

COVID-19 e Indicadores Socioeconômicos nos Municípios Paulistas: Uma Análise na Perspectiva da Estatística Espacial

RODRIGO MENDES FERNANDES

ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO (FGV-EAESP)

EDUARDO DE REZENDE FRANCISCO

ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO (FGV-EAESP)

COVID-19 e Indicadores Socioeconômicos nos Municípios Paulistas: Uma Análise na Perspectiva da Estatística Espacial

1. INTRODUÇÃO

O mundo foi extremamente impactado pela pandemia do Novo Coronavírus em 2020. No começo do mencionado ano, noticiava-se apenas que uma onda de pneumonia ou gripe mais forte assolava a China, promovendo um grande número de internações e estressando fortemente o sistema de saúde. Lá foi realizado um severo *lockdown*, ação que consistiu em o governo proibir as pessoas de saírem de casa, com o objetivo de impedir a disseminação do vírus. Aos poucos, mais foi se conhecendo sobre esse vírus, bem como ele foi se alastrando por todo o globo. A China deixou de ser o epicentro da doença e a Europa passou a ser a mais atingida. Depois disso passou a ser os EUA. No Brasil, a doença chegou primeiro na cidade de São Paulo, e aos poucos foi se alastrando por todo o Estado, assim como por todo o país (OMS, 2020).

Por se tratar de uma doença, é de se esperar que a área mais afetada e de destaque seja a saúde. Porém, devido às medidas de distanciamento social, paralisação da economia e demais restrições, o vírus levou a impactos em diversos setores da sociedade, envolvendo um enorme esforço de diversos segmentos do setor público para serem mitigados. Isso destacou ainda mais a enorme desigualdade presente no Brasil. Ao passo que alguns puderam passar a trabalhar de forma remota e sem se expor desnecessariamente ao vírus, outros não puderam fazer o mesmo, como os profissionais de saúde ou de atividades de menor qualificação profissional. Houve ainda quem perdeu o emprego e teve de partir para a informalidade. Isso sem falar dos que já estavam na informalidade antes disso, que tiveram sua renda ainda mais diminuída. Enfim, para além dos mortos para a doença, muitos outros também foram muito atingidos, em especial os mais vulneráveis (CASTRO *et al.*, 2021).

Dentro desse cenário, é crucial que o poder público seja capaz de tomar as melhores decisões relacionadas à COVID-19, para que a doença seja controlada o quanto antes, diminuindo o número de vidas perdidas e o impacto social e econômico resultante. Para isso, é fundamental a utilização de novas tecnologias, *big data*, ferramentas estatísticas e técnicas e conhecimentos utilizados pela academia e setor privado. Faz-se necessária uma grande capacidade gerencial para articular diferentes *stakeholders* e propiciar um ambiente capaz de fazer com que essas tecnologias e conhecimentos sejam efetivas no controle da pandemia. Além disso, o uso de dados e do conhecimento é crucial para a sociedade civil, para que possa atuar em paralelo ao governo nas medidas de controle da doença, além de cobrar as atitudes mais assertivas dos governantes.

Objetivos de pesquisa e hipóteses do modelo

O presente trabalho visa compreender a situação e a evolução da COVID-19 nos municípios paulistas. A proposta é explorar como os casos de COVID-19 estão dispostos pelo Estado e verificar se indicadores socioeconômicos e características populacionais possuem influência no número de casos ou na severidade da doença. Toma-se como hipótese inicial que indicadores socioeconômicos, aqui representados pelas condições de vida de determinada localidade, possuem influência positiva ou negativa no número de casos e/ou severidade da doença. Além disso, outra hipótese central é a de que os casos de COVID-19 não estão dispostos aleatoriamente no território. Dessa forma, pretende-se analisar que fatores socioeconômicos e territoriais têm influência na COVID-19 nos municípios paulistas.

É importante pontuar que a presente pesquisa foi realizada durante a pandemia e é certo que os números aqui apresentados e analisados são diferentes na data da leitura desse artigo. A data de coleta dos dados é o dia 08 de novembro de 2020.

Até o dia 11/11/2020, 1.117.174 casos foram diagnosticados positivamente e 39.311 moradores perderam suas vidas para a doença no Estado de São Paulo (BRASIL.IO, 2020a). Doravante, até o dia 02/06/2021, data da finalização deste artigo, os casos subiram consideravelmente, para 3.314.631, e os óbitos para 112.927 (BRASIL.IO, 2021).

2. REFERENCIAL TEÓRICO

Espaço Geográfico e Epidemiologia

A relação entre epidemiologia e o espaço geográfico não é recente. Um exemplo célebre de pesquisador que vislumbrou essa interseção é o médico inglês John Snow, considerado como um dos fundadores da epidemiologia moderna. Ele foi um dos primeiros a solucionar um problema de saúde pública com recursos geográficos. Em 1854 ocorria uma epidemia de cólera em Londres e não se sabia de que forma ela era transmitida. Snow decidiu colocar em um mapa da cidade pontos que simbolizavam os endereços onde moravam os que haviam morrido pela doença. Após analisar o mapa, pôde-se chegar à conclusão de que as mortes estavam fortemente agrupadas no entorno de bombas de distribuição de água. Após a suspeita de que esse era o vetor de transmissão, bombas foram fechadas e os casos caíram vertiginosamente (MCLEOD, 2000).

Desde então, os conhecimentos sobre epidemiologia, geografia e estatística evoluíram bastante, e surgiram ferramentas, digitais, sofisticadas e robustas. Hoje em dia, é possível alcançar conclusões a partir de dados mais diversos e complexos. Estudos recentes chegaram a conclusão de que o CEP de alguém, isso é, onde essa pessoa mora, é mais relevante para a saúde do que marcadores genéticos (GRAHAM, 2016). Adler e Ostrove (1999), Roux e Mair (2010) e Meijer (2012) também chegaram a resultados que indicam que o espaço geográfico é um forte preditor das condições de saúde de um indivíduo. Nesse sentido, cabe a reflexão de que para além de fatores biológicos, inerentes ao indivíduo, questões construídas socialmente, como o modo de vida, possuem grande impacto na relação saúde-doença. Alimentação, exercícios físicos, exposição a fatores ambientais e saneamento básico, por exemplo, possuem um grande impacto na saúde e afetam as possibilidades de alguém contrair algum tipo de enfermidade.

Esses fatores construídos socialmente possuem, por sua vez, uma forte relação com o espaço geográfico. É no espaço geográfico e no território que as pessoas vivem suas vidas. Que vão ao trabalho, estudam, produzem e consomem. Para o Geógrafo Milton Santos:

“O espaço geográfico é o espaço social onde se dão as relações humanas, é um espaço relacional. Assim sendo, a localização geográfica mostra-se relevantíssima para o entendimento desses fatores que afetam diretamente a saúde das pessoas. Mais especificamente, quando pensamos pelo olhar da epidemiologia, isso é, em doenças infecciosas, o espaço geográfico e suas condições de vida ganham um papel de destaque ainda maior” (SANTOS, 1979)

Essa ideia dialoga com o conceito de se realizar uma “abordagem ecológica”. Nesse tipo de abordagem, leva-se em consideração as características da população como um todo (ou um grupo de pessoas em específico) e do ambiente na qual se inserem (um contexto) para fazer inferências a respeito de um indivíduo em específico que esteja dentro desse mesmo grupo. (MINISTÉRIO DA SAÚDE, 2006). Barrozo *et al.* (2019) mostram um exemplo dessa questão. Segundo eles é raro que prontuários médicos contenham informações a respeito do status socioeconômico ou outras questões que tragam a informação relevantes relacionadas a aspectos de vulnerabilidade que alguém possa estar exposto. Assim sendo, saber a qual população esse indivíduo pertence, onde ele mora ou de onde veio, por exemplo, pode trazer informações relevantes a respeito dessas questões.

Por fim, ainda tratando sobre Epidemiologia e Espaço Geográfico, faz-se mister relacionar a unidade de análise com as análises de abordagem ecológicas, definidas anteriormente. A

unidade de análise é “o nível de agregação que irá se trabalhar o problema apontado de uma investigação” (MINISTÉRIO DA SAÚDE, 2006). Ou seja, é o espaço geográfico no qual está se investigando determinado fenômeno e o qual as variáveis analisadas se relacionam. Por exemplo, bairros, municípios, regiões metropolitanas, estados e até países podem ser as unidades de análise de abordagem ecológica. Todavia, faz-se necessário atentar que quanto maior a unidade de análise, maiores serão as diferenças e desigualdades dentro delas. Dentro de uma cidade, por exemplo, podem existir regiões com excelentes indicadores socioeconômicos e outras extremamente vulneráveis.

Indicadores Socioeconômicos e Saúde

Aspectos de natureza socioeconômica impactam mais a saúde dos indivíduos do que a própria biologia (GRAHAM, 2016). No caso de doenças infecciosas mais especificamente, é de se esperar que quanto menor o grau de vulnerabilidade que alguém está exposto, menor seja o risco de contrair doenças transmissíveis. Outro exemplo é que pessoas que possuam maior grau educacional sejam mais bem informadas e possam ter mais acesso de como se proteger de doenças (BARROZO *et al.*, 2019).

Para além dos indicadores socioeconômicos de maneira geral, existem também alguns que são relacionados diretamente a aspectos de saúde. Mortalidade infantil, expectativa de vida ao nascer, índice de infecção por doenças transmissíveis são alguns exemplos. Estes por sua vez possuem uma forte correlação com os indicadores socioeconômicos. Quanto melhor o Status socioeconômico, menor tende a ser seus problemas de saúde (BARROZO *et al.*, 2019).

Ainda assim, existem críticas quanto ao uso de indicadores socioeconômicos em estudos na área de saúde. O IDH é um exemplo, pois foi criado com o objetivo de auferir a qualidade de vida de uma pessoa e não para estudos relacionados a saúde. Por mais que exista a versão do IDH que dá mais peso a dimensão longevidade, ainda assim não é o ideal, pois como é um índice composto por diferentes aspectos, acaba ocorrendo redundância matemática, que pode atrapalhar a análise de questões de saúde. Para auxiliar nesse aspecto, foi criado o GeoSES, que busca ser um índice adequado para estudos em saúde (BARROZO *et al.*, 2019). Enfim, indicadores nunca serão perfeitos para medir determinados fenômenos sociais e novos indicadores surgirão, também com todas as limitações, para dar uma resposta a isso ou ao menos serem mais adequados.

Estatística Espacial e Regressões

Por conta do progresso tecnológico, cada vez mais existe a disponibilidade de Sistemas de Informação Geográfica (*Geographic Information Systems – GIS*), popularizando e permitindo estudos e análises que abarquem o contexto espacial (FRANCISCO, 2010). Uma das formas de incorporar o espaço geográfico nos modelos estatísticos é integrar termos autorregressivos espaciais baseados em proximidade ou adjacência. Em outras palavras, é avaliar se determinadas variáveis ou eventos estudados estão distribuídos aleatoriamente no espaço ou se existe algum tipo de padrão (clusterização ou dispersão). Para realizar essa análise, o modelo mais usual é criando uma matriz de vizinhança (W), que pode ser estabelecida via contiguidade ou distância de seus vizinhos. Uma vez definida a matriz de vizinhança, é possível calcular a autocorrelação espacial de uma variável. A principal medida de autocorrelação espacial é o Índice I de Moran, que correlaciona o valor de uma variável observada em uma região e os valores dessa mesma variável em regiões vizinhas. O indicador varia de -1 (perfeita dispersão) a 1 (perfeita associação), passando pelo valor de padrão espacial totalmente aleatório ($-1/n$, aproximadamente 0 para bases grandes) (MORAN, 1948). Cada observação (polígono) possui uma contribuição local para compor o indicador global de dependência espacial. O conjunto de

contribuições locais, denomina-se *Local Indicators of Spatial Association* (LISA), o qual pode ser visualizado espacialmente e testado para significância estatística para prover uma indicação de padrões de concentração espacial da variável em estudo na área (CÂMARA *et al.*, 2003). Já as técnicas de regressão são utilizadas para determinar como diferentes fatores considerados se relacionam através de uma equação. Basicamente, o objetivo é prever (ou estimar os coeficientes β_i) o valor de uma variável dependente (y) com base em valores de uma ou mais variáveis independentes (x), conforme fórmula clássica abaixo.

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p + \varepsilon$$

Dois conceitos são fundamentais quando se fala em regressão. O primeiro é o coeficiente de determinação (R^2), que indica o quão bem a regressão consegue explicar o fenômeno observado. É um valor que varia entre 0 e 1, sendo que zero indica que o modelo pouco explica a realidade observada e 1 que existe uma relação perfeita entre a reta de regressão e a realidade observada. Outro conceito importante é o valor-P, que informa se os coeficientes das variáveis independentes são estatisticamente significantes para a regressão. Como convenção costuma se utilizar que caso o valor-P seja inferior a 5%, o que implica em 95% de confiança, a variável é estatisticamente significativa. Caso contrário a variável não é adequada para integrar o modelo. (ANDERSON, SWEENEY e WILLIAMS, 2021).

No modelo Espacial Auto-regressivo (SAR), uma matriz de vizinhança (W) é incorporada como uma variável dependente espacialmente defasada (Wy), para a qual se estima o coeficiente (espacial) ρ , conforme abaixo.

$$y = \rho Wy + X\beta + \varepsilon$$

Pode-se interpretar a fórmula acima da seguinte maneira: W é a matriz de proximidade espacial, o produto Wy representa a dependência espacial em y e ρ é o coeficiente espacial autorregressivo. A hipótese nula para a não existência de autocorrelação, é que ρ é igual a zero. A ideia básica deste modelo é incorporar a auto correlação espacial como componente do modelo. O modelo também é conhecido como *Spatial Lagged Autoregressive Model* (FRANCISCO, 2010; ANSELIN; BERA, 1998).

3. METODOLOGIA

Dados Coronavírus

Levando em consideração o objetivo do presente trabalho e os dados disponíveis sobre o vírus, são duas as variáveis relacionadas à COVID-19 que serão utilizadas: Taxa de Incidência e Taxa de Mortalidade. A taxa de incidência é o número de casos por 100 mil habitantes, ao passo que a taxa de mortalidade mede a letalidade da doença nos casos confirmados. A primeira é calculada dividindo o número de casos em determinado espaço geográfico (no caso, os municípios) pela população total. Em seguida esse número é multiplicado por 100 mil. Já a taxa de mortalidade é calculada dividindo o número de mortes pela doença pelo número de casos confirmados. Certas fontes podem chamar também de taxa de letalidade.

Os dados referentes à COVID-19 foram obtidos do *Brasil.IO*, uma iniciativa sem fins lucrativos que tem como objetivo democratizar o uso de dados no Brasil, de forma tratar os dados públicos para que possam ser utilizados mais facilmente por qualquer um para qualquer fim. (BRASIL.IO, 2020b). Diariamente é consolidada em uma única base de dados as informações relacionadas à COVID-19, disponibilizadas pelas secretarias estaduais de saúde. Faz-se necessário um lembrete: o presente trabalho está sendo realizado durante o período da pandemia. Logo, os dados disponíveis são de até a data de elaboração da pesquisa. No caso, os dados vão até o dia 02 de novembro de 2020, que corresponde a 45^a semana epidemiológica do vírus. Para o tratamento desses dados, foi utilizada a linguagem Python e o pacote *Pandas*.

Como não é o propósito central do presente trabalho realizar uma análise temporal do comportamento do vírus, mas ainda sim pode ser relevante ter dados de momentos diferentes da pandemia, foram selecionadas 8 datas específicas para que seus dados fossem salvos e juntados em um mesmo *dataset* posteriormente. As datas foram selecionadas de forma: (I) possuir a última data disponível; (II) escolher a primeira data de forma que ao menos 2 municípios já registrassem casos de COVID-19; (III) que as seis datas intermediárias fossem pontos que dividissem igualmente o intervalo de tempo entre a primeira e a última data. A Tabela 1 mostra quais foram essas datas e a que semana epidemiológica se referem. Uma vez tendo escolhido esses 8 recortes temporais, eles foram georreferenciados com o *software* GeoDa.

Tabela 1. Recorte temporal escolhido para o estudo

Semana Epidemiológica de 2020	Início	Final
10	01 / março	07 / março
15	05 / abril	11 / abril
20	10 / maio	16 / maio
25	14 / junho	20 / junho
30	19 / julho	25 / julho
35	23 / agosto	29 / agosto
40	27 / setembro	03 / outubro
45	01 / novembro	07 / novembro

Fonte: Os autores

Dados Socioeconômicos

Como colocado no referencial teórico, há muito tempo se utiliza indicadores e índices socioeconômicos em análises de saúde. Faz bastante sentido aplicá-los a COVID-19 pois por se tratar de uma doença infecciosa, é de se esperar que aspectos socioeconômicos e de acesso a saúde influenciem no comportamento da doença (BARROZO *et al.*, 2019). A Tabela 2 contém a relação completa dos 22 índices e indicadores utilizados na presente pesquisa, juntamente de suas fontes originárias e a que ano se referem. Assim como foi feito com os dados relacionados à COVID-19, após coletados, esses dados foram georreferenciados com o *software* GeoDa. Posteriormente, todos os dados utilizados foram aglutinados em um mesmo arquivo.

Tabela 2. Dados e Indicadores Socioeconômicos e Demográficos coletados e utilizados

Nome da Variável	Dado	Fonte	Ano
IDHM	Índice de Desenvolvimento Humano Municipal - Geral		
IDHM_E	Índice de Desenvolvimento Humano Municipal - Dimensão Escolaridade	PNUD e IPEA	2010
IDHM_L	Índice de Desenvolvimento Humano Municipal - Dimensão Longevidade		
IDHM_R	Índice de Desenvolvimento Humano Municipal - Dimensão Renda		
IPRS_R	Índice Paulista de Responsabilidade Social - Dimensão Renda		
IPRS_L	Índice Paulista de Responsabilidade Social - Dimensão Longevidade	Seade	2018
IPRS_E	Índice Paulista de Responsabilidade Social - Dimensão Escolaridade		
den_demo	Densidade Demográfica	Seade	2020
ind_envl	Índice de Envelhecimento	Seade	2020
taxa_urb	Taxa de Urbanização	Seade	2020
pibpercapi	PIB <i>per capita</i>	Seade	2017
gini	Índice de Gini	IBGE	2010
GEOSES	GeoSES - Índice Principal		
GEO_educ	GeoSES - Dimensão Educação		
GEO_pobre	GeoSES - Dimensão Pobreza	Ministério da Saúde	2010
GEO_priva	GeoSES - Dimensão Privação Material		
GEO_rique	GeoSES - Dimensão Riqueza		
GEO_renda	GeoSES - Dimensão Renda		
GEO_segre	GeoSES - Dimensão Segregação		
dim_finan	Dimensão Finanças, trabalho e renda na plataforma Longeiver	Plataforma Longeiver	Múltiplos
dim_cidad	Dimensão Cidadania, saúde, bem-estar e segurança na plataforma Longeiver		
dim_habita	Dimensão Habitat do Idoso na plataforma Longeiver		

Fonte: Os autores, a partir das fontes descritas para cada variável

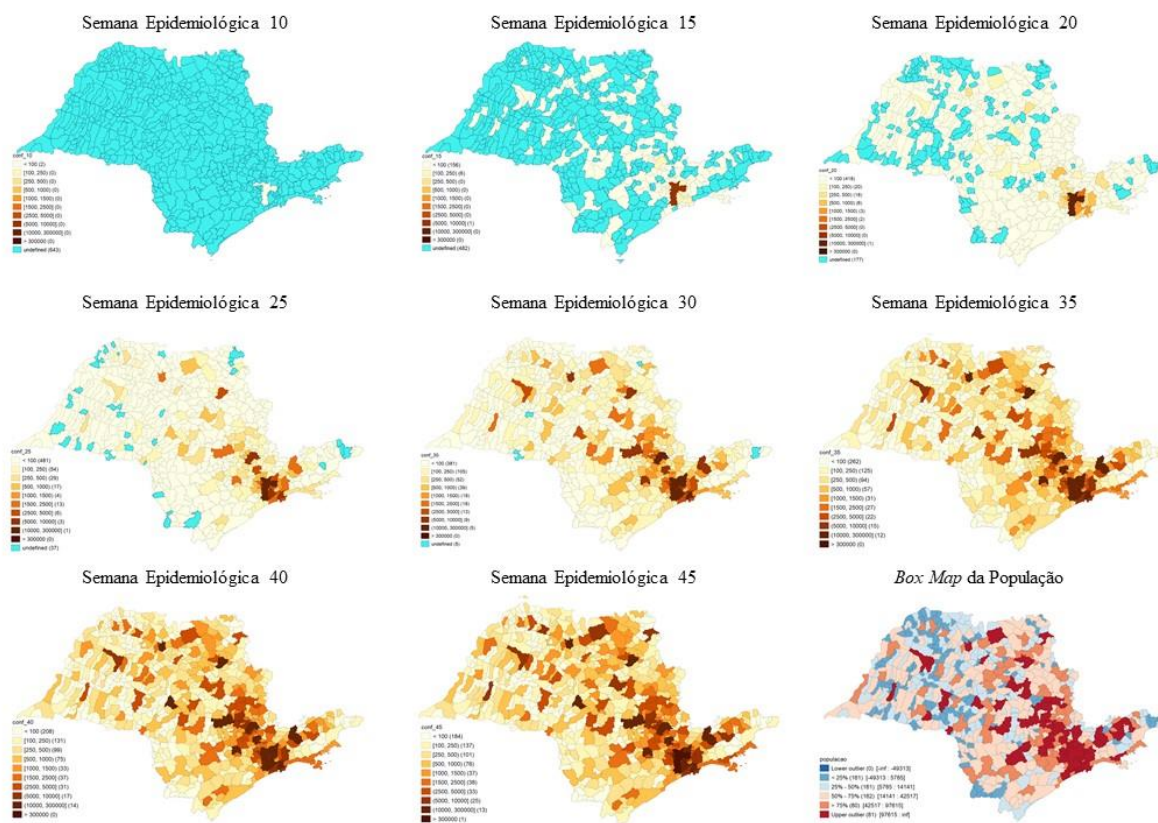
4. ANÁLISES EXPLORATÓRIAS E REGRESSÕES

A presente seção analisa descritivamente os dados referentes aos números da pandemia da COVID-19 nos municípios paulistas no recorte temporal adotado, além de apresentar os resultados das regressões realizadas com o objetivo de compreender a relação entre a COVID-19 e indicadores socioeconômicos, sob a ótica da estatística espacial.

O primeiro caso do Novo Coronavírus no Estado de São Paulo foi confirmado na capital paulista no dia 25 de fevereiro de 2020. Em 07 de março de 2020, 2 municípios apresentaram ao menos um caso positivo; em 11 de abril, 161 municípios; em 16 de maio, 460 municípios; em 20 de junho, 599 municípios; em 25 de julho, 638 municípios e em 29 de agosto, 644 municípios, de um total de 645 municípios no Estado de São Paulo. A Figura 1 mostra a evolução dos casos de COVID-19 no recorte temporal adotado. Os municípios em azul ciano são aqueles que não possuem casos confirmados. As cores mais escuras representam maior contingente (absoluto) de casos confirmados da doença, o que ocorre gradativamente ao longo do período estudado, com mais força em algumas localidades específicas do Estado. Por meio dos mapas é possível perceber que a doença começou no município de São Paulo e logo foi se espalhando para outros municípios de todo o Estado, principalmente aqueles da região metropolitana e mais próximos da capital, além dos municípios mais populosos do interior. O

último mapa do diagrama da Figura 1 (canto inferior direito) apresenta um *box map* da população do Estado de São Paulo. Por meio dele é possível perceber que os municípios do interior e afastados da capital que primeiro apresentaram casos positivos da doença são aqueles de maior população.

Figura 1. Diagrama de mapas dos municípios de SP de acordo com o número de casos confirmados de COVID-19 no recorte temporal adotado

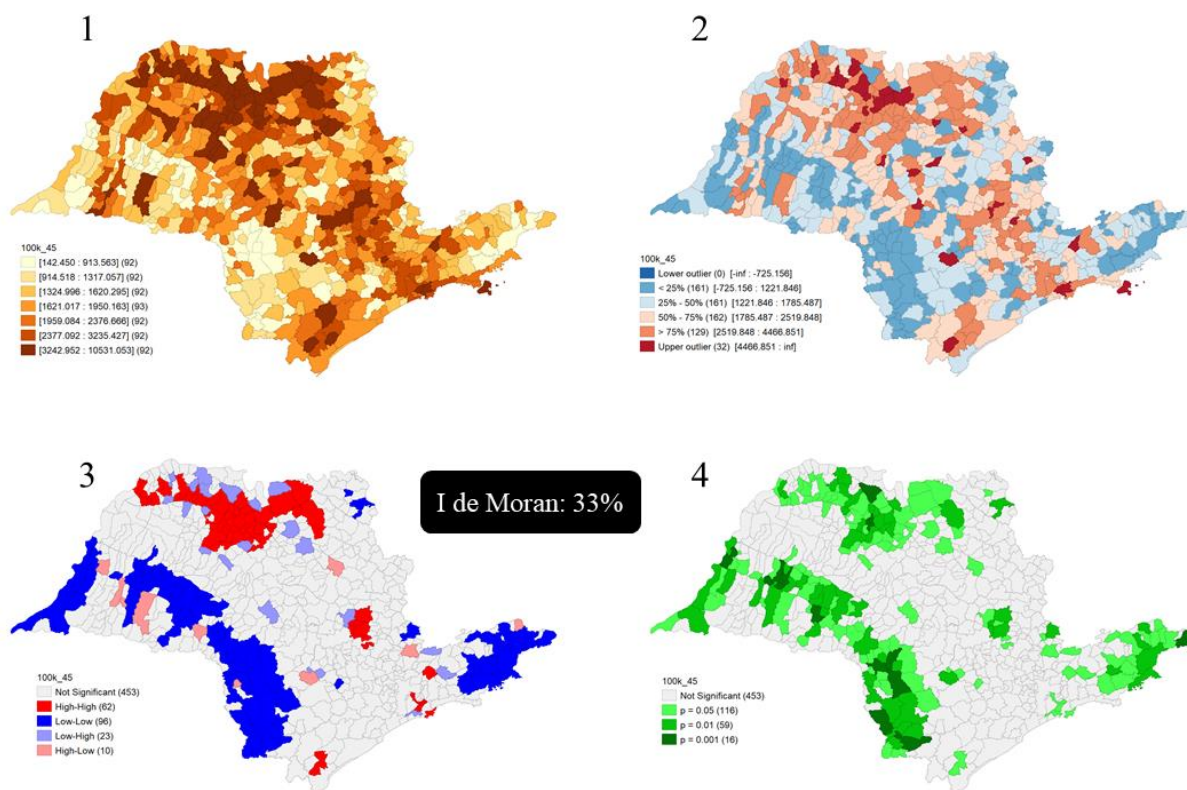


Fonte: Brasil.IO (2020a); Elaborado com o *software* GeoDA 1.18 (ANSELIN; IBNU; YOUNGIHN, 2006)
 Nota: Recorte temporal contempla março a novembro 2020 – Semanas Epidemiológicas 10 a 45 (Tabela 1)

Esse movimento de dispersão do vírus primeiramente nos centros urbanos e depois nas cidades mais interioranas provavelmente se deve ao fato que os municípios mais populosos possuem um maior fluxo de pessoas e mercadorias entre si, utilizando a malha rodoviária como vetor de transmissão. Levando em consideração que esse é um vírus que se dispersa por meio das relações e o contato social, essa dinâmica contribui para as cidades mais populosas e perto de cidades que já possuem casos da doença estejam mais suscetíveis a entrar em contato com o vírus. Além disso, também cabe destacar novamente que os casos da doença somente são confirmados e entram para as estatísticas oficiais após a realização dos testes – principalmente no começo na pandemia, eles eram mais escassos. Logo, há a possibilidade que municípios do interior e com menos recursos só tenham tido a capacidade de testar seus municípios tardiamente.

Além da dispersão do vírus pelos municípios do Estado, também é relevante analisar a intensidade que os municípios apresentaram casos positivos da doença. A métrica utilizada para avaliar o quanto os municípios foram afetados pelo vírus é a taxa de incidência (número de casos confirmados por 100 mil habitantes). A Figura 2 mostra 4 mapas relevantes para a análise dessa variável.

Figura 2. Mapa de Quantis (1), *Box Map* (2), *LISA map* (3) e *Significance map* (4) para a taxa de incidência da COVID-19 no recorte temporal mais recente (Semana Epidemiológica 45 de 2020)



Fonte: Brasil.IO (2020a); Elaborado com o *software* GeoDA 1.18 (ANSELIN; IBNU; YOUNGIHN, 2006)

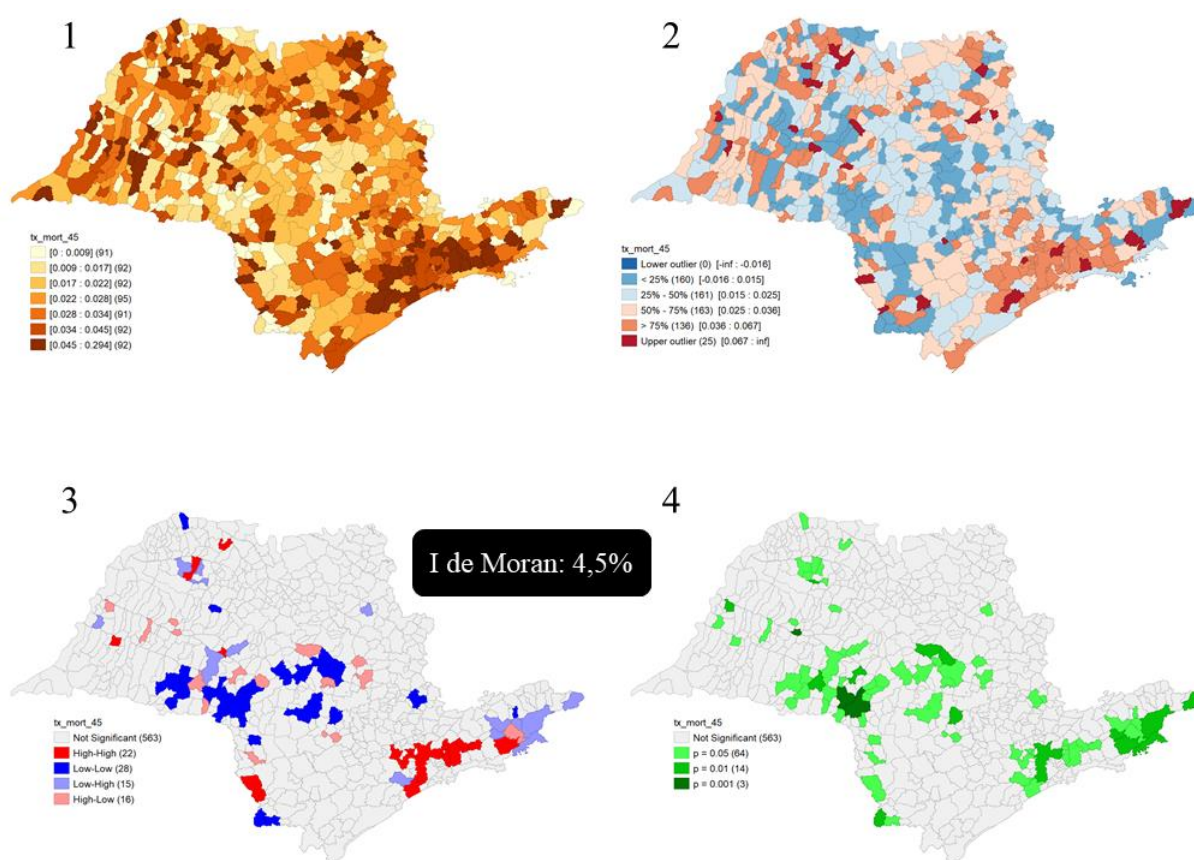
Nota: $n = 645$ municípios. Para as análises de Autocorrelação Espacial, $n = 644$, pois a Matriz de Vizinhaça adotada foi de Contiguidade de Primeira Ordem tipo *Queen* (município de Ilha Bela não apresenta vizinhos contíguos)

No Mapa de Quantis (1), que divide os municípios em 7 diferentes classes com o mesmo número de municípios em cada uma delas, quanto mais escuro aparece o município, maior é o número de casos de COVID-19 por 100 mil habitantes. O *Box Map* traz em um mapa as mesmas classificações de um *boxplot* (quartis). Por meio dele é possível notar que não há a presença de *outliers* inferiores e existem 32 *outliers* superiores, ou seja, municípios que apresentam taxas de incidência significativamente superiores do que a média. Além disso, também é relevante notar que visualmente pode-se perceber uma concentração de municípios com números altos desse indicador mais ao norte do Estado, além de uma concentração de municípios com números mais baixos do que a média nas regiões sudoeste, oeste e litoral norte. Assim sendo, para averiguar se existe algum tipo de tendência a clusterização dessa variável em relação ao espaço, pode-se realizar a análise da autocorrelação espacial. Os municípios que aparecem em vermelho no *LISA map* (3) são justamente aqueles que possuem números altos dessa variável e são vizinhos de municípios que também possuem números altos, a um nível de significância estatística de 95%. Já os municípios em azul escuro são aqueles que possuem números baixos para esse indicador e também são vizinhos de outros municípios que também apresentam números baixos. Todos os municípios coloridos em algum tom de verde no *significance map* (4) possuem significância estatística nessa análise. Por fim, o *I* de Moran indica que 33,0% do valor dessa variável é relacionado a localização geográfica. Um número bastante razoável e não desprezível.

É importante que se realize um apontamento quanto a matriz de contiguidade utilizada. A mesma análise foi realizada para outras ordens de contiguidade. As ordens 2, 3 e 4 obtiveram um *I* de Moran de 27%, 22% e 18%, respectivamente. O fato de à medida que a ordem de contiguidade aumenta a Autocorrelação Espacial diminuir corrobora a conclusão da tendência da clusterização da variável em relação ao espaço – quanto mais aumentamos o número de vizinhos (subindo a ordem), mais diferenças são apontadas entre cada município e a média dos vizinhos, diminuindo, pois, a auto-correlação espacial.

A Figura 3 apresenta a mesma estrutura da Figura 2, a menos do indicador apresentado – taxa de mortalidade da doença, também no recorte temporal mais recente.

Figura 3. Mapa de Quantis (1), *Box Map* (2), *LISA map* (3) e *significance map* (4) para a taxa de mortalidade da COVID-19 no recorte temporal mais recente (Semana Epidemiológica 45 de 2020)



Fonte: Brasil.IO (2020a); Elaborado com o *software* GeoDA 1.18 (ANSELIN; IBNU; YOUNGIHN, 2006)

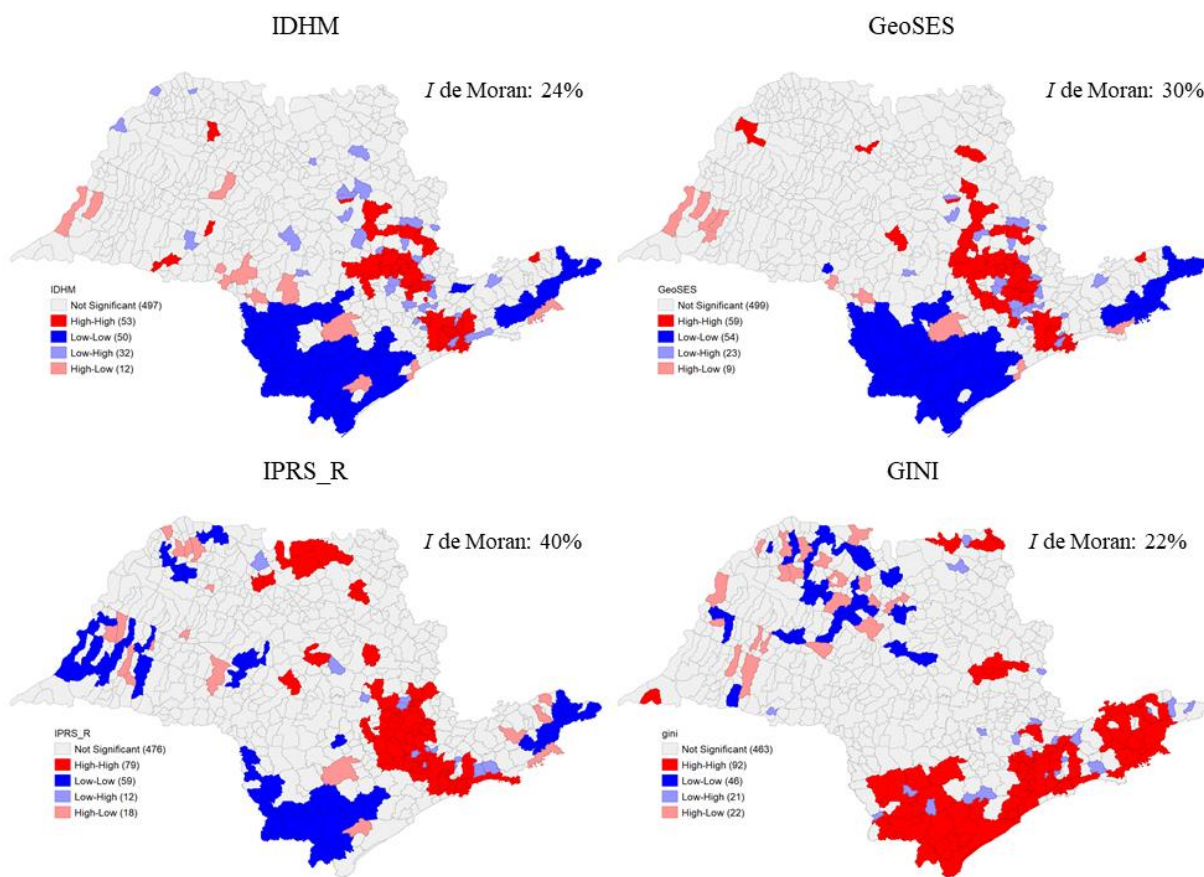
Nota: $n = 645$ municípios. Para as análises de Autocorrelação Espacial, $n = 644$, pois a Matriz de Vizinhaça adotada foi de Contiguidade de Primeira Ordem tipo *Queen* (município de Ilha Bela não apresenta vizinhos contíguos)

Já para a variável de mortalidade, é possível perceber por meio do Mapa de Quantis (1) e *box Map* (2) que não aparenta haver uma ou mais regiões no Estado com municípios próximos e com números similares para a taxa de mortalidade por COVID-19. Nesse caso, também não existem *outliers* inferiores e existem 25 *outliers* superiores. Para confirmar a hipótese de que aparentemente essa variável é aleatória em relação ao espaço e não apresenta clusterização, foi realizada a análise de autocorrelação espacial. Nesse caso, o *I* de Moran foi de 4,5%, um valor significativamente baixo, indicando que de fato essa variável é aleatória em relação ao espaço e não tende à clusterização. Observando o *LISA Map* (3) e *Significance Map* (4), também

podemos observar um número baixo de municípios significantes, além da não-existência de concentrações de municípios, como vimos para a variável de incidência.

Também foi realizada a análise de dependência espacial para as variáveis socioeconômicas coletadas. A Figura 4 mostra o *LISA map* juntamente dos valores do *I* de Moran para 4 dessas variáveis. Vale destacar como a região do Vale do Ribeira, no sul do Estado, apresenta uma grande quantidade de municípios com indicadores socioeconômicos e de qualidade de vida consideravelmente baixos. A Figura 5 mostra um mapa destacando os municípios que compõe a região do Vale do Ribeira.

Figura 4. *LISA Maps* e *I* de Moran para as variáveis Índice de Desenvolvimento Humano Municipal (IDHM), GeoSES, Índice Paulista de Responsabilidade Social – Renda (IPRS_R) e Gini



Fonte: Brasil.IO (2020a); Elaborado com o *software* GeoDA 1.18 (ANSELIN; IBNU; YOUNGIHN, 2006)

Nota: $n = 644$ – a Matriz de Vizinhaça adotada foi de Contiguidade de Primeira Ordem tipo *Queen* (município de Ilha Bela não apresenta vizinhos contíguos)

Foram realizadas regressões Lineares e Espaciais, para compreender se os indicadores socioeconômicos e a matriz espacial (variáveis independentes) possuem influência nas variáveis relacionadas à COVID-19 já exploradas acima (variáveis dependentes). Quanto as variáveis independentes, foram coletadas 22 de diversas fontes. A seguir serão apresentados apenas os resultados finais das regressões realizadas. É importante apontar que os critérios para as regressões foram: (I) contar apenas com variáveis independentes estatisticamente significantes (valor-P < 5%); e (II) sem a presença de forte multicolinearidade ($VIF < 10$) (DAOUD, 2017). Para isso, foi utilizada a metodologia *stepwise backward*, em que são retiradas e incorporadas uma a uma as variáveis de maior valor-P até se chegar a um conjunto de modelos com coeficientes de determinação (R^2 ou R^2 -ajustado) relativamente altos.

Figura 5. Mapa com destaque para os municípios da região do Vale do Ribeira em SP



Fonte: Elaborado com o *software* GeoDA 1.18 (ANSELIN; IBNU; YOUNGIHN, 2006)

Abordando primeiramente a variável de mortalidade da COVID-19, a Regressão Linear (OLS) contou com 4 variáveis: (I) Índice Paulista de Responsabilidade Social – Escolaridade, (II) Densidade demográfica, (III) Índice de envelhecimento e (IV) Dimensão Finanças, trabalho e renda na plataforma Longeviver. O R^2 ajustado foi de 4,05%, considerado bastante baixo. Isso indica que os indicadores socioeconômicos coletados não possuem um poder explicativo sobre a taxa de mortalidade da COVID-19. A seguir, realizou-se a regressão espacial (SAR) com as mesmas variáveis. Nesse caso, a própria variável espacial não contou com um valor-P significativo, indicando que o espaço também não é relevante ao se observar essa variável. Dado que o I de Moran visto anteriormente era de apenas 4,5% para essa variável, isso também já era esperado.

Realizando o mesmo procedimento para a taxa de incidência como variável dependente, obtivemos resultados mais significativos. A Regressão Linear (OLS) com as variáveis socioeconômicas como variáveis independentes contou com 4 variáveis estatisticamente significantes: (I) Índice Paulista de Responsabilidade Social – Longevidade, (II) PIB *per capita*, (III) Gini e (IV) GeoSES. O R^2 ajustado foi de 12,00%.

Inserindo o termo espacial auto-regressivo como variável dependente, isso é, realizando uma regressão espacial SAR, a variável Índice Paulista de Responsabilidade Social – Longevidade deixou de ser significativa, fazendo com o que o modelo ficasse com 3 variáveis independentes, além da própria variável espacial acrescentada. Agora, o R^2 saltou para 27,35%, aumentando o poder explicativo do modelo se comparado a regressão linear.

5. ANÁLISE DOS RESULTADOS

A partir do que foi apresentado e explorado anteriormente, é possível afirmar que existe um padrão na distribuição de casos de COVID-19 no Estado de São Paulo. É possível observar que no início da pandemia os primeiros municípios a observar casos positivos da doença foram aqueles de maior população e mais próximos a região metropolitana de São Paulo, em especial a capital. Visto que a capital foi o início e epicentro da pandemia no Estado e que essa é uma doença que se espalha por meio do contato social, faz sentido esse movimento. Municípios de maior população são também aqueles de maior atividade econômica, o que implica em maior fluxo de pessoas e mercadorias com outros centros urbanos.

Quanto a autocorrelação espacial da taxa de incidência da COVID-19 no Estado, pudemos avaliar que o território possui influência nessa variável. Obteve-se um I de Moran de 33% para a matriz de contiguidade *queen* de primeira ordem. Levando em conta que 644 municípios foram levados em consideração na análise, pode-se afirmar que esse evento não é distribuído aleatoriamente no espaço. Quando aumentamos a ordem da matriz, a autocorrelação espacial diminuiu (27%; 22% e 18% para as ordens 2, 3 e 4, respectivamente), reforçando que esse

evento não é aleatoriamente distribuído no espaço. Logo, faz sentido que o poder público leve em consideração o espaço quando pensar em planos de ação contra a pandemia. O Governo do Estado de São Paulo fez isso e em seu plano para combater a COVID-19. No “Plano SP” (PLANO SP, 2020), criado para o combate à pandemia, o Estado foi dividido em regiões que só puderam reabrir a economia (comércios, escolas, escritórios e outros) a depender de seus índices da COVID-19.

Quanto a taxa de mortalidade da doença, tivemos resultados diferentes se comparado a taxa de incidência. Sobre a autocorrelação espacial dessa taxa, obtivemos um *I* de Moran de apenas 4,5% na matriz de contiguidade tipo *Queen* de primeira ordem. Logo, isso indica que essa variável tende a ser aleatoriamente distribuída do espaço.

Também observamos a autocorrelação espacial de algumas das variáveis socioeconômicas coletadas. Chamou a atenção os baixos índices de desenvolvimento humano, renda, igualdade e escolaridade para os municípios do litoral sul de São Paulo, na região do Vale do Ribeira. Isso indica a necessidade de uma política governamental focalizada. O Governo do Estado de São Paulo possui um programa que foca especificamente nessa região, chamado “Vale do Futuro”. Esse programa tem como objetivo levar a região a um desenvolvimento sustentável, envolvendo as secretarias de desenvolvimento econômico, educação e outras (VALE DO FUTURO, 2020). Para compreender quais variáveis socioeconômicas e demográficas melhor explicam os casos e mortalidade de COVID-19 nos municípios do Estado de São Paulo, realizamos dois tipos de regressão, a Regressão Linear (*OLS*) e a espacial (*SAR*). Por meio da comparação dos coeficientes de determinação (R^2 e R^2 -ajustado) em ambos os tipos de regressão, podemos além de observar o quanto da variável dependente é explicada pelas variáveis independentes (socioeconômicas), ver o quanto que o território possui de impacto nas variáveis. Abordando primeiramente a Taxa de Mortalidade da doença, a regressão linear apenas com as variáveis estatisticamente significantes apresentou um R^2 ajustado muito baixo, de apenas 4,05%. Realizando então a regressão espacial (*SAR*), a variável espacial obteve um valor-P demasiadamente alto, indicando que a variável espaço não contribui para esse modelo. Em suma, concluímos que nem o território nem as variáveis socioeconômicas escolhidas são capazes de explicar a taxa de mortalidade por COVID-19 nos municípios paulistas.

Já quanto a variável de incidência da COVID-19, tivemos resultados significativos para a análise. Quanto a regressão linear múltipla final, isso é, apenas com as variáveis estatisticamente significantes e sem problemas de multicolinearidade, obtivemos um coeficiente de determinação ajustado (R^2 ajustado) de 12,00% e 4 variáveis independentes. Ou seja, a um nível de confiança de 95%, 12,00% do valor da variável dependente é explicado pelas variáveis independentes. As variáveis independentes são: (I) Índice Paulista de Responsabilidade Social - Dimensão Longevidade; (II) PIB *per capita*; (III) *gini*; (IV) *GeoSES*. A equação da regressão está abaixo:

$$\text{Taxa de Incidência} = 2544 + 18,91 * \text{IPRS_L} + 0,00657 * \text{pibpercapi} - 4068 * \text{gini} + 1055 * \text{GeoSES}$$

Chama a atenção os sinais dos estimadores de cada variável. *IPRS_L*, *pibpercap* e *GeoSES* apresentam sinais positivos. Ou seja, um aumento neles tende a levar a um aumento na taxa de incidência da COVID-19 nos municípios paulistas. Ao olhar do referencial teórico e das informações relacionadas à COVID-19 trazidas anteriormente, isso faz todo o sentido, pois esses três indicadores são maiores justamente nos maiores e mais ricos municípios do Estado. Visto que a doença se espalha melhor em ambientes urbanos e que chegou ao Estado pelo município de São Paulo, faz sentido que os municípios mais afetados pela doença sejam também os mais desenvolvidos. Já o Índice de Gini possui uma relação negativa com a taxa de incidência da doença. Ou seja, caso ele seja maior, o número de casos de COVID-19 tende a ser menor. Lembrando que quanto maior o Gini maior é a desigualdade. Logo, também

percebemos que quanto maior a desigualdade menor tende a ser o número de casos. Mais uma vez, olhando para o referencial teórico e dados analisados anteriormente, percebemos que os municípios que mais foram impactados pelo vírus também são os mais desiguais.

Partindo para a regressão espacial (SAR), porém sem a variável *IRPS_L*, que deixou de ser significativa, obtivemos resultados bastante relevantes. O R^2 saltou para 27,35%, reforçando a informação que o espaço é uma variável relevante quando se trata da taxa de incidência de COVID-19 nos municípios paulistas. Abaixo a equação da regressão espacial apenas com as variáveis significantes:

$$\text{Taxa de Incidência} = 2115,25 + 0,4803 * Wy + 0,00594 * \text{pibpercapi} - 2395,86 * \text{gini} + 778,845 * \text{GeoSES}$$

A Tabela 3 mostra os coeficientes de determinação (R^2) das regressões finais realizadas para ambas as variáveis dependentes. O R^2 da regressão espacial para a variável da taxa de mortalidade está destacado pois nela não houve significância estatística para alfa de 5%.

Tabela 3. Coeficientes de determinação (R^2) das as regressões finais

	Regressão Linear Múltipla (OLS)	Regressão Espacial (SAR)
Taxa de Incidência	$R^2 = 12,00\%$	$R^2 = 27,35\%$
Taxa de Mortalidade	$R^2 = 4,05\%$	$R^2 = 4,74\%$

Fonte: Os autores

Obs: Destaque em vermelho para a regressão com termo espacial não estatisticamente significativa (valor- $P > 0,05$).

6. CONSIDERAÇÕES FINAIS

Trouxemos no presente trabalho uma série de análises a respeito da relação entre os dados da COVID-19 nos municípios paulistas e sua relação com índices socioeconômicos e demográficos. Além disso, incorporamos na análise o aspecto territorial, com o objetivo de entender se e de que forma o espaço geográfico influencia as variáveis escolhidas. Concluimos que a taxa de mortalidade da COVID-19 nos municípios paulistas tende a ser uniforme e não correlacionada com os dados socioeconômicos e demográficos coletados. Além disso, também concluimos que essa taxa tende a ser aleatoriamente distribuída no espaço, visto que apresenta uma autocorrelação espacial muito baixa. Já a taxa de incidência da doença apresenta uma correlação mais forte com alguns indicadores socioeconômicos e não é aleatoriamente dispersa no território. Para essa variável, houve um aumento significativo do coeficiente de determinação (R^2) quando incluímos o território como variável independente, isso é, quando comparada a regressão linear (OLS) com a regressão espacial (SAR).

Como trazido no referencial teórico, ao realizar estudos e ações contra doenças infecciosas, utilizar a variável espaço melhora a análise e permite que se chegue a conclusões mais assertivas. Barrozo *et al.* (2019) apontaram que local onde alguém vive diz muito mais sobre sua saúde e suscetibilidade em contrair doenças infecciosas do que fatores genéticos ou biológicos. Os resultados do presente trabalho corroboram essa ideia. Isso reforça a necessidade que autoridades públicas que lidem com a gestão da saúde e controle epidemiológico embarquem conhecimentos de Sistemas de Informação Geográfica (GIS) em seu cotidiano. Para além do controle de doenças que já se mostram um problema, como é o caso da COVID-19, o uso de GIS pode ser útil para identificar locais mais vulneráveis e suscetíveis a determinadas doenças. Esse planejamento e conhecimentos podem tanto salvar vidas como gerar grandes economias para o sistema público de saúde.

Por meio dos resultados das regressões realizadas, notamos como a COVID-19 é uma doença que afeta principalmente as cidades mais populosas e de maior desenvolvimento econômico.

Caso voltemos ao início da pandemia, os primeiros casos ocorreram entre os setores mais abastados da sociedade, visto que a doença veio de fora do país, trazida por quem realizou viagens internacionais. Todavia, por conta de a unidade de análise ter sido os municípios do Estado, não pode ser investigado se a doença segue infectando mais fortemente as camadas mais ricas da população ou se as camadas mais pobres - que possuem empregos de menor qualificação profissional e não podem trabalhar de forma remota em sua maioria - passaram a ser mais fortemente impactados (CASTRO *et al.*, 2021). Como apontado no referencial teórico (MINISTÉRIO DA SAÚDE, 2006), quanto maior a unidade de análise, maior a desigualdade presente dentro dessa unidade. Como mostra o próprio índice de Gini destacado anteriormente, a desigualdade intramunicipal nos municípios de São Paulo é bastante grande. Assim sendo, não pode ser verificado, por meio dos dados utilizados, se dentro dos municípios existe uma diferença relevante na taxa de incidência ou de mortalidade entre a população mais rica e mais pobre.

Por mais que não seja um dos focos do trabalho, também pode-se realizar uma análise exploratória dos dados socioeconômicos e demográficos coletados. É notável a diferença regional presente no Estado, indicada pela autocorrelação espacial que foi bastante significativa para essas variáveis. Municípios tendem a estar próximos de outros municípios de características similares, seja em características urbanas, de renda média ou acesso a serviços públicos (JANNUZZI, 2002). Logo, essa tendência a clusterização é esperada. Como um exemplo, a região do Vale do Ribeira concentra os piores índices socioeconômicos do Estado – seja de renda, desigualdade ou educação. Isso pode ser comprovado por meio da autocorrelação espacial. Como apontado na seção anterior, é plausível que gestores públicos comecem a incorporar a variável espacial na proposição de políticas públicas de desenvolvimento local e redução de desigualdades.

Este não é um trabalho de epidemiologia, mas pode ser lido e gerar contribuições para diversos setores da sociedade. É bom que gestores públicos saibam quais fatores estão mais correlacionados com doenças infecciosas como a COVID-19 e possam assim se planejar melhor. Para a sociedade civil, os dados geram um grande empoderamento, pois assim podem cobrar que o poder público realize ações que sejam de fato assertivas e corretas quanto ao combate e gestão da pandemia. Já o setor privado de maneira geral pode usar as informações aqui presentes para saber em quais regiões a pandemia deve ser mais ou menos severa e assim planejar investimentos, logística e a saúde de seus colaboradores.

O aperfeiçoamento e o aprofundamento da presente pesquisa podem ser realizados de diversas maneiras. Novas variáveis podem ser analisadas, além das duas variáveis da COVID-19: Taxa de Incidência e Taxa de Mortalidade. Existem muitos outros dados referentes a doença que sem dúvida valem a pena serem estudados. Dentre elas, a Taxa de Transmissão, que informa para quantas pessoas um infectado está passando a doença. Caso esse número seja maior do que 1, a doença está acelerando e caso seja menor do que 1, a doença está diminuindo. Seria relevante, por exemplo, avaliar quais fatores afetam essa taxa ao longo do tempo. Além disso, para a presente pesquisa foram escolhidos apenas 8 recortes temporais. É possível realizar análises mais aprofundadas caso os números da doença sejam realizados dia a dia, principalmente quanto aos fatores da transmissão e disseminação entre os municípios, e incorporar períodos mais recentes ao estudo. Perguntas como “O território possui influência na taxa de transmissão da COVID-19?” ou “O que tem em comum os municípios que contraíram a doença por último?” podem ser estudadas. Por fim, destaca-se que dados como as mortes sobressalentes ou pesquisas sorológicas também podem ser analisados. Em suma, tendo acesso a mais dados referentes à COVID-19, principalmente se esses forem georreferenciados (sabendo, por exemplo, o CEP de moradia e trabalho de cada infectado e morte), análises mais profundas podem ser realizadas. Mudar a unidade de análise também pode ser extremamente relevante. Conforme discutido anteriormente, quanto maior a unidade de análise, maior as desigualdades intra-unidades. Nesse

sentido, tendo todos os dados disponíveis, seria relevante realizar uma análise dos distritos mais atingidos na cidade de São Paulo, por exemplo. Avaliar se nesse caso os distritos mais atingidos também são os mais ricos e também se dentro da cidade a taxa de mortalidade é igual em todos os distritos ou não.

Por fim, destaca-se que outros dados socioeconômicos e demográficos podem ser acrescentados na análise. Para a definição deles, é relevante que se ouça profissionais da área da epidemiologia, pois certamente são capazes de apontar quais variáveis são mais ou menos relevantes na análise de doenças infecciosas.

REFERÊNCIAS BIBLIOGRÁFICAS

- ADLER, N. E., OSTROVE J. M. **Socioeconomic Status and Health: What We Know and What We Don't**. Ann N Y Acad Sci, 1999.
- ANDERSON, David R.; SWEENEY, Dennis J.; WILLIAMS, Thomas A. **Estatística aplicada à administração e economia**: subtítulo do livro. 5 ed. Boston: CENGAGE LEARNING, 2021.
- ANSELIN, Luc, IBNU Syabri, YOUNGIHN Kho. **GeoDa: An Introduction to Spatial Data Analysis**. Geographical Analysis 38 (1), 5-22, 2006.
- ANSELIN, Luc, BERA, Anil K. Introduction to spatial econometrics. **Handbook of applied economic statistics**, v. 237, 1998.
- BARROZO, Ligia *et al.* **GeoSES – Um índice Socioeconômico para estudos de saúde no Brasil**. Universidade de São Paulo, 2019.
- BRASIL.IO. **COVID-19**: Boletins informativos e casos do Coronavírus por município por dia, 2020a. Disponível em: <https://brasil.io/dataset/covid19/caso_full/>. Acesso em 11 de novembro de 2020.
- BRASIL.IO. **Manifesto BRASIL.IO**, 2020b. Disponível em: <<https://brasil.io/manifesto/>>. Acesso em 11 de novembro de 2020.
- BRASIL.IO. **COVID-19**: Boletins informativos e casos do Coronavírus por município por dia, 2021. Disponível em: <https://brasil.io/dataset/covid19/caso_full/>. Acesso em 03 de junho de 2021.
- CÂMARA, Gilberto *et al.* Tutorial Análise Geográfica. In: CONGRESSO GEOBRASIL 2003, 2003, São Paulo.
- CASTRO, M. C., KIM, S., BARBERIA, L., RIBEIRO, A. F., GURZENDA, S., RIBEIRO, K, B., ABBOTT, E., BLOSSOM, J., RACHE, B., SINGER, B. H. Spatiotemporal pattern of COVID-19 spread in Brazil. **Science** 10.1126/science.abh1558, 2021.
- DAOUD, Jamal. **Multicollinearity and Regression Analysis**. J. Phys.: Conf. Ser. 949 012009, 2017.
- FRANCISCO, E. R. **Indicadores de Renda baseados em Consumo de Energia Elétrica: Abordagens Domiciliar e Regional na Perspectiva da Estatística Espacial**. Tese (Doutorado) - Escola de Administração de Empresas de São Paulo, 2010.
- GRAHAM, G. N. **Why Your ZIP Code Matters More Than Your Genetic Code: Promoting Healthy Outcomes from Mother to Child**. Breastfeed Med, 2016.
- JANNUZZI, P. **Considerações sobre o uso, mau uso e abuso dos indicadores sociais na formulação e avaliação de políticas públicas municipais**. RAP. Rio de Janeiro, 2002
- MCLEOD, Kari. **Our sense of Snow: the myth of John Snow in medical geography**. Social Science & Medicine. Volume 50, Issues 7–8, Pages 923-935, 2000.
- MEIJER M; *et al.* **Do neighborhoods affect individual mortality? A systematic review and meta-analysis of multilevel studies**. Soc Sci Med [Internet], 2012

MINISTÉRIO DA SAÚDE. **Abordagens espaciais na Saúde Pública**: Volume 1. Brasília: Secretaria de Vigilância em Saúde do Ministério da Saúde e Fundação Oswaldo Cruz, 2006.

MORAN, Patrick A. P. The interpretation of statistical maps. *Journal of the Royal Statistical Society, Series B (Methodological)*, v. 10, n. 2, p. 243-251, 1948.

OMS. **Timeline of WHO's response to COVID-19**. Publicado em 29 de junho de 2020. Disponível em: <https://www.who.int/news/item/29-06-2020-covidtimeline>. Acesso em 23 de outubro de 2020.

PLANO SP. **Informações sobre o Programa de Combate à COVID-19**, 2020. Governo do Estado de São Paulo. Disponível em: < <https://www.saopaulo.sp.gov.br/planosp/>>. Acesso em 15 de outubro de 2020.

ROUX, A. V.; MAIR, C. **Neighborhoods and health**. *Ann N Y Acad Sci*, 2010.

SANTOS, M. **O espaço dividido: Os Dois Circuitos da Economia Urbana nos Países Subdesenvolvidos**. Rio de Janeiro: Ed. Francisco Alves, 1979

VALE DO FUTURO. **Informações sobre o programa de desenvolvimento do Vale do Ribeira**, 2020. Governo do Estado de São Paulo. Disponível em: < <https://www.saopaulo.sp.gov.br/valedofuturo/>>. Acesso em 15 de outubro de 2020.