Modelos preditivos da COVID-19 nos municípios brasileiros: correlação entre a densidade demográfica e os casos confirmados

Isadora C. R. Carneiro, Janaina C. da Silva, Eloiza D. Ferreira, Guilherme Soares, Guilherme F. Silveira\*

Instituto Carlos Chagas, Fiocruz/PR

\* Corresponde autor: [gfsilveira@gmail.com](mailto:gfsilveira@gmail.com)

# RESUMO:

# INTRODUÇÃO:

Vírus:

Infecção/Epidemia mundo:

Infecção/Epidemia Brasil:

# M&M:

Para o desenvolvimento de soluções computacionais utilizou-se linguagem de programação Python, com a justificativa de que essa linguagem obtém diferentes bibliotecas como ferramenta para análise de dados estatísticos. O conjunto de dados foram trabalhados em ambiente Anaconda (IDE Jupyter Lab) para a análise de dados a partir do número de casos confirmados de COVID-19 por município e a correlação a partir da densidade demográfica da população brasileira.

BANCO DE DADOS: O banco de dados utilizado para a gerar o modelo preditivo é um conjunto de informações geradas por Álvaro Justen e seus colaboradores a partir dos dados da Secretarias de Saúde Estaduais e pode ser acessado através de seu website brasil.io. Nesse banco de dados contém informação a partir de cada munícipio brasileiro que houve registro de casos confirmados por COVID –19. Também foi levantado e gerado um banco de dados contendo informações da densidade demográfica, sendo a população de pessoas por município e estado, faixa etária e gênero. Essas informações foram coletadas a partir de dados fornecidos pelo IBGE (Instituto Brasileiro de Geografia e Estatística).

CORRELAÇÃO: Para gerar a correlação entre a densidade demográfica populacional e o número de casos em municípios confirmados com COVID – 19, foi necessário importar diferentes bibliotecas para análise dos dados, como *pandas, numpy* e para gerar gráficos, foi importado *seaborn* e *matplotlib*. A partir de diferentes bancos de dados, foi criado um *DataFrame* contendo informações populacionais e casos confirmados, e assim poder gerar uma correlação entre eles. (MELHORAR)

MODELO ARIMA: Para gerar os resultados dos modelos preditivos, foi necessário importar a classe ARIMA a partir da biblioteca *statsmodels*. Ao usar o modelo ARIMA utilizamos os parâmetros p,q e d, em que p representa n° de termos auto regressivos, q o n° da média móvel e d o n° de diferenças não sazonais. A fim de buscar o melhor parâmetro para o modelo, foi testado diferentes ajustes. Para gerar o modelo, foi utilizado os parâmetros (1,2,2), pois apresentou melhor resposta e justaposição entre previsão e medição.

# RESULTADOS:

|  |
| --- |
|  |
| **Figure 01. The distribution of the analyzed data is not Gaussian.** The values of (A) population, (B) confirmed cases and (C) deaths are log-normal distribution. |

|  |
| --- |
|  |
| **Figure 02. Population has a positive correlation (R = 0.6) with the confirmed cases.** Spearman correlation between population, confirmed cases, deaths, confirmed/100k habitants and death rate in all 542 cities (A) or 540 cities with population lowest of 6.000.000 (B). |

|  |
| --- |
|  |
| **Figure 03. Positive correlation (R = 0.6994) between population and confirmed cases occurs only in cities with population up to 296.844 habitants.** Spearman correlation and linear regression model (95% confidence) between population and confirmed cases in 5 quartiles distribution of population. (A) 1149 – 23.286, (B) 23.286 – 56.428, (C) 56.428 – 132.709, (D) 132.709 – 296.844 and (E) 296.844 – 12.252.023 population |

|  |
| --- |
|  |
| **Figure 04. Positive correlation (R = 0.6449) between population and deaths occurs only in cities with population up to 296.844 habitants.** Spearman correlation and linear regression model (95% confidence) between population and deaths in 5 quartiles distribution of population. (A) 1149 – 23.286, (B) 23.286 – 56.428, (C) 56.428 – 132.709, (D) 132.709 – 296.844 and (E) 296.844 – 12.252.023 population |
|  |

|  |
| --- |
|  |
| **Figure 05. The number of confirmed cases and ratio of confirmed / 100k inhabitants shows a clear upward trend.** Decomposition of the time series of the daily values of number of confirmed cases (blue), deaths (orange) and ratio of confirmed / 100k inhabitants (green), in components (A) raw data, (B) trends, (C) seasonality and (D) randomness. |

|  |
| --- |
|  |
| **Figure 06. Average estimate of 63,638 confirmed cases in 30 days.** ARIMA model (1,2,2) of forecast of confirmed cases until 05/10/2020. Confirmed cases (blue), forecast (orange), model fit analysis (green) and forecast interval with 95% confidence (gray). Up to the end date, between 56,829 and 70,447 cases are expected. |

# DISCUSSÃO:

Existe correlação entre a transmissão do COVID-19 e a densidade demográfica da população à nível municipal e estadual no Brasil?

Uma vez que as medidas de circulação tem sido o principal meio de controle, será que isso varia de acordo com o número de pessoas?

Quais é o nível de trânsito médio (distância e periodicidade) das viagens intermunicipais?

# BIBLIOGRAFIA: