**Casos de COVID-19 em municípios brasileiros, eleição presidencial de 2018 e nível de desigualdade socioeconômico**

Alison Rolemberg de Lima1\*;Elton Gean Araújo2

1Analista de Negócios. Av. Dr. Chucri Zaidan, 1240 – Chácara Santo Antônio; 04711-130 São Paulo, São Paulo, Brasil

2 Universidade Federal de Mato Grosso do Sul, Campus de Paranaíba. Av. Pedro Pedrossian, 725 – Universitário; 79500-000 Paranaíba, Mato Grosso do Sul, Brasil

\*autor correspondente: alisonlima1902@gmail.com

**Casos de COVID-19 em municípios brasileiros, eleição presidencial de 2018 e nível de desigualdade socioeconômico**

**Resumo**

Nos últimos anos o mundo enfrentou uma grande pandemia causada pela COVID-19, infectando milhares de pessoas. Durante esse tempo, diversos estudos buscaram entender a heterogeneidade de casos da doença no Brasil, baseado não apenas nas características do sistema de saúde de uma determinada região, mas também de aspectos sociais e políticos. Ao longo do primeiro ano da pandemia, estudos mostraram que o nível de desigualdade socioeconômica e o candidato eleito na última eleição presidencial tem relação com a quantidade de infectados pela doença. Assim, esse trabalho avaliou, por meio de análise exploratória, se esse comportamento persistiu no final de 2021, frente as mudanças que a pandemia demonstrou desde o seu início. Em seguida, através de técnicas de agrupamento, verificou a possibilidade de separar as cidades brasileiras em grupos distintos entre si e observar o perfil de cada um. Após analisar os dados, demonstrou que a relação entre o número de casos de COVID-19 por 100 mil habitantes com o índice de Gini e a porcentagem do presidente eleito nos municípios brasileiros diminuiu comparado com os primeiros estudos, porém foi possível dividir as cidades em quatro clusters, sendo que em cada grupo a média da porcentagem para cada candidato, o nível de desigualdade e a quantidade de contaminados por COVID-19 se mostraram diferentes entre si.

**Palavras-chave:** COVID; coronavírus; eleições; Gini; agrupamento.

**COVID-19 cases in Brazilian municipalities, 2018 presidential election and level of socioeconomic inequality**

**Abstract**

In recent years the world has faced a major pandemic caused by COVID-19, infecting thousands of people. During this time, several studies sought to understand the heterogeneity of cases of the disease in Brazil, based not only on the characteristics of the health system in each region, but also on social and political aspects. Over the first year of the pandemic, studies have shown that the level of socioeconomic inequality and the candidate elected in the last presidential election is related to the number of people infected with the disease. Thus, this work evaluated, through exploratory analysis, whether this behavior persisted at the end of 2021, given the changes that the pandemic has shown since its beginning. Then, through grouping techniques, he verified the possibility of separating Brazilian cities into distinct groups and verifying the profile of each one. After analyzing the data, it showed that the relationship between the number of COVID-19 cases per 100,000 inhabitants with the Gini index and the percentage of the president-elect in Brazilian municipalities decreased, compared to the first studies, but it was possible to divide the cities into four clusters, and in each group the average percentage for each candidate, the level of inequality and the number of people infected by COVID-19 were different from each other.

**Keywords:** COVID; coronavirus; election; Gini; clustering.

**Introdução**

No início de 2020 a humanidade passou a enfrentar um dos maiores desafios de saúde pública já registrado, através do surgimento de uma doença causada por um novo tipo de coronavirus, chamado COVID-19 (do inglês “coronavirus disease 2019”). Essa doença, que teve os seus primeiros relatos na China no fim de 2019, tomou proporções inimagináveis se alastrando pelo mundo inteiro e infectando milhões de pessoas. Em out. 2021, por exemplo, o número de casos positivos pelo COVID-19 chegou à marca de mais de 240 milhões e quase 5 milhões de mortes. Os locais com maior número de casos em termos absolutos estão concentrados principalmente em regiões e países com grande população, como os Estados Unidos com mais de 38 milhões de casos e a Índia com mais de 34 milhões. (Demenech et al., 2020; Our World in Data, 2021).

Entretanto, ao analisar em números relativos, percebe-se uma heterogeneidade maior em quantidade de casos confirmados, ou seja, países poucos populosos como Montenegro, com 230 mil casos por milhão, lideram junto com países com grandes populações, como Estados Unidos com 138 mil e Reino Unido com 134 mil casos por milhão. Isso ocorre não apenas entre países, mas também dentro de um mesmo território. No Brasil, por exemplo, em termos absolutos as cidades com maior número de casos são as que possuem grande população, como São Paulo com mais de 970 mil casos. Porém, proporcionalmente cidades pequenas e grandes ocupam as primeiras posições, como Porteirão em Goiás com 47 mil casos por 100 mil habitantes (Brasil.io, 2021; Our World in Data, 2020).

Desse modo, pode-se observar que a quantidade de pessoas que habitam um país ou uma cidade não é o único fator que pode levar ao aumento de casos de COVID-19 em termos relativos, ou seja, existem outras variáveis que levam a essa diferença no número de casos entre as regiões. Ademais, embora trate-se de uma doença, essas variáveis não são apenas ligadas à saúde, isto é, podem estar associadas a outras áreas, como questões políticas, sociais e econômicas.

Uma delas é relatado em um estudo publicado por Cabral et al. (2021), no qual analisaram que os municípios que elegeram Jair Bolsonaro em 2018 como presidente do Brasil apresentaram maior número de casos de COVID-19 em número relativo comparado com cidades em que ele não foi eleito. Segundo o artigo, isso ocorre pois, visto que o presidente apresenta uma postura negacionista da doença e das medidas de prevenção, como ser contrário a vacinas, uso de máscaras e ao distanciamento social, muitos dos seus apoiadores seguem sua opinião e, assim, se recusam a seguir as medidas que evitam a contaminação por COVID-19 (Cabral et al., 2021).

Outra variável agora de caráter socioeconômica foi analisada em uma pesquisa elaborada por Demenech et al. (2020). De acordo com esse estudo, os estados brasileiros que apresentam maior taxa de desigualdade econômica, comparado pelo índice de Gini, tiveram maior número de infecções e de mortes por COVID-19 proporcionalmente, do que os estados menos desiguais. Nos estados com maiores taxas de incidência da doença, como Amapá com quase 36 mil casos por milhão de habitantes (dados registrados até dia 7 de julho de 2020), está entre os que possuem os maiores índices de Gini (0,61). Já os estados que tiveram as menores taxas de casos, como Paraná (3 mil casos por milhão) é um dos estados com os menores índices de Gini (0,54) (Demenech et al., 2020).

Isso pode ocorrer por efeitos distintos chamados de absoluto e contextual. No efeito absoluto a distribuição desigual de renda afeta diretamente no acesso à saúde, por exemplo entre a população mais pobre a grande maioria não tem plano de saúde e o número de leitos na Unidade de Tratamento Intensico [UTI] é quase cinco vezes menor para usuários do Sistema Único de Saúde, comparado com aqueles que possuem acesso à rede privada de saúde (Kawachi e Subramanian, 2014).

Em relação ao efeito contextual, por sua vez, evidencia que em localidades desiguais, piores são as estruturas públicas de saúde, segurança, saneamento e urbanismo, condições essas que degradam a qualidade de vida de todos, mas que impactam de forma mais severa os menos favorecidos. Assim, pessoas que estão em condições de desigualdade socioeconômica maior, estão mais expostas ao vírus COVID-19, visto que a maioria vive em habitações de pior qualidade e com maior número de moradores, necessita utilizar meios de transporte público com maior aglomeração, apresenta uma qualidade alimentar nutricional pior, é mais suscetível ao estresse psicológico e entre outros aspectos (Kawachi e Subramanian, 2014).

Ademais, é importante destacar que o trabalho elaborado por Demenech et al. (2020), teve como metodologia, além da análise correlacional das variáveis, o uso de modelos estatísticos, como regressão linear simples e regressão espacial (modelo de defasagem). Isso demonstra a vantagem de utilizar esses modelos em estudos que relacionam os casos de COVID-19, com outros fatores não diretamente ligados à própria doença.

Portanto, o trabalho de conclusão de curso uniu essas duas variáveis de caráter político e socioeconômico, ou seja, presidente eleito na última eleição e taxa de desigualdade nos municípios brasileiros, e verificou se há associação em termos relativos com o número de casos de COVID-19. Ademais, esse projeto buscou complementar os trabalhos citados anteriormente, visto que, além de colocar as duas variáveis em uma mesma análise, utilizou dados municipais e fez uma análise não somente descritiva, como médias e correlações, mas também trabalhou com modelagem de clusterização, visando observar se é possível formar grupos com características distintas entre si.

**Material e Métodos**

Primeiramente, para a coleta dos dados de casos de COVID-19, presidente eleito na eleição de 2018 em cada município e dados de desigualdade socioeconômica, foram obtidos em diferentes fontes. Em relação à primeira variável, obteve-se no repositório do Brasil.io (2021), em que, dentre os diversos tipos de dados sobre a pandemia no Brasil presente no site, teve como foco o número de casos de COVID-19 por 100 mil habitantes nos municípios brasileiros, acumulados até o dia 30 nov. 2021.

Em seguida, para os dados de eleição presidencial de 2018, foram extraídos através do repositório do Tribunal Superior Eleitoral (2018). Nessa base observou-se o candidato vencedor no segundo turno e a porcentagem de votos do presidente eleito para cada município, de modo a verificar se as cidades que tiveram uma diferença grande entre os candidatos apresentaram mais casos da doença proporcionalmente, comparado com outras cidades em que a diferença foi menos expressiva.

A terceira característica a ser analisada foi o nível de desigualdade de um município. Para isto, utilizou-se o Índice de Gini da renda domiciliar per capta de cada cidade brasileira, obtido no censo realizado em 2010 pelo Instituto Brasileiro de Geografia e Estatística. Esse indicador varia de zero a um e quanto mais alto o seu valor maior é o grau de concentração de renda dos municípios.

Após coleta e consolidação dos dados, passou-se para a etapa de análise dos dados, tendo como principal ferramenta o software R. A primeira atividade foi realizar o cruzamento das bases, a fim de identificar as cidades presentes nas três bases, e fazer o tratamento e limpeza dos dados, como filtros, renomeação das variáveis e presença de valores ausentes. Para isso, utilizou-se as funções presentes no pacote “dplyr”.

Em seguida, foi realizado uma análise exploratória dos dados, de modo a conhecer algumas características das variáveis. Dentre vários indicadores observou-se médias, medianas, desvio padrão, percentis, presença de outliers, correlação, skewness e curtose (funções presentes nos pacotes “stats” e “moments”). Além disso, trabalhou-se com regressão logística, do pacote “stats”, a fim de verificar a chance de ser uma cidade que elegeu Jair Bolsonaro conforme o aumento do número de casos de COVID-19.

Outro procedimento realizado durante a análise exploratória foi dividir a base em duas, sendo uma contendo os municípios que elegeram o candidato Fernando Haddad e outra com as cidades onde Jair Bolsonaro venceu a eleição. O objetivo foi verificar, por meio de gráficos de dispersão e de indicador de correlação, se as variáveis apresentavam maior associação entre si e se as bases se diferenciavam significativamente.

Por fim, foi utilizado os dados dos municípios brasileiros para formar grupos distintos entre si, por meio de modelagem não supervisionada. Para isso, trabalhou-se com a métrica de distância de Gower, em que mede a dissimilaridade de dois itens que apresentam variáveis numéricas e não numéricas. Após ter construído as medidas das distâncias, utilizou-se o algoritmo PAM (Partitioning Around Medoids), pois além de ser capaz de utilizar as medidas da métrica de Gower, ele se assemelha ao algoritmo K-means, porém com algumas diferenças. Em vez de buscar os centroides dos grupos, o algoritmo PAM procura os medoides dos clusters, ou seja, as medianas dos grupos (Kaufman e Rousseeuw, 1990). As funções da distância do Gower e do algoritmo PAM estão presentes no pacote “cluster” do R.

**Resultados e Discussão**

**Análise exploratória**

Para iniciar o processo de análise, primeiramente foi realizado a importação de três bases de dados contendo as variáveis necessárias ao trabalho. Em seguida, ocorreu o tratamento dos dados, por meio de filtros, junção das informações das três bases e seleção das principais variáveis. Dessa forma, gerou-se uma base contendo 5565 municípios e 4 variáveis: número de casos de COVID-19 por 100 mil habitantes, Índice de Gini do município, candidato a presidente vencedor no segundo turno e porcentagem de votos que esse candidato recebeu.

O próximo passo foi verificar o comportamento das variáveis, tanto individualmente quanto entre si, através de algumas técnicas estatísticas como análise da média, mediana, quartis, curtose, assimetria, correlações e análise gráfica. O primeiro aspecto observado foi a distribuição dos dados da variável número de casos de COVID-19 por 100 mil habitantes, na qual apresentou uma distribuição assimétrica à direita, dado que a média se mostrou maior que a mediana (10158 casos contra 9737 casos, respectivamente). Para confirmar esse comportamento, foi calculado o coeficiente de assimetria, utilizando o pacote “moments” e, a partir do resultado gerado, concluiu-se que a variável de fato possui assimetria positiva, pois o valor do coeficiente foi de 0,796.

Além disso, outra medida utilizada para estudar o comportamento dessa variável foi a curtose. Com base também do pacote “moments”, verificou-se que a distribuição dos dados apresentou uma frequência mais alongada que a curva normal, dado que o valor da curtose foi de 5,12. No histograma abaixo (Figura 1) é possível verificar também essa assimetria de forma visual.

Outro ponto levantado sobre a variável de casos de COVID-19 foi durante a análise de seus quartis. Conforme pode ser visto no gráfico Box-Plot (Figura 1), foi verificado a presença de outliers, mais especificamente de dados maiores que uma vez e meia a amplitude interquartil. Isso indica, dentre várias interpretações, que houve municípios impactados significativamente com a pandemia, comparado com a maioria das outras cidades, como é o caso de Severiano Melo-RN com quase 39 mil casos por 100 mil habitantes e Santa Cecília do Sul-RS com 35 mil casos. Além disso, isso explica, em certa medida, o comportamento da distribuição assimétrica positiva da variável.

Chart, histogram

Description automatically generated

Chart

Description automatically generated with medium confidence

Figura 1. Histograma e Box-Plot do número de casos de COVID-19 por 100 mil habitantes, indicando o comportamento assimétrico da variável e a presença de outliers

Fonte: Brasil.io (2021)

Além disso, outra variável que apresentou um comportamento em destaque foi o índice de Gini. Ao analisar a presença de outliers, verificou-se a ocorrência de dados maiores e menores que uma vez e meia o intervalo interquartil, como poder ser visto em seu gráfico de Box-Plot (Figura 2). Isso é um indicativo da diferença socioeconômica entre regiões do Brasil, ou seja, apesar de estarem em um mesmo país é possível encontrar regiões com níveis de desigualdade social muito diferentes entre si.

Chart, box and whisker chart

Description automatically generated

Figura 2. Box-Plot do Índice de Gini demonstra a presença de outliers, ou seja, há regiões com níveis de desigualdade social muito diferentes entre si.

Fonte: IBGE (2010)

Em relação ao modo como as variáveis se relacionaram entre si, um dos primeiros pontos observados foi através da correlação de Spearman entre as variáveis numéricas. Vale ressaltar que foi utilizado esse tipo de correlação, pois se adere melhor a variáveis que não apresentam uma distribuição normal. Como pode ser visto no gráfico abaixo (Figura 3), verificou-se uma fraca correlação entre elas, uma vez que seus valores estão próximos de 0 (Pontes, 2010).

Chart, waterfall chart

Description automatically generated

Figura 3. Correlações das variáveis são baixas, visto que estão próximos de zero.

Fonte: Resultados originais da pesquisa

Ademais, no caso da relação entre a variável número de casos de COVID-19 por 100 mil habitantes e o índice de Gini, o *rho* foi apenas de -0.22 (com um p-valor igual a 2,2\*10-16). Essa observação foi divergente ao apontado no estudo Demenech et al. (2020), visto que a correlação entre elas foi forte e positiva (*rho* = 0,69). Uma das possíveis explicações dessa diferença é que quando foi realizado a extração dos dados para sua pesquisa (julho de 2020), a pandemia estava em seu início e afetando as regiões mais desiguais. No entanto, no período em que a pandemia foi analisada para esse trabalho (novembro de 2021), a doença já havia afetado fortemente todas as regiões do país, independentemente do seu índice de Gini, devido, entre vários motivos, ao surgimento de variantes mais transmissíveis e à flexibilização das medidas de isolamento. Outro ponto importante é que no estudo de Demenech et al. (2020) foi analisado os dados dos estados federativos, diferentemente desse trabalho que focou nos municípios brasileiros, o que pôde também ter gerado resultados divergentes.

Em seguida, buscou-se analisar o comportamento da variável categórica “Candidato\_Eleito” em relação às demais. Um dos primeiros passos foi verificar a mediana das variáveis para cada candidato e confirmar se são estatisticamente diferentes entre si. Para a variável índice de Gini, observou-se que o grupo das cidades que elegeram Jair Bolsonaro apresentou mediana inferior ao grupo que Fernando Haddad venceu (0,47 contra 0,52, respectivamente). Posteriormente, a fim de garantir que essas medianas não são iguais, foi utilizado o Teste Mann-Whitney e, a partir dele, verificou-se que as medidas são estatisticamente diferentes entre si, dado que o p-valor do teste foi menor que 0,05. Assim, as cidades que Haddad foi eleito tiveram um índice de Gini maior e estatisticamente diferentes das que elegeram Bolsonaro.

Por sua vez, em relação à variável número de casos de COVID-19 por 100 mil habitantes, observou-se que a mediana do grupo de Jair Bolsonaro foi superior ao grupo de Fernando Haddad (11770 versus 7606 casos, respectivamente). O Teste Mann-Whitney, indicou também que ambas as medidas são diferentes entre si, dado que o p-valor foi menor que 0,05. Isso mostrou que, com base nas medianas, as cidades que apoiaram Jair Bolsonaro foram mais afetadas pela pandemia, comparado com as que elegeram Fernando Haddad.

Contudo, foi realizado outra análise para verificar se é possível afirmar que quanto maior o número de casos de COVID-19, maior é a chance de ser uma cidade que elegeu Jair Bolsonaro. Para isso, foi elaborada uma regressão logística utilizando a variável categórica “Candidato\_Eleito” como variável dependente e o número de casos de COVID-19 por 100 mil habitantes e índice de Gini como variáveis independentes. Após esse procedimento, foi observado a cada aumento de uma unidade no número de casos, espera-se um aumento de 0,0002% na chance de Bolsonaro ter vencido. Em outras palavras, não é possível garantir que quanto maior o número de casos, maior a probabilidade de ser uma cidade que apoiou Jair Bolsonaro, mesmo apresentando uma mediana de casos de COVID-19 superior ao de Fernando Haddad.

O próximo passo realizado foi analisar mais detalhadamente o comportamento das variáveis para cada candidato e verificar se há uma diferença relevante. Para isso, dividiu-se a base de dados em duas, sendo uma contendo só os municípios que elegeram Bolsonaro e outra com os que tiveram Haddad vencedor. O primeiro procedimento realizado, após essa separação, foi visualizar os dados através de gráficos de dispersão, comparando o número de casos de COVID-19 proporcionalmente com os índices de Gini e a porcentagens dos candidatos eleitos.

De início, observou-se que na primeira comparação (número de casos de COVID-19 e índice de Gini) os dados se comportaram de modo muito semelhante entre as bases de cada candidato e com o aumento do índice não houve necessariamente um aumento de infectados por COVID-19, como pode ser visto na Figura 4. Já ao comparar casos da doença com a porcentagem dos políticos, verificou-se que também não houve uma grande diferença entre eles e o número de casos não acompanhou o aumento da porcentagem, apenas apresentou uma leve inclinação no gráfico referente ao Jair Bolsonaro (Figura 5).

Dessa forma, foi analisado que, pelo menos visualizando os gráficos, não foi possível afirmar que houve uma relação significativa de casos de COVID-19 com as outras variáveis, para nenhuma das bases. Além disso, verificou-se que as bases se comportaram de maneira muito parecida entre si, ou seja, dividir a base por candidato não indicou uma diferença relevante no comportamento dos dados.

Chart, scatter chart

Description automatically generated

Figura 4. Comparação índice de Gini versus casos de COVID-19 entre os candidatos - comportamento muito semelhante entre as bases de cada candidato

Fonte: Resultados originais da pesquisa

Chart, scatter chart

Description automatically generated

Figura 5. Comparação porcentagem de votos versus casos de COVID-19 entre os candidatos - o número de casos não acompanhou significativamente o aumento da porcentagem de votos

Fonte: Resultados originais da pesquisa

Para confirmar essa análise, foi verificado a correlação das variáveis, para cada base dos candidatos. Como pode ser visto nos gráficos abaixo (Figura 6), percebeu-se que as correlações entre as variáveis são fracas e que os valores são muito parecidos entre as bases. Com isso, contribuiu com o pensamento de que o número de casos de COVID-19 em termos relativos não indicou uma associação significativa com o índice de Gini ou com a porcentagem da eleição, independente do candidato.

Chart, waterfall chart

Description automatically generated

Chart, waterfall chart

Description automatically generated

Figura 6. Correlação das variáveis dos municípios que elegeram Jair Bolsonaro e Fernando Haddad são fracas e que os valores são muito parecidos entre as bases.

Fonte: Resultados originais da pesquisa

Esses últimos resultados divergem, em certa medida, com o que foi apresentado por Cabral et al. (2021), visto que, segundo sua pesquisa, quanto maior a porcentagem de votos ao Jair Bolsonaro, maior o número de casos de COVID-19 proporcionalmente. Em termos numéricos em sua análise, a cada dez por cento de votos para Bolsonaro, há uma adição de 129 casos novos de COVID-19. Uma possível explicação dessa diferença seria também em relação à data de análise, dado que no estudo de Cabral et al. (2021), a extração dos dados foi entre fev. 2020 a fev. 2021, período em que ainda encontrava regiões mais afetadas que outras. Já no período desse trabalho, com novas variantes e flexibilização das medidas de isolamento em todo o país, essa relação enfraqueceu.

Ademais, essa separação das bases reforçou a divergência de resultados em relação ao estudo Demenech et al. (2020). Até o momento, não foi encontrado uma relação indicando que quanto pior o índice de Gini de uma região (mais próximo de 1), maior foi o número de casos de COVID-19, mesmo separando a base entre os candidatos.

**Análise não supervisionada**

O segundo processo analítico foi verificar a possibilidade de separar os municípios brasileiros em grupos que são significativamente divergentes entre si e observar as suas características. Para isso, foram utilizadas as mesmas variáveis da análise exploratória, ou seja, número de casos de COVID-19 por 100 mil habitantes, índice de Gini das cidades brasileiras, candidato à presidente eleito nos municípios e a porcentagem de votos do candidato vencedor. Ademais, o método utilizado para medir as distâncias das observações foi a distância de Gower (por meio do pacote “cluster” do R), pois ela é capaz de medir a dissimilaridade de dois itens que apresentam variáveis numéricas e não numéricas (Kaufman e Rousseeuw, 1990).

Em seguida, para separar os dados em grupos foi utilizado o algoritmo Partitioning Around Medoids [PAM]. Esse algoritmo além de ser capaz de trabalhar com matriz de dissimilaridade, ele tem como procedimento buscar os medoides dos grupos, ou seja, as medianas dos clusters, algo um pouco diferente do algoritmo K-means, que procura os centroides (Kaufman e Rousseeuw, 1990).

Após o uso dessas técnicas, a próxima etapa foi selecionar o número de grupos que seriam divididas as cidades e para auxiliar na escolha foi utilizado o método da Silhueta. Assim, a partir dos resultados apresentados, a quantidade de grupos que apresentou a maior Silhueta foi igual à 2 (com um coeficiente igual a 0,81). No entanto, notou-se que os dois grupos gerados seriam basicamente a divisão entre os municípios que elegeram cada candidato, algo que já foi realizado na análise exploratória. Portanto, foi escolhido separar as cidades em 4 clusters, pois agregaria mais na análise dos dados e os grupos ainda se apresentaram distintos entre si, visto que o coeficiente de Silhueta foi igual à 0,4. Por meio do gráfico abaixo (Figura 7) é possível visualizar de modo bidimensional a divisão dos clusters.

Chart, scatter chart

Description automatically generated

Figura 7. Visualização bidimensional dos grupos gerados.

Fonte: Resultados originais da pesquisa

Desse modo, observou-se que o grupo 1 foi composto por cidades que elegeram Jair Bolsonaro (com 1321 itens) e teve uma média de 73,82% de votos. Ademais, possuiu na média o menor nível de desigualdade entre os grupos, com um índice de Gini de 0,46 e o que obteve o maior número de casos de COVID-19 por 100 mil habitantes, com 13280 infecções.

O segundo cluster também foi formado por municípios que elegeram Bolsonaro, com um total de 1434 membros, porém apresentou uma média menor de votos (58,72%). Apresentou um índice de Gini um pouco mais elevado que o primeiro grupo, com uma média de 0,49, e menor número de contaminados por COVID-19 (11070 casos).

O grupo 3, por sua vez, foi composto por cidades que elegeram Fernando Haddad, contendo 1154 itens, e com uma média de votos de 60%. Além disso, possuiu um nível de desigualdade maior que os dois primeiros grupos, com um índice de Gini de 0,51, e uma média de casos de COVID-19 igual a 9353.

Por fim, o quarto cluster também constou apenas municípios que elegeram Haddad, com 1656 membros, porém com uma porcentagem maior de votos na média (81%). Apresentou o maior índice de Gini entre os grupos (0,54), sendo, portanto, o grupo com maior desigualdade socioeconômica, e teve a menor média de número de casos de COVID-19 em termos relativos, com 7438 contaminados por 100 mil habitantes.

Outro modo de analisar os grupos foi visualizando como eles se distribuíram geograficamente. Como pode ser visto no mapa abaixo (Figura 8), os grupos 1 e 2 estão localizados principalmente nas regiões Sul, Sudeste, Centro-Oeste e parte do Norte. Enquanto isso, os clusters 3 e 4 estão nas regiões Nordeste (principalmente o grupo 1 no interior do Nordeste) e parte do Norte. Ademais, percebeu-se que um determinado núcleo de cidades está relacionado de certa forma com os municípios ao redor, visto que as regiões onde se concentraram os grupos 1 e 4 apresentaram entorno delas cidades que fazem parte dos grupos com características mais semelhantes entre si, que no caso seriam os grupos 2 e 3, respectivamente (Tobler, 1970).

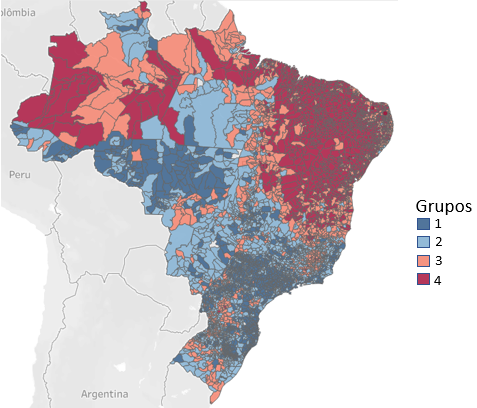


Figura 8. Visualização geográfica dos grupos gerados. Os grupos 1 e 2 estão localizados principalmente nas regiões Sul, Sudeste, Centro-Oeste e parte do Norte e os clusters 3 e 4 estão nas regiões Nordeste e parte do Norte.

Fonte: Resultados originais da pesquisa

Portanto, observou-se que os grupos que elegeram Bolsonaro estão em regiões com menor nível de desigualdade e que apresentaram mais casos de COVID-19, comparado com os outros grupos, e o primeiro cluster, que teve em média a maior porcentagem de votos ao atual presidente, foi o que apresentou a menor taxa de índice de Gini e o maior número de casos da doença em termo relativos. Já os grupos que escolheram Fernando Haddad, tiveram um comportamento contrário, ou seja, estão em regiões com piores níveis de desigualdade e com menor taxa de infectados por COVID-19, sendo que o grupo 4, que teve em média maior porcentagem a favor do petista, apresentou o maior índice de Gini e o menor número de contaminados proporcionalmente.

**Considerações Finais**

O trabalho de conclusão obteve pontos relevantes sobre o comportamento da pandemia nos municípios brasileiros, baseado em certas características sociais e políticas. Após uma análise exploratória nos dados e testes estatísticos, observou-se que as cidades que elegeram Fernando Haddad tiveram uma mediana menor de casos de COVID-19 por 100 mil habitantes e maior desigualdade socioeconômica, comparado com as que Jair Bolsonaro venceu.

No entanto, ao verificar a correlação das variáveis, não foi possível afirmar que houve uma relação (positiva ou negativa) significativa entre o número infectados por COVID-19 e o índice de Gini ou a porcentagem dos candidatos eleitos, diferentemente do que foi apontado em alguns estudos. Uma das possíveis justificativas seria o período de análise dos casos de COVID-19, pois, enquanto as pesquisas levantadas observaram a pandemia em seu primeiro ano, no período estudado desse trabalho a pandemia já havia atingido um alto número de pessoas e estava presente em quase todas as regiões do país, devido à presença de novas variantes mais transmissíveis e da flexibilização das medidas de isolamento.

Ademais, através de técnicas de clusterização, notou-se que os dois clusters que tiveram apenas cidades que elegeram Jair Bolsonaro, apresentaram em média menor índice de Gini e mais casos de COVID-19, sendo que o grupo que teve maior porcentagem de votos a favor do presidente, obteve o menor nível de desigualdade e o maior número de casos da doença. Já os dois grupos restantes que possuíam apenas municípios em que Fernando Haddad foi o vencedor, tiveram um índice de Gini maior e menos casos de COVID-19, tendo o grupo com maior porcentagem a favor de Haddad o maior nível de desigualdade socioeconômica e o menor número de infectados pela doença.

Portanto, o estudo não encontrou uma relação significativa entre o número de casos de COVID-19 em termos relativos com a porcentagem do candidato eleito e o índice de Gini nas cidades do Brasil, baseado principalmente pelos valores de correlação. No entanto, tanto pela análise exploratória quanto pelo processo de agrupamento, verificou-se que os municípios que elegeram Jair Bolsonaro são em média menos desiguais e foram mais afetados pela pandemia, comparado com as cidades que tiveram maior apoio à Fernando Haddad.

**Agradecimento**

Gostaria de agradecer à minha esposa Jaine, por estar ao meu lado todos os dias e me apoiar nos momentos mais importantes e aos meus pais Marcos e Maria e irmã Brenda por contribuírem de diversas formas no meu desenvolvimento pessoal.

**Referências**

Brasil.io. 2021. Boletins informativos e casos do coronavírus por município por dia. Disponível em <https://brasil.io/dataset/covid19/caso\_full/>. Acesso em: 17 nov. 2021.

Cabral, S.; Ito N. C.; Pomgeluppe L. 2021. The disastrous effects of leaders in denial: evidence from the COVID-19 crisis in Brazil. SSRN 3836147: 1-30.

Dasa. 2021. Dados COVID-19. Disponível em: <https://dadoscoronavirus.dasa.com.br/>. Acesso em: 17 nov. 2021.

Demenech, L. M.; Dumith S. C.; Viera M. E. C. D.; Silva L. N. 2020. Desigualdade econômica e risco de infecção e morte por COVID-19 no Brasil. Revista Brasileira de Epidemiologia 23: 1-12

Fávero, L. P.; Belfiore, P. 2017. Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®. 1ed. Elsevier, Rio de Janeiro, RJ, Brasil.

Gower, J. C. 1971. A general coefficient of similarity and some of its properties*.*Biometrics. 27: 857–874.

Instituto Brasileiro de Geografia e Estatística [IBGE]. 2010. Disponível em: <http://abnet.datasus.gov.br/cgi/ibge/censo/cnv/giniuf.def>. Acesso em: 17 nov. 2021.

Kaufman, L.; Rousseeuw, P.J. 1990. Finding Groups in Data: An Introduction to Cluster Analysis*.* 1 ed. John Wiley & Sons, Nova Iorque, Nova Iorque. Estados Unidos.

Kawachi I, Subramanian S. V. 2014. Income inequality. Social Epidemiology. 126-52.

McCaffrey, J. D. 2020. Example of Calculating the Gower Distance. Disponível em:< https://jamesmccaffrey.wordpress.com/2020/04/21/example-of-calculating-the-gower-distance/>. Acesso em 10 maio 2021.

Our World in Data. 2021. Coronavirus (COVID-19) Cases. Disponível em: <https://ourworldindata.org/covid-cases>. Acesso em: 17 nov. 2021.

Pontes, A. C. F. Ensino da correlação de posto no ensino médio. In: 19º SINAPE, 2010, São Pedro, São Paulo, Brasil. Anais... p. 1-8.

Shendre, S. 2020. Clustering datasets having both numerical and categorical variables. Disponível em:<https://towardsdatascience.com/clustering-datasets-having-both-numerical-and-categorical-variables-ed91cdca0677?gi=8bc0f26e5e67>. Acesso em 10 maio 2021.

Tobler, W. R. 1970. A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography. 46: 234-240.

Tribunal Superior Eleitoral [TSE]. 2018. Disponível em: <https://www.tse.jus.br/eleicoes/eleicoes-2018/votacao-e-resultados/resultados-eleicoes-2018>. Acesso em: 17 nov. 2021.