

GESTÃO DE DADOS DURANTE A PANDEMIA: UM CASO DE ESTUDO

F.C. Batista*,

(1) Escola Superior de Tecnologia e Gestão do Politécnico de Leiria,
*fernando.batista@ipleiria.pt, ORCID 0000-0002-1494-4163

Resumo

O aparecimento de vários casos positivos com coronavírus SARS-CoV-2 em vários países da Europa anteriores à primeira notificação oficial em Portugal (2 de março de 2020), permitiu o levantamento da suspeição que o vírus já estaria em circulação em Portugal. Neste sentido, o presente trabalho insere-se no esforço de tentar compreender melhor a dinâmica inicial da COVID-19 em Portugal utilizando os vários tipos de dados fornecidos pelas autoridades do país.

Palavras-chaves: COVID-19 em Portugal, SARS-CoV-2, Modelo epidêmico, Regressão logística

1 INTRODUÇÃO

Os primeiros surtos na Europa foram em Itália na região da Lombardia onde o primeiro caso oficial detetado foi no dia 20 de fevereiro. Um estudo feito nesta região italiana a 20 de março apontava para que o COVID-19 tivesse chegado muito antes de 20 de fevereiro, porque encontraram um caso com início de sintomas a 1 de janeiro de 2020 (Cereda et al. 2020). Em França, um estudo retrospectivo dos testes de PCR realizados em 58 doentes internados com pneumonia, entre 2 de dezembro de 2019 e 16 de janeiro de 2020, foi detetado o vírus SARS-CoV-2 a um doente de 42 anos que foi internado de emergência a 27 de dezembro de 2019 e já tinha sintomas há 4 dias (Deslandes et. al., 2020).

O primeiro caso positivo oficial em Portugal foi registado no dia 2 de março na região do Porto (<https://www.dssg.pt/>), nesta data já existiam vários casos positivos nos países vizinhos. Dia 11 março de 2020 foi diagnosticado COVID-19 a dois idosos no hospital de Santa Maria em Lisboa, um dos idosos já estava internado há uma semana no hospital (Diário de Notícias, 2020). O que demonstra, pelo tempo de incubação e o aparecimento de sintomas que já haveria infeção na população junto deste idoso pelo menos desde meados de fevereiro de 2020.

Existe, portanto, um conjunto de evidências de que o vírus causador da COVID-19 já estava na Europa muito antes dos primeiros casos positivos oficiais. A infeção provocada pelo vírus já estaria provavelmente a provocar casos clínicos e talvez mortes, mas por alguma razão isso não provocou nenhum alarme clínico.

Onde está então, a verdadeira data de início da epidemia de COVID-19 em Portugal? Procurando uma possível resposta, o presente trabalho utiliza uma abordagem estatística paramétrica, baseada num regressor logístico, que ajusta os dados oficiais de registo de casos positivos. Três indicadores são utilizados para determinar uma curva epidémica que bem ajusta os casos de COVID-19 registados em Portugal, sem, contudo, levar em consideração atrasos ou subnotificações. São eles o rácio entre testes positivos e testes negativos, a estabilização dos sintomas dos casos positivos registados e o índice de transmissão R_t . Esse regressor permite fazer uma projeção para trás no tempo, inferindo um intervalo para a provável data de início da epidemia em Portugal.

2 DADOS

Todos os dados utilizados foram extraídos dos boletins da DGS (figura 1) e foram agregados no serviço web Github (<https://github.com/dssg-pt/covid19pt-data>) pela organização não governamental, Data

Science for Social Good Portugal (<https://www.dssg.pt/>). Foram usados os dados dos casos confirmados do dia 2 de março de 2020 a 6 de outubro de 2020.

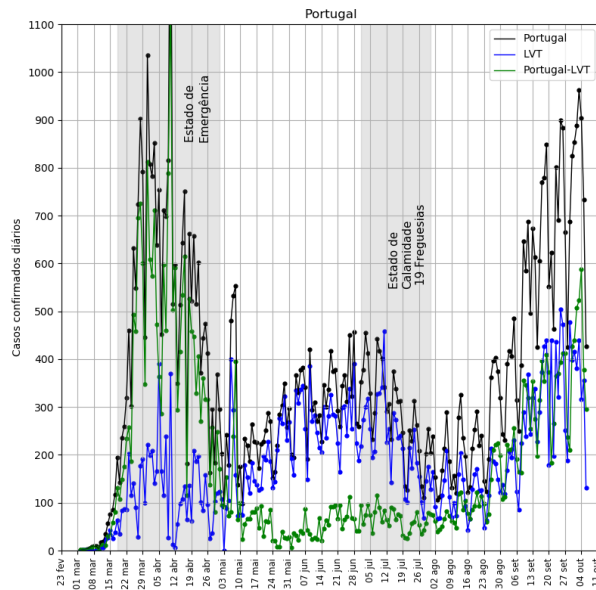


Figura 1. Incidência da COVID-19 em Portugal.

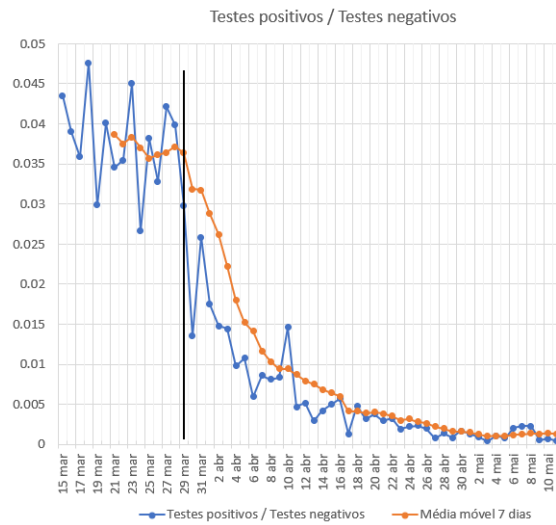


Figura 2. Rácio entre testes positivos e testes negativos

3 INDICADORES

O aumento do número de casos só estabilizou no fim do mês de março e houve a respetiva descida ao longo do mês de abril. Naturalmente esta descida já fez parte do lado descendente da curva epidémica que bem ajusta a evolução natural da epidemia. Pretendemos saber em que dia foi encontrada essa curva. Os três indicadores convergem para o dia 29 de março de 2020.

3.1 Rácio entre testes positivos e testes negativos

A mudança brusca de tendência do rácio entre testes positivos e testes negativos é reveladora de que encontramos os pontos que fazem parte da curva epidémica procurada (Figura 2). Ou seja, antes do dia 29 de março de 2020 não estávamos a testar verdadeiramente uma amostra significativa da população que desenvolveu a doença. Após esse momento verificou-se um aumento considerável de testes negativos face ao número de testes positivos e o rácio caiu.

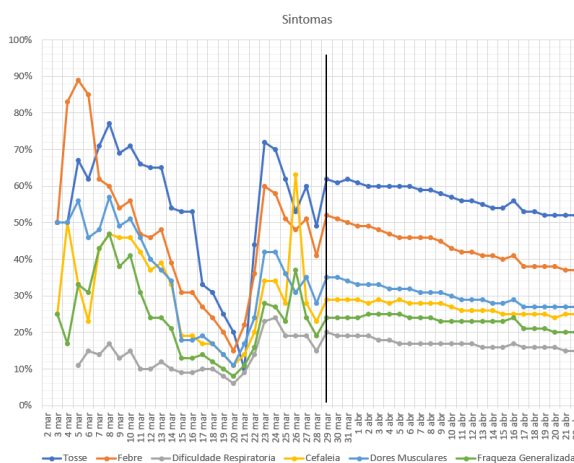


Figura 3. Percentagem de sintomas dos casos positivos.

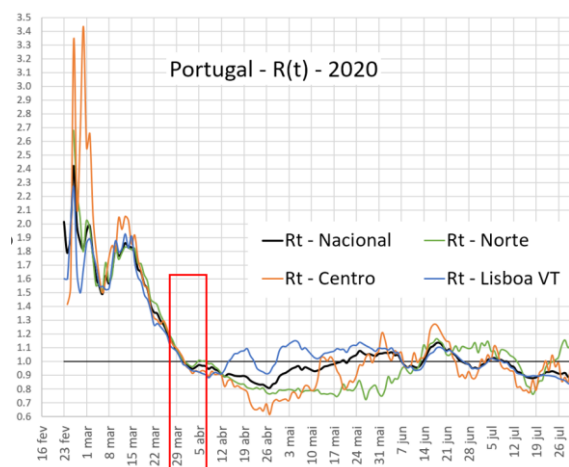


Figura 4. Índice de transmissão $R(t)$

3.2 Estabilização dos sintomas

No boletim diário da DGS é reportado diariamente informação percentual relativa a sintomas da grande maioria dos casos confirmados. Os sintomas reportados são; febre, tosse, dificuldade respiratória, cefaleia, dores musculares e fraqueza generalizada (Figura 3). Foi verificado que a partir do dia 29 de março os valores estabilizaram de forma mais ou menos linear. Esta estabilização parece ser indicadora de que, a partir do dia 29 de março, já estamos a testar uma amostra significativa da população que está a desenvolver a doença e diariamente todos os casos tem percentualmente os mesmos sintomas que nos dias anteriores, com uma queda suave dia após dia.

3.3 Índice de transmissão R(t)

O cálculo do R(t) tem como base a incidência registada. Só a partir do dia 29 de março é que surgiram valores coerentes com taxa de transmissão deste vírus (Figura 4).

4 MÉTODO

Para analisar corretamente os dados necessitamos de ferramentas estatísticas e de algum conhecimento de epidemiologia.

4.1 Media Móvel

Os dados brutos da Figura 1 têm presente a flutuação diária inerente ao processo de testagem e registo que existe ao longo do tempo. Para suavizar estas oscilações é frequente usar médias móveis de 3, 5 ou de 7 dias. No presente caso há um ciclo de aquisição de dados (testagem) com uma frequência semanal, neste sentido foi usado uma média móvel de 7 dias (mm7).

4.2 Curva logística

Toda a problemática da análise de dados da área da epidemiologia carece de uma utilização de ferramentas de análise estatística avançada. Existem modelos bastante complexos para abordar todo o problema, neste trabalho será utilizado um dos mais simples. O modelo da curva de crescimento acumulado de Richards (Richards, 1959), a chamada curva logística generalizada (equação 1), é um modelo de curva para estudos populacionais. Essa curva é utilizada em epidemiologia para previsão em tempo real de surtos de doenças.

$$C(t) = \frac{K}{1+e^{-r(t-t_m)}}, \quad (1)$$

A curva de incidência (curva epidêmica), é obtida pela derivada da curva de crescimento acumulado:

$$I(t) = \frac{rKe^{-r(t-t_m)}}{(1+e^{-r(t-t_m)})^2} \quad (2)$$

Ao longo do tempo da infeção de uma população, o vírus, pela dinâmica de movimentação dos infetados, dos recuperados e dos suscetíveis, vai infetar partes da população em tempos diferentes. Isto vai provocar ao longo do tempo alguma ondulação na curva epidêmica. É possível que o confinamento e o desconfinamento de partes da população num determinado momento, ajudem a explicar estas várias ondas. Ou seja, ao longo do tempo temos uma soma de curvas.

5 RESULTADOS

Na Figura 5 podemos ver que há uma alteração de tendência da mm7 no dia 15 de maio. Para estimar a “curva epidêmica estacionária” inicial vamos usar os dados entre 29 de março e 15 de maio. O regressor (equação 2) utiliza 3 parâmetros r , t_m e K , como estamos do lado direito da curva foi admitido 3 cenários com 3 picos diferentes (t_m), 29, 24 e 19 de março. O 1º cenário está representado na Figura 5 e podemos verificar com IC de 95% que a data de 8 de fevereiro é uma data provável do início da curva epidêmica. Fazendo uma média dos 3 cenários encontrou-se a data de 27 de janeiro (Figura 6).

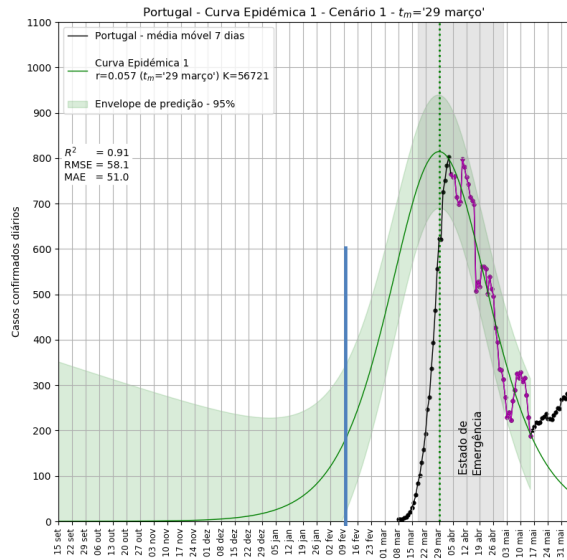


Figura 5. - Curva Epidêmica 1 - Cenário 1 - Pico 29 de março - Início 8 fevereiro

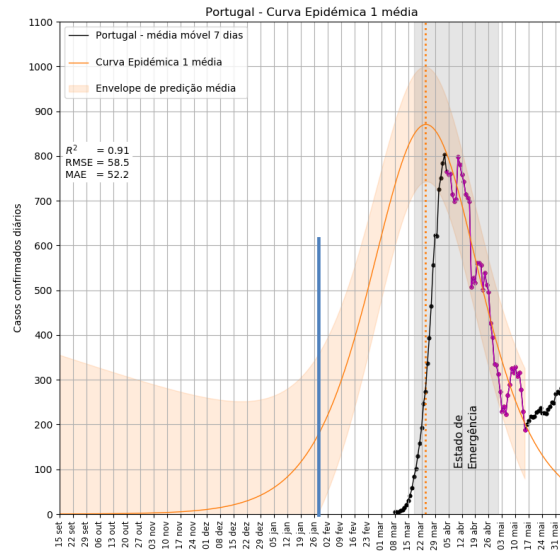


Figura 6. - Média dos três cenários da Curva Epidêmica 1 - - Início 27 janeiro

5.1 Outros autores

Delatorre, Edson et al analisaram os óbitos na Europa e concluíram que os primeiros contágios se deram nos últimos 15 dias de janeiro em Itália, França e Espanha (Figura 6).

Borges, V. et al analisaram o genoma do vírus em março e projetaram o primeiro contágio para 2 de fevereiro (Figura 7).

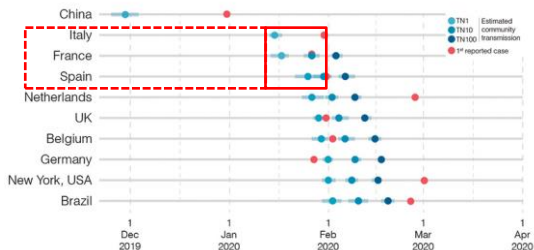


Figura 6. – Contagios com 21 dias antes do óbito

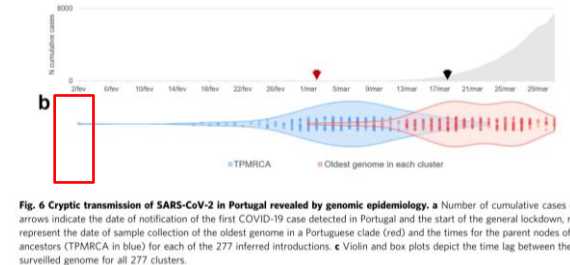


Fig. 6 Cryptic transmission of SARS-CoV-2 in Portugal revealed by genomic epidemiology. a Number of cumulative cases or arrows indicate the date of notification of the first COVID-19 case detected in Portugal and the start of the general lockdown, represent the date of sample collection of the oldest genome in a Portuguese clade (red) and the times for the parent nodes of t ancestors (TPMRCA in blue) for each of the 277 inferred introductions. c Violin and box plots depict the time lag between the surveilled genome for all 277 clusters.

Figura 7. – “Sequência genômica projetada”

6 CONCLUSÃO

Apesar do ajuste da curva epidêmica não ser feita com as notificações de sintomas, a data estimada com a incidência (27 janeiro) está em consonância com outros autores que usaram outros métodos.

REFERÊNCIAS

Borges, V., Isidro, J., Trovão, N.S. et al. SARS-CoV-2 introductions and early dynamics of the epidemic in Portugal. *Commun Med* 2, 10, 2022.

Cereda, D, and Other, The early phase of the COVID-19 outbreak in Lombardy, Italy, arXiv: Populations and Evolution, 2020.

Delatorre, Edson et al., Tracking the onset date of the COVID-19 outbreak in Lombardy, Italy, arXiv: Populations and Evolution, 2020.

Deslandes A., Berti V., Tandjaoui-Lambotte Y., Alloui C., Carbonnelle E., Zahar J.R., 2020.

Richards F. A. Flexible growth function for empirical use. *Journal of experimental Botany*, 1959.