Fundação Getúlio Vargas
Escola de Administração de Empresas de São Paulo
Unidade Brasília
MBA Executivo em Business Analytics e Big Data

Alexandre de Almeida Fonseca
Ana Paula Garutti
Davi Khoury Oliveira
Eduardo Campedelli Kavamoto
Geraldo Rodrigues Júnior
José Ricardo Almeida de Britto Filho
Marcelo Semerene Farah

Influências de variavéis sociodemográficas e políticas em hospitalizações por Síndrome Respiratória Aguda Grave causadas por Covid-19 em municípios brasileiros: uma abordagem em Estatística Espacial

Artigo

Brasília
2021

Alexandre de Almeida Fonseca
Ana Paula Garutti
Davi Khoury Oliveira
Eduardo Campedelli Kavamoto
Geraldo Rodrigues Júnior
José Ricardo Almeida de Britto Filho
Marcelo Semerene Farah

Influências de variavéis sociodemográficas e políticas em hospitalizações por Síndrome Respiratória Aguda Grave causadas por Covid-19 em municípios brasileiros: uma abordagem em Estatística Espacial

Artigo apresentado no programa de MBA Executivo em Business Analytics e Big Data, como parte dos requisitos necessários à conclusão de curso.

Orientador: Eduardo de Rezende Francisco
Coorientador: Álvaro Teixeira Villarinho

Brasília
2021

Agradecimentos

Não importa o quanto nos esforcemos para sermos justos ao lembrar de todas as pessoas que colaboraram de alguma forma para a concepção deste trabalho: sempre alguns serão esquecidos (e por isso pedimos desculpas antecipadamente!).

Contudo, esse trabalho não seria sequer cogitado sem o empenho e solicitude do prof. Dr. Eduardo de Rezende Francisco. Sua dedicação nas aulas e a disposição em orientação em horários extraclasse foram primordiais para que esse trabalho saísse do forno sem “solar”.

Da mesma maneira, o prof.MSc. Álvaro Teixeira Villarinho sempre deu contribuições e *insights* a respeito de eventuais problemas e sugestões de soluções.

Por fim, agradecemos ao coordenador e professor PhD Jose Luiz Carlos Kugler, por sempre ouvir reclamações e sugestões da nossa (barulhenta) turma e nos atender sempre que possível.

Os autores.

Resumo

O presente estudo tem como objetivo avaliar a influência de variáveis na contaminação grave por Covid-19.

Foram utilizadas variáveis sociodemográficas dos municípios e o resultado, também por município, do segundo turno das eleições presidenciais de 2018. Foram utilizadas bases públicas do Instituto Brasileiro de Geografia e Estatística (IBGE), Instituto de Pesquisa Econômica Aplicada (IPEA), Atlas BR, Departamento de Informática do SUS (DATASUS) e Tribunal Superior Eleitoral (TSE).

Foram utilizadas técnicas de análise descritiva e regressão linear múltipla, modelo autorregressivo espacial (Spatial Auto-Regressive model, SAR), e modelo de regressão geograficamente ponderada (Geographically Weighted Regression, GWR) para definição do modelo preditivo. Através do uso do Modelo GWR, obteve-se uma explicação de 53.18% da variação média da taxa por 1000 habitantes de contaminados graves durante o período entre os meses de julho de 2020 a janeiro de 2021.

Tais achados permitem endereçar melhor, no âmbito da gestão pública, políticas mais eficientes, eficazes e assertivas.

Palavras-chave: Estatística Espacial. Covid-19. Regressão Espacial.

Abstract

The present study aims to assess the influence of variables on severe contamination by Covid-19.

Sociodemographic variables of the counties were used, and the result, also by county, of the second round of the 2018 presidential elections. Public bases of the Brazilian Institute of Geography and Statistics (IBGE), Institute of Applied Economic Research (IPEA), Atlas BR, Department of Informatics (DATASUS) and Superior Electoral Court (TSE).

Descriptive analysis techniques and multiple linear regression, spatial autoregressive model (SAR), and geographically weighted regression model (GWR) were used to define the predictive model. Using the GWR Model, an explanation of 53.18% of the average rate variation per 1000 inhabitants of serious contaminants was obtained during the period from July 2020 to January 2021.

Such findings make it possible to better address, within the scope of public management, more efficient, effective and assertive policies.

Keywords: Spatial Statistics. Covid-19. Spatial regression.

Lista de ilustrações

Lista de tabelas

Sumário

1 Introdução

A doença causada pelo novo coronavírus, Covid-19, tem impactado o cenário mundial. Atingiu os cinco continentes do planeta e, em março de 2020, foi classificada pela Organização Mundial da Saúde como uma pandemia. Os surtos da doença em várias regiões do mundo causaram milhares de mortes, o que gera a todo instante numerosa produção intelectual sobre abalos econômicos, sociais e políticos da doença, além de teorias e reflexões sobre o futuro das sociedades impactadas pela atual crise sanitária.

A Covid-19 expôs o Brasil a um desafio sem precedentes. A doença acirrou ainda mais uma guerra ideológica, causando a politização e a ideologização do vírus. A consequência disso gera incerteza acerca da eficiência do impacto das políticas públicas e o crescimento do número de informações falsas, contribuindo assim com o pânico generalizado.

Diante do cenário acima exposto, foi escolhido como objeto da análise proposta para este estudo a avaliação da influência de variáveis socio-demográficas e políticas na contaminação grave por Covid-19.

2 Objeto e Objetivo do Estudo

A pandemia da Covid-19, uma das espécies de Síndrome Respiratória Aguda Grave (SRAG), foi principal responsável das causas de morte no Brasil em 2020. Conforme estudo da Fiocruz, mais de 70% de mortes por SRAG em 2020 foram causadas por Covid-19 (Saraiva, 26/02/2021).

Para conter a pandemia, foram implementadas medidas para retardar a propagação do vírus na tentativa de evitar uma sobrecarga no sistema de saúde, com um grande número de pacientes em estado grave.

Ainda neste contexto, notícias associadas a Covid-19 afirmam haver correlação entre apoiadores do atual Presidente da República, Jair Bolsonaro, com a piora da pandemia:

- "Pandemia é pior nas cidades governadas por apoiadores de Bolsonaro, diz pesquisa"[1].
- "Covid no DF: mulheres são maior parte de infectados, mas homens morrem mais"[2].

Neste contexto, o objetivo desta pesquisa é de avaliar quais variáveis explicam a quantidade de pessoas com casos graves de Covid-19 no Brasil. A pesquisa é realizada no nível municipal, e visa identificar qual o poder de explicação da taxa da população que contraiu Covid-19 e evoluiu para o quadro de SRAG no período de julho de 2020 a janeiro de 2021 a partir de modelos distintos comparados: (i) regressão linear múltipla; (ii) modelo autorregressivo espacial (*Spatial Auto-Regressive model*, SAR), e (iii) modelo de regressão geograficamente ponderada (*Geographically Weighted Regression*, GWR).

Nesse sentido, como podemos afirmar que as políticas públicas foram eficientes no combate ao Covid-19? O que influencia uma maior contaminação pelo Covid-19 e quais variáveis impactam na contaminação? A resposta destas perguntas visam avaliar com maior precisão a eficiência das políticas públicas, além de proporcionar um cotejamento entre notícias e dados com a finalidade de validar inferências jornalísticas.

---

[1] Fonte: Correio Braziliense (https://www.correiobraziliense.com.br/brasil/2020/10/4881890-pandemia-e-pior-nas-cidades-governadas-por-apoiadores-de-bolsonaro-diz-pesquisa.htmL, recuperado em 21, abril, 2021)

[2] Fonte: Correio Braziliense (https://www.correiobraziliense.com.br/app/noticia/cidades/2020/07/24/interna_cidadesdf,874833/covid-no-df-mulheres-sao-maior-parte-de-infectados-mas-homens-morrem.shtml, recuperado em 21, abril, 2021)

3 Fundamentação Teórica

3.1 Modelos de Regressão Múltipla

Um Modelo de Regressão Linear é uma das várias técnicas estatísticas de dependência. Com ela pode-se verificar a relação entre uma única variável dependente ou predita métrica e outra (ou diversas outras) variáveis ou covariáveis, independentes ou preditoras, métricas ou dicotômicas. O resultado é uma equação que traz pesos para as variáveis preditoras e que correspondem à contribuição que cada uma delas dá à variável predita.

Sua formulação básica é a seguinte:

Yi = β0 + β1Xi + β2Xi + Ei

em que:

Yi = é a variável predita;

β0 = é o intercepto. É o valor da variável predita caso as covariáveis sejam nulas;

β1, β2 = são os coeficientes da regressão para cada uma das covariáveis. É o parâmetro que mostra o impacto que aquela covariável tem na variável predita;

Xi = covariáveis;

Ei = erro aleatório do modelo que segue um distribuição normal e são independentes.

Neste trabalho, a utilizaremos da seguinte forma:

“Taxa de hospitalizações por Síndrome Respiratória Aguda Grave-Período 2” = β0 + β1 * “Taxa de hospitalizações por Síndrome Respiratória Aguda Grave-Período 1” + β2 * “Taxa de votos em Jair Bolsonaro” + β3 * “Renda percapita” + β4 * “IVS Infraestrutura Urbana” + β5 * “Taxa populacional Feminina” + Ei

Para comprovar a dependência que a infecção respiratória aguda grave causada pelo coronavírus SARS-CoV-2 possui perante variáveis sociodemográficas, econômicas e políticas, foi utilizado uma técnica de dependência, o Modelo de Regressão Linear Múltipla.

Conforme (Hair Jr., Black, Babin, Anderson, & Tatham, 2005) “as técnicas de dependência são baseadas no uso de um conjunto de variáveis independentes para prever e explicar uma ou mais variáveis dependentes”, portanto, é possível estimar a taxa municipal de infectados pela síndrome respiratória aguda grave, dependendo de variáveis sociodemográficas e políticas existe ali. Pode-se também verificar qual das variáveis influência mais na taxa de infectados, ou seja, comparar o grau de impacto dentre as características sociais, demográficas e políticas de cada município.

3.2 Modelos de Regressão Espacial

A incorporação de variáveis espaciais em um Modelo de Regressão, demonstrando a dependência espacial entre dados, é uma técnica recente. Apesar de recente, ela é uma característica inerente às representações de dados em divisões territoriais. Ela acrescenta a autocorrelação espacial existente, incorporando entre as variáveis independentes uma variável espacial.

Neste estudo foram utilizados dois Modelos espaciais, o *Geographically Weighted Regression Model* (GWR) e o *Spatial Autoregressive Model* (SAR), mas antes de falarmos dos dois, discorreremos sobre dois pontos fundamentais em uma Regressão Espacial, o I de Moran e o *Local Indicator Spatial Association* (LISA).

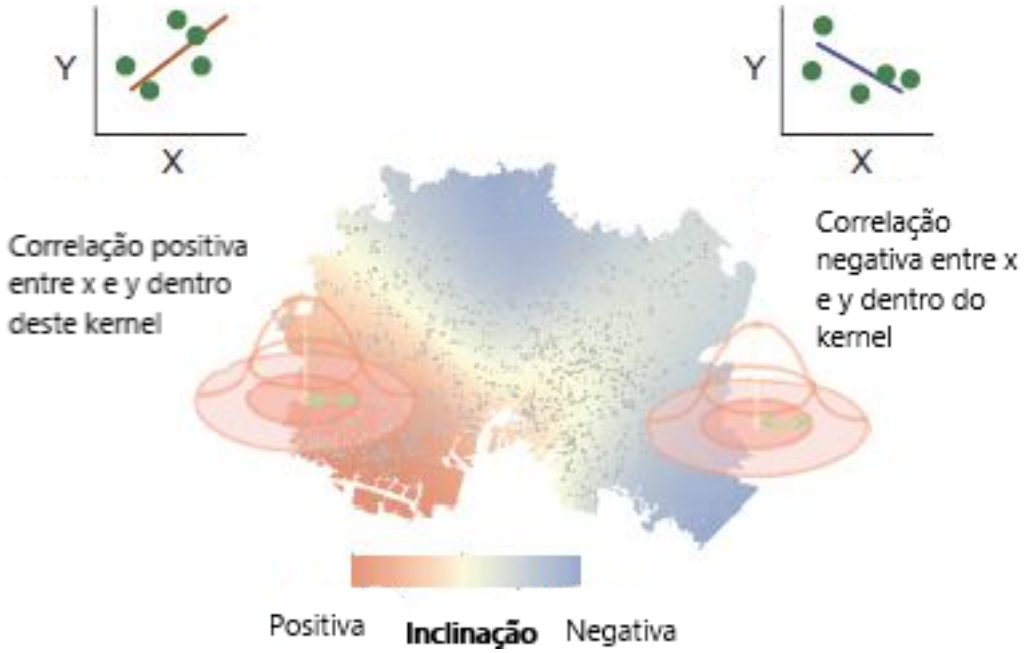
O *I* de Moran mede a autocorrelação espacial, ou seja, neste nosso caso, ele mede a correlação da taxa por 1000 habitantes de contaminados graves durante o 2º período da amostragem (jul20 a jan21) dos municípios brasileiros com seus vizinhos. O valor pode variar de -1 a +1. Quanto mais alto seu valor (mais próximo de 1), sinal de que os municípios com maior taxa estarão mais próximos dos de maior taxa, em outras palavras, os mais parecidos estão mais próximos entre si.

O mapa que melhor representa a associação espacial é o LISA. Ele agrupa as áreas analisadas calculando *I* de Moran Local para cada uma delas gerando uma significância estatística. Visualmente é a melhor ferramenta para análise e compreensão da associação de variáveis no espaço.

3.2.1 *Geographically Weighted Regression Model* (GWR)

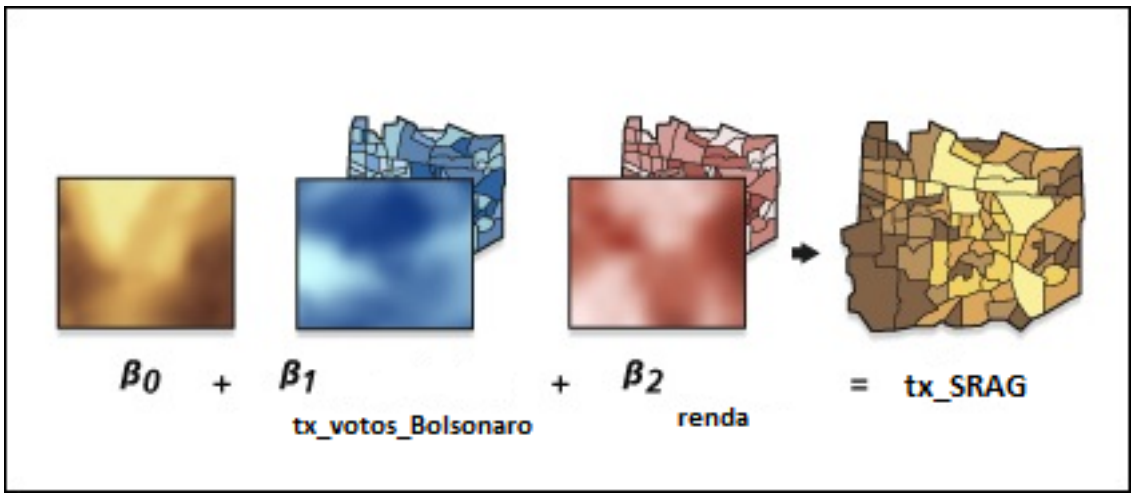
Segundo (Fotheringham, Brunsdon, & Charlton, 2002), o modelo GWR foi um método alternativo desenvolvido para análises locais espaciais de conjuntos de dados multivariados. Ela é baseada na estrutura tradicional regressão linear, mas que incorpora relações espaciais locais de forma intuitiva.

Figura 1 – Conceito da abordagem no GWR.



No GWR, as observações são ponderadas de acordo com seu ponto de proximidade *I* (determinado pelo tamanho do *kernel* utilizado) (USDA-ARS JORNADA EXPERIMENTAL RANGE, BLM-AIM PROGRAM, & IDAHO CHAPTER OF THE NATURE CONSERVANCY, 2012). Isso garante que o peso de uma observação não permanece constante em uma calibração, ao contrário, ele varia com *I*. Em suma, o GWR utiliza um *kernel* (também chamado de *janela* ou *banda*) que se move sobre a área de estudo e procura os melhores resultados para encaixe em determinada sub-área. Além disso, o *kernel* estabelece a razão decrescente de influência sobre cada coeficiente à medida em que a distância aumenta.

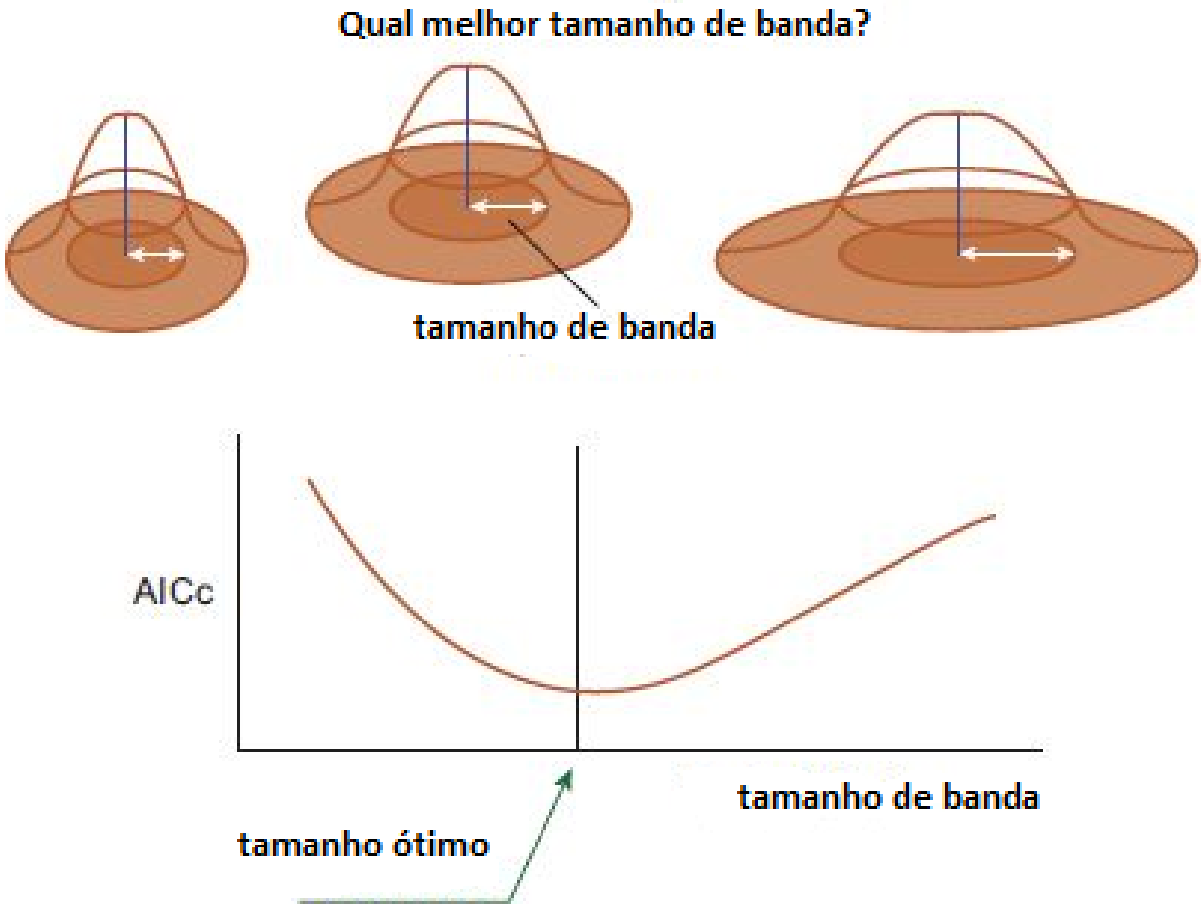
Figura 2 – Estrutura do GWR com variáveis independentes resultando em uma variável dependente (taxa de hospitalizado por SRAG Covid-19).



Conforme preceitua (Nakaya, 2015), o tamanho do *kernel* regula a complexidade do modelo GWR uma vez que controla o grau de variabilidade dos coeficientes estimados. Um tamanho de *kernel* muito pequeno implica na repetição do encaixe do modelo local para um subconjunto de áreas muito pequenas, ocasionando que as estimativas dos coeficientes sejam boas, porém não confiáveis, pois apresentariam alta variância devido à falta de graus de liberdade do modelo local. Por outro lado, quando utiliza-se uma janela muito grande, faz com que sejam ignoradas variações espaciais importantes nos coeficientes. Logo, um modelo com janela larga produz estimativas enviesadas. Por fim, deve-se optar por um compromisso entre o quão bom haja o encaixe da função e os graus de liberdade.

Dentre as sugestões de soluções para este problema, o estado da arte elenca alguns modelos de seleção de critérios como o a validação cruzada (*cross-validation* ou CV), critério de informação de Akaike (AIC) ou AIC corrigido para uma pequena amostra (AICc). O conceito é ilustrado na figura abaixo.

Figura 3 – Decidindo qual o tamanho ótimo do kernel.



O presente trabalho julgou com que AICc proveu melhores resultados para o caso em estudo. Sua fórmula é dada por:

$$\text{AICC} = \text{AIC} + \frac{2p(p+1)}{N-p-1}$$

Em que:

$$\text{AIC} = 2k - 2\ ln(\hat{L})$$

De forma que *k* é o número estimado de parâmetros do modelo, $\hat{L}$ é o máximo valor da função de probabilidade do modelo, *p* é o número de parâmetros e *N* é o tamanho da amostra.

Por fim, a formulação do modelo GWR é dada por:

$$y(g) = \beta_0 + \beta_1(g)x_1 + \beta_2(g)x_2 + ... + \beta_p(g)x_p + \varepsilon$$

que é explicitada por (Francisco, 2010, 123):

> onde *g* é um vetor dos *n* pontos, no espaço bidimensional, os parâmetros do vetor β(g) são específicos para cada observação i de localização $g_i = (u_i, v_i)$ e o termo de erro ε é suposto independente e de comportamento $\varepsilon \sim$ N(0,$\sigma^2 I$) . Temos, na realidade, um conjunto de *n* regressões diferentes, uma para cada ponto g_i do espaço.

3.2.2 *Spatial Autoregressive Model* (SAR)

O Modelo SAR, diferentemente do GWR, é um modelo global, logo seus indicadores são sumarizados para a região estuda inteira. Seus parâmetros não são mapeáveis e, consequentemente, inadequados para visualização. Destaca similaridades no espaço, ou seja, busca regularidades e não é adequado para não estacionariedade espacial.

"O modelo SAR puroinforma que a variável dependente y é influenciada por tal variável dependente, observada nas regiões vizinhas (W)"(Almeida, 2012).

O Modelo é representado pela seguinte equação:

y = W + X + ε

em que,

y = como em qualquer regressão, é a variável dependente, mas neste caso, um vetor. Neste estudo, é a taxa por 1000 habitantes de contaminados graves durante o 2º período da amostragem (jul20 a jan21);

ρ = é o coeficiente autoregressivo espacial. Ele pode variar entre -1 e 1;

Wy = vetor de defasagem espacial para a variável dependente (taxa por 1000 habitantes de contaminados graves durante o 2º período da amostragem (jul20 a jan21));

X = variável(is) independente(s);

β = coeficiente(s) da regressão. É o peso de cada variável independente no Modelo;

ε = é o erro, ou seja, tudo o que não pode ser explicado pelo Modelo.

4 Resultados e Discussão

4.1 Resultados

4.1.1 Base de Dados

4.1.1.1 Características Gerais

A granularidade da base de dados utilizada neste trabalho tem como unidade as informações por município. Mais especificamente, as informações estão distribuídas em 5.570 municípios brasileiros (o escopo deste trabalho se restringe somente ao Brasil).

Em virtude de ter sido usada uma base do censo de 2010 e pelo fato de, nos últimos 10 anos, 05 municípios brasileiros terem se desmembrado em outros, este ajuste se fez necessário na base de dados. Sendo assim, para as variáveis Valor Renda per capita, IVS e IDH os dados destes 05 municípios foram replicados para os outros 05 desmembrados.

Adicionalmente, em virtude de problemas ocorridos no uso de dados referente a municípios sem vizinhos (ilhas brasileiras), no Modelo SAR, os mesmos foram retirados da amostra.

Figura 4 – Municípios desmembrados e excluídos

4.1.1.2 IBGE – Instituto Brasileiro de Geografia e Estatística

Foram utilizadas informações do Censo Demográfico do IBGE de 2010. Por meio desta fonte foi possível obter várias informações econômicas e sócio-demográficas dos municípios brasileiros, entre as quais podemos destacar Densidade demográfica e Valor do rendimento nominal médio mensal das pessoas de 10 anos ou mais de idade (com e sem rendimento).

4.1.1.3 IPEA – Instituto de Pesquisa Econômica Aplicada

Para obter as informações sobre o Índice de Desenvolvimento Humano (IDH) e Índice de Vulnerabilidade Social (IVS) dos municípios brasileiros, foram utilizadas bases de dados a partir do site do Instituto de Pesquisa Econômica Aplicada (IPEA). Foram utilizadas as seguintes dimensões destes índices:

- **IDH** – Educação, Renda e Longevidade.
- **IVS** – Infraestrutura Urbana, Capital Humano, Renda e Trabalho.

4.1.1.4 DATASUS – Departamento de Informática do SUS

Para cada município brasileiro, foram obtidas as taxas de casos graves de contaminação por Covid-19. Obs.: por casos graves, entendam-se aqueles que evoluíram para o quadro de SRAG (Síndrome Respiratória Aguda Grave).

Adicionalmente, para fins comparativos, foram obtidas taxas de diferentes períodos da pandemia no Brasil, o que será mais bem detalhado na seção 4.1.2.

4.1.1.5 TSE – Tribunal Superior Eleitoral

Foram obtidas também informações sobre as taxas dos votos, em cada município, referentes ao 2º turno das eleições presidenciais do Brasil em 2018. Obs.: o motivo da utilização desta base de dados é em virtude de algumas matérias, publicadas por veículos de comunicação, alegarem que, um dos fatores que contribuem para o aumento de casos de Covid-19 no Brasil, é a influência do atual presidente do país (Jair Bolsonaro) sobre seus apoiadores, estimulando-os ao não seguimento de certos protocolos para a prevenção da doença.

4.1.1.6 Agrupamento dos Dados

Por fim, todos esses dados foram reunidos em um arquivo de dados em formato .csv. Por meio do código do município do IBGE foi realizado um *join* com a *layer* de municípios também do IBGE criando um arquivo no formato *shapefile*, para que pudessem ser aplicados em modelos de regressão diferentes, conforme detalhado mais adiante.

4.1.2 Seleção das Variáveis

O processo de criação do Modelo de Regressão foi iniciado a partir do uso de 12 (doze) variáveis independentes. Dentre estas variáveis, foram selecionadas aquelas que atenderam

determinados requisitos, a partir da aplicação de algumas técnicas, conforme detalhado a seguir. Obs.: estes procedimentos foram realizados por meio do uso de script em linguagem R (R-Studio).

4.1.2.1 Variáveis Iniciais

Inicialmente, foram escolhidas as seguintes variáveis para compor o Modelo de Regressão:

4.1.2.1.1 Variável Dependente

- **Taxa de Casos Graves 2º Período**: refere-se à taxa da população que contraiu Covid-19 e evoluiu para o quadro de SRAG (Síndrome Respiratória Aguda Grave), no período de julho/2020 a jan/2021.

4.1.2.1.2 Variável Independente

- **Taxa de Casos Graves 1º Período**: refere-se à taxa da população que contraiu Covid-19 e evoluiu para o quadro de SRAG (Síndrome Respiratória Aguda Grave), no período de fevereiro/2020 a junho/2020.
- **Taxa de Votos Jair Bolsonaro**: refere-se à taxa de votos válidos para o então candidato a presidência Jair Bolsonaro, no 2º turno das eleições presidenciais de 2018.
- **Valor Renda**: valor da renda per capita de cada município brasileiro.
- **IDHM Educação (Índice de Desenvolvimento Humano Municipal de Educação)**: refere-se à quantidade média de anos de estudo da população de cada município.
- **IDHM Renda (Índice de Desenvolvimento Humano Municipal de Educação)**: mede-se o valor médio do rendimento dos cidadãos de cada município, com base na média do Produto Interno Bruto (PIB).
- **IDHM Longevidade**: refere-se ao número médio de anos que as pessoas viveriam a partir do nascimento, mantidos os mesmos padrões de mortalidade observados no ano de referência.
- **IVS Infraestrutura Urbana**: dimensão do Índice de Vulnerabilidade Social que procura refletir as condições de acesso a serviços de saneamento básico e de mobilidade urbana.

- **IVS Capital Humano**: dimensão do Índice de Vulnerabilidade Social que envolve dois tipos de ativos que, de acordo com Schultz (1962), determinam as perspectivas de futuro dos indivíduos: suas condições de saúde e seu acesso a educação.
- **IVS Renda e Trabalho**: dimensão do IVS que envolve o acesso ao trabalho e a forma de inserção (formal ou não) de determinada população.
- **Taxa Populacional Feminina**: refere-se à taxa da população feminina por município.
- **Taxa Populacional de Idosos**: taxa da população referente a indivíduos com idade a partir de 60 anos.
- **Densidade Demográfica:** índice demográfico que calcula o número de habitantes por quilômetro quadrado.

4.1.2.2 Matriz de Correlação

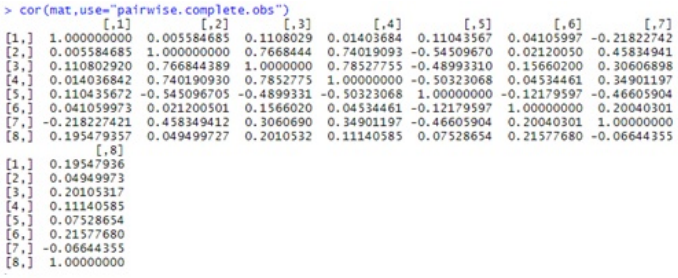
Após a escolha das variáveis independentes, foram verificadas quais apresentam alto nível de correlação entre si. Para isto, aplicaremos o teste da Matriz de Correlação, retirando, gradativamente, as variáveis que apresentem nível de correlação acima de **0,8**.

Figura 5 – Exemplo de aplicação da Matriz de Correlação

```
> cor(mat,use="pairwise.complete.obs")
             [,1]         [,2]       [,3]        [,4]       [,5]       [,6]       [,7]
 [1,]  1.000000000  0.005584685  0.1108029  0.09576383  0.07465244  0.01403684  0.11043567
 [2,]  0.005584685  1.000000000  0.7668444  0.68741546  0.82227491  0.74019093 -0.54509670
 [3,]  0.110802920  0.766844389  1.0000000  0.79053869  0.93449440  0.78527755 -0.48993310
 [4,]  0.095763831  0.687415460  0.7905387  1.00000000  0.81902063  0.70400677 -0.46429936
 [5,]  0.074652444  0.822274911  0.9344944  0.81902063  1.00000000  0.83382632 -0.56688866
 [6,]  0.014036842  0.740190930  0.7852775  0.70400677  0.83382632  1.00000000 -0.50323068
 [7,]  0.110435672 -0.545096705 -0.4899331 -0.46429936 -0.56688866 -0.50323068  1.00000000
 [8,]  0.009759426 -0.800732393 -0.8172276 -0.86431304 -0.87498343 -0.80839150  0.59200856
 [9,] -0.059981111 -0.796709159 -0.8536784 -0.80525427 -0.89822627 -0.79298597  0.52473639
[10,]  0.041059973  0.021200501  0.1566020  0.23793722  0.14507223  0.04534461 -0.12179597
[11,] -0.218227421  0.458349412  0.3060690  0.32464837  0.37207141  0.34901197 -0.46605904
[12,]  0.195479357  0.049499727  0.2010532  0.18638976  0.16107898  0.11140585  0.07528654
             [,8]         [,9]       [,10]       [,11]       [,12]
 [1,]  0.009759426 -0.059981111  0.04105997 -0.21822742  0.19547936
 [2,] -0.800732393 -0.796709159  0.02120050  0.45834941  0.04949973
 [3,] -0.817227572 -0.853678431  0.15660200  0.30606898  0.20105317
 [4,] -0.864313039 -0.805254271  0.23793722  0.32464837  0.18638976
 [5,] -0.874983428 -0.898226271  0.14507223  0.37207141  0.16107898
 [6,] -0.808391497 -0.792985971  0.04534461  0.34901197  0.11140585
 [7,]  0.592008562  0.524736391 -0.12179597 -0.46605904  0.07528654
 [8,]  1.000000000  0.843765123 -0.13792052 -0.50503279 -0.12457065
 [9,]  0.843765123  1.000000000 -0.08701945 -0.29995617 -0.13251163
[10,] -0.137920519 -0.087019451  1.00000000  0.20040301  0.21577680
[11,] -0.505032787 -0.299956171  0.20040301  1.00000000 -0.06644355
[12,] -0.124570651 -0.132511631  0.21577680 -0.06644355  1.00000000
```

Após sucessivos refinamentos, permaneceram as variáveis **Taxa de Casos Graves 1º período**, **Taxa de Votos Jair Bolsonaro**, **Valor Renda**, **IDHM Longevidade, IVS Infraestrutura Urbana**, **Taxa Populacional Feminina Taxa Populacional de Idosos** e **Densidade Demográfica**.

Figura 6 – Resultado Final da aplicação da Matriz de Correlação

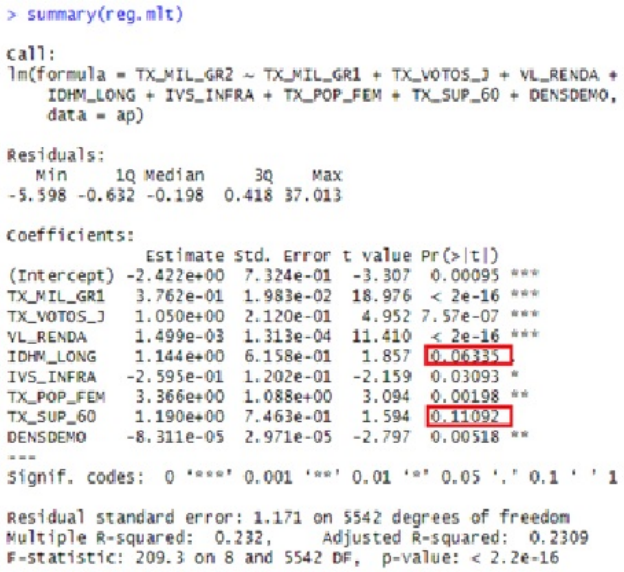


```
> cor(mat,use="pairwise.complete.obs")
             [,1]         [,2]       [,3]        [,4]        [,5]        [,6]        [,7]
[1,]  1.000000000  0.005584685  0.1108029  0.01403684  0.11043567  0.04105997 -0.21822742
[2,]  0.005584685  1.000000000  0.7668444  0.74019093 -0.54509670  0.02120050  0.45834941
[3,]  0.110802920  0.766844389  1.0000000  0.78527755 -0.48993310  0.15660200  0.30606898
[4,]  0.014036842  0.740190930  0.7852775  1.00000000 -0.50323068  0.04534461  0.34901197
[5,]  0.110435672 -0.545096705 -0.4899331 -0.50323068  1.00000000 -0.12179597 -0.46605904
[6,]  0.041059973  0.021200501  0.1566020  0.04534461 -0.12179597  1.00000000  0.20040301
[7,] -0.218227421  0.458349412  0.3060690  0.34901197 -0.46605904  0.20040301  1.00000000
[8,]  0.195479357  0.049499727  0.2010532  0.11140585  0.07528654  0.21577680 -0.06644355
             [,8]
[1,]  0.19547936
[2,]  0.04949973
[3,]  0.20105317
[4,]  0.11140585
[5,]  0.07528654
[6,]  0.21577680
[7,] -0.06644355
[8,]  1.00000000
```

4.1.2.3 Teste de Hipótese (*p-value*)

Após o teste de Matriz de Correlação, foi construído um Modelo de Regressão Linear com as variáveis selecionadas e verificado o *p-value* de cada uma delas. O critério escolhido foi manter somente as variáveis com *p-value* abaixo de ***0,05***.

Figura 7 – Variáveis com *p-value* acima de 0,05.



```
> summary(reg.mlt)

Call:
lm(formula = TX_MIL_GR2 ~ TX_MIL_GR1 + TX_VOTOS_J + VL_RENDA +
    IDHM_LONG + IVS_INFRA + TX_POP_FEM + TX_SUP_60 + DENSDEMO,
    data = ap)

Residuals:
   Min     1Q Median     3Q    Max
-5.598 -0.632 -0.198  0.418 37.013

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.422e+00  7.324e-01  -3.307  0.00095 ***
TX_MIL_GR1   3.762e-01  1.983e-02  18.976  < 2e-16 ***
TX_VOTOS_J   1.050e+00  2.120e-01   4.952 7.57e-07 ***
VL_RENDA     1.499e-03  1.313e-04  11.410  < 2e-16 ***
IDHM_LONG    1.144e+00  6.158e-01   1.857  0.06335 .
IVS_INFRA   -2.595e-01  1.202e-01  -2.159  0.03093 *
TX_POP_FEM   3.366e+00  1.088e+00   3.094  0.00198 **
TX_SUP_60    1.190e+00  7.463e-01   1.594  0.11092
DENSDEMO    -8.311e-05  2.971e-05  -2.797  0.00518 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.171 on 5542 degrees of freedom
Multiple R-squared:  0.232,     Adjusted R-squared:  0.2309
F-statistic: 209.3 on 8 and 5542 DF,  p-value: < 2.2e-16
```

Foi verificado que as variáveis **IDHM Longevidade** e **Taxa Populacional de Idosos** apresentaram *p-value* igual 0,06335 e 0,11092, respectivamente. Portanto, foram eliminadas do modelo.

4.1.2.4 Análise do Coeficiente das Variáveis

Foi identificado também que o coeficiente da variável independente **Densidade Demográfica** apresentou um coeficiente negativo.

Figura 8 – Coeficiente da variável Densidade Demográfica

```
> summary(reg.mlt2)

Call:
lm(formula = TX_MIL_GR2 ~ TX_MIL_GR1 + TX_VOTOS_J + VL_RENDA +
    IVS_INFRA + TX_POP_FEM + DENSDEMO, data = bp)

Residuals:
   Min     1Q Median     3Q    Max
-1.571 -0.627 -0.202  0.417 37.006

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.588e+00  5.277e-01  -3.010 0.002627 **
TX_MIL_GR1   3.675e-01  1.919e-02  18.954  < 2e-16 ***
TX_VOTOS_J   1.258e+00  1.947e-01   6.459 1.14e-10 ***
VL_RENDA     1.596e-03  1.118e-04  14.018  < 2e-16 ***
IVS_INFRA   -3.351e-01  1.117e-01  -2.897 0.003787 **
TX_POP_FEM   3.580e+00  1.053e+00   3.401 0.000677 ***
DENSDEMO    -8.413e-05  2.967e-05  -2.836 0.004592 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.172 on 5544 degrees of freedom
Multiple R-squared:  0.2311,    Adjusted R-squared:  0.2303
F-statistic: 277.8 on 5 and 5544 DF,  p-value: < 2.2e-16
```

Isto faria com que, neste modelo de regressão, quanto maior a densidade demográfica menor a **Taxa de Casos Graves 2º Período** (variável dependente). Entretanto, isto se apresenta com sendo uma **correlação espúria** [1], uma vez que, de acordo com dados divulgados pela imprensa (como veremos mais adiante), as cidades brasileiras com maior densidade demográfica são aquelas que mais possuem maiores taxas de casos graves de Covid-19. Por este motivo, esta variável foi retirada do modelo.

[1] Relação estatística existente entre duas variáveis, onde não existe nenhuma relação causa-efeito entre elas (pt.wikipedia.org).

Figura 9 – Modelo de Regressão com a remoção da variável Densidade Demográfica

```
> summary(reg.mlt3)

Call:
lm(formula = TX_MIL_GR2 ~ TX_MIL_GR1 + TX_VOTOS_J + VL_RENDA +
    IVS_INFRA + TX_POP_FEM, data = ap)

Residuals:
   Min     1Q Median     3Q    Max
-5.449 -0.626 -0.204  0.414 37.064

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.2787525  0.5166078  -2.475 0.013343 *
TX_MIL_GR1   0.3600175  0.0192206  18.731  < 2e-16 ***
TX_VOTOS_J   1.2963204  0.1943864   6.669 2.83e-11 ***
VL_RENDA     0.0015261  0.0001112  13.720  < 2e-16 ***
IVS_INFRA   -0.3883422  0.1142197  -3.400 0.000679 ***
TX_POP_FEM   3.0169831  1.0344503   2.917 0.003554 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.172 on 5545 degrees of freedom
Multiple R-squared:   0.23,     Adjusted R-squared:  0.2293
F-statistic: 331.3 on 5 and 5545 DF,  p-value: < 2.2e-16
```

4.1.2.5 Multicolinearidade

Por fim, foram verificadas as taxa de multicolinearidade das variáveis. Para realizar esta medição, foi utilizado o **fator de inflação da variância** (**VIF**), que avalia o quanto a variância de um coeficiente de regressão estimado aumenta se as suas preditoras estiverem correlacionadas. Como critério de corte, seriam eliminadas todas as variáveis que apresentassem o VIF com valor acima de **0,5**.

Após a aplicação do teste, foi constatado que todas as variáveis apresentaram VIF abaixo de 0,5, conforme apresentado na tabela abaixo. Deste modo, não foi necessária a eliminação de mais variáveis.

Tabela 1 – VIF das variáveis do Modelo de Regressão Linear

Taxa Casos 1º Periodo	**Taxa Votos Bolsonaro**	**Valor Renda (per capita)**	**IVS Infaes-trutura**	**Taxa População Feminina**
1,1	2,8	2,7	1,5	1,1

4.1.3 Análise Geoespacial

Como demonstrado por Anselin (1995), Indicador Local para Associação Espacial - LISA (Local Indicator for Sapatial Association), permite a decomposição de indicadores globais

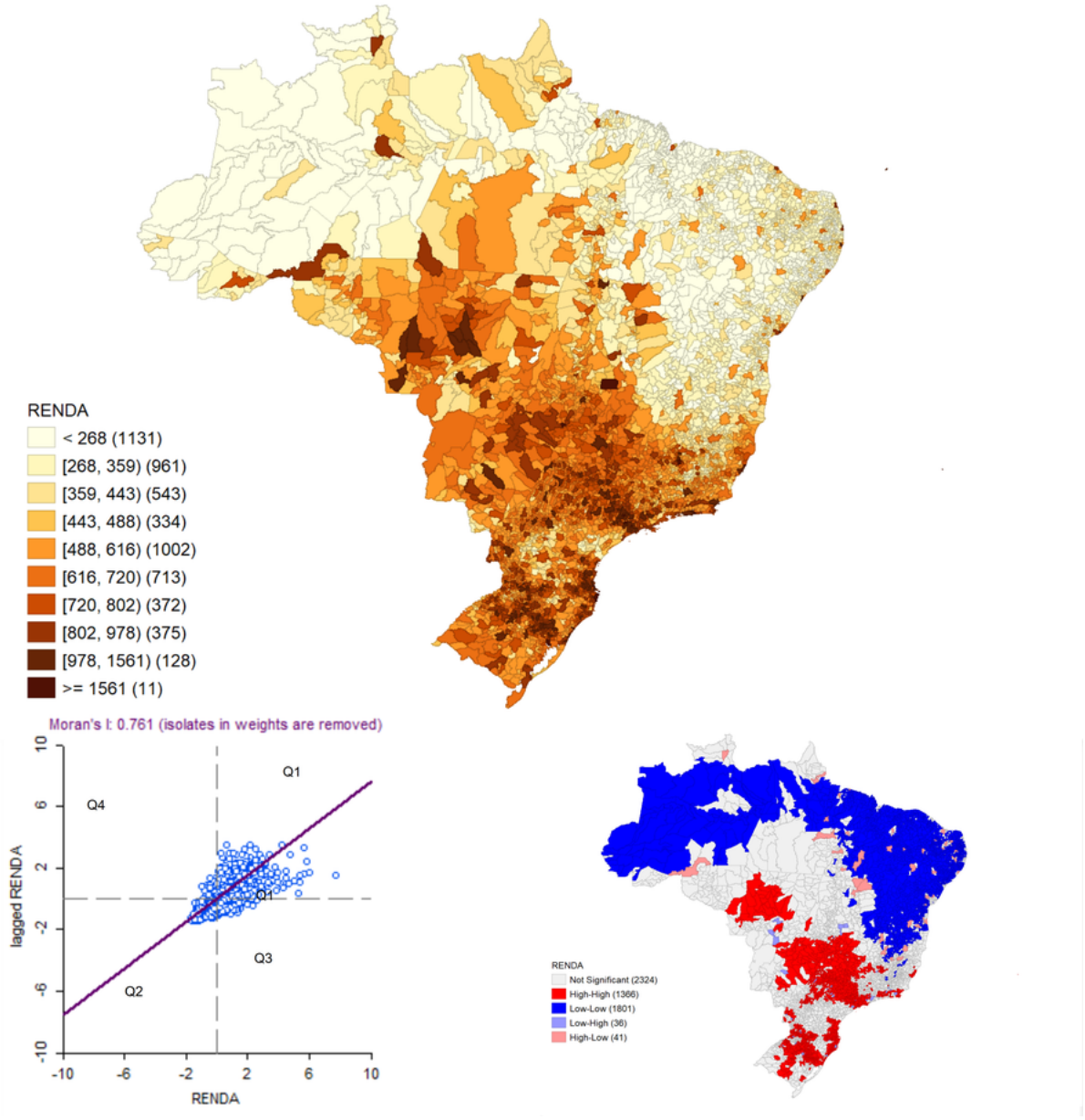
em indicadores locais possibilitando a identificação de *clusters* regionais e/ou *hotspots* por semelhança de valor, baseado em nível de significância estatística. Além disso, a soma dos valores das autocorrelações locais são diretamente proporcionais ao indicador global da associação espacial.

Para fazer a análise de dependência espacial das variáveis usamos o I de Moran Local e o Mapa de Cluster LISA associado, estatística mais comumente aceita e usada. Para fazer tal análise usamos uma matriz de vizinhança W do tipo Queen´s de ordem 1.

Para a variável VALOR RENDA, apresentada na Figura10 , usamos método de classificação de Quebras Naturais de Jenks (*Jenks Natural Breaks*) em 10 quebras. No mapa é possível observar a nítida diferença na distribuição da capacidade de renda dos municípios. Nota-se uma maior renda nas regiões Centro-Oeste, Sudeste e Sul e de menor renda nas regiões Norte e Nordeste.

Na análise do índice I de Moran Local, o gráfico de dispersão demonstra forte correlação espacial e o Mapa de Clusters LISA aponta as regiões em que há força. Em azul, o *cluster low-low* - valores baixos da variável vizinhos de valores baixos, em grande parte das regiões Norte e Nordeste, estão representado os pontos contidos no quadrante Q2 do gráfico de dispersão de Moran. Os três grandes focos em vermelho nas regiões Centro-Oeste Sudeste e Sul participantes do *cluster high-high* - valores altos da variável vizinhos de valores altos, representados no quadrante Q1 do gráfico de dispersão da Figura10 . Esses dois quadrantes determinam uma relação direta ou positiva na associação espacial. Os demais quadrantes Q3 e Q4 representam vizinhos diferentes com relação espacial negativa e/ou distinta.

Figura 10 – Distribuição espacial de VALOR RENDA (per capta), escala em quebras naturais de Jenks (*Natural Breaks),* Gráfico de dispersão do I de Moran Local e mapa LISA para a variável VALOR RENDA.



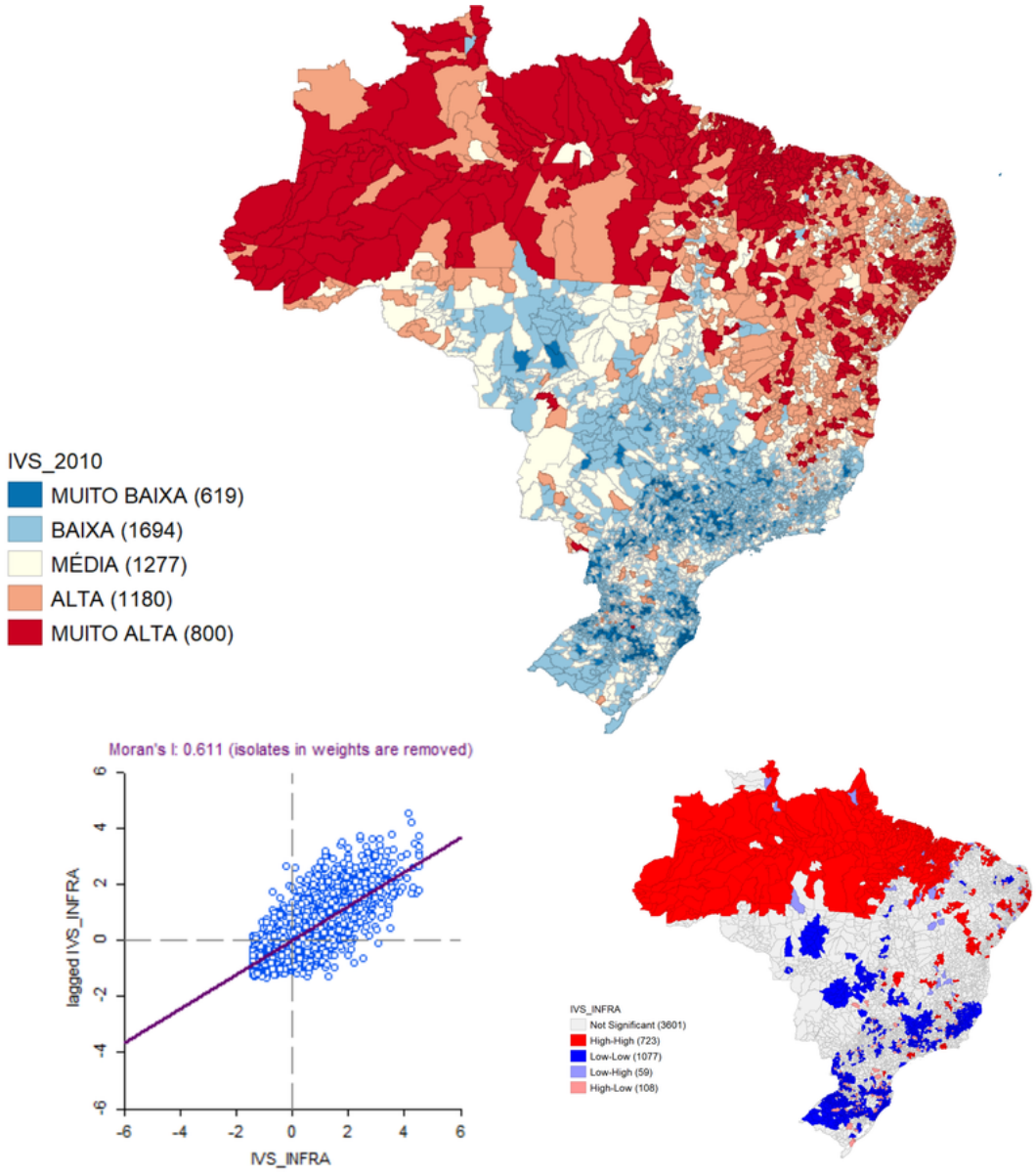
Fonte: IBGE, Gráficos de elaboração própria dos autores.

A variável IVS INFRAESTRUTURA, na mesma linha da variável renda, demonstra a fragilidade e a falta de desenvolvimento das regiões mais pobres do país, novamente dividindo o Brasil em dois clusters. Na Figura 11 nota-se as faixas do índice gerado pelo IPEA – com a escala em faixas como sugerida pela instituição. As regiões de menor renda geralmente têm menor nível de desenvolvimento e infraestrutura.

No Mapa de Cluster verificamos a região Norte e parte da região Nordeste como vizinhos

de alto índice na variável e alguns pontos de baixo índice nas demais regiões. No gráfico de dispersão, o I de Moran de 61,1% aponta a boa autocorrelação espacial direta para esta variável.

Figura 11 – IVS_INFRA dos municípios em faixas,
Gráfico de dispersão e I de Moran e Mapa de Cluster LISA para IVS INFRAESTRUTURA.



Fonte: IPEA, Gráficos de elaboração própria dos autores.

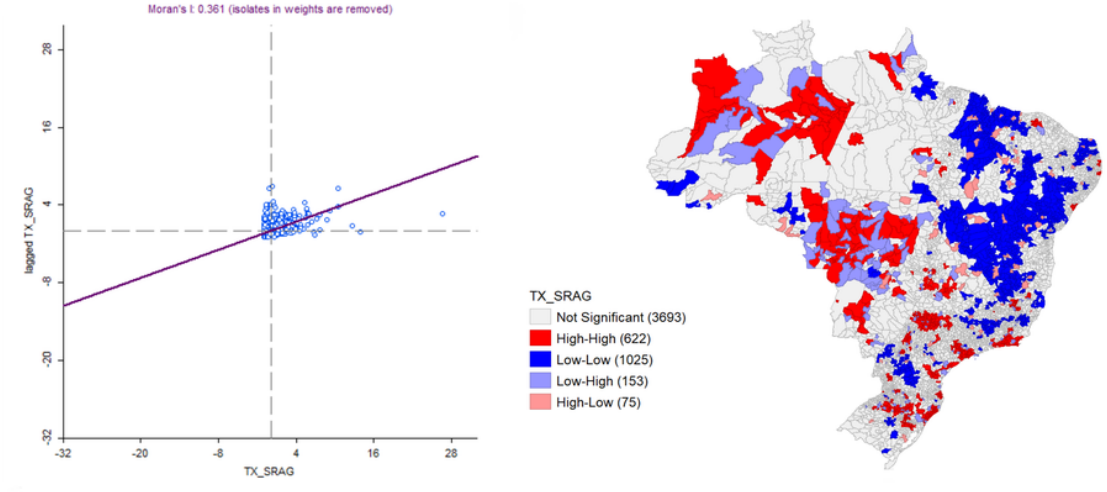
Figura 12 – Mapas de casos de SRAG-COVID-19 em casos/mil habitantes. Evolução: fev-jun/20, jul-dez/20 e total, respectivamente.

Fonte: DATASUS e IBGE, Gráficos de elaboração própria dos autores.

Os casos de SRAG provocados pela Covid-19, da variável TAXA DE CASOS GRAVES, estão colocados em três mapas na Figura12 , classificados em 10 escalas de Natural Breaks, considerando os dois períodos definidos neste trabalho, de fevereiro de 2020 a junho de 2020 e de julho de 2020 a dezembro de 2020, e o total de casos ou o período integral de fevereiro a dezembro de 2020. Nota-se a evolução geográfica da doença e um curioso desempenho da região amazônica, caso não abordado neste trabalho.

O I de Moran de 36,1% assim como Mapa de Cluster LISA indica alguma correlação espacial para a variável – considerando o total de casos, mesmo que mais dispersa, Figura13 . O mapa aponta alguns pequenos focos de high-high em todas as regiões com a exceção da Nordeste, que apresenta uma mais homogênea autocorrelação em low-low. Esses focos em vermelho retratam os grandes centros urbanos e regiões metropolitanas enquanto que a massa maior em azul representa o semiárido da região Nordeste.

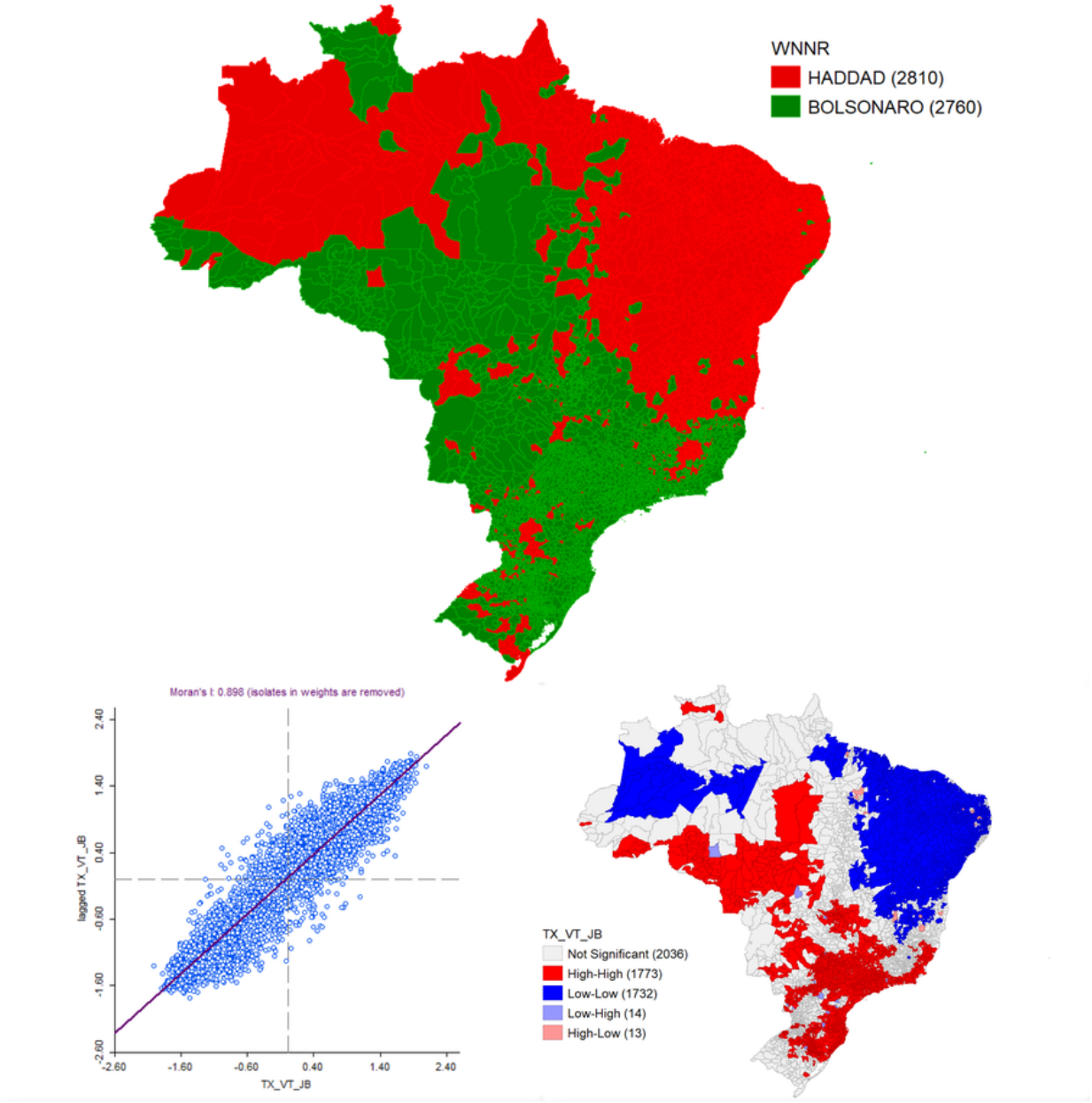
Figura 13 – I de Moran Local e Mapa de Clusters para a variável TAXA DE CASOS GRAVES



Fonte: DATASUS e IBGE, Gráficos de elaboração própria dos autores.

Já para a variável TAXA DE VOTOS JAIR BOLSONARO, a taxa de votos do então candidato Jair Bolsonaro - variável eleitoral extraída dos dados das Eleições Presidenciais de 2018 – 2º turno, dados do TSE; a forte relação geográfica se apresenta nitidamente no mapa da Figura14 que preconiza o vencedor no município no 2º turno das eleições presidenciais de 2018, independente da diferença entre primeiro e segundo colocado. Como já era de se esperar, observa-se um melhor desempenho do então candidato Jair Bolsonaro nas regiões em que outras variáveis apresentam melhores índices como VALOR RENDA e IVS INFRAESTRUTURA. Da mesma forma, o então candidato Fernando Haddad teve predominância nas regiões mais carentes do Estado brasileiro com menores índices de desempenho nas variáveis VALOR RENDA e IVS INFRAESTRUTURA. Esse fenômeno é bem observado no I de Moran da variável TAXA DE VOTOS JAIR BOLSONARO, que apresentou valor de 89,8% indicando forte correlação geoespacial e no Mapa de Cluster LISA, em que se pode observar os clusters bem definidos.

Figura 14 – Mapa dos vencedores – por município, do 2º turno das eleições presidenciais de 2018. I de Moran Local da Variável TAXA DE VOTOS JAIR BOLSONARO e o Mapa de Cluster LISA.

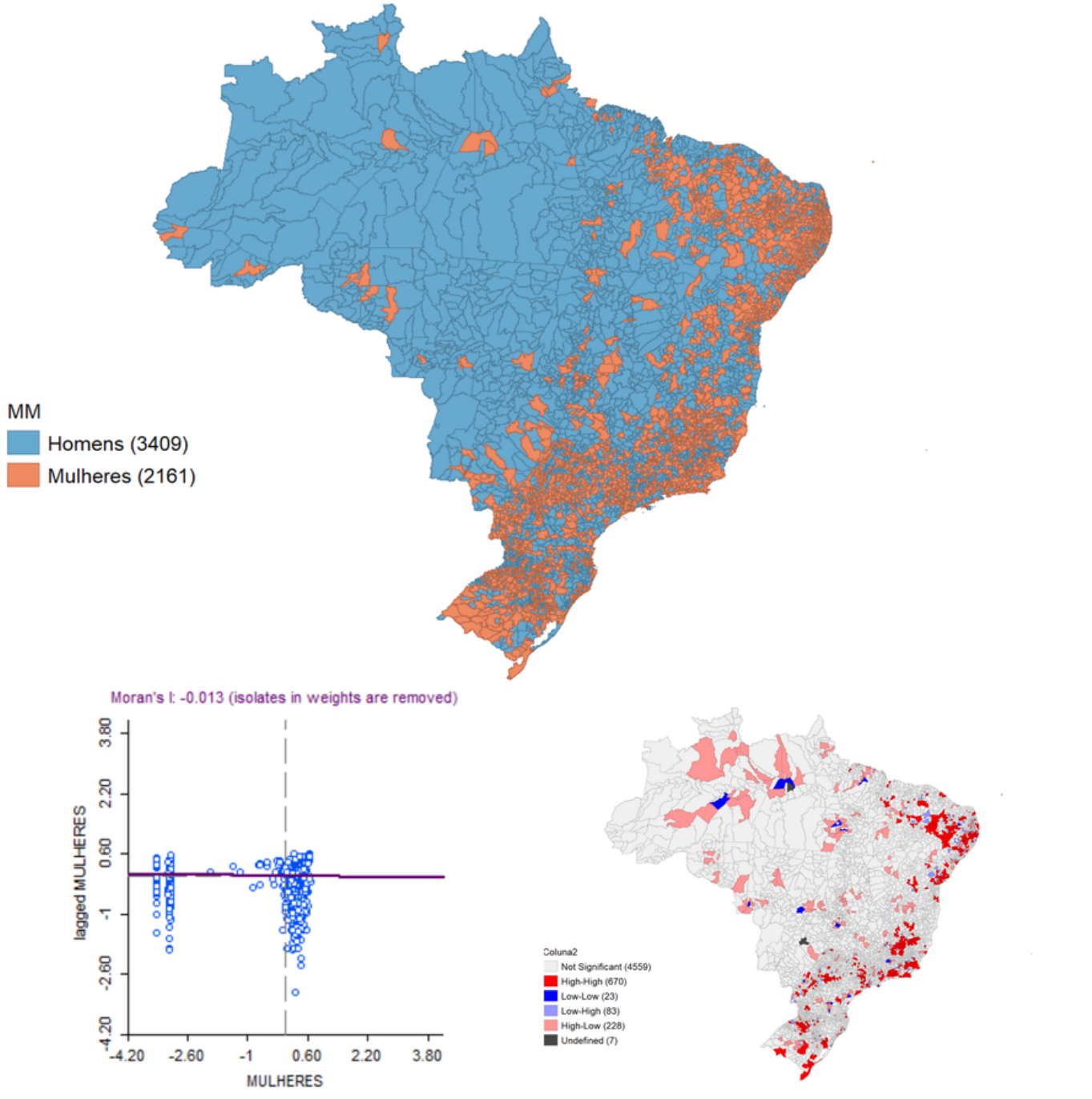


Fonte: TSE e IBGE, Gráficos de elaboração própria dos autores.

A última variável selecionada para nosso estudo representa a relação entre o número de mulheres e homens do município. O mapa central da Figura 15 apresenta os municípios que têm maior número de mulheres ou de homens. Verifica-se uma distribuição majoritária de mulheres nas regiões Sul, Sudeste e Nordeste ou mais consistentemente na metade leste do país e a predominância de homens no interior do Brasil. A análise o I de Moran indica pouca ou

nenhuma correlação espacial para esta variável.

Figura 15 – Mapa de municípios com predominância de mulheres ou homens. I e Moran local e mapa de clusters LISA para a variável MULHERES



Fonte: IBGE, Gráficos de elaboração própria dos autores.

4.1.4 Elaboração dos Modelos de Regressão Linear e Espacial

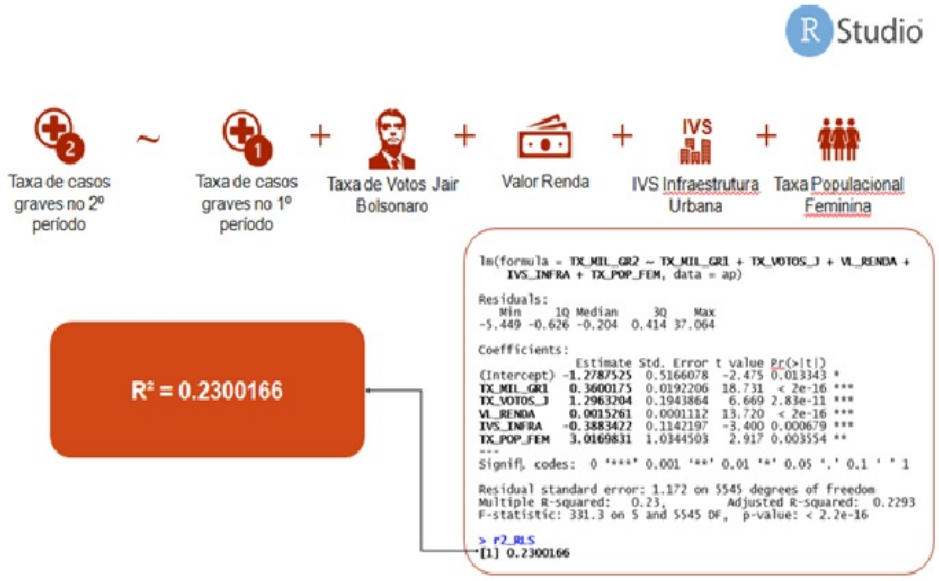
Após a seleção das variáveis, para fins de comparação, foram criados: **Modelo de Regressão Linear**, **Modelo Espacial Autorregressivo (SAR)** e **Modelo de Regressão Ponderada Geográfica (GWR)**, conforme detalhado a seguir. Obs.: os modelos foram gerados por meio de funções da linguagem R (R-Studio) e da ferramenta GeoDA.

4.1.4.1 Modelo de Regressão Linear

Este modelo apresentou um R^2 com o valor de 0.23, indicando que 23% da variação média da taxa por 1000 habitantes de contaminados graves durante o 2º período da amostragem (jul20 a jan21) é explicada pelo mesmo.

É possível notar que todas as variáveis deste modelo possuem coeficiente positivo, com exceção do **IVS de Infraestrutura Urbana**. Adicionalmente, notamos que a variável **Taxa Populacional Feminina** é a que apresenta o coeficiente de maior valor (3.0169).

Figura 16 – Modelo de Regressão Linear



4.1.4.2 Modelo Espacial Autorregressivo (SAR)

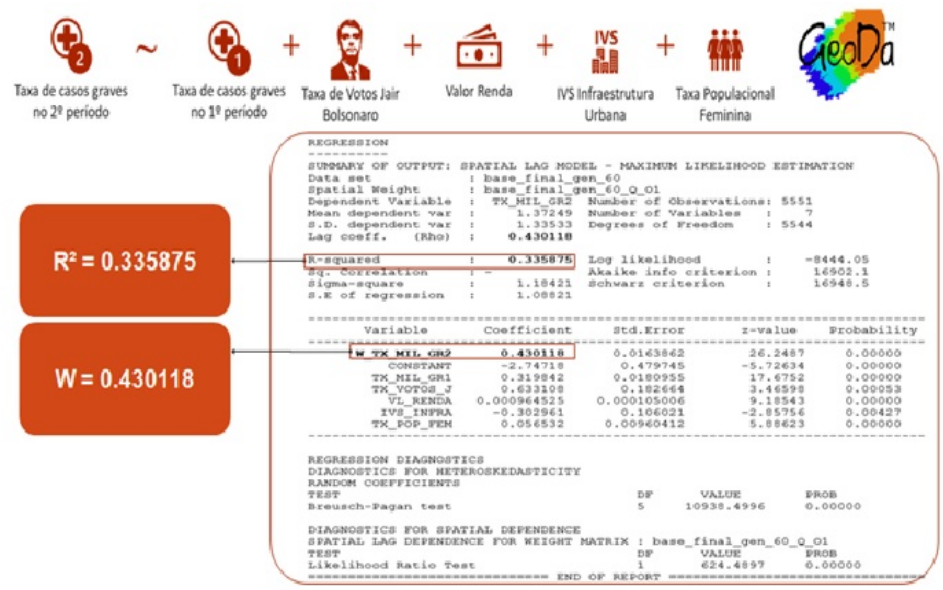
Depois de utilizadas as mesmas variáveis em Modelo SAR (*Spatial Autoregressive Lag Model*), este apresentou um **R^2 com o valor de 0.3358**, indicando que 33.58% da variação média da taxa por 1000 habitantes de contaminados graves durante o 2º período da amostragem (julho/2020 a janeiro/2021) é explicada pelo mesmo. Este valor mais elevado de R^2, quando

comparado ao Modelo de Regressão Linear, indica uma influência espacial na explicação da variação da taxa de contaminações graves por covid entre os municípios brasileiros.

Para este modelo foi utilizada uma **Matriz de Vizinhança de Ordem 01 do Tipo Queen** (a que forneceu ao modelo o melhor R^2), sendo o valor do **Coeficiente Espacial Autoregressivo (Lag Coeffcient)** igual a **0.4301**.

Vale destacar que o sinal aritmético dos coeficientes permaneceu idêntico aos do Modelo de Regressão Linear. Entretanto, a variável independente que passou a ter o **maior valor de coeficiente** passou a ser a **Taxa de Votos Jair Bolsonaro (0.6331).**

Figura 17 – Modelo Espacial Autoregressivo (SAR)



4.1.4.3 Modelo de Regressão Ponderada Geográfica (GWR)

Por fim, foi criado um Modelo GWR, utilizando as mesmas variáveis dos modelos anteriores. Este apresentou um **R^2 com o valor de 0.5318**, indicando que 53.18% da variação média da taxa por 1000 habitantes de contaminados graves durante o 2º período da amostragem (julho/2020 a janeiro/2021) é explicado por este modelo. Na figura abaixo, é apresentada uma comparação entre os valores do R^2 de cada modelo:

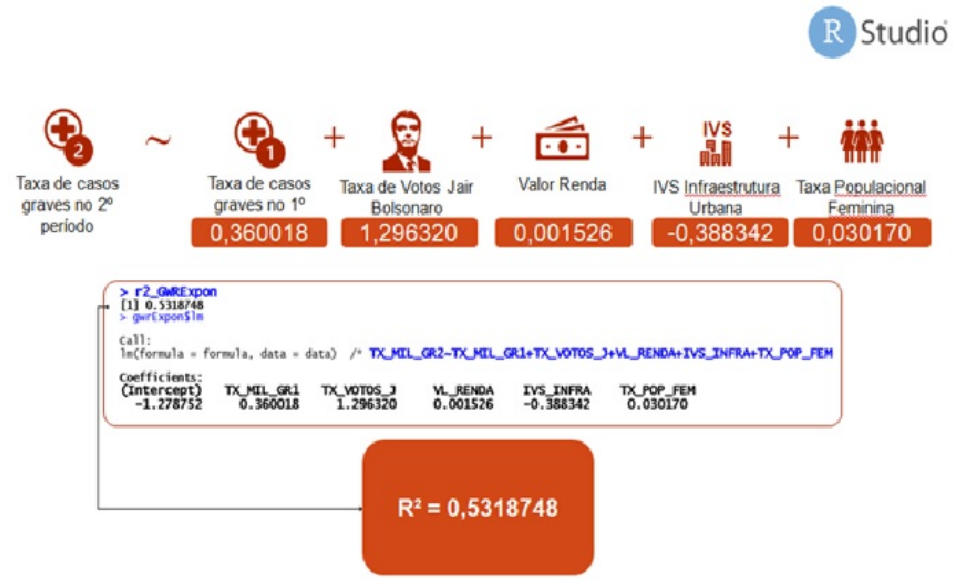
Tabela 2 – Comparação entre os R² dos Modelos de Regressão

Modelo de Regressão	**Valor R²**	**Valor R² (%)**
Modelo Regressão Linear Simples	0.2300166	23,00
Modelo Espacial SAR	0.3358828	33,59
Modelo Espacial GWR (Gaussiana)	0.4968948	49,69
Modelo Espacial GWR (Exponencial)	0.5318748	53,19
Modelo Espacial GWR (Biquadrada)	0.4315257	43,15
Modelo Espacial GWR (*Tricube*)	0.465158	46,52
Modelo Espacial GWR (*Boxcar*)	0.4282822	42,83

Para obtenção deste modelo, foram experimentadas as seguintes funções *Kernel*: Gaussiana, Exponencial, Biquadrada, *Tricube* e *Boxcar*. Aquela que contribuiu para o maior R² foi a **Exponencial**. A métrica de distancia aplicada foi a **Euclidiana**.

O sinal aritmético dos coeficientes permaneceu idêntico aos dos demais modelos. A variável independente, assim como no Modelo SAR, que apresentou o **maior valor de coeficiente** continuou sendo a **Taxa de Votos Jair Bolsonaro (1.2963).**

Figura 18 – Melhor resultado: R^2 do modelo GWR com kernel exponencial



4.2 Discussão

4.2.1 Análise das Variáveis Independentes

Nesta seção, é apresentada uma análise dos coeficientes das variáveis que compõem os modelos de regressão criados. Entretanto, para o escopo desta análise, utilizamos somente o Modelo GWR por ter sido aquele que apresentou o maior R^2.

4.2.1.1 Taxa de Casos Graves 1º Período

O coeficiente desta variável possui um sinal aritmético **positivo**. Isto indica que a taxa de casos graves por município, no período de fevereiro/2020 a junho/2020, influencia na taxa de casos graves do período de julho/2020 a janeiro/2021, em uma **relação proporcional direta**.

Esta variável possui um coeficiente de valor igual a **0.3600**. Isto significa que, a cada **01 ponto** de elevação no valor desta variável, há um **acréscimo** de **0.36** pontos na variável dependente **Taxa de Casos Graves 2º Período**.

4.2.1.2 Taxa de Votos Jair Bolsonaro

O coeficiente desta variável possui um sinal aritmético **positivo**. Isto indica que esta variável também possui uma **relação proporcional direta** com a variável dependente do modelo de regressão.

Esta variável possui um coeficiente de valor igual a **1.2963**. Isto significa que, a cada **01 ponto** de elevação no valor desta variável, há um **acréscimo** de, aproximadamente, **1.30** pontos na variável dependente **Taxa de Casos Graves 2º Período**.

Estes números convergem com o divulgado na pesquisa realizada pela UFRJ em outubro/2020, publicada à época por vários canais de notícias, segundo a qual destaca que *"a COVID-19 tem causado mais impactos nos municípios mais favoráveis ao presidente*" (Correio Braziliense, 13/10/2020)[2].

Vale ressaltar que, devido à impossibilidade de acesso aos dados das pesquisas de aprovação do atual governo (em especial, essas realizadas entre setembro e outubro/2020), neste trabalho foram utilizados os dados do 2º turno das Eleições Presidenciais 2018, para identificação da taxa de votos, por município, para o candidato (e atual presidente) Jair Bolsonaro.

[2] https://www.correiobraziliense.com.br/brasil/2020/10/4881890-pandemia-e-pior-nas-cidades-governadas-por-apoiadores-de-bolsonaro-diz-pesquisa.html

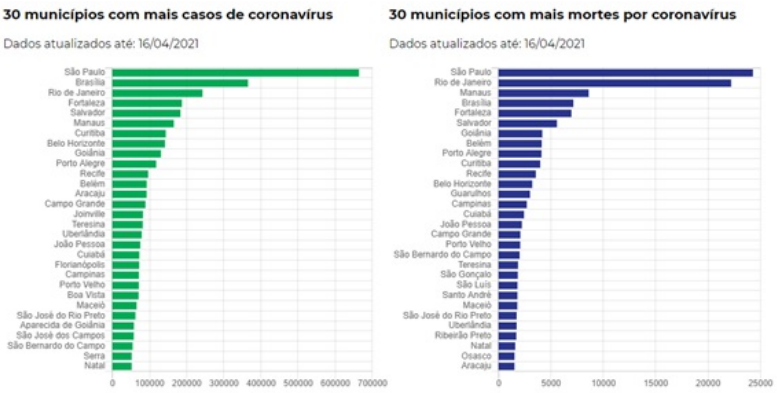
Ressaltamos também que, ao se tentar utilizar a variável **Taxa de Votos Haddad** (concorrente ao cargo de presidente no 2º turno das Eleições 2018), no modelo de regressão, esta apresentou **coeficiente com sinal negativo**, indicando uma coerência lógica já que a mesma possui um "contexto oposto" à variável **Taxa de Votos Jair Bolsonaro**. Obs.: De todo modo, não foi possível fazer uso desta variável já que a mesma apresentou um *p-value* acima de 0.05.

4.2.1.3 Valor Renda (per capita)

O coeficiente desta variável possui sinal aritmético **positivo**, com valor igual a **0.001526**. Isto significa que, a cada **01 ponto** de elevação no valor desta variável, há um **acréscimo** de **0.36** pontos na variável dependente na **Taxa de Casos Graves 2º Período**.

Assim sendo, quanto maior o valor da renda per capita do município, maior será a taxa de casos graves. Isso leva a concluir que as taxas de casos graves são mais elevadas nos grandes centros urbanos do Brasil. De fato, isto converge com noticias que publicadas em canais de comunicação, segundo as quais são, principalmente, nas grandes capitais em que se encontram as maiores taxas de casos de contaminação por Covid-19 [3].

Figura 19 – Municípios brasileiros mais afetados por Covid-19



[3] https://congressoemfoco.uol.com.br/covid19/municipios/index.html (dados computados em 16/04/2021)

4.2.1.4 IVS Infraestrura Urbana

Esta é a única variável do modelo cujo coeficiente possui sinal aritmético **negativo**, tendo valor igual a -**0.388342**. Isto significa que a cada **01 ponto** de elevação no valor desta variável, há um **decréscimo** de **0.39** pontos na variável dependente **Taxa de Casos Graves 2º Período**.

Assim sendo, quanto **maior** índice de vulnerabilidade social do município, no quesito infraestrutura urbana, **menor** será a taxa de casos graves. Em outras palavras, municípios com melhor infraestrutura (conseqüentemente, as principais capitais brasileiras), tendem a apresentar maiores taxas de casos graves. Isto também converge com os dados apresentados na seção anterior.

4.2.1.5 Taxa Populacional Feminina

Esta variável possui sinal aritmético **positivo**, com valor igual a **0.0301**. Isto significa que, a cada **10%** de elevação no valor desta variável, há um **acréscimo** de **03 pontos** na variável dependente na **Taxa de Casos Graves 2º Período**.

Isto coincide com algumas notícias divulgadas na mídia, onde se afirma que "*dos contaminados, a maioria é mulher*" (Correio Braziliense, 24/07/2020) [4].

4.2.2 Coeficiente de Determinação (R^2)

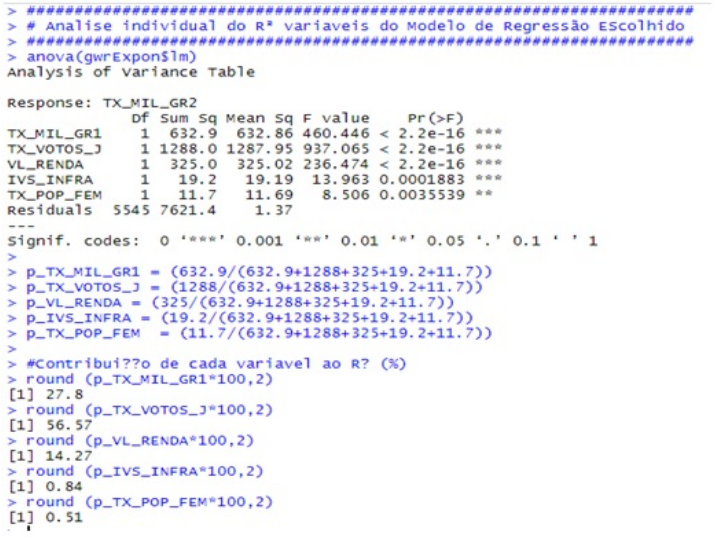
Por meio da aplicação da **Análise de Variância (ANOVA)** [5], foi obtida a decomposição da soma dos quadrados para cada fonte de variação (variável) do Modelo de Regressão.

Em seguida, dividindo-se a soma dos quadrados de cada variável independente pela Soma dos Quadrados Total, obteve-se o percentual de contribuição de cada uma delas para o R^2 do Modelo de Regressão.

[4] https://www.correiobraziliense.com.br/app/noticia/cidades/2020/07/24/interna_cidadesdf,874833/covid-no-df-mulheres-sao-maior-parte-de-infectados-mas-homens-morrem.shtml

[5] A **análise de variância** conhecida como **ANOVA** é uma técnica estatística ou um procedimento utilizado para fazer comparações entre três ou mais grupos em amostras independentes. Permitindo assim, fazer afirmações sobre as médias das populações baseado na **análise de variâncias** amostrais (https://ejeconsultoria.com.br/2019/10/31/importancia-da-anova-analise-de-variancia/)

Figura 20 – Execução da Análise de Variância (ANOVA)



```
> ########################################################################
> # Analise individual do R² variaveis do Modelo de Regressão EScolhido
> ########################################################################
> anova(gwrExpon$lm)
Analysis of Variance Table

Response: TX_MIL_GR2
            Df Sum Sq Mean Sq F value    Pr(>F)
TX_MIL_GR1   1  632.9  632.86 460.446 < 2.2e-16 ***
TX_VOTOS_J   1 1288.0 1287.95 937.065 < 2.2e-16 ***
VL_RENDA     1  325.0  325.02 236.474 < 2.2e-16 ***
IVS_INFRA    1   19.2   19.19  13.963 0.0001883 ***
TX_POP_FEM   1   11.7   11.69   8.506 0.0035539 **
Residuals 5545 7621.4    1.37
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
> p_TX_MIL_GR1 = (632.9/(632.9+1288+325+19.2+11.7))
> p_TX_VOTOS_J = (1288/(632.9+1288+325+19.2+11.7))
> p_VL_RENDA = (325/(632.9+1288+325+19.2+11.7))
> p_IVS_INFRA = (19.2/(632.9+1288+325+19.2+11.7))
> p_TX_POP_FEM  = (11.7/(632.9+1288+325+19.2+11.7))
>
> #Contribui??o de cada variavel ao R? (%)
> round (p_TX_MIL_GR1*100,2)
[1] 27.8
> round (p_TX_VOTOS_J*100,2)
[1] 56.57
> round (p_VL_RENDA*100,2)
[1] 14.27
> round (p_IVS_INFRA*100,2)
[1] 0.84
> round (p_TX_POP_FEM*100,2)
[1] 0.51
```

Abaixo, são apresentadas as Somas dos Quadrados de cada variável e sua contribuição para o R² do Modelo de Regressão:

Tabela 3 – Soma dos quadrados e contribuição das variáveis para o R²

Variáveis Independentes	**Soma Quadrados**	**Contribuição p/ R² (%)**
Taxa de Casos Graves 1º Período	632,9	27,80
Taxa Votos Bolsonaro	1.288	56,57
Valor Renda (per capita)	325	14,27
IVS Infraestrutura	19,2	0,84
Taxa Populacional Feminina	11,7	0,51
TOTAL	2.276,80	100

5 Conclusão e Considerações Finais

A presente pesquisa se dedicou a identificar e mensurar quais variáveis demonstraram maior poder preditivo e explicativo na taxa de evolução de casos graves de Covid-19 nos municípios brasileiros (*target*), no segundo semestre de 2020. Nesse ímpeto, utilizou-se inicialmente de instrumental teórico clássico de regressões múltiplas para definição do modelo preditivo em função de variáveis socioeconômicas, demográficas e eleitorais. Aprimorou-se em seguida o modelo originalmente desenvolvido, dotando-o de ferramentas de estatística espacial, tanto em modelo autorregressivo espacial (SAR) quanto em modelos geo-espaciais locais (GWR).

A adoção de *features* espaciais aprimorou a qualidade preditiva do modelo e nos possibilitou com maior acurácia aferir e mensurar aquelas variáveis que evidenciaram maior poder preditivo e contribuíram de maneira mais relevante na predição de casos graves de Covid-19.

A pertinência e atualidade da matéria é inequívoca.

O tema é indubitavelmente o cerne de grandes esforços dispensados pela comunidade científica internacional. É, portanto, nessa vertente em que se insere nossa modesta contribuição, de forma que se obtenha uma adequada compreensão do fenômeno no Brasil. Contribuindo na adoção de políticas públicas que permitam de maneira mais efetiva a mitigação dos desfechos mais severos da doença no sistema de saúde publico e privado nacional.

O estudo não pretende exaurir as inúmeras frentes de pesquisas possíveis, mas suscitar a adensar a discussão de posse de dados reais e ferramentas tanto clássicas quanto contemporâneas de modelagem preditiva.

A situação dos casos graves no primeiro semestre, a taxa de votos na chapa vencedora no segundo turno do certame eleitoral para o executivo federal e a renda per capita do município foram as *features* que apresentaram maior relevância e contribuiram mais significativamente na predição dos casos graves de Covid-19 no Brasil durante o segundo semestre de 2020.

6 Bibliografia

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Base de Informações por Setor Censitário Censo 2010 - Universo Novo.** Rio de Janeiro, 2011.

SARAIVA, A. (2021, fevereiro 2). **Mais de 70% de mortes por SRAG em 2020 foram causadas por covid-19, diz Fiocruz** [versão eletrônica]. Valor Econômico, Recuperado em 21 de abril, 2021, de https://valor.globo.com/brasil/noticia/2021/02/26/mais-de-70percent-de-mortes-por-sindrome-respiratoria-aguda-grave-no-pais-em-2020-foram-causadas-por-covid-19-diz-fiocruz.ghtml

SILVA, B. F. A., **Coesão Social, Desordem percebida e Vitimização em Belo Horizonte, Minas Gerais e Brasil**. Dissertação (Mestrado em Sociologia) - Faculdade de Filosofia e Ciências Humanas, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.

Referências

Almeida, E. (2012). Econometria Espacial Aplicada. *Livro*.

Anselin, L. (1995, 4). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, *27*(2), 93 – 115. Disponível em https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Rregression* (1st ed.). West Sussex: John Wiley & Sons.

Francisco, E. (2010). *Indicadores de renda baseados em consumo de energia elétrica* (Administração de Empresas, Fundação Getúlio Vargas). Disponível em http://hdl.handle.net/10438/8158

Hair Jr., J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2005). *Análise Multivariada de Dados* (5th ed.). Porto Alegre: Bookman.

Nakaya, T. (2015). Geographically weighted generalised linear modelling. In C. Brunsdon & A. Singleton (Eds.), *Geocomputation - a pratical primer* (1st ed., chap. 12). London: SAGE Publications Ltd.

Saraiva, A. (26/02/2021). *Mais de 70% de mortes por SRAG em 2020 foram causadas por covid-19, diz Fiocruz.* Versão eletrônica. Disponível em https://valor.globo.com/brasil/noticia/2021/02/26/mais-de-70percent-de-mortes-por-sindrome-respiratoria-aguda-grave-no-pais-em-2020-foram-causadas-por-covid-19-diz-fiocruz.ghtml

USDA-ARS JORNADA EXPERIMENTAL RANGE, BLM-AIM PROGRAM, & IDAHO CHAPTER OF THE NATURE CONSERVANCY. (2012). *Geographically Weighted Regression.* Disponível em https://wiki.landscapetoolbox.org/doku.php/spatial_analysis_methods:geographically_weighted_regression