

Modelagem e Predição das quantidades de casos confirmados da COVID-19 no estado de São Paulo, Brasil

Erlandson Ferreira Saraiva^a e Carlos Alberto de Bragança Pereira^b

^aInstituto de Matemática, Universidade Federal de Mato Grosso do Sul, Brasil

^b Instituto de Matemática e Estatística, Universidade de São Paulo

Abstract

Nos últimos meses, a COVID-19 têm sido um dos principais problemas de saúde do mundo, causando um enorme ônus social e econômico. Neste cenário, modelos matemáticos e/ou estatísticos são extremamente importantes para se entender a dinâmica de propagação da doença e previsão das quantidades de novos casos. Neste relatório técnico, apresentamos os resultados de uma modelagem matemática/estatística das quantidades de casos confirmados da COVID-19 no estado de São Paulo, Brasil. Como as quantidades confirmadas da COVID-19 ao longo do tempo apresentou heterogeneidade temporal, adotamos o ajuste de um modelo por partes. Baseados no modelo ajustado apresentamos as previsões para a máxima quantidade de notificações e a data em que as taxas de notificação começarão a decrescer, chamado de ponto de inflexão. Na prática, o ponto de inflexão nos mostra quando as atitudes de combate estarão surtindo efeito.

Keywords: COVID-19, Modelagem, Predição.

1. Introdução

No mês de Dezembro do ano de 2019, um novo coronavírus foi descoberto na cidade de Wuhan, China. A Organização Mundial da Saúde (OMS) nomeou oficialmente esse coronavírus como COVID-19 (Wu and McGoogan, 2020; Ahmadi *et al.*, 2020). Desde a sua descoberta, o vírus se espalhou rapidamente pelo mundo; sendo atualmente um dos principais problemas de saúde, causando um enorme ônus social e econômico.

Neste cenário, análises quantitativas e o ajuste de modelos matemáticos e/ou estatísticos aos dados referentes as quantidades de notificações são extremamente importantes para se entender a dinâmica de propagação da doença, previsão e projeção das quantidades de novos casos. Além disso, os modelos matemáticos permitem a realização de simulações computacionais com o objetivo de obter a tendência futura das quantidades de casos confirmados da doença.

Neste relatório técnico apresentamos os resultados de uma modelagem matemática/estatística referente as quantidades de casos confirmados da COVID-19 no estado de São Paulo, Brasil. Ajustamos aos dados três modelos de crescimento. São eles: Modelo Exponencial (Abramowitz and Stegun, 1965), Modelo Logístico (Blumberg, 1968) e Modelo Gompertz (Gompertz, 1825; Tjorve and Tjorve, 2017). As estimativas para os parâmetros dos modelos foram obtidas utilizando o software R (R Core Team, 2009) e o comando *nls* do pacote *nlme* (Pinheiro *et al.*, 2020).

Os modelos foram comparados, utilizando como critério a medida de ajuste R^2 e os critérios de seleção de modelos AIC (Akaike, 1974; Bozdogan, 1987) e BIC (Schwarz, 1978). De acordo com estes três critérios, o modelo Gompertz é o modelo que melhor explica os dados. Utilizando o modelo selecionado, estimamos a quantidade máxima de notificações, a data que isto irá ocorrer e as quantidades de pacientes que precisarão de atendimento em leitos clínicos e de UTI.

O restante do relatório técnico está organizado da seguinte maneira. Na seção 2, apresentamos a descrição dos dados utilizados. Na seção 3, apresentamos o modelo ajustados e os principais resultados de interesse. Maiores informações sobre os modelos de crescimento considerados e o procedimento de estimação dos parâmetros podem ser obtidas via email aos autores.

2. Dados registrados

Os dados utilizados neste relatório técnico são referentes as quantidades de casos confirmados da COVID-19 no estado de São Paulo no período de 26/02/20 à 11/04/2021. Nestes 411 dias houve um total de 2.643.534 casos confirmados e 82.917 óbitos (Fonte: <https://www.seade.gov.br/coronavirus>).

A Figura 1 mostra o gráfico das quantidades de casos confirmados no dia t , para $t = 0, 1, \dots, n = 410$, onde $t = 0$ representa a data em que houve o primeiro caso confirmado (26/02/20) e $t = 410$ representa a data 11/04/21. A maior quantidade de casos confirmados em um único dia foi de 26.567 casos e ocorreu no dia 400 (01/04/20). No ano de 2020 foram confirmados um total de 1.462.297 em 310 dias; resultando em um média de 4.717 casos por dia. No ano de 2021 já foram confirmados um total de 1.181.237 em 101 dias; resultando em uma média de 11.695 casos por dia. Ou seja, nestes 101 dias do ano de 2021, temos uma média de casos diários 2,5 vezes maior que a média de casos diários registrados em todo o ano de 2020.

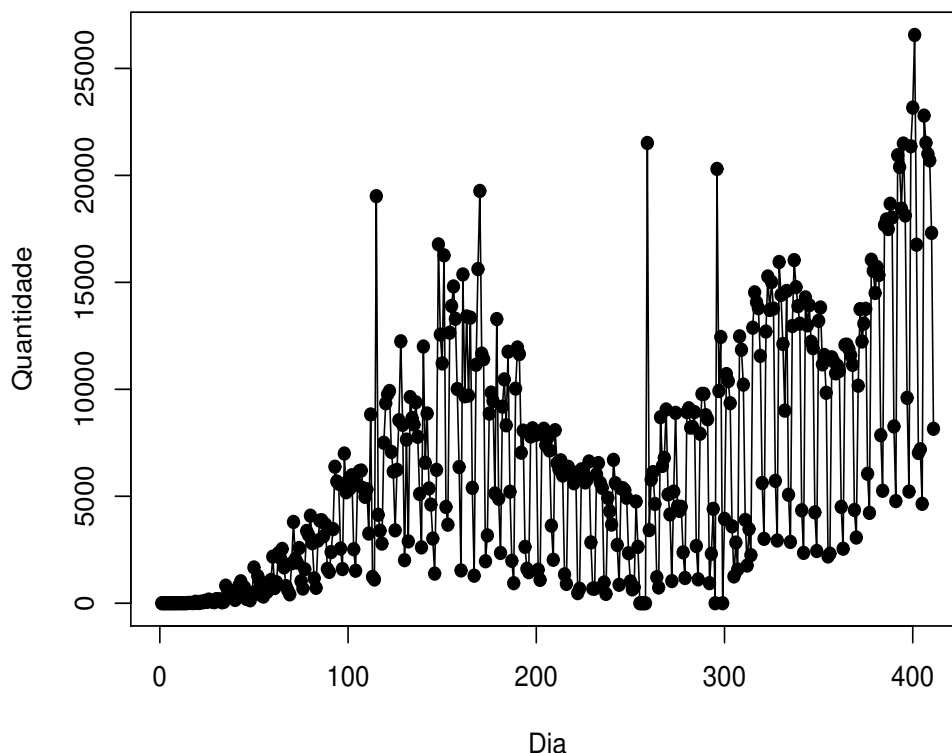


Figura 1: Quantidades de notificações por dia.

A Figura 2 mostra a distribuição de frequências das quantidades de casos confirmados por mês. Do total de casos confirmados, 4,15% (109.698 casos) foram registrados nos meses de Fevereiro, Março, Abril e Maio, 6,49% (171.682 casos) no mês de Junho, 9,87% (260.924 casos) no mês de Julho, 9,91% (262.038 casos) no mês de Agosto, 6,86% (181.286 casos) no mês de Setembro, 4,94% (130.499 casos) no mês de Outubro, 4,75% (125.526 casos) no mês de Novembro, 8,35% (220.644 casos) no mês de Dezembro, 11,92% (315.071 casos) no mês de Janeiro de 2021, 10% (264.260 casos) no mês de Fevereiro de 2021; 16,20% (428.221 casos) no mês de Março de 2021 e 6,57% (173.685 casos) nos primeiros onze dias do mês de Abril.

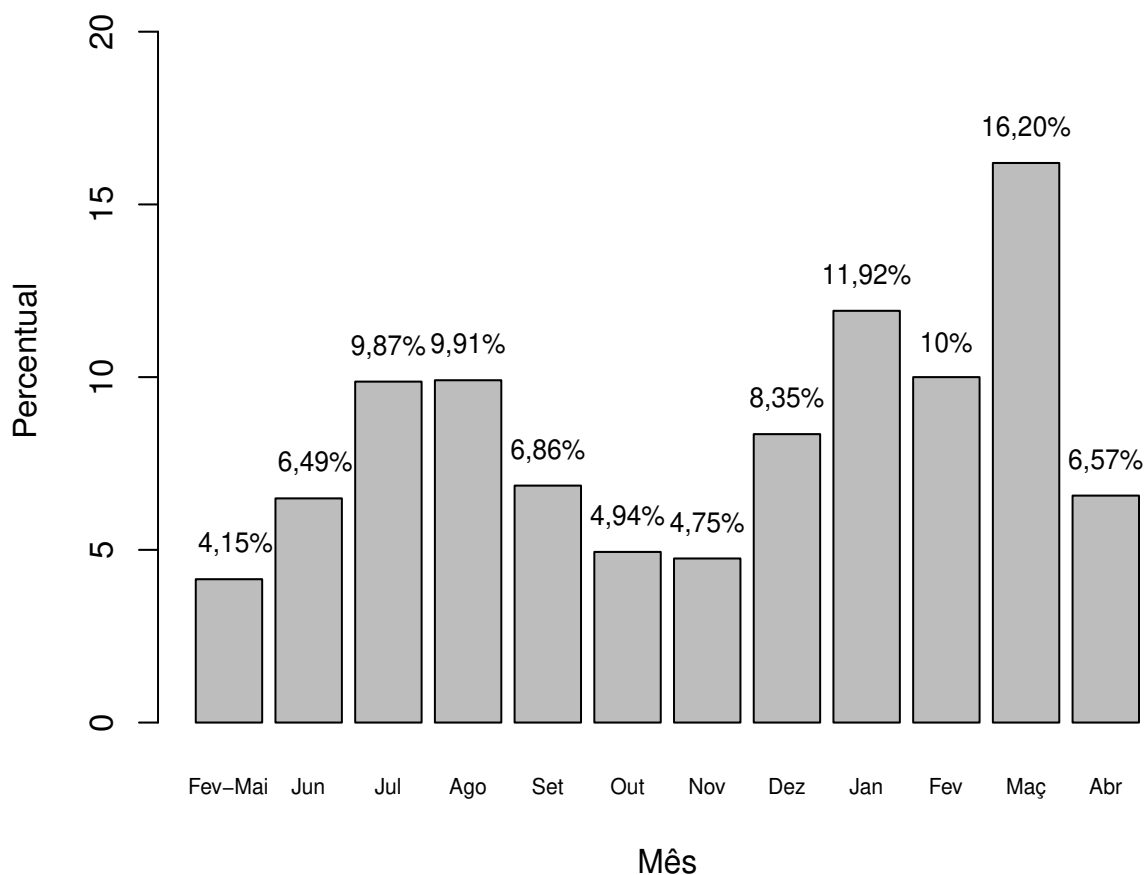


Figura 2: Percentual de casos confirmados por mês.

Note na Figura 2, que os dois meses com os maiores percentuais de casos confirmados ocorreram no ano de 2021 (Janeiro e Março); superando o primeiro pico da pandemia que ocorreu no mês de Agosto de 2020. Nos últimos quatorze dias (de 29/03 à 11/04) foram registrados 223.434 novos casos da COVID-19. Com relação aos últimos sete dias (de 29/03 à 04/04), foram confirmados 116.134 novos casos.

A Tabela 1 mostra as estatísticas resumo das quantidades de casos confirmados. O valor mediano é de 189,50 casos, com média de 225,60 casos confirmados por dia e desvio-padrão de 213,01.

Tabela 1: Estatísticas descritivas.

Data	Mínimo	1° Quartil	Mediana	Média	D.P.	3° Quartil	Máximo
11/04/21	0	1.572	5.327	6.432	5.616,38	9.884	26.567

A Figura 3 mostra o gráfico das médias móveis de 7 dias para o número de casos confirmados. O maior valor de média móvel (17.933), ocorreu no dia 400 (01/04/20). Após este pico, ocorreu 8 dias com queda nos valores das médias móveis. Porém, nos últimos dois dias o valor da média móvel voltou a aumentar. O valor da média móvel em 11/04 foi de 16.590,57 casos. Comparado ao valor da média móvel de sete dias atrás (15.328,57 em 04/04), tivemos um aumento de 8,23% no valor médio do número de casos confirmados. Comparado ao valor da média móvel de quatorze dias atrás (16.253,43 em 28/03), tivemos uma aumento de 2,07% no valor médio do número de casos confirmados.

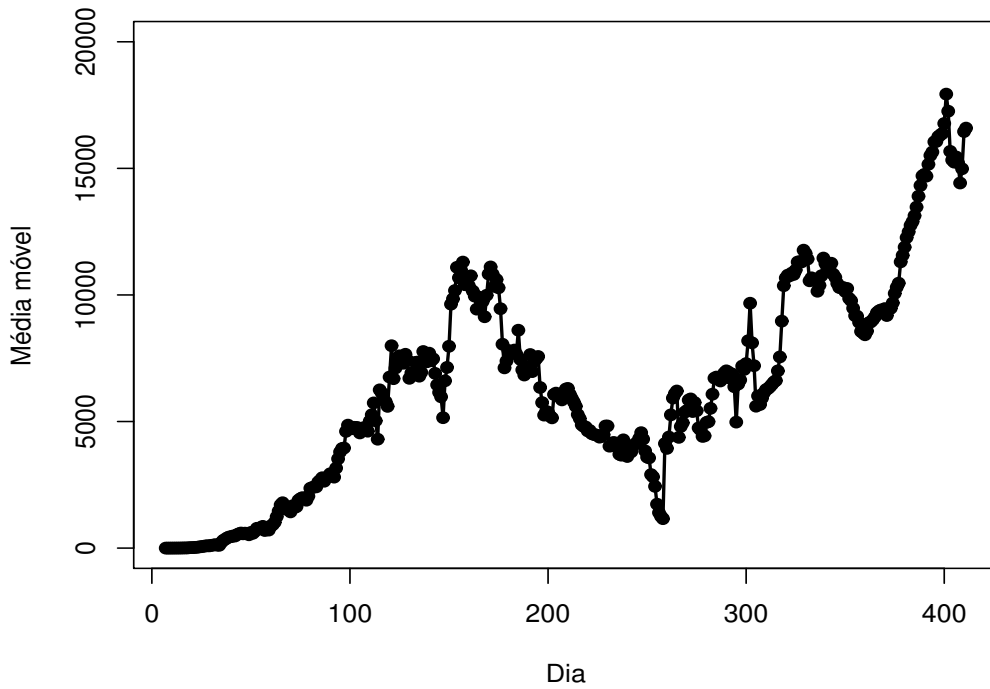


Figura 3: Média móvel de 7 dias do número diários de casos confirmados.

A Figura 4 mostra o gráfico das médias móveis de 7 dias para o número de óbitos. O maior valor de média móvel para o número de óbitos também ocorreu no dia 400 (01/04). Após este pico, ocorreram 8 dias com queda destes valores; porém nos últimos dois dias o valor da média móvel voltou a aumentar, ultrapassando o valor de 800 mortes diárias. O valor da média móvel em 11/04 foi de 842,43 óbitos. Comparado ao valor de sete dias atrás (718,43 em 04/04), temos um aumento de 17,26% no número médio de óbitos. Comparado ao valor da média móvel de quatorze dias atrás (633,29 em 28/03), temos um aumento de 33,02% no número médio de óbitos.

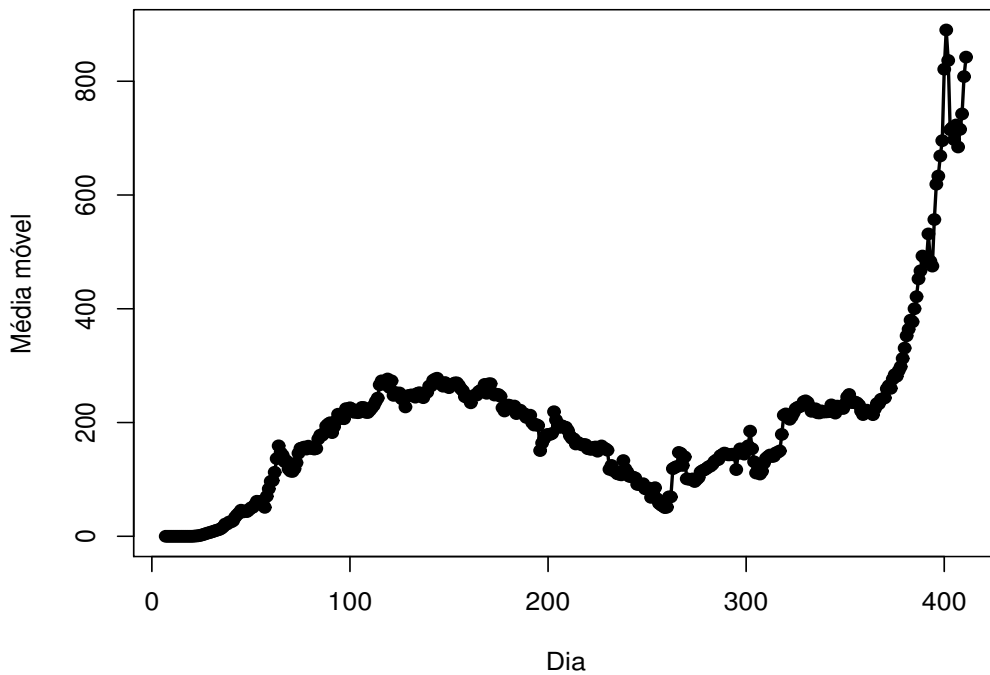
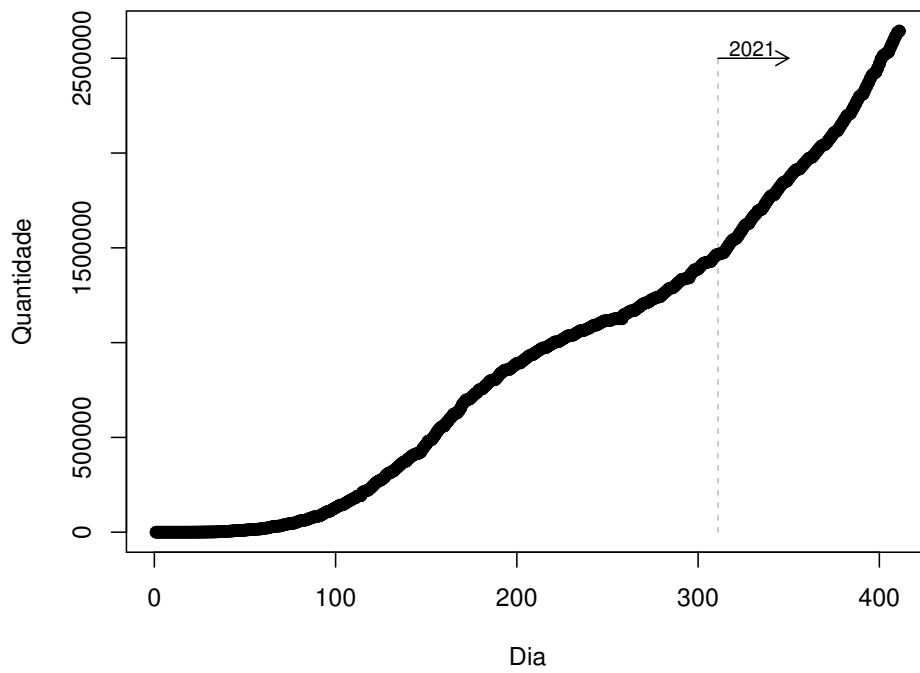
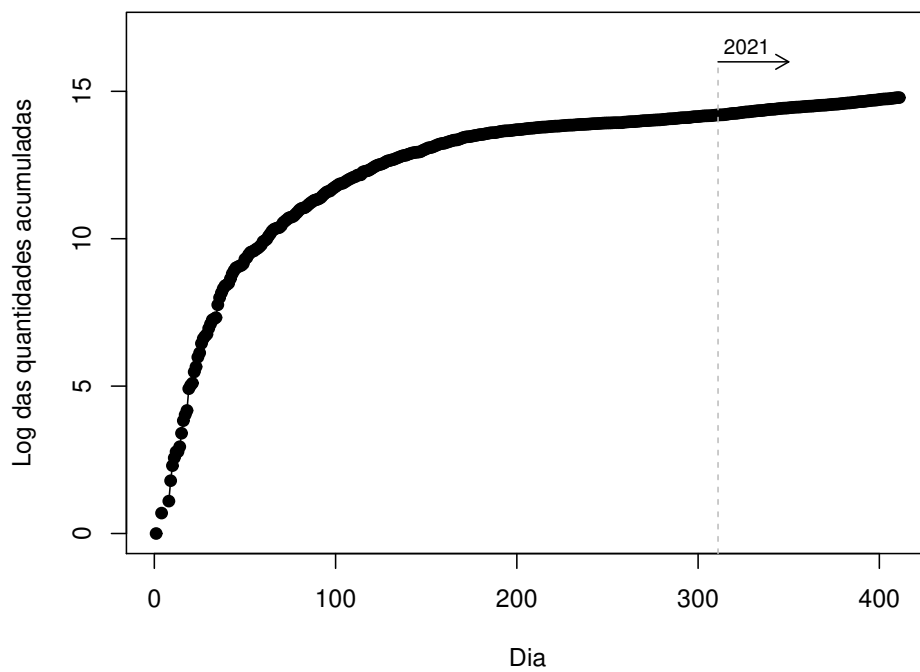


Figura 4: Média móvel de 7 dias do número diários de óbitos.

A Figura 5(a) mostra a quantidade total de casos confirmados até o dia t e a Figura 5(b) mostra o mesmo gráfico na escala logarítmica, para $t = 0, 1, \dots, n$.



(a) Quantidade total de notificações.



(b) Quantidade total de notificações - escala log.

Figura 5: Quantidades total de notificações até o dia t , para $t = 0, \dots, n$.

3. Resultados

Como o comportamento das quantidades de casos confirmados da COVID-19 apresentou heterogeneidade ao longo do tempo, adotamos o ajuste de um modelo por partes aos dados registrados. De acordo com nossas análises, as quantidades registradas no estado de SP apresentou 5 fases com distintas taxas de crescimento. A Figura 6 mostra os dados observados e o modelo ajustado para um período de 441 dias, sendo 411 dias de registros e 30 dias de projeção, até o dia 11/05/21.

Note que o modelo ajustado indica que estamos em uma fase, em que, o esperado é o crescimento do número de casos. Além disso, o modelo ajustado ainda não indica a ocorrência do pico. Ou seja, ainda não é possível prever quando as quantidades de casos começarão a diminuir.

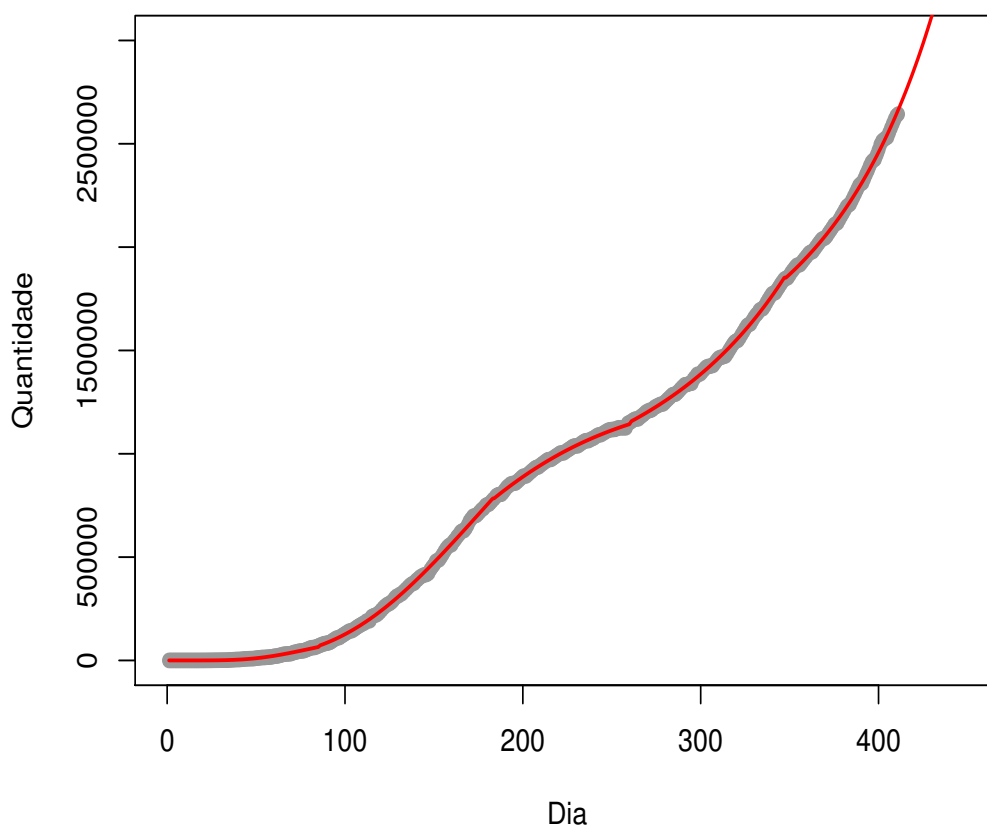


Figura 6: Modelo Ajustado para a quantidade de casos confirmados.

A Tabela 2 mostra os valores projetados pelo modelo para a última semana e os valores registrados. A Tabela também mostra as projeções do modelo para a quantidade total de casos que serão registrados nos próximos sete dias (de 12/04/21 à 18/04/21). A última linha da Tabela mostra o erro percentual absoluto entre o valor registrado e o valor projetado pelo modelo. O maior erro percentual cometido foi de 1,50%.

Tabela 2: Projeções do modelo ajustado para a quantidade de casos confirmados.

Dia	05/03	06/03	07/03	08/04	09/04	10/04	11/04	12/04	13/04	14/04	15/04	16/04	17/04	18/04
Projeção	2.565.293	2.584.114	2.603.253	2.622.715	2.642.507	2.662.638	2.683.113	2.703.940	2.725.128	2.746.683	2.768.614	2.790.929	2.813.637	2.836.745
Valor	2.532.047	2.554.841	2.5763.62	2.597.366	2.618.067	2.635.378	2.643.534	-	-	-	-	-	-	-
erro %	1,31	1,15	1,04	0,98	0,93	1,03	1,50	-	-	-	-	-	-	-

3.1. Modelo para o número de óbitos

A Figura 7 mostra o modelo ajustado para o número de óbitos. Note que, o modelo ajustado também indica que estamos em uma fase, em que, o esperado é o crescimento do número de óbitos. O modelo ajustado ainda não mostra o data de ocorrência do pico; ou seja, ainda não é possível projetar quando o número de óbitos começará a diminuir.

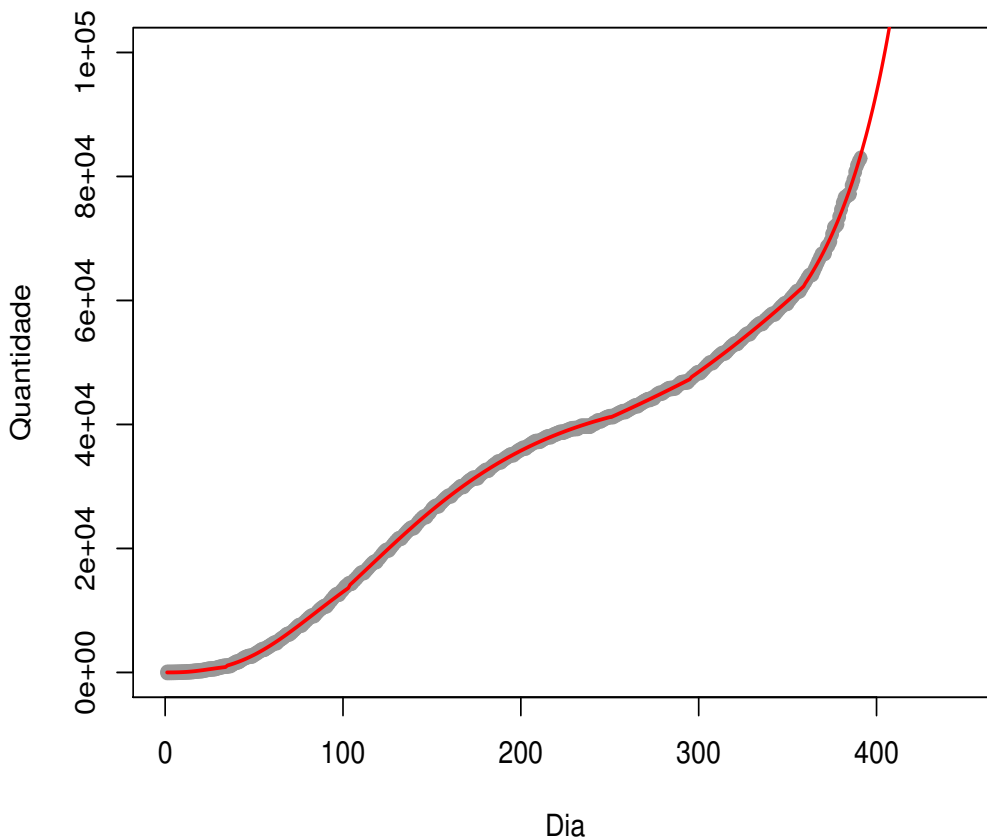


Figura 7: Modelo Ajustado para o número de óbitos.

A Tabela 3 mostra os valores projetados pelo modelo para o número de óbitos para a última semana e os valores registrados. A Tabela também mostra as projeções do modelo para a quantidade total de casos que serão registrados nos próximos sete dias (de 12/04/21 à 18/04/21). A última linha da Tabela mostra o erro percentual absoluto entre o valor registrado e o valor projetado pelo modelo. O maior erro percentual cometido foi de 1,03%.

Tabela 3: Projeções do modelo ajustado para o número de óbitos.

Dia	05/03	06/03	07/03	08/04	09/04	10/04	11/04	12/04	13/04	14/04	15/04	16/04	17/04	18/04
Projeção	77.963	78.806	79.673	80.565	81.482	82.426	83.398	84.399	85.429	86.490	87.583	88.709	89.870	91.066
Valor	77.165	78.554	79.443	80.742	81.750	82.407	82.917	-	-	-	-	-	-	-
erro %	1,03	0,32	0,29	0,22	0,33	0,02	0,58	-	-	-	-	-	-	-

A Figura 8 mostra o gráfico da Figura 7 juntamente com o modelo ajustado com os dados registrados até o dia 10/03/21 (32 dias atrás). Note que, após o dia 10/03/21 iniciou-se um “novo” período de crescimento do número de óbitos com taxa de crescimento mais agressiva (curva mais inclinada).

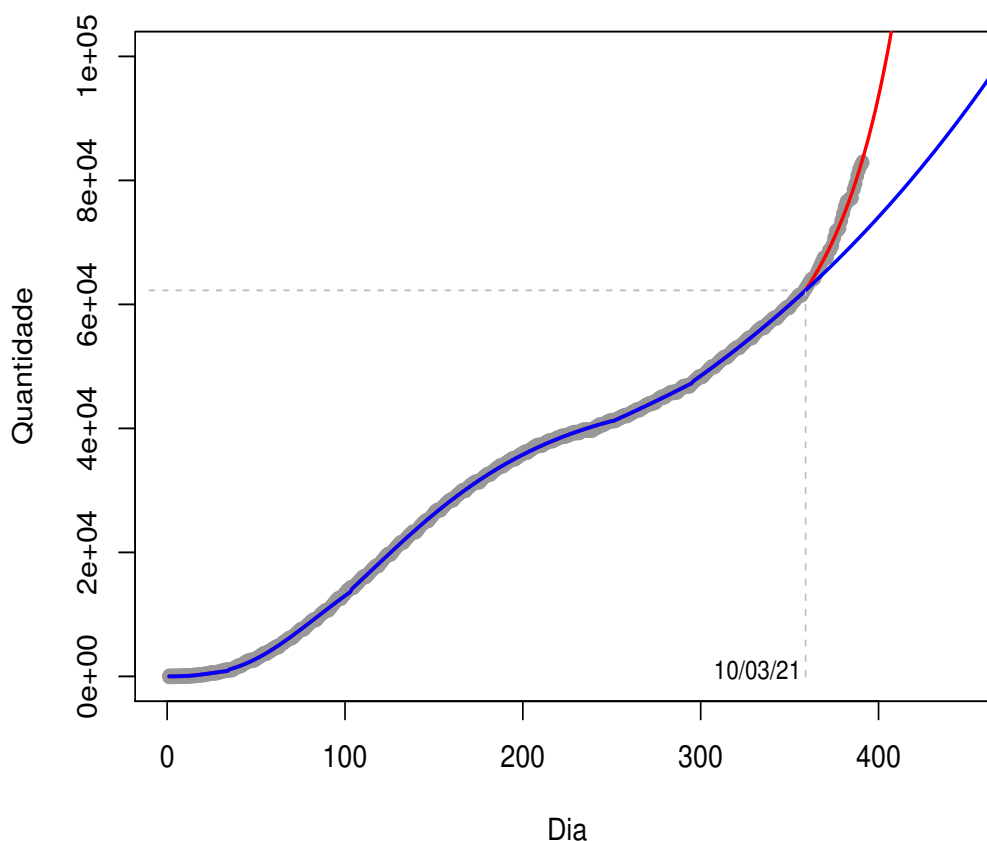


Figura 8: Modelo Ajustado.

Os resultados mostram a importância da população aderir as medidas restritivas impostas pelo governo estadual para se manter o isolamento social sempre que possível. Neste momento, este procedimento é extremamente necessário para que as quantidades registradas em uma semana estejam sempre abaixo da curva estimada. Pois somente desta maneira obteremos o desejado “achatamento” da curva.

Referências

- Abramowitz, M, Stegun, I. M., Handbook of Mathematical Functions, Dover, New York, 1965.
- Akaike, H. A. (1974). New look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716-723. MR0423716 <https://doi.org/10.1109/tac.1974.1100705>
- Ahmadi, A., Fadai, Y., Shirani, M., Rahmani, F., Modeling and Forecasting Trend of COVID-19 Epidemic in Iran, 2020, <https://www.medrxiv.org/content/medrxiv/early/2020/03/20/2020.03.17.20037671.full.pdf>.
- Blumberg, A. A., Logistic Growth Rate Functions. Journal of Theoretical Biology 1968, 42-44.
- Bozdogan, H. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika, v.52, p.345-370, 1987.
- Farias, C. M., Medronho, R. A. e Travassos, G. H. Avaliação do comportamento da COVID-19 no estado do Rio de Janeiro e seus municípios com base em R_0 calculado a partir das evoluções anteriores de R dos casos notificados à Secretaria de Estado de Saúde-RJ. Disponível em Covidimetro.

- Gompertz, B., On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 1825, 182:513?85.
- Pinheiro, J., Bates, D., DebRoy, S., Sakar, D. and R Core Team, *Linear and Nonlinear Mixed*. R package version 3.1-147, 2020. URL <https://CRAN.R-project.org/package=nlme>.
- R Development Core Team (2009). *R: A language and environment for statistical computing*. R foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://R-project.org>.
- Schwarz, G. E. (1978). Estimating the dimension of a model. *The Annals of Statistics* 6, 461?464.
- Tjorve, K. M.C., Tjorve, E., The use of Gompertz models in growth analyses, and new Gompertz-model approach: An addition to the Unified-Richards family. *PLoS ONE* 2017, 12, (6): e0178691. <https://doi.org/10.1371/journal.pone.0178691>.
- Winsor, C. P. THE GOMPERTZ CURVE AS A GROWTH CURVE. *Proceedings of the national academy of sciences*, **18**(1), 1932.
- Wu, Z. McGoogan, J. M, Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the Chinese Center for Disease Control and Prevention. *Jama*, 2020.