

UNIVERSIDADE DO ESTADO DO AMAZONAS
CENTRO DE ESTUDOS SUPERIORES DE TABATINGA
LICENCIATURA EM MATEMÁTICA

ELVIS DA SILVA RODRIGUES

ANÁLISE DA EFICIÊNCIA DO MODELO SIR QUANDO CONFRONTADO COM
DADOS REAIS

Prof. Orientador: Me. Edfram Rodrigues Pereira
Prof. Coorientador: Dr. Edilson de Carvalho Filho

Tabatinga-AM
2023

ELVIS DA SILVA RODRIGUES

ANÁLISE DA EFICIÊNCIA DO MODELO SIR QUANDO CONFRONTADO COM
DADOS REAIS

Prof. Orientador: Me. Edfram Rodrigues Pereira
Prof. Coorientador: Dr. Edilson de Carvalho Filho

Artigo científico apresentado como Trabalho de Conclusão de Curso para obtenção de nota parcial na disciplina Trabalho de Conclusão de Curso II, ministrada pela Prof.^a Dr.^a. Karem Keyth de Oliveira Marinho, do Curso de Licenciatura em Matemática do Centro de Estudos Superiores de Tabatinga da Universidade do Estado do Amazonas.

Tabatinga-AM
2023

Ficha Catalográfica

Ficha catalográfica elaborada automaticamente de acordo com os dados fornecidos pelo(a) autor(a).
Sistema Integrado de Bibliotecas da Universidade do Estado do Amazonas.

R696aa Rodrigues, Elvis da Silva
Análise da Eficiência do Modelo SIR Quando
Confrontado com Dados Reais / Elvis da Silva Rodrigues.
Manaus : [s.n], 2023.
29 f.: color.; 29 cm.

TCC - Graduação em Matemática - Licenciatura -
Universidade do Estado do Amazonas, Manaus, 2023.
Inclui bibliografia
Orientador: Pereira, Edfram Rodrigues
Coorientador: Filho, Edilson de Carvalho

1. modelo SIR. 2. eficiência. 3. COVID-19. 4.
subnotificação. 5. python. I. Pereira, Edfram Rodrigues
(Orient.). II. Filho, Edilson de Carvalho (Coorient.). III.
Universidade do Estado do Amazonas. IV. Análise da
Eficiência do Modelo SIR Quando Confrontado com
Dados Reais

Elaborado por Jeane Macelino Galves - CRB-11/463

ELVIS DA SILVA RODRIGUES

ANÁLISE DA EFICIÊNCIA DO MODELO SIR QUANDO CONFRONTADO COM
DADOS REAIS

Artigo científico apresentado como Trabalho de Conclusão de Curso para obtenção de nota parcial na disciplina Trabalho de Conclusão de Curso II, ministrada pela Prof.^a Dr.^a. Karem Keyth de Oliveira Marinho, do Curso de Licenciatura em Matemática do Centro de Estudos Superiores de Tabatinga da Universidade do Estado do Amazonas.

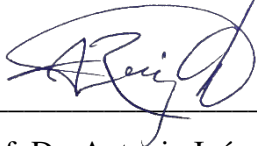
Data de aprovação: 28 de março de 2023.



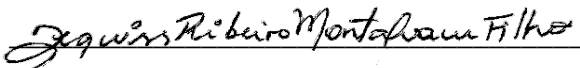
Prof. Me. Edfram Rodrigues Pereira (Presidente da Banca avaliadora / Orientador – UEA)



Prof. Dr. Edilson de Carvalho Filho (Coorientador – UEA)



Prof. Dr. Antonio Iván Ruiz Chaveco (Examinador Interno – UEA)



Prof. Esp. Zequias Ribeiro Montalvam Filho (Examinador Interno – UEA)

AGRADECIMENTOS

Agradeço primeiramente a Deus pelo fôlego de vida e força necessária para terminar esta pesquisa.

Agradeço, especialmente, aos meus pais José Ferreira Rodrigues e Elisete da Silva Rodrigues por todo o apoio desde a minha infância até este momento, e por sempre acreditarem e me incentivarem a buscar ir além.

À minha noiva, Natali Façanha de Souza, por toda a motivação e incentivo que, desde o início, me proporcionou.

Agradeço também aos meus professores, em especial, Edfram Rodrigues Pereira, Edilson de Carvalho Filho e Zequias Ribeiro Montalvam por todo o conhecimento passado ao longo destes 4 anos de graduação, além das conversas incentivadoras a respeito de projetos futuros.

Agradeço ainda aos amigos que fiz ao longo da graduação, em especial, Geovane da Silva Dias, que além de ser um grande incentivador, também me proporcionou, em nossas conversas, importantes momentos de descontração e reflexão a respeito de Deus e seu grande amor, temas que tem grande importância em minha vida.

RESUMO

Tendo em vista a grande importância do estudo de modelos matemáticos que auxiliam na compreensão da dinâmica de transmissão de doenças infecciosas, em especial o modelo SIR, este trabalho se propõe a investigar a eficiência do modelo SIR aplicado a dados de COVID-19 no estado do Amazonas. Para isso, foram realizados estudos a respeito do modelo SIR, também acerca da COVID-19 e, além disso, foi criado um algoritmo em linguagem Python para realizar as análises necessárias, comparando com dados reais fornecidos pela Fundação de Vigilância em Saúde (FVS). Como resultado, vimos que o modelo SIR simples não se mostra eficiente frente aos dados reais fornecidos pelos órgãos oficiais, já que conforme as aplicações, enquanto o modelo estima picos muito altos, os dados reais mostram picos bem menores. Deste modo, podemos compreender a discrepância como sendo causada, principalmente, pela subnotificação, já que o SIR simples considera que todos os casos são diagnosticados. Com isso apresentamos neste trabalho uma extensão do modelo SIR para considerar também a subnotificação.

Palavras-chave: modelo SIR; eficiência; COVID-19; subnotificação; python.

RESUMEN

Dada la gran importancia de estudiar modelos matemáticos que ayuden a comprender la dinámica de transmisión de enfermedades infecciosas, en particular el modelo SIR. Este trabajo propone investigar la eficiencia del modelo SIR aplicado al COVID-19 en el estado de Amazonas. Para ello se realizaron estudios sobre el modelo SIR, también sobre COVID-19, se creó un algoritmo en lenguaje Python para luego realizar los análisis necesarios, comparando con datos reales proporcionados por la Fundación de Vigilancia de la Salud (FVS). Como resultado, vimos que el modelo SIR simple no es eficiente frente a los datos reales proporcionados por los organismos competentes, ya que según las aplicaciones, mientras el modelo estima picos muy altos, los datos reales muestran picos mucho más pequeños. Así, podemos entender que la discrepancia se debe, principalmente, al subregistro, ya que el SIR simple considera que todos los casos están diagnosticados. Por lo tanto, en este trabajo presentamos una extensión del modelo SIR para considerar también el subregistro.

Palabras-clave: modelo SIR; COVID-19; eficiencia; subdeclaración; python.

SUMÁRIO

1 INTRODUÇÃO.....	8
2 FUNDAMENTAÇÃO TEÓRICA	9
2.1 A COVID-19 no Amazonas	9
2.3 Modelo SIR	11
2.4 Determinação do parâmetro β	14
2.5 Taxa de reprodutibilidade basal.....	15
3 METODOLOGIA.....	17
3.1 Aplicação do modelo	18
4 RESULTADOS E DISCUSSÃO	20
4.1 Mudança do Modelo SIR adicionando a subnotificação	21
4.2 Validação do modelo	23
4.3 Aplicação do modelo aos dados reais de infectados.....	25
5 CONSIDERAÇÕES FINAIS	29
REFERÊNCIAS	30

1 INTRODUÇÃO

Como sabemos, a COVID-19 foi um dos maiores desafios já enfrentados no estado do Amazonas. Por aqui a doença se mostrou particularmente desafiadora, já que em muitos locais não havia hospitais voltados ao tratamento de casos graves da doença. Em meio a este fato, este trabalho surgiu com a proposta de analisar a eficiência do modelo SIR na predição do comportamento da doença durante todo o ciclo pandêmico e, assim, contribuir para o entendimento deste modelo, da própria doença e o real impacto da mesma no estado.

O modelo que será trabalhado é muito importante para o estudo de epidemias e serve de base para diversos outros modelos que buscam, por meio de ferramentas matemáticas, compreender como ocorre a dinâmica de transmissão de uma pandemia, entendendo seu ciclo e os fatores (parâmetros) associados.

Compreender este modelo é essencial para o desenvolvimento de outros modelos. Precisamos investigar até onde o modelo é eficiente para que então possamos desenvolver formas de melhorá-lo. Se tratando de um tema tão importante, pois lida diretamente com a questão do bem estar das pessoas e da sua sobrevivência, quanto mais estudarmos maneiras de compreender as epidemias, melhor lidaremos com desastres dessa natureza.

O modelo SIR foi proposto inicialmente por Kermack e Mckendrick em 1927, estando fortemente ligado a ideia da cinética química. Mas no lugar de interações entre elementos, o modelo propõe uma interação entre caixas nas quais se subdivide uma população, as caixas de infectados, suscetíveis e recuperados. O modelo considera na caixa de recuperados tanto as pessoas que sobreviveram ao vírus quanto as que faleceram em decorrência do mesmo. Por este fato, essa caixa é também chamada de caixa de removidos.

Ao longo deste trabalho mostraremos o que é o modelo SIR Clássico. E a maneira como ele será aplicado. O mesmo será confrontado com dados reais, onde buscaremos os melhores parâmetros que tornam as projeções do modelo mais próximas dos dados reais. E análise da subnotificação por meio de uma modificação do modelo SIR de forma a incluir esta variável, já que o modelo SIR simples não considera este fator.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 A COVID-19 no Amazonas

Com o sistema público de saúde despreparado e com uma estrutura para o tratamento de casos graves aquém das demandas surgidas durante o desenvolvimento da pandemia, o Amazonas foi umas das federações que mais sofreram ao longo de mais de 3 anos da doença no país. O coronavírus (SARS-COV-2) surgiu na China no final de 2019 e o que era inicialmente apenas um pequeno problema de saúde pública local daquele país logo se espalhou para o mundo gerando impactos desastrosos e pondo a prova os sistemas de saúde de todo o planeta.

No Amazonas o primeiro caso detectado ocorreu no dia 13 de março de 2020, contabilizado 17 dias depois do primeiro caso registrado no Brasil. Inicialmente, o epicentro de casos no estado foi a capital Manaus, tendo, após 30 dias de transmissão local no estado, cerca de 88% dos casos registrados acumulados. E o município do interior com maior agravamento era Manacapuru, que neste período, registrava mais da metade dos casos de COVID-19 no interior. Até esse momento haviam sido registrados no Amazonas cerca de 1050 casos da doença (MONTEIRO, XAVIER e MAZZARARI, 2022).

Os casos a partir desse período não pararam de crescer e o sistema de saúde local sofreu com a grande demanda provocada pelo avanço da doença. Cerca de 6 meses depois, em 08 de agosto, a situação no estado se inverteu e os casos acumulados no interior representavam cerca de 64% do total de 106.428 (MONTEIRO, XAVIER e MAZZARARI, 2022).

É importante dizer que as cidades do interior do Amazonas, como Tabatinga por exemplo, não possuíam condições estruturais para o tratamento de casos graves da doença. Deste modo, se fez necessário o deslocamento destes pacientes para a capital do estado. Assim, notamos que a qualidade da estrutura em saúde pública no Estado se mostra bastante desequilibrada, já que a capacidade de atendimento para doenças graves no interior é muito menor do que na capital, e isso ficou ainda mais evidente durante a pandemia.

O Amazonas pode ser considerado um dos Estados que mais causou preocupação em relação ao número de casos. O primeiro aumento exponencial de casos da doença ocorreu durante os primeiros meses, terminando em maio de 2020. Esta primeira onda se deu em decorrência da disseminação das linhagens B.1.19 e B.1.1.28, esta última tendo surgido no final da primeira onda (NAVECA, NASCIMENTO, SOUZA et al, 2021).

Após o período de intenso crescimento do número de infectados, o estado experimentou uma fase de estabilidade, o que contribuiu para a flexibilização das medidas de segurança biológica.

Então, na metade do mês de dezembro do ano de 2020, após 7 meses de estabilidade com pequenas variações, surge uma nova variante denominada P.1. Esta variante, juntamente com o descumprimento das medidas de contenção da doença, causou uma explosão de casos de COVID-19 (NAVECA, NASCIMENTO, SOUZA et al, 2021).

Esta explosão de casos colapsou o sistema de saúde da capital e a alta demanda fez com que se implantasse em Manaus uma das maiores crises na área de saúde já vista no país, com recordes de internações por COVID-19, situação agravada em função da falta de oxigênio nos hospitais (G1-Amazonas).

Em meio a esta realidade, vimos que existe cada vez mais a necessidade de se utilizar ferramentas como a modelagem dentro da epidemiologia matemática a fim de analisar e dimensionar a dinâmica de propagação de uma doença tal como a COVID-19. Essas ferramentas podem inclusive auxiliar os governos sobre quais medidas devem ser tomadas para conter o avanço da doença, fazendo com que compreendam a dimensão do problema a ser enfrentado.

2.2 Epidemiologia Matemática

Ao longo da história a humanidade enfrentou diversas epidemias, tais como a varíola e a peste negra, como consequência, uma quantidade muito grande de pessoas perderam as suas vidas em decorrência dessas doenças infecciosas. Apesar do contágio por estas doenças ser enfraquecido com o tempo, e as mesmas serem controladas naturalmente antes que a população suscetível seja completamente infectada, o custo do ponto de vista de vidas humanas pode ser bastante alto até que se atinja esta fase naturalmente.

Deste modo, estudar e prever como será o desenvolvimento e a dinâmica de transmissão de doenças deste tipo tem importância fundamental para a proteção da vida e a garantia da sobrevivência. Desde o início, a humanidade buscou maneiras de garantir a sua sobrevivência e, diante de doenças provocadas por seres invisíveis, como os vírus, esta preocupação se torna ainda maior. Afinal, como controlar e eliminar o que se conhece pouco ou nada?

Neste contexto, surge a ideia de associar conceitos matemáticos à dinâmica de transmissão e disseminação destas doenças. Um dos primeiros a introduzir conceitos

matemáticos na análise de doenças infecciosas foi Daniel Bernoulli, tendo publicado em 1766 um artigo onde descrevia os efeitos da vacinação contra varíola na expectativa de vida (MAGEL e RUAN, 1936).

Na busca por analisar matematicamente como ocorre a transmissão de doenças infecciosas é importante ter em mente que a mesma depende de diversos fatores, inclusive o geográfico. Assim, a maneira como uma enfermidade contagiosa se propaga em diferentes regiões, estados ou cidades pode ser distinta.

Isso ocorre devido a uma dependência entre o número de pessoas que se infectam, a quantidade de pessoas já infectadas e a quantidade de pessoas sujeitas a serem infectadas, além da maneira como ocorre a dinâmica social de cada local (a movimentação das pessoas em cada localidade).

Em meio a isso, um dos conceitos mais importantes da epidemiologia matemática é o chamado princípio da ação das massas, que nos diz que a propagação de uma doença infecciosa está relacionada de forma direta com o número de indivíduos suscetíveis, infectados e, ainda, a taxa de contato entre os mesmos, em outras palavras, é proporcional à quantidade de indivíduos contaminados e indivíduos saudáveis (HAMER, 1906 apud UGALDE, I.A. A; DÍAZ, R.G; LEZCA, W. M, 2020).

2.3 Modelo SIR

O modelo epidemiológico SIR (Suscetível, Infectados, Recuperados) é um dos modelos mais importantes no estudo de doenças infecciosas. Este modelo, proposto inicialmente por Kermack e McKendrick no ano de 1927, serviu de base para diversos outros modelos a partir do século XX. Ele se baseia em uma relação dinâmica entre três caixas, ou compartimentos, (a qual está dividida uma determinada população). Cada uma dessas caixas contém uma fração da população total, sendo esta fração variável em relação ao tempo, ou seja, são funções dependentes do tempo (KERMACK; MCKENDRIK, 1927)

$S(t)$: Caixa de Suscetíveis

$I(t)$: Caixa de Infectados

$R(t)$: Caixa de Recuperados

A fração populacional na caixa $S(t)$ é composta por indivíduos que estão vulneráveis a contrair a doença, na caixa $I(t)$ estão os indivíduos que contraíram e estão se recuperando e, a caixa $R(t)$ contém todos os indivíduos que já recuperaram da doença (MAGEL

e RUAN, 1936). Busca-se com esse modelo, compreender de que maneira está ocorrendo a transferência de indivíduos de uma caixa para a outra.

Essa dinâmica de transferência de indivíduos entre caixas depende de alguns fatores. Um destes fatores é a probabilidade de infecção p , que representa a probabilidade de um indivíduo contrair o agente infeccioso ao entrar em contato com outro indivíduo já infectado. Outro fator é a taxa de contato β_1 , ela determina a frequência de contato entre um indivíduo infectado e indivíduos saudáveis (FARIAS, 2017). O produto desses fatores determina a taxa de transmissão

$$\beta = p\beta_1$$

Como os indivíduos infectados podem apenas infectar indivíduos suscetíveis, podemos escrever a capacidade de infecção como:

$$\beta \frac{S}{N}$$

sendo $N = S(t) + I(t) + R(t)$, a este termo chamamos razão de infecção.

Analisando mais a fundo, temos que $\frac{S(t)}{N}$ nos dá a probabilidade de um infectado encontrar um indivíduo saudável, assim $\beta_1 \frac{S}{N}$ é o número médio de pessoas contactadas por um indivíduo infectado. Como existe uma dependência entre o número de indivíduos infectados e as novas infecções (cada indivíduo com o vírus pode potencialmente transmiti-lo com a mesma capacidade transmissão), podemos escrever o número de novos infectados como

$$\beta \frac{S}{N} I$$

dessa forma, como está saindo esta quantidade de indivíduos da caixa dos suscetíveis ($S(t)$), então a variação nessa caixa é descrita por

$$\frac{dS(t)}{dt} = -\beta \frac{S}{N} I$$

Além disso, ao passo em que existem indivíduos sendo infectados, também terão outros se recuperando da doença a uma taxa de recuperação μ , sendo esta o inverso do tempo de duração da infecção em um indivíduo.

Com isso é possível determinar a quantidade de infectados que se recuperam pelo termo

$$\mu I$$

dessa forma podemos definir a variação na caixa de infectados em relação ao tempo como

$$\frac{dI(t)}{dt} = \beta \frac{S}{N} I - \mu I$$

Por fim, os indivíduos infectados que se recuperam passam a ser recuperados e, portanto, a variação na caixa dos recuperados ($R(t)$) é dada por

$$\frac{dR(t)}{dt} = \mu I$$

com isso obtemos as três equações diferenciais ordinárias (EDO) nas quais se baseia o modelo SIR, juntas elas formam o seguinte sistema de EDOs

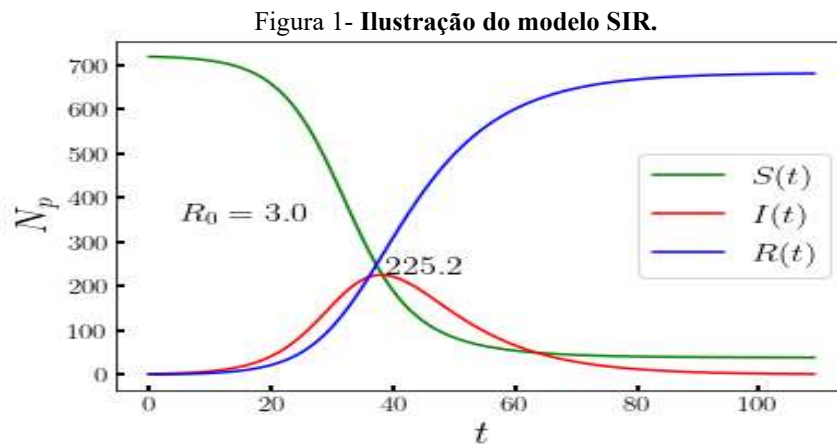
$$\begin{cases} \frac{dS(t)}{dt} = -\beta \frac{S}{N} I \\ \frac{dI(t)}{dt} = \beta \frac{S}{N} I - \mu I \\ \frac{dR(t)}{dt} = \mu I \end{cases}$$

com este sistema diferencial, se define matematicamente o modelo epidemiológico SIR.

Notemos que neste modelo SIR clássico, não se considera dinâmica vital, ou seja, a população total se mantém constante, $N = \text{constante}$. Podemos ver facilmente que variação populacional ($\frac{dN}{dt}$) ao longo do tempo é nula, já que

$$\frac{dN}{dt} = \frac{dS(t)}{dt} + \frac{dI(t)}{dt} + \frac{dR(t)}{dt} = -\beta \frac{S}{N} I + \beta \frac{S}{N} I - \mu I + \mu I = 0$$

logo, o que ocorre é apenas uma redistribuição de indivíduos de uma caixa para outra. É considerado ainda, que o indivíduo uma vez recuperado não volta a ser infectado (KERMACK E MCKENDRICK, 1927) e além disso, a caixa de recuperados armazena tanto indivíduos que resistiram à doença e estão imunes a mesma, quanto indivíduos que não resistiram e vieram a óbito. A seguir é apresentado o gráfico ilustrativo do modelo considerando uma taxa de transmissão igual a 0.3 e uma taxa de recuperação igual a 0.1, uma população de 720 pessoas e 100 dias de transmissão.



A ilustração acima mostra a quantidade total de indivíduos em cada caixa variando com o tempo, assim, na caixa $I(t)$ temos o total de pessoas que ainda estão infectadas até o respectivo dia. O pico, nesta ilustração, ocorre no dia 38 com um total de aproximadamente 225 pessoas na caixa de infectados.

2.4 Determinação do parâmetro β

O beta, como visto anteriormente, representa a taxa de transmissão e é o parâmetro que determina a transição de indivíduos da caixa $S(t)$ para a caixa $I(t)$. A análise deste parâmetro é essencialmente importante, já que o mesmo possui impacto direto no número absoluto de infectados. É ele que, predominantemente, influencia o pico de infecção, além da quantidade de indivíduos infectados ao atingir este pico.

O método que será utilizado para este fim será o Método dos Mínimos Quadrados (MMQ), criado por Friedrich Gauss, proposto inicialmente para determinar as órbitas dos planetas. Porém, com o passar dos séculos, se tornou uma importante ferramenta para análise de diversos outros fenômenos (ALMEIDA, 2015, pg. 17). Este método tem como fundamento, minimizar a soma do quadrado dos resíduos, sendo este definido como

$$M(\bar{\alpha}, f) = \sum_{i=1}^n (y_i - \hat{y})^2$$

em que $\bar{\alpha}$ e f são, respectivamente, um conjunto de parâmetros e a função associados ao modelo. Além disso, y_i é o valor observado (através da análise temporal do fenômeno estudado), \hat{y} é o valor estimado pelo modelo. A função f descreve, em teoria, o fenômeno estudado, sendo ele de qualquer natureza (ALMEIDA, 2015, pg. 31).

Quando se trata deste método, podemos nos perguntar porque o quadrado e não simplesmente a soma dos resíduos, a soma dos cubos ou a soma dos módulos dos resíduos. O resíduo representa a diferença entre os dados observados e o previsto pelo modelo, sendo assim, parece razoável fazer a soma dos resíduos tender a zero, porém poderia acontecer da soma dar zero simplesmente por haver simetria entre resíduos (mesmo valor em módulo, porém com sinais opostos).

Notamos que apesar de resolvermos os dois problemas descritos anteriormente utilizando qualquer potência par, o quadrado ainda é melhor, pois é a menor potência possível de ser utilizada considerando a diferenciabilidade e evitar que o resíduo se aproxime de zero apenas por simetria (ALMEIDA, 2015).

A maneira que o método foi aplicado para a determinação do β utiliza diretamente a expressão acima. Para o caso não linear, a expressão é chamada de R-quadrado. Ele nos mostra o quão bom é o ajuste, ou seja, se ao minimizar os resíduos considerando os parâmetros e a função, conseguimos um ajuste satisfatório.

Esta forma de buscar o melhor ajuste é essencialmente útil aqui pelo fato de não conhecermos características suficientes dos dados de forma a darmos um palpite preciso de qual função melhor descreve os dados, como no caso de ajuste linear por exemplo.

2.5 Taxa de reprodutibilidade basal

O R_0 representa a quantidade de contaminação secundária, ou seja, o número médio de pessoas para quais um infectado pode transmitir a doença (ROSS, 1911; KEELING; ROHANI, 2008 apud VILCHES, 2019). Resolvendo a segunda equação diferencial do modelo por separação de variáveis, considerando $S \approx N$, temos

$$\frac{dI(t)}{dt} = \beta \frac{S}{N} I - \mu I$$

$$I(t) = ce^{(\beta - \mu)t} = ce^{\mu(\frac{\beta}{\mu} - 1)t}$$

a expressão $\frac{\beta}{\mu}$ é chamada de R_0 . Logo temos as seguintes condições pelas quais é possível determinar o intervalo onde buscar o parâmetro β :

- $R_0 > 1$ - está ocorrendo a transmissão com um aumento exponencial de infectados (epidemia);
- $R_0 < 1$ - os casos da doença tendem a diminuir com o tempo, ou seja, a infecção está sob controle;

- $R_0 = 1$ - temos um equilíbrio epidêmico.

3 METODOLOGIA

Para o desenvolvimento deste trabalho, inicialmente foi feito um estudo sobre o modelo SIR, a fim de conhecer o mesmo de forma mais detalhada. Foi realizado ainda um estudo a respeito da COVID-19 e seu desenvolvimento no estado do Amazonas, já que conforme estabelecido, esta pesquisa está voltada ao estudo o modelo SIR aplicado a COVID-19.

Com base nas equações diferenciais já apresentadas foi criado um algoritmo em linguagem de programação Python para resolver o sistema de equações diferenciais associado ao modelo. Depois de criado, foram realizadas uma série de testes variando os parâmetros de forma a analisar as curvas de maneira a verificar se apresentavam o resultado gráfico esperado.

Após estes procedimentos, fez-se a busca de dados da COVID-19 no estado do Amazonas, do início da pandemia até o mês de janeiro de 2023. Os dados foram obtidos através do Painel COVID-19 Amazonas da Fundação de Vigilância em Saúde (FVS) vinculada a Secretaria Estadual de Saúde do Amazonas.

Em seguida, foi dado início as análises dos dados reais comparando com as previsões do modelo. Nesta fase podemos ver uma grande discrepância entre os casos diários preditos pelo modelo e os dados reais. Isso nos instigou a realizar uma adaptação do modelo SIR, tal adaptação será descrita ao longo da apresentação dos resultados.

Destaco ainda que o intervalo onde o parâmetro β é tomado está diretamente associado ao R_0 . Conforme vimos, a contaminação é crescente quando o valor deste é maior que 1. Logo, se queremos analisar com foco na pandemia da COVID-19, onde notavelmente tivemos uma contaminação crescente, devemos considerar este caso. Assim, temos que $R_0 > 1$, onde

$$R_0 = \frac{\beta}{\mu} \Rightarrow \beta = R_0 \mu$$

definindo um intervalo conveniente para o R_0 , de forma que se tenha a garantia de infecção com potencial de causar uma epidemia, escolhemos

$$R_0 \in (1, 6)$$

é importante dizer que o intervalo superior definido para o conjunto em que foi tomado o R_0 foi escolhido com base em estudos a respeito do mesmo. A média do R_0 para a COVID-19 ficou em torno de 3 pessoas (OXFORD-BRAZIL EBM ALLIANCE, 2022). Com isso tomamos, como forma de garantir que o algoritmo em Python chegaria na melhor estimativa

para o parâmetro β , um limite superior igual a 6. Com base nessas considerações e tomando uma taxa de recuperação fixa igual a $\frac{1}{10}$, adotamos

$$\beta \in (0.1, 0.6)$$

o algoritmo varreu este intervalo, com um passo de 0.001, de forma que para cada β calculamos um resíduo (R^2), no final o resultado é uma lista de resíduos e o menor deles está associado ao melhor ajuste.

Após a implementação do algoritmo, levando em conta as considerações anteriores, foi possível obter os dados necessários para as análises do modelo SIR, colocando os dados do modelo em confronto com os dados reais de órgãos de saúde.

3.1 Aplicação do modelo

Para implementação do modelo será utilizada a expansão em série de Taylor até a primeira ordem (Aproximação Linear). Para mais detalhes consulte LIMA (2019) e STEWART (2013). A partir do valor inicial para cada caixa, podemos aproximar os próximos valores pelas seguintes expressões

$$\begin{cases} S(t+1) = S(t) - \beta \frac{S}{N} I + O(t) \\ I(t+1) = I(t) + \beta \frac{S}{N} I - \mu I + O(t) \\ R(t+1) = R(t) + \mu I + O(t) \end{cases}$$

onde temos sempre um erro associado, que é da ordem do intervalo de tempo que estamos considerando. Como consideramos variações pequenas de tempo, podemos desprezar o erro e as equações ganham a seguinte forma

$$\begin{cases} S(t+1) = S(t) - \beta \frac{S}{N} I \\ I(t+1) = I(t) + \beta \frac{S}{N} I - \mu I \\ R(t+1) = R(t) + \mu I \end{cases}$$

É importante frisar que o modelo SIR sozinho não nos dá diretamente a quantidade de infecções diárias, mas a soma de todos os indivíduos que ainda estão na caixa I no dia t . Analisando a variação diária na caixa I temos que

$$I(1) = I(0) + \beta \frac{S(0)}{N} I(0) - \mu I(0)$$

$$I(2) = I(1) + \beta \frac{S(1)}{N} I(1) - \mu I(1)$$

·
·
·

$$I(t + 1) = I(t) + \beta \frac{S(t)}{N} I(t) - \mu I(t)$$

Deste modo, a quantidade de pessoas que entram, de fato no dia atual, na caixa de infectados (incidência diária) é dada por

$$\beta \frac{S(t)}{N} I(t) = I_{cd}(t) = I(t + 1) + I(t)(\mu - 1)$$

em que I_{cd} representa os casos diários e é com esta expressão que obtemos os dados diários preditos pelo modelo, necessários para o procedimento de ajuste.

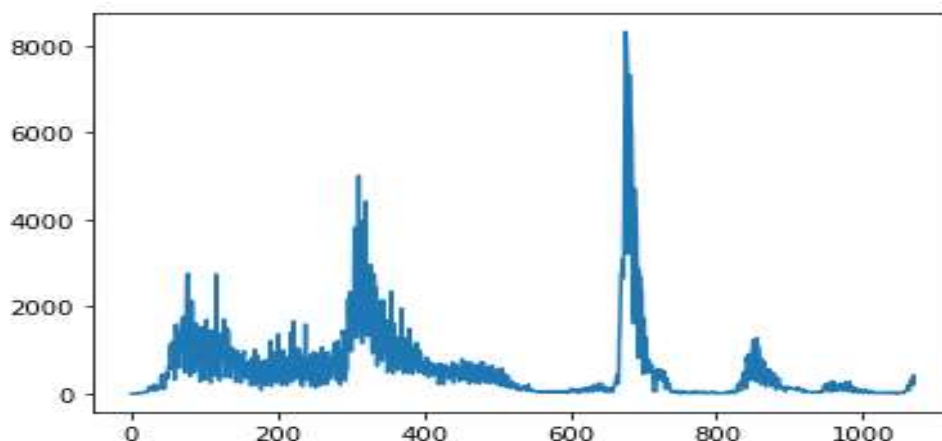
4 RESULTADOS E DISCUSSÃO

Nesta seção serão apresentados os resultados das análises do modelo SIR simples e também justificaremos a necessidade da modificação deste modelo, já que o mesmo não considera a subnotificação.

É importante destacar que os resultados das análises serão apresentados em 4 partes, considerando os quatro períodos de crescimento exponencial da doença no estado. O primeiro causado pelas linhagens B.1.19 e B.1.1.28, o segundo causado pela variante P.1(gama). Mostraremos a análise modelo simples na primeira fase e então justificaremos a necessidade de uma modificação neste modelo, já que o mesmo não consegue compensar a subnotificação.

Acrescento ainda que para os dois últimos picos, o critério de escolha do dia inicial será com base na análise da estabilidade de infecção diária mostrados na série temporal de casos. A seguir temos a imagem com a série no período 13/03/2020 a 19/02/2023:

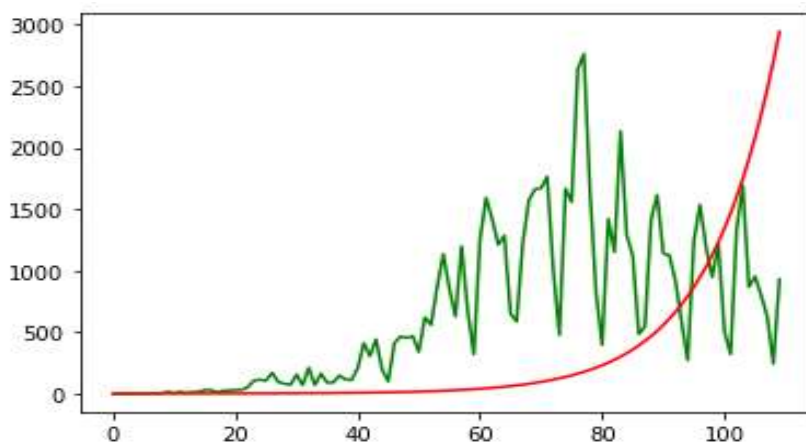
Figura 2 - Série temporal de casos diários de COVID-19 entre os anos de 2020 e 2023.



Fonte: elaborado pelo autor (2023)

A curva do modelo associada ao número de infectados tem a característica de ser suave e sem ruídos, enquanto a curva dos dados reais notificados é bastante ruidosa. Aplicando o modelo com o método de minimização de resíduo para 110 dias, obtemos $R_0 = 1,93$ como a melhor taxa de reprodução basal com um resíduo de $6.33187418 \times 10^{-7}$. A seguir podemos ver o resultado gráfico após o ajuste

Figura 3- Ilustração do ajuste do modelo SIR aos dados reais. Na imagem vemos a curva gerada tomando o melhor beta, escolhido por minimização de resíduos.



Fonte: elaborado pelo autor (2023)

Apesar do resultado ser melhor segundo o processo de minimização de resíduos, temos problemas significativos. O dia em que ocorre o pico da doença, segundo os dados reais, não está sendo previsto pelo modelo de forma satisfatória. O mesmo está predizendo o pico para depois do centésimo dia, enquanto que os dados reais mostram que ocorreu no dia 77 após o início da doença. Deste modo, o ajuste mostrado, apesar de ser o melhor possível, considerando os fatores já mencionados, está longe de ser o ideal.

O fato apresentado sugere fortemente que o modelo não foi eficiente no ajuste devido a grande subnotificação, fazendo com que a minimização leve a um menor resíduo antes do modelo atingir seu pico. Isso se deve ao fato de que no início do desenvolvimento da curva de infectados no modelo SIR os casos diários são menores que no pico e, por tanto, estão mais próximos dos dados notificados.

4.1 Mudança do Modelo SIR adicionando a subnotificação

O fato de o modelo SIR clássico não considerar a subnotificação, algo que na realidade ocorre, faz com que o mesmo não seja suficiente para modelar os dados reais. Na verdade, existe uma taxa das pessoas que se infectam que não são contados, pois não procuraram os hospitais para realizar o teste ou devida a capacidade de testagem e até mesmo a quantidade de testes disponíveis serem insuficientes.

Portanto, existe uma taxa que denotaremos por ρ , que indica o percentual de pessoas infectadas que são contabilizadas pelos órgãos responsáveis. Também existe uma outra taxa que aqui denotaremos por ϵ e esta indica o percentual de indivíduos que contraem a doença e não passam por nenhum diagnóstico, ou seja, não são contabilizados.

Logo, considerando a incidência diária de casos $\beta \frac{S}{N} I$, temos a quantidade de pessoas que são efetivamente contadas pelos órgãos de saúde diariamente dada por $\rho \beta \frac{S}{N} I$, do mesmo modo, a quantidade de pessoas que não são diagnosticadas é dada por $\epsilon \beta \frac{S}{N} I$. Aqui vamos chamar de I_d a caixa de novos casos diagnosticados e I_{nd} a caixa de novos casos não diagnosticados. Assim, temos

$$I = I_d + I_{nd}$$

com isso podemos escrever a variação na caixa I como

$$\frac{dI(t)}{dt} = \frac{d(I_d + I_{nd})}{dt} = \rho \beta \frac{S}{N} I - \mu I_d + \epsilon \beta \frac{S}{N} I - \mu I_{nd}$$

ou seja, a variação na caixa I é a combinação das variações nas caixas I_d e I_{nd} . Como na caixa I do SIR tradicional, ao mesmo tempo em que tem gente entrando por estarem contaminadas, também tem gente saindo da caixa devido estarem recuperados, a uma mesma taxa μ tanto em I_n quanto em I_{nd} . É importante notarmos que

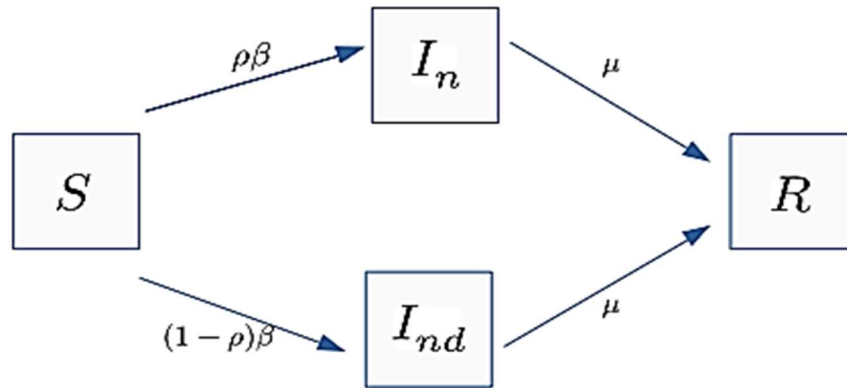
$$\rho \beta \frac{S}{N} I + \epsilon \beta \frac{S}{N} I - \mu I_d - \mu I_{nd} = \beta \frac{S}{N} I - \mu I \quad \Rightarrow \quad \rho \beta + \epsilon \beta = \beta$$

ou seja, $\rho + \epsilon = 1$, daí concluímos que $\rho \leq 1$, a mesma observação vale para ϵ . Utilizando as considerações anteriores podemos obter um novo sistema de equações diferenciais da seguinte forma

$$\begin{cases} \frac{dS(t)}{dt} &= -\beta \frac{S}{N} (I_d + I_{nd}) \\ \frac{dI_d(t)}{dt} &= \rho \beta \frac{S}{N} (I_d + I_{nd}) - \mu I_d \\ \frac{dI_{nd}(t)}{dt} &= (1 - \rho) \beta \frac{S}{N} (I_d + I_{nd}) - \mu I_{nd} \\ \frac{dR(t)}{dt} &= \mu (I_d + I_{nd}) \end{cases}$$

Para compreendermos melhor como está ocorrendo a interação entre as caixas consideradas neste sistema temos o seguinte diagrama

Figura 4- Ilustração da dinâmica entre as caixas do modelo SIR modificado.



Fonte: elaborado pelo autor (2023)

a aproximação por série de Taylor de primeira ordem das novas caixas fica

$$I_d(t+1) = I_d(t) + \rho\beta \frac{S}{N} (I_d(t) + I_{nd}(t)) - \mu I_d(t)$$

$$I_{nd}(t+1) = I_{nd}(t) + (1-\rho)\beta \frac{S}{N} (I_d(t) + I_{nd}(t)) - \mu I_{nd}(t)$$

deste modo a expressão para casos diários notificados tem a forma

$$\rho\beta \frac{S(t)}{N} (I_d(t) + I_{nd}(t)) = I_{ad}(t) = I_d(t+1) + I_d(t)(\mu - 1)$$

em que I_{ad} é o número de casos atuais notificados.

4.2 Validação do modelo

Para validação do modelo, utilizamos $R_0 = 3$ e uma taxa de notificação $\rho = 1$, ou seja, o caso em que todos infectados são diagnosticados para produzir dados simulados. Após isso, foi adicionado um ruído com distribuição normal a esses dados. O ruído foi gerado considerando uma média da distribuição normal igual ao número de infectados no dia de infecção considerado e um desvio padrão como uma taxa dessa média. Desta forma obtivemos uma curva que tenha um comportamento semelhante a curva de casos reais notificados.

Apresentamos 4 testes para diferentes ruídos, os valores para o σ e a taxa de notificação estão relacionados seguir:

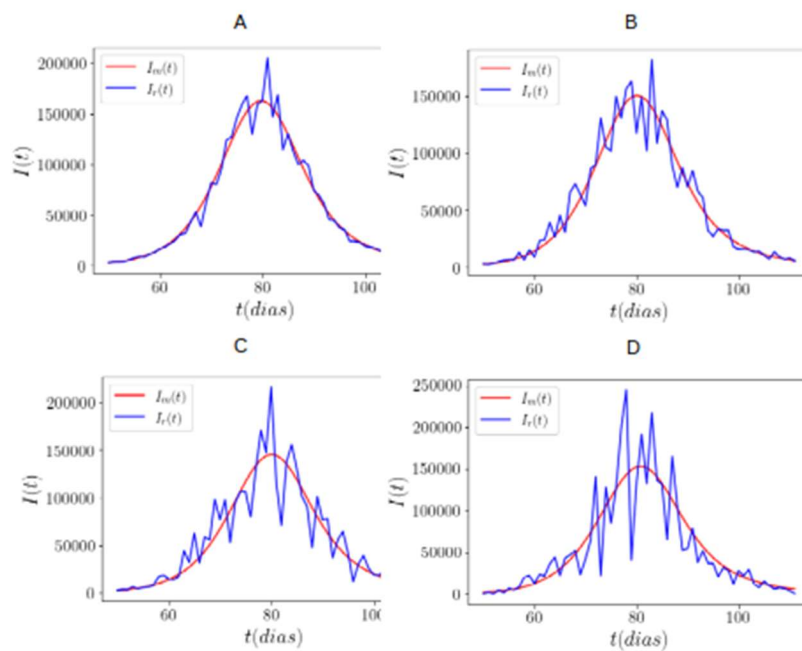
Tabela com valores utilizados na validação do modelo SIR Modificado

Desvio padrão	Taxa de notificação	Taxa de reprodução basal	Resíduo
---------------	---------------------	--------------------------	---------

10% da média	3.0	0.99	5.7×10^9
20% da média	2.99	0.919	11.3×10^9
30% da média	2.99	0.889	$24,3 \times 10^9$
40% da média	2.97	0.949	$55,6 \times 10^9$

A seguir, nas imagens A, B, C e D, apresentamos os resultados gráficos dos ajustes para os valores de desvio padrão apresentados na tabela e média tomada de acordo a descrição acima.

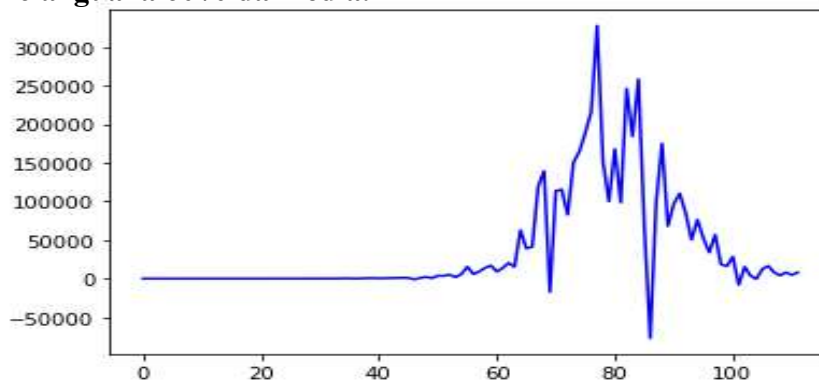
Figura 5- Ajustes utilizando dados simulados.



Fonte: elaborado pelo autor (2023)

Logo, podemos concluir que o algoritmo funciona para a predição dos parâmetros ideais que melhor se ajustam aos dados. Ressalto que não é apresentado testes para variâncias maiores pelo fato de que a partir de 50% da média, a curva ruidosa perde a característica esperada para uma série temporal de uma epidemia. A seguir apresentamos um gráfico a título de demonstração

Figura 6- Curva do modelo SIR modificado após a adição do ruído com variância igual a 60% da média.



Fonte: elaborado pelo autor (2023)

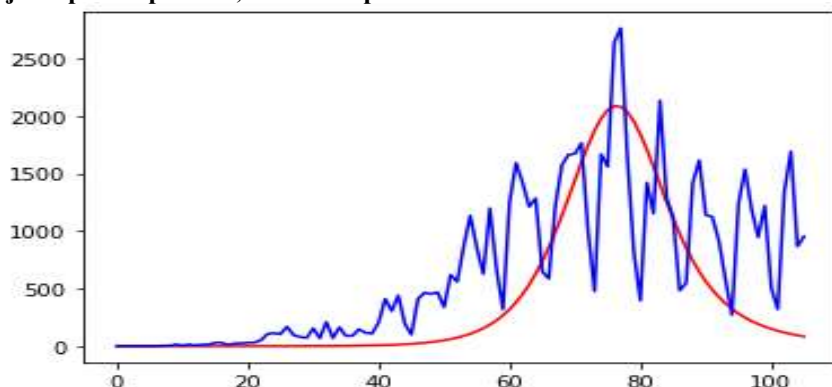
o gráfico acima é obtido com variância de 60% da média e, notavelmente, se distancia do que se espera em uma pandemia, já que não pode haver número negativos de casos.

4.3 Aplicação do modelo aos dados reais de infectados

Como já foi dito, o Amazonas experimentou 4 períodos de crescimento exponencial da doença, por isso dividiremos os dados casos em quatro partes de forma a analisar cada pico individualmente. Essa forma de aplicar o método se deu pelo fato de o modelo simular apenas um pico, desconsiderando novas ondas de infecção.

Período 1. Aqui foram consideradas as infecções ocorridas entre o primeiro dia e o centésimo sexto dia, de forma a incluir o primeiro pico da doença. Ressaltando que o ajuste é sempre feito através da minimização de resíduos, considerando as saídas do modelo SIR modificado. O pico de acordo com os dados Fundação de Vigilância em Saúde (FVS) ocorre no dia 77 após o início da infecção no estado, com um total de 2763 casos. Após o ajuste do modelo modificado aos dados reais obtermos o seguinte resultado,

Figura 7 - Ilustração do ajuste aos dados reais. Em vermelho temos a curva de ajuste para o período, fornecida pelo modelo SIR modificado.

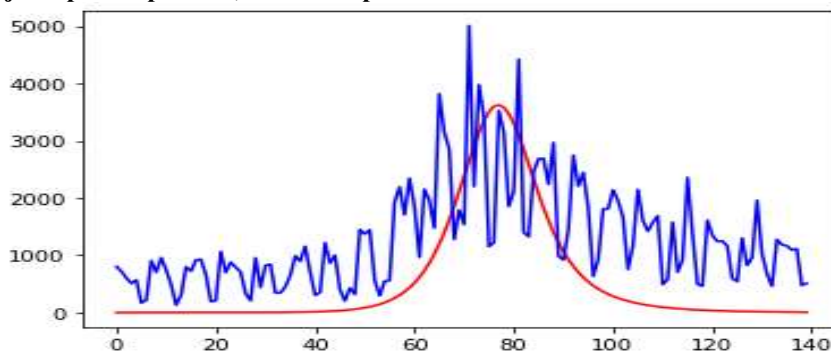


Fonte: elaborado pelo autor (2023)

o modelo SIR modificado prevê o pico de infecção para o dia 76, com um total de aproximadamente 2086 casos, com $R_0 = 3.09$ e uma taxa de subnotificação, $(1 - \rho)$, de aproximadamente 0.988. Com isso temos um percentual aproximado de 98.8% de casos não notificados.

Período 2. Neste período foi tomado um intervalo de 140 dias entre os dias 240 e 380, de forma a pegar todo desenvolvimento da segunda onda da doença no estado. O gráfico do ajuste para este período é mostrado a seguir.

Figura 8- Ilustração do ajuste aos dados reais. Em vermelho temos a curva de ajuste para o período, fornecida pelo modelo SIR modificado.

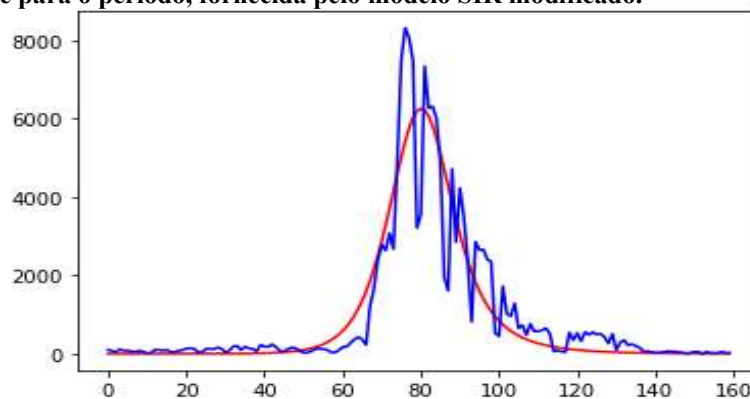


Fonte: elaborado pelo autor (2023)

Os dados reais datam o pico no dia 311 após o início da infecção (o dia zero neste gráfico é 240), com um total de 5009 infectados. O modelo prevê o pico para o dia 317, com 3652 infectados. O R_0 estimado pelo modelo é de 3.08 e uma taxa de notificação de 2.12%. É importante ressaltar que na realidade a taxa de diagnóstico não é fixa, tendo dias em que a testagem é bem maior que em outros e, portanto, mais pessoas são diagnosticadas. Este fato pode explicar a diferença entre a quantidade de pessoas no pico dadas pelo modelo e a quantidade mostrada pelos dados reais coletados.

Período 3. O período em que o modelo será aplicado agora está entre os dias 600 e 760 após o início da doença, de forma a pegar o novo crescimento exponencial da COVID-19 no estado nesta fase da pandemia. O gráfico com a curva predita pelo modelo e dados reais para este período é apresentado a seguir.

Figura 9 - Ilustração do ajuste aos dados reais. Em vermelho temos a curva de ajuste para o período, fornecida pelo modelo SIR modificado.



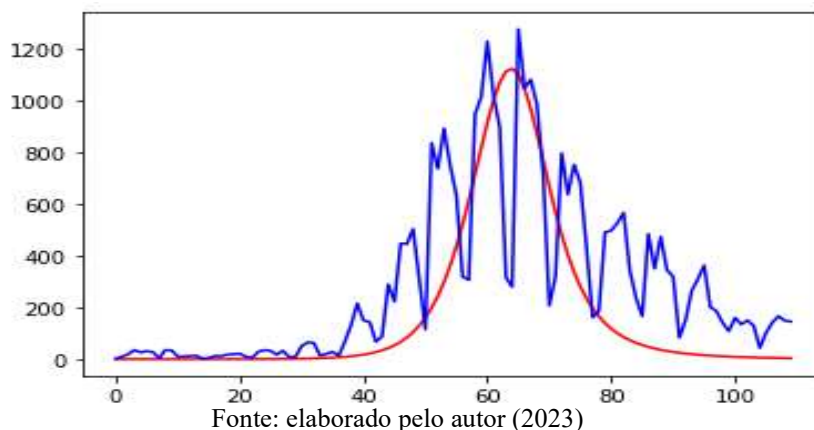
Fonte: elaborado pelo autor (2023)

O dia zero do gráfico corresponde ao dia 600 após o início da infecção, do mesmo modo, o dia 760 é o dia 160 no gráfico. Aplicação do modelo modificado prevê o pico de infecção para este período de 680 após o início da pandemia, com um total de 6257 infectados e $R_0 = 2.99$.

Além disso, o modelo mostra para esse período uma taxa de subnotificação de 96.16 %, ou seja, apenas aproximadamente 3.83% dos casos foram notificados. Os dados reais mostram que o pico ocorreu no dia 676, ou seja, o modelo está ajustando o pico para 4 dias depois do pico obtido diretamente dos dados. Isto pode ser explicado pelo fato de a quantidade de testes diários ser variável, enquanto o modelo aqui trabalhado considera que todos os dias se testa a uma mesma taxa ρ . Este foi o período de maior crescimento do número de casos diários de infecção.

Período 4. Para este período utilizamos o intervalo de dias entre 790 e 940 após início da infecção. Nesta fase, depois de um tempo de estabilidade, o Amazonas experimentou um novo crescimento exponencial de casos. Vejamos o gráfico com o ajuste utilizando o modelo.

Figura 10- Ilustração do ajuste aos dados reais. Em vermelho temos a curva de ajuste para o período, fornecida pelo modelo SIR modificado.



Os dados neste período mostram que pico da COVID-19 provocado pela cepa dominante ocorreu no dia 855 após o início da infecção do estado (dia 65 no gráfico acima) com um total de 1276 infectados. O modelo SIR modificado prevê o pico para o dia 64, com aproximadamente 1123 casos. Este resultado é obtido com $R_0 = 3.57$ e uma taxa de notificação de 0.051, ou seja, o modelo está indicando que 99.49% dos casos neste período não foram notificados.

5 CONSIDERAÇÕES FINAIS

Notamos no decorrer deste trabalho que o modelo SIR simples não descreve bem com dados subnotificados. Sabendo que o modelo considera que todos os casos diários são notificados, porém na realidade isto não ocorre. Dessa forma, o modelo acaba se distanciando muito dos dados diários fornecidos pelos órgãos de saúde.

Cada um dos parâmetros utilizados no modelo SIR simples ou SIR modificado têm importante significado epidemiológico, determinando todo o comportamento do modelo. Desta forma estimar o melhor parâmetro de ajuste frente aos dados reais nos possibilita compreender melhor a dinâmica de transmissão de doenças infecciosas como a COVID-19. Os parâmetros β e μ , por exemplo, determinam o R_0 , ou seja, o número médio de pessoas que são infectadas por cada indivíduo infeccioso e através dos ajustes, por meio de minimização de resíduos, podemos estimá-lo, obtendo a taxa de transmissão.

Além destes parâmetros mencionados a adição dos parâmetros ρ e ϵ , de notificação e subnotificação respectivamente, é de suma importância para adequar o modelo a realidade, já que a testagem não foi eficiente. Assim é possível estimar qual foi a real dimensão da doença, ou seja, a adição destas taxas deixa mais clara o grau de infecção que o estado viveu no período pandêmico.

O modelo modificado que é apresentado foi desenvolvido de maneira independente, isto é, sem a consulta de outros trabalhos. Assim, considero, que no futuro ao trazer considerações de trabalhos já realizados, podemos obter resultados muito mais eficientes. Por fim, pudemos compreender neste estudo que a falta de testagem da população pode comprometer o monitoramento do real estado da infecção.

REFERÊNCIAS

- ARONSON, J.K; BRASSEY, J; MAHTANI, K.R. **“Quando isso acabará?”: Uma introdução aos números de reprodução viral, R0 e Re.** Oxford-Brazil EBM Alliance, 2022. Disponível em: <<https://oxfordbrazilebm.com/index.php/quando-isso-acabara-uma-introducao-aos-numeros-de-reproducao-viral-r0-e-re/>>. Acesso em: 12 de mar. De 2022.
- FARIAS, A.V. **Um estudo da modelagem epidemiológica SIR usando conceitos de derivadas de ordem inteira e fracionária.** 2017, pág. 17 a 22. Dissertação (Bacharel em Matemática Aplicada) - Universidade Federal do Rio Grande, Rio Grande, 2017.
- FUNDAÇÃO DE VIGILÂNCIA EM SAÚDE DO AMAZONAS. c2023. Painel Covid-19 Amazonas. Disponível em <<http://saude.am.gov.br/painel/corona/>>. Acesso em: 19 de fev. de 2023.
- GAZEL, A.S; CRUZ, V. **G1 Amazonas, 2022.** Disponível em: <<https://g1.globo.com/am/amazonas/noticia/2022/01/14/crise-do-oxigenio-no-amazonas-completa-um-ano-com-impunidade-e-incerteza-causada-pela-omicron.ghtml>>. Acesso em: 14 de dez. de 2022.
- Kermack, W., and McKendrick, A. (1927): **A Contribution to the Mathematical Theory of Epidemics.** Proc. Roy. Soc. A 115, 700 – 721.
- LIMA, Elon Lages. **Curso de Análise.** 15.ed. Rio de Janeiro, IMPA. 2019.
- MAGAL, Pierre e RUAN, Shigui. **Structured Population Models in Biology and Epidemiology.** Springer, 1936.
- MONTEIRO, T. R; XAVIER, D. S; MAZZARI, A. S. Epidemiologia da COVID-19 no Amazonas, Brasil. BEPA, 2020.
- NAVECA, F.G; NASCIMENTO, V.; DE SOUZA, V.C. et al. COVID-19 in Amazonas, Brazil, was driven by the persistence of endemic lineages and P.1 emergence, **Nature medicine**, v.27, pg.1230 -1238, 2021. <https://doi.org/10.1038/s41591-021-01378-7>.
- STEWART, J. **Cálculo.** Volume 1, 7 ed. São Paulo. Cengage Learning, 2013.
- UGALDE, I.A. A; DÍAZ, R.G; LEZCA, W. M. **El modelo SIR básico y políticas antiepidémicas de salud pública para la COVID-19 en Cuba.** Revista Cubana de Salud Pública, 2020;46 (Supl. especial):e2597.
- VILCHES, T. N. **Epidemiologia matemática em redes complexas.** 2019. Tese (Doutorado em Biometria) - Universidade Estadual Paulista “Júlio de Mesquita Filho”, São Paulo, 2019.