

# Relatório 2º Exercício Programa MAC0209 - IME - USP

Daniel Silva Lopes da Costa - N°USP 11302720

16 de outubro de 2020

## Resumo

Por meio de ferramentas computacionais e técnicas estatísticas procurou-se prever o comportamento do Covid-19 em uma determinada população, foi analisada como essas curvas se desenvolvem em relação ao número total de infectados e o número diário de novos casos, e como esses índices se relacionam com os parâmetros utilizados. Como resultado foi implementado dois modelos dinâmicos, pelo método de Euler que simulam situações reais, um com parâmetros fixos e outro com cinco populações distintas com parâmetros aleatorizados, os quais se transformam a cada passagem. Tais informações são de suma importância para entender o comportamento do vírus e gerar estratégias de combate.

## Sumário

<b>1</b>	<b>Introdução</b>	<b>2</b>
1.1	Motivação . . . . .	2
1.2	Objetivos . . . . .	2
<b>2</b>	<b>Materiais e métodos</b>	<b>3</b>
<b>3</b>	<b>Resultados Experimentais</b>	<b>5</b>
3.1	Parte 1 . . . . .	5
3.1.1	Análise gráfica . . . . .	6
3.1.2	Simulação dinâmica . . . . .	7
3.2	Parte 2 . . . . .	7
3.2.1	Análise gráfica . . . . .	8
3.2.2	Simulação dinâmica . . . . .	9
3.3	Parte adicional . . . . .	10
3.3.1	Análise gráfica . . . . .	11
3.3.2	Simulação dinâmica . . . . .	12
<b>4</b>	<b>Conclusão</b>	<b>13</b>

# 1 Introdução

O COVID-19 é uma doença infecciosa causada pelo coronavírus da síndrome respiratória aguda grave 2. Os sintomas mais comuns são febre, tosse seca, cansaço e perda do paladar e olfato. Por ainda não existir um tratamento específico - sem remédios ou vacina com eficiência comprovada - as principais formas de combate adotadas foram o distanciamento social, medidas de higiene e uso de máscaras, o que atingiu quase a totalidade da população do globo de alguma forma, atualmente (14/10/20) temos aproximadamente 38.317.777 casos confirmados e mais de um milhão de mortos.

Diante desses índices alarmantes, uma ferramenta de suma importância são as previsões estatísticas, as quais ajudam a entender o comportamento do vírus e fazer previsões sobre o número de infectados. Foi utilizada como metodologia para desenvolvimento do trabalho, uma parte de pesquisa sobre o Modelo matemático e artigo base, em seguida foi implementado os algoritmos e desenvolvidos os gráficos e animações no Jupyter Notebook, em seguida foi escrito o artigo com os métodos e resultados encontrados, e por fim uma etapa de revisão sobre os resultados encontrados. Com isso chegou-se a vários resultados que simulam a propagação do vírus em uma população, o que foi analisado por meio de tabelas, gráficos e animações, permitindo uma melhor compreensão do comportamento do vírus.

## 1.1 Motivação

Diante dessa nova ordem mundial, alarmante e urgente, uma ferramenta de extrema importância são os modelos estatísticos que visam entender o comportamento do vírus para que assim os governos e a sociedade possam agir da melhor maneira possível no combate à doença e na manutenção do bem-estar social. Portanto, a motivação para essa análise é procurar conhecer mais sobre esse tipo de modelagem de extrema importância para entender o comportamento do vírus e combatê-lo, ganhando experiência prática por meio de simulações que se aproximem da realidade.

## 1.2 Objetivos

O objetivo deste trabalho é implementar um desses modelos estatísticos sobre a curva de contágio do COVID-19 de maneira simplificada, a fim de identificar o comportamento das curvas da doença e entender como os modelos matemáticos podem ser utilizados em situações reais, o modelo que será utilizado encontra-se no artigo do pesquisador Giorgio SONNINO da Universidade Livre de Bruxelas. Serão implementados gráficos e animações a fim de entender a natureza do contágio. Tal modelo matemático será implementado sobre duas perspectivas, em primeiro para uma única população com parâmetros bem definidos. Em seguida, será implementado para um conjunto maior de populações, com parâmetros aleatórios. Outro objetivo, é analisar o modelo adicional SIR que foi construído de forma semelhante ao caso anterior, a fim de estabelecer comparações e entender mais sobre o funcionamento desse tipo de modelagem.

## 2 Materiais e métodos

### Organização

Foram utilizadas diversas ferramentas para que pudéssemos analisar o modelo proposto. Nesse sentido, a organização de todo o trabalho foi de suma importância para sua finalização, para tanto foi utilizado a ferramenta Mapa de Gantt, a qual permite a organização de um projeto com prazo e tarefas específicas, divididas entre os integrantes do projeto. Como esse trata-se um projeto individuo seu principal uso foi para a manutenção dos prazos.

Na Figura 1, temos a tabela e o gráfico de Gantt, o projeto foi desenvolvido durante o mês de junho principalmente, houve um pequeno erro de entendimento quanto ao artigo que foi implementado junto com o Jupyter Notebook, por isso essa tarefa consta como ainda incompleta. O desenvolvimento do Trabalho se desenvolveu em três partes: pesquisa, implementação do Notebook, escrita do artigo.

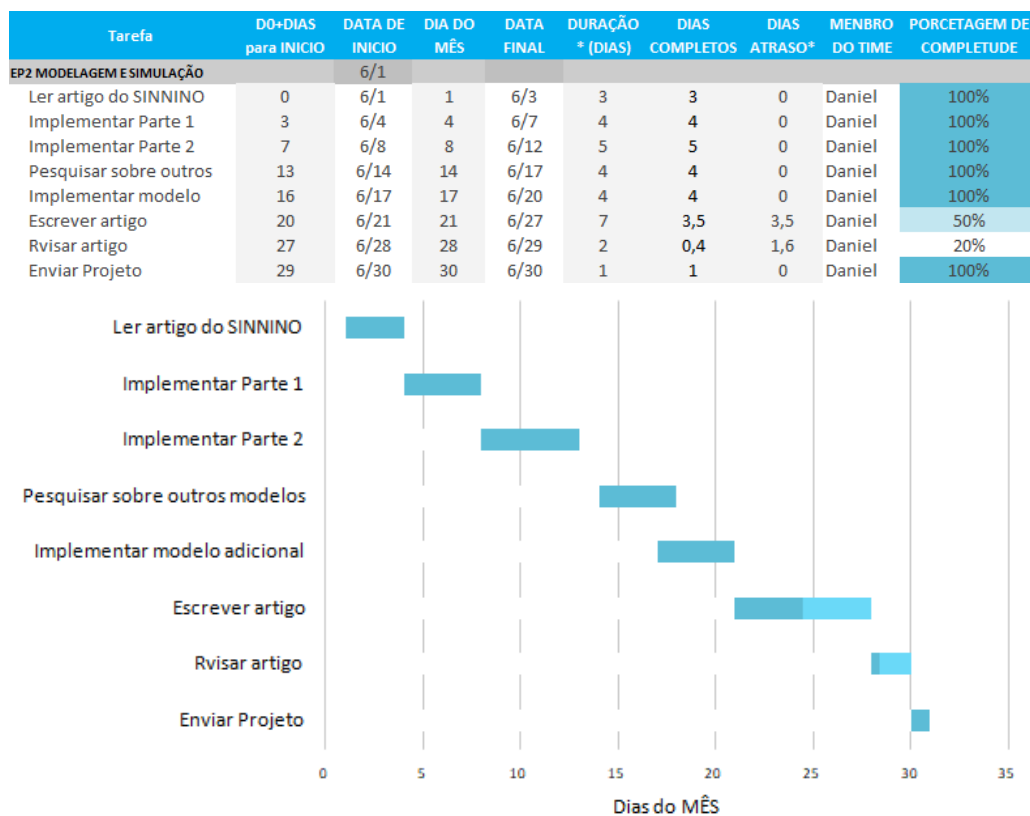


Figura 1: Gantt chart adotado para o desenvolvimento do projeto.

## Jupyter Notebook

Além disso como ferramenta principal para implementação do projeto foi utilizado o Jupyter Notebook, um ambiente computacional web para criação de documentos, que permite a implementação de programas na linguagem python, possui bibliotecas gráficas e possibilita a intersecção entre código e texto(Markdown), o que permitiu uma melhor documentação de todos os processos.

Sobre as bibliotecas do python foram utilizadas:

- **Math** - Para cálculo de funções trigonométricas e exponenciais;
- **Matplotlib.pyplot e Matplotlib** - A qual possibilitou a geração de gráficos e animações;
- **Random** - Para a geração de número aleatórios.

## Modelo matemático

Como afirmado o modelo que será usado encontra-se no artigo do pesquisador Giorgio SONNINO da Universidade Livre de Bruxelas. A equação diferencial que é:

$$\frac{dN}{dt} = \alpha(1 - \frac{N}{\eta t^2})N - (\frac{2\lambda t^2 - 1}{t} - \frac{\lambda t_0^2}{t} e^{-\lambda(t-t_0)^2})N$$

Como base nisso temos que:

- $N$  é número total de casos, ou seja a soma de todos os novos casos diários, o menor valor que  $n$  pode assumir é zero que seria quando a população estivesse completamente livre da doença.
- $dN$  será representado no algoritmo pelo variável  $V$ , representada a velocidade de contágio, ou seja o número de novos casos diários, Tal índice pode assumir valores negativos se o número de pessoas curadas ultrapassar o número de novos casos.
- $dt$  é a variação do tempo no qual são contados os novos casos, seguindo os exemplos reais e teóricos esse índice será atualizado uma vez por dia, portanto  $dt=1$  dia no algoritmo.
- $t_0$  é tempo inicial para que medidas de restrição sejam impostas, como uso de máscaras, distanciamento social ou até mesmo lockdown.
- $t$  será o tempo em dias desde o primeiro caso da doença foi registrado.
- $\alpha$  que será representado no algoritmo pela variável  $a$ , é fator de crescimento.
- $\lambda$  é o inverso do parâmetro  $\sigma$  o qual depende do vírus e das condições de restrições, tal parâmetro será implementado no algoritmo como a variável  $h$ .

- $\eta$  é um parâmetro diretamente proporcional ao dobro do parâmetro  $A$  que é subjetivo e depende de cada população e é inversamente proporcional ao  $t_0$ , será representado no algoritmo pela variável  $m$ .

Para implementar o algoritmo foi usado o método de euler, onde o próximo número é gerado com base no caso anterior, no modelo acima foi usado a seguinte equação:

$$N_{n+1} = N_n + V \cdot \Delta t$$

A implementação dessa simulação foi dividida em dois casos, primeiramente será analisada uma ilha isolada com parâmetros indicados pelo programador, para que se possa analisar a formato da curva e o efeito inicial de cada um dos parâmetros na implementação do modelo, em seguida será desenvolvido o mesmo modelo para cinco ilhas com parâmetros gerados de maneira pseudo-aleatória pelo computador, dentro de uma margem estabelecida, para que assim possa-se comparar diferentes formatos de curvas e possa-se identificar de maneira ainda mais clara como cada um dos parâmetros afeta os índices, por fim será feita uma comparação entre esse modelo e o modelo epidemiológico SIR.

## 3 Resultados Experimentais

### 3.1 Parte 1

Para a modelagem de uma ilha isolada foram utilizados parâmetros encontrados no artigo base, tais parâmetros fazem referência a curva de contágio da Itália e foram usados com o intuito de testar a corretude do cálculo do algoritmo com base no modelo de Euler. Para tanto adotou-se:

- $\alpha = 0.261 \text{ dias}^{-1}$ ;
- $t_0 = 70.6 \text{ dias}$ ;
- $\lambda = 0.00185 \text{ dias}^{-2}$ ;
- $\eta = 61.10 \text{ dias}^{-2}$

Os quais foram o valores teóricos adotados para a Itália pelos pesquisadores, com tais parâmetros o algoritmo implementado no Notebook conseguiu chegar a uma curva praticamente idêntica a curva teórica mostrada no artigo, o que comprova que o algoritmo está sendo implementado corretamente. Além disso adotou-se como o primeiro dia, aquele com o primeiro caso registrado e foram datadas o número de casos diários e número de casos total em dois vetores, vale ressaltar também que adotou-se o dia 73 como parada dos cálculo por ser o dia adotado no artigo.

### 3.1.1 Análise gráfica

Pelo gráfico sobre o número total de casos, representado na Figura 2, é possível perceber a autenticidade da simulação uma vez que no artigo base para esse exercício existe esse gráfico sobre a quantidade de casos total e é perceptível que os dois gráficos apresentam valores muito próximos o que demonstra a veracidade da simulação implementada. Além disso, vale pontuar que o número de novos caso diário, na Figura 3, se aproxima muito dos valores reais uma vez que no pico da epidemia na Itália tivemos um pouco mais de 6000 novos caso por dia.

É possível identificar ainda, que o número de novos casos diários começou a cair a partir do 70º, tal informação é de suma importância pois permite detectar o pico de contágio e mostra que as ferramentas de combate a doenças estão surtindo efeito. Sobre as curvas é possível perceber seu carácter dinâmico, com a subida exponencial do números de casos no primeiro gráficos diretamente relacionada ao aumento do número que casos diários.

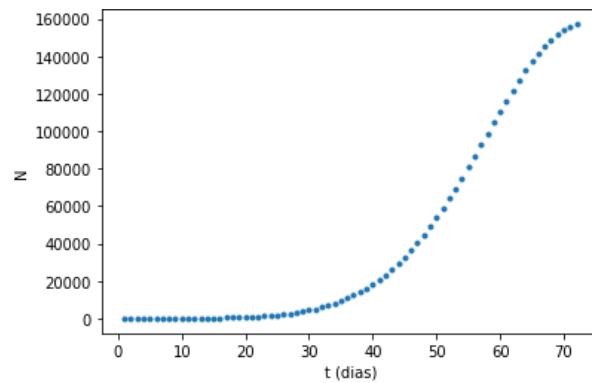


Figura 2: Número de infectados Itália.

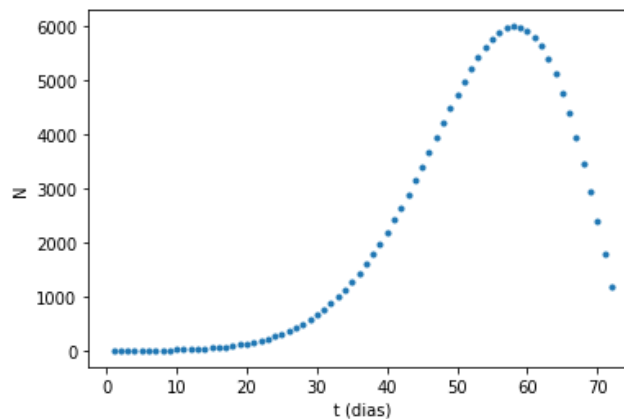


Figura 3: Número de novos casos Itália.

### 3.1.2 Simulação dinâmica

Por meio da simulação dinâmica do número total de casos pelo tempo é possível perceber a velocidade exponencial e chocante que o número de casos aumenta, atingindo patamares extremos em pouco tempo, contaminando mais de 150 mil pessoas em menos de três meses. É possível visualizar a animação completa no Jupiter Notebook.

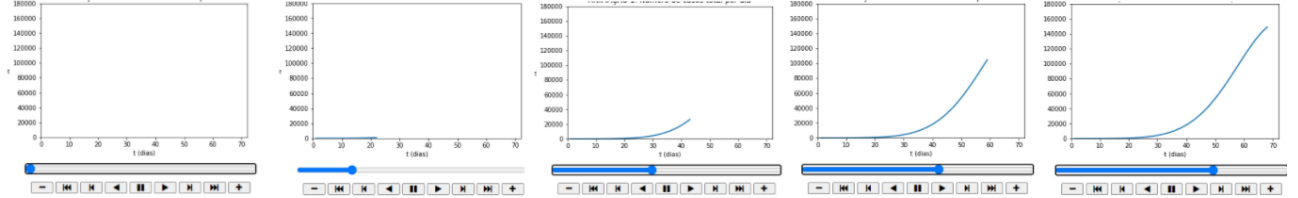


Figura 4: Número de casos total por dia.

## 3.2 Parte 2

Na segunda parte desta análise foi a implementação da simulação para cinco ilhas distintas, onde os parâmetros foram gerados de maneira pseudo-aleatória pelo computador. Nesse sentido temos que:

- $0 < \alpha < 1 \text{ dias}^{-1}$  o qual foi gerado pela função `random()` que seleciona um valor float entre 0 e 1;
- $10 < t_0 < 100 \text{ dias}$  tal domínio foi estabelecido após sucessivos testes e análise de valores teóricos já utilizados, foi usado a função `randint(x,y)` que devolve um inteiro entre x e y;
- $0.00185 < \lambda < 0.00285 \text{ dias}^{-2}$  tal parâmetro possui um limitado domínio pois ele pode provocar anomalias nas curvas e dentro dessa linha de valores temos curvas bem comportadas e que atendem as necessidades para análise, tais números são gerados pela função `uniform(w,z)` que devolve valores do tipo float entre w e z;
- $50 < \eta < 70 \text{ dias}^{-2}$  para esse parâmetro foi utilizado novamente a função `randint`, e o domínio foi estabelecido com base em testes e análise de valores utilizados no artigo base.

Vale pontuar que essa estratégia de gerar curvas como base em parâmetros aleatórios é base de métodos de suma importância como o Método de Monte Carlos que permite chegar aos parâmetros reais de uma determinada população.

Por serem cinco ilhas foram utilizados vetores de estados, onde o vetor N, possui cinco lista e cada uma se refere ao número total de cada uma das ilhas e de maneira análoga o vetor V, que contempla os números de novos casos de cada uma das cinco situações. O tempo máximo de análise é de 300 dias, tal valor foi escolhido com base em sucessivos testes, que comprovaram

que a maioria das curvas realizavam todo o seu ciclo em um tempo inferior a trezentos dias, além disso dependendo da natureza de cada curva se todas as ilhas terminarem o ciclo de contaminação antes de trezentos, o algoritmo tem um mecanismo que para o tempo antes de se completar os 300 dias o que permite um melhor enquadramento de gráficos e animações.

### 3.2.1 Análise gráfica

Temos aqui as curvas das cinco ilhas comparadas, é possível perceber a variedade de formatos dependendo dos parâmetros, temos ilhas em que mais 500 mil pessoas foram infectadas enquanto em outra menos de mil pessoas, o que realmente deixa claro a importância do meio e das medidas de prevenção na contenção da doença e como esse tipo de medida influencia para que a doença cresça de maneira mais exponencial ou não. Por ser um modelo com base em parâmetros aleatórios podemos realizar milhares de simulações diferentes e comparar e analisar diferentes gráficos a cada implementação. O resultado de uma dessas simulações está expresso nas Figuras 5 e 6.

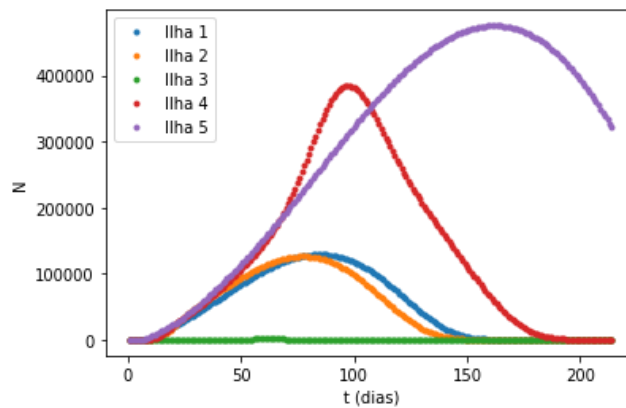


Figura 5: Número de infectados nas Ilhas.



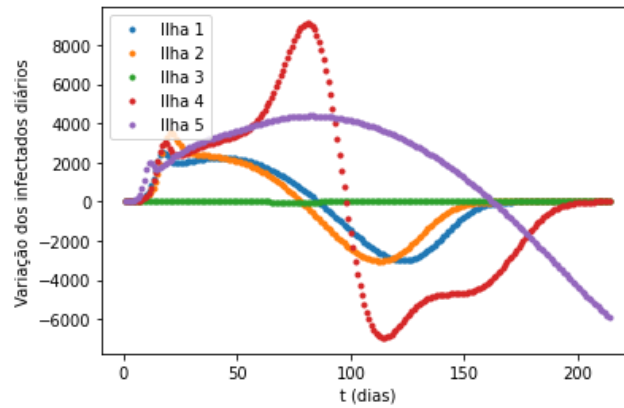


Figura 6: Número de novos infectados nas Ilhas.

A seguir temos a comparação entre os números totais de casos das ilhas comparadas uma a uma, o que resultou em alguns gráficos semelhantes às asas de borboletas das equações de Lorenz. Como os fatores podem ser bem distintos alguns gráficos apresentaram um aspecto muito deformado, justamente pela distinção entre os fatores que pode provocar a contaminação de menos de 5 mil pessoas ou a contaminação de mais de 400 mil. Por meio da análise desses gráficos é possível perceber como diferentes localidades podem apresentar distinções mesmo estando no mesmo período da doença, tudo depende das características de cada população e da forma como as medidas de combate são implementadas.

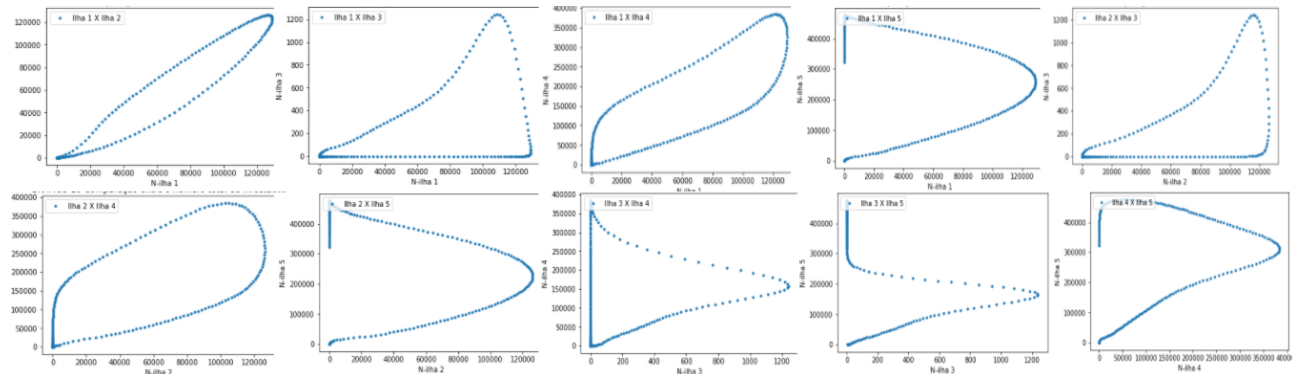


Figura 7: Comparação entre as ilhas, duas a duas.

### 3.2.2 Simulação dinâmica

Por meio da animação dinâmica é possível ver de maneira análoga ao caso 1, a curva com o número total de pessoas infectadas, o que permite ver de maneira clara o avanço da doença

em função do tempo assim como o decrescimento desta. Por serem cinco ilhas e os parâmetro podem ser alterados com frequência o algoritmo sempre vai fazer a animação dinâmica da ilha com maior número de casos a fim de se ter o melhor enquadramento com tempo, além disso vale destacar que como é possível girar o algoritmo inúmeras vezes podemos ter milhares de simulações para parâmetros e casos distintos. A animação está disponível no arquivo Jupyter Notebook do projeto.

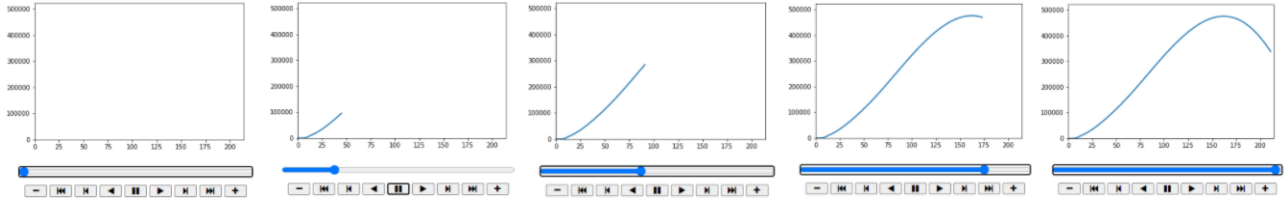


Figura 8: Número de casos total por dia.

### 3.3 Parte adicional

O objetivo dessa última parte é implementar outro modelo e compará-lo com o implementado acima. O modelo escolhido foi o modelo SIR. O modelo epidemiológico SIR foi desenvolvido em 1927, é atualmente um dos modelos mais utilizados para representação de doenças infecciosas. O modelo divide a população em três compartimentos ou classes, cada compartimento indica o estado atual da doença nos indivíduos da população.

- $S$ =Suscetíveis: pessoas que podem se contaminar;
- $I$ =Infectados: pessoas que estão contaminadas;
- $R$ =Recuperados: pessoas que já se recuperaram, estão imunes e não transmitem mais o vírus - No caso do novo coronavírus ainda não se sabem se as pessoas recuperadas estão realmente imunes, mas existem evidências que comprovam nesse sentido e iremos adotar essa premissa como verdadeira, do contrário haveria apenas dois grupos os suscetíveis e os infectados;
- $N$ =Número de indivíduos de determinada população.

Uma diferença clara entre esse modelo e o elaborado por Sonnino, é que esse modelo leva em consideração a quantidade total de indivíduos, no modelo usado acima a número de indivíduos da população deve estar obviamente ligada aos parâmetros  $\sigma$  e  $A$  que foram utilizados, mas nesse modelo o tamanho da população é um valor conhecido e importante para a implementação.

Temos as seguinte equações básicas:

$$\frac{dS}{dt} = -\frac{\beta IS}{N}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

Além disso temos também que:

$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0$$

$$S(t) + I(t) + R(t) = N$$

Outro conceito muito importante e que se relaciona diretamente com os parâmetros apresentados é a razão de reprodutividade basal  $R_0$ , o qual representa o número médio de infecções secundárias quando um indivíduo infectado é introduzido em uma população de hospedeiros completamente suscetíveis. Nesse modelo temos que:

$$R_0 = \frac{\beta}{\gamma}$$

- $\gamma$  é um parâmetro inversamente proporcional ao tempo que a pessoa fica infectada pelo vírus, no caso do covid-20 não se sabe ao certo quantos dias elas permanecem com o vírus, consideraremos portanto que o tempo de infecção é de 20 dias e portanto  $\frac{1}{20 \cdot 10} = \gamma = 0,005$  e será representado no algoritmo pela variável  $\gamma$ .
- $\beta$  é um parâmetro que depende de  $\gamma$  e de  $R_0$  portanto, se considerarmos o  $R_0 = 3$  -índice real de várias cidades brasileiras no início da pandemia - temos que  $\beta = 3 \cdot 0,005 = 0,015$

Com essas informações em mãos já podemos implementar nosso modelo, vamos considerar uma cidade com 100 habitantes e analisar a curva de crescimento do número de infectados durante um ano.

### 3.3.1 Análise gráfica

Um dos principais problemas desse modelo é que ele não considera as medidas de distanciamento social, apesar de ser possível implementar mais parâmetros a fim de aproximá-lo da realidade para fins de descrição teórica esse modelo é mais do que suficiente para que possa-se ter uma noção melhor de como os modelos estatísticos modelam as informações sobre algum tipo de epidemia. Podemos ver pelo gráfico a correlação direta entre os números de pessoas suscetíveis, contaminadas e curadas, e perceber como a curva de infecção é parecida com o do modelo anterior, tendo uma subida exponencial. Além disso é importante notar como uma população reagiria sem as medidas de distanciamento social e combate a doença, nessa população com 100 pessoas mais 80% foi infectado pela doença no período de um ano.

Pelo gráfico sobre os número de novos caso diários, na Figura 10, é possível perceber a explosão no número de novos casos logo nos primeiros dias, e logo em seguida com o aumento de pessoas curadas e imunes a propagação do vírus tende a cair. Com o tempo a grande massa de pessoas imunes formam uma barreira natural o que dificulta a propagação do vírus.

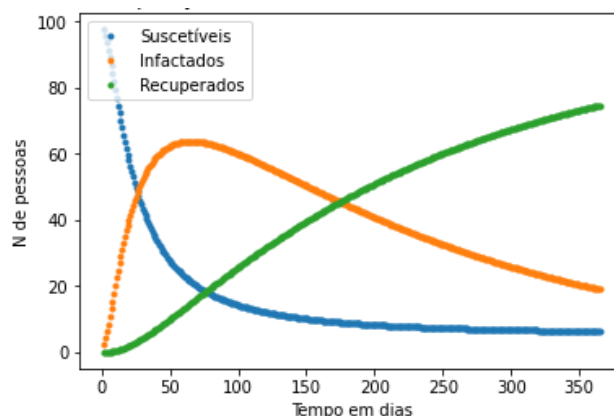


Figura 9: Comparação entre número de Infectados, Sucetíveis e Curados.

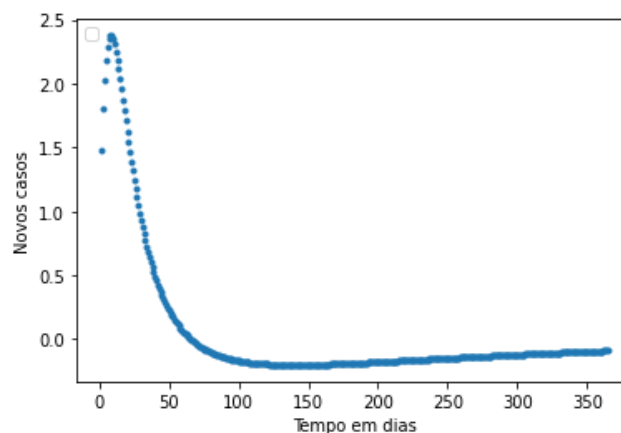


Figura 10: Novos casos de infectados.

### 3.3.2 Simulação dinâmica

Assim como nos modelos anteriores foi desenvolvida uma animação simples a fim de entender melhor o comportamento da curva de infectados. Por meio dela é possível perceber a explosão inicial no número de casos, assim como a queda gradativa em função do tempo e assim entender melhor a dinâmica de um determinada epidemia.

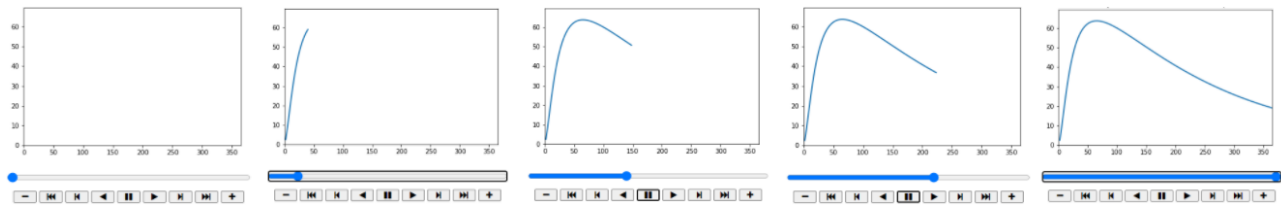


Figura 11: Novos de casos total por dia.

## 4 Conclusão

Diante de tudo o que se foi apresentado, é cabível pontuar sobre a importância do algoritmo de Euler que funcionou de maneira satisfatória tanto no modelo inicial quanto na modelagem extra, o qual possibilitou simular o modelo dinâmico do COVID-19 resolvendo numericamente. Além disso, vale destacar as inúmeras ferramentas de cálculo, e exposição de imagens e informações que a ferramenta utilizada Jupiter Notebook permitiu, como a comparação entre diferentes curvas e recursos lúdicos que enriqueceram o debate e o entendimento sobre o assunto. Outro ponto importante é sobre a definição dos parâmetros utilizados nessa modelagem, que foram em sua maioria estimados de maneiras superficial, na força bruta e sem muito embasamento em casos reais, mas isso poderia ter sido feito com modelos de estatísticos estudados em aula, como o Método de Monte Carlos que estima parâmetros de uma dinâmica. O qual depende de amostragem aleatória repetida para encontrar resultados numéricos, ou seja basicamente o que foi realizado com as cinco ilhas neste relatório mas de maneira mais sofisticada e com o uso ainda mais ferramentas matemáticas. Por fim, gostaria de ressaltar a importância desse tipo de simulação, principalmente em meio a uma pandemia, onde a informação é uma das maiores armas de combate a propagação do vírus, usada para conscientizar a população e alertar a autoridades competentes.

## Contribuições dos Autores

Todo o trabalho, foi desenvolvido por Daniel Silva Lopes da Costa, responsável pela implementação do Notebook, com os modelos matemáticos, os gráficos e animações e também pela escrita do artigo.

## Referências

- [Quintas, 2019] Quintas, Thiago. (2019). **Métodos para estimação de parâmetros em um modelo SIR**. <https://www.ime.usp.br/map/tcc/2019/ThiagoQuintasV1.pdf>
- [Sonnino, 2020] Sonnino, Giorgio. (2020). **Dynamics of the COVID - Comparison between the Theoretical Predictions and the real Data**. <https://arxiv.org/pdf/2003.13540v4.pdf>.