Predictive model of COVID-19 incidence in Brazilian municipalities: correlation between population, MHDI and demographic density

Isadora C. R. Carneiro1 #, Eloiza D. Ferreira1#, Janaina C. da Silva2 #, Guilherme Soares1,3, Guilherme F. Silveira\*1

1. Grupo de Pesquisa e Desenvolvimento em Análise de Dados - DataLab, Instituto Carlos Chagas - Fiocruz/PR, Curitiba, PR
2. Laboratório de Biologia de Tumores, Universidade Estadual do Oeste do Paraná, UNIOESTE, Francisco Beltrão-PR.
3. Laboratórios De Referência Em Viroses Emergentes E Reemergentes, Instituto Carlos Chagas - Fiocruz/PR, Curitiba, PR

\* Corresponding author: gfsilveira@gmail.com, guilherme.silveira@fiocruz.br

# These authors contributed equally to the work.

Na minha opinião, olhar para a correlação entre número absoluto de casos confirmados e outras características do município não parece a escolha mais adequada para estudos epidemiológicos baseados em dados agregados. Meu raciocínio é o seguinte: um município com mais habitantes ter mais casos e um município com menos habitantes ter menos casos é esperado e não implica que eles tenham riscos diferentes para a doença (ou seja, podem ter taxas de incidência parecidas). Por isso é importante olhar a taxa, ela vai nos mostrar municípios fora do padrão esperado para eles. Portanto, a resposta de interesse, na minha opinião, deveria ser a taxa de incidência de COVID-19/ 100.000 habitantes (e não apenas a quantidade de casos confirmados ou) e a letalidade. Então, a correlação dessas variáveis (taxa de incidência e letalidade) com características socioeconômicas e demográficas dos municípios poderia ser explorada, tendo também o mesmo cuidado com essas variáveis, no sentido de utilizar indicadores – na forma de proporção ou taxas.Características/indicadores demográficas e socioeconômicas (renda per capita, % de idosos, % de homens, densidade domiciliar, residentes em favela/habitantes, etc) e relacionadas a assistência à saúde (leitos hospitalares/habitantes, médicos/habitantes, equipamentos de suporte básico de vida/habitantes, número de equipes do programa saúde da família/habitantes, etc) poderiam ser interessantes de se relacionar com a taxa de incidência e de letalidade.

# ABSTRACT:

Coronaviruses are enveloped viruses that are part of a large family of single-stranded RNA viruses with positive polarity genome who cause respiratory, gastrointestinal, liver and neurological diseases. On February 11, 2020, after phylogenetic and pathophysiological analyzes, the new coronavirus (2019-nCoV) was named SARS-CoV-2 due to its similarity with the SARS-CoV. In Brazil, the official data from the Ministry of Health indicate 28,320 infected and 1,736 deaths until 16/04/2020. It has already been shown that measures of conduct and social distance alone may not be enough to prevent the spread of COVID-19. In this context, the present work aims to contribute to the decision-making process through exploratory data analysis of COVID-19 cases from Brazilian municipalities. It was noticed a positive correlation between the number of confirmed cases and the following characteristics: size of the population, for municipalities with more than 295,985 residents, Municipal Human Development Index (MHDI) and demographic density, regardless of the number of residents. Additionally, we adjusted an autoregressive integrated moving average (ARIMA) model for the number of cases which suggests, with 95% confidence, that at 05/10/2020, if the conditions do not change, we will have between 56,829 and 70,447 cases in Brazil. These findings add information that may be relevant for (…), thus expanding the possibilities of decision making at the micro and macro-regional level.

Keywords: COVID-19, COVID-19 in Brazil, Exploratory data analysis, ARIMA, computational model.

# INTRODUCTION:

Coronaviruses are enveloped viruses that are part of a large family of single-stranded RNA viruses with positive polarity genome who cause respiratory, gastrointestinal, liver and neurological diseases. These can infect many species of animals, including humans, causing acute and chronic illnesses. (Chang et al., 2012; Weiss, 2011). It is known that most infections caused by coronaviruses in humans induce a mild form of the disease, where the patient usually has flu-like symptoms. However, after the discovery of the Severe Acute Respiratory Syndrome (SARS), a greater contagion capacity and potential lethality of this viral family was evidenced (Weiss; Leibowitz, 2011). The etiologic agent for SARS, SARS-CoV, was identified in mid-2003, after an outbreak of the disease in November 2002, in Guangdong province, China, where 8,700 cases were confirmed with 774 deaths (Contini et al., 2020).

The so-called new coronavirus, initially described as 2019-nCoV, was first described when a group of patients reported symptoms of pneumonia of unknown cause in the city of Wuhan, Hubei province, China, in December 2019 (ZHU, 2020). On February 11, 2020, after phylogenetic and pathophysiological analyzes, the new coronavirus (2019-nCoV) was named SARS-CoV-2 due to the similarity it had with the SARS-CoV, as announced by the Coronavirus Study Group (CSG) of the International Virus Taxonomy Committee in accordance with the 2015 World Health Organization nomenclature guidelines (Gorbalenya et al., 2020). The pathology caused by the infection by SARS-COV-2 received the name of COVID-19, being characterized by a flu-like condition associated with fever and cough, which can progress to a stage of pneumonia and dyspnea in more severe cases (CHAN et al., 2020). The disease incubation period varies from 2 to 14 days, and in approximately 80% of cases, individuals who become infected remain asymptomatic, however, unlike influenza infection, these patients are competent for viral transmission (Contini et al., 2020). Also, according to Contini (2020) his mechanism of contagion is direct, that is, through contact with the infected person, through handshakes, saliva droplets, sneezing, coughing, aerosols, among others. Other studies show that SARS-CoV-2 can survive in the air for more than 3 hours and on surfaces such as plastics and metals for up to 3 days (Van Doremalen et al., 2020). Currently, there are no vaccines to combat the disease, reinforcing the need for prophylactic measures, the main ones being: the correct hygiene of hands, environments and surfaces and social distance.

The most recent World Health Organization (WHO) estimates of the status of the pandemic, considering the date of 04/15/2020, indicate 1914,916 confirmed cases and more than 123,010 deaths. In Brazil, the official data from the Ministry of Health indicate 28,320 infected and 1736 deaths until 4/16/2020, with the state of São Paulo concentrating most of the notifications reaching 11,043 cases and 778 deaths. Still, according to the national perspective of the disease, the states of Amazonas, Amapá, Distrito Federal, Ceará, São Paulo, Rio de Janeiro and Roraima are the states of emergency, that is, they need to redouble care in relation to disease prevention for being 50% above the national incidence (Ministry of Health, 2020). It has already been shown that the measures of conduct and social distance alone may not be enough to prevent the spread of COVID-19, and the global impact of this viral infection is of great concern (Sohrabi et al., 2020). Additional research is needed to help define the exact rates and mechanisms of human-to-human transmission, as well as knowledge of additional factors that may guide containment actions. The COVID-19 pandemic requires strict surveillance and continuous monitoring to accurately track and predict potential host adaptation, evolution, transmissibility and pathogenicity (Sohrabi et al., 2020). Different approaches are being used in order to better understand the transmission dynamics of SARS-CoV-2 in order to apply pandemic prevention and control measures. In this context, the present study aims to contribute to the decision-making process from the exploratory data analysis (EDA) and predictive computational model of cases in Brazilian municipalities, seeking statistically significant correlations between confirmed cases and mortality with demographic data and the municipal human development index (MHDI), thus expanding the possibilities of decision making at the micro and macro-regional level.

Sugestão: a seção de métodos e técnicas poderia contemplar os seguintes tópicos:

OK - Desenho do estudo: estudo ecológico

OK - Unidade de análise: população pertencente a uma área geográfica definida (municípios ou Estados)

- Período (como os dados são públicos, cada um corresponde a períodos diferentes) - perguntar – MODELO ARIMA – Adicionar o período análisado.

OK - Fontes de informação

- Indicadores calculados - perguntar

- Medidas utilizadas para descrever o perfil da taxa de incidência de COVID-19. - Descrever pós novos resultados

OK - Ferramenta utilizada/Pacotes.

# M&M:

Desenho do estudo:

For the exploratory data analysis (EDA) and the development of computational solutions, Python programming language was used, which allows the use of diverse libraries, specific for this purpose. The data sets were worked in an Anaconda environment (IDE Jupyter Notebook), platforms widely applied in the field of data science, offering users robust and established tools and libraries. The project can be accessed through the github page, <https://github.com/gfsilveira/covid>.

Database: COVID-10 confirmed cases records at municipal level were obtained through a set of information from Health Departments of the Federative Units compiled by Álvaro Justen and his collaborators available on https://brasil.io/dataset/covid19/caso. The code license is LGPL3 and the converted data Creative Commons Attribution ShareAlike. Demographic and socioeconomic characteristics publicly available at the municipal level, such as demographic density, Municipal Human Development Index (MHDI) and total area in km² (…) were obtained from Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística – IBGE) for542 municipalities.

Correlation: For the Spearman’s correlation tests, different libraries were used for data analysis, such as pandas, numpy and for graphics generation, the seaborn and matplotlib libraries were used. The different databases were imported into python, using the pandas library, in the DataFrame data structure format, which contained municipal information such as: number of inhabitants in the population, confirmed cases, deaths, confirmed cases for 100k inhabitants and case ratios confirmed and deaths, official area of the municipality in km², inhabitants in population density data, population density, and Municipal Human Development Index (MHDI).

ARIMA model: For the development of the predictive model we used the Autoregressive Integrated Moving Average (ARIMA) models from the statsmodels library. When using the ARIMA model, different values were tested for p, q and d parameters, where p represents the number of auto-regressive terms, q the number of the moving average and d the number of non-seasonal differences. In order to search for the parameters that best fit the model to the data obtained, all p, q, and d configurations accepted by the algorithm were tested. The configuration chosen was (1,2,2) because it presents better response and juxtaposition between forecast and measurement.

M&M – Correção.

Utilizando o conceito de Estudo ecológico, no presente trabalho foi utilizado esse método para compreender como se dá a correlação entre distribuição populacional por municípios brasileiros e a taxa de contaminação de COVID-19 por área geográfica. Considerando as 542 cidades presentes em um banco de dados alimentado diariamente, que tiveram confirmação do novo Coronavírus, sua área territorial, densidade demográfica, idade média entre os habitantes, gênero, dados socioeconômicos e IDHM (Índice de Desenvolvimento Humano Municipal).

Para a análise exploratória de dados (EDA) e o desenvolvimento de soluções computacionais, foi utilizada a linguagem de programação Python, que permite o uso de diversas bibliotecas, específicas para esse fim. Os conjuntos de dados foram trabalhados em um ambiente Anaconda (IDE Jupyter Notebook), plataformas amplamente aplicadas no campo da ciência de dados, oferecendo aos usuários ferramentas e bibliotecas robustas e estabelecidas. Junto da linguagem Python, foi necessário importar diferentes pacotes e bibliotecas, como *Pandas,* *Numpy e Scipy*, utilizadas para organização e estruturação dos dados. Para cálculos estatísticos, foi importado *Statsmodels (statsmodels.tsa.seasonal e* sm.tsa.statespace.SARIMAX)*,* um modulo usado em Python utilizado para gerar as séries temporais e a classe ARIMA para as previsões. M*atchplotlib e Seaborn,* utilizadas para gerar gráficos de duas dimensões (2D). O projeto pode ser acessado através da página do github, <https://github.com/gfsilveira/covid>.

Base de dados: Os registros de casos confirmados do COVID-19 em nível municipal foram obtidos por meio de um conjunto de informações diários dos Departamentos de Saúde das Unidades Federativas compiladas por Álvaro Justen e seus colaboradores disponíveis em <https://brasil.io/dataset/covid19/caso>. Características demográficas e socioeconômicas disponíveis publicamente a nível municipal, como densidade demográfica, Índice Municipal de Desenvolvimento Humano (IDHM) e área total em km² (…) foram obtidas do Instituto Brasileiro de Geografia e Estatística (IBGE) a partir do censo demográfico realizado em 2010. Já os dados referentes da idade média entre os habitantes e gênero, foram obtidas a partir do censo demográfico realizado em 2015.

Correlação: Para os testes de correlação de Spearman, diferentes bibliotecas foram usadas para análise de dados, como pandas, numpy e para geração de gráficos, as bibliotecas seaborn e matplotlib. Os diferentes bancos de dados foram importados para python, usando a biblioteca pandas, no formato de estrutura de dados DataFrame, que continha informações municipais como: número de habitantes na população, casos confirmados, óbitos, casos confirmados para 100 mil habitantes e proporções confirmadas e óbitos , área oficial do município em km², habitantes em dados de densidade populacional, densidade populacional e Índice de Desenvolvimento Humano Municipal (IDHM). (ESPERANDO NOVOS RESULTADOS A PARTIR DAS SUGESTÕES)Medidas utilizadas para descrever o perfil da taxa de incidência de COVID-19: O perfil de incidência foi considerada através da correlação entre municípios de maior população com o número de casos confirmados por covid-19. DESCREVER APOS RESULTADOS

Séries Temporais: A Série Temporal foi gerada a partir do banco de dados de casos confirmados por COVID-19. O período analisado é de 25 de janeiro até o dia 22 de abril de 2020, demonstrando o número de casos em municípios brasileiros a partir de dados diários, verificando a tendência, sazonalidade dos dados e apresentando o ruído que não foi incorporada a série.

Modelo (S)ARIMA : Para o desenvolvimento do modelo preditivo, usamos os modelos ARIMA (*Autoregressive Integrated Moving Average*) que utiliza os parâmetros p, q e d, em que p representa o número de termos auto-regressivos, q o número da média móvel e d o número de diferenças não sazonais, porém, não obtivemos bons resultados ao testar diferentes parâmetros, portanto, foi incorporado a Sazonalidade da série temporal (SAMIRA), que adiciona novos três hiperparâmetros: SAMIRA(p,d,p)x(P,D,Q)m em que m representa número de etapas de tempo para um único período sazonal. Foi gerado uma lista de diferentes parâmetros e filtrado o de menor valor AIC (Critério de Informação Akaike), que nos resultou o melhor ajuste de AIC = 397.846 e o valor de P>|Z| (..descrever..) foi igual a zero (0.00). O parâmetro utilizado com melhor ajuste foi SARIMA (2,2,2)x(1,0,1)30 em que 30 representa o intervalo de dias. Após determinar os parâmetros, foi testado os ajustes do modelo (Figura 01). O modelo de previsão está ajustado dentro do período analisado para gerar a previsão para os próximos meses foi o mesmo para gerar nossa série temporal. Já o período reservado para avaliar a previsão com IC95% foi a partir do dia 23 de macho a 20 de abril de 2020.

Tela de celular com publicação numa rede social

Descrição gerada automaticamente

**Figure 01. Parcelas Residuais para testar os parâmetros utilizados**. (A) Gráfico de Resíduos: os erros residuais estão próximos de uma média de zero (linha). (B) Gráfico de Histograma: este gráfico de densidade sugere distribuição normal dos dados a partir do modelo ajustado. (C) Gráfico QQ Normal: os pontos (quantil) estão alinhados com a linha vermelha, testando o modelo. (D) Gráfico ACF: apresenta os erros residuais que não são correlacionados automaticamente. Testando o parâmetro d (número de diferenças não sazonais que foi utilizado no modelo). (E) Ajuste da previsão e dos dados que foram observados.

# RESULTS:

1. The number of inhabitants has a positive correlation with the number of confirmed cases.

Since COVID-19 is a pathology caused by coronavirus that is transmitted directly from human to human, we seek to understand whether the characteristics of the affected towns can help in understanding the pandemic. The database analyzed has 542 municipalities in the 26 states of the federation. Seeking to understand the distribution of values referring to the population (number of inhabitants), confirmed cases and deaths confirmed by COVID-19, we analyzed the histograms of the samples (Figure 1). We observed a log-normal distribution of the data, which indicates the need to use the Spearman correlation test, seeking to understand the relationship between the different characteristics.

|  |
| --- |
|  |
| **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. |

Spearman's correlation in population values, confirmed cases, deaths, confirmed cases per 100k inhabitants and ratio of confirmed cases and deaths, was tested for all cities (Figure 2A) or cities with less than 6x106 inhabitants (Figure 2B). As expected, we observed a strong negative correlation between Case-Fatality Rate when compared to confirmed cases (R = -0.93), or the population (R = -0.77) or deaths (R = -0.39), in both conditions. Medium positive correlation values were observed between deaths and confirmed cases in both conditions (R = 0.45), as well as Incidence rate/ 100.000 residents and confirmed cases (R = 0.39). Additionally, we detected an medium positive correlation (R = 0.63) between the confirmed cases and the population of inhabitants, which is maintained when removing the most populous municipalities (R = 0.62).

|  |
| --- |
|  |
| **Figure 02. Population has a positive correlation (R = 0.63) 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). |

The results show a correlation between the number of inhabitants in the population and the number of confirmed cases, either in the entire sample, or in cities with less than 6x106 inhabitants. In order to better understand this correlation, confirmed cases were analyzed according to the population sample, divided according to their quartiles, with 5 groups from 1,149 to 22,809 (Figure 3A), 22,809 to 56,250 (Figure 3B), 56,250 to 132,079 (Figure 3C), 132,079 to 295,985 (Figure 3D) and 295,985 to 12,252,023 (Figure 3E) inhabitants. It was observed that for the first 2 quartiles, the correlation is non-significant neutral, with a value of R = 0.0373 and 0.04 and *p* = 0.5857 and 0.5579, respectively for the first and second groups. In the intermediate quartiles, there is a clear weak positive correlation, with R = 0.2468 and 0.1897 and *p* = 0.0002 and 0.0364. The last quartile, containing the municipalities with more than 295,985 inhabitants, showing a strong positive correlation with R = 0.6877 and *p* <0.0001.

|  |
| --- |
|  |
| **Figure 03. Positive correlation (R = 0.6994) between population and confirmed cases occurs only in cities with population up to 295.985 habitants.** Spearman correlation and linear regression model (95% confidence) between population and confirmed cases in 5 quartiles distribution of population. (A) 1149 – 22.809, (B) 22.809 – 56.250, (C) 56.250 – 132.079, (D) 132.079 – 295.985 and (E) 295.985 – 12.252.023 population |

Additionally, the same stratification of municipalities by the number of inhabitants was carried out to analyze the correlation between population and deaths due to COVID-19 (Figure 4). It was observed that the first, second and fourth strata showed a non-significant neutral correlation, with R = 0.0251, 0.0451 and 0.0994 and *p* = 0.7141, 0.5088 and 0.2762. It was not possible to calculate metrics for the third tier. The last quartile, with the municipalities that have more than 295,985 inhabitants, presented a strong positive correlation with R = 0.6813 and *p* <0.0001.

|  |
| --- |
|  |
| **Figure 04. Positive correlation (R = 0.6449) between population and deaths occurs only in cities with population up to 295.985 habitants.** Spearman correlation and linear regression model (95% confidence) between population and deaths in 5 quartiles distribution of population. (A) 1149 – 22.809, (B) 22.809 – 56.250, (C) 56.250 – 132.079, (D) 132.079 – 295.985 and (E) 295.985 – 12.252.023 population. |
|  |

Municipal data referring to the number of inhabitants in the population, confirmed cases, deaths, confirmed cases per 100k inhabitants and the ratio of confirmed cases and deaths, allows us to observe a positive correlation between population size in inhabitants and confirmed cases. This correlation is significant only in municipalities with more than 295,985 inhabitants. This suggests that other characteristics of these locations may be correlated with the pathology.

1. The Municipal Human Development Index (MHDI) and population density correlates with the number of confirmed cases of COVID-19

It is known that different conditions present in the municipalities can influence the spread of the epidemic of different pathogens. The results obtained here show a positive correlation between population and confirmed cases in municipalities with more than 295,985 inhabitants. Seeking to better understand the factors that may explain this correlation and, considering that COVID-19 is transmitted from person to person, different characteristics of the 542 municipalities were analyzed, with new population data now obtained from the Brazilian Institute of Geography and Statistics (IBGE). For that, Spearman correlation tests were performed with the number of confirmed cases (Confirmed), deaths, confirmed per 100k inhabitants (Confirmed/100k), ratio of confirmed cases and deaths per confirmed cases (Reason for Death), official area of ​​the municipality in km² (Area), inhabitants in population density data (Population\_Dens), population density (Density), and Municipal Human Development Index (MHDI). The total data (Figure 5A) were stratified in municipalities with fewer (Figure 5B) or more (Figure 5C) of 295,985 inhabitants.

|  |
| --- |
|  |
| **Figure 05. The Municipalities Human Development Index (MHDI) and the demographic density of the municipalities are positively correlated with the number of confirmed cases.** Spearman's correlation for the different characteristics of municipalities with confirmed cases. (A) For all municipalities, positive correlation for demographic density (R = 0.56) and MHDI (R = 0.43). (B) For all municipalities with less than 300,000 inhabitants, a positive correlation for demographic density (R = 0.45) and MHDI (R = 0.34). (C) For all municipalities with more than 300,000 inhabitants, a positive correlation for demographic density (R = 0.37) and MHDI (R = 0.43). |

The results for new data base confirm the analyzes previously carried out in relation to the negative correlation between the ratio of confirmed cases and deaths and confirmed cases (R = -0.88), population (R = -0.74) or deaths (R = -0.27). The medium positive correlation values between deaths and confirmed cases were also confirmed (R = 0.43), as well as confirmed cases for 100k inhabitants and confirmed cases (R = 0.46) and the medium positive correlation between confirmed cases and the population of inhabitants of the previous base (Population\_COVID \_ R = 0.71) and the IBGE base (Population\_Dens \_ R = 0.7). It was observed that, when the municipalities with less than 300,000 inhabitants were analyzed, the negative correlation between the ratio of confirmed cases and deaths and confirmed cases (R = -0.93), population (R = -0.69) or deaths (R = -0.13), remains. However, in cities with a population over 300,000 inhabitants, we observed a variation with a negative correlation between the ratio of confirmed cases and deaths and lower confirmed cases (R = -0.52), null correlation for population (R = -0.072) and positive for deaths (R = 0.21). This characterizes as an inversion of the conditions of confirmed cases for 100k inhabitants in the cities with bigger population. Additionally, for the medium positive correlation between confirmed cases and the population in the previous base (Population\_COVID - R = 0.57) and the IBGE base (Population\_Dens - R = 0.55) in municipalities with less than 300,000 inhabitants and, (Population\_COVID \_ R = 0.76) and the IBGE base (Population\_Dens \_ R = 0.76) in municipalities with more than 300,000 inhabitants. Again, we observed differences between the locations with the largest or smallest population, reinforcing the hypothesis that municipal characteristics must be observed in order to understand the pandemic.

In this contest, we analyzed three characteristics of the municipalities that are: the official area of the municipality in km² (Area), population density (Density), and Municipal Human Development Index (MHDI). In the total data, confirmed cases have zero correlation with the Area (R = -0.07) and medium with Density (R = 0.56) and MHDI (R = 0.43). This is maintained in the group of municipalities with a population of less than 300.00 inhabitants, with a weak negative correlation with the Area (R = -0.12) and a positive medium with Density (R = 0.45) and MHDI (R = 0.34). In cities with more than 300.00 inhabitants, it was possible to observe a null correlation with Area (R = -0.0093) and medium with Density (R = 0.37) and MHDI (R = 0.43). Therefore, as well as the relation of the number of inhabitants, the Density and the MHDI, it also has a positive relationship with the number of confirmed cases of COVID-19.

In order to better understand these correlations, we performed a new Spearman’s test, now only with data from MHDI (Figure 6) and Demographic density (Figure 7). For MHDI, it was possible to show a medium positive correlation with confirmed cases, with R = 0.3396 and *p* <0.0001, in cities with fewer inhabitants (Figure 6A) and R = 0.4293 and *p* <0.0001 in cities with the largest number of inhabitants (Figure 6B).

|  |
| --- |
|  |
| **Figure 06. The Municipalities Human Development Index (MHDI) has a positive correlation with the number of positive cases.** Spearman's correlation for MHDI for all (A) municipalities with less than 300,000 inhabitants (R = 0.3396), or (B) municipalities with more than 300,000 inhabitants (R = 0.4293). |

As for demographic density, an medium positive correlation was observed with confirmed cases, with R = 0.4474 and p <0.0001, in cities with fewer inhabitants (Figure 7A) and R = 0.369 and p <0.0001 in cities with the largest number of inhabitants (Figure 7B).

|  |
| --- |
|  |
| **Figure 07. Demographic density has a positive correlation with the number of positive cases.** Spearman correlation for demographic density for all (A) municipalities with less than 300,000 inhabitants (R = 0.4474), or (B) municipalities with more than 300,000 inhabitants (R = 0.369). |

Therefore, we can conclude that the positive correlation between cities with more than 295,985 inhabitants and the number of confirmed cases (Figure 3) or deaths (Figure 4) is corroborated by the positive correlation with demographic density (Figure 7B) and the MHDI (Figure 6B). However, in cities with less than 295,985 inhabitants, which have not significant correlation with the number of confirmed cases, they have a positive correlation with demographic density (Figure 7A) and MHDI (Figure 6A). This suggests that demographic density and MHDI are factors that should be considered as factors for understanding the epidemic in all 594 municipalities studied. Still, no difference was observed between the two indexes, which suggests that, at least regarding viral spread (confirmed cases), municipalities with a denser population and with better socioeconomic status, regardless of population size, are at higher risk.

1. Until 05/10/2020, the evolution model allows predicting 56,829 to 70,447 confirmed cases.

The results presented suggest that municipal characteristics should be considered regarding the current epidemiological condition. However, due to the current level of infection in the cities analyzed, the scarcity of data does not allow the development of a robust predictive model for cases confirmed at the municipal level. Seeking to understand the condition of the infection at the national level, we analyzed the time series of accumulated data for confirmed cases (blue), deaths (orange) and confirmed for 100k inhabitants (green) from 02/25/2020 to 04/10/2020 (Figure 8A). The decomposition of the time series allows to perceive a clear tendency of increase in the number of confirmed cases and deaths, however, still a stationary condition for confirmed for 100k inhabitants (Figure 8B). As for seasonality (Figure 8C) and the random component (Figure 8D), the variation in the number of confirmed cases is significantly greater than the other data analyzed, as expected.

|  |
| --- |
|  |
| **Figure 08. 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. |

Once the components of the daily time series of the data were determined, where we observed a clear upward trend in the number of cases, we sought to determine a computational model for predict the evolution of COVID-19 in Brazil. The best adjusted model for the forecast was ARIMA (1,2,2), where we reached the forecast of 56,829 to 70,447, up to 05/10/2020, with 95% confidence (Figure 9).

|  |
| --- |
|  |
| **Figure 09. 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. |

# Conclusions:

1) Positive correlation between the size of the population and the number of confirmed cases, only in cities with more than 295,985 inhabitants.

2) Positive correlation between the MHDI (Municipal Human Development Index) and the number of confirmed cases, regardless of the number of inhabitants.

3) Positive correlation between demographic density and the number of confirmed cases, regardless of the number of inhabitants.

4) There is no difference between the correlation between the MHDI (Municipal Human Development Index) and the demographic density, with the number of confirmed cases.

5) Predictive model for the number of cases suggests with 95% confidence that in 30 days, if the conditions do not change, we will have between 56,829 and 70,447.

# Discussão:

O Brasil possui, atualmente, 5570 municípios distribuídos desparamente em uma área total de 8.511.000 km² que conta com 26 estados. Desse total de munícipios, apenas 95 possuem população superior a 295.955 habitantes, sendo que todas as capitais estaduais e mais o Distrito Federal se enquadram nesta situação. Em porcentagem, estes 95 municípios representam apenas 1,71% do total de cidades do país (IBGE,2019). A maior parte das cidades brasileiras, 94,22% (5.245 munícipios), possui população menor ou igual a 100.000 habitantes (IBGE,2019).

Ter conhecimento da distribuição demográfica da população pelo território nacional se torna ainda mais importante quando almeja-se compará-la com os resultados obtidos e apresentados no presente artigo. Anteriormente, concluiu-se que os munícipios com população superior a 295.955 habitantes apresentavam uma correlação positiva entre o tamanho da população e o número de casos confirmados. Logo, baseado nos resultados do presente artigo, é possível notar que a transmissão humano-humano é mais suscetível de ocorrer em menos de 2% dos municípios brasileiros, contudo, é imprescindível reforçar que estas 95 cidades, juntas, abrigam 83.951.535 habitantes, o que representa 40% do total da população do Brasil, que atualmente conta com densidade demográfica de 209,5 milhões de pessoas (IBGE, 2018), logo, negligenciar as recomendações da Organização Mundial da Saúde acerca do isolamento e conduta social, em época de pandemia, será um atentado contra a saúde pública do Brasil, por mais que o presente artigo apresente a correlação positiva entre casos e população municipal acima de 295.955 habitantes.

Quando analisados os dados com relação aos estados da Federação, nota-se que segue na mesma linha obtida em relação aos munícipios, sendo o estado de São Paulo o mais populoso do país e por sua vez sendo o estado com maior número de casos confirmados de COVID-19, 11.043 casos até a presente data, segundo o Ministério da Saúde.

Por meio de estudos de dados descritos no presente artigo, também é possível ter-se uma análise mais refinada a respeito de que as novas medidas de conduta e isolamento social empregados no país têm sido o principal meio de controle da pandemia.

Quando levamos em consideração os dados obtidos, é possível dizer que sim, o isolamento social é uma medida válida a ser aplicada naqueles munícipios que possuem população residente superior a 295.000 habitantes e que nestes munícipios, quanto mais intensificada a medida for, maior será o achatamento da curva de transmissão, possibilitando assim que hospitais e unidades de saúde consigam ter um maior controle sob a demanda de pacientes que inspiram cuidado especializado.

Entretanto, é importante citar que não é possível concluir a importância real do isolamento social nos municípios com população inferior ao número supracitado, contudo, ao que indicam os resultados, existe uma relação negativa entre a densidade demográfica e o número de casos nessas cidades, o que, em teoria, indicaria uma falta de relação entre estes aspectos.

Estudos anteriores a este e realizados em outro país indicam que apenas o isolamento social é insuficiente para controlar o surto, sendo necessárias novas intervenções para alcançar o controle (Hellewell et. al.,202), contudo, o mesmo artigo demonstra que o isolamento pode contribuir para reduzir o tamanho geral de um surto ou controlá-lo por um maior período de tempo (Hellewell et. al.,202). Levando isso em consideração, torna-se possível avaliar sobre a importância de medidas de isolamento social, inclusive, para os munícipios com menor população, concluído que esta medida possui sim grande importância, mas que deve ser intensificada nas cidades mais populosas e não negligenciada nas cidades com menos de 295.000 habitantes.

Como forma de auxiliar no entendimento dos problemas, é interessante avaliarmos o trânsito médio da população brasileira. Do total de passageiros embarcados em 2018, 5 milhões viajaram para destinos dentro do Brasil, 77,1% (Associação Brasileira das Operadoras de Turismo, 2019), sendo que a maioria maciça destes embarques se originam em cidades mais estruturadas e populosa. Tal dado permite concluir que as pessoas das cidades com maior número de habitantes, quando optam por não cumprirem o isolamento social e ainda decidem se locomover, apresentam maior risco e probabilidade de contaminar pessoas durante o trânsito e até mesmo na cidade-destino, uma vez que comprovada a relação entre as cidades mais populosas e o contágio pessoa-pessoa.

Seria interessante, como perspectiva futura, aprofundar os estudos a fim de desenvolver um sistema de equações que indicasse a existência ou não um fator proporcional de relação entre população vs a taxa de letalidade, tentando correlacionar em quantas vezes a letalidade de uma cidade mais populosa é maior que a de uma cidade menos populosa. Com o intuito de mostrar que a medida que você dobra ou triplica o número de habitantes de uma cidade, a letalidade, não necessariamente dobra ou triplica, indicando a não existência de linearidade nos casos.

Por fim, é imprescindível discutir o fato de que a contaminação pelo novo corona vírus não se dá pelo simples fato de se residir em uma cidade mais ou menos populosa, é importante ressaltar que existem diferentes condições, tanto nos municípios quanto na estância socioeconômica da população residente de cada cidade e que esses fatores influenciam na dispersão, não somente do vírus SARS-CoV-2, mas de muitos outros agentes patógenos.

Deve-se ressaltar ainda que o isolamento social demonstrou ser uma medida de diminuição de contaminação pessoa-pessoa nos municípios mais populosos, o que faz com que esta medida deva sim ser aplicada e incentivada à população como uma opção no combate à epidemia.

Por sua vez, nos munícipios menores, não foi possível concluir eficazmente a importância do isolamento social, mas deve-se ressaltar que esse não deve ser descartado como medida de combate ao avanço futuro do vírus.

# Acknowledgments:

We would like to thank the Creative Commons Attribution ShareAlike, and Álvaro Justen and collaborators from Brasil.IO for confirmed case data base.

# Bibliografia:

JUSTEN, Álvaro e colaboradores. Boletins informativos e casos do coronavírus por município por dia. Brasil.io, 2020. Disponível em: <<https://brasil.io/dataset/covid19/caso>>. Accessed on: 10, April 2020.