

ANÁLISE DE DADOS DA COVID-19 NA PARAÍBA E REGIÕES

Roberto Mendes P. F. Mendonça – UNIESP – (robertompfm@gmail.com), Alana Marques de Moraes – UNIESP – (alanamm.prof@gmail.com), Aline Marques de Moraes – UNIESP – (alinemm.prof@gmail.com)

RESUMO

Um novo tipo de coronavírus, causador da doença COVID-19, vem se espalhando rapidamente pelo mundo, e a análise de dados tem sido fundamental para compreender melhor como esse vírus tem se espalhado nas diferentes regiões. Porém a análise de dados só pode gerar resultados representativos se a região estudada estiver monitorando adequadamente as pessoas contaminadas. Esse trabalho teve como objetivo estudar os dados da COVID-19 na Paraíba e suas regiões, avaliando o crescimento no número de casos e de óbitos confirmados, a quantidade de testes realizados e a distribuição de testes rápidos. Verificou-se que o ritmo de crescimento de novos casos e novos óbitos vem diminuindo, e a região de João Pessoa aparece como epicentro da doença no estado. No entanto, as análises mostraram que os dados disponíveis não representam de forma precisa a real condição da COVID-19 no estado devido a insuficiência de testes. E por fim, foi investigado se havia alguma correlação entre a distribuição de testes rápidos e a taxa de letalidade nas regiões da Paraíba, porém não foram encontrados coeficientes de correlação significativos.

Palavras-chave: análise de dados. covid-19. coronavírus. paraíba.

ABSTRACT

A new kind of coronavirus, that causes the COVID-19 disease, has been spreading all around the World, and Data Analysis has been very important to better comprehend the spread of this virus in different regions. However, Data Analysis can only generate representative results if the studied region has been properly monitoring the number of people infected. This work aimed to study the data of the COVID-19 in the state of Paraíba, evaluating the growth in the number of cases and deaths, the number of tests done, and the distribution of quick tests. It was found that the growth rate of new cases and new deaths has been decreasing, and that the Joao Pessoa region is the epicenter of the disease in the state. However, the analysis have shown that the available data do not represent the real situation of COVID-19 in the state due to insufficient testing. And, it was investigated if there was some correlation between the distribution of quick tests and the case fatality rate in Paraíba's regions, but the correlation coefficients found were not significant.

Keywords: data analysis. covid-19. coronavirus. paraiba.

1 INTRODUÇÃO

A COVID-19 é uma doença causada por um novo tipo de coronavírus e que vem se espalhando rapidamente pelo mundo, colocando em risco os sistemas de saúde e desafiando a capacidade de reação dos países atingidos. Existe um esforço global em compreender melhor essa doença e tentar frear o avanço desse novo vírus. Para isso, a ciência de dados vêm sendo utilizada para monitorar o avanço da doença, compreender suas características epidemiológicas e auxiliar na tomada de decisões. Porém, para interpretar dados que representem de forma

precisa o número de infectados de uma região, é necessário que sejam realizados testes suficientes, o que nem sempre é possível.

O presente artigo tem como objetivo buscar respostas por meio da análise de dados para três questionamentos distintos sobre a COVID-19 na Paraíba:

- O que os dados públicos dizem sobre o avanço do novo coronavírus no estado da Paraíba e em suas regiões?
- Os dados públicos disponíveis representam de forma precisa o número de contaminados e de falecimentos causados pelo novo coronavírus na Paraíba e em suas regiões?
- Existe correlação entre o número de testes rápidos distribuídos pelo Governo do Estado e a letalidade nas diferentes regiões Paraibanas?

Para isso, foram analisados o que os dados públicos dizem sobre o avanço no número de casos confirmados e mortes no estado e em suas regiões. Além disso, foram analisados dois parâmetros de confiabilidade dos dados: o número de testes realizados por casos confirmados e a taxa de letalidade dos casos. E, por fim, foi verificado se existe algum tipo de correlação entre o número de testes rápidos distribuídos pelo governo do estado aos municípios e as diferentes taxas de letalidade dos casos.

Em relação ao avanço da doença, os resultados mostraram que o número de casos confirmados permanece crescendo no estado, porém o ritmo de crescimento vem diminuindo. Outro dado relevante é que o número de novos óbitos registrados, que também vinha crescendo, teve uma redução pela primeira vez na última semana epidemiológica registrada (semana 25); a próxima seção irá discutir em detalhes essas semanas. Os dados mostraram que o epicentro de estado, ou seja, a região mais afetada pelo novo coronavírus, é a região de João Pessoa. Os parâmetros utilizados para avaliar a confiabilidade dos dados sugerem que os dados disponíveis não refletem de forma precisa a real situação epidemiológica do estado e suas regiões. Por fim, não foi encontrada correlação significativa entre a quantidade de testes rápidos distribuídos aos municípios e a taxa de letalidade de casos nas diferentes regiões.

As próximas seções apresentam a fundamentação teórica na qual este estudo se baseia e que são fundamentais ao entendimento do artigo; a descrição da metodologia adotada; o desenvolvimento das análises e seus resultados; as conclusões obtidas; e ao final, uma lista das referências bibliográficas citadas.

2 FUNDAMENTAÇÃO TEÓRICA

A presente seção apresenta uma discussão do cenário da pandemia de COVID-19 no mundo e, especificamente, na Paraíba. Além disso, são discutidos também conceitos relevantes relacionados a análise de dados (Data Analysis) conduzida neste estudo. Os temas apresentados tem o intuito de proporcionar uma fundamentação teórica para o estudo realizado.

2.1 ORIGEM E ESPALHAMENTO DA COVID-19

Em dezembro de 2019, hospitais da cidade de Wuhan, capital da província de Hubei, na China, passaram a receber casos de pacientes com pneumonia de causa desconhecida. No dia 31 de dezembro de 2019, 27 casos dessa doença não identificada foram reportados a

Organização Mundial da Saúde (OMS) (SOHRABI, 2020). Dentre os sintomas mais notáveis, apresentados pelos pacientes, estavam: tosse seca, dificuldade de respirar, febre e infiltrações nos pulmões. No dia 7 de janeiro de 2020, o agente causador foi identificado como um novo tipo de coronavírus semelhante aos coronavírus encontrados em morcegos, esse vírus foi então nomeado como Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), ou Coronavírus da Síndrome Respiratória Aguda Grave 2 em português. A doença causada pelo SARS-CoV-2 foi então batizada de COVID-19, pela OMS (SOHRABI, 2020; SINGHAL, 2020).

O novo coronavírus se mostrou facilmente transmissível de pessoa para pessoa, podendo ser transmitido inclusive a partir de pessoas que não manifestaram sintomas da COVID-19. O que fez com que o vírus se espalhasse de forma exponencial e chegasse rapidamente a outros países (SINGHAL, 2020). A OMS, no dia 30 de janeiro de 2020, declarou o surto de COVID-19 como Emergência de Saúde Pública de Importância Internacional, e posteriormente, no dia 11 de março de 2020, o surto foi classificado como uma pandemia (CUCINOTTA, 2020; SINGHAL, 2020).

No Brasil, o primeiro caso de COVID-19 foi confirmado no dia 26 de fevereiro e se tratava de um homem de 61 anos que havia viajado para a Itália (BRASIL, 2020a). Já na Paraíba, o paciente zero foi confirmado no dia 18 de março e era um homem de 60 anos e que havia retornado da Europa, mas era residente da capital João Pessoa. Ainda no mesmo mês, no dia 31 de março, foi registrado o primeiro óbito no Estado, um homem de 36 anos, diabético, residente do município de Patos (PARAÍBA; 2020e, 2020h).

Atualmente o Brasil superou a marca de 1 milhão de infectados pelo novo coronavírus, tendo registrado no dia 21 de junho um total de 1.085.038 casos confirmados e 50.617 mortes causadas pela doença (BRASIL, 2020c). A Paraíba, na mesma data, tinha registrado ao todo 36.784 casos e 766 óbitos (PARAÍBA, 2020a).

2.2 A SEVERIDADE DA COVID-19

A COVID-19 se mostra moderada na maior parte dos infectados, que tendem a se recuperar por conta própria, sem a necessidade de assistência médica. No entanto, parte dos infectados desenvolve complicações sérias que requerem tratamento e internação hospitalar, podendo ser fatais, como choque séptico, edema pulmonar e pneumonia severa. Os casos mais severos, que requerem tratamento intensivo, costumam acontecer com maior frequência em pacientes mais velhos e que já possuem outras comorbidades antes da infecção (SOHRABI, 2020).

Entre as medidas de avaliação da severidade de uma doença está a taxa de letalidade, que é a razão entre o número de mortos e o número de casos da doença. A taxa de letalidade pode variar devido a diversos fatores, como faixa etária, tratamento médico, hábitos e doenças pré-existentes. Sendo assim a taxa de letalidade de uma doença não é única, ela varia em diferentes contextos (ROSER, 2020).

Muitas vezes a taxa de letalidade dos casos, que é a razão entre o número de mortes confirmadas e o número de casos confirmados, é utilizada para tentar estimar a taxa de letalidade real da doença. No entanto, a letalidade dos casos não é necessariamente um bom indicativo da letalidade real. Existem dois fatores que fazem com que a taxa de letalidade de casos se distancie da letalidade real: o número de pessoas que são infectadas, mas não foram

diagnosticadas; e o intervalo de tempo entre o diagnóstico e o óbito do paciente (ROSER, 2020; VERITY; 2020).

Como parte dos infectados não costuma desenvolver sintomas graves, e a disponibilidade de testes diagnósticos para um novo agente patológico costuma ser pequena, existe uma tendência a focar os testes nos casos mais graves. Fazendo com que casos mais leves fiquem sem diagnóstico. Isso é especialmente verdadeiro para a COVID-19, que se espalhou rapidamente pelo planeta, com a maior parte dos países despreparados, e que grande parte dos infectados não apresenta sintomas. Esse fenômeno faz com que a taxa de letalidade dos casos gere valores superestimados (ROSER, 2020; VERITY; 2020).

Já o intervalo entre o diagnóstico e o óbito dos infectados pode causar o efeito inverso, fazendo com que a taxa de letalidade dos casos gere valores subestimados. Esse fenômeno costuma ocorrer com mais intensidade no início dos surtos, enquanto o número de infectados cresce exponencialmente. Dessa forma, quando se utiliza o número de casos e obtidos registrados até o momento, não se está levando em consideração os óbitos das pessoas que se infectaram recentemente, que ainda irão ocorrer. Esse efeito deve ser considerado para os dados que temos sobre a COVID-19, uma vez que a doença possui uma alta transmissibilidade, e que costuma haver um período de 2 a 3 semanas entre o início dos sintomas e a conclusão do estado clínico (recuperação ou óbito) (ROSER, 2020; VERITY; 2020).

Por conta das variáveis mencionadas anteriormente, os dados coletados até então vem apresentando valores de taxa de letalidade dos casos bem variados nos diferentes países e regiões. Estimativas sugerem que a taxa de letalidade real seja próxima de 1% (DORIGATTI, 2020; KIM, 2020; RUSSELL, 2020). No dia 13 de junho, foi divulgado que a Paraíba estava com a menor taxa de letalidade dos casos entre os estados do nordeste, com 2,3% de letalidade, também menor que a média nacional de 5,05% (PARAÍBA, 2020g).

2.3 A QUALIDADE DO MONITORAMENTO DA COVID-19

A subnotificação de casos da COVID-19 pode gerar dados que mostrem resultados distantes da realidade, comprometendo assim a compreensão do cenário real e a tomada de decisões. Roser (2020) cita dois indicadores de qualidade dos dados, são eles: o número de testes por mil habitantes, e o número de testes por casos confirmados.

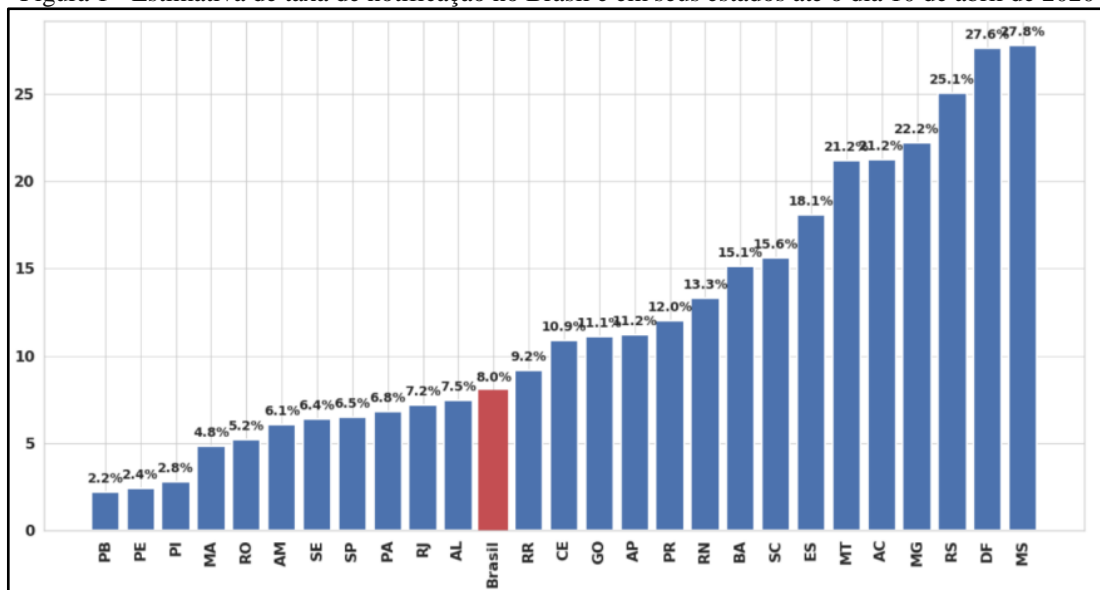
De uma forma geral, um número maior de testes por habitante permite que mais pessoas sejam testadas, aumentando assim o tamanho da amostragem. Além disso, uma maior capacidade de testagem pode implicar numa redução da necessidade de racionalizar os testes, reduzindo assim o viés de testar apenas casos mais severos (ROSER, 2020).

Porém, apenas dividir o número de testes pelo número de habitante não leva em consideração o estágio de expansão da doença na região. Por exemplo, se uma determinada região está no início do surto, não há necessidade de realizar muitos testes, pois ainda não existem muitas pessoas infectadas, o que faria com que o número de testes por habitante não fosse o melhor indicador. A medida que o número de casos aumenta, a necessidade de realizar mais testes também aumenta. Por isso, é útil observar também o número de testes por casos confirmados, que permite avaliar a qualidade dos dados, independente do grau de espalhamento da doença (ROSER, 2020).

Existe uma preocupação em relação a capacidade do Brasil em realizar testes para a COVID-19. Em um estudo analisando os dados disponíveis até o dia 10 de abril de 2020, Prado (2020) fez uma estimativa de taxa de notificação. No estudo foi encontrada uma taxa de

notificação de apenas 8,0% para o país. Para o estado da Paraíba, a taxa de notificação encontrada foi a menor entre os estados, apenas 2,2%, como mostrado no Figura 1.

Figura 1 - Estimativa de taxa de notificação no Brasil e em seus estados até o dia 10 de abril de 2020



Fonte: Prado (2020)

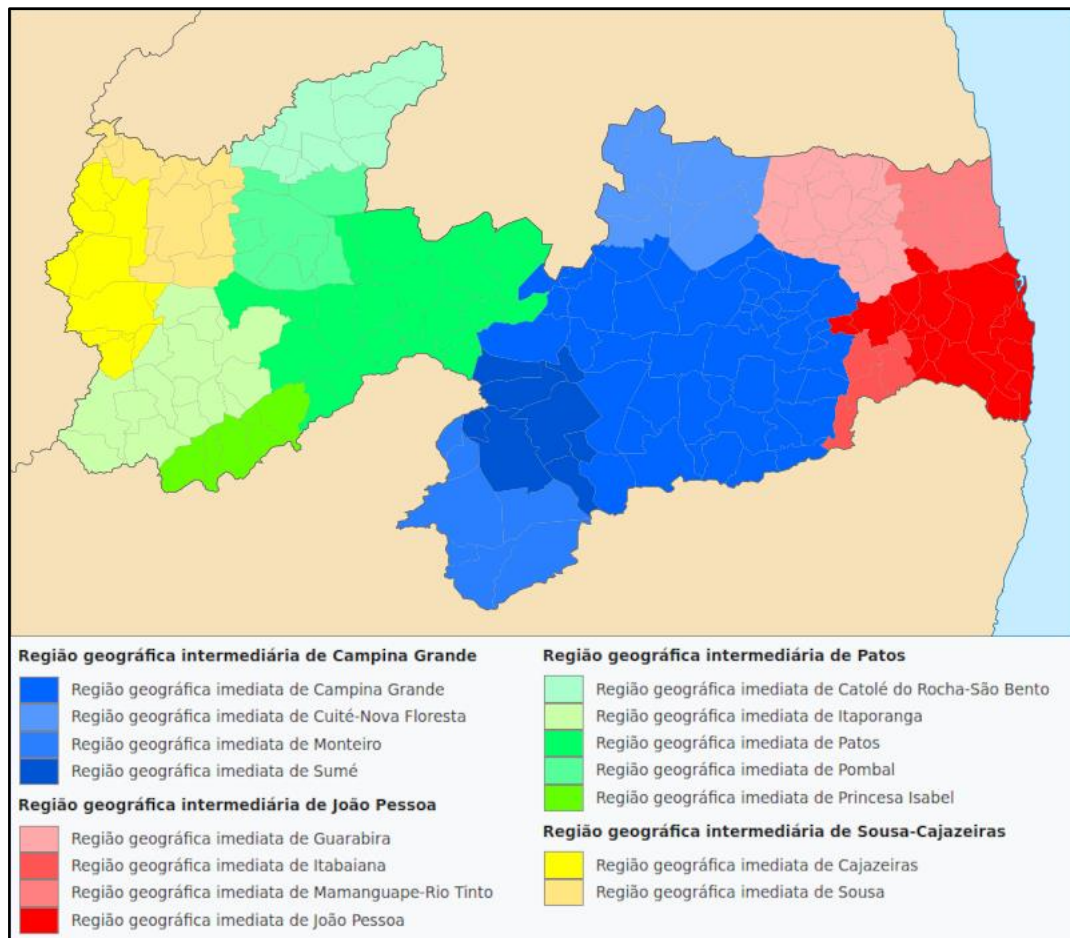
No dia 26 de maio, foi divulgado que a Paraíba estava em sexto lugar entre os estados que mais realizaram testes, tendo realizado 32.702 testes ao todo, juntando as duas modalidades de testes aplicadas: testes rápidos e método SWAB (PARAÍBA, 2020f).

Foram adquiridos pelo Estado 310 mil testes rápidos, além de 103 mil que foram doados pelo Ministério da Saúde. A distribuição dos testes rápidos tem sido feita de forma gradual, o primeiro lote de 20 mil unidades começou a ser distribuído no dia 24 de abril. Até o dia 26 de maio foram distribuídos 109.265 unidades (PARAÍBA, 2020c; PARAÍBA; 2020f).

2.4 DIVISÃO REGIONAL DA PARAÍBA

O IBGE subdivide os estados brasileiros em micro e mesorregiões, chamadas de Regiões Geográficas Imediatas e Regiões Geográficas Intermediárias, respectivamente (IBGE, 2017). As regiões geográficas intermediárias e imediatas do estado da Paraíba são exibidas na Figura 2.

Figura 2 - Regiões geográficas intermediárias e imediatas do estado da Paraíba



Fonte: Júnior (2018)

2.5 CALENDÁRIO EPIDEMIOLÓGICO

Existe um padrão internacional em epidemiologia de divisão dos anos em semanas, que são contadas de domingo a sábado. A primeira semana epidemiológica do ano é a primeira semana que possui a maioria dos dias em janeiro. Analogamente, a última semana epidemiológica do ano é a última semana que possui a maioria dos dias em dezembro (GOIÁS, 2020). A Tabela 1 mostra o calendário epidemiológico de 2020.

O presente estudo levou em consideração o período da semana 12 (primeiro caso registrado na Paraíba) até a semana 25 (última semana concluída antes do dia 21 de junho de 2020).

Tabela 1 - Calendário epidemiológico de 2020

Semana	Início	Término	Semana	Início	Término	Semana	Início	Término
1	29/12/2019	04/01/2020	19	03/05/2020	09/05/2020	37	06/09/2020	12/09/2020
2	05/01/2020	11/01/2020	20	10/05/2020	16/05/2020	38	13/09/2020	19/09/2020
3	12/01/2020	18/01/2020	21	17/05/2020	23/05/2020	39	20/09/2020	26/09/2020
4	19/01/2020	25/01/2020	22	24/05/2020	30/05/2020	40	27/09/2020	03/10/2020
5	26/01/2020	01/02/2020	23	31/05/2020	06/06/2020	41	04/10/2020	10/10/2020
6	02/02/2020	08/02/2020	24	07/06/2020	13/06/2020	42	11/10/2020	17/10/2020
7	09/02/2020	15/02/2020	25	14/06/2020	20/06/2020	43	18/10/2020	24/10/2020
8	16/02/2020	22/02/2020	26	21/06/2020	27/06/2020	44	25/10/2020	31/10/2020

9	23/02/2020	29/02/2020	27	28/06/2020	04/07/2020	45	01/11/2020	07/11/2020
10	01/03/2020	07/03/2020	28	05/07/2020	11/07/2020	46	08/11/2020	14/11/2020
11	08/03/2020	14/03/2020	29	12/07/2020	18/07/2020	47	15/11/2020	21/11/2020
12	15/03/2020	21/03/2020	30	19/07/2020	25/07/2020	48	22/11/2020	28/11/2020
13	22/03/2020	28/03/2020	31	26/07/2020	01/08/2020	49	29/11/2020	05/12/2020
14	29/03/2020	04/04/2020	32	02/08/2020	08/08/2020	50	06/12/2020	12/12/2020
15	05/04/2020	11/04/2020	33	09/08/2020	15/08/2020	51	13/12/2020	19/12/2020
16	12/04/2020	18/04/2020	34	16/08/2020	22/08/2020	52	20/12/2020	26/12/2020
17	19/04/2020	25/04/2020	35	23/08/2020	29/08/2020	53	27/12/2020	02/01/2021
18	26/04/2020	02/05/2020	36	30/08/2020	05/09/2020			

Fonte: Brasil (2020b)

2.6 FERRAMENTAS PARA ANÁLISE DE DADOS

Os dados são muito importantes para compreender a pandemia de COVID-19. Apenas utilizando dados completos e confiáveis é possível avaliar de forma adequada a disseminação e características da doença (ROSER, 2020). Dificilmente os dados vêm prontos para análise, na maioria das vezes eles precisam de ajustes até que estejam devidamente prontos para uso.

Mesmo quando se utiliza fontes que disponibilizam dados bem estruturados e em formato correto, ainda assim é necessário limpar e manipular esses dados para se realizar as análises. Dentre as ferramentas mais populares para manipulação, tratamento e análise de dados está o pandas, que é uma biblioteca de estrutura de dados e ferramentas estatísticas para a linguagem Python (MCKINNEY, 2010).

Outra parte importante da análise de dados é a visualização dos dados gerados. E dentre as ferramentas de visualização de dados, está o matplotlib, que também é uma biblioteca Python, e que foi desenvolvida para gerar representações visuais com qualidade de publicação (HUNTER, 2007).

3 METODOLOGIA

Nesta seção, será apresentada a metodologia utilizada na elaboração deste estudo, executada de acordo com as seguintes etapas: elaboração de questionamentos; revisão literária; definição de critérios de avaliação; coleta de dados; tratamento dos dados; análise exploratória; interpretação dos resultados. As etapas mencionadas foram reproduzidas de modo sequencial, como representado na Figura 3.

Figura 3 - Fluxograma da metodologia empregada



Fonte: Autor (2020)

3.1 ELABORAÇÃO DE QUESTIONAMENTOS

A primeira etapa da pesquisa é constituída pela elaboração de questionamentos concisos e mensuráveis que são a base para a definição dos objetivos específicos da análise de dados. Os questionamentos levantados nesta etapa no presente estudo foram:

- O que os dados públicos dizem sobre o avanço do novo coronavírus no estado da Paraíba e em suas regiões?
- Os dados públicos disponíveis representam de forma precisa o número de contaminados e de falecimentos causados pelo novo coronavírus na Paraíba e em suas regiões?
- Existe correlação entre o número de testes rápidos distribuídos pelo Governo do Estado e a letalidade nas diferentes regiões Paraibanas?

3.2 REVISÃO LITERÁRIA

A segunda etapa corresponde a uma pesquisa exploratória em artigos científicos, livros, relatórios e matérias jornalísticas sobre temas relacionados aos questionamentos propostos. A revisão literária tem o intuito de construir um arcabouço teórico que irá fundamentar a elaboração desse estudo.

O resultado da revisão literária deste estudo foi apresentado na seção anterior, fundamentação teórica. A maior parte do material pesquisado foi encontrada utilizando a ferramenta de busca do Google para conteúdo acadêmico (<https://scholar.google.com/>). Já as notícias utilizadas foram buscadas nas seções de notícias do Ministério da Saúde (<https://www.saude.gov.br/noticias>) e do Governo do Estado da Paraíba (https://paraiba.pb.gov.br/diretas/saude/coronavirus/noticias?b_start:int=0). A presente revisão selecionou materiais publicados de junho de 2007 até a data 21 de junho de 2020.

3.3 DEFINIÇÃO DE CRITÉRIOS DE AVALIAÇÃO

A terceira etapa elenca quais critérios são utilizados para responder os questionamentos mencionados na etapa inicial. Com os critérios definidos, é possível saber quais tipos de dados devem ser coletados para realização da análise.

Foi decidido que a expansão do novo coronavírus será avaliada levando em consideração a progressão do número de casos confirmados e do número de óbitos confirmados causados pela doença.

Para avaliar a representatividade dos dados disponíveis, foram definidos os seguintes parâmetros como critério de avaliação: o número de testes pelo número de casos confirmados e a taxa de letalidade dos casos. De modo que, quanto menor o número de testes por casos confirmados menos representativos são os dados disponíveis. Já em relação a taxa de letalidade, valores acima de 1,0% são um indicativo de subnotificação de casos.

Por fim, para avaliar a distribuição dos testes rápidos, foram adotados o número de testes distribuídos por 100.000 habitantes e a taxa de letalidade dos casos como parâmetros para avaliação. Considerando que quanto mais acima de 1,0% os valores de taxa de letalidade se apresentam, maior deve ser a subnotificação e conseqüentemente maior a carência de testes. A idéia é a verificação de correlação entre os dois parâmetros mencionados. Para isso, a análise levou em consideração os dados disponíveis no dia 21 de junho de 2020.

3.4 COLETA DE DADOS

Essa etapa representou a fase da busca e coleta de dados seguindo os critérios definidos na etapa anterior. Os dados utilizados têm como fonte a Secretária de Saúde do Estado da Paraíba e o Instituto Brasileiro de Geografia e Estatística (IBGE), e foram obtidos através de diferentes portais.

Nas análises que levaram em consideração o estado da Paraíba como um todo, foram utilizados os dados de boletim diário da Secretaria de Saúde do Estado, disponibilizados em plataforma do Governo do Estado da Paraíba (2020d). O boletim apresenta para cada dia, o número de casos acumulados, casos novos, descartados, óbitos acumulados, óbitos novos e a taxa de letalidade dos casos. Os dados podem ser acessados pelo endereço <https://superset.plataformatarget.com.br/superset/dashboard/microdados/>.

Para as análises que consideram os diferentes municípios, foi utilizado o banco de dados disponibilizado pelo portal Brasil.io (JUSTEN, 2020), que reúne os mesmos dados disponibilizados na plataforma do Governo do Estado da Paraíba, porém nele os dados estão divididos entre os municípios. Além disso, esse portal também possui os códigos e a população estimada de cada município segundo o IBGE. Esses dados podem ser acessados por meio do endereço https://brasil.io/dataset/covid19/caso_full/.

Também foi necessário utilizar os dados de divisão regional do Brasil do IBGE (2017) para possibilitar o agrupamento dos municípios em regiões imediatas ou regiões intermediárias (<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/divisao-regional/15778-divisoes-regionais-do-brasil.html?=&t=acesso-ao-produto>).

E por fim, o último conjunto de dados utilizado foi a relação de testes rápidos distribuídos e testes rápidos aplicados por município, disponibilizados pelo Governo do Estado da Paraíba (2020b). Os dados podem ser acessados pelo endereço: <https://superset.plataformatarget.com.br/superset/dashboard/55/>.

3.5 TRATAMENTO DOS DADOS

A etapa de tratamento dos dados executa o pré processamento dos dados para a etapa seguinte que é da análise exploratória em si. Dentre os processos realizados nessa etapa estão: renomeação de linhas e colunas, junção de conjuntos de dados diferentes, criação de novas colunas, remoção de missing values e transformação dos dados.

Foi utilizada a linguagem de programação Python junto da biblioteca para análise de dados pandas e da biblioteca para visualização de dados matplotlib.

3.6 ANÁLISE EXPLORATÓRIA

Com os dados pré processados, é possível realizar a etapa de análise exploratória. Nela os dados são visualizados de formas diferentes, utilizando tabelas, diferentes tipos de gráficos, combinando diferentes parâmetros, realizando agrupamentos, utilizando filtros. Essas operações têm o intuito de gerar informações visuais ou números que permitam compreender de forma mais simples o comportamento dos dados e assim buscar as respostas para as questões propostas.

A análise exploratória nesse estudo foi subdividida em duas etapas, primeiro foi realizada uma análise do estado da Paraíba como um todo, e em seguida foi feita uma análise considerando as diferentes regiões.

3.7 INTERPRETAÇÃO DOS RESULTADOS

A etapa de interpretação dos resultados consiste em uma análise crítica dos gráficos e valores obtidos na etapa anterior. Nessa etapa, procura-se descrever, encontrar causas, implicações e significados no comportamento dos dados. É necessário que as interpretações sejam feitas com bastante cautela e devidamente embasadas na literatura científica. É nessa etapa em que se verifica o que os dados permitem dizer sobre os questionamentos propostos.

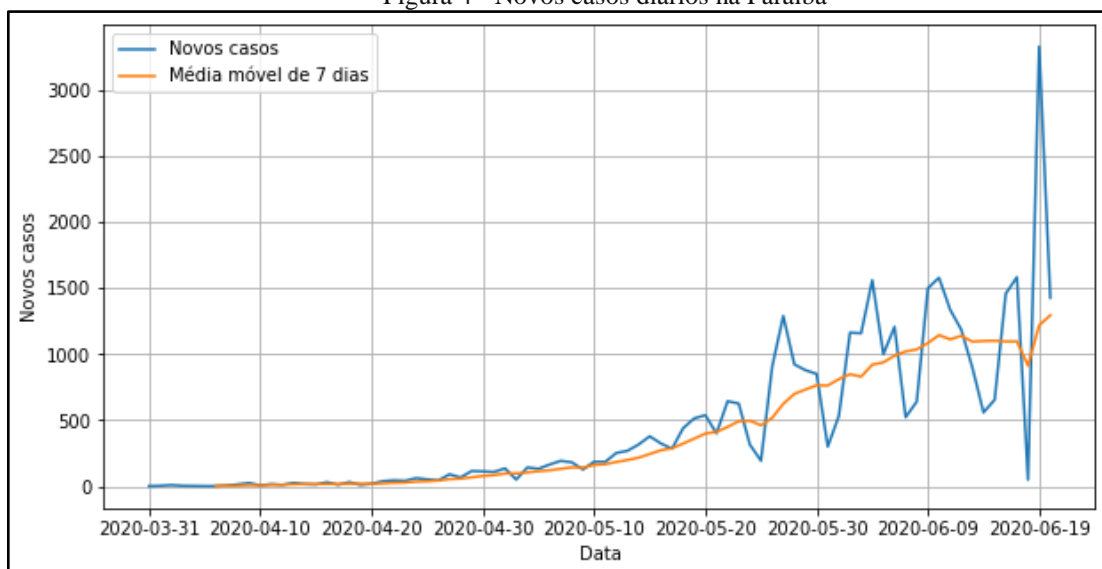
4 RESULTADO E DISCUSSÃO

Essa seção apresenta uma descrição da análise exploratória realizada, junto com uma interpretação dos resultados, contextualizada com os achados na literatura. A análise foi dividida em: análise da COVID-19 no estado da Paraíba; e análise da COVID-19 nas regiões do estado da Paraíba.

4.1 ANÁLISE DA COVID-19 NO ESTADO DA PARAÍBA

A análise foi iniciada verificando-se o crescimento da COVID-19 na Paraíba por meio do número de novos casos e do número de novos óbitos. A Figura 4 apresenta o número de novos casos confirmados por dia no estado da Paraíba (linha azul), foi utilizada uma média móvel de sete dias (linha laranja) para tentar suavizar as oscilações e tentar enxergar possíveis tendências

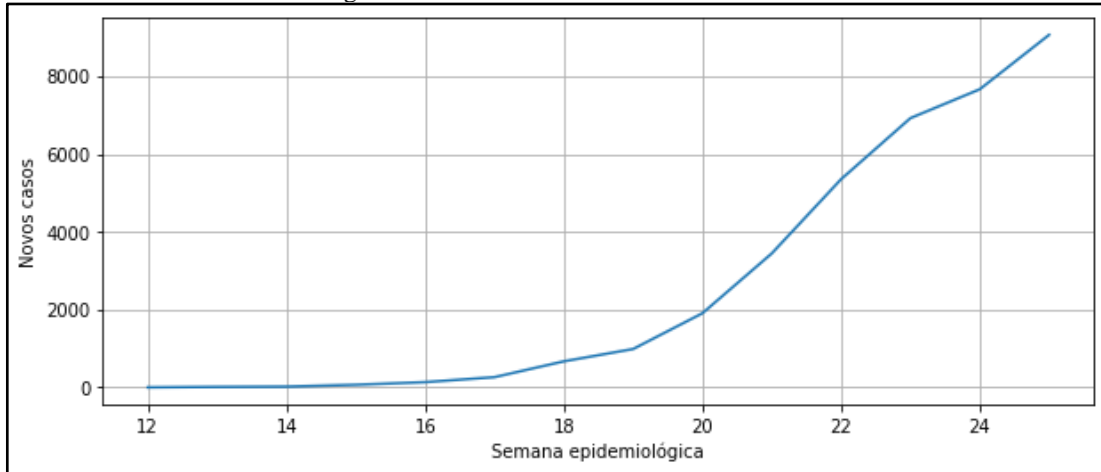
Figura 4 - Novos casos diários na Paraíba



Fonte: Autor (2020)

Chama atenção na figura o pico repentino, bem acima dos valores anteriores, no dia 19 de janeiro (penúltimo dia registrado). Mesmo com o auxílio da média móvel, é difícil observar uma tendência clara, por conta de constantes oscilações. É possível observar que as oscilações aparentam seguir ciclos semanais, provavelmente por conta de finais de semana e feriados. Optou-se então por fazer um agrupamento dos dados em semanas epidemiológicas, somando-se os novos casos diários. A Figura 5 a seguir mostra os novos casos semanais.

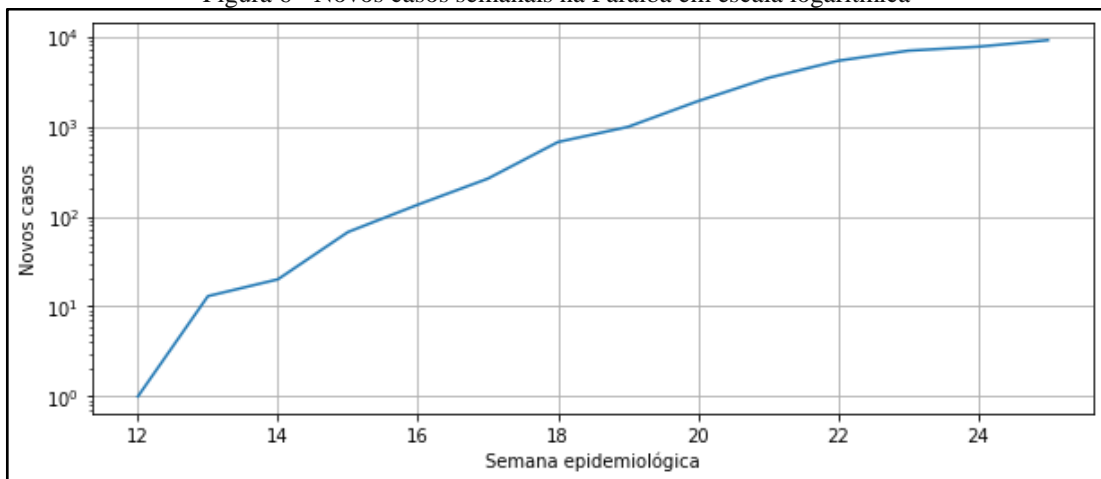
Figura 5 - Novos casos semanais na Paraíba



Fonte: Autor (2020)

O gráfico semanal (Figura 5) mostra uma curva mais suave, nela é possível verificar um crescimento inicial acelerado que vai aos poucos se tornando linear. Essa desaceleração pode ser verificada de forma mais clara em uma versão do mesmo gráfico com o eixo dos novos casos em escala logarítmica, mostrado a seguir (Figura 6).

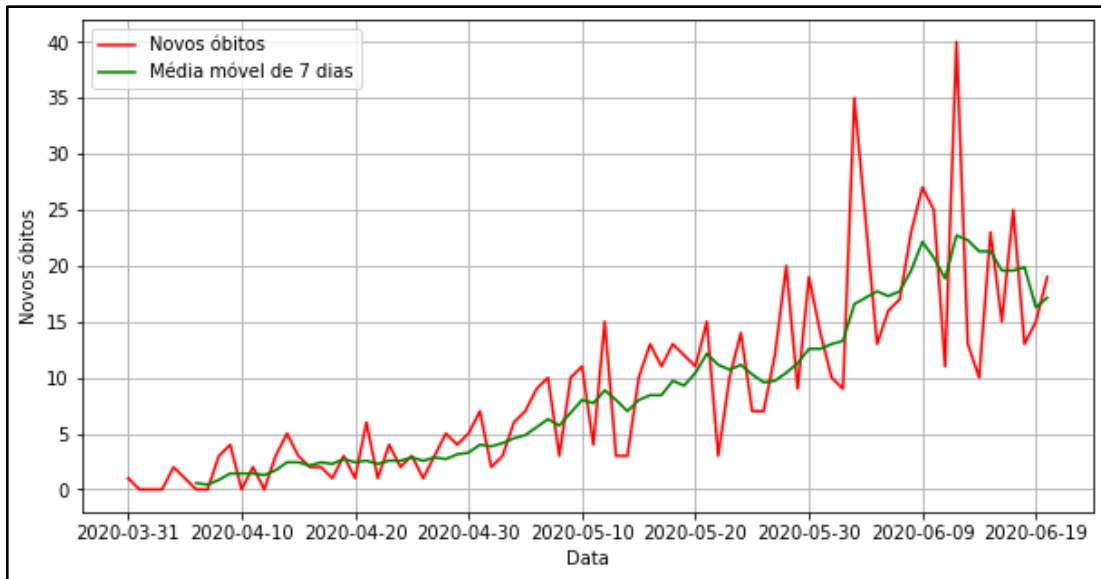
Figura 6 - Novos casos semanais na Paraíba em escala logarítmica



Fonte: Autor (2020)

A curva mostrada na Figura 6 em escala logarítmica se torna menos íngreme ao longo das semanas, indicando uma desaceleração no crescimento dos novos casos. Após concluída a análise dos novos casos no estado da Paraíba, repetiu-se o mesmo processo, dessa vez analisando o número de novos óbitos (Figura 7).

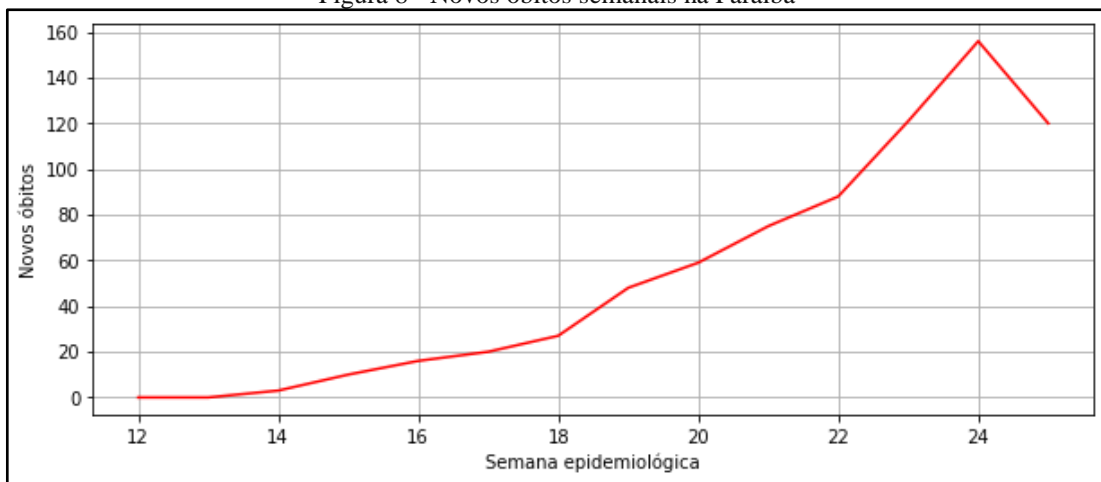
Figura 7 - Novos óbitos diários na Paraíba



Fonte: Autor (2020)

A Figura 7 mostra os novos óbitos diários (linha vermelha) e uma média móvel de sete dias dos novos óbitos (linha verde). Os dados de novos óbitos diários mostram um aumento no número de novos óbitos até aproximadamente o dia 12 de junho, quando o número de novos óbitos começa a cair. Como os dados diários apresentam oscilações, foi realizada também uma análise semanal, mostrada na figura seguinte (Figura 8).

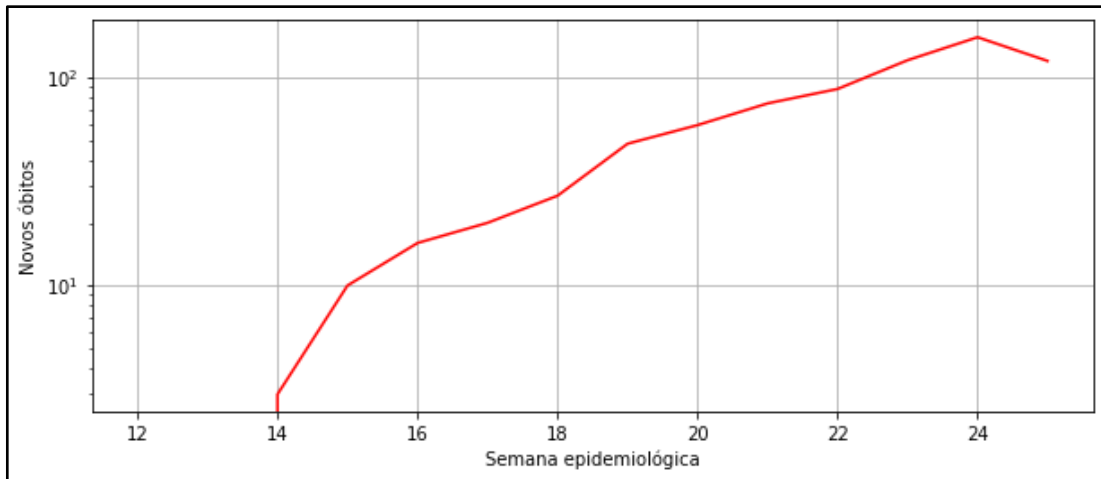
Figura 8 - Novos óbitos semanais na Paraíba



Fonte: Autor (2020)

Analisando o gráfico semanal (Figura 8), é possível observar um aumento no número de novos óbitos até a semana epidemiológica 24, e logo após, uma súbita redução no número de novos óbitos. Essa mudança repentina, também é observada no gráfico da Figura 9 em escala logarítmica, mostrando a seguir.

Figura 9 - Novos óbitos semanais na Paraíba em escala logarítmica

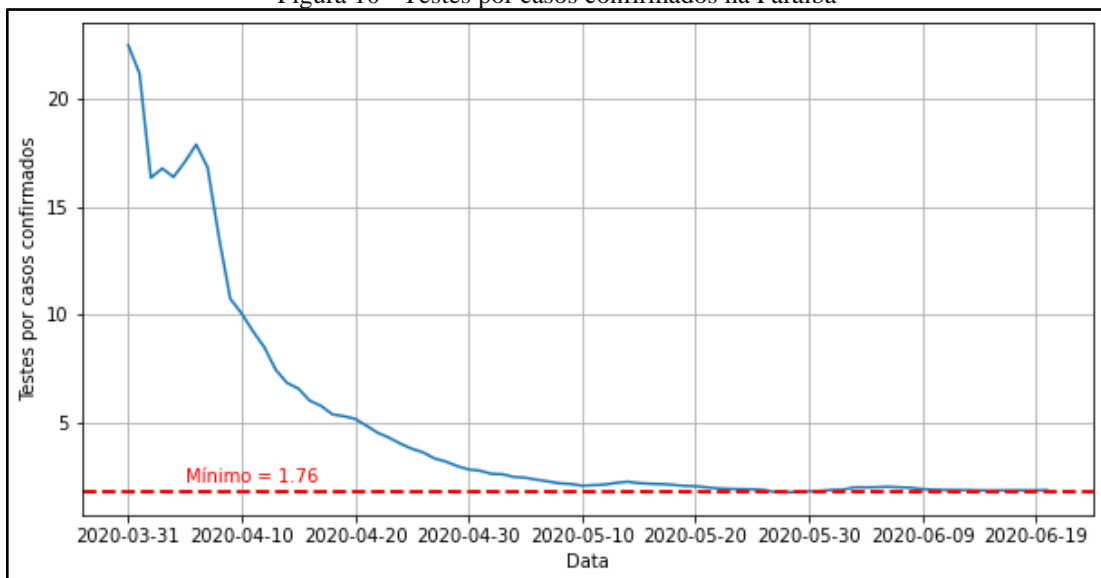


Fonte: Autor (2020)

O número de óbitos tende a representar melhor a disseminação da doença em uma região, uma vez que a testagem acaba por focar mais nos casos mais severos. É importante observar que o número de óbitos não é capaz de informar a situação atual de disseminação da doença, já que existe um intervalo de tempo entre a contaminação e o óbito. Apesar de ser um parâmetro mais confiável que o número de casos confirmados, uma carência no número de testes disponíveis também pode afetar a confiabilidade desses dados.

Para avaliar melhor se a testagem vem sendo feita de forma ampla, foi criado um gráfico mostrando a razão testes realizados por casos confirmados ao longo dos dias (Figura 10). O número de testes realizados foi obtido somando o número acumulado de casos confirmados com o número acumulado de casos descartados. O gráfico obtido é mostrado a seguir (Figura 10).

Figura 10 - Testes por casos confirmados na Paraíba



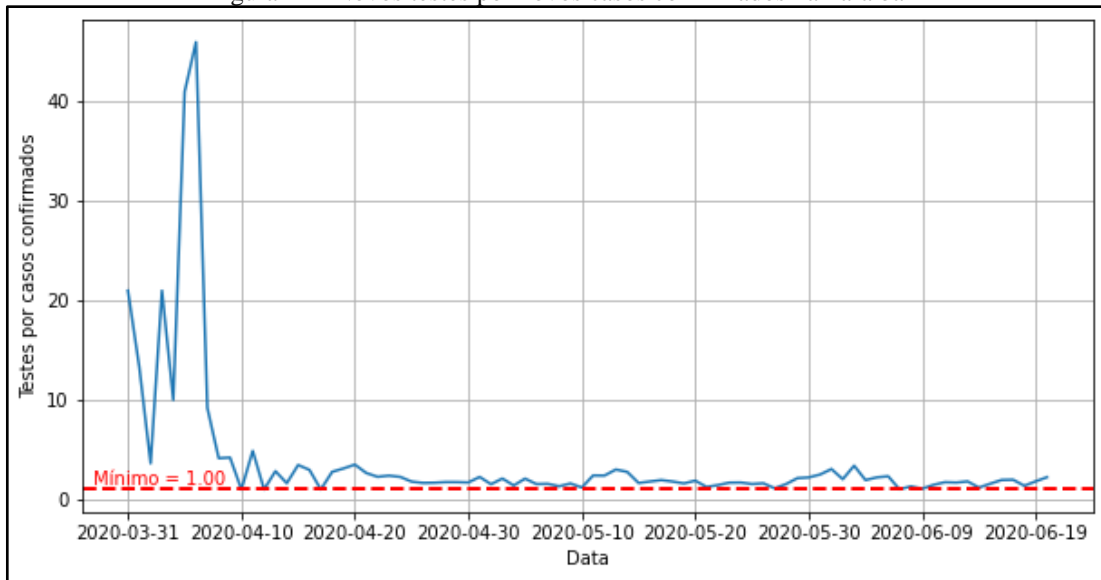
Fonte: Autor (2020)

A Figura 10 mostra que no início do surto do novo coronavírus na Paraíba o número de testes por casos confirmados (linha azul) era elevado, sugerindo que havia uma ampla disponibilidade de testes, o que é esperado no início de um surto. Porém rapidamente esse número cai, e se estabiliza próximo de um valor mínimo de 1,76 (linha vermelha tracejada).

Indicando que apenas uma pequena parte dos que foram testados não estavam contaminados, sendo assim, um indício de que estão sendo realizados poucos testes e que muitas pessoas que estão contaminadas não estão sendo testadas. A Figura 11 apresenta isso de forma mais clara, nela são mostrados os valores diários de novos testes por novos casos confirmados (linha azul) e seu valor mínimo (linha vermelha tracejada).

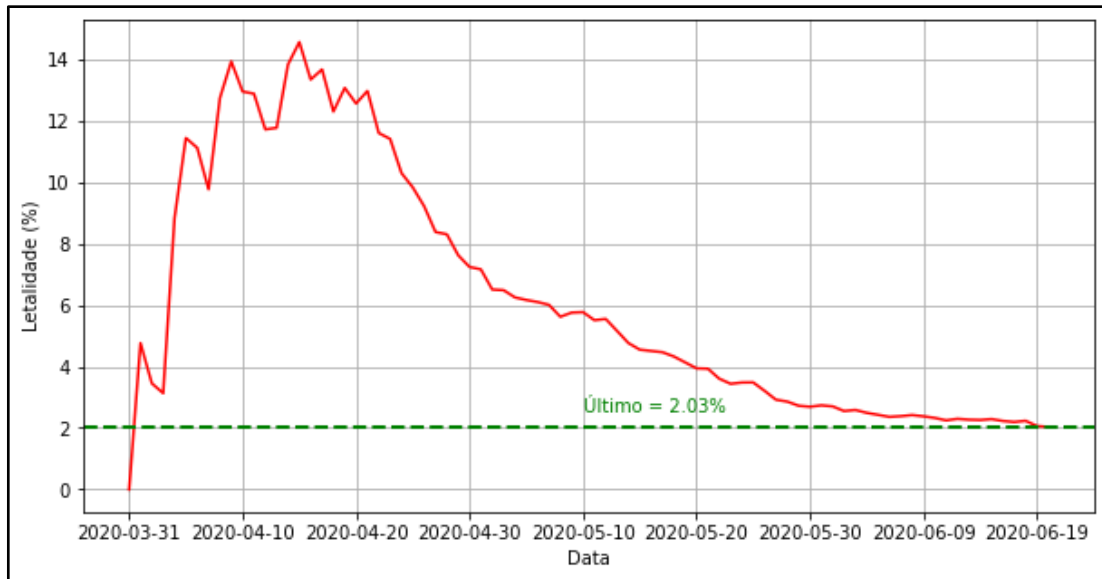
A Figura 11 demonstra que desde o dia 10 de abril, apenas uma pequena parte dos testados não estava contaminada. Outro parâmetro utilizado para avaliar possíveis subnotificações foi a taxa de letalidade. A Figura 12 exibe a taxa de letalidade dos casos de COVID-19 na Paraíba.

Figura 11 - Novos testes por novos casos confirmados na Paraíba



Fonte: Autor (2020)

Figura 12 - Taxa de letalidade dos casos na Paraíba

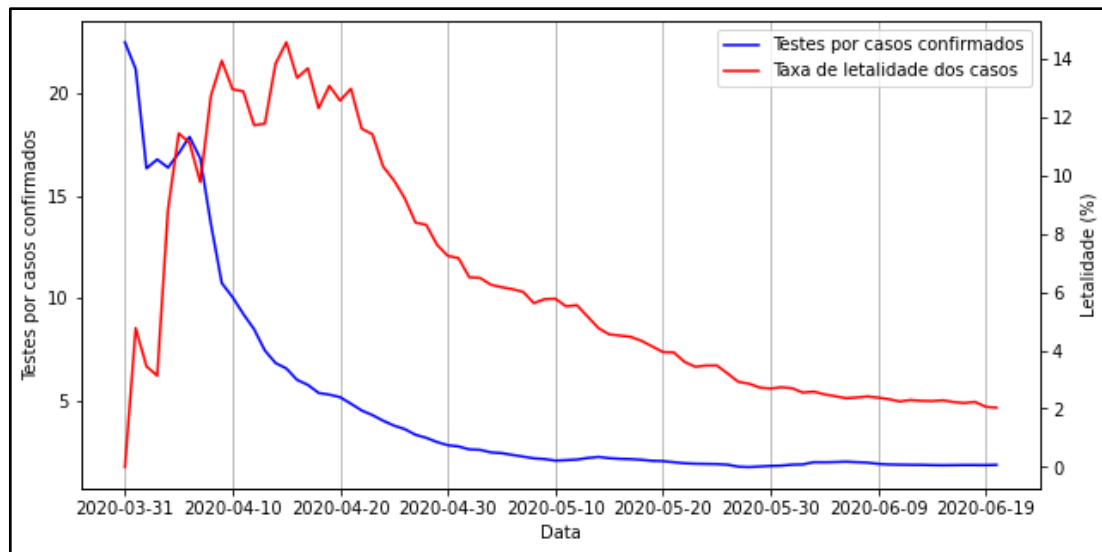


Fonte: Autor (2020)

A taxa de letalidade dos casos cresce rapidamente no início do surto e após um tempo começa a cair de forma gradativa e atualmente está atingindo um valor próximo de 2% (Figura 12). O que é um comportamento curioso, o esperado seria que a taxa de letalidade apresentasse um comportamento semelhante ao comportamento dos testes por casos confirmados (Figura 10).

A Figura 13 exhibe as duas curvas, sendo os testes por casos confirmados representados pela linha azul e a taxa de letalidade dos casos representada pela linha vermelha. É possível observar que durante as primeiras semanas as duas curvas apresentam comportamentos opostos, como esperado. Porém, após o dia 20 de abril as curvas passam a apresentar um comportamento semelhante de queda, o que é bastante intrigante. Dentre as possibilidades que podem explicar esse comportamento estão: existe um grande índice de subnotificação dos óbitos causados pela COVID-19; ou a doença vem se tornando menos letal. Porém essas suposições não podem ser avaliadas com os dados utilizados nesta pesquisa. Ainda assim, o baixo número de testes por casos confirmados junto a taxa de letalidade dos casos de 2%, nos permite afirmar que os dados públicos da COVID-19 na Paraíba não representam de forma precisa a real situação do espalhamento da doença no estado.

Figura 13 - Testes por casos confirmado e taxa de letalidade dos casos na Paraíba



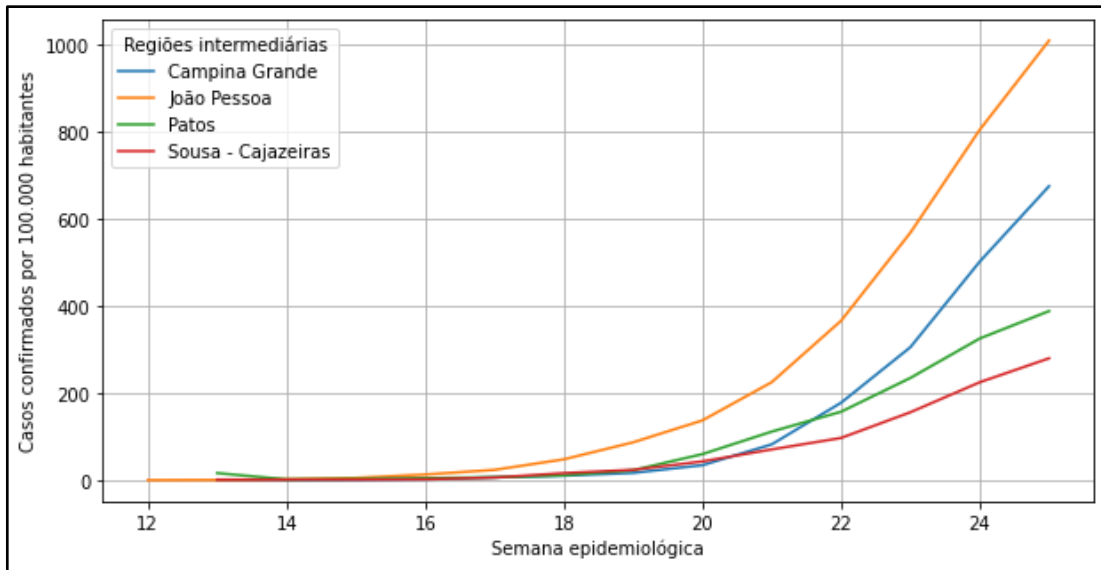
Fonte: Autor (2020)

4.2 ANÁLISE DA COVID-19 NAS REGIÕES DO ESTADO DA PARAÍBA

A análise das regiões foi iniciada comparando os números de casos confirmados. Como a população das diferentes regiões difere de forma significativa, é natural que regiões com maior população apresentem um maior número de casos, para compensar esse viés, foram utilizados então dados de casos confirmados por 100.000 habitantes, desse modo é possível avaliar de forma proporcional a difusão da doença nas diferentes regiões. O gráfico a seguir mostra o número de casos confirmados por 100.000 habitantes nas regiões geográficas intermediárias do estado da Paraíba (Figura 14).

Na Figura 14 podemos ver que as regiões de João Pessoa (linha amarela) e de Campina Grande (linha azul) possuem uma curva mais íngreme comparadas às curvas das regiões de Patos (linha verde) e Sousa - Cajazeiras (linha vermelha), indicando que nessas regiões os casos confirmados estão crescendo de forma mais rápida. Também é possível ver que a região de João Pessoa é a que possui mais casos confirmados proporcionalmente, seguida pelas regiões de Campina Grande, Patos e Sousa - Cajazeiras, seguindo a ordem de proximidade do litoral.

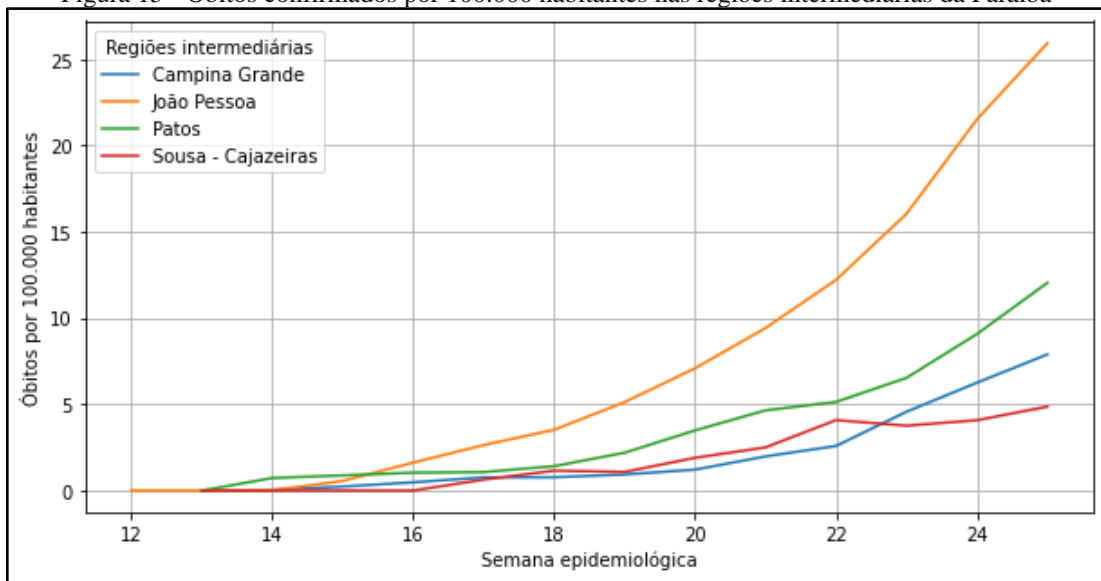
Figura 14 - Casos confirmados por 100.000 habitantes nas regiões intermediárias da Paraíba



Fonte: Autor (2020)

Após analisar o número de casos por 100.000 habitantes, foi analisado o número de óbitos por 100.000 habitantes. A figura 15 mostra o número de óbitos por 100.000 habitantes nas regiões intermediárias do estado.

Figura 15 - Óbitos confirmados por 100.000 habitantes nas regiões intermediárias da Paraíba



Fonte: Autor (2020)

As curvas de óbitos por 100.000 habitantes mostradas na Figura 15 não seguem o mesmo padrão observado pelas curvas de casos por 100.000 habitantes observadas na Figura 14. A região de João Pessoa (linha amarela) também aparece com os valores mais elevados, mas dessa vez se apresenta mais distante das demais regiões. A ordem também está diferente, a região de João Pessoa é seguida pelas regiões de Patos (linha verde), Campina Grande (linha azul) e Sousa-Cajazeiras (linha vermelha).

Após a análise dos óbitos por 100.000 habitantes, procurou-se por uma correlação entre a quantidade de testes rápidos distribuídos pelo governo estadual e a taxa de letalidade nas diferentes regiões. Porém nenhum coeficiente de correlação significativo foi encontrado, tanto

considerando o agrupamento dos municípios em regiões imediatas, quanto considerando o agrupamento em regiões intermediárias. Os coeficientes de correlações obtidos estão nos Quadros 1 e 2 a seguir.

Quadro 1 - Coeficientes de correlação considerando as regiões imediatas

	Taxa de letalidade dos casos	Testes distribuídos por 100.000 habitantes	Testes aplicados por 100.000 habitantes
Taxa de letalidade dos casos	1.000000	0.282199	-0.100721
Testes distribuídos por 100.000 habitantes	0.282199	1.000000	0.417515
Testes aplicados por 100.000 habitantes	-0.100721	0.417515	1.000000

Fonte: Autor (2020)

Quadro 2 - Coeficientes de correlação considerando as regiões intermediárias

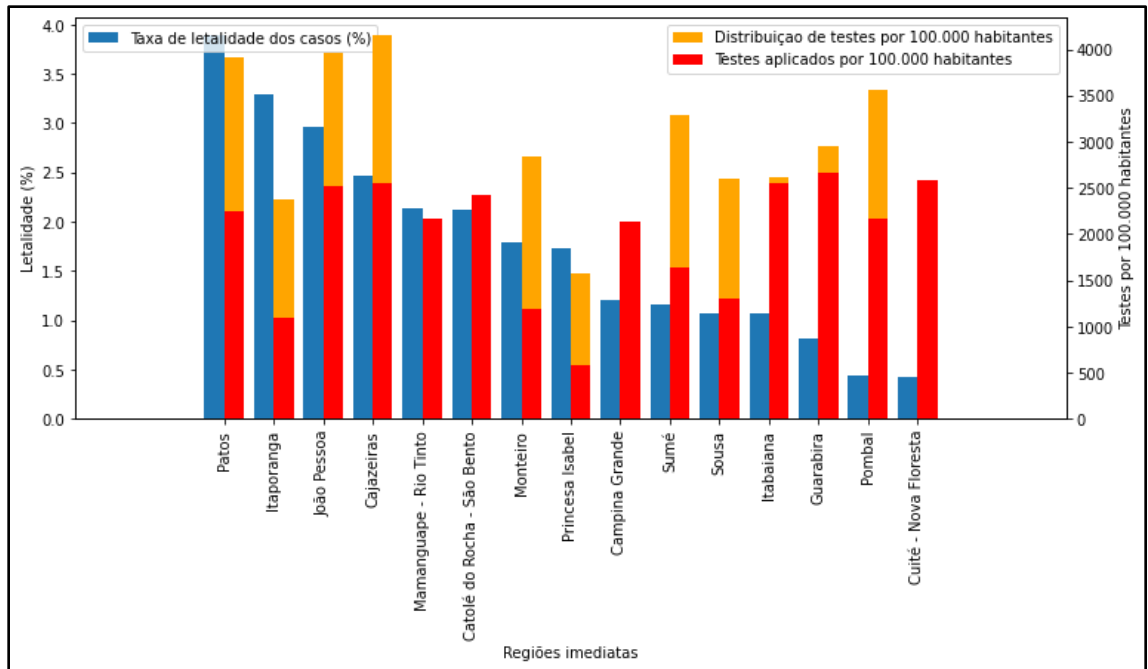
	Taxa de letalidade dos casos	Testes distribuídos por 100.000 habitantes	Testes aplicados por 100.000 habitantes
Taxa de letalidade dos casos	1.000000	0.554489	0.082417
Testes distribuídos por 100.000 habitantes	0.554489	1.000000	0.337178
Testes aplicados por 100.000 habitantes	0.082417	0.337178	1.000000

Fonte: Autor (2020)

Essa falta de correlação também pode ser observada nas Figuras 16 e 17, que mostram a letalidade, o número de testes rápidos distribuídos por 100.000 habitantes e o número de testes rápidos aplicados por 100.000 habitantes para as regiões geográficas imediatas e intermediárias, respectivamente.

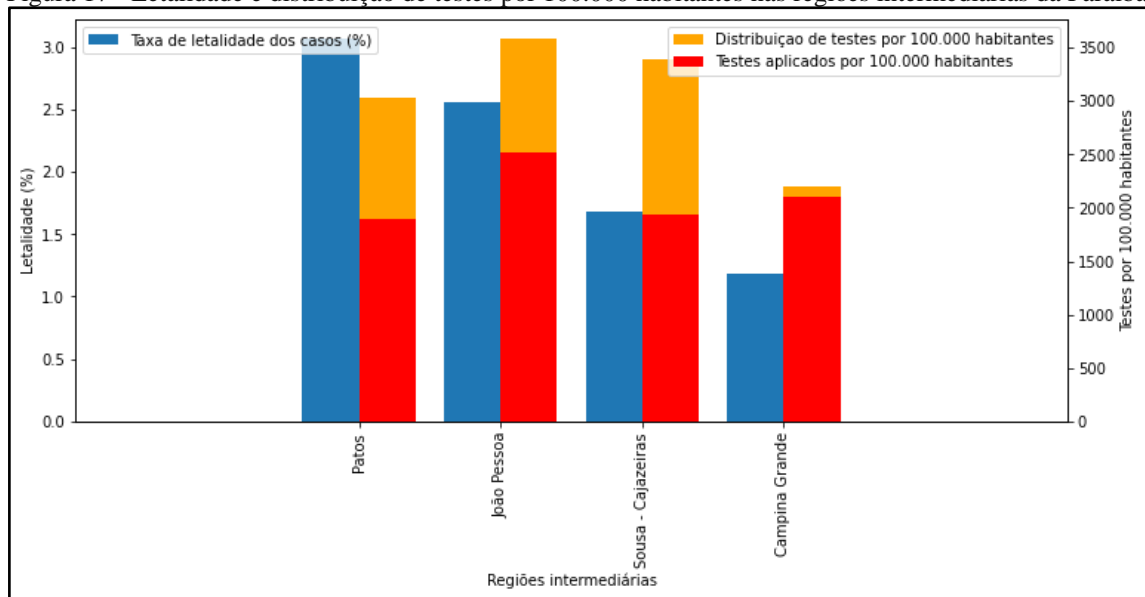
As Figuras 16 e 17 também mostram que a maioria das regiões possuem valores elevados de taxa de letalidade dos casos. Apenas as regiões imediatas de Guarabira, Pombal e Cuité - Nova Floresta apresentaram valores menores que 1%.

Figura 16 - Letalidade e distribuição de testes por 100.000 habitantes nas regiões imediatas da Paraíba



Fonte: Autor (2020)

Figura 17 - Letalidade e distribuição de testes por 100.000 habitantes nas regiões intermediárias da Paraíba



Fonte: Autor (2020)

5 CONSIDERAÇÕES FINAIS

O presente artigo buscou, por meio da análise de dados, encontrar respostas para três questionamentos levantados, relacionados a COVID-19 no estado da Paraíba. Os questionamentos são:

- O que os dados públicos dizem sobre o avanço do novo coronavírus no estado da Paraíba e em suas regiões?

- Os dados públicos disponíveis representam de forma precisa o número de contaminados e de falecimentos causados pelo novo coronavírus na Paraíba e em suas regiões?
- Existe correlação entre o número de testes rápidos distribuídos pelo Governo do Estado e a letalidade nas diferentes regiões Paraibanas?

Foram utilizados dados dos números de casos e óbitos confirmados para analisar o espalhamento do novo coronavírus pelo estado da Paraíba e suas regiões. Os resultados mostraram que os novos casos confirmados de COVID-19 permanecem crescendo no estado, no entanto esse crescimento vem se tornando cada vez menor. Os dados de novos óbitos confirmados, que também vinham crescendo, apresentaram pela primeira vez uma redução ao final da semana epidemiológica mais recente (semana 25). A região geográfica intermediária que vem sendo mais afetada pelo novo coronavírus é a região de João Pessoa, enquanto que a região que até então foi menos afetada é a região de Sousa - Cajazeiras.

Para analisar a confiabilidade dos dados, foram utilizados dois critérios de avaliação, o número de testes realizados por casos confirmados e a taxa de letalidade dos casos. Era esperado que esses dois parâmetros apresentassem comportamentos opostos ao longo do tempo, porém isso só ocorreu durante as primeiras semanas da doença na Paraíba. Após as primeiras semanas, os dois parâmetros passaram a se comportar de forma semelhante, o que é bastante intrigante. Apesar disso, o baixo número de testes por caso confirmado e a taxa de letalidade atual de aproximadamente 2%, nos leva a concluir que existem muitos casos de COVID-19 que não estão sendo diagnosticados, o que faz com que os dados disponíveis não representem de forma precisa a real situação do estado.

Por fim, tentou-se encontrar correlações entre a distribuição dos testes rápidos por 100.000 habitantes, o número de testes rápidos utilizados por 100.000 habitantes e a taxa de letalidade dos casos nas regiões imediatas e intermediárias. Porém não foram encontrados valores significativos de coeficiente de correlação.

Por meio das análises realizadas foi possível compreender de forma mais detalhada a situação da pandemia de COVID-19 no estado da Paraíba até o dia 21 de junho de 2020. Foi possível avaliar: o avanço da doença, que parece crescer em um ritmo cada vez menor no estado; a qualidade dos dados disponíveis, que de acordo com o estudo não representa de forma precisa o número real de infectados e de óbitos; e, a distribuição de testes rápidos entre os municípios, que não apresentou correlação significativa com a taxa de letalidade de casos nos municípios.

É importante observar que o conhecimento científico sobre a COVID-19 não está consolidado e vem sendo constantemente atualizado. Além disso, a disseminação da doença é afetada significativamente pelos hábitos da população e pelas políticas públicas estabelecidas. Sendo assim, é importante que análises semelhantes às apresentadas neste trabalho sejam repetidas futuramente.

Seria interessante a realização de novos estudos que busquem uma explicação para a redução observada na taxa de letalidade dos casos após as primeiras semanas do surto na Paraíba, Utilizando também nas análises, dados relacionados a óbitos causados por outras comorbidades.

Também seria interessante a utilização de técnicas de aprendizagem de máquina para tentar prever o avanço do novo coronavírus na região.

REFERÊNCIAS

BRASIL. Brasil confirma primeiro caso da doença. **Ministério da Saúde**, Brasília, DF, Coronavírus (COVID-19), Últimas notícias 26 fev. 2020a. Disponível em: <https://www.saude.gov.br/noticias/agencia-saude/46435-brasil-confirma-primeiro-caso-de-novo-coronavirus>. Acesso em: 19 jun. 2020.

_____. Calendário Epidemiológico 2020. **Sistema de Informação de Agravos de Notificação**, 2020b. Disponível em: <http://portalsinan.saude.gov.br/calendario-epidemiologico>. Acesso em: 19 jun. 2020.

_____. Número de curados do coronavírus é 13,2% maior que os casos ativos no Brasil. **Ministério da Saúde**, Brasília, DF, Coronavírus (COVID-19), Últimas notícias 21 jun. 2020c. Disponível em: <https://www.saude.gov.br/noticias/agencia-saude/47101-numero-de-curados-do-coronavirus-e-13-2-maior-que-os-casos-ativos-no-brasil>. Acesso em: 29 jun. 2020.

CUCINOTTA, Domenico; VANELLI, Maurizio. WHO declares COVID-19 a pandemic. **Acta bio-medica: Atenei Parmensis**, v. 91, n. 1, p. 157-160, 2020. DOI: 10.23750/abm.v91i1.9397. Disponível em: <https://europepmc.org/article/med/32191675>. Acesso em: 19 jun. 2020.

DORIGATTI, Ilaria et al. Report 4: severity of 2019-novel coronavirus (nCoV). **Imperial College London**, London, 2020. Disponível em: <http://www.corona-portal.de/ImpColl.pdf>. Acesso em: 19 jun. 2020.

GOIÁS. Calendário Epidemiológico. **Secretaria de Estado de Saúde de Goiás**, Vigilância Epidemiológica, 2020. Disponível em: <https://www.saude.go.gov.br/sesg/comissao-de-integracao-ensino-servico/712-suvisa/vigil%C3%A2ncia-epidemiol%C3%B3gica/8412-calend%C3%A1rio-epidemiol%C3%B3gico>. Acesso em: 29 jun. 2020

HUNTER, John D. Matplotlib: A 2D graphics environment. **Computing In Science & Engineering**, v. 9, n. 3, p. 90-95, 2007. DOI: 10.1109/MCSE.2007.55. Disponível em: https://www.researchgate.net/publication/3422921_Matplotlib_A_2D_Graphics_Environment. Acesso em: 15 jul. 2020.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Divisão regional do Brasil em regiões geográficas imediatas e regiões geográficas intermediárias : 2017**. Rio de Janeiro, IBGE, Coordenação de Geografia, 2017. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv100600.pdf>. Acesso em: 20 jun. 2020.

JÚNIOR, Marcos E. de O.. File:Brazil Paraiba RegioesGeograficasImediatas.svg. **Wikimedia Commons**, 2018. Disponível em: https://brasil.io/dataset/covid19/caso_full/. Acesso em: 16 jun. 2020.

JUSTEN, Álvaro. COVID-19 Boletins informativos e casos do coronavírus por município por dia. **Brasil.io.**, 2020. Disponível em: https://brasil.io/dataset/covid19/caso_full/. Acesso em: 16 jun. 2020.

KIM, Dong-Hyun et al. Understanding and interpretation of case fatality rate of coronavirus disease 2019. **Journal of Korean Medical Science**, v. 35, n. 12, 2020. DOI: <https://doi.org/10.3346/jkms.2020.35.e137>. Disponível em: <https://synapse.koreamed.org/Synapse/Data/PDFData/0063JKMS/jkms-35-e137.pdf>. Acesso em: 19 jun. 2020.

MCKINNEY, Wes. Data structures for statistical computing in python. **Proceedings of the 9th Python in Science Conference**. 2010. p. 51-56. Disponível em: <http://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf>. Acesso em: 19 jun. 2020.

PARAÍBA. Atualização Covid-19 | 21/06/2020. **Governo do Estado da Paraíba**, Paraíba, notícias, 21 jun. 2020a. Disponível em: <https://paraiba.pb.gov.br/diretas/saude/coronavirus/noticias/atualizacao-covid-19-21-06-2020>. Acesso em: 29 jun. 2020.

_____. Dados Epidemiológicos Covid-19 Paraíba. **Governo do Estado da Paraíba**, Paraíba, 2020b. Disponível em: <https://superset.plataformatarget.com.br/superset/dashboard/55/>. Acesso em: 21 jun. 2020.

_____. Governo distribui primeiro lote de Testes Rápidos de Covid-19 para 36 municípios a partir desta sexta-feira. **Governo do Estado da Paraíba**, Paraíba, notícias, 24 abr. 2020c. Disponível em: <https://paraiba.pb.gov.br/noticias/governo-distribui-primeiro-lote-de-testes-rapidos-de-covid-19-para-36-municipios-a-partir-desta-sexta-feira>. Acesso em: 29 jun. 2020.

_____. Microdados. **Governo do Estado da Paraíba**, Paraíba, 2020d. Disponível em: <https://superset.plataformatarget.com.br/superset/dashboard/microdados/>. Acesso em: 21 jun. 2020.

_____. Paraíba confirma primeiro caso de coronavírus. **Governo do Estado da Paraíba**, Paraíba, notícias, 18 mar. 2020e. Disponível em: <https://paraiba.pb.gov.br/noticias/paraiba-confirma-primeiro-caso-de-coronavirus>. Acesso em: 19 jun. 2020.

_____. Paraíba é um dos estados que mais testa para coronavírus. **Governo do Estado da Paraíba**, Paraíba, notícias, 26 jun. 2020f. Disponível em: <https://paraiba.pb.gov.br/noticias/paraiba-e-um-dos-estados-que-mais-testa-para-coronavirus-3>. Acesso em: 29 jun. 2020.

_____. Paraíba tem a menor taxa de letalidade por Covid-19 do Nordeste. **Governo do Estado da Paraíba**, Paraíba, notícias, 13 jun. 2020g. Disponível em: <https://paraiba.pb.gov.br/noticias/paraiba-tem-a-menor-taxa-de-letalidade-por-covid-19-do-nordeste>. Acesso em: 29 jun. 2020.

_____. Paraíba tem primeiro óbito confirmado por Covid-19. **Governo do Estado da Paraíba**, Paraíba, notícias, 31 mar. 2020h. Disponível em: <https://paraiba.pb.gov.br/noticias/paraiba-tem-primeiro-obito-confirmado-por-covid-19>. Acesso em: 19 jun. 2020.

PRADO, Marcelo et al. Análise de subnotificação do número de casos confirmados da COVID-19 no Brasil. **Núcleo de Operações e Inteligência em Saúde (NOIS)**, PUC-Rio, Rio

de Janeiro, 2020. Disponível em: <https://sites.google.com/prod/view/nois-pucrio/publica%C3%A7%C3%B5es?authuser=0>. Acesso em: 19 jun. 2020.

ROSER, Max; RITCHIE, Hannah; ORTIZ-OSPINA, Esteban; HASELL, Joe. Coronavirus Pandemic (COVID-19). **OurWorldInData.org**, 2020 . Disponível em: <https://ourworldindata.org/coronavirus>. Acesso em: 19 jun. 2020.

RUSSELL, Timothy W. et al. Estimating the infection and case fatality ratio for coronavirus disease (COVID-19) using age-adjusted data from the outbreak on the Diamond Princess cruise ship, February 2020. **Eurosurveillance**, v. 25, n. 12, p. 2000256, 2020. DOI: <https://doi.org/10.2807/1560-7917.ES.2020.25.12.2000256>. Disponível em: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.12.2000256?utm_source=sootoday.com&utm_campaign=sootoday.com&utm_medium=referral. Acesso em: 19 jun. 2020.

SINGHAL, Tanu. A review of coronavirus disease-2019 (COVID-19). **The Indian Journal of Pediatrics**, p. 1-6, 2020. DOI: <https://doi.org/10.1007/s12098-020-03263-6>. Disponível em: <https://rdcu.be/b436B>. Acesso em: 19 jun. 2020.

SOHRABI, Catrin et al. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). **International Journal of Surgery**, 2020. DOI: <https://doi.org/10.1016/j.ijssu.2020.02.034>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1743919120301977>. Acesso em: 19 jun. 2020.

VERITY, Robert et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. **The Lancet infectious diseases**, 2020. DOI: [https://doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7). Disponível em: <https://www.sciencedirect.com/science/article/pii/S1473309920302437>. Acesso em: 19 jun. 2020.