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)

# Abstract:

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.

# INTRODUÇÃO:

Vírus:

Infecção/Epidemia mundo:

Infecção/Epidemia Brasil:

# M&M:

Para a análise exploratória de dados (AED) e 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) que é uma das plataformas mais conhecidas para ciência de dados, oferecendo aos usuários diversas ferramentas de bibliotecas que auxiliaram nossas análises.

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:

1. O número de habitantes tem correlação positiva com a número de casos confirmados.

Uma vez que a COVID-19 é uma patologia causada por coronavírus que é transmitido diretamente de humano para humano, buscamos entender se as características dos munícipios atingidos podem auxiliar no entendimento da pandemia. O banco de dados analisado conta com 542 municípios dos 26 estados da federação. Buscando entender a distribuição dos valores referentes à população (número de habitantes), casos confirmados e mortes confirmadas pela COVID-19, analisamos os histogramas das amostras (Figura 1). Observamos uma distribuição log-normal dos dados, o que indica a necessidade do uso de teste de correlação Spearman, buscando entender a relação entre as diferentes características.

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

A correlação de Spearman nos valores de população, casos confirmados, mortes, casos confirmados para 100k habitantes e razão de casos confirmados e mortes, foi testada para todas as cidades (Figura 2A) ou cidades com menos de 6x106 habitantes (Figura 2B). Como esperado, observamos uma forte correlação negativa entre razão de casos confirmados e mortes e casos confirmados (R = -0.93), população (R = -0.77) ou mortes (R = -0.39), em ambas as condições. Foi observado valores de correlação positiva média entre as mortes e os casos confirmados em ambas as condições (R = 0.45), assim como os casos confirmados para 100k habitantes e os casos confirmados (R = 0.39). Adicionalmente, detectamos uma correlação positiva média (R = 0.63) entre os casos confirmados e a população de habitantes, o que se mantem ao retirar os municípios mais populosos (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). |

Os resultados mostram uma correlação entre número de habitantes na população e número de casos confirmados, seja em toda a amostra, seja nas cidades com menos de 6x106 habitantes. Buscando melhor entender esta correlação, os casos confirmados foram analisados segundo a amostra da população, dividida segundo seus quartis, com 5 grupos de 1149 a 22.809 (Figura 3A), 22.809 a 56.250 (Figura 3B), 56.250 a 132.079 (Figura 3C), 132.079 a 295.985 (Figura 3D) e 295.985 a 12.252.023 (Figura 3E) habitantes. Foi observado que para os 2 primeiros quartis, a correlação é neutra não significativa, com valor de R = 0.0373 e 0.04 e *p* = 0.5857 e 0;5579, respectivamente ao primeiro e ao segundo grupo. Já nos quartis intermediários, corre uma correlação franca positiva, com R = 0.2468 e 0.1897 e *p* = 0.0002 e 0.0364. O último quartil, contendo os municípios com mais de 295.985 habitantes, mostrando uma forte correlação positiva com R = 0.6877 e *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 |

A mesma estratificação de municípios pelo número de habitantes foi realizada para a análise de correlação entre população e mortes por COVID-19 (Figura 4). Foi observado que o primeiro, segundo e quarto estrato apresentou correlação neutra não significativa, com R = 0.0251, 0.0451 e 0.0994 e *p* = 0.7141, 0.5088 e 0.2762. Não foi possível calcular métricas para o terceiro estrato. Já o último quartil, com os municípios que possuem mais de 295.985, apresentou uma forte correlação positiva com R = 0.6813 e *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. |
|  |

Os dados municipais referentes à número de habitantes na população, casos confirmados, mortes, casos confirmados para 100k habitantes e razão de casos confirmados e mortes, permite observar uma correlação positiva entre tamanho da população em habitantes e casos confirmados. Esta correlação é significativa apenas em municípios com mais de 295.985 de habitantes. Isso sugere que outras características destas localidades possam ser correlacionadas com a patologia.

1. O Índice de Desenvolvimento Humano Municipal (MHDI) e a densidade demográfica tem correlação com o número de casos confirmados de COVID-19

É conhecido que diferentes condições presentes nos municípios podem influenciar na dispersão da epidemia de distintos patógenos. Os resultados aqui obtidos mostram uma correlação positiva entre população e casos confirmados em municípios com mais de 295.985 habitantes. Buscando melhor entender os fatores que possam explicar esta correlação e, tendo em vista que a COVID-19 é transmitida de pessoa para pessoa, foram analisadas diferentes características dos 542 municípios, com novos dados populacionais agora obtidos no Instituto Brasileiro de Geografia e Estatística (IBGE). Para tanto, ensaios de correlação Spearman foram performados com o número de casos confirmados (Confirmados), mortes, confirmados para 100k habitantes (Confirmados/100k), razão de casos confirmados e mortes e casos confirmados (Razão de Morte), área oficial do município em km² (Área), habitantes em dados de densidade populacional (População Dens), densidade populacional (Densidade), e Índice de Desenvolvimento Humano Municipal (MHDI). Os dados totais (Figura 5A) foram estratificados em municípios com menos (Figura 5B) ou mais (Figura 5C) de 295.985 habitantes.

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

Os resultados confirmam as análises realizadas anteriormente em relação a correlação negativa entre razão de casos confirmados e mortes e casos confirmados (R = -0.88), população (R = -0.74) ou mortes (R = -0.27). Foi confirmado também os valores de correlação positiva média entre as mortes e os casos confirmados (R = 0.43), assim como os casos confirmados para 100k habitantes e os casos confirmados (R = 0.46) e a correlação positiva média entre os casos confirmados e a população de habitantes da base anterior (População COVID - R = 0.71) e a base IBGE (População Dens - R = 0.7). Foi observado que, quando analisados os municípios com menos de 300.000 habitantes, a correlação negativa entre razão de casos confirmados e mortes e casos confirmados (R = -0.93), população (R = -0.69) ou mortes (R = -0.13), se mantem. Porém, nas cidades com população maior de 300.000 habitantes, observamos uma variação com correlação negativa entre razão de casos confirmados e mortes e casos confirmados inferior (R = -0.52), correlação nula para população (R = -0.072) e positiva para mortes (R = 0.21). Isso caracteriza como uma inversão das condições de casos confirmados para 100k habitantes nas cidades mais populosas analisadas. Adicionalmente, para a correlação positiva média entre os casos confirmados e a população de habitantes da base anterior (População COVID - R = 0.57) e a base IBGE (População Dens - R = 0.55) nos municípios com menos de 300.000 habitantes e, (População COVID - R = 0.76) e a base IBGE (População Dens - R = 0.76) nos municípios com mais de 300.000 habitantes. Novamente, observamos diferenças entre as localidades com maior ou menor população, reforçando a hipótese de que características municipais devem ser observadas para o entendimento da pandemia.

Neste contesto, analisamos três características dos municípios que são a área oficial do município em km² (Área), densidade populacional (Densidade), e Índice de Desenvolvimento Humano Municipal (MHDI). Nos dados totais, casos confirmados tem correlação nula com a Área (R = -0.07) e média com Densidade (R = 0.56) e MHDI (R = 0.43). Isso se mantem no grupo de municípios com população menor que 300.00 habitantes, com correlação fraca negativa com a Área (R = -0.12) e média positiva com Densidade (R = 0.45) e MHDI (R = 0.34). Já nas cidades com mais de 300.00 habitantes, foi possível observar correlação nula com a Área (R = -0.0093) e média com Densidade (R = 0.37) e MHDI (R = 0.43). Portanto, assim, como a relação do número de habitantes, a densidade populacional e o MHDI, também tem relação positiva com o número de casos confirmados de COVID-19.

Buscando melhor entender estas correlações, realizamos nova análise de Spearman, agora apenas com os dados de MHDI (Figura 6) e Densidade demográfica (Figura 7). Para MHDI, foi possível mostrar uma correlação positiva média com os casos confirmados, com R = 0.3396 e *p* < 0.0001, nas cidades com menos habitantes e R = 0.4293 e *p* < 0.0001.

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

Já para densidade demográfica, foi observada uma correlação positiva média com os casos confirmados, com R = 0.4474 e *p* < 0.0001, nas cidades com menos habitantes e R = 0.369 e *p* < 0.0001.

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

Assim, podemos concluir que a correlação positiva entre cidades com mais de 295.985 habitantes e o número de casos confirmados (Figura 3) ou mortes (Figura 4) é corroborado pela correlação positiva com a densidade demográfica (Figura 7) e o MHDI (Figura 6). Porém, nas cidades com menos de 295.985 habitantes, que não apresentam correlação com o número de casos confirmados, apresentam correlação positiva com a densidade demográfica (Figura 7) e o MHDI (Figura 6). Isso permite sugerir que a densidade demográfica e o MHDI são fatores que devem ser considerados como fatores para o entendimento da epidemia em todos os 594 municípios estudados. Ainda, não foi observada diferença entre os dois índices, o que sugere que, ao menos quanto ao espalhamento viral (casos confirmados), municípios com população mais densa e com melhor condição socioeconômica, independentemente do tamanho da população, apresentam maior risco.

1. A série temporal dos confirmados permite prever de 56,829 a 70,447 de casos em 30 dias.

Estas informações sugerem que características municipais devem ser consideradas quanto a condição epidemiológica atual. Porém, buscando entender a condição da infecção em nível nacional, analisamos a série temporal de casos acumulados desde 25/02/2020 a 10/04/2020 (Figura 8).

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

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

# 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:

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>>. Acesso em: 04, abril de 2020.