Método Computacional para Determinação do Parâmetro R_t e sua Relevância no Controle da Epidemia Covid-19 no Contexto da Cidade de Bragança Paulista

Francisco Fambrini[1], Vânia Gayer[2,3]

[1]Faculdade de Tecnologia de Limeira – Universidade Estadual de Campinas

R. Paschoal Marmo, 1888 - Limeira - SP, 13484-332 – SP – Brazil

[2]Fundação de Ensino Superior de Bragança Paulista (FESB)
Bragança Paulista, SP.
[3]Faculdade de Extrema (FAEX)

Estr. Mun. Pedro Rosa da Silva, 303 - Vila Rica, 37640-000– Extrema, MG – Brazil

fambrini@unicamp.br, vania.gayer@hotmail.com

***Abstract**: This article describes a statistical-epidemiological model underlying a computer program to estimate a measure of the spread of Covid-19 disease, caused by a coronavirus called SARS-Cov2. The program is applied to data from the city of Bragança Paulista (SP) between April 1st and August 3rd, 2020. The results show that the current value of the R_t parameter around 1.5 does not allow the feasibility of the quarantine measures and social isolation, under penalty of the spread of the epidemic, reaching up to 60% of the population of this municipality. The computational methods used here can be extended to other cities in the state of São Paulo and Brazil and constitute an important tool for the statistical analysis of the progression of the epidemic in Brazil.*

***Resumo**: O presente artigo descreve um modelo estatístico-epidemiológico subjacente a um programa computacional para estimar uma medida de propagação da doença Covid-19, provocada por um coronavírus denominado SARS-Cov2. O programa é aplicado a dados da cidade de Bragança Paulista (SP) entre os dias 01 de abril e 03 de agosto de 2020. Os resultados mostram que o valor atual do parâmetro R_t em torno de 1.5 não permite a fexibilização das medidas de quarentena e isolamento social, sob pena da propagação da epidemia, atingindo até 60% da população deste município. Os métodos computacionais aqui empregados podem ser extendidos a outras cidades do estado de São Paulo e do Brasil e se constituem em uma importante ferramenta para a análise estatística da progressão da epidemia no Brasil.*

1. Introdução

Muitas epidemias de diversas doenças com efeitos importantes sobre a população humana e animal tem se propagado ao longo dos séculos. Uma das mais conhecidas foi a peste negra na Europa no século XIV, que matou entre 25 a 75 milhões de pessoas, aproximadamente um terço da população européia. Na época do Brasil-colônia, a propagação de doenças européias, como por exemplo o sarampo e a varíola, teve um impacto desastroso sobre certas populacões

indígenas que não tinham desenvolvido resistência a essas doenças. Nos dias de hoje são epidemias importantes: hepatites B e C, HIV (vírus causador da AIDS), o vírus Ébola, o Zika vírus, os 4 subtipos do vírus da Dengue e mais recentemente, o SARS-Cov2. O entendimento de como uma doença se propaga numa determinada população permite conter a epidemia, através de vacinação ou quarentena [Systrom 2020]. Algumas doenças são mais contagiosas do que outras. Sarampo e gripe são altamente contagiosas, enquanto a febre glandular é muito menos. Dois parâmetros matemáticos se destacam para medir o grau de infectividade de um determinado agente patológico e controlar o avanço de uma epidemia: R_0 e R (ou R_t). Diversos epidemiologistas argumentam que a única maneira verdadeira de combater o Covid-19 é entender e gerenciar o parâmetro R_t [Sanche S., Lin Y., Xu C. et al. 2020]. Este texto discute brevemente estas métricas no contexto da atual pandemia do coronavirus SARS-Cov2 que atingiu o mundo no ano de 2020 e oferece uma ferramenta computacional eficiente para o cálculo do parãmetro R_t em tempo real. O artigo está organizado da seguinte forma: ítem 1 é a introdução; o ítem 2 descreve a definição e importância do parâmetro R_0 ; o ítem 3 introduz o conceito do parâmetro R_t e sua importância para o monitoramento das epidemias, o ítem 4 descreve o modelo SIR para epidemias, o ítem 5 apresenta os materiais e métodos utilizados, o ítem 6 apresenta os resultados da aplicação do *software*, através de tabelas e gráficos e finalmente, o item 7 encerra com as conclusões.

2. Significado do parâmetro R_0

R_0 é o número de reprodução básico de cada vírus. É também conhecido na literatura como Razão de Reprodutibilidade Basal [Qun L.M., Xuhua G. et al. 2020]. R_0 representa o número médio de contágios causados por cada pessoa infectada, em uma população onde todos os indivíduos são suscetíveis. A gripe sazonal, por exemplo, tem um número R_0 entre 1 e 2, a varicela tem um R_0 de cerca de 5, enquanto a catapora está em 10 e 12. Apesar de R_0 ser uma métrica estática, atrelada ao vírus, espera-se que efetivamente a taxa de infecção diminua quando a população entra em quarentena e com as medidas de distanciamento social, uso de máscaras, lavar as mãos com maior frequência e com a utilização do álcool em gel, entre outras ações [James, H. J. 2007]. Enquanto valores de R_0 situados no intervalo entre [3 ; 4] significam que o vírus se espalhará para quase todas as pessoas da população, um valor de 1.5 significa que tal patógeno pode se espalhar para cerca de 60% da população (dependendo das estimativas dos parâmetros) [James, H. J. 2007]. De qualquer forma, o objetivo deve ser reduzir R_0 para o menor valor possível antes de encerrar as medidas de isolamento social. A Tabela 1, adaptada de [Zaparolli, D. 2020], mostra alguns valores de R_0 para diferentes viroses humanas.

Tabela I – Valor de R_0 para algumas viroses. Adaptado de [Zaparolli, D. 2020].

Número básico de reprodução R_0 para diferentes viroses humanas	
Sarampo	15,6
Catapora	11,3
Caxumba	8,1
Rubéola	7
Varíola	5,2
Poliomielite	4,9
Covid-19	de 2,5 a 3
Gripe espanhola	de 1,5 a 4
H1N1 (2009)	1,46
Gripe comum	1,2

3. Número de Reprodução Eficaz (R ou R_t)

Para levar em conta as mudanças não capturadas pelo parâmetro R_0 , pode-se utilizar um outro parâmetro, o Número de Reprodução Eficaz (denotado por R ou R_t) ou em inglês *Effective*

Reproduction Number, que tem definição similar à do R_0 , com a diferença de que é uma métrica dinâmica. O R_t é o R_0 efetivo em um determinado momento, levando-se em conta as mudanças que podem afetar o R_0 , como o distanciamento social, por exemplo. R_t é o R_0 exposto às condições reais de evolução da doença. Quando se consegue estimar esta métrica, pode-se desenhar políticas para controlar a epidemia: com um R_t maior do que 1, a epidemia crescerá em ritmo exponencial. Somente se o R_t for menor do que 1, a epidemia irá diminuir de tamanho até desaparecer em determinada região [James, H.J. 2007]. Não é possível calcular diretamente o parâmetro R_t [Zaparolli, D. 2020], por isto este parâmetro deve ser estimado [Hefferman, J., Smith, R. and Wahl, L. 2005][Zaparolli, D. 2020]. Existem diversos métodos para estimar R_t. Um método específico foi descrito por [Bettencourt, L. and Ribeiro R.M. 2008] em seu artigo de 2008, *"Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases*". Este algoritmo utiliza estatística bayesiana para estimar o valor mais provável de R_t e também retorna um intervalo de confiança para o valor verdadeiro de R_t. O que se segue é uma aplicação do processo de Bettencourt e Ribeiro (2008) (com uma modificação importante introduzida por Kevin Systrom (2020)) aos dados Covid-19 da cidade de Bragança Paulista. O algoritmo original [Bettencourt, L. and Ribeiro R.M. 2008] para estimar R_t é uma função de quantos novos casos aparecem todos os dias. A relação entre o número de casos ontem e o número de casos hoje nos dá uma idéia do valor de R_t para determinada população [Zaparolli, D. 2020]. No entanto, não se pode confiar demasiadamente em um único dia para tentar prever o valor de R_t, pois a contagem diária de casos é imperfeita devido à alterações da capacidade de teste, defasagens nos relatórios de dados e notificações imperfeitas a respeito dos resultados dos testes. Apesar disso, aplicando-se adequadamente o Teorema de Bayes [Teorema de Bayes: o que é e qual sua aplicação 2018], pode-se usar as novas informações que foram obtidas a partir da contagem de casos a cada dia para ajustar a expectativa do valor de R_t, aproximando-o do valor real à medida que mais dados diários se tornam disponíveis. Para compreender como se calcula o parâmetro R_t é necessário analisar o modelo SIR para epidemias, um dos mais simples e mais utilizados na prática.

4. O Modelo SIR

No modelo SIR, os indivíduos são divididos nos seguinte grupos:

S: representa o número de indivíduos suscetíveis à doença (não imunizados);

I: representa o número de individuos infectados;

R: é o número de indivíduos que são resistentes a doença, por qualquer motivo.

N: é o número total de individuos da população.

Existem outros modelos matemáticos para epidemias, que são mais elaborados e mais complexos do que o modelo SIR. Por exemplo: SEIR (suscetíveis - expostos - infectados - recuperados) e SIRS (suscetíveis – infectados – recuperados - suscetíveis) [Amaku, M. 2001]. Este trabalho levará em conta apenas o modelo SIR. O modelo SIR parte de algumas hipóteses, que no caso da Covid-19 são aproximações teóricas:

1. Todos os indivíduos nascem suscetíveis;
2. O tamanho da população N é constante. Isto significa que a taxa de natalidade e a taxa de mortalidade μ da população são iguais;
3. A interação entre as populações dos compartimentos se dá de forma homogênea;
4. Os indivíduos recuperados ficam imunes por toda a vida;
5. A população está isolada, isto é, não possui contato com membros de outras populações;
6. A taxa de letalidade da doença é nula.

Em particular, os itens (5) e (6) citados são aproximações nem sempre verdadeiras para este caso, mas que simplificam as equações e permitem um modelamento matemático mais acessível [Cristovão, R. B. 2015]. Matematicamente, as funções S=S(t), I=I(t) e R=R(t) são todas dependentes do tempo, e sendo N o número total de indivíduos de uma determinada população (exemplo: para a região bragantina, tem-se N=169000 de acordo com [Dados do IBGE,

Bragança Paulista 2020] pode-se estabelecer as seguintes relações diferenciais entre as grandezas [Bettencourt, L. and Ribeiro R.M. 2008]:

$$\frac{dS}{dt} = -\beta \frac{S}{N} I \qquad (1)$$

$$\frac{dI}{dt} = \beta \frac{S}{N} I - \gamma I \qquad (2)$$

$$S(t) + I(t) + R(t) = N(t) \qquad (3)$$

onde β e γ são parâmetros que devem ser determinados experimentalmente, a partir dos dados coletados em cada região onde ocorre a epidemia. Pode-se demonstrar que os parâmetros R_0 e R_t se relacionam com os parâmetros das equações (1) e (2) através das seguintes relações, de acordo com [Bettencourt, L. and Ribeiro R.M. 2008]:

$$R_0 = \frac{\beta}{\gamma} \qquad (4) \qquad \text{e} \qquad R_t = \frac{S(t)}{N(t)} \times R_0 \qquad (5)$$

A dificuldade matemática está justamente em obter os parâmetros β e γ, porque são totalmente dependentes dos dados epidemiológicos empíricos compartilhados pelas autoridades de saúde [Cristovão, R. B. 2015].

5. Materiais e Métodos

Para determinar o parâmetro R_t, foi utilizado neste trabalho o algoritmo proposto por [Bettencourt, L. and Ribeiro R.M. 2008] e que foi implementado em linguagem Python no artigo de Kevin Systrom (2020). O código-fonte em linguagem Python foi disponibilizado gratuitamente e pode ser encontrado no repositório Github, através da referência bibliográfica. Os detalhes sobre este código-fonte podem ser obtidos diretamente em [Systrom, K . 2020].

O mesmo código foi também implementado em linguagem R, por [Vaidyanathan, R. 2020]. Outras ferramentas para o cálculo do valor de R_t em tempo real estão descritas em [Tianzhi W., Xijin G., Guang C.Y. and Erqiang H. 2020].

5.1 Distribuição de Poisson

O modelo descrito parte da Distribuição de Poisson para estimar a probabilidade de um determinado individuo se contaminar [Amaku, M. 2001]. Sendo λ a taxa média de infecções por dia, a probabilidade de ocorrer k novos casos em um dia é dada por:

$$P(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \qquad (6)$$

De acordo com a eq. (6), o software obtido em [Systrom, K . 2020] constroi a distribuição de probabilidade de novos casos para um conjunto de diferentes valores de λ. O software usa a função de cruzamento de purrr [Crossing function from purrr 2018] para calcular as densidades de probabilidade através de uma combinação de valores de k e λ distintos, além de usar o teorema dos antecedentes de Bayes para estimar a probabilidade de R_t.

6. Resultados

Neste artigo foi utilizada a base de dados fornecida pela Secretaria de Saúde da cidade de Bragança Paulista, com dados enviados entre 1 de abril a 3 de Agosto de 2020. Inicialmente, eram emitidos dois boletins diários com poucas informações. A partir do dia 18/05, as informações passaram a ser fornecidas de modo mais completo. Estas informações alimentaram o software [Systrom, K . 2020] para obter os gráficos mostrados nas figuras a seguir. A Figura 1

mostra o número de pessoas infectadas (confirmadas através de algum exame laboratorial) entre os meses de Abril a Agosto de 2020. Em cinza claro são plotados os valores reais referentes ao número de casos positivos para Covid-19 por dia. Em azul, é mostrada uma curva média suavizada destes pontos. A Tabela II mostra os valores de R_t médio, máximo e mínimo para os dias 30/07 até 03/08 de 2020, também calculados através de [Systrom, K . 2020].

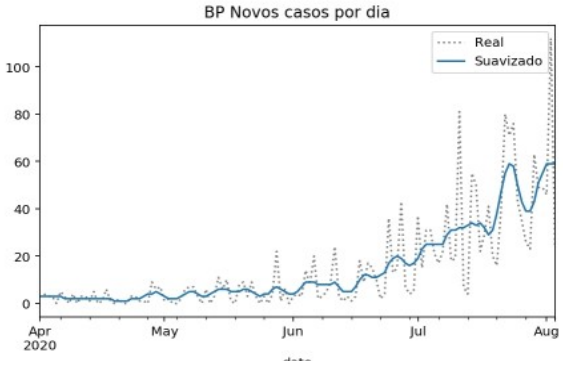

Tabela II – valores prováveis de Rt para os últimos 5 dias do conjunto de dados.

Data (2020)	Valor médio de Rt	Rt mínimo	Rt máximo
30/07	1.23	0.37	1.98
31/07	1.30	0.47	2.05
01/08	1.35	0.51	2.06
02/08	1.26	0.45	1.96
03/08	1.19	0.40	1.89

Figura 1- Números de pessoas infectadas a cada dia em Bragança Pta., entre os dias 01 de abril e 03 de agosto de 2020.

O software [Systrom, K. 2020] é capaz de plotar gráficos que mostram os valores de R_t em tempo real, dia após dia. Este gráfico é mostrado na Figura 2. A faixa em cinza, ao redor da linha do gráfico, mostra a densidade de probabilidade sobre os valores de R_t estimados.

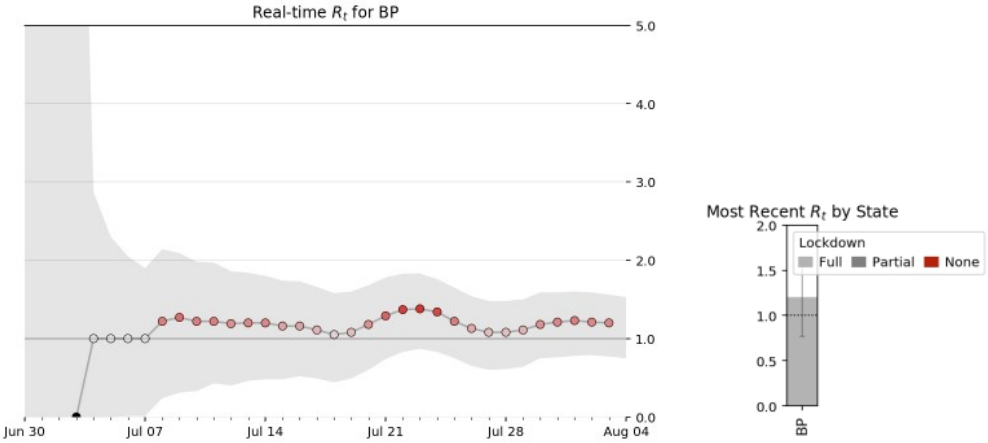

Figura 2- Valores do parâmetro R_t em tempo real, a cada dia, entre os meses de Julho a Agosto de 2020. À direita, gráfico em barras mostrando o valor médio de R_t. Obtido através de [Systrom, K. 2020].

7. Conclusão

Através da observação do gráfico mostrado na Figura 2, pode-se concluir que o valor de R_t está acima de 1.0 entre os meses de julho e agosto de 2020, o que mostra que a epidemia está avançando em Bragança Paulista de modo exponencial. Os autores esperam que, ao compartilhar este trabalho com a comunidade, que as autoridades sanitárias da cidade de Bragança Paulista considerem o parâmetro R_t como a métrica que pode guiar a análise e a tomada de decisões sobre a flexibilização do isolamento social e a retomada das atividades não essenciais, garantindo a segurança da população na ausência de um medicamento eficaz e de vacinas capazes de alterar o curso da doença. O *software* em estudo mostrou-se uma ferramenta eficiente para o cálculo do R_t em tempo real e é disponibilizado gratuitamente pelos autores.

8. Referências

Systrom, K. (2020). "The Metric We Need to Manage Covid-19". Available from: http://systrom.com/blog/the-metric-we-need-to-manage-Covid-19/

Sanche S, Lin Y, Xu C., et al. (2020). "High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2" . *Emerging Infectious Diseases*.26(7):1470-1477. doi:10.3201/eid2607.200282. Available from: https://wwwnc.cdc.gov/eid/article/26/7/20-0282_article

Qun L, M., Xuhua , G. et al. (2020). "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia". https://www.nejm.org/doi/full/10.1056/NEJMoa2001316

James, H. J. (2007). "Notes On R_0 ". Department of Anthropological Sciences, Stanford University, May 1. https://web.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf

Hefferman, J., Smith, R. and Wahl, L.. (2005). "Perspectives on the basic reproduction ratio". Journal of the Royal Society Interface 2, 281–293.

Bettencourt, L. and Ribeiro R.M. (2008). "Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases". PLoS ONE 3(5): e2185. doi.org/10.1371/journal.pone.0002185 Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0002185

Zaparolli, D. (2020). "O Desafio de Calcular o r". Revista FAPESP, edição 293. Available from: https://revistapesquisa.fapesp.br/o-desafio-de-calcular-o-r/

"Teorema de Bayes: o que é e qual sua aplicação?" (2018). Available from: https://www.voitto.com.br/blog/artigo/teorema-de-bayes

Cristovão, R. B. (2015). "Modelo SIR: Uma Aplicação à Hepatite A". Trabalho de Conclusão de Curso. Available from: www.ime.usp.br/~map/tcc/2015/Rafael%20Belmiro.pdf

"Dados do IBGE, Bragança Paulista" (2020). Available from: https://cidades.ibge.gov.br/brasil/sp/braganca-paulista/panorama

Systrom, K . (2020). "Estimating Covid-19's R in Real Time". April, 17. Source Code in https://github.com/k-sys/Covid-19/blob/master/Realtime%20R0.ipynb

Amaku, M. (2001). "Estudos de parâmetros epidemiológicos através de modelamento matemático: aspectos estacionários, espaciais e temporais". Tese (Doutorado em ciências). Universidade de São Paulo, USP, São Paulo.

Vaidyanathan, R . (2020). "Estimating Covid-19's Rt in Real-Time (Replicating in R)". April, 17. https://www.datacamp.com/community/tutorials/replicating-in-r-Covid19

Tianzhi W., Xijin G., Guang C.Y. and Erqiang H. (2020). "Open-source analytics tools for studying the Covid-19 coronavirus outbreak" doi.org/10.1101/2020.02.25.20027433. Available from: www.medrxiv.org/content/10.1101/2020.02.25.20027433v2

"Crossing function from purrr" (2018) . Available from: https://purrr.tidyverse.org/reference/cross.html