

UNIVERSIDADE FEDERAL DO PARANÁ
STEFANI VICOSKI MARQUES

**MÉTODO DOS QUADRADOS MÍNIMOS EM
ÉPOCA DA PANDEMIA DE COVID-19**

Curitiba
2020

Stefani Vicoski Marques

MÉTODO DOS QUADRADOS MÍNIMOS EM ÉPOCA DA PANDEMIA DE COVID-19

Trabalho de conclusão de curso apresentado ao Curso de Licenciatura em Matemática da Universidade Federal do Paraná como requisito parcial para obtenção do grau de Licenciado em Matemática.

Orientadora: Profa. Dra. Elizabeth Wegner Karas.

**Curitiba
2020**

Agradecimentos

A Deus, sem ele nada seria possível.

À minha mãe, Eliana Vicoski, pelo carinho, afeto, dedicação e cuidado ao longo de toda minha vida.

À Profa. Dra. Elizabeth Wegner Karas por ser uma constante fonte de motivação e incentivo ao longo do desenvolvimento deste projeto.

Ao Prof. Dr. Ademir Alves Ribeiro pelas discussões e sugestões que nos motivaram no desenvolvimento teórico do trabalho.

À Profa. Dra. Nara Bobko por sua palestra “Desafios da Modelagem Matemática aplicada no COVID-19”, ministrada na “Semana Acadêmica da Matemática de 2020” na Universidade Federal do Paraná, que nos direcionou nos experimentos numéricos e pelas suas contribuições para melhoria do trabalho.

Resumo

Este trabalho foi conduzido em meio à pandemia gerada pela COVID-19. Visto a necessidade de estudar modelos que se ajustassem aos números de casos e óbitos pela doença decidimos trabalhar com o Método dos Quadrados Mínimos, cientes da existência de modelos mais apropriados. A escolha por este método é devida à sua aplicação em diversas situações e pela possibilidade de revisar conceitos de Cálculo e Álgebra Linear de forma aplicada. Este método consiste em modelar uma função f que se ajuste a um conjunto de pontos de coordenadas (x_i, y_i) de maneira que a soma dos quadrados da diferença entre $f(x_i)$ e y_i seja o menor possível. Apresentamos este método e aplicamos aos números de casos e óbitos por COVID-19 em Curitiba, Paraná e no Brasil.

Palavras-chave: Método dos Quadrados Mínimos, modelagem matemática, COVID-19.

Sumário

	Sumário	5
1	INTRODUÇÃO	7
2	MÉTODO DOS QUADRADOS MÍNIMOS	9
2.1	Modelo Linear	10
2.2	Modelo Exponencial	14
2.3	Modelo Geométrico	17
2.4	Modelo Polinomial	19
2.4.1	Modelo Quadrático	19
2.4.2	Solução Matricial	25
2.5	Modelo exponencial de uma quadrática	28
3	APLICAÇÃO À MODELAGEM DOS DADOS DA COVID-19	31
3.1	Curitiba	31
3.1.1	Número de casos confirmados	31
3.1.2	Número de óbitos	34
3.1.3	Erros relativos	38
3.2	Paraná	38
3.2.1	Número de casos confirmados	39
3.2.2	Número de óbitos	42
3.2.3	Erros relativos	44
3.3	Brasil	45
3.3.1	Número de casos confirmados	46
3.3.2	Número de óbitos	49
3.3.3	Erros relativos	52
4	EXPERIÊNCIA DIDÁTICA EM TEMPOS DE QUARENTENA	53
5	CONCLUSÃO	57
	Referências	59

1 Introdução

A COVID-19 é uma doença causada pelo coronavírus denominado SARS-CoV-2, que foi identificado pela primeira vez em Dezembro de 2019, em Wuhan na China. O surto dessa doença aqui no Brasil iniciou em Março de 2020 [7].

Diante dos números alarmantes relacionados ao número de casos e óbitos pela doença, decidimos estudar o método de Quadrados Mínimos para buscar modelos que se ajustassem a eles. Estamos cientes de que existem modelos mais apropriados como SI, SIR, SIRS e SEIR [1, 6, 5] usados em epidemiologia. Escolhemos o método de Quadrados Mínimos por sua facilidade de compreensão e entendimento para modelagem de dados de modo geral.

O método de Quadrados Mínimos [8] constitui em modelar uma função f que se ajuste a um conjunto de pontos dados de coordenadas (x_i, y_i) de maneira que o erro da soma dos quadrados da diferença entre $f(x_i)$ e y_i seja o menor possível. As abscissas podem representar semanas, por exemplo, e as ordenadas o número de novos casos da COVID-19 numa determinada localidade na semana correspondente.

Nossa intenção não é usar esses modelos para previsões. Nossa motivação foi rever conceitos estudados nas disciplinas de Cálculo e Álgebra Linear e aplicá-los na fundamentação do Método dos Quadrados Mínimos, tendo como plano de fundo o momento vivido pela pandemia. Assim, o objetivo deste trabalho é, estudar o método de Quadrados Mínimos e aplicá-lo aos dados da COVID-19.

Seguindo as ideias de [2, 8], no Capítulo 2 apresentamos a fundamentação do Método de Quadrados Mínimos. Iniciamos com o modelo linear usando ferramentas de Cálculo Diferencial de várias variáveis. A partir do caso linear obtemos o modelo exponencial para a modelagem do conjunto de pontos dados. Generalizamos o caso linear para modelos polinomiais tendo como base conceitos de Álgebra Linear. Discutimos os casos em que é possível garantir a unicidade do modelo desejado e quando há uma infinidade de soluções o que depende da disposição dos pontos dados.

O Capítulo 3 é dedicado à aplicação do Método dos Quadrados Mínimos para modelar o número de casos confirmados e mortes pela COVID-19, no período de março a agosto de 2020, em Curitiba, Paraná e Brasil, com dados obtidos de [11], [10] e [7], respectivamente.

A pandemia impôs à sociedade restrições de convívio social levando a substituição de atividades presenciais por remotas. Com a suspensão do calendário acadêmico na Universidade Federal do Paraná participamos da oferta remota do curso de extensão intitulado *Modelagem Matemática em época de quarentena* com o módulo de Quadrados Mínimos. Essa experiência está registrada no Capítulo 4.

Finalmente, conclusões do trabalho são apresentadas no Capítulo 5.

2 Método dos Quadrados Mínimos

Dados n pares de pontos $(x_i, y_i) \in \mathbb{R}^2, i = 1, \dots, n$, desejamos obter uma função $f : \mathbb{R} \rightarrow \mathbb{R}$ de modo que $f(x_i)$ seja o mais próximo de y_i . Nosso trabalho aqui será encontrar esta f . Para isso vamos utilizar o Método dos Quadrados Mínimos, que consiste em minimizar a soma do quadrado das distâncias entre os pontos (x_i, y_i) e $(x_i, f(x_i))$. Note que esses pares de pontos têm abscissas coincidentes, e diferem nas ordenadas. Considerando o vetor cujas componentes são os valores de $f(x_i)$ e um outro vetor y com coordenadas y_i , o problema é equivalente a minimizar a norma da diferença entre estes vetores, ou seja, minimizar:

$$D = \sqrt{\sum_{i=1}^n (f(x_i) - y_i)^2}. \quad (2.1)$$

Podemos considerar D como o erro absoluto dessa aproximação. Como minimizar $\|u\|$ é o mesmo que minimizar $\|u\|^2$, minimizar D equivale a minimizar

$$E = \sum_{i=1}^n (f(x_i) - y_i)^2, \quad (2.2)$$

que é a soma do quadrado das distâncias entre os pontos (x_i, y_i) e $(x_i, f(x_i))$. Na expressão do erro (2.2) poderíamos pensar em usar o valor absoluto de cada parcela do somatório ao invés do quadrado, no entanto, isso tornaria a expressão não diferenciável.

O erro relativo da aproximação é definido como a razão entre o erro absoluto D e a norma do vetor y , ou seja:

$$E_r = \frac{D}{\|y\|} = \frac{\sqrt{\sum_{i=1}^n (f(x_i) - y_i)^2}}{\sqrt{\sum_{i=1}^n y_i^2}}. \quad (2.3)$$

O erro relativo reflete melhor a qualidade da aproximação da função f em relação aos pontos dados pois leva em consideração a ordem de grandeza dos dados.

Neste capítulo discutimos diferentes modelos para a função f , tendo como principais referências [2, 8]. Deduzimos as expressões dos coeficientes para os modelos afim, exponencial, geométrico, polinomial e exponencial de uma quadrática. Para essas deduções usaremos a minimização do erro (2.2), enquanto o erro relativo (2.3) é usado para a análise dos resultados. Apresentamos um exemplo para cada modelo discutido, utilizando o software Excel para determinar os coeficientes a partir das expressões deduzidas e traçar os gráficos dos modelos.

2.1 Modelo Linear

Primeiro iremos trabalhar esta aproximação por:

$$f(x) = a_0 + a_1 x, \quad (2.4)$$

ou seja, vamos supor que ao analisar estes pares de pontos notamos que eles podem ser ajustados graficamente por uma reta. Neste caso o erro (2.2) é da forma:

$$E(a_0, a_1) = \sum_{i=1}^n (a_0 + a_1 x_i - y_i)^2. \quad (2.5)$$

Queremos então encontrar os coeficientes a_0 e a_1 de modo a minimizar $E(a_0, a_1)$. Um candidato a minimizar o erro deve anular as derivadas parciais de $E(a_0, a_1)$, ou seja,

$$\begin{aligned} \frac{\partial}{\partial a_0} E(a_0, a_1) &= \sum_{i=1}^n 2(a_0 + a_1 x_i - y_i) = 0 \\ \frac{\partial}{\partial a_1} E(a_0, a_1) &= \sum_{i=1}^n 2(a_0 + a_1 x_i - y_i)(x_i) = 0. \end{aligned}$$

Provamos, na Seção 2.4.2, que de fato esse candidato minimiza a função. Reescrevendo como um sistema nas variáveis a_0 e a_1 temos:

$$\begin{cases} n a_0 + \sum_{i=1}^n x_i a_1 = \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i a_0 + \sum_{i=1}^n x_i^2 a_1 = \sum_{i=1}^n x_i y_i \end{cases}. \quad (2.6)$$

Este sistema pode ser escrito na forma matricial

$$Az = b, \quad (2.7)$$

em que

$$A = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix} \quad z = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \quad \text{e} \quad b = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix}. \quad (2.8)$$

Resolvendo o sistema (2.7) através do método de Gauss-Jordan [9], vemos que se

$$\alpha = \det(A) = n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \neq 0, \quad (2.9)$$

a solução do sistema é única e é dada por:

$$a_0 = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{\alpha} \quad \text{e} \quad a_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\alpha}. \quad (2.10)$$

Apresentamos a seguir um exemplo de ajuste linear de um conjunto de pontos. Após o exemplo, iremos discutir o caso em que $\alpha = 0$.

Exemplo 1 A partir da Tabela 1 vamos determinar a reta que melhor se ajusta aos dados. Para obter a aproximação linear, basta determinar a_1 e a_0 pelas igualdades (2.10). Para tanto

x	1	2	3	4	5	6	7	8	9	10
y	1	3	4	3	7	8	10	12	10	15

Tabela 1 – Dados do Exemplo 1.

calculamos:

$$\sum_{i=1}^{10} x_i = 55, \quad \sum_{i=1}^{10} y_i = 73, \quad \sum_{i=1}^{10} x_i^2 = 385 \quad \text{e} \quad \sum_{i=1}^{10} x_i y_i = 520;$$

logo:

$$a_0 = -0,6 \quad \text{e} \quad a_1 = 1,436364.$$

Sendo assim o modelo linear que se ajusta aos dados pelo método dos quadrados mínimos é da forma:

$$f(x) = -0,6 + 1,436364x. \quad (2.11)$$

Os pontos dados na Tabela 1 estão representados por círculos na Figura 1. A curva contínua é o gráfico do modelo linear (2.11) que se ajusta aos dados. Os pontos de coordenadas $(x_i, f(x_i))$, para $i = 1, \dots, 10$, estão indicados por triângulos.

Das expressões (2.1) e (2.3) obtemos:

$$D = 3,727051 \quad \text{e} \quad E_r = 0,139189.$$

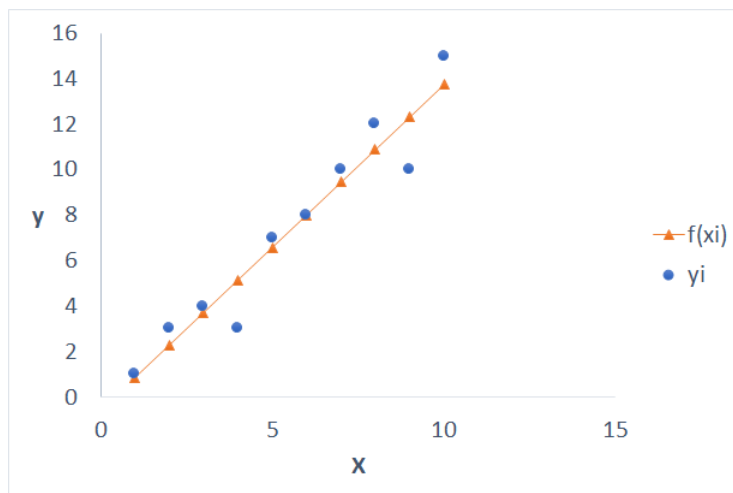


Figura 1 – Modelo linear para os dados da Tabela 1.

Caso em que há infinitas soluções.

Vimos que a solução do sistema (2.7) é única quando $\alpha \neq 0$. Veremos a seguir que esta condição não é satisfeita, se e somente se, as abscissas de todos os pontos coincidem. Note que na prática ter distintas medidas da ordenada para uma mesma abscissa, em geral, não acontece. Mas como nosso interesse está na formalização matemática, iremos provar este e outros resultados, que dependem de alguns conceitos de Álgebra Linear como veremos a seguir.

Definição 1. *Seja A uma matriz $n \times m$. O número de linhas não nulas após seu escalonamento é chamado de posto, e denotado por $\text{posto}(A)$. Ou seja, o posto de uma matriz é a dimensão do seu espaço linha.*

Segue desta definição que

$$\text{posto}(A) = \text{posto}(A^T A). \quad (2.12)$$

Definição 2. *O núcleo da matriz $A \in \mathbb{R}^{n \times m}$ é um subconjunto de \mathbb{R}^m formado pelas soluções de $Ax = 0$, e será denotado por $\mathcal{N}(A)$.*

Segue destas definições que

$$\mathcal{N}(A) = \mathcal{N}(A^T A) \quad \text{e} \quad \dim(\mathcal{N}(A)) + \text{posto}(A) = m. \quad (2.13)$$

A dimensão do núcleo coincide com o grau de liberdade do sistema. As demonstrações dos resultados (2.12) e (2.13) podem ser vistas em [4].

Lema 1. *O sistema (2.7) pode ser reescrito da forma*

$$X^T X z = X^T y,$$

onde

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad e \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}. \quad (2.14)$$

Demonstração. Trivial. □

O lema anterior será útil para provar os seguintes resultado.

Lema 2. *Considere o escalar α definido em (2.9). Assim $\alpha = 0$, se e somente se, $x_1 = x_2 = \dots = x_n$.*

Demonstração. Escalonando a matriz A temos:

$$\frac{1}{n} \begin{bmatrix} n & \sum_{i=1}^n x_i \\ 0 & \alpha \end{bmatrix}.$$

Então $\alpha = 0$ implica em $\text{posto}(A) = 1$. Pelo Lema 1 e (2.12), temos que

$$\text{posto}(A) = \text{posto}(X^T X) = \text{posto}(X).$$

Como os elementos da primeira coluna da matriz X são todos 1 e os elementos da segunda coluna são as abcissas dos pontos, temos que todas as abcissas coincidem. Para provar a volta basta notar que se $x_1 = x_2 = \dots = x_n$ pela definição de α dada pela expressão (2.9), temos $\alpha = 0$. □

O próximo lema mostra que quando as abcissas são coincidentes o sistema (2.7) é possível e há uma infinidade de retas que satisfazem a condição de quadrados mínimos. Além disso todas essas retas passam pelo ponto cuja ordenada é a média das ordenadas dos pontos dados.

Lema 3. *Suponha que todos os pontos (x_i, y_i) , $i = 1, \dots, n$, possuem a mesma abcissa, ou seja, $x_1 = x_2 = \dots = x_n$. Então, existem infinitas retas cujos coeficientes (a_0, a_1) minimizam o erro (2.5). Além disso todas estas retas passam pelo ponto (x_1, M) , onde $M = \frac{1}{n} \sum_{i=1}^n y_i$ é a ordenada média de todos os pontos considerados.*

Demonstração. Pela hipótese temos que $\alpha = 0$, pelo Lema 2 o escalonamento da matriz ampliada do sistema (2.7) resulta em:

$$\left[\begin{array}{cc|c} 1 & x_1 & \frac{\sum_{i=1}^n y_i}{n} \\ 0 & 0 & 0 \end{array} \right].$$

Ou seja, seu posto é 1 e conseqüentemente o grau de liberdade do sistema também é 1 indicando uma infinidade de soluções (a_0, a_1) que satisfazem $a_0 = M - a_1 x_1$. Substituindo na expressão do modelo (2.4), temos que

$$f(x) = M + a_1(x - x_1)$$

que são retas que passam pelo ponto (x_1, M) , para todo $a_1 \in \mathbb{R}$. □

2.2 Modelo Exponencial

Suponha que todos os valores de $y_i, i = 1, \dots, n$, têm o mesmo sinal. Neste caso, talvez seja interessante aproximar os pontos dados por um modelo exponencial da forma

$$f(x) = a_0 e^{a_1 x}. \quad (2.15)$$

Suponha, sem perda de generalidade, que todas as ordenadas y_i são positivas. Neste caso, a_0 será positivo e $f(x) > 0$, para todo $x \in \mathbb{R}$.

Desejamos determinar expressões para os coeficientes a_0 e a_1 de modo a minimizar o erro (2.2) dado, neste caso, por

$$E(a_0, a_1) = \sum_{i=1}^n (a_0 e^{a_1 x_i} - y_i)^2. \quad (2.16)$$

A fim de aproveitar o que foi desenvolvido na seção anterior, aplicamos logaritmo neperiano a ambos os lados de (2.15), obtendo:

$$\ln f(x) = \ln(a_0 e^{a_1 x}) = \ln a_0 + a_1 x.$$

Veja que recaímos em (2.4) com $\ln f$ no lugar de f e com coeficiente linear $\ln a_0$ ao invés de a_0 . Se $x_1 = \dots = x_n$, como visto no Lema 3 temos infinitas exponenciais que satisfazem o ajuste de quadrados mínimos. Caso contrário obtemos $\ln a_0$ e a_1 das igualdades (2.10), com

$\ln y_i$ no lugar de y_i e α definido em (2.9):

$$\ln a_0 = \frac{\sum_{i=1}^n \ln y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i \ln y_i}{\alpha} \quad \text{e} \quad (2.17)$$

$$\ln a_0 = \frac{\sum_{i=1}^n \ln y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i \ln y_i}{\alpha} .$$

Exemplo 2 A partir da Tabela 2 vamos determinar a curva exponencial que melhor se ajusta aos dados.

x	1	2	3	4	5
y	1	3	6	10	15

Tabela 2 – Dados do Exemplo 2.

Para tanto calculamos:

$$n = 5, \sum_{i=1}^5 x_i = 15, \sum_{i=1}^5 \ln y_i = 7,901007, \sum_{i=1}^5 x_i^2 = 55 \quad \text{e} \quad \sum_{i=1}^5 x_i \ln y_i = 30,32309.$$

Portanto das expressões (2.17):

$$a_1 = 0,662007 \quad \text{e} \quad \ln a_0 = -0,40582,$$

então $a_0 = 0,66643$. Substituindo em (2.15), temos o modelo exponencial que se ajusta aos dados pelo método dos quadrados mínimos:

$$f(x) = 0,66643 e^{0,662007x}. \quad (2.18)$$

Os pontos dados na Tabela 2 estão representados na Figura 2 por círculos. A curva contínua

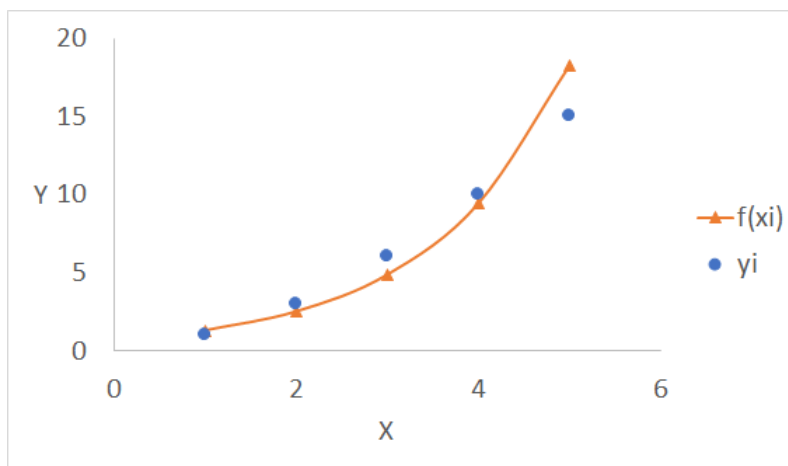


Figura 2 – Modelo exponencial para os dados da Tabela 2.

é o gráfico do modelo exponencial (2.18) que se ajusta aos dados. Os pontos de coordenadas $(x_i, f(x_i))$, para $i = 1, \dots, 5$, estão indicados por triângulos. Das expressões (2.1) e (2.3) obtemos os erros

$$D = 3,542772 \quad \text{e} \quad E_r = 0,183931.$$

Modelos exponenciais equivalentes

Discutimos a seguir modelos exponenciais equivalentes a (2.15). Considere o modelo exponencial com base qualquer, ou seja, da forma

$$f(x) = b_0 b_1^x. \quad (2.19)$$

Desejamos obter os termos b_0 e b_1 pelo método dos quadrados mínimos. Para tanto, basta notar que (2.19) pode ser reescrito como

$$f(x) = b_0 b_1^x = a_0 e^{a_1 x},$$

com $b_0 = a_0$ e $b_1 = e^{a_1}$. Assim $\ln b_1 = a_1$ e dessa forma, por (2.17):

$$\ln b_0 = \frac{\sum_{i=1}^n \ln y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i \ln y_i}{\alpha} \quad e \quad (2.20)$$

$$\ln b_1 = \frac{n \sum_{i=1}^n x_i \ln y_i - \sum_{i=1}^n \ln y_i \sum_{i=1}^n x_i}{\alpha},$$

com α definido em (2.9).

2.3 Modelo Geométrico

Suponha agora que todos os valores de x_i são positivos e y_i têm o mesmo sinal. Neste caso o ajuste dos pontos (x_i, y_i) pode ser feito por um modelo geométrico da forma:

$$f(x) = a_0 x^{a_1}. \quad (2.21)$$

Suponha, sem perda de generalidade, que todas as ordenadas y_i são positivas. Neste caso, a_0 será positivo e $f(x) > 0$ para todo $x \in \mathbb{R}$.

Precisamos determinar os coeficientes a_0 e a_1 de modo a minimizar o erro (2.2) dado, neste caso, por:

$$E(a_0, a_1) = \sum_{i=1}^n (a_0 x_i^{a_1} - y_i)^2. \quad (2.22)$$

Como na seção anterior aplicamos logaritmo neperiano a ambos os lados de (2.21), obtendo:

$$\ln f(x) = \ln a_0 + a_1 \ln x.$$

Veja que recaímos em (2.4) com $\ln f$ e $\ln x$ no lugar de f e x , respectivamente e com $\ln a_0$ ao invés de a_0 . Minimizando esse erro, como \ln é injetora, pelo Lema 3, temos que se $x_1 =$

$\dots = x_n$, então haverá infinitas curvas geométricas que satisfazem o ajuste de quadrados mínimos. Caso contrário obtemos $\ln a_0$ e a_1 das igualdades (2.10) com $\ln x_i$ e $\ln y_i$, no lugar de x_i e y_i , respectivamente:

$$\begin{aligned} \ln a_0 &= \frac{\sum_{i=1}^n \ln y_i \sum_{i=1}^n (\ln x_i)^2 - \sum_{i=1}^n \ln x_i \sum_{i=1}^n \ln x_i \ln y_i}{n \sum_{i=1}^n (\ln x_i)^2 - \left(\sum_{i=1}^n \ln x_i \right)^2}, \\ a_1 &= \frac{n \sum_{i=1}^n \ln x_i \ln y_i - \sum_{i=1}^n \ln y_i \sum_{i=1}^n \ln x_i}{n \sum_{i=1}^n (\ln x_i)^2 - \left(\sum_{i=1}^n \ln x_i \right)^2}. \end{aligned} \quad (2.23)$$

Exemplo 3 Considere os dados da Tabela 3. Vamos determinar a curva geométrica que melhor se aproxima a esses dados no sentido de quadrados mínimos.

x	1	2	3	4	5	6	7	8	9	10
y	0,5	1	3	4	4	5	8	10	14	15

Tabela 3 – Dados do Exemplo 3.

Neste caso $n = 10$. Para determinar os coeficientes a_0 e a_1 da curva (2.21), calculamos os somatórios auxiliares:

$$\begin{aligned} \sum_{i=1}^{10} \ln x_i &= 15,104412, & \sum_{i=1}^{10} \ln y_i &= 14,516626, \\ \sum_{i=1}^{10} (\ln x_i)^2 &= 27,650244, & \sum_{i=1}^{10} \ln x_i \ln y_i &= 29,112256. \end{aligned}$$

Substituindo em (2.23), obtemos

$$\ln a_0 = -0,792720 \quad \text{e} \quad a_1 = 1,485912,$$

e conseqüentemente

$$a_0 = 0,452612.$$

Sendo assim o modelo geométrico para os dados da Tabela 3 é da forma

$$f(x) = 0,452612 x^{1,485912}. \quad (2.24)$$

A Figura 3 mostra em círculos os pontos dados na Tabela 3 que foram ajustados pelo modelo geométrico (2.24) cujo gráfico é representado pela curva contínua. Os pontos de coordenadas $(x_i, f(x_i))$, para $i = 1, \dots, 10$, estão indicados por triângulos.

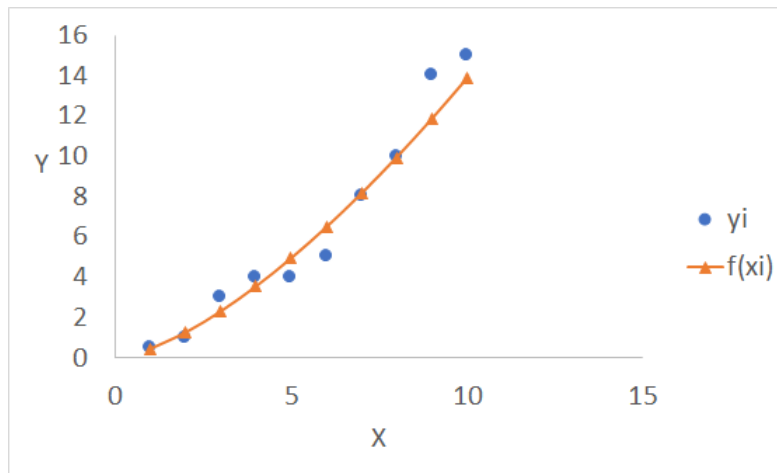


Figura 3 – Modelo geométrico para os dados da Tabela 3.

Das expressões (2.1) e (2.3) obtemos o valor dos erros:

$$D = 3,133082 \quad \text{e} \quad E_r = 0,122677$$

2.4 Modelo Polinomial

Nesta seção discutimos o modelo polinomial para ajuste dos pontos dados (x_i, y_i) , $i = 1, \dots, n$, através do método de quadrados mínimos. Para $m < n$, considere o modelo polinomial de grau m da forma:

$$f(x) = a_0 + a_1x + a_2x^2 + \dots + a_mx^m. \quad (2.25)$$

Desejamos determinar os coeficientes a_0, a_1, \dots, a_m de modo a minimizar o erro

$$E(a_0, a_1, \dots, a_m) = \sum_{i=1}^n (f(x_i) - y_i)^2. \quad (2.26)$$

Calculando as derivadas parciais $\frac{\partial E}{\partial a_j}$ e as igualando a zero, obtemos um sistema com $(m+1)$ equações. Observe que no caso linear $m = 1$ e portanto nosso sistema recaía em 2 equações. Vamos discutir a seguir o caso em que f é um polinômio quadrático. Na Seção 2.4.2 generalizamos para um polinômio de grau $m \geq 1$, usando uma linguagem matricial.

2.4.1 Modelo Quadrático

O modelo quadrático é da forma:

$$f(x) = a_0 + a_1x + a_2x^2. \quad (2.27)$$

De modo análogo ao caso linear queremos minimizar o erro (2.26), que se escreve como

$$E(a_0, a_1, a_2) = \sum_{i=1}^n (a_0 + a_1 x_i + a_2 x_i^2 - y_i)^2. \quad (2.28)$$

O candidato a minimizar o erro deve anular as derivadas parciais de E , assim obtemos o sistema:

$$\begin{cases} \sum_{i=1}^n (a_0 + a_1 x_i + a_2 x_i^2 - y_i) = 0 \\ \sum_{i=1}^n (a_0 + a_1 x_i + a_2 x_i^2 - y_i) x_i = 0 \\ \sum_{i=1}^n (a_0 + a_1 x_i + a_2 x_i^2 - y_i) x_i^2 = 0 \end{cases}.$$

Reescrevendo na forma matricial temos

$$Az = b, \quad (2.29)$$

onde

$$A = \begin{bmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \end{bmatrix}, \quad z = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \quad \text{e} \quad b = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i^2 y_i \end{bmatrix}. \quad (2.30)$$

O lema a seguir será útil para reescrevermos esse sistema de forma conveniente.

Lema 4. *Considere as matrizes $A \in \mathbb{R}^{m \times m}$ e $X \in \mathbb{R}^{n \times m}$, os vetores $b \in \mathbb{R}^m$ e $y \in \mathbb{R}^n$ e os conjuntos $S_1 = \{z \in \mathbb{R}^m \mid Az = b\}$ e $S_2 = \{w \in \mathbb{R}^m \mid Xw = y\}$. Se $A = X^T X$ e $b = X^T y$, então $S_1 = S_2$.*

Demonstração. Sendo $A = X^T X$, por (2.13) temos que

$$\mathcal{N}(A) = \mathcal{N}(X^T X) = \mathcal{N}(X).$$

Se t e $w \in S_2$, então $(t - w) \in \mathcal{N}(X) = \mathcal{N}(A)$, ou seja

$$\begin{aligned} A(t - w) &= 0 \\ At &= Aw \\ At &= X^T X w \\ At &= X^T y \\ At &= b, \end{aligned}$$

logo se $t \in S_2$ então $t \in S_1$, ou seja $S_2 \subseteq S_1$.

Reciprocamente se $z \in S_1$ e $w \in S_2 \subseteq S_1$ então $A(z - w) \in \mathcal{N}(A) = \mathcal{N}(X)$, logo

$$\begin{aligned} X(z - w) &= 0 \\ Xz &= Xw \\ Xz &= y. \end{aligned}$$

Portanto $S_1 \subseteq S_2$, isso implica em $S_1 = S_2$. \square

Note que

$$A = X^T X \quad \text{e} \quad b = X^T y$$

com,

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \quad \text{e} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}. \quad (2.31)$$

Disto e do Lema 4, temos que o sistema (2.29) é equivalente ao sistema

$$Xz = y.$$

O próximo resultado apresenta o caso em que a solução do sistema é única.

Lema 5. *O sistema (2.29) é possível e determinado, se e somente se, existem pelo menos três abscissas distintas entre os pontos dados.*

Demonstração. O sistema (2.29) é possível e determinado, se e somente se, o posto da matriz A e da matriz ampliada do sistema é 3. Mas

$$\text{posto}(A) = \text{posto}(X^T X) = \text{posto}(X) = 3,$$

o que equivale a existir pelo menos três abscissas distintas. \square

Os dois próximos lemas discutem a situação em que há infinitas soluções. O lema a seguir mostra que se as abscissas de todos os pontos dados coincidem, então existem infinitas quadráticas que satisfazem o ajuste de quadrados mínimos, sendo que todas elas passam pelo ponto desta abscissa coincidente e ordenada média dos pontos dados.

Lema 6. *Suponha que os pontos (x_i, y_i) , $i = 1, \dots, n$, possuem a mesma abscissa, ou seja, $x_1 = \dots = x_n$. Então existem infinitas curvas quadráticas cujos coeficientes (a_0, a_1, a_2) minimizam o erro (2.28), ou seja, o sistema (2.29) é possível e indeterminado, com solução*

$$a_0 = M - a_1 x_1 - a_2 x_1^2 \quad \text{com} \quad M = \frac{1}{n} \sum_{i=1}^n y_i,$$

e a_1, a_2 arbitrários. Além disso, todas as quadráticas passam pelo ponto (x_1, M) .

Demonstração. Pelos Lema 4 para encontrar as soluções de $Az = b$ basta olhar para o sistema $Xz = y$, pois ambos tem o mesmo conjunto de soluções. Por hipótese $\text{posto}(X) = 1$ então, podemos reescrever o sistema $Az = y$ da seguinte forma:

$$\begin{aligned} n(a_0 + a_1 x_1 + a_2 x_1^2) &= \sum_{i=1}^n y_i \\ a_0 + a_1 x_1 + a_2 x_1^2 &= M. \end{aligned} \quad (2.32)$$

Portanto o sistema possui infinitas soluções com

$$a_0 = M - a_1 x_1 - a_2 x_1^2,$$

e o modelo quadrático é da forma:

$$M + a_1(x - x_1) + a_2(x^2 - x_1^2) = y, \quad (2.33)$$

e todas as curvas passam por (x_1, M) . \square

O lema a seguir garante que se os pontos estão distribuídos em apenas duas abscissas distintas, então existem infinitas quadráticas que satisfazem o ajuste de quadrados mínimos e que passam pelos pontos dessas abscissas e ordenadas médias correspondentes.

Lema 7. *Suponha que os pontos (x_i, y_i) , $i = 1, \dots, n$, estão em duas abscissas, ou seja, $x_1 = \dots = x_p \neq x_{p+1} = \dots = x_n$. Então existem infinitas curvas quadráticas cujos coeficientes (a_0, a_1, a_2) minimizam o erro (2.28), ou seja o sistema é possível e indeterminado, com solução*

$$a_0 = M_2 - W x_n + a_2 x_1 x_n \quad e \quad a_1 = W - a_2(x_1 + x_n).$$

Além disso todas estas curvas passam pelos pontos (x_1, M_1) e (x_n, M_2) , onde $M_1 = \frac{1}{p} \sum_{i=1}^p y_i$ e $M_2 = \frac{1}{n-p} \sum_{i=p+1}^n y_i$ são as ordenadas médias dos pontos considerados.

Demonstração. Como vimos, pelo Lema 4 os sistemas $Az = b$ e $Xz = y$ têm o mesmo conjunto de soluções. Por hipótese $\text{posto}(X) = 2$, logo o sistema $Xz = y$ equivale a:

$$\begin{cases} a_0 + x_1 a_1 + x_1^2 a_2 = \frac{1}{p} \sum_{i=1}^p y_i \\ a_0 + x_n a_1 + x_n^2 a_2 = \frac{1}{n-p} \sum_{i=p+1}^n y_i \end{cases} \quad (2.34)$$

Portanto se $x_1 = \dots = x_p \neq x_{p+1} = \dots = x_n$, o sistema é possível e indeterminado. Sendo $W = \frac{M_1 - M_2}{x_1 - x_n}$ as soluções do sistema $Xz = y$ são

$$a_0 = M_2 - W x_n + a_2 x_1 x_n \quad \text{e} \quad a_1 = W - a_2(x_1 + x_n),$$

e o modelo quadrático é da forma

$$y = (M_2 - W x_n + a_2 x_1 x_n) + (W - a_2(x_1 + x_n))x + a_2 x^2 \quad (2.35)$$

Ou seja, todas as soluções passam por (x_1, M_1) e (x_n, M_2) . \square

A seguir, vamos apresentar as expressões dos coeficientes a_0, a_1, a_2 em todos os casos possíveis. Para tanto, considere a matriz ampliada $[A|b]$ escalonada

$$\frac{1}{n} \begin{bmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \vdots & \sum_{i=1}^n y_i \\ 0 & \alpha & \beta & \vdots & \theta \\ 0 & \beta & \gamma & \vdots & \eta \end{bmatrix}, \quad (2.36)$$

em que α está definido em (2.9) e

$$\begin{aligned} \beta &= n \sum_{i=1}^n x_i^3 - \sum_{i=1}^n x_i \sum_{i=1}^n x_i^2, & \gamma &= n \sum_{i=1}^n x_i^4 - \left(\sum_{i=1}^n x_i^2 \right)^2, \\ \theta &= n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \quad \text{e} & \eta &= n \sum_{i=1}^n x_i^2 y_i - \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i. \end{aligned} \quad (2.37)$$

Temos alguns casos a considerar:

Caso 1 $\alpha = 0$. Pela expressão (2.9), segue que o quadrado da segunda coluna é igual ao produto da primeira e terceira colunas da matriz (2.36). De onde segue que $\beta = 0$. Vamos provar que $\gamma = 0$. Suponha por contradição que $\gamma \neq 0$. Neste caso, $\text{posto}(A) = 2$ e como $A = X^T X$, $\text{posto}(X) = 2$. Sendo assim $x_1 = \dots = x_p \neq x_{p+1} = \dots = x_n$. Usando isto e o fato de $\alpha = 0$, temos que

$$0 = \alpha = n(p x_1^2 + (n-p)x_n^2) - (p x_1 + (n-p)x_n)^2 = (np - p^2)(x_1 - x_n)^2,$$

de onde segue que $x_1 = x_n$ e conseqüentemente $x_i = x_1$, para todo $i = 1, \dots, n$ e $\text{posto}(X) = 1$, o que é uma contradição. Assim, nesse caso, $\alpha = \beta = \gamma = 0$, e segue do Lema 6 que o sistema tem infinitas soluções da forma

$$a_0 = \frac{1}{n} \sum_{i=1}^n y_i - a_1 x_1 - a_2 x_1^2,$$

com a_1, a_2 arbitrários.

Caso 2 $\alpha \neq 0$. A matriz ampliada (2.36) se reescreve como:

$$\frac{1}{n} \begin{bmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \vdots & \sum_{i=1}^n y_i \\ 0 & 1 & \frac{\beta}{\alpha} & \vdots & \frac{\theta}{\alpha} \\ 0 & 0 & \gamma - \frac{\beta^2}{\alpha} & \vdots & \eta - \frac{\theta \beta}{\alpha} \end{bmatrix}. \quad (2.38)$$

- Se $\gamma - \frac{\beta^2}{\alpha} = 0$, neste caso $\text{posto}(A) = 2$ e por (2.12) $\text{posto}(X) = 2$. Segue do Lema 7 que há infinitas soluções da forma

$$a_0 = M_2 - W x_n + a_2 x_1 x_n \quad \text{e} \quad a_1 = W - a_2(x_1 + x_n),$$

com $W = \frac{M_1 - M_2}{x_1 - x_n}$ e a_2 arbitrário.

- Se $\gamma - \frac{\beta^2}{\alpha} \neq 0$ o sistema é possível e determinado. Temos então:

$$a_2 = \frac{\alpha \eta - \theta \beta}{\alpha \gamma - \beta^2}, \quad a_1 = \frac{\theta - a_2 \beta}{\alpha} \quad \text{e} \\ a_0 = n \sum_{i=1}^n y_i - \frac{n \theta}{\alpha} \sum_{i=1}^n x_i - a_2 \left(\sum_{i=1}^n x_i^2 - \frac{\beta}{\alpha} \sum_{i=1}^n x_i \right). \quad (2.39)$$

Apresentamos a seguir um exemplo de ajuste quadrático de um conjunto de pontos.

Exemplo 4 Considere os dados da Tabela 4 para que sejam ajustados por um polinômio quadrático.

x_i	0	0,25	0,5	0,75	1
y_i	2,7183	1,284	1	2	2,7183

Tabela 4 – Dados do Exemplo 4.

Neste caso $m = 2$ e $n = 5$. Calculamos os somatórios auxiliares:

$$\sum_{i=1}^5 x_i = 2,5, \quad \sum_{i=1}^5 y_i = 9,7206, \quad \sum_{i=1}^5 x_i^2 = 1,875, \quad \sum_{i=1}^5 x_i^3 = 1,5625, \\ \sum_{i=1}^5 x_i^4 = 1,382812, \quad \sum_{i=1}^5 x_i y_i = 5,0393 \quad \text{e} \quad \sum_{i=1}^5 x_i^2 y_i = 4,17355.$$

Das igualdades (2.39) temos

$$a_0 = 2,599377, \quad a_1 = -6,101257 \quad \text{e} \quad a_2 = 6,387657.$$

Portanto o modelo quadrático que se ajusta aos dados é da forma

$$f(x) = 2,599377 - 6,101257x + 6,387657x^2. \quad (2.40)$$

A Figura 4 mostra em círculos os pontos dados na Tabela 4 que foram ajustados pelo modelo quadrático (2.40) cujo gráfico é representado pela curva contínua. Os pontos de coordenadas $(x_i, f(x_i))$, para $i = 1, \dots, 5$, estão indicados por triângulos.

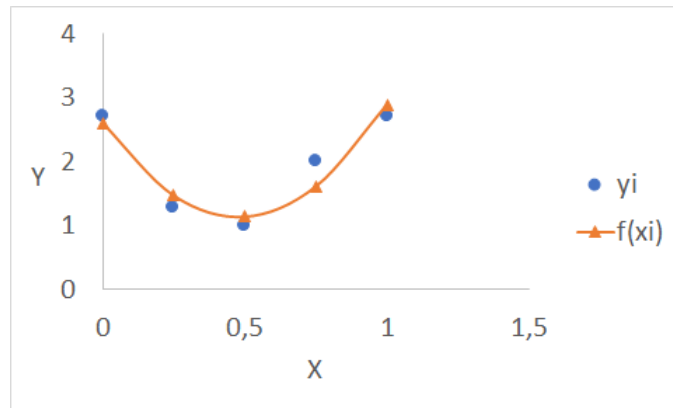


Figura 4 – Modelo quadrático para os dados da Tabela 4.

De (2.1) e (2.3) obtemos os valores dos erros dessa aproximação:

$$D = 0,496305 \quad \text{e} \quad E_r = 0,053213.$$

2.4.2 Solução Matricial

Nesta seção, generalizamos a obtenção dos coeficientes do polinômio (2.25) para $m \geq 1$, usando uma linguagem matricial. Temos n pontos (x_i, y_i) com $x_i, y_i \in \mathbb{R}$ e queremos determinar uma função polinomial $f : \mathbb{R} \rightarrow \mathbb{R}$ de grau $m < n$ que se aproxime destes n pontos, sendo:

$$f(x) = a_0 + a_1x + \dots + a_mx^m.$$

Desejamos determinar os coeficientes a_0, a_1, \dots, a_m de modo a minimizar o erro

$$E(a_0, a_1, \dots, a_m) = \sum_{i=1}^n (f(x_i) - y_i)^2.$$

Considerando:

$$X = \begin{bmatrix} 1 & x_1 & \cdots & x_1^m \\ 1 & x_2 & \cdots & x_2^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^m \end{bmatrix}, \quad z = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} \quad \text{e} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

a expressão do erro E se escreve na forma matricial como

$$E = \|Xz - y\|^2, \quad (2.41)$$

em que $\|\cdot\|$ representa a norma euclidiana. Como $\|v\|^2 = v^T v$, para qualquer vetor v , temos que

$$\begin{aligned} E &= \|Xz - y\|^2 \\ &= (Xz - y)^T (Xz - y) \\ &= (z^T X^T - y^T) (Xz - y) \\ &= y^T y - y^T Xz - z^T X^T y + z^T X^T Xz \\ &= z^T X^T Xz - 2(X^T y)^T z + y^T y. \end{aligned}$$

A última igualdade segue do fato da comutatividade do produto interno, ou seja, $y^T Xz = (X^T y)^T z$ e $z^T X^T y = (Xz)^T y = y^T Xz$; pois (Xz) e y são vetores do \mathbb{R}^n .

Para determinar os candidatos a minimizadores da função erro E basta calcular os pontos que anulam o gradiente de E , sendo assim:

$$\nabla E = 2X^T Xz - 2X^T y = 0$$

Se $X^T X$ é invertível, temos que o minimizador global de E é dado por

$$z = (X^T X)^{-1} X^T y. \quad (2.42)$$

Assim os coeficientes do modelo polinomial são obtidos pela expressão matricial acima.

Abordagem Geométrica

Vimos que uma maneira de encontrar z é calculando os pontos que anulam o gradiente E . Faremos uma abordagem geométrica para compreender o que acontece com os vetores e verificar se a solução encontrada é mínima. Vamos revisar, inicialmente, alguns conceitos de Álgebra Linear.

Definição 3. O conjunto dos vetores Ax é o espaço imagem de A , denotado por $Im(A)$.

$$Im(A) = \{Ax \mid x \in \mathbb{R}^m\}.$$

Dada $X \in \mathbb{R}^{n \times m}$ e $y \in \mathbb{R}^n$, queremos determinar $z \in \mathbb{R}^m$ que minimize $\|Xz - y\|$. Em outras palavras, queremos determinar o vetor do espaço $Im(X)$ de forma que o vetor $(Xz - y)$ seja de mínima norma. Basta então projetar y no espaço $Im(X)$. A Figura (5) ilustra a projeção de um vetor $y \in \mathbb{R}^3$ no espaço imagem de uma matriz X representado por um plano.

Note que o vetor $(Xz - y)$ é ortogonal ao espaço $Im(X)$, e portanto ortogonal a todos os vetores da $Im(X)$. Cada coluna w_k de X pertence a $Im(X)$, para provar basta

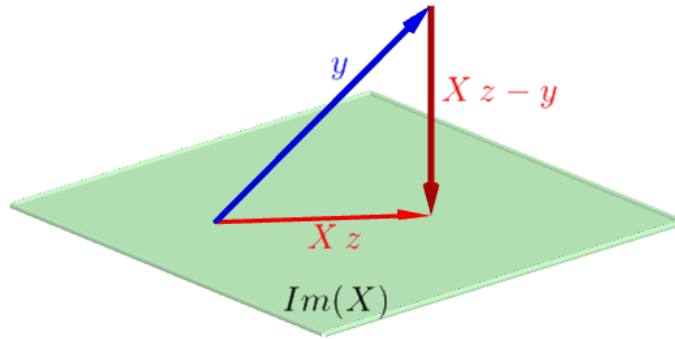


Figura 5 – Projeção do vetor $y \in \mathbb{R}^3$ na imagem de $X \in \mathbb{R}^{3 \times 2}$, com $z \in \mathbb{R}^2$.

tomar v_k , o vetor com 1 na posição k e 0 nas demais, o produto $X \cdot v_k$ está na $Im(X)$ e resulta na coluna w_k . Assim, $(Xz - y)$ deve ser ortogonal a todas as colunas de X . Ser ortogonal implica no produto interno usual entre $(Xz - y)$ e w_k igual a zero. Como $w_k \in \mathbb{R}^n$, temos para todo $k = 1, \dots, m$,

$$w_k^T (Xz - y) = 0.$$

O que equivale a

$$\begin{aligned} X^T (Xz - y) &= 0 \\ X^T Xz &= X^T y. \end{aligned}$$

Se $X^T X$ é invertível, recaímos na expressão (2.42), ou seja $z = (X^T X)^{-1} X^T y$. Abrindo a expressão:

$$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \left(\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^m & x_2^m & \cdots & x_n^m \end{bmatrix} \cdot \begin{bmatrix} 1 & x_1 & \cdots & x_1^m \\ 1 & x_2 & \cdots & x_2^m \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^m \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^m & x_2^m & \cdots & x_n^m \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Em particular no caso linear:

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \left(\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix},$$

ou seja,

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix};$$

que recai nas expressões (2.10). De forma análoga podemos achar os coeficientes para equação quadrática, e também chegaríamos nas igualdades (2.39).

O próximo lema mostra quando $X^T X$ é invertível.

Lema 8. *Seja X uma matriz $n \times m + 1$. Temos que $X^T X$ é invertível, se e somente se, $\text{posto}(X) = m + 1$.*

Demonstração. Temos que $X^T X$ é invertível, se e somente se, $X^T X z = 0$ admite apenas a solução trivial. Ou seja, $\mathcal{N}(X^T X) = \{0\}$. Consequentemente, por (2.13) $\mathcal{N}(X) = \{0\}$, o que equivale a $\text{posto}(X) = m + 1$. \square

Como vimos nas Seções 2.1 e 2.4.1, para o $\text{posto}(X)$ ser menor que $m + 1$ vai depender da disposição dos pontos. Se o número de linhas linearmente independentes da matriz X for menor que $(m + 1)$ a matriz não será invertível e teremos infinitas soluções que satisfazem o ajuste de quadrados mínimos.

2.5 Modelo exponencial de uma quadrática

Nesta seção discutimos o modelo exponencial de uma quadrática para ajuste dos pontos dados, da forma

$$f(x) = e^{a_0 + a_1 x + a_2 x^2}. \quad (2.43)$$

Do mesmo modo que no caso da exponencial, aplicamos o logaritmo na base e :

$$\begin{aligned} \ln f(x) &= \ln e^{a_0 + a_1 x + a_2 x^2} \\ &= a_0 + a_1 x + a_2 x^2 \end{aligned}$$

Assim recaímos no modelo quadrático (2.27), mas com $\ln f$ no lugar de f . Cabe ressaltar que pelo Lema 8, se $\text{posto}(X) \leq 2$ teremos infinitas curvas que vão satisfazer o ajuste. Mas se $\text{posto}(X) = 3$ temos que a solução do do sistema é única. Neste caso, obtemos as expressões dos coeficientes através de (2.39), substituindo y_i por $\ln y_i$. Os coeficientes são dados por

$$\begin{aligned} a_2 &= \frac{\alpha \eta - \theta \beta}{\alpha \gamma - \beta^2}, \quad a_1 = \frac{\theta - a_2 \beta}{\alpha} \quad e \\ a_0 &= n \sum_{i=1}^n \ln y_i - \frac{n \theta}{\alpha} \sum_{i=1}^n x_i - a_2 \left(\sum_{i=1}^n x_i^2 - \frac{\beta}{\alpha} \sum_{i=1}^n x_i \right), \end{aligned} \quad (2.44)$$

em que β, γ, θ e η são definidos em (2.37).

Exemplo 5 Considere os dados da Tabela 5 para o ajuste a uma exponencial de uma quadrática.

x_i	1	2	3	4	5	6	7	8	9	10
y_i	1	1,3	1,5	1,7	2	1,5	2,4	2,7	3	4,5

Tabela 5 – Dados do Exemplo 5.

Calculamos os somatórios auxiliares:

$$\sum_{i=1}^{10} x_i = 55, \quad \sum_{i=1}^{10} \ln y_i = 6,768480, \quad \sum_{i=1}^{10} x_i^2 = 385, \quad \sum_{i=1}^{10} x_i^3 = 3025,$$

$$\sum_{i=1}^{10} x_i^4 = 25333, \quad \sum_{i=1}^{10} x_i \ln y_i = 48,764743 \quad \text{e} \quad \sum_{i=1}^{10} x_i^2 \ln y_i = 390,975535;$$

Substituindo nas expressões (2.44) temos

$$a_0 = 0,198882, \quad a_1 = 0,000161 \quad \text{e} \quad a_2 = 0,012392.$$

Sendo assim o modelo exponencial da quadrática que se ajusta aos pontos é da forma:

$$f(x) = e^{0,198882+0,000161x+0,012392x^2}. \quad (2.45)$$

A Figura 6 apresenta em círculos os pontos dados na Tabela 5. A curva contínua representa o modelo (2.45) e os triângulos representam os pontos de coordenadas $(x_i, f(x_i))$ para $i = 1, \dots, 10$.

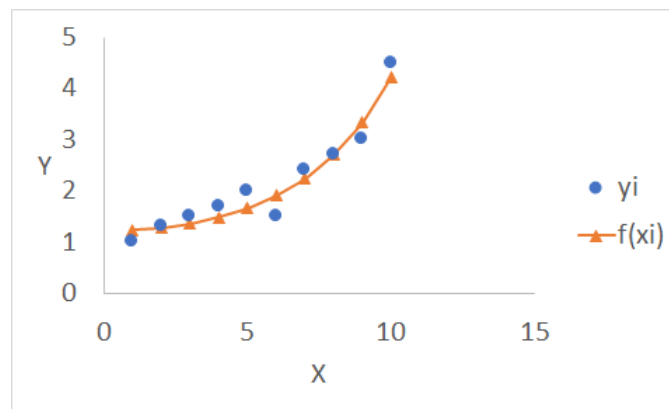


Figura 6 – Modelo exponencial de uma quadrática aos dados da Tabela 5.

De (2.1) e (2.3) obtemos os erros:

$$D = 0,782871 \quad \text{e} \quad E_r = 0,010871.$$

3 Aplicação à modelagem dos dados da COVID-19

Neste capítulo, aplicamos os ajustes de curvas pelo método dos quadrados mínimos apresentados no Capítulo 2 para modelar o número de casos confirmados e de mortes semanais relacionados à COVID-19 em Curitiba, Paraná e no Brasil.

Os dados da cidade de Curitiba foram extraídos da página da Secretaria Municipal da Saúde de Curitiba [11], os dados do estado foram retirados da página da Secretaria da Saúde do Estado do Paraná [10] e os dados do país foram obtidos da página do Ministério da Saúde [7]. Foram considerados os dados semanais visto que os diários são afetados sobretudo pela falta de registro durante os finais de semana, provocando um aumento irreal no número de casos e mortes no início da semana.

Assim, neste estudo, a abscissa x_i corresponde à i -ésima semana considerada, enquanto a ordenada y_i corresponde ao dado daquela semana. Para obter os coeficientes dos modelos: linear (2.4), exponencial (2.15), geométrico (2.21), quadrático (2.27) e exponencial de uma quadrática (2.43) para ajuste dos dados pelo método dos quadrados mínimos, foi utilizado o software Excel.

3.1 Curitiba

Nesta seção vamos aplicar os ajustes de curvas aos dados semanais de COVID em Curitiba. Na primeira subseção discutimos os modelos para o número de casos confirmados; na segunda, para o número de óbitos e encerramos a seção com a análise dos erros relativos obtidos e as expressões dos modelos que melhor se ajustaram aos dados.

3.1.1 Número de casos confirmados

A Figura 7, extraída de [11], exibe os números oficiais de casos confirmados de COVID-19 na cidade de Curitiba até 29/08/2020. Como as informações da última semana aparecem incompletas, consideramos as 25 semanas do período de 01/03 a 22/08/2020. Com base nesses dados, discutimos os modelos em duas situações. Na primeira situação, consideramos os dados de todas essas semanas, enquanto na segunda situação foram adotados os dados das últimas 15 semanas.

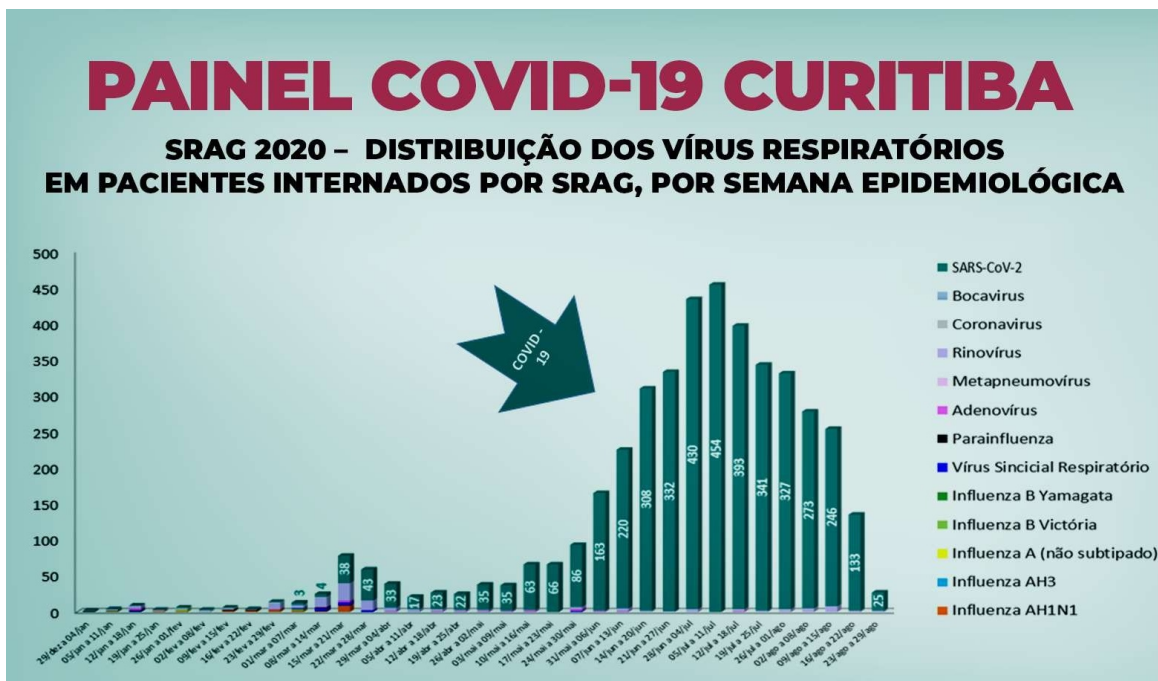


Figura 7 – Número de casos confirmados de COVID-19 em Curitiba até 29/08/2020 [11].

A Figura 8 mostra os modelos obtidos para o número de casos semanais confirmados por COVID-19 na cidade de Curitiba no período de 01/03 a 22/08/2020. As abscissas representam as semanas e as ordenadas, o número de casos semanais. Os pontos em azul indicam os dados fornecidos pela Figura 7. Em cada um dos gráficos está representado um dos modelos estudados que melhor se ajustou aos dados pelo método dos quadrados mínimos.

A Figura 9 mostra os modelos obtidos para o número de casos semanais confirmados por COVID-19 na cidade de Curitiba para as 15 últimas semanas, ou sejam, no período de 10/05 a 22/08/2020. As abscissas representam as semanas e as ordenadas, o número de casos semanais. Os pontos em azul indicam os dados fornecidos pela Figura 7. Em cada um dos gráficos está representado um dos modelos estudados que melhor se ajustou aos dados pelo método dos quadrados mínimos.

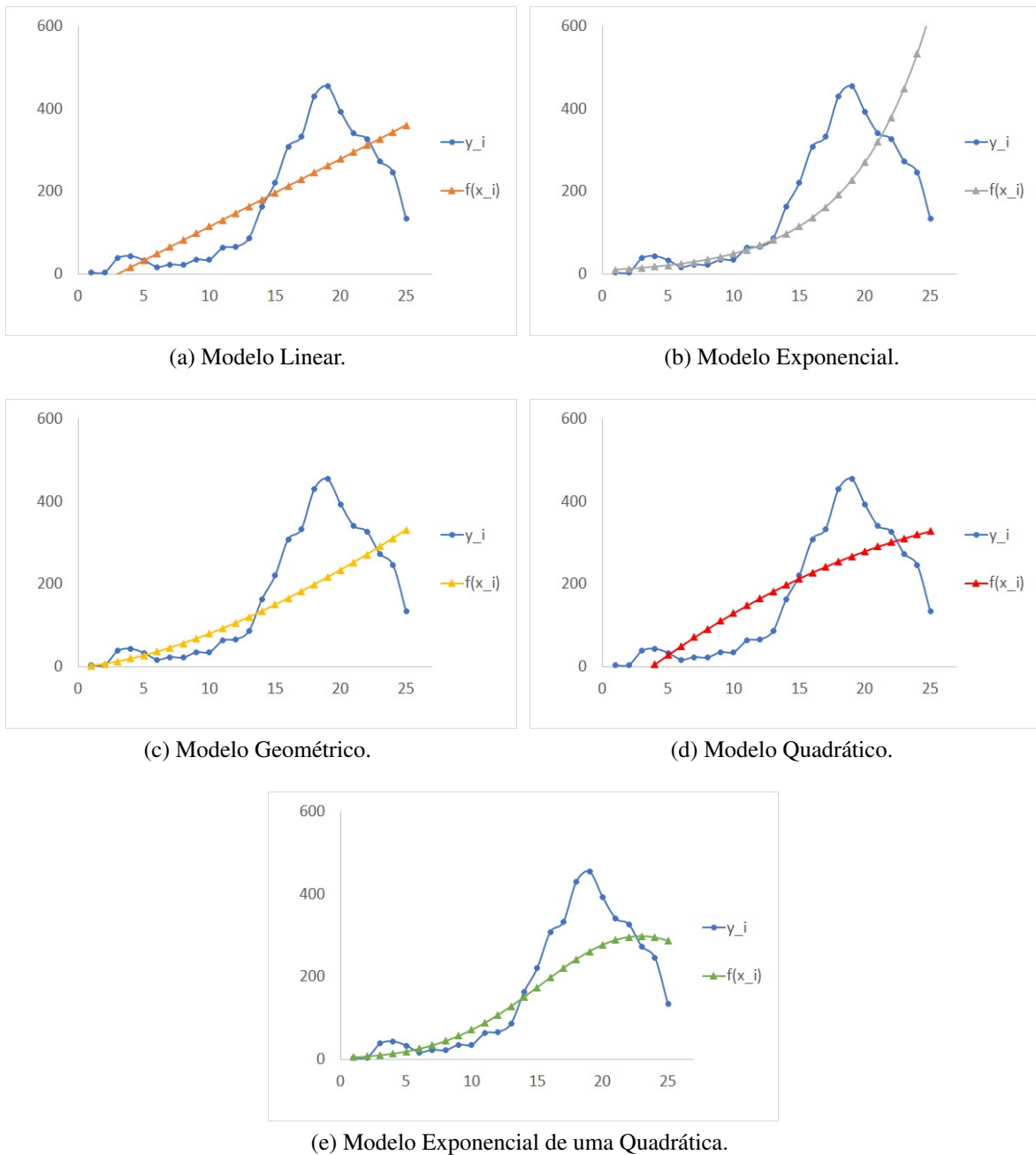


Figura 8 – Ajuste de curvas aos dados semanais de casos confirmados de COVID-19 em Curitiba.

Comparando os gráficos das Figuras 8 e 9, vemos que o modelo que melhor se adequou ao número de casos confirmados em Curitiba foi a exponencial de uma quadrática, representado no item (e) da Figura 9.

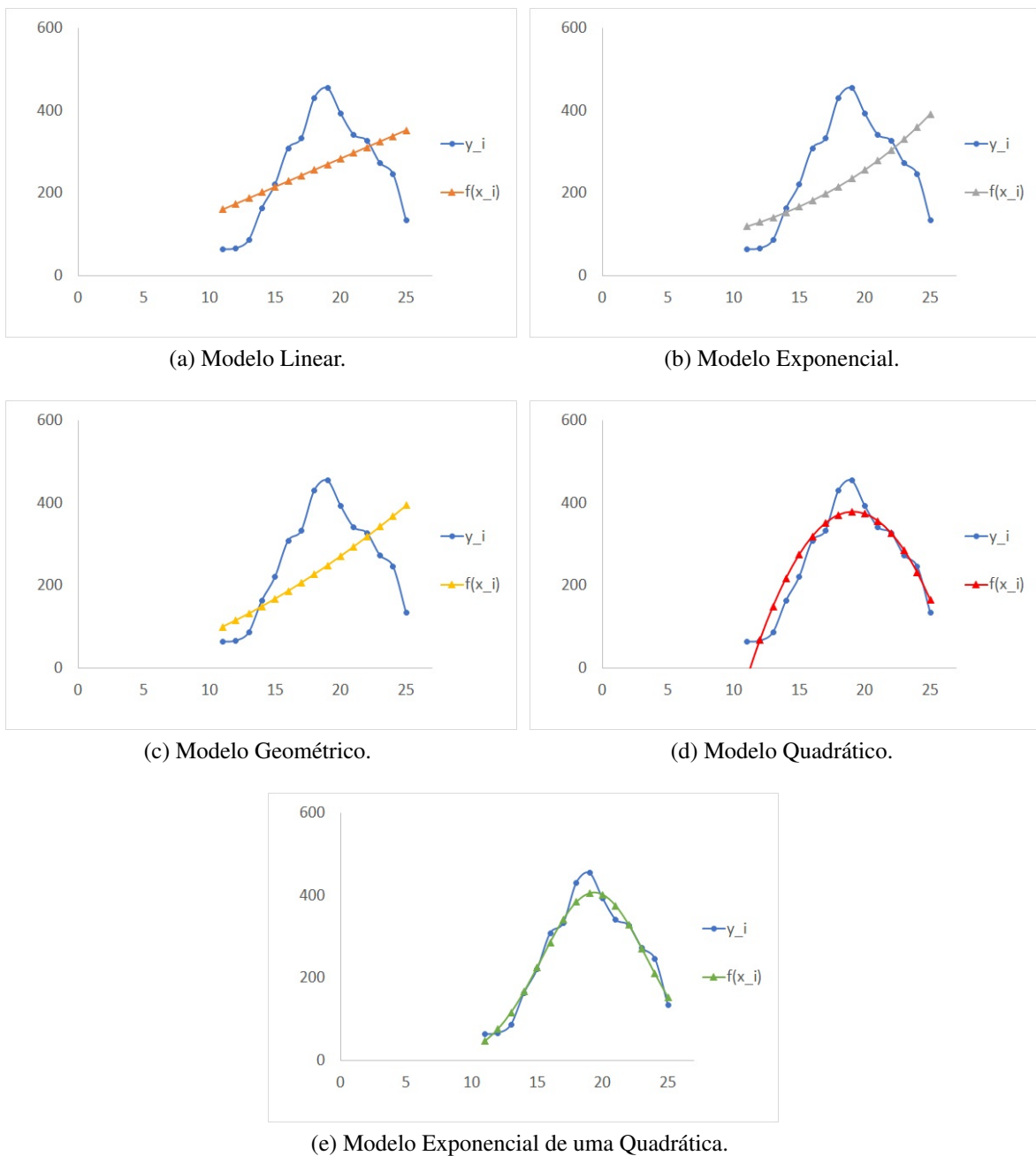


Figura 9 – Ajuste de curvas de casos de COVID-19 das 15 últimas semanas em Curitiba.

3.1.2 Número de óbitos

A Figura 10, extraída de [11], exibe os números oficiais de óbitos semanais por COVID-19 na cidade de Curitiba até 05/09/2020. Como as informações da última semana aparecem incompletas, foram considerados os dados das 21 semanas compreendidas no período de 05/04 a 29/08/2020.

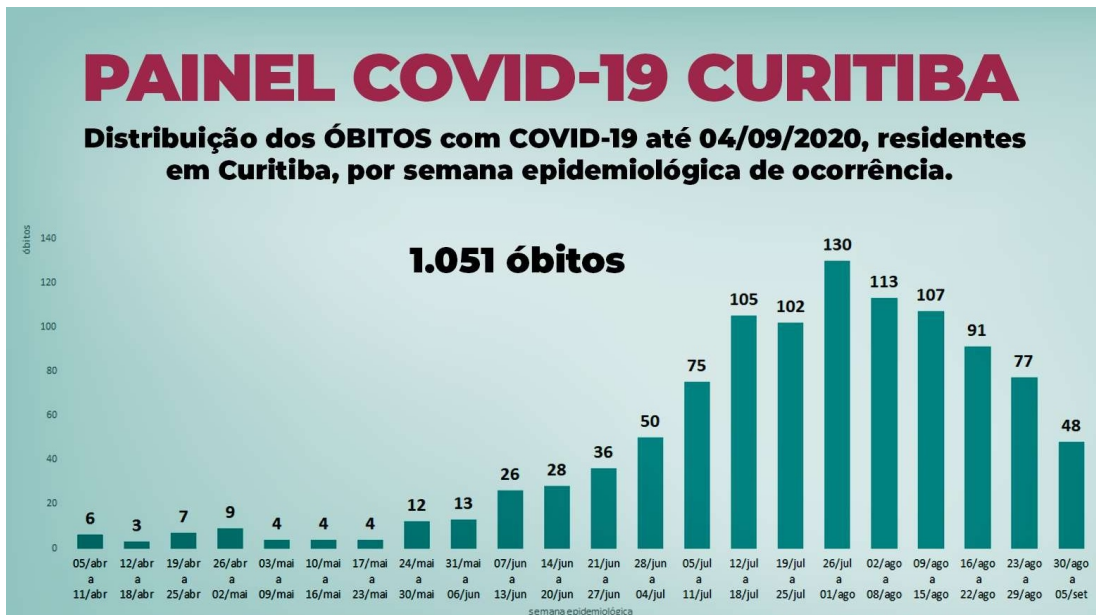


Figura 10 – Número semanal de óbitos por COVID-19 em Curitiba [11].

Com base nesses dados, discutimos os modelos em duas situações. Na primeira situação, consideramos os dados de todas essas semanas, enquanto na segunda situação foram adotados os dados das últimas 15 semanas.

A Figura 11 mostra os modelos para o número de óbitos em Curitiba para as 21 semanas de 05/04 a 29/08/2020. Os modelos se ajustam mais adequadamente aos dados das primeiras semanas, no entanto, nas últimas eles apresentam uma previsão muito além da realidade.

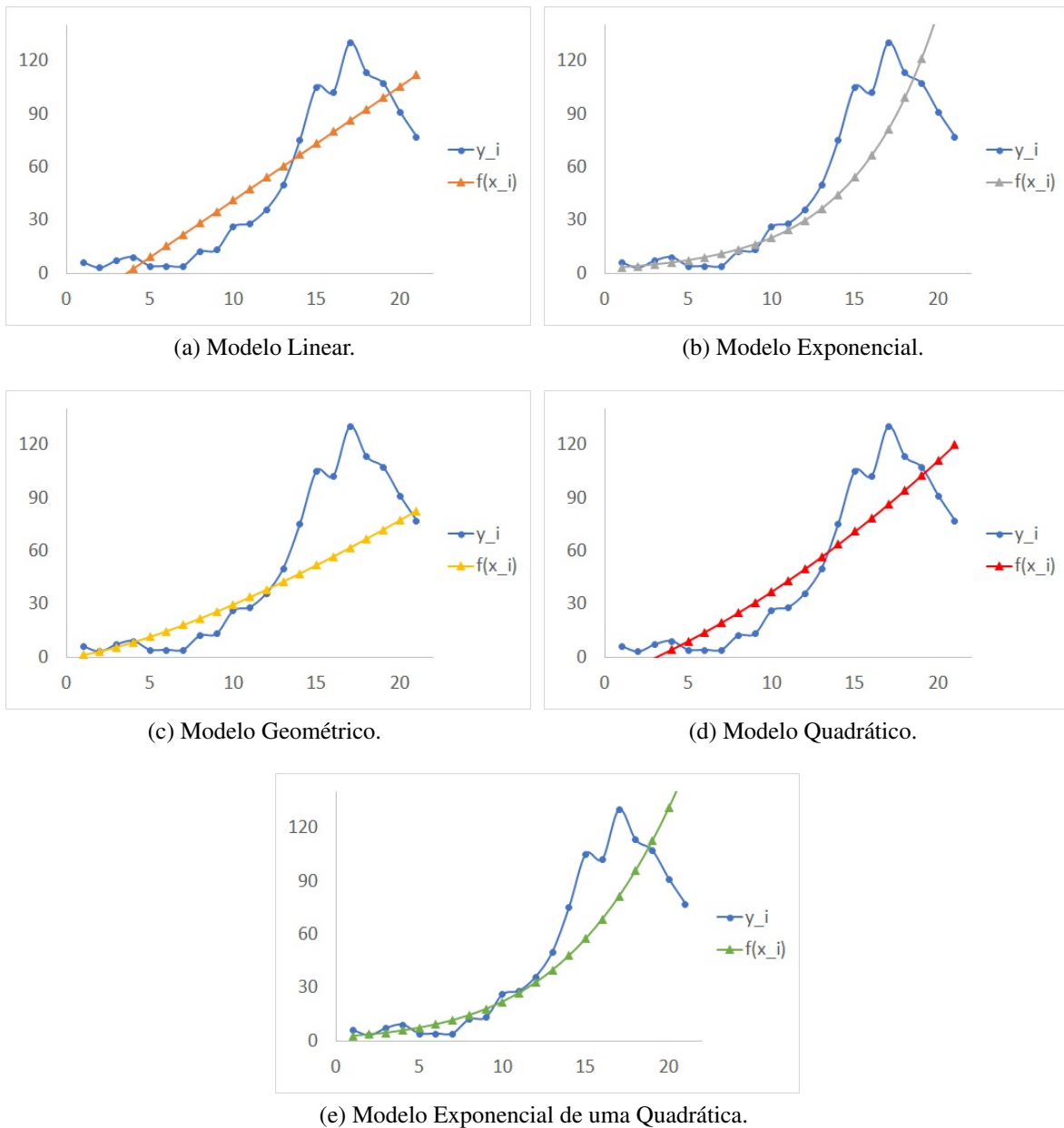


Figura 11 – Ajuste de curvas aos dados de óbitos semanais por COVID-19 em Curitiba.

A Figura 12 mostra os modelos para o número de óbitos em Curitiba para as últimas 15 semanas de 17/05 a 29/08/2020.

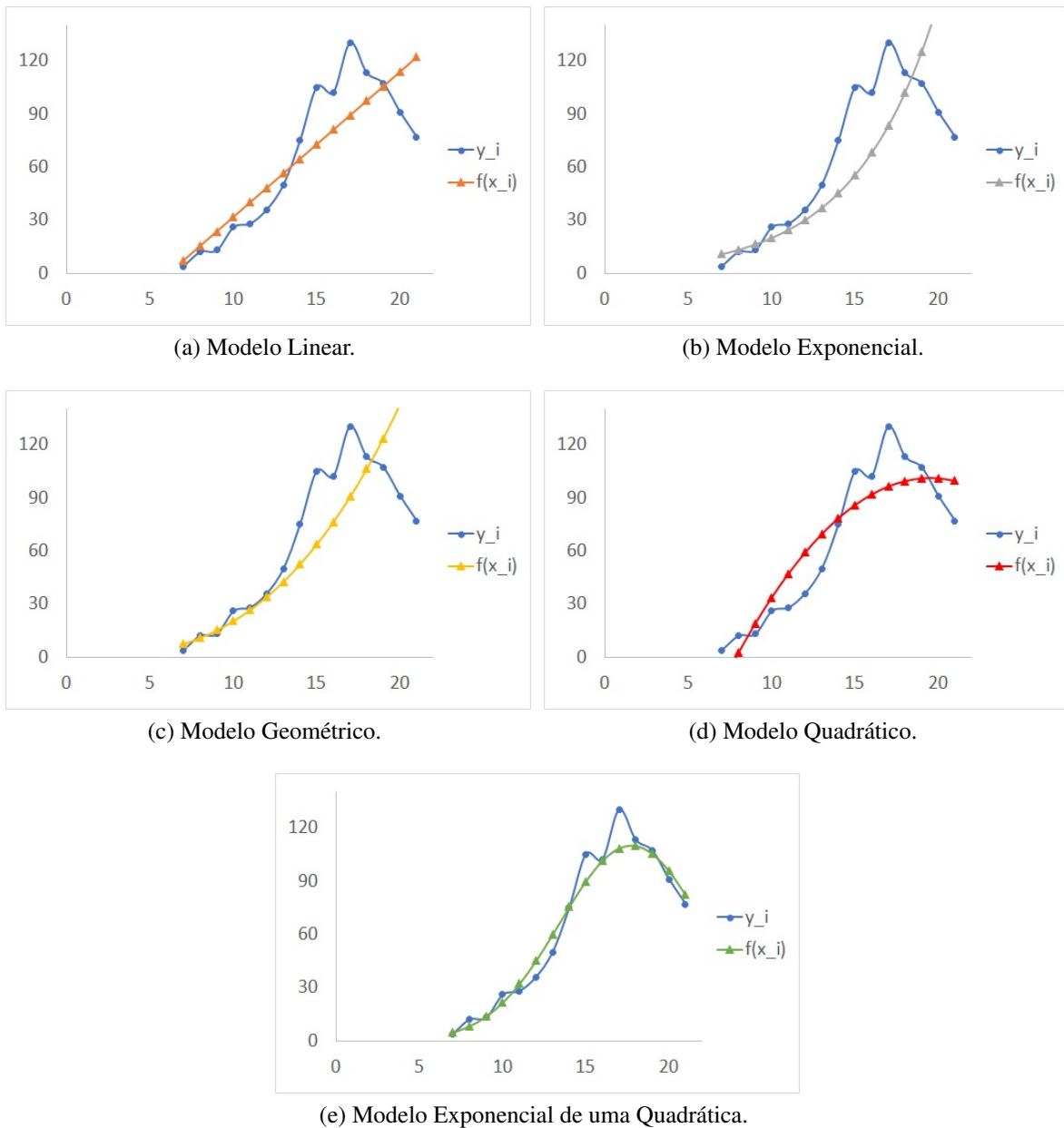


Figura 12 – Ajuste de curvas aos óbitos de COVID-19 das 15 últimas semanas em Curitiba.

Observamos, visualmente que o modelo que melhor se ajustou ao número de óbitos em Curitiba é o exponencial de uma quadrática, mostrado no item (e) da Figura 12.

3.1.3 Erros relativos

A Tabela 6 mostra o erro relativo de cada um dos modelos considerados tanto para o número de casos confirmados como para o número de óbitos por COVID-19 em Curitiba, para o período total considerado e para o período que abrange apenas as 15 últimas semanas. Em destaque estão os menores erros relativos.

Modelos	Período total considerado		15 últimas semanas	
	casos	óbitos	casos	óbitos
Linear	0,412926602	0,310562026	0,389541087	0,272714768
Exponencial	0,677809514	0,499587428	0,450409628	0,516300031
Geométrico	0,450665296	0,407732467	0,432636575	0,413170102
Quadrático	0,406347367	0,304284718	0,155887599	0,220464123
Exp. da Quadrática	0,352329054	0,403260833	0,087947213	0,106114569

Tabela 6 – Erro relativo para os modelos de número de casos confirmados e óbitos por COVID-19 em Curitiba.

Os menores erros ocorreram quando foram considerados os dados das últimas 15 semanas. Tanto para o número de casos confirmados como para o número de óbitos, o modelo que melhor se ajustou foi o da exponencial da quadrática. Este fato pode também ser inferido a partir das Figuras 8-12. A expressão do modelo que melhor se ajustou ao número de casos confirmados em Curitiba é:

$$f(x) = e^{-5,573909666+1,195269825x-0,030840812x^2}$$

em que x representa o número da semana, sendo que a semana 25 corresponde ao período de 16 a 22/08/2020. Por outro lado, a expressão do modelo que melhor se ajustou ao número de óbitos em Curitiba é:

$$f(x) = e^{-3,832128219+0,962451749x-0,027140408x^2}$$

em que x representa o número da semana, sendo que a semana 25 corresponde ao período de 23 a 29/08/2020.

3.2 Paraná

Nesta seção vamos aplicar os ajustes de curvas aos dados semanais de COVID-19 no Paraná. Na primeira subseção discutimos os modelos para o número de casos confirmados; na

segunda, para o número de óbitos e encerramos a seção com a análise dos erros relativos obtidos e as expressões dos modelos que melhor se ajustaram aos dados.

3.2.1 Número de casos confirmados

A Figura 13 exibe os números oficiais de casos semanais confirmados por COVID-19 no estado do Paraná de 08/03 a 31/08/2020. Nela as semanas estão numeradas de 11 a 36, essa numeração vem da contagem das semanas epidemiológicas disponíveis no Calendário Epidemiológico 2020 [3]. As semanas do gráfico que foram consideradas são de 11 a 35, pois a semana 36 está incompleta. Com base nesses dados, discutimos os modelos em duas situações. Na primeira situação, consideramos os dados de todas essas semanas, enquanto na segunda situação foram adotados os dados das últimas 15 semanas, denotadas no gráfico da Figura 13 de semana 21 a 35.

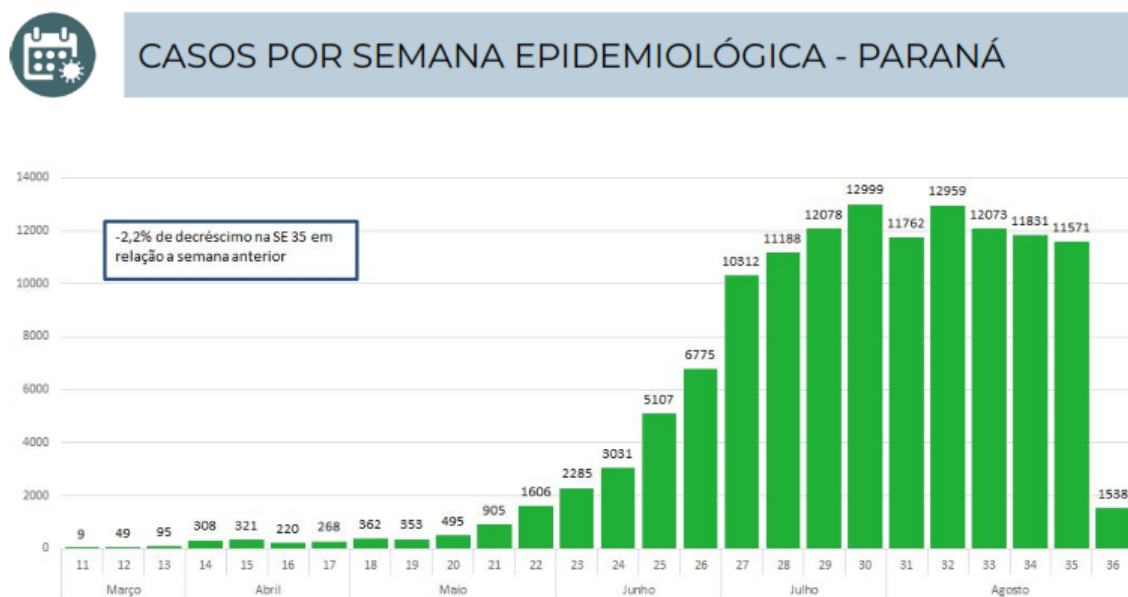


Figura 13 – Número de casos confirmados de COVID-19 no Paraná até agosto [10].

A Figura 14 mostra os modelos obtidos para o número de casos semanais confirmados por COVID-19 no estado do Paraná de 08/03 a 29/08/2020. Consideramos a semana 11 do gráfico da Figura 13 como a primeira. As abscissas representam as semanas e as ordenadas o número de casos semanais. Os pontos em azul indicam os dados fornecidos pela Figura 13. Em cada um dos gráficos está representado um dos modelos estudados que melhor se ajustou aos dados pelo método dos quadrados mínimos.

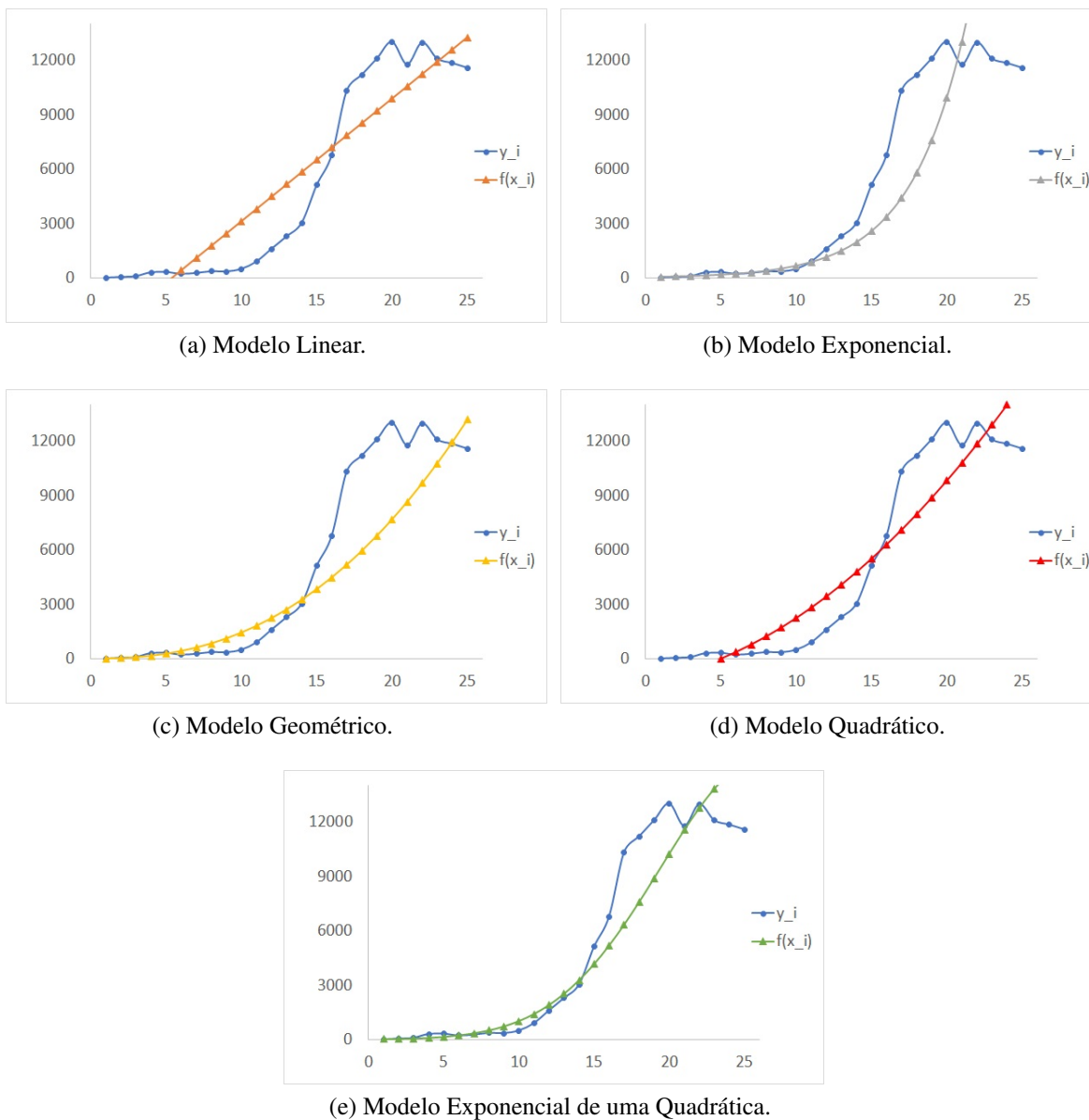


Figura 14 – Ajuste de curvas aos dados de casos semanais confirmados de COVID-19 no Paraná até agosto.

A Figura 15 mostra os modelos obtidos para o número de casos semanais confirmados por COVID-19 no estado do Paraná para as últimas 15 semanas consideradas, que consiste no período de 17/05 a 29/08/2020. As abscissas representam as semanas e as ordenadas, o número de casos semanais. Os pontos em azul indicam os dados fornecidos pela Figura 13. Em cada um dos gráficos está representado um dos modelos estudados que melhor se ajustou aos dados pelo método dos quadrados mínimos.

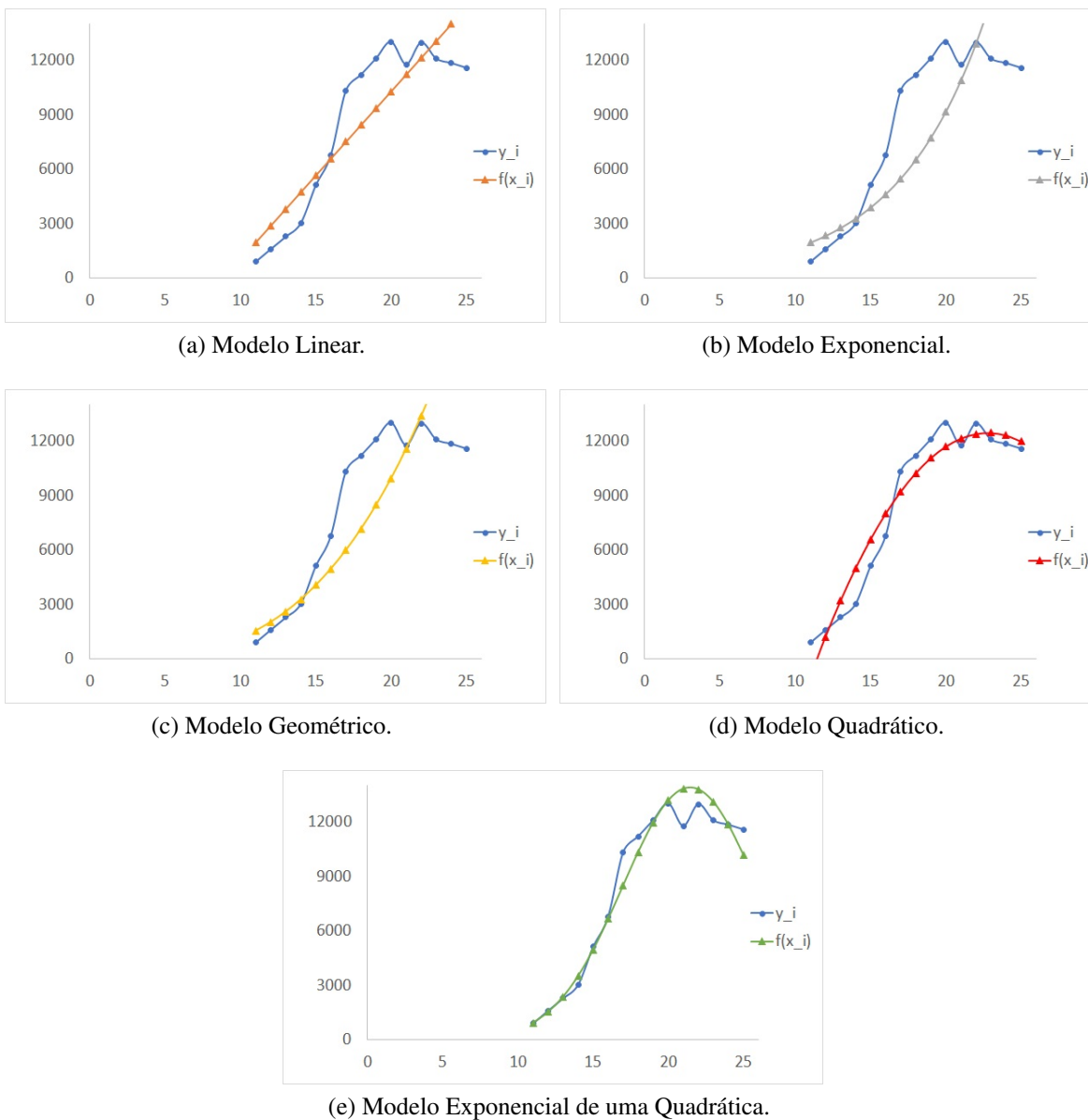


Figura 15 – Ajuste de curvas de casos de COVID-19 das 15 últimas semanas no Paraná.

Comparando os gráficos das Figuras 14 e 15, vemos que o modelo que melhor se adequou ao número de casos confirmados no Paraná foi a exponencial de uma quadrática, representado no item (e) da Figura 15.

3.2.2 Número de óbitos

A Figura 16 exibe os números oficiais de óbitos semanais por COVID-19 no estado do Paraná de 08/03 a 31/08/2020. Como as informações da última semana parecem incompletas e as duas primeiras semanas não possuem óbitos, foram considerados os dados das 23 semanas numeradas no gráfico de 13 a 35. Como na subseção anterior, discutimos os modelos em duas

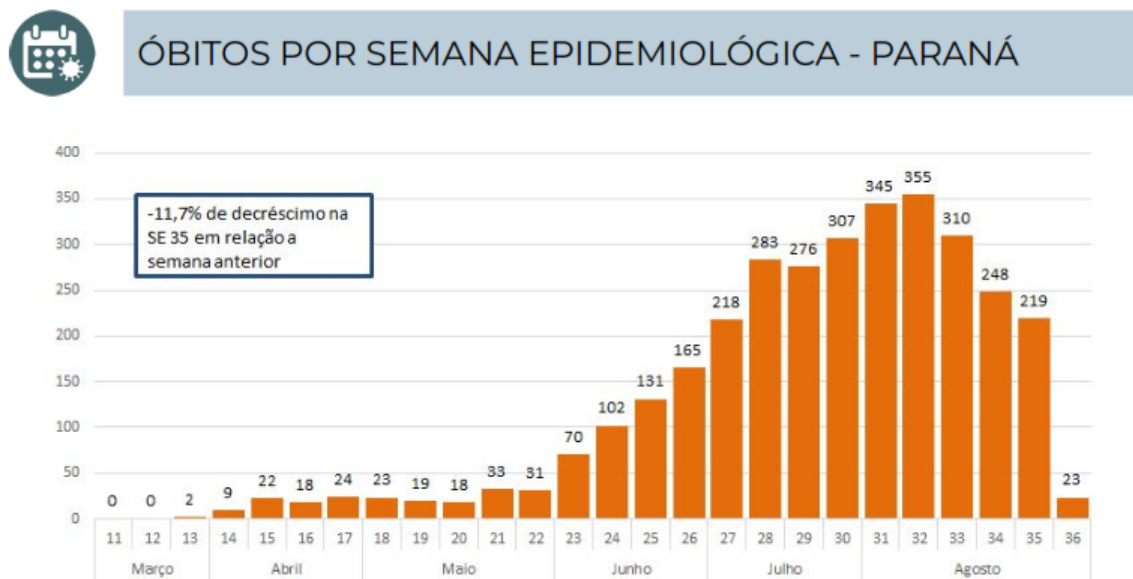


Figura 16 – Óbitos semanais de COVID-19 no Paraná até Agosto. [10]

situações. Na primeira situação, consideramos os dados de todas essas semanas, enquanto na segunda situação foram adotados os dados das últimas 15 semanas, denotadas no gráfico da Figura 16 de semana 21 a 35.

A Figura 17 mostra os modelos para o número de óbitos no Paraná para as 23 semanas consideradas. Esse período vai de 22/03 a 29/08/2020. Os modelos se ajustam mais adequadamente aos dados das primeiras semanas, no entanto, nas últimas eles apresentam uma previsão muito além da realidade.

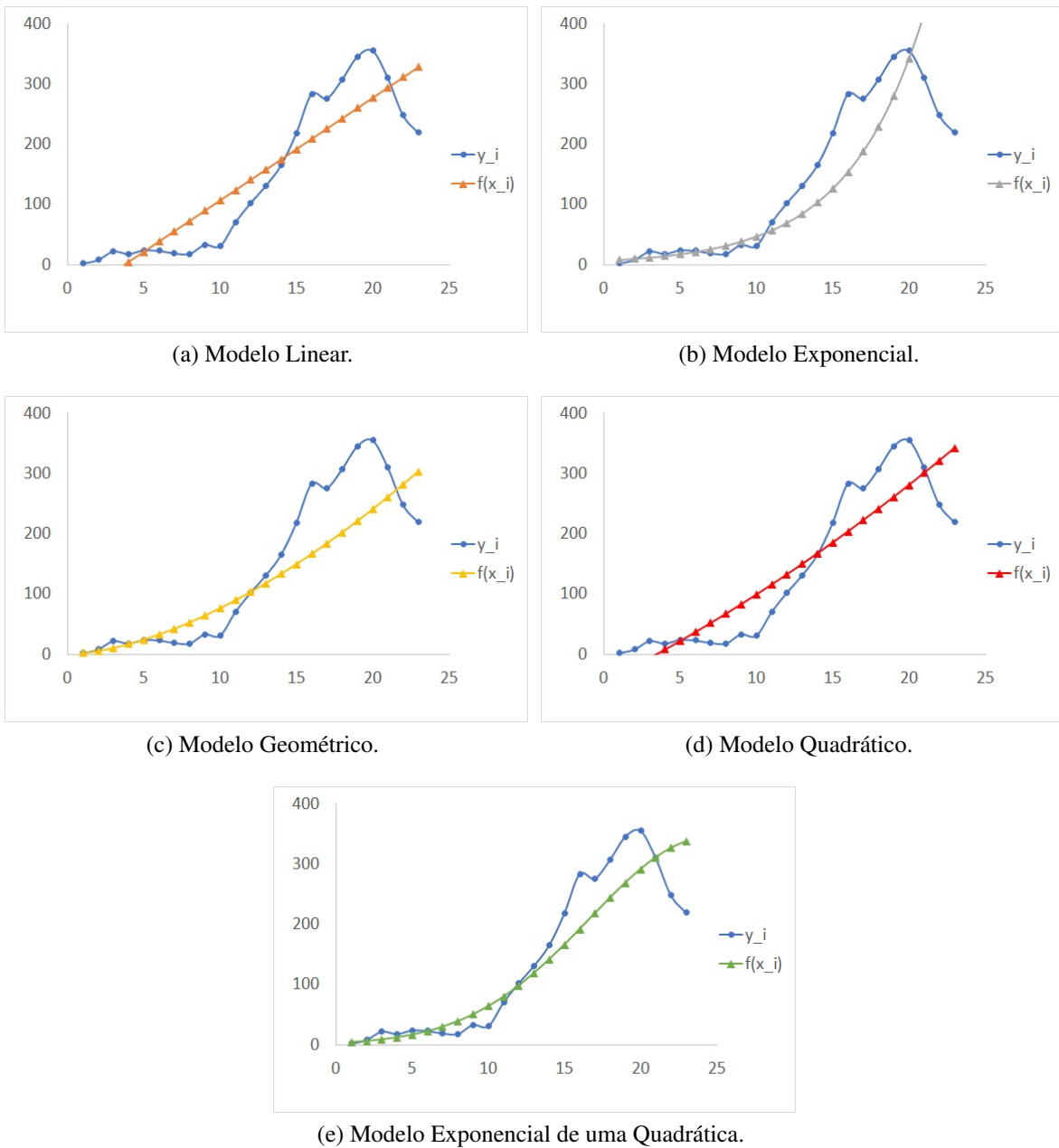


Figura 17 – Ajuste de curvas aos dados semanais de óbitos por COVID-19 no Paraná.

A Figura 18 mostra os modelos para o número de óbitos no Paraná para as últimas 15 semanas, denotadas no gráfico da Figura 16 por semana 21 a 35. Esse período vai de 17/05 a 23/08/2020. Visualmente, o modelo que melhor se ajustou ao número de óbitos no Paraná é o exponencial de uma quadrática, mostrado no item (e) da Figura 18.

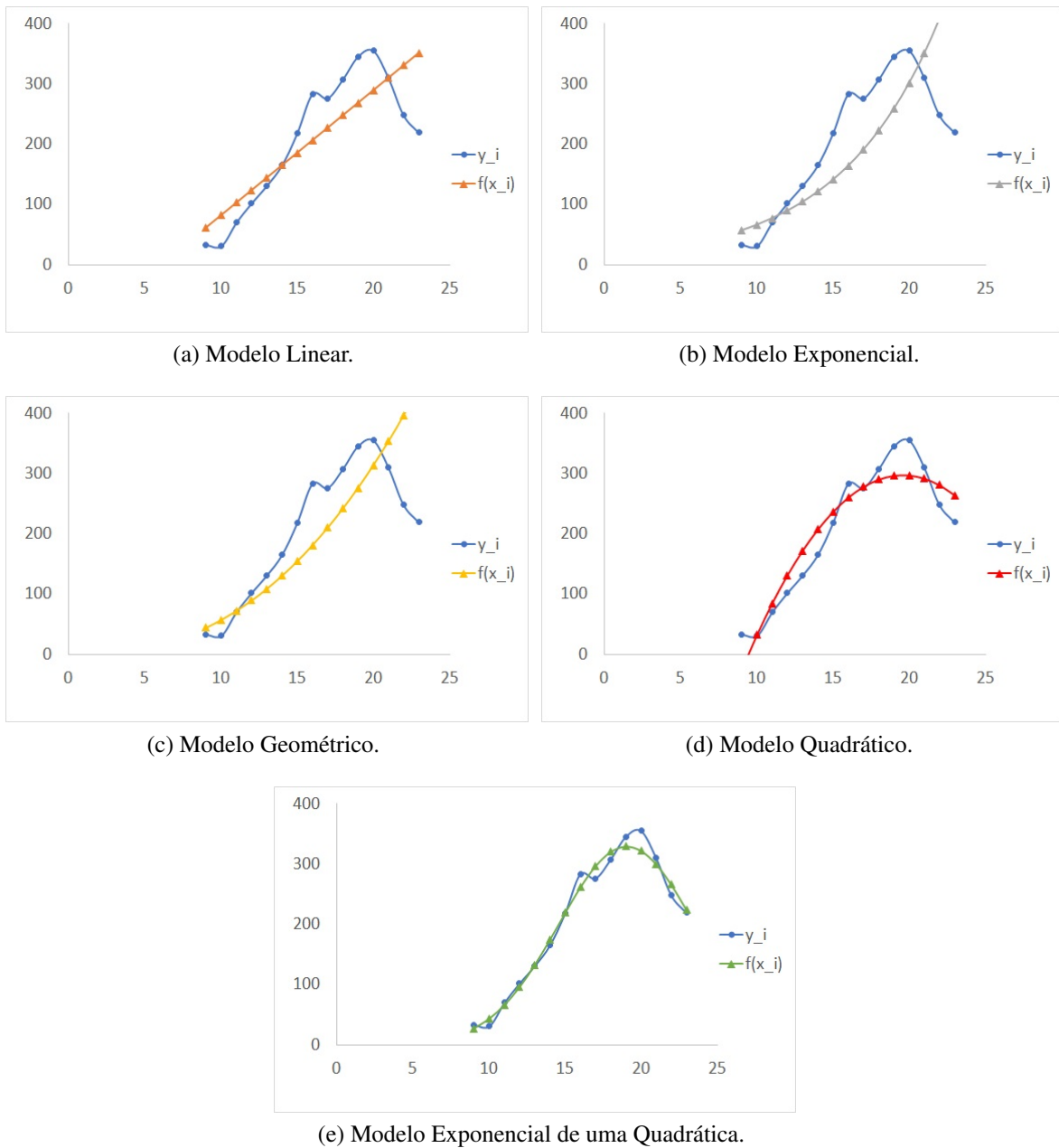


Figura 18 – Ajuste de curvas aos óbitos por COVID-19 das 15 últimas semanas no Paraná.

3.2.3 Erros relativos

A Tabela 7 mostra o erro relativo de cada um dos modelos considerados tanto para o número de casos confirmados como para o número de óbitos por COVID-19 no Paraná, para o período total considerado e para o período que abrange apenas as 15 últimas semanas. Em destaque estão os menores erros relativos.

Modelos	Período total considerado		15 últimas semanas	
	casos	óbitos	casos	óbitos
Linear	0,279860696	0,28511128	0,203238535	0,254461658
Exponencial	0,959863571	0,598992161	0,418382518	0,419408083
Geométrico	0,326550393	0,321203296	0,359222196	0,360958634
Quadrático	0,248064702	0,282516336	0,114571303	0,150446242
Exp. da Quadrática	0,236125467	0,251723827	0,094850993	0,062573571

Tabela 7 – Erro relativo para os modelos de número de casos confirmados e óbitos por COVID-19 no Paraná.

Os menores erros ocorreram quando foram considerados os dados das últimas 15 semanas. Tanto para o número de casos confirmados como para o número de óbitos, o modelo que melhor se ajustou foi o da exponencial da quadrática. Este fato pode também ser inferido a partir das Figuras 14-18. A expressão do modelo que melhor se ajustou ao número de casos confirmados no Paraná é:

$$f(x) = e^{-1,859476774+1,062376078x-0,024755107x^2}$$

em que x representa o número da semana, sendo que a semana 25 representa a semana de 23/08 a 29/08/202. Por outro lado, a expressão do modelo que melhor se ajustou ao número de óbitos no Paraná é:

$$f(x) = e^{-3,222535031+0,946480426x-0,024830157x^2}$$

em que x representa o número da semana, sendo que a semana 25 representa a semana de 23/08 a 29/08/2020.

3.3 Brasil

Nesta seção vamos aplicar os ajustes de curvas aos dados semanais de COVID-19 no Brasil. Na primeira subseção discutimos os modelos para o número de casos confirmados; na segunda, para o número de óbitos e encerramos a seção com a análise dos erros relativos obtidos e as expressões dos modelos que melhor se ajustaram aos dados.

3.3.1 Número de casos confirmados

A Figura 19 exibe os números oficiais de casos semanais confirmados por COVID-19 no Brasil de 23/02 a 31/08/2020. Nela as semanas estão numeradas de 9 a 36, essa numeração vem da contagem das semanas epidemiológicas disponíveis no Calendário Epidemiológico 2020 [3]. As semanas do gráfico que foram consideradas são de 10 a 35, pois a semana 36 está incompleta e a semana 9 não possui casos. Com base nesses dados, discutimos os modelos em duas situações. Na primeira situação, consideramos os dados de todas essas semanas, enquanto na segunda situação foram adotados os dados das últimas 15 semanas, denotadas no gráfico da Figura 13 de semana 21 a 35.

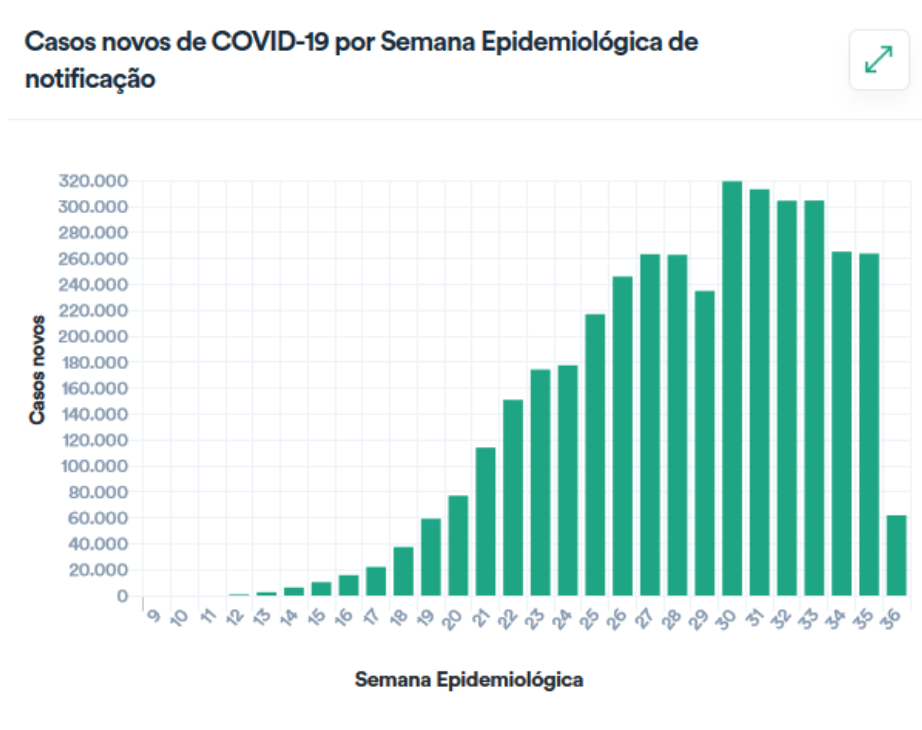


Figura 19 – Número de casos confirmados de COVID-19 no Brasil de 23/02 a 31/08/2020 [7].

A Figura 20 mostra os modelos obtidos para o número de casos semanais confirmados por COVID-19 no Brasil de 23/02 a 31/08/2020. Consideramos a semana 10 do gráfico da Figura 19 como a primeira. As abscissas representam as semanas e as ordenadas o número de casos semanais. Os pontos em azul indicam os dados fornecidos pela Figura 19. Em cada um dos gráficos está representado um dos modelos estudados que melhor se ajustou aos dados pelo método dos quadrados mínimos.

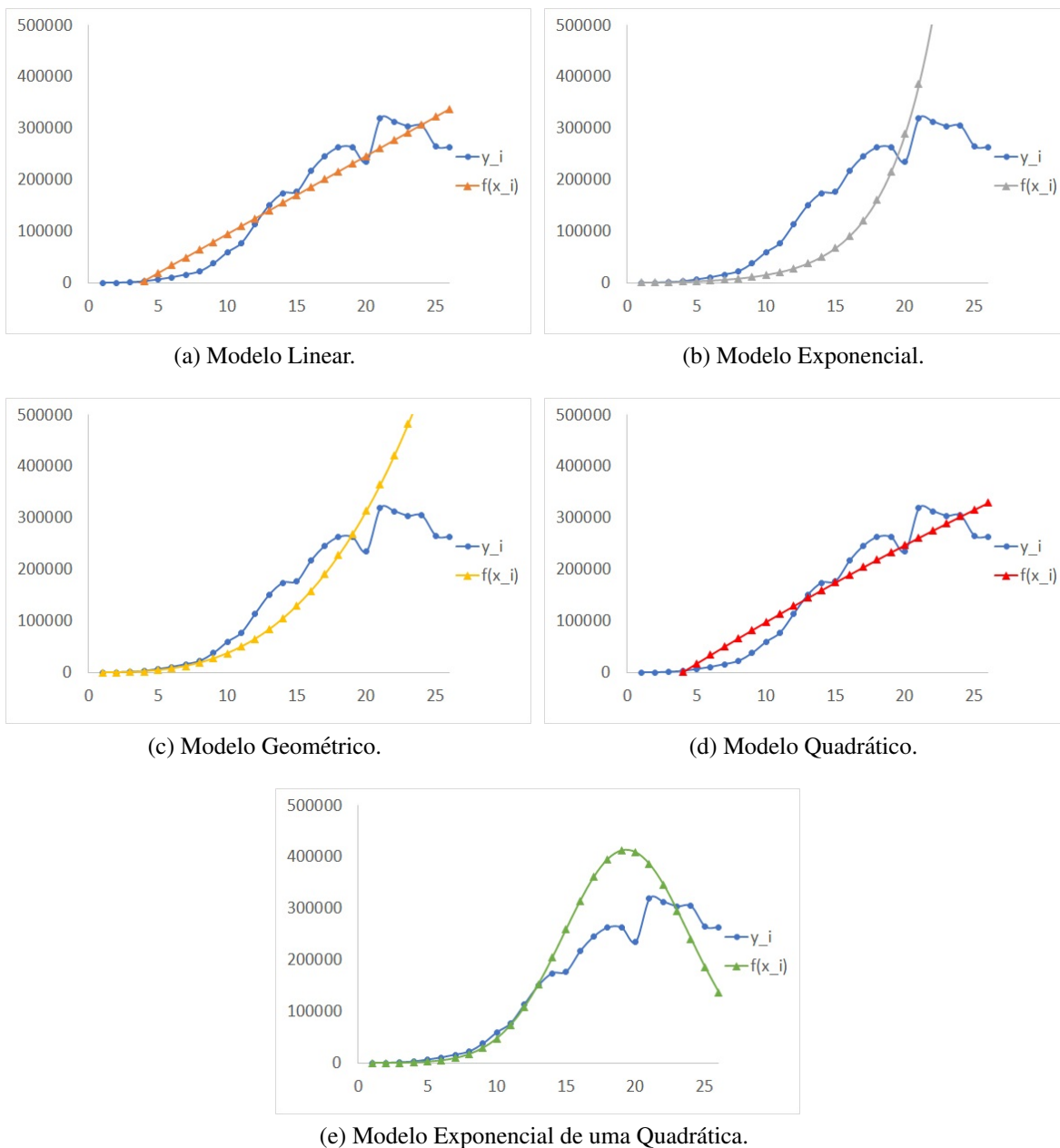


Figura 20 – Ajuste de curvas aos dados semanais de casos confirmados de COVID-19 no Brasil.

A Figura 21 mostra os modelos obtidos para o número de casos semanais confirmados por COVID-19 no Brasil para as últimas 15 semanas consideradas. Esse período vai de 17/05 a 31/08/2020. As abscissas representam as semanas e as ordenadas, o número de casos semanais. Os pontos em azul indicam os dados fornecidos pela Figura 19. Em cada um dos gráficos está representado um dos modelos estudados que melhor se ajustou aos dados pelo método dos quadrados mínimos.

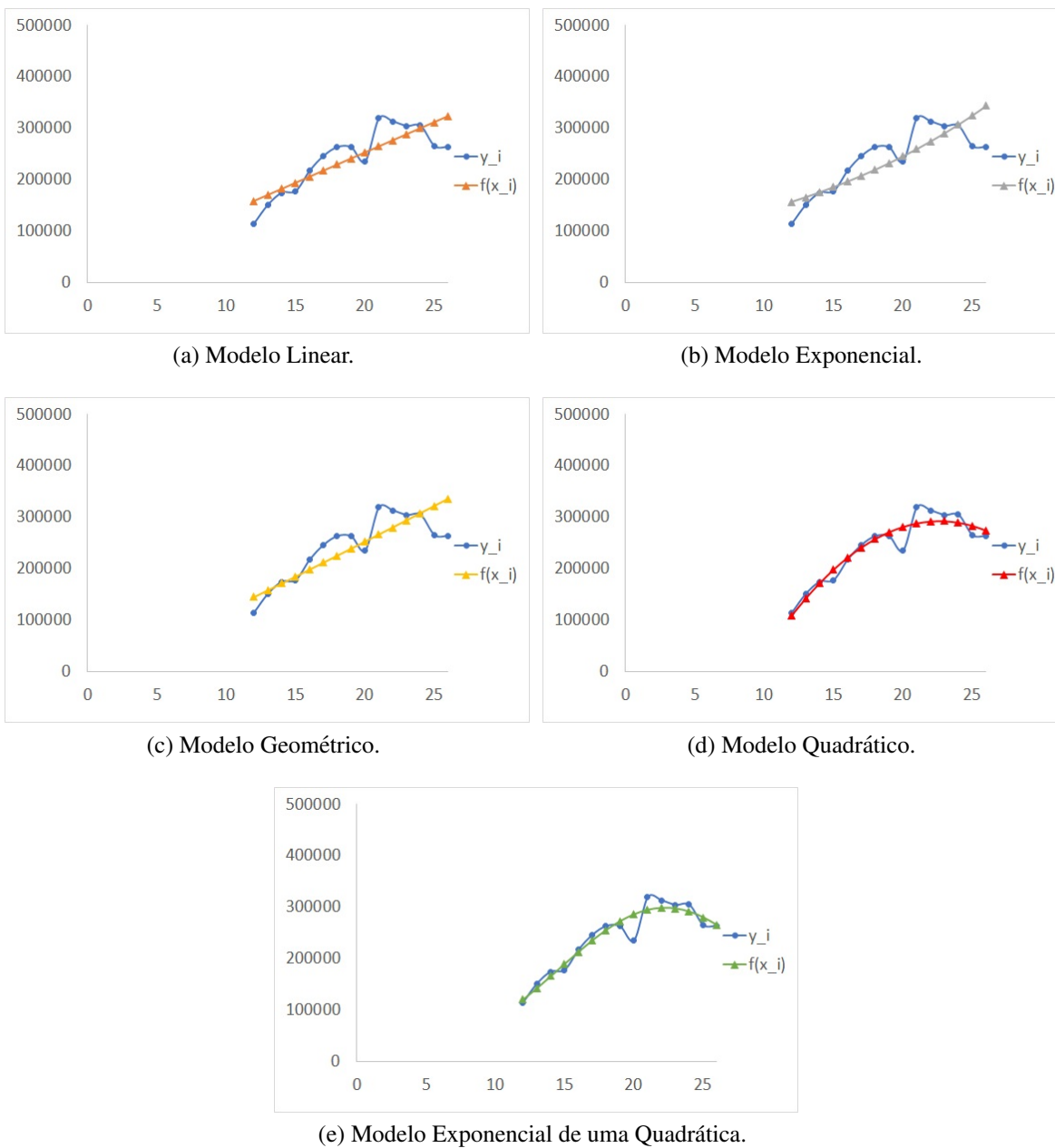


Figura 21 – Ajuste de curvas de casos de COVID-19 das 15 últimas semanas no Brasil.

Comparando os gráficos das Figuras 20 e 21, vemos que o modelo que melhor se adequou ao número de casos confirmados no Brasil foi a exponencial de uma quadrática, representado no item (e) da Figura 21.

3.3.2 Número de óbitos

A Figura 22 exibe os números oficiais de óbitos semanais por COVID-19 no Brasil de 23/02 a 31/08/2020. Como as informações da última semana parecem incompletas e as três primeiras semanas não possuem óbitos, foram considerados os dados das 24 semanas numeradas no gráfico de 12 a 35. Com base nesses dados, discutimos os modelos em duas situações.

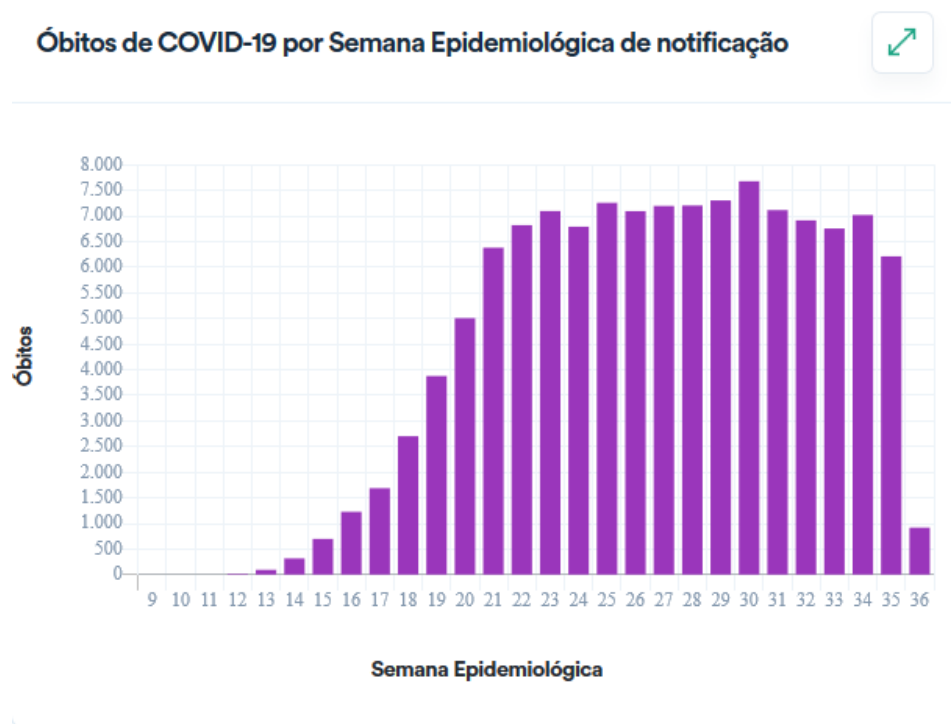


Figura 22 – Óbitos semanais de COVID-19 no Brasil 23/02 a 31/08/2020 [7].

Na primeira situação, consideramos os dados de todas essas semanas, enquanto na segunda situação foram adotados os dados das últimas 15 semanas, denotadas no gráfico da Figura 22 de semana 21 a 35.

A Figura 23 mostra os modelos para o número de óbitos no Brasil para as 24 semanas consideradas. Esse período vai de de 15/03 a 31/08/2020. A maioria dos modelos se ajustam mais adequadamente aos dados das primeiras semanas, no entanto, nas últimas eles apresentam uma previsão muito além da realidade. O único modelo que se ajusta quase por completo aos dados é o quadrático representado no item (d) da Figura 23.

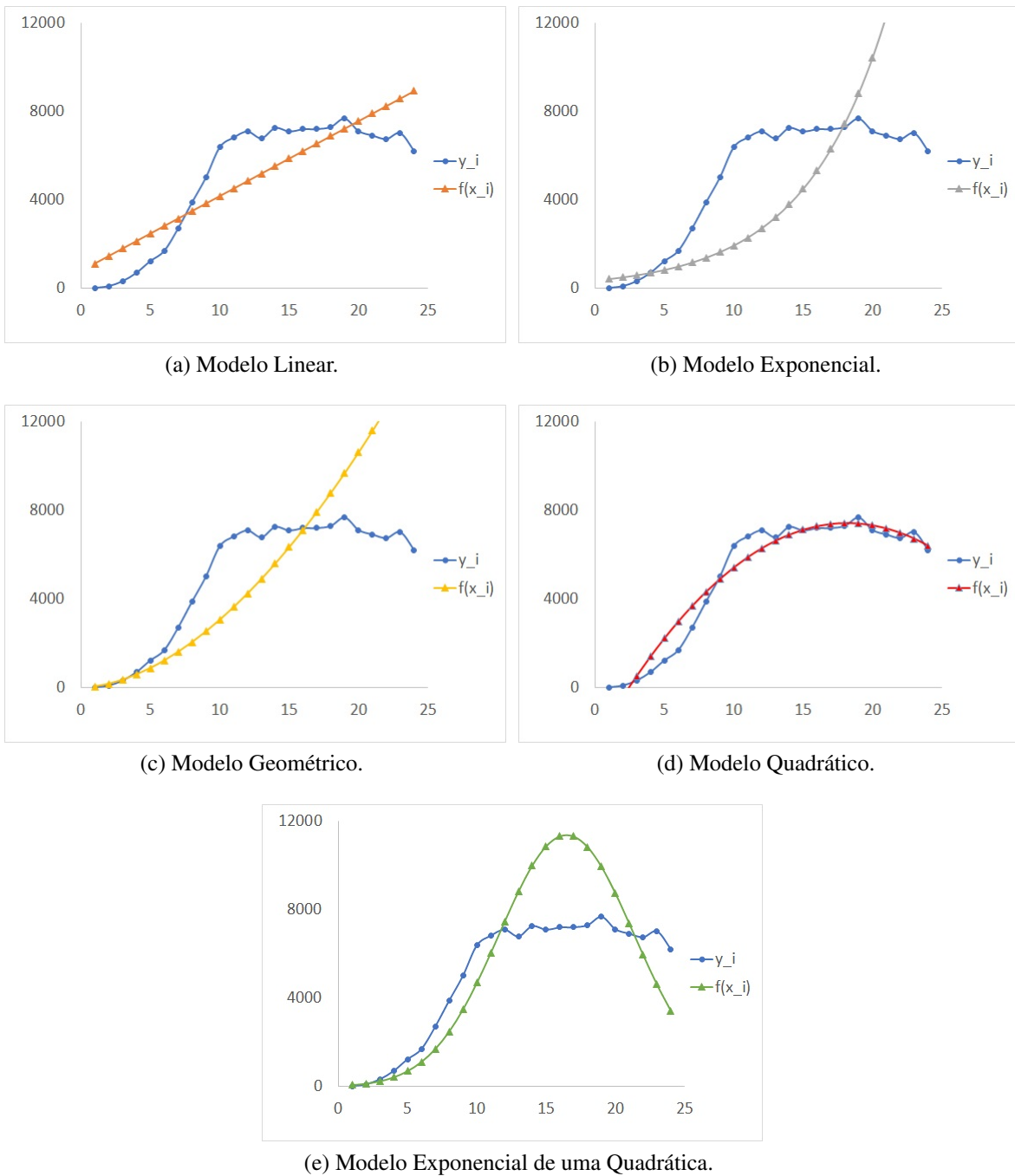


Figura 23 – Ajuste de curvas aos dados semanais de óbitos por COVID-19 no Brasil.

A Figura 24 mostra os modelos para o número de óbitos no Brasil para as últimas 15 semanas, denotadas no gráfico da Figura 22 por semana 21 a 35. Esse período vai de de 17/05 a 31/08/2020.

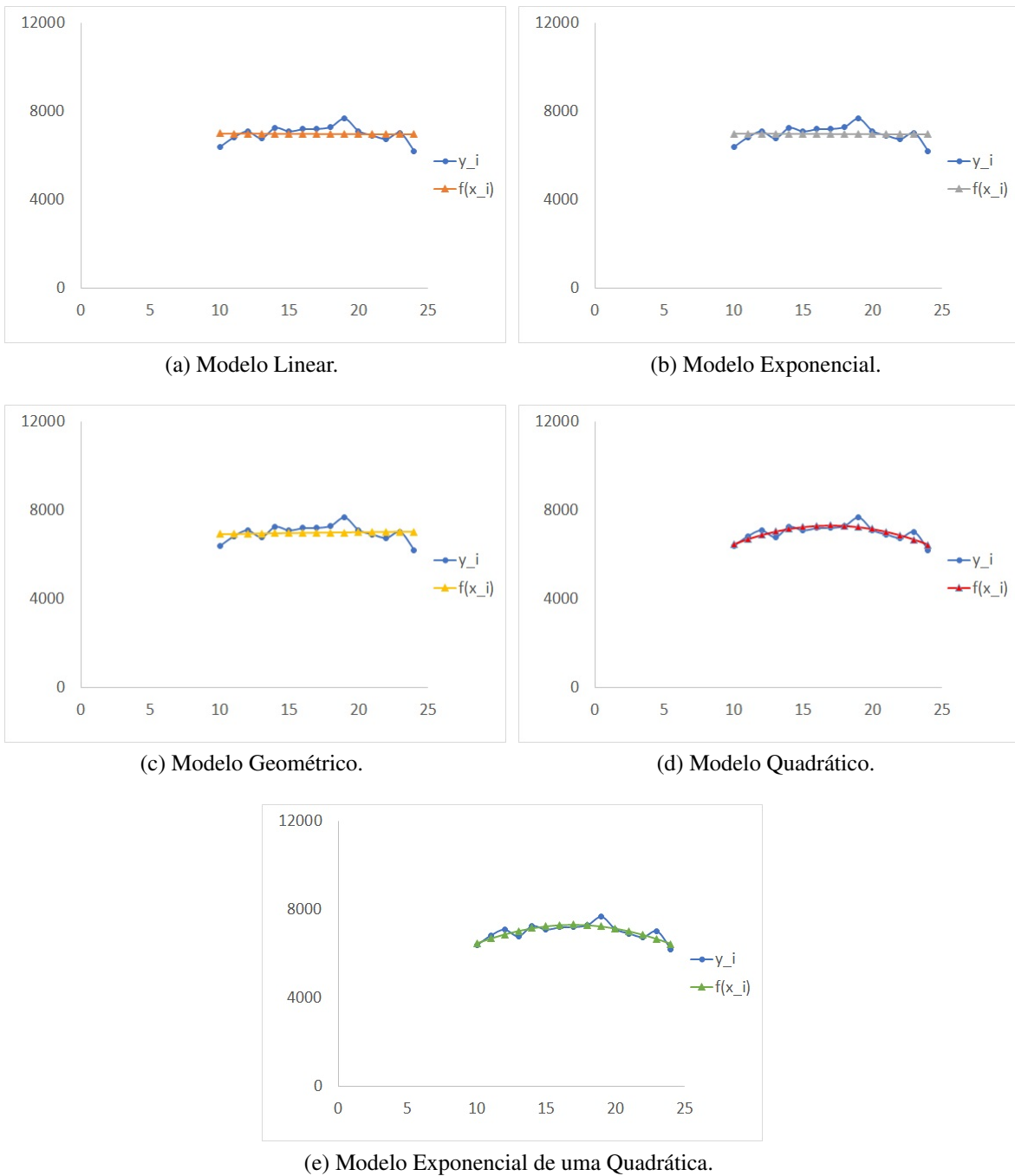


Figura 24 – Ajuste de curvas aos óbitos por COVID-19 das 15 últimas semanas no Brasil.

Visualmente, os modelos que apresentam a maior aproximação são o quadrático e o exponencial de uma quadrática, mostrado no item (d) e (e) da Figura 24. Veremos, pelo erro relativo, qual foi o modelo que melhor se ajusta.

3.3.3 Erros relativos

A Tabela 8 mostra o erro relativo de cada um dos modelos considerados tanto para o número de casos confirmados como para o número de óbitos por COVID-19 no Paraná, para o período total considerado e para o período que abrange apenas as 15 últimas semanas. Em destaque estão os menores erros relativos.

Modelos	Período total considerado		15 últimas semanas	
	casos	óbitos	casos	óbitos
Linear	0,181522424	0,249912686	0,131410845	0,050501103
Exponencial	1,923998618	0,819622838	0,15500427	0,0505199
Geométrico	0,696283429	0,551068588	0,136850097	0,050364971
Quadrático	0,180266946	0,111187726	0,074113243	0,028010908
Exp. da Quadrática	0,377030624	0,363812576	0,069766872	0,028200833

Tabela 8 – Erro relativo para os modelos de casos confirmados e mortes por COVID-19 no Brasil.

Os menores erros ocorreram quando foram considerados os dados das últimas 15 semanas. Para o número de casos confirmados o modelo que melhor se ajustou foi o exponencial de uma quadrática e para o número de óbitos foi o quadrático. Este fato pode também ser inferido a partir das Figuras 20-24. A expressão do modelo que melhor se ajustou ao número de casos confirmados no Brasil é:

$$f(x) = e^{8,349656776+0,382360029x-0,008587315x^2}$$

em que x representa o número da semana, sendo que a semana 25 representa a semana de 23/08 a 29/08/2020. Por outro lado, a expressão do modelo que melhor se ajustou ao número de óbitos no Brasil é:

$$f(x) = 2216,957466 + 601,5755495x - 17,73068843x^2$$

em que x representa o número da semana, sendo que a semana 25 representa a semana de 23/08 a 29/08/2020.

4 Experiência Didática em Tempos de Quarentena

O calendário acadêmico da UFPR - Universidade Federal do Paraná foi suspenso a partir do dia 15 de março de 2020, por conta da pandemia causada pelo COVID-19. Com o objetivo de propor horas formativas aos alunos do curso de Matemática Industrial, a Coordenação do curso aprovou a oferta de uma atividade à distância com cinco módulos com o tema “Modelagem Matemática em Tempos de Quarentena”. A atividade foi ofertada para todos os alunos do curso, sem exigência de pré-requisitos. Tendo em vista que o tema deste TCC - Trabalho de Conclusão de Curso se encaixava na proposta da atividade, propusemos ofertar o primeiro módulo intitulado “Método dos Quadrados Mínimos”, sob responsabilidade da minha orientadora, Professora Elizabeth Wegner Karas, e eu.

Os responsáveis por cada um dos módulos tiveram autonomia para decidir sobre a metodologia adotada. A coordenação da atividade criou uma sala no “Google Classroom” para disponibilizar os materiais dos cinco módulos. Houve no total 47 inscritos entre alunos do curso e egressos.

Para o módulo de “Método dos Quadrados Mínimos”, sob nossa responsabilidade, optamos em preparar slides tendo como base o material apresentado no Capítulo 2 deste TCC. Essa preparação dos slides foi fundamental para uma releitura e reescrita do próprio capítulo. Para disponibilizar aos alunos optamos pela filmagem dos slides com gravação de explicações em áudio. Esta gravação foi feita pela professora com duração de 49 minutos. O material foi disponibilizado no dia 14/05/2020 na “Google Classroom” e no “YouTube” através do link <https://www.youtube.com/watch?v=1VArqkJEMN8>.

Como meio de avaliação foi solicitado um relatório de aplicação dos modelos estudados numa base de dados escolhida pelo próprio aluno. A base de dados não precisava ter relação com a pandemia. Com o objetivo de discutir e esclarecer eventuais dúvidas os participantes foram convidados para um momento síncrono pelo “Discord”, cinco dias após a publicação do material na “Classroom”. Essa discussão ocorreu no dia 19/05/2020, no horário marcado, com a participação de 18 alunos e da Professora Ailin Ruiz Zarate Fabregas, responsável pelo segundo módulo. Os participantes eram de diferentes períodos desde o primeiro semestre ao último ano. A conversa no “Discord” foi coordenada pela professora Elizabeth. Os alunos comentaram que gostaram da apresentação dos slides e demonstraram interesse na discussão. Questionaram a obtenção das fórmulas (2.39) dos coeficientes do modelo quadrático. Explicamos que a obtenção dessas fórmulas segue da resolução de um

sistema linear com 3 variáveis. No entanto exige vários passos de cálculos e que os alunos tinham toda razão no questionamento. Em seguida foi compartilhada a tela com os trabalhos já postados até aquele momento. Em um deles havia a apresentação do erro relativo (2.3) dos modelos. A importância deste erro para a análise dos modelos foi evidenciada na discussão. Dada sua relevância decidimos incluir esta discussão no trabalho e nos slides. Os alunos também foram orientados a discutir os erros relativos nos relatórios. Finalizamos a conversa no “Discord” após duas horas com a solicitação de “feedbacks”, os quais foram muito positivos e incentivadores como mostra a Fig. 25.

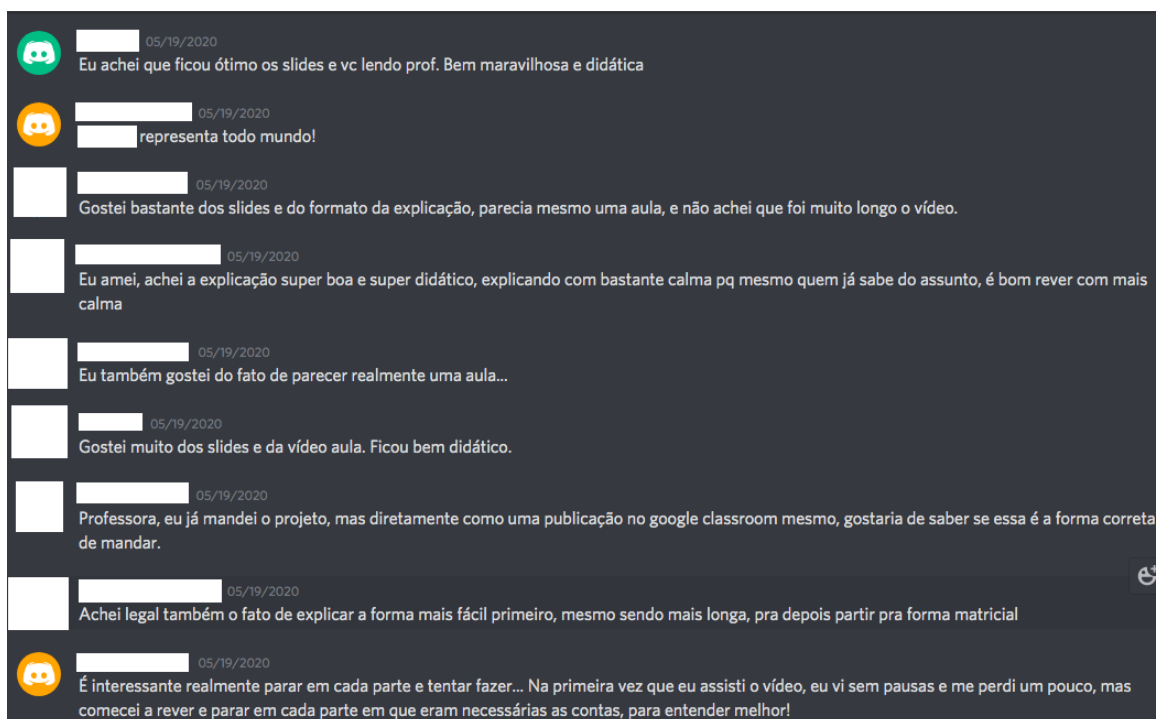


Figura 25 – Tela com “feedbacks” recebidos no “Discord”.

Recebemos através da “Classroom” o relatório de 13 participantes. Alguns alunos usaram mais de um banco de dados e os temas abordados foram variados, como mostra a Tabela 9 que apresenta uma síntese dos trabalhos entregues. A primeira coluna (GRR) refere-se ao ano de ingresso do aluno no curso. A segunda (Número) representa o tema escolhido para a base de dados, onde (A) representa “por ano”, (T) “por trimestre”, (M) “por mês”, (S) “por semana”, (D) “por dia” e (H) “por hora”. A terceira coluna (Prog.) indica o software ou a linguagem de programação utilizada pelo aluno. E finalmente a quarta (f) coluna indica o modelo que melhor se ajustou aos dados do problema. O modelo está indicado pela letra (L) no caso linear (2.4), (Q) quadrático (2.27), (E) exponencial (2.15), (EQ) exponencial de uma quadrática (2.43) e (O) indica outros.

GRR	Número	Prog.	f
2016	(A) aplicativos na B3 de 2002 a 2020.	Excel	EQ
2016	(S) casos de Dengue no Paraná de 2015 á 2016	Julia	O
2018	(D) casos de COVID na França, de 18/03 a 18/05/2020.	Octave	Q
2018	(A) população brasileira em milhões, de 1872 a 2010.	Octave	Q
2018	(D) casos de COVID nos EUA de 04/03 a 03/04/2020.	Oct/Exc	Q
2018	(A) feminicídios no Brasil, de 2015 a 2019.	Excel	Q
2018	(M) 0 a 100, de popularidade do veganismo e do vegetarianismo, de 03/2004 a 12/2019.	Julia	Q Q
2018	(T) porcentagem de desemprego no Brasil, de 2017 á 2020	Excel	EQ
2019	(A) mortes por desnutrição em Curitiba 1979 a 2017.	Excel	Q
2019	(D) casos de Dengue no Paraná 08/2019 a 05/2020.	Excel	EQ
2019	(D) mortes por COVID na Itália em abril de 2020.	Octave	EQ
2020	(A) população Brasileira de 1500 a 2020 (H) bactérias em uma amostra.	Python	E EQ
2020	(A) mortes no período neonatal no Haiti, de 1960 a 2018.	Excel	Q

Tabela 9 – Síntese dos Projetos.

Como mostra a Tabela 9, dos 13 participantes que entregaram o relatório, 2 ingressaram no curso em 2016, 6 em 2018, 3 em 2019 e 2 em 2020. Cabe ressaltar que um dos alunos que ingressou em 2020 já é formado em Estatística. Dos 12 trabalhos, 7 utilizaram o “Excel” para realizar os cálculos e fazer os gráficos, 2 programaram em “Julia”, 4 em “Octave” e 1 em “Python”. Alguns alunos realizaram os cálculos tanto da foram estudada nas Seções 2.1 e 2.4.1 como na forma matricial vista na Seção 2.4.2. Um dos alunos também inseriu em seu trabalho mais dois modelos de aproximação a exponencial de seno e cosseno, que para o caso dele foram os com menor erro. Inferimos dos relatórios que os participantes conseguiram aplicar a metodologia estudada a diferentes bancos de dados. Não houveram erros recorrentes. Tivemos apenas três trabalhos com algum equívoco no cálculo de erros que não condiziam com os gráficos apresentados; na exponencial da quadrática, usando logaritmo ao invés da própria exponencial ou na plotagem dos gráficos. Esses equívocos foram solucionados após nosso “feedback” individual ao aluno.

Como licencianda a experiência de prática à distância contribui de maneira significativa para minha graduação, pois possibilita uma forma de ensino que complementa a que vivenciamos em nossa formação. Além disso muitos dos temas abordados são de grande relevância para nossa sociedade e abrem espaço para discussões de extrema importância para o cenário atual. Em um tempo de crise onde todos tivemos que nos reinventar, acredito que este módulo foi de grande aprendizagem, não apenas para os alunos mas também para mim como futura professora.

5 Conclusão

O surto da COVID-19 aqui no Brasil iniciou em meados de março e gerou uma série de dilemas em nossa sociedade. Dentro deste cenário é que foi construído este trabalho.

O objetivo deste trabalho foi de estudar o método de quadrados mínimos e aplicá-lo no ajuste de dados de casos confirmados e óbitos pela doença em Curitiba, Paraná e Brasil.

Iniciamos o Capítulo 2 partindo do modelo mais simples de compreensão, o linear, e generalizando na forma matricial se baseando nos livros [2, 8]. Com a necessidade de conteúdos de Álgebra Linear e de Cálculo para a compreensão dos conceitos envolvendo solução de sistemas, propriedades das matrizes, posto, imagem, derivadas parciais e projeção de vetores, outras referências, como [1, 4, 9, 12], complementaram o estudo. Ao aplicá-los neste trabalho tive a oportunidade de rever tais conteúdos com mais atenção e cuidado, além de vislumbrar aplicações a conceitos que podem passar despercebidos durante a graduação, contribuindo assim com um dos objetivos de um TCC.

Outro fator que contribuiu para a finalização do Capítulo 2 foi a criação e aplicação do módulo “Método dos Quadrados Mínimos” descrito no Capítulo 4. Com a participação dos alunos e a discussão sobre o erro relativo (2.3), decidimos incorporá-lo a este trabalho, dada a sua facilidade na comparação dos erros. Essa prática de aula à distância agregou à minha graduação como licencianda, visto que o momento propiciou uma maneira de dar aula além da convencional.

No Capítulo 3 aplicamos os modelos estudados pelo método de quadrados mínimos ao número de casos e óbitos por COVID-19. Em um primeiro momento trabalhamos com o número de casos e mortes diários. Durante a Semana da Matemática, proporcionada pelo PET Matemática da UFPR, assisti a palestra “Desafios da Modelagem Matemática aplicada no COVID-19” conduzida pelos professores Roberto Ribeiro e Elias Gudiño da UFPR e Nara Bobko da UTFPR. Nesta palestra foi comentado que os dados diários são afetados pela falta de registro durante os finais de semana, provocando um aumento irreal no número de casos e mortes no início da semana. Esse fato gera uma irregularidade nos dados. Isso foi crucial na nossa decisão em trabalhar com os dados semanais.

Dos modelos aplicados, o que se destacou foi o exponencial de uma quadrática, por apresentar o menor erro na maioria dos casos. Esse fato condiz com o crescimento exponencial da doença. Tais modelos podem ser considerados como modelos de projeção [5] já que para previsões a curto prazo o erro é pequeno. Mas como foi abordado, o foco deste trabalho não foi fazer previsões mas apresentar o Método dos Quadrados Mínimos e os conteúdos envolvidos que são estudados durante a graduação.

Referências

- [1] W. E. Boyce e R. C. DiPrima. Equações Diferenciais Elementares e Problemas de Valores de contorno LTC, 2006.
- [2] R.L. Burden e J.D. Faires. Análisis numérico. Thomson Learning, Sétima edição, 2002.
- [3] Calendário Epidemiológico 2020. <http://portalsinan.saude.gov.br/calendario-epidemiologico-2020>. Acessado em 01/08/2020.
- [4] A. Howard e C. Rorres. Álgebra Linear com Aplicações. Bookman, Oitava edição, 2001.
- [5] I. Holmdahl e C. Buckee. Wrong but Useful — What Covid-19 Epidemiologic Models Can and Cannot Tell Us The New England Journal of Medicine, Julho 2020.
- [6] W. O. Kermack e A. G. McKendrick. A Contribution to the Mathematical Theory of Epidemics, Proceedings of the Royal Society, Vol. 115A, p.700–721, 1927.
- [7] Ministério da Saúde. <https://coronavirus.saude.gov.br/>. Acessado em 01/09/2020.
- [8] M.A.G. Ruggiero and V.L.R. Lopes. Cálculo Numérico: Aspectos Teóricos e Computacionais. Pearson, Segunda edição, 2000.
- [9] R. J. Santos. Um Curso de Geometria e Álgebra Linear. Imprensa Universitária da UFMG, Julho 2010.
- [10] Secretaria da Saúde do Estado do Paraná. <http://www.saude.pr.gov.br/Pagina/Coronavirus-COVID-19>. Acessado em 01/08/2020.
- [11] Secretaria Municipal da Saúde de Curitiba. <https://coronavirus.curitiba.pr.gov.br/>. Acessado em 01/09/2020.
- [12] S. Treil. Linear Algebra Done Wrong. 2017. <https://sites.google.com/a/brown.edu/sergei-treil-homepage/linear-algebra-done-wrong?authuser=0>. Acessado em 20/08/2020.