

## **GeoAnalytics Aplicado a Estratégia Eleitoral: o Caso do Partido NOVO nas Eleições Presidenciais de 2018 no Brasil**

**AFONSO HENRIQUE BARROS MACHADO**  
FGV MANAGEMENT

**ÁQUILA ESTEVÃO DA SILVA CAMPOS**  
FGV MANAGEMENT

**FELIPE GONTIJO FONSECA**  
FGV MANAGEMENT

**MARCOS ANTONIO ALVES BEZERRA JUNIOR**  
FGV MANAGEMENT

**EDUARDO DE REZENDE FRANCISCO**  
ESCOLA DE ADMINISTRAÇÃO DE EMPRESAS DE SÃO PAULO (FGV-EAESP)

# **GeoAnalytics Aplicado a Estratégia Eleitoral: o Caso do Partido NOVO nas Eleições Presidenciais de 2018 no Brasil**

**Palavras-chave:** Eleições, Regressão Espacial, *GeoAnalytics*

## **INTRODUÇÃO**

A Geografia Eleitoral é um campo com mais de cem anos de existência, porém com um número relativamente pequeno de publicações quando comparado a disciplinas de Economia e Ciência Política. No Brasil, até 2008 era exíguo o número de publicações relacionadas à geografia das eleições subsequentes ao regime militar (O’Loughlin, 2003; Terron, 2012). Porém, o interesse por esse campo de conhecimento vem crescendo.

Os avanços tecnológicos, que possibilitam cada vez mais o surgimento de novas técnicas de análise espacial, aliados à necessidade de conhecer melhor o impacto do espaço brasileiro na escolha do candidato, motivam a realização deste estudo.

Foi escolhido como objeto da análise proposta para este estudo a eleição presidencial do partido NOVO em 2018. O partido NOVO apresenta ideologia bem definida em seu estatuto e os valores do partido foram bem difundidos na sua primeira eleição, fatores determinantes para a escolha realizada. Carreirão (2002) confirma a identificação da identidade ideológica, porém adiciona a escolaridade do eleitor como fator que determina o impacto da identidade ideológica.

## **PROBLEMA DE PESQUISA E OBJETIVO**

Estudos de Ciência Política ressaltam que algumas variáveis eleitorais possuem significância em relação ao desempenho eleitoral. Por exemplo, Pereira & Rennó (2007), na previsão de determinantes da reeleição para câmara de deputados, encontraram significância em variáveis eleitorais, tais como: percentual de emendas executadas, base eleitoral, votação da eleição anterior, concentração eleitoral, presença de mesa diretora, relator de comissão e projetos aprovados.

Nesse sentido, seria possível identificar que características socioeconômicas levam os municípios do Brasil a se identificarem com o partido NOVO? Qual a importância da variável geográfica nesse contexto? Há dependência espacial, ou influência da região geográfica, no contexto nacional, para essa identificação? Desse modo, espera-se encontrar um padrão socioeconômico, bem como geográfico, entre os municípios que mais depositaram confiança no partido.

Neste contexto, o objetivo desta pesquisa é analisar o poder de explicação do desempenho eleitoral a partir de modelos distintos comparados: (i) regressão linear múltipla, que não considera a variável geográfica; e modelos que incluam a variável geográfica, (ii) em aspecto global (*Spatial Auto-Regressive model*, SAR), e (iii) em abordagem local, que diferencie os parâmetros por meio das características de cada região – *clusters* (*Geographically Weighted Regression*, GWR).

Burnham e Anderson (2004) ressaltam a importância da seleção de modelos em evidências científicas. Dentre as alternativas disponíveis escolhemos o modelo de critérios de informação de Akaike (AIC) para comparar o desempenho dos modelos.

## **FUNDAMENTAÇÃO TEÓRICA**

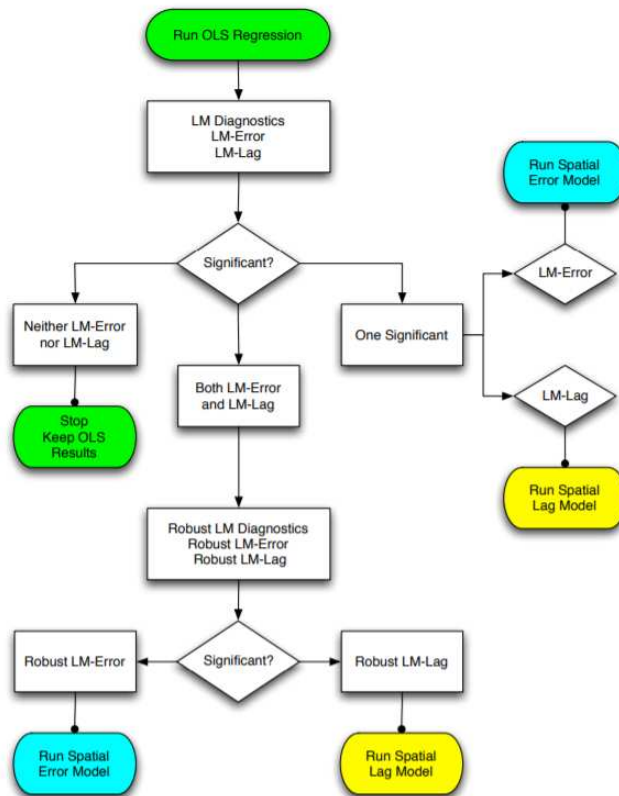
Em 1913, representações cartográficas e estatísticas foram utilizadas em conjunto por André Siegfried, na França, mais especificamente para avaliar e compreender fenômenos políticos e sociais que guiavam as orientações de votos com a relação causal ao solo no Oeste da França

na Terceira República. A hipótese de Siegfried, era de que existia correlação direta entre as estruturas sociais pré-existentes no local e o comportamento dos eleitores. Tal sugestão foi abordada em sua obra republicada em 1995, intitulada *Tableau Politique de la France de l'Ouest sous la IIIe République* (Siegfried, 1995). Adicionalmente, Singer (1999) identifica a identidade ideológica entre eleitor e candidato como um fator central na decisão do voto.

Dadas as proporções continentais do Brasil, assim como características socioeconômicas que variam bastante ao longo do território, é de se esperar que o espaço exerça influência na votação para presidente de um partido. Isso está de acordo com a primeira lei da geografia, que afirma, em tradução livre, que “tudo está relacionado com tudo, mas coisas mais próximas estão mais relacionadas entre si que as mais distantes” (Tobler, 1970).

O que determina de forma objetiva a pertinência da utilização de modelos espaciais são os multiplicadores de Lagrange. Esse teste estatístico verifica se vale ou não a pena testar o modelo espacial com os dados pretendidos, a partir de conceitos de significância e robustez. A análise dos multiplicadores de Lagrange, pode ser feita pelo modelo *Spatial Regression Decision Process* (Anselin, p.198, 2005), considerando o fluxograma da Figura 1.

**Figura 1.** Fluxograma de *Spatial Regression Decision Process*



Fonte: Anselin (2005, p. 198).

O trabalho foi realizado a partir de bases de dados secundárias provenientes do IBGE e TSE, agregadas em 5565 municípios com eleições realizadas em 2018. Foi realizada a regressão linear múltipla do total de votos para Presidente do Partido NOVO. Posteriormente, foi realizada uma regressão espacial, que incorpora aspectos geográficos ao modelo com o intuito de aumentar o poder preditivo. Basicamente, a técnica pode ser aplicada de maneira global ou de forma local. Segundo Francisco (2010), a primeira maneira utiliza indicadores sumarizados para região como um todo, destacando as similaridades no espaço e “leis” gerais. A segunda

forma utiliza a desagregação local das estatísticas globais, buscando as diferenças no espaço, procurando as exceções ou *hot spots* locais.

Com relação aos modelos globais (SAR), foram aplicadas duas técnicas: *Spatial Lag Model* e o *Spatial Error Model*. Com relação aos modelos locais, aplicou-se a GWR (*Geographically Weighted Regression*), ou Regressão Ponderada Geograficamente.

### ***Spatial Lag Model***

O ponto principal dos modelos SAR (*Spatial Auto-Regressive models*) é que a variável dependente de um elemento do modelo, possui influência no valor da variável dependente dos vizinhos. Segundo Carvalho Ywata e Albuquerque (2011), “a ideia dos modelos SAR é utilizar a mesma ideia dos modelos AR (autorregressivos) em séries temporais, por meio da incorporação de um termo de lag entre os regressores da equação”.

Na sua forma mais simples, o modelo SAR tem expressão:

$$y = \rho W y + X \beta + \varepsilon \quad (1)$$

Segundo Francisco (2010), 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 ( $\rho$ ) é o coeficiente espacial autorregressivo. A hipótese nula para a não existência de autocorrelação, é que ( $\rho$ ) 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*.

### ***Spatial Error Model***

No *Spatial Error*, o “erro” de um elemento influencia nos erros dos vizinhos. A ideia é que existe uma “variável escondida” que não se consegue descobrir qual, pois não se possui o dado ou informação, mas ela se distribui no espaço e tentou-se simular por meio dos erros da vizinhança. A formulação é a seguinte:

$$y = X \beta + u \quad , \text{ onde } u = \lambda W u + \varepsilon \quad (2)$$

Analisando a fórmula acima, Carvalho Ywata e Albuquerque (2011) teceram o seguinte comentário:

O vetor de resíduos  $\varepsilon$  possui distribuição normal multivariada, com média nula e matriz de covariância  $\sigma^2 I$ . O coeficiente escalar  $\lambda$  indica a intensidade da autocorrelação espacial entre os resíduos da equação observada. Mais especificamente, esse parâmetro mensura o efeito médio dos erros dos vizinhos em relação ao resíduo da região em questão. Note-se que, ao contrário dos modelos SAR, os modelos SEM não apresentam a variável resposta como uma função direta dos seus lags espaciais. A autocorrelação espacial nos modelos SEM aparece nos termos de erro. Outra diferença dos modelos SEM em relação aos modelos SAR é que os coeficientes no vetor  $\beta$ , podem ser estimados consistentemente via mínimos quadrados ordinários. (Rev. Bras. Biom., São Paulo, v.29, n.2, p.273-306, 2011.).

### ***GWR – Geographically Weighted Regression***

A ideia principal do modelo GWR, é que os parâmetros vão mudando ao longo do mapa. O foco é achar as diferenças no Mapa, com o intuito de prover um modelo mais eficiente.

Segundo Francisco (2010) a regressão ponderada geograficamente descreve uma gama de modelos de regressão em que os coeficientes, parâmetros ( $\beta$ ), variam de acordo com a localidade. Ela ajusta um modelo de regressão a cada ponto observado, ponderando todas as

demais observações como função da distância (ou de qualquer medida de vizinhança) deste ponto. Em outras palavras, têm-se regressões diferentes para cada observação, assim a contribuição (valor do parâmetro) de cada variável explicativa ao modelo é diferente para cada ponto.

$$y(g) = \beta_0(g) + \beta_1(g)x_1 + \beta_2(g)x_2 + \dots + \beta_p(g)x_p + \varepsilon \quad (3)$$

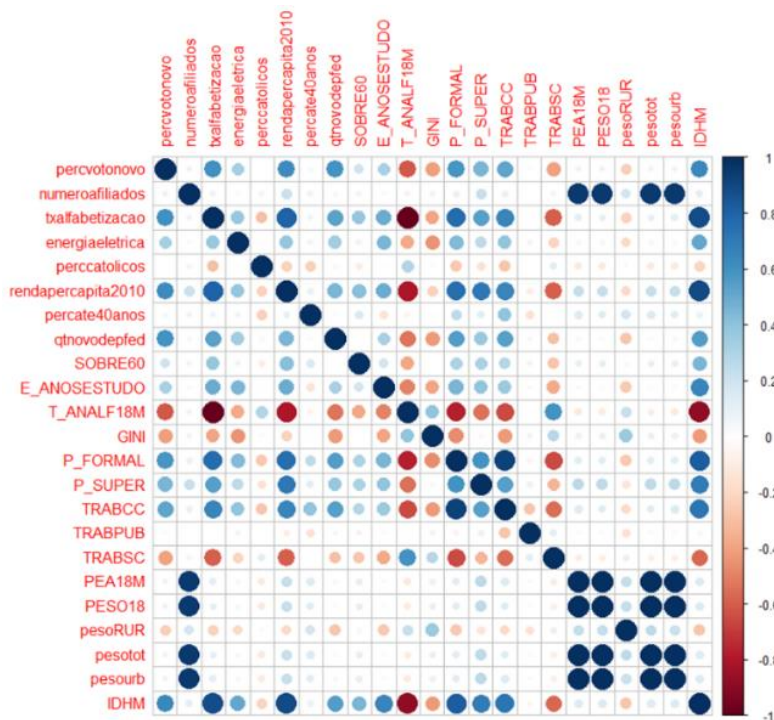
Segundo Francisco (2010, p.123), “ $g$  é um vetor dos  $n$  pontos, no espaço bidimensional, os parâmetros do vetor  $\beta(g)$  são específicos para cada observação  $i$  de localização  $g(i)=(u_i, v_i)$  e o termo de erro  $\varepsilon$  é suposto e independente e de comportamento  $\varepsilon \sim N(0, \sigma^2 I)$ . Temos na realidade, um conjunto de  $n$  regressões diferentes, um para cada ponto  $g_i$  no espaço”, ponderadas pela distância geográfica ao ponto de estimação  $i$ , central à amostra local.

## DISCUSSÃO

Em primeiro lugar, foram escolhidas variáveis que expressam os valores do partido, comunicados em seu estatuto, utilizando a hipótese de identidade ideológica como fator decisivo do voto. Para descobrir as variáveis que poderiam ser utilizadas para o modelo, foram testadas suas correlações com a variável dependente que é percentual de votos no NOVO por município.

Percebe-se que algumas variáveis apresentaram correlações mais significativas, merecendo ser consideradas para uma regressão. Entretanto, muitas dessas variáveis possuem correlação elevada entre elas mesmas, como é o caso de taxa de alfabetização e taxa de analfabetismo, renda per capita e IDHM, entre outras. Para evitar que essa correlação entre variáveis explicativas cause multicolinearidade, a escolha das variáveis a serem utilizadas levou isso em conta.

**Figura 2.** Matriz de correlação



Fonte: Elaboração própria dos autores, a partir do software R.

Verificada a importância de se considerar o espaço geográfico para a análise de determinantes do voto no candidato do partido NOVO para presidente, dividiu-se a análise em três partes:

- Regressão linear: antes da verificação do efeito espacial, foram testadas diversas variáveis que pudessem prever a votação em João Amoedo, candidato do partido NOVO para presidência em 2018;
- Regressão espacial global: inclusão do efeito do espaço, pelo teste da dependência espacial e utilização de *lag* espacial ou erro espacial, de acordo com a significância;
- Regressão geograficamente ponderada: feita uma análise espacial com regressões estimadas por região, ponderada pela distância.

### Resultados da regressão linear múltipla

Os resultados da regressão linear múltipla com as variáveis renda per capita, total de deputados federais candidatos ao partido NOVO, taxa de analfabetismo e o índice de GINI, podem ser observados na Figura 3.

**Figura 3.** Regressão linear múltipla

```
lm(formula = perc_votonovo ~ rendapercapita2010 + qtnovodepfed +
  T_ANALF18M + GINI, data = base)

Residuals:
    Min       1Q   Median       3Q      Max
-0.025459 -0.004961 -0.001157  0.002871  0.161870

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.662e-02  1.241e-03  13.396 < 2e-16 ***
rendapercapita2010  2.114e-05  8.202e-07  25.773 < 2e-16 ***
qtnovodepfed    2.730e-04  9.854e-06  27.703 < 2e-16 ***
T_ANALF18M     -6.791e-05  2.194e-05  -3.095  0.00198 **
GINI           -3.713e-02  2.217e-03 -16.752 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.009432 on 5560 degrees of freedom
Multiple R-squared:  0.5441, Adjusted R-squared:  0.5437
F-statistic: 1659 on 4 and 5560 DF, p-value: < 2.2e-16
```

Fonte: Elaboração própria dos autores, a partir do software R.

Para esse modelo, o valor-P associado com a estatística F do modelo é menor do que o  $\alpha$  de 0,05, ou seja, o modelo é significativo para explicar a variável dependente. Verifica-se também, que todas as variáveis são significativas para o modelo a um nível de significância de 0,01. Ainda com essas variáveis, também foi testado o VIF, que pode ser verificado na Figura 4 abaixo:

**Figura 4.** Fator de Inflação da Variância – VIF.

rendapercapita2010	qtnovodepfed	T_ANALF18M	GINI
2.983086	1.532750	3.447281	1.341407

Fonte: Elaboração própria dos autores, a partir do software R.

Uma vez que o VIF quantifica qual a correlação de um preditor com outros preditores de um modelo, quanto maior o VIF, maior a probabilidade de ocorrer multicolinearidade. Para o

modelo em questão, nenhuma variável apresentou fator maior que 4, o que aponta para uma baixa correlação entre as variáveis explicativas.

Devido ao tamanho da amostra analisada (5.565 municípios oficiais do Brasil em 2018), não será testada a normalidade já que, para grandes amostras, até uma pequena variação vai levar a um resultado significativo e que rejeita a normalidade (Field, 2009). Isso não será um problema pois como o Teorema do Limite Central afirma, a média amostral converge para uma distribuição normal.

Realizou-se a verificação de existência de autocorrelação, dado o valor do teste de *Durbin-Watson* próximo de 2, que revela ausência de autocorrelação, assim como o valor-P que não rejeita a hipótese nula de ausência de autocorrelação. Esse resultado já era esperado, já que não há ligação entre observações sem a dimensão tempo.

Foi então testado se o modelo apresenta heterocedasticidade, problema comum em dados transversais, que é quando o erro apresenta variância desigual. Por essa razão, foi feita a verificação por meio do teste de *Breusch-Pagan* (Figura 5).

**Figura 5.** Teste de *Breusch-Pagan*

```
studentized Breusch-Pagan test
BP = 153.05, df = 4, p-value < 2.2e-16
```

Fonte: Elaboração própria dos autores, a partir do software R.

Considerando a hipótese nula de que a variância do erro é homocedástica, para o valor-P significativo que foi obtido podemos rejeitar a hipótese nula, ou seja, há presença de heterocedasticidade. Ainda segundo Gujarati (2019), a heterocedasticidade não altera as propriedades dos estimadores de MQO de serem não viesados e consistentes, mas eles deixam de ser de variância mínima ou eficientes.

Essa heterocedasticidade identificada pode ter sua raiz no espaço geográfico, já que para um país de grandes dimensões e diversificado como o Brasil, possuindo um elevado número de municípios, é de se supor a grande variação entre as unidades territoriais mais distantes.

Após essas verificações, chegou-se em um modelo conforme demonstrado abaixo:

$$\text{perc}votonov = \beta_0 + \beta_1 * \text{rendapercapita2010} + \beta_2 * \text{qtnovodepfed} + \beta_3 * T_{ANALF18M} + \beta_4 * GINI + \varepsilon$$

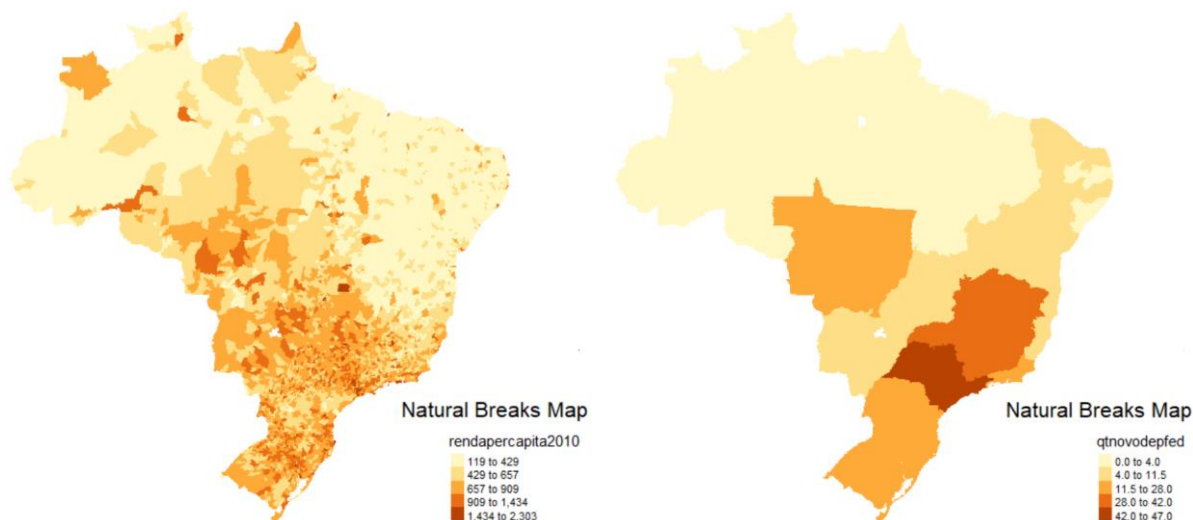
## **Análise Geoespacial**

Com as variáveis selecionadas, utilizamos a otimização de *Natural Breaks*, método de classificação de quebras naturais de Jenks (1977), para observar a distribuição e o agrupamento de dados no mapa. Pelo mapa da Figura 6 é possível observar as classes de maior renda per capita concentradas nas regiões Centro-Oeste, Sudeste e Sul do Brasil. Semelhante a distribuição de renda per capita, a quantidade de deputados federais disputando um cargo pelo partido NOVO teve maior concentração nos Estados de São Paulo, Minas Gerais, Rio de Janeiro, Mato Grosso, Paraná, Santa Catarina e Rio Grande do Sul.

Já o coeficiente de GINI, índice criado pelo italiano Conrado Gini (1912) para medir a desigualdade de renda de uma população, costumeiramente sendo medido em 0 para igualdade e 1 para desigualdade, observado no mapa da (Figura 7) é possível verificar uma maior distribuição de renda nas regiões Sudeste e Sul, tendo as regiões Norte e Nordeste apresentado maiores índices de desigualdade de renda. Em seguida, a taxa de analfabetismo da população

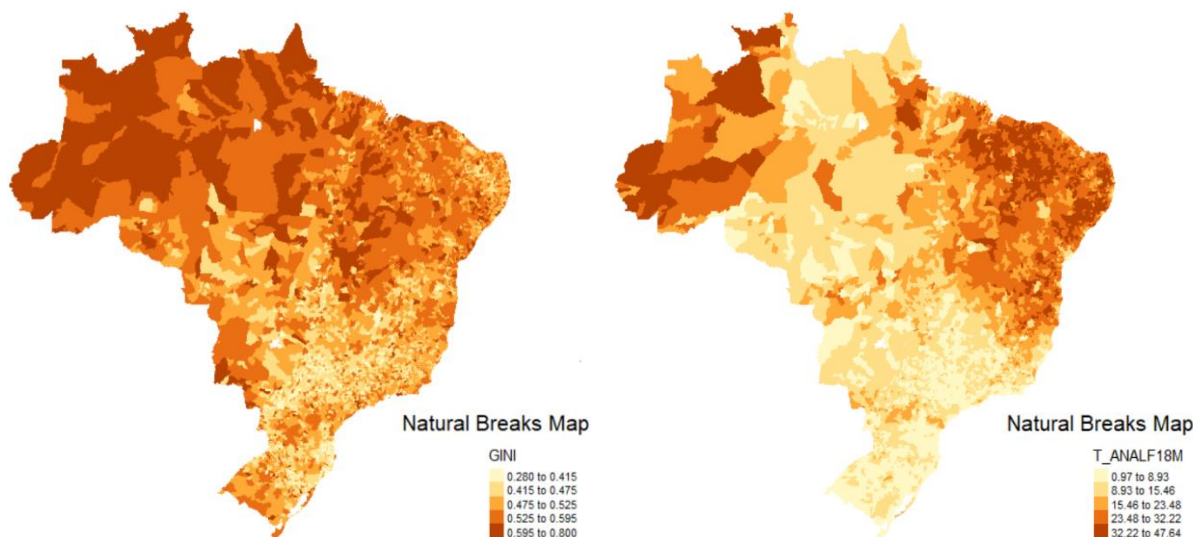
de 18 anos ou mais de idade está mais elevada no extremo Norte e na região do Nordeste, indicando a falta de instrução de uma maior parte da população dessas regiões. Visualmente, é possível verificar que, nas regiões com maior renda per capita, maior quantidade de deputados federais pelo partido NOVO, menor coeficiente de GINI e menores taxas de analfabetismo, estão as maiores taxas de distribuições de votos para o partido NOVO, ou seja, nas regiões Centro-Oeste, Sul e principalmente Sudeste do país (Figura 8), indicando assim, correlação entre as variáveis explicativas e a variável independente.

**Figura 6.** Distribuição Espacial das variáveis Renda per capita por município (esquerda) e Quantidade de deputados federais pelo partido NOVO por UF (direita)



Fonte: IBGE e TSE – Elaboração própria dos autores, a partir do software R.

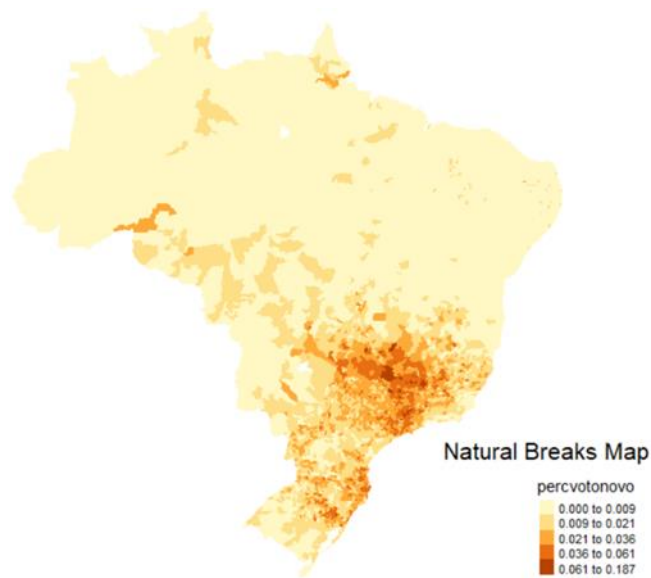
**Figura 7.** Distribuição Espacial das variáveis índice de GINI (esquerda) e taxa de analfabetismo da população de 18 anos ou mais de idade (direita)



Fonte: IBGE e TSE – Elaboração própria dos autores, a partir do software R.



**Figura 8.** Distribuição Espacial da variável Percentual de votos no partido NOVO

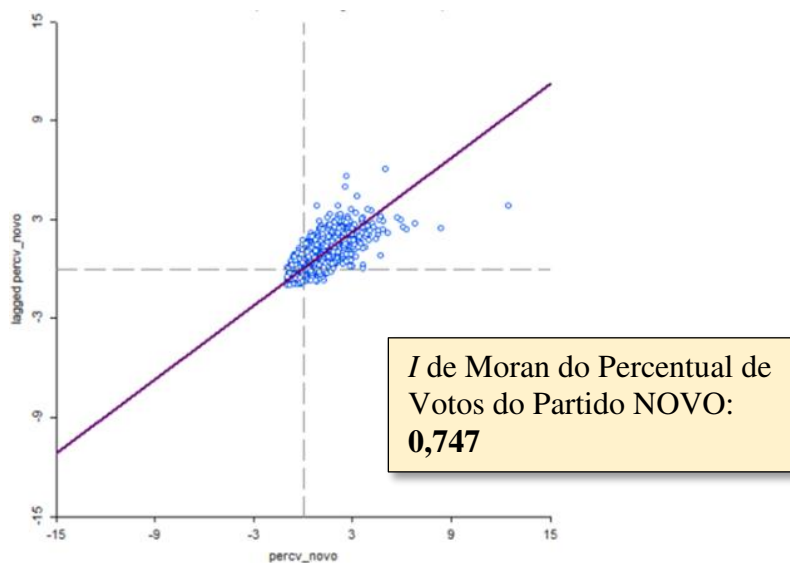


Fonte: IBGE e TSE – Elaboração própria dos autores, a partir do software R.

Para explorar a existência de autocorrelação espacial, realizou-se análise espacial por meio do teste de  $I$  de Moran Univariado Local. Utilizou-se o local, porque ele mostra onde estão os *clusters*, ao contrário do global que responde se existem *clusters* espaciais.

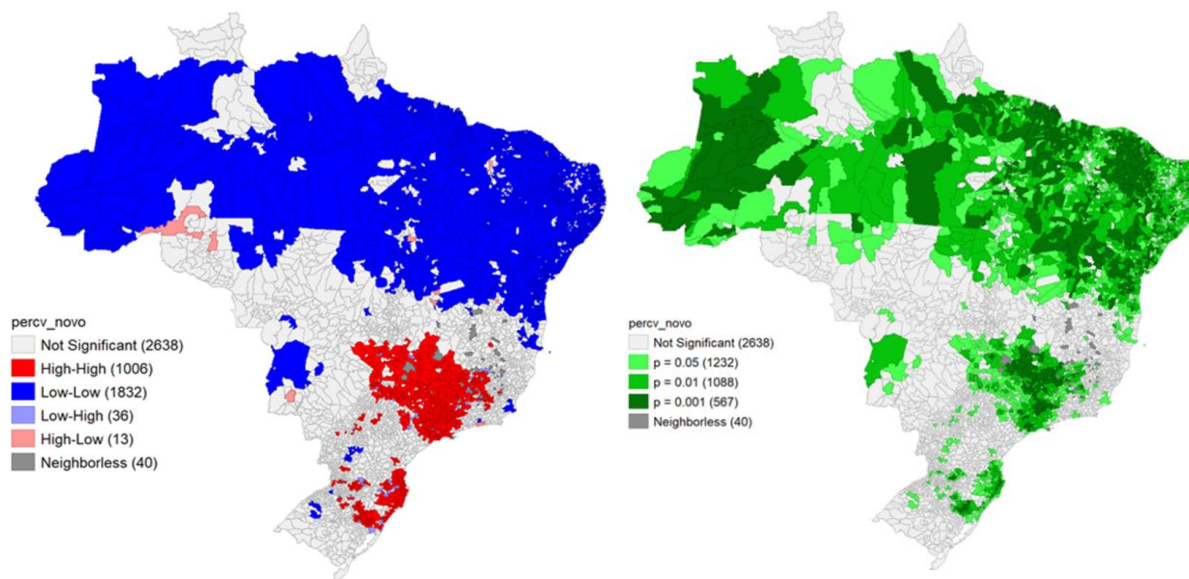
O  $I$  de Moran obtido, a partir de uma matriz de vizinhança do tipo Rook de Ordem 1, está demonstrado no mapa de dispersão do Gráfico 1 e foi de 0,747, o que demonstra uma autocorrelação forte positiva, ou seja, o percentual de votos no NOVO no Brasil tem algum tipo de padrão espacial. Percebe-se grande concentração das observações no 1º quadrante, ou seja, de um município alto vizinho de outro município alto.

**Gráfico 1.** Moran *Scatterplot* da variável percentual de votos no partido NOVO



Fonte: TSE – Elaboração própria dos autores, a partir do software GeoDa 1.14

**Figura 9.** Mapa *Local Moran's I* (LISA) e Mapa de Significância da variável Percentual de votos no partido NOVO

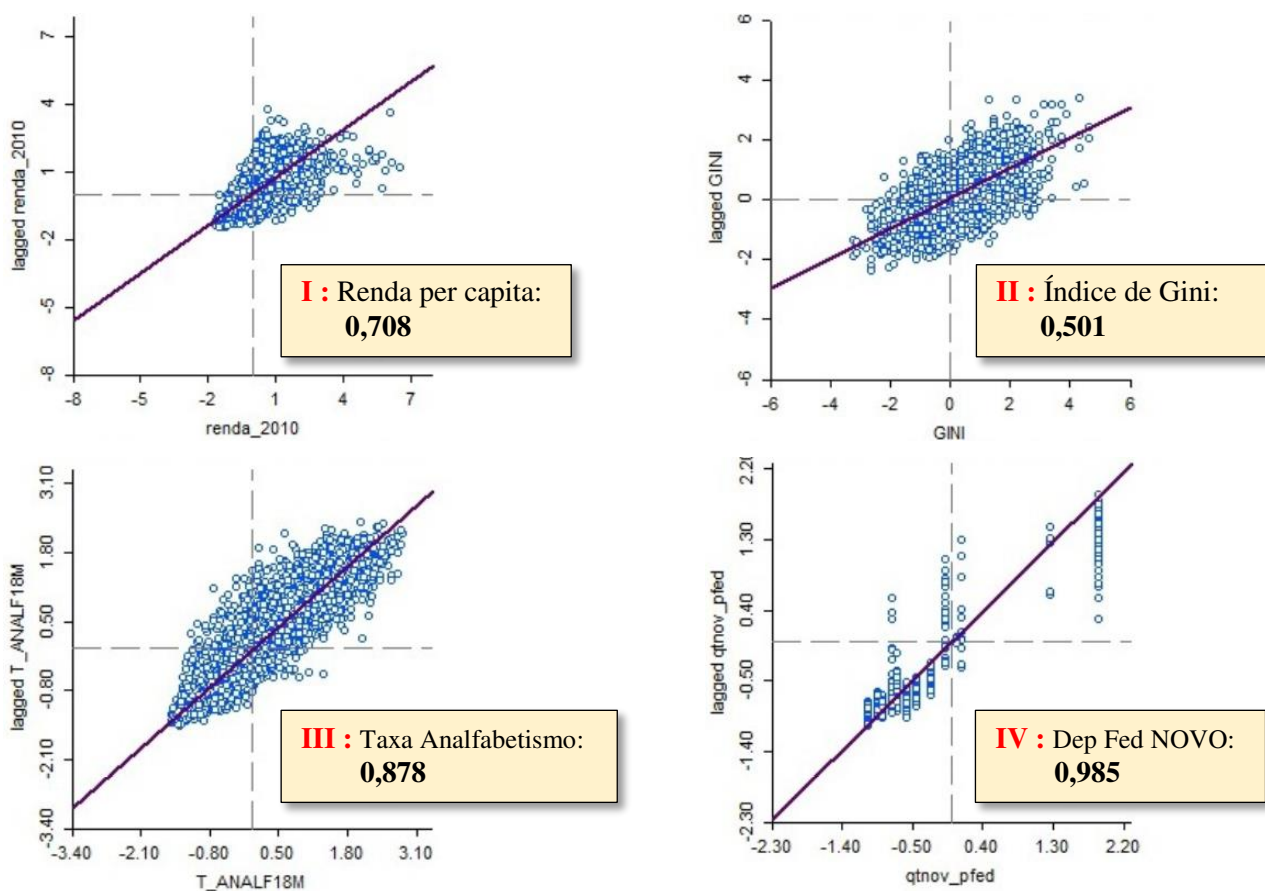


Fonte: IBGE e TSE – Elaboração própria dos autores, a partir do software GeoDa 1.14.

Foi elaborado, além disso, o Mapa de Cluster – LISA (*Local Indicator Spatial Association*), evidenciando os clusters existentes para o Brasil, conforme a (Figura 9, esquerda). Visualmente, é possível perceber a presença de um *cluster low-low* (valores baixos da variável avizinhados por valores também baixos) englobando grande parte das regiões Norte e Nordeste, enquanto as regiões Sudeste e litoral Sul indicam a presença de clusters *high-high* (valores altos da variável avizinhados por valores altos). Esse *cluster* do Sudeste é em linha com o que se podia esperar, já que as ações do partido foram mais focadas ali, com um candidato a Governador do partido, inclusive, sendo eleito (Governador Romeu Zema, em Minas Gerais). Também, foi realizado o Mapa de Significância – LISA para as observações, conforme a (Figura 9, direita), evidenciando que os municípios significantes para a análise espacial estão concentrados no Norte, Nordeste, Sudeste e litoral Sul.

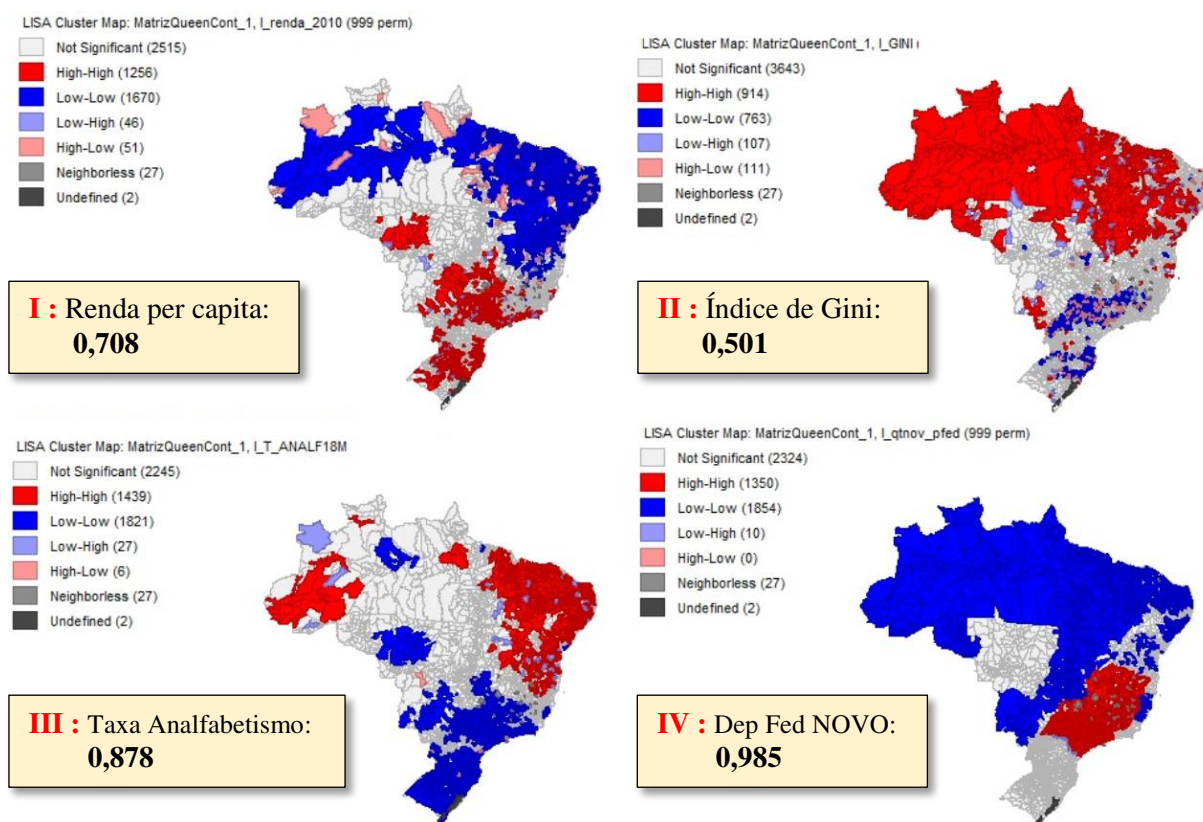
Pelo Gráfico 2, é possível verificar o *I* de Moran obtido nas variáveis renda per capita (0,708), índice de GINI (0,501), quantidade de deputados federais pelo partido NOVO (0,985) e taxa de analfabetismo da população de 18 anos ou mais de idade (0,878) também demonstram uma autocorrelação positiva e significativa nas variáveis explicativas. Também foi elaborado o Mapa de Cluster – LISA de cada variável, evidenciando os clusters existentes, conforme a Figura 10.

**Gráfico 2.** I de Moran das variáveis Renda per capita (I), índice de Gini (II), Taxa de analfabetismo da população de 18 anos ou mais de idade (III) e Quantidade de Deputados Federais pelo partido NOVO (IV)



Fonte: Elaboração própria dos autores, a partir do software GeoDa 1.14.

**Figura 10.** Mapa *Local Moran's I* das variáveis Renda per capita (I), índice de Gini (II), Taxa de analfabetismo da população de 18 anos ou mais de idade (III) e Quantidade de Deputados Federais pelo partido NOVO (IV)



Fonte: Elaboração própria dos autores, a partir do software GeoDa 1.14.

### Resultados das Regressões Espaciais

A primeira tentativa de melhorar a performance dos resultados foi por meio da execução do Modelo Espacial Auto Regressivo (SAR). Houve uma melhora significativa no resultado do ( $R^2$ ), com o valor chegando ao patamar de 0,7259 e o AIC atingindo o valor de -38403,3. Dessa forma, o *Spatial Lag* aumentou a qualidade do poder de explicação do modelo. Avançando para a execução do *Spatial Error*, obteve-se uma pequena melhora em relação ao SAR com um ( $R^2$ ) chegando ao valor de 0,7467 e o AIC atingindo -38690,9. Em relação às variáveis utilizadas no estudo, permaneceram significativas nos três modelos executados, com exceção da variável “Taxa de analfabetismo da população de 18 anos ou mais de idade”, que perdeu a significância na execução do modelo *spatial lag*. Para comparar a importância das variáveis em cada modelo será utilizado o coeficiente padronizado, ou seja, dividem-se os coeficientes pelo desvio padrão, conforme apresentado na Tabela 1.

**Tabela 1.** Coeficientes padronizados comparados dos modelos de Regressão

	<i>Regressão Simples</i>	<i>Spatial Lag</i>	<i>Spatial Error</i>
<i>Constante</i>	13,4074	1,4368	11,9513
<i>Renda_2010</i>	25,77	19,6493	23,0105
<i>qnov_pfed</i>	27,70	10,2904	14,0443
<i>GINI</i>	-16,7515	-6,9342	-12,4137
<i>T_ANALF18M</i>	-3,0959	2,4236	-4,1887
<i>LAMBDA</i>			62,0659

Fonte: Elaboração própria dos autores, a partir do software R.

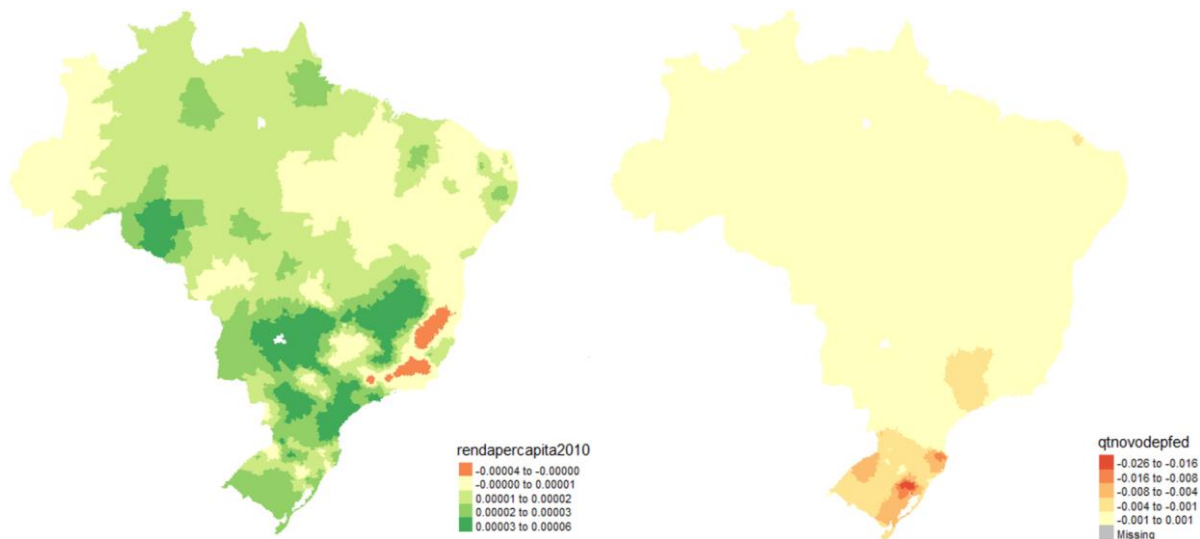
A partir da Tabela 1, verifica-se que as variáveis, quando sujeitas a alguma técnica de incorporação de variável espacial perdem um pouco a sua importância.

Analisando os bons resultados, surge a pergunta: é necessário rodar o modelo GWR? Infelizmente não foi possível manter a hipótese de homocedasticidade nos modelos globais executados. Por essa razão e por evidências de que existem relações regionais no fenômeno estudado, prosseguimos com a execução do modelo local, o GWR.

A Abordagem do modelo GWR é um pouco diferente daquela que vinha sendo adotada com os modelos Globais. Enquanto que nas regressões utilizando técnicas globais os parâmetros valem para o mapa inteiro, nos modelos locais, como o GWR, os parâmetros vão mudando ao longo do Mapa. Dessa forma, as análises são diferentes. Vamos verificar que, em determinadas regiões o parâmetro renda tem mais influência, em outras a variável quantidade de deputados estaduais irá influenciar mais, ou seja, há uma questão regional a ser explorada. Na Figura 11 (esquerda), que apresenta o Mapa com o desempenho da variável de renda per capita, verifica-se que nas regiões sudeste e sul a variável possui mais influência nas outras regiões do Brasil. Por outro lado, é também na região sudeste que se nota a menor influência, mais especificamente na região serrana do Rio de Janeiro e no leste do estado de Minas Gerais. Em relação ao total de Deputados Federais candidatos pelo partido NOVO, praticamente só conseguiu ter influência no Sul do país, conforme ilustrado na Figura 11 (direita).

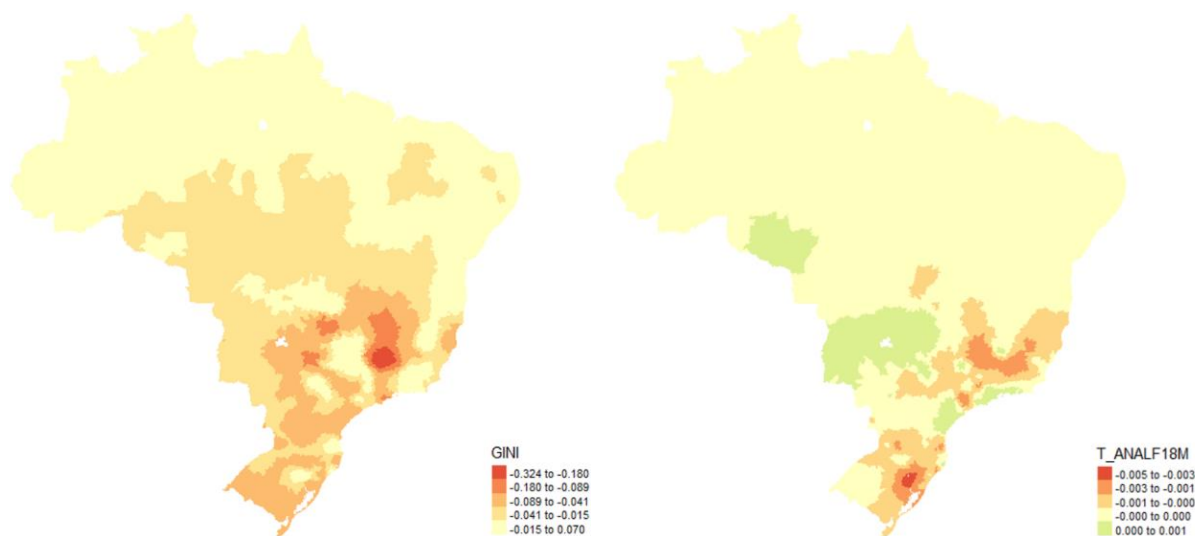
No tocante ao coeficiente de GINI e a taxa de analfabetismo (Figura 12), as variáveis conseguem explicar melhor no Sul e Sudeste. Já o poder de influência das variáveis diminui quando se avança em direção as regiões Norte e Nordeste, conforme verificado nas figuras abaixo. Finalmente, analisando o ( $R^2$ ) ao longo do Mapa (Figura 13) verifica-se que o modelo tem um melhor desempenho na região Norte, com um valor de ( $R^2$ ) variando entre 0,727 e 0,996.

**Figura 11.** Mapa da Distribuição Espacial dos Coeficientes das variáveis Renda per capita (esquerda) e Quantidade de Deputados Federais pelo partido NOVO (direita) no Modelo GWR



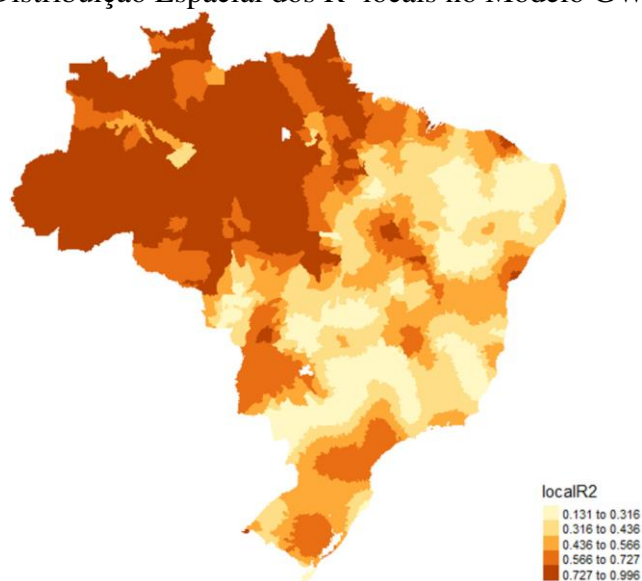
Fonte: Elaboração própria dos autores, a partir do software R.

**Figura 12.** Mapa Distribuição Espacial dos Coeficientes das variáveis Índice de GINI (esquerda) e taxa de analfabetismo da população de 18 anos ou mais de idade (direita) no modelo GWR



Fonte: Elaboração própria dos autores, a partir do software R.

**Figura 13.** Mapa da Distribuição Espacial dos  $R^2$  locais no Modelo GWR



Fonte: Elaboração própria dos autores, a partir do software R.

Com relação à metodologia técnica utilizada no GWR, Albuquerque (2017) destaca que as duas principais funções de ponderação encontradas na literatura são a função Gaussiana (normal) e a função Biquadrática, sendo que para cada uma delas existem dois tipos de expressões para cada uma das funções Gaussiana e Biquadrática, que se diferenciam por meio da escolha do parâmetro  $b$  (*bandwidth*) a ser utilizado (se fixo ou variável). Esse parâmetro é a faixa de distância ou número de vizinhos usado para cada regressor local.

Sendo assim, foram testados os modelos para as duas funções, sendo verificado o parâmetro ( $b$ ) fixo e o variável para a função Gaussiana e o fixo para a função Biquadrática, que obteve pior desempenho, conforme pode ser verificado na Tabela 2 abaixo:

**Tabela 2.** Comparativo de funções, GWR Gaussiano Fixo, GWR Gaussiano Variável e GWR Biquadrática Fixo

Função	AICC	AIC	$\sum$ Quadrados Residual	R <sup>2</sup> Quasi-global
GWR Gaussiano Fixo	-38662,5	-39197,9	0,2622	0,7583
GWR Gaussiano Variável	-39337,7	-39921,4	0,2288	0,7891
GWR Biquadrático Fixo	-36558,3	-36583,6	0,4531	0,5824

Fonte: Elaboração própria dos autores, a partir do software R.

É possível perceber que, para uma regressão geograficamente ponderada, a que ofereceu melhor resposta foi a que utilizou a função Gaussiana de parâmetro variável, tanto na comparação pelo critério de Akaike (AIC), quanto na comparação pelo coeficiente de determinação (R<sup>2</sup>).

### Comparações dos modelos

Tendo-se obtido os resultados para todos os modelos, a Tabela 3 apresenta a comparação entre seus AICs para se verificar qual o melhor modelo para explicar a variável dependente, que é percentual de votos no partido NOVO para presidente por município.

**Tabela 3.** Comparativo AIC, Regressão Linear Múltipla, Regressão Spatial Error, Regressão Spatial Lag e Regressão GWR.

	Regressão Linear Múltipla	Regressão Spatial Error	Regressão Spatial Lag	Regressão GWR (kernel variável, decaimento Gaussiano)
<i>Akaike Info Criterion</i>	-36108,2	-38690,9	-38403,3	-39921,4

Fonte: Elaboração própria dos autores, a partir do software R.

Verifica-se que o modelo GWR, foi o que apresentou o melhor resultado segundo o critério de comparação definido.

Os parâmetros estimados pelos modelos OLS e GWR foram comparados na Tabela 4. A variação dos coeficientes, de negativo no mínimo a positivo no máximo, mostra que, de acordo com a localidade, há variação significativa dos coeficientes. Considerando a mediana, apenas a variável “qtnovodepfed”, que representa a quantidade de candidatos a deputado federal em um estado, apresentou sinal diferente do valor de coeficiente obtido para OLS.

**Tabela 4.** Distribuição dos coeficientes do Total de Deputados Federais candidatos pelo partido NOVO, dos modelos OLS e GWR

Variáveis	OLS	GWR Gaussiano Variável				
		Mínimo	1º quartil	Mediana	3º quartil	Máximo
Intercepto	1,66E-02	-2,54E-02	3,49E-03	1,23E-02	4,31E-02	5,13E-01
rendapercapita2010	2,11E-05	-4,22E-05	9,72E-06	1,43E-05	2,32E-05	5,76E-05
qtnovodepfed	2,73E-04	-2,60E-02	-4,92E-04	-7,48E-05	1,36E-04	1,45E-03
T_ANALF18M	-6,79E-05	-5,34E-03	-4,76E-04	-8,48E-05	1,42E-05	1,31E-03
GINI	-3,71E-02	-3,24E-01	-4,00E-02	-1,68E-02	-4,40E-03	6,95E-02

Fonte: Elaboração própria dos autores, a partir do software R.

A razão para os valores aparentemente baixos para os coeficientes é que a variável dependente também tem valor baixo, já que se trata da porcentagem de votos do partido NOVO por município, podendo ser, por isso, no máximo 1, além do fato de que o partido novo obteve apenas 2,5% dos votos totais (2,5E-02).

Realizando a interpretação dos coeficientes pelo resultado da mediana, verifica-se, primeiro, que o valor do intercepto de 1,23E-02 representa o valor esperado para a variável dependente quando todas as variáveis explicativas são 0.

Já a renda per capita, aponta que um aumento de R\$ 1,00 na renda de 2010 de um município tem o efeito de aumentar o percentual de voto no novo em 1,43E-05. As outras variáveis (qtnovodepfed, T\_ANALF18M e GINI), considerando a mediana, têm efeito negativo na variável dependente, sendo que a redução de uma unidade para cada uma delas, tem efeito de reduzir o percentual de votos no NOVO em -7,48E-05, -8,48E-05 e -1,68E-02 respectivamente.

## CONCLUSÃO E CONSIDERAÇÕES FINAIS

O objetivo principal deste estudo foi verificar a influência de variáveis socioeconômicas no comportamento eleitoral brasileiro, na perspectiva analítica tradicional e sob as lentes da distribuição espacial. O tema “determinantes do voto” vem tomando cada vez mais espaço nos estudos eleitorais. Entender e mapear o mecanismo de escolha do eleitor não somente ajuda os partidos políticos na tomada de decisão, mas também ajuda a compreender a evolução da democracia no Brasil.

Ao início do processo investigativo e seleção de variáveis, utilizando a verificação de correlação para determinantes de votos, levou-se em consideração na seleção, a correlação entre variáveis explicativas, etapa cuidadosamente imprescindível para evitar multicolinearidade. Desta maneira, foi possível chegar nas variáveis renda per capita, total de Deputados Federais candidatos pelo partido NOVO, índice de GINI e taxa de analfabetismo da população de 18 anos ou mais de idade como preditoras e determinantes na escolha do candidato pelo NOVO.

Para verificar a existência de autocorrelação espacial, com o objetivo de considerar o espaço geográfico no modelo preditivo, foi analisado o  $I$  de Moran local da variável dependente, em que foi obtida uma autocorrelação positiva e significativa de 0,747, o que constatou fortes evidências para existência de um padrão espacial dos votos destinados ao partido NOVO, indicando também, a consideração desse parâmetro no modelo.

O modelo de regressão linear múltipla mostrou ser significativo para explicar a variável dependente, permitindo-se realizar a análise espacial para incorporação e melhora do poder de explicação do modelo. Em síntese, todos os testes para diagnóstico de dependência espacial apresentaram resultados significativos, sendo necessário a execução dos modelos *Spatial Error* e *Spatial Lag* e GWR para verificar qual dos dois modelos seria o melhor. Sendo assim, a comparação do desempenho dos modelos pelo AIC indicou o GWR como melhor modelo espacial, mais especificamente com a função Gaussiana de parâmetro variável (-39.921,4), sendo também o melhor modelo pelo coeficiente de determinação  $R^2$  (0,79). Mais ainda, os modelos GWR permitem avaliar a variação espacial da influência das variáveis na quantidade de votos para Presidente da República no Partido NOVO no território nacional, o que deve direcionar a estratégia do Partido para seu crescimento a partir de futuros pleitos.

Desse modo, ao levarmos em consideração o espaço geográfico, obtivemos uma melhora significativa no modelo preditivo e determinante para agremiação de votos para presidente do partido NOVO. No aspecto analítico e político, sempre há espaço para incorporação de variáveis e parâmetros ainda não identificados e não explicados pelo modelo, que possam futuramente contribuir ainda mais para a tomada de decisões e ações a serem seguidas pelo partido.



Estima-se que este estudo possa contribuir com a discussão do uso de modelos de regressão espacial no estudo de fenômenos ligados ao comportamento do cidadão brasileiro em suas escolhas em pleitos eleitorais e, mais ainda, no estímulo ao uso de *GeoAnalytics* na gestão e promoção de políticas públicas no país.

## BIBLIOGRAFIA

- ALBUQUERQUE, Pedro Henrique Melo; MEDINA, Fabio Augusto Scalet; SILVA, Alan Ricardo da. Regressão Logística Geograficamente Ponderada Aplicada a Modelos de Credit Scoring. *Rev. contab. finanç.*, São Paulo, v. 28, n. 73, p. 93-112, Apr. 2017. Available from <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1519-70772017000100093&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1519-70772017000100093&lng=en&nrm=iso)>. access on 30 June 2020.
- ANSELIN, L., *Exploring Spatial Data with GeoDa™: A Workkbook*, University of Illinois, 2005.
- BURNHAM, Kenneth P.; ANDERSON, David R. Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, v. 33, n. 2, p. 261-304, 2004.
- CARREIRO, Y. D. S. (2002). Identificação ideológica e voto para presidente. *Opinião Pública*, 8(1), 54-79.
- CARVALHO, Alexandre Xavier Ywata; ALBUQUERQUE, Pedro Henrique Melo. Tópicos em econometria espacial para dados cross-section. *Texto para Discussão*, 2010.
- FIELD, A. *Discovering statistics using SPSS*. 3 ed. London: SAGE publications Ltd; 2009. p. 822.
- FRANCISCO, Eduardo de Rezende. Indicadores de renda baseados em consumo de energia elétrica: abordagens domiciliar e regional na perspectiva da estatística espacial. 2010. Tese de Doutorado.
- GUJARATI, D.N. *Econometria: princípios, teoria e aplicações práticas*. Tradução de Cristina Yamagami; Revisão técnica de Salvatore Benito Virgillito. São Paulo: Saraiva Educação, 2019.
- KAISER, B., *Targeted: My Inside Story of Cambrige Analytica and How Trump, Brexit and Facebook Broke Democracy*, 2019.
- O'LOUGHLIN, J. (2003). Spatial analysis in political geography. A companion to political geography, (Feb), 30-46.
- PEREIRA, Carlos; RENNÓ, Lúcio. O que é que o reeleito tem? O retorno: o esboço de uma teoria da reeleição no Brasil. *Brazilian Journal of Political Economy*, v. 27, n. 4, p. 664-683, 2007.
- SIEGFRIED, A., *Tableau Politique de la France de l'Ouest sous la IIIe République*. Paris, Imprimerie National, 1995 (Republicação).
- SINGER, A. (1999). Esquerda e direita no eleitorado brasileiro: a identificação ideológica nas disputas presidenciais de 1989 e 1994. *Edusp*.
- TERRON, S. L. (2012). Geografia eleitoral em foco. *Em Debate: Periódico de Opinião Pública e Conjuntura Política*: ano 4, n. 2 (maio 2012).
- TOBLER W., (1970) "A computer movie simulating urban growth in the Detroit region". *Economic Geography*, 46(Supplement): 234-240.