

MÉTODO DE VALIDAÇÃO DE MODELO MATEMÁTICO PARA PANDEMIA DO SARS-Cov-2 A PARTIR DE MINERAÇÃO DE DADOS E ANÁLISE ESTATÍSTICA.

Rafael Pereira Santana*¹, Anderson Lupo Nunes¹ e Pedro Maia Salomone¹

1. Grupo de Pesquisa em Física e Ensino de Física, Laboratório de Física Nilton de Souza Medeiros, Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro, Campus Duque de Caxias, RJ, Brasil. e-mail: rafael.santana@ifrj.edu.br

Resumo: O mundo hoje, no ano de 2020, vive a mais grave pandemia de caráter global que se tem notícia desde a Gripe Espanhola, ocorrida entre 1918 e 1920. Segundo registros da Organização Mundial de Saúde (OMS), a pandemia provocada pelo vírus Sars-Cov-2, teve início em dezembro de 2019 e está presente em todos os continentes e quase todos os países, com mais de 79 milhões de infectados e mais de 1,7 milhões de mortos até dezembro de 2020. Diversos modelos matemáticos aplicados à Epidemiologia foram adotados ao longo do tempo. Um dos mais adotados é o SIR (Suscetíveis, Infectados e Removidos), desenvolvido na Índia, por Kermack e McKendrick, em 1927. Em nossa pesquisa foi feita uma ampla coleta de dados acerca da pandemia do SARS-Cov-2 a partir de relatórios da OMS, da Dadax Limited (empresa de dados chinesa) e da Universidade Johns Hopkins nos Estados Unidos da América (EUA). Foram coletados dados de diversos países, relativos ao número de casos confirmados, recuperados e mortos. Neste trabalho, construímos um modelo matemático que descreve a evolução da pandemia a partir de uma similaridade com modelos já adotados na área da física nuclear. Para validação do modelo matemático, escolhemos os dados da Alemanha, devido à confiabilidade dos dados disponibilizados. Desta forma, foi realizada uma análise estatística para qualificar o desempenho do método e o caráter preditivo do modelo matemático. Até o momento foram coletados 11.716 dados brutos, dos quais realizamos uma mineração dos dados relevantes para esta pesquisa.

Palavras-chave: Análise estatística, Mineração de dados, Epidemiologia; Modelos matemáticos; SARS-Cov-2; Novo Coronavírus, Pandemia; Equações diferenciais.

Adesão ao escopo da BJEDIS: Neste trabalho, utilizamos de modelagem matemática para analisar um número muito grande de dados relacionados à contaminação e propagação de um vírus altamente contagioso em populações humanas. Utilizamos planilhas de cálculo para mineração dos dados relevantes e produção de novos dados reconhecer tendências e fazer previsões. Acreditamos que nosso trabalho pode contribuir de forma relevante tanto para a revista como para a comunidade científica. Informando e atualizando pesquisadores e fortalecendo as ciências voltadas para evolução de epidemias.

*Endereço de correspondência para o autor: Laboratório de Física Nilton de Souza Medeiros, Campus Duque de Caxias, Instituto Federal do Rio de Janeiro, CEP: 25.050-100, Duque de Caxias, RJ, Brasil; Tel: ++55(21)99721-2726; E-mail: rafael.santana@ifrj.edu.br



1. INTRODUÇÃO

Em dezembro de 2019, um surto de pneumonia de causa desconhecida foi relatado pelas autoridades de saúde em Wuhan, na China. Resultados de investigação laboratorial identificaram um novo coronavírus como responsável pelo surto. O novo coronavírus foi nomeado, pelo Comitê Internacional de Taxonomia de Vírus (ICTV), como coronavírus 2, SARS-Cov-2. A Organização Mundial da Saúde (OMS) denominou como COVID-19 a doença causada pelo SARS-Cov-2 (CHAVES, 2020) (1).

A pandemia mais recente, que teve impactos até agora maiores que a do SARS-COV-2, foi a Gripe Espanhola, ocorrida entre janeiro de 1918 e dezembro de 1920. Estimativas avaliam que foram 500 milhões de infectados, 50 milhões de mortos e atingiu, através das duas primeiras ondas em 1918, os Estados Unidos, Europa, Ásia, Américas Central e do Sul. Segundo Goulart (2005) (2), as primeiras notícias acerca de uma epidemia na Espanha chegaram aos jornais do Rio de Janeiro em agosto de 1918. A alcunha de Gripe Espanhola surgiu porque, neste país, as informações acerca da epidemia eram amplamente disseminadas na imprensa, diferentemente de outros países onde houve uma tentativa de encobrir a doença. Estas primeiras informações acerca daquela pandemia foram recebidas pela população fluminense com desdém e descrença (GOULART, 2005) (2). Tanto em pandemias anteriores como na atual, esta descrença pode atrapalhar a realização de procedimentos efetivos de controle da propagação e contaminação das doenças virais.

De acordo com o relatório de 29/12/2020 da Organização Mundial da Saúde (WHO, 2020) (3), até dezembro de 2020, o país com maior número de mortos e infectados com o SARS-Cov-2, são os EUA, com 18.648.989 infectados e 328.014 mortos. Em segundo lugar vem o Brasil, com 7.448.560 infectados e 190.488 mortos. A gravidade da situação sanitária em nosso país já seria suficiente para justificar uma pesquisa sobre a disseminação do vírus.

Luiz (2012) (4) define epidemia como uma doença infecciosa que se alastra em grandes proporções, com grande número de mortos, em um tempo muito reduzido. A modelagem matemática na área da epidemiologia é feita através do estudo de equações, geralmente equações diferenciais ordinárias, que descrevem a interação entre a população e o ambiente, o que resulta em uma análise pormenorizada a respeito da doença infecciosa.

Neste trabalho, utilizamos um modelo matemático para descrever a evolução da pandemia. Assim, podemos evidenciar os impactos que a disseminação do vírus SARS-Cov-2 tem imposto ao planeta, tanto pela imensa quantidade de mortos, quanto pelos devastadores reflexos na rotina das pessoas, assim como na economia global (COSTA, 2020) (5) e (MACEDO, 2020) (6).

Esta pesquisa mostra-se importante, não apenas por conta da atualidade e relevância da temática, mas também pela possibilidade de benefícios decorrentes da construção e do aperfeiçoamento dos modelos matemáticos aplicados à epidemiologia e que poderão ser utilizados em epidemias futuras. Sendo assim, quanto maior for o conhecimento a respeito da doença e do modo como ela se propaga, maior será a eficiência dos métodos adotados para impedir sua transmissão, e até mesmo o estudo de ações preventivas, quando houver.

Devido a uma gigantesca quantidade de dados, precisamos fazer a mineração dos dados relevantes dos quais utilizamos uma modelagem matemática. A partir desta modelagem, equações foram adicionadas à uma planilha de cálculos para a produção de novos dados para futura análise estatística com objetivo reconhecer tendências e de fazer provisões.

A BJEDIS é uma revista com objetivo de realizar a ampla divulgação de pesquisas relacionadas à análise e mineração de dados, acreditamos que nosso trabalho pode contribuir de forma relevante tanto para a revista quanto para a comunidade científica, informando e atualizando pesquisadores e fortalecendo as ciências voltadas para o estudo sobre evolução de epidemias com a utilização de modelagem matemática e mineração de dados.

1.1. Modelo Matemático SIR

Diversos modelos matemáticos aplicados à área da Epidemiologia foram adotados ao longo do tempo na literatura científica. Cristovão (2015) (7), Okhuuse (2020) (8), Costa (2020) (9), Ciufolini (2020) (10) e Fokas (2020) (11) são alguns exemplos de modelos mais recentes que são utilizados para descrever qualquer epidemia ou, mais especificamente, a atual pandemia do SARS-Cov-2.



Um dos modelos matemáticos mais adotados em epidemiologia é o SIR (Suscetíveis, Infectados e Removidos). Kermack e McKendrick (1927) (12), desenvolveram o modelo SIR na Índia em 1927. É o primeiro e um dos mais bem sucedidos modelos epidemiológicos desenvolvidos. Neste modelo matemático, sobre as doenças que se espalham na população, propuseram uma divisão em classes disjuntas, denotadas por:

- a) Suscetíveis, $S(t)$, que representam a classe dos indivíduos saudáveis, ou seja, aqueles indivíduos que estão expostos a uma eventual infecção.
- b) Infectados, $I(t)$, que representam a classe dos indivíduos que estão infectados e que são possíveis causadores de novas infecções, ou seja, indivíduos infecciosos.
- c) Recuperados, $R(t)$, que representam os indivíduos recuperados das doenças, passando assim a se tornarem imunes a uma nova infecção.

De acordo com Oliveira (2018) (13), seus postulados são:

- i. Toda a pessoa que faz parte da população e ainda não foi infectada, é Suscetível, $S(t)$.
- ii. Quando um suscetível contrai a doença, torna-se um Infectado, $I(t)$.
- iii. Aquele indivíduo que evolui para cura da doença ou para morte, torna-se um Removido, $R(t)$.
- iv. A taxa de natalidade e a taxa de mortalidade são iguais, o que implica que a população total é constante.
- v. A população total, $N(t)$, é a soma dos Suscetíveis, Infectados e Removidos.

$$N(t) = S(t) + I(t) + R(t) \quad (\text{eq.1})$$

Assim, segundo Cristóvão (2015) (7) e Oliveira (2018) (13), as equações que relacionam a evolução no tempo do número de suscetíveis, infectados e removidos são:

$$\frac{dS}{dt} = -\beta \cdot S(t) \cdot I(t) \quad (\text{eq.2})$$

$$\frac{dI}{dt} = \beta \cdot S(t) \cdot I(t) - \gamma \cdot I(t) \quad (\text{eq.3})$$

$$\frac{dR}{dt} = \beta \cdot S(t) \cdot I(t) - \gamma \cdot I(t) \quad (\text{eq.4})$$

Onde a equação (eq.2) corresponde a taxa de variação do número de suscetíveis; a equação 3 (eq.3) corresponde a taxa de variação do número de infectados e a equação (eq.4) corresponde a taxa de variação do número de removidos.

Os coeficientes β e γ correspondem, respectivamente à:

$$\begin{array}{ll} \beta & \rightarrow \text{Coeficiente de transmissão} \\ \gamma & \rightarrow \text{Taxa de recuperação} \end{array}$$

Note que este modelo pode ser bastante aprimorado a partir de dados confiáveis e mais condizentes com a realidade.

A gripe espanhola, por exemplo, foi provocada pelo vírus da influenza e tornou-se uma doença que os especialistas chamam de endêmica, ou seja, permanece como uma doença infecciosa que atinge uma grande quantidade de indivíduos, mas com uma letalidade muito baixa, sem nunca deixar de existir (GOULART, 2005) (2).

Existe uma grande probabilidade de que a pandemia do SARS-Cov-2 torne-se também endêmica. Assim, o uso de modelos matemáticos pode continuar sendo relevante por muito tempo. Enquanto não for desenvolvida uma vacina eficaz, que previna a evolução da doença COVID-19 em sua forma mais letal, o único meio capaz de proteger os indivíduos é o isolamento social aliado a muita informação acerca da pandemia, conforme recomendações da WHO (2020) (3).

Os modelos matemáticos em epidemiologia trazem um conhecimento sólido acerca do contágio e permite fazer previsões de cenários futuros da situação epidemiológica em uma população impondo ações de prevenção, aumentando ou diminuindo o distanciamento social. Este tipo de trabalho torna-se perfeitamente justificado e, mais do que isso, torna-se impositivo diante da relevância do tema. Assim, verifica-se muitas publicações relacionadas à modelagem matemática para o estudo do avanço da pandemia do novo coronavírus, tais quais em Cruz (2021) (14), Neto (2021) (15), Kamrujjaman (2020) (16) e Mallapaty (2020) (17) de reconhecida relevância. Desta forma, apresentaremos um modelo matemático estatístico que, através do recurso de planilhas de cálculo, pode ser mais uma contribuição visando o mesmo objetivo.

1.2. Modelo matemático proposto

O modelo proposto neste trabalho é inspirado na teoria da cinética pontual de um reator nuclear. Esta teoria é descrita em diversos trabalhos na literatura, tais quais, Duderstadt (1987) (18), Hertrick (1971) (19), Akcasu (1971) (20), Nunes (2015) (21) e Palma (2016) (22). Esta teoria trata-se de um estudo na variação temporal da concentração de nêutrons no núcleo do reator nuclear, parâmetro fundamental e diretamente proporcional à potência gerada e as temperaturas internas do reator. A tabela 1 nos apresenta um quadro comparativo entre o modelo da cinética pontual de um reator nuclear e o modelo matemático para a pandemia do SARS-Cov-2. Esta tabela tem o objetivo de mostrar similaridades entre os modelos, justificando sua utilização.

Tabela 1: Quadro comparativo entre modelo da cinética pontual e modelo para pandemia da SARS-Cov-2.

Cinética Pontual da Física Nuclear	Modelo Matemático para o SARS-Cov 2
Para ocorrer a fissão nuclear o núcleo de urânio (U235) absorve um nêutron.	Para ocorrer a infecção por SARS-Cov-2, o indivíduo é contaminado pelo vírus.
Cada nêutron é muito menor que o núcleo do U235 na proporção de 1/235.	Cada vírus é muito menor que a célula do indivíduo infectado especificamente na proporção de 125 mm para 30.000 mm, ou seja, na proporção 1/240.
Ao absorver o nêutron, o núcleo de U235 fica instável e pode sofrer a fissão nuclear liberando 2 ou 3 nêutrons.	O indivíduo infectado contamina uma média entre 2 e 3 indivíduos.
Sem as medidas de controle do reator nuclear, como as barras de controle e a adição de reagentes absorvedores de nêutrons, o número de nêutrons cresce exponencialmente.	Sem as medidas de prevenção e isolamento social, uso de máscara e constante higienização das mãos o número de infectados cresce exponencialmente.

O modelo da cinética pontual trata da variação ao longo do tempo da concentração de nêutrons no núcleo de um reator nuclear (DUDERSTADT, 1987) (18) e (PALMA, 2016) (22). Neste modelo a localização do nêutron não é relevante, mas importa saber se o nêutron é gerado imediatamente após a fissão nuclear (são os chamados nêutrons prontos) ou gerado com um tempo relativamente longo (são os chamados nêutrons retardados) após a fissão por causa do decaimento radiativo dos fragmentos da fissão. Assim, as equações da cinética pontual, em um reator nuclear, têm como única variável o tempo e são apresentadas a seguir:

$$\frac{dn(t)}{dt} = \frac{[\rho(t) - \beta(t)]}{\Lambda} \cdot n(t) + \sum_{i=1}^6 \lambda_i \cdot C_i(t) \quad (\text{eq.5})$$

$$\frac{dC_i(t)}{dt} = \frac{\beta_i}{\Lambda} \cdot n(t) - \lambda_i \cdot C_i(t) \quad (\text{eq.6})$$

Onde $n(t)$ é a densidade de nêutrons prontos, ou seja, nêutrons gerados no ato da fissão nuclear e $C_i(t)$ são os nêutrons retardados, ou seja, nêutrons gerados a partir do decaimento radioativo dos produtos de fissão. Também são chamados de nêutrons precursores.

Os parâmetros nucleares são listados na tabela 2. Nesta tabela são apresentados parâmetros nucleares que permitem entender os fatores que influem no funcionamento de um reator nuclear e seu monitoramento. Estes fatores são a reatividade, a fração de nêutrons retardados e a constante de decaimento radioativo.

Tabela 2: Listagem dos parâmetros nucleares da cinética pontual e suas respectivas descrições.

Parâmetros nucleares	Descrição
$\rho(t)$	Reatividade = parâmetro relacionado ao fator multiplicativo da reação em cadeia, ou seja, o quanto a concentração de nêutrons aumenta, diminui ou se mantém constante.
$\beta(t)$	Fração de nêutrons retardados, considerando todos os grupos de precursores, com relação ao total de nêutrons.
$i(t)$	Fração de nêutrons retardados de cada grupo de precursores i com relação ao total de nêutrons.
λ_i	Constante de decaimento radioativo para cada um dos grupos de precursores.

1.3. Fundamentos do modelo matemático para a pandemia de SARS-Cov-2

Sabendo que o modelo matemático para a pandemia deve possuir algumas semelhanças com o modelo nuclear, porém, também deve possuir suas peculiaridades, é importante frisar algumas premissas que são necessárias para a elaboração do modelo:

- O grupo dos Suscetíveis é formado por toda a população do país que ainda não foi infectada.
- Todos os Infectados (Sintomáticos ou Assintomáticos) podem transmitir o vírus para indivíduos Suscetíveis.
- Todos os Infectados Sintomáticos são testados e compõem os dados oficiais divulgados pela OMS no país escolhido para validar o método.
- Nenhum dos Infectados Assintomáticos são testados e compõem os dados oficiais divulgados pela OMS no país escolhido para validar o método.
- A pessoa Recuperada do SARS-Cov-2 não pode mais infectar outros indivíduos.
- A pessoa Recuperada do SARS-Cov-2 não pode mais ser reinfectada no período de vigência deste estudo.
- Todos os Infectados Sintomáticos que morrem tem sua causa da morte determinada pela SARS-Cov-2.
- Todos os Infectados Assintomáticos que morrem tem sua causa da morte determinada por causas que não tem relação com a SARS-Cov-2. A taxa de mortalidade deste seguimento é mesma do restante da população que não está infectada.
- Os Recuperados são divididos em dois grupos, a saber: Recuperados dos Sintomáticos (que eram Infectados Sintomáticos) e Recuperados dos Assintomáticos (que eram Infectados Assintomáticos)
- Considera-se que a População Total do país é variável durante a pandemia.

A partir dos fundamentos do modelo, apresentamos na tabela 3 suas variáveis e parâmetros.

Tabela 3: Listagem de variáveis e parâmetros do modelo nuclear para a pandemia do SARS-Cov-2.

Variáveis do modelo	Coefficientes ou parâmetros do modelo
$S(t)$ = Suscetíveis	$\rho(t)$ = Parâmetro de infecção específico
$I_S(t)$ = Infectados Sintomáticos	$\alpha(t)$ = Taxa de natalidade do país
$I_A(t)$ = Infectados Assintomáticos	$\mu(t)$ = Taxa de mortalidade do país
$R_S(t)$ = Recuperados dos Sintomáticos	$\beta(t)$ = Fração de mortos por minuto de COVID-19
$R_A(t)$ = Recuperados dos Assintomáticos	$\gamma(t)$ = Taxa de imigração do país
População Total = $N(t)$	$S(t)$ = Fração de recuperados dos sintomáticos
-----	$A(t)$ = Fração de recuperados dos assintomáticos

As equações do modelo são:

$$\frac{dN(t)}{dt} = \alpha \cdot N(t) + \gamma \cdot N(t) - \mu \cdot [I_A(t) + R_A(t) + R_S(t) + S(t)] - \beta I_S(t) \quad (\text{eq.7})$$

$$\frac{dS(t)}{dt} = \alpha \cdot N(t) + \gamma \cdot N(t) - \rho(t) \frac{I_S(t)}{N(t)} S(t) - \rho(t) \frac{I_A(t)}{N(t)} S(t) - \mu S(t) \quad (\text{eq.8})$$

$$\frac{dI_S(t)}{dt} = \rho(t) \frac{I_S(t)}{N(t)} S(t) - \lambda_S R_S(t) - \beta I_S(t) \quad (\text{eq.9})$$

$$\frac{dI_A(t)}{dt} = \rho(t) \frac{I_A(t)}{N(t)} S(t) - \lambda_A R_A(t) - \mu I_A(t) \quad (\text{eq.10})$$

$$\frac{dR_S(t)}{dt} = \lambda_S R_S(t) - \mu R_S(t) \quad (\text{eq.11})$$

$$\frac{dR_A(t)}{dt} = \lambda_A R_S(t) - \mu R_A(t) \quad (\text{eq.12})$$

$$N(t) = S(t) + I_S(t) + I_A(t) + R_A(t) + R_S(t) \quad (\text{eq.13})$$

2. METODOLOGIA

Sabemos que nenhum país terá os seus dados acerca da pandemia do SARS-Cov-2 sem nenhum desvio ou imprecisão. Apesar disso, para validar o modelo é necessária a escolha de uma região ou país com dados relativamente confiáveis para realizar estes testes. Além disso, também vale salientar que, existe uma subnotificação natural decorrente do fato de que um grande percentual dos infectados desenvolve a doença de forma mais branda, quase sem sintomas, ou mesmo com total ausência de sintomas. São os chamados pacientes assintomáticos. Estes raramente vão descobrir que contraíram o vírus, porque sem sintomas não farão o teste e ficarão fora da estatística. Mesmo assim, estes indivíduos podem ser agentes transmissores do vírus (MALLAPATY, 2020) (17).

A coleta de grande quantidade de dados provenientes de fontes confiáveis é um fator importante para o sucesso do modelo matemático desenvolvido neste projeto. Torna-se fundamental ampliar a quantidade de países investigados de modo que fatores políticos, como os expostos anteriormente, possam ser considerados desprezíveis no resultado final do modelo.

Dois parâmetros são fundamentais nesta análise do nível de confiabilidade dos dados oficiais de cada país. O primeiro é o número de testes realizados para cada cem mil habitantes daquele país. Quanto menor for este número, menos confiáveis serão os dados oficiais. O segundo parâmetro é a taxa de letalidade da COVID-19 naquele país (DONSIMONI, 2020) (23)

A princípio, a taxa de letalidade deve ser a mesma em países que tenham similaridade com relação a dados acerca da distribuição demográfica dos grupos de risco e de situações ambientais que possam ser relevantes. Confirmando esta hipótese, e havendo a similaridade das populações, espera-se que a taxa de letalidade seja similar. Caso um dos países tenha uma taxa de letalidade alta em comparação com o outro similar isso pode indicar que o número de infectados é maior do que os dados oficiais indicam (CRUZ, 2021) (14).

Vale ressaltar que, a sobrecarga e a qualidade do sistema de saúde, de qualquer país, também irão influenciar na sua taxa de letalidade, seja por conta da COVID-19, ou por qualquer outra doença. Tudo isso irá determinar uma metodologia científica extremamente criteriosa para a escolha do país, ou países, que vão ser utilizados para validar o modelo matemático que será desenvolvido.

Discrepâncias entre os resultados do modelo matemático e os dados oficiais confiáveis vão determinar importantes ajustes no próprio modelo. Desta forma, a Alemanha foi o país escolhido para a validação do modelo nuclear para a pandemia do SARS-Cov-2. Os critérios para esta escolha estão na tabela 4. O estudo acerca da pandemia na Alemanha foi desenvolvido por Donsimoni (2020) (23) e Barbarossa (2020) (24).

Tabela 4: Critérios que resultaram na escolha da Alemanha como objeto de validação do modelo.

Critérios de Escolha do país para a validação do modelo
1) Grande número de testes de COVID-19 realizados por milhão de habitantes.
2) Baixa taxa de mortalidade por COVID-19.
3) Bom percentual de recuperados.
4) Longo período de evolução da pandemia o país, chegando perto de uma normalidade.
5) Elevada densidade demográfica.
6) Grande número de infectados.
7) Taxa de ocupação dos leitos de UTI inferior a 95% durante todo o período.

Na tabela 5 temos uma pequena amostra dos dados coletados para a Alemanha: (levando-se em consideração que a tabela completa tem, aproximadamente, 100 linhas). Estes dados completos acerca da pandemia na Alemanha foram coletados na WHO (2020) (25), John Hopkins (2020) (26) e CountryMeter (2020) (27).

Tabela 5: Amostragem dos dados oficiais para Alemanha.

Data (2020)	Dia	Casos Confirmados	Infectados	Mortos	Recuperados	Taxa de Mortalidade
10/03	50	1112	1095	2	15	0,180%
20/03	60	8198	8001	20	177	0,244%
21/03	61	14138	13887	45	206	0,318%
05/04	76	85778	55739	1342	28697	1,565%
07/04	78	95391	57706	1607	36078	1,685%
15/04	86	127584	51733	3254	72597	2,550%
16/04	87	130450	49884	3569	76997	2,736%
28/04	99	156337	33027	5913	117397	3,782%
29/04	100	157641	31129	6115	120397	3,879%

A partir das equações do modelo nuclear para a pandemia do SARS-Cov-2 e do conjunto completo dos dados coletados, foi possível obter a seguinte função para os Infectados Sintomáticos:

$$I_S(t) = A \cdot e^{at^2+bt+c} + B \cdot \text{sen}(w_1 t) + C \quad (\text{eq.14})$$

Onde:

$$A = 0,8, \quad B = 6, \quad C = 4, \\ a = -8,87058 \times 10^{-6}, \quad b = 0,03504, \quad c = -23,37399 \quad \& \quad w_1 = 0,8.$$

Para obter a equação acima utilizamos várias aproximações e técnicas básicas de resolução de equações diferenciais, pois verificou-se que esta função não possui solução analítica. Note que a equação acima é uma solução particular para a equação (eq.14), da derivada de $I_S(t)$, levando-se em conta as condições iniciais e de contorno estabelecidas pelos resultados da tabela 5.

Estas mesmas condições permitem a obtenção de outros parâmetros e variáveis. Assim, todas as variáveis são obtidas em conformidade com o modelo. Vale ressaltar que todo modelo matemático, seja ele aplicado em qualquer área do conhecimento, pressupõe o uso de hipóteses simplificadoras, a fim de que seja possível identificar as variáveis e parâmetros mais relevantes para que aquele problema tenha uma solução. Esta solução é chamada de analítica quando encontramos uma função para cada variável que satisfaz o sistema de equações diferenciais do modelo.

Na impossibilidade ou dificuldade em se obter uma solução analítica, pode-se obter uma solução numérica utilizando métodos computacionais. A solução numérica obtém valores que permitem a construção de tabelas e gráficos que descrevem plenamente a evolução das variáveis, mas sem que tenhamos, necessariamente, uma função analítica correspondente.

O método numérico adotado para determinar os coeficientes do modelo foi o método de diferenças finitas. Foi feito um teste com a evolução dos dados por minuto e por hora. Como diferença entre ambos os resultados foi inferior a 5%, por questão de celeridade, resolvemos considerar o intervalo de tempo entre cada iteração (Δt) igual a uma hora. Esta hipótese é razoável porque nosso universo amostral é de aproximadamente 100 dias.

Portanto, o método de diferenças finitas foi usado para calcular o parâmetro de infecção específico, a fração de mortos por minuto de COVID-19, a fração de recuperados dos sintomáticos, fração de recuperados dos assintomáticos, recuperados dos sintomáticos e recuperados dos assintomáticos.

O cálculo da taxa de mortalidade, taxa de crescimento populacional, taxa de imigração e taxa de natalidade, todos da Alemanha, foram obtidos diretamente dos dados demográficos da Alemanha, disponíveis em tempo real no site da CountryMatters (2020) (27).

Como neste trabalho a quantidade de dados coletados foi imensa, bem como a produção de dados por interpolação, torna-se necessária uma metodologia específica para o tratamento destes dados. Segundo Trafimow (2016) (28) o mais usual é um pesquisador calcular as médias do grupo e usar o procedimento de teste de significância da hipótese nula para tirar conclusões sobre as populações das quais os grupos foram retirados. Em nosso trabalho é possível garantir que os grupos constituintes da população do país sob estudo são dinâmicos e seu estado é variável com o tempo. Isto permite que metodologias sejam utilizadas para eliminar ou reduzir as fontes de incertezas no processo (SIMONSOHN, 2020) (29).

Vergura (2009) (30) descreve uma metodologia de monitoramento de uma usina fotovoltaica na Itália por meio de inferência estatística. A quantidade de variáveis reais envolvidas neste processo se assemelha à evolução da pandemia do novo coronavírus, o que permitiria uma análise estatística real em nosso trabalho. Por conta das próprias dificuldades inerentes à situação atual da pandemia e da necessidade básica de isolamento social este processo de análise estatística foi bem reduzido, conforme será visto na próxima seção.

3. RESULTADOS E DISCUSSÃO

A partir do exposto na seção anterior e dos dados completos acerca da pandemia de SARS-Cov-2 COVID-19 na Alemanha temos os gráficos apresentados nas Figuras (1) até (6).

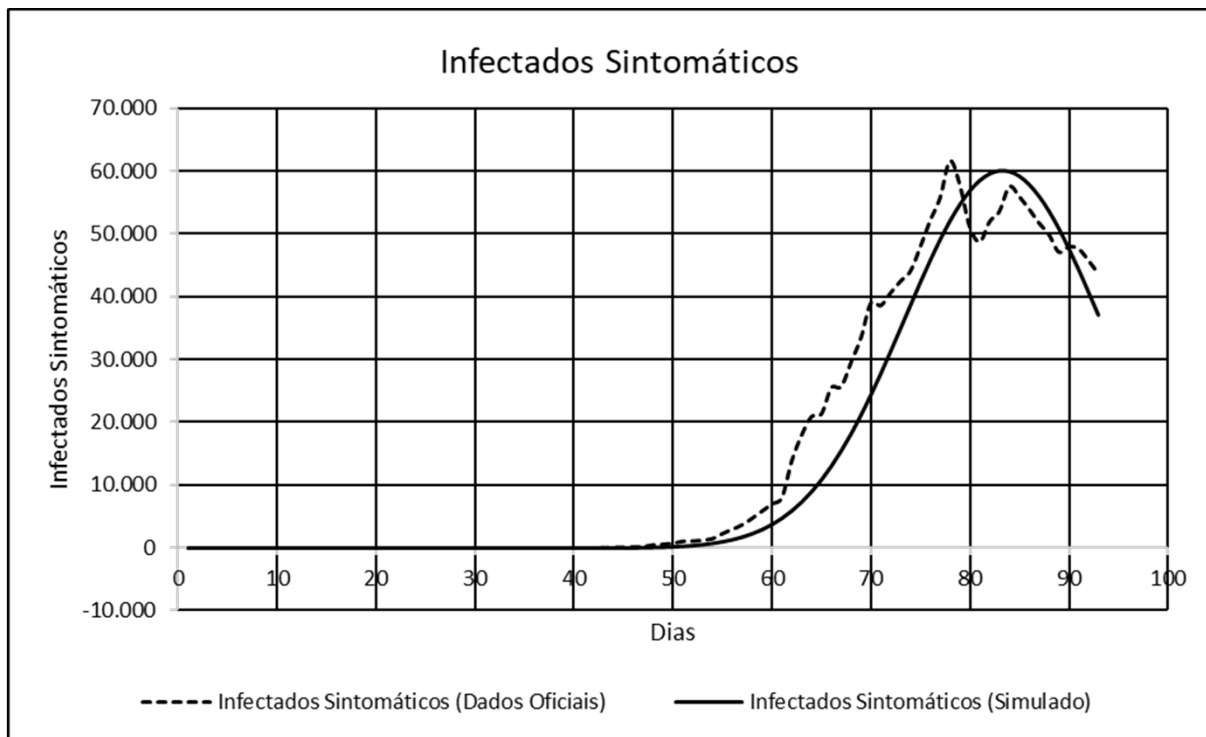


Figura 1. Infectados Sintomáticos. Dados Oficiais em relação aos resultados obtidos na equação (eq. 14).



Figura 2. Infectados Assintomáticos obtido através do modelo.

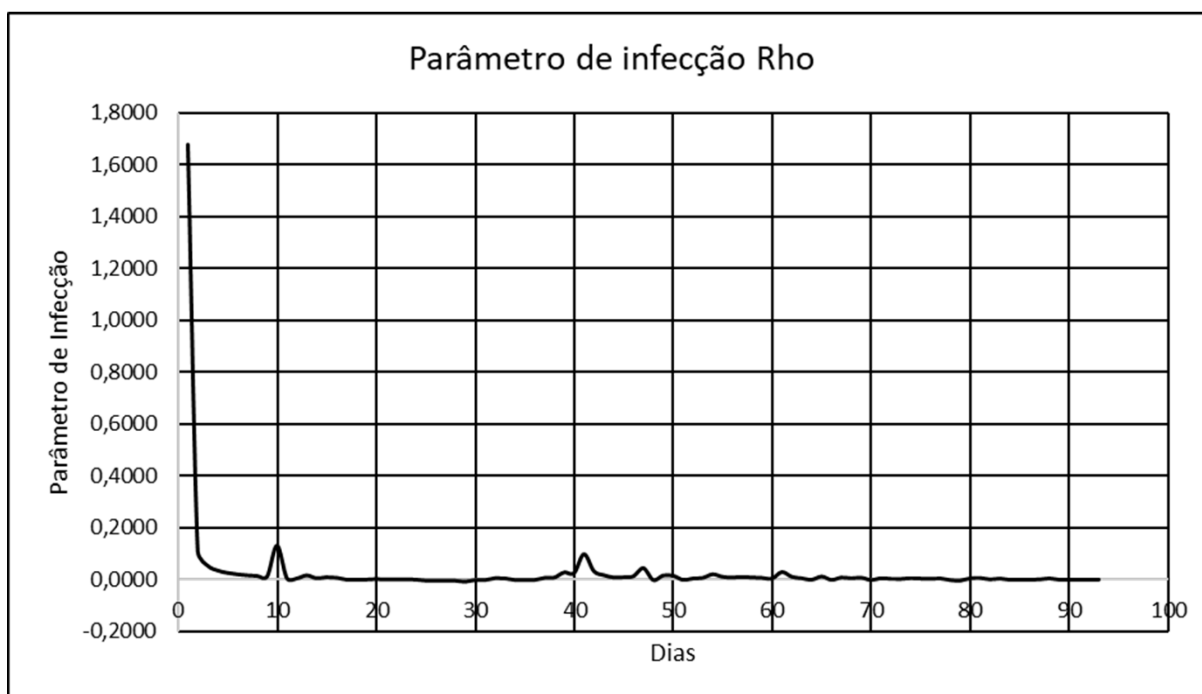


Figura 3. Parâmetro de infecção Rho obtido através do modelo.



Figura 4. Recuperados Sintomáticos e Assintomáticos, obtidos, respectivamente, por dados oficiais e equações do modelo.

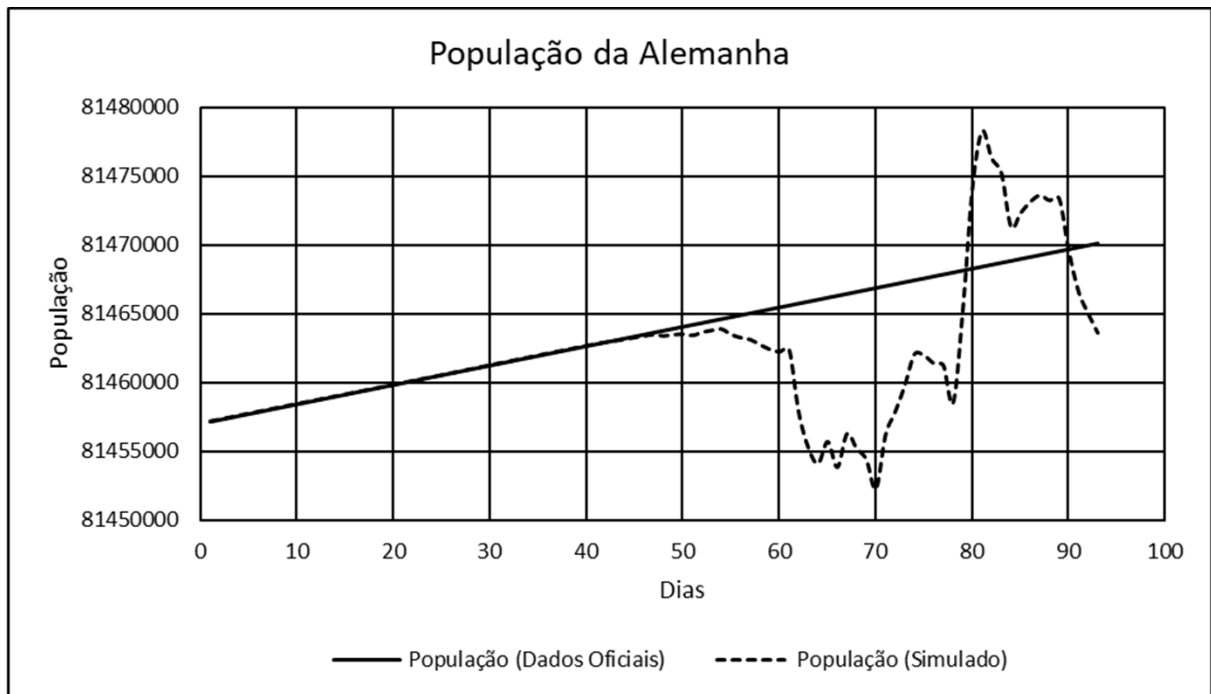


Figura 5. População da Alemanha, em comparação, os dados obtidos pelo modelo e dados oficiais.

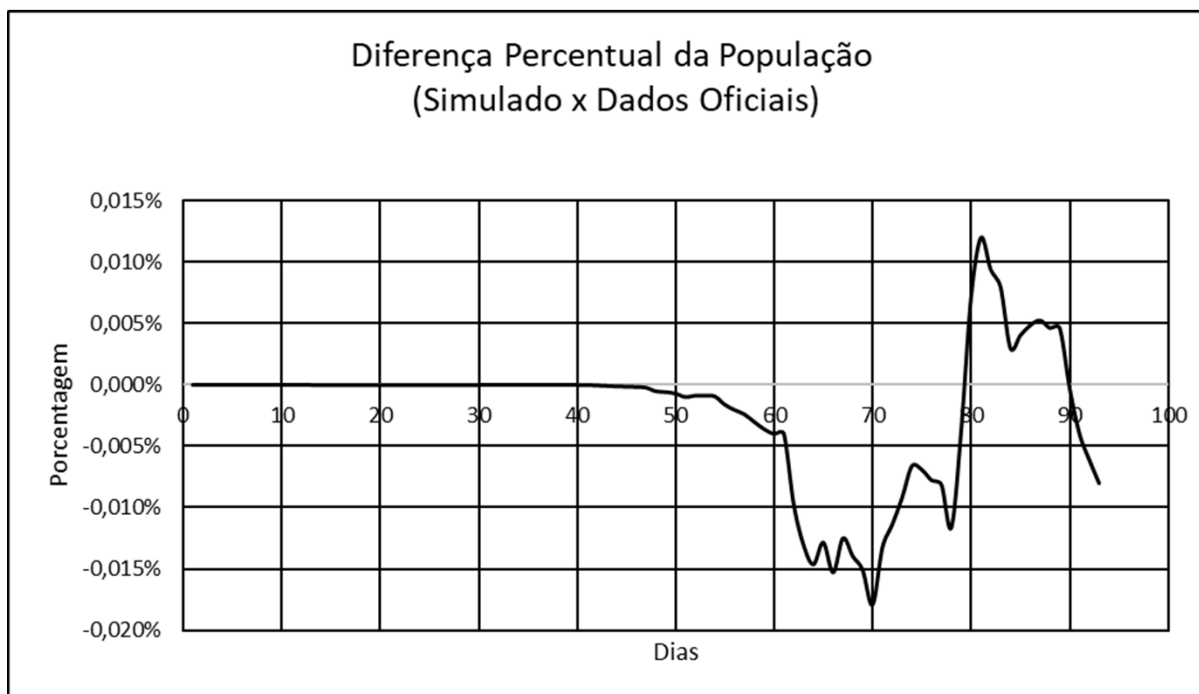


Figura 6. Diferença percentual entre os dados dos gráficos da figura 5.

A seguir temos um recorte das tabelas com uma amostra dos respectivos resultados. As tabelas originais, produzidas em planilhas de cálculo, não puderam ser colocadas na íntegra neste trabalho devido a imensa quantidade de dados mensurados, com aproximadamente 2 400 linhas para o cálculo por hora e 144 000 linhas para o cálculo por minuto.

Vale ressaltar que foi feito um processo de interpolação dos dados a fim de que fosse feita uma variação gradual dos dados, variando de hora em hora, até completarmos um dia inteiro. Os dados oficiais foram divulgados diariamente, mas fizemos sua interpolação para escrevermos o incremento de cada hora. Este procedimento foi feito para todas as variáveis do modelo matemático, infectados, recuperados, população etc. Assim, conseguimos ampliar a quantidade de dados, partindo da hipótese de um crescimento linear dentro daquele dia, considerando que em cada hora houve um mesmo crescimento. Isso também vale quando ocorre uma diminuição do parâmetro, por exemplo quando o número de infectados diminui de um dia para o outro.

Também fizemos este incremento por minuto, mas a discrepância com relação ao resultado obtido por hora foi quase nula, o que determinou o incremento por hora. Em um único dia temos $24 \times 60 = 1440$ minutos. Como a análise foi feita em 90 dias, segue que teremos 129 600 dados para cada parâmetro. Isso torna a operação do aplicativo muito mais trabalhosa, para praticamente nenhum resultado superior a interpolação feita por hora. Nesta interpolação feita por hora teremos ao longo dos 90 dias de observação um total de 2160 dados por variável do modelo, o que demonstrou ser muito mais fácil de operar.

Tabela 6: Amostragem dos dados obtidos através do modelo.

Data (2020)	Dia	Infectados Assintomáticos $I_A(t)$	Parâmetro de Infecção $\rho(t)$	Recuperados dos Sintomáticos $R_S(t)$	Recuperados dos Assintomáticos $R_A(t)$
10/03	50	293	0,0008	21	9
20/03	60	4754	0,0310	177	71
21/03	61	5937	0,0125	206	103
05/04	76	49092	0,0062	28697	7403
07/04	78	54701	-0,0020	36078	10965
15/04	86	56028	0,0023	72597	35130
16/04	87	53669	0,0065	76997	38811

Tabela 7: Comparativo entre os dados oficiais e o modelo durante e além da amostra coletada.

Data (2020)	Dia	Infectados Sintomáticos a partir dos dados oficiais	Infectados Sintomáticos a partir do modelo	Discrepância Percentual
20/03	60	8001	4754	40,3%
05/04	76	55739	49092	11,9%
07/04	78	57706	54701	5,2%
20/05	121	13361	7335	45,0%
10/06	142	6966	6890	1,1%
30/07	192	8432	5989	29,0%

Pelo gráfico apresentado na Figura 1, podemos observar que o modelo matemático implementado é compatível com os dados coletados no período compreendido entre 21 de janeiro de 2020 até 29 de abril de 2020. Este fato de maneira isolada não é suficiente para garantir que o modelo nuclear proposto para a pandemia do SARS-Cov-2 seja válido no período posterior ao da amostra de dados. Entretanto, esta verificação é feita a partir da análise estatística apresentada na tabela 7.

Nota-se uma discrepância percentual que varia de 1% até 45% no período posterior a amostragem dos dados mostra que se faz necessário refinar o processo e melhorar ainda mais a qualidade do modelo proposto. A flutuação da discrepância percentual é bastante similar com o período da coleta da amostra de dados, ou seja, o período compreendido entre janeiro e abril.

Uma componente oscilatória da curva de evolução periódica das variáveis da pandemia no modelo matemático, quando comparada com os dados reais ressalta o fato de que o período de oscilação não é regular para os dados reais. Assim, os ciclos de variação dos parâmetros da pandemia não são iguais, resultando em uma discrepância percentual de até 45% no máximo. Uma forma de reduzir esta discrepância é utilizar os dados para prever um período de poucos dias para frente. De fato, é muito difícil fazer uma previsão para 90 dias a frente, porque são muitas variáveis envolvidas no modelo matemático.

Baseado nos resultados obtidos, é possível apontar que os dados coletados no início da pandemia em uma determinada localidade são suficientes para prever o seu comportamento até cerca de três meses após o término da data de coleta da amostra inicial de dados, com discrepâncias percentuais diminuindo a medida que a quantidade de dias a ser prevista for menor, ou seja, se usarmos o modelo para prever os próximos 10 dias os resultados serão bem melhores do que se fossemos fazer uma previsão para os próximos 90 dias. Espera-se, na situação atual de desenvolvimento do modelo nuclear para a pandemia do SARS-Cov-2 uma discrepância percentual de no máximo 45%. Com o objetivo de fazer uma comparação, em artigo publicado por Donsimoni (2020) (23) foi encontrado uma discrepância percentual de até 50% utilizando um modelo de Markov de tempo contínuo.

Este trabalho apresenta os seus resultados muito similares ao artigo de Bärwolff (2020) (31). O comportamento dos gráficos das figuras 1 até 4 encontram uma similaridade com aqueles apresentados no que concerne a evolução da pandemia do novo coronavírus na Alemanha. Note que isso implica numa maior confiabilidade dos dados obtidos a partir do modelo nuclear para a epidemiologia que aqui foi apresentado.

Nos trabalhos de Donsimoni (2020) (23) e Barbarossa (2020) (24) também encontramos uma razoável similaridade com os resultados, embora os métodos matemáticos adotados por eles sejam bem diversos daquele desenvolvido aqui neste trabalho. A capacidade de previsibilidade do modelo nuclear para a pandemia do novo coronavírus é confirmada a partir, também, de uma comparação com estes artigos que apresentam resultados para a Alemanha.

Vale ressaltar que este trabalho inicial se resume apenas à validação do método. Apesar disso, o objetivo para um trabalho futuro consiste em aplicar este modelo matemático em uma situação no Brasil, seja no país em geral ou em algum estado ou município.

CONCLUSÃO

A partir da discussão do trabalho e dos resultados obtidos, levando-se em conta a situação epidemiológica da Alemanha, o modelo nuclear para a pandemia do SARS-Cov-2 foi considerado válido. Apesar disso, a precisão do modelo ainda pode ser aperfeiçoada, uma vez que este trabalho é parte de um projeto de pesquisa institucional no IFRJ e se encontra em suas fases iniciais.

Uma forma de verificar a eficiência do método aqui proposto consiste na análise feita e apresentada no gráfico da figura 5. Ali pode-se observar o cálculo da população a partir dos dados demográficos em comparação com o cálculo a partir do nosso modelo, onde usamos infectados sintomáticos, infectados assintomáticos, recuperados sintomáticos, recuperados assintomáticos e suscetíveis para obter de dentro do modelo a população. Assim, notem que a discrepância deste resultado da população é mínima, o que de forma definitiva comprova a consistência do modelo.

Desta forma, podemos considerar que os resultados para validação do modelo aqui apresentados são preliminares. Apesar disso, o modelo matemático apresentado é muito promissor. Este modelo pode e irá ser revisitado com o objetivo de ser aplicado futuramente para prever a evolução da pandemia no Brasil, regionalmente e nacionalmente.

LISTA DE ABBREVIATURAS

COVID-19 = Corona Vírus Disease (2019).

EUA = Estados Unidos da América.

IFRJ = Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro

OMS = Organização Mundial de Saúde.

Sars-Cov-2 = Vírus da Síndrome Respiratória Aguda Grave (Novo Coronavírus).

SIR = Suscetíveis, Infectados e Removidos (modelo matemático em epidemiologia).

U235 = Urânio 235.

USA = United States of America

WHO = World Health Organization.

CONFLICT DE INTERESSE

Este trabalho é fruto do projeto de pesquisa “**Construção e aperfeiçoamento de Modelos Matemáticos descritivos da Pandemia do SARS-Cov-2**”, que foi aprovado no Edital Integrado de Ensino, Pesquisa, Inovação e Extensão nº 01 e 02/2020 do Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro (IFRJ). Este projeto de pesquisa foi contemplado com uma bolsa de iniciação científica, têm anuência da direção geral do *Campus Duque de Caxias* do IFRJ e têm vigência de agosto de 2020 à julho de 2021.

AGRADECIMENTOS

Agradecemos à Equipe de Física - (Física IFRJ CDuC) pela contribuição com ideias e pela utilização do Laboratório de Física Nilton de Souza Medeiros como espaço de orientação, trabalho e reuniões de pesquisa. Gostaríamos de agradecer também à Coordenação de Pesquisa e Inovação (CoPI), ao Programa de Bolsas de Iniciação Científica - PIBIC/IFRJ e à Pró-Reitora de Pós-graduação, Pesquisa e Inovação - PROPPI, pela aprovação do projeto e concessão de bolsa de iniciação científica. Um grande agradecimento ao Instituto Federal do Rio de Janeiro - *Campus Duque de Caxias* (IFRJ-CDuC), pela oportunidade de atuar como professores e pesquisadores nesta instituição. Finalmente, gostaríamos de manifestar um agradecimento especial à toda equipe organizadora da 1ª Conferência Brasileira de Planejamento Experimental e Análise de Dados (ConBraPa), pela oportunidade de apresentar os resultados iniciais de nosso projeto de pesquisa, num evento de tamanha relevância.

Créditos dos autores

Rafael Pereira Santana é coordenador e orientador do projeto de pesquisa, responsável pela escrita do artigo e pela revisão do texto, tabelas e gráficos. **Anderson Lupo Nunes** é colaborador e co-orientador do projeto, responsável pela escrita do artigo, interpolação dos dados e ajuste das curvas. **Pedro Maia Solomone** é bolsista de iniciação científica do projeto e responsável pela interpolação dos dados, elaboração de tabelas e ajuste de curvas.

REFERENCIAS

1. CHAVES, T. S. S.; BELLEI, N. SARS-COV-2: The new coronavirus: a reflection about “One Health” and the importance of travel medicine when new pathogens emerge. **Revista de Medicina**. v. 1, p. 99, i-iv. 2020 DOI: dx.doi.org/10.11606/issn.1679-9836.v99i1pi-iv.
2. GOULART, A. D. C. Revisiting the Spanish flu: the 1918 influenza pandemic in Rio de Janeiro. **História, Ciência e saúde – Manguinhos**. v. 12, n. 1, p. 101-142. 2005. Available from: www.nescon.medicina.ufmg.br/biblioteca/imagem/0915.pdf.
3. WHO. COVID-19 Weekly Epidemiological Update in 29 Dez. 2020. **Worlds Health Organization**. 2020. Available from: www.who.int/publications/m/item/weekly-epidemiological-update---29-december-2020.
4. LUIZ, M. H. R. **Modelo-s Matemáticos em Epidemiologia**. UNESP, 2012. Available from: repositorio.unesp.br/bitstream/handle/11449/94348/luiz_mhr_me_rcla.pdf;jsessionid=3A9005292828646A6A77FB5BF8C73F43?sequence=1.
5. COSTA, S. D. S. Pandemia e desemprego no Brasil. **Revista de Administração Pública**. v. 54, n. 4, p. 969-978, 2020. DOI: dx.doi.org/10.1590/0034-761220200170.
6. MACEDO, L. D. D.; MACEDO, J. R. D. D. A pandemia de Covid-19: aspectos do seu impacto na sociedade globalizada do século XXI. **Caderno de Ciências Sociais Aplicadas**. v. 17, n. 30, p. 40-43. 2020. DOI: doi.org/10.22481/ccsa.v17i30.7315.

7. CRISTÓVÃO, R. B. **Modelo SIR: Uma Aplicação à Hepatite A**. USP, 2015. Available from: www.ime.usp.br/~map/tcc/2015/Rafael%20Belmiro.pdf.
8. OKHUESE, V. A. Mathematical Predictions for COVID-19 as a Global Pandemic. **MedRXiv: The Preprint Server for Health Sciences**. p. 01-16. 2020. DOI: doi.org/10.1101/2020.03.19.20038794.
9. COSTA, G. S.; COTA, W.; FERREIRA, S. C. Metapopulation modeling of COVID-19 advancing into the countryside: an analysis of mitigation strategies for Brazil. **MedRXiv: The Preprint Server for Health Sciences**. p. 01-13. 2020. DOI: doi.org/10.1101/2020.05.06.20093492.
10. CIUFOLINI, I.; PAOLOZZI, A. Mathematical prediction of the time evolution of the COVID-19 pandemic in Italy by a Gauss error function and Monte Carlo simulations. **The European Physical Journal Plus**. v. 135, n. 355, p. 1-8. 2020. DOI: doi.org/10.1140/epjp/s13360-020-00383-y.
11. FOKAS, A. S.; DIKAIOS, N.; KASTIS, G. A. Mathematical models and deep learning for predicting the number of individuals reported to be infected with SARS-CoV-2. **Journal of The Royal Society Interface**. v. 17, p. 01-12. 2020. DOI: doi.org/10.1098/rsif.2020.0494.
12. KERMACK, W. O.; MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. **Proceedings of the Royal Society A Mathematical, Physical and Engineering Sciences**. p. 700-721, 1927. DOI: doi.org/10.1098/rspa.1927.0118.
13. OLIVEIRA, M. H. D. Análise do modelo SIR: Comportamento da curva de infectados em relação à inclusão de novas semanas epidemiológicas. **Tabalho de Conclusão de Curso - Instituto de Matemática e Estatística - USP**. 2018.
14. CRUZ, P. A. D.; CREMA-CRUZ, L. C.; CAMPOS, F. S. Modeling transmission dynamics of severe acute respiratory syndrome coronavirus 2 in São Paulo, Brazil. **Revista da Sociedade Brasileira de Medicina Tropical**. v. 54, p. 1-8. 2021. DOI: doi.org/10.1590/0037-8682-0553-2020.
15. NETO, O. P. et al. Mathematical model of COVID-19 intervention scenarios for São Paulo - Brazil. **Nature Communications**. v. 12, p. 418. 2021. DOI: doi.org/10.1038/s41467-020-20687-y.
16. KAMRUJJAMAN, M.; JUBYREA, J.; ISLAM, M. S. Data analysis and mathematical model: control measures and prediction to prevent COVID-19 outbreak. **Arabian Journal of Medical Sciences**. v. 3, n. 2, p. 5-9. 2020. Available from: www.researchgate.net/publication/341448410.
17. MALLAPATY, S. The Mathematical Strategy that could transform coronavirus testing. **Nature**. v. 583, p. 504. 2020. Available from: www.nature.com/articles/d41586-020-02053-6.
18. DUDERSTADT, J. J. **Nuclear Reactor Analysis**. New York: Wiley & Sons. 1987. Available from: documents.pub/document/nuclear-reactor-analysis-56075ae4de6bc.html.
19. HETRICK, D. L. **Dynamics of nuclear reactors**. [S.I.]: The University of Chicago Press Ltda, 1971. Available from: https://openlibrary.org/books/OL1425869M/Dynamics_of_nuclear_reactors.
20. AKCASU, Z.; LELLOUCHE, G. S.; SHOTKIN, L. M. **Mathematical Methods in Nuclear Reactor Dynamics**. [S.I.]: Academic Press, 1971. Available from: www.sciencedirect.com/book/9780120471508/mathematical-methods-in-nuclear-reactor-dynamics.
21. NUNES, A. L. et al. A New Formulation to the Point Kinetics Equations Considering the Time Variation of the Neutron Currents. **World Journal of Nuclear Science and Technology**. v. 5, p. 57-71. 2015. DOI: [dx.doi.org/10.4236/wjnst.2015.51006](https://doi.org/10.4236/wjnst.2015.51006).

22. PALMA, D. A. P.; LUPONUNES, A.; MARTINEZ, A. S. Effect of the time variation of the neutron current density in the calculation of the reactivity. **Annals of Nuclear Energy**. v. 96, p. 204-211. 2016. DOI: doi.org/10.1016/j.anucene.2016.05.022.
23. DONSIMONI, J. R. et al. Projecting the spread of COVID-19 for Germany. **De Gruyter**. v. 21, n. 2, p. 181-216. 2020. DOI: doi.org/10.1515/ger-2020-0031.
24. BARBAROSSA, M. V. et al. Modeling the spread of COVID-19 in Germany: Early assessment and possible scenarios. **Journal Plos One**. v. 15, n. 9, p. 01-22. 2020. Doi: doi.org/10.1371/journal.pone.0238559.
25. WHO. Coronavirus Disease (COVID-19) Situation Report in 21 Set. 2020. **Worlds Health Organization**, 2020. Available from: www.who.int/emergencies/diseases/novel-coronavirus-2019.
26. JOHNS HOPKINS UNIVERSITY. Mortality in the most affected countries. **MORTALITY ANALYSES**. Johns Hopkins, 2020. Available from: coronavirus.jhu.edu/data/mortality.
27. COUNTRYMETERS. **Relógio de Mortalidade da População Mundial**. 2020. Available from: countrymeters.info/pt.
28. TRAFIMOW, D.; MACDONALD, J. A. Performing Inferential Statistics Prior to Data Collection. **Educational and Psychological Measurement**. v. 77, n. 2. 2016. DOI: <https://doi.org/10.1177%2F0013164416659745>.
29. SIMONSOHN, U.; SIMMONS, J. P.; NELSON, L. D. SUPPLEMENTARY MATERIALS FOR: Specification Curve: Descriptive and Inferential Statistics On All Reasonable Specifications. **ResearchBOX: Open Research Made Easy**. p. 01-19. 2019. Available from: urisohn.com/sohn_files/wp/wordpress/wp-content/uploads/Supplement-Specification-Curve-2019-10-29.pdf.
30. VERGURA, S. et al. Descriptive and Inferential Statistics for Supervising and Monitoring the Operation of PV Plants. **IEEE Transactions on Industrial Electronics**. v. 56, n. 11, p. 4456-4464. 2009. DOI: doi.org/10.1109/TIE.2008.927404.
31. BÄRWOLFF, G. Mathematical Modeling and Simulation of the COVID-19 Pandemic. **Life in the Time of a Pandemic: Social, Economic, Health and Environmental Impacts of COVID-19 - Systems Approach Study**. v. 8, n. 24, p. 01-12. 2020. DOI: doi.org/10.3390/systems8030024.