

Modelagem Robusta de Dados de COVID-19 no Brasil

Luciano Salomão e Márcio Laurini
FEARP-USP

Dados da pandemia de COVID-19 sobre território brasileiro, em seu processo de divulgação e estudo, foram distorcidos no processo de coleta das informações. Curvas de casos e óbitos a nível nacional, amplamente expostas pela mídia durante a crise sanitária, mostraram-se temporalmente errôneas em comparação com informações posteriormente consolidadas. Os dados diariamente divulgados durante a pandemia eram de responsabilidade do Ministério da Saúde e baseavam-se no município e na data de notificação, sofrendo de um considerável atraso e alto grau de sazonalidade. Por sua vez, a base de dados de Síndrome Respiratória Aguda Grave, que engloba casos graves de COVID-19, era organizada pelo SIVEP e atualizada semanalmente. Possuía informações mais robustas para análises de curto prazo, uma vez que fornecia a data de início dos sintomas e a data de ocorrência dos óbitos, além de apresentar um comportamento muito similar ao das informações do Sistema de Informação sobre Mortalidade, administrado pelo Sistema Único de Saúde e fonte oficial de dados obituários.

Utilizando métodos de decomposição estrutural, foi exposto como a dinâmica de baixa frequência da série de óbitos por data de notificação, representada por componentes de tendência para frequência semanal, possui uma nítida defasagem temporal em relação a série de óbitos por data de ocorrência, principalmente em momentos de alta de casos e óbitos. Também foi realizada uma análise estatística baseada em *frameworks* de distribuições para dados de contagem, através de métodos Bayesianos, assumindo que a log intensidade varia ao longo do tempo através da soma dos componentes de tendência, ciclo e sazonalidade. Para tais estimações, foi utilizado o arcabouço do INLA (*Integrated Nested Laplace Approximations*) usando a razão de óbitos e casos graves da base SIVEP como variável dependente. O componente de tendência foi formulado como um processo de *Random Walk* de 2ª ordem, dada a natureza de rápido crescimento das variáveis pandêmicas; para o processo cíclico utilizamos um processo Autorregressivo latente de 2ª ordem, e um componente de sazonalidade semanal foi utilizado para controlar os erros de medida. Os resultados indicam que a estrutura proposta é uma alternativa viável e mais robusta em relação ao filtro de Médias Móveis utilizado para a extração da tendência.