para construir esse, e os demais gráficos, foram utilizados os 10 primeiros dias de observações para gerar os parâmetros iniciais de monitoramento, calculados utilizando o algoritmo de adaptação descrito nesta seção. A linha central e os limites de controle foram projetados para os dias seguintes, e daí em diante a regra de deslocamento foi sendo empregada sistematicamente. A Fase 1, por exemplo, foi determinada ao selecionar o ponto correspondente à data de 4 de maio, por ser a primeira data após uma sequência de sete pontos acima da linha central, observados nos sete dias imediatamente anteriores a 4 de maio. O algoritmo de adaptação foi então empregado aos dados da Fase 1, e seus limites projetados para as datas futuras, repetindo-se a análise. Ao final, para o Brasil, são identificadas 5 fases, a quinta delas fornecendo parâmetros para o monitoramento mais recente.

Figura 6 - Gráfico de Shewhart para mortes por COVID-19 no Brasil



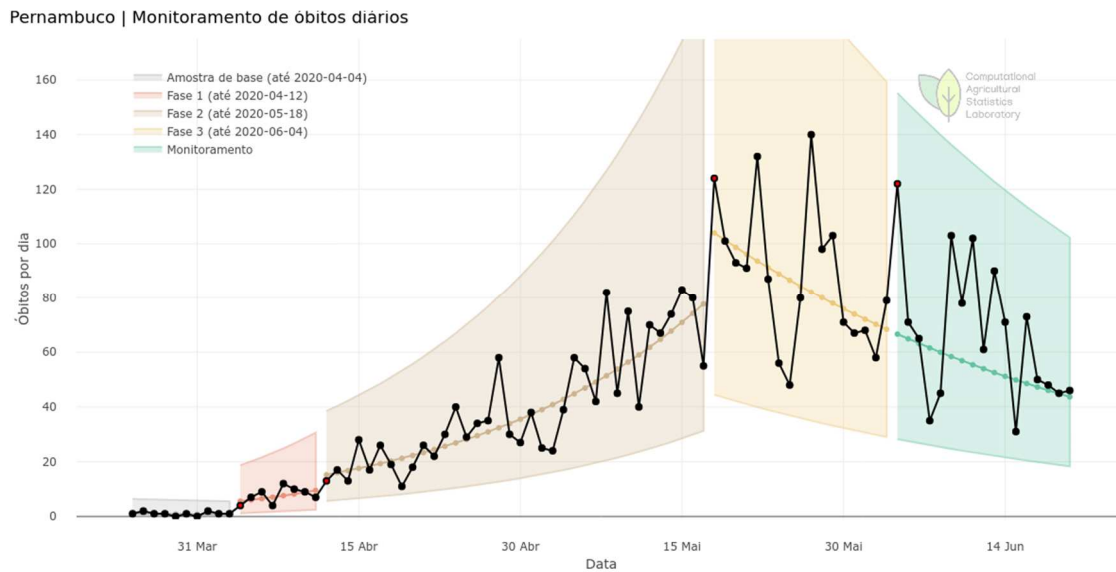
É possível visualizar uma pequena queda no nível de crescimento de mortes, da terceira para a quarta fase, embora a tendência ainda seja de crescimento. É importante registrar que, a partir do dia 7 de junho, o Ministério da Saúde adotou números oficiais que não tem a mesma definição dos anteriores, sem, no entanto, descrever qual definição está sendo utilizada. Nesse dia, inicialmente foi divulgado um valor superior a 1.000 mortes, enquanto ao final, esse número foi reduzido para o patamar de pouco mais de 500 mortes.

O Gráfico de Shewhart para monitoramento de óbitos em Pernambuco, apresentado na Figura 7, apresenta um padrão de comportamento em que se identifica uma tendência decrescente, a partir de meados de maio, até a data em que este artigo é finalizado. O período a que se refere a Fase 3, que está servindo de base para monitoramento subsequente, coincide com o período em que foi determinado um regime de *lockdown* na capital e região metropolitana.

A Figura 8 mostra a situação no estado de São Paulo. Até a data em que este artigo é encerrado, percebe-se a tendência de crescimento estável no número de óbitos diários.

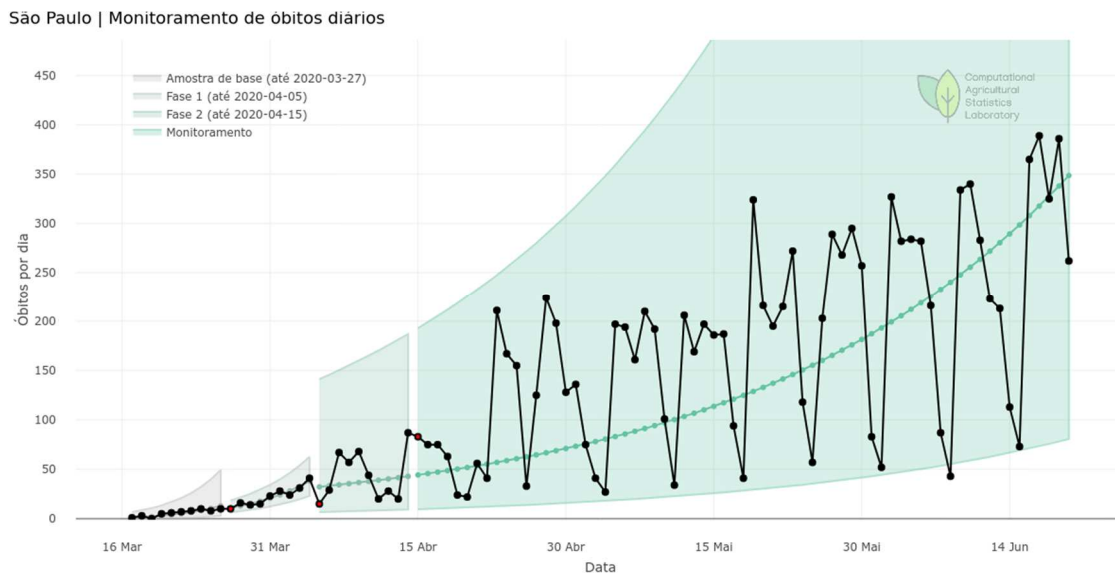
Embora os gráficos de Shewhart indiquem a situação mais recente do Brasil e dos estados da Federação, a velocidade com que o SARS-Cov-2 é transmitido tem papel de grande influência nas variações constantes de tendência.

Figura 7 - Gráfico de Shewhart para mortes por COVID-19 em Pernambuco



Dessa forma, detectar que o momento atual é de decrescimento (ou crescimento) no número médio diário de óbitos não significa que necessariamente esse padrão permanecerá de forma estável por um longo período de tempo. Monitoramento constante é necessário.

Figura 8. Gráfico de Shewhart para mortes por COVID-19 em São Paulo



4.1 PLATAFORMA WEB

Os exemplos de gráficos de Shewhart mostrados foram construídos usando a plataforma web covid19.castlab.org. Nesta plataforma, usuários são convidados a fazer o monitoramento da série de óbitos diários fornecida pelo Ministério de Saúde (covid.saude.gov.br). Na plataforma, pontos são escolhidos clicando sobre eles diretamente no gráfico, ou clicando sobre datas no calendário, determinando fases de monitoramento.

A cada ponto escolhido, o gráfico ajusta-se automaticamente para refletir novos parâmetros de monitoramento (linha central e limites). É possível voltar atrás de uma escolha clicando novamente sobre ela (o ponto escolhido). Após construir um gráfico, o usuário tem opções de salvar a imagem e baixar os dados utilizados na sua construção.

A versão da plataforma disponível ao término deste artigo permite analisar dados do Brasil, estados e municípios. Também permite que o usuário use uma base de dados própria, desde que seja observada a mesma estrutura para a tabela de dados. Aprimoramentos na plataforma continuarão a ser feitos pelo Laboratório CAST (CASTLab) e seus parceiros.

5. COMENTÁRIOS FINAIS

Gráficos de Shewhart, adaptados para monitoramento de óbitos diários por COVID-19 foram apresentados como uma ferramenta de grande potencial para ser agregada às análises de dados relativos aos efeitos do SARS-Cov-2 na população. Não obstante sua utilidade, ações de combate a epidemia devem ser embasadas em um conjunto de fatores, não sendo razoável ter apenas o número de óbitos diários como referência para tomadas de decisão como relaxamento de medidas de isolamento social. A gravidade da doença, como descrita na introdução, exige mecanismos de monitoramento e vigilância constantes da situação, que incluam também indicadores como número diário de infectados, capacidade de leitos de UTI, número de chamadas do SAMU, dentre outros. Todos esses indicadores podem ser monitorados através de Gráficos de Shewhart, devidamente adaptados.

Por fim, é importante considerar as limitações de alcance dos dados oficiais, como os aqui considerados, bem como de outras fontes. Subnotificações relacionadas a falta de testes para diagnósticos se fazem presentes tanto em número de casos confirmados quanto em número de óbitos, por COVID-19. Por outro lado, existe ainda a necessidade de levantamento de indicadores essenciais, que tenham abrangência populacional. Poucas tem sido as

iniciativas de realização de pesquisas por amostragem probabilística de populações cujo plano amostral permita gerar estimativas de prevalência de casos de COVID-19, incluindo casos assintomáticos, de forma eficiente, abrangendo diferentes domínios de estimação. Aumentar esforços para produzir pesquisas por amostragem probabilística longitudinais, combinados com estratégias de monitoramento que façam uso de Gráficos de Shewhart, são medidas que ainda podem ser utilizadas no combate a COVID-19 no Brasil.

6. REFERÊNCIAS BIBLIOGRÁFICAS

Bedford, J., Enria, D., Giesecke, J., Heymann, D.L., Ihekweazu, C., Kobinger, G., Lane, H.C., Memish, Z., Oh, M.D., Schuchat, A. & Ungchusak, K. (2020). COVID-19: towards controlling of a pandemic. *The Lancet*, 395(10229), 1015-1018.

Rothan, H. A., & Byrareddy, S. N. (2020). The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *Journal of autoimmunity*, 102433.

Longo, D. L., Kasper, D. L., Jameson, J. L., Fauci, A. S., Hauser, S. L., & Loscalzo, J. (2013). *Medicina interna de Harrison*. In *Medicina interna de Harrison* (pp. 1796-1796).

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N. & Bi, Y, Y. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, 395(10224), 565-574.

Holmes, K. V. (2003). SARS-associated coronavirus. *New England Journal of Medicine*, 348(20), 1948-1951.

Park, S. E. (2020). Epidemiology, virology, and clinical features of severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2; Coronavirus Disease-19). *Clinical and experimental pediatrics*, 63(4), 119.

Wilder-Smith, A., Chiew, C. J., & Lee, V. J. (2020). Can we contain the COVID-19 outbreak with the same measures as for SARS?. *The Lancet Infectious Diseases*, 20(5), e102-e107.

Baud, D., Qi, X., Nielsen-Saines, K., Musso, D., Pomar, L., & Favre, G. (2020). Real estimates of mortality following COVID-19 infection. *The Lancet infectious diseases*.

World Health Organization. (2020). Coronavirus disease 2019 (COVID-19): situation report, 94. [https://www.who.int/emergencies/diseases/novel-](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/)

[coronavirus-2019/situation-reports/](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/)

Santarpia, J.L., Rivera, D.N., Herrera, V., Morwitzer, M.J., Creager, H., Santarpia, G.W., Kevin, K., Brett-Major, D., Schnaubelt, E., Broadhurst, M.J. & Lawler, J.V., (2020). Transmission Potential of SARS-CoV-2 in Viral Shedding Observed at the University of Nebraska Medical Center. <https://www.medrxiv.org/content/10.1101/2020.03.23.20039446v2.full.pdf>

Van Doremalen, N., Bushmaker, T., Morris, D.H., Holbrook, M.G., Gamble, A., Williamson, B.N., Tamin, A., Harcourt, J.L., Thornburg, N.J., Gerber, S.I. & Lloyd-Smith,

- J.O., (2020). Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. *New England Journal of Medicine*, 382(16), pp.1564-1567.
- Chin, A.W., Chu, J.T., Perera, M.R., Hui, K.P., Yen, H.L., Chan, M.C., Peiris, M. & Poon, L.L., (2020). Stability of SARS-CoV-2 in different environmental conditions. *The Lancet Microbe*, 1(1), p.e10.
- Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y. & Zhou, Q., (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*, 367(6485), pp.1444-1448.
- Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T. & Velesler, D., (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*, 180, pp.1-12.
- Yuen, K.S., Ye, Z.W., Fung, S.Y., Chan, C.P. & Jin, D.Y., (2020). SARS-CoV-2 and COVID-19: The most important research questions. *Cell & bioscience*, 10(1), pp.1-5.
- Kermack, M. & McKendrick, A., (1927). Contributions to the mathematical theory of epidemics. Part I. *Proc. r. soc. a*, 115(5), pp.700-721.
- Vasconcelos G.L., Macêdo A.M.S., Ospina R., Almeida F.A.G., Duarte-Filho G.C., Brum A.A. & Souza I.C.L. (2020). "Modelling fatality curves of COVID-19 and the effectiveness of intervention strategies." *PeerJ*, 8: e9421. DOI: <https://doi.org/10.7717/peerj.9421>
- Shewhart, W. A., & Deming, W. E. (1986). *Statistical method from the viewpoint of quality control*. Courier Corporation.
- Perla, J.R., Provost, S.M., Parry, G.J., Little, K., Provost, L.P. (2020) "Understanding Variation in COVID-19 Reported Deaths with a novel Shewhart Charts application." *International Journal for Quality in Health Care*. DOI: <https://doi.org/10.1093/intqhc/mzaa069>.
- Nelson, Lloyd S. (1983) *The Shewhart Control Chart – Tests for Special Causes*, *Journal of Quality Technology*, Vol. 16, 1984.
- Montgomery, D.C. (2016). *Introdução ao Controle Estatístico da Qualidade*; 7ª ed. Editora LTC.
- Vardeman, S.B. and Jobe, M. (2016) *Statistical Methods for Quality Assurance*; 2nd edition, Springer.

AGRADECIMENTOS e COLABORAÇÕES

Esse trabalho foi parcialmente apoiado pelo Conselho Nacional de Desenvolvimento Científico através dos processos de bolsa de produtividade de N.º 305305/2019-0 (RO).

Os autores agradecem os comentários e sugestões do editor e avaliador, que muito contribuíram para melhoria do texto, no sentido de proporcionar uma exposição ainda mais clara sobre o assunto.

MODELAGEM E PREDIÇÃO DAS QUANTIDADES DE CASOS CONFIRMADOS DA COVID-19 NO ESTADO DO MATO GROSSO DO SUL

Erlandson Ferreira Saraiva

erlandson.saraiva@ufms.br

Universidade Federal do Mato Grosso do Sul

Leandro Sauer

leandro.sauer@ufms.br

Universidade Federal do Mato Grosso do Sul

Resumo: Nos últimos meses, a COVID-19 têm sido um dos principais problemas de saúde do mundo, causando um enorme ônus social e econômico. Neste cenário, modelos matemáticos ou estatísticos são extremamente importantes para se entender a dinâmica de propagação da doença e previsão das quantidades de novos casos. Neste artigo, apresentamos os resultados de uma modelagem matemática/estatística das quantidades de casos confirmados da COVID-19 no estado do Mato Grosso do Sul. Apresentamos as previsões para a máxima quantidade de notificações e a data em que as taxas de notificações começarão a decrescer, chamado de ponto de inflexão. Na prática, o ponto de inflexão nos mostra quando as atitudes de combate estarão surtindo efeito. Além disso, apresentamos as projeções referentes as quantidades de leitos clínicos e de UTI que serão necessários para atendimento dos pacientes diagnosticados com a COVID-19.

Palavras-chave: COVID-19, Modelos de crescimento, Estimação, Seleção de modelos, Predição.

Abstract: In the last few months, COVID-19 has been one of the main health problems in the world, causing enormous social and economic costs. In this scenario, mathematical/statistical models are extremely important to understand the dynamics of the spread of the disease and predict the quantities of new cases. In this article, we present the results of a mathematical/statistical modeling of the quantities of confirmed cases of COVID-19 in the state of Mato Grosso do Sul. We present the prediction for the maximum amount of notifications and the date when the notification rates will start to decrease called an inflection point. In practice, the tipping point shows us when combat attitudes will be having an effect. In addition, we project the number of patients who will need treatment in intensive care units.

Keywords: COVID-19, Growth models, Estimation, Model selection, Prediction.

1. INTRODUÇÃO

No mês de Dezembro do ano de 2019, um novo coronavírus foi descoberto na cidade de Wuhan, China. A Organização Mundial da Saúde (OMS) nomeou oficialmente esse coronavírus como COVID-19 (WU & McGoogan, 2020; Ahmadi *et al.*, 2020). Desde a sua descoberta, o vírus se espalhou rapidamente pelo mundo, sendo atualmente um dos principais problemas de saúde, causando um enorme ônus social e econômico.

Neste cenário, análises quantitativas e o ajuste de modelos matemáticos ou estatísticos aos dados referentes as quantidades de casos confirmados são extremamente importantes para se entender a dinâmica de propagação da doença, previsão e projeção das quantidades de novos casos. Além disso, os modelos permitem a realização de simulações computacionais com o objetivo de obter a tendência futura das quantidades de casos confirmados da doença.

A partir dos primeiros casos confirmados da COVID-19 na cidade de Campo Grande, MS, a secretária municipal de saúde da cidade solicitou a Universidade Federal de Mato Grosso do Sul (UFMS) um estudo matemático/estatístico que projete semanalmente a quantidade de casos confirmados na cidade de Campo Grande e no estado do Mato Grosso do Sul (MS). O interesse da prefeitura neste estudo foi para embasar as decisões, tais como, ampliação de leitos hospitalares, compras de equipamentos, relaxamento ou não do isolamento social, entre outros.

A partir da solicitação, iniciou-se o procedimento de modelagem das quantidades totais de casos confirmados da COVID-19 na cidade de Campo Grande e no estado do MS. Semanalmente, a toda segunda-feira, foi disponibilizado a representantes da prefeitura e do estado um relatório técnico com os resultados. Até a data de submissão deste artigo foram disponibilizados um total de 18 relatórios.

Neste artigo, descrevemos de forma sucinta o procedimento de modelagem adotado e os resultados da análise dos dados registrados no estado do MS no período de 14/03/20 à 09/08/2020. Para a análise dos dados, consideramos o ajustes de três modelos de crescimento. São eles: Modelo

Exponencial (Abramowitz & Stegun 1965), Modelo Logístico (Blumberg, 1968) e o Modelo Gompertz (Gompertz, 1825; Tjorve & Tjorve, 2017).

Devido ao aumento significativo das quantidades de casos confirmados registrados a partir do dia 08/05/20 não foi possível ajustar um único modelo. Devido a isto, consideramos uma divisão dos dados observados em dois sub-conjuntos de dados, denominados de D_1 e D_2 . O sub-conjunto de dados D_1 é composto pelas quantidades totais registradas até o dia 58 (10/05/20); e o sub-conjunto D_2 é composto pelas quantidades totais do dia 59 (11/05/20) até o dia 148 (09/08/20). A escolha do dia 59 como ponto de separação foi feita utilizando como critério o erro quadrático médio de modelos ajustados para diferentes dias de separação. Ajustamos os três modelos aos sub-conjuntos de dados D_1 e D_2 .

Os modelos foram comparados, utilizando como critério o erro padrão residual e os critérios de seleção de modelos AIC (Akaike, 1974; Bozdogan, 1987) e BIC (Schwarz, 1978). De acordo com estes três critérios, o modelo Gompertz é o modelo que melhor explica os dados. Com base no modelo selecionado, estimamos a quantidade máxima de notificações, a data que isto irá ocorrer e as quantidades de pacientes que precisarão de atendimento em leitos clínicos e de UTI.

O restante do artigo está organizado da seguinte maneira. Na seção 2, apresentamos os dados utilizados e os três modelos de crescimento considerados. Na seção 3, apresentamos os modelos ajustados e os resultados de interesse. Na seção 4 apresentamos as considerações finais.

2.DADOS E MODELOS

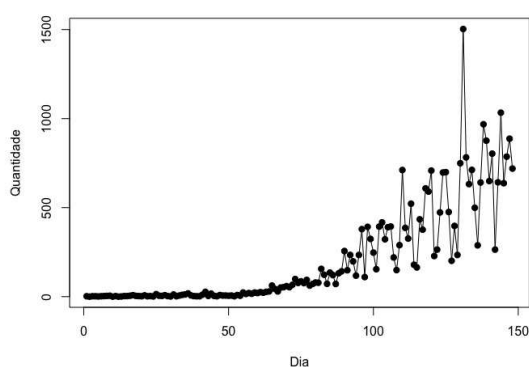
Os dados utilizados neste artigo são referentes as quantidades de casos confirmados da COVID-19 no estado de Mato Grosso do Sul (MS), no período de 14/03/20 à 09/08/2020. Nestes 149 dias houve um total de 31.344 casos confirmados e 509 mortes (Fonte: www.coronavirus.ms.gov.br). Do total de casos confirmados, 25.195 são considerados recuperados e 5.132 estão em isolamento domiciliar. Além disso, há um total de 514 pacientes diagnosticados com a COVID-19 internados, sendo 298 em leitos clínicos e 216 em leitos de UTI.

Dos 31.344 casos confirmados, 17.512 casos foram registrados nas duas maiores cidades do estado, Campo Grande e Dourados. Este valor representa 55,87% do total de casos registrados. Campo Grande registrou 12.934 casos (41,26%) e Dourados registrou 4.578 casos (14,61%).

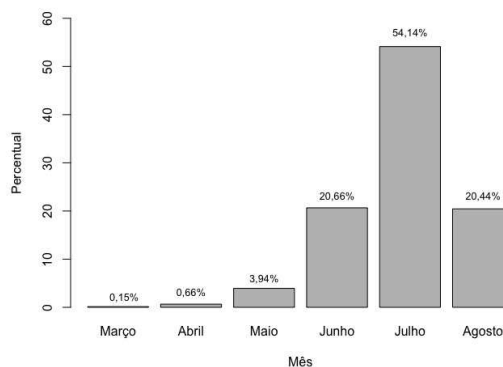
A Figura 1(a) mostra as quantidades de casos confirmados no dia t , para $t=0,1,\dots,n=148$, onde $t=0$ representa a data em que houve o primeiro caso confirmado (14/03/20) e $t=148$ representa a data 09/08/20. A maior quantidade de casos confirmados em um dia foi 1.503 casos e ocorreu no dia 130 (22/07/20).

A Figura 1(b) mostra a distribuição de frequências das quantidades de casos confirmados por mês. Do total de casos confirmados, 0,15% (48 casos) foram registrados no mês de Março, 0,66% (207 casos) no mês de Abril, 3,94% (1.234 casos) no mês de Maio, 20,66% (6.476 casos) no mês de Junho, 54,14% (16.971) no mês de Julho e 20,44% (6.408) nos nove primeiros dias no mês de Agosto.

Figura 1 - Quantidades de notificações por dia e distribuição de frequências



(a) Quantidades por dia



(b) Distribuição de Frequências

O mês de Junho apresentou um aumento significativo de casos confirmados em relação aos três meses anteriores; passando de 1.489 casos em 31/05/20 (dia 78) para 7.965 casos em 30/06/20 (dia 108). Um aumento de 434,93%. Infelizmente, o mês de Julho apresentou um cenário de crescimento maior do que o mês de Junho; passando de 7.965 casos em 30/06/20 para 24.936 casos em 31/07/20 (dia 139). Um aumento de 213,07%. O número de

óbitos passou de 76 (31/07) para 376 (09/08). Um aumento de 394,74%. Somente nos nove primeiros dias do mês de Agosto ocorreram 133 óbitos.

A Tabela 1 mostra as estatísticas resumo das quantidades de casos confirmados. O valor mediano é de 71 casos diários, com média de 207,50 casos confirmados por dia e desvio-padrão de 281,49.

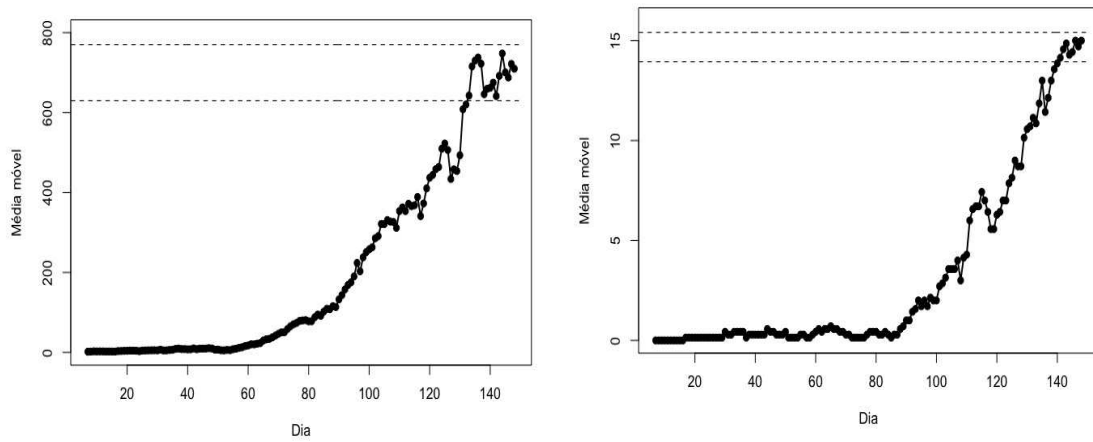
Tabela 1 – Estatísticas Descritiva

Mín.	1o Q.	Median	Média	D.P.	3o Q.	Máx
0	6	71	207,50	281,49	324,50	1.503

A Figura 2 mostra o gráfico das médias móveis de sete dias para a quantidade de casos confirmados e para o número de óbitos. Nos últimos 15 dias, o valor das médias móveis do número de casos confirmados esteve dentro do intervalo de $\pm 10\%$ da média dos valores das médias móveis neste período (linhas tracejadas no gráfico). Se considerarmos este intervalo como sendo um indicativo de estabilização, então temos uma estabilização em torno de 700 casos diários, que é um valor muito elevado e preocupante.

Com relação as médias móveis do número de óbitos, nos últimos nove dias os valores das médias móveis estiveram dentro de um intervalo de $\pm 5\%$ da média dos valores das médias móveis neste período. Neste período, há em média 15 óbitos por dia.

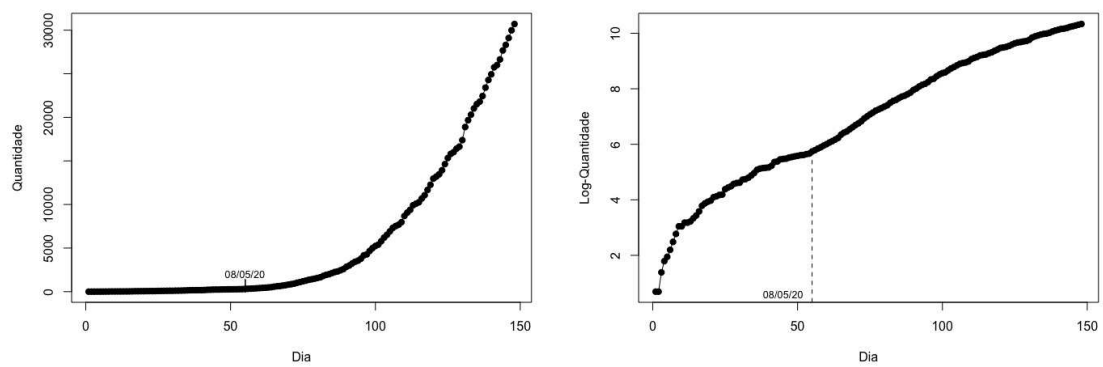
Figura 2 - Média móvel de sete dias para as quantidades de notificações e de óbitos



(a) Casos diários

A Figura 3(a) mostra a quantidade total de notificações até o dia t e a Figura 3(b) mostra o mesmo gráfico na escala logarítmica, para $t=0,1,\dots,n$. Nosso objetivo é modelar as quantidades acumuladas de casos da COVID-19 no MS.

Figura 3 - Quantidades total de casos confirmados até o dia t , para $t=0,1,\dots,n$



(a) Escala original

(b) Escala Logarítmica

2.1 Modelos

Para modelar o crescimento do número de casos confirmados da COVID-19 no estado do MS, consideramos três modelos de crescimento. O

primeiro modelo é o modelo Exponencial. A expressão matemática para a projeção de casos deste modelo é dada por

$$N_t = \alpha_1 \exp(\alpha_2 t), \quad (1)$$

onde N_t é a quantidade total de casos confirmados no tempo t , α_1 e α_2 são os parâmetros do modelo, para $t \geq 0$. O valor de $N_{t \neq t'}$ representa a quantidade no tempo inicial $t=0$ e o valor de α_2 é a taxa de crescimento.

Tomando o logaritmo (base e) em ambos os lados da Equação (1), obtemos o modelo Log-Exponencial,

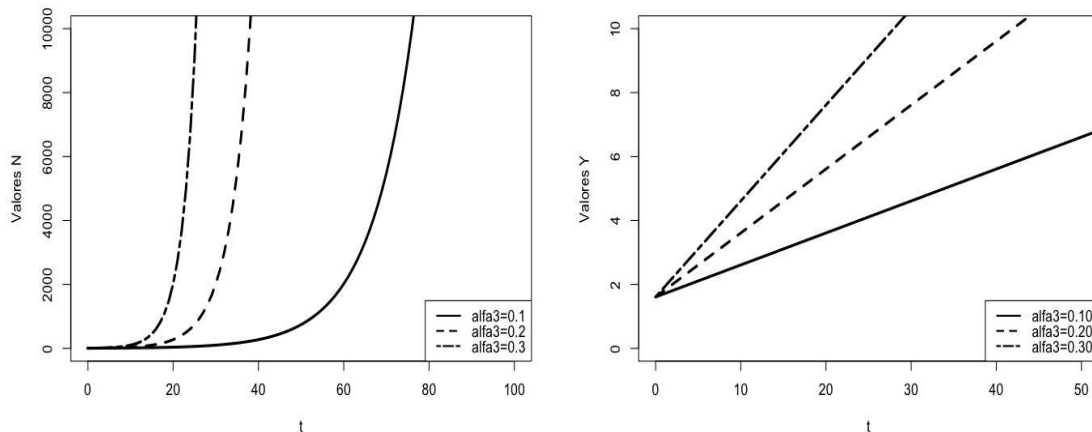
$$Y_t = \log(N_t) = \alpha_1 + \alpha_2 t, \quad (2)$$

para $t \geq 0$.

A Figura 4(a) mostra o gráfico do modelo exponencial para $\alpha_1 = 5$ e $\alpha_2 = \{0,10; 0,20; 0,30\}$. A Figura 4(b) mostra o mesmo gráfico na escala logarítmica. Note que, aumentando o valor de α_2 (taxa de crescimento) mais rápido é o crescimento da curva. Além disso, neste modelo, a taxa de crescimento permanece a mesma durante todo o intervalo de tempo do estudo, independentemente do tamanho da população.

Como este modelo apresenta um crescimento ilimitado, isto pode ser um tanto irrealístico do ponto de vista prático, pois as quantidades em estudo serão em algum momento limitadas pela disponibilidade de recursos, tais como, espaço, alimento, tamanho da população, entre outras.

Figura 4 - Modelo exponencial e log-exponencial



(a) Exponencial

(b) Log-exponencial

Devido a isto, consideramos como segundo modelo o modelo de crescimento Logístico. Este modelo é caracterizado por acréscimos crescentes no período inicial do estudo seguido de acréscimos decrescentes a medida que se aproxima do valor máximo. A expressão matemática do modelo Logístico é dada por

$$N_t = \frac{\alpha_1}{1 + \alpha_2 \exp(-\alpha_3 t)}, \quad (3)$$

onde N_t é a quantidade total de casos confirmados no tempo t e α_1 , α_2 e α_3 são os parâmetros do modelo, para $t \geq 0$.

O comportamento gráfico deste modelo é uma curva em forma de S, em que, o parâmetro α_1 representa a assíntota superior. No caso do estudo da COVID-19 este valor representa a projeção para a máxima quantidade de casos confirmados. O parâmetro α_3 é a taxa média de crescimento (afeta a inclinação da curva). O parâmetro α_2 está associado as coordenadas do ponto de inflexão da curva, *i.e.*, quando as taxas de crescimento começam a diminuir.

As coordenadas do ponto de inflexão é dada por (T, N_T) , para $T = \frac{\log(\alpha_2)}{\alpha_3}$ e $N_T = \frac{\alpha_1}{2}$, onde T é o tempo que irá ocorrer a mudança de comportamento da curva.

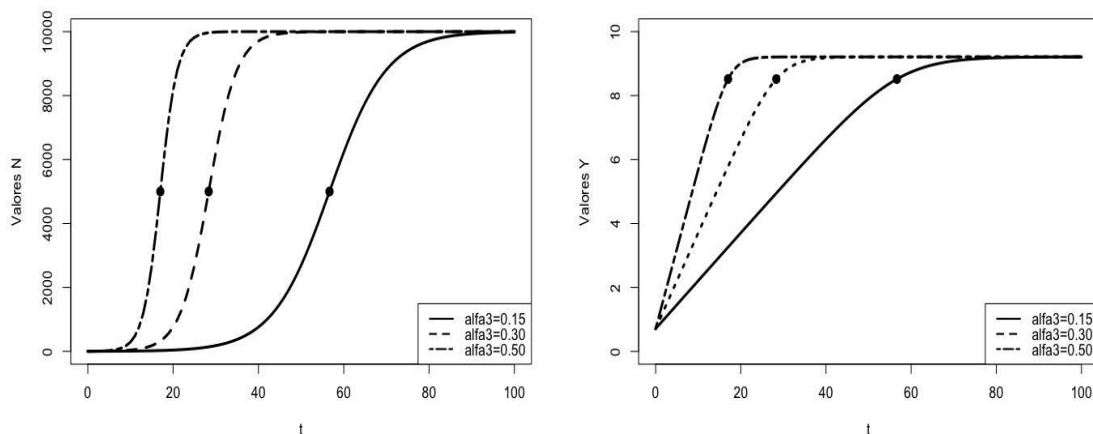
Tomando o logaritmo em ambos os lados da Equação (3), obtemos o modelo Log-Logístico, dado por

$$Y_t = \log(N_t) = \log(\alpha_1) - \log(1 + \exp(-\alpha_3 t)), \quad (4)$$

para $t \geq 0$.

A Figura 5 mostra o gráfico do Modelo Logístico e Log-Logístico para $\alpha_1 = 10.000$, $\alpha_2 = 8,5$ e $\alpha_3 \in \{0,15; 0,30; 0,50\}$. Ou seja, as figuras mostram o comportamento do modelo Logístico para um valor máximo $\alpha_1 = 10.000$. Os símbolos \bullet nos gráficos representam os pontos de inflexão. Como podemos notar nesta Figura os gráficos são simétricos em relação ao ponto de inflexão. Aumentando o valor de α_3 (taxa de crescimento) e mantendo α_1 fixo, mais rápido é o crescimento da curva.

Figura 5 - Modelo Logístico e Log-Logístico



(a) Logístico

(b) Log-Logístico

Como terceiro modelo, consideramos o modelo Gompertz. Este modelo possui um comportamento similar ao modelo Logístico, porém não é simétrico

em relação ao ponto de inflexão. A expressão matemática do modelo Gompertz é dada por

$$N_t = \alpha_1 \exp(-\alpha_2 \exp(-\alpha_3 t)), \quad (5)$$

onde N_t é a quantidade total de casos confirmados no tempo t e α_1 , α_2 e α_3 são os parâmetros do modelo, para $t \geq 0$. De maneira similar ao modelo Logístico, no modelo Gompertz o valor do parâmetro α_1 é a assintota superior. O parâmetro α_3 é a taxa média de crescimento e o parâmetro α_2 determina as coordenadas do ponto de inflexão através das expressões $T = \frac{\log(\alpha_2)}{\alpha_3}$ e N_T .

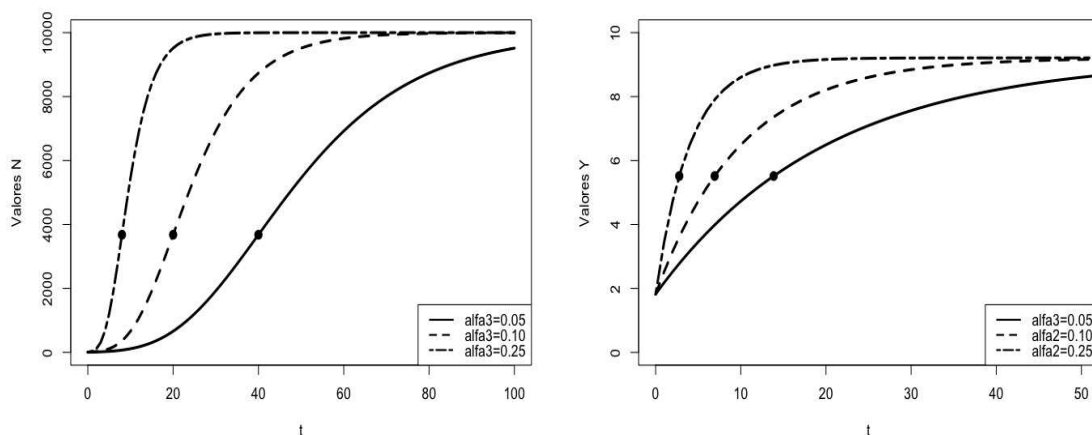
Tomando o logaritmo em ambos os lados da Equação (5), obtemos o modelo Log-Gompertz, dado por

$$Y_t = \log(N_t) = \log(\alpha_1) - \alpha_2 \exp(-\alpha_3 t), \quad (6)$$

para $t \geq 0$.

A Figura 6 mostra o gráfico do modelo Gompertz e log-Gompertz para $\alpha_1 = 10.000$, $\alpha_2 = 150$ e $\alpha_3 = \{0,10; 0,15; 0,20\}$. Os pontos de inflexão são destacados nos gráficos pelos símbolos \bullet . Aumentando o valor de α_3 (taxa de crescimento) e mantendo α_1 fixo, mais rápido é o crescimento da curva.

Figura 6 - Modelo Gompertz e Log-Gompertz



(a) Gompertz

(b) Log-Gompertz

3.RESULTADOS

Para modelar o crescimento do número de casos confirmados da COVID-19 no estado do MS consideramos os três modelos de crescimento descritos na seção anterior, na escala logarítmica. Além disso, como os valores observados não se comportam exatamente como os modelos, consideramos os modelos acrescidos de um erro aleatório. Ou seja, assumimos que

$$Y_t = f(t/\theta) + \varepsilon_t,$$

(7)

onde $f(t/\theta)$ é dada pelas funções em (2), (4) ou (6) e ε_t , é um erro aleatório assumido como sendo proveniente de uma distribuição Normal com média 0 e variância σ^2 , i.e., $Y_t \sim N(0, \sigma^2)$, com $\sigma^2 < \infty$ e $Cov(\varepsilon_t, \varepsilon_{t'}) = 0$, para $t, t' \geq 0$ e $t \neq t'$.

Para obtenção das estimativas dos parâmetros dos três modelos, utilizamos o software R (R Core Team 2018) e o comando *nls* do pacote *nlstools* (Baty et al., 1995). O pacote *nlstools* pode ser obtido no website <http://CRAN.R-project.org/package=nlstools>. Ao ajustar os três modelos aos dados, as observações 1 a 7 foram identificadas como *outliers*. Para identificar os valores *outliers* utilizamos o comando *boxplot.stats* do software R. Removemos estas observação e ajustamos os modelos novamente.

A Figura 7 mostra os gráfico dos valores observados (log-transformados) e as curva ajustada dos três modelos. A Tabela 2 mostra os valores dos critérios de seleção de modelos AIC e BIC e o erro padrão residual dos modelos. Os menores valores estão destacado em negrito. Como podemos notar, o modelo Logístico apresenta os menores valores pelos três critérios. Ou seja, o modelo Logístico é o modelo que melhor explica os dados observados dentre os três modelos considerados.

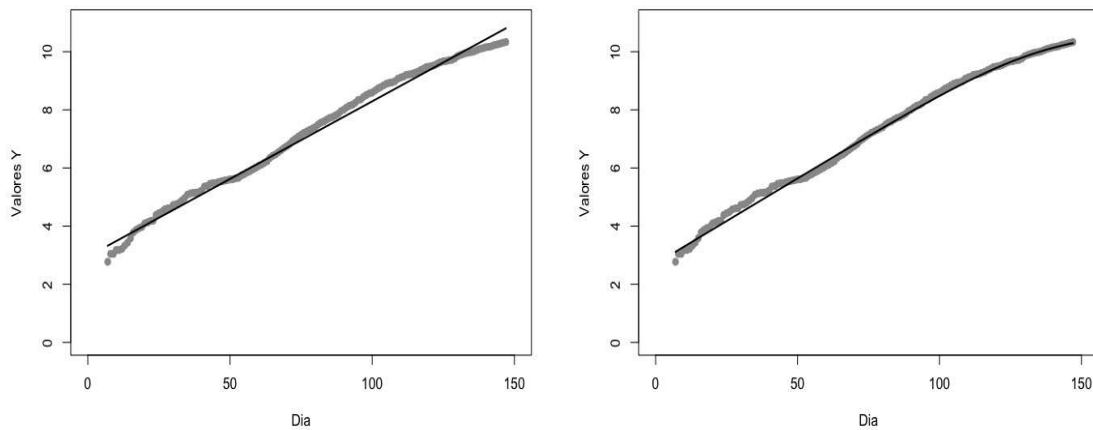
Tabela 2 – Valores AIC, BIC e Erro padrão residual

Modelo	Critério		
	AIC	BIC	Erro padrão
Exponencial	-32,2712	-23,4037	0,2130

Logístico	-183,0166	-171,1933	0,1248
Gompertz	-130,3640	-118,5407	0,1502

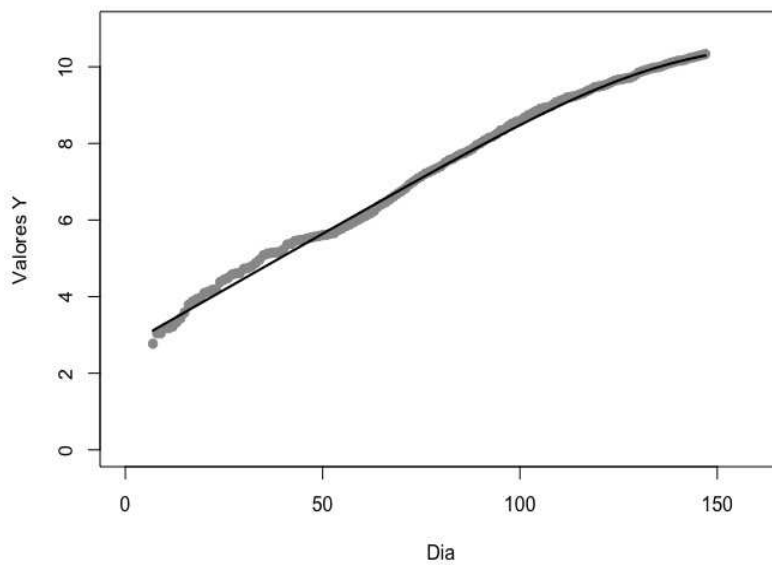
Porém, os valores referentes ao erro-padrão não indicam um ajuste satisfatório. Isto ocorre devido ao aumento significativo das quantidades registradas a partir do dia 59 (12/05/20). Além disso, o modelo Logístico ajustado não satisfaz as suposições de normalidade e homocedasticidade feitas na Equação (7), conforme mostrado na Figura 13 do Apêndice 1.

Figura 7 - Modelos Ajustados



(a) Log-Exponencial

(b) Log-Logístico



(c) Log-Gompertz

Assim, considere a divisão dos dados observados em dois subconjuntos de dados, denominados de D_1 e D_2 . O subconjunto de dados D_1 é composto pelas quantidades totais registradas até o dia 58 (11/05/20); e o subconjunto D_2 é composto pelas quantidades totais do dia 59 (12/05/20) até o dia 149 (09/08/20). No Apêndice 2, apresentamos uma discussão sobre a escolha do dia 59 como ponto de separação de D_1 e D_2 . Para o ajuste dos modelos também descartamos as observação 1 a 7, como feito na modelagem anterior.

Ajustamos os três modelos aos conjuntos de dados D_1 e D_2 . A Tabela 3 mostra os valores dos critérios de seleção de modelo AIC e BIC e o erro padrão residual dos modelos para os dois conjuntos de dados. Os menores valores dos três critérios para cada subconjunto de dados estão destacado em negrito. O modelo Gompertz é o modelo que melhor explica os dois conjuntos de dados.

Tabela 3 – Valores AIC, BIC e Erro padrão residual

Modelo	Dados D_1			Dados D_2		
	AIC	BIC	Erro padrão	AIC	BIC	Erro padrão
Exponencial	-27,4055	-21,5518	0,1790	-38,4820	-30,9825	0,1911
Logístico	-134,5260	-126,7310	0,0633	-288,1247	-278,1255	0,0475
Gompertz	-151,0293	-143,2243	0,0540	-348,9111	-338,9118	0,0339

A Tabela 4 mostra as estimativas e os intervalos de confiança (95%) para os parâmetros dos modelos Gompertz ajustados aos conjuntos de dados D_1 e D_2 . Para obtenção dos intervalos utilizamos o comando *confinte2* do software R. Como mostrado nos gráficos da Figura 14, Apêndice 1, não há razões para duvidarmos que as suposições feitas no modelo da Equação (7) estejam satisfeitas.

Tabela 4 – Estimativas para os parâmetros

Dados	Critério		
	$\log(\alpha_1)$	α_2	α_3
D1	6,8246	4,0271	0,0285
	(6,6305; 7,0188)	(3,8583; 4,1961)	(0,0259; 0,0311)
D2	12,5865	6,6652	0,0125
	(12,4302; 12,7429)	(6,5227; 7,8077)	(0,0120; 0,0129)

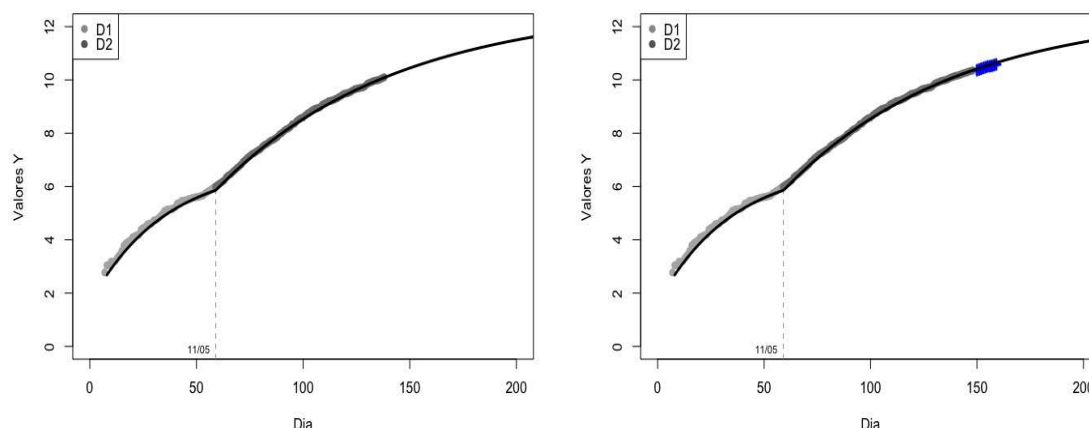
O modelo ajustado é dado pelo seguinte modelo por partes:

$$\hat{Y}_t = \begin{cases} 6,8246 - 4,0271 \exp(-0,0285t) & \text{para } 0 \leq t < 59 \\ 12,5865 - 6,6652 \exp(-0,0125t) & \text{para } t \geq 59 \end{cases} \quad (8)$$

A Figura 8(a) mostra os dados observados (símbolos ●) e o modelo ajustado considerando um período de 200 dias, sendo 149 dias de registros e 61 dias de projeção (até 30 de Setembro). Note que, as projeções futuras são dadas pelo modelo ajustado ao conjunto de dados D₂. O ponto de inflexão é projetado para o dia 204 (04/10/20). Isto mostra que, se as ações tomadas com o objetivo de conter a proliferação da COVID-19 não surtirem efeito, teremos mais 56 dias de aumento de casos até atingirmos o "pico" da pandemia. A estimativa para a máxima quantidade de casos confirmados é 292.582.

A Figura 8(b), mostra o gráfico do modelo ajustado (Figura 8a) e as quantidades registradas nos próximos 10 dias (10/09/20 à 19/06/20), indicados pelos símbolos + na cor azul. A Tabela 5 mostra os valores registrados e os valores projetados pelo modelo para este período de 10 dias. A quarta linha desta Tabela mostra o erro percentual entre o valor registrado e o valor projetado pelo modelo ajustado. Os erros percentuais são todos inferiores a 5%. O maior erro percentual cometido foi de 4,11% e ocorreu no dia 17/08.

Figura 8 - Modelo Gompertz ajustado



(a) Modelo ajustado

(b) Dados de 10/08 a 19/08

Como a pandemia da COVID-19 ainda esta em curso, apresentamos as projeções somente para 10 dias, pois o objetivo é mostrar aos governantes do MS o comportamento da pandemia neste período para basear as tomadas de decisões. Tomadas as decisões com o objetivo de controlar a proliferação da doença; se estas realmente funcionarem, então é esperado que os valores registrados após 09/08 não se comportem conforme as projeções do modelo, mas sim abaixo dos valores projetados.

Tabela 5 – Valores registrados e valores projetados pelo modelo ajustado

Data	10/08	11/08	12/08	13/08	14/08
Registrado	31.739	32.562	33.509	34.559	35434
Projeção	32.425	33.320	34.229	35.150	36.084
Erro %	2,16	2,33	2,15	1,71	1,83
Data	15/08	16/08	17/08	18/08	19/08
Registrado	36.542	36836	37.425	38.393	39.381

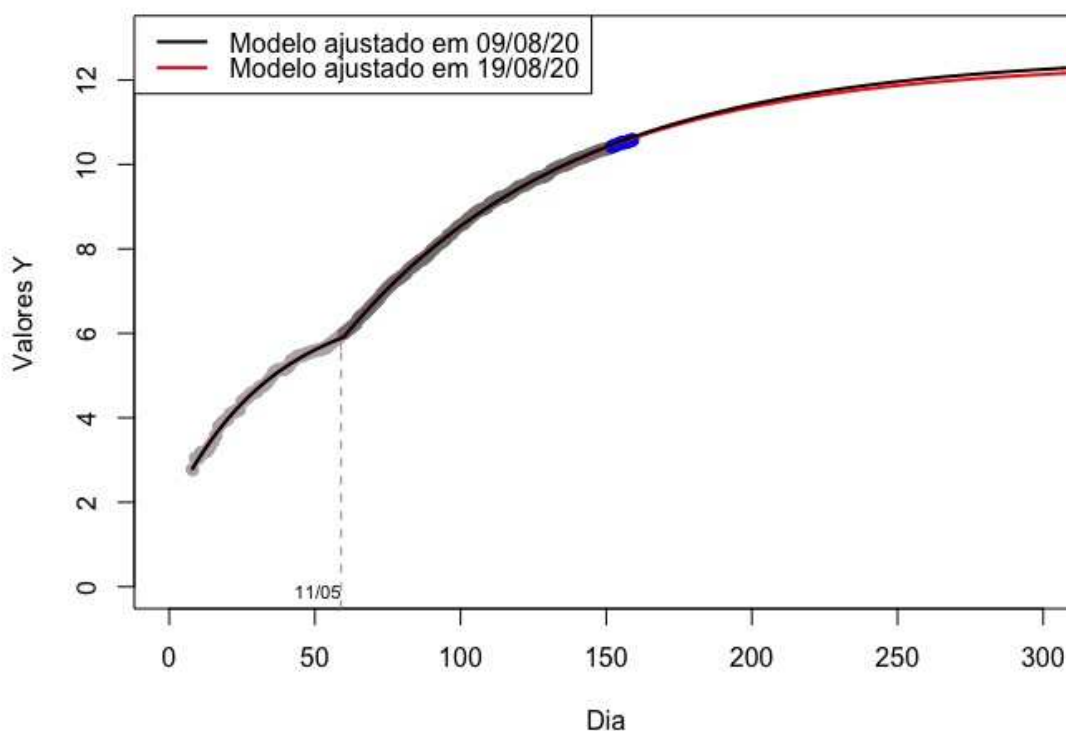
Projeção	37.031	37.990	38.963	39.947	40.944
Erro %	1,34	3,14	4,11	4,05	3,97

Como os valores registrados foram todos abaixo dos valores projetados pelo modelo ajustado em 09/08/20; reajustamos o modelo inserindo os valores registrados nos dias 10/08/20 à 19/08/20. A Figura 9 mostra os dados observados e os modelos ajustados em 09/08/20 e em 19/08/20. Note que, temos um leve "achatoamento" da curva. Neste novo cenário, o pico é estimado para o dia 197 (27/09/20). A estimativa para a máxima quantidade de casos confirmados é de 248.886. Uma redução de 43.696 casos em relação a projeção do modelo ajustado em 09/08/20.

Estes resultados indicam que podemos estar iniciando um período de "achatoamento" da curva. O comportamento das pessoas nas próximas semanas nos mostrará em qual cenário realmente estamos.

Se as medidas de isolamento social continuarem a serem seguidas, podemos manter este cenário de "achatoamento". Porém, com o relaxamento das medidas de isolamento social poderemos ter um aumento das notificações de casos confirmados e voltarmos a ter uma inclinação da curva.

Figura 9 - Modelos ajustados em 09/08 e 19/08



3.1 ESTIMATIVAS DE USO DE LEITOS CLÍNICOS E DE UTI

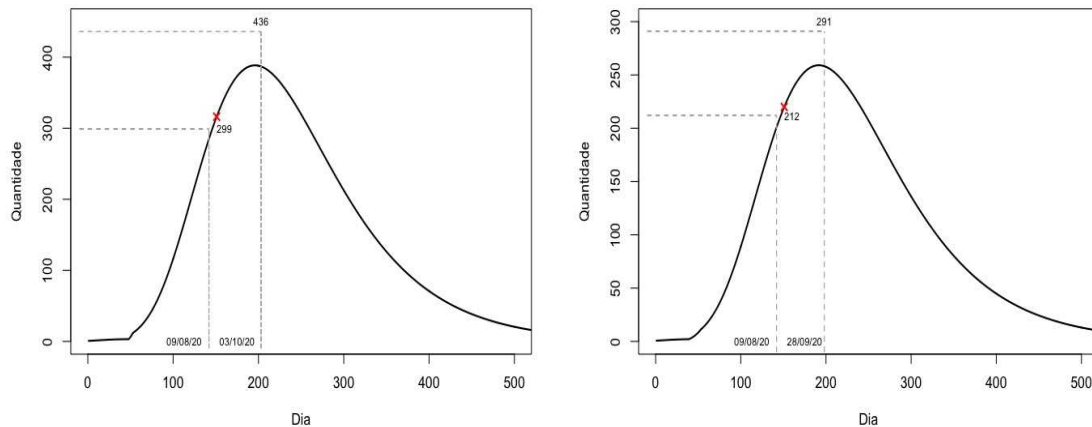
O estado do MS conta com 739 leitos clínicos e 321 leitos de UTI. De acordo com o boletim da secretária estadual de saúde do MS publicado no dia 09/08/20, 514 indivíduos diagnosticados com a COVID-19 estão internados, sendo 298 indivíduos internados em leitos clínicos e 216 em leitos de UTI.

Utilizando as quantidades registradas nos últimos nove dias (01/08/20 à 09/08/20), temos que, aproximadamente 6,5% dos casos confirmados precisam de atendimento em leitos clínicos e 1% precisam de atendimento em leitos de UTI. De acordo com a associação de medicina intensiva brasileira (AMIB) a permanência em leitos de UTI varia entre 14 e 21 dias.

Utilizando o modelo Gompertz ajustado na seção anterior, obtemos as estimativas para as quantidades de indivíduos diagnosticados com a COVID-19 que precisarão de atendimento em leitos clínicos e de UTI. Para isto, consideramos que 6,5% dos pacientes diagnosticados com a COVID-19 precisam de internação em leitos clínicos; e que 1% precisam de atendimento em unidades de terapia intensiva (UTI). Além disso, consideramos um procedimento de somas móveis, em que, a quantidade estimada de leitos clínicos e de UTI ocupados em um dia d qualquer passa a estar disponível após 5 e 14 dias, respectivamente.

A Figura 10, mostra o gráfico das quantidades estimadas de pacientes que precisarão de atendimento em leitos clínicos e de UTI. Para o dia 148 (09/08/20) eram esperados 299 pacientes internados em leitos clínicos e 212 em leitos de UTI. Estes valores estão destacados nos gráficos da Figura 9 pelo símbolo x na cor vermelho. Os valores registrados foram 298 em leitos clínicos e 216 em leitos de UTI.

Figura 10 - Projeções de ocupação de leitos clínicos e de UTI



(a) Leitos clínicos

(b) Leitos de UTI

A projeção para a máxima quantidade de pacientes que precisarão de atendimento em leitos clínicos é de 436 e a previsão é que ocorra no dia 203 (03/10/20). Com relação aos leitos de UTI, a projeção para a máxima quantidade de atendimentos é de 291 e está previsto para ocorrer no dia 198 (28/09/20). Ou seja, as projeções não indicam o colapso do sistema público de saúde do MS.

Neste ponto é importante ressaltar que as datas estimadas levam em consideração apenas a disponibilidade de leitos públicos. Como alguns pacientes receberão tratamento em hospitais privados, a quantidade de pacientes internados nos leitos públicos será menor. Por outro lado, os leitos públicos não são ocupados apenas por pacientes com a COVID-19, mas também por pacientes com outros tipos de doenças e/ou que sofreram algum tipo de acidente. Estes casos podem elevar a ocupação dos leitos públicos. Contudo, o cenário apresentado juntamente com os percentuais de lotação dos hospitais privados e o percentual médio de internações de casos que não sejam da COVID-19 podem auxiliar os especialistas da secretária de saúde do MS na definição das estratégias para melhor atendimento dos pacientes.

As informações descritas acima foram repassadas a representantes da secretária de saúde da prefeitura municipal da cidade de Campo Grande e da secretaria estadual de saúde do MS, nos relatórios dos dias 09/08/20 e 19/08/20.

3.2 SUB-NOTIFICAÇÃO

Nesta seção, apresentamos um estudo de simulação desenvolvido para obter a projeção do número acumulado de casos em um cenário em que há sub-notificação de casos. Consideramos quatro cenários, representando as situações em que há em média 5, 10 e 20 sub-notificações diárias.

Para simular os dados, assumimos que $R_t = N_t + U_t$ é a quantidade acumulada de casos até o dia t , onde N_t representa a quantidade registrada e U_t é a quantidade sub-notificada, para $t \geq 0$. Considere que u_t é um valor de uma variável aleatória U com distribuição Uniforme discreta no conjunto $A = \{0, \dots, 2k\}$, *i.e.*, $U \sim Unif(A)$, para $k = \{5, 10, 20\}$ e $t \geq 0$.

O procedimento de simulação foi desenvolvido através dos seguintes passos:

1. Para $t = 1, \dots, n$:
 - Geramos $U \sim Unif(A)$ e calculamos $R_t = N_t + U_t$;
 - Estimamos os parâmetros dos modelos Gompertz considerando abordagem com os subconjuntos de dados D_1 e D_2 ;
1. Repetimos o passo (1), $L=1.000$ vezes. Seja $\hat{\alpha}^{(l)} = (\hat{\alpha}_1^{(l)}, \hat{\alpha}_2^{(l)}, \hat{\alpha}_3^{(l)})$ as estimativas dos parâmetros do modelo Gompertz para o l -ésimo conjunto de dados gerado;
2. Assuma como estimativa para os parâmetros dos modelos Gompertz, a média das L estimativas, *i.e.*,

$$-\alpha_1 = \sum_{l=1}^L \hat{\alpha}_1^{(l)}, -\alpha_2 = \sum_{l=1}^L \hat{\alpha}_2^{(l)}, -\alpha_3 = \sum_{l=1}^L \hat{\alpha}_3^{(l)},$$

são as estimativas para os parâmetros α_1, α_2 e α_3 , respectivamente.

A Tabela 6 mostra as estimativas para os parâmetros dos modelos Gompertz ajustados aos dados D_1 e D_2 para o cenário com média de k sub-notificações diárias, para $k = \{5, 10, 20\}$.

Tabela 6 – Estimativas para os parâmetros

k	Dados D_1			Dados D_2		
	$\log(\alpha_1)$	α_2	α_3	$\log(\alpha_1)$	α_2	α_3
5	6,4366	4,6055	0,0490	12,1164	6,6004	0,0117
10	6,6043	4,7571	0,0493	12,8555	6,5908	0,0110
20	6,8704	4,9894	0,0501	13,1164	6,5771	0,0100

A Tabela 7 mostra a quantidade acumulada de casos no dia 148 (09/08/20), o valor estimado \hat{N}_{148} dado pelo modelo por partes da Equação 8, e os valores estimados $\hat{N}_{148}^{(k)}$ dado pelos modelos ajustados para cada cenário com média de k sub-notificações diárias. A Terceira linha desta Tabela mostra o número de casos a mais que os valores estimados estão do valor real. Para o pior cenário, média de $k=20$ sub-notificações diárias, é estimado que 2.600 pessoas estão infectadas, mas não serão contabilizadas nos números oficiais da secretária de saúde do MS.

Tabela 7 – Valor registrado e estimativas para o dia 148.

Real	\hat{N}_{148}	$\hat{N}_{148}^{(5)}$	$\hat{N}_{148}^{(10)}$	$\hat{N}_{148}^{(20)}$
31.344	31.544	31.651	32.432	33.944
	+200	+307	+1.088	+2.600

A Figura 11 mostra os valores observados (símbolos ●) e as curva ajustada dos modelos Gompertz. O período considerado é de 292 dias, até 31/12/20. A curva em negrito mostra a curva do modelo descrito na Equação 8. A curva na cor azul, mostra a projeção das quantidades de notificações para o

caso, em que, há em média 3 sub-notificações diárias. As curvas nas cores vermelho e rosa, mostram as projeções das quantidades de notificações para os casos em que há, em média, 10 e 20 sub-notificações diárias. Os valores 208.336, 217.842, 238.414 e 273.487 apresentados no gráfico, são as projeções para a quantidade acumulada de casos confirmados até o dia 292. A Figura 12 mostra o mesmo gráfico para um período de 1.000 dias, destacando as estimativas para as máximas quantidades de notificações.

Figura 11 - Projeções para os casos de sub-notificações em 292 dias

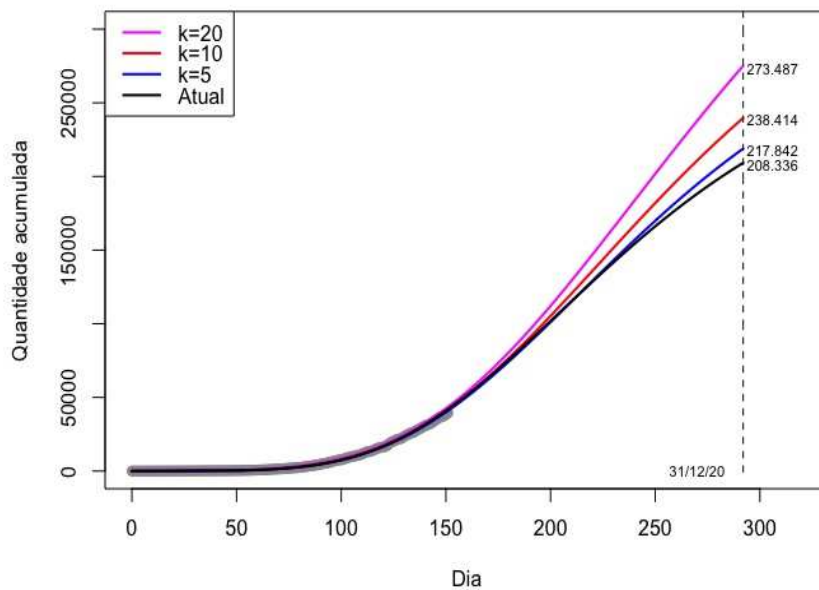
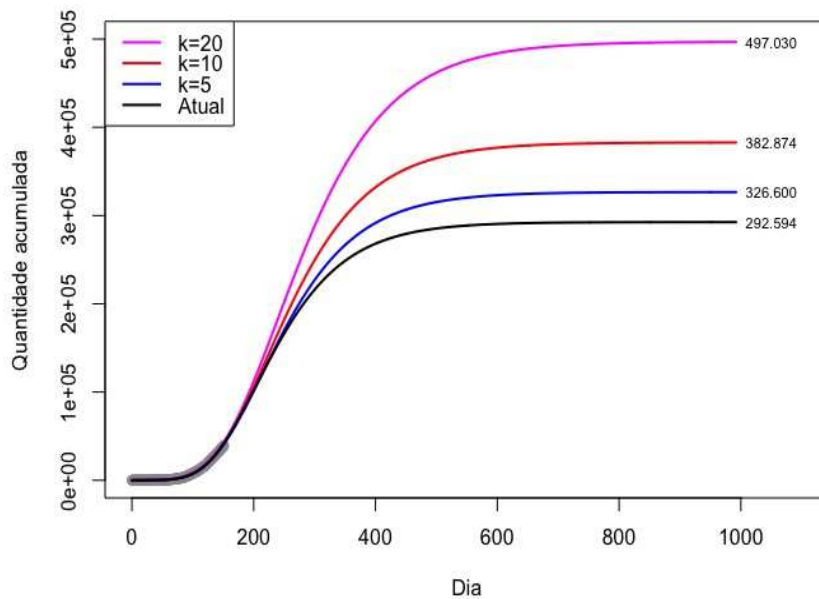


Figura 12
Projeções
casos de

-
para os
sub-

notificações em 1.000 dias



4. CONSIDERAÇÕES FINAIS

Neste artigo, descrevemos um procedimento de modelagem estatística adotado para análise dos dados referentes a quantidade de casos confirmados da COVID-19 no estado do Mato Grosso do Sul.

Esta modelagem iniciou-se a partir de uma solicitação da secretária municipal de saúde da cidade de Campo Grande, MS, para o desenvolvimento de um estudo matemático/estatístico capaz de projetar a quantidade de casos confirmados da COVID-19 e embasar as decisões dos governantes em ações para redução dos impactos que a COVID-19 causaria na cidade.

Os resultados apresentados neste artigo são referentes a análise dos dados registrados no estado do MS no período de 14/03/20 à 14/06/2020. Optamos por apresentar os resultados deste período, pois o modelo ajustado continuou explicando de forma satisfatório os dados registrados nas próximas duas semanas (15/06/20 à 28/06/2020). O último relatório enviado às secretarias de saúde do município e do estado, antes da submissão deste artigo, foi em 28/06/2020.

Além da análise dos dados registrados no MS, também foi desenvolvido a modelagem para a quantidade de casos confirmados da COVID-19 somente na cidade de Campo Grande. O modelo ajustado somente aos dados registrados na Campo Grande não mostram risco de colapso do sistema público de Saúde, *i.e.*, a quantidade estimada de paciente que precisarão de atendimento em leitos clínicos e de UTI não são maiores do que a quantidade de leitos disponíveis em Campo Grande. Contudo, como Campo Grande é a cidade referência do estado do MS, seu sistema de saúde pode entrar em colapso devido a pacientes provenientes das cidades do interior do MS. A partir do conhecimento deste estudo, o ministério público estadual da cidade de Três Lagoas solicitou que a modelagem também fosse feita para a quantidade registradas na cidade de Três Lagoas.

Os principais resultados do estudo, em especial os gráficos dos modelos ajustados, foram utilizados por diversos meios de comunicação para informar e alertar a população sobre a importância de continuar seguindo a

orientação de especialistas da área da saúde para se manter o isolamento social sempre que possível. Pois desta maneira, as quantidades registradas estariam abaixo da curva ajustada e isto levaria ao desejado "achamento" da curva.

Embora o artigo não apresente nenhum resultado matemático e/ou estatístico inovador, os resultados se mostraram muito importantes para os governantes, em especial, para justificar as ações de combate a proliferação da doença. Além disso, devido a necessidade de uma rápida resposta, o uso de modelos consolidados e bem descritos na literatura se mostrou a melhor alternativa para o momento; visto que o desenvolvimento e validação de novos modelos pode necessitar de um período longo de pesquisa.

Todas as análises estatísticas descritas foram implementadas no software R. Todos os relatórios e os códigos em R podem ser obtidos via email aos autores.

5. REFERÊNCIAS BIBLIOGRÁFICAS

Abramowitz, M. and Stegun, I. M. *Handbook of Mathematical Functions*. Dover, New York, 1965.

Ahmadi, A., FADAY, Y., SHIRANI, M. & RAHMANI, F. (2020). Modeling and Forecasting Trend of COVID-19 Epidemic in Iran, 2020. <https://www.medrxiv.org/content/medrxiv/early/2020/03/20/2020.03.17.2003767/1.full.pdf>.

Akaike, H. A. (1974). New look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723.

Baty, F., Ritz, C., Charles, S., Brutsche, M., Flanndrois, J. P. & Delignette-Muller, M. L. (2015). A Toolbox for Nonlinear Regression in R: The Package nlstools. *Journal of Statistical Software*, 66(5).

Blumberg, A. A. (1968). Logistic Growth Rate Functions. *Journal of Theoretical Biology*, 42-44.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52, 345-370.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 182, 513-85.

Pinheiro, J., Bates, D., Debroy, S., Sakar, D. & R Core Team (2020). Linear and Nonlinear Mixed. R package version 3.1-147, 2020. URL <https://CRAN.R-project.org/package=nlme>.

R CORE TEAM. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL: <https://www.R-project.org/>.

Schwarz, G. E. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 461-464.

Tjorve, K. M. C. & Tjorv, E. (2017). The use of Gompertz models in growth analyses, and new Gompertz- model approach: An addition to the Unified-Richards family. *PLoS ONE*, 12(6): e0178691

Winsor, C. P. (1932). The Gompertz curve as a growth curve. *Proceedings of the national academy of sciences*, 18(1).

Wu, Z. & McGoogan, J. M. (2020). Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID- 19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA*,

<https://jamanetwork.com/journals/jama/fullarticle/2762130>.

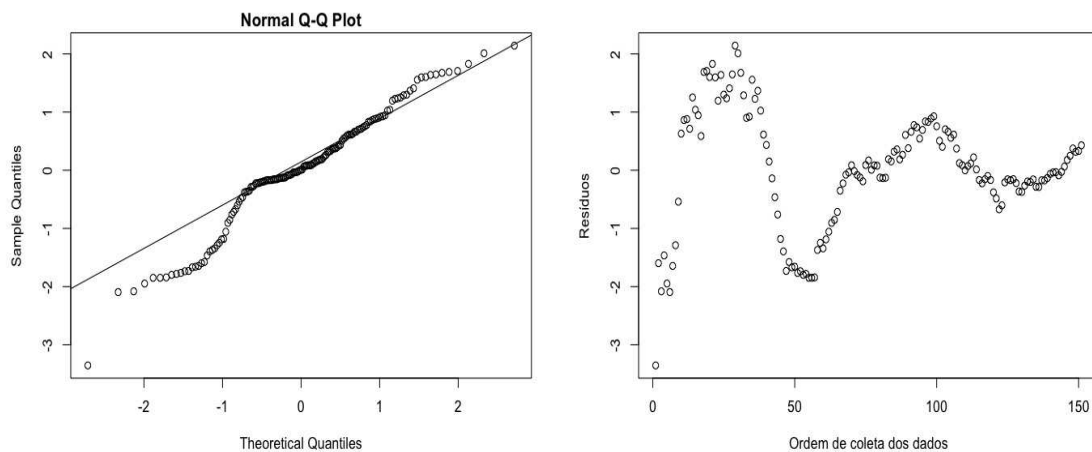
APÊNDICE 1. RESÍDUOS

Neste Apêndice, apresentamos os gráficos referentes a análise dos resíduos dos modelos ajustados. A Figura 13, mostra o gráfico qq-plot e o gráficos dos resíduos *versus* valores ajustados do modelo Logístico ajustado aos dados. Note que, os gráficos indicam que as suposições de normalidade e de homocedasticidade não são satisfeitas. Também, aplicamos o teste de normalidade de Shapiro-Wilks aos resíduos. O p-valor do teste foi de 8.528e-

10. Ou seja, não temos evidências de normalidade dos resíduos. Devido a isto, adotamos o ajuste de um modelo por partes, conforme descrito na seção 3.

A Figura 14, mostra o gráfico qq-plot e o gráficos dos resíduos *versus* valores ajustados do modelo por partes da Equação (8). Como citado na seção 3, os gráficos mostram que não há razões para duvidar que as suposições de normalidade e homocedasticidade feitas no modelo da Equação (7) estejam satisfeitas. Também, aplicamos o teste de normalidade de Shapiro-Wilks aos resíduos. O p-valor do teste de normalidade de Shapiro-Wilks foi de 0,8446. Ou seja, não temos evidências para rejeitar a normalidade dos resíduos

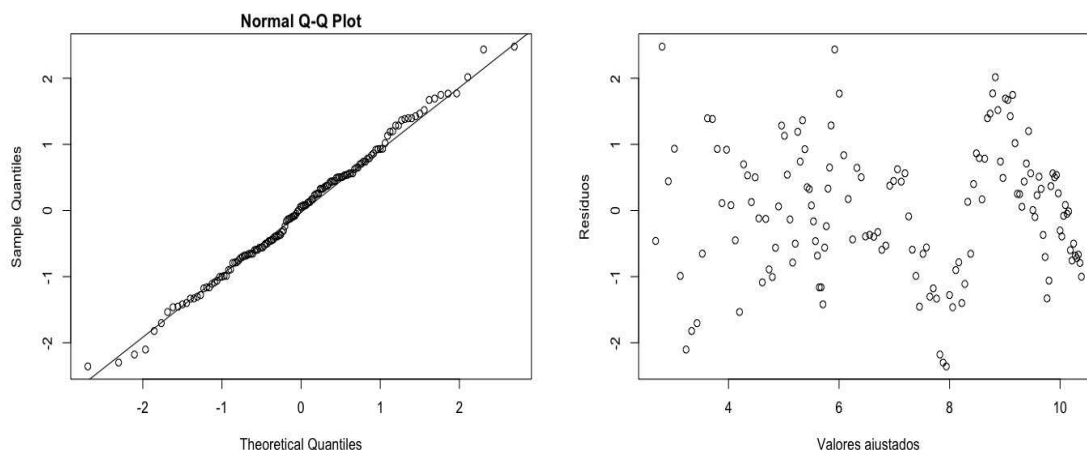
Figura 13 - qq-plot gráfico dos resíduos *versus* valores ajustados



(a) qq-plot

(b) resíduos x valores ajustados

Figura 14 - qq-plot e gráfico dos resíduos *versus* valores ajustados



(a) qq-plot

(b) resíduos x valores ajustados

APÊNDICE 2. ESCOLHA DO DIA 59

Para obter um ponto d que melhor separe os subconjuntos de dados D_1 e D_2 , adotamos a seguinte abordagem. Considere $G=\{50,\dots,70\}$ um gride do valor 50 até o valor 70 com incrementos de tamanho 1. Utilizando G , definimos os seguintes 20 cenários: $D_1=\{8,\dots,d\}$ e $D_2=\{d+1,\dots,158\}$, para d pertencente a G , em que, D_1 inicia no dia 8 devido o descarte dos 7 primeiros dias, conforme descrito na seção 3.

Para cada um dos 20 cenários, ajustamos os três modelos de crescimento a D_1 e D_2 . Selecionamos o melhor modelo utilizando os critérios de seleção de modelos AIC e BIC. Para todos os cenários o modelo por partes têm a seguinte configuração: Gompertz para D_1 e Gompertz para D_2 .

Ajustados os modelos, definimos o “melhor” ponto d como sendo o ponto que o modelo ajustado apresenta menor erro quadrático médio (EQM). Como ilustração, apresentamos na Tabela 8 os valores EQM dos modelos ajustados aos subconjuntos de dados D_1 e D_2 para $d=\{56,\dots,64\}$. O ponto $d=59$ é o pontos com menor EQM. Devidos a isto, optamos pela separação D_1 e D_2 com $d=59$.

Tabela 8 – Erro quadrático médio

d	EQM
56	0,002148
57	0,002034
58	0,001962
59	0,001930
60	0,001940
61	0,001999
62	0,002094
63	0,002231
64	0,002409

AGRADECIMENTOS e COLABORAÇÕES

Os autores agradecem o apoio recebido da UFMS.

MODELANDO ÓBITOS POR COVID-19 EM CENÁRIOS DE HETEROGENEIDADE ESPACIAL: UMA ABORDAGEM UTILIZANDO MODELOS HIERÁRQUICOS BAYESIANOS

Ricardo Junqueira de Souza

rjunqueira@id.uff.br

Universidade Federal Fluminense

Patrícia Viana da Silva

patriciaviana@ufu.br

Universidade Federal de Uberlândia

Jony Arrais Pinto Junior

jarrais@id.uff.br

Universidade Federal Fluminense

Resumo: A evolução da pandemia da COVID-19 no estado do Rio de Janeiro tem ocorrido de forma espacialmente heterogênea uma vez que a mesma atingiu os municípios em diferentes momentos no tempo e cada município desenvolveu suas próprias políticas de combate. Os modelos mais populares na literatura tratam a superfície de risco de forma homogênea, fazendo com que não sejam a melhor escolha para um cenário de alta complexidade como o visto nos dados da COVID-19. Sendo assim, estratégias de modelagem para este cenário devem levar em consideração a possível existência de variações locais na autocorrelação espacial. Para abordar este problema foi considerado o modelo CAR dissimilaridade, este permite a identificação de variações locais na superfície de risco a partir de métricas de dissimilaridade baseadas em covariáveis (Lee & Mitchell, 2012). Mesmo sob um cenário de subnotificação, o modelo permitiu a estimação de uma superfície de risco para o número de óbitos confirmados por COVID-19 e a identificação de descontinuidades.

Palavras-chave: Estatística Espacial; CAR Dissimilaridade, COVID-19

Abstract: The development of COVID-19 in Rio de Janeiro state has been occurring in a spatially heterogenous pattern, as different municipalities confirmed its first case on different times and each of them developed their own policies. The most popular models in the literature treat the risk surface as homogenous, making them not the most suited option for a high complexity scenario as the one seen in COVID-19 data. Considering this, the modelling strategy must consider the possible existence of local variations in the spatial autocorrelation. The CAR dissimilarity model was considered to address this issue. This model identifies local variations in the risk surface via the use of dissimilarity metrics based on covariates (Lee & Mitchell, 2012). Even under an

underreporting scenario, the model allowed the estimation of the risk surface for the number of deaths by COVID-19 and identified boundaries between the municipalities.

Keywords: Spatial Statistics, CAR Dissimilarity, COVID-19

1. INTRODUÇÃO

A COVID-19 é uma síndrome respiratória causada pelo vírus SARS-CoV-2. O vírus foi identificado inicialmente na província de Wuhan, na China, no final do ano de 2019 e desde então se espalhou rapidamente pelo mundo (Zi et al., 2020). A maior parte dos indivíduos infectados pelo vírus são assintomáticos, porém há indivíduos que apresentam sintomas similares aos da gripe e também aqueles que desenvolvem um quadro de pneumonia severa, que pode levar a morte (Guan et al., 2020; Huang et al., 2020; Zhou et al., 2020). Dados da Organização Mundial da Saúde (OMS) indicam que cerca de 80% dos indivíduos infectados podem ser assintomáticos e que 20% dos casos podem requerer atendimento hospitalar (Ministério da Saúde, 2020). Estas informações encontram-se disponíveis no portal do Ministério da Saúde sobre a COVID-19, ele pode ser acessado em: <https://coronavirus.saude.gov.br/sobre-a-doenca>.

No Brasil, o primeiro caso confirmado da COVID-19 foi notificado em 26 de fevereiro de 2020 no estado de São Paulo e o primeiro óbito em 17 de março de 2020 no mesmo estado. No dia 20 de março de 2020 foi confirmada a transmissão comunitária e, até o fim de junho de 2020, o país já possuía 1.294.725 casos e 54.670 mortes confirmadas (Ministério da Saúde, 2020). Tendo como principais focos da epidemia os estados de São Paulo, Rio de Janeiro, Ceará, Pará e Amazonas (Ministério da Saúde, 2020). Os dados agregados por estado sobre a COVID-19 estão disponíveis em: <https://covid.saude.gov.br/>.

A análise se concentrará nos municípios que compõem o estado do Rio de Janeiro. O primeiro caso da COVID-19 confirmado na região foi registrado no município de Barra Mansa em 05 de março de 2020 e o primeiro óbito em 19 de março no município de Miguel Pereira (Secretaria de Saúde do Rio de Janeiro, 2020). Até 30 de junho de 2020 os municípios do Rio de Janeiro já haviam confirmado 115.278 casos e 10.198 óbitos pela COVID-19, sendo a maior parte destes concentrados nos municípios da região metropolitana do estado (Ministério da Saúde, 2020). O número de casos e de óbitos confirmados por COVID-19 nos municípios do estado do Rio de Janeiro estão disponíveis em: <http://painel.saude.rj.gov.br/monitoramento/covid19.html#>.

Áreas como a epidemiologia e a saúde pública tem por objetivo ao longo dos anos mapear a incidência e a prevalência de doenças em populações humanas (Koch, 2005). Muitas vezes, o objetivo é quantificar o padrão espacial do risco de doenças em uma região geográfica com base na contagem de casos observados em sub-regiões do espaço de interesse, utilizando algumas informações potencialmente relevantes, como por exemplo, o número de indivíduos em risco e covariáveis.

Na literatura estatística, o termo "mapeamento de doenças" (do inglês, *disease mapping*) refere-se a uma coleção de métodos estatísticos que estimam efeitos locais utilizando características do cenário espacial e assumindo correlação espacial positiva entre as observações (Waller & Carlin, 2010). De um modo geral, estes métodos se caracterizam por "pegar emprestado" informações dos vizinhos, de modo que são utilizadas mais informações das sub-regiões vizinhas do que das sub-regiões distantes, e também por suavizar as taxas locais.

O termo "mapeamento de doenças" é oriundo de Clayton & Kaldor (1987), que definiram métodos bayesianos empíricos baseados na regressão de Poisson com interceptos aleatórios com uma estrutura espacial. Esta abordagem de modelos hierárquicos tornou-se muito popular sob a ótica Bayesiana, pois fornece uma estrutura conveniente, na qual se induz correlação espacial (positiva) entre as taxas estimadas de doenças por meio de uma atribuição de distribuições *a priori* condicionais autoregressivas para os efeitos aleatórios.

Dentre os modelos que utilizam distribuições condicionais autoregressivas, pode-se citar o CAR Intrínseco (Besag, 1974), o Besag-York-Mollié - BYM (Besag et al., 1991) e o Leroux (Leroux et al., 2000). Entretanto, todos estes são modelos de suavização global, ou seja, eles induzem efeitos aleatórios espaciais com uma suavização de toda a área geográfica e não levam em consideração variações locais.

No entanto, em cenários complexos, como o de interesse neste trabalho, é possível que o risco da doença exiba uma estrutura mais localizada, incluindo sub-regiões que variam suavemente ao longo do espaço, bem como sub-regiões com mudanças abruptas do seu valor. Tais mudanças nos riscos são conhecidas na literatura como limites ou descontinuidades e podem ocorrer por diversos motivos, incluindo a existência de sub-regiões ricas e pobres como vizinhas.

Cada município do estado do Rio de Janeiro apresentou seu primeiro caso confirmado da doença em um momento diferente, fazendo com que estejam em diferentes estágios no desenvolvimento da epidemia. Além disso, os municípios adotaram diferentes diretrizes de combate à pandemia e possuem

contextos econômicos e sociais muito distintos. Devido a essa heterogeneidade, estratégias de modelagem que consideram a existência de autocorrelação espacial local tornam-se atrativas para abordar este cenário.

Diversos autores já propuseram modelos espaciais capazes de identificar *clusters* de unidades de área que apresentam comportamento similar para a variável resposta, isto é, identificar superfícies de descontinuidade. Um dos primeiros trabalhos existentes foi proposto por Lu & Carlin (2005), que classificava duas sub-regiões como descontínuas se as mesmas possuíam diferença entre os riscos estimados acima de uma constante c_1 ou ainda se pertencia aos $c_2\%$ maiores valores observados das diferenças obtidas entre todos os pares de sub-regiões. Tal abordagem sofreu diversas críticas por atribuir a escolha de c_1 e c_2 ao pesquisador. Mais recentemente, Lee & Mitchell (2012) apresentaram o CAR dissimilaridade, uma extensão do modelo Leroux que é capaz de identificar descontinuidades na superfície de risco. Posteriormente, Adin et al. (2019) propuseram um modelo em dois estágios capaz de estimar a superfície de risco e identificar descontinuidades e *clusters* locais tanto no tempo quanto no espaço.

Para parte dos modelos disponíveis, a identificação dos *clusters* é feita por meio de dados históricos da doença em análise, o que impossibilita sua aplicação aos dados da COVID-19 (Anderson et al., 2014). O CAR dissimilaridade possibilita a identificação das fronteiras na superfície de risco por meio de medidas de dissimilaridade calculadas para covariáveis.

Como a COVID-19 é uma doença nova e pouco se sabe sobre a mesma em todo o mundo, entender como funciona a dinâmica da superfície de risco da sua mortalidade é de suma importância. Apesar das subnotificações de casos, que ocorrem nos municípios brasileiros uma vez que o Brasil possui uma das mais baixas taxas de testes por 1.000 habitantes do mundo, o objetivo deste trabalho é estimar as superfícies de riscos de mortalidade considerando uma abordagem global e uma local (Roser et al., 2020). Para a abordagem global será utilizado o modelo Leroux e para a local será utilizado o modelo CAR Dissimilaridade. A ideia é comparar as superfícies de riscos estimadas pelos dois métodos e apresentar as particularidades de cada análises. Deste modo, pretende-se mostrar o impacto das covariáveis no risco médio de cada região por meio do modelo global e mostrar se existem e onde se encontram as possíveis descontinuidades na superfície de risco estimada, por meio do modelo local. Vale ressaltar que, a identificação destas fronteiras permite indicar quais municípios, que apesar de vizinhos, apresentam grandes diferenças em seus

riscos para a doença, possibilitando o ajuste nas políticas de enfrentamento à pandemia.

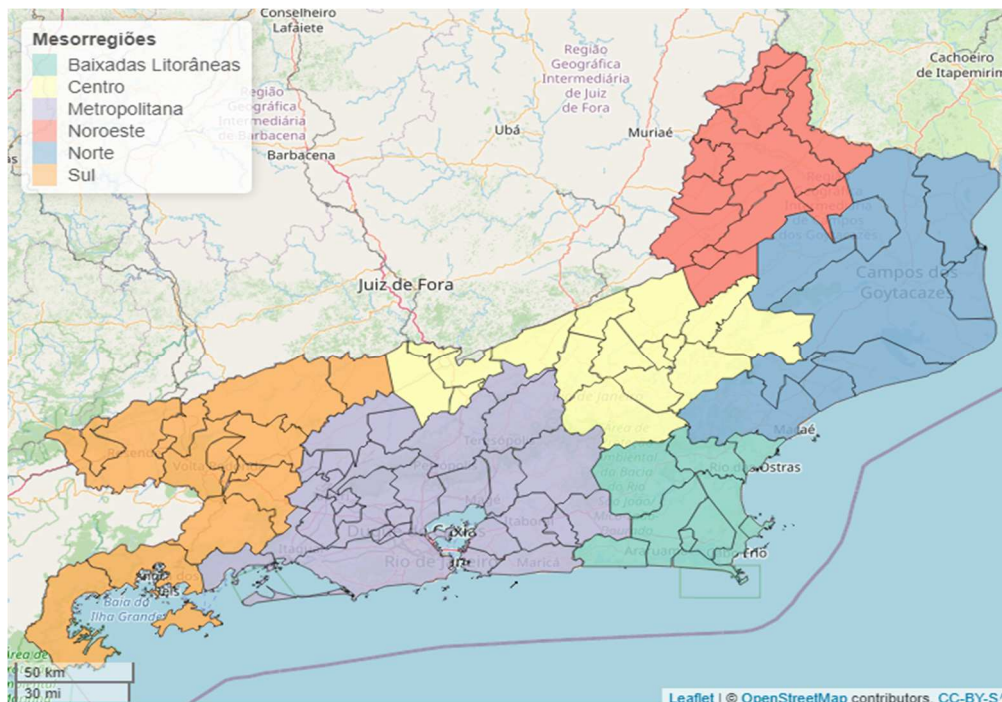
O artigo é composto por mais 3 seções, além desta introdução. Na Seção 2, são apresentados e discutidos a região de estudo, os dados que serão utilizados e os modelos Leroux e CAR dissimilaridade. A Seção 3 apresenta os resultados encontrados pelo ajuste dos modelos para a mortalidade da COVID-19 no estado do Rio de Janeiro. Por fim, na Seção 4 algumas conclusões dos resultados encontrados são apresentadas.

2.MATERIAIS E MÉTODOS

2.1 CONTEXTUALIZAÇÃO DA REGIÃO DE ESTUDO

O estado do Rio de Janeiro é composto por 92 municípios. Estes são divididos em 6 mesorregiões: Norte, Noroeste, Centro, Baixadas Litorâneas, Sul e Metropolitana, como apresentado na Figura 1. Até 30 de junho de 2020 os municípios do Rio de Janeiro já confirmaram 115.278 casos e 10.198 óbitos por COVID-19, sendo a maior parte destes concentrados nos municípios da região metropolitana do estado (Ministério da Saúde, 2020).

Figura 1 – Divisão do Rio de Janeiro em mesorregiões.



Fonte – O autor

As estimativas populacionais dos municípios do Rio de Janeiro no ano de 2019 do Instituto Brasileiro de Geografia e Estatística - IBGE indicam que naquele ano a população do estado era de 17.286.074 habitantes, com 77,83% destes morando nos municípios da região metropolitana (IBGE, 2019). A segunda mesorregião mais populosa é a Sul, que contém 6,71% dos habitantes

do estado. As estimativas populacionais utilizadas estão disponíveis em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?edicao=25272&t=resultados> .

Durante o período abordado neste trabalho o epicentro da epidemia esteve na região metropolitana. Entre as demais mesorregiões, os municípios de Angra dos Reis e Volta Redonda, na região Sul e Macaé e Campos dos Goytacazes, na região Norte, apresentaram um número de casos confirmados acima da média de suas respectivas mesorregiões.

2.2 DADOS

Os totais acumulados até o dia 30 de junho de 2020 de óbitos confirmados por COVID-19 nos 92 municípios do estado foram obtidos na plataforma do Ministério da Saúde, eles estão disponíveis no site: <https://susanalitico.saude.gov.br/extensions/covid-19.html/covid-19.html.html> . Esse ponto de corte faz com que os dados compreendam aproximadamente os três primeiros meses da pandemia no estado.

Para calcular o número esperado de óbitos nos municípios foi necessário utilizar dados do tamanho estimado da população dos municípios do Rio de Janeiro fornecidas pelo IBGE. Como as estimativas não são fornecidas por sexo, foram calculados os percentuais de cada sexo de acordo com o último Censo disponível e então calculado os tamanhos da população de cada sexo considerando a estimativa fornecida para o ano de 2019.

A escolha das covariáveis para a identificação de descontinuidades justificou-se com base em resultados apresentados por outros autores. Abuelgasim et al. (2020) mostraram que populações minoritárias no Reino Unido estavam mais vulneráveis ao vírus devido às condições socioeconômicas relacionadas, Gold et al. (2020) descreveram uma amostra colhida em um hospital nos Estados Unidos em que houve super-representação de indivíduos negros e Laurencin & McClinton (2020) apresentaram evidências de que a população negra apresentou maiores taxas de incidência e óbito por COVID-19 quando comparadas às demais populações étnicas dos Estados Unidos.

Desta forma, foram consideradas nas análises realizadas as covariáveis Índice de Desenvolvimento Humano Municipal (IDHM), número de pessoas não-brancas por município (NB), proporções de idosos (ID) (IPEA, 2020a), Índice de Vulnerabilidade Social (IVS) (IPEA, 2020b), número de pessoas cobertas pela rede de Atenção Básica (AB), número de pessoas registradas no Cadastro Único (CAD) (Instituto Votorantim, 2020), número de respiradores (NR) e número de leitos (LE) do Sistema Único de Saúde por 100.000 habitantes (DATASUS,

2020). Estas variáveis foram escolhidas por representarem a vulnerabilidade, as condições socioeconômicas e a capacidade de tratamento dos municípios.

Nas análises realizadas, O IDHM e o IVS são indicadores que variam de 0 a 1. No caso do primeiro, valores mais elevados indicam um maior nível de desenvolvimento do município e, no caso do IVS, valores mais próximos de 1 indicam maior vulnerabilidade.

2.3 ANÁLISE EXPLORATÓRIA PARA DADOS ESPACIAIS

A análise exploratória dos dados consistiu na obtenção de medidas de tendência central para a variável resposta número total de óbitos confirmados pela COVID-19 e também para as covariáveis listadas na seção anterior.

Em seguida foi realizada uma análise exploratória para dados espaciais. Por se tratar de dados de área, o primeiro passo consistiu na construção de mapas coropléticos para verificar visualmente a existência de um padrão espacial global para o número total de óbitos confirmados pela COVID-19. Esta visualização permite obter uma intuição inicial sobre a existência de possíveis *clusters* de municípios com comportamentos similares.

Para a quantificação de uma possível correlação espacial global e local, bem como para a etapa de modelagem, foi necessário estabelecer uma estrutura de vizinhança entre os municípios do Rio de Janeiro por meio de uma matriz de vizinhança W . Esta estrutura de vizinhança pode ser determinada de diversas formas, incluindo distância ou contiguidade (Bivand et al., 2013). Sob o critério de contiguidade, duas sub-regiões são consideradas vizinhas se suas fronteiras se tocam, a matriz de vizinhança de dimensão $N \times N$ possui elementos definidos da seguinte forma:

$$w_{ij} = \begin{cases} 1, & \text{se } i \sim j \\ 0, & \text{c. c.} \end{cases}, \quad (1)$$

em que $w_{ij} = 1$ se as unidades de área i e j são vizinhas, $w_{ij} = 0$ caso contrário e N é o número de sub-regiões. Apenas foi necessário adicionar uma relação de vizinha entre os municípios de Niterói e Rio de Janeiro já que sua ligação é feita por meio de uma ponte e possuem um fluxo significativo de pessoas diariamente.

Proposto por Anselin (1995), o Índice de Moran Global foi calculado para quantificar a autocorrelação espacial global. Este índice é uma medida que quantifica a força da dependência espacial global de uma variável e é definido como:

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

em que N é o número de sub-regiões na região de estudo, y_i e \bar{y} representam o valor observado e a média da variável de interesse, respectivamente e w_{ij} é o elemento da matriz de vizinhança W correspondentes a linha i e coluna j .

A versão local do Índice de Moran permite detectar *clusters* de sub-regiões espacialmente dependentes em relação a uma variável de interesse, caracterizando a relação de uma sub-região com seus vizinhos (Anselin, 1995). O Índice de Moran Local é definido da seguinte forma:

$$I_i = \frac{y_i - \bar{y}}{S_i^2} \sum_{j=1, j \neq i}^N w_{ij} (y_j - \bar{y}), \text{ com } S_i^2 = \frac{\sum_{j=1, j \neq i}^N (y_j - \bar{y})^2}{N-1}, i = 1, \dots, N, \quad (3)$$

em que y_i e y_j são os valores da variável de interesse y nas sub-regiões i e j , N é o número de municípios na região de estudo, \bar{y} é a média da variável, w_{ij} é o elemento da matriz de vizinhança W correspondentes a linha i e coluna j .

2.4 MODELAGEM

Os modelos hierárquicos utilizados neste trabalho são definidos com base nos $N = 92$ municípios que compõem o estado do Rio de Janeiro. Para as especificações dos modelos, Y_i é o número total de óbitos confirmados pela COVID-19 até o dia 30 de junho de 2020 no i -ésimo município, $x_i = (1, x_{1i}, \dots, x_{pi})^T$ é o vetor com p variáveis explicativas e E_i é o número esperado de óbitos no município i , $i = 1, \dots, N$.

O cálculo de E_i , usualmente, leva em consideração os números de óbitos e o tamanho da população para alguns estratos da população de interesse que são definidos por características como sexo, idade, entre outras. Neste trabalho, adotou-se o cálculo do número esperado considerando somente o sexo, pois a variável apresentou aproximadamente 0,02% de valores faltantes enquanto para a variável idade esta proporção foi de 78,46%. Por este valor ser substancialmente baixo, os indivíduos com valores faltantes para a variável sexo foram descartados.

Seguindo López-Abente et al. (2014), o número de óbitos esperado no município i foi obtido da seguinte forma:

$$E_i = \frac{\sum_{i=1}^N pm_i Ym_i + \sum_{i=1}^N pf_i Yf_i}{\sum_{i=1}^N pt_i}, \quad (4)$$

em que pm_i e pf_i são o tamanho das populações masculinas e femininas, respectivamente, Ym_i e Yf_i são o número de óbitos confirmados para homens e mulheres, respectivamente, e pt_i é o total populacional do i -ésimo município.

A seguir serão definidos dois modelos para estimar a superfície de risco da mortalidade da COVID-19 no estado do Rio de Janeiro. O primeiro, o modelo

Leroux, é um modelo que considera uma suavização espacial global dos riscos e o segundo, o modelo CAR Dissimilaridade, é um modelo que permite identificar possíveis descontinuidades na superfície de risco.

Modelo Leroux

O modelo Leroux é definido da seguinte maneira:

$$Y_i | \Lambda_i \sim \text{Poisson}(\Lambda_i), \quad i = 1, \dots, N, \quad (5)$$

$$\Lambda_i = E_i R_i, \quad (6)$$

$$\log(R_i) = \mathbf{x}_i^T \boldsymbol{\beta} + \phi_i, \quad (7)$$

$$\beta_k \sim N(0, 100.000), \quad k = 1, \dots, p + 1, \quad (8)$$

$$\phi_i | \boldsymbol{\phi}_{-i}, W, \tau, \rho \sim N\left(\frac{\rho \sum_{j=1}^N w_{ij} \phi_j}{\rho \sum_{j=1}^N w_{ij} + 1 - \rho}, \frac{1}{\tau \rho \sum_{j=1}^N w_{ij} + 1 - \rho}\right), \quad (9)$$

$$\tau \sim \text{Gama}(1; 0,01), \quad (10)$$

$$\rho \sim \text{Uniforme}(0; 1), \quad (11)$$

em que $\boldsymbol{\beta}$ é o vetor com $p+1$ parâmetros de regressão associados ao vetor das variáveis explicativas \mathbf{x}_i e ϕ_i é o efeito espacial no município i . O número esperado de óbitos E_i é utilizado com o *offset* do modelo, ou seja, é utilizado para corrigir o efeito do tamanho da população nas contagens observadas para cada município.

Note que, no primeiro nível a variável resposta possui distribuição Poisson com média Λ_i , no segundo nível é considerado que a média se decompõe de forma multiplicativa, em que E_i é um componente determinístico e R_i é um componente estocástico, o risco. No terceiro nível, é proposto um modelo log-linear para R_i que considera variáveis explicativas e um efeito aleatório espacial. Foram atribuídas distribuições *a priori* pouco informativas para $\boldsymbol{\beta}$, ρ e τ .

A distribuição condicional autorregressiva proposta por Leroux et al. (2000) para os efeitos espaciais (9) depende dos hiperparâmetros ρ e τ , o primeiro representa a força da autocorrelação espacial global e o segundo, a precisão. Banerjee et al. (2015) mostraram que a matriz de precisão dos efeitos espaciais (Q) no modelo Leroux pode ser escrita da seguinte forma:

$$Q = \tau(\rho(D - W) + (1 - \rho)I), \quad (12)$$

em que D é uma matriz diagonal quadrada de dimensão igual ao número de unidades de área e com elementos $\sum_{i=1}^N w_{ij}$, ou seja, o número de vizinhos da unidade de área i , W é a matriz de vizinhança e I é uma matriz identidade de ordem igual ao número de unidades de área. Quando $\rho = 0$, pode-se perceber

facilmente que o modelo trata os efeitos espaciais como independentes e, quando $\rho = 1$, o modelo é simplificado para o CAR Intrínseco, um modelo que apresenta bom desempenho apenas em cenários de autocorrelação espacial global forte.

Modelo CAR Dissimilaridade

A especificação do modelo CAR Dissimilaridade, proposto por Lee & Mitchell (2012), é realizada em duas etapas. Na primeira delas, é adotado um modelo Leroux sem a presença de covariáveis, considerando somente o intercepto e definindo $\rho = 0,99$. Além disso, é assumido que os componentes da matriz de vizinhança W dependem de um vetor de hiperparâmetros $\alpha = (\alpha_1, \dots, \alpha_p)^T$, que serão responsáveis por determinar a autocorrelação local. Na segunda, são especificados os elementos da matriz de vizinhança W , de modo que os mesmos dependam das covariáveis.

O modelo é definido da seguinte forma:

$$Y_i | \Lambda_i \sim \text{Poisson}(\Lambda_i), \quad i = 1, \dots, N, \quad (13)$$

$$\Lambda_i = E_i R_i, \quad (14)$$

$$\log(R_i) = \beta_0 + \phi_i, \quad (15)$$

$$\beta_0 \sim N(0, 100.000) \quad (16)$$

$$\phi_i | \phi_{-i}, W, \tau \sim N\left(\frac{0,99 \sum_{j=1}^N w_{ij}(\alpha) \phi_j}{0,99 \sum_{j=1}^N w_{ij}(\alpha) + 0,01}, \frac{1}{\tau 0,99 \sum_{j=1}^N w_{ij}(\alpha) + 0,01}\right), \quad (17)$$

$$\tau \sim \text{Gama}(1; 0,01), \quad (18)$$

em que β_0 é o intercepto do modelo, τ é a precisão e $w_{ij}(\alpha)$ são os elementos da matriz de vizinhança transformada pela presença de covariáveis. Assim como no modelo Leroux, o número esperado de óbitos E_i foi utilizado como *offset* do modelo.

É natural esperar nestas abordagens utilizadas que sub-regiões mais próximas apresentem um comportamento de risco mais similar do que sub-regiões mais distantes entre si. Entretanto, fatores físicos, socioeconômicos e populacionais podem fazer com que sub-regiões vizinhas apresentem comportamentos extremamente distintos para o risco, o que representaria um limite ou descontinuidade.

A presença ou não destes limites é modelada na proposta do modelo CAR dissimilaridade por meio de medidas não-negativas de dissimilaridade para cada uma das p covariáveis $\mathbf{z}_{ijk} = (z_{ij1}, \dots, z_{ijp})^T$, em que $z_{ijk} = |x_{ik} - x_{jk}| / \sigma_k$, $k = 1, \dots, p$. Note que, z_{ijk} representa a k -ésima medida de dissimilaridade e é definida em função do módulo da diferença da k -ésima covariável do vetor x entre as sub-regiões i e j , que é então reescalada por σ_k , que representa o

desvio padrão de $|x_{ik} - x_{jk}|$ para os pares de sub-regiões vizinhas. Os autores indicam a definição de z_{ijk} desta forma para melhorar a mixagem e a convergência das cadeias no algoritmo MCMC. Usando essas medidas, os elementos de W são modelados da seguinte forma:

$$w_{ij}(\alpha) = \begin{cases} 1, & \text{se } \exp(-\sum_{k=1}^p z_{ijk} \alpha_k) \geq 0,5 \text{ e } i \sim j \\ 0, & \text{c. c.} \end{cases} \quad (19)$$

$$\alpha_k \sim \text{Uniforme}(0, M_k) \text{ para } k = 1, \dots, p, \quad (20)$$

em que M_k é o limite superior da distribuição *a priori* para α e é definido com base na distribuição das métricas de dissimilaridade z_{ijk} . A especificação sob esta forma faz com que a distribuição seja pouco informativa (Lee, 2013).

O vetor de parâmetros α é não-negativo, indicando que quanto maior for a medida de dissimilaridade entre duas sub-regiões vizinhas, é mais provável que exista um limite entre estas sub-regiões. Deste modo, quando se dispõe de p medidas de dissimilaridade, uma descontinuidade entre duas sub-regiões i e j é identificada se $\prod_{k=1}^p \exp(-z_{ijk} \alpha_k) < 0,5$.

A contribuição da k -ésima covariável na identificação de descontinuidades é avaliada pela comparação entre o intervalo de credibilidade de 95% para α_k , que representa o efeito da k -ésima covariável na identificação de descontinuidades, e a quantidade α_k^{min} que é o limiar de identificação de fronteiras para a k -ésima covariável. Lee e Mitchell (2012) mostram que a partir da Equação (19) obtém-se $\alpha_k^{min} = -\ln(0,5)/z_k^{max}$, em que z_k^{max} denota o maior valor de métrica de dissimilaridade observada. Considera-se que uma covariável não teve efeito significativo para a identificação quando o intervalo de credibilidade de 95% para α_k está completamente abaixo de α_k^{min} . Caso o intervalo esteja completamente acima de α_k^{min} , considera-se que a covariável apresentou uma contribuição significativa. Quando α_k^{min} está contido no intervalo de credibilidade, considera-se que a covariável apresentou contribuição não-nula para a identificação.

Para realizar o procedimento de inferência nos parâmetros e hiperparâmetros dos modelos apresentados, precisou-se recorrer a métodos de aproximação como os algoritmos MCMC (Gelman & Lopes, 2006) e o INLA (Rue et al., 2009). Para o modelo CAR Dissimilaridade, a configuração do MCMC para o ajuste do modelo continha cadeias de tamanho 500.000, com período de aquecimento de 100.000 e espaçamento de tamanho 200, resultando em uma amostra final de tamanho 2.000. A autocorrelação das cadeias dos parâmetros e hiperparâmetros e suas convergências foram avaliadas de forma visual. O ajuste do modelo Leroux foi feito utilizando o INLA, que estima as distribuições *a posteriori* por meio da aproximação de Laplace.

3.RESULTADOS

Inicialmente foram calculadas medidas descritivas para o número de óbitos confirmados pela COVID-19 sumarizadas na Tabela 1. Percebe-se que até o dia 30 de junho de 2020, ainda existem municípios que não apresentavam casos de óbitos confirmados e o número médio de óbitos confirmados por município era de aproximadamente 110 óbitos. O município do Rio de Janeiro possui o maior registro de óbitos, totalizando 6.550 óbitos acumulados. Vale ressaltar que este valor corresponde a 64,98% dos óbitos confirmados no estado.

Tabela 1 – Medidas descritivas dos óbitos confirmados por COVID-19 até 30 de junho de 2020

Variável	Média	Desvio Padrão	Mínimo	Máximo
Óbitos	109,565	684,604	0	6.550

Fonte – O autor

A Tabela 2 apresenta medidas resumos para as covariáveis Índice de Desenvolvimento Humano Municipal (IDHM), o Índice de Vulnerabilidade Social (IVS), a proporção de idosos (ID), número de pessoas não-brancas por município (NB), número de respiradores (NR), número de pessoas registradas no Cadastro Único (CAD), número de pessoas cobertas pela rede de Atenção Básica (AB) e a taxa de leitos do Sistema Único de Saúde por 100 mil habitantes (LE). Percebe-se que os municípios possuem um IDHM médio de 0,708, que o município com o maior percentual de idosos possui 21,71% da sua população idosa e que existem municípios que não possuem leitos do Sistema Único de Saúde.

Tabela 2 – Medidas descritivas das covariáveis utilizadas nos modelos

Variável	Média	Desvio Padrão	Mínimo	Máximo
IDHM	0,708	0,037	0,611	0,837
IVS	0,276	0,055	0,191	0,437
ID	15,683%	2,858%	9,032%	21,711%
NB	91.324,8	333.165,6	2.786	3.085.634
NR	89,478	166,900	0	861
CAD	47.593,8	140.929,8	3.084,247	1.279.951

AB	129.250,6	401.513,2	4.157,258	3.675.240
LE	176,921	153,610	0	1.171,171

Fonte – O autor

A autocorrelação espacial global presente no número de óbitos confirmados por COVID-19 foi estimada por meio do Índice de Moran Global, conforme apresentado na Tabela 3. Ainda que a intensidade seja baixa, há evidências da existência de dependência espacial significativa. Deve-se levar em consideração que a pandemia ainda está em desenvolvimento e que os municípios da região Metropolitana foram afetados primeiro que os municípios do interior. Assim, é possível que a baixa autocorrelação global estimada seja um reflexo da concentração dos casos em uma única região do estado.

Tabela 3 – Resultados do Índice de Moran Global

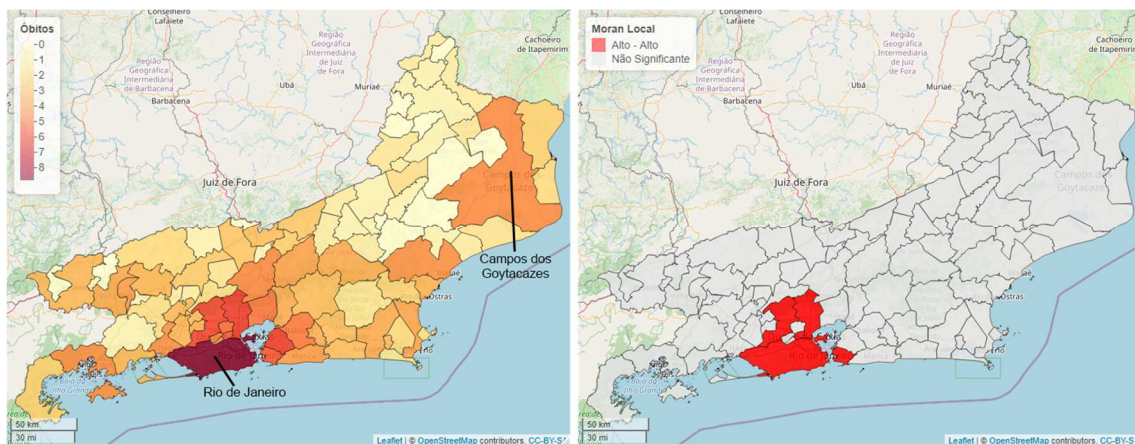
Variável	Índice de Moran	P-valor
Óbitos	0,054	<0,001

Fonte – O autor

Analisando o mapa da esquerda na Figura 2, tem-se o comportamento do número de óbitos acumulados por município até o dia 30 de junho de 2020. Foi utilizado a escala logarítmica acrescida de uma unidade para uma melhor visualização, uma vez que os dados possuem grande dispersão e até o momento em que os dados foram coletados 8 municípios ainda não haviam confirmado óbitos. Nota-se que os municípios com maior número de óbitos se concentraram na região Metropolitana, com um destaque especial para o município do Rio de Janeiro. No extremo norte do estado, a cidade de Campos dos Goytacazes apresenta um maior número de óbitos do que os seus municípios vizinhos.

Já o mapa da direita na Figura 2, apresenta a significância das estimativas obtidas para os Índices de Moran local. O comportamento espacial da variável resposta nos municípios neste tipo de análise, usualmente, é caracterizado em ‘Alto - Alto’ quando um município apresenta autocorrelação espacial positiva e valores acima da média da variável na região de estudo, ‘Baixo – Baixo’, quando apresenta autocorrelação espacial negativa e valores abaixo da média da região de estudo conforme descrito por Anselin (1995). Percebe-se que todos os municípios que apresentaram dependência espacial significativa apresentaram um único tipo e fazem parte da região Metropolitana e, com exceção do próprio, todos são vizinhos do município que apresentou as maiores contagens, o Rio de Janeiro.

Figura 2 – Análise espacial do número de óbitos confirmados pela COVID-19 no Rio de Janeiro. No mapa a esquerda é apresentado o log do número de óbitos mais uma unidade e no mapa a direita são apresentadas as significâncias das estimativas dos índices de Moran local.



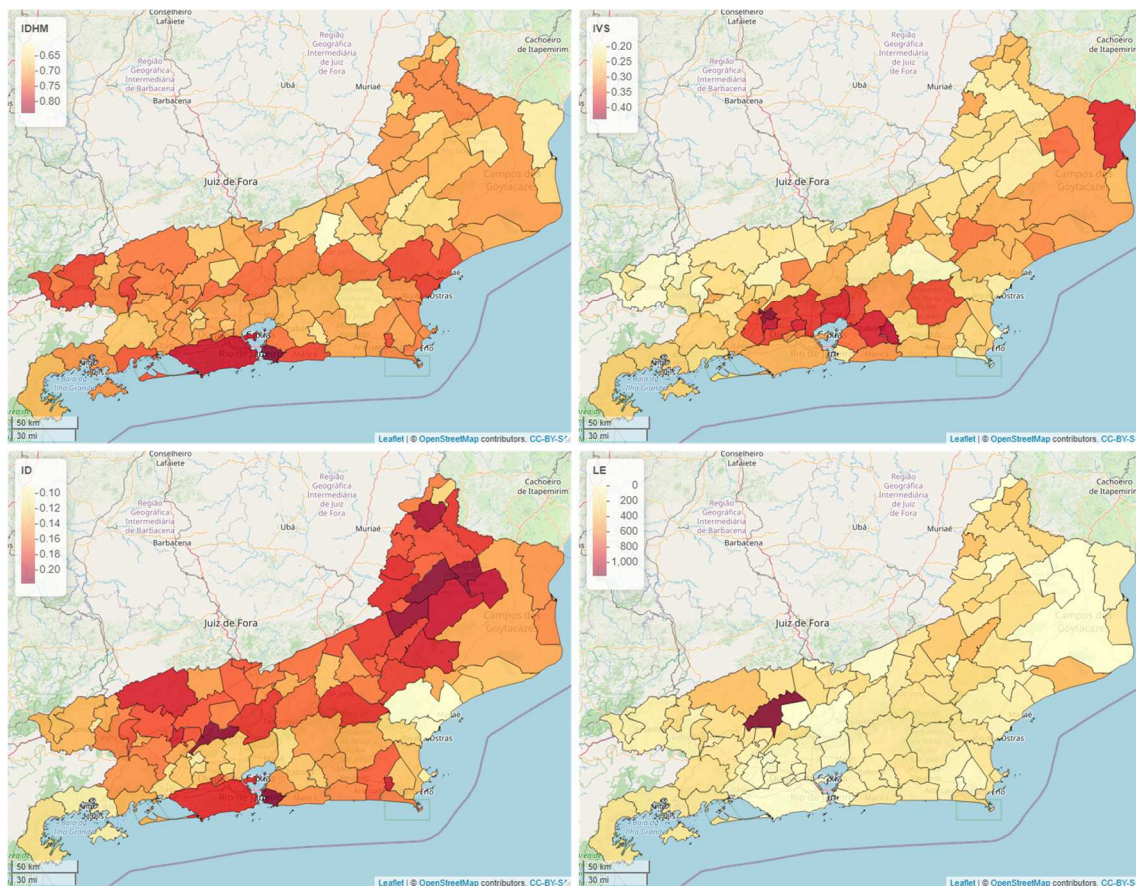
Fonte – O autor

Após a realização da análise exploratória, verificou-se a existência de uma correlação espacial global baixa, porém significativa e a existência de efeitos locais. A seguir foram construídos mapas coropléticos para as covariáveis selecionadas para compor os modelos finais investigados neste trabalho. Conhecer o comportamento espacial destas variáveis é um passo importante para entender a maneira com que as discontinuidades são identificadas pelo CAR Dissimilaridade.

Os mapas coropléticos para as quatro covariáveis são apresentados na Figura 3. A figura superior esquerda apresenta o mapa do IDHM, nela pode-se perceber que IDHMs mais elevados estão concentrados na Região Metropolitana do estado, com exceção de alguns poucos municípios isolados. O mapa coroplético para o IVS permite observar claramente uma faixa de municípios com valores mais elevados de vulnerabilidade na Região Metropolitana, além destes se destaca um único município na Região Norte do estado. Para a Proporção de Idosos nota-se uma concentração de municípios com quantidades elevadas de idosos nas regiões Norte e Noroeste, porém em grande parte estas quantidades são similares. Nas demais regiões do estado notam-se municípios isolados com proporções maiores do que seus vizinhos. Por fim, para os Leitos do SUS por 100 mil habitantes percebe-se um comportamento bastante homogêneo na Região Metropolitana, nas demais regiões existem municípios isolados com um número de leitos ligeiramente maior que seus vizinhos, o que possivelmente é causado pela forma com que o sistema de saúde local é estruturado.

Em seguida, foram ajustados os dois modelos apresentados na Subseção 2.4 para estimar as superfícies de riscos. Foi realizado uma seleção de covariáveis, por meio da avaliação do DIC (Spiegelhalter et al., 2014). Os ajustes apresentados abaixo são os que apresentaram melhores valores de DIC: 519,155 (Leroux) e 605,342 (Dissimilaridade). Este resultado sugere que o modelo Leroux possui um melhor ajuste aos dados do que o modelo CAR Dissimilaridade. No seu trabalho, Lee & Mitchel (2012) já ressaltaram que este resultado não traz nenhuma surpresa, pois o modelo Leroux usa as covariáveis na estrutura da média para fornecer uma melhor descrição dos dados, enquanto no modelo de detecção de limites as covariáveis são utilizadas para identificar limites por meio da estrutura de covariância.

Figura 3 – Mapas coropléticos para as variáveis: Índice de Desenvolvimento Humano Municipal (IDHM), Índice de Vulnerabilidade Social (IVS), Proporção de Idosos (ID) e Leitos do SUS (LE) por 100 mil habitantes.



Fonte – O autor

A Tabela 4 apresenta a estimativa pontual e o intervalo de credibilidade para os efeitos das covariáveis selecionadas e hiperparâmetros resultantes do ajuste do modelo Leroux. Neste cenário, é possível estimar a superfície de risco (R_i) e quantificar o impacto das covariáveis na superfície.

Tabela 4 – Sumário do modelo Leroux para o número de óbitos confirmados por COVID-19.

Variável	Estimativa	Intervalo de Credibilidade	
		2,5%	97,5%
Intercepto	-11,415	-16,300	-6,839
IDHM	12,584	7,355	18,219
IVS	8,308	4,433	12,322
ID	-2,462	-8,851	3,683
LE	0,000	-0,001	0,002
τ	1,830	1,120	2,970
ρ	0,168	0,067	0,199

Fonte – O autor

Das quatro covariáveis selecionadas para o modelo final, apenas o IDHM e o IVS apresentaram estimativas significativas. Seria natural esperar que a contribuição do IDHM fosse negativa, pois municípios com melhores IDHMs teriam maior capacidade de tratar os doentes. Entretanto, os municípios do estado do Rio de Janeiro com maiores IDHMs também são aqueles que apresentam as maiores quantidades de óbitos pela doença, possivelmente por possuírem uma melhor infraestrutura, isto é, melhores sistemas de saúde e capacidade de testagem.

O IVS também apresentou estimativa positiva, um comportamento esperado visto que quanto mais próximo de 1 o valor do IVS, maior a vulnerabilidade do município, logo, maior vulnerabilidade está associada a um maior risco de óbitos por COVID-19. Um aumento de 0,1 no IVS dobra o risco relativo de óbito por COVID-19 em um município.

No ajuste do modelo CAR Dissimilaridade, é possível estimar a superfície de risco (R_i) e identificar quais covariáveis são úteis na identificação de possíveis descontinuidades nos riscos. Na Tabela 5 são apresentados os limites de identificação de descontinuidades de cada variável, seus intervalos de credibilidade de 95% e o número de descontinuidades identificadas.

Tabela 5 – Sumário do modelo CAR Dissimilaridade para o número de óbitos confirmados por COVID-19.

Variável	α^{min}	Intervalo de Credibilidade	
		2,5%	97,5%
IDHM	0,113	0,011	0,269
IVS	0,155	0,049	0,271
ID	0,156	0,246	0,462
LE	0,091	0,016	0,189
Descontinuidades	79	-	-

Fonte – O autor

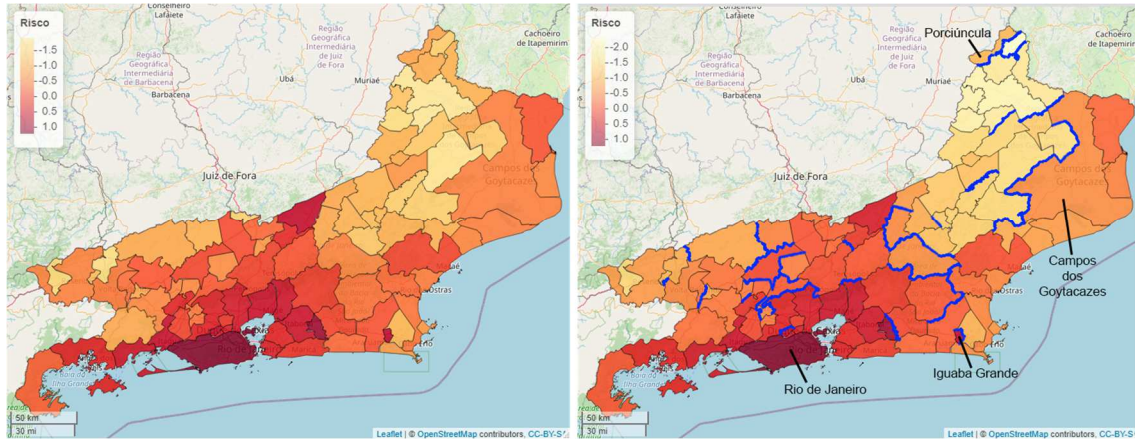
Pelo ajuste do modelo, percebe-se que apenas a covariável proporção de idosos (ID) apresentou um efeito significativo para a identificação de limites de descontinuidades da superfície de risco, uma vez que o limite inferior do intervalo de credibilidade do regressor α associado a esta covariável é maior que o limiar de identificação α_k^{min} . O IDHM, o IVS e a taxa de leitos do SUS por 100 mil habitantes (LE) apresentaram intervalos que continham α_k^{min} , de forma que se considerou que estas variáveis contribuíram de alguma forma na identificação dos limites. Ao todo foram identificadas 79 descontinuidades na região de estudo, o que equivale a 34,93% de todas as relações de vizinhança existentes entre os municípios do estado.

A Figura 4 apresenta as superfícies de risco estimadas na escala logarítmica pelos modelos Leroux, a esquerda, e CAR Dissimilaridade, a direita. Novamente foi utilizado a escala logarítmica para uma melhor representação visual dos resultados. A análise do CAR Dissimilaridade, também inclui as descontinuidades identificadas pelo modelo em azul.

Nota-se que o comportamento espacial do logaritmo do risco estimado nos municípios foi bastante similar entre ambos os modelos, apresentando uma correlação entre as superfícies de riscos estimada de 0,98. Resultado semelhante ao encontrado por Lee & Mitchel (2012) na aplicação considerada em seu artigo. Nota-se também que em ambas as análises, a cidade do Rio de Janeiro é apontada como a cidade com o maior risco relativo de óbito pela COVID-19, sendo o risco estimado de 3,34 no modelo Leroux e de 3,33 no

modelo CAR Dissimilaridade. Após a capital, os municípios que compõem a Região Metropolitana possuem os maiores riscos.

Figura 4 – Log dos riscos estimados pelos modelos Leroux (esquerda) e CAR Dissimilaridade (direita).



Fonte – O autor

Entretanto, vale ressaltar que o modelo CAR Dissimilaridade aponta as possíveis fronteiras que indicam uma descontinuidade na superfície de risco marcadas como linhas azuis. A maioria dos limites estimados parece corresponder a mudanças na superfície de risco, sugerindo que a covariável proporção de idosos parece ser uma boa métrica de dissimilaridade para detectar descontinuidades. De forma geral, como as pessoas idosas fazem parte do grupo de risco é possível que o risco de uma região esteja fortemente associado a esta variável. Ao comparar a localização das descontinuidades com o mapa apresentado na Figura 3, pode-se perceber que, de uma forma geral, elas coincidem com as fronteiras dos municípios que apresentaram proporções de idosos substancialmente distintas de seus vizinhos, o que se torna especialmente claro ao observar as regiões dos municípios de Porciúncula, Campos dos Goytacazes e Iguaba Grande. Entretanto, vale ressaltar que alguns limites identificados não mostram fortes evidências de separação das sub-regiões, isto é, sub-regiões com diferentes riscos, como alguns limites indicados a noroeste do estado.

Destaca-se ainda que a Região Noroeste do estado apresentou os menores riscos, região mais clara na Figura 3. Vale a pena ressaltar as diferenças observadas para o município de Porciúncula pertencente a esta região nos dois ajustes. Enquanto o modelo global ajustou riscos mais semelhantes para os vizinhos de Porciúncula, o modelo local apresentou as fronteiras deste município devidamente identificadas como limites, indicando que a superfície de risco possui uma descontinuidade nesta localidade. De forma

geral, as variáveis utilizadas no CAR Dissimilaridade apresentaram um bom desempenho dado as circunstâncias que envolvem os dados da COVID-19.

4.CONCLUSÃO

Neste trabalho foram apresentadas aplicações de modelos espaciais para o número de óbitos confirmados pela COVID-19 nos 92 municípios do estado do Rio de Janeiro. A pandemia se desenvolveu de forma espacialmente heterogênea no estado, se concentrando inicialmente na região metropolitana, o que foi corroborado pelo fato de que todos os municípios que apresentaram dependência espacial local significativa faziam parte da mesma. Isto faz com que estratégias comuns como o uso de modelos com distribuições *a priori* condicionais autorregressivas para os efeitos espaciais possam não ser as mais adequadas, pois podem gerar uma superfície de risco excessivamente suavizada.

Para lidar com este cenário foi proposto o uso do CAR Dissimilaridade, um modelo de suavização local capaz de identificar discontinuidades na superfície de risco. Com base no uso de covariáveis o modelo identifica quais fronteiras entre municípios da região de estudo são discontinuidades, ou seja, quais municípios apresentam riscos significativamente distintos.

O modelo foi aplicado aos dados compreendendo o período desde o começo da epidemia até o dia 30 de junho de 2020. Foram utilizadas covariáveis que buscavam caracterizar os municípios, os índices de Desenvolvimento Humano e de Vulnerabilidade Social; caracterizar as populações vulneráveis, a proporção de idosos, e caracterizar a capacidade dos municípios de cuidar dos enfermos, a taxa de leitos do SUS por 100 mil habitantes.

No ajuste do modelo de autocorrelação global, o Leroux, percebeu-se que as variáveis IDHM e IVS estavam associadas ao risco de óbito pela COVID-19. A análise indicou que um incremento de 10% do IVS implica em duplicar o risco relativo do município, mantendo as demais covariáveis iguais. Outra constatação é que existe um efeito positivo para o IDHM, indicando municípios com maiores valores de IDHM possuem maiores riscos, sendo este um possível reflexo de melhores sistemas de saúde, capacidade de testagem e o próprio ciclo de propagação da doença que se inicia nas capitais, com maior fluxo de movimentação, e só depois vai se interiorizando.

Já no ajuste do modelo local, o CAR Dissimilaridade, verificou-se que todas as covariáveis apresentaram alguma contribuição para a identificação das discontinuidades. Destas, a única que apresentou contribuição substancial foi a proporção de idosos. As superfícies de riscos estimadas apresentaram alta

correlação, mas diferenças foram observadas ao se avaliar a suavidade de determinadas localidades. Pode-se observar que a região Noroeste possui os menores riscos do estado, mas foi estimada de forma mais suave no modelo Leroux do que no modelo CAR Dissimilaridade. Este último identificou uma fronteira completa para o município mais extremo desta região, Porciúncula, que possui um risco maior quando comparado aos demais municípios que compõem a região, uma possível explicação para este resultado é que o fluxo das pessoas desta localidade pode estar mais associado à cidades fronteiras do estado vizinho.

Fatores como a existência de comorbidades incluindo diabetes e doenças cardiovasculares aumentam o risco de óbito, e que o isolamento social pode interferir na quantidade de casos, que naturalmente pode sobrecarregar o sistema de saúde e influenciar numa maior quantidade de óbitos. Entretanto, estas variáveis não estavam disponíveis para os municípios do estado do Rio de Janeiro, sendo esta uma limitação deste estudo.

5.REFERÊNCIAS BIBLIOGRÁFICAS

Abuelgasim, E., Saw, L. J., Shrike, M., Zeinah, M. & Harke, A. (2020). COVID-19: Unique public health issues facing Black, Asian and minority ethnic communities. *Current problems in cardiology*, 45(8).

Adin, A., Lee, D., Goicoa, T. & Ugarte, M. D. (2019). A two-stage approach to estimate spatial and spatio-temporal disease risks in the presence of local discontinuities and clusters. *Statistical Methods in Medical Research*, 28(9), 2595-2613.

Anderson, C., Lee, D. & Dean, N. (2014). Identifying clusters in Bayesian disease mapping. *Biostatistics*, 15(3), 457-469.

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis*, 27(2), 93-115.

Banerjee, S., Carlin, B. P. & Gelfand, A. E. (2015). *Hierarchical modelling and analysis for spatial data* (2a ed.). Boca Raton: Chapman and Hall/CRC Press.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion) *J. Roy. Statist. Soc., Ser. B*;36:192–236.

Besag, J., York, J. & Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43, 1-20.

Bivand, R. S., Pebesma, E. & Gómez-Rubio, V. (2013). *Applied spatial data analysis with R* (2a ed.). Nova Iorque: Springer.

Clayton, D. & Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 43(3), 671–681.

COVID-19 no Brasil. (2020). Ministério da Saúde. Disponível em: <https://susanalitico.saude.gov.br/extensions/covid-19_html/covid-19_html.html>. Acesso em: 27 de ago. de 2020.

Estimativas Populacionais. (2019). Instituto Brasileiro de Geografia e Estatística (IBGE). Disponível em: <<https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?edicao=25272&t=resultados>>. Acesso em: 27 de ago. de 2020.

Gamerman, D. & Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference (2 ed.)*. Boca Raton: Chapman and Hall/CRC Press.

Gold, J. A. W., Wong, K. K., Szablewski, C. M., Patel, P. R. et al. (2020). Characteristics and clinical outcomes of adult patients hospitalized with COVID-19 - Geórgia, March 2020. *Morbidity and Mortality Weekly Report*, 69(18), 545-550.

Guan, W., Ni, Z., Hu, Y., Liang, W., Ou, C., et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. *New England Journal of Medicine*, 382(18), 1708-1720.

Huang, X., Wei, F., Hu, L., Wen, L. & Chen, K. (2020). Epidemiology and clinical characteristics of COVID-19. *Archives of Iranian Medicine*, 23(4), 268-271.

Índice de Vulnerabilidade Municipal - COVID-19. (2020). Instituto Votorantim. Disponível em: <<http://institutovotorantim.org.br/municipioscontraocorona/ivm/>>. Acesso em: 17 de jun. de 2020.

Informações de Saúde (TABNET) - DATASUS. (2020). DATASUS. Disponível em: <<https://datasus.saude.gov.br/informacoes-de-saude-tabnet/>>. Acesso em: 17 de jun. de 2020.

IPEA. (2020). Atlas do Desenvolvimento Humano no Brasil. Disponível em: <<http://atlasbrasil.org.br/>>. Acesso em: 10 de jun. de 2020.

IPEA. (2020). Atlas da Vulnerabilidade Social. Disponível em: <<http://ivs.ipea.gov.br/index.php/pt/>>. Acesso em: 29 de ago. de 2020.

Koch T. (2005). *Cartographies of Disease: Maps, Mapping, and Medicine*. Redlands, CA: ESRI Press.

Laurencin, C. T. & McClinton, A. (2020). The COVID-19 Pandemic: A call to action to identify and address racial and ethnic disparities. *Journal of Racial and Ethnic Health Disparities*, 7, 398-402.

Lee, D. (2013). CARBayes: An R Package for Bayesian spatial modeling with conditional autoregressive priors. *Journal of Statistical Software*, 55(13), 1-24.

Lee, D. & Mitchell, R. (2012). Boundary detection in disease mapping studies. *Biostatistics*. 13(3), 415-426.

Leroux, B. G., Lei, X. & Breslow N. (2000). Estimation of disease rates in small areas: A new mixed model for spatial dependence. *Statistical models in epidemiology, the environment, and clinical trials*, 116, 179-191.

López-Abente, G., Aragonés, N., García-Pérez, J., & Fernández-Navarro, P. (2014). Disease mapping and spatio-temporal analysis: importance of expected-case computation criteria. *Geospatial Health*, 9(1), 27-35.

Lu, H. & Carlin, B. P. (2005). Bayesian areal wombling for geographical boundary analysis. *Geographical Analysis*, 37, 265–285.

Painel Coronavírus. (2020). Ministério da Saúde. Disponível em: <<http://covid.saude.gov.br>>. Acesso em: 19 de jun. de 2020.

Painel Coronavírus COVID-19. (2020). Secretaria de Saúde do Rio de Janeiro. Disponível em: < <http://painel.saude.rj.gov.br/monitoramento/covid19.html#>>. Acesso em: 19 de set. de 2020.

Coronavirus Pandemic (COVID-19). (2020). Roser, M., Ritchie, H., Ortiz-Ospina, E. & Hasell, J. Disponível em: <<https://ourworldindata.org/coronavirus-testing#brazil>>. Acesso em: 18 de set. de 2020.

Rue, H., Martino, S. & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 71(2), 319-392.

Sobre a doença. (2020). Ministério da Saúde. Disponível em: <<https://coronavirus.saude.gov.br/sobre-a-doenca#o-que-e-covid>>. Acesso em: 17 de jun. de 2020.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., van der Linde, A. (2014). The deviance information criterion: 12 years on. *Journal of the Royal Society of Statistics, Series B (Statistical Methodology)*, 76(3), 485-493.

Waller, L. A. & Carlin, B. P. (2010). Disease Mapping. *Chapman & Hall/CRC handbooks of modern statistical methods*, 1, 217-243.

Zi, Y. Z., Meng, D. J., Peng P. X., et al. (2020). Coronavirus disease 2019 (COVID-19): A perspective from China. *Radiology*, 296(2), 15-25.

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 359(10229), 1054-1062.

AGRADECIMENTOS e COLABORAÇÕES

Os autores agradecem ao Programa Institucional de Bolsas de Iniciação Científica (PIBIC) da Universidade Federal Fluminense (UFF) pela bolsa do Ricardo Junqueira de Souza e aos revisores anônimos pelas contribuições que tornaram o trabalho mais compreensível.

ESTIMAÇÃO DO NÚMERO DE REPRODUÇÃO DA PANDEMIA POR COVID-19 EM SÃO PAULO

Chang Chiann

chang@ime.usp.br

Universidade de São Paulo

Michel H. Montoril

Universidade Federal de São Carlos

Pedro A. Morettin

Universidade de São Paulo

Resumo: A partir do início da pandemia devida à COVID-19, vários estudos têm sido desenvolvidos com a finalidade de contribuir para um melhor conhecimento a respeito desse novo vírus, até então desconhecido. Em uma das frentes de pesquisa, diversos modelos visam estudar o que se chama de número de reprodução, que mede a velocidade de contágio do vírus. Neste trabalho, consideramos duas formas de se estimar essa quantidade. As metodologias utilizadas são aplicadas a dados do município e do estado de São Paulo, e seus resultados são úteis para descrever o perfil de contágio ao longo do tempo.

Palavras-chave: Covid-19; escalograma; número de reprodução; médias móveis; ondaletas

Abstract: Since the beginning of the pandemic due to COVID-19, several studies have been developed aiming to contribute to a better knowledge about the virus. Among the researches, many models intend to study the reproduction number, which measures how fast is the virus spreading. In this work, we consider two ways of estimating the reproduction number. The employed methodologies are applied to the daily data of new cases in the state and the city of São Paulo. The results are useful to describe the contagion profile along time.

Keywords: Covid-19; scalogram; reproduction number; moving average; wavelets

1. INTRODUÇÃO

Desde que começou a pandemia devida à COVID-19, diversos modelos têm sido utilizados (publicados ou divulgados) para descrever fenômenos associados a esse tipo de evento. Sob o risco de simplificação, há os modelos epidemiológicos (SIR, SEIR e modificações, como o da EMAP/FGV (2020)), modelos matemáticos, modelos estatísticos etc.

Um modelo bastante divulgado é o do Imperial College (IC), que é um modelo bayesiano hierárquico. Esse modelo estima o número de infectados a partir de um modelo simples, que depende, em cada t (dia), de $R(t)$, o número de reprodução (do inglês *reproduction number*) e de uma soma ponderada de infectados nos $t-1$ dias anteriores.

O número de reprodução $R(t)$ representa o número esperado de contágio de cada pessoa infectada no dia t . Esse número caracteriza o status da epidemia, no sentido que $R(t) < 1$ indica retração, enquanto $R(t) > 1$ indica intensificação da epidemia. Logo, $R(t)$ está sujeito a intervenções na população, como quarentenas parciais ou totais.

Ainda sobre estudos do IC, a dificuldade está em estimar $R(t)$ e os pesos. Para isso, usam informações sobre diversos índices de intervenção (6 no total), como o fechamento de escolas e universidades, intervenções governamentais (isolamento e distanciamento social) etc. Em outras palavras, $R(t)$ é estimado via uma função constante por partes, produto de indicadores. Alguns índices valem para todos os países, outros variam segundo o país. Para mais detalhes, veja Flaxman et al. (2020).

Dentre os modelos propostos, a maioria tem por objetivo prever valores de séries de infectados, mortos e recuperados. Nossa opinião é que fazer previsões pode ser uma tarefa complicada, devido à quantidade e pobreza dos dados. Centenas de modelos foram considerados na literatura e previsões geralmente são ruins ao prever picos, final da epidemia etc., especialmente para nossos dados (não há picos e sim platôs). Além disso, podemos observar que muitos modelos, especialmente os epidemiológicos, apresentam previsões unimodais, o que não se observa em muitos conjuntos de dados e especialmente nos nossos.

Nosso objetivo nesse artigo é estimar $R(t)$ para dados do estado e do município de São Paulo, e o interesse é estudar a tendência da série. Para tanto, vamos utilizar dois modelos para estimar $R(t)$: (i) um modelo ingênuo, que parte de uma média móvel simples de sete dias para as séries de infectados no estado e na capital para, em seguida, ajustar um modelo

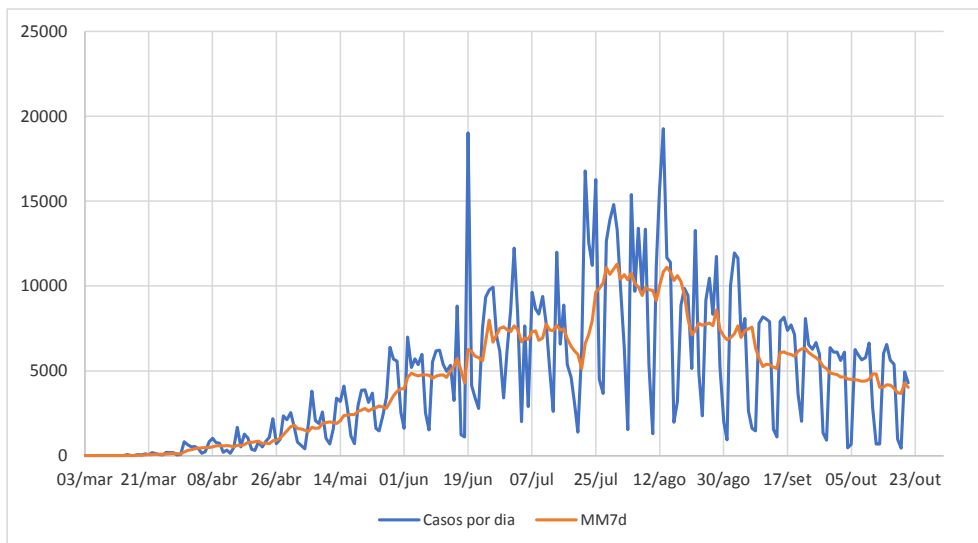
autorregressivo de ordem um, cujo coeficiente variando no tempo é $R(t)$. A ideia é que o número de infectados no instante t depende de $R(t)$ e do número de infectados até o instante $t-1$. Esse coeficiente é estimado via bases de ondaletas. Poder-se-ia considerar um processo autorregressivo de ordem p , com coeficientes variando no tempo, mas seria difícil caracterizar $R(t)$ nesse caso, portanto nesse artigo vamos considerar $p=1$; (ii) utilizamos um segundo modelo, no qual a série de infectados é decomposta em componentes não observáveis, a saber, a tendência e um erro aleatório. A determinação dessas duas componentes é feita novamente com expansões em ondaletas, tomando como base a metodologia proposta por Ariño et al. (2004).

A organização do artigo é a seguinte: na Seção 2 descrevemos os dados usados no artigo e, na Seção 3, apresentamos uma breve introdução às ondaletas e a metodologia a ser usada. Na Seção 4 aplicamos os modelos aos dados do estado de São Paulo e da capital. Finalmente, na Seção 5 fazemos alguns comentários adicionais.

2. OS DADOS

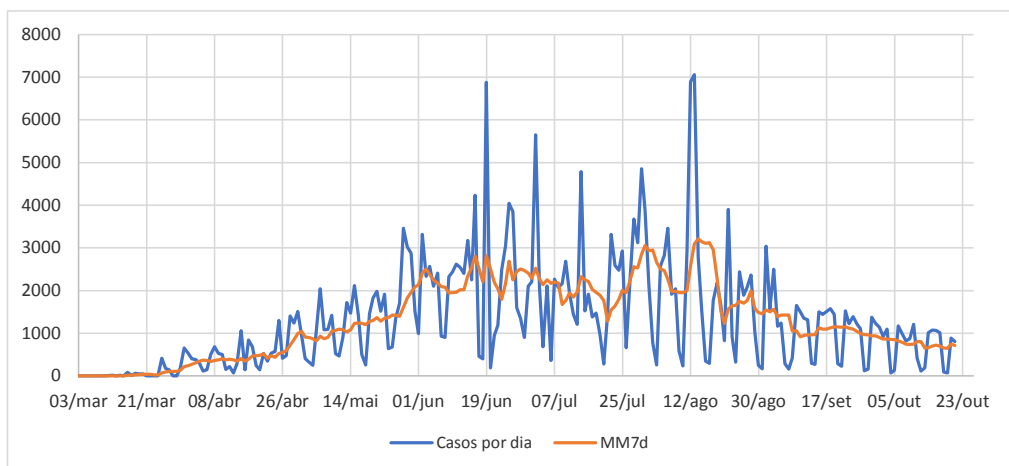
Os dados sobre o número de infectados e mortes no estado de São Paulo, dos demais estados e do Brasil, são pouco confiáveis. Para se ter uma ideia, entre os dias 26/02/2020 e 21/10/2020 tivemos 1.073.261 total de infectados confirmados e 38.371 mortes no estado de São Paulo, segundo dados divulgados no sítio da Fundação SEADE (<https://www.seade.gov.br/coronavirus>). Entre os dias registrados, há períodos de dados atípicos. Entre os dias 15/06/2020 e 20/06/2020, por exemplo, os casos confirmados foram 3.258, 8.825, 1.232, 1.111, 19.030, 4.135, respectivamente. Note que entre os dias 15 e 16 a variação foi superior a 5 mil casos, representando um aumento superior a 170%. Já entre os dias 18/06/2020 e 19/06/2020, a variação foi superior a 9 mil casos (ultrapassando 1.612%). A Figura 1 mostra com mais detalhe essas variações atípicas durante o período estudado.

Figura 1 - Série diária de novos casos de COVID-19 no estado de São Paulo e média móvel de sete dias.



Os dados sobre o número de infectados e mortes no município de São Paulo seguem a mesma tendência que os do estado, presença de variações atípicas conforme Figura 2.

Figura 2 - Série diária de novos casos de COVID-19 no município de São Paulo e média móvel de sete dias



3. CONCEITOS BÁSICOS

Nesta Seção descrevemos brevemente noções sobre ondaletas e modelos autorregressivos variando no tempo.

3.1 ONDALETAS

Seja Z o conjunto dos inteiros. A partir de duas funções, a função escala $\phi(x)$ e a ondaleta mãe $\psi(x)$, podemos definir infinitas coleções de translações e dilatações, $\phi_{j,k}(x) = 2^{j/2} \phi(2^j x - k)$, $\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$, $j, k \in Z$.

Supomos que $\{\phi_{\ell,k}(\cdot)\}_{k \in Z} \cup \{\psi_{j,k}(\cdot)\}_{j \geq \ell, k \in Z}$ forma uma base ortonormal de $L_2(\mathcal{R})$, para algum ℓ inteiro. O ponto chave (Daubechies, 1992) é a possibilidade de se construir funções de ondaletas com suporte compacto. Essas ondaletas geram um sistema ortonormal com boa localização em tempo-escala, que permite representações parcimoniosas para amplas classes de funções.

Assim, para qualquer função $f \in L_2(\mathcal{R})$, podemos expandi-la em uma série ortonormal

$$f(x) = \sum_{k \in Z} \alpha_{\ell,k} \phi_{\ell,k}(x) + \sum_{j \geq \ell} \sum_{k \in Z} \beta_{j,k} \psi_{j,k}(x), \quad (1)$$

para algum ℓ , com os coeficientes de ondaletas dados por

$$\alpha_{\ell,k} = \int f(x) \phi_{\ell,k}(x) dx, \quad \beta_{j,k} = \int f(x) \psi_{j,k}(x) dx. \quad (2)$$

Considere a expansão (1). Admitimos que a seguinte suposição seja válida:

S1. ϕ e ψ são tais que:

(a) $\int \phi(t) dt = 1$;

(b) $\int \psi(t) dt = 0$;

(c) as funções ϕ e ψ têm suporte compacto com $r-1$ momentos nulos, para algum $r \geq 1$; além disso, $\int |x^r \phi(x)| dx < \infty$ e $\int |x^r \psi(x)| dx < \infty$.

3.2 ESCALOGRAMA

O escalograma é análogo ao periodograma da análise espectral clássica de séries temporais. Como o periodograma produz uma decomposição de energia de um sinal em diferentes frequências, o escalograma decompõe a energia em diferentes escalas.

Seja $x = (x_0, x_1, \dots, x_{T-1})$, um vetor de observações com $T = 2^n$. Para $j = 0, 1, \dots, n-1$ e $k = 0, \dots, 2^j - 1$, definimos a transformada discreta de ondaletas com respeito a ψ como

$$d_{j,k} = \sum_{t=0}^{T-1} x_t \psi_{j,k}(t/T), \quad (3)$$

e a transformada discreta de ondaletas com respeito a ϕ como

$$c_{l,k} = \sum_{t=0}^{T-1} x_t \phi_{l,k}(t/T). \quad (4)$$

Na prática, as transformadas discretas não são calculadas por (3) e (4), mas usando um algoritmo piramidal, devido a Mallat. Veja Morettin et al. (2017) para detalhes.

Utilizaremos $\ell = 0$ como nível mais grosso de resolução. Se $\mathbf{d} = (c_{0,0}, d_{0,0}, d_{1,0}, d_{1,1}, \dots, d_{n-1}, 2^{n-1} - 1)$ for o vetor de coeficientes da transformada discreta de ondaletas de \mathbf{x} , a energia $E(j)$ de \mathbf{d} no nível j é definida como

$$E(j) = \sum_{k=0}^{2^j-1} d_{j,k}^2 \quad \text{para } j = 0, 1, \dots, n-1. \quad (5)$$

Usando o escalograma da decomposição em ondaletas correspondente, poderemos dividir a decomposição da ondaleta em duas partes. Reconstruiremos os dois componentes usando escalograma. Para mais detalhes, veja Ariño et al. (2004).

3.3 MODELOS AUTORREGRESSIVOS COM COEFICIENTES VARIANDO NO TEMPO

Diz-se que uma série temporal X_t segue um modelo autorregressivo de ordem p , com coeficientes variando no tempo, se satisfizer

$$X_t = a_1(t)X_{t-1} + \dots + a_p(t)X_{t-p} + \varepsilon_t, \quad (6)$$

em que p é um inteiro positivo, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é uma sequência de variáveis aleatórias independentes identicamente distribuídas (iid) com média zero e variância σ^2 , de tal modo que ε_t seja independente de $\{X_{t-i}\}_{i>0}$. Utilizaremos a notação $tvAR(p)$ para esse modelo.

O objetivo é estimar os parâmetros funcionais $a_i(\cdot)$, $i = 1, \dots, p$, a partir das observações $\{X_t, t = 1, \dots, T\}$ do processo.

Expandimos $a_i(\cdot)$, $i = 1, \dots, p$ como

$$a_i(t) = \alpha_{0,0}^{(i)} \phi_{0,0}(t) + \sum_{j=0}^{\infty} \sum_k \beta_{j,k}^{(i)} \psi_{j,k}(t), \quad (7)$$

em que

$$\alpha_{0,0}^{(i)} = \int a_i(t) \phi_{0,0}(t) dt, \quad \beta_{j,k}^{(i)} = \int a_i(t) \psi_{j,k}(t) dt. \quad (8)$$

Em (7), $k \in I_j = \{k: k = 0, \dots, 2^j - 1\}$ e $j = 0, \dots, J^T - 1$ no segundo termo, para alguma nível J^T dependendo de T . Em geral, assumimos $T = 2^J$ e $J^T \leq J$. Os coeficientes de ondaletas podem ser estimados por meio de uma regressão linear, utilizando o método de mínimos quadrados.

Aqui, admitimos a seguinte suposição:

S2. A função $a(t)$ é quadrado integrável e possui segunda derivada contínua, para $i = 1, \dots, p$. Para detalhes referentes ao procedimento de estimação, veja Chiann & Morettin (2005).

4. APLICAÇÕES

4.1 CASOS CONFIRMADOS DE COVID-19 NO ESTADO DE SÃO PAULO

A Figura 1 apresenta novos casos confirmados de Covid-19 no estado de São Paulo, no período de 26/02/2020 a 21/10/2020, resultando em uma série com $T = 239$ observações. Nosso objetivo é estimar o número de reprodução $R(t)$.

4.1.1 MÉTODO 1: AJUSTE DE UM MODELO tvAR(1) PARA A SÉRIE DE MÉDIAS MÓVEIS

Como podemos observar pela Figura 1, os dados apresentam muitas variações atípicas. Inicialmente, suavizamos a série de infectados no estado com médias móveis de sete dias e, em seguida, ajustamos um modelo autorregressivo de ordem 1, com coeficiente variando no tempo, que corresponderá a $R(t)$. A relação direta de $R(t)$ com o modelo tvAR(1), bem como a flexibilidade dessa classe de modelos, motivaram essa escolha. Esse coeficiente é estimado via ondaletas de Haar.

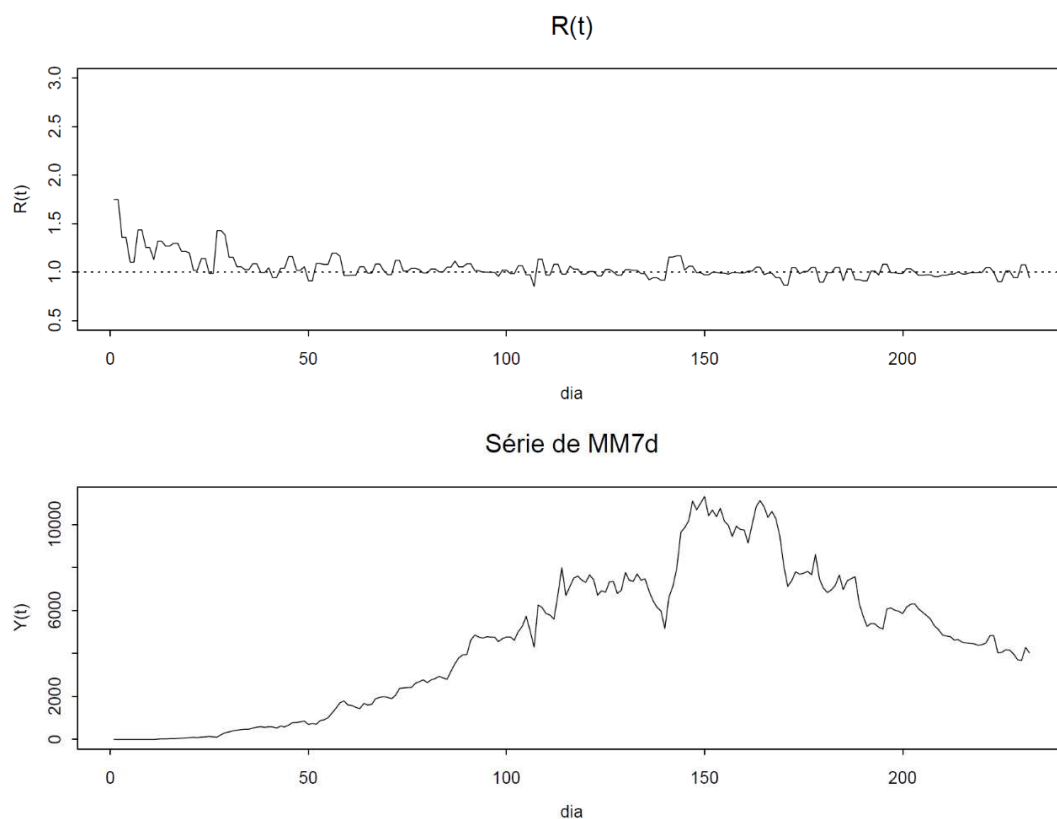
O modelo a ser ajustado é do tipo

$$X_t = R(t)X_{t-1} + \varepsilon_t, \quad (9)$$

em que $\{X_t, t = 1, \dots, T\}$ corresponde à série de médias móveis de sete dias da série de casos confirmados no estado.

A partir de (9), estimamos $R(t)$ usando uma expansão em ondaletas de Haar com $J = 7$. A estimativa é apresentada na Figura 3, juntamente com a curva de médias móveis.

Figura 3 - Estimativa do número de reprodução $R(t)$ para a série diária de novos casos no estado de São Paulo (gráfico superior), juntamente com a série de médias móveis (gráfico inferior). Resultados obtidos por meio do método 1 (Haar, com $J = 7$)



Observamos que o comportamento de $R(t)$ é condizente com o das médias móveis, no sentido de que ao passo em que $R(t)$ diminui, o crescimento das médias móveis desacelera. No final do período, a curva de infectados parece já ter atingido um platô, que poderá ser confirmado nos próximos dias, enquanto $R(t)$ oscila ao redor de um. Lembremos que a epidemia ficará controlada quando $R(t) < 1$.

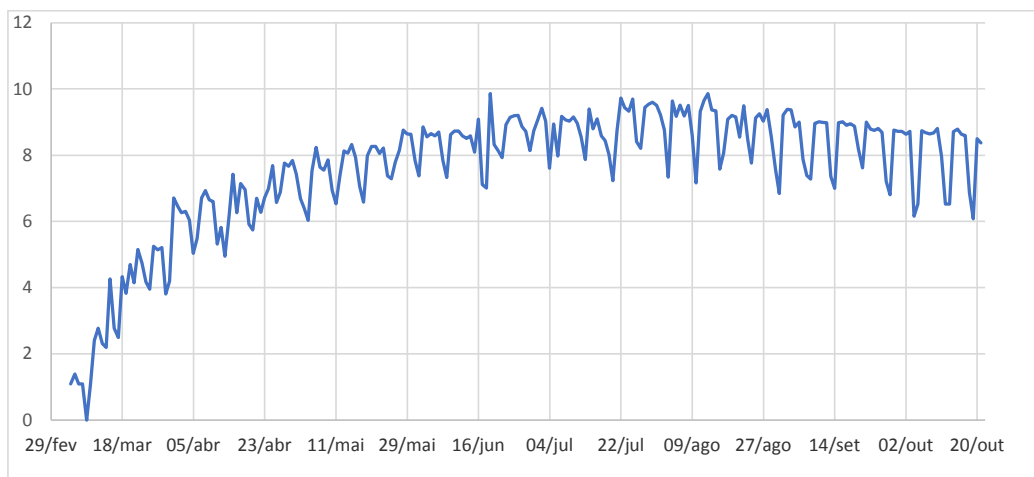
Vale destacar que as estimativas de $R(t)$ apresentadas ao longo do texto apresentam um comportamento de “escada”. Isso se dá devido ao fato de estarmos usando uma base de Haar, que naturalmente possui comportamento descontínuo. Outras bases de ondaletas (mais suaves) poderiam ser usadas. Todavia, análises não reportadas neste trabalho, utilizando os mesmos dados e diferentes bases de ondaletas, indicaram que o uso da base de Haar tende a apresentar resultados mais satisfatórios.

4.1.2 MÉTODO 2: USO DO ESCALOGRAMA E DO MODELO $tvAR(1)$

A seguir, usando o escalograma, decompomos a série de infectados em componentes não observáveis, a saber, tendência e erro aleatório.

Denotemos a série de novos casos por x_t . Em primeiro lugar, a fim de estabilizar a variância da série, trabalharemos com o logaritmo de x_t , apresentado na Figura 4.

Figura 4 - Logaritmo da série diária de novos casos de COVID-19 no estado de São Paulo



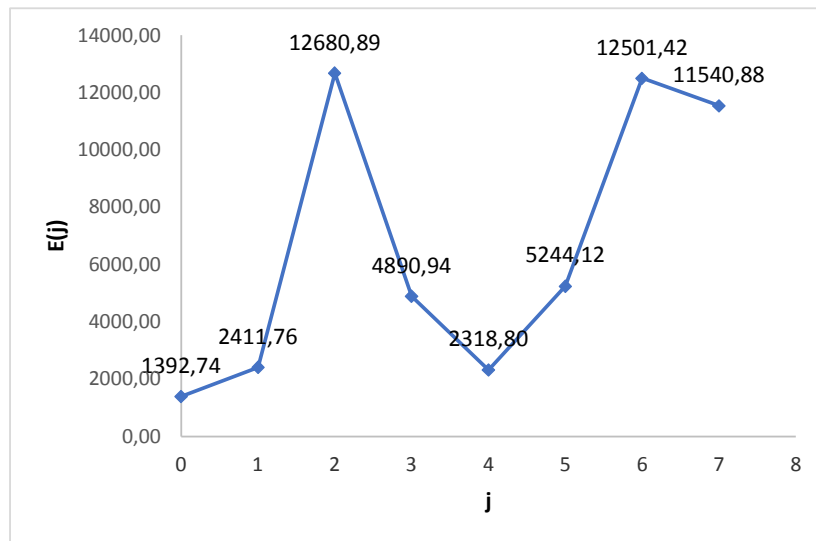
Adicionalmente, em um primeiro estágio, ajustamos uma parábola a $\log(x_t)$, para remover a tendência e estudar os resíduos, que resultou em $y_t = \log x_t - (-0,749 + 0,11705t - 0,000340t^2)$. (10)

Note que y_t pode ser interpretada como uma série filtrada de x_t . Desse modo, seja F qualquer função invertível (filtro) que transforma x_t em y_t (ou seja, $y_t = F(x_t)$), com a finalidade de eliminar a tendência da série. A fim de analisar o escalograma, vamos considerar uma expansão em ondaletas de Haar de y_t .

Para fazer a decomposição, o tamanho da série utilizada deve ser uma potência de 2. Desse modo, para considerarmos um tamanho $T = 256 = 2^8$, é necessário completar os dados com mais 17 pontos. Nesse trabalho, isso é feito acrescentando uns no início dos dados originais. A série $\log(x_t)$ está apresentada na Figura 4.

Os coeficientes de ondaletas de Haar, e o escalograma em cada nível, foram calculados de acordo com as equações (3), (4) e (5). O escalograma é apresentado na Figura 5.

Figura 5 - Escalograma da série y_t no estado de São Paulo, usando Haar com $J=8$



Pela Figura 5, há um pico no nível $j = 2$. Propomos dividir a decomposição de ondaletas d da série y_t em duas novas decomposições, denotadas por $d_{j,k}^{(1)}$ e $d_{j,k}^{(2)}$, da seguinte maneira:

$$d_{j,k}^{(1)} = d_{j,k}, \text{ para } j = 0, 1, 2, 3, 4 \text{ e } d_{j,k}^{(1)} = 0, \text{ para } j = 5, 6, 7;$$

$$d_{j,k}^{(2)} = 0, \text{ para } j = 0, 1, 2, 3, 4 \text{ e } d_{j,k}^{(2)} = d_{j,k}, \text{ para } j = 5, 6, 7.$$

A fim de separar as componentes de tendência e ruído, aplicamos a transformada discreta inversa de ondaletas para $d^{(1)}$ e $d^{(2)}$, obtendo as séries das duas componentes separadas $y_t^{(1)}$ e $y_t^{(2)}$:

$$y_t^{(1)} = c_{0,0}^{(1)} \Phi_{0,0}(t/T) + \sum_{j=0}^6 \sum_{k=0}^{2^j-1} d_{j,k}^{(1)} \psi_{j,k}(t/T), \quad (11)$$

$$y_t^{(2)} = \sum_{j=0}^6 \sum_{k=0}^{2^j-1} d_{j,k}^{(2)} \psi_{j,k}(t/T). \quad (12)$$

As componentes da série original $\{x_t\}$ são extraídas usando a transformada de ondaleta inversa. Essa série pode ser decomposta da forma

$$x_t = x_t^{(1)} + x_t^{(2)} + e_t. \quad (13)$$

A primeira componente, $x_t^{(1)}$, que captura a tendência da série original, é representada por

$$x_t^{(1)} = F^{-1}(y_t^{(1)}). \quad (14)$$

Uma vez que

$$e_t = x_t - F^{-1}(y_t^{(1)} + y_t^{(2)}), \quad (15)$$

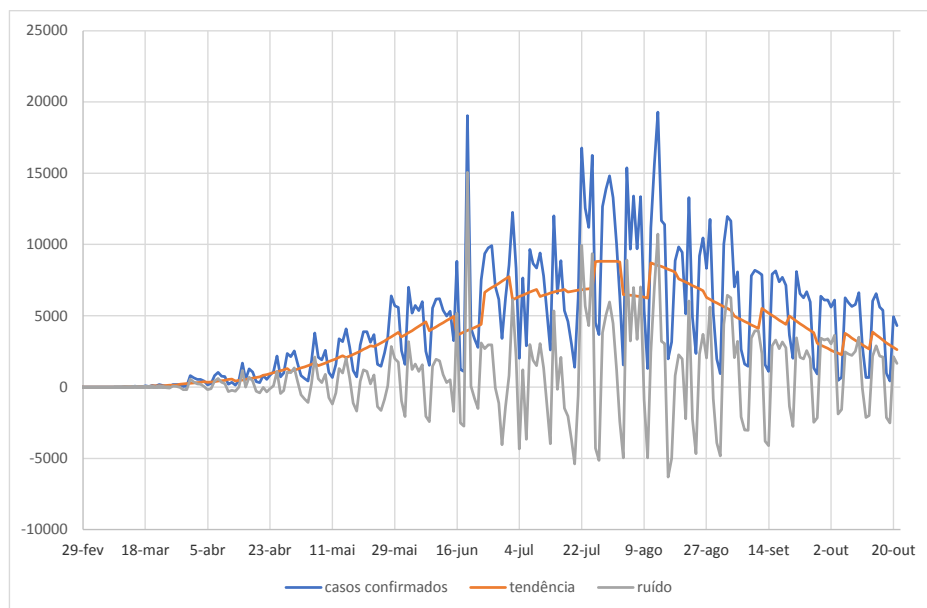
finalmente a segunda componente é obtida fazendo

$$x_t^{(2)} = x_t - x_t^{(1)} - e_t, \quad (16)$$

que captura o ruído da série original.

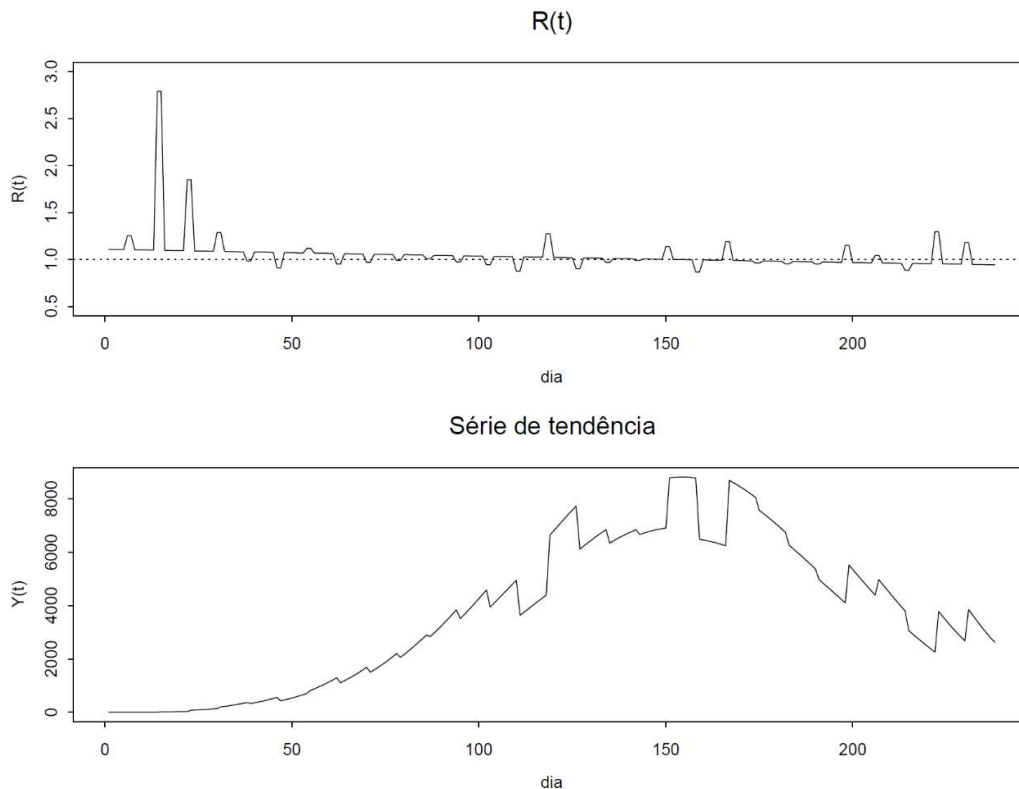
Série do número de novos casos no estado de São Paulo (em azul) e suas componentes de tendência (em vermelho) e ruído (em cinza) são apresentadas na Figura 6.

Figura 6 - Série do número de novos casos no estado de São Paulo (em azul) e suas componentes de tendência (em vermelho) e ruído (em cinza)



Para estimar $R(t)$, utilizamos (6) e (7) na série $\{x_t^{(1)}\}$ usando ondaletas de Haar com $J = 7$. As estimativas de $R(t)$ podem ser observadas na Figura 7. Aqui valem as mesmas observações feitas para o primeiro método. Notamos que $R(t)$ apresenta um comportamento condizente ao da tendência, oscilando ao redor de um no final (detalhes, ver Tabela 1).

Figura 7 - Estimativa do número de reprodução $R(t)$ para a série diária de novos casos no estado de São Paulo (gráfico superior), juntamente com a série de tendência (gráfico inferior), obtida através do escalograma. Resultados obtidos por meio do método 2 (Haar, com $J = 7$)



Na Tabela 1 temos as estimativas de $R(t)$ de 01/10/2020 a 21/10/2020. Notamos, obviamente, que diferentes métodos (e modelos) levam a resultados diferentes, mas indicam que os valores de $R(t)$ oscilam, ora abaixo, ora acima de 1, o menor valor da ordem de 0,94 (são necessárias 100 pessoas para infectar 94). Há uma tendência de queda, mas ainda será necessário observar algumas semanas para se confirmar o arrefecimento da epidemia no estado de São Paulo.

Tabela 1 - Novos casos diários e $R(t)$ calculado pelos 2 métodos no período de 01/10/2020 a 21/10/2020 no estado de São Paulo

Data	Novos casos	R(t) Método 1	R(t) Método 2
01/out	6097	0,9815	0,9576
02/out	5608	0,9815	0,9576
03/out	6096	1,0040	0,9563
04/out	473	0,9827	0,9563
05/out	677	0,9827	1,2971
06/out	6260	0,9951	1,2971
07/out	5916	0,9951	0,9537
08/out	5649	0,9956	0,9537
09/out	5786	0,9956	0,9524
10/out	6626	1,0469	0,9524
11/out	2844	1,0469	0,9511
12/out	684	1,0002	0,9511
13/out	685	0,9044	1,1816
14/out	6031	0,9044	1,1816
15/out	6553	1,0127	0,9486
16/out	5627	1,0127	0,9486
17/out	5394	0,9456	0,9473
18/out	968	0,9456	0,9473
19/out	437	1,0769	0,9460
20/out	4923	1,0769	0,9460
21/out	4299	0,9421	0,9450

4.2 CASOS CONFIRMADOS DE COVID-19 NO MUNICÍPIO DE SÃO PAULO

Os mesmos procedimentos foram aplicados para analisar os dados do município de São Paulo, durante o mesmo período.

A Figura 2 apresenta os dados diários de novos casos confirmados de Covid-19 no município de São Paulo, no período de 26/02/2020 e 21/10/2020, resultando em $T = 239$ observações, assim como nos dados do estado. Nosso objetivo é o mesmo, estimar o número de reprodução utilizando as duas abordagens já mencionadas (método 1 e método 2).

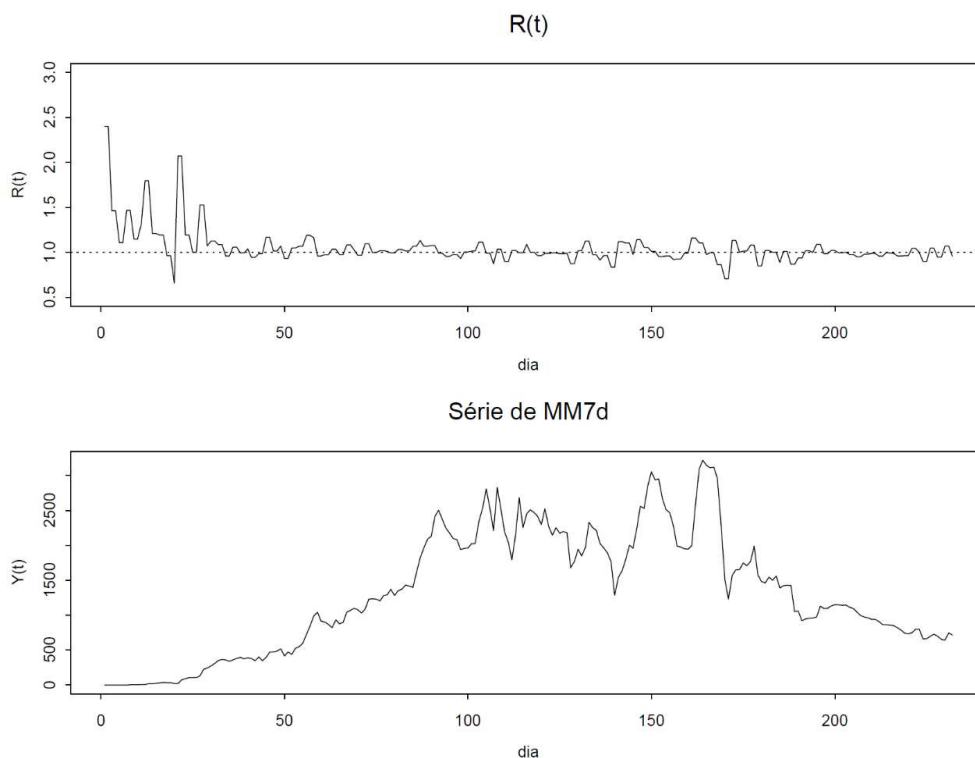
4.2.1 MÉTODO 1: MÉDIAS MÓVEIS SEGUIDA DE tvAR(1)

Assim como nos dados dos estados de São Paulo, podemos observar na Figura 2 que a curva de infectados apresenta irregular, com variações voláteis.

Inicialmente, aplicamos um modelo que parte de uma média móvel de sete dias para a série de novos casos no município. Em seguida, ajustamos o modelo tvAR(1). Novamente, usamos a ondaleta de Haar com $J = 7$ para

estimar $R(t)$. A estimativa é apresentada na Figura 8. A série suavizada mostra um comportamento similar ao do estado, mas em níveis mais baixos e uma queda a partir de 04 de setembro ($t = 185$). O gráfico de $R(t)$ acompanha essa tendência, e as quedas e subidas são mais acentuadas que no caso do estado. Valem observações similares sobre o comportamento no final do período.

Figura 8 - Estimativa do número de reprodução $R(t)$ para a série diária de novos casos no município de São Paulo (gráfico superior), juntamente com a série de médias móveis (gráfico inferior). Resultados obtidos por meio do método 1 (Haar, com $J = 7$)



4.2.2. MÉTODO 2: ESCALOGRAMA SEGUIDO DE tvAR(1)

Os procedimentos descritos na Seção 4.1.2 foram utilizados para a série de novos casos de COVID-19 no município de São Paulo.

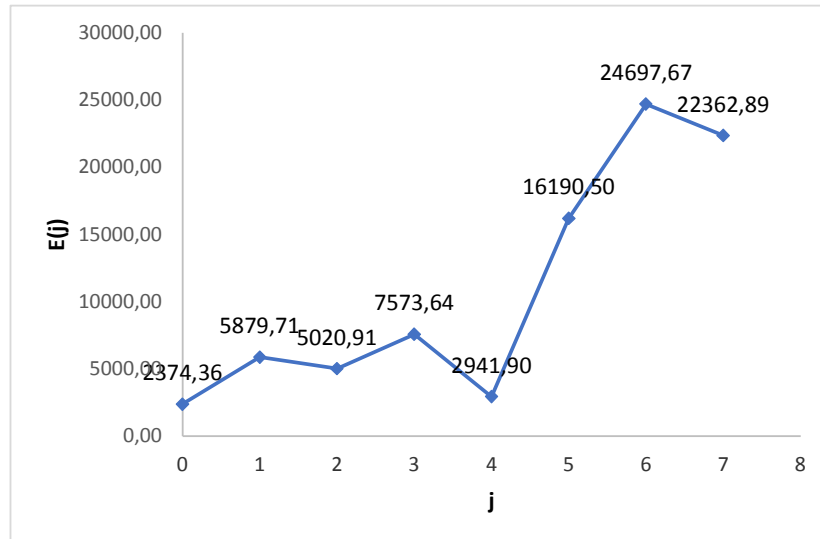
Ajustamos uma parábola para $\log(x_t)$, com o objetivo de remover a tendência e estudar os resíduos, o que resultou em

$$y_t = \log x_t - (-0,935 + 0,10973t - 0,000338t^2). \quad (17)$$

A fim de analisar o escalograma de y_t , escolhemos a ondaleta de Haar para decompor a série, pelos motivos já mencionados anteriormente. Uma vez que o tamanho da série deve ser uma potência de 2, consideramos $T = 256 = 2^8$. Os coeficientes de ondaletas de Haar e o escalograma, em cada nível de

resolução, foram calculados de acordo com as equações (3), (4) e (5). O escalograma está apresentada na Figura 9.

Figura 9 - Escalograma da série y_t no município de São Paulo, usando Haar com $J=8$



Pela Figura 9, há um pico no nível $j = 3$. Propomos dividir a decomposição de ondaletas \mathbf{d} da série y_t em duas novas decomposições, denotadas por $d_{j,k}^{(1)}$ e $d_{j,k}^{(2)}$, da seguinte maneira:

$$d_{j,k}^{(1)} = d_{j,k}, \text{ para } j = 0, 1, 2, 3, 4 \text{ e } d_{j,k}^{(1)} = 0, \text{ para } j = 5, 6 \text{ e } 7;$$

$$d_{j,k}^{(2)} = 0, \text{ para } j = 0, 1, 2, 3, 4 \text{ e } d_{j,k}^{(2)} = d_{j,k}, \text{ para } j = 5, 6 \text{ e } 7.$$

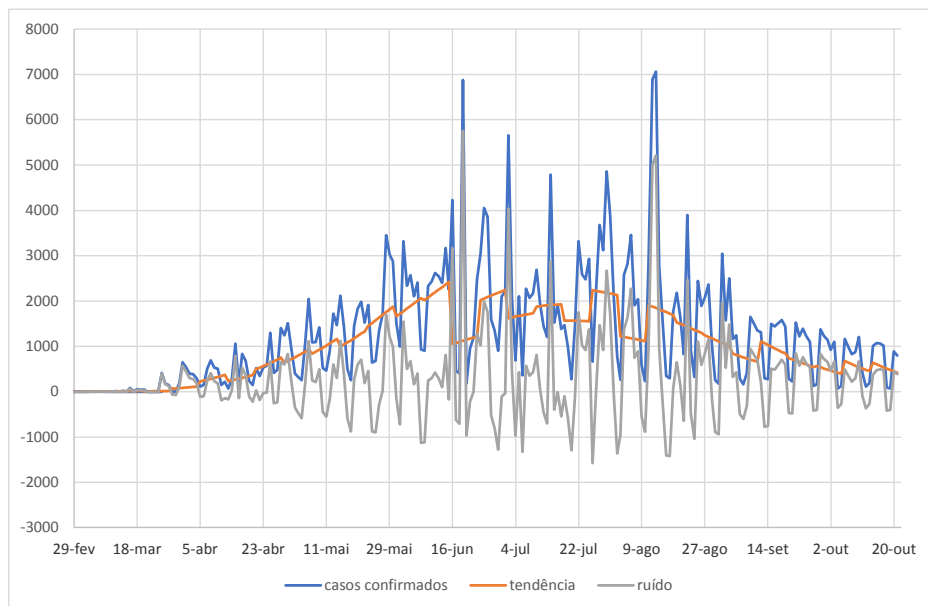
A fim de separar as componentes de tendência e ruído, aplicamos a transformada discreta inversa de ondaletas para $\mathbf{d}^{(1)}$ e $\mathbf{d}^{(2)}$, obtendo as séries das duas componentes separadas $y_t^{(1)}$ e $y_t^{(2)}$.

Como feito para o caso dos dados do estado de São Paulo, as componentes da série original x_t são extraídas usando a inversa do filtro F , utilizado na obtenção de y_t . Essa série pode ser decomposta da forma

$$x_t = x_t^{(1)} + x_t^{(2)} + \epsilon_t. \quad (18)$$

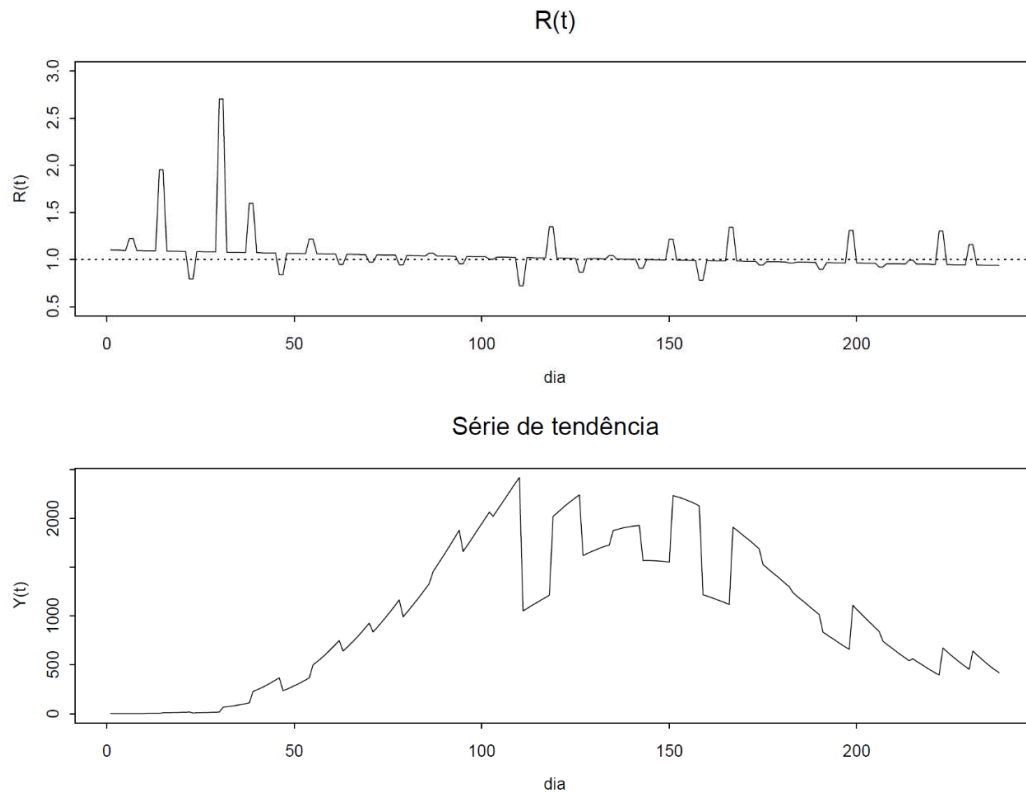
A primeira componente, $x_t^{(1)}$, que captura a tendência da série original, e a segunda componente, $x_t^{(2)}$, que captura o ruído da série original, são obtidas através das equações (14), (15) e (16). Essas séries estão apresentadas na Figura 10.

Figura 10 - Componentes de tendência (em vermelho) e ruído (em cinza) para a série de novos casos de COVID-19 do município de São Paulo (em azul)



Para estimar o número de reprodução $R(t)$, aplicamos as equações (6) e (7) a $x_t^{(1)}$, de modo a obter a estimativa de $R(t)$, apresentada na Figura 11. A interpretação de $R(t)$ é similar a feita pelo método anterior.

Figura 11 - Estimativa do número de reprodução $R(t)$ para a série diária de novos casos no município de São Paulo (gráfico superior), juntamente com a série de tendência (gráfico inferior), obtida através do escalograma. Resultados obtidos por meio do método 2 (Haar, com $J = 7$)



Na Tabela 2 temos os 21 últimos valores de $R(t)$ estimados pelos dois métodos. O comportamento da série é similar àquela do estado. As conclusões são semelhantes, e o efetivo decréscimo da epidemia vai depender das próximas semanas. Ainda não há evidências de que a epidemia está contida. Para isso, a série estimada de $R(t)$ deverá se manter abaixo de um, com valores menores do que os atuais, por algumas semanas.

Tabela 2 - Novos casos diários e $R(t)$ calculado pelos 2 métodos no período de 01/10/2020 a 21/10/2020 no município de São Paulo

Data	Novos casos	$R(t)$ Método 1	$R(t)$ Método 2
01/out	1143	0,9585	0,9576
02/out	925	0,9585	0,9576
03/out	1099	0,9992	0,9563
04/out	67	0,9925	0,9563
05/out	124	0,9925	1,2971
06/out	1168	0,9622	1,2971
07/out	990	0,9622	0,9537
08/out	822	0,9644	0,9537
09/out	872	0,9644	0,9524
10/out	1210	1,0447	0,9524
11/out	418	1,0447	0,9511
12/out	116	0,9986	0,9511
13/out	188	0,8979	1,1816
14/out	1014	0,8979	1,1816
15/out	1075	1,0467	0,9486
16/out	1066	1,0467	0,9486
17/out	1015	0,9481	0,9473
18/out	92	0,9481	0,9473
19/out	69	1,0714	0,9460
20/out	888	1,0714	0,9460
21/out	802	0,9594	0,9450

5. CONCLUSÕES

O objetivo do trabalho é estimar o número de reprodução $R(t)$ para as séries de infectados no município e no estado de São Paulo. Para isso, consideramos duas metodologias: (i) um modelo autorregressivo com coeficiente variando no tempo aplicado à série suavizada por médias móveis de sete dias; e (ii) aplicando uma decomposição das séries em tendência e ruído, de forma a extrair essas componentes usando o procedimento de Ariño et al. (2004). Esse procedimento não foi explicitado, e o leitor pode consultar a referência supracitada para mais detalhes.

As duas metodologias apresentaram comportamentos similares para $R(t)$ e sugerem que as séries de infectados atingiram um platô no qual os números ainda são grandes, tanto para o estado como para o município, e indicam que a epidemia ainda não se encontra contida. A série estimada de $R(t)$ na cidade de São Paulo mostra um declínio, embora os dados disponíveis não permitam concluir que haverá uma queda mais pronunciada nos próximos dias ou semanas. Os dados e códigos deste trabalho podem ser obtidos dos autores.

6. REFERÊNCIAS BIBLIOGRÁFICAS

Ariño, M. A., Morettin, P. A. & Vidakovic, B. (2004). On wavelet scalograms and their applications in economic time series. *Brazilian Journal of Probability and Statistics*, 18, 37–51.

Boletim de Epidemiologia Matemática Número 1 (2020), FGV EMAP.

Chiann, C. & Morettin, P. A. (2005). Time-domain nonlinear estimation of time-Varying linear systems. *Journal of Nonparametric Statistics*, bf 17, 365–383.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia, SIAM.

Flaxman, S. et al. (2020). Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. Report 13, IC.

Morettin, P. A. Pinheiro, A. S. and Vidakovic, B. (2017). *Wavelets in Functional Data Analysis*. Springer Briefs in Mathematics.

AGRADECIMENTOS e COLABORAÇÕES

Os autores agradecem o suporte parcial da FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), por meio do Projeto Temático 2018/04654-9.

EVOLUÇÃO DO NÚMERO DE TESTES PARA COVID-19 E RECOMENDAÇÕES PARA O NÚMERO MÍNIMO DE TESTES EM DIFERENTES CENÁRIOS. ESTUDO DE CASO: SERGIPE - BRASIL

Thiago de Jesus dos Santos

thiago.dejesus0705@gmail.com

Universidade Federal de Sergipe

Jesy Karolayne Sales dos Santos

jesy.sales.comunic@gmail.com

Universidade Federal de Sergipe

Carlos Raphael Araújo Daniel

raphael_crad@yahoo.com.br

Universidade Federal de Sergipe

Rita de Cássia de Lima Idalino

rita@ufpi.edu.br

Universidade Federal do Piauí

Resumo: A subnotificação dificulta o combate ao avanço da pandemia da COVID-19. Uma das primeiras etapas para evitar que o vírus se espalhe é identificar, através dos testes, qual a parcela contaminada da população. Realizando testes suficientes, é possível isolar os casos confirmados e rastrear os contatos próximos, portanto é interessante conhecer um número mínimo de testes a serem realizados de modo a conseguir monitorar adequadamente a situação. A partir do número de óbitos foi possível obter uma aproximação para o número total de casos em Sergipe e propor uma quantidade mínima de testes semanais, considerando ainda os casos de outras síndromes respiratórias, de acordo com os municípios e levando em conta diferentes cenários de contaminação. Até o final de junho, o número de testes realizados no estado seria suficiente para aproximadamente 3% da população, mas independentemente da condição avaliada, as recomendações são pelo menos quatro vezes maiores que os números oficiais, e servem como direcionamento para as próximas semanas. Além da capital Aracaju, as cidades de Nossa Senhora do Socorro, São Cristóvão, Itabaiana e Lagarto necessitam dos maiores volumes de teste.

Palavras-chave: COVID-19; testes; Sergipe; SRAG; subnotificação

Abstract: Under-reporting makes it difficult to fight the advance of the COVID-19 pandemic. One of the first steps to prevent the virus from spreading is to identify, through testing, what is the

contaminated portion of the population. With enough tests, it is possible to isolate the confirmed cases and track their close contacts, so it is interesting to know a minimum number of tests to be carried out in order to properly monitor the situation. Based on the number of deaths, it was possible to estimate the total number of cases in Sergipe and to propose a minimum number of weekly tests, also considering the cases of other respiratory syndromes, according to the cities and taking into account different contamination scenarios. By the end of June, the number of tests carried out in the state would be sufficient for approximately 3% of the population, but regardless of the condition evaluated, the recommendations are at least four times higher than the official numbers, and serve as a guide for the coming weeks. In addition to the capital Aracaju, the cities of Nossa Senhora do Socorro, São Cristóvão, Itabaiana and Lagarto require the largest amount of tests.

Keywords: COVID-19; testing, Sergipe; SARS; under-reporting

1. INTRODUÇÃO

A COVID-19 (*Corona Virus Disease 2019*) é uma doença contagiosa pertencente à classe das SRAGs (Síndrome Respiratória Aguda Grave, do inglês *Severe Acute Respiratory Syndrome - SARS*) cujos sintomas demoram de 2 a 14 dias para se manifestar, podendo em alguns casos ser assintomática e, portanto, é provável que apresente um ritmo de contágio acelerado sem que seja facilmente detectada (World Health Organization, 2020a; Centers for Disease Control and Prevention, 2020). Os primeiros relatos sobre a doença ocorreram na China em dezembro de 2019, dando início à maior pandemia dos últimos 100 anos (Chen et al., 2020a; Huang et al., 2020).

Como ainda não existe vacina comprovadamente eficaz, uma estratégia recomendada é evitar o contato social tanto quanto possível para dificultar a transmissão de uma pessoa para outra. Um estudo realizado na Itália evidenciou que o isolamento pode diminuir a taxa de contágio, alcançando uma queda de 90% no número de sintomáticos em 10 dias após o isolamento total dos residentes de uma vila (Day, 2020).

A Organização Mundial de Saúde (2020b) sugere que a proporção de infecções verdadeiramente assintomáticas não está clara, mas é relativamente difícil de ser identificada e não parece ser o principal fator de transmissão. Devido à dificuldade em isolar toda a população, é importante identificar e priorizar o isolamento dos indivíduos acometidos pelo vírus, o que é alcançado através dos testes (Lai et al., 2020; Song et al., 2020).

A baixa realização de testes para identificar os casos de COVID-19, resulta na ausência de um indicador preciso sobre o estágio em que a pandemia se encontra. Impedindo a adesão de algumas medidas e dificultando o

planejamento, administração de recursos, insumos e equipamentos de saúde com base na necessidade territorial do nível de contágio. Reprimindo a tomada de decisão gerencial sobre as medidas de um cenário de isolamento social, lockdown, até mesmo de reabertura do comércio e de retorno às atividades (Fernandes, 2020; Li et al., 2020; Emanuel et al., 2020; Salathé et al., 2020).

Tendo em vista o alto grau de propagação do vírus, a testagem torna-se uma medida primordial para identificar pessoas infectadas, pois além do período que um portador leva para demonstrar os primeiros sintomas, ainda existe a realidade das pessoas assintomáticas, que contribuem positivamente para o aumento da transmissibilidade. Investigar massivamente a população é considerada uma ação de contenção e até mesmo de mitigação do número de futuros casos, uma vez que ao identificar um indivíduo infectado, deve-se direcioná-lo ou até forçá-lo ao distanciamento social, evitando ao máximo o contato com pessoas e, por conseguinte, reduzindo o número de contágios (Cohen & Kupferschmidt, 2020; Castro et al., 2020; Day, 2020; Rivett, 2020).

Kretzschmar et al. (2020) analisaram o impacto de diferentes estratégias para isolamento de casos confirmados e rastreamento dos contatos, avaliando quantitativamente como atrasos e mudanças na proporção de cobertura da população com testes em diferentes cenários influenciam no controle da transmissão da COVID-19, sugerindo que a utilização de aplicativos para otimizar o rastreamento poderia evitar até 80% das transmissões.

Cada país reagiu de uma maneira diferente à pandemia (Tanne, 2020). A Nova Zelândia, Eslováquia e Tailândia monitoraram bem a contaminação desde o início ou foram rápidos na investigação após o início do surto. Tornando-se capazes de reduzir a quantidade de casos confirmados, enquanto aumentavam a quantidade de testes. Já países que testavam relativamente pouco em relação à quantidade de casos do vírus, como o México, Reino Unido, Suécia e Brasil, diariamente relataram números altos de casos confirmados e óbitos (Ritchie et al., 2020a).

O grande investimento em testes e em medidas organizadas para avaliação da população com o intuito de rastrear massivamente os infectados foi a chave do sucesso na Coreia do Sul. Os Estados Unidos iniciaram tardiamente os testes em comparação com a Coreia do Sul, e quando o número de testes começou a crescer superando outros países, os casos confirmados também aumentaram, a ponto de chegar a um número 50 vezes maior do que o pico da Coreia (Cohen & Kupferschmidt, 2020; Hasell, 2020).

No cenário brasileiro, o primeiro teste positivo para COVID-19 ocorreu em São Paulo, no dia 26 de fevereiro de 2020. Após cinco dias, outro caso foi confirmado e em apenas onze dias haviam 25 testes positivos (Macedo et al., 2020). Após a evolução para o contágio comunitário, a quantidade de testes para pessoas com suspeitas da doença, nas unidades do Sistema Único de Saúde (SUS), tornou-se insuficiente. Assim, a pandemia expõe as desigualdades do SUS, sobretudo da sua limitação de produção e de realização de testes que identifiquem a doença (Oliveira et al., 2020).

Até o dia 11 de julho, que marcou o final da 28ª semana epidemiológica, o Brasil tinha um total de mais de 1,8 milhão de casos confirmados e mais 71 mil óbitos registrados de acordo com os números oficiais (Ministério Da Saúde, 2020a), que poderiam ter sido reduzidos através de uma testagem massiva e contenção dos infectados, além de uma melhor administração dos recursos da saúde (Day, 2020; Rivett, 2020).

O Brasil era o país com o segundo maior número de casos confirmados de SARS-CoV-2 ao final da 28ª semana, atrás tão somente dos Estados Unidos que tinha mais 3,3 milhões de casos e mais de 175 mil óbitos. Entre os estados brasileiros com maior número de infectados estava São Paulo com mais de 370 mil casos e mais de 17 mil óbitos, resultando em uma taxa de letalidade de 4,8%, seguido do Ceará que registrou mais de 136 mil casos e 6,9 mil óbitos, constituindo uma taxa de letalidade de 5,06%. Em um Ranking de casos confirmados. Sergipe estava na 18ª colocação entre os 27 estados (Fundação Sead, 2020; Comitê Científico de Combate ao Coronavírus, 2020).

A taxa de letalidade do vírus no Brasil era de 3,9% até 11 de julho, contudo havia estados que estavam muito acima dessa taxa, sobretudo o Rio de Janeiro com uma taxa de 8,69%, Pernambuco que apresentava 7,7% e Ceará com 5,1%. Além disso, tinha estados com uma taxa de letalidade muito menor como Santa Catarina com 1,2%, e Mato Grosso do Sul com 1,2%. Até então, Sergipe tinha uma taxa de letalidade de 2,6% (Comitê Científico de Combate ao Coronavírus, 2020).

O primeiro caso confirmado em Sergipe foi no dia 14 de março de 2020 e os primeiros óbitos foram registrados dia 02 de abril na capital do estado. Em três meses já haviam sido confirmados mais 13 mil casos e mais de 330 mortes. Constituindo uma taxa de letalidade anteriormente citada e uma taxa de mortalidade de 14,4 por 100 mil habitantes. Sergipe aumentou a quantidade de testagem no dia 18 de abril, entretanto a diferença da quantidade de testes só foi expressiva a partir do dia 07 de maio (Observatório de Sergipe, 2020).

Diversas dificuldades com o registro e controle de casos de COVID-19 vem sendo enfrentadas em Sergipe devido a várias peculiaridades, como a economia local, proporção da população atendida pelo sistema público deficitário, densidade demográfica (Freitas & Sandes, 2020) e a quantidade de testes realizados que dificultam o rastreamento dos infectados e na estimação da dimensão da doença no estado.

Até o fim do mês de junho, Sergipe testou menos de 60.000 pessoas, sendo uma média diária de 639 testes, e com uma proporção de rastreamento de infectados de 54% dos testes. Até o final da 28ª semana epidemiológica, o município de Aracaju era o que apresentava maior incidência de casos, sendo em números absolutos mais de 20,5 mil casos, um total de 3129,3 casos por 100.000 habitantes, formando um percentual de 15,65% da quantidade de casos do estado. É importante destacar que a maioria das cidades de Sergipe tinham menos de 400 casos confirmados de COVID-19, algumas delas como Pedra Mole e Cumbe tinham menos de 15 casos confirmados em seu território. As maiores incidências de casos além de Aracaju, foram registradas em Nossa senhora do Socorro (mais de 2,1 mil), Itabaiana (mais de 1,8 mil), São Cristóvão (mais de 1,2 mil), Estância (mais de 1,1 mil), Lagarto (mais de 900), Barra dos coqueiros (mais de 700), Propriá (mais de 600) e Nossa senhora da Glória (mais de 480) (Secretaria de Estado da Saúde, 2020a).

Nos locais em que o número de testes é muito reduzido, surge o problema da subnotificação, ou seja, a quantidade de indivíduos acometidos pela doença é desconhecida e superior aos números oficiais, porque os testes realizados são insuficientes para contabilizar de maneira adequada e permitir estimativas com certo grau de confiabilidade para o total de casos (Lau et al., 2020; Salathé et al., 2020). O número de óbitos também está sujeito à subnotificação, pois nem sempre a causa da morte é corretamente identificada ou pode demorar a ser esclarecida (Oliveira, 1997; Silva & Paes, 2018), porém o número de óbitos é proporcionalmente muito inferior ao de casos e, se o paciente foi a óbito, então a doença estava em um estágio mais avançado, possivelmente necessitando de maiores cuidados e apresentando os sintomas característicos, portanto é menos provável que permaneça não detectada (Baud et al., 2020; Lachmann et al., 2020).

No entanto, é importante destacar que a COVID-19 não é a única SRAG e pode ser erroneamente diagnosticada como outra síndrome respiratória por apresentar alguns dos mesmos sintomas. Assim, os pacientes com qualquer uma das outras SRAGs também necessitam passar pelos mesmos testes e,

ainda que comparativamente os números não sejam tão expressivos, os casos de outras síndromes respiratórias não podem ser ignorados na distribuição dos testes (Holmes, 2003; Dias et al., 2020).

Tendo em vista a importância da testagem, surge a necessidade de avaliar se o número de testes que vem sendo realizado é suficiente para conseguir acompanhar o avanço da pandemia. Cleavelly et al. (2020) apontam a abordagem estratificada para os testes como a melhor estratégia para reduzir a quantidade de contágios, de modo que os mesmos estariam mais concentrados nos locais com a taxa de reprodução maior. Jha et al. (2020) sugeriram uma forma de estimar o número mínimo de testes recomendado para cada um dos estados dos EUA, com base no número de óbitos e na ocorrência de outros problemas respiratórios.

O presente estudo propõe uma adaptação da abordagem de Jha et al. (2020) como estudo de caso para o cenário epidemiológico de Sergipe, possibilitando uma avaliação da quantidade de testes realizados nas últimas semanas e sugerindo um modelo para recomendar o número mínimo de testes que deveriam ser realizados na semana seguinte, com base na aceleração exponencial que é característica da curva de contágio. Por fim, o número de testes é particionado por município, de acordo com as respectivas populações e pelo quanto cada município já foi atingido pelo vírus, considerando número de óbitos e casos confirmados, garantindo que, mesmo que houvesse locais em que possivelmente a contaminação ainda não tivesse chegado, eles receberiam também uma parcela de testes para contemplar as outras gripes e agir preventivamente com relação à COVID-19.

2. MATERIAIS E MÉTODOS

Nesse estudo, o delineamento epidemiológico foi um processo observacional analítico e transversal em que o surgimento dos dados sob análise ocorreu naturalmente como consequência da pandemia (Lima-Costa & Barreto, 2003; Nunes et al., 2013). Partindo da teoria proposta por Jha et al. (2020), foram considerados o número de óbitos (O_t) diários registrados para o estado de Sergipe e três valores para a taxa de letalidade (Let). Foi possível obter diferentes aproximações para o número de casos já existentes duas semanas antes (ND_{Ant}). Além da taxa de letalidade, a proporção de indivíduos sintomáticos (S_{nt}) também foi avaliada para diferentes cenários, permitindo estimar o número de indivíduos que teoricamente apresentariam os sintomas (NS_{Ant}) e, portanto, precisariam se submeter ao teste. Considerando ainda o

número de pessoas com as quais cada indivíduo doente teve contato próximo (Cnt), foi calculado um conjunto de cenários para o total de pessoas que deveriam ficar sob observação devido à suspeita de contaminação e serem testadas (NC_{Ant}). Por fim, é necessário considerar os casos de gripes e outros problemas respiratórios confirmados no estado (Grp_{Ant}), que apresentam sintomas semelhantes e que necessariamente deveriam passar pelo teste. O número mínimo de testes (NT_{Ant}) é dado pela soma dos indivíduos sintomáticos, aqueles com os quais tiveram contato, e os casos de outras doenças respiratórias. Assim, temos que:

$$ND_{Ant} = \frac{O_t}{Let} \quad (1)$$

$$NS_{Ant} = ND_{Ant} \times Snt \quad (2)$$

$$NC_{Ant} = NS_{Ant} \times Cnt \quad (3)$$

$$NT_{Ant} = NS_{Ant} + NC_{Ant} + Grp_{Ant} \quad (4)$$

Em que:

O_t corresponde ao número de óbitos no tempo t ;

ND_{Ant} é o número de pessoas doentes duas semanas antes do instante t ;

Let representa a taxa de letalidade, isto é, a proporção entre o total de óbitos e o de doentes;

NS_{Ant} é o número de pessoas sintomáticas duas semanas antes do tempo t em que foram observados os O_t óbitos;

Snt é a proporção de indivíduos que desenvolvem os sintomas da doença;

NC_{Ant} indica quantas pessoas tiveram contato com pessoas infectadas pelo vírus duas semanas antes do tempo de referência t ;

Cnt representa o número médio de pessoas com as quais alguém infectado teve contato;

NT_{Ant} é o número mínimo de testes que seria recomendado duas semanas antes do instante t ; e

Grp_{Ant} é o número de casos de outros problemas respiratórios duas semanas antes do momento t .

Com exceção do número de casos de outras gripes, todos os outros termos (número de casos, número de indivíduos sintomáticos, e pessoas com as quais houve contato próximo) são calculados sequencialmente a partir do número de óbitos com um atraso de duas semanas. Dessa forma, se em um determinado dia não houver nenhum óbito, o cálculo vai sugerir que

anteriormente não havia nenhum caso, o que não é uma conclusão realista. Não apenas isso, mas também o fato de a capital Aracaju concentrar a totalidade dos óbitos por um período, enquanto diversos municípios demoraram para registrar as primeiras ocorrências, faz com que as estimativas diárias para o número de indivíduos necessitando de testes tenha um comportamento instável, principalmente nos estágios iniciais da curva e nos municípios menores.

Como o cálculo depende do número de óbitos observado, quando esse número é muito pequeno, o resultado pode apresentar distorções e, no início da pandemia era comum que durante vários dias não houvesse nenhum registro de óbito, portanto inicialmente os dados de todo o estado foram agrupados por semana para que seja viável obter uma visão geral de quantos testes seriam necessários para serem realizados no estado em cada semana epidemiológica e quantos testes efetivamente estavam sendo realizados. De posse desse número, pode-se utilizar critérios para alocar os testes de uma maneira mais eficiente nos municípios e sugerir uma quantidade mínima para cada semana sem que esse número precise mudar constantemente, facilitando inclusive o planejamento e questões operacionais.

Foram considerados quatro cenários que partiram dos valores utilizados em Jha et al. (2020):

- um cenário base (cenário 0), com taxa de letalidade de 1%, proporção de sintomáticos de 75% e o número de indivíduos sob suspeita de contágio igual a 10 para cada caso confirmado, como em (Jha et al. , 2020);
- um pessimista (cenário A), no qual a taxa de letalidade é a mais alta (6%) entre os cenários considerados, a proporção de sintomáticos é maior (85%) e o número de contatos sob suspeita também é maior (12);
- um intermediário (cenário B), em que a taxa de letalidade não é tão alta (4%), e os valores para proporção de sintomáticos (75%) e número de contatos próximos necessitando atenção (10) foram mantidos iguais aos do estudo que serviu de base para esse trabalho (Jha et al. , 2020); e
- um otimista (cenário C), no qual a taxa de letalidade é menor (2%), a proporção de sintomáticos é também mais reduzida (65%) e o número de indivíduos que tiveram contato com o vírus é menor (8).

Os cenários avaliados partiram dos valores utilizados em Jha et al. (2020). Porém a taxa de letalidade (1%) considerada nos cálculos para estimar o número de testes recomendado nos EUA em maio, atualmente parece muito otimista, mesmo comparando com países que testaram muito e alcançaram bons

resultados como Alemanha (4,6%), Coreia do Sul (2,1%) e Taiwan (1,7%) (Comitê Científico de Combate ao Coronavírus, 2020). Pequenas variações nesses parâmetros podem provocar diferenças da ordem dos milhares no resultado final, portanto os valores sugeridos para proporção de sintomáticos e número de contatos não divergiram muito dos utilizados no trabalho original, que são os mesmos do cenário intermediário, como mostra a Tabela 1.

Ao decidir trabalhar com os dados semanais, foi feito um cálculo de médias móveis para a série semanal, pois no intervalo analisado houveram mudanças como a implementação do isolamento social que ocorreu de maneira distinta em cada município, a chegada de um maior número de testes (Secretaria de Estado Saúde, 2020b), e outras fontes de variação que dificultam a compreensão do comportamento da curva. Foi utilizada a média móvel de cada semana com as duas anteriores para suavizar essas oscilações, de forma que a curva resultante adquirisse um padrão de crescimento exponencial como é frequentemente observado para a COVID-19 (Piareto & Gomes, 2020).

Tabela 1- Parâmetros considerados na definição dos cenários para o número mínimo recomendado de testes para COVID-19

Cenário	Taxa de Letalidade (<i>Let</i>)	Proporção de Sintomáticos (<i>Snt</i>)	Número de Contatos (<i>Cnt</i>)
0 – Base (JHA et al.,2020)	1%	75%	10
A – Pessimista	6%	85%	12
B – Intermediário	4%	75%	10
C – Otimista	2%	65%	8

Inicialmente essa série foi comparada com o número de testes realizados no estado nas respectivas semanas para avaliar como seria o desempenho em cada um dos cenários hipotéticos. Em seguida, esses números foram tomados como base para obter projeções para as semanas seguintes. A quantidade reduzida de pontos restringe a utilização de técnicas de séries temporais (Antunes & Cardoso, 2015; Morettin & Tolo, 2018), mas o comportamento exponencial já conhecido e bem característico, permite construir um modelo de regressão linear simples em que a variável explicativa é o tempo e a variável resposta é o logaritmo do número mínimo de testes (Modelo 1). Não é recomendável utilizar modelos de regressão para tentar estimar valores que

extrapolam muito os limites estudados para as variáveis independentes (Cox, 1972; Gomes & Rocha, 2020), portanto o estudo se limitou a fazer apenas recomendações para as duas semanas imediatamente subsequentes, ou seja, conseguindo apenas reverter a defasagem do método e alcançando a semana atual.

Modelo 1:

$$\log(NT_t) = \beta_0 + \beta_1 t + \varepsilon_t \quad (5)$$

Em que:

NT_t é o número mínimo de testes no tempo t ;

β_0 corresponde ao intercepto do modelo;

β_1 representa o coeficiente angular do modelo; e

ε é uma variável aleatória com Normal com média zero no tempo t .

Como os resultados dos cálculos com base no número de óbitos sempre estão defasados em duas semanas, então na verdade o modelo foi utilizado para obter dois pontos correspondentes às duas semanas mais recentes já observadas, para as quais já estavam disponíveis os números de casos, óbitos e também das SRAGs durante a elaboração deste estudo (semanas epidemiológicas 26 e 27). Se por um lado, essa abordagem acaba exigindo mais do modelo ao tentar descrever o número de casos sem ter acesso ao número de óbitos que só estarão disponíveis duas semanas no futuro, por outro lado permite que as previsões das semanas 26 e 27 sejam imediatamente comparadas com o que foi realizado, dando uma perspectiva de como o modelo se comporta ao tentar descrever valores futuros.

Em seguida, o número mínimo de testes para o estado foi dividido entre os municípios levando em consideração o tamanho da população, o número de casos confirmados e de óbitos em cada um deles, numa tentativa de direcionar mais testes para os locais em que a situação está mais grave, porém garantindo que os municípios menores não ficariam desassistidos. Apesar de os casos relacionados a outras síndromes respiratórias entrarem no cálculo para a recomendação mínima de testes no estado, os números das SRAGs não foram separados de acordo com o município de origem, porém os casos de outras gripes são proporcionalmente muito menores e o peso atribuído à população de cada município é suficiente para assegurar que o número de testes atenderia também as síndromes respiratórias.

A partição dos testes foi feita de forma que 40% do total foi distribuído proporcionalmente de acordo com o tamanho da população de cada município, 30% dos testes foi dividido de acordo com o número acumulado de casos confirmados em cada cidade, e os últimos 30% foram direcionados de acordo com o número de óbitos acumulados nos respectivos municípios, como descrito a seguir:

$$\widehat{NT}_{t,h} = \widehat{NT}_t \times \left(0,4 \times \frac{Pop_h}{Pop} + 0,3 \times \frac{CA_h}{CA} + 0,3 \times \frac{OA_h}{OA} \right) \quad (6)$$

Em que:

\widehat{NT}_t é o número mínimo de testes recomendado para o estado no tempo t ;

$\widehat{NT}_{t,h}$ representa o número mínimo de testes recomendado para o município h no instante de tempo t ;

Pop é a população do estado;

Pop_h corresponde à população do município h ;

CA representa o número acumulado de casos confirmados no estado;

CA_h é o número acumulado de casos confirmados do município h ;

OA corresponde ao número acumulado de óbitos no estado; e

OA_h é o número acumulado de óbitos do município h ;

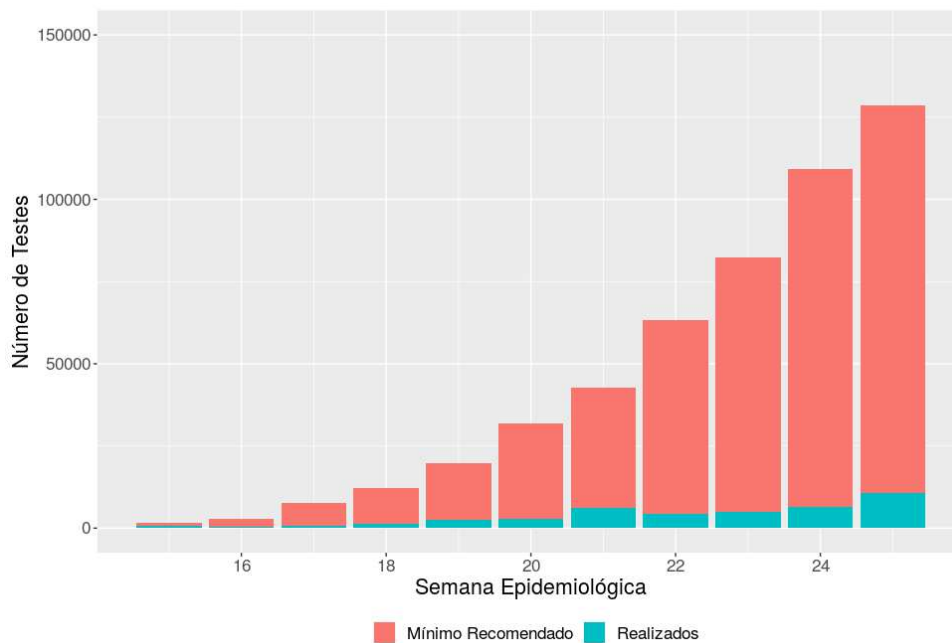
Por fim, foi analisado qual o percentual da população teria sido testado até a semana 27 supondo que fossem realizados os números sugeridos em cada cenário. Como eventualmente algumas pessoas precisariam submeter-se ao teste mais de uma vez, e sabendo que isso é mais provável à medida que a testagem ocorre mais lentamente no estado, então a expectativa seria de que a testagem conseguisse abranger todo o estado tão rápido quanto possível.

Os dados referentes aos óbitos semanais por município utilizados neste estudo são provenientes do repositório de dados mantido por Justen (2020), os números da SRAG foram obtidos no site do Ministério da Saúde (2020b) e as informações da quantidade de testes realizados no estado de Sergipe foram alcançadas mediante a Secretaria de Estado da Saúde (2020c). Todas as análises e gráficos foram feitos utilizando a linguagem de programação R sob o ambiente de desenvolvimento integrado R Studio, versão 1.1.463.

3. RESULTADOS E DISCUSSÃO

Considerando os valores do cenário base, os mesmos utilizados originalmente por Jha et al. (2020), nota-se na Figura 1 que a exigência é muito discrepante em comparação com a realidade que vem sendo registrada no estado de Sergipe. A aplicação dos pressupostos definidos por Jha et al. (2020) é difícil de implementar imediatamente na realidade da administração pública e privada da saúde sergipana, começando pela quantidade de testes alocados para o estado (Secretaria de Estado da Saúde, 2020b).

Figura 1- Comparação entre o número de testes realizados em Sergipe até a 25ª semana epidemiológica e o mínimo de testes semanais recomendados no Cenário Base (Taxa de Letalidade 1%, Contágio 75%, Contatos 10)



Fonte: Autoria própria

A diferença entre a quantidade mínima de testes recomendada no cenário base e a que vem sendo realizada ocorre devido à baixa taxa de letalidade. A princípio pode parecer contra intuitivo que uma taxa de letalidade menor leve a um elevado número de testes, porém como o cálculo envolve o número de óbitos dividido pelo total de casos, então a relação entre número de casos e a taxa de letalidade é inversamente proporcional. Portanto, como o número de óbitos foi mantido fixo nos cálculos, esse comportamento é esperado, indicando que para atingir um número de óbitos predeterminado num contexto de baixa letalidade seriam observados mais casos e, conseqüentemente, seriam recomendados mais testes. Em outras palavras, quando a taxa de letalidade é

menor, o número de casos é proporcionalmente maior em comparação com o de óbitos, tornando necessário um maior número de exames para que se tenha dimensão da quantidade de infectados.

Por outro lado, considerando o cenário pessimista (Figura 2A), com uma taxa de letalidade de 6%, supondo que a proporção de sintomáticos seja de 85%, e que os mesmos tenham contato com outras 12 pessoas, o número mínimo recomendado de testes fica abaixo de 30 mil na semana epidemiológica de número 25, ou seja, aproximadamente quatro vezes menor que o cenário base, mas ainda assim excedendo bastante o número de testes realizados.

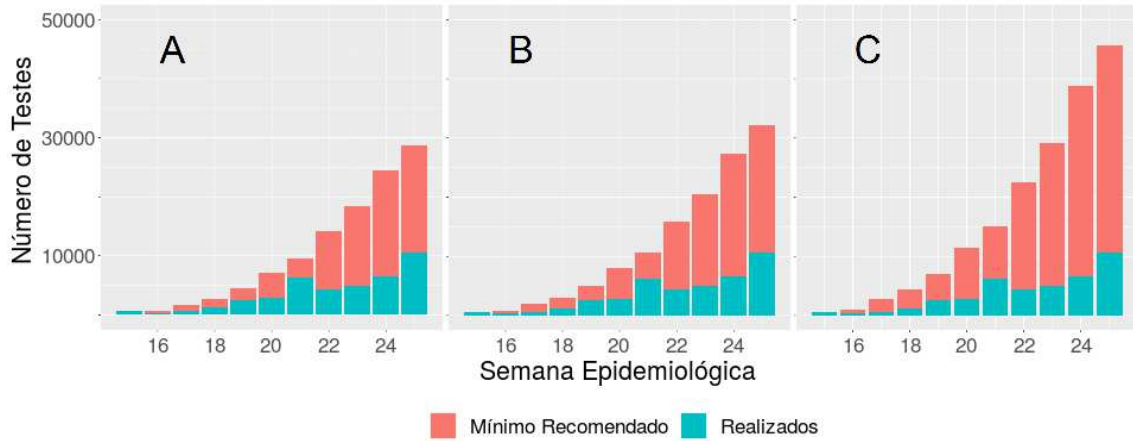
No caso intermediário (Figura 2B), a recomendação é de um número maior de testes em todas as semanas, e superando a exigência de pelo menos 30 mil testes na 25ª semana, pois, apesar de ter uma proporção de contágio menor e cada indivíduo ter contato com menos pessoas, a taxa de letalidade é menor, levando a uma maior demanda por testes.

Analisando o cenário otimista (Figura 2C), no qual a doença leva proporcionalmente menos pessoas ao óbito (taxa de letalidade de 2%), tem menor probabilidade para manifestar sintomas (proporção de sintomáticos de 65%), e o vírus tem menos oportunidades de contaminar a população (cada infectado tem contato com outras 8 pessoas), percebe-se como o número de testes é maior mas ainda não chega nem à metade do recomendado pelo cenário base.

Para avaliar a suposição de normalidade dos resíduos do modelo, foi feita a comparação com base na previsão do número de óbitos, já que é impossível observar o verdadeiro valor do número adequado de testes, e este último é calculado em função dos óbitos. O teste de Kolmogorov-Smirnov forneceu p-valor igual a 0,29 e o teste de Shapiro-Wilk obteve p-valor de 0,41, ambos não rejeitam a hipótese de normalidade dos resíduos.

Os resultados sugerem que o estado deve aumentar a quantidade de testes, corroborando com trabalhos de Kretzschmar et al. (2020) e Hellewell et al. (2020) que analisaram a propagação do vírus e avaliaram como um rastreamento eficiente de casos e contatos influencia na redução da velocidade de reprodução do vírus.

Figura 2 - Comparação entre o número de testes realizados em Sergipe até a 25ª semana epidemiológica e o mínimo de testes semanais recomendados para os cenários Pessimista (A), Intermediário (B) e Otimista (C)



Fonte: Autoria própria

Quanto à alocação dos testes entre os 75 municípios do estado, há uma discrepância na distribuição decorrente do tamanho da população entre os municípios e do estágio da pandemia em cada um deles. As Tabelas 2 e 3 apresentam, respectivamente, a 25ª e 27ª semanas epidemiológicas com os 5 municípios de maior e os 5 de menor número de testes recomendados, sendo que a capital Aracaju aparece com o maior número em todos os cenários para as duas semanas, como esperado, pois tem a maior população, foi o primeiro município a apresentar casos de COVID-19, e é responsável pela maioria dos óbitos e dos casos confirmados no estado.

Entre as cidades com maior número de testes recomendados, destaca-se Nossa Senhora do Socorro, São Cristóvão, Itabaiana e Lagarto. O município de Nossa Senhora do Socorro tem a segunda maior população e é vizinho à capital, apresentando também muitos casos e óbitos. Itabaiana e Lagarto não fazem fronteira com Aracaju, mas possuem a terceira e quarta maiores populações do estado, com aproximadamente 100 mil habitantes cada uma. Em contrapartida, São Cristóvão tem uma população um pouco menor mas, pela proximidade com a capital, foi sugerido um número maior de testes, pois o número de casos e óbitos confirmados no município foi maior. Esses resultados estão alinhados com a ideia de estratificação sugerida por Cleavelly et al. (2020), pois nos locais que tiveram maiores índices de propagação do vírus foi sugerido uma maior quantidade de diagnósticos.

Tabela 2- Distribuição de testes recomendados pelo modelo para os quatro cenários considerados na semana 25, para os 5 municípios de Sergipe com maior e menor alocação

Municípios	População	Semana 25			
		0	A	B	C
Aracaju	657.013	108.476	37.822	26.392	23.468
N. Sra. do Socorro	183.628	22.230	7.751	5.409	4.809
São Cristóvão	90.072	11.786	4.109	2.868	2.550
Itabaiana	95.427	10.379	3.619	2.525	2.245
Lagarto	104.408	8.039	2.803	1.956	1.739
São Francisco	3.724	190	66	46	41
Cumbe	3.987	190	66	46	41
Telha	3.227	156	54	38	34
Pedra Mole	3.261	151	53	37	33
Amparo de São Francisco	2.374	148	51	36	32
Sergipe	2.298.696	256.096	89.292	62.308	55.405

Fonte: Autoria própria

No outro extremo estão municípios cuja população é pequena e nos quais não foram confirmados muitos casos. Amparo de São Francisco seria o município com menor alocação de testes, com cerca de 2.300 habitantes, menos de 50 casos confirmados e nenhum óbito. Por serem municípios pequenos, poucas ocorrências já podem representar um sinal de alerta, por isso algumas das menores cidades do estado em termos populacionais não aparecem entre aquelas com menos testes recomendados, pois devido ao número de casos e óbitos, seria interessante direcionar mais testes para avaliar o avanço da pandemia nesses locais.

É também possível verificar na comparação entre a Tabela 2 e a Tabela 3 o aumento exponencial na demanda por testes entre as semanas epidemiológicas 25 e 27, sendo que a quantidade sugerida de testes na semana 27 chega a ser duas vezes maior do que o recomendado na semana 25. Esse crescimento ocorre devido à evolução exponencial da doença e, conseqüentemente, requer um maior número de exames que identifiquem os possíveis infectados. Além disso, os números sugeridos no cenário base ressaltam a sua inviabilidade, de tal forma que é quase o triplo do recomendado pelo Cenário A, mesmo sendo este o segundo mais exigente quanto à aplicação de exames.

Tabela 3- Distribuição de testes recomendados pelo modelo para os quatro cenários considerados na semana 27, para os 5 municípios de Sergipe com maior e menor alocação

Municípios	População	Semana 27			
		0	A	B	C
Aracaju	657.013	272.322	94.153	65.353	57.990
N. Sra. do Socorro	183.628	55.807	19.295	13.393	11.884
São Cristóvão	90.072	29.589	10.230	7.101	6.301
Itabaiana	95.427	26.055	9.008	6.253	5.548
Lagarto	104.408	20.181	6.977	4.843	4.297
São Francisco	3.724	477	165	114	101
Cumbe	3.987	476	165	114	101
Telha	3.227	391	135	94	83
Pedra Mole	3.261	380	131	91	81
Amparo de São Francisco	2.374	371	128	89	79
Sergipe	2.298.696	642.914	222.282	154.289	136.906

Fonte: Autoria própria

É importante ainda destacar que essas recomendações devem servir como orientação qualitativa para o direcionamento de recursos, permitindo uma avaliação de quais locais estão mais atrasados no que diz respeito aos testes, mas os números não devem ser tomados como metas definitivas e, se viável, a busca deve ser por testar toda a população. É mais uma questão de “por onde começar” do que “quando parar”.

Sobre o número acumulado de testes sugeridos em relação a cada cenário, observa-se na Tabela 4 que, se a sugestão do cenário 0 fosse implementada, o número de testes aplicados até a 27^a semana epidemiológica no estado seria de aproximadamente 1,7 milhões, ou seja, 76% da sua população. A natureza dinâmica da pandemia faz com que uma pessoa possa precisar realizar mais de um teste, seja para confirmar um resultado, ou porque o contágio pode ter ocorrido entre o momento de um teste e outro, portanto o monitoramento do contágio seria muito eficiente se os municípios tivessem condições de realizar tantos testes assim em tão pouco tempo. Porém, no início da pandemia não havia ainda tanta disponibilidade de testes e, mesmo com várias semanas de enfrentamento, a realidade observada parece ainda muito distante desse número.

Tabela 4 - Número acumulado de testes sugeridos até a 27ª semana epidemiológica por cenário, para os 5 municípios de Sergipe com maior e menor alocação

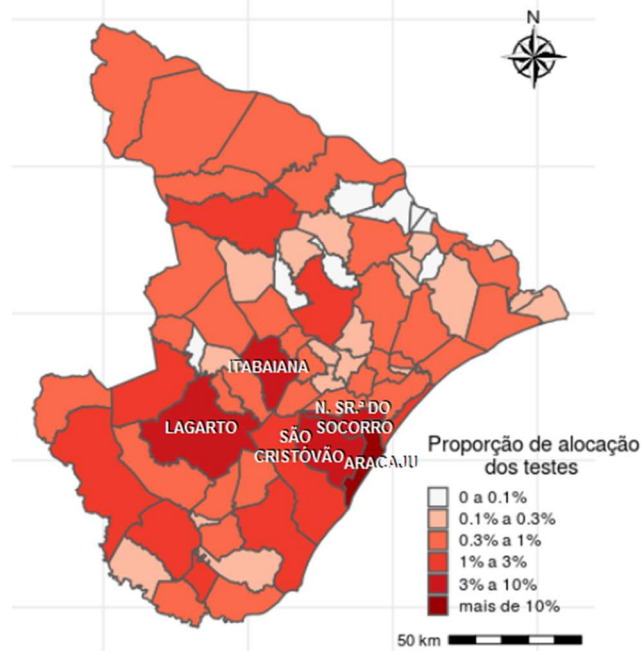
Municípios	População	Número Acumulado de Testes Sugeridos (Percentual Testado da População)			
		0	A	B	C
Aracaju	657.013	736.416 (112%)	256.428 (39%)	178.793 (27%)	158.939 (24%)
N. Sr. ^a do Socorro	183.628	150.914 (82%)	52.551 (29%)	36.640 (20%)	32.571 (18%)
São Cristóvão	90.072	80.014 (89%)	27.861 (31%)	19.427 (22%)	17.270 (19%)
Itabaiana	95.427	70.457 (74%)	24.535 (26%)	17.107 (18%)	15.206 (16%)
Lagarto	104.408	54.575 (52%)	19.002 (18%)	13.251 (13%)	11.778 (11%)
São Francisco	3.724	1.291 (35%)	450 (11%)	313 (8%)	278 (7%)
Cumbe	3.987	1.289 (32%)	450 (12%)	313 (8%)	279 (7%)
Telha	3.227	1.058 (33%)	369 (11%)	257 (8%)	229 (7%)
Pedra Mole	3.261	1.027 (31%)	356 (11%)	249 (8%)	220 (7%)
Amparo de São Francisco	2.374	1.002 (42%)	348 (15%)	245 (10%)	215 (9%)
Sergipe	2.298.696	1.738.580 (76%)	605.390 (26%)	422.109 (18%)	375.231 (16%)

Fonte: Autoria própria

Analisando os cenários mais factíveis (A, B e C), percebe-se que até a 27ª semana epidemiológica, o estado de Sergipe conseguiria testar entre 16% e 26% da sua população, sendo que os municípios mais atingidos teriam uma cobertura maior. O município de Aracaju, por exemplo, teria aproximadamente 40% no cenário B. Além disso, cidades como São Cristóvão e Nossa Senhora do Socorro tiveram uma grande quantidade de testes sugeridos, não somente pelo tamanho de sua população, mas também pelo grande número de casos e óbitos, possivelmente devido à proximidade com a capital, que pode ser considerada o epicentro da pandemia no estado e talvez tenha facilitado o acesso aos testes em um primeiro momento. A identificação de infectados é de grande valia para evitar o colapso do sistema de saúde sergipano, em acordo com o que sugere Aleta et al. (2020) identificando que a redução das restrições de isolamento, se não alinhadas com aplicações de testes rápidos e rastreamento de indivíduos infectados poderiam levar um sobrecarregamento do sistema de saúde.

A Figura 3 traz o mapa do estado e como ficaria a distribuição recomendada dos testes de acordo com os municípios. Independente do cenário considerado, uma fração do total de testes iria para cada município de modo que as cidades maiores e mais atingidas pela pandemia receberiam uma quantidade proporcionalmente maior, representada por uma tonalidade cada vez mais escura de vermelho. Aracaju concentra sozinha um percentual de mais de 10% dos testes de todo o estado, e identifica-se que apenas os municípios de Nossa Senhora do Socorro e São Cristóvão, vizinhos à capital, assim como Lagarto e Itabaiana, seriam os únicos em que esse percentual fica entre 3% e 10%.

Figura 3 - Distribuição dos testes por município de acordo com a alocação recomendada



Fonte: Autoria própria

Embora as recomendações do número mínimo de testes feitas possam ser consideradas elevadas em comparação com a quantidade que estado vem aplicando, Salathé et al. (2020) destacam que o custo pelo crescimento das testagens são altos no curto prazo, entretanto, a longo prazo, por meio do controle e ações preventivas tomadas, os custos econômicos e sociais serão reduzidos. O cenário C seria o que menos demandaria investimento imediato, porém é difícil estimar a economia que seria proporcionada em qualquer um dos cenários, em termos de custos hospitalares, força de trabalho parada, e vidas interrompidas, se o estado pudesse identificar antecipadamente, isolar e tratar os casos da doença sem que estivessem circulando e disseminando o vírus. Até

o fim do mês de junho, Sergipe tinha aplicado menos de 60.000 testes para COVID-19, o que corresponde a aproximadamente 2,61% da sua população.

4. CONCLUSÃO

O presente estudo teve como objetivo avaliar no estado de Sergipe a abordagem apresentada por Jha et al. (2020) para estimar a quantidade mínima de testes que deveriam ser realizados, de modo a ter uma dimensão mais precisa do alcance do contágio no território. Em todos os cenários considerados, o número de testes que vêm sendo realizados no estado não alcançam a quantidade mínima recomendada.

Com relação à recomendação para as semanas seguintes, que a abordagem original não permitia alcançar, a quantidade mínima de testes deve aumentar não apenas para tentar acompanhar o avanço da pandemia, mas também para recuperar o atraso. De fato, para seguir essa sugestão seria necessário um investimento muito maior em testes, pois o estado precisaria realizar em média entre 5 e 10 vezes mais testes do que tem efetuado por dia.

A alocação recomendada entre os municípios direciona um grande número de testes para a capital Aracaju, não somente pela população, mas também por concentrar a maior parte dos casos confirmados e óbitos no estado. Nossa Senhora do Socorro e São Cristóvão, que são vizinhas de Aracaju, também se encontram em um estágio avançado da pandemia e precisam de mais acompanhamento. Na sequência, Itabaiana e Lagarto compõem as cinco cidades com maior necessidade de testes. Todas as demais cidades teriam realizado pelo menos 215 testes ao final da 27ª semana epidemiológica, considerando o cenário menos exigente.

Países como Alemanha e Coreia do Sul começaram cedo a testar sua população e conseguiram conter o avanço da pandemia aplicando mais de 28 testes para cada mil pessoas. Por outro lado, em países que demoraram mais a reagir, como Estados Unidos e Itália, foi necessário um esforço muito maior, chegando a mais de 100 testes para cada mil habitantes. Para efeito de comparação, o Brasil, que também iniciou tarde os testes, realizou apenas 12 testes para cada mil pessoas (Ritchie et al., 2020b).

O aumento no número de exames para a COVID-19 é amplamente defendido pela comunidade epidemiológica, pois ajudaria nas medidas de isolamento social e possibilitaria a flexibilização em alguns locais. No entanto, vale ressaltar que diversos setores são atingidos direta e indiretamente pelos efeitos da pandemia e não tinham nenhum preparo para lidar com as questões

socioeconômicas que acometeram a população. Existem ainda muitos fatores desconexos nas políticas públicas em relação à gestão de recursos públicos, que refletem diretamente numa maior lentidão para adquirir e distribuir os testes e dimensionar de forma eficiente os investimentos no setor de saúde.

As sugestões apresentadas neste trabalho não consideram dificuldades como ausência de recursos e mão de obra para tornar viável a alocação dos testes, porém funcionam como um guia para apontar quais locais necessitam maior atenção. Mesmo que eventualmente esses números sejam alcançados, não significa que a situação esteja sob controle e serve apenas para permitir uma maior clareza sobre a real situação em que o estado se encontra, já que a subnotificação dificulta todas as estimativas, entre elas a verdadeira taxa de letalidade no estado. Uma testagem minimamente adequada é, na verdade, apenas um dos primeiros passos para um controle bem planejado das ações no combate à pandemia.

5. REFERÊNCIAS BIBLIOGRÁFICAS

Antunes, J. L. F., & Cardoso, M. R. A. (2015). Uso da análise de séries temporais em estudos epidemiológicos. *Epidemiologia e Serviços de Saúde*, 24, 565-576.

Baud, D., Qi, X., Nielsen-Saines, K., Musso, D., Pomar, L., & Favre, G. (2020). Real estimates of mortality following COVID-19 infection. *The Lancet infectious diseases*, 20, 773.

Castro, M. C., de Carvalho, L. R., Chin, T., Kahn, R., Franca, G. V., Macario, E. M., & de Oliveira, W. K. (2020). Demand for hospitalization services for COVID-19 patients in Brazil. *MedRxiv*.

Centers for Disease Control and Prevention. Symptoms of Coronavirus. (n.d.). Recuperado em 05 de Julho, 2020, de <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., ... & Yu, T. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet*, 395(10223), 507-513.

Cleevely, M., Susskind, D., Vines, D., Vines, L., & Wills, S. (2020). A workable strategy for COVID-19 testing: stratified periodic testing rather than universal random testing. *Oxford Review of Economic Policy*, 36(Supplement_1), S14-S37.

Cohen, J., & Kupferschmidt, K. (2020). Countries test tactics in 'war' against COVID-19. *Science* (New York, NY), v. 367, n. 6484, p. 1287, 2020.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187-202.

Comitê Científico de Combate ao Coronavírus. Covid-19 no Nordeste Brasileiro. (n.d.). Recuperado em 05 de Julho, 2020, de <https://app.powerbi.com/view?r=eyJrIjoiaNzRhMWE5ZmEtMjA0My00YjQ5LTlkMTgtZjh mYjFiMjAwN2VlliwidCI6ImFmZjU2ODdkLTlkZDUtNDJkMS04ZTMzLTljMTlyODk5NzN kOSJ9>

Day, M. (2020). Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. *BMJ: British Medical Journal (Online)*, 368.

Dias, G. H., Silva, C. D., Bezerra, M. R., & da Silva Peixoto, F. (2019). Análise da distribuição espacial da covid-19 e subnotificação de casos novos e óbitos no estado do Rio Grande do Norte, Brasil. *PENSAR GEOGRAFIA*, 3(2), 51-67.

Emanuel, E. J., Persad, G., Upshur, R., Thome, B., Parker, M., Glickman, A., ... & Phillips, J. P. (2020). Fair allocation of scarce medical resources in the time of Covid-19.

Fernandes, J. L. (2020). Covid-19 no Brasil: Aprendendo a Andar no Escuro sem Deixar Nada para Trás. *Arq Bras Cardiol*, 114(6), 988-991.

Fundação SEAD (2020). SP Contra o Coronavírus: Boletim Completo. Recuperado em 05 de Julho, 2020, de <https://www.seade.gov.br/coronavirus/>

Gomes, S. C., Monteiro, I. O., & Rocha, C. R. Modelagem Dinâmica da Covid-19 com Aplicação a Algumas Cidades Brasileiras.

Hasell, J. (2020). Testing early, testing late: Four countries' approaches to COVID-19 testing compared. Recuperado em 03 de Julho, 2020, de <https://ourworldindata.org/covid-testing-us-uk-korea-italy>

Hellewell, J., Abbott, S., Gimma, A., Bosse, N. I., Jarvis, C. I., Russell, T. W., ... & Flasche, S. (2020). Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*.

Holmes, K. V. (2003). SARS-associated coronavirus. *New England Journal of Medicine*, 348(20), 1948-1951.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., ... & Cheng, Z. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*, 395(10223), 497-506.

Jha, A. K., Jacobson, B., Friedhoff, S., & Tsai, T. (2020, Maio 25). HGHI and NPR publish new state testing targets. *Harvard Global Health Institute*. Recuperado em 15 de maio, 2020, de <https://globalepidemics.org/2020/05/07/hghi-projected-tests-needed-may15/>.

Jonnerby, J., Lazos, P., Lock, E., Marmolejo-Cossío, F., Ramsey, C. B., Shukla, M., & Sridhar, D. (2020). Maximising the Benefits of an Acutely Limited Number of COVID-19 Tests. *arXiv preprint arXiv:2004.13650*.

Justen, A. (2020). Brasil.io. Boletins informativos e casos do coronavírus por município por dia (Covid-19). Recuperado em 05 de Julho, 2020, de https://brasil.io/dataset/covid19/caso_full/

Kretzschmar, M. E., Rozhnova, G., Bootsma, M. C., van Boven, M., van de Wijgert, J. H., & Bonten, M. J. (2020). Impact of delays on effectiveness of contact tracing strategies

for COVID-19: a modelling study. *The Lancet Public Health*.

Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., & Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. *International journal of antimicrobial agents*, 105924.

Lachmann, A., Jagodnik, K., Giorgi, F. M., Ray, F. (2020). Correcting under-reported COVID-19 case numbers: estimating the true scale of the pandemic. *Preprint medRxiv*, 14.

Lau, H., Khosrawipour, T., Kocbach, P., Ichii, H., Bania, J., & Khosrawipour, V. (2020). Evaluating the massive underreporting and undertesting of COVID-19 cases in multiple global epicenters. *Pulmonology*.

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., & Shaman, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490), 489-493.

Lima-Costa, M. F., & Barreto, S. M. (2003). Tipos de estudos epidemiológicos: conceitos básicos e aplicações na área do envelhecimento. *Epidemiologia e serviços de saúde*, 12(4), 189-201.

Macedo, Y. M., Ornellas, J. L., & do Bomfim, H. F. (2020). COVID-19 NO BRASIL: o que se espera para população subalternizada?. *Revista Encantar-Educação, Cultura e Sociedade*, 2, 01-10.

Ministério da Saúde. (2020a). Painel de Casos de Doença pelo Coronavírus 2019 (Covid-19) no Brasil. Recuperado em 05 de Julho, 2020, de <https://covid.saude.gov.br/>.

Ministério da Saúde. (2020b). DATASUS. Recuperado em 05 de Julho, 2020, de <https://opendatasus.saude.gov.br/dataset/bd-srag-2020>

Morettin, P. A., & Toloi, C. M. (2018). *Análise de séries temporais: modelos lineares univariados*. Editora Blucher.

Nunes, L. N., Camey, S. A., Guimarães, L. S. P., Mancuso, A. C. B., & Hirakata, V. N. (2013). Os principais delineamentos na Epidemiologia. *Revista HCPA. Porto Alegre. Vol. 33, no. 2 (2013), p. 178-183*.

Observatório de Sergipe. (2020). Covid-19: Sergipe e Território Nacional (3a ed.). Recuperado em 21 de Junho, 2020, de <http://docs.observatorio.se.gov.br/wl/?id=QxchLBhjhKncqUYyel3Izhh1uYPdviHD>

Oliveira, H. D., & Pereira, I. P. D. A. (1997). Estatísticas de mortalidade e nascidos vivos: considerações sobre principais problemas. *Informe Epidemiológico do SUS*, 6(3), 15-19.

Oliveira, W. K. D., Duarte, E., França, G. V. A. D., & Garcia, L. P. (2020). Como o Brasil pode deter a COVID-19. *Epidemiologia e Serviços de Saúde*, 29, e2020044.

Platero, K., & Gomes, F. (2020). Números estatísticos e realidades: Uma proposta de reflexão sobre a pandemia de Covid-19 no Brasil. *DILEMAS: Revista de Estudos de Conflito e Controle Social-Reflexões na Pandemia*, 1-11.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rivett, L., Sridhar, S., Sparkes, D., Routledge, M., Jones, N. K., Forrest, S., ... & Torok, M. E. (2020). Screening of healthcare workers for SARS-CoV-2 highlights the role of asymptomatic carriage in COVID-19 transmission. *Elife*, *9*, e58728.

Ritchie, H., Ortiz-Ospina, E., Beltekian, D., Mathieu, E., Hasell, J., ... & Crawford, J. (2020a). Is the world making progress against the pandemic? We built the chart to answer this question. *Our World in Data*. Recuperado em 03 de Julho, 2020, de <https://ourworldindata.org/epi-curve-covid-19.T>

Ritchie, H., Ortiz-Ospina, E., Beltekian, D., Mathieu, E., Hasell, J., ... & Crawford, J. (2020b). Coronavirus (COVID-19) Testing. *Our World in Data*. Recuperado em 27 de Julho, 2020, de <https://ourworldindata.org/coronavirus-testing>

Salathé, M., Althaus, C. L., Neher, R., Stringhini, S., Hodcroft, E., Fellay, J., ... & Eckerle, I. (2020). COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. *Swiss medical weekly*, *150*(11-12), w20225.

Sandes, S., & Freitas, A. S. (2020). Modelo SIR com taxa de exposição para estudo da projeção de casos de COVID-19 no estado de Sergipe. doi: 10.1590/scielopreprints.311

Secretaria de Estado da Saúde (2020a). Boletim Sergipe Contra o Coronavírus para o Covid-19. Recuperado em 05 de Julho, 2020, de <https://todoscontraocorona.net.br/>

Secretaria de Estado da Saúde (2020b) [Decretos]. Sergipe Contra o Coronavírus. Recuperado em 16 de Julho, 2020, de <https://todoscontraocorona.net.br/decretos/>

Secretaria de Estado da Saúde (2020c, Julho 16). Governo de Sergipe Aumenta o número de Testagem para a Covid-19. Aracaju. Recuperado em 16 de Julho, 2020, de <https://www.saude.se.gov.br/?p=42249>

Silva, E. S. A., & Paes, N. A. (2018). Programa Bolsa Família e Mortalidade Infantil no Brasil: Revisão Integrativa. *HOLOS*, *34*(1), 201-211.

Song, F., Shi, N., Shan, F., Zhang, Z., Shen, J., Lu, H., ... & Shi, Y. (2020). Emerging 2019 novel coronavirus (2019-nCoV) pneumonia. *Radiology*, *295*(1), 210-217.

Tanne, J. H., Hayasaki, E., Zastrow, M., Pulla, P., Smith, P., & Rada, A. G. (2020). Covid-19: how doctors and healthcare systems are tackling coronavirus worldwide. *Bmj*, *368*.

World Health Organization (2020a). Coronavírus disease 2019 (Covid-19) Situation Report - 46. Recuperado em 16 de Julho, 2020, de <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf>

World Health Organization. (2020b). Report of the WHO-China joint mission on coronavirus disease 2019 (COVID-19). Recuperado em 23 de Outubro, 2020, de <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf>

APÊNDICE 1

A seguir, apresenta-se algumas tabelas contendo as estimativas para todos os municípios do estado de Sergipe.

Tabela 5- Distribuição de testes recomendados pelo modelo para os quatro cenários considerados na semana 25 para os municípios de Sergipe

Município	População	Cenários			
		A	B	C	D
Sergipe	2.298.696	256.096	89.292	62.308	55.405
Aracaju	657.013	108.476	37.822	26.392	23.468
Nossa Senhora do Socorro	183.628	22.230	7.751	5.409	4.809
São Cristóvão	90.072	11.786	4.109	2.868	2.550
Itabaiana	95.427	10.379	3.619	2.525	2.245
Lagarto	104.408	8.039	2.803	1.956	1.739
Estância	69.184	6.714	2.341	1.633	1.452
Umbaúba	25.294	4.052	1.413	986	877
Simão Dias	40.484	3.969	1.384	966	859
Tobias Barreto	52.191	3.704	1.292	901	801
Itabaianinha	41.928	3.625	1.264	882	784
Itaporanga d'Ajuda	34.356	3.569	1.244	868	772
Nossa Senhora da Glória	36.924	3.495	1.219	850	756
Barra dos Coqueiros	30.407	3.264	1.138	794	706
Nossa Senhora das Dores	26.629	2.749	958	669	595
Propriá	29.626	2.468	860	600	534
Maruim	17.213	2.359	822	574	510
Poço Redondo	34.775	2.157	752	525	467
Capela	34.213	2.126	741	517	460
Laranjeiras	29.826	2.120	739	516	459
Canindé de São Francisco	29.900	2.111	736	514	457
Areia Branca	18.542	2.033	709	495	440
Porto da Folha	28.596	2.029	708	494	439
Salgado	19.998	1.907	665	464	413
Boquim	26.816	1.878	655	457	406
Neópolis	18.719	1.767	616	430	382
Riachão do Dantas	19.805	1.696	591	413	367

Continua

Município	População	Cenários			
		A	B	C	D
Carira	22.082	1.560	544	380	337
Ribeirópolis	18.652	1.532	534	373	332
Poço Verde	23.728	1.498	522	364	324
Araúá	10.056	1.468	512	357	318
Indiaroba	17.957	1.460	509	355	316
Aquidabã	21.563	1.377	480	335	298
Santo Amaro das Brotas	12.102	1.348	470	328	292
Japaratuba	18.743	1.246	434	303	269
Carmópolis	16.634	1.211	422	295	262
Campo do Brito	18.109	1.164	406	283	252
Moita Bonita	11.335	1.113	388	271	241
Pacatuba	14.428	1.089	380	265	236
Frei Paulo	15.421	1.062	370	258	230
Malhador	12.618	1.056	368	257	229
Pirambu	9.280	977	341	238	211
Monte Alegre de Sergipe	15.031	955	333	232	207
Cristinápolis	17.874	916	319	223	198
Rosário do Catete	10.855	876	305	213	190
Nossa Senhora de Lourdes	6.483	835	291	203	181
Pinhão	6.576	833	290	203	180
São Domingos	11.137	811	283	197	175
Gararu	11.604	790	275	192	171
Santa Luzia do Itanhy	14.035	685	239	167	148
Tomar do Geru	13.536	657	229	160	142
Japoatã	13.434	623	217	151	135
Santana do São Francisco	7.780	608	200	140	124
Riachuelo	10.213	575	190	133	118
Siriri	8.893	546	172	120	107
Ilha das Flores	8.520	493	162	113	100
General Maynard	3.346	464	160	112	99
Cedro de São João	5.897	460	155	108	96
Pedrinhas	9.602	446	155	108	96
Graccho Cardoso	5.818	444	153	107	95

Continua

Município	População	Cenários			
		A	B	C	D
Nossa Senhora Aparecida	8.796	440	152	106	94
Muribeca	7.625	394	137	96	85
Brejo Grande	8.309	388	135	94	84
Divina Pastora	5.138	384	134	93	83
Macambira	6.919	350	129	90	80
Malhada dos Bois	3.682	307	122	85	76
Feira Nova	5.584	279	107	75	67
São Miguel do Aleixo	3.930	229	97	68	60
Itabi	4.903	224	80	56	50
Santa Rosa de Lima	3.913	198	78	55	49
Canhoba	4.008	191	66	46	41
São Francisco	3.724	190	66	46	41
Cumbe	3.987	190	66	46	41
Telha	3.227	156	54	38	34
Pedra Mole	3.261	151	53	37	33
Amparo de São Francisco	2.374	148	51	36	32

Fonte: Autoria própria

Tabela 6- Distribuição de testes recomendados pelo modelo para os quatro cenários considerados na semana 27 para os municípios de Sergipe

Município	População	Cenários			
		A	B	C	D
Sergipe	2.298.696	642.914	222.282	154.289	136.906
Aracaju	657.013	272.322	94.153	65.353	57.990
Nossa Senhora do Socorro	183.628	55.807	19.295	13.393	11.884
São Cristóvão	90.072	29.589	10.230	7.101	6.301
Itabaiana	95.427	26.055	9.008	6.253	5.548
Lagarto	104.408	20.181	6.977	4.843	4.297
Estância	69.184	16.855	5.827	4.045	3.589
Umbaúba	25.294	10.171	3.517	2.441	2.166
Simão Dias	40.484	9.965	3.445	2.391	2.122

Continua

Município	População	Cenários			
		A	B	C	D
Tobias Barreto	52.191	9.300	3.215	2.232	1.980
Itabaianinha	41.928	9.100	3.146	2.184	1.938
Itaporanga d'Ajuda	34.356	8.960	3.098	2.150	1.908
Nossa Senhora da Glória	36.924	8.774	3.033	2.106	1.868
Barra dos Coqueiros	30.407	8.195	2.833	1.967	1.745
Nossa Senhora das Dores	26.629	6.900	2.386	1.656	1.469
Propriá	29.626	6.195	2.142	1.487	1.319
Maruim	17.213	5.922	2.047	1.421	1.261
Poço Redondo	34.775	5.415	1.872	1.300	1.153
Capela	34.213	5.337	1.845	1.281	1.137
Laranjeiras	29.826	5.322	1.840	1.277	1.133
Canindé de São Francisco	29.900	5.300	1.833	1.272	1.129
Areia Branca	18.542	5.105	1.765	1.225	1.087
Porto da Folha	28.596	5.094	1.761	1.223	1.085
Salgado	19.998	4.788	1.655	1.149	1.020
Boquim	26.816	4.715	1.630	1.132	1.004
Neópolis	18.719	4.435	1.533	1.064	944
Riachão do Dantas	19.805	4.257	1.472	1.022	906
Carira	22.082	3.916	1.354	940	834
Ribeirópolis	18.652	3.847	1.330	923	819
Poço Verde	23.728	3.760	1.300	902	801
Araúá	10.056	3.686	1.275	885	785
Indiaroba	17.957	3.664	1.267	879	780
Aquidabã	21.563	3.457	1.195	830	736
Santo Amaro das Brotas	12.102	3.384	1.170	812	721
Japarutuba	18.743	3.127	1.081	750	666
Carmópolis	16.634	3.041	1.051	730	648
Campo do Brito	18.109	2.921	1.010	701	622
Moita Bonita	11.335	2.793	966	670	595
Pacatuba	14.428	2.734	945	656	582
Frei Paulo	15.421	2.665	922	640	568
Malhador	12.618	2.652	917	636	565
Pirambu	9.280	2.454	848	589	523
Monte Alegre de Sergipe	15.031	2.397	829	575	510

Continua

Município	População	Cenários			
		A	B	C	D
Cristinápolis	17.874	2.299	795	552	490
Rosário do Catete	10.855	2.200	760	528	468
Nossa Senhora de Lourdes	6.483	2.096	725	503	446
Pinhão	6.576	2.091	723	502	445
São Domingos	11.137	2.036	704	489	434
Gararu	11.604	1.984	686	476	422
Santa Luzia do Itanhy	14.035	1.720	595	413	366
Tomar do Geru	13.536	1.649	570	396	351
Japoatã	13.434	1.563	540	375	333
Santana do São Francisco	7.780	1.526	499	346	307
Riachuelo	10.213	1.442	474	329	292
Siriri	8.893	1.370	428	297	264
Ilha das Flores	8.520	1.238	403	279	248
General Maynard	3.346	1.165	399	277	246
Cedro de São João	5.897	1.154	387	269	238
Pedrinhas	9.602	1.119	386	268	238
Graccho Cardoso	5.818	1.116	382	265	235
Nossa Senhora Aparecida	8.796	1.104	378	262	233
Muribeca	7.625	988	342	237	210
Brejo Grande	8.309	975	337	234	208
Divina Pastora	5.138	965	333	231	205
Macambira	6.919	879	322	223	198
Malhada dos Bois	3.682	772	304	211	187
Feira Nova	5.584	700	267	185	164
São Miguel do Aleixo	3.930	575	242	168	149
Itabi	4.903	564	199	138	122
Santa Rosa de Lima	3.913	498	195	135	120
Canhoba	4.008	478	165	115	102
São Francisco	3.724	477	165	114	101
Cumbe	3.987	476	165	114	101
Telha	3.227	391	135	94	83
Pedra Mole	3.261	380	131	91	81
Amparo de São Francisco	2.374	371	128	89	79

Fonte: Autoria própria

Tabela 7 - Número acumulado de testes sugeridos até a 27ª semana epidemiológica por cenário para os municípios de Sergipe

Município	População	Cenários			
		A	B	C	D
Sergipe	2.298.696	1.738.580 (76%)	605.390 (26%)	422.109 (18%)	375.231 (16%)
Aracaju	657.013	736.416 (112%)	256.428 (39%)	178.793 (27%)	158.939 (24%)
Nossa Senhora do Socorro	183.628	150.914 (82%)	52.551 (29%)	36.640 (20%)	32.571 (18%)
São Cristóvão	90.072	80.014 (89%)	27.861 (31%)	19.427 (22%)	17.270 (19%)
Itabaiana	95.427	70.457 (74%)	24.535 (26%)	17.107 (18%)	15.206 (16%)
Lagarto	104.408	54.575 (52%)	19.002 (18%)	13.251 (13%)	11.778 (11%)
Estância	69.184	45.578 (66%)	15.871 (23%)	11.065 (16%)	9.836 (14%)
Umbaúba	25.294	27.506 (109%)	9.578 (38%)	6.678 (26%)	5.939 (23%)
Simão Dias	40.484	26.947 (67%)	9.383 (23%)	6.543 (16%)	5.816 (14%)
Tobias Barreto	52.191	25.148 (48%)	8.758 (17%)	6.105 (12%)	5.428 (10%)
Itabaianinha	41.928	24.608 (59%)	8.569 (20%)	5.973 (14%)	5.311 (13%)
Itaporanga d'Ajuda	34.356	24.230 (71%)	8.437 (25%)	5.882 (17%)	5.228 (15%)
Nossa Senhora da Glória	36.924	23.727 (64%)	8.263 (22%)	5.760 (16%)	5.121 (14%)
Barra dos Coqueiros	30.407	22.160 (73%)	7.717 (25%)	5.381 (18%)	4.784 (16%)
Nossa Senhora das Dores	26.629	18.661 (70%)	6.497 (24%)	4.530 (17%)	4.026 (15%)
Propriá	29.626	16.752 (57%)	5.832 (20%)	4.067 (14%)	3.615 (12%)
Maruim	17.213	16.014 (93%)	5.575 (32%)	3.888 (23%)	3.458 (20%)
Poço Redondo	34.775	14.642 (42%)	5.099 (15%)	3.558 (10%)	3.162 (9%)
Capela	34.213	14.434 (42%)	5.026 (15%)	3.506 (10%)	3.114 (9%)
Laranjeiras	29.826	14.391 (48%)	5.010 (17%)	3.495 (12%)	3.106 (10%)

Continua

Município	População	Cenários			
		A	B	C	D
Canindé de São Francisco	29.900	14.332 (48%)	4.991 (17%)	3.481 (12%)	3.095 (10%)
Areia Branca	18.542	13.804 (74%)	4.807 (26%)	3.352 (18%)	2.981 (16%)
Porto da Folha	28.596	13.774 (48%)	4.795 (17%)	3.346 (12%)	2.974 (10%)
Salgado	19.998	12.948 (65%)	4.506 (23%)	3.143 (16%)	2.794 (14%)
Boquim	26.816	12.750 (48%)	4.440 (17%)	3.096 (12%)	2.752 (10%)
Neópolis	18.719	11.993 (64%)	4.174 (22%)	2.911 (16%)	2.588 (14%)
Riachão do Dantas	19.805	11.512 (58%)	4.007 (20%)	2.795 (14%)	2.482 (13%)
Carira	22.082	10.591 (48%)	3.687 (17%)	2.570 (12%)	2.285 (10%)
Ribeirópolis	18.652	10.401 (56%)	3.624 (19%)	2.528 (14%)	2.246 (12%)
Poço Verde	23.728	10.170 (43%)	3.540 (15%)	2.468 (10%)	2.195 (9%)
Araúá	10.056	9.969 (99%)	3.471 (35%)	2.421 (24%)	2.151 (21%)
Indiaroba	17.957	9.909 (55%)	3.450 (19%)	2.405 (13%)	2.137 (12%)
Aquidabã	21.563	9.350 (43%)	3.255 (15%)	2.271 (11%)	2.019 (9%)
Santo Amaro das Brotas	12.102	9.152 (76%)	3.187 (26%)	2.222 (18%)	1.975 (16%)
Japaratuba	18.743	8.457 (45%)	2.943 (16%)	2.052 (11%)	1.824 (10%)
Carmópolis	16.634	8.223 (49%)	2.862 (17%)	1.997 (12%)	1.774 (11%)
Campo do Brito	18.109	7.901 (44%)	2.750 (15%)	1.918 (11%)	1.706 (9%)
Moita Bonita	11.335	7.553 (67%)	2.631 (23%)	1.833 (16%)	1.629 (14%)
Pacatuba	14.428	7.393 (51%)	2.576 (18%)	1.793 (12%)	1.597 (11%)
Frei Paulo	15.421	7.207 (47%)	2.511 (16%)	1.749 (11%)	1.556 (10%)
Malhador	12.618	7.174 (57%)	2.497 (20%)	1.741 (14%)	1.547 (12%)
Pirambu	9.280	6.637 (72%)	2.312 (25%)	1.612 (17%)	1.433 (15%)

Continua

Município	População	Cenários			
		A	B	C	D
Monte Alegre de Sergipe	15.031	6.482 (43%)	2.258 (15%)	1.574 (10%)	1.401 (9%)
Cristinápolis	17.874	6.218 (35%)	2.164 (12%)	1.510 (8%)	1.341 (8%)
Rosário do Catete	10.855	5.948 (55%)	2.071 (19%)	1.444 (13%)	1.285 (12%)
Nossa Senhora de Lourdes	6.483	5.668 (87%)	1.974 (30%)	1.375 (21%)	1.225 (19%)
Pinhão	6.576	5.654 (86%)	1.970 (30%)	1.372 (21%)	1.221 (19%)
São Domingos	11.137	5.505 (49%)	1.918 (17%)	1.336 (12%)	1.189 (11%)
Gararu	11.604	5.366 (46%)	1.868 (16%)	1.301 (11%)	1.158 (10%)
Santa Luzia do Itanhy	14.035	4.651 (33%)	1.620 (12%)	1.129 (8%)	1.003 (7%)
Tomar do Geru	13.536	4.461 (33%)	1.551 (11%)	1.086 (8%)	962 (7%)
Japoatã	13.434	4.227 (31%)	1.471 (11%)	1.026 (8%)	914 (7%)
Santana do São Francisco	7.780	4.126 (53%)	1.356 (13%)	947 (9%)	839 (8%)
Riachuelo	10.213	3.900 (38%)	1.289 (14%)	901 (10%)	799 (9%)
Siriri	8.893	3.704 (42%)	1.166 (14%)	812 (10%)	723 (8%)
Ilha das Flores	8.520	3.346 (39%)	1.098 (33%)	765 (23%)	680 (20%)
General Maynard	3.346	3.151 (94%)	1.087 (18%)	758 (13%)	673 (11%)
Cedro de São João	5.897	3.122 (53%)	1.054 (11%)	736 (8%)	652 (11%)
Pedrinhas	9.602	3.026 (32%)	1.050 (18%)	734 (13%)	653 (7%)
Graccho Cardoso	5.818	3.017 (52%)	1.041 (12%)	723 (8%)	646 (7%)
Nossa Senhora Aparecida	8.796	2.987 (34%)	1.031 (26%)	717 (18%)	638 (16%)
Muribeca	7.625	2.672 (35%)	930 (12%)	651 (9%)	576 (8%)
Brejo Grande	8.309	2.637 (32%)	918 (11%)	640 (8%)	570 (7%)

Continua

Município	População	Cenários			
		A	B	C	D
Divina Pastora	5.138	2.608 (51%)	908 (18%)	632 (12%)	564 (11%)
Macambira	6.919	2.378 (34%)	875 (11%)	610 (8%)	541 (7%)
Malhada dos Bois	3.682	2.086 (57%)	827 (12%)	577 (8%)	513 (7%)
Feira Nova	5.584	1.893 (34%)	726 (20%)	507 (14%)	451 (12%)
São Miguel do Aleixo	3.930	1.554 (40%)	660 (12%)	459 (8%)	409 (7%)
Itabi	4.903	1.524 (31%)	541 (14%)	377 (10%)	336 (9%)
Santa Rosa de Lima	3.913	1.346 (34%)	530 (11%)	371 (8%)	330 (7%)
Canhoba	4.008	1.294 (32%)	451 (11%)	314 (8%)	280 (7%)
São Francisco	3.724	1.291 (35%)	450 (11%)	313 (8%)	278 (7%)
Cumbe	3.987	1.289 (32%)	450 (12%)	313 (8%)	279 (7%)
Telha	3.227	1.058 (33%)	369 (11%)	257 (8%)	229 (7%)
Pedra Mole	3.261	1.027 (31%)	356 (11%)	249 (8%)	220 (7%)
Amparo de São Francisco	2.374	1.002 (42%)	348 (15%)	245 (10%)	215 (9%)

Fonte: Autoria própria

DISTRIBUIÇÃO GEOGRÁFICA E DIFUSÃO ESPACIAL DOS CASOS DE COVID-19 NO ESTADO DE MATO GROSSO

Luiz Geraldo Mendes

anton36@gmail.com

Gizelle Prado da Silva Fonseca

gizelleprado@gmail.com

José Manuel Cañizal Salceda

josecanizalsalceda@gmail.com

Adriana Queiroz Nascimento Pinhorati

professoraadrianaqueiroz@gmail.com

UFMT

Resumo: O estudo apresenta um procedimento metodológico para analisar a difusão espacial da pandemia Covid-19 através dos mapas de vetores de mobilidade espacial do centro geográfico ponderado, nas Regiões Geográficas Intermediárias do Estado de Mato Grosso e módulos de fluxos rodoviários. As informações por município sobre casos e óbitos confirmados em decorrência do vírus e divulgados pela Secretaria de Saúde Pública do Estado de Mato Grosso foram consideradas para as 15 primeiras semanas de epidemia. A média móvel de casos e óbitos permitiu constatar visível avanço da doença, desde início de junho. As análises do deslocamento dos Centros Geográficos Ponderados possibilitou identificar as maiores concentrações de casos e a sua dispersão, constituindo a Região Intermediária de Cuiabá a com maior número de casos, e menor raio de distância média ponderada, menor dispersão, padrão que se repete nas regiões de Cáceres e Rondonópolis, já as regiões de Sinop e Barra do Garças apresentaram maiores distâncias médias ponderadas e deslocamento alinhando-se com o trajeto das rodovias BR-163, no centro e MT-158 na parte leste do Estado. Este tipo de análise fornece uma nova abordagem para avaliar a tendência da dinâmica espacial a partir do ponto de vista estatístico com o uso de SIG.

Palavras-chave: Covid-19; média móvel; centro geográfico ponderado; análise espacial; SIG.

Abstract: The study presents a methodological procedure to go over the spatial diffusion of Covid-19 pandemic through Weighted Geographic Centre maps of spatial mobility vectors, in the Intermediate Geographic Regions of the State of Mato Grosso and motorway flows modules. Municipal information of cases and deaths confirmed as due to the virus and made public by the Public Health Department of the State of Mato Grosso were considered for the first 15 weeks since the beginning of the epidemic. The mobile average of cases and deaths allowed us to notice the visible spread of the illness, since the beginning of June. The assess of the displacement of Averaged Geographic Centres allowed us to identify the major cases concentration and its dispersion, being the Intermediate Region of Cuiabá the one with more number of cases, shorter

radius of weighted average distance and smaller dispersion, model repeated in the regions of Cáceres and Rondonópolis; meanwhile the regions of Sinop and Barra do Garças showed longer weighted average distances and displacement aligned with the motorways BR-163 in the center and MT-158 in the East side of the State. This kind of analyses offers a new approach to review the spatial dynamic tendency from a statistic point of view using GIS.

Keywords: COVID-19; moving average; weighted geographical center; spatial analysis; GIS.

1. INTRODUÇÃO

Percorridos aproximadamente seis meses da descoberta de um novo tipo de coronavírus no continente asiático, com a primeira alerta à Organização Mundial da Saúde emitida em dezembro de 2019, referente a ocorrência dos primeiros casos na China, a partir de janeiro já se disseminando rapidamente a outros países, com a declaração de emergência de saúde pública de importância internacional, a doença Covid-19, provocada pelo coronavírus 2019, passou a ser classificada em março de 2020 como uma pandemia, presente nos diferentes continentes do mundo, devido a sua transmissibilidade rápida, com resultados graves à saúde e em muitos casos causando a morte.

No Brasil o primeiro caso confirmado da doença ocorreu em São Paulo, no mês de fevereiro, a partir dessa data os casos foram aumentando e avançando rapidamente pelo país, com a transmissão comunitária reconhecida em todo o território brasileiro a partir do mês de março.

A Região Centro-Oeste no mês de abril de 2020, ainda registrava o menor número de diagnósticos de Covid-19, se comparado às outras regiões do Brasil e os índices mais baixos de óbitos pela doença. Fato que poderia estar relacionado ao menor fluxo de pessoas em direção à região, provenientes de outras localidades do país e do mundo, também se comparado às outras regiões do país, o que denota um certo isolamento, somado a algumas medidas tomadas inicialmente, com restrições propícias ao controle da disseminação do vírus no Estado.

Até o dia 10 de abril de 2020 se registravam 955 casos no Centro-Oeste, enquanto na Região Norte apresentavam 1.505 casos, Nordeste 3.528, Sudeste 11.678 e Sul 1.972 casos (Boletim Epidemiológico do Ministério de Saúde). Em se tratando do Estado de Mato Grosso, o baixo número de ocorrência de casos e óbitos inicialmente e o tardio avanço, provavelmente estavam ligados às medidas de enfrentamento à pandemia, que seguia as ações do cenário internacional, com a emissão, pelo Governo do Estado, do Decreto nº407 de 16 de março de 2020, estabelecendo, dentre outras ações, medidas como a suspensão das atividades escolares, orientações sobre o isolamento,

quarentena, orientações para a diminuição de aglomerações humanas, implantação do teletrabalho (home office), quarentena obrigatória às pessoas que chegassem de viagem com sintomas. O referido decreto reproduziu ações, com o mesmo objetivo, relacionado ao controle da Pandemia, nos diferentes municípios do Estado, com a emissão de aproximadamente 780 registros de atos, Decretos, Portarias, Leis e similares em nível municipal, de acordo com o levantamento realizado pelo Grupo de Pesquisa da UFMT: A Geografia das ações políticas de enfrentamento à Covid-19 em Mato Grosso; uma análise de março de 2020, divulgado por meio de Relatório Técnico. (<http://geografiaufmt.com.br/index.php/pt-br/COVID-geografia-acoes-politicas> , acessado em 25/07/2020).

No entanto, a partir do mês de abril/2020 houveram precocemente flexibilizações que permitiram abertura da rede comercial levando a maior circulação de pessoas e a mudanças significativas no aumento do número de casos, fato que no final do mês de junho provocou a ausência de leitos de UTI, por mais de 2 semanas.

Ao se comparar o número de casos e óbitos acumulados até o final do mês de junho de 2020, do Estado de Mato Grosso com os de outros Estados brasileiros, Mato Grosso se encontrava entre os cinco com menor quantitativo de casos e entre os nove com menor quantitativo de óbitos, contudo, apresentou um avanço alarmante em três meses, considerando que entre o período de 12/04/2020 a 30/06/2020 o número de casos progrediram de 123 para 15.636, conforme os dados disponibilizados pelo Departamento de Estatística da UFMT (<https://www.ufmt.br/covid> , acessado em 25/07/2020).

Na busca por tentar explicar os fatos que levaram ao avanço do número de casos no estado, esta pesquisa busca identificar correlações estatísticas e por meio dos diferentes fluxos rodoviários a dispersão zonal do vírus no estado. É apresentado a análise da difusão espacial da pandemia causada pelo Covid-19 através dos mapas de vetores de mobilidade espacial do centro geográfico ponderado por semana nas Regiões Geográficas Imediatas e Intermediárias do Estado de Mato Grosso e das comunidades encontradas na análise do fluxo rodoviário.

2.METODOLOGIA

Ao refletir sobre a difusão espacial, teoria discutida pelo geógrafo sueco, Torsten Hägerstrand (1952), pode-se discutir à luz da sua teoria, a difusão do vírus no Estado de Mato Grosso. As etapas teórico-metodológicas por ele proposta, afirma que o processo de difusão espacial ocorre em 4 etapas: a

primeira delas contraste, a segunda parte de uma lógica centro/periferia saindo do ponto central chegando a lugares mais distantes, a terceira ocorre quando as diferenças são diminuídas, gerando uma aproximação entre o que era central com o periférico e por último ocorre o processo de saturação sendo este o maior nível.

Alinhado com essa ideia de Hägerstrand, temos a proposta por Raffestin quando discute esta relação centro-periferia (RAFFESTIN, 1993). Nela podemos entender uma das relações presentes nas cidades em que registram as ocorrências da Covid, onde muitas vezes partem das cidades com maiores fluxos, por exemplo, de transporte e pessoas, para as demais com menores fluxos.

Analisando os dados sobre a evolução da Covid-19 em Mato Grosso foi estabelecido um caminho metodológico que envolveu a teoria de Hägerstrand (1952) sobre difusão espacial, que também envolve além da relação centro-periferia, a logística de deslocamento rodoviário, aéreo e hidroviário, fatores importantes para o processo de difusão do vírus pelo Estado.

Sob a lógica da Teoria da Difusão Espacial, para a espacialização das informações, relacionadas a disseminação da Covid-19 foram adotadas como base as Regiões Geográficas Imediatas e Intermediárias definidas pelo IBGE em 2017, quando reorganizou o território dos Estados brasileiros. As Regiões Geográficas Imediatas se estruturam a “partir de centros urbanos próximos, para a satisfação das necessidades imediatas das populações, tais como: compras de bens de consumo duráveis e não duráveis” (IBGE,2017,p.20) e as Regiões Geográficas Intermediárias “organizam o território, articulando as Regiões Geográficas Imediatas por meio de um polo de hierarquia superior diferenciado, a partir dos fluxos de gestão privado e público e da existência de funções urbanas de maior complexidade” (IBGE, 2017, p.20).

Assim, para o Estado de Mato Grosso, as cidades que organizam o território mato-grossense são: Cuiabá, Rondonópolis, Cáceres, Barra do Garças e Sinop, as mesmas que compõem os centros urbanos com maior quantitativo de casos de Covid-19 no Estado.

De acordo com Corrêa (2016), “as interações espaciais variam segundo a duração, a velocidade, a frequência, o ritmo e o período de ocorrência”. E dentre os diversos tipos de redes geográficas, para que se articulem são necessárias determinadas estruturas de conexão (SOUZA et al.,2020). A identificação das infraestruturas de conexão entre cidades, como rodovias, aeroportos e portos tem auxiliado no estudo de difusão da Covid-19 pelo mundo e “a circulação e a conectividade entre diferentes lugares têm peso tão

importante como a localização territorial no processo de difusão espacial de fenômenos de todo o tipo” (SPOSITO; GUIMARÃES, 2020).

Já existem alguns estudos de análise de propagação do vírus influenza A/H1N1 (SHOBOUGAWA et al.,2012), influenza aviária A/H7N9 (DONG et al.,2017), insuficiência respiratória aguda SARS (FANG et al.,2009) e da COVID-19 usando dados geográficos (CAVALCANTE et al.,2020), (COVID-19 BR,MAIA et al.,2020). Neste estudo, foi analisada a difusão e o padrão de distribuição da Covid-19 usando um indicador espacial, o centro médio geográfico (C_g) e distância média ponderada (d_w), que nos permite indicar o sentido de difusão e medir o nível de compactação da distribuição espacial e da Covid-19 no Estado de MT.

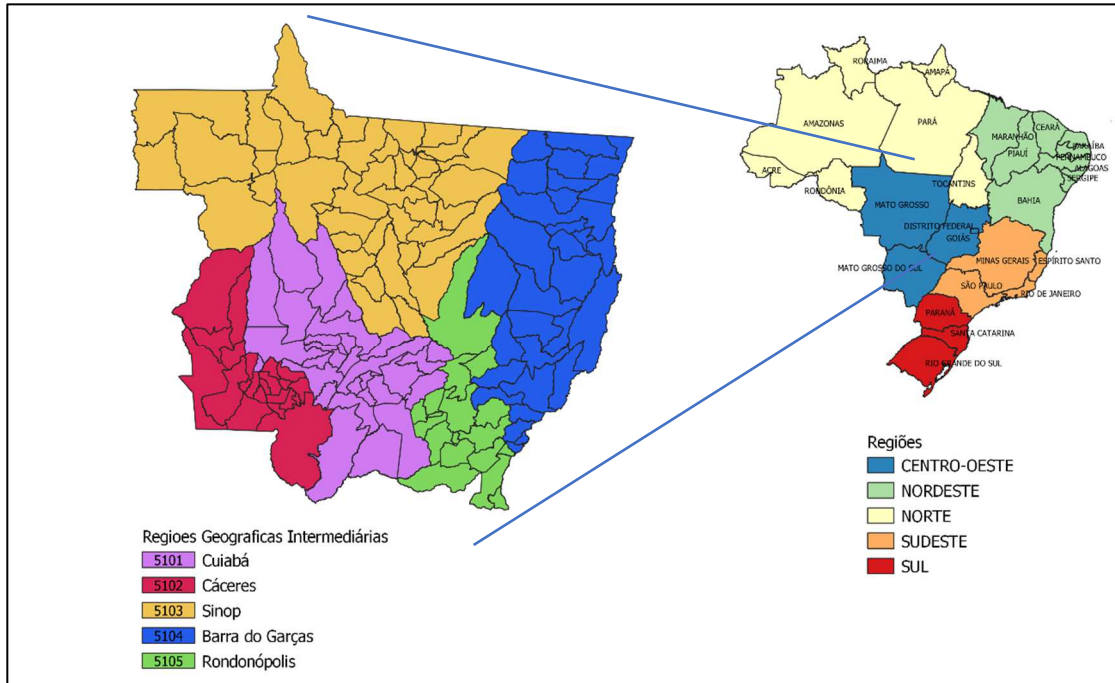
Todos os dados usados neste estudo estão disponíveis ao público nos sites do Ministério da Infraestrutura (<https://www.infraestrutura.gov.br>), IBGE (<http://www.ibge.gov.br>) e da SES-MT (<http://www.saude.mt.gov.br/informe/584>).

A análise da distribuição espacial da Covid-19 no Estado de Mato Grosso (MT) foi realizada com a coleta de dados de casos confirmados até a data de 05/07/2020 sendo escolhido o período entre 20/03/2020 (primeira divulgação de caso confirmado no Estado) e 03/07/2020 totalizando 15 semanas de casos confirmados no Estado. Foram utilizados os seguintes indicadores:

- 1) Número acumulado de casos e óbitos por Covid-19, confirmados e corrigidos até 03/07/2020.
- 2) Evolução diária do número de casos e óbitos confirmados até 03/07/2020, para os municípios de MT com o maior número de casos acumulados à data de coleta. As séries temporais foram analisadas em nível e na escala logarítmica.
- 3) Evolução da taxa de crescimento dos óbitos diários e da taxa de letalidade para os municípios de MT com maior ocorrência de óbitos, até 03/07/2020.
- 4) Evolução da taxa de crescimento dos óbitos diários e da taxa de letalidade por faixa etária, no Estado de MT, até 03/07/2020.
- 5) Evolução da média móvel de casos e óbitos para os principais municípios de MT, Estado de MT e comparação com alguns Estados.
- 6) Centro geográfico médio e distância padrão ponderada do número de casos no período de 20/03/2020 a 03/07/2020.

O Estado de Mato Grosso localiza-se na Região Centro-Oeste do Brasil, sendo dividido em 141 municípios, com área aproximada de 903.207,019 km². Possui aproximadamente 3.484.466 habitantes (IBGE).

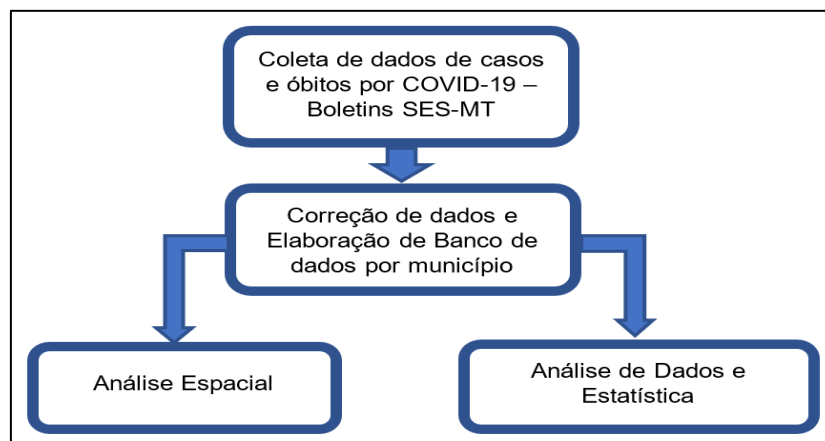
Figura 1 - Localização do Estado de Mato Grosso/Brasil



Fonte de dados: IBGE,2017. Sem escala. Editado pelos autores

Na Figura 2 apresentamos as etapas metodológicas utilizadas neste trabalho.

Figura 2 - Etapas metodológicas



Foi utilizado o software Excel 365 da Microsoft e os softwares livres GEPHI para análise, visualização de redes e cálculo de comunidades e QGIS 3.12 para a análise espacial e elaboração de mapas temáticos.

2.1. ANÁLISE ESPACIAL

Para avaliar o comportamento da difusão do Covid-19 no Estado de Mato Grosso, nas áreas de estudo e na escala de tempo semanal, utilizou-se a técnica estatística de suporte do **centro médio ou geográfico ponderado** (C_g) que é definido de acordo com as equações 1 e 2.

$$\bar{X}_g = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (1)$$

$$\bar{Y}_g = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i} \quad (2)$$

Onde x_i e y_i , são as coordenadas (x,y) de um ponto i e w é o peso (número de casos). O índice i faz referência a data de ocorrência do número de casos.

Desta forma, as médias das coordenadas x (\bar{X}_g) e das coordenadas y (\bar{Y}_g) de um determinado arranjo de pontos formarão um par de coordenadas espaciais que indicarão o centro da distribuição dos casos de Covid-19 para cada área de interesse.

O próximo passo, foi o cálculo da **distância padrão ponderada** (d_w), definida pela equação 3.

$$d_w = \sqrt{\frac{\sum_{i=1}^n w_i (x_i - \bar{X}_g)^2 + \sum_{i=1}^n w_i (y_i - \bar{Y}_g)^2}{\sum_{i=1}^n w_i}} \quad (3)$$

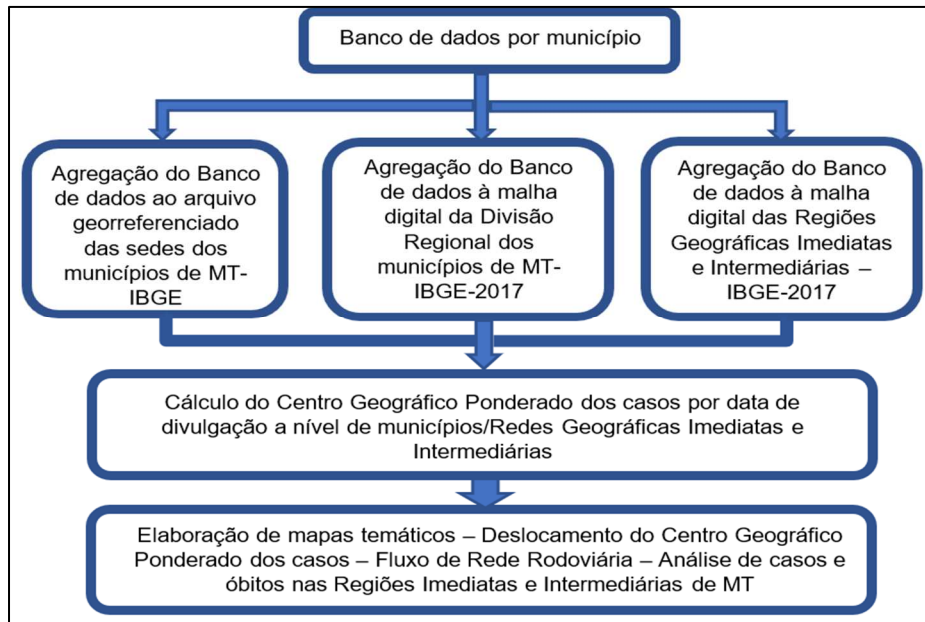
Este raio padrão, d_w , considera a distância de cada ponto i até o centro geográfico ponderado de distribuição. Dessa maneira, o raio padrão também auxilia no entendimento dos arranjos espaciais de pontos, onde os maiores valores de d_w são relativos ao maior espalhamento dos pontos em relação ao centro geográfico ponderado; e os menores valores de d_w relacionam-se à maior agregação dos pontos (clusters) em relação ao centro geográfico ponderado da distribuição espacial.

Consideramos apenas as datas de início com número de casos ≥ 3 para o cálculo dos C_g e d_w , por serem mais significativas e em muitos casos, o C_g encontra-se localizado em um único município com $d_w=0$ para número de casos <3 .

Na Figura 3, apresentamos as etapas de análise espacial. A análise de fluxo rodoviário foi feita usando a base de dados de ligações rodoviárias e hidroviárias 2016 – IBGE, selecionando apenas as cidades do Estado de MT. Foi usado o software GEPHI para a análise de rede, caracterizando as comunidades pela modularidade, usando o algoritmo de Blondel et al. (2008). Este algoritmo funciona em duas etapas. Na primeira, é realizado o agrupamento de nós em

comunidades. E na segunda etapa cada comunidade é tratada como um nó em uma nova rede e o processo é iniciado novamente até que não seja possível detectar alterações na rede obtendo um valor máximo para a modularidade.

Figura 3 - Etapas da análise espacial

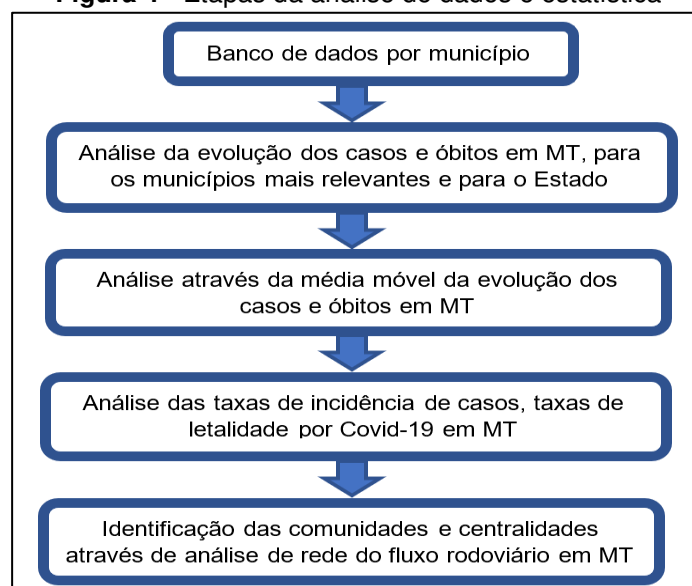


Para a elaboração dos mapas temáticos: Deslocamento do Centro Geográfico Ponderado dos casos – Fluxo de Rede Rodoviária – Análise de casos e óbitos nas Regiões Imediatas e Intermediárias de MT foi usado o software QGIS 3.12.

2.2. ANÁLISE DE DADOS E ESTATÍSTICA

A análise de dados e estatística segue as etapas apresentada na Figura 4.

Figura 4 - Etapas da análise de dados e estatística



A análise de dados foi realizada com os softwares EXCEL e GEPHI para a identificação das comunidades através da modularidade com o algoritmo de Blondel, que trabalha bem com redes extensas, otimizando o particionamento de um grafo em comunidades (modularidade) e com rápido tempo de computação.

3. RESULTADOS

A análise da distribuição espacial dos casos de Covid-19 no Estado de Mato Grosso demonstrou desigualdades na difusão espacial, com maior incidência nos Centros Sub Regionais - A, e cidades polos das Regiões Geográficas Intermediárias e Imediatas, apresentando correlação positiva entre a expansão dos casos, com as cidades mais populosas do Estado e as que apresentam maiores fluxos rodoviário e aéreo e maior conectividade com a capital do Estado e demais cidades do módulo rodoviário a qual fazem parte.

O Estado de Mato Grosso encontra-se, de certa forma, mais isolado, no interior do território Brasileiro, com a sua capital classificada na hierarquia dos centros urbanos (IBGE, 2018) como Capital Regional, com alta concentração de atividades de gestão, porém, com alcance menor em termos de região de influência que as cidades classificadas como metrópole. Neste caso, considerada apenas Região Metropolitana, se somada a população da conurbação, Cuiabá-Várzea Grande, pode ter apresentado maior complexidade nos processos de difusão espacial do coronavírus, que se propaga também em função das heterogeneidades do espaço geográfico do Estado.

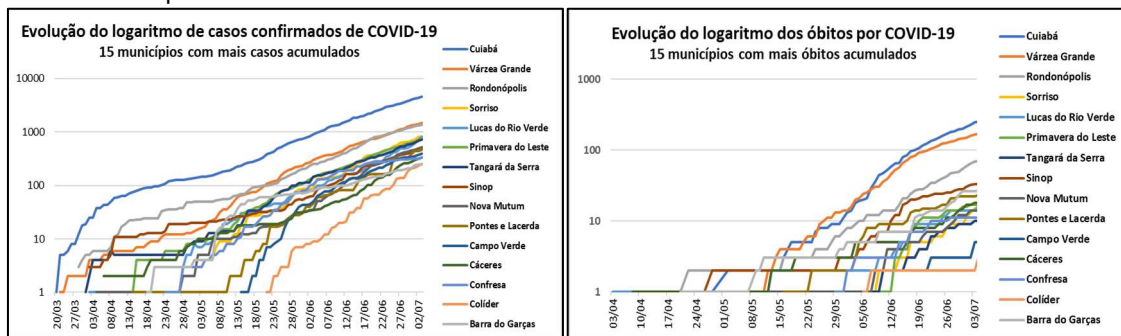
3.1. ANÁLISE DE DADOS E ESTATÍSTICA

Ao analisar a evolução dos casos e óbitos em escala logarítmica no Estado de Mato Grosso, percebe-se que a partir do final do mês de abril o número de casos, na maior parte dos municípios presentes no gráfico passaram a apresentar comportamento exponencial, com o crescimento do número de casos multiplicados por dez a cada mês, considerando os intervalos entre 28/04/2020 a 28/05/2020 e 02/06/2020 a 02/07/2020. Com relação aos óbitos a partir da primeira dezena do mês de maio os municípios passaram a apresentar comportamento exponencial, com destaque no aumento de óbitos nos municípios de Cuiabá, Várzea grande, Rondonópolis e Sinop.

No período de aproximadamente dois meses, entre o dia 20/03/2020 a 23/05/2020 as 15 principais cidades do Estado, que se destacam por serem as mais populosas e de maior influência econômica, constituídas pela capital do Estado, os principais Centros Sub Regionais (REGIC, 2018) e as principais cidades polos das Regiões Intermediárias e Imediatas (IBGE,2017) passaram a

apresentar casos da doença. São cidades que atuam como polos atrativos, com maior concentração e circulação de pessoas, possuem aeroportos, com conexões aéreas com o aeroporto Marechal Rondon (Várzea Grande), o principal do Estado de Mato Grosso, que por sua vez, está interconectado com aeroportos das outras regiões do país, prováveis causas de serem as primeiras cidades com a incidência da doença e do aumento exponencial de casos e óbitos, entre os meses de abril e junho de 2020, conforme é demonstrado na Figura 5.

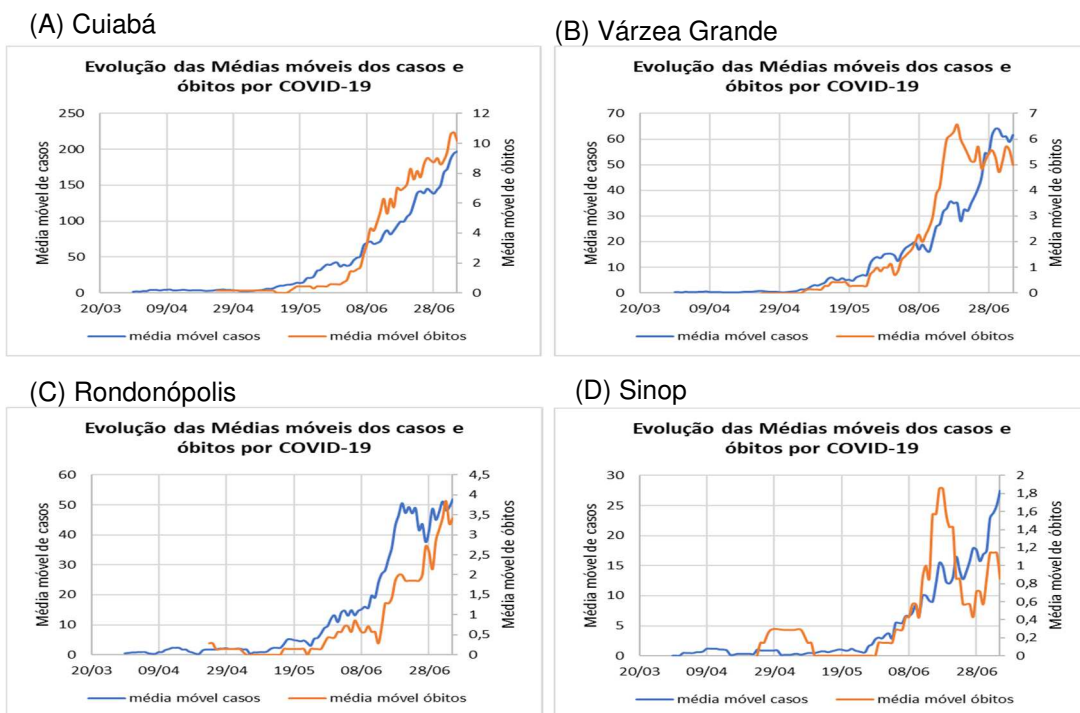
Figura 5 - Evolução na escala logarítmica dos casos e óbitos confirmados de COVID-19 nos 15 municípios de MT com maior número de casos acumulados até 03/07/2020 – MT



Fonte: SES-MT. Elaborado pelos autores

A análise da série temporal de casos e óbitos por Covid-19 é melhor entendida através do modelo estatístico da média móvel em 7 dias, onde podemos verificar o comportamento da difusão da pandemia e fazer estimativas.

Figura 6 - Evolução da Média móvel dos casos e óbitos confirmados de Covid-19 nos municípios de MT com maior número de casos e óbitos até 03/07/2020



Fontes: COTA,W e SES-MT. Elaborado pelos autores.

Dos municípios do Estado de MT com maior número de casos e óbitos por Covid-19 (Figura 5), destacam-se os municípios de Cuiabá, Várzea Grande, Rondonópolis e Sinop com as maiores médias móveis (Figura 6).

Nos municípios de Cuiabá, Várzea Grande e Rondonópolis percebeu-se certa estabilidade com relação ao avanço dos casos e óbitos até a primeira dezena do mês de maio, às medidas de restrição impostas pelo Decreto emitido em 16 de março pelo Governo do Estado pode ter contribuído, para a evolução de forma mais lenta na fase inicial, pois sabe-se que a evolução da transmissibilidade da doença e da ocorrência de óbitos, representados pela média móvel, depende também do comportamento da sociedade e das medidas que os governantes tomam.

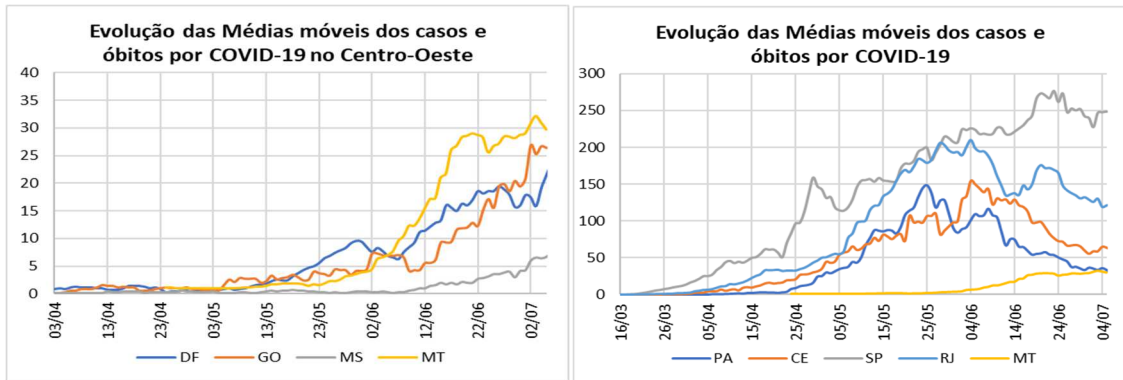
A curva passou a aumentar muito nos três municípios e demais municípios do Estado, a partir do final da primeira dezena do mês de junho, demonstrando menor controle da doença no Estado, podendo ter forte relação com as medidas que levaram a flexibilização da abertura dos comércios. Já no município de Sinop, o comportamento manifestou-se com pequena diferença, tendo o aumento da curva ocorrendo mais tardiamente, a partir no início do mês de junho.

Sabe-se que as variações diárias muito grandes nem sempre tem a ver com avanço do vírus, mas devido a atrasos nos registros, falta de testes, entre outros problemas, o que aconteceu de forma diferenciada nos municípios do Estado, que nem sempre possuíam o quantitativo de testes que demonstraria a real quantidade de pessoas contaminadas pela doença ou que passou a realizá-los de forma tardia, além de toda a heterogeneidade do espaço geográfico. No entanto, a Média Móvel permite a equalização dos dados, chegando o mais próximo da real situação, não demonstrando prováveis inconsistências provenientes, por exemplo, do acúmulo do resultado de testes após o final de semana ou a subnotificação devido a falta da realização dos mesmos.

Nos quatro municípios verifica-se as maiores médias móveis ocorrendo a partir da segunda quinzena do mês de junho, considerando o período analisado, entre o dia 20/03/2020 a 03/07/2020.

Na avaliação da série temporal da média móvel de casos e óbitos por Covid-19, o Estado de MT se destaca dentro da Região Centro-Oeste e comparando com os Estados que flexibilizaram a abertura de funcionamento de comércio e serviços quando a média móvel começou a diminuir ou estabilizar, o Estado de MT ainda não apresenta uma tendência de baixa para permitir essa flexibilização (Figura 7).

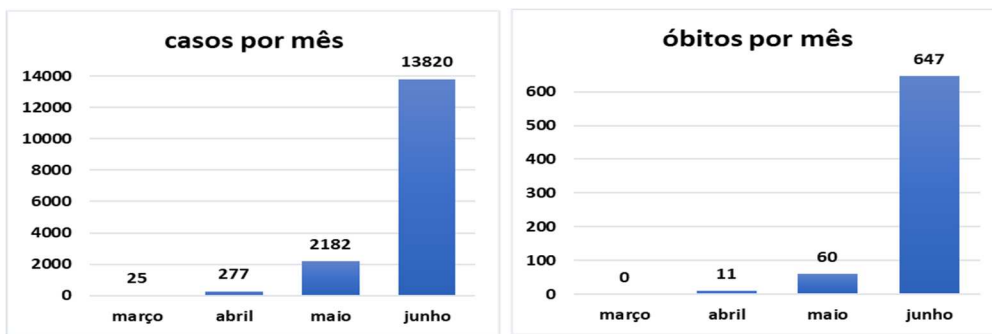
Figura 7 - Comparação da evolução da Média móvel dos casos e óbitos confirmados de Covid-19 na Região Centro-Oeste e alguns Estados até 03/07/2020



Fontes: COTA,W e SES-MT. Elaborado pelos autores

Analisando o número de casos e óbitos confirmados por mês comprovamos esse aumento exponencial de 633,36% no número de casos e de 1078,33% no número de óbitos entre os meses de maio e junho (Figura 8).

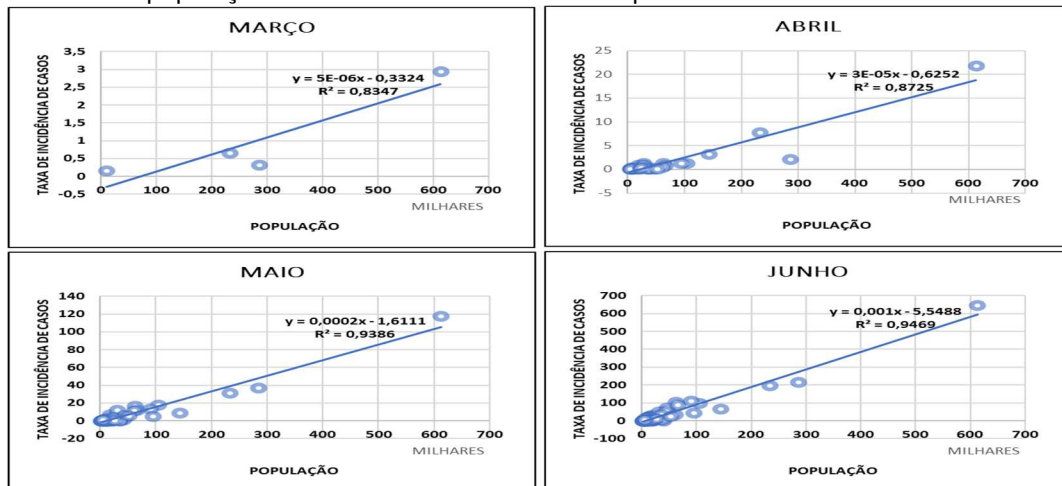
Figura 8 - Evolução do número de casos e óbitos confirmados por COVID-19 por mês no Estado de MT



Fonte: SES-MT. Elaborado pelos autores

Com a difusão do vírus para os outros municípios do Estado, há uma forte correlação entre a taxa de incidência de casos por Covid-19 com a população estimada dos municípios por mês (Figura 9).

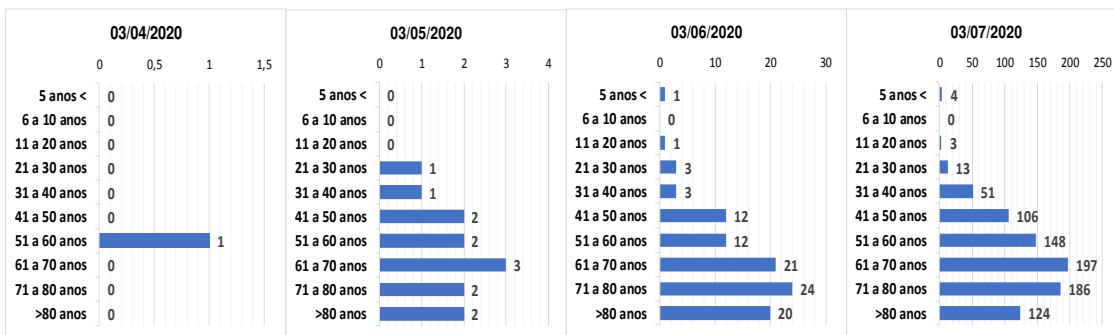
Figura 9 - Gráficos de correlação entre a taxa de incidência de casos por COVID-19 com a população estimada em 2019 dos municípios de MT em cada mês



Fonte: SES-MT, IBGE. Elaborado pelos autores

Podemos avaliar como a evolução do número de óbitos atinge a população por faixa etária (Figura 10).

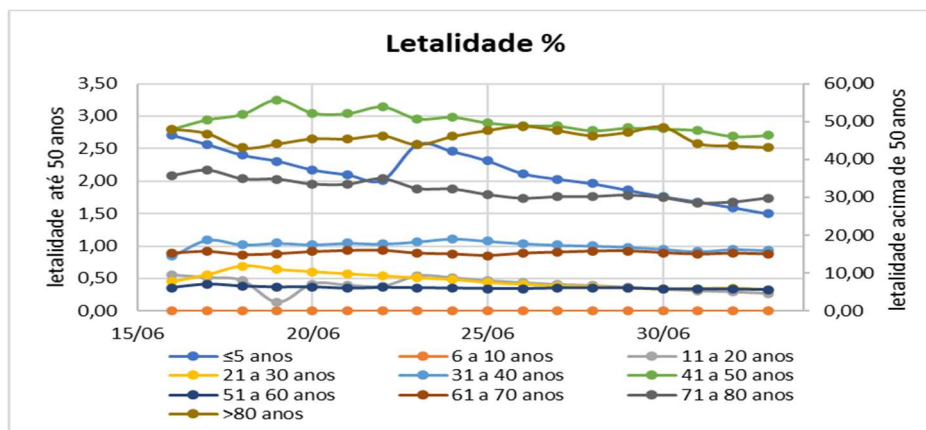
Figura 10 - Evolução do número de óbitos confirmados por Covid-19, por faixa etária, no Estado de MT



Fonte: SES-MT. Elaborado pelos autores

Até a data de consolidação dos dados (03/07/2020), verificamos que o número de óbitos atinge mais a faixa acima de 51 anos, não afetando muito as faixas etárias até os 20 anos. Isso pode ser entendido como uma baixa propagação do vírus nesta faixa (faixa escolarizável) com a não propagação do vírus, o que pode estar condicionado ao não funcionamento das escolas, em forma presencial, desde o início da pandemia no Estado de Mato Grosso, com a emissão de Decreto Estadual a partir do mês de março.

Figura 11 - Evolução da Taxa de letalidade por Covid-19, por faixa etária, no período de 19/06/2020 a 03/07/2020 no Estado de MT



Fonte: SES-MT. Elaborado pelos autores

Pode-se verificar que no Estado de Mato Grosso, os dados de mortalidade encontrados seguem o padrão dos dados nacional, se confirma as maiores taxas de letalidades na população idosa, nas faixas etárias acima dos 51 anos de idade, que representa um dos grupos mais sujeitos a infecção e sintomatologia por Covid-19, com risco de morte mais elevado a partir da quinta década de vida (Figura 11).

3.2. ANÁLISE ESPACIAL

Ao analisar a evolução de casos de Covid-19 em Mato Grosso, podemos perceber que ao se tratar da relação centro-periferia, ocorre partindo de cidades com maior intensidade de relações comerciais e intensas atividades de fluxos sejam rodoviários e aeroportuários.

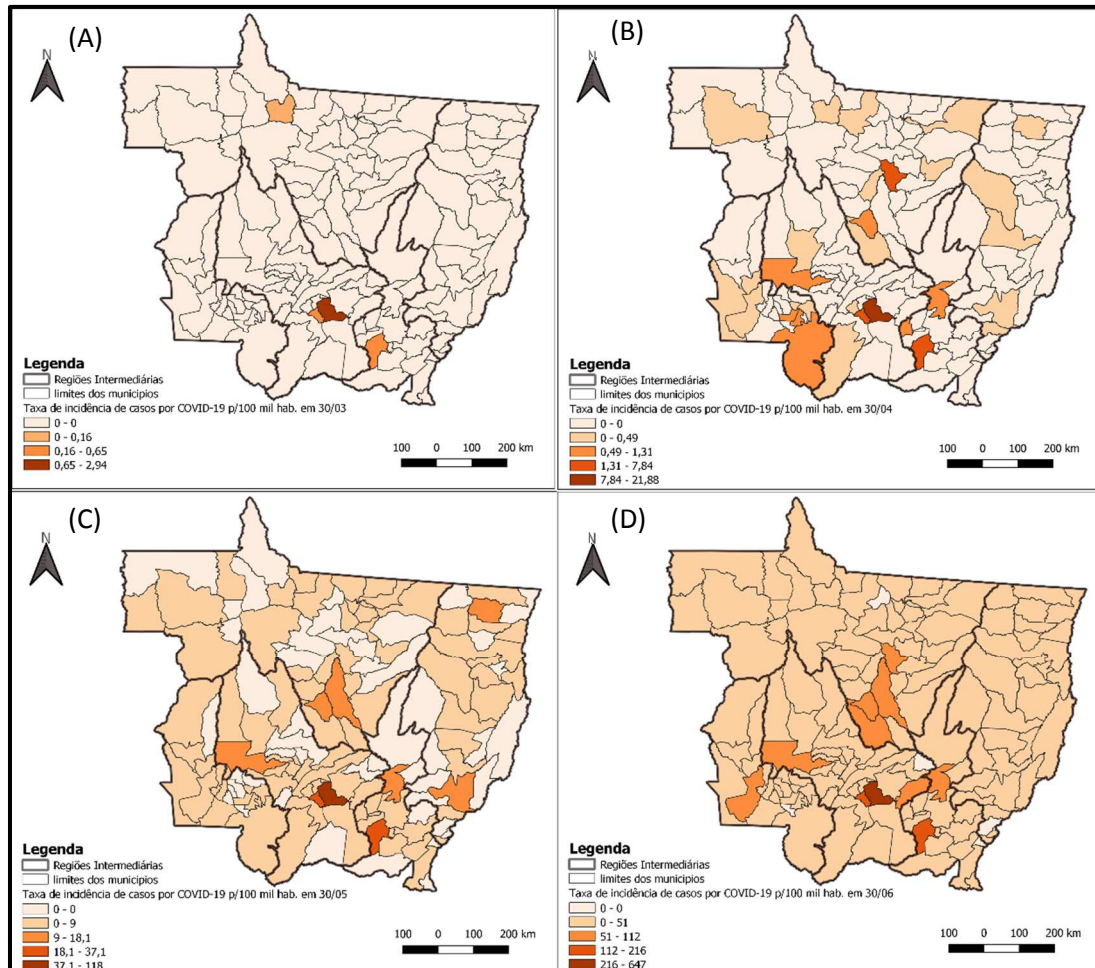
Em março de 2020, quando datam os primeiros registros da Pandemia no estado, verifica-se ocorrências em 3 Regiões Geográficas Intermediárias do Estado: **Cuiabá** (Cuiabá e Várzea Grande), **Sinop** (Alta Floresta) e **Rondonópolis** (Rondonópolis). No caso da cidade de Alta Floresta, o foco ocorreu dentro de uma cadeia pública da cidade, onde 54 detentos foram infectados pelo Covid-19. Nas demais cidades que fazem parte da Região Geográfica Intermediária de Sinop, que constituem principalmente os polos econômicos da mesma (Sinop, Sorriso e Lucas do Rio Verde, as quais possuem ligações através de estradas apenas com as regiões norte e centro-sul, pela rodovia BR-163, por onde toda safra é escoada, os primeiros casos estavam vinculados à um perfil de pessoas com maior poder aquisitivo e que haviam realizado viagens recentes à países em que foram registrados casos de Covid-19, locais com maiores fluxos aeroportuários e rodoviários, que passaram a ter aumento dos casos, já com transmissividade comunitária dos casos e com visível dispersão às cidades contíguas, como podem ser vistas nos mapas, a incidência nos meses posteriores (abril, maio e junho) (Figura 12).

Em abril de 2020, foram notificados casos de transmissão comunitária em duas dessas regiões intermediárias, a de **Cuiabá** (Cuiabá e Várzea Grande) e **Rondonópolis** (Rondonópolis), nessas cidades não era possível saber a origem da contaminação. Similar a uma difusão radial, surgem incidências de casos de Covid-19 acima de 0,49 casos por 100 mil habitantes nas cidades com maior densidade populacional no estado: Cuiabá, Várzea Grande, Rondonópolis, Sinop, Tangará da Serra, Cáceres, Lucas do Rio Verde e Primavera do Leste. A única cidade que em abril também registrou alta incidência e não é uma das mais populosas foi Jaciara (0,97 casos por 100 mil habitantes), mas está localizada nas margens da BR 364 que faz ligação entre os dois principais focos da Covid-19 no estado, as cidades de Cuiabá e Rondonópolis.

Também é possível verificar a expansão do número de casos, além das cidades polos da Região Intermediária de Cuiabá (Cuiabá e Várzea Grande) mais populosas do Estado e com maior conectividade rodoviária e aeroportuária, em cidades do Centro Norte do Estado, dispostas ao longo da Rodovia BR-163

e que compõem a Região Intermediária de Sinop, cidades com economia fortemente vinculadas ao agronegócio.

Figura 12 - Mapa de incidência de casos por COVID-19 nos municípios de MT nos meses de março (A), abril (B), maio (C) e junho (D)



Fonte: SES-MT. Elaborado pelos autores

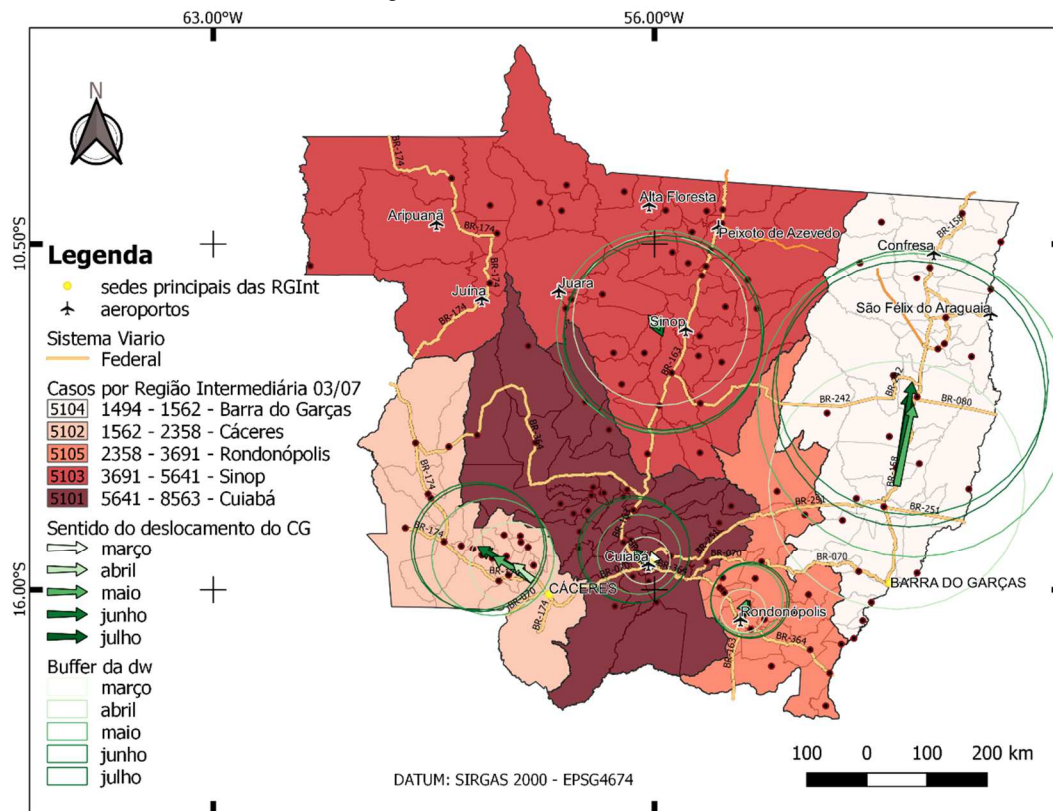
Em maio, registra-se a permanência de uma radial que parte da mesma centralidade verificada em abril. No entanto, verifica-se a redução de intensidade em Sinop, Cáceres e Jaciara. Uma nova região geográfica intermediária ganha destaque no mapa, a de **Barra do Garças**, apresentando uma taxa de incidência maior que 11,91 casos por 100 mil habitantes. Essa região possui ligação de fluxo muito intensa com o estado de Goiás, desenvolvendo atividades econômica vinculadas ao agronegócio (cadeia produtiva agrícola e pecuária). O aumento na taxa de incidência é também um fato preocupante pois está em uma região com diversas reservas indígenas.

Outro fato relevante em relação a incidência da pandemia no Estado de Mato Grosso, foi o avanço do número de casos, se espalhando por praticamente todo o Estado, onde a variação ficou em até 9 casos por 100 mil habitantes. Cabe

destacar os dados da taxa de incidência em Cuiabá, que em abril registrava 21,87 casos por 100 mil habitantes, em maio passou a registrar 118 casos por 100 mil habitantes, passando a evidenciar a taxa de crescimento geométrico de casos na cidade.

Em junho foram registradas taxas de incidência que variavam de 0 – 51 casos por 100 mil habitantes em mais de 92% dos municípios do Estado. Em Cuiabá, pode-se verificar que a taxa de incidência foi de 647,13 casos por 100 mil habitantes, classificada como risco de contaminação muito alta, esses dados são seguidos pelas cidades com as quais mantém maior fluxo de vias rodoviárias e aeroportuárias: Várzea Grande e Rondonópolis (Figura 12).

Figura 13 - Mapa de sentido deslocamento do Centro Geográfico Ponderado e do buffer da distância padrão ponderada do número de casos confirmados de COVID-19 nas Regiões Geográficas Intermediárias em MT



Fonte: SES-MT, IBGE. Elaborado pelos autores

Na análise de sentido de deslocamento dos centros geográficos ponderados, C_g , nas Regiões Geográficas Intermediárias de Mato Grosso (Figura 13), verifica-se a difusão da Covid-19 para o interior do Estado e um aumento do espalhamento considerando o valor da distância padrão ponderada, d_w , indicando a interação espacial entre as redes urbanas, sob influência do fluxo rodoviário (IBGE,2016).

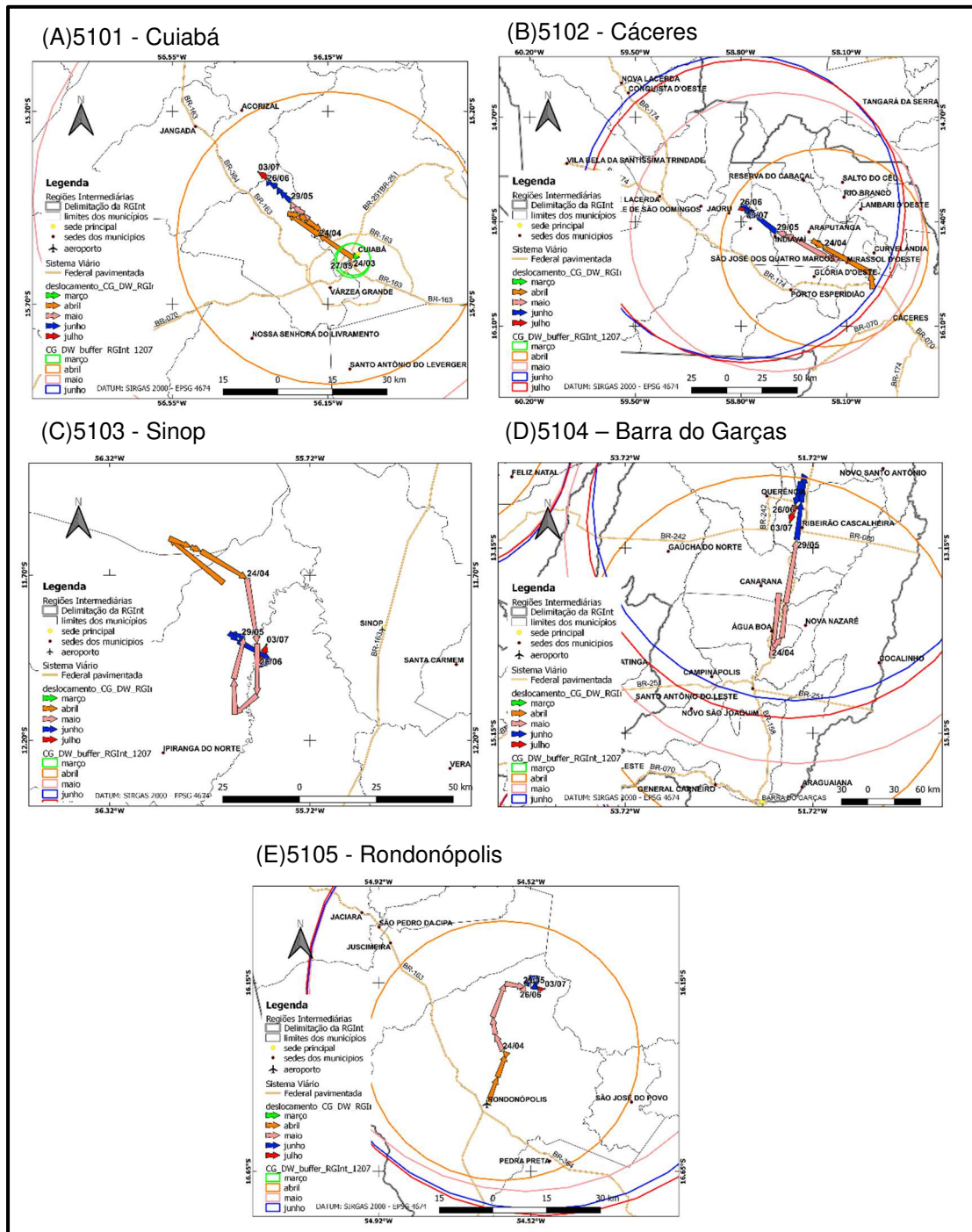
Os Centros Geográficos Ponderados e os pontos de início da disseminação da doença Covid-19 no Estado de Mato Grosso estão visivelmente dispostos, na maior parte dos casos, sobre os principais Centros Urbanos e polos econômicos de cada uma das Regiões Geográficas Intermediárias do Estado de Mato Grosso ou seguindo um alinhamento a partir destes, Cuiabá, Sinop, Cáceres e Rondonópolis e das rodovias que os interligam.

O padrão de distribuição espacial da disseminação da Covid-19 ocorreu nas porções Centro Sul, Centro Norte, Sudeste, Sudoeste e Nordeste, não constando apenas na porção Noroeste do Estado. Cabe destaque à região intermediária de Cuiabá, com maior concentração de casos nos municípios de Cuiabá e Várzea Grande, mas atingindo também outros municípios da região do Vale do Rio Cuiabá, próximos à capital, com raio de distância média ponderada menor, assim como nas regiões de Cáceres e Rondonópolis, o que demonstra menor dispersão dos casos nessas regiões. As Regiões Intermediárias com maiores raios de distância média ponderada são constituídas por Sinop e Barra do Garças, a última tendo o início dos casos nas proximidades do município de Água Boa com direcionamento para o Nordeste do Estado (Figura 14).

Ao observar a trajetória temporal e espacial da Covid-19 (Figura 14), fez-se o cálculo do centro médio da distribuição espacial da doença, ponderando o número de casos pelas coordenadas geográficas das sedes municipais, em escala de maior detalhe, possibilitando verificar o estágio da pandemia em Mato Grosso que estava em franca expansão, entre os meses de maio, junho e julho, com os maiores raios de distância média ponderada também nesses mesmos meses.

Na Região de Cuiabá (5101) o centro médio ponderado no mês de março, ficou centralizado entre Cuiabá e Várzea Grande, onde surgiram também os primeiros casos, com o deslocamento em direção noroeste, no sentido das cidades de Acorizal e Jangada a partir do mês de abril, maio e junho. Na Região de Cáceres (5102) o deslocamento se deu também no mesmo sentido, noroeste, partindo da cidade de Cáceres em direção às cidades de Mirassol do Oeste e Pontes e Lacerda, a partir do mês de abril com o deslocamento ocorrendo até o mês de julho e julho, completando as 15 semanas de análise. Na Região de Sinop (5103) a trajetória temporal e espacial se dá com o centro médio ponderado sendo direcionado nas proximidades dos municípios de Sinop, Sorriso e Lucas do Rio Verde, com o deslocamento circular, próximo a essas localidades, ocorrendo entre os meses de abril e julho.

Figura 14 - Mapa ampliado do deslocamento do Centro Geográfico Ponderado e do buffer da distância padrão ponderada do número de casos confirmados de COVID-19 nas Regiões Geográficas Intermediárias em MT



Fonte: SES-MT, IBGE. Elaborado pelos autores

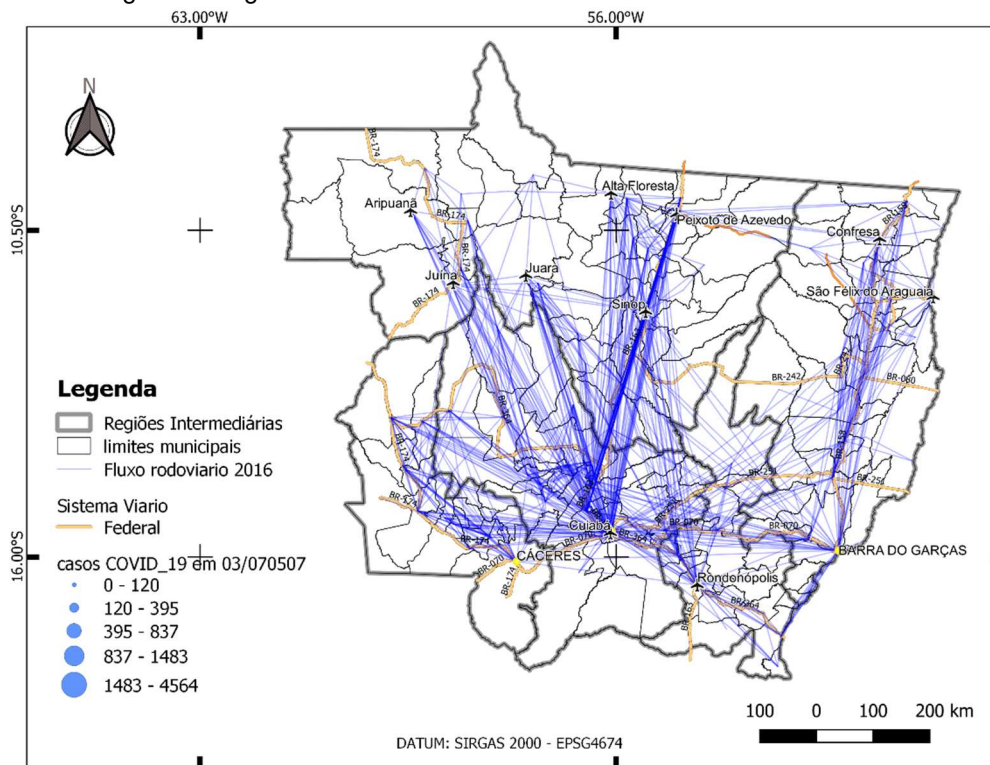
Na Região de Barra do Garças (5104) o deslocamento do centro médio ponderado se deu no mês de maio, a partir do município de Água Boa, em direção aos municípios de Querência e Ribeirão Cascalheira, nos meses de junho e julho. A Região de Rondonópolis teve o seu centro médio se deslocando do próprio município de Rondonópolis a partir do mês de abril, em sentido

nordeste, deslocando-se em direção ao município de Poxoréu nos meses de maio, junho e julho.

Outra maneira de analisar essa tendência é através da análise da rede de fluxo rodoviário (2016) com a identificação de aglomerados de cidades com maior interação (comunidades).

Os deslocamentos no Estado de Mato Grosso são frequentemente realizados utilizando a rede rodoviária, seja por transporte público ou particular, meio pelo qual o vírus encontra um meio para a sua expansão, principalmente a partir da Região Metropolitana (Cuiabá-Várzea Grande), onde o trânsito e o comércio com as demais regiões do Estado são muito intensos em que o índice de infecção são maiores (Figura 15).

Figura 15 - Mapa da rede de fluxo rodoviário (2016) e a relação entre casos de COVID-19 nas Regiões Geográficas Intermediárias de Mato Grosso e o Sistema Viário

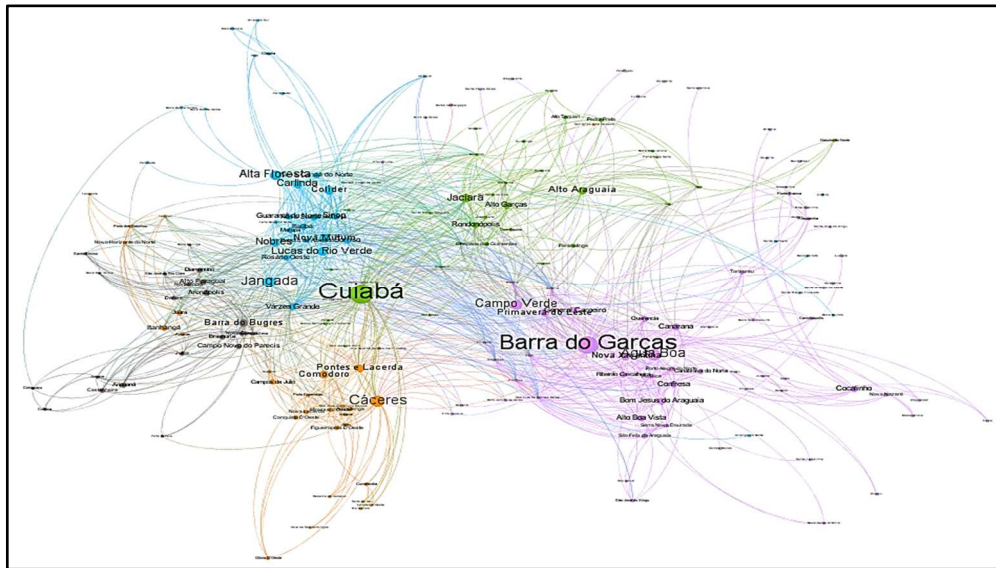


Fonte: SES-MT, IBGE.Elaborado pelos autores

Analisando o fluxo rodoviário (IBGE,2016) com o número de ligações entre as cidades, com o uso do algoritmo de Blondel et al. (2008) por trabalhar com redes complexas e baixo tempo de computação, identificamos as comunidades através da análise por modularidade. Com a Composição de módulos rodoviários, ao agrupar os municípios com mesma cor, identificou-se os

que são interconectados por fluxos rodoviários e portanto, com grande potencial para apresentar dinâmica epidemiológica relativamente sincronizadas.

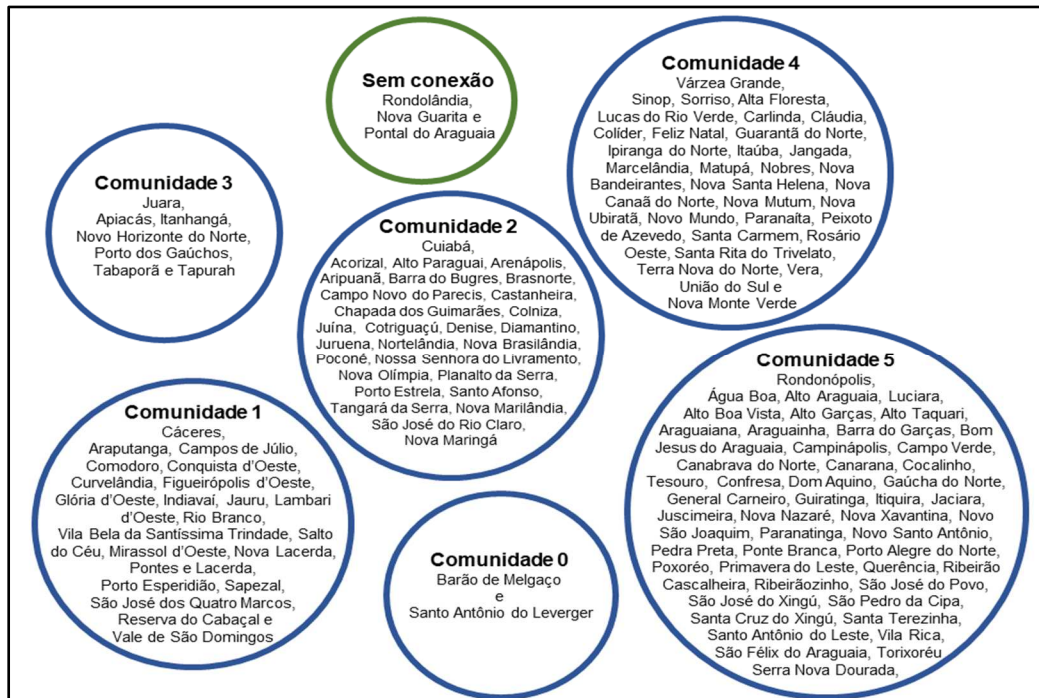
Figura 16 - Grafo da rede de fluxo rodoviário entre as cidades do Estado de Mato Grosso (2016)



Fonte: IBGE. Elaborado pelos autores com o software GEPHI

Na análise de rede de fluxo rodoviário, com a identificação das Comunidades através do algoritmo de Blondel utilizamos parâmetros de forma a manter uma continuidade na forma da Comunidade.

Figura 17 - Municípios aglomerados por comunidades com uso do algoritmo de Blondel no software GEPHI



Fonte – Obtido através do software GEPHI

Na análise de deslocamento dos centros geográficos ponderados, C_g , e das distâncias médias ponderadas, d_w , nas Comunidades identificadas na análise de redes de fluxo rodoviário, ratifica o comportamento da indicação da direção de disseminação da Covid-19 para o interior do Estado de Mato Grosso. A análise do fluxo rodoviário (IBGE,2016) com a interação entre as cidades indicam uma forte correlação das Regiões Geográficas Imediatas (Figura 18).

Verifica-se que a dispersão da doença no Estado ocorreu partindo-se da capital Cuiabá, com o avanço da Covid-19 nas direções Centro Norte, ao longo da BR-163, constituindo-se clusters de municípios, além da parte central do Estado, nas proximidades de Sinop, Sorriso e Lucas do Rio Verde, onde os raios de distância média ponderada atingiram municípios mais distantes. Na parte sudeste do Estado é visível um agrupamento de cidades próximas à Barra do Garças e Rondonópolis, onde houve o avanço de casos, principalmente a partir do mês de abril. E na parte sudoeste do estado o agrupamento de cidades com o aumento de casos de Covid-19 nas proximidades do município de Cáceres com menor raio de distância média ponderada.

4.CONCLUSÕES

Com a análise a partir da média móvel de casos no Estado de Mato Grosso foi possível constatar visível avanço da doença no Estado a partir da primeira dezena do mês de junho. Como resultado das análises do deslocamento dos Centros Geográficos Ponderados, uma das Regiões Geográficas Intermediárias mais atingidas foi a de Cuiabá, com maior número de casos, mas com um menor raio de distância média ponderada, ou seja, com a concentração da incidência da doença na capital e cidades próximas. Os maiores raios de distância média ponderada foram apresentados nas Regiões Intermediárias de Sinop, claramente alinhando-se com o trajeto da Rodovia BR-163 e a de Barra do Garças, com o direcionamento seguindo a Rodovia MT-158, nas proximidades das cidades de Água Boa e Ribeirão Cascalheira. A Região Intermediária de Rondonópolis seguiu o mesmo padrão da de Cuiabá, menor raio de distância média ponderada e a de Cáceres apresentou maior dispersão na direção Oeste do Estado.

A análise da evolução dos casos e óbitos, com dados acumulados do mês de março ao início do mês de julho, nos diferentes municípios do Estado de Mato Grosso demonstrou que as maiores incidências de contaminação e óbitos ocorreram em cidades mais populosas, de maior influência econômica, geralmente polos econômicos da região a qual faz parte, destacando-se a capital, os principais Centros Sub Regionais do Estado (REGIC, 2018), principais

idades polos das Regiões Intermediárias e Imediatas (IBGE,2017), constituídas pelas cidades de Várzea Grande, Sinop, Rondonópolis, Cáceres, Barra do Garças, Sorriso, Lucas do Rio Verde, Tangará da Serra, Primavera do Leste e Pontes e Lacerda. São cidades que atuam como polos atrativos, com maior concentração e circulação de pessoas. Fato que também corroborou para a disseminação da doença Covid-19 pelo Estado de Mato, foram as flexibilizações das legislações quanto às medidas restritivas da circulação de pessoas, que haviam sido emitidas ainda no mês de março e abril, seguindo os preceitos das Legislações Federais e orientações da OMS.

Na perspectiva socioeconômica comprovou-se que a população idosa representa um dos grupos mais sujeitos à infecção e sintomas da Covid-19, com maior quantidade de óbitos a partir da quinta década de vida no Estado de Mato Grosso.

Os dados de casos e óbitos analisados neste estudo tiveram como base os boletins da Secretaria Estadual de Saúde - SES/MT, divulgados diariamente, conforme dados repassados pelos municípios, de acordo com o número de testes realizados, o que nem sempre corresponde ao número ideal de testagem, ou por falta de condições financeiras para adquirir os testes ou por outros motivos, e em alguns casos resultando na subnotificação, o que naturalmente refletiu na representação em gráficos e espacialização da informação. Contudo, a partir dessas análises foi possível identificar as regiões do Estado que apresentam maiores concentrações de casos e como estão se dispersando ao longo do tempo e do espaço nas mesmas, servindo como suporte na compreensão da dinâmica dessa pandemia no Estado, pois se trata de um fenômeno recente, que precisa de diferentes estudos que viabilizem também o entendimento de como se dispersa no espaço geográfico.

O método por meio do Centro Geográfico Ponderado, C_g , e da distância padrão ponderada, d_w , auxilia no entendimento do padrão de difusão espacial da Covid-19. Este tipo de análise fornece uma nova abordagem para avaliar a tendência da dinâmica espacial a partir do ponto de vista estatístico podendo ser aplicado para o estudo de outras doenças infecciosas. O uso de estatísticas espaciais usando SIG pode ser uma ferramenta poderosa para analisar e estudar a disseminação de doenças.

5.REFERENCIAS BIBLIOGRÁFICAS

IBGE. Divisão regional do Brasil em regiões geográficas imediatas e regiões geográficas intermediárias. Rio de Janeiro: IBGE, 2017.

COTA, W. (2020). Monitoring the number of COVID-19 cases and deaths in brazil at municipal and federative units level. SciELOPreprints:362, 10.1590/scielopreprints.362.

BLONDEL, Vincent D.; GUILLAUME, Jean-Loup, LAMBIOTTE, Renaud; LEFEBVRE, Etienne (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment,n.10, P10008.

DINIZ, Alexandre Magno Alves (2000). Estatística Espacial. (Apostila). UFMG. Disponível em: <http://csr.ufmg.br/geoprocessamento/publicacoes/estespacial.pdf>. Acesso em 1/06/2020.

RAFFESTIN, Claude (1993). Por uma geografia do poder. Tradução de Maria Cecília França. São Paulo: Ed. Ática.

SHOBUGAWA, Y; WIAFE, AS; SAITO, R, et al. (2012). Novel measurement of spreading pattern of influenza epidemic by using weighted standard distance method: retrospective spatial statistical study of influenza, Japan, 1999-2009. Int J Health Geogr. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3495731/pdf/1476-072X-11-20.pdf> . Acesso em 1/6/2020.

DONG, W., YANG, K., XU, Q. et al. (2017). Spatio-temporal pattern analysis for evaluation of the spread of human infections with avian influenza A(H7N9) virus in China, 2013–2014. BMC Infect Dis 17, 704. Disponível em: <https://bmcinfectdis.biomedcentral.com/track/pdf/10.1186/s12879-017-2781-2>. Acesso em 1/6/2020.

FANG, L. Q., DE VLAS, S. J., FENG, D., LIANG, S., XU, Y. F., ZHOU, J. P., RICHARDUS, J. H., CAO, W. C. (2009). Geographical spread of SARS in mainland China. Tropical medicine & international health : TM & IH, 14 Suppl 1(Suppl 1), 14–20. Disponível em: <https://doi.org/10.1111/j.1365-3156.2008.02189.x>. Acesso em 1/6/2020.

FERREIRA, M. C. (2003). Procedimento metodológico para modelagem cartográfica e análise regional de epidemias de dengue em sistema de informação geográfica. (Tese de Livre-Docência). DGEO-IG/UNICAMP, Campinas.

FERREIRA, M. C. (2014). Iniciação à Análise Geoespacial: Teoria, Técnicas e Exemplos para Geoprocessamento (1a ed.). São Paulo: Editora da UNESP.

SOUZA, Marcus Vinicius Mariano de Souza; FERREIRA JÚNIOR, D.B. (2020). Rede urbana, interações espaciais e a geografia da saúde: análise da trajetória da COVID-19 no estado do Pará. ESPAÇO E ECONOMIA, v. IX, p. 1-12. Disponível em: <https://doi.org/10.4000/espacoeconomia.13146> . Acesso em 1/06/2020.

SPOSITO, Maria Encarnação Beltrão; GUIMARÃES, Raul Borges (2020). Por que a circulação de pessoas tem peso na difusão da pandemia. Disponível em: <https://www2.unesp.br/portal#!/noticia/35626/por-que-a-circulacao-de-pessoas-tem-peso-na-difusao-da-pandemia> . Acesso em 1 junho de 2020.

COVIDMETRIKA: APLICATIVOS EM SHINY PARA MONITORAMENTO DA COVID-19

Gustavo Machado Utpott

gustavo.utpott@gmail.com

Juliana Sena de Souza

julianass.estadistica@gmail.com

Gabriel Holmer Saul

gabrielholmersaul@gmail.com

Márcia Helena Barbian

mhbarbian@ufrgs.br

Rodrigo Citton Padilha dos Reis

citton.padilha@ufrgs.br

Universidade Federal do Rio Grande do Sul

Resumo: Com o surgimento do novo Coronavírus, profissionais e pesquisadores em todo o mundo e de diversas áreas do conhecimento se mobilizaram a fim de mitigar os danos causados pela pandemia. Nesse contexto, um grupo de estudantes e professores do Departamento de Estatística e Programa de Pós-Graduação em Epidemiologia da Universidade Federal do Rio Grande do Sul decidiram formar o CovidMetrika, com o objetivo de descrever os casos de COVID-19 temporal e espacialmente, assim como monitorar a situação dos leitos hospitalares no estado do Rio Grande do Sul. O grupo criou uma série de painéis de monitoramento da doença, com o objetivo de auxiliar gestores de saúde na tomada de decisão. Os painéis e aplicativos foram desenvolvidos com ferramentas de código aberto, e todos os resultados do grupo estão disponíveis em repositórios públicos.

Palavras-chave: Análise Espacial, COVID-19, Monitoramento, R, Shiny

Abstract: With the emergence of the newly discovered Coronavirus, professionals and researchers worldwide and from different fields of expertise have mobilized towards reducing the damage caused by the pandemic. In this context, a group of students and professors from the Department of Statistics and Post-Graduate Program in Epidemiology at the Federal University of Rio Grande do Sul decided to form CovidMetrika, with the objective of describing the cases of COVID-19 temporally and spatially, as well as how to monitor the situation of hospital beds in the state of Rio Grande do Sul. The group created a series of dashboards to monitor the disease with the purpose to help health managers in the decision making process. The applications and the

dashboards were developed with open source tools, hence all the results of the group are available in public repositories.

Keywords: Spatial Analysis, COVID-19, Monitoring R, Shiny

1. INTRODUÇÃO

Em dezembro de 2019 foram identificados pacientes com pneumonia por causas desconhecidas em Wuhan, província de Hubei, na China. Posteriormente, em 10 de Janeiro de 2020, foi noticiada a primeira morte relacionada a esses casos. Através do uso de sequenciamento das amostras dos pacientes foi-se descoberto um *betacoronavírus* até então desconhecido, o SARS-CoV-2 (World Health Organization, 2020b, Zhu et al., 2020), que causa a COVID-19 (do inglês *Coronavirus Disease 2019*), uma doença infecciosa causada pelo novo coronavírus da síndrome respiratória aguda grave (SRAG) (World Health Organization, 2020a). Em termos de taxonomia, pertence aos coronavírus (Gorbalenya et al., 2020), um grupo de vírus de ácido ribonucléico (*RiboNucleic Acid*, RNA) de cadeia simples e positiva, com um dos maiores genomas entre os vírus de RNA (Lai and Cavanagh, 1997) e que pode causar doenças em humanos e animais. Em humanos, nos últimos 20 anos se tem registros de outras duas doenças de origem zoonótica, além da COVID-19, causadas pelo coronavírus: a síndrome respiratória aguda grave (SARS, do inglês *Severe Acute Respiratory Syndrome*) e a síndrome respiratória do Oriente Médio (MERS, do inglês *Middle East respiratory syndrome coronavirus*) (Meo et al., 2020).

As características clínicas da COVID-19 variam individualmente, podendo existir indivíduos que são assintomáticos ou que possuem sintomas de doença respiratória leves a graves. Entre as características biológicas, o tempo de incubação da doença varia entre 2 a 14 dias, com média de 5,2 dias (Meo et al., 2020). Dados clínicos e epidemiológicos do Centro de Controle e Prevenção de Doenças chinês e sobre 72.314 registros de casos (Wu and McGoogan, 2020) apontaram que cerca de 20% dos pacientes desenvolvem forma grave da doença, com dificuldades respiratórias e necessidade de hospitalização. O grupo de alto risco para desenvolver formas graves da doença são pacientes com problemas pré-existentes, como por exemplo, hipertensão, diabetes, câncer ou problemas do coração, além de pacientes idosos (World Health Organization, 2020a).

Por ser uma doença infecciosa, a transmissão pode ocorrer por três métodos básicos: pela fonte de infecção, rota de transmissão e suscetibilidade (Sudan et al., 2020). A propagação da COVID-19 é alta (Meo et al., 2020) e heterogênea, o que indica que a maioria das transmissões se dá por uma pequena parcela dos casos, os *superspreaders* (Beldomenico, 2020). Até a data de 14 de Julho, globalmente, foram identificados 12.964.809 casos confirmados e 570.288 óbitos relacionadas à doença (World Health Organization, 2020c), com uma taxa de letalidade de 4,4%. Destes, 1.926.824 e 74.133 casos confirmados e óbitos, respectivamente, são no Brasil (Ministério da Saúde, 2020).

Com o crescente número de casos no país, nos últimos meses pesquisadores de diversas áreas do conhecimento vem realizando um enorme esforço em pesquisas e projetos com o objetivo de trazer informações acerca da doença, a nível nacional e regional. Em Rezende et al. (2020) foi estimado que entre 53 a 86 milhões de adultos Brasileiros apresentam pelo menos um fator de risco para a doença, com alta prevalência de condições graves entre os idosos. Souza et al. (2020) argumentam sobre a importância da Atenção Primária à Saúde (APS) no enfrentamento da pandemia e Prata et al. (2020) discorrem acerca da relação entre a temperatura e número de casos confirmados nas capitais do Brasil.

Devido à falta de tratamentos e prevenções, a pandemia deve atingir números ainda mais preocupantes que, se não forem controlados, podem causar colapsos nos sistemas de saúde, além da perda de vidas. Logo, é preciso traçar estratégias de combate ao surto, para colaborar na elaboração de políticas que auxiliem a tomada de decisão. Tais estratégias devem se basear em informação, por isso os métodos estatísticos, combinando dados e modelos, são ferramentas valiosas no combate à doença, já que tais métodos geram informações e descrevem a incerteza com respeito aos seus achados.

Entre as várias possibilidades que podem ser estudadas, uma importante característica a ser considerada na estimação do número de casos e na alocação de recursos é a diversidade dos municípios e de suas características, como abordado em relatório divulgado pela Fundação Oswaldo Cruz (Codeço et al., 2020), que aponta o fluxo aéreo como fator importante na estimação do número de casos. Outras características a serem avaliadas é a proporção de pessoas em idade de risco, tamanho da população e proporção de

aglomerações urbanas nos municípios. Como a doença é contagiosa, cidades em que a população esteja mais concentrada terá maior risco de contágio.

O fator espacial também é relevante quando consideramos recursos de assistência à saúde. Centros urbanos possuem um maior número de leitos complementares (unidade de tratamento intensivo e unidade intermediária), sendo referência para cidades próximas e sem a mesma infraestrutura. Nesse sentido, a análise espacial já vindo sendo comumente utilizada em análises epidemiológicas, em que a variação geográfica em taxas de incidência ou de mortalidade podem ser realizadas através do mapeamento de doenças. Esses mapas são instrumentos relevantes para apontar associações entre fontes potenciais de contaminação e áreas de risco elevado, sugerir determinantes locais de doenças e visualizar a distribuição geográfica da mesma. Assim como, importante ferramenta descritiva de monitoramento.

Com isto, dado o início da quarentena e o agravamento da pandemia da COVID-19 no Brasil, estudantes e professores do Departamento de Estatística e do Programa de Pós-Graduação em Epidemiologia da Universidade Federal do Rio Grande do Sul decidiram formar o **CovidMetrika**, um grupo de estudos cujo objetivo é desenvolver e implementar um sistema de vigilância espacial de casos de COVID-19. Foram desenvolvidos aplicativos para web e celular que fornecem visualizações da dinâmica da doença de forma espacial, temporal e espaço-temporal a nível municipal, estadual e nacional. Um dos focos dos aplicativos é auxiliar os gestores públicos a entender o espalhamento da doença e como os números de casos e a ocupação de leitos de UTI são distribuídos, o que colabora nas políticas de distribuição de recursos para cuidados e atenção à saúde. Outro objetivo é fornecer à população em geral uma informação qualificada do espalhamento de casos e ocupação de leitos em seu estado ou região.

2. MÉTODOS

Os projetos do **CovidMetrika** foram desenvolvidos em código R Core Team (2019), sob a interface do RStudio (2018).

2.1 FONTE DOS DADOS

Para a confecção dos painéis do **CovidMetrika** utilizou-se diversas bases de dados, tanto para acompanhar o avanço da doença nas regiões de interesse, quanto para avaliar a situação dos leitos de unidade de tratamento intensivo (UTI) no Rio Grande do Sul (RS) e em Porto Alegre.

2.1.1 PREFEITURA DE PORTO ALEGRE

Inicialmente, a lista de casos positivos para COVID-19 era extraída do boletim epidemiológico disponibilizado pela Prefeitura de Porto Alegre (2020a). O arquivo em formato *pdf* possui uma lista de dados referentes às seguintes variáveis: identificação, data de notificação, sexo, idade e fonte de notificação. Portanto, tinha-se os dados do avanço da doença a nível de indivíduo, com informação da fonte notificadora, o que permitiu apresentar uma visualização espacial dos casos na cidade. O boletim epidemiológico de Porto Alegre teve sua primeira versão em 23 de março e utilizou-se desta fonte para extração dos dados sobre a doença de 30 de março até 11 de maio.

Além dos casos, a prefeitura também disponibiliza os dados sobre a situação das emergências (Prefeitura de Porto Alegre, 2020b) e UTIs (Prefeitura de Porto Alegre, 2020c), as informações são referentes aos números de leitos ocupados, disponíveis e quantidade de leitos ocupados com pacientes com COVID-19 em cada hospital da capital.

2.1.1 BRASIL.IO

O Brasil.io (2020) é um repositório de dados públicos disponível em <https://brasil.io/home/>, em que bancos de dados relacionados a assuntos de interesse público estão disponíveis de forma aberta, acessível e estruturada. Com o início do surto no Brasil, o Brasil.io (2020) tem organizado dados relacionados à doença, compilando diariamente informações retiradas dos boletins epidemiológicos das 27 Secretarias Estaduais de Saúde (SES), concatenando a série histórica de casos e óbitos confirmados por município no Brasil. Outras bases de dados disponíveis são a população dos municípios afetados e óbitos suspeitos registrados em cartório.

Para o Painel do Brasil utilizou-se de dois conjuntos de dados do repositório em questão. O primeiro é a base com casos e óbitos confirmados de COVID-19 por município. O segundo é a base com óbitos registrados nos cartórios, disponíveis no Portal da Transparência do Registro Civil.

2.1.2 SECRETARIA ESTADUAL DE SAÚDE DO RIO GRANDE DO SUL

A SES-RS (2020c) começou a disponibilizar os dados da COVID-19 no estado a nível de indivíduo a partir do início de junho, através desse banco é possível ter acesso à faixa etária, sexo, data de confirmação, bairro, município,

tipo de teste utilizado, sintomas mais frequentes apresentados e comorbidades dos casos confirmados.

Além dos casos a nível de indivíduo, há um painel com dados dos hospitais do estado (SES-RS, 2020b), fornecendo informações acerca dos leitos clínicos e de UTI. Os dados de leitos de UTI adulto foram extraídos dessa fonte entre abril e fim de junho. Mais recentemente, o estado reuniu todas essas informações de forma estruturada, em um novo site, tornando possível a automatização do processo de coleta do banco (SES-RS, 2020a).

2.2 TRATAMENTO DOS DADOS

Como os dados são extraídos de diversas fontes e formatos, não existe uma padronização nas variáveis, portanto é necessário realizar um tratamento dessas informações para unificação dos conjuntos de dados.

2.2.1 BOLETIM EPIDEMIOLÓGICO DE PORTO ALEGRE

Para a leitura dos dados do boletim da prefeitura utilizou-se funções disponíveis no pacote *pdftools* (Ooms, 2020), que auxilia na extração de texto de arquivos em formato *pdf*. Além disso, um arquivo com as respectivas latitudes e longitudes das fontes notificadoras foi elaborado pelos autores, para que essas informações pudessem ser visualizadas no mapa da cidade.

As maiores dificuldades no tratamento desse conjunto de dados eram as mudanças ocasionais no *layout* da página, que afetavam diretamente a leitura do algoritmo, fazendo-se necessária a verificação manual de alguns detalhes do código diariamente. Outra adversidade era a inconsistência de nomes usados para as fontes notificadoras, pois algumas das unidades de saúde eram registradas de várias maneiras.

2.2.2 PAINÉIS SITUAÇÃO HOSPITALAR DE PORTO ALEGRE

As informações dos painéis da situação hospitalar em Porto Alegre são extraídas diariamente de forma não automatizada, o site com essas informações é formado por uma tabela com os nomes das unidades de saúde, o número de leitos ocupados e disponíveis. Não há série histórica dessas quantidades, visto que os dados são atualizados diariamente pelos responsáveis de cada hospital. Assim como nos dados do boletim da prefeitura, utilizou-se do mesmo arquivo

criado manualmente com latitudes e longitudes das unidades de saúde, permitindo a visualização espacial do número de leitos ocupados.

2.2.3 BASE DE DADOS DO BRASIL.IO

Os dados provenientes do Brasil.io têm seu processo de leitura e tratamento simplificado e automatizado, os bancos são estruturados e disponibilizados em formato *csv*. Os dados foram agrupados por municípios e estados, para a visualização espacial utilizou-se os *shapefiles* das malhas geográficas disponíveis no site do Instituto Brasileiro de Geografia e Estatística (IBGE, 2015).

2.2.4 PAINEL CORONAVÍRUS SES-RS

Os dados da COVID-19 disponibilizados pela SES-RS também são, desde junho, de fácil acesso e em formato *csv*. Além dos *shapefiles* extraídos do (IBGE, 2015), os dados foram agrupados de acordo com as regiões COVID (Governo do Rio Grande do Sul, 2020), uma reformulação das regiões de saúde, desenvolvida pelo estado para monitoração do distanciamento controlado do Rio Grande do Sul. Recentemente foi adicionado uma nova informação aos dados da Secretaria de Saúde do Rio Grande do Sul, referente ao bairro do indivíduo infectado, que foi agregada ao *dashboard* da cidade de Porto Alegre.

2.2.5 PAINEL SITUAÇÃO HOSPITALAR SES-RS

A ocupação de leitos de UTI Adulto por hospital no Rio Grande do Sul era extraída do site (SES-RS, 2020b) no formato *csv*. A SES-RS não disponibilizava a série histórica da ocupação de leitos, portanto era possível obter apenas os dados do dia de acesso ao banco. As coordenadas referentes à latitude e longitude de cada hospital do estado foram obtidas no Cadastro Nacional de Estabelecimentos de Saúde (CNES, 2020).

Atualmente o painel da SES-RS fornece um arquivo em formato *csv*, que contém uma série histórica diária de ocupação de leitos, que é composta pelos dados referentes aos 14 dias anteriores à data de acesso ao site.

2.3 PAINÉIS E HOSPEDAGEM PELO SHINY.IO

Destaca-se como principais pacotes do R necessários para a criação dos aplicativos desenvolvidos: *shiny* (Chang et al., 2019) para o desenvolvimento dos dashboards; *ggplot2* (Wickham, 2016) e *plotly* (Sievert, 2020) utilizados na confecção dos gráficos interativos e o pacote *leaflet* (Cheng et al., 2018) para gerar os mapas apresentados nos painéis. Além dos pacotes citados acima, diversas bibliotecas do R são carregadas nos algoritmos dos aplicativos desenvolvidos pelo **CovidMerika** (ver Apêndice 1, Tabela 1), indicando em qual dos painéis eles foram utilizados, se no **Painel de Porto Alegre** (POA), **Painel do Rio Grande do Sul** (RS) ou **Painel do Brasil** (BR).

A publicação *online* dos aplicativos é feita pelo *shinyapps.io*, uma plataforma do Rstudio (2018) que permite aos usuários do R hospedarem suas aplicações em um servidor *web*. Para utilizar a ferramenta, não é necessário conhecimento de linguagens de desenvolvimento *web*, apenas saber programar em R e utilizar o pacote *Shiny* (Chang et al., 2019). Uma facilidade da plataforma é a capacidade do servidor, possibilitando que vários usuários acessem o aplicativo simultaneamente. O *shinyapps.io* possui uma conta gratuita em que é possível hospedar até 5 *websites* com uso limitado de até 25 horas mensais de uso. A plataforma também disponibiliza planos pagos por meio de uma assinatura mensal ou anual em que é possível aumentar o número de horas, aumentar o número de aplicativos disponíveis, mudar as configurações do servidor, além de outros serviços.

Durante a pandemia de COVID o *shinyapps.io* ofereceu 6 meses do plano básico gratuitamente para aplicativos relacionados à doença, esse plano também possuía número de horas de acesso ilimitado. O **CovidMetrika**, aderiu a esse plano especial e atualmente está hospedando os seus 3 painéis em *websites* do *shinyapps.io*.

3. COVIDMETRIKA

Até o momento, dentro do **CovidMetrika** foram desenvolvidos três painéis voltados para a descrição dos casos de COVID-19 de forma temporal, espacial e espaço-temporal a nível municipal, regional e nacional: Painel de Porto Alegre, Painel do Rio Grande do Sul e Painel do Brasil.

3.1 PAINEL DE PORTO ALEGRE

O **Painel de Porto Alegre**¹ foi o primeiro projeto desenvolvido por alguns participantes do **CovidMetrika**, anterior à criação do grupo, que foi posteriormente incorporado ao portfólio do mesmo. O painel é composto por cinco abas: leitos adultos (Figura 1), casos por bairro (Figura 2), leitos pediátricos, fontes de dados e informações do grupo.

Em relação aos leitos adultos, os usuários podem selecionar as informações que desejam visualizar. Na Figura 1 é possível identificar três tipos de filtros: os hospitais de interesse, a data de atualização e o tipo de leito - UTI ou emergência. Para este último, há a possibilidade de selecionar as duas opções. O padrão é considerar todos os hospitais, a data mais recente de atualização disponível e os leitos de UTI.

Figura 1 - Aba de Leitos - Adultos do Painel de Porto Alegre



A segunda aba traz a distribuição dos casos de COVID-19 na cidade de Porto Alegre, de acordo com os bairros da capital. Pela Figura 2 é possível perceber que há oito informações distintas que o usuário pode ter acesso. O primeiro filtro é em relação à variável que será apresentada, em que é possível escolher entre número total de casos confirmados, ou selecionar as variáveis que dizem respeito a evolução da doença, como número de óbitos, número de pessoas recuperadas e em acompanhamento. No segundo filtro é possível selecionar se os resultados serão mostrados em termos de números absolutos ou taxa por 100.000 habitantes.

¹ Disponível em: https://mhbarbian.shinyapps.io/covid19_poa/

Figura 2 - Aba de Casos por Bairro do Pannel de Porto Alegre



3.2 PAINEL DO RIO GRANDE DO SUL

O **Painel do Rio Grande do Sul**² é similar ao de Porto Alegre, pois ele é composto por duas abas principais: leitos UTI - Adulto (Figura 3) e COVID-19 (Figura 5).

Na aba de leitos de UTI - Adulto é possível visualizar as principais informações relacionadas à situação dos leitos, como número de leitos totais, leitos disponíveis, porcentagem de ocupação e leitos de pacientes confirmados com COVID-19 (Figura 3). Além disso, é permitido ao usuário escolher entre hospital, município, ou região COVID - caso o usuário deseje que o *dashboard* mostre os dados de alguma região específica.

Por exemplo, é possível selecionar somente a região Bagé-R22, conforme indicado na Figura 4, fazendo com que o aplicativo desconsidere os dados da capital e do resto do estado.

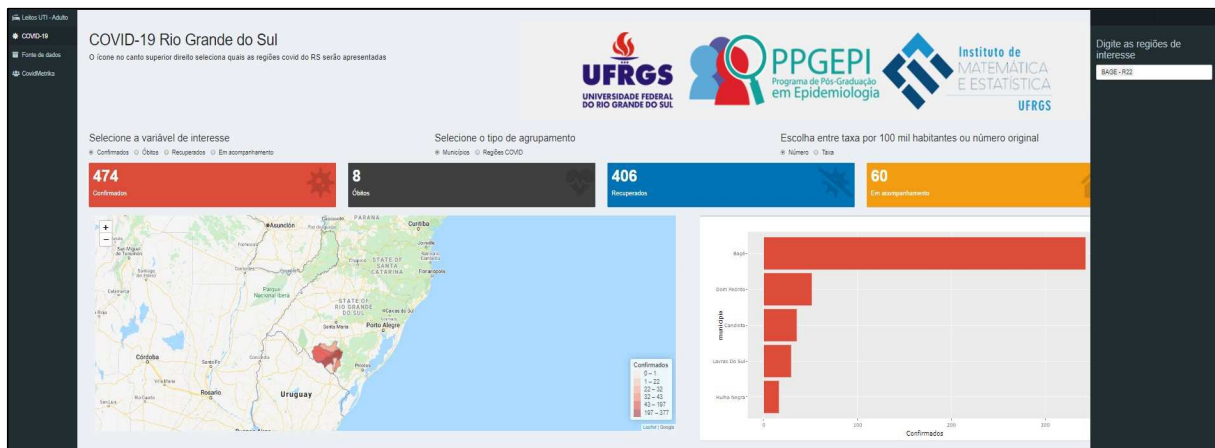
² Disponível em: https://mhbarbian.shinyapps.io/covid19_rs/

Figura 3 – Aba de Leitos - Adultos do Painel do Rio Grande do Sul



Na segunda aba são apresentados dados sobre a COVID-19 por município, em que há possibilidade do usuário escolher entre o tipo de variável (confirmados, óbitos, recuperados e em acompanhamento), o agrupamento (municípios e regiões COVID) e número absoluto ou taxa por 100.000 habitantes, como indicado na Figura 5.

Figura 4 – Aba de Regiões COVID do Painel do Rio Grande do Sul



As duas abas restantes, assim como no painel de Porto Alegre, trazem informações das fontes de dados utilizadas e do grupo **CovidMetrika**.

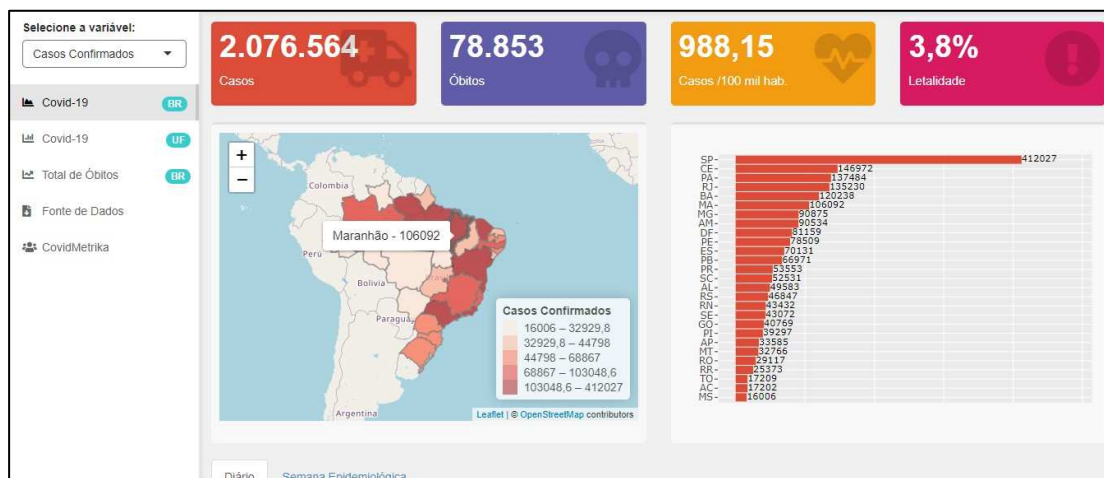
Figura 5 – Aba de COVID-19 do Painel do Rio Grande do Sul



3.3 PAINEL DO BRASIL

O **Painel do Brasil**³ traz dados sobre a COVID-19 em território nacional. O painel é composto por três abas principais: COVID-19 (BR) (Figura 6), COVID-19 (UF) (Figura 7) e total de Óbitos (Figura 8).

Figura 6 – Aba de COVID-19 no país do Painel do Brasil

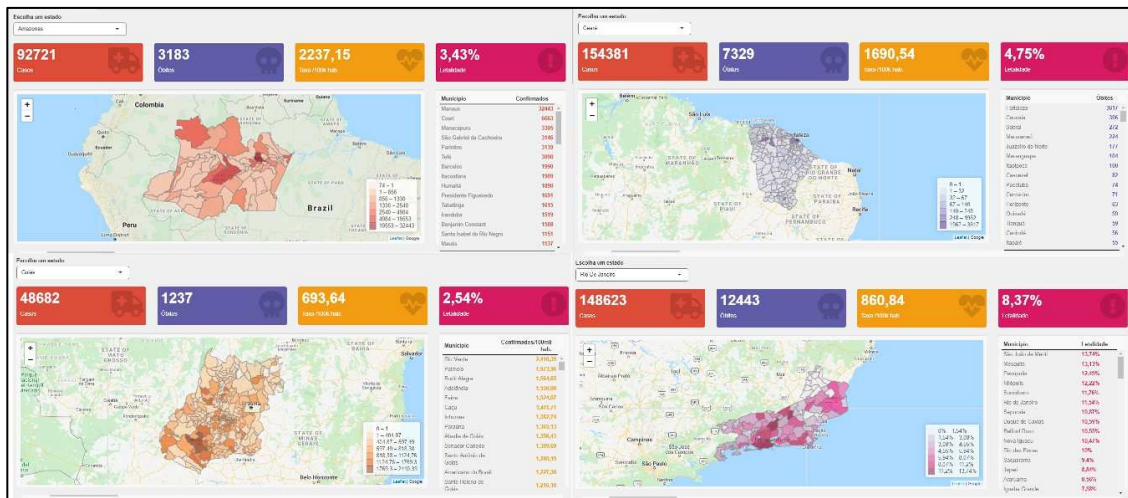


Na primeira aba é possível visualizar informações relacionadas à doença, agrupadas por estado, como o número total de casos confirmados, óbitos, casos confirmados por 100 mil habitantes e letalidade. A série histórica é apresentada em termos de dados diários e por semana epidemiológica, a contar do primeiro caso confirmado no país.

³ Disponível em: <https://mhbarbian.shinyapps.io/CovidMetrika/>

Na segunda aba é possível escolher o estado a qual o usuário deseja ter informações à nível municipal, conforme Figura 7.

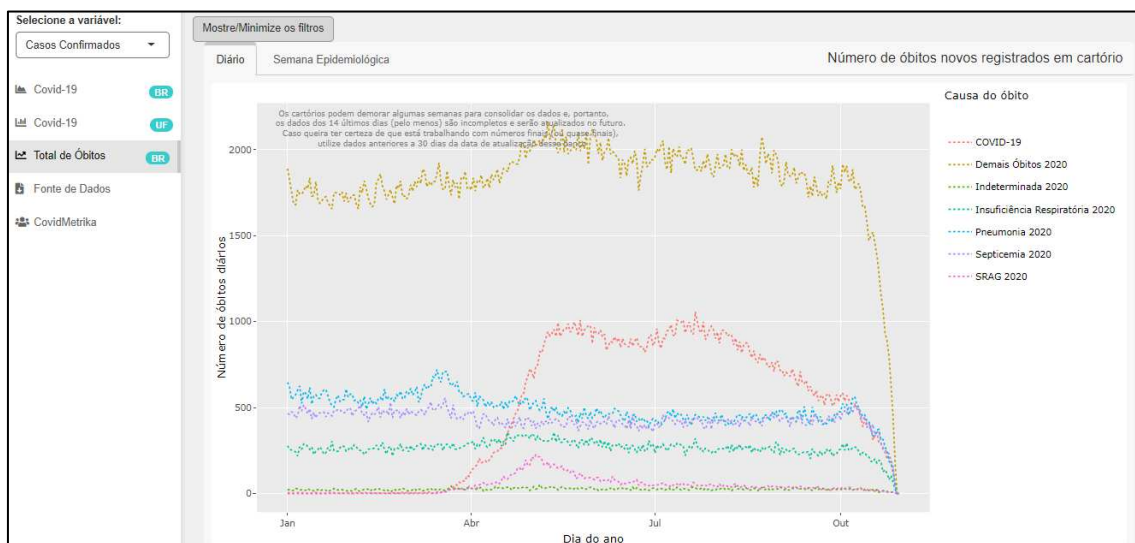
Figura 7 – Aba de COVID-19 nos estados do Painel do Brasil



Em total de óbitos é apresentado os dados sobre as causas de mortes relacionadas à óbitos registrados nos cartórios de registro civil no ano de 2019 e 2020, conforme Figura 8. Foi inserido um filtro que permite selecionar as principais causas de morte relacionadas a doenças respiratórias, outras causas e o número total de óbitos. Além disso, também é possível filtrar os estados de interesse do usuário.

É importante frisar que os cartórios podem demorar algumas semanas para consolidar os dados. Portanto, os óbitos registrados dos últimos 14 dias poderão sofrer alterações, de acordo com novas atualizações nos cartórios.

Figura 8 – Aba de óbitos registrados em cartório do Painel do Brasil



4. ACESSÓRIOS

O **CovidMetrika** utiliza de outras plataformas para ajudar na divulgação e monitoramento dos projetos desenvolvidos pelo grupo.

4.1 SITE COVIDMETRIKA

O website do CovidMetrika⁴ está hospedado em um domínio da Universidade Federal do Rio Grande do Sul (UFRGS) e foi construído utilizando o pacote *blogdown* (Xie et al., 2017) do *software* R, que permite a construção de *sites* utilizando o *rmarkdown* (Xie et al., 2018) e o gerador de sites estáticos Hugo⁵.

4.2 APLICATIVO PARA CELULAR

Além do *website*, o grupo também conta com um aplicativo para celular chamado **CovidMetrika**, disponível no Google Play⁶. O aplicativo foi construído utilizando o *MIT App Inventor*⁷. Nele, criou-se um projeto com apenas um bloco *WebView*, de modo que toda a tela do dispositivo é preenchida. Na *HomeUrl* desse bloco foi inserida a *url* do site do **CovidMetrika**, então quando o usuário acessa o aplicativo no celular ele é redirecionado para a página do grupo. Essa é uma forma simples de fazer com que aplicações hospedadas no *shinyapps.io* possam ser transformados em aplicativos para celular.

O aplicativo deverá ser salvo no formato *.apk* para ser disponibilizado em serviços de distribuição digital de aplicativos. Para obter esse arquivo por meio do *MIT App Inventor*, selecione a aba *Build* no projeto e escolha a opção *App (save.apk to my computer)*, com isso é possível salvar localmente o arquivo *.apk*.

Para publicar o aplicativo no *Play Store* é necessário criar uma conta de desenvolvedor na plataforma e submeter o arquivo *.apk* em um projeto no *Google Play Console*. Após submeter o código do aplicativo, o *Google* possui um processo de avaliação do arquivo *.apk*, para garantir a integridade dos usuários. No caso de apps relacionados à COVID-19, a política de avaliação de disponibilização é diferente. Nesse protocolo é necessário fornecer um documento que comprovava o vínculo do desenvolvedor com uma entidade

⁴ Disponível em: <https://www.ufrgs.br/covidmetrika>

⁵ <https://gohugo.io/>

⁶ Disponível em:

https://play.google.com/store/apps/details?id=appinventor.ai_covidmetrika.CovidMetrika

⁷ <https://appinventor.mit.edu/>

governamental, no caso do **CovidMetrika**, foi enviado um documento que mostrava o vínculo do grupo com a UFRGS.

4.3 GOOGLE ANALYTICS

É muito importante ter conhecimento da influência e alcance dos aplicativos desenvolvidos. Entre suas ferramentas, o *shinyapps.io* disponibiliza um painel ao usuário com relatórios de acesso, como horas de uso ou número de conexões por hora do *app*. Outra plataforma que possibilita o monitoramento de acessos é o *Google Analytics*, como pode ser visto na Figura 9, que disponibiliza o número total de visitantes, país de acesso, se o acesso é pelo celular ou pelo computador, quantidade de acessos semanais, tempo que cada usuário ficou logado, país da consulta, evolução semanal da quantidade de acessos, etc. A sincronização do *Google Analytics* com o *shinyapps.io* é bem simplificada. Um guia com os passos necessários pode ser consultado em Shiny from RStudio (2020). O site desenvolvido com o pacote *blogdown* (Xie et al., 2017) também pode ser monitorado.

Figura 9 - Painel do *Google Analytics* apresentando algumas métricas do aplicativo de POA



5. CONCLUSÃO

Os painéis foram desenvolvidos com o intuito de descrever os casos de COVID-19 e monitorar a situação hospitalar dos leitos de forma temporal e espacial, apresentando um painel de monitoramento simples e de fácil visualização que ainda não era disponibilizado pelo governo no início da

pandemia. Dessa forma, se promoveu o acesso aos dados relacionados à pandemia para a população em geral e, além disso, possibilita que gestores de saúde possam utilizar as ferramentas para auxiliar na tomada de decisões de medidas públicas.

Logo no início do projeto, surgiram veículos de informação que se mostraram interessados em divulgar o trabalho feito. Dentre os mais relevantes, vale ressaltar as notícias escritas no G1, gaúchaZH, UFRGS, blog do André Machado e também a matéria feita pelo programa televisivo RBS notícias.

Tudo que foi utilizado para a confecção dos painéis, como os bancos de dados e códigos, está disponibilizado de forma transparente e pública no *GitHub*⁸. O pacote *shiny* possui uma função que permite rodar os aplicativos localmente através da função *runGitHub*, sem a necessidade de baixar os arquivos do repositório. Para isso, é necessário ter o pacote *shiny* instalado na máquina e escrever no console do R *runGitHub(nome-do-repositório, nome-do-usuário-github)*, por exemplo, *runGitHub(repo = "dashboard_brasil", username = "CovidMetrika", filetype = ".tar.gz")*. Ainda, é possível que outras pessoas façam sugestões de mudanças no código, criem notificações para alerta de erros, ou para a solicitação de novos recursos nos painéis. Isso vai de encontro com a filosofia do grupo, de ciência colaborativa, em que os bancos de dados e códigos produzidos ficam disponíveis, para acesso de todos que queiram colaborar, utilizar as ferramentas apresentadas ou que queiram fazer uso do material desenvolvido em algum outro projeto.

5. REFERÊNCIAS BIBLIOGRÁFICAS

Appelhans, T. and Detsch, F. (2019). *leafpop: Include Tables, Images and Graphs in Leaflet Pop-Ups*. R package version 0.0.1. <https://CRAN.R-project.org/package=leafpop>.

Arnold, J. B. (2019). *ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'*. R package version 4.2.0. [://CRAN.R-project.org/package=ggthemes](https://CRAN.R-project.org/package=ggthemes).

Attali, D. (2018). *shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds*. R package version 1.0. <https://CRAN.R-project.org/package=shinyjs>.

Attali, D. and Edwards, T. (2020). *shinyalert: Easily Create Pretty Popup Messages (Modals) in 'Shiny'*. R package version 1.1. <https://CRAN.R-project.org/package=shinyalert>.

⁸ <https://github.com/CovidMetrika>

Bailey, E. (2015). *shinyBS: Twitter Bootstrap Components for Shiny*. R package version 0.61. <https://CRAN.R-project.org/package=shinyBS>.

Beldomenico, P. (2020). *Do superspreaders generate new superspreaders? a hypothesis to explain the propagation pattern of covid-19*. International Journal of Infectious Diseases, 96.

Bivand, R. and Rundel, C. (2019). *rgeos: Interface to Geometry Engine - Open Source ('GEOS')*. R package version 0.5-2. <https://CRAN.R-project.org/package=rgeos>.

Brasil.io (2020). *Boletins epidemiológicos da COVID-19 por município por dia*. <https://brasil.io/covid19/>. Acesso em: 16 jul. 2020.

Chang, W. and Borges Ribeiro, B. (2018). *shinydashboard: Create Dashboards with 'Shiny'*. R package version 0.7.1. <https://CRAN.R-project.org/package=shinydashboard>.

Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. (2019). *shiny: Web Application Framework for R*. R package version 1.3.2. <https://CRAN.R-project.org/package=shiny>.

Cheng, J., Karambelkar, B., and Xie, Y. (2018). *leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*. R package version 2.0.2. <https://CRAN.R-project.org/package=leaflet>.

CNES (2020). *Extração de estabelecimento*. <http://cnes.datasus.gov.br/pages/estabelecimentos/extracao.jsp>. Acesso em: 15 jul. 2020.

Codeço, C. T., Villela, D., Coelho, F., Bastos, L. S., Gomes, M. F. C., Cruz, O. G., Lana, R. M., Piontti, A. P. y., Vespignani, A., and Davis, J. T. (2020). *Estimativa de risco de espalhamento da COVID-19 no Brasil e o impacto no sistema de saúde e população por microrregião*. FIOCRUZ/PROCC, Rio de Janeiro.

Garnier, S. (2018). *viridis: Default Color Maps from 'matplotlib'*. R package version 0.5.1. <https://CRAN.R-project.org/package=viridis>.

Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L., Samborskiy, D., Sidorov, I. A., Sola, I., and Ziebuhr, J. (2020). *Severe acute respiratory syndrome-related coronavirus : The species and its viruses – a statement of the coronavirus study group*.

Governo do Rio Grande do Sul (2020). *Modelo de distanciamento do Rio Grande do Sul*. <https://distanciamentocontrolado.rs.gov.br/>. Acesso em: 16 jul. 2020.

Granjon, D. (2018). *shinyEffects: Customize Your Web Apps with Fancy Effects*. R package version 0.1.0. <https://CRAN.R-project.org/package=shinyEffects>.

Granjon, D. (2019). *shinydashboardPlus: Add More 'AdminLTE2' Components to 'shinydashboard'*. R package version 0.7.0. <https://CRAN.R-project.org/package=shinydashboardPlus>.

Grolemund, G. and Wickham, H. (2011). *Dates and times made easy with lubridate*. Journal of Statistical Software, 40(3):1–25. <http://www.jstatsoft.org/v40/i03/>.

IBGE (2015). Malhas digitais. <https://mapas.ibge.gov.br/bases-e-referenciais/bases-cartograficas/malhas-digitais>. Acesso em: 16 jul. 2020.

Lai, M. M. C. and Cavanagh, D. (1997). *The molecular biology of coronaviruses*. In Maramorosch, K., Murphy, F. A., and Shatkin, A. J., editors, *Advances in Virus Research*, volume 48, pages 1–100. Academic Press.

Lilovski, N. (2020). *dashboardthemes: Customise the appearance of Shiny dashboard applications using themes*. R package version 1.0.6. <https://github.com/nik01010/dashboardthemes>.

Meo, S., Alhowikan, A., Al-khlaiwi, T., Meo, I., Halepoto, D., Iqbal, M., Usmani, A., Hajjar, W., and Ahmed, N. (2020). *Novel coronavirus 2019-ncov: prevalence, biological and clinical characteristics comparison with sars-cov and mers-cov*. *European review for medical and pharmacological sciences*, 24:2012–2019.

Ministério da Saúde (2020). *Painel de casos de doença pelo coronavírus 2019 (COVID-19) no Brasil pelo Ministério da Saúde*. <https://covid.saude.gov.br/>. Acesso em: 15 jul. 2020.

Müller, K. (2017). *here: A Simpler Way to Find Your Files*. R package version 0.1. <https://CRAN.R-project.org/package=here>.

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes*. R package version 1.1-2. <https://CRAN.R-project.org/package=RColorBrewer>.

Ooms, J. (2019). *curl: A Modern and Flexible Web Client for R*. R package version 4.3. <https://CRAN.R-project.org/package=curl>.

Ooms, J. (2020). *pdftools: Text Extraction, Rendering and Converting of PDF Documents*. R package version 2.3.1.

Pebesma, E. (2018). *Simple Features for R: Standardized Support for Spatial Vector Data*. *The R Journal*, 10(1):439–446. <https://doi.org/10.32614/RJ-2018-009>.

Prata, D. N., Rodrigues, W., and Bermejo, P. H. (2020). *Temperature significantly changes COVID-19 transmission in (sub)tropical cities of Brazil*. 729:138862.

Prefeitura de Porto Alegre (2020a). *Boletins coronavirus*. http://www2.portoalegre.rs.gov.br/sms/default.php?p_secao=1027. Acesso em: 16 jul. 2020.

Prefeitura de Porto Alegre (2020b). *Situação das Emergências do Hospitais de Porto Alegre*. <http://bit.ly/EmergenciasSaudePOA>. Acesso em: 16 jul. 2020.

Prefeitura de Porto Alegre (2020c). *Situação das UTIs dos Hospitais de Porto Alegre*. http://bit.ly/monitoramento_uti. Acesso em: 16 jul. 2020.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. Version 3.6.0 - R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.

Rezende, L. F. M., Thome, B., Schweitzer, M. C., Souza-Júnior, P. R. B. d., and Szwarcwald, C. L. (2020). *Adults at high-risk of severe coronavirus disease-2019 (Covid-19) in Brazil*. 54:50.

RStudio (2018). *RStudio: Integrated Development Environment for R*. Version 1.1.463 - RStudio, Inc., Boston, MA. <http://www.rstudio.com/>.

SES-RS (2020a). *Monitoramento covid-19: Internações hospitalais*. <https://covid.saude.rs.gov.br/>. Acesso em: 16 jul. 2020.

SES-RS (2020b). *Monitoramento hospitalais*. <https://ti.saude.rs.gov.br/covid19/leitos/dashboard.php>. Acesso em: 16 jul. 2020.

SES-RS (2020c). *Painel Coronavírus RS*. <https://ti.saude.rs.gov.br/covid19/>. Acesso em: 16 jul. 2020.

Shang, H. L. and Hyndman, R. (2019). *rainbow: Bagplots, Boxplots and Rainbow Plots for Functional Data*. R package version 3.6. <https://CRAN.R-project.org/package=rainbow>.

Shiny from RStudio (2020). *Add google analytics to a shiny app*. <https://shiny.rstudio.com/articles/google-analytics.html>. Acesso em: 16 jul. 2020.

Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Chapman and Hall/CRC. <https://plotly-r.com>.

Souza, C. D. F. d., Gois-Santos, V. T. d., Correia, D. S., Martins-Filho, P. R., and Santos, V. S. (2020). *The need to strengthen Primary Health Care in Brazil in the context of the COVID-19 pandemic*. 34:e047.

Sudan, P., Goswami, M., and M, A. (2020). *Novel coronavirus (covid-19)-a critical review*. International Journal of Research in Pharmaceutical Sciences, 11:43– 47.

Trecenti, J., Correa, F., and Lente, C. (2019). *abjutils: Useful Tools for Jurimetrical Analysis Used by the Brazilian Jurimetrics Association*. R package version 0.2.3. <https://CRAN.R-project.org/package=abjutils>.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. R package version 3.3.2. <https://ggplot2.tidyverse.org>.

Wickham, H. (2019a). *forcats: Tools for Working with Categorical Variables (Factors)*. R package version 0.4.0. <https://CRAN.R-project.org/package=forcats>.

Wickham, H. (2019b). *httr: Tools for Working with URLs and HTTP*. R package version 1.4.1. <https://CRAN.R-project.org/package=httr>.

Wickham, H. (2019c). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.4.0. <https://CRAN.R-project.org/package=stringr>.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K.,

Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). *Welcome to the tidyverse*. Journal of Open Source Software, 4(43):1686.

Wickham, H. and Bryan, J. (2019). *readxl: Read Excel Files*. R package version 1.3.1. <https://CRAN.R-project.org/package=readxl>.

Wickham, H. and Seidel, D. (2019). *scales: Scale Functions for Visualization*. R package version 1.1.0. <https://CRAN.R-project.org/package=scales>.

World Health Organization (2020a). *Q&A on coronaviruses (COVID-19)*. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses>. Acesso em: 22 jun. 2020.

World Health Organization (2020b). *Timeline: Who's covid-19 response*. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline>. Acesso em: 22 set. 2020.

World Health Organization (2020c). *WHO coronavirus disease (COVID-19) dashboard*. <https://covid19.who.int/>. Acesso em: 14 jul. 2019.

Wu, Z. and McGoogan, J. M. (2020). Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. JAMA, 323(13):1239–1242.

Xie, Y., Allaire, J., and Golemund, G. (2018). *R Markdown: The Definitive Guide*. Chapman and Hall/CRC, Boca Raton, Florida. ISBN 9781138359338.

Xie, Y., Cheng, J., and Tan, X. (2020). *DT: A Wrapper of the JavaScript Library 'DataTables'*. R package version 0.13. <https://CRAN.R-project.org/package=DT>.

Xie, Y., Hill, A. P., and Thomas, A. (2017). *blogdown: Creating Websites with R Markdown*. Chapman and Hall/CRC, Boca Raton, Florida. <https://github.com/rstudio/blogdown>.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G. F., and Tan, W. (2020). *A Novel Coronavirus from Patients with Pneumonia in China, 2019*. New England Journal of Medicine, 382(8):727–733. PMID: 31978945.

APÊNDICE 1**Tabela 1** - Lista de pacotes utilizados na criação dos aplicativos

Pacote	Autor	Painéis
shiny	Chang et al. (2019)	POA, RS, BR
shinydashboard	Chang and Borges Ribeiro (2018)	POA, RS, BR
shinydashboardPlus	Granjon (2019)	POA, RS, BR
dashboardthemes	Lilovski (2020)	BR
shinyEffects	Granjon (2018)	POA, RS, BR
shinyalert	Attali and Edwards (2020)	POA, RS, BR
shinyBS	Bailey (2015)	POA, RS
shinyjs	Attali (2018)	BR
leaflet	Cheng et al. (2018)	POA, RS, BR
ggplot2	Wickham (2016)	POA, RS, BR
ggthemes	Arnold (2019)	RS, BR
RColorBrewer	Neuwirth (2014)	POA, RS, BR
Plotly	Sievert (2020)	POA, RS, BR
Viridis	Garnier (2018)	POA, RS
Rainbow	Shang and Hyndman (2019)	POA, RS
Lubridate	Grolemund and Wickham (2011)	POA, RS, BR
Scales	Wickham and Seidel (2019)	POA, RS, BR
DT	Xie et al. (2020)	POA, RS, BR
tidyverse	Wickham et al. (2019)	POA, RS, BR
sf	Pebesma (2018)	POA, RS, BR
stringr	Wickham (2019c)	POA, RS
readxl	Wickham and Bryan (2019)	POA, RS, BR
here	Müller (2017)	BR
rgeos	Bivand and Rundel (2019)	POA, RS

leafpop	Appelhans and Detsch (2019)	POA, RS
httr	Wickham (2019b)	POA, BR
curl	Ooms (2019)	POA, RS
abjutils	Trecenti et al. (2019)	POA, RS
forcats	Wickham (2019a)	BR

AGRADECIMENTOS

Os autores agradecem à Secretaria de Planejamento, Orçamento e Gestão do Departamento de Planejamento Governamental do Rio Grande do Sul (DEPLAN-SEPLAG-RS) por conceder os shapefiles das vinte regiões de saúde COVID do estado do Rio Grande do Sul.

CONTEXTUALIZAÇÃO DO CENÁRIO DE COVID-19 NA REGIÃO DO BAIXO TOCANTINS-PARÁ

Carla Braga Pereira

carlabpereira10@gmail.com

Universidade Federal do Pará, Campus Universitário do Tocantins/Cametá

Luisa Dias Barros

luisadbars@gmail.com

Universidade da Amazônia

Estevão José da Silva Barbosa

estevaobarbos@yahoo.com.br

Universidade Federal do Pará, Campus Universitário de Ananindeua

Resumo: O noticiário sobre o primeiro registro da Covid-19 na cidade de Wuhan, China, em dezembro de 2019 simboliza de forma peculiar o ano de 2020 em função da pandemia do novo coronavírus. Deste modo a sociedade, em nível mundial, teve que se reorganizar para evitar a transmissão da doença, alterando o hábito de vida em diversas escalas geográficas na Terra. Nesse contexto, o modo de contágio exponencial alçou o Brasil, estado do Pará e assim a parte interiorana no referido estado, atingindo a região do Baixo Tocantins, alvo de estudo do presente artigo. Em decorrência desse alcance, nota-se a importância de registrar a espacialidade geográfica do cenário pandêmico na região do Baixo Tocantins, e assim evidenciar o arcabouço desenhado pela continua passagem da Covid-19. A partir de dados de junho de 2020, o estudo será norteado em função de informações sobre casos confirmados, óbitos, letalidade, números de leitos e médicos permitindo esboçar o panorama espacial da presente pandemia, por meio de representações estatísticas e cartográficas.

Palavras-chave: Pandemia, Covid-19, Baixo Tocantins, Pará

Abstract:

The news about the first Covid-19 record in the city of Wuhan, China, in December 2019 symbolizes the peculiar shape of the year 2020 due to the new coronavirus pandemic. Thus, a society, at a global level, had to reorganize itself to prevent the transmission of the disease, altering the habit of living on different geographical scales on Earth. In this context, the exponential contagion mode reached Brazil, the state of Pará and thus the interior part of that state, reaching the region of Baixo Tocantins, the target of this article. As a result of this reach, it is noted the importance of registering the geographical spatiality of the pandemic scenario in the region of Baixo Tocantins, and thus highlighting the framework requested by the continuous passage of Covid-19. From June 2020 data, the study will be guided by information on confirmed cases, deaths, lethality, numbers of beds and doctors, allowing the spatial panorama of this pandemic to be sketched through statistical and cartographic representations.

Keywords: Pandemic, Covid-19, Baixo Tocantins, Pará

1. INTRODUÇÃO

O ano de 2020 será marcado pela pandemia do coronavírus (SARS-COV-2), caracterizada por uma síndrome respiratória aguda grave (Covid-19), cujo contágio é transmitido de humano para humano (Croda e Garcia, 2020). Por sua vez, a capacidade de transmissão a partir de casos assintomáticos, torna mais peculiar a proliferação da Covid-19 (Munster *et al.*, 2020).

O novo coronavírus foi descoberto na China, continente asiático, no final do mês de dezembro de 2019, mais precisamente na cidade de Wuhan, esta ficou conhecida na mídia internacional em virtude dos elevados casos de uma pneumonia, que até então era desconhecida, somada a preocupação eminente resultou no alerta do Organização Mundial de Saúde (OMS) (Belforte *et al.*, 2020).

Em 11 de março de 2020, a Covid-19 foi notificada a nível de pandemia, uma vez que foram registrados casos confirmados da doença em inúmeros países (Cardoso *et al.*, 2020). Infere-se o alcance do contágio, a nível global, em virtude da rede de mobilidade e migração humana. Esse Panorama sustenta-se em face à circulação aérea, que favoreceu sua difusão por vários continentes. Deste modo imperou no Brasil, primeiramente em São Paulo, de onde irradiou-se para as regiões mais vulneráveis do país (Sposito e Guimarães, 2020). Nesse contexto, Albuquerque e Pedrosa (2020) afirmam que a distribuição do COVID-19, não ocorre de forma homogênea no território brasileiro, pois os primeiros casos foram notificados nas capitais brasileiras, e posteriormente detectados em regiões mais longínquas, em detrimento da transmissão comunitária. Yang (2020) acredita que a propagação da doença irá perdurar por um longo em tempo, sustentado pela velocidade exponencial de propagação do vírus, alcance imensurável de infectados a partir da fonte primária e a dificuldade em amenizar o contágio em um grande contingente populacional.

Em pouco tempo a pandemia de Covid-19 impôs consequências devastadoras e imensuráveis para a humanidade, resultando em expressivos números de pessoas mortas pela doença, além dos impactos econômicos e sociais que são incalculáveis (Albuquerque e Pedrosa, 2020).

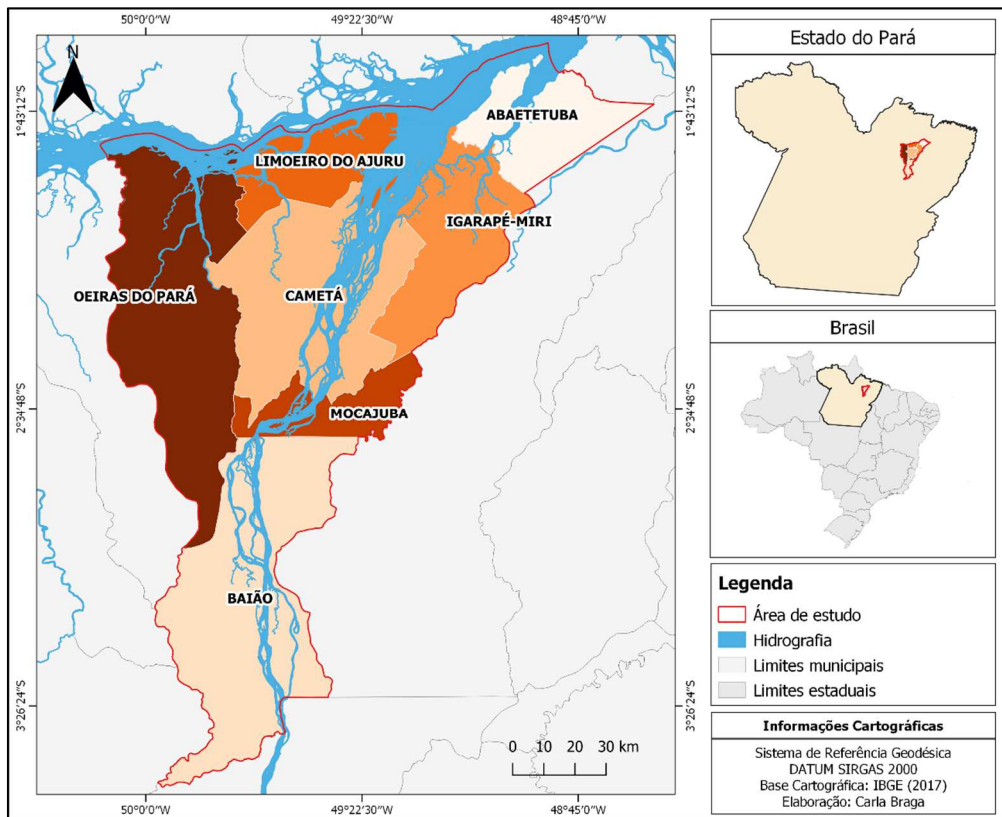
No Brasil o primeiro contágio pelo novo coronavírus foi datado em 26 de fevereiro, 21 dias depois, no dia 18 de março ocorreu a notificação da presença da Covid-19 no estado do Pará de acordo com a informação repassada pela Secretária de Saúde Pública (SESPA), enquanto isso a nível nacional já havia outros 528 casos confirmados (G1 Globo; PARÁ, 2020). O Pará foi um dos últimos estados a contar casos de contaminação no país, mais precisamente o 20º, além do Distrito Federal (BRASIL, 2020; G1 Globo, 2020).

Barbosa (2020) explica que não há como precisar a entrada do novo coronavírus no Pará, mas somente fazer indicações com base nos primeiros casos na capital do estado, Belém. Infere-se que os primeiros registros foram de casos importados de São Paulo – SP e Rio de Janeiro. A interiorização dos casos confirmados pode estar relacionada com fluxos rodoviários, aéreos e fluviais. Deste modo, todas as regiões do estado do Pará possuem registros da Covid-19, incluindo a região do Baixo Tocantins (Figura 1). Sugere-se que a proliferação neste território alcançou por meio da hidrografia, na qual os rios são tradicionais e importantes fluxos neste processo de interação das redes estaduais e municipais, bem como estradas e fluxos aéreos. A propagação por meio dos rios também é alimentada informalmente pelas inúmeras embarcações clandestinas, que apesar das restrições por decretos estaduais, são atividades comuns, pois são essenciais para atividades econômicas formais e informais, servindo como elo de interação entre municípios (Machado, 2020).

Por vários meses, o Baixo Tocantins destacou-se no cenário estadual devido os números exponenciais da propagação do novo coronavírus. O estado do Pará registrava 18 casos confirmados quando ocorreu a primeira notificação na região, datando do dia 29 de março no município de Oeiras do Pará (PARA, 2020), a Sespa não confirmou se a pessoa contraiu o vírus de caso importado ou por transmissão comunitária (G1 GLOBO, 2020).

A atual pandemia evidencia as diferentes faces resultantes da propagação de Covid-19 na região, seja no âmbito da precariedade da infraestrutura hospitalar, no número dos casos confirmados e óbitos, no comportamento do vírus nos municípios que constituem a área em estudo, entre outras variáveis que mudaram ou revelaram o cotidiano da sociedade local. Em face do exposto o presente artigo visa estudar e analisar o comportamento e a espacialidade geográfica dos dados coletados, análise da infraestrutura hospitalar da região, para poder ajudar a compreender a dinâmica do vírus e suas consequências no atual panorama de propagação do mesmo no Baixo Tocantins. A metodologia foi norteada a partir da fonte de dados coletada na base da Secretaria Estadual do Pará (SESPA), boletins epidemiológicos das prefeituras municipais, DataSus, Fiocruz, tais informações serviram de arcabouço para embasar os pontos analisados e interpretados no presente artigo. Posteriormente, parte dos dados foi espacializado no Sistema de Informação Geográfica (SIG) via software QGIS e computados gráficos utilizando o programa excel.

Figura 1 - Mapa de localização da área de estudo, compreendendo a região do Baixo Tocantins, a qual é constituída pelos municípios de Abaetetuba, Baião, Cametá, Igarapé-Miri, Limoeiro do Ajuru, Mocajuba e Oeiras do Pará



2. CONTEXTUALIZAÇÃO GEOGRÁFICA A PARTIR DOS DADOS DA COVID-19

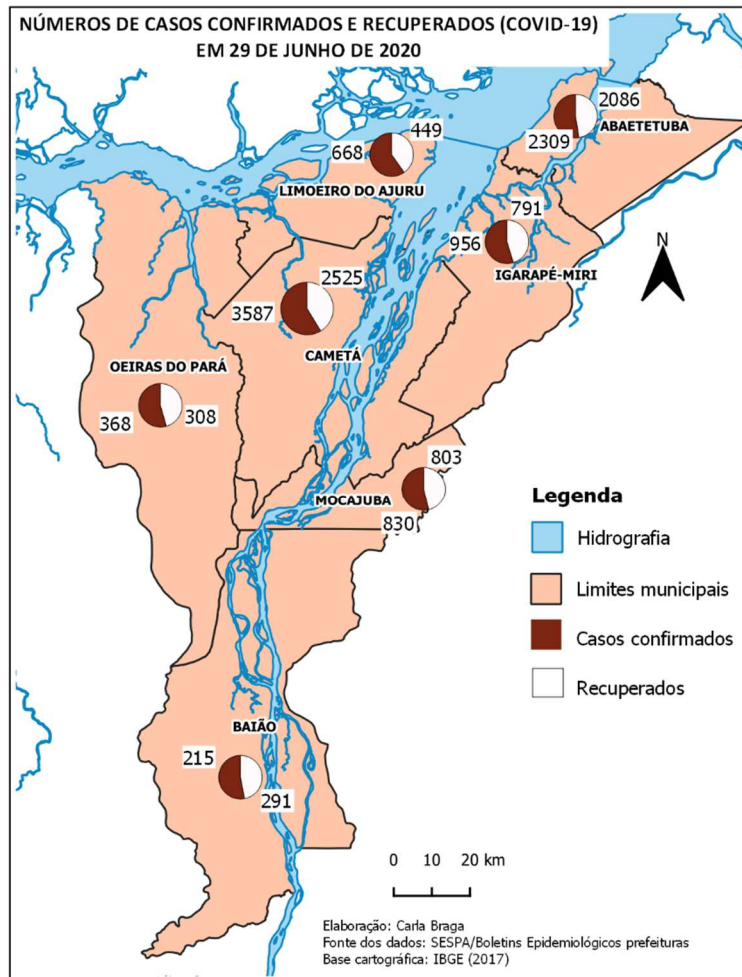
A geoinformação oriunda de números contabilizados pelos diversos efeitos da Covid-19 traz uma gama de representações por meio de mapas, gráficos, tabulações, tabelas etc. favorecendo uma comunicação mais didática com o público. Tais aparatos pedagógicos favorecem a inclusão da análise qualitativa e quantitativa do espaço geográfico e a visualização da escala de atuação do fenômeno como atores essenciais para repassar dados oficiais.

Por vezes, somente quantificar os números de infectados e vítimas da doença não se caracteriza de um “linguajar” necessário para contrapor as variadas perguntas necessárias para as tomadas de decisão (Cardoso, *et al.* 2020). Deste modo a utilização de métodos visuais, a exemplo de mapas, trazem uma resposta mais lúdica para a exposição dos dados disponíveis.

Visando atender o que foi discorrido anteriormente a Figura 2 expõe a situação dos casos confirmados e recuperados da Covid-19 nos municípios da região do Baixo Tocantins. Abaetetuba apresenta o maior número de

recuperados com 90% (2086) em relação à ordem de 2.086 casos confirmados. Por outro lado o menor valor dispõe em Limoeiro do Ajuru correspondendo a 67% de recuperados traduzidos em 449 casos, em detrimento dos 668 confirmados

Figura 2 - Municípios do Baixo Tocantins apresentando os números de casos confirmados e recuperados (Covid-19) em 29 de junho de 2020



O município de Cametá, uma das cidades polos da região, apresentava no dia 29 de junho de 2020 o maior número de pessoas infectadas, na ordem de 3587 casos. O panorama atual no estado indica que os novos pontos críticos da pandemia da Covid-19, que estavam concentrados na capital Belém e Região Metropolitana estão surgindo no interior do estado (ALEPA, 2020). Nesse contexto, Cametá e Abaetetuba posicionam-se entre os 10 municípios com maiores casos no Pará (PARÁ, 2020).

Baião apresenta 291 de casos confirmados e 215 de recuperados; Oeiras do Pará sinaliza 368 de infectados e 306 de curados, Mocajuba corresponde a 830 e 803, respectivamente de casos confirmados e recuperados, e por fim Igarapé- Miri concentra 956 de casos confirmados e 791 de recuperado.

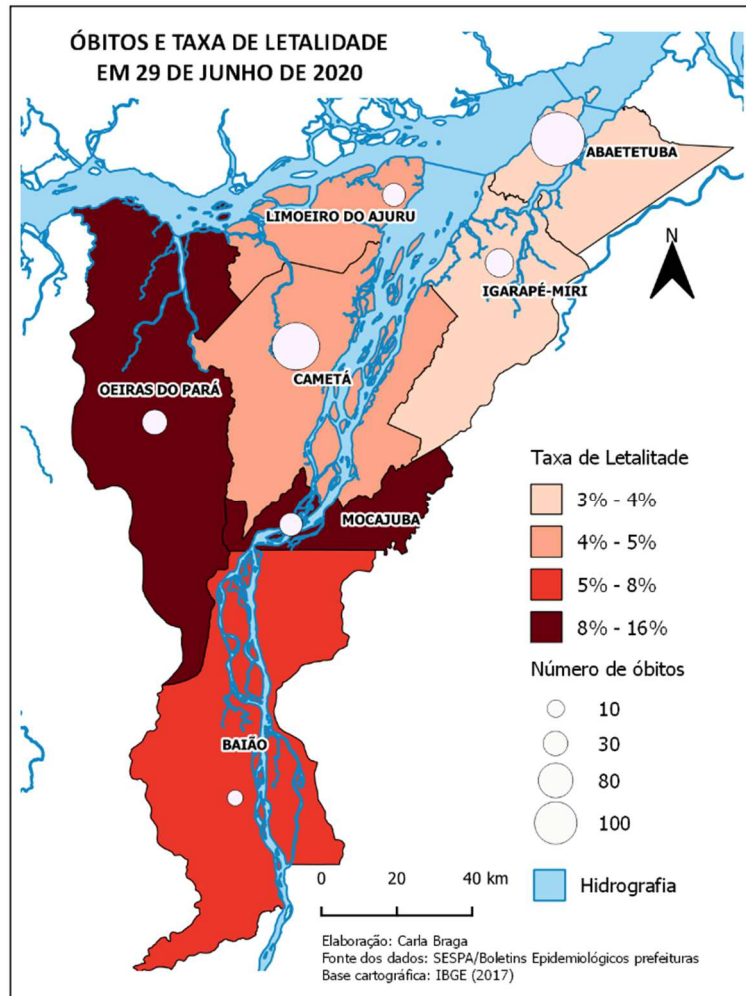
De um modo geral os dados mostram a média de pessoas recuperadas pela doença está em torno de 79%. Portanto, até 29 de junho de 2020, existiam 9.007 casos totais confirmados nos 7 municípios, sendo 2.309 casos (25%) no município de Abaetetuba.

Ao analisar as variáveis de óbitos e letalidades, a Figura 3 exibe algumas informações, a taxa de letalidade causada pela doença mostra que os municípios de Oeiras do Pará e Mocajuba apresentam maiores índices (entre 8 e 16%) seguido por Baião (entre 5 e 8%), Cametá e Limoeiro do Ajuru (4 e 5%), Igarapé-Miri e Abaetetuba, com taxa entre 3 e 4%. Conceitualmente, a taxa de letalidade equivale a proporção de óbitos por determinada doença em relação ao número total de doentes multiplicado por 100. Feng, et al. (2020) ressaltam que a letalidade pela Covid-19 é norteadas pelas características intrínsecas dos indivíduos infectados (idade, doenças prévias, hábitos de vida) quanto pela oferta/disponibilidade de recursos terapêuticos (leitos hospitalares, equipes de saúde, ventiladores mecânicos e medicamentos).

Deste modo, vários fatores devem ser analisados ao interpretar a taxa de letalidade, esta variável pode ser influenciada pela subnotificação da doença e de óbitos (Souza, 2020). Levar em consideração que estados com pouca testagem e nos quais os testes são realizados apenas em pacientes graves tendem a gerar taxas de letalidade maiores, desconsiderando o número real de doentes. Além de contabilizar o elevado número de óbitos que não tiveram sua causa confirmada, afetando a confiabilidade das notificações (COVID-19 Brasil, 2020).

No âmbito de números de óbitos, os municípios apresentam em ordem decrescente os seguintes dados: Abaetetuba apresenta 99 óbitos, Cametá perfaz de 76, Igarapé-Miri apresenta 28, Oeiras do Pará relata 21, Mocajuba a ordem de 18 e Baião corresponde a 9, este apresentado o menor valor entre os locais estudados.

Figura 3 - Mapa do número de óbitos e taxa de letalidade por Covid-19 em 29 de junho de 2020, nos municípios do Baixo Tocantins



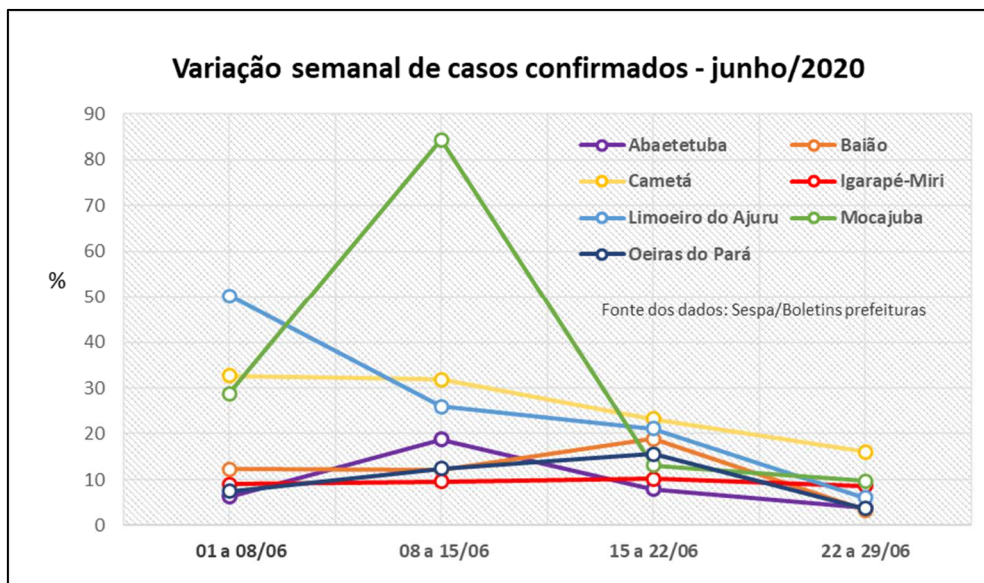
A Figura 4 sinaliza a correlação de dados de casos confirmados do dia 01 a 29 de junho de 2020, foi possível analisar a porcentagem de variação de 4 semanas, diante disso foi possível determinar a ascendência e o declínio de casos confirmados em todos os municípios do Baixo Tocantins. Tal variável reflete pelos fatores de aglomeração, distanciamento, quarentena e cuidados emergenciais em relação a saúde. Em função da referida variação semanal do mês de junho, percebe-se que a segunda semana do mês representou um pico ascendente de casos no município de Mocajuba, com uma variação de mais 50%, aumentando os casos. Por outro lado, comportamento contrastando com os dados referentes ao município de Limoeiro do Ajuru com uma queda de 20% nos casos confirmados; Mocajuba e Igarapé-Miri tiveram uma estabilidade dos números de casos entre a variação da primeira e segunda semana de junho.

A partir da variação semanal dos dias 08-15 de junho de 2020 comparado com o período de 15-22 de junho de 2020, municípios de Igarapé-

Miri, Oeiras do Pará e Baião apresentaram um singelo aumento dos casos confirmados. No entanto, Cametá, Abaetetuba e Limoeiro do Ajuru apresentaram decaimento dos valores. Ao todo, destaca-se Mocajuba que no início do mês teve uma variação ascendente de aproximadamente 60%, na semana posterior exibiu uma decadência de 70%, assinalando um caso anômalo entre os dados analisados em cada município da região do Baixo Tocantins.

No geral percebe-se uma queda dos casos registrados a partir da 2ª metade do mês de junho de 2020, fomentado pelo baixo valor de transmissão do novo corona vírus.

Figura 4 - Gráfico da variação semanal no mês de junho de 2020



3. ANÁLISE DA INFRAESTRUTURA HOSPITALAR

O Sistema Único de Saúde (SUS) possui o o número de leitos de UTI por 100 mil habitantes muito abaixo do inferior para atender o básico necessário à população, em 72% das regiões de saúde do país, sem considerar as necessidades colocadas pelo Covid-19. Frente á baixa quantidade de leitos adultos de UTI, cabe analisar a capacidade das regiões de saúde de comportar o fluxo de hospitalizações adicionais em decorrência de casos críticos em decorrência da transmissão do novo coronavírus (RACHER, *et al.* 2020).

A nota técnica ENSP/Fiocruz destaca a essencialidade da disposição do sistema de saúde em abarcar leitos e equipamentos suficientes, e desta maneira analisar à estrutura hospitalar de diversos municípios do país, imprescindíveis para amenizar problemas na capacidade de atenção e cuidados a pacientes e reduzir índices de letalidade da doença (Said e Silva, 2020).

As condições de infraestrutura dos hospitais do Baixo Tocantins são precárias e alarmantes. A disponibilidade dos números de habitantes e médicos nos municípios da região está disposto na figura 5. Pereira (2020) ao apresentar dados analíticos da infraestrutura hospitalar dos municípios do Baixo Tocantins, no que diz respeito a leitos hospitalares e médicos por 1.000 habitantes (Figura 6), apresentou dados preocupantes, uma vez que todos estão abaixo do nível nacional e estadual. Scheffer, *et al.* (2018) indicaram uma taxa de 2,5 médicos/1000 habitantes no Brasil, nesse viés, o Pará tem a razão de 0,97.

Nesse contexto, Limoeiro do Ajuru que compreende a ordem de 0,03 e Oeiras do Pará (0,09). O município de Abaetetuba indica a taxa de 0,60, maior valor em relação aos demais, em segundo destaca-se Cameté por manifestar a taxa de 0,32 médico por mil habitantes.

Figura 5 - Gráfico exibindo os números de habitantes e de leitos disponíveis em cada município do Baixo Tocantins

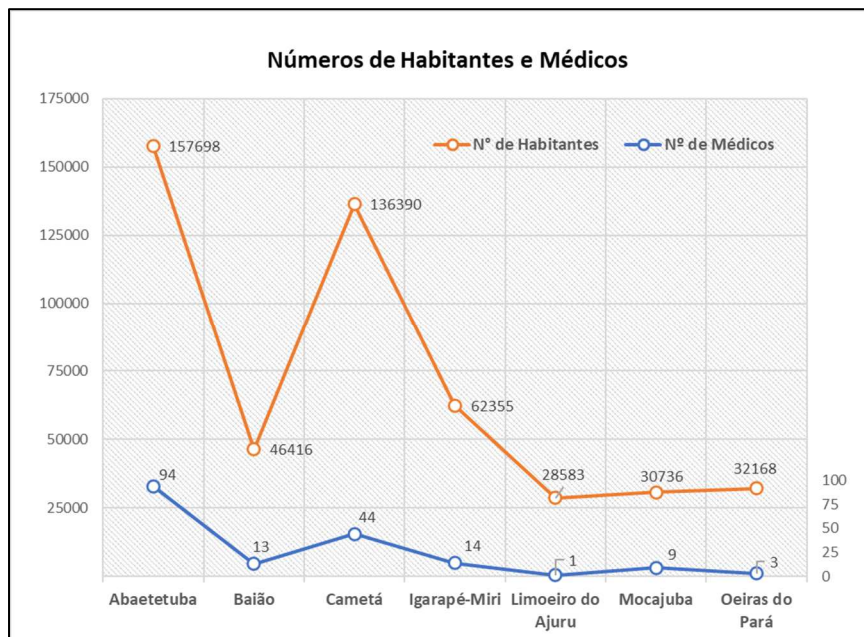
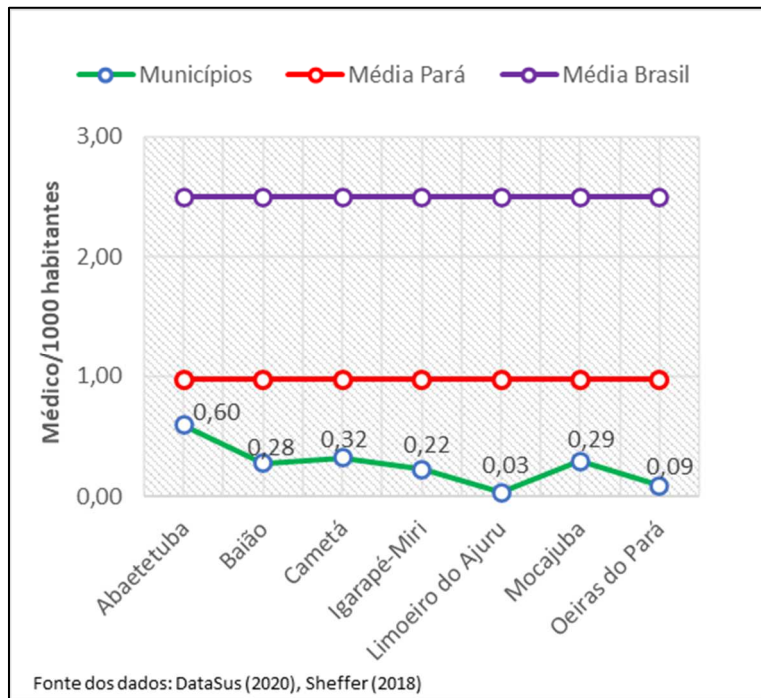


Figura 6 - Razão entre o número de médico por mil habitantes dos municípios do Baixo Tocantins em comparação com a média do estado do Pará e Brasil



Os números de leitos e habitantes por municípios do Baixo Tocantins podem ser observados na Figura 7, dessa forma é possível observar a proporção entre tais variáveis.

Em referência aos leitos por 1.000 habitantes, os municípios de Abaetetuba e Igarapé-Miri revelam taxas acima da média nacional (2,00), ainda assim estão abaixo do ideal do definido pelo Ministério da Saúde (MS), de 3 leitos/1000 habitantes (FIOCRUZ, 2019). Todos os municípios da região estão bem abaixo desse patamar, em destaque: Baião (0,6), Limoeiro do Ajuru (0,73) e Oeiras do Pará (0,5) por apresentarem marcas muito inferiores ao recomendado pelo órgão público (Figuras 7 e 8).

A região do Baixo Tocantins, assim como as análises apresentadas por Machado (2020) e; Dagnino e Freitas (2020) da região Amazônica, precisa do fortalecimento de ações visando ampliar o atendimento em saúde à população, como o caso de implantação com urgência de um hospital de campanha. Vimos que a fragilidade da infraestrutura dos municípios que compõem do Baixo Tocantins mostra que o Sistema de Saúde já não é suficiente para os casos que não estão relacionados a pandemia, situação esta que será agravada profundamente pela interiorização do vírus no estado do Pará.

Figura 7 - Gráfico do número de habitantes e número de leitos disponível em cada município

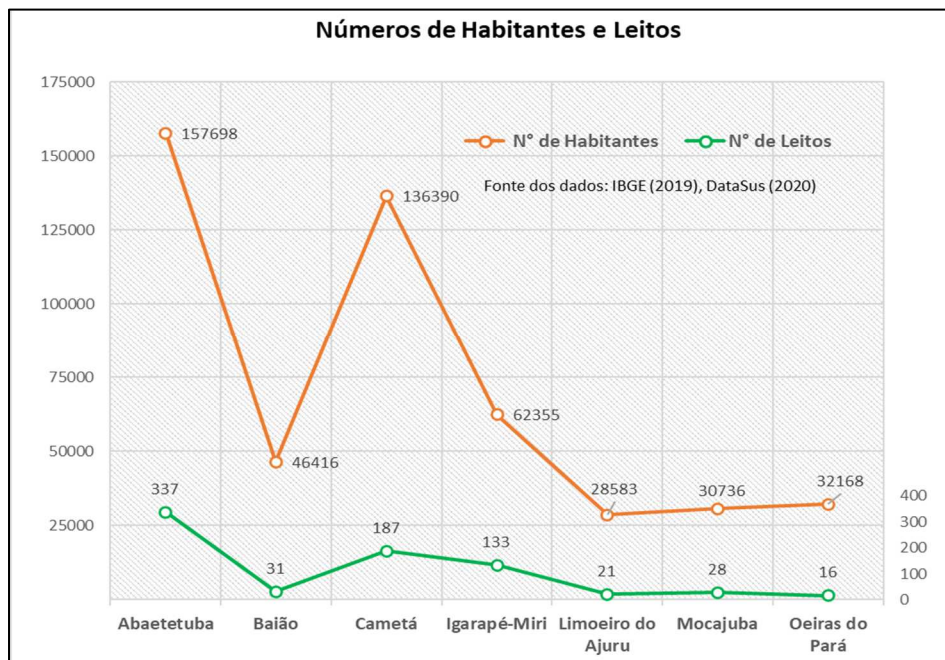
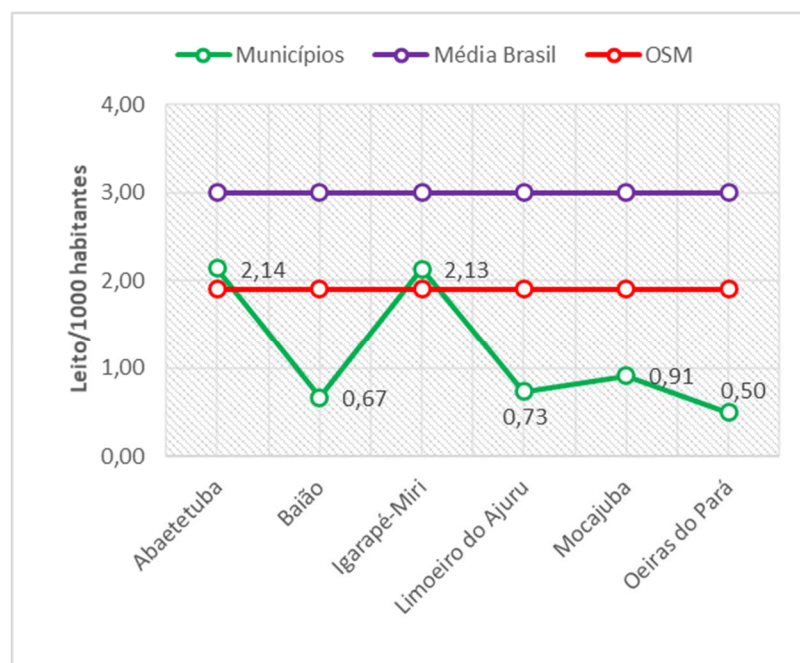


Figura 8 - Gráfico da relação entre o número de leito por mil habitantes de cada município do Baixo Tocantins, comparados a média brasileira e da recomendação pela OMS



Diante da explanação sobre a infraestrutura hospitalar do Baixo Tocantins precisa-se do fortalecimento de ações visando ampliar o atendimento em saúde à população, além do período atípico.

4. CONCLUSÃO

O presente cenário esboçado a partir dos dados da Covid-19 e as condições hospitalares na região do Baixo Tocantins mostra a recorte recorrente entre a maioria dos municípios brasileiros, relacionada ao alcance da contaminação do novo coronavírus assim como a precariedade no sistema de saúde.

A espacialização geográfica arquitetada à escala regional e municipal mostrou que as formas mais severas de contaminação em função da quantidade de óbitos e conseqüentemente a taxa de letalidade evidencia o lado mais cruel da doença que vem atingido os municípios mais frágeis para enfrentar essa pandemia.

No baixo Tocantins encontra-se os dois municípios entre os dez com maiores casos no estado do Pará. Algo alarmante ao ser comparado dentre os 144 municípios que compõem a conjuntura estadual. Em face do exposto, o presente artigo discorreu sobre dados pertinentes e mostrou a contextualização geográfica, revelando a dinâmica da Covid-19 a partir do panorama local, e assim entender o comportamento da novo coronavírus em uma das regiões interioranas do Pará.

5.REFERÊNCIAS BIBLIOGRÁFICAS

ALEPA. Instalação de Hospital de Campanha na Região do Baixo Tocantins é defendida por Dilvanda Faro. 20/05/2020. Disponível em > <https://www.alepa.pa.gov.br/noticiadep/4011/115> <. Acesso: 23 maio 2020.

Albuquerque, N.; Pedrosa, N. Análise espacial dos casos de COVID-19 e leitos de terapia intensiva no estado do Ceará, Brasil. SciELO Preprints, 2020. DOI:<https://doi.org/10.1590/SciELOPreprints.181>. Disponível em: <https://preprints.scielo.org/index.php/scielo/preprint/view/181>. Acesso em maio 2020.

Belforte, L. C. M.; Reis, R. S. P.; Da Silva, G. P.; Cavalcante, M. M. A. C. Leitura geográfica no contexto da Covid-19 em Rondônia ao norte do Brasil. Rev. Tamoios, São Gonçalo (RJ), ano 16, n. 1, Especial COVID-19. pág. 145-153, maio 2020.

Cardoso, P. V.; Seabra, V. S.; Bastos, I. B.; Costa, E. C. P. A importância da análise espacial para tomada de decisão: um olhar sobre a pandemia de Covid-19. Rev. Tamoios, São Gonçalo (RJ), ano 16, n. 1, Especial COVID-19. pág. 125-137, maio 2020.

COVID-19 Brasil . Monitoramento e análises da situação do Coronavírus no Brasil. Ribeirão Preto. Disponível em: <https://ciis.fmrp.usp.br/>. 2020.

Croda, J.; Garcia, L. Resposta imediata da Vigilância em Saúde à epidemia da COVID-19. Epidemiologia e Serviços de Saúde, 29(1), e2020002, 2020. <https://doi.org/10.5123/s1679-4974202000010002>.

DATASUS. Disponível em > <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?cnes/cnv/leiintbr.def>. < Acesso em 22 maio 2020.

Dagnino, R. S.; Freitas, M. W. D. Casos de Covid-19 nos municípios do estado do Amazonas, Brasil. Relatório técnico atualizado em 04/04/2020 com dados até 2/04/2020. Relatório Covid-19 – Amazonas. Atualizado: 4 abr. 2020

FIOCRUZ. ICITC. Levantamento do Proadess mostra a redução do número de leitos em hospitais no Brasil.22/04/2019. Disponível em: <https://www.icict.fiocruz.br/content/levantamento-do-proadess-mostra-redu%C3%A7%C3%A3o-do-n%C3%BAmero-de-leitos-em-hospitais-no-brasil> Acesso em 30 junho 2020

Feng, Y., Ling, Y., Bai, T., Xie, Y., Huang, J., Li, J., ... & Lu, Y. (2020). COVID-19 with different severities: a multicenter study of clinical features. *American journal of respiratory and critical care medicine*, 201(11), 1380-1388.

G1 Globo: Bem estar. Casos de coronavírus no Brasil em 18 de março de 2020. 18/03/2020. Disponível em: > <https://g1.globo.com/bemestar/coronavirus/noticia/2020/03/18/casos-de-coronavirusno-brasil-em-18-de-marco.ghtml><. Acesso: 15 maio 2020.

PARÁ. Secretaria de Estado de Saúde. Coronavírus no Pará. Atualização: 23 maio 2020. Disponível em: ><https://www.Covid-19.pa.gov.br/#/><. Acesso:23 de maio 2020(b).

Pereira C. B. Propagação da COVID-19 na região do Baixo Tocantins – Estado do Pará. Boletim de Divulgação do Observatório Geográfico da Covid-19. Nº. 04, 24/05/2020. Universidade Federal do Pará – UFPA.

Machado, R.L.M. (2020). Cenários da Pandemia na Amazônia: análise para além de números. Boletim COVID-19. Nº.06, 01/-6/2020. Ananindeua-PA ISSN: aguardando.

Munster, V. J., Koopmans, M., Van Doremalen, N., Van Riel, D., De Wit, E. A novel coronavirus emerging in China—key questions for impact assessment. *New England Journal of Medicine*, 382(8), 692-694. 2020.

PARÁ. Secretaria de Estado de Saúde. Confirmado o primeiro caso de Covid-19 no Estado do Pará. 18/03/2020. Disponível em: <http://www.saude.pa.gov.br/2020/03/18/confirmado-o-primeiro-caso-de-Covid-19-no-para/> Acesso: 15 maio 2020(a).

Racher, B., Rocha, R., Nunes, L., Spinola, P., Malik, A. M., Massuda, A. Necessidades de Infraestrutura do SUS em Preparo ao COVID19: Leitos de UTI, Respiradores e Ocupação Hospitalar. IEPS – Instituto de Estudos para Políticas de Saúde. Nota Técnica n. 3. 2020.

Said, A. M.; Silva, J. C. A espacialidade da COVID-19 e o pandemônio em meio a pandemia nas comunidades tradicionais da Cajuíba – Paraty. *Rev. Tamoios, São Gonçalo (RJ)*, ano 16, n. 1, Especial COVID-19. pág. 206-231, maio 2020

Sheffer, M., Cassenote, A., Guilloux, A. G. A., Biancarelli, A., Miotto, B. A., Mainardi, G. M. *Demografia Médica no Brasil 2018*. São Paulo, SP : FMUSP, CFM, Cremesp. 286p. ISBN 978-85-8707-7-55-4. 2018.

Souza, C. D. F., Paiva, J. P. S., Leal, T. C., Silva, L. F., Lucas, L. G. Evolução espaço-temporal da letalidade por COVID-19 no Brasil. *J Bras Pneumol*. 2020;46(4):e20200208. <https://dx.doi.org/10.36416/1806-3756/e20200208>. 2020.

Sposito, M.; Guimarães, R. Por que a circulação de pessoas tem peso na difusão da pandemia: Difusão da Covid-19 no país segue modelo relacionado a interações espaciais na rede urbana. UNAN - Unesp Agência de Notícias, Unesp, 26/03/2020. Disponível em <<https://unan.unesp.br/destaques/35626/por-que-a-circulacao-de-pessoas-tem-peso-na-difusao-da-pandemia>> Acesso em junho 2020.

Yang W. Z. Thoughts of the COVID-19 outbreak phases changed from emergency response to combination of emergency response and regular prevention and control. Chin J Epidemiol [Internet]. 2020 [acessado 2020 Apr 18]; 41. doi:10.3760/cma.j.cn112338-20200404-00516.

ISSN 2675-3243

volume 78

número 245

julho/dezembro 2020