

Universidade de São Paulo
Instituto de Matemática e Estatística
Bacharelado em Matemática Aplicada e Computacional

**Análise do modelo SIR:
Comportamento da curva de infectados em relação à
inclusão de novas semanas epidemiológicas**

Milena Heidecher de Oliveira

Orientador: Prof. Dr. Sérgio Muniz Oliva Filho

São Paulo, Dezembro de 2018

Milena Heidecher de Oliveira

**Análise do modelo SIR:
Comportamento da curva de infectados em relação à inclusão de novas semanas
epidemiológicas**

Trabalho de Conclusão de Curso submetido à Universidade de São Paulo, como requisito necessário para obtenção do grau de Bacharel em Matemática Aplicada e Computacional.

São Paulo, Dezembro de 2018

Resumo

Os estudos na área da saúde envolvendo modelos matemáticos trouxeram avanços significativos na análise e combate à algumas doenças.

O objetivo deste trabalho é analisar o comportamento do modelo epidemiológico SIR ao incluir novas semanas epidemiológicas no estudo.

Com os dados de incidência de infectados por Dengue, da cidade do Rio de Janeiro/RJ entre os anos 2012 e 2013, foram feitas simulações da trajetória de infectados pela doença, comparando aos dados reais.

Com isso, foram feitas análises de como a inclusão de cada semana epidemiológica interfere na previsão dessas trajetórias, calculando o erro cometido pelas simulações em comparação com a trajetória real, levando em conta as limitações do modelo e a não utilização de dinâmicas vitais.

Abstract

Health studies containing mathematical models have brought significant advances in analyzing and combating some diseases.

The objective of this study is analyzing the behavior of the SIR epidemiological model by including new epidemiological weeks in the study.

Based on incidence data of Dengue infected, from the city of Rio de Janeiro/RJ between the years 2012/2013, trajectory simulations of infected by the disease were made, comparing to the real data.

Thus, the analysis of how the inclusion of each epidemiological week interferes in the process of predicting these trajectories, by calculating the error generated by the simulations in comparison with the real trajectory, taking into account the model limitations and the non-use of vital dynamics.

Agradecimentos

À minha mãe pela oportunidade de estudar, por todo o incentivo, por me aguentar nos dias difíceis e por sempre acreditar no meu potencial, sem ela nada disso seria possível

Ao meu pai por me ajudar quando necessário.

À toda minha família pelo apoio emocional e financeiro.

Ao meu namorado por estar por perto nos momentos difíceis, por revisar parte desta monografia e por me apoiar e incentivar sempre.

À minha sogra por sempre me ajudar no que fosse possível.

Aos meus amigos que trago do colégio por sempre acreditarem que seria possível.

Às minhas amigas Gabriela, Juliana por existirem na minha vida.

Aos meus colegas e amigos do ICMC-USP e do alojamento pelo companheirismo e pela vivência.

À todas as mulheres que estão ou estiveram presentes na minha vida pessoal e acadêmica, por me mostrarem força e determinação, por me inspirarem em todos os aspectos possíveis, sou grata por existirem.

À minha avó materna por sempre me incentivar a aprender e gostar de matemática, sem ela nada disso seria possível.

Ao professor Sergio Muniz Oliva Filho, por aceitar me orientar, pelo aprendizado, pelo apoio e principalmente pelos conselhos.

Ao professor Marcos Nascimento Magalhães pelos dois anos de projeto e por me ouvir e aconselhar sempre que precisei.

À professora Sônia Regina, pelo ano de iniciação científica, pelos conselhos e aprendizados e por me compreender nos momentos difíceis.

Aos poucos amigos que fiz nesse instituto, Ana, Amanda, Hanna, Camila, Karina, Gabriel, Gino, Millena, Martin, Thay, Fernanda e Isabela.

Ao DJ, por revisar esta monografia com tanto cuidado.

"Essentially, all models are wrong, but some are useful."

- George E. P. Box

Sumário

Resumo	i
Abstract	iii
Agradecimentos	v
Lista de Figuras	ix
Lista de Tabelas	xi
1 Introdução	1
1.1 Epidemiologia e o conceito de epidemia	1
1.2 A Dengue no Brasil	2
1.3 A importância de estudos sobre Dengue	2
2 Objetivos	3
2.1 Métodos e Procedimentos	3
3 O modelo SIR	5
3.1 Propriedades do modelo SIR	7
3.2 Exemplo do impacto do ajuste dos parâmetros	8
4 Métodos utilizados	11
4.1 Dados Utilizados	11
4.2 Método de <i>Runge-Kutta</i> de Quarta Ordem (RK4)	12
4.3 Método de Monte Carlo	12
4.3.1 Pseudo-Código do Método de Monte Carlo	13
4.4 Método dos Mínimos Quadrados (MMQ)	13
5 Metodologia	15
6 Resultados	17
6.1 Impacto da inclusão de semanas epidemiológicas na estimação do erro	17
6.2 Previsão do pico da doença	21
7 Conclusão	25
7.1 Propostas de Continuidade	25

Lista de Figuras

2.1	Fluxograma do trabalho	4
3.1	Esquema Modelo SIR	5
3.2	Impacto de diferentes parâmetros no modelo SIR	8
4.1	Ideia genérica do Método de Monte Carlo	13
4.2	Ilustração das diferenças entre os dados observados e os valores estimados (dados fictícios)	14
6.1	Curva de infectados do modelo SIR em relação à inclusão de novas semanas epidemiológicas	21
6.2	Curvas das simulações que se aproximam de alguma forma dos dados reais	23
6.3	Gráfico com todas as curvas de previsões de pico da doença	23

Lista de Tabelas

3.1	Parâmetros, condições iniciais e erro (ϵ) de algumas simulações	9
4.1	Semanas epidemiológicas analisadas	11
4.2	Descrição das inclinações	12
6.1	Erros (ϵ) cometidos a cada semana epidemiológica incluída na simulação .	18
6.2	Comparação dos dados das simulações com os dados reais	22

Capítulo 1

Introdução

1.1 Epidemiologia e o conceito de epidemia

A história da epidemiologia tem seus primeiros registros já na Grécia antiga (ano 400 a.C.), quando *Hipócrates*, em um trabalho clássico denominado **Dos Ares, Águas e Lugares**. Ele buscou apresentar explicações, com base no racional e não no sobrenatural, em relação a ocorrência de doenças na população [1].

No século XIV, as rotas marítimas entre o Oriente Médio e a Europa eram constantemente assombradas por epidemias de pequeno ou grande porte. Os barcos, além de cargas, transportavam um estoque razoável de hospedeiros transmissores de doenças, sendo esses barcos responsáveis pela propagação de grandes epidemias.

Segundo *Waldman* (1998) [20], o inglês *John Graunt*, no século XVII, foi o primeiro a quantificar os padrões de ocorrência de doenças, natalidade e mortalidade, identificando características importantes nesses eventos.

Já os primeiros desenvolvimentos em Epidemiologia Matemática parecem ter sido realizados por *Daniel Bernoulli* na última metade do século XVIII [2]. No entanto, somente a partir da metade do século XIX, com o avanço dos conhecimentos médicos sobre as causas das doenças infecciosas, desenvolveram-se teorias matemáticas para fenômenos de larga escala, em contraposição às descrições empíricas [4].

Contudo, nem mesmo com os avanços da medicina moderna foi possível impedir o surgimento de novas epidemias.

Mas como pode-se definir uma epidemia?

Segundo Forattini (2005) [10], epidemia é o "*nome dado ao estado de incidência ou agravo à saúde, além do normalmente esperado dentro da faixa de endemidade, em determinada área ou grupo populacional*".

A propagação de doenças infecciosas continua sendo preocupante e por isso gera muitos estudos na área da Saúde Pública, sendo a matemática uma ferramenta fundamental para auxiliar esses avanços, a qual, através de modelos e números, descobre maneiras de diminuir ou prevenir surtos. A longo prazo pode contribuir para erradicação da doença.

Percebe-se que por muitos anos a humanidade tem lutado contra diversos tipos de endemia, mas os avanços mais significativos vieram do século XX, já que nessa época os progressos da medicina desencadearam no desenvolvimento das primeiras vacinas. Apesar

dos avanços e estudos para criação de novas vacinas, como no caso da Dengue, ainda há um caminho longo a ser percorrido até erradicar-se qualquer doença por completo.

1.2 A Dengue no Brasil

O mosquito transmissor da dengue é originário do Egito (África), e vem se espalhando pelas regiões tropicais e subtropicais do planeta desde o século XVI, por meio de navios que traficavam escravos. O vetor foi retratado cientificamente pela primeira vez em 1762, mas o seu nome definitivo, *Aedes aegypti*, só foi estabelecido em 1818.

No Brasil, os primeiros relatos de dengue são do século XIX, em Curitiba/PR e do início do século XX, em Niterói/RJ. Durante esses séculos o mosquito já era um problema, mas não por conta da dengue, a principal preocupação era a transmissão da febre amarela [19].

Na década de 40 do século passado, o Brasil passava por processos de industrialização e urbanização. Estes processos, de forma acelerada, favoreceram a eclosão de novos criadouros do mosquito, com auxílio da indústria automobilística, através de pneus e ferros-velhos [14].

Apenas em 1942 foi aprovada a erradicação do *Aedes aegypti* em caráter oficial, após os resultados dos trabalhos realizados em 1940 pelo serviço de controle da Febre Amarela no Recife. Este trabalho foi nomeado como "**Marcha para a erradicação do *Aedes aegypti***", que foi reconhecido internacionalmente [18].

A partir de 2013, para a OMS, com mais de 300 casos por 100 mil habitantes o país é considerado em **estado de epidemia**.

1.3 A importância de estudos sobre Dengue

A partir de estudos sobre Dengue, pode ser possível identificar padrões nos surtos, regiões mais suscetíveis à doença e também encontrar situações que estimulam a proliferação do mosquito. Nesse sentido, estudos sobre outras doenças auxiliaram o desenvolvimento de vacinas e campanhas governamentais para o controle de surtos.

Isto posto, este trabalho pretende analisar um modelo matemático e verificar as aproximações em relação aos dados reais, apesar das limitações do modelo. Sendo assim, tal trabalho está estruturado da seguinte forma: Objetivos, O Modelo SIR, Métodos Utilizados, Metodologia, Resultados e Conclusão.

Capítulo 2

Objetivos

Este trabalho tem como proposta analisar o modelo matemático **SIR** (Suscetíveis, Infectados e Removidos), com o auxílio de um método que busca ajustar os parâmetros β e γ e as condições iniciais (S_0 , I_0 e R_0) do modelo sem dinâmicas vitais, descrito pelo sistema

$$\frac{dS}{dt} = -\beta \cdot S(t) \cdot I(t) \quad (2.1)$$

$$\frac{dI}{dt} = \beta \cdot S(t) \cdot I(t) - \gamma \cdot I(t) \quad (2.2)$$

$$\frac{dR}{dt} = \gamma \cdot I(t) \quad (2.3)$$

e assim comparar a estimativa do valor máximo de casos de incidência de infectados por Dengue na cidade do Rio de Janeiro/RJ, com uma quantidade determinada de semanas epidemiológicas, incluídas uma a uma.

Pretende-se mostrar brevemente as diferentes trajetórias que a mudança de parâmetros ocasiona no ajuste do modelo.

Existem diversas abordagens de estudos envolvendo modelos epidemiológicos, como, por exemplo, estudar o envelhecimento de uma população. Neste trabalho a abordagem escolhida foi a análise de como a inclusão de cada semana epidemiológica interfere na projeção do valor máximo de infectados pela doença.

2.1 Métodos e Procedimentos

Existem muitos métodos possíveis para ajustar os parâmetros de modelos matemáticos, dentre eles: *Simplex*, *Quasi-Newton*, *HyperCube*, entre outros [22]. O escolhido para esse caso foi o método de **Monte Carlo (MMC)**. Como ferramenta, foi utilizado o *software R* [9] e o algoritmo de *Runge-Kutta* de quarta ordem (RK4) para resolver o sistema de equações.

Foram feitas aproximações com os dados de incidência de infectados por Dengue no Rio de Janeiro/RJ entre os anos de 2012 e 2013, por semana epidemiológica (início na 46ª de 2012 e final na 23ª de 2013), utilizando o modelo matemático **SIR** de equações diferenciais ordinárias, para comparar o valor máximo de infectados estimado, a cada semana epidemiológica incluída no modelo, verificando se é coerente com o valor máximo de infectados

dos dados reais.

Para explicar rapidamente o trabalho, tem-se o fluxograma abaixo. Todos os métodos serão explicados com maior detalhamento no capítulo 4.

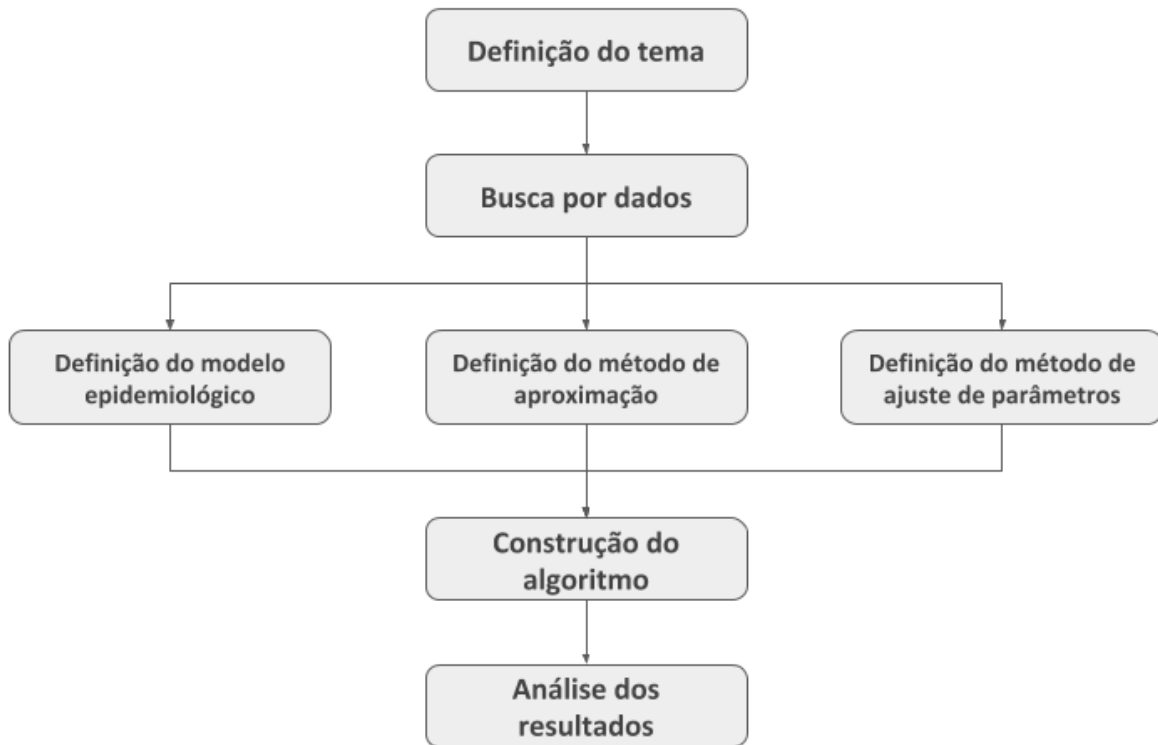


Figura 2.1: Fluxograma do trabalho

Este fluxograma pode ser utilizado para estudos similares com outros tipos de doenças.

Capítulo 3

O modelo SIR

Modelos matemáticos tornaram-se importantes ferramentas para análise e controle de doenças infecciosas [21]. O modelo **SIR** foi proposto por *Kermack* e *McKendrick*, em 1927 [13].

O modelo consiste em três populações:

- **SUSCETÍVEIS (S)**: Pessoas que podem pegar a doença caso ocorra algum contato com o vírus ou bactéria;
- **INFECTADOS (I)**: Pessoas que estão doentes;
- **REMOVIDOS (R)**: Pessoas que se recuperaram da doença ou morreram decorrente dela.

O modelo descreve a dinâmica da divisão da população em três partes, como mostra o esquema da figura abaixo.



Figura 3.1: Esquema Modelo SIR

Como citado no Capítulo 2, o modelo é descrito pelas equações

$$\frac{dS}{dt} = -\beta \cdot S(t) \cdot I(t) \quad (3.1)$$

$$\frac{dI}{dt} = \beta \cdot S(t) \cdot I(t) - \gamma \cdot I(t) \quad (3.2)$$

$$\frac{dR}{dt} = \gamma \cdot I(t) \quad (3.3)$$

sendo β o coeficiente de transmissão e γ a taxa de recuperação [6], que serão ajustados ao modelo conforme o **Método de Monte Carlo (MMC)**, que será descrito na seção 4.3.

Existem muitas derivações desse modelo, como, por exemplo, o modelo SIRS (Suscetível, Infectado, Removido, Suscetível). A versão do modelo SIR utilizada nesse trabalho é mais simples, pois despreza os nascimentos, mortes que não sejam decorrentes da doença (se for decorrente da doença, é considerado removido) e fenômenos migratórios [3]. Portanto, o número total da população a ser considerada é constante (k), como apresentado na equação 3.4

$$N = s + i + r = k \quad (3.4)$$

e como resultado disto, pode-se afirmar que a soma das taxas de mudança de **Suscetíveis, Infectados e Removidos** é zero [5], como mostrado abaixo.

$$\frac{ds}{dt} + \frac{di}{dt} + \frac{dr}{dt} = 0 \quad (3.5)$$

Neste trabalho, utilizaram-se taxas, portanto

$$S = \frac{s}{N} \quad (3.6)$$

$$I = \frac{i}{N} \quad (3.7)$$

$$R = \frac{r}{N} \quad (3.8)$$

assim,

$$S + I + R = 1 \quad (3.9)$$

será importante para geração uniforme das condições iniciais nas simulações.

Esse sistema também pode ser representado diagramaticamente por

$$\boxed{S \xrightarrow{\beta IS} I \xrightarrow{\gamma I} R} \quad (3.10)$$

Algumas considerações sobre epidemia são importantes para interpretar o modelo [16]:

- A doença ou a epidemia sempre acaba antes que toda a população morra;

- A doença parece agir como limitante populacional;
- Em comparação com o tempo de vida do hospedeiro, a duração da doença é curta;
- Uma epidemia ocorre somente se existir um número mínimo de indivíduos, portanto, em populações suficientemente pequenas, não ocorrem epidemias.

As hipóteses do modelo SIR [17] também devem ser levadas em consideração:

- H₁** Devido a numerosidade da população em análise ignora-se efeitos aleatórios pertinentes às diferenças entre os indivíduos e outros fatores;
- H₂** Assume-se que, em qualquer momento, a população é homogeneamente misturada, ou seja, os infectados e suscetíveis estão sempre distribuídos aleatoriamente na área em que a população reside;
- H₃** A doença é transmitida por proximidade ou contato entre infectado/suscetível;
- H₄** Um suscetível fica infectado logo após a transmissão, ou seja, o período de latência da doença é ignorado, fixando-o igual a zero;
- H₅** Os infectados possivelmente se recuperam e, uma vez recuperados, ficam para sempre imunes (ao menos durante o período em análise);
- H₆** Não são considerados nascimentos, mortes ou efeitos migratórios.

3.1 Propriedades do modelo SIR

Propriedade 1. *Modelo SIR sem taxas vitais:*

- 1 Sendo $S'(t) \leq 0$ e $R'(t) \geq 0$, $S(t)$ é uma função decrescente e $R(t)$ é uma função crescente. Assim, $0 \leq S(t) \leq S(0)$ e $0 \leq R(t) \leq R(0)$. Segue que $S(\infty)$, $R(\infty)$ e, portanto,

$$I(\infty) = N(0) - S(\infty) - R(\infty)$$

existem e são finitos.

- 2 A doença sempre desaparece, ou seja, todas as condições iniciais levam a $I(\infty) = 0$. Se não, a equação 3.3 implica que para t suficientemente grande,

$$R'(t) > \gamma \frac{I(\infty)}{2}$$

isso implica que $R(\infty) = 0$, **contradição**.

Teorema 1. *Dinâmicas do Modelo SIR:*

- 1 Se $R_0 < 1$, então $I(t)$ decresce monotonicamente para zero em $t \rightarrow \infty$ (com taxa exponencial).

2 Se $R_0 > 1$, então $I(t)$ começa a crescer, atinge seu máximo e então decresce para zero em $t \rightarrow \infty$.

3 Se $R_0 = 1$, então $I(t)$ decresce para zero em $t \rightarrow \infty$.

Demonstração primeira afirmação Teorema 1. A equação (3.1) e a Propriedade 1.1 implicam que

$$I'(t) = (\beta S(t) - \gamma) \cdot I(t) \leq (\beta S(0) - \gamma) \cdot I(t) < 0$$

para $R_0 < 1$, junto com a propriedade 1.2, fica provado a **primeira afirmação do teorema**. \square

Demonstração segunda afirmação Teorema 1. A equação (3.1) implica

$$I'(0) = (\beta S(0) - \gamma) \cdot I(0) > 0$$

para $R_0 > 1$. Assim $I(t)$ está aumentando em $t = 0$. A equação (3.1) também implica que $I(t)$ tem apenas um ponto crítico não nulo. Juntando essas informações com a propriedade 1.2, prova-se a **segunda afirmação do teorema**. \square

Demonstração terceira afirmação Teorema 1. A equação (3.1), junto com a propriedade 1.1 implicam

$$I'(t) = (\beta S(t) - \gamma) \cdot I(t) \leq (\beta S(0) - \gamma) \cdot I(t) \leq 0$$

para $R_0 = 1$. Juntando essas informações com a propriedade 1.2, prova-se a **terceira afirmação do teorema**. \square

Uma doença infecciosa é chamada endêmica se $I(\infty) > 0$. O **Teorema 1** implica que o modelo SIR ignora doenças endêmicas.

3.2 Exemplo do impacto do ajuste dos parâmetros

Conforme altera-se os intervalos dos parâmetros, foram obtidas diferentes curvas de aproximação.

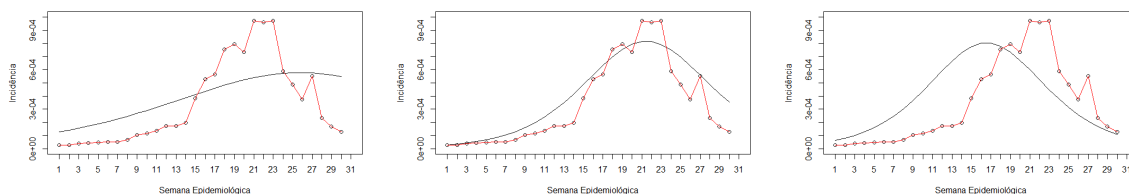


Figura 3.2: Impacto de diferentes parâmetros no modelo SIR

Os pontos são os dados reais utilizados neste trabalho, a linha vermelha foi gerada através do comando *lines* do *software* R e a curva em preto é a aproximação obtida pela simulação.

A cada simulação, alterou-se os intervalos de β e γ , afim de chegar cada vez mais próximo do melhor ajuste.

A seguir, tabela com os valores de β , γ , condições iniciais (que nesse caso também são números aleatórios de uma distribuição uniforme) e o erro (ϵ) de algumas simulações realizadas

β	γ	S_0	I_0	R_0	Erro (ϵ)
0.706224	0.386446	0.7095272	0.00003179698	0.2904410	0.001851287
7.608440	5.085827	0.7004463	0.00006175068	0.2994919	0.001508861
6.845251	4.945399	0.7565661	0.00002715233	0.2434067	0.000702913

Tabela 3.1: Parâmetros, condições iniciais e erro (ϵ) de algumas simulações

fica evidente que as alterações nos parâmetros e condições iniciais têm forte impacto no erro (ϵ).

Capítulo 4

Métodos utilizados

Neste capítulo serão apresentados todos os métodos utilizados para a realização do trabalho. Para isso, todas as partes do fluxograma 2.1 iniciadas com "**definir**" serão detalhadas a seguir.

4.1 Dados Utilizados

Conforme citado na seção 2.1, os dados utilizados são de incidência de infectados por Dengue na cidade do Rio de Janeiro/RJ entre os anos de 2012 e 2013, por semana epidemiológica, conforme tabela

Semana Epidemiológica Inicial	Semana Epidemiológica Final
46 ^a de 2012	23 ^a de 2013

Tabela 4.1: Semanas epidemiológicas analisadas

lembrando que cada ano é formado por 52 semanas epidemiológicas, o que significa que foram analisadas 30 semanas epidemiológicas com dados de incidência de infectados por Dengue.

Todos os dados de incidência de dengue (x_i , com $1 \leq i \leq 30$), por semana epidemiológica, foram transformados em taxas

$$\frac{x_i}{6320446} \quad (4.1)$$

sendo 6320446 a população total do município do Rio de Janeiro/RJ no período das semanas epidemiológicas utilizadas.

4.2 Método de *Runge-Kutta* de Quarta Ordem (RK4)

O método de *Runge-Kutta* de 4ª ordem é dado por

$$y_{k+1} = y_k + \frac{h}{6} (k_1 + 2k_2 + 2k_3 + k_4) \quad (4.2)$$

em que

$$k_1 = f(x_k, y_k) \quad (4.3)$$

$$k_2 = f\left(x_k + \frac{h}{2}, y_k + k_1 \frac{h}{2}\right) \quad (4.4)$$

$$k_3 = f\left(x_k + \frac{h}{2}, y_k + k_2 \frac{h}{2}\right) \quad (4.5)$$

$$k_4 = f(x_k + h, y_k + hk_3) \quad (4.6)$$

então 4.2 é determinado pelo valor atual $y(k)$ somado com o produto de uma inclinação estimada com o tamanho do intervalo (h), no qual a inclinação é a média ponderada de inclinações:

Sigla da inclinação	Descrição
k_1	inclinação no início do intervalo
k_2	inclinação no ponto médio do intervalo, usando k_1 para determinar o valor de y no ponto $t_k + \frac{h}{2}$ pelo método de <i>Euler</i>
k_3	inclinação no ponto médio do intervalo, mas agora usando a inclinação k_2 para determinar o valor de y
k_4	inclinação no final do intervalo, com seu valor y determinado usando k_3

Tabela 4.2: Descrição das inclinações

Esse método serve para resolver numericamente equações diferenciais ordinárias [7]. Ser de 4ª ordem significa que o erro por passo é da ordem de h^5 , enquanto o erro total acumulado é da ordem de h^4 .

Este tipo de técnica foi desenvolvida por volta de 1900 pelos matemáticos **Carl David Tolmé Runge** e **Martin Wilhelm Kutta**.

4.3 Método de Monte Carlo

Como primeira tentativa, foi utilizado o método *Hypercube*, mas por percorrer todos os possíveis valores, se tornou inviável de concluir pela demanda de tempo que o algoritmo necessitou. Portanto, o método de *Monte Carlo* (MMC) surgiu como uma boa solução para o problema.

Este método se baseia em amostragens aleatórias em grande quantidade para obtenção de resultados numéricos, ou seja, repetir simulações um número massivo de vezes, para

calcular probabilidades de modo heurístico, como se registrassem os resultados reais em jogos de cassino (por isso o nome do método) [12].

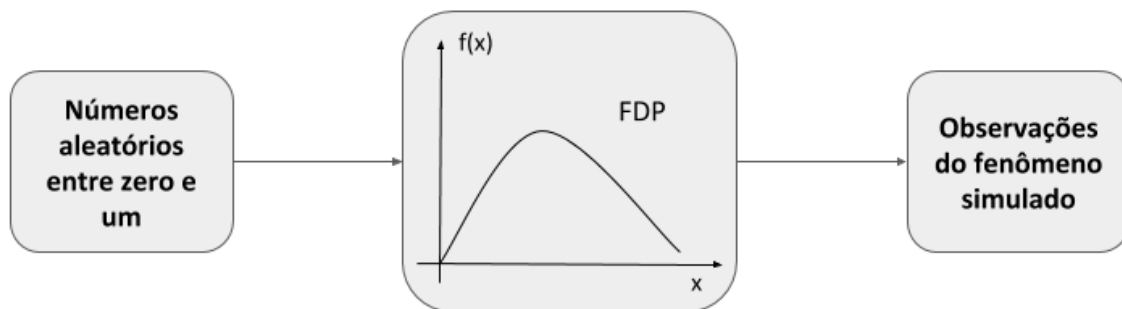


Figura 4.1: Ideia genérica do Método de Monte Carlo

Este tipo de método é utilizado em simulações estocásticas com inúmeras aplicações em áreas como a física e biologia [11]. Também é bastante utilizado como forma de obter aproximações numéricas de funções em que não é viável, ou é mesmo impossível, obter uma solução analítica, como é o caso da abordagem desse trabalho.

O Método de *Monte Carlo* foi formalizado em 1949, por intermédio do artigo “**Monte Carlo Method**”, publicado por *John Von Neumann* e *Stanislav Ulam* [8].

4.3.1 Pseudo-Código do Método de Monte Carlo

No pseudo-código abaixo, $[a_i, b_j]$ são intervalos fechados.

```

definir intervalos para os parâmetros
definir número de iterações
enquanto (i in 1:número de iterações)
{
  gerar parâmetro  $\beta$  uniformemente (1,  $a_1$ ,  $b_1$ )
  gerar parâmetro  $\gamma$  uniformemente (1,  $a_2$ ,  $b_2$ )
  gerar Infectados_iniciais ( $I_0$ ) uniformemente (1,  $a_3$ ,  $b_3$ )
  gerar Suscetíveis_iniciais ( $S_0$ ) uniformemente (1,  $a_4$ ,  $b_4$ )
  gerar Removidos_iniciais ( $R_0$ ) uniformemente (1,  $a_5$ ,  $b_5$ )
  rodar modelo SIR pelo método RK4 com os parâmetros ( $\beta$ ,  $\gamma$ ,  $S_0$ ,  $I_0$ ,  $R_0$ )
  calcular estatística que quero minimizar por Mínimos Quadrados (MMQ)
}

```

4.4 Método dos Mínimos Quadrados (MMQ)

Calcular o erro por MMQ significa calcular a soma dos quadrados das diferenças entre os dados observados e os valores estimados pela simulação, conforme a fórmula

$$\epsilon = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (4.7)$$

sendo n o número de observações, x_i o número real de infectados e y_i o número de infectados obtidos em uma dada simulação.

Os parâmetros do modelo SIR (β e γ) serão ajustados de forma a minimizar ε .

Abaixo, representação gráfica do MMQ.

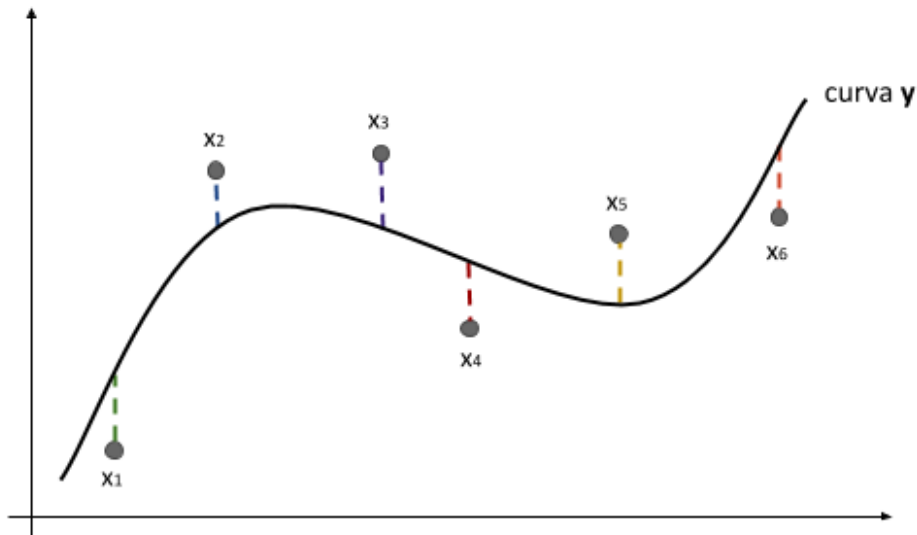


Figura 4.2: Ilustração das diferenças entre os dados observados e os valores estimados (dados fictícios)

Essas diferenças (linhas tracejadas coloridas na Figura 4.2)

$$(x_i - y_i)^2 \quad (4.8)$$

são chamadas de resíduos.¹

¹**Curiosidade:** Historicamente, *Carl Friedrich Gauss* é creditado como o desenvolvedor das bases fundamentais do Método dos Mínimos Quadrados (MMQ), quando tinha apenas dezoito anos, em 1795. No entanto, *Adrien-Marie Legendre* foi o primeiro a publicar o método, em 1805, em seu *Nouvelles méthodes pour la détermination des orbites des comètes*. Enquanto *Gauss* só publicou suas conclusões em 1809 [15].

Capítulo 5

Metodologia

Para entender e interpretar os resultados do estudo, são necessárias as seguintes considerações:

- A semana epidemiológica inicial (46^a de 2012) foi nomeada como **semana epidemiológica 1** do estudo, e decorrente disso, a semana epidemiológica final (23^a de 2013) foi nomeada como **semana epidemiológica 30** do estudo;
- A primeira simulação foi feita com as **cinco** primeiras semanas epidemiológicas e a cada nova simulação, uma semana epidemiológica foi adicionada;
- O número de iterações definido para o algoritmo foi de **100.000**, levando em consideração a baixa variabilidade do erro (ϵ) de uma simulação para outra;
- O valor de I_0 utilizado foi a taxa de incidência de infectados da **semana epidemiológica 1** do estudo

$$I_0 = \frac{170}{6320446}$$

já que após alguns testes, quanto mais próximo desse valor era sorteado I_0 , menor o erro (ϵ);

- A geração uniforme do valor de S_0 foi feita no intervalo $[0.7, 1 - I_0]$, ou seja

$$S_0 \in \left[\frac{4424312}{6320446}, \frac{6320276}{6320446} \right]$$

então

$$R_0 = 1 - S_0 - I_0$$

ficando de acordo com a equação (3.9);

- Intervalos utilizados para β e γ

$$\beta \in [5, 8]$$

$$\gamma \in [4, 5.5]$$

pois obtiveram melhores aproximações em relação à intervalos maiores ou muito restritos;

- O método escolhido não leva em conta a população de mosquitos, assim como o contato mosquito/mosquito e mosquito/humano;
- As análises, comparações e simulações foram feitas apenas com os dados de incidência de **INFECTADOS**;
- Os gráficos 6.2 e 6.3 consideram o período de um ano, ou seja, 52 semanas epidemiológicas a partir da **semana epidemiológica 1** do estudo.

Capítulo 6

Resultados

Este capítulo apresentará os resultados obtidos após as simulações e os estudos teóricos sobre o modelo SIR.

6.1 Impacto da inclusão de semanas epidemiológicas na estimação do erro

A análise do erro ocorreu a cada inclusão de semana epidemiológica, já que um novo valor foi adicionado e, com isso, deve-se ajustá-lo a simulação. Na tabela a seguir, todos os erros obtidos a cada semana epidemiológica inclusa na simulação.

Número da simulação	Quantidade de semanas epidemiológicas simuladas	Erro (ϵ)
1	5	0.52166×10^{-5}
2	6	0.67016×10^{-5}
3	7	0.93968×10^{-5}
4	8	1.03149×10^{-5}
5	9	2.67023×10^{-5}
6	10	2.93925×10^{-5}
7	11	3.05061×10^{-5}
8	12	3.61551×10^{-5}
9	13	3.73172×10^{-5}
10	14	3.82072×10^{-5}
11	15	12.69012×10^{-5}
12	16	20.03124×10^{-5}
13	17	21.68887×10^{-5}
14	18	26.58112×10^{-5}
15	19	27.49438×10^{-5}
16	20	28.52509×10^{-5}
17	21	29.68196×10^{-5}
18	22	29.92236×10^{-5}
19	23	29.88748×10^{-5}
20	24	46.09902×10^{-5}
21	25	53.72622×10^{-5}
22	26	59.02231×10^{-5}
23	27	59.09545×10^{-5}
24	28	62.73481×10^{-5}
25	29	65.45625×10^{-5}
26	30	67.17216×10^{-5}

Tabela 6.1: Erros (ϵ) cometidos a cada semana epidemiológica incluída na simulação

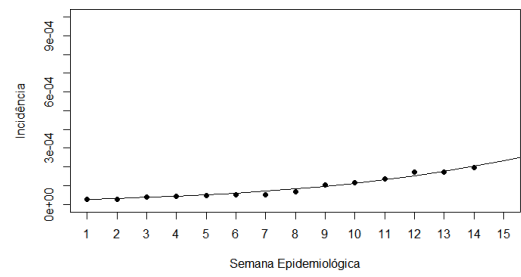
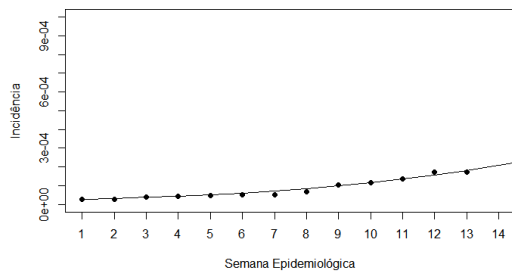
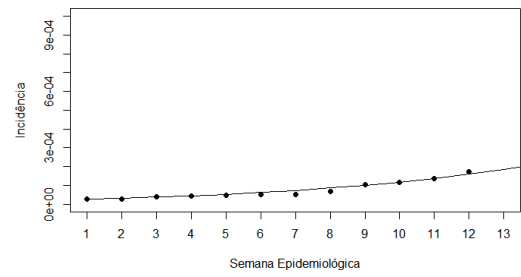
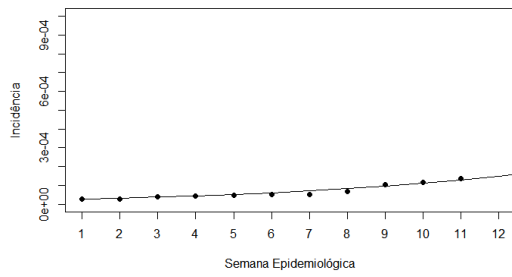
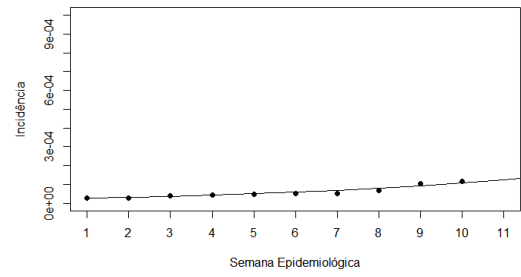
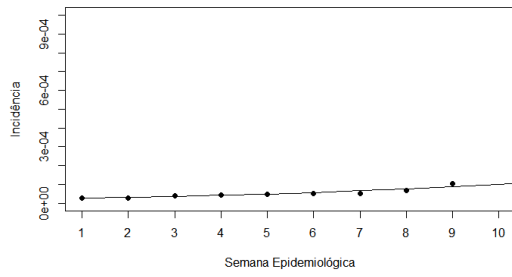
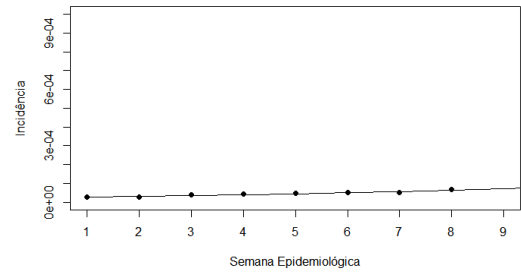
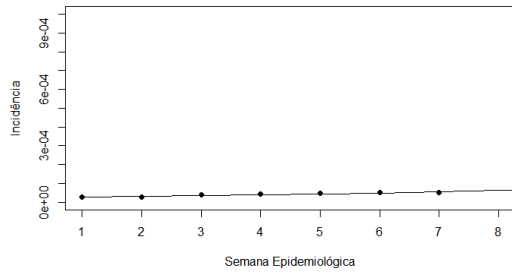
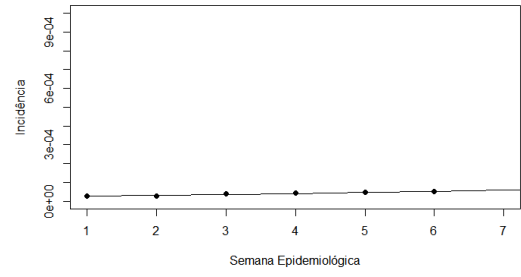
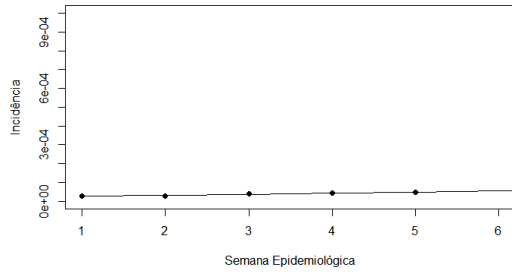
Lembrando que os dados reais tiveram interferências não consideradas na análise pelo modelo SIR, como por exemplo nascimentos e mortes por outras causas, inviabilizando um ajuste com maior precisão.

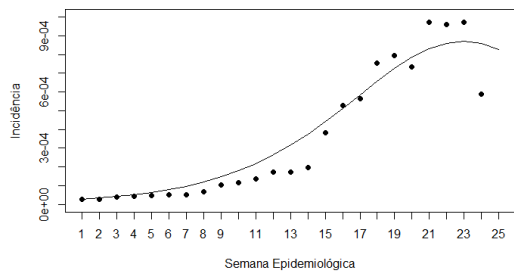
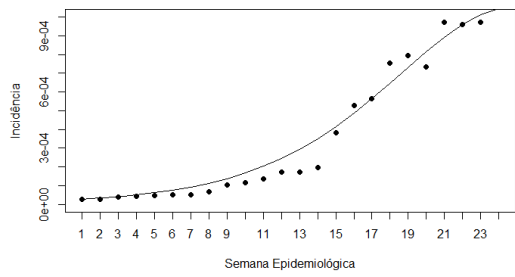
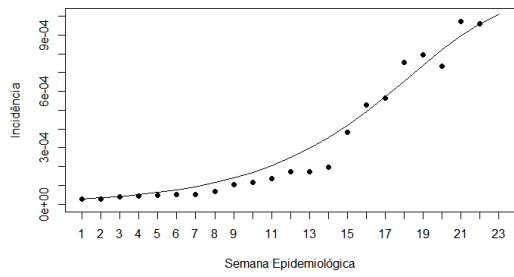
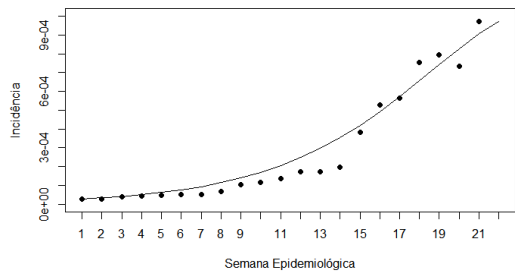
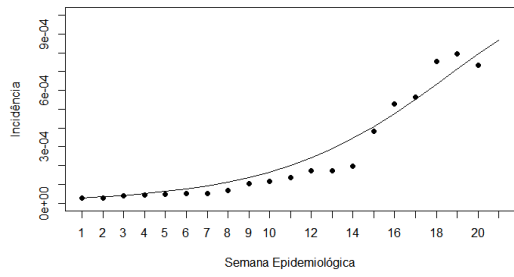
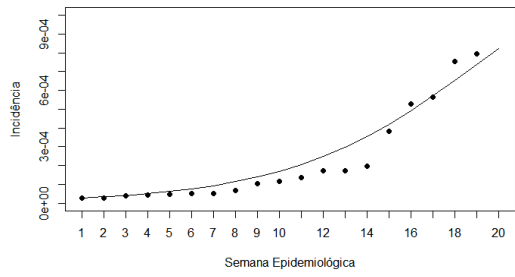
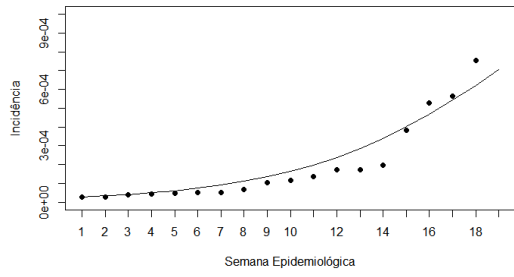
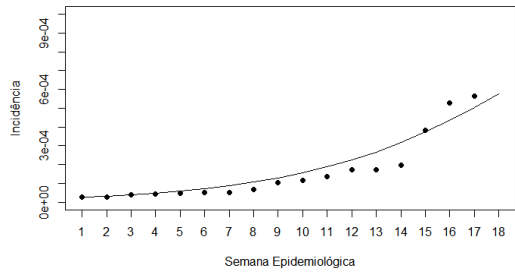
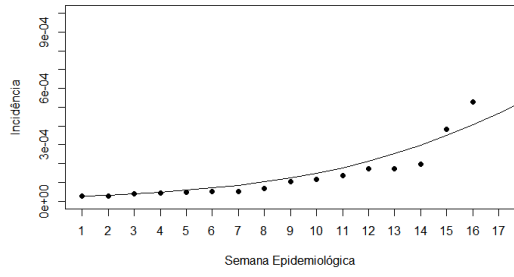
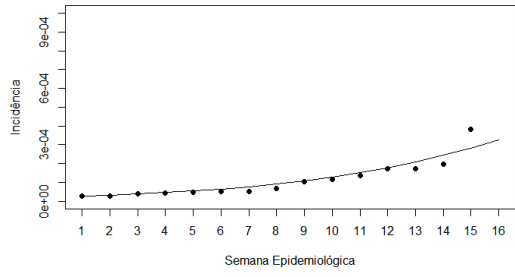
Fica evidente que a inclusão de novas semanas faz com que o erro (ϵ) aumente. Comparando o erro cometido na **simulação 1**, com o erro cometido na **simulação 26**

$$\epsilon(\text{semana } 26) \approx 128 \times \epsilon(\text{semana } 1) \quad (6.1)$$

percebe-se o grande impacto da inclusão de novas semanas epidemiológicas na simulação.

A seguir, os gráficos que mostram os ajustes aos dados reais, por semana epidemiológica incluída.





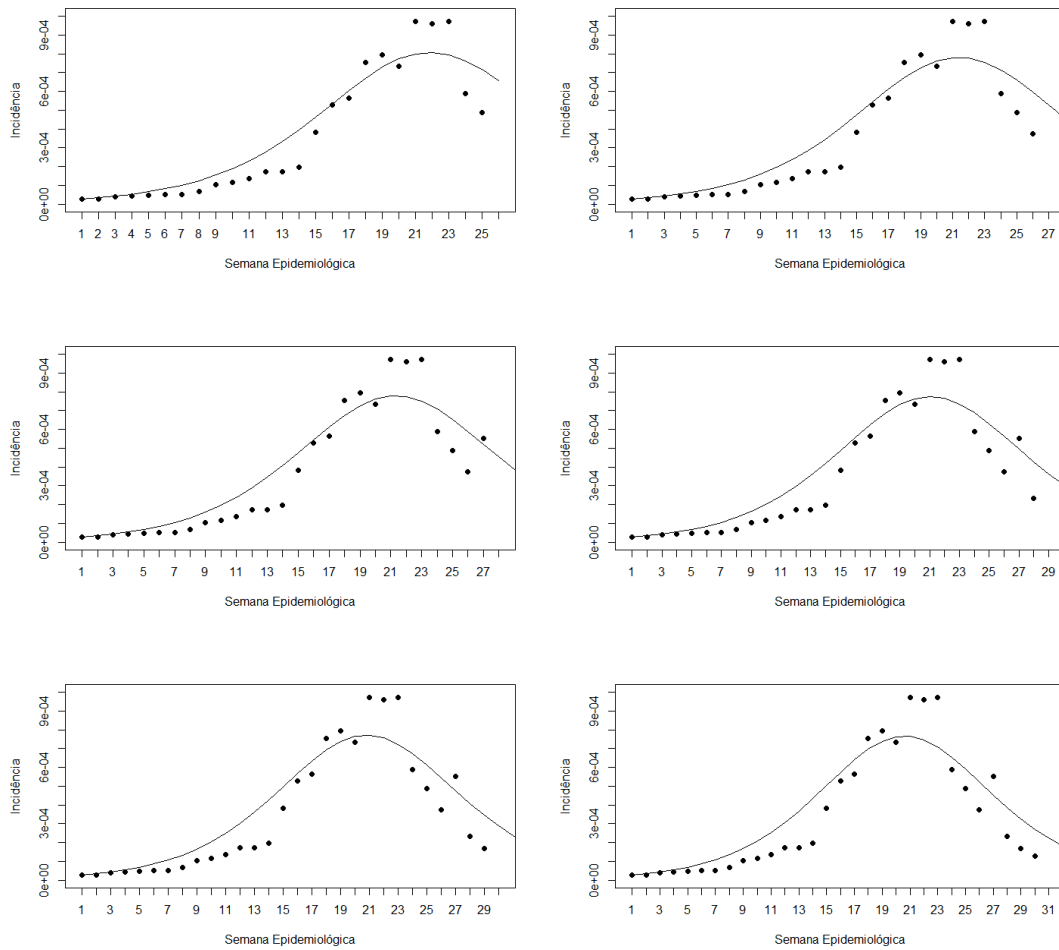


Figura 6.1: Curva de infectados do modelo SIR em relação à inclusão de novas semanas epidemiológicas

Os pontos representam os dados reais e as curva representam cada simulação realizada, em ordem crescente.

6.2 Previsão do pico da doença

Analisando os dados reais de incidência de infectados por dengue no Rio de Janeiro/RJ, tem-se que:

- Semana epidemiológica do estudo em que a doença atinge o número máximo de infectados: **23**
- Máximo de infectados durante o período de análise, em números brutos: **6146**

A tabela a seguir faz um comparativo entre cada simulação com os dados reais, relacionando a semana em que a doença atinge o máximo de infectados e qual o valor bruto de infectados.

Número da simulação	Semana em que atingiu o máximo	Diferença entre a semana de pico máximo (23)	Valor máximo de infectados	Diferença entre o valor máximo de infectados (6146)
1	25	2	1933	4213
2	25	2	1500	4646
3	26	3	1327	4819
4	26	3	1426	4720
5	30	7	3574	2572
6	29	6	3976	2170
7	29	6	4093	2053
8	28	5	4262	1884
9	28	5	4218	1928
10	28	5	3924	2222
11	27	4	4986	1160
12	26	3	5950	196
13	26	3	6174	28
14	25	2	6423	277
15	25	2	6646	500
16	25	2	6440	294
17	25	2	6739	593
18	25	2	6583	437
19	25	2	6659	513
20	23	0	5497	649
21	22	1	5110	1036
22	21	2	4932	1214
23	21	2	4912	1234
24	21	2	4899	1247
25	21	2	4878	1268
26	21	2	4834	1312

Tabela 6.2: Comparação dos dados das simulações com os dados reais

O gráfico a seguir permite uma comparação visual das simulações destacadas na tabela.

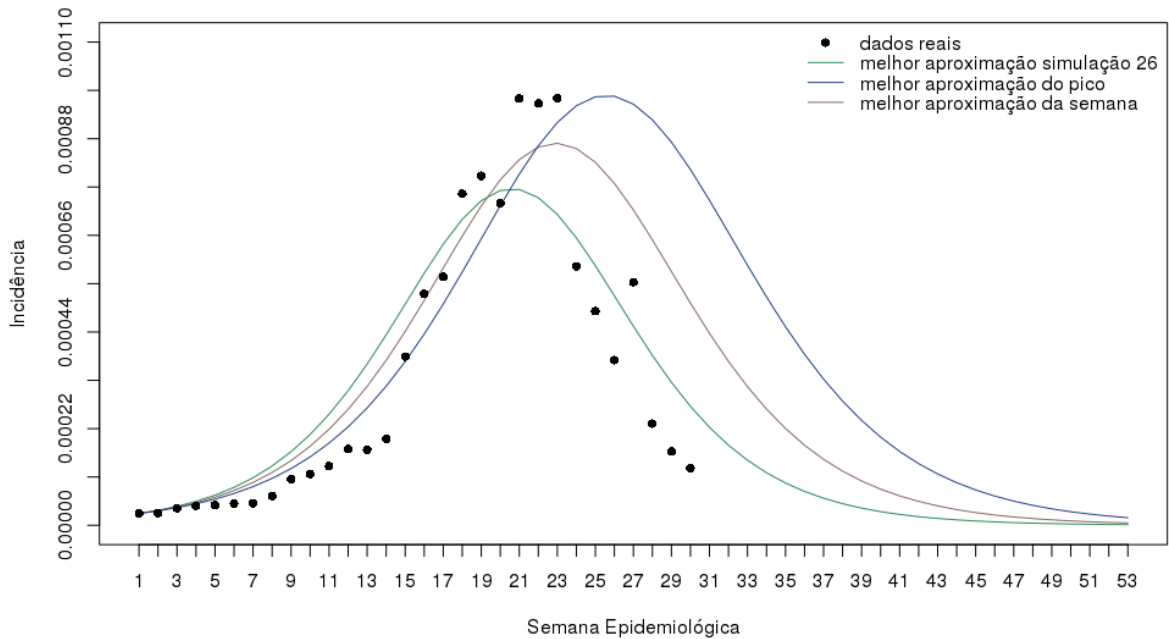


Figura 6.2: Curvas das simulações que se aproximam de alguma forma dos dados reais

As simulações em destaque na Tabela 6.2 mostram as melhores aproximações aos dados reais, a **simulação 13** é a que melhor se aproxima do número máximo de incidência, a **simulação 20** é a que acerta a semana epidemiológica em que ocorre o número máximo de incidência e a **simulação 26** a melhor aproximação envolvendo os 30 dados.

O gráfico abaixo mostra as curvas previstas em cada uma das 26 simulações.

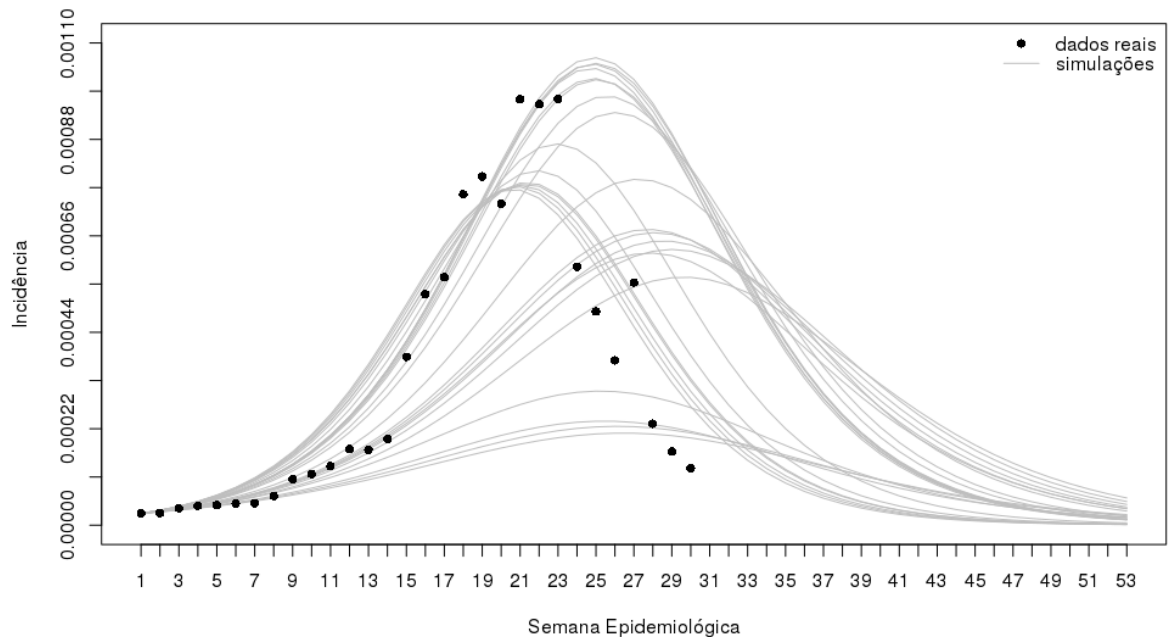


Figura 6.3: Gráfico com todas as curvas de previsões de pico da doença

Capítulo 7

Conclusão

Diante das dificuldades para criar um bom algoritmo de aproximação para o ajuste de parâmetros, pode-se concluir que o Método de Monte Carlo (MMC) gera aproximações coerentes com os dados reais, ressaltando que o modelo utilizado não leva em consideração interferências externas como nascimentos, fenômenos migratórios e mortes decorrente de outras causas, nem o contato mosquito/mosquito, que incluiria uma nova população ao modelo.

O objetivo era comparar as diferentes possíveis trajetórias da doença em relação a inclusão de novas semanas epidemiológicas, o resultado obtido mostrou que novos valores impactam muito nas previsões, como mostra o gráfico 6.3.

Em relação as previsões obtidas nas simulações, percebeu-se que quanto mais semanas epidemiológicas eram incluídas, maior era o erro cometido, já que a incidência de infectados por dengue não segue a distribuição de uma curva perfeita. Contudo, mesmo com as diversas limitações envolvidas, a aproximação de algumas simulações alcançou valores próximos tanto do máximo de infectados como da semana de ocorrência deste evento, conforme gráfico 6.2.

7.1 Propostas de Continuidade

Sugere-se aprimorar o modelo utilizado, ou utilizar um modelo mais sofisticado para ter melhores resultados, além de incluir a população de mosquitos, que sem dúvidas melhoraria as aproximações simuladas.

Também seria interessante realizar o mesmo tipo de análise para outros modelos, com os mesmo dados, e verificar como diferentes modelos matemáticos mostram aproximações para o mesmo conjunto de dados.

Referências Bibliográficas

- [1] N. d. Almeida Filho and M. Z. Rouquayrol. Introdução à epidemiologia moderna. In *Introdução à epidemiologia moderna*. 1992.
- [2] R. M. Anderson and R. M. May. *Infectious diseases of humans: dynamics and control*. Oxford University Press, 1992.
- [3] H. Andersson and T. Britton. *Stochastic epidemic models and their statistical analysis*, volume 151. Springer Science & Business Media, 2012.
- [4] N. T. Bailey et al. *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.
- [5] R. C. Bassanezi. *Ensino-aprendizagem com modelagem matemática: uma nova estratégia*. Editora Contexto, 2002.
- [6] N. F. Britton. *Essential mathematical biology*. Springer Science & Business Media, 2012.
- [7] J. C. Butcher. *Numerical methods for ordinary differential equations*. John Wiley & Sons, 2016.
- [8] R. Eckhardt. Stan ulam, john von neumann, and the monte carlo method. *Los Alamos Science*, 15(131-136):30, 1987.
- [9] A. Field, J. Miles, and Z. Field. *Discovering statistics using R*. Sage publications, 2012.
- [10] O. P. Forattini. *Conceitos Básicos de Epidemiologia Molecular Vol. 64*. EdUSP, 2005.
- [11] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.
- [12] J. Hromkovič. *Algorithmics for hard problems: introduction to combinatorial optimization, randomization, approximation, and heuristics*. Springer Science & Business Media, 2013.
- [13] W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics—i. *Bulletin of mathematical biology*, 53(1-2):33–55, 1991.
- [14] J. S. Silva, I. Scopel, et al. A dengue no brasil e as políticas de combate ao aedes aegypti: da tentativa de erradicação às políticas de controle. *Hygeia: Revista Brasileira de Geografia Médica e da Saúde*, 4(6), 2008.
- [15] S. M. Stigler. *The history of statistics: The measurement of uncertainty before 1900*. Harvard University Press, 1986.

- [16] Y. Takeuchi, Y. Iwasa, and K. Sato. *Mathematics for life science and medicine*. Springer Science & Business Media, 2007.
- [17] J. N. Tavares. Modelo sir em epidemiologia. *Revista de Ciência Elementar*, 5(2):rce2017–020, 2017.
- [18] M. TEIXEIRA. *Dengue e Espaços Intra-Urbanos: Dinâmica de Circulação Viral e Efetividade de Ações de Combate Vetorial*. 2000. 189 p. PhD thesis, Tese (Doutorado em Saúde Coletiva). Instituto de Saúde Coletiva da Universidade Federal da Bahia, Salvador, 2000.
- [19] D. Valle et al. *Aedes aegypti: introdução aos aspectos científicos do vetor*. 2013.
- [20] E. A. Waldman and T. E. da Costa Rosa. *Vigilância em saúde pública*, volume 7. Universidade de Sao Paulo. Faculdade de Saude Publica, 1998.
- [21] H. Weiss. *A Mathematical Introduction to Population Dynamics*. IMPA, 2009.
- [22] A. I. Yuceer M. and B. R. *An Integration Based Optimization Approach for Parameter Estimation in Dynamic Models*. European Symposium on Computer Aided Process Engineering-15., 2005.