

Fernando José Pessoa de Andrade

**Desenvolvimento do Processo de Difusão Gama
Modificado e Aplicações**

Recife

21 de março de 2023

*Este trabalho é dedicado às minhas avós queridas,
Nina e Menininha, que foram ao encontro do Senhor
enquanto essa dissertação era escrita.*

Agradecimentos

Em primeiro lugar agradeço a Deus pelo dom da vida e pela sua permissão em chegar até aqui. Agradeço à minha esposa, Jéssica, que me incentivou desde o primeiro momento e não me deixou abater por desânimo algum. Aos meus pais, cujo estímulo aos estudos sempre foi a prioridade máxima. À minha irmã, que compartilhou comigo a jornada da pós-graduação e que por isso compreendeu com empatia cada dificuldade.

Agradeço à Universidade Federal Rural de Pernambuco - UFRPE, instituição pela qual tenho profunda gratidão e respeito, e que foi minha casa também durante a graduação. Agradeço ao Departamento de Informática e Estatística - DEINFO, desde o seu corpo docente pela acolhida e disponibilidade até os funcionários e servidores pela prestabilidade.

Um agradecimento especial ao professor Dr. Cláudio Tadeu Cristino, meu orientador, por toda sua paciência ao longo do mestrado. Sem sua ajuda certamente esse trabalho não teria sido concluído. Obrigado por toda disponibilidade.

Agradeço à Gerência de Epidemiologia de Campo - GECAMP da Secretaria de Saúde do Distrito Federal, nas pessoas das Sr.as Priscilleyne Reis e Carolina Musso, pela disponibilização de Boletins Epidemiológicos tão bem elaborados, que foram importantes fontes de consulta para a elaboração dessa pesquisa. Agradeço, inclusive, ao Sr. Cássio Peterka, quem viabilizou o contato com a GECAMP.

Resumo

A pandemia de covid-19 é a maior emergência sanitária desde a gripe espanhola, no início do século passado. A intensidade com a qual essas e outras doenças se espalham na população é determinada pelo Número Reprodutivo Efetivo (R_e). O objetivo dessa pesquisa é verificar a adequação do modelo Gama Condicional Modificado para descrever a dinâmica do R_e da covid-19. Para tanto, mostra-se que a evolução temporal do R_e pode ser matematicamente caracterizada como um processo estocástico difusivo. A base de dados utilizada diz respeito aos casos de covid-19 no Brasil entre os anos de 2020 e 2022, obtidos na plataforma online brasileira OpenDataSUS, sendo a “Data dos Primeiros Sintomas” o dado primário para todas as análises dessa pesquisa. Os dados são tratados e analisados de acordo com parâmetros indicados pela Organização Pan-Americana de Saúde (OPAS). Ao aplicar o modelo aos dados, verifica-se o comportamento dos parâmetros ξ e λ para diferentes cenários e períodos. Em seguida, é testada a capacidade preditiva do modelo a partir de gráficos de envelope. A indicação de uma boa capacidade preditiva do modelo em análise é o ponto central da pesquisa, uma vez que fornece informações úteis aos responsáveis pela tomada de decisão. Por fim, são indicadas algumas melhorias no modelo Gama Condicional Modificado e as possibilidades de extensão da pesquisa em parceria com pesquisadores da área da saúde.

Palavras-chaves: Processos de Difusão. Covid-19. Número Reprodutivo Efetivo. Capacidade preditiva.

Abstract

The covid-19 pandemic is the biggest health emergency since the Spanish flu at the beginning of the last century. The extent to which these and other diseases spread in the population is determined by the Effective Reproductive Number (R_e). The objective of this research is to verify the suitability of the Modified Conditional Gamma model to describe the R_e dynamics of covid-19. Therefore, it is shown that the temporal evolution of R_e can be mathematically characterized as a diffusive stochastic process. The database used concerns cases of covid-19 in Brazil between the years 2020 and 2022, obtained from the Brazilian online platform OpenDataSUS, with the “Date of First Symptoms” being the primary data for all analyzes in this research. Data are treated and analyzed according to parameters indicated by the Pan American Health Organization (PAHO). When applying the model to the data, the behavior of the parameters ξ and λ is verified for different scenarios and periods. Then, the predictive ability of the model is tested using envelope plots. The indication of a good predictive capacity of the model under analysis is the central point of the research, since it provides useful information to those responsible for decision-making. Finally, some improvements are indicated in the Modified Conditional Gamma model and possibilities for research extension in partnership with health researchers.

Key-words: Diffusion processes. Covid-19. Effective Reproductive Number. Predictive ability.

Lista de Figuras

- Figura 1 – Resultado para três cenários de 15 lançamentos de um dado, representados pelas cores vermelho, azul e verde. Cada cenário s_i , $i = \{1, 2, 3\}$, é um elemento distinto do espaço de estados Ω . Fixando $s_1 \in \Omega$, o processo estocástico dos lançamentos torna-se a série temporal destacada em vermelho. Fixando a observação $t = 7$, destacada na linha vertical preta, o processo estocástico é a variável aleatória $X(7, s)$ cujo espaço amostral é o conjunto $\Phi = \{1, 2, 3, 4, 5, 6\}$, idêntico ao espaço amostral de apenas um lançamento do dado. 4
- Figura 2 – Incidência diária da covid-19 no Brasil. As retas vermelhas destacam, respectivamente, a data em que o Ministério da Saúde declarou Emergência em Saúde Pública de Importância Nacional (03/02/2020), a data do primeiro caso confirmado no Brasil (25/02/2020) e a data em que a Organização Mundial da Saúde classificou a covid-19 como Pandemia (11/03/2020). 9
- Figura 3 – Número Reprodutivo Efetivo (R_e) da covid-19 no Brasil. 10
- Figura 4 – Modelo para $t = 1$ e $\epsilon = \{0.5, 1, 2\}$. A estimação dos parâmetros fornece $\xi = 626.4871$ e $\lambda = 628.4288$. No gráfico, a reta preta é a esperança condicional $\mathbb{E}(R_e(\tau)|R_e(\sigma))$, o envelope azul escuro representa um desvio padrão e o envelope azul claro dois desvios padrão. Os pontos vermelhos são os pares $(R_e(d - 1), R_e(d))$ em que d representa o dia observado. 16
- Figura 5 – Modelo para $t = 7$ dias. O valor de ϵ é igual a 0.5, 1 e 2, respectivamente. Apesar da representação gráfica ser praticamente idêntica para os três valores de ϵ , obtêm-se diferentes valores de parâmetros: $(\xi, \lambda) = (149.2549, 151.283)$, $(394.8913, 400.2572)$ e $(2764.239, 2801.8)$. Nos gráficos, as retas pretas são a esperança condicional $\mathbb{E}(R_e(\tau)|R_e(\sigma))$, os envelopes azul escuro representam um desvio padrão e os envelopes azul claro dois desvios padrão. Os pontos vermelhos são os pares $(R_e(d - 7), R_e(d))$ em que d representa o dia observado. 17

- Figura 6 – Modelo para $t = 15$ dias. O valor de ϵ é igual a 0.5, 1 e 2, respectivamente. Apesar da representação gráfica ser praticamente idêntica para os três valores de ϵ , obtêm-se diferentes valores de parâmetros: $(\xi, \lambda) = (137.9355, 140.5156)$, $(534.2217, 544.2146)$ e $(8013.326, 8163.218)$. Nos gráficos, as retas pretas são a esperança condicional $\mathbb{E}(R_e(\tau)|R_e(\sigma))$, os envelopes azul escuro representam um desvio padrão e os envelopes azul claro dois desvios padrão. Os pontos vermelhos são os pares $(R_e(d - 15), R_e(d))$ em que d representa o dia observado. 18
- Figura 7 – Número Reprodutivo Efetivo (R_e) da covid-19 no Brasil durante a primeira onda de transmissão, de 23 de fevereiro de 2020 a 07 de novembro de 2020. 21
- Figura 8 – Número Reprodutivo Efetivo (R_e) da covid-19 no Brasil durante a segunda onda de transmissão, de 08 de novembro de 2020 a 25 de dezembro de 2021. 22
- Figura 9 – Número Reprodutivo Efetivo (R_e) da covid-19 no Brasil durante a terceira onda de transmissão, de 26 de dezembro de 2021 a 21 de maio de 2022. 23
- Figura 10 – Em (a), o período 1 vai de 25 de fevereiro de 2020 a 25 de setembro de 2022, totalizando 943 dias; o período 2, com 90 dias, vai de 26 de setembro de 2022 a 25 de dezembro de 2022. Em (b), o período 1 vai de 25 de fevereiro de 2020 a 25 de outubro de 2022, totalizando 973 dias; o período 2, com 60 dias, vai de 26 de outubro de 2022 a 25 de dezembro de 2022. Em (c), o período 1 vai de 25 de fevereiro de 2020 a 24 de novembro de 2022, totalizando 1003 dias; o período 2, com 30 dias, vai de 25 de novembro de 2022 a 25 de dezembro de 2022. Em todos os casos o primeiro dia do período 1 corresponde à data de confirmação do primeiro caso de covid-19 no Brasil. Por outro lado, o último dia de cada período 2 corresponde ao último dia de registro disponível na base de dados trabalhada. 25

- Figura 11 – Modelo para $t = 7$ dias e $\epsilon = 1$ usando os parâmetros $\xi = 547.2682$ e $\lambda = 554.5022$, calculados com base nos primeiros 943 dias da pandemia. O modelo foi capaz de capturar 84.47% dos dados. A reta preta é a esperança condicional $\mathbb{E}(R_e(\tau)|R_e(\sigma))$, o envelope azul escuro representa um desvio padrão e o envelope azul claro dois desvios padrão. Os pontos vermelhos são os pares $(R_e(d - 7), R_e(d))$ dos últimos 90 dias da base de dados e d representa o dia observado. 26
- Figura 12 – Modelo para $t = 7$ dias e $\epsilon = 1$ usando os parâmetros $\xi = 558.8363$ e $\lambda = 565.9619$, calculados com base nos primeiros 973 dias da pandemia. O modelo foi capaz de capturar 78.72% dos dados. A reta preta é a esperança condicional $\mathbb{E}(R_e(\tau)|R_e(\sigma))$, o envelope azul escuro representa um desvio padrão e o envelope azul claro dois desvios padrão. Os pontos vermelhos são os pares $(R_e(d - 7), R_e(d))$ dos últimos 60 dias da base de dados e d representa o dia observado. 27
- Figura 13 – Modelo para $t = 7$ dias e $\epsilon = 1$ usando os parâmetros $\xi = 550.9037$ e $\lambda = 557.5795$, calculados com base nos primeiros 1003 dias da pandemia. O modelo foi capaz de capturar 35.29% dos dados. A reta preta é a esperança condicional $\mathbb{E}(R_e(\tau)|R_e(\sigma))$, o envelope azul escuro representa um desvio padrão e o envelope azul claro dois desvios padrão. Os pontos vermelhos são os pares $(R_e(d - 7), R_e(d))$ dos últimos 30 dias da base de dados e d representa o dia observado. 28
- Figura 14 – Ilustração do modelo compartimental SI. b é a proporção da população que sai do estado suscetível (S) para infectado (I) e a é a proporção da população que sai do estado infectado (I) para suscetível (S). 37
- Figura 15 – Gráficos de $S(t)$ e $I(t)$ para $a > b$ (taxa de infectado para suscetível maior que taxa de contaminação). 39

Lista de tabelas

Tabela 1 – Vetores de parâmetros para o conjunto de dados, razão ξ/λ e desvio padrão.	19
Tabela 2 – Vetores de parâmetros da primeira onda de covid-19, razão ξ/λ e desvio padrão.	21
Tabela 3 – Vetores de parâmetros da segunda onda de covid-19, razão ξ/λ e desvio padrão.	22
Tabela 4 – Vetores de parâmetros da terceira onda de covid-19, razão ξ/λ e desvio padrão.	23
Tabela 5 – Tabela comparativa dos vetores de parâmetros do modelo Gama Condicional Modificado para a covid-19 no Brasil, dada uma janela de $t = 7$ dias.	24
Tabela 6 – Tabela comparativa da capacidade preditiva do modelo Gama Condicional Modificado.	28

Sumário

1	Introdução	1
2	Objetivos	3
2.1	Objetivo Geral	3
2.2	Objetivos Específicos	3
3	Conceitos fundamentais	4
3.1	Processos estocásticos e Cadeias de Markov	4
3.2	Processos de difusão	5
3.3	Covid-19 no Brasil	7
4	Metodologia	12
4.1	Modelo estocástico	12
4.2	Base de dados	13
4.3	Tratamento dos dados	14
4.4	Janela de dias	15
5	Aplicação do modelo	16
5.1	Aplicação do modelo ao conjunto completo de dados	16
5.2	Parâmetros da primeira onda de covid-19	20
5.3	Parâmetros da segunda onda de covid-19	21
5.4	Parâmetros da terceira onda de covid-19	22
5.5	Comparação de parâmetros	23
5.6	Capacidade preditiva do modelo	25
5.6.1	Capacidade preditiva do modelo para os últimos 90 dias	25
5.6.2	Capacidade preditiva do modelo para os últimos 60 dias	26
5.6.3	Capacidade preditiva do modelo para os últimos 30 dias	27
5.6.4	Comparação da capacidade preditiva	28
6	Conclusões	30
6.1	Valor fixo para ϵ	30
6.2	R_e como variável aleatória modelada	30
6.3	Adequação do modelo Gama Condicional Modificado	31

Referências Bibliográficas	32
Apêndices	36
APÊNDICE A R_0 no modelo compartimental SIS	37
APÊNDICE B Código em R para a construção dos gráficos de envelope . .	41

1 Introdução

A epidemiologia é um ramo das ciências médicas que se ocupa em estudar como as doenças se espalham em uma população (GORDIS, 2013). O enfoque trivial ao se pensar em uma epidemia é o médico-sanitário, isto é, maneiras de evitar a propagação da doença e de combatê-la nos indivíduos já acometidos. Contudo, dada a complexidade do fenômeno, muitas outras áreas da ciência contribuem para a compreensão de uma epidemia. Este trabalho tem o objetivo de contribuir com o conhecimento matemático-estatístico da propagação da Síndrome Respiratória Aguda Grave (SRAG) no Brasil, incluindo a covid-19, nos anos de 2020 a 2022. Para tanto, será verificada a adequação da distribuição de probabilidade Gama Condicional modificada (SCHREIBER, 2015) para modelar o Número Reprodutivo Efetivo (R_e).

O Número Reprodutivo Efetivo (ou Número Efetivo de Reprodução) R_e de uma doença é a medida de força de sua propagação, também conhecido como virulência. Trata-se do número esperado de novas infecções causadas por um agente infectado em uma população onde parte dos indivíduos está suscetível e outra não (GOSTIC et al., 2020). O fato de que a população é composta por indivíduos suscetíveis e não suscetíveis é determinante para a definição de R_e , pois uma outra medida de interesse na epidemiologia é o Número Reprodutivo Básico (R_0), definido como o número esperado de novas infecções causadas por um agente infectado em uma população onde todos os indivíduos são suscetíveis (GUERRA et al., 2017). A diferença crucial entre essas medidas é que R_0 caracteriza a fase inicial da epidemia, enquanto R_e é a intensidade da propagação ao longo do tempo.

Essa dissertação adota a modelagem compartimental para a descrição da dinâmica de uma doença. Tal abordagem consiste em dividir a população sob análise em categorias de interesse, como suscetíveis (S), indivíduos que não estão doentes mas podem vir a tornar-se; infectados (I), indivíduos que contraíram a doença e recuperados (R), indivíduos que após serem infectados foram curados e adquiriram imunidade à doença (BRAUER; CASTILLO-CHAVEZ; FENG, 2019). A depender da doença, é possível definir outros compartimentos de modo a representar adequadamente seu comportamento, por exemplo: expostos (E), pessoas que estão em contato com a doença; imunizados passivamente (M), recém nascidos cuja imunidade é herdada da mãe e quarentena (Q), pessoas que ficam

isoladas após o contato com alguém infectado (ROCHA, 2012).

Os compartimentos de cada modelo se relacionam entre si por taxas. Tomando, por exemplo, o modelo Suscetível-Infectado-Suscetível (SIS), a parte da população que sai da categoria suscetível para infectado pode ser representada como taxa de infecção, enquanto a parcela da população que deixa a categoria de infectado e passa a estar novamente suscetível é a taxa de recuperação. Essas taxas são simplesmente a proporção da população que está mudando de compartimento. Uma pergunta possível é a seguinte: qual a vantagem de modelar o número efetivo de reprodução em vez da taxa de infecção? A resposta é que o número efetivo de reprodução é uma medida que consolida todas as taxas do modelo (taxas de infecção, de recuperação, de morte causada pela doença, de morte por causas diversas e etc, a depender do modelo sob análise).

Os modelos compartimentais são tratados como sistemas de equações diferenciais ordinárias. À medida que aumenta a quantidade de compartimentos do modelo, aumenta também a complexidade do sistema de equações. No apêndice A, mostra-se o tratamento do modelo compartimental SIS como um sistema de duas equações diferenciais ordinárias, que culmina com a determinação do número reprodutivo básico R_0 em função das taxas de contaminação e de recuperação, a partir de algumas suposições iniciais em relação às proporções de infectados e suscetíveis da população.

O primeiro modelo compartimental da literatura é o SIR, publicado em 1927 por Kermack e McKendrick (KERMACK; MCKENDRICK, 1927). Nesse modelo, é desenvolvido um sistema de duas equações diferenciais ordinárias que representam a variação da quantidade de indivíduos suscetíveis e infectados ao longo do tempo. Algumas suposições básicas precisam ser feitas: um agente infectado é capaz de contaminar um número médio de pessoas por unidade de tempo em uma população de tamanho fixo; uma certa proporção de infectados muda para a categoria recuperados e não há entrada ou saída de novos indivíduos na população, exceto, no último caso, por morte pela doença (BRAUER, 2005).

A covid-19, doença causada pelo vírus SARS-CoV-2 e foco desse trabalho, tem sua dinâmica explicada pelo modelo Suscetível-Infectado-Isolado-Removido. Tais compartimentos podem ainda ser subdivididos: infectados sintomáticos e assintomáticos, isolados em quarentena ou hospitalizados e removidos por cura ou morte. A relação entre esses compartimentos não é rígida, isto é, os indivíduos podem mudar de categoria de forma não-sequencial. Por exemplo, uma pessoa infectada pode mudar para o compartimento Removido sem ter sido isolada (WANG et al., 2020).

2 Objetivos

2.1 Objetivo Geral

Descrever completamente a dinâmica do Número Reprodutivo Efetivo (R_e) como um processo de difusão, considerando-se a distribuição Gama Condicional Modificada e estabelecer as condições para sua aplicação.

2.2 Objetivos Específicos

- Reescrever a dinâmica da virulência considerando R_e ;
- Caracterizar processos estocásticos markovianos e processos de difusão, considerando suas propriedades;
- Aplicar o modelo Gama Condicional Modificado aos dados de covid-19 no Brasil, fazendo recortes de interesse em períodos críticos; e
- Verificar a capacidade preditiva do modelo Gama Condicional Modificado, de modo que possa ser útil futuramente no acompanhamento de outras doenças.

3 Conceitos fundamentais

3.1 Processos estocásticos e Cadeias de Markov

Processos estocásticos são uma extensão do conceito de variável aleatória. Formalmente, dizemos que um processo estocástico é uma família de variáveis aleatórias $\{X(t, s) | t \in T, s \in \Omega\}$, onde geralmente o parâmetro t representa o tempo e Ω é o espaço de estados (IBE, 2013). É possível simplificar a notação suprimindo a variável s do espaço de estados Ω , de modo que $X(t, s) = X(t) = X_t$. Fixando o estado $s \in \Omega$, o processo estocástico pode ser entendido como uma função do tempo. Por outro lado, fixando um valor $t \in T$, o processo estocástico pode ser compreendido como uma variável aleatória. Para exemplificar, seja $\{X(t, s) | t \in T, s \in \Omega\}$ o processo estocástico do lançamento de um dado 15 vezes. O espaço de estados Ω representa cada possível resultado do conjunto de 15 lançamentos. Já o conjunto T individualiza cada lançamento, seja qual for o elemento $s \in \Omega$. Em outras palavras, $t = n$ representa o resultado observado no n -ésimo lançamento para qualquer $s \in \Omega$ e $n \in \mathbb{Z}$ tal que $1 \leq n \leq 15$.

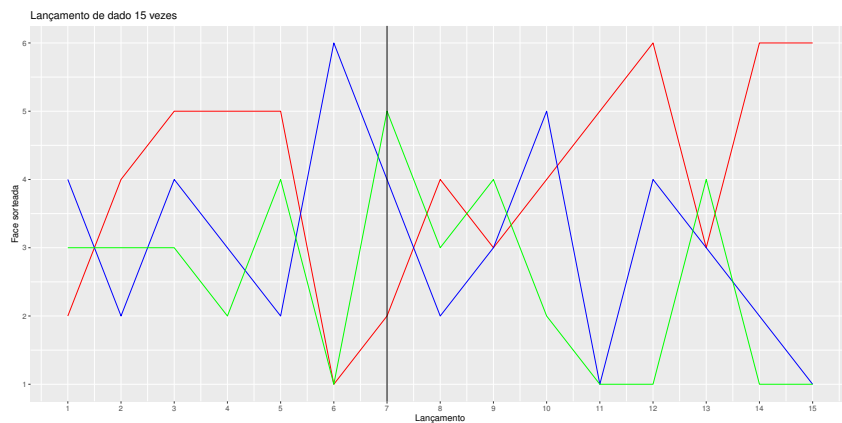


Figura 1 – Resultado para três cenários de 15 lançamentos de um dado, representados pelas cores vermelho, azul e verde. Cada cenário s_i , $i = \{1, 2, 3\}$, é um elemento distinto do espaço de estados Ω . Fixando $s_1 \in \Omega$, o processo estocástico dos lançamentos torna-se a série temporal destacada em vermelho. Fixando a observação $t = 7$, destacada na linha vertical preta, o processo estocástico é a variável aleatória $X(7, s)$ cujo espaço amostral é o conjunto $\Phi = \{1, 2, 3, 4, 5, 6\}$, idêntico ao espaço amostral de apenas um lançamento do dado.

Processos estocásticos podem ser classificados de acordo com o parâmetro T ou com o espaço de estados Ω . Considerando o parâmetro T como sendo o tempo (caso mais usual), dizemos que um processo estocástico é de tempo contínuo ou de tempo discreto. Analogamente, se o espaço de estados Ω é contínuo classificamos o processo estocástico como sendo de espaço contínuo e se o espaço de estados Ω é discreto classificamos o processo estocástico como sendo de espaço discreto (IBE, 2013).

Um processo estocástico $\{X(t, s) | t \in T, s \in \Omega\}$ é chamado de processo markoviano de primeira ordem se para qualquer $t_0 \leq t_1 \leq \dots \leq t_n$ a probabilidade condicional de $X(t_n)$ dados $X(t_0), X(t_1), \dots, X(t_{n-1})$ depende apenas de $X(t_{n-1})$. Isto é,

$$\begin{aligned} P(X(t_n) = x_n | X(t_{n-1}) = x_{n-1}, X(t_{n-2}) = x_{n-2}, \dots, X(t_0) = x_0) \\ = P(X(t_n) = x_n | X(t_{n-1}) = x_{n-1}). \end{aligned}$$

Essa característica é conhecida como propriedade markoviana e representa a falta de memória de uma cadeia de Markov, denominação dada aos processos estocásticos de espaço discreto, tanto no caso do tempo contínuo quanto no caso do tempo discreto. Assim, dizemos que o estado futuro de uma cadeia de Markov não depende dos estados passados, mas apenas do estado presente (SALSA, 2016). No caso de processos estocásticos com espaço de estados contínuo, classificados como processos de Markov, a definição é semelhante:

$$\begin{aligned} P(X(t_n) \in S | X(t_{n-1}) \in K_{n-1}, X(t_{n-2}) \in K_{n-2}, \dots, X(t_0) \in K_0) \\ = P(X(t_n) \in S | X(t_{n-1}) \in K_{n-1}), \end{aligned}$$

com $S, K_i \subset \Omega$ ($i = 1, \dots, n$).

3.2 Processos de difusão

Difusão é o processo de espalhamento da matéria por um sistema através do movimento aleatório de suas partículas (CRANK, 1979). Muitos fenômenos naturais podem ser modelados via processos de difusão, como a transmissão de calor em um sólido metálico, o espalhamento de um gás em uma câmara e a propagação de uma doença na população, sendo este último exemplo o caso particular tratado ao longo dessa dissertação.

Diz-se que um processo estocástico X_t é um processo de difusão se for um processo de Markov (ou seja, um processo estocástico de espaço de estados contínuo) com trajetórias

quase certamente ¹ contínuas tal que X_t é contiuamente diferenciável até a segunda ordem e satisfaz, com convergência uniforme para todo $x \in \mathbb{R}$ e $t \in [0, d)$, as seguintes condições (BRAUMANN, 2005):

$$\lim_{\Delta \rightarrow 0^+} \frac{1}{\Delta} P(|X_{t+\Delta} - x| > \epsilon | X_t = x) = 0 \quad (\forall \epsilon > 0), \quad (3.1)$$

$$\lim_{\Delta \rightarrow 0^+} \mathbb{E} \left[\frac{X_{t+\Delta} - x}{\Delta} | X_t = x \right] = a(t, x), \quad (3.2)$$

$$\lim_{\Delta \rightarrow 0^+} \mathbb{E} \left[\frac{(X_{t+\Delta} - x)^2}{\Delta} | X_t = x \right] = b(t, x). \quad (3.3)$$

A equação (3.1) significa que dado um pequeno intervalo de tempo, não se espera grandes variações no processo. No caso do número reprodutivo efetivo, trata-se de uma condição bastante razoável, visto que seu crescimento ou decrescimento acontece de forma controlada, sem saltos repentinos. Matematicamente, quer dizer que $R_e(\sigma + \Delta) \approx R_e(\sigma)$ quando $\Delta \rightarrow 0^+$. A equação (3.2) é o coeficiente de tendência (ou termo inglês *drift*) e a equação (3.3) é o coeficiente de difusão (BRAUMANN, 2005).

As equações (3.2) e (3.3) são os termos da equação diferencial estocástica genérica dada por

$$dX_t = a(t, x)dt + \sqrt{b(t, x)}dW_t. \quad (3.4)$$

Na equação (3.4), W_t é o processo de Wiener, definido como segue:

Definição 1 (Processo de Wiener). Seja $(W_t)_{t \in [0, +\infty)}$ um processo estocástico definido no espaço de probabilidade (Ω, \mathcal{F}, P) . W_t é chamado processo de Wiener se satisfaz as seguintes propriedades (BRAUMANN, 2005):

- $W(0)=0$ quase certamente;
- Os incrementos $W(t) - W(s)$ ($s < t$) têm distribuição normal com média 0 e variância $t - s$;

¹ Definição de Convergência Quase Certa (MAGALHÃES, 2006): Considere o espaço de probabilidade (Ω, \mathcal{F}, P) e X uma variável aleatória. A seqüência $\{X_n : n \geq 1\}$ tem convergência quase certa para X se existir um conjunto $A \in \mathcal{F}$ tal que $P(A) = 0$ e $X_n \rightarrow X$ em A^c , quando $n \rightarrow \infty$.

- Em intervalos de tempo $[s_i, t_i)$ não sobrepostos, os incrementos $W(t_i) - W(s_i)$ são variáveis aleatórias independentes.

Supondo $X_0 = x_0$, a solução da equação (3.4) é dada por

$$X_t = x_0 + \int_0^t a(s, x) ds + \int_0^t \sqrt{b(s, x)} dW_s. \quad (3.5)$$

Enquanto a integral $\int_0^t a(s, x) ds$ pode ser determinada com o cálculo tradicional, baseado nas somas de Riemann, a integral $\int_0^t \sqrt{b(s, x)} dW_s$ exige a utilização do cálculo estocástico, desenvolvido por Stratonovich e por Itô (OKSENDAL, 2013).

Uma vez caracterizado o processo de difusão, é possível definir o operador de difusão \mathcal{D} em função dos termos $a(t, x)$ e $b(t, x)$ como

$$\mathcal{D} = \frac{\partial}{\partial x} a(t, x) + \frac{1}{2} \frac{\partial^2}{\partial x^2} b(t, x).$$

Com algumas suposições iniciais a respeito de $a(t, x)$ e $b(t, x)$ e a definição conveniente de novas funções h e u relacionadas pela esperança condicional, o operador de difusão \mathcal{D} satisfaz as condições para as equações de Kolmogorov, cujas soluções são dadas a partir de equações diferenciais parciais. Esse fato é muito importante, pois permite a solução de fenômenos complexos tanto em termos probabilísticos quanto de forma determinística (SALSA, 2016).

3.3 Covid-19 no Brasil

Conforme a Sociedade Brasileira de Imunizações, a covid-19 é a doença causada pelo vírus SARS-CoV-2, pertencente à família dos Coronavírus (CoV) (SBIM, 2021). Os primeiros casos de covid-19 surgiram em Wuhan, na China, em dezembro de 2019, e rapidamente se espalharam pelo mundo (ZHU; WEI; NIU, 2020). A respeito da origem do SARS-CoV-2, a hipótese mais aceita até o momento é de que seu hospedeiro primário seja o morcego. Entretanto, a infecção para humanos não se deu de forma direta, sendo desconhecido o agente intermediário causador das primeiras infecções animal-humano (LAUXMANN; SANTUCCI; AUTRÁN-GÓMEZ, 2020).

À medida que a covid-19 se propagou na população mundial, o vírus SARS-CoV-2 sofreu mutações que deram origem às chamadas variantes (COSAR et al., 2022). Apesar da maioria das mutações não se refletir em mudanças na dinâmica da doença, algumas

revelam essa capacidade e por isso merecem atenção especial. Nesse sentido, a Organização Mundial da Saúde classifica as variantes como Variantes de Interesse (VOIs, sigla da expressão em inglês *Variants of Interest*) e Variantes de Preocupação (VOCs, sigla para a expressão em inglês *Variants of Concern*) (WHO, 2022). As VOIs englobam as mutações que alteram o fenótipo do vírus e que tenham sido identificadas como causadoras da transmissão comunitária da covid-19, de múltiplos casos ou que tenham sido identificadas em vários países. Já as VOCs são as variantes que representam algum impacto na saúde pública. A caracterização como VOC pode ser feita em razão do aumento da transmissibilidade ou virulência da covid-19, mudança na sua manifestação clínica ou diminuição da eficácia das medidas de combate usuais, dos métodos de diagnóstico e dos métodos terapêuticos (incluindo a eficácia de vacinas e imunizantes) (PAHO, 2021). As principais variantes de interesse e de preocupação identificadas no Brasil são a Alfa, Beta, Gama, Zeta e Delta/Kapa (COVID, 2021) ².

O primeiro caso confirmado de covid-19 no Brasil ocorreu em 25 de fevereiro de 2020, em um idoso de 61 anos que visitou a Itália (AQUINO et al., 2020). Antes disso, em 03 de fevereiro de 2020, o Ministério da Saúde declarou Emergência em Saúde Pública de Importância Nacional através da Portaria nº 188/2020. No cenário internacional, os primeiros casos de covid-19 foram confirmados ainda mais cedo: nos Estados Unidos em 21 de janeiro de 2020 (CONTROL; CDC, 2020) e na Europa em 24 de janeiro de 2020 (LESCURE et al., 2020). Já no continente africano o primeiro caso confirmado ocorreu no Egito, em 14 de fevereiro de 2020 (LONE; AHMAD, 2020), enquanto na Oceania o primeiro caso foi confirmado em 25 de janeiro de 2020, na Austrália (IBRAHIM et al., 2020). Diante da rápida propagação da doença, em 30 de janeiro de 2020 a Organização Mundial de Saúde declarou Emergência em Saúde Pública de Importância Internacional, o que viria a agravar-se em 11 de março de 2020, com a declaração de Pandemia, que é a transmissão sustentada de uma doença a nível mundial (FIOCRUZ, 2021).

A disseminação da covid-19 no Brasil ocorreu rapidamente, conforme o gráfico de incidência diária apresentado na FIGURA 2, a seguir.

² Acrescente-se à essa lista a variante Ômicron, que surgiu posteriormente à publicação do artigo referenciado.

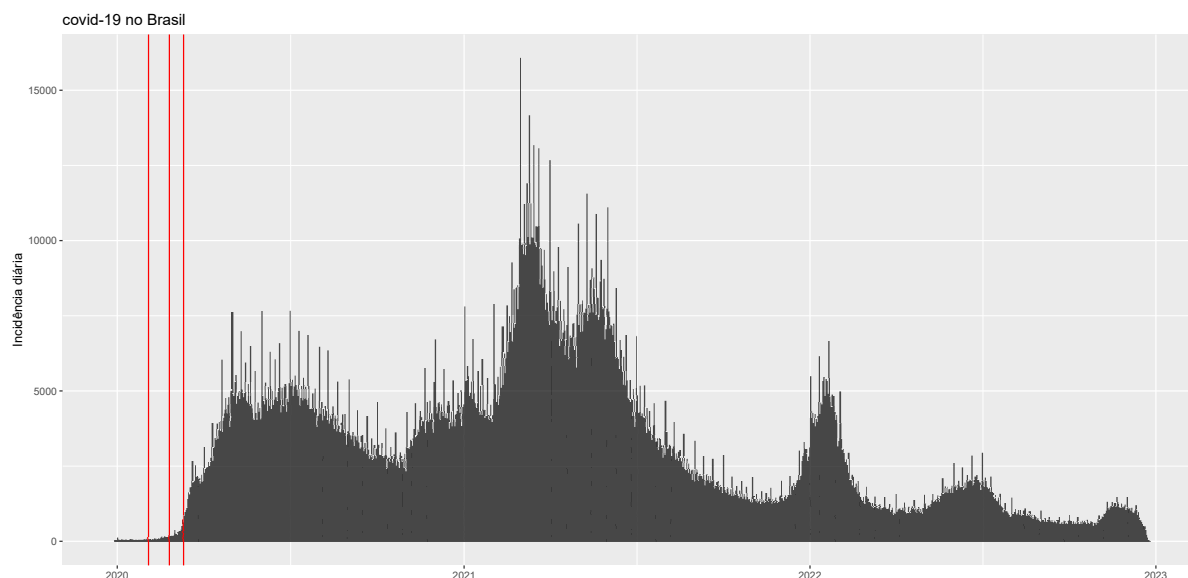


Figura 2 – Incidência diária da covid-19 no Brasil. As retas vermelhas destacam, respectivamente, a data em que o Ministério da Saúde declarou Emergência em Saúde Pública de Importância Nacional (03/02/2020), a data do primeiro caso confirmado no Brasil (25/02/2020) e a data em que a Organização Mundial da Saúde classificou a covid-19 como Pandemia (11/03/2020).

A base de dados utilizada para a elaboração do gráfico de incidência diárias é relativa à Síndrome Respiratória Aguda Grave (SRAG) e engloba tanto as doenças causadas por outros vírus respiratórios quanto a covid-19, causada pelo novo coronavírus (SARS-CoV-2). Sendo assim, os casos apresentados no gráfico anteriormente à confirmação do primeiro caso de covid-19 devem ser compreendidos como as doenças respiratórias comuns na população brasileira antes do início da pandemia. Mais detalhes a respeito da base de dados utilizada e seu tratamento podem ser consultados nas seções 4.2 e 4.3.

A incidência diária de novos casos está fortemente relacionada com o Número Efetivo de Reprodução (R_e), objeto de estudo dessa pesquisa. O número efetivo de reprodução serve para medir a transmissibilidade de uma doença e é definido como o número esperado de novas infecções causadas por um agente infectado em um ambiente onde apenas uma parte da população é suscetível (GOSTIC et al., 2020). Trata-se de uma medida abrangente e que consolida as várias taxas de um fenômeno epidemiológico: a taxa de infecção, a taxa de cura e a taxa de mortalidade. Em contraste ao número efetivo de reprodução, há o Número Básico de Reprodução (R_0), definido como o número esperado de novas infecções causadas por um agente infectado em uma população onde todos os indivíduos são suscetíveis (GUERRA et al., 2017). O R_0 está relacionado à

gênese de uma epidemia, momento em que há o primeiro doente e todos os demais indivíduos da população estão vulneráveis. Há pesquisas que apontam, ainda, outra maneira de quantificar a transmissibilidade de uma doença por meio da individualização da capacidade de transmissão de um agente infectado. Em outras palavras, em vez de descrever a média de novas infecções causadas por um agente infectado (como acontece no número efetivo de reprodução), tenta-se descrever a capacidade que cada indivíduo infectado possui de transmitir a doença, partindo do pressuposto de que nem todo infectado transmitirá a doença para a mesma quantidade de pessoas (LLOYD-SMITH et al., 2005). Trata-se de uma hipótese bastante razoável e que pode ser claramente observada na própria pandemia de covid-19: um indivíduo acometido por essa doença e que não respeita as regras de distanciamento social e uso de máscara certamente a transmitirá para uma quantidade de pessoas maior do que aquele indivíduo infectado que adota o isolamento social pelo período indicado pelas autoridades sanitárias. Apesar de ser uma premissa válida, na prática é muito difícil individualizar a transmissibilidade de cada agente infectado, especialmente no caso de doenças respiratórias, já que não dependem do contato direto entre suscetível e infectado, sendo possível que a simples presença em um ambiente comum seja suficiente para a contaminação (LLOYD-SMITH et al., 2005).

Na FIGURA 3, construída a partir da data dos primeiros sintomas, é possível observar a evolução temporal do número reprodutivo efetivo da covid-19 no Brasil.

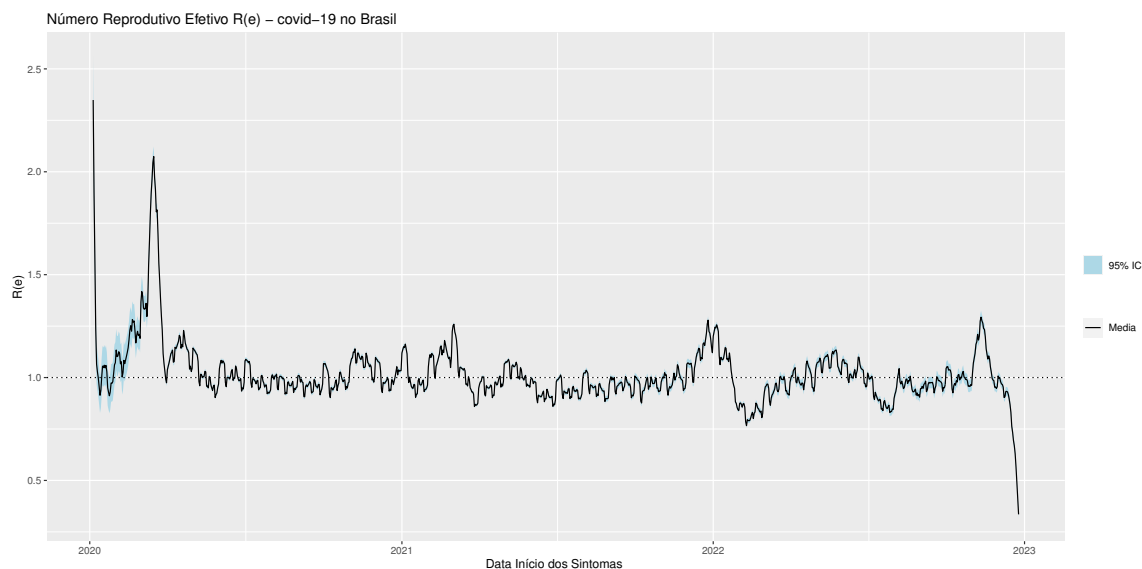


Figura 3 – Número Reprodutivo Efetivo (R_e) da covid-19 no Brasil.

Conforme a FIGURA 3, o número reprodutivo efetivo alcançou picos em alguns momentos. Esses picos podem ser interpretados como surtos da doença e há algumas hipóteses que os justificam de acordo com o momento em que eles ocorrem. No caso do início de uma epidemia, dada a grande quantidade de indivíduos suscetíveis e a ausência de medidas sociais e sanitárias de controle da doença, a inclusão de um agente infectado na população tende a causar mais impacto do que sua inclusão em um ambiente onde já ocorre a transmissão comunitária. Por outro lado, nos picos seguintes aos períodos de transmissão regular, é possível o surgimento de agentes cuja capacidade de contaminação exceda muito à transmissibilidade média dos demais infectados (VESPIGNANI *et al.*, 2020). Há, ainda, a possibilidade de surgimento de novas variantes com maior virulência, decorrentes de mutações do vírus.

Essa pesquisa está concentrada na compreensão da dinâmica do número reprodutivo efetivo (R_e). Nesse sentido, será verificada a adequação do modelo Gama Condicional Modificado para a descrição de R_e como uma variável estocástica e sua caracterização como um Processo de Difusão, o que pressupõe que as condições das equações de Kolmogorov estão satisfeitas.

4 Metodologia

4.1 Modelo estocástico

As distribuições de probabilidade podem ser discretas ou contínuas. Dentre as contínuas, há a distribuição Gama(α, β), cuja função densidade de probabilidade é dada por $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta x} x^{\alpha-1}$, com $\alpha > 0$, $\beta > 0$ e $x > 0$. Além disso, $\mathbb{E}(X) = \frac{\alpha}{\beta}$ e $\text{Var}[X] = \frac{\alpha}{\beta^2}$, quando $X \sim \text{Gama}(\alpha, \beta)$ (MAGALHÃES, 2006). O fato de Gama(α, β) ter suporte no intervalo $[0, \infty)$ é adequado para a modelagem do número reprodutivo efetivo, uma vez que tal medida evolui ao longo do tempo. Para tornar os parâmetros α e β mais representativos para o que se quer modelar, que é o número reprodutivo da covid-19, faz sentido observar algumas premissas do fenômeno:

- (a) O valor do número reprodutivo efetivo de um dia qualquer é influenciado pelo valor do número reprodutivo efetivo do dia anterior. Generalizando um pouco mais: fixado um período de interesse (dias, semanas, meses, etc), o período atual é dado em termos proporcionais do período anterior.
- (b) A distância entre os períodos observados é capaz de influenciar a capacidade de previsão do comportamento da variável estocástica sob análise. Em outras palavras, o valor do número reprodutivo efetivo de hoje é mais bem explicado pelo número reprodutivo efetivo de ontem do que pelo número reprodutivo de 15 dias atrás.
- (c) É possível que a influência do tempo no processo seja mais ou menos intensa. Assim, é interessante pensar em alguma medida de controle para essa variável.

As três premissas apresentadas acima podem ser traduzidas em termos matemáticos da seguinte forma: por (a), toma-se $\alpha = x^2 \xi$ e $\beta = x \lambda$, de modo que ambos são expressos em função do valor do número reprodutivo efetivo, que é x . Por (b), acrescenta-se a informação sobre a distância entre os períodos observados: $\alpha = \frac{x^2 \xi}{\tau - \sigma}$ e $\beta = \frac{x \lambda}{\tau - \sigma}$, com $\sigma < \tau$. Por fim e com o intuito de aplicar uma medida de controle à variação do tempo, tem por (c) que $\alpha = \frac{x^2 \xi}{(\tau - \sigma)^\epsilon}$ e $\beta = \frac{x \lambda}{(\tau - \sigma)^\epsilon}$.

As reflexões acima foram utilizadas para construir o modelo Gama condicional modificado proposto por [Schreiber \(2015\)](#). Assim, esta dissertação utiliza o modelo estocástico Gama condicional modificado, cuja função de distribuição acumulada é definida por ([SCHREIBER, 2015](#)):

$$P(R_0^\#(\tau) \in S | R_0^\#(\sigma) = x) = \frac{\left(\frac{x\lambda}{(\tau-\sigma)^\epsilon}\right)^{\left(\frac{x^2\xi}{(\tau-\sigma)^\epsilon}\right)}}{\Gamma\left(\frac{x^2\xi}{(\tau-\sigma)^\epsilon}\right)} \int_{y \in S} z^{\left(\frac{x^2\xi}{(\tau-\sigma)^\epsilon}\right)-1} e^{-\left(\frac{x\lambda}{(\tau-\sigma)^\epsilon}\right)z} dz, \quad (4.1)$$

onde $P(R_0^\#(\tau) \in S | R_0^\#(\sigma) = x)$ é a probabilidade de que o número reprodutivo básico alcance um valor no conjunto S no instante τ dado que no instante σ o valor do número reprodutivo era x ; $\Gamma(\alpha)$ é a função Gama dada por $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$ e $x > 0, y \geq 0, \lambda > 0, \xi > 0, \epsilon > 0, 0 \leq \sigma \leq \tau$. Com o objetivo de diferenciar a variável estocástica que está sendo tratada nesse trabalho, que é o Número Reprodutivo Efetivo (R_e), utiliza-se a notação $P(R_e(\tau) \in S | R_e(\sigma) = x)$. A função densidade de probabilidade da distribuição Gama Condicional Modificada é dada por

$$p(R_0^\#(\tau) \in S | R_0^\#(\sigma) = x) = \frac{\left(\frac{x\lambda}{(\tau-\sigma)^\epsilon}\right)^{\left(\frac{x^2\xi}{(\tau-\sigma)^\epsilon}\right)}}{\Gamma\left(\frac{x^2\xi}{(\tau-\sigma)^\epsilon}\right)} y^{\left(\frac{x^2\xi}{(\tau-\sigma)^\epsilon}\right)-1} e^{-\left(\frac{x\lambda}{(\tau-\sigma)^\epsilon}\right)y}. \quad (4.2)$$

Adaptando a função acima à variável estocástica R_e , tem-se que $p(R_0^\#(\tau) \in S | R_0^\#(\sigma) = x) = p(R_e(\tau) \in S | R_e(\sigma) = x)$. As condições que tornam $p(R_e(\tau) \in S | R_e(\sigma) = x)$ uma função densidade de probabilidade estão demonstradas em ([SCHREIBER, 2015](#)) e por isso não serão detalhadas.

Conforme o método dos momentos, tem-se que $\mathbb{E}(R_e(\tau) | R_e(\sigma)) = \frac{\xi}{\lambda} R_e(\sigma)$ e $\text{Var}[R_e(\tau) | R_e(\sigma)] = \frac{\xi}{\lambda^2} (\tau - \sigma)^\epsilon = \frac{\xi}{\lambda^2} t^\epsilon$.

4.2 Base de dados

As bases de dados utilizadas para aplicação do modelo consistem em três arquivos *comma-separated-values* (.csv) obtidos no openDataSUS, plataforma online brasileira de disponibilização de dados relativos à saúde da população. Os arquivos se referem à Síndrome Respiratória Aguda Grave (SRAG), gênero que engloba a covid-19, nos anos de 2020, 2021 e 2022 e somados contêm 3.469.123 linhas com os elementos mencionados na lista abaixo:

DT_NOTIFC	SEM_NOT	DT_SIN_PRI	SEM_PRI	SG_UF
ID_REGIONA	CO_REGIONA	ID_MUNICIP	CO_MUN_NOT	CS_SEXO
CO_UNI_NOT	ID_UNIDADE	DT_NASC	NU_IDADE_N	TP_IDADE
COD_IDADE	COD_IDADE	CS_GESTANT	CS_RACA	ID_PAIS
CS_ESCOL_N	CO_RG_RESI	SG_UF_NOT	ID_RG_RESI	CO_PAIS
ID_MN_RESI	CO_MUN_RES	CS_ZONA	SURTO_SG	

A definição de cada elemento pode ser consultada no Dicionário de Dados, documento disponibilizado também no OpenDataSUS, junto a cada arquivo *.csv*. O campo `DT_SIN_PRI` refere-se à data dos primeiros sintomas, informação que serve de base para todas as análises feitas nessa pesquisa. A identidade do paciente é preservada, pois não há menção ao seu nome nem CPF.

4.3 Tratamento dos dados

Os dados foram processados com o software R ([R Core Team, 2021](#)), pelo ambiente de desenvolvimento integrado RStudio ([RStudio Team, 2020](#)). Merece especial destaque o pacote EpiEstim ([CORI, 2021](#)), que estima o Número Reprodutivo Efetivo (R_e) de uma epidemia. Para tanto, é necessário um tratamento intermediário da informação `DT_SIN_PRI` dos bancos de dados em análise, de modo que seja contabilizada a quantidade total de casos por data. Essa etapa foi feita com o uso do pacote Incidence ([JOMBART et al., 2020](#)). Conhecida a incidência de casos por data, essa informação é usada como argumento na função `estimate_R` do pacote EpiEstim ([CORI, 2021](#)). A função `estimate_R` possui a seguinte sintaxe:

```
estimate_R(incidência,
method = c("non_parametric_si", "parametric_si", "uncertain_si",
"si_from_data","si_from_sample"),
si_data = NULL,
  si_sample = NULL,
config = make_config(incid = incid, method = method))
```

Conforme orientação da Organização Pan-Americana de Saúde ([ORGANIZATION, 2020](#)), devem ser adotados, para a estimação do Número Reprodutivo Efetivo (R_e) da covid-19, a média de 4,8 dias e desvio padrão de 2,3 ([LIU et al., 2020](#); [NISHIURA; LINTON; AKHMETZHANOV, 2020](#); [PEAK et al., 2020](#)). Sendo assim, foi adotado o

método "parametric_si" na função estimate_R, conforme detalha a documentação do pacote EpiEstim. O código a ser executado passa a ser o seguinte:

```
estimate_R(incidência,
           method="parametric_si",
           config=make_config(list(
             mean_si= 4.8,
             std_si= 2.3)))
```

4.4 Janela de dias

O uso do pacote EpiEstim fornece uma série com os valores estimados do Número Reprodutivo Efetivo (R_e), indexados ao dia da observação e calculados a partir dos sete dias anteriores (média móvel de R_e). Esse conjunto de dados é a base que fornecerá os valores $R_e(\tau) = y$, em que $y \in S$, e $R_e(\sigma) = x$ (com $\sigma < \tau$) para aplicação do modelo Gama Condicional Modificado (GCM). O termo $\tau - \sigma$ das equações 4.1 e 4.2 será definido como a janela de dias utilizada e, por simplificação, será chamado de t .

A menor janela de tempo adotada nessa dissertação é de apenas 1 dia. Isto significa que está sendo considerada a diferença $\tau - \sigma = 1$. Dessa forma, é possível definir um par genérico de valores de R_e para essa janela de tempo como sendo $(R_e(d-1), R_e(d))$, onde d é o dia da observação. Dessa forma, o par genérico de valores de R_e pode ser escrito como $(R_e(\sigma), R_e(\tau)) = (R_e(d-t), R_e(d))$.

Cumprе destacar uma nuance dessa metodologia de determinação dos pares $(R_e(\sigma), R_e(\tau))$. Supondo que o pacote EpiEstim forneça uma série com n estimações de R_e , o n -ésimo valor pode ser tomado como $R_e(\tau)$, o que forneceria $R_e(\sigma) = R_e(n-1)$. Porém, não seria possível tomar $R_e(\sigma) = R_e(n)$, uma vez que não existe $R_e(n+1)$ que se possa atribuir a $R_e(\tau)$. Como consequência, o modelo GCM poderá ser reproduzido $n-1$ vezes. É possível generalizar esse resultado para qualquer valor da janela de tempo $(\tau - \sigma) = t$ adotada, de modo que um vetor com n estimações de R_e poderá ser reproduzido $n-t$ vezes pelo modelo. A consequência desse fato é que à medida que se aumenta a janela de tempo a ser analisada, menos dados estão disponíveis para aplicação do modelo. Janelas de tempo muito grandes e poucas estimações de R_e podem prejudicar a capacidade preditiva do modelo. Nesse trabalho, além de uma grande quantidade de estimações de R_e (pouco menos que 1.100), a maior janela de tempo adotada foi $t = 15$, de modo que não há redução significativa dos dados para processamento.

5 Aplicação do modelo

5.1 Aplicação do modelo ao conjunto completo de dados

Nesta seção será discutida a aderência do modelo aos dados. Para tanto, serão testadas diferentes janelas de tempo t iguais a 1 dia, 7 dias e 15 dias. À variável ϵ serão atribuídos três valores distintos: $\epsilon = \{0.5, 1, 2\}$.

A estimação dos parâmetros ξ e λ é feita conforme (SCHREIBER, 2015), em que as variáveis t e ϵ se relacionam sempre como t^ϵ . Por consequência, $(t, \epsilon) = \{(1, 0.5), (1, 1), (1, 2)\}$ fornece os mesmos valores dos parâmetros ξ e λ , o que possibilita representar os três casos em apenas um gráfico de envelopes, como explicitado na FIGURA 4.

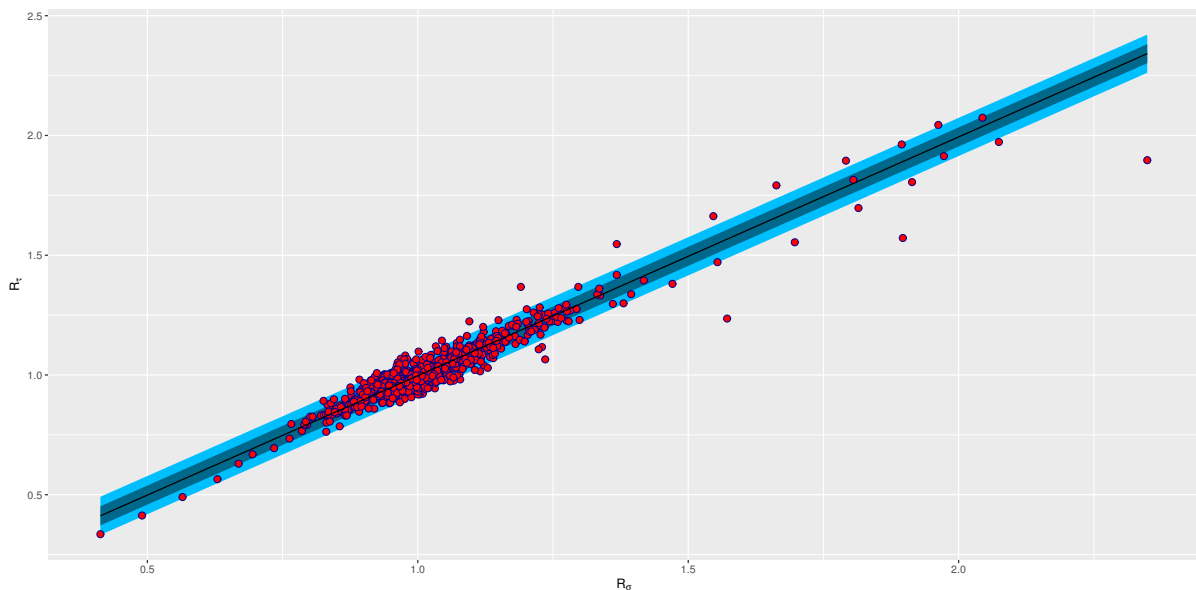


Figura 4 – Modelo para $t = 1$ e $\epsilon = \{0.5, 1, 2\}$. A estimação dos parâmetros fornece $\xi = 626.4871$ e $\lambda = 628.4288$. No gráfico, a reta preta é a esperança condicional $\mathbb{E}(R_e(\tau)|R_e(\sigma))$, o envelope azul escuro representa um desvio padrão e o envelope azul claro dois desvios padrão. Os pontos vermelhos são os pares $(R_e(d-1), R_e(d))$ em que d representa o dia observado.

Com a janela de $t = 1$ dia, o modelo captura 96.12% dos dados, revelando uma boa aderência. O gráfico de envelope é influenciado pela variação de t , mas não pela

variação de ϵ , como poderá ser visto nas FIGURAS 5 e 6, ao aplicar $t = 7$ e $t = 15$.

A FIGURA 5 ilustra a aplicação do modelo a uma janela de $t = 7$ dias, o período mais utilizado nos boletins epidemiológicos de acompanhamento da covid-19 para o cálculo da média móvel ponderada. O modelo captura 97.59% dos dados nessa janela.

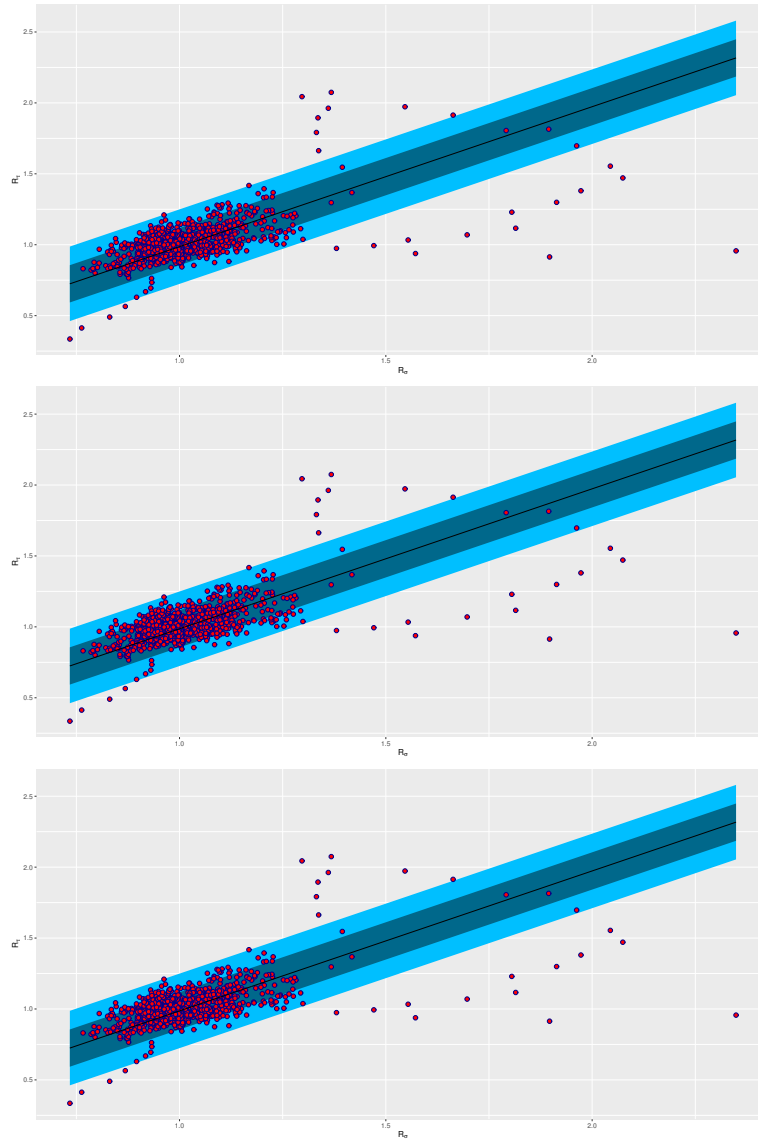


Figura 5 – Modelo para $t = 7$ dias. O valor de ϵ é igual a 0.5, 1 e 2, respectivamente. Apesar da representação gráfica ser praticamente idêntica para os três valores de ϵ , obtêm-se diferentes valores de parâmetros: $(\xi, \lambda) = (149.2549, 151.283)$, $(394.8913, 400.2572)$ e $(2764.239, 2801.8)$. Nos gráficos, as retas pretas são a esperança condicional $\mathbb{E}(R_e(\tau)|R_e(\sigma))$, os envelopes azul escuro representam um desvio padrão e os envelopes azul claro dois desvios padrão. Os pontos vermelhos são os pares $(R_e(d-7), R_e(d))$ em que d representa o dia observado.

A seguir, a FIGURA 6 representa a aplicação do modelo a uma janela de 15 dias, em que 96.91% dos dados são capturados. O APÊNDICE B reproduz o código utilizado no software R para criar os gráficos de envelope, tomando como exemplo $t = 15$ e $\epsilon = 1$.

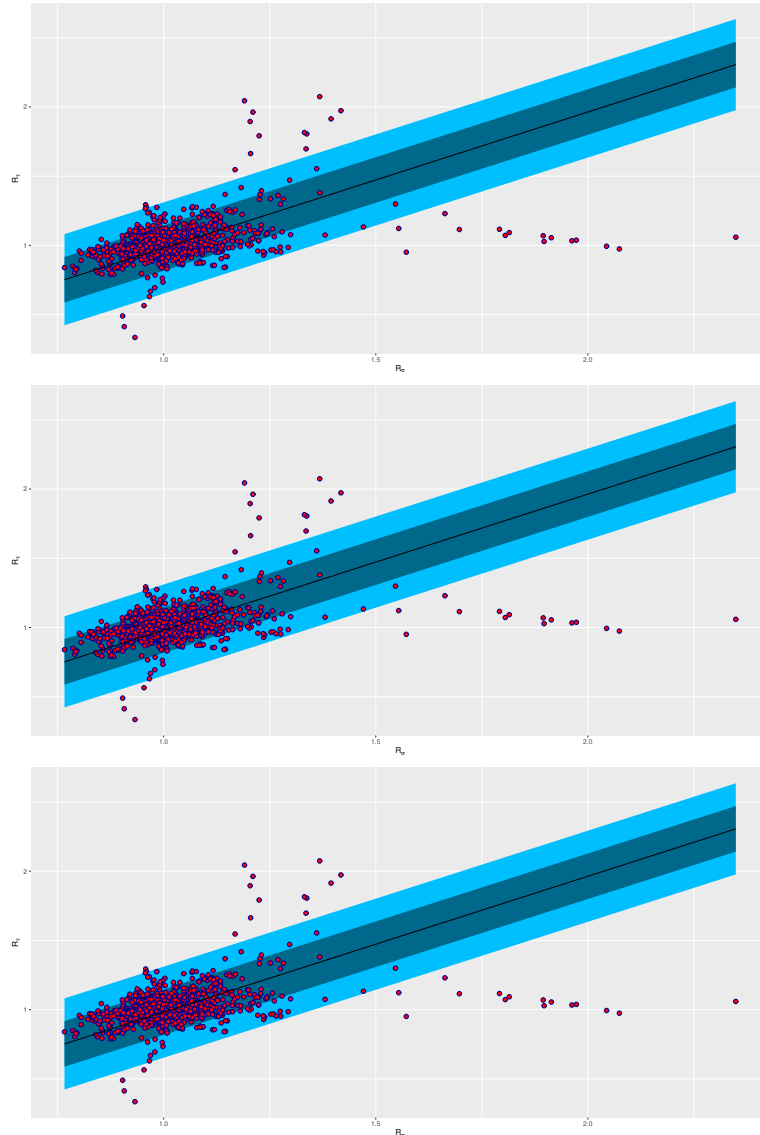


Figura 6 – Modelo para $t = 15$ dias. O valor de ϵ é igual a 0.5, 1 e 2, respectivamente. Apesar da representação gráfica ser praticamente idêntica para os três valores de ϵ , obtêm-se diferentes valores de parâmetros: $(\xi, \lambda) = (137.9355, 140.5156)$, $(534.2217, 544.2146)$ e $(8013.326, 8163.218)$. Nos gráficos, as retas pretas são a esperança condicional $\mathbb{E}(R_e(\tau)|R_e(\sigma))$, os envelopes azul escuro representam um desvio padrão e os envelopes azul claro dois desvios padrão. Os pontos vermelhos são os pares $(R_e(d - 15), R_e(d))$ em que d representa o dia observado.

As informações dos gráficos anteriores podem ser resumidas na seguinte tabela, à qual foi adicionada a razão $\frac{\xi}{\lambda}$ e o desvio padrão.

Tabela 1 – Vetores de parâmetros para o conjunto de dados, razão ξ/λ e desvio padrão.

t	ϵ	ξ	λ	ξ/λ	Desvio Padrão
1	0.5, 1 ou 2	626.4871	628.4288	0.9969102	0.0398290
	0.5	149.2549	151.2830	0.9865940	0.1313556
7	1	394.8913	400.2572	0.9865939	0.1313556
	2	2764.239	2801.800	0.9865940	0.1313556
15	0.5	137.9355	140.5156	0.9816383	0.1644888
	1	534.2217	544.2146	0.9816379	0.1644888
	2	8013.326	8163.218	0.9816381	0.1644888

É possível perceber que a razão ξ/λ é constante até pelo menos a quinta casa decimal para cada valor de t , mostrando uma dependência praticamente irrelevante do valor de ϵ . Isso justifica a semelhança dos gráficos de envelope, considerando-se a mesma janela de tempo. Tal razão é de particular interesse porque influencia diretamente o cálculo da esperança condicional, dada por $\mathbb{E}(R_e(\tau)|R_e(\sigma)) = \frac{\xi}{\lambda}R_e(\sigma)$ (SCHREIBER, 2015). Também é possível depreender da TABELA 1 que o desvio padrão não sofre influência do valor de ϵ , apenas de t .

Nas aplicações acima, o modelo foi capaz de capturar pelo menos 96% dos dados independentemente da janela de tempo t adotada, o que revela uma boa aderência aos dados. Por outro lado, a janela de tempo t influencia significativamente o cálculo do desvio padrão, que é dado por $\sqrt{\frac{\xi}{\lambda^2}(\tau - \sigma)^\epsilon}$ (SCHREIBER, 2015). Essa influência é refletida nos gráficos de envelope anteriores, nos quais as regiões azul escuro e azul claro (que representam um e dois desvios padrão, respectivamente) aumentam à medida em que se aumenta a janela de dias adotada. Em outras palavras, quanto maior a janela de dias adotada entre as observações, maior o intervalo de variação do valor do número efetivo de reprodução, pois

$$R_e(\tau) \in \left[\frac{\xi}{\lambda}R_e(\sigma) - 2\sqrt{\frac{\xi}{\lambda^2}(\tau - \sigma)^\epsilon}, \frac{\xi}{\lambda}R_e(\sigma) + 2\sqrt{\frac{\xi}{\lambda^2}(\tau - \sigma)^\epsilon} \right].$$

O intervalo acima é particularmente útil para os tomadores de decisão, pois fornece uma previsão confiável (nesse caso específico, de 96%) do número efetivo de reprodução. O intervalo pode ser ajustado de acordo com o período considerado ideal para a adoção de

medidas sanitárias, tomando-se o cuidado de alterar não apenas o valor de $t = \tau - \sigma$, mas também os valores de ξ e λ , que mudam de acordo com t , conforme se nota na TABELA 1.

O surgimento de novas variantes do vírus SARS-CoV-2 causou a ocorrência de surtos de covid-19 que ficaram popularmente conhecidos como novas ondas da doença. A primeira onda foi aquela causada pelo vírus em seu estado selvagem e que deu início à pandemia. No Brasil, a segunda e terceira onda estão associadas mais fortemente às variantes de preocupação Gama (FARIA et al., 2021) e Ômicron (OLIVEIRA et al., 2022), respectivamente. Nas próximas seções, serão analisados os parâmetros ξ e λ de cada onda de contaminação por covid-19 no Brasil, com o objetivo de verificar se diferem significativamente daqueles obtidos com a aplicação do modelo ao total dos dados. Para tanto, será considerado que a primeira onda ocorreu de 23 de fevereiro de 2020 a 07 de novembro de 2020, a segunda de 08 de novembro de 2020 a 25 de dezembro de 2021 e a terceira de 26 de dezembro de 2021 até 21 de maio de 2022 (MOURA et al., 2022). Outros recortes de interesse podem ser feitos, como os parâmetros do modelo por ano, por semestre, por percentual de doses aplicadas e etc.

5.2 Parâmetros da primeira onda de covid-19

A FIGURA 7 mostra o comportamento do número reprodutivo efetivo da covid-19 ao longo da primeira onda de contaminação, causada pelo vírus em sua forma selvagem.

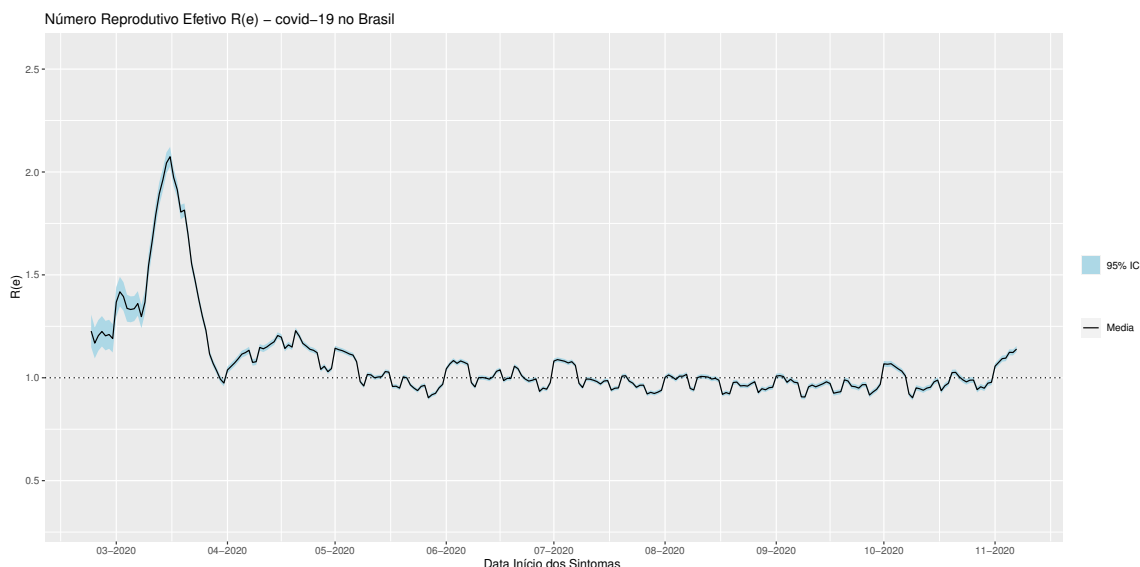


Figura 7 – Número Reprodutivo Efetivo (R_e) da covid-19 no Brasil durante a primeira onda de transmissão, de 23 de fevereiro de 2020 a 07 de novembro de 2020.

A TABELA 2 sintetiza os valores dos parâmetros ξ e λ , das variáveis t e ϵ , a razão $\frac{\xi}{\lambda}$ e o desvio padrão, todos relativos à primeira onda de covid-19.

Tabela 2 – Vetores de parâmetros da primeira onda de covid-19, razão ξ/λ e desvio padrão.

t	ϵ	ξ	λ	ξ/λ	Desvio Padrão
1	0.5, 1 ou 2	301.4839	304.5889	0.989806	0.0570056
	0.5	64.83124	67.47393	0.960834	0.1941023
7	1	171.5274	178.5192	0.960834	0.1941023
	2	1200.691	1249.635	0.960833	0.1941023
15	0.5	76.58206	82.28283	0.930717	0.2093038
	1	296.6011	318.6800	0.930718	0.2093038
	2	4449.016	4780.200	0.930718	0.2093038

5.3 Parâmetros da segunda onda de covid-19

A segunda onda de covid-19 no Brasil teve origem na cidade de Manaus, estado do Amazonas, e logo se espalhou pelo restante do país (SOUZA et al., 2021). Foi causada principalmente pela variante de preocupação Gama. A FIGURA 8 mostra a dinâmica do número reprodutivo efetivo nesse período e a TABELA 3 exibe os valores dos parâmetros ξ e λ , das variáveis t e ϵ , a razão $\frac{\xi}{\lambda}$ e o desvio padrão desse período.

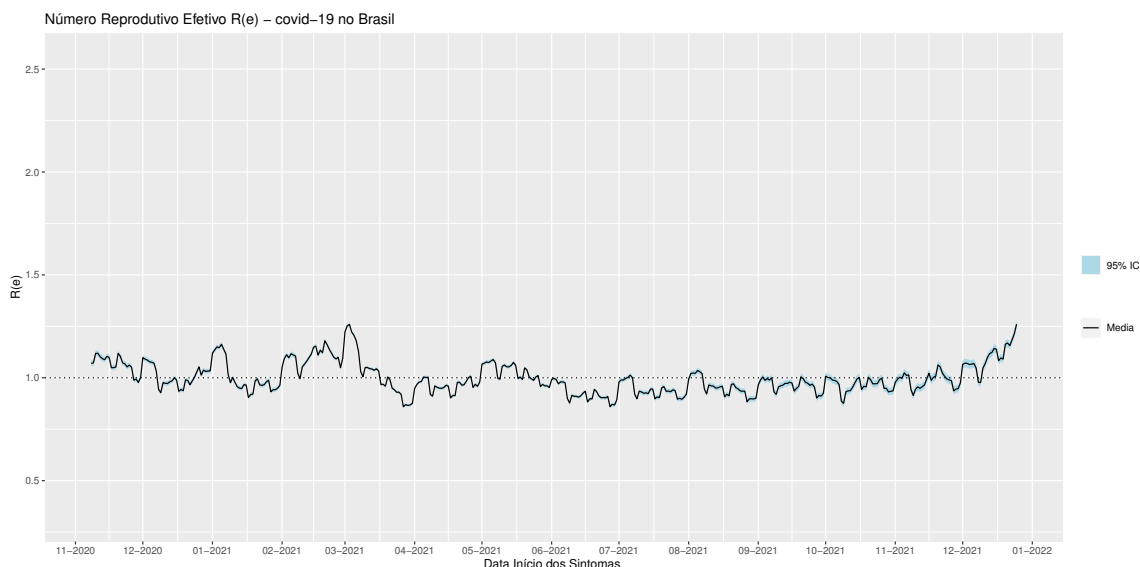


Figura 8 – Número Reprodutivo Efetivo (R_e) da covid-19 no Brasil durante a segunda onda de transmissão, de 08 de novembro de 2020 a 25 de dezembro de 2021.

Tabela 3 – Vetores de parâmetros da segunda onda de covid-19, razão ξ/λ e desvio padrão.

t	ϵ	ξ	λ	ξ/λ	Desvio Padrão
1	0.5, 1 ou 2	553.0552	555.5073	0.995585	0.0423345
	0.5	226.4473	229.0813	0.988501	0.1068485
7	1	599.1233	606.0921	0.988502	0.1068485
	2	4193.863	4242.645	0.988502	0.1068485
15	0.5	273.2187	276.2961	0.988861	0.1177344
	1	1058.172	1070.090	0.988862	0.1177344
	2	15872.57	16051.35	0.988862	0.1177344

5.4 Parâmetros da terceira onda de covid-19

A terceira onda de covid-19 está associada à variante Ômicron e foi menos letal que as ondas anteriores, principalmente devido ao avanço da vacinação na população (MARTINS-FILHO; SANTOS, 2022). A evolução do número reprodutivo efetivo nessa fase pode ser analisada na FIGURA 10, enquanto os parâmetros ξ e λ , as variáveis t e ϵ , a razão $\frac{\xi}{\lambda}$ e o desvio padrão na TABELA 4.

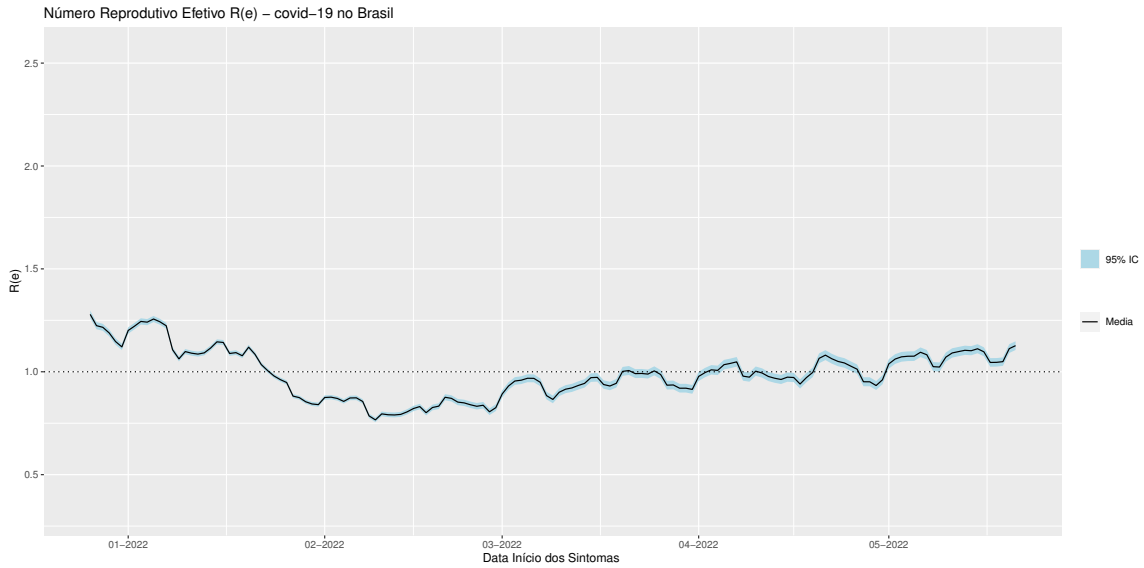


Figura 9 – Número Reprodutivo Efetivo (R_e) da covid-19 no Brasil durante a terceira onda de transmissão, de 26 de dezembro de 2021 a 21 de maio de 2022.

Tabela 4 – Vetores de parâmetros da terceira onda de covid-19, razão ξ/λ e desvio padrão.

t	ϵ	ξ	λ	ξ/λ	Desvio Padrão
1	0.5, 1 ou 2	287.8002	292.9697	0.9823548	0.05790591
	0.5	68.78504	72.84327	0.9442882	0.1851962
7	1	181.9881	192.7252	0.9442880	0.1851962
	2	1273.917	1349.076	0.9442885	0.1851962
15	0.5	86.00884	91.90690	0.9358257	0.1985849
	1	333.1108	355.9539	0.9358257	0.1985849
	2	4996.662	5339.308	0.9358258	0.1985849

5.5 Comparação de parâmetros

Além dos dados já apresentados anteriormente, nesta seção será indicada a variância condicional do modelo Gama Condicional Modificado para cada onda de transmissão da covid-19 no Brasil. A variância condicional do modelo é dada por (SCHREIBER, 2015):

$$\text{Var} [R_e(\tau)|R_e(\sigma)] = \frac{\xi}{\lambda^2}(\tau - \sigma)^\epsilon = \frac{\xi}{\lambda^2}t^\epsilon. \quad (5.1)$$

Conforme a equação 5.1, nota-se que a variância depende dos parâmetros ξ e λ e das variáveis t e ϵ . Porém, vale destacar que a dependência em relação a ϵ é apenas

aparente, pois a mudança nos valores de ξ e λ causada pela variação de ϵ anula os efeitos no cálculo final da variância. Para exemplificar, basta fixar $t = 7$ de modo que

$$\text{Var} [R_e(\tau)|R_e(\sigma)]_{(t=7, \epsilon=0.5, \xi=149.2549, \lambda=151.2830)} = \frac{149.2549}{151.2830^2} (7)^{0.5} = 0.017254 ;$$

$$\text{Var} [R_e(\tau)|R_e(\sigma)]_{(t=7, \epsilon=1, \xi=394.8913, \lambda=400.2572)} = \frac{394.8913}{400.2572^2} (7)^1 = 0.017254 ; e$$

$$\text{Var} [R_e(\tau)|R_e(\sigma)]_{(t=7, \epsilon=2, \xi=2764.239, \lambda=2801.800)} = \frac{2764.239}{2801.800^2} (7)^2 = 0.017254 .$$

Os valores das variáveis t e ϵ e dos parâmetros ξ e λ utilizados nos cálculos acima podem ser consultados na TABELA 1. Como o desvio padrão é a raiz quadrada da variância, naturalmente também não será influenciado por variações de ϵ .

Para facilitar a comparação dos dados, a TABELA 5 resume os valores de ξ , λ , ξ/λ e desvio padrão, dado $t = 7$ fixo. Além disso, especifica o período, a quantidade de dias a que se refere e a variância condicional $\text{Var} [R_e(\tau)|R_e(\sigma)]$.

Tabela 5 – Tabela comparativa dos vetores de parâmetros do modelo Gama Condicional Modificado para a covid-19 no Brasil, dada uma janela de $t = 7$ dias.

Período	Nº de dias	ξ	λ	ξ/λ	Desvio Padrão	Variância
Total	1092	394.891	400.257	0.98659	0.131355	0.017254
Primeira onda	258	171.527	178.519	0.96083	0.194102	0.037675
Segunda onda	412	599.123	606.092	0.98850	0.106848	0.011416
Terceira onda	146	181.988	192.725	0.94428	0.185196	0.034297

A partir da leitura da TABELA 5, percebe-se que a segunda onda de covid-19 apresenta os maiores valores individuais de ξ , λ e ξ/λ (neste último caso, apenas levemente superior à mesma razão do período total). A segunda onda de covid-19 concentra também as piores estatísticas de número absoluto de mortes, número de mortes por 100.000 habitantes e pico de mortes em relação às outras ondas (MOURA et al., 2022). Por outro lado, não parece haver relação entre os valores de ξ e λ e a quantidade de dias considerados, o que contribui positivamente para a comparabilidade de surtos com diferentes durações. Diante disso, é possível estabelecer que valores elevados de ξ e λ estão associados a fases críticas da propagação da doença, na qual os indicadores de letalidade atingem seus piores índices.

5.6 Capacidade preditiva do modelo

O objetivo desta seção é verificar se o modelo Gama Condicional Modificado é capaz de prever com eficiência o valor futuro do número reprodutivo efetivo. Para tanto, são simulados três cenários de recorte dos dados, nos quais os parâmetros ξ e λ são calculados em relação aos dados iniciais. Em seguida verifica-se a porcentagem dos dados finais capturada pelo modelo com a utilização dos parâmetros referidos.

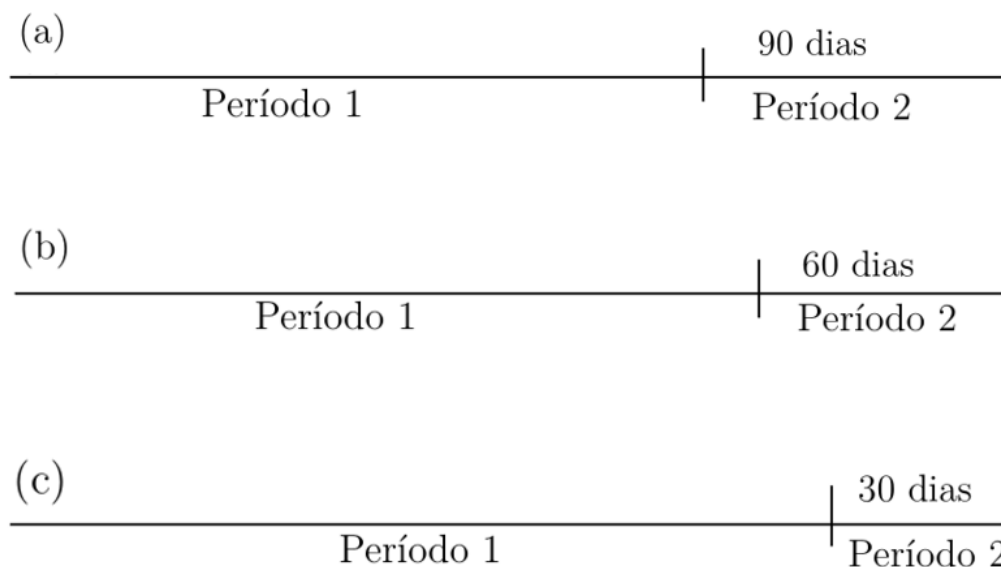


Figura 10 – Em (a), o período 1 vai de 25 de fevereiro de 2020 a 25 de setembro de 2022, totalizando 943 dias; o período 2, com 90 dias, vai de 26 de setembro de 2022 a 25 de dezembro de 2022. Em (b), o período 1 vai de 25 de fevereiro de 2020 a 25 de outubro de 2022, totalizando 973 dias; o período 2, com 60 dias, vai de 26 de outubro de 2022 a 25 de dezembro de 2022. Em (c), o período 1 vai de 25 de fevereiro de 2020 a 24 de novembro de 2022, totalizando 1003 dias; o período 2, com 30 dias, vai de 25 de novembro de 2022 a 25 de dezembro de 2022. Em todos os casos o primeiro dia do período 1 corresponde à data de confirmação do primeiro caso de covid-19 no Brasil. Por outro lado, o último dia de cada período 2 corresponde ao último dia de registro disponível na base de dados trabalhada.

5.6.1 Capacidade preditiva do modelo para os últimos 90 dias

Os dados dos 943 dias iniciais da pandemia fornecem os valores de $\xi = 547.2682$ e $\lambda = 554.5022$. Ao usar tais valores para a construção do gráfico de envelopes, 84.47% dos pontos $(R_e(\sigma), R_e(\tau))$ referentes aos últimos 90 dias são capturados. A FIGURA 11 abaixo ilustra o que acontece.

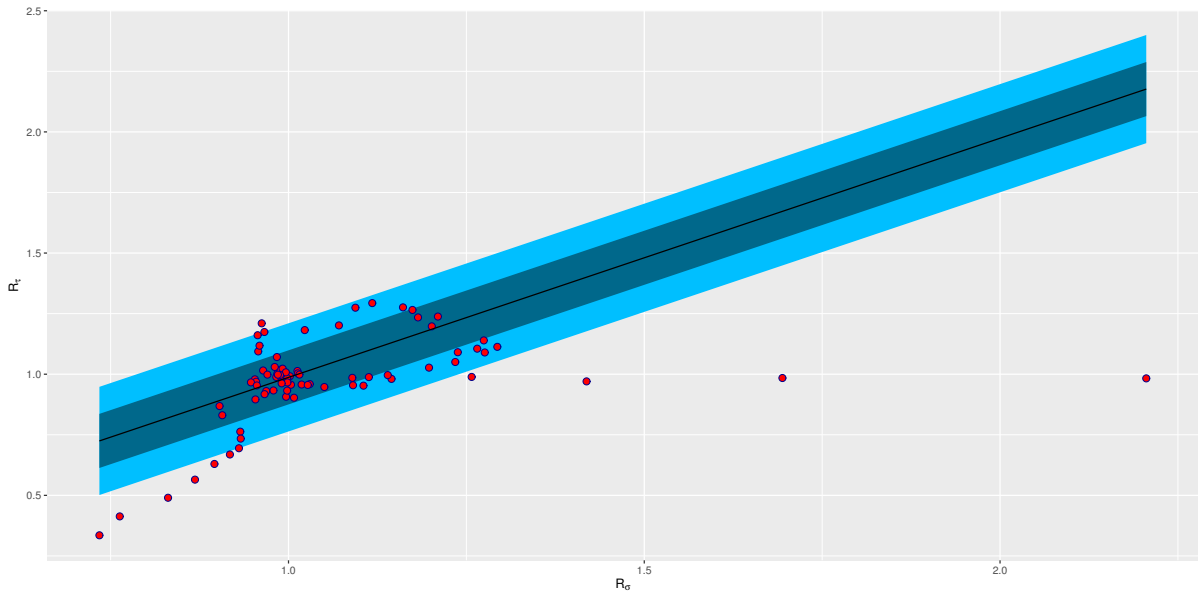


Figura 11 – Modelo para $t = 7$ dias e $\epsilon = 1$ usando os parâmetros $\xi = 547.2682$ e $\lambda = 554.5022$, calculados com base nos primeiros 943 dias da pandemia. O modelo foi capaz de capturar 84.47% dos dados. A reta preta é a esperança condicional $\mathbb{E}(R_e(\tau)|R_e(\sigma))$, o envelope azul escuro representa um desvio padrão e o envelope azul claro dois desvios padrão. Os pontos vermelhos são os pares $(R_e(d - 7), R_e(d))$ dos últimos 90 dias da base de dados e d representa o dia observado.

5.6.2 Capacidade preditiva do modelo para os últimos 60 dias

Os dados dos 973 dias iniciais da pandemia fornecem os valores de $\xi = 558.8363$ e $\lambda = 565.9619$. Ao usar tais valores para a construção do gráfico de envelopes, 78.72% dos pontos $(R_e(\sigma), R_e(\tau))$ referentes aos últimos 60 dias são capturados. A FIGURA 12 abaixo ilustra o que acontece.

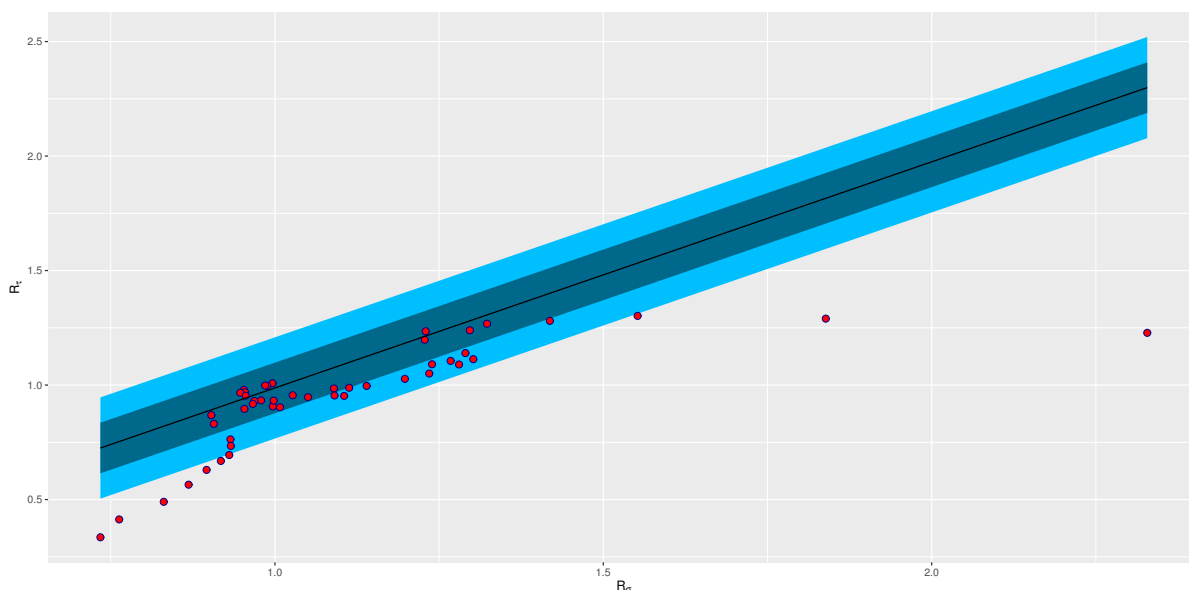


Figura 12 – Modelo para $t = 7$ dias e $\epsilon = 1$ usando os parâmetros $\xi = 558.8363$ e $\lambda = 565.9619$, calculados com base nos primeiros 973 dias da pandemia. O modelo foi capaz de capturar 78.72% dos dados. A reta preta é a esperança condicional $\mathbb{E}(R_e(\tau)|R_e(\sigma))$, o envelope azul escuro representa um desvio padrão e o envelope azul claro dois desvios padrão. Os pontos vermelhos são os pares $(R_e(d - 7), R_e(d))$ dos últimos 60 dias da base de dados e d representa o dia observado.

5.6.3 Capacidade preditiva do modelo para os últimos 30 dias

Os dados dos 1003 dias iniciais da pandemia fornecem os valores de $\xi = 550.9037$ e $\lambda = 557.5795$. Ao usar tais valores para a construção do gráfico de envelopes, 35.29% dos pontos $(R_e(\sigma), R_e(\tau))$ referentes aos últimos 30 dias são capturados. A FIGURA 13 abaixo ilustra o que acontece.

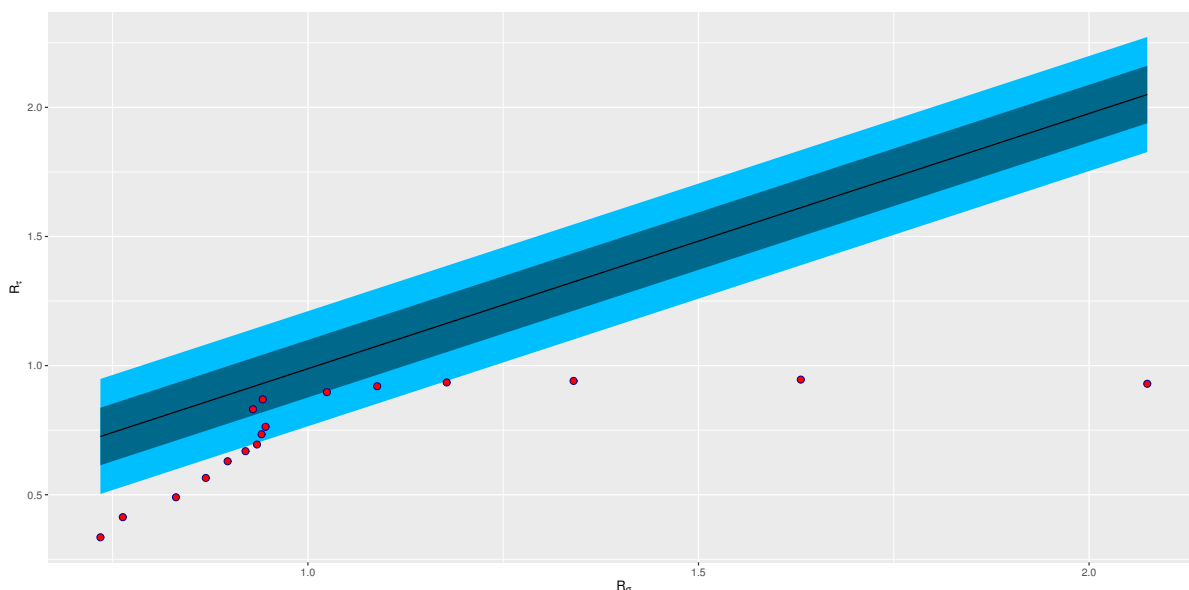


Figura 13 – Modelo para $t = 7$ dias e $\epsilon = 1$ usando os parâmetros $\xi = 550.9037$ e $\lambda = 557.5795$, calculados com base nos primeiros 1003 dias da pandemia. O modelo foi capaz de capturar 35.29% dos dados. A reta preta é a esperança condicional $\mathbb{E}(R_e(\tau)|R_e(\sigma))$, o envelope azul escuro representa um desvio padrão e o envelope azul claro dois desvios padrão. Os pontos vermelhos são os pares $(R_e(d - 7), R_e(d))$ dos últimos 30 dias da base de dados e d representa o dia observado.

5.6.4 Comparação da capacidade preditiva

Os dados obtidos nas seções 5.6.1, 5.6.2 e 5.6.3 estão relacionados na TABELA 6.

Tabela 6 – Tabela comparativa da capacidade preditiva do modelo Gama Condicional Modificado.

Referência para o cálculo de ξ e λ	Período a ser analisado	% predição
25/02/2020 a 25/09/2022 (943 dias)	26/09/2022 a 25/12/2022 (90 dias)	84.47%
25/02/2020 a 25/10/2022 (973 dias)	26/10/2022 a 25/12/2022 (60 dias)	78.72%
25/02/2020 a 24/11/2022 (1003 dias)	25/11/2022 a 25/12/2022 (30 dias)	35.29%

Com base na TABELA 6, é possível perceber que os valores de ξ e λ calculados para o período de 25 de fevereiro de 2020 a 24 de novembro de 2022 não informa com qualidade a dinâmica de R_e nos últimos 30 dias. A principal hipótese para explicar esse fato é que os 1003 dias levados em consideração para o cálculo dos parâmetros ξ e λ envolvem características muito distintas da transmissibilidade da doença em relação

ao período dos 30 dias finais. Dentre essas características, pode-se destacar o fato de que no início da pandemia R_e assumiu seus maiores valores, pois praticamente toda a população estava suscetível; o surgimento de variantes de preocupação capazes de gerar novas ondas de contaminação e o crescimento lento e gradual do número de pessoas vacinadas. Em contrapartida, todos esses fatores são mitigados no período dos 30 dias finais, fase em que a propagação da doença está aparentemente sob controle, com uma grande porcentagem da população vacinada e sem a circulação de novas variantes de preocupação responsáveis por novo surto. É provável que a capacidade preditiva do modelo para os 30 dias finais melhore ao se tomar uma quantidade de dias anteriores menor do que os 1003 ora analisados. A título de exemplo, ao estimar ξ e λ com base nos dados do período de 26 de agosto de 2022 a 24 de novembro de 2022 (90 dias), o gráfico de envelopes do modelo captura 88.23% dos dados de R_e entre 25 de novembro de 2022 e 25 de dezembro de 2022, apresentando uma melhora nítida em relação aos 35.29% apresentados na TABELA 6. Entretanto, esse é um recorte intuitivo. Sem levar em consideração fatores médico-biológicos específicos da propagação da doença, a tarefa se resume ao processamento aleatório de períodos. Dito de outra forma, apenas com a contribuição de especialistas na área da saúde será possível estabelecer uma janela ótima de recorte e previsão de dados.

6 Conclusões

Com o intuito de aperfeiçoar o modelo Gama Condicional Modificado proposto por [Schreiber \(2015\)](#), sugere-se a adoção de $\epsilon = 1$ e a mudança da variável aleatória descrita pelo modelo, que deixa de ser o Número Reprodutivo Básico (R_0) e passa a ser o Número Reprodutivo Efetivo (R_e).

6.1 Valor fixo para ϵ

Uma das variáveis do modelo Gama Condicional Modificado é ϵ . Ao aplicar o modelo aos dados, é possível observar na TABELA 1 que a variação de ϵ influencia os valores individuais dos parâmetros ξ e λ , mas não influencia a variância condicional $\text{Var}[(R_e(\tau)|R_e(\sigma))]$ nem a razão ξ/λ . Essa razão é indiretamente refletida nos gráficos de envelope, especificamente nas retas pretas que são a esperança condicional $\mathbb{E}(R_e(\tau)|R_e(\sigma))$, já os valores individuais dos parâmetros ξ e λ não o são. Por consequência, a variação de ϵ não determina a melhoria na quantidade de dados capturados pelo modelo. Dessa maneira, considera-se uma simplificação útil para o modelo a adoção de $\epsilon = 1$.

6.2 R_e como variável aleatória modelada

Os conceitos de Número Reprodutivo Básico (R_0) e Número Reprodutivo Efetivo (R_e) não se confundem. Enquanto aquele é o número esperado de novas infecções causadas por um agente infectado em uma população onde todos os indivíduos são suscetíveis ([GUERRA et al., 2017](#)), este é o número esperado de novas infecções causadas por um agente infectado em uma população onde parte dos indivíduos está suscetível e outra não ([GOSTIC et al., 2020](#)).

O fato de R_0 ser definido em função de uma população onde há apenas um indivíduo infectado e todos os demais são suscetíveis indica que essa medida caracteriza o início da propagação de uma doença. Esse valor de R_0 é estimado com base nos dados disponíveis e é fixo para cada doença ([ANDERSON; MAY, 1992](#)).

Por outro lado, a definição de R_e considera uma população na qual há vários indivíduos suscetíveis e vários infectados. Essa situação é a que bem representa a propagação de uma doença infecciosa ao longo do seu ciclo de existência. Trata-se, portanto, de uma medida dinâmica. Dessa forma, sugere-se que a função distribuição 4.1 e a função densidade de probabilidade 4.2 sejam dadas, respectivamente, por

$$P(R_e(\tau) \in S | R_e(\sigma) = x) = \frac{\left(\frac{x\lambda}{(\tau-\sigma)^\epsilon}\right) \left(\frac{x^2\xi}{(\tau-\sigma)^\epsilon}\right)}{\Gamma\left(\frac{x^2\xi}{(\tau-\sigma)^\epsilon}\right)} \int_{y \in S} z \left(\frac{x^2\xi}{(\tau-\sigma)^\epsilon}\right)^{-1} e^{-\left(\frac{x\lambda}{(\tau-\sigma)^\epsilon}\right)z} dz \quad (6.1)$$

e

$$p(R_e(\tau) \in S | R_e(\sigma) = x) = \frac{\left(\frac{x\lambda}{(\tau-\sigma)^\epsilon}\right) \left(\frac{x^2\xi}{(\tau-\sigma)^\epsilon}\right)}{\Gamma\left(\frac{x^2\xi}{(\tau-\sigma)^\epsilon}\right)} y \left(\frac{x^2\xi}{(\tau-\sigma)^\epsilon}\right)^{-1} e^{-\left(\frac{x\lambda}{(\tau-\sigma)^\epsilon}\right)y}. \quad (6.2)$$

6.3 Adequação do modelo Gama Condicional Modificado

Com base nos ensaios computacionais realizados nessa dissertação, é possível considerar o modelo Gama Condicional Modificado adequado para a modelagem do Número Reprodutivo Efetivo (R_e) de doenças. O modelo demonstra que a propagação de doenças em uma população atende às condições matemáticas para sua caracterização como um Processo de Difusão. Naturalmente, essa pesquisa não esgota o tratamento do assunto. Por exemplo, não foi determinada explicitamente a equação de Kolmogorov com base nas equações 3.1, 3.2 e 3.3. Em relação à utilização do modelo em políticas públicas de acompanhamento de doenças transmissíveis, há espaço para a valiosa contribuição de pesquisadores na área das Ciências Biológicas e da Saúde no que diz respeito à definição de recortes ótimos para a previsão de dados, conforme exposto na seção 5.6.4. Afinal, a transmissão de uma doença está relacionada às características intrínsecas tanto do organismo causador (vírus, bactéria, parasita e etc) quanto do organismo infectado, ambos objeto de estudo daquelas áreas.

Referências Bibliográficas

- ANDERSON, R. M.; MAY, R. M. **Infectious diseases of humans: dynamics and control**. [S.l.]: Oxford university press, 1992.
- AQUINO, E. M. et al. Medidas de distanciamento social no controle da pandemia de covid-19: potenciais impactos e desafios no brasil. **Ciência & Saúde Coletiva**, SciELO Public Health, v. 25, n. suppl 1, p. 2423–2446, 2020.
- BRAUER, F. The kermack–mckendrick epidemic model revisited. **Mathematical Biosciences**, v. 198, n. 2, p. 119–131, 2005. ISSN 0025-5564. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0025556405001331>>.
- BRAUER, F.; CASTILLO-CHAVEZ, C.; FENG, Z. **Mathematical models in epidemiology**. [S.l.]: Springer, 2019. v. 32.
- BRAUMANN, C. A. Introdução aos equações diferenciais estocásticas e aplicações. **Sociedade Portuguesa de Estatística, Ericeira, Portugal**, p. 190, 2005.
- CONTROL, C. for D.; CDC, P. **First Travel-related Case of 2019 Novel Coronavirus Detected in United States**. 2020. <<https://www.cdc.gov/media/releases/2020/p0121-novel-coronavirus-travel-case.html>>. Acessado em 11/01/2023.
- CORI, A. Epiestim: Estimate time varying reproduction numbers from epidemic curves. 2021. R package version 2.2-4. Disponível em: <<https://CRAN.R-project.org/package=EpiEstim>>.
- COSAR, B. et al. Sars-cov-2 mutations and their viral variants. **Cytokine & growth factor reviews**, Elsevier, v. 63, p. 10–22, 2022.
- COVID, E. E. Principais variantes do sars-cov-2 notificadas no brasil. **RBAC**, v. 53, n. 2, p. 109–116, 2021.
- CRANK, J. **The mathematics of diffusion**. [S.l.]: Oxford university press, 1979.
- FARIA, N. R. et al. Genomics and epidemiology of the p. 1 sars-cov-2 lineage in manaus, brazil. **Science**, American Association for the Advancement of Science, v. 372, n. 6544, p. 815–821, 2021.
- FIOCRUZ, F. O. C. **O que é uma Pandemia?** 2021. <<https://www.bio.fiocruz.br/index.php/br/noticias/1763-o-que-e-uma-pandemia>>. Acessado em 10/01/2023.
- GORDIS, L. **Epidemiology E-Book**. [S.l.]: Elsevier Health Sciences, 2013.

GOSTIC, K. M. et al. Practical considerations for measuring the effective reproductive number, r_t . **PLoS computational biology**, Public Library of Science San Francisco, CA USA, v. 16, n. 12, p. e1008409, 2020.

GUERRA, F. M. et al. The basic reproduction number (r_0) of measles: a systematic review. **The Lancet Infectious Diseases**, Elsevier, v. 17, n. 12, p. e420–e428, 2017.

IBE, O. **Markov processes for stochastic modeling**. [S.l.]: Newnes, 2013.

IBRAHIM, L. F. et al. Sars-cov-2 testing and outcomes in the first 30 days after the first case of covid-19 at an australian children's hospital. **Emergency Medicine Australasia**, Wiley Online Library, v. 32, n. 5, p. 801–808, 2020.

JOMBART, T. et al. incidence: Compute, handle, plot and model incidence of dated events. 2020. R package version 1.7.3. Disponível em: <<https://doi.org/10.5281/zenodo.2584018>>.

KERMACK, W. O.; MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. **Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character**, The Royal Society London, v. 115, n. 772, p. 700–721, 1927.

LAUXMANN, M. A.; SANTUCCI, N. E.; AUTRÁN-GÓMEZ, A. M. The sars-cov-2 coronavirus and the covid-19 outbreak. **International braz j urol**, SciELO Brasil, v. 46, p. 6–18, 2020.

LESCURE, F.-X. et al. Clinical and virological data of the first cases of covid-19 in europe: a case series. **The Lancet Infectious Diseases**, Elsevier, v. 20, n. 6, p. 697–706, 2020.

LIU, Y. et al. The contribution of pre-symptomatic infection to the transmission dynamics of covid-2019. **Wellcome Open Research**, The Wellcome Trust, v. 5, 2020.

LLOYD-SMITH, J. O. et al. Superspreading and the effect of individual variation on disease emergence. **Nature**, Nature Publishing Group, v. 438, n. 7066, p. 355–359, 2005.

LONE, S. A.; AHMAD, A. Covid-19 pandemic—an african perspective. **Emerging microbes & infections**, Taylor & Francis, v. 9, n. 1, p. 1300–1308, 2020.

MAGALHÃES, M. N. **Probabilidade e variáveis aleatórias**. [S.l.]: Edusp, 2006.

MARTINS-FILHO, P. R.; SANTOS, V. S. Relationship between living conditions in neighborhoods in the northeast brazil and covid-19 estimates during the third wave associated with the omicron variant. **Biodemography and Social Biology**, Taylor & Francis, v. 67, n. 3-4, p. 224–228, 2022.

MOURA, E. C. et al. Covid-19: temporal evolution and immunization in the three epidemiological waves, brazil, 2020–2022. **Revista de Saúde Pública**, SciELO Brasil, v. 56, 2022.

NISHIURA, H.; LINTON, N. M.; AKHMETZHANOV, A. R. Serial interval of novel coronavirus (2019-ncov) infections. **medRxiv**, 2020.

OKSENDAL, B. **Stochastic differential equations: an introduction with applications**. [S.l.]: Springer Science & Business Media, 2013.

OLIVEIRA, E. A. et al. Clinical outcomes of omicron variant (b. 1.1. 529) infection in children and adolescents hospitalized with covid-19 in brazil with observational data on the efficacy of the vaccines in adolescents. **The Pediatric Infectious Disease Journal**, LWW, p. 10–1097, 2022.

ORGANIZATION, P. A. H. Covid-19 modeling exercise, technical guide, methodology and parameters. 2020.

PAHO, P. A. H. O. **OMS anuncia nomenclaturas simples e fáceis de pronunciar para variantes de interesse e de preocupação do SARS-CoV-2**. 2021. <<https://www.paho.org/pt/noticias/1-6-2021-oms-anuncia-nomenclaturas-simples-e-faceis-pronunciar-para-variantes-interesse-e>>. Acessado em 10/01/2023.

PEAK, C. M. et al. Comparative impact of individual quarantine vs. active monitoring of contacts for the mitigation of covid-19: a modelling study. **MedRxiv**, Cold Spring Harbor Laboratory Preprints, 2020.

R Core Team. R: A language and environment for statistical computing. Vienna, Austria, 2021. Disponível em: <<https://www.R-project.org/>>.

ROCHA, D. I. C. Modelos matemáticos aplicados à epidemiologia. Faculdade de Economia da Universidade do Porto, 2012.

RStudio Team. Rstudio: Integrated development environment for r. Boston, MA, 2020. Disponível em: <<http://www.rstudio.com/>>.

SALSA, S. **Partial differential equations in action: from modelling to theory**. [S.l.]: Springer, 2016. v. 99.

SBIM, S. B. de I. **Seção sobre covid-19**. 2021. <<https://sbim.org.br/covid-19>>. Acessado em 10/01/2023.

SCHREIBER, K. P. Modelagem do número reprodutivo básico transformado de doenças infecciosas via a função de distribuição gama condicional modificada. 2015.

SOUZA, F. S. H. de et al. Second wave of covid-19 in brazil: younger at higher risk. **European journal of epidemiology**, Springer, v. 36, n. 4, p. 441–443, 2021.

VESPIGNANI, A. et al. Modelling covid-19. **Nature Reviews Physics**, Nature Publishing Group, v. 2, n. 6, p. 279–281, 2020.

WANG, T. et al. A four-compartment model for the COVID-19 infection—implications on infection kinetics, control measures, and lockdown exit strategies. **Precision Clinical Medicine**, v. 3, n. 2, p. 104–112, 05 2020. ISSN 2096-5303. Disponível em: <<https://doi.org/10.1093/pmedi/pbaa018>>.

WHO, W. H. O. **Tracking SARS-CoV-2 variants**. 2022. <<https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>>. Acessado em 10/01/2023.

ZHU, H.; WEI, L.; NIU, P. The novel coronavirus outbreak in wuhan, china. **Global health research and policy**, Springer, v. 5, n. 1, p. 1–3, 2020.

Apêndices

APÊNDICE A – R_0 no modelo compartimental SIS

Para exemplificar o uso dos modelos compartimentais no cálculo de R_0 , considere o modelo SIS cuja dinâmica é dada pela seguinte cadeia (markoviana):

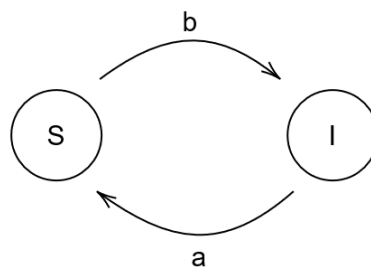


Figura 14 – Ilustração do modelo compartimental SI. b é a proporção da população que sai do estado suscetível (S) para infectado (I) e a é a proporção da população que sai do estado infectado (I) para suscetível (S).

Além de nomear os compartimentos, S é o número de indivíduos suscetíveis na população e I é o número de indivíduos infectados. Para alinhar as ordens de grandeza trabalhadas, toma-se $s = \frac{S}{N}$ e $i = \frac{I}{N}$, onde N é o número total de indivíduos na população. A variação da proporção de suscetíveis e infectados ao longo do tempo t pode ser representada pelo seguinte sistema de equações diferenciais:

$$\begin{cases} ds/dt = -bs + ai \\ di/dt = bs - ai. \end{cases} \quad (\text{A.1})$$

A equação A.1 pode ser reescrita na forma matricial como $X' = MX$, onde

$$X' = \begin{bmatrix} ds/dt \\ di/dt \end{bmatrix}, M = \begin{bmatrix} -b & a \\ b & -a \end{bmatrix} \text{ e } X = \begin{bmatrix} s \\ i \end{bmatrix}. \quad (\text{A.2})$$

Lema 1. *Seja $X' = MX$ com M uma matriz de ordem $m \times n$ e $X \in \mathbb{R}^n$. Para cada autovalor real λ de M e cada autovetor $E_\lambda = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$ pertencente ao autoespaço de λ , a*

função $X_\lambda = E_\lambda e^{\lambda t}$ é uma solução de A.2. Além disso, soluções formadas desta maneira a partir de autovalores distintos são linearmente independentes no espaço das funções reais contínuas $C(\mathbb{R})$.

Demonstração. Como E_λ é autovetor relativo a λ , temos que $ME_\lambda = \lambda E_\lambda$. Portanto,

$$MX_\lambda = \lambda E_\lambda e^{\lambda t} = \lambda X_\lambda.$$

Por outro lado,

$$X'_\lambda = \lambda E_\lambda e^{\lambda t} = \lambda X_\lambda.$$

Logo, $X'_\lambda = MX_\lambda$, como em A.2.

Agora, sejam $\lambda_1, \dots, \lambda_n$ autovalores reais distintos de M com respectivos autovetores $E_{\lambda_1}, \dots, E_{\lambda_n}$ e suponha que $c_1 E_{\lambda_1} e^{\lambda_1 t} + \dots + c_n E_{\lambda_n} e^{\lambda_n t} = \vec{0}$. Então, fazendo $t = 0$, temos $c_1 E_{\lambda_1} + \dots + c_n E_{\lambda_n} = 0$ e a independência linear dos E_{λ_i} ($i = 1, 2, \dots, n$) em \mathbb{R}^n implica que $c_1 = c_2 = \dots = c_n = 0$. Logo, as funções $E_{\lambda_i} e^{\lambda_i t}$ são linearmente independentes. \square

Corolário 1. Se M tem n autovalores reais distintos $\lambda_1, \dots, \lambda_n$ e $E_{\lambda_1}, \dots, E_{\lambda_n}$ são autovetores relativos a cada autovalor, então a solução geral do sistema normal de primeira ordem $X' = MX$ é

$$X(t) = c_1 E_{\lambda_1} e^{\lambda_1 t} + \dots + c_n E_{\lambda_n} e^{\lambda_n t},$$

em que c_1, \dots, c_n são constantes arbitrárias.

No caso específico do modelo SI, em que $X = \begin{bmatrix} s \\ i \end{bmatrix}$ e $M = \begin{bmatrix} -b & a \\ b & -a \end{bmatrix}$ temos primeiramente que os autovalores são obtidos a partir de

$$\det \begin{bmatrix} -b - \lambda & a \\ b & -a - \lambda \end{bmatrix} = 0.$$

Dessa forma, $\lambda_1 = 0$ (caso em que a matriz M é não inversível) e $\lambda_2 = -(a + b)$. O autoespaço dado por λ_1 é $S_{\lambda_1} = \left\{ (x, y) \in \mathbb{R}^2 : y = \frac{b}{a}x \right\} = \left\langle \left(1, \frac{b}{a} \right) \right\rangle$, que implica

$E_{\lambda_1} = \begin{bmatrix} 1 \\ \frac{b}{a} \end{bmatrix}$. O autoespaço dado por λ_2 é $S_{\lambda_2} = \{(x, y) \in \mathbb{R}^2 : x = -y\} = \langle (1, -1) \rangle$,

que implica $E_{\lambda_2} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

Assim, a solução geral dos sistemas A.1 ou A.2 é dada por

$$\begin{aligned}
 \begin{bmatrix} S(t) \\ I(t) \end{bmatrix} &= c_1 E_{\lambda_1} e^{\lambda_1 t} + c_2 E_{\lambda_2} e^{\lambda_2 t} \\
 &= c_1 E_{\lambda_1} e^{0t} + c_2 E_{\lambda_2} e^{-(a+b)t} \\
 &= c_1 \begin{bmatrix} 1 \\ \frac{b}{a} \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} e^{-(a+b)t} \\
 &= \begin{bmatrix} c_1 + c_2 e^{-(a+b)t} \\ c_1 \frac{b}{a} - c_2 e^{-(a+b)t} \end{bmatrix}
 \end{aligned} \tag{A.3}$$

Faz sentido supor $S(0)=1$ e $I(0)=0$, pois o instante inicial $t = 0$ de uma nova doença contagiosa é marcado por uma população onde todos são suscetíveis e ninguém foi infectado ainda. Usando essas condições iniciais em A.3, temos que

$$c_1 = \frac{a}{a+b} \quad \text{e} \quad c_2 = 1 - \frac{a}{a+b} = \frac{b}{a+b}.$$

Aplicando os valores encontrados para c_1 e c_2 em A.3, é possível descrever $S(t)$ e $I(t)$ como

$$S(t) = \frac{a + be^{-(a+b)t}}{a+b} \quad \text{e} \quad I(t) = \frac{b(1 - e^{-(a+b)t})}{a+b}.$$

Tomando um valor t_k suficientemente grande, com $k \in \mathbb{N}$, parte dos termos acima torna-se insignificante e é possível considerar as seguintes aproximações:

$$S(t_k) \approx \frac{a}{a+b} \quad \text{e} \quad I(t_k) \approx \frac{b}{a+b}.$$

A aproximação também é válida para todo $t > t_k$. A dinâmica da evolução de $S(t)$ e $I(t)$ pode ser verificada na FIGURA 15.

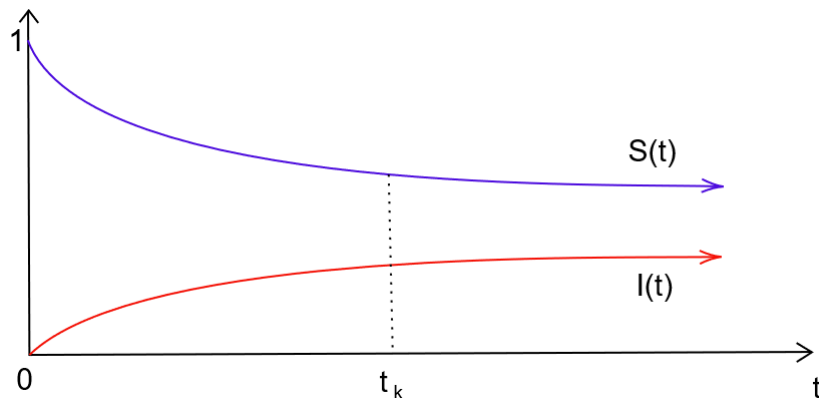


Figura 15 – Gráficos de $S(t)$ e $I(t)$ para $a > b$ (taxa de infectado para suscetível maior que taxa de contaminação).

Pela simplicidade do modelo, a variável aleatória de interesse poderia ser $R = \frac{b}{a+b}$, entendendo-se $b = b(t)$ e $a = a(t)$ variáveis aleatórias positivas. Note-se, por fim, que enquanto a e b são taxas, R é adimensional.

APÊNDICE B – Código em R para a construção dos gráficos de envelope

Com o intuito de tornar as simulações reproduzíveis, transcreve-se abaixo o código utilizado no software R para a construção dos gráficos de envelope, com os comentários pertinentes.

```
#Carregando os dados
dados<-read.csv2('dados-abertos.csv',stringsAsFactors = FALSE)

#Atribuindo a coluna de interesse
#"data dos primeiros sintomas" a uma variável
#também em formato de data:
DPS<- dmy(dados$dataPrimeirosSintomas, "%d-%m-%Y")

#Atribuindo a quantidade de ocorrências por data
#com o uso do pacote incidence
i<- incidence(DPS)

#Estimando o valor de R com o pacote EpiEstim
#e os parâmetros mean_si e std_si definidos pela OPAS
n_reprodutivo<- estimate_R(i,
                           method="parametric_si",
                           config=make_config(list(
                               mean_si= 4.8,
                               std_si= 2.3)))

#visualizando o gráfico do R
R_estimado<-plot(n_reprodutivo, legend = TRUE, 'R')+
  ggtitle("Número Reprodutivo Efetivo R(e)")+
  scale_y_continuous(name="R(e)")+
```

```

scale_colour_manual("", values="black", labels="Media")+
scale_fill_manual("", values="lightblue", labels="95% IC")+
scale_x_date(name="Data Inicio dos Sintomas ", date_labels = "%m-%Y",
date_breaks = "3 month")
R_estimado

#atribuindo o valor médio de R ao objeto Rest
Rest<- n_reprodutivo$R$`Mean(R)`

#SIMULAÇÕES PARA t E \epsilon VARIANDO:
t<-15
e<-1

#base de dados (varia de acordo com a janela adotada):
x<-Rest[1:(length(Rest)-t)]
y<-Rest[(1+t):length(Rest)]

#Para verificar se os vetores de dados x e y
#criados acima têm o mesmo tamanho, usa-se o comando abaixo,
#o qual deve retornar TRUE.
#Caso esses vetores possuam tamanhos diferentes, não será possível
#estabelecer os pares (R(\sigma),R(\tau))
length(x)==length(y)

#Calculando a função k(x,y,t), que será necessária para
#estimar os parâmetros \psi e \lambda
fator_r<- sum(t**e/6*x**2)

#Criando o fator S(x^2)/S(xy), que também será necessário
#para estimar os parâmetros \psi e \lambda
S<- (sum((x**2)/t**e))/(sum((x*y)/(t**e)))

#criando a função k(x,y,t), último cálculo
#auxiliar para a determinação dos parâmetros \psi e \lambda
fator_k<- sum(((x**2)/(t**e))*(log(1/x)+log(S)+log(y)))

```

```

#estimando o parâmetro \qsi:
n<-length(x)
fator_k
fator_r
qsi<- (-n-sqrt(n**2 - 8*fator_k*fator_r))/(4*fator_k)

#estimando \lâmbda:
lbda<- qsi*S

#definindo a esperança condicional
EC<- qsi*x/lbda

#completando a base de dados com a esperança condicional
bd<- data.frame(x,y,EC)

#calculando o desvio padrão para o gráfico de envelope
variancia<-(qsi/lbda**2)*t**e
dp<-sqrt(variancia)

#Criando um contador para a quantidade de pares
#(R(\sigma),R(\tau)) fora dos envelopes em até
#dois desvios padrão
count<-0
for (i in 1:length(y)){
  if (y[i]>EC[i]+2*dp |
      y[i]<EC[i]-2*dp){
    count=count+1
  }
}
1-count/length(y)

#criando e exibindo o gráfico de envelope
ggplot(bd, aes(x=x,y=y))+
  geom_ribbon(aes(ymin=EC-2*dp,ymax=EC+2*dp),fill='deepskyblue')+

```

```
geom_ribbon(aes(ymin=EC-dp,ymax=EC+dp),fill='deepskyblue4')+  
geom_point(col='blue4',size=2.5,pch=21,bg='red')+  
geom_line(aes(x=x,y=EC))+  
labs(x=expression("R"[sigma]), y=expression("R"[tau]))
```