

## **Organização partidária e votos no Legislativo: Estudo das organizações municipais do PT e PSDB no estado de São Paulo a partir da composição e influência geográfica**

### ***Party organization and votes in the legislature: Study of the municipal organizations of the PT and PSDB in Sao Paulo state from the composition and geographical influence***

**Artur Junqueira Lascala\*, Bruno Mitio Assano Silva, Eduardo de Rezende Francisco**

Escola de Administração de Empresas de São Paulo, Fundação Getulio Vargas (FGV-EAESP), SP, Brasil

#### **RESUMO**

A forma de organização partidária em um município (diretórios ou comissões provisórias) pode afetar a quantidade de votos atribuídos a determinado partido? Fatores socioeconômicos e demográficos se relacionam com o desempenho eleitoral dos partidos? Para responder a essas questões, este estudo utilizou dados das eleições municipais de 2010 para o estado de São Paulo. Analisou-se a distribuição espacial da quantidade de votos atribuída a Deputado Federal, menos suscetível a influências de coligações partidárias do que eleições para Prefeito, a partir da configuração da organização partidária e do Índice Paulista de Vulnerabilidade Social (IPVS) de cada município. Os partidos “polo” PT e PSDB foram analisados. Trata-se de análise georreferenciada acerca da potencial influência da organização partidária localmente instalada e do perfil social do eleitor, no município, sobre o resultado da eleição. Técnicas geográficas exploratórias e modelos de Estatística Espacial, como o Spatial Auto-Regressive model (SAR) e Geographically Weighted Regression (GWR) foram utilizados. Os resultados encontrados demonstram a grande importância dos aspectos geográficos como explicação dos fenômenos políticos locais, e apontam para a importância da variável espaço sobre a explicação do desempenho dos partidos no processo eleitoral.

**PALAVRAS-CHAVE:** Organização Partidária; Eleições; Inteligência Geográfica; Estatística Espacial; IPVS.

#### **ABSTRACT**

*Does the type of party organization in a municipality (directories or commissions provisórias) affect the number of votes attributed to a given party? Socio-economic-demographic factors are related to the electoral performance of the parties? Seeking to answer these questions, this study used data from the 2010 municipal elections for the State of São Paulo. It was analyzed the spatial distribution of the amount of votes attributed to Congressman, less susceptible to influences of party coalitions than elections for Mayor, from the configuration of party organization and IPVS - Paulista Social Vulnerability Index of each municipality. The parties "polo" PT and PSDB were analyzed. This is a georeferenced analysis of the potential influence of locally installed partisan organization and voter social profile in the municipality on the result of the election. Exploratory geographic techniques and Spatial Statistics models, such as Spatial Auto-Regressive model (SAR) and GWR (Geographically Weighted Regression) were used. The results show the great importance of the geographic aspects as an explanation of the local political phenomena, and point to the importance of the variable space on the explanation of the performance of the parties in the electoral process.*

**KEYWORDS:** Party Organization; Elections; Geographical Intelligence; Spatial Statistics; IPVS.

Submissão: 28 dezembro 2017

Aprovação: 30 maio 2018

#### **\*Artur Junqueira Lascala**

Mestre em Gestão e Políticas Públicas pela FGV-EAESP. Analista de Políticas Públicas e Gestão Governamental na Prefeitura de São Paulo.

Endereço: Viaduto do Chá, 15, 01020-900, Centro, São Paulo, SP, Brasil.

E-mail: arturlascala@gmail.com

#### **Bruno Mitio Assano Silva**

Mestre em Administração Pública e Governo pela FGV-EAESP. Professor auxiliar do curso de economia da FECAP, onde leciona matérias de Microeconomia.

E-mail: brunomitio@gmail.com

#### **Eduardo de Rezende Francisco**

Doutor em Administração de Empresas pela FGV-EAESP. Professor do Mestrado Profissional em Gestão e Políticas Públicas da FGV-EAESP e do Mestrado Profissional em Comportamento do Consumidor da ESPM. Sócio-fundador do Meia Bandeirada, sistema de gestão de *car sharing* e mobilidade urbana e Sócio-fundador do GisBI, think tank em Big Data e Geointeligência.

E-mail: eduardo.francisco@fgv.br

## 1 INTRODUÇÃO

Entre alguns autores da literatura contemporânea como Guarnieri (2009) e Sousa Braga, Rodrigues-Silveira e Borges (2012), a eleição para o Legislativo no Brasil é, em geral, um jogo em nível local. A forma como as bases institucionais se articula localmente é fundamental para se entender como o “jogo de poder” é distribuído entre os atores e como estes se organizam estrategicamente para disputar território no espaço eleitoral. E, neste cenário, como se enquadra a figura do partido político?

Guarnieri (2009) observa, a partir da análise dos estatutos, que os partidos se organizam por meio de suas bases filiadas. Ou seja, convenções zonais e municipais escolhem as convenções regionais, que escolhem a convenção nacional. Cada convenção determina um diretório que define a direção do partido no dia a dia, quando da não reunião da convenção, e arbitra sobre listas de candidatura e as coligações em sua respectiva esfera (municipal, estadual e nacional). Nesse sentido, a organização partidária se daria “de baixo para cima”.

Contudo, como observam Guarnieri (2009), Sousa Braga et al. (2012), Sousa Braga e Pimentel (2013) e Ribeiro (2013), a atual legislação brasileira e os estatutos partidários têm permitido um mecanismo de dissolução das comissões e diretórios locais para a instauração de uma comissão provisória, com duração média em torno de 90 dias (dependendo do estatuto de cada partido), nomeada a partir de um diretório ou comissão nacional. Ou seja, neste segundo cenário, o controle é feito “de cima para baixo”. As lideranças, que compõem a comissão nacional, determinam as comissões provisórias que, por exemplo, selecionam os candidatos em cada esfera.

No entanto, como bem apontam os estudos acima citados, sobretudo em Guarnieri (2009), as comissões provisórias não vêm sendo utilizadas de forma exclusiva ou homogênea entre os partidos. O que sugere, por sua vez, diferentes estratégias de organização partidária, que atribuem aos diretórios ou as comissões provisórias, decisões relevantes, como a formação da lista de candidatos ou do uso de recursos do fundo partidário.

Alguns resultados apontados por Sousa Braga e Pimentel (2013) mostram maior incidência dos diretórios em partidos com maior representação na Câmara. Qual o motivo disso? Segundo os autores, há uma correlação positiva entre a concentração de diretórios e o sucesso eleitoral para cada partido, pois tais tipos de organizações seriam localmente mais institucionalizados e com maior capacidade de mobilizar o eleitorado local. Porém, tal correlação tem sido mais forte em partidos com maior representação. Esta relação pode ajudar a entender o resultado encontrado por Sousa Braga et al. (2012), que apontam uma diferença na evolução organizativa dos partidos “maiores” e “menores” – com relação ao número de afiliados – no período pós Constituição de 1988. Partidos “maiores” tenderiam a dispersar sua base filiada (processo de cobertura territorial) para logo após adensar localmente sua base (processo de “massificação”). Ao passo que os partidos “menores” realizaram os dois estágios quase que simultaneamente.

Outra variável relevante parece ser o tamanho do município. Sousa Braga e Pimentel (2013) apontam uma correlação positiva entre municípios com maior número de habitantes e aqueles com maior incidência de diretórios. De fato, Sousa Braga et al. (2012) identificam três regiões com maior concentração de diretórios no Brasil, nas seguintes regiões: Rio Grande do Sul, no eixo São Paulo, Minas Gerais e Espírito Santo, e no eixo Paraíba e Alagoas. Uma hipótese preliminar que Sousa Braga e Pimentel (2013) indicam, é de que o tamanho do partido (em termos de representação na Câmara e de base filiada) e o tamanho do município (em número de habitantes) são variáveis relevantes e que determinam diferentes tipos de estratégias. Em outras palavras, a organização partidária em termos de diretórios, parece ter muito mais efeito sobre o sucesso eleitoral para partidos grandes em municípios maiores.

## 2 OBJETIVOS

A literatura recente aponta diferentes formas de organização como indício de diferentes estratégias eleitorais e de estrutura de poder entre os partidos. Essa visão parece enfraquecer a visão clássica dos

partidos puramente “fracos” e abre o precedente para algumas questões ainda não respondidas. Dentre elas, como o espaço geográfico, em suas idiosincrasias, influencia o comportamento dos partidos políticos? Se o componente local tem sua importância, por que a avaliação geográfica ainda não tem sido efetivamente analisada na literatura?

Na análise descrita neste artigo, inclui-se uma variável que não tem sido verificada em profundidade na literatura recente sobre dinâmicas partidárias: o espaço geográfico. Isto é, não apenas incluir a variável no debate, como avaliá-la criticamente em modelos estatísticos. De fato, como aponta Francisco (2010), este viés não é uma exclusividade para a análise dos partidos. Uma “abordagem mais tradicional” da avaliação dos dados sociais, econômicos e ambientais, em geral, utiliza-se de técnicas úteis e válidas de análise, como os testes de hipótese e as análises de variância. Porém, excluem justamente os aspectos locais desses fenômenos. Uma das premissas deste artigo, é que os aspectos geográficos – o “onde” ocorrem os fatos políticos – importam e são relevantes para a sua compreensão.

Assim, busca-se um padrão de distribuição da organização partidária municipal no espaço. A hipótese é que, em municípios onde a organização partidária se encontre na forma em que a literatura considere mais “institucionalizada”, ou seja, com um diretório instalado, esta consiga mobilizar um maior nível de eleitorado local. Desse modo, o objetivo é analisar o poder de explicação do desempenho eleitoral a partir de modelos distintos comparados: (i) regressão clássica, 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) ou Regressão Ponderada Geograficamente.

Com grande potencial de adoção para estudos de fenômenos ligados a políticas públicas, são ainda incipientes os usos efetivos de técnicas de Estatística Espacial nesse contexto. Mais recentemente, Morandi et al. (2016) utilizaram-se de técnicas de autocorrelação espacial e regressão espacial para avaliar a relação entre transporte público sobre trilhos, renda e tempo médio de deslocamento.

Se o debate recente estiver correto, espera-se que a organização partidária municipal instituída por meio de diretórios possua melhor desempenho quando comparada a municípios e partidos de trajetória ou estrutura similar. Mais ainda, espera-se que esta relação (organização x desempenho eleitoral) tenha efeitos distintos entre os diferentes tipos de partido e de municípios. Por fim, busca-se analisar o resultado eleitoral sob a perspectiva de uma eventual associação a outro índice, de cunho social, o Índice Paulista de Vulnerabilidade Social (IPVS), visando a aferir potencial relação dessa variável ao resultado eleitoral.

### 3 REVISÃO DA LITERATURA

Como forma de inclusão do aspecto geográfico em estudos sociais, Mittal, Kamakura e Govind (2004) avaliam quais motivos explicam a satisfação (variável dependente) dos clientes de uma fábrica de automotivos, a partir das variáveis experiência do vendedor, qualidade do veículo e o serviço da concessionária. Porém, o grande diferencial foi considerar o aspecto geográfico nestes modelos. Os autores consideraram que tais aspectos, como vendedor, veículo e concessionária não tinham o mesmo efeito sobre a satisfação dos clientes em todas as regiões dos EUA. Esta conclusão pôde ser feita por meio da construção do índice I de Moran, que verifica a autocorrelação das variáveis no espaço. A presença de autocorrelação encontrada pelos autores significa que as variáveis não são aleatórias no espaço. Em outras palavras, o espaço é uma variável importante na representação e entendimento da satisfação de clientes com o setor automotivo. A partir desta informação, utilizaram o modelo GWR, técnica apropriada para espaços heterogêneos e não estacionários, e realizaram regressões locais para estimar a satisfação de cada distrito norte-americano, considerando suas respectivas amostras locais e as variáveis independentes dos vizinhos. Os autores obtiveram um modelo preditivo bastante eficiente, conseguiram uma boa solução para sustentar a hipótese de exogeneidade dos erros, tomando a informação geográfica como informação relevante. O resultado disso foi um modelo preditivo bastante eficiente.

Francisco, Fagundes, Ponchio e Zambaldi (2010) concebeu um modelo com algum poder preditivo sobre a propensão de fraudes sobre o uso de energia elétrica pela AES Eletropaulo, considerando variáveis socioeconômicas (como o Índice Paulista de Vulnerabilidade Social) e variáveis ligadas ao processo de operação e inspeção de fraudes pela empresa. O modelo espacial utilizado neste caso foi o SAR que inclui a própria variável dependente como termo autorregressivo (defasado espacialmente) na explicação do fenômeno, além de outras variáveis independentes.

Diferentemente do modelo de regressão linear por mínimos quadrados (MQO), onde apenas as variáveis dentro do recorte geográfico a que pertence a unidade de análise são consideradas, os modelos GWR e SAR incluem o efeito das variáveis independentes (ou dependente) de seus vizinhos sobre a variável dependente analisada, por meio do espaço geográfico.

O mero “rodar o modelo”, que já despertou interesses e excessos no passado, é apenas um pequeno detalhe na avaliação das abordagens que se propõem empíricas. Para que o estimador calculado realmente represente o efeito isolado, ou seja, ao buscar-se simular o efeito *ceteris paribus*, tão comum na teoria, deve-se certificar que determinadas hipóteses estejam garantidas. Caso contrário, tem-se apenas um indicador de estatística descritiva e não uma explicação científica de causa, nem um preditor estatisticamente confiável.

## 4 METODOLOGIA

Para o atingimento dos objetivos, foi avaliada a relação entre a existência da organização partidária no município e o desempenho eleitoral, isto é, quantos votos o partido obteve no mesmo município quando comparados com o total de votos obtidos em toda a federação (o “*share*” de votos do partido no município). Para isso, o ano de 2010 foi utilizado como período de análise, pelo advento das eleições para a Câmara dos Deputados.

As bases desta análise são:

- Abrangência: municípios do estado de São Paulo;
- Eleições do ano de 2010 para o cargo de deputado federal;
- Base partidária: partidos PT e PSDB;
- Organização possível: diretório municipal, comissão provisória municipal, nenhuma organização;
- IPVS de 2010.

O escopo territorial selecionado foi o estado de São Paulo, sendo os municípios as unidades de análise. Escolheu-se apenas esse estado como uma forma preliminar de averiguação acerca das hipóteses presentes na literatura. O projeto contou com os dados sobre organização partidária disponíveis no site do Tribunal Superior Eleitoral (TSE) (2010). Para o resultado das eleições, foram utilizados os resultados eleitorais agregados por município disponíveis no site do Centro de Política e Economia do Setor Público da Fundação Getúlio Vargas (Cepesp/FGV) (2010), inicialmente para as eleições de 2010.

A escolha do cargo de Deputado Federal se deu em razão de que, para este, a questão das coligações partidárias não tem efetiva influência, como ocorre em cargos de eleição majoritária (executivo e senado), e também porque, tendo conotação federal, de certa forma se afasta de influências regionais não mensuráveis. A adoção de métricas apenas quanto a PT e PSDB advém do fato de que, nessas eleições em particular, e até de forma geral, são partidos considerados “polo”, hábeis a promover polarização de interesses em torno de si. Assim, acredita-se estar capturando melhor uma heterogeneidade de preferências entre os eleitores. Desse modo, entende-se que resultados de análise georreferenciada envolvendo apenas esses dois partidos têm o condão de apresentar resultados capazes de espelhar boa parte da realidade eleitoral.

Para a realização de tais avaliações, partiu-se de um modelo linear clássico (MQO). Acrescentaram-se dois novos modelos com o intuito de incluir o aspecto espacial na explicação do fenômeno. Os modelos são o SAR, que tem poder de explicação global, a partir de uma matriz de

vizinhança, e o GWR, que efetua um modelo de regressão local para cada município. Os dois métodos espaciais podem ser detalhadamente observados em Francisco (2010). Além do aspecto espacial, incluiu-se o IPVS, índice de alta confiança disponível em bases de dados públicas, hábil a configurar o perfil médio dos habitantes – e eleitores – em cada município paulista, viabilizando essa importante análise de potencial influência da camada social em que o eleitor se encontra, no resultado eleitoral. O IPVS de cada município foi o resultado da média do IPVS de cada distrito, ponderada pelo número de domicílios.

Em resumo, a análise consiste de quatro etapas, conforme detalhado a seguir:

#### 1) Modelo Clássico (MQO)

$$y = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \varepsilon$$

Em que:

$y$ : é o *share* de votos;

$d_1$ : é a variável *dummy* que representa a presença ou não de diretórios;

$d_2$ : é a variável *dummy* que representa a presença ou não de comissões provisórias.

#### 2) Modelo Espacial (SAR)

$$y = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \rho W y + \varepsilon$$

Em que:

$y$ : é o *share* de votos;

$d_1$ : é a variável *dummy* que representa a presença ou não de diretórios;

$d_2$ : é a variável *dummy* que representa a presença ou não de comissões provisórias;

$W$ : é a matriz de vizinhança (contiguidade: *queen case*), que considera a influência (peso) do  $y$  de cada um dos outros distritos (vizinhos) no modelo.

#### 3) Modelo Espacial (SAR) com a inclusão da variável IPVS

$$y = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \rho W y + \beta_3 IPVS + \varepsilon$$

Em que:

$y$ : é o *share* de votos;

$d_1$ : é a variável *dummy* que representa a presença ou não de diretórios;

$d_2$ : é a variável *dummy* que representa a presença ou não de comissões provisórias;

$W$ : é a matriz de vizinhança (contiguidade: *queen rook*), que considera a influência (peso) do  $y$  de cada um dos outros distritos (vizinhos) no modelo;

$IPVS$ : é a média do IPVS de todos os distritos no município.

#### 4) Modelo Espacial (GWR) com a inclusão do IPVS:

$$y_i = \beta_{0i} + X(g_i)\beta(g_i) + \varepsilon$$

Em que:

$y_1$ : é o *share* de votos calculado para todos os municípios;



$X(g_i)$  : é o vetor de variáveis independentes (*dummy* da presença de diretórios, *dummy* da presença de comissões provisórias, e IPVS) dos municípios vizinhos, para cada ponto  $g_i$  (centroide do distrito) do estado de São Paulo.

A técnica empregada considerou a criação de pontos baseados nos centroides dos municípios para realizar a GWR. Ao final da pesquisa, pretendeu-se avaliar: (i) o impacto da organização partidária (que pode ser um diretório, comissão provisória ou nenhuma organização) sobre o número de votos que os partidos PT e PSDB receberam em cada município do estado de São Paulo; e (ii) a influência da condição social média, do eleitorado de cada município paulista, aferida pelo IPVS, no total de votos que os referidos partidos receberam em 2010.

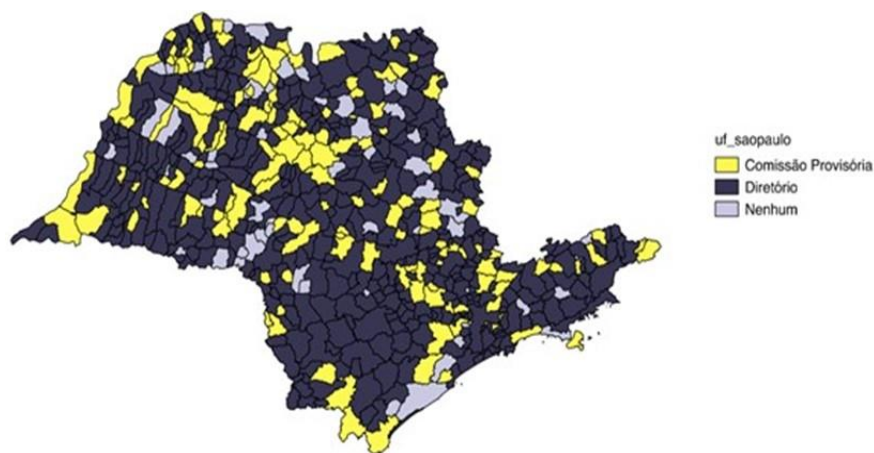
## 5 ANÁLISE EXPLORATÓRIA DOS DADOS

Por meio da utilização das ferramentas geográficas (ArcView GIS, ArcGIS, QGIS, GeoDA e R), foram produzidas algumas imagens capazes de sugerir que há uma correlação, em termos também geográficos, entre o tipo de organização partidária e o desempenho eleitoral. Primeiramente, foi realizado um mapeamento dos tipos de organização político-partidária para os dois principais partidos do estado de São Paulo, o PT e o PSDB. Também, o mapeamento do *share* de cada partido no resultado da eleição municipal para deputado federal. Os resultados estão demonstrados nas Figuras 1 a 4.

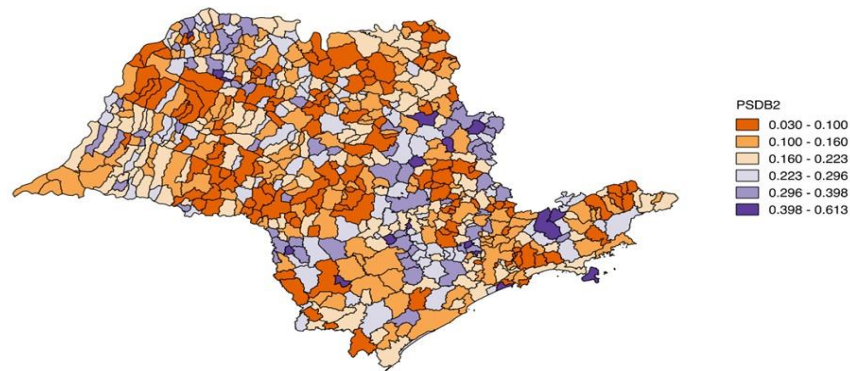
As Figuras 1 e 2 apresentam a organização partidária e o desempenho eleitoral do PSDB (neste caso, o percentual de votos válidos). Notam-se algumas áreas onde a concentração do *share* de votos é maior do que em outras. Isto é, aparentemente a concentração de votos não se dá de forma homogênea ou aleatória no espaço. As Figuras 3 e 4, por sua vez, apresentam a organização partidária e o desempenho eleitoral do PT (neste caso, o percentual de votos válidos).

É nítido o predomínio de diretórios nos dois partidos observados, o partido PT com 462 e o partido PSDB com 442. Trata-se, portanto, de partidos bem organizados no estado de São Paulo, com baixo número de municípios com comissões (o partido PT com 154 e o partido PSDB com 146) ou nenhuma organização municipal (PT, 29 e PSDB, 57). Uma análise preliminar dos dados parece indicar maiores níveis em relação à variável *share* de votos para municípios com diretórios constituídos, e não com meras comissões.

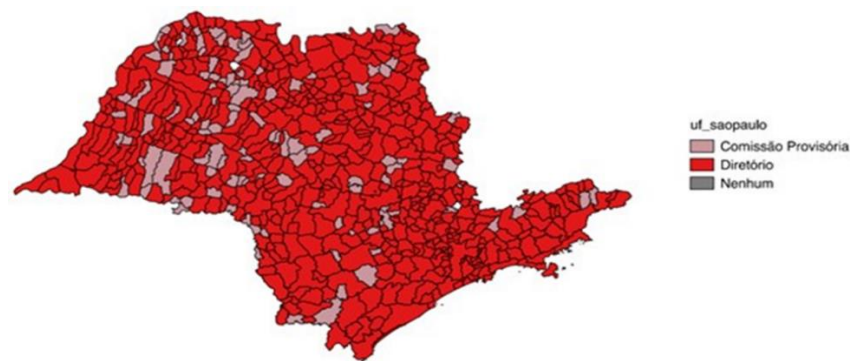
Nota-se que as áreas onde o partido PT domina, em geral, não são áreas onde o partido PSDB tem maior importância em termos de proporção da distribuição dos seus votos. Estes são os resultados esperados.



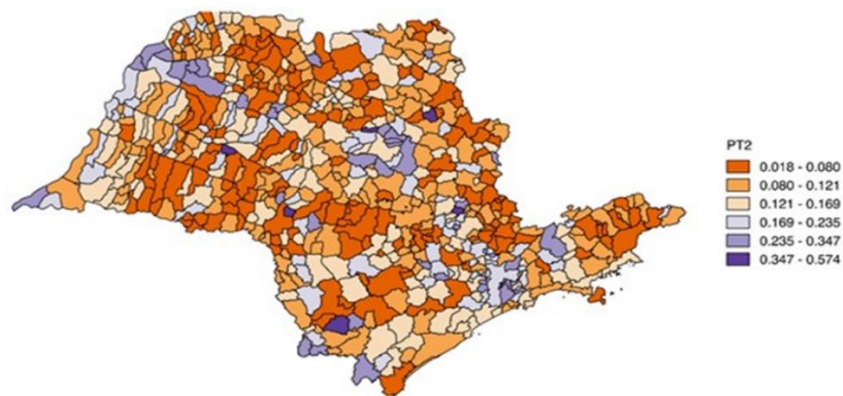
**Figura 1** - Organização Partidária do PSDB em 2010  
Fonte: Os autores, com o uso da ferramenta QGIS.



**Figura 2** – Desempenho Eleitoral (percentual de votos válidos) do PSDB para Deputado Federal em 2010  
Fonte: Os autores, com o uso da ferramenta QGIS.



**Figura 3** – Organização Partidária do PT em 2010  
Fonte: Os autores, com o uso da ferramenta QGIS.



**Figura 4** – Desempenho Eleitoral (percentual de votos válidos) do PT para Deputado Federal em 2010  
Fonte: Os autores, com o uso da ferramenta QGIS.

## 6 AUTOCORRELAÇÃO ESPACIAL DOS VOTOS E INCLUSÃO DA VARIÁVEL ESPAÇO NOS MODELOS

Não é possível, contudo, prosseguir com a análise dos modelos sem antes fazer uma referência à distribuição espacial da variável dependente. Francisco (2010) aponta que, em regra, a análise espacial aborda duas classes de dados: ambientais e socioeconômicos. Dados ambientais, basicamente, resultantes de levantamentos de recursos e fenômenos da natureza, justamente por sua característica de estacionariedade, perenidade, são tratados no âmbito da Geoestatística. Não se trata deste caso.

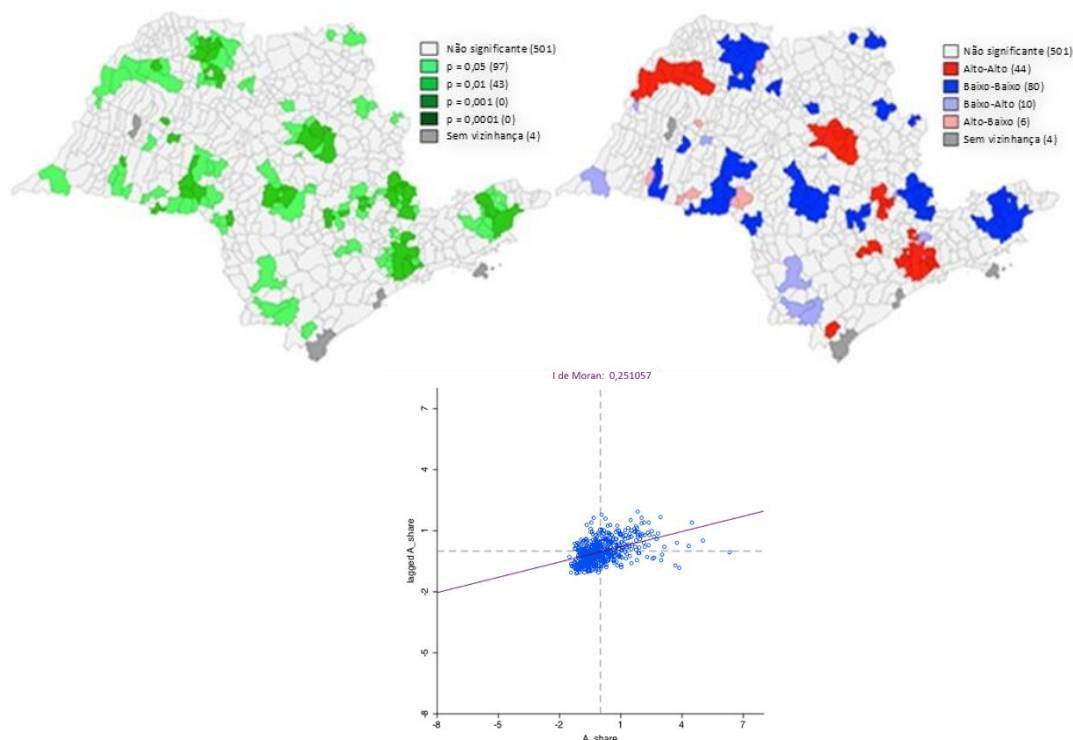
Dados socioeconômicos, aqueles que podem incluir as organizações partidárias, as eleições e seus resultados – são de interpretação bem mais complexa: advêm de censos e amostras sobre indivíduos

e coletividades heterogêneas e em contínua alteração. Por este motivo, têm de ter tratamento distinto, mais acurado, sob as técnicas da Estatística Espacial. Nessa linha, Francisco (2010) assevera que os dados socioeconômicos podem ter duas abordagens: (i) como um conjunto de polígonos homogêneos, disjuntos e adjacentes, cobrindo toda a área em estudo, com atributos descritivos para cada região, e (ii) como um conjunto de amostras (onde cada amostra está associada a um ponto [centroide, sede ou arbitrário] da unidade de levantamento), revelando a complexidade de seu tratamento e análise.

Nesse sentido a diferença fundamental entre a Geoestatística e a Estatística Espacial diz respeito ao conceito de estacionariedade espacial. Um processo aleatório estacionário é, em semelhança às análises que envolvam o estudo da distribuição da variável no tempo (séries temporais), identificado por apresentar propriedades estatísticas (média, variância, distribuição, correlações etc.) que não variam no espaço. Assim, a abordagem Geoestatística representa um conjunto de técnicas para se construir uma superfície com base numa caracterização da similaridade espacial entre as amostras, dado que estuda fenômenos espaciais contínuos e estacionários no espaço. A Estatística Espacial considera a heterogeneidade da distribuição da variável no espaço, sendo, portanto, mais apropriada para grande parte dos fenômenos sociais e socioeconômicos (Francisco, 2010).

Isso posto, para que se possa entender como os votos são influenciados em cada local, deve-se observar como estão espacialmente distribuídos, visando detectar as correlações possíveis dessa variável com ela mesma, a correlação entre um valor de uma observação e o valor das observações próximas dessa mesma variável e ainda o grau em que as características de uma localização são semelhantes – ou distintas – daquelas nas proximidades. As Figuras 5 e 6 apresentam alguns elementos da autocorrelação espacial do *share* de votos de cada partido e de sua significância estatística.

É possível verificar um Índice I de Moran do desempenho eleitoral entre baixo e moderado para os dois partidos (25,1% para PT e 23,8% para PSDB). Porém, nota-se a existência de alguns agrupamentos (*clusters*) onde o padrão de votação parece ser mais comum. Esse comportamento diferenciado para cada *cluster* encontrado pode explicar o baixo Índice de Moran global. Este também parece ser um indício de que modelos locais de regressão espacial apresentarão melhor resultado.

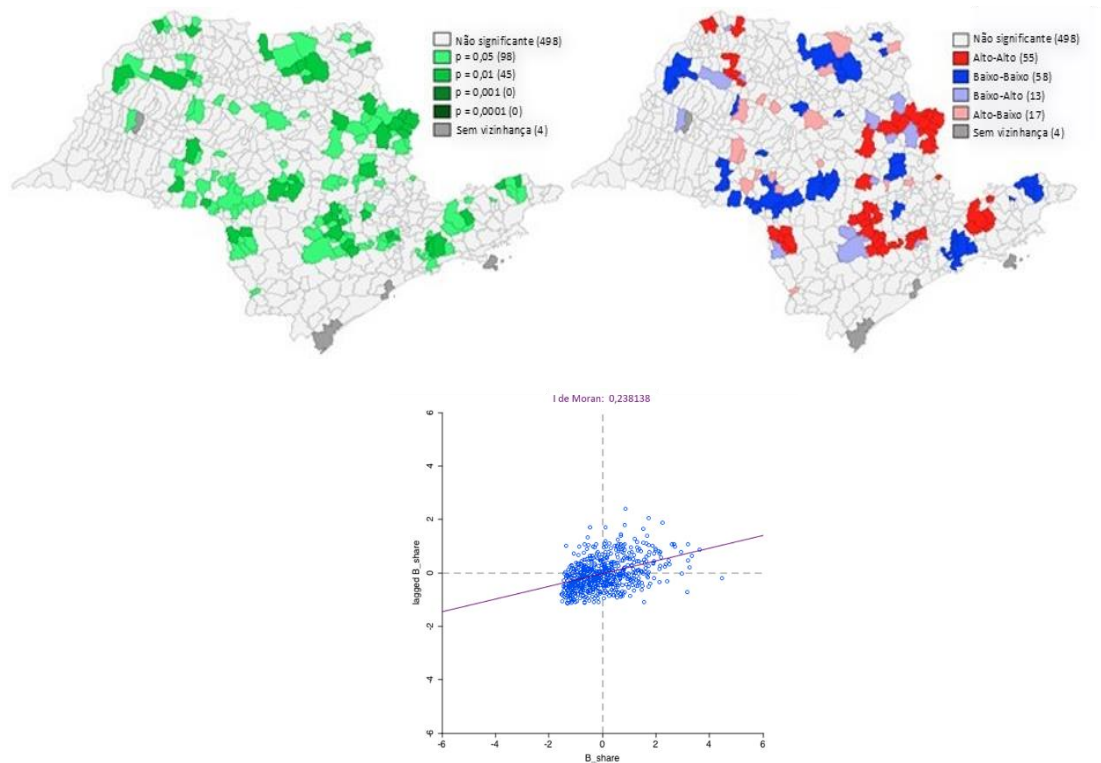


**Figura 5** – Autocorrelação do Share de Votos para Deputado Federal do PT em 2010: Significância (esquerda), Categoria (direita) e I de Moran (centro)

Fonte: Os autores, com o uso da ferramenta GeoDA.



Algumas considerações, entretanto, merecem referência nesta ocasião. Por que se acredita ser necessária a adoção da autocorrelação nessas análises? Deve-se retornar um pouco e fazer-se referência ao método de Mínimos Quadrados Ordinários (MQO) linear clássico, sem dúvidas, o método estatístico de análise mais utilizado nas Ciências Sociais. Tal método constrói uma relação (inicialmente teórica) entre a variável dependente (que se pretende explicar) e a variável explicativa (exógena ao modelo e, como o pressuposto da racionalidade, não se pretende inicialmente explicar).



**Figura 6** – Autocorrelação do Share de Votos para Deputado Federal do PSDB em 2010: Significância (esquerda), Categoria (direita) e I de Moran (centro)

Fonte: Os autores, com o uso da ferramenta GeoDA.

Cabe frisar aqui que o método MQO procura definir uma curva entre todos os pontos observados da variável independente que se relaciona com todos os pontos observados da variável dependente. Esta curva é construída a partir da “minimização do erro”, isto é, a mínima distância desta com todos os pontos observados da variável dependente. Porém, o simples cálculo da curva não necessariamente conduz a um modelo explicativo e preditivo da variável dependente. O mero cálculo leva a uma correlação entre variáveis, uma informação que pode ter bastante relevância nesta análise, mas não necessariamente conduz aos objetivos principais a que se fez referência no início deste texto. Cabe, a esta altura, citar o ensinamento mais valioso da econometria em forma de jargão: “correlação não implica causalidade”, pois isolar a causalidade tem sido um dos grandes desafios das Ciências Sociais.

Voltando ao modelo de OLS, o que garante a relação causal entre variáveis não é a matemática e sim a garantia de determinados pressupostos teóricos que garantam a validade estatística do modelo. A base teórica mais famosa e importante, são os pressupostos de *Gauss-Markov*. Alguns desses pressupostos garantem apenas a estruturação do modelo, como a linearidade nos parâmetros (e não nas variáveis, o que dá a liberdade de construir modelos mais precisos) e aleatoriedade na amostragem. Quando se observa mais de uma variável independente, é preciso garantir que as variáveis também não sejam perfeitamente relacionadas (multicolineares), de forma que a variável dependente entenda apenas o efeito de uma variável, mas também é necessário que a variável explicativa “varie”, isto é, não tenha variância igual a zero.

Os pressupostos mencionados não são inteiramente problemáticos nem têm sido os maiores desafios nos estudos sociais. Porém, outros pressupostos são mais problemáticos. Talvez o mais

problemático seja o que exige que a variável dependente não seja correlacionada com os erros do modelo. Isso significa que, caso se tenha uma correlação entre erro e a explicativa, significa que o modelo possui algum tipo de viés, isto é, alguma outra variável afeta a variável, mas não foi considerada no modelo. A forma de tentar “adequar” o modelo, é considerar esta variável. Ou seja, transformá-la em uma variável independente para retirar seu efeito do termo de erro. É o que se chama de “variável controle”. O problema disso é que, em geral, é muito difícil controlar todos os efeitos sobre a variável dependente, ou seja, dispor de todas as variáveis de controle necessárias.

Quando se trabalha com amostragem, isto é, coleta-se apenas parte dos dados de forma aleatória, de tal forma que este conjunto represente toda a população de dados, tem-se que assumir mais dois pressupostos quanto à inferência dos estimadores (ou seja, se os parâmetros estimados representam mesmo a população). O primeiro deles, é que o conjunto amostral tem distribuição normal. Como pelo Teorema do Limite Central tem-se que, em geral, conforme o conjunto de dados aumente, a distribuição de dados tende à distribuição normal, esta hipótese, em geral, se resolve pela quantidade da amostra. Outro pressuposto, mais complexo, é de que os dados possuem dispersão constante (homocedasticidade). Alguns testes mais complexos verificam se o modelo tem variância constante; e alguns modelos (também complexos) adaptam modelos com variância não constante (heterocedásticos).

Considere-se agora uma hipótese nada improvável de que o espaço afete a variável dependente. O que aconteceria com o modelo de MQO? Mesmo garantindo todos os outros pressupostos, existiria o viés da variável não controlada (o espaço). É, pois, a autocorrelação espacial quem avalia e analisa o grau de dependência entre observações no espaço. Em geral, relações de dependência espacial e seus efeitos são estimados a partir de matrizes de vizinhança das áreas estudadas em um mapa. Diversos são os critérios utilizados para definir as matrizes de vizinhança. Uma forma é determinar a matriz com base nas fronteiras, por meio de uma variável dicotômica (*dummy*) que representaria a presença ou ausência de vizinhança entre as localidades. O índice I de Moran é um dos indicadores mais utilizados como aproximação da medida de autocorrelação espacial calculada a partir da matriz de vizinhança. Neste caso, indica a correlação entre o valor do *share* de votos de uma variável observada em uma região e os valores desta mesma variável observados em suas regiões vizinhas (Francisco, 2010).

Dado a constatação da presença de autocorrelação espacial, conforme as Figuras 5 e 6 parecem indicar, torna-se possível a inclusão de dois modelos. A primeira forma é o modelo espacial autorregressivo ou Spatial Auto-Regressive model (SAR), que utiliza a incorporação da matriz de vizinhança espacial como parte das variáveis independentes do modelo de regressão. Por fim, outra técnica adotada é o Geographically Weighted Regression (GWR) ou Regressão Ponderada Geograficamente que descreve “uma família de modelos de regressão em que os coeficientes, parâmetros  $\beta$ , podem variar espacialmente”. O modelo estima uma regressão para cada ponto observado, ponderando as demais observações em função da distância deste ponto. Neste estudo, considerou-se o centroide de cada município. O resultado do modelo é a estimação de diferentes parâmetros betas para cada município, com diferentes valores (Francisco, 2010).

## 7 RESULTADO DOS MODELOS DE REGRESSÃO

Relacionam-se nas tabelas a seguir, o resultado dos modelos (i) MQO, (ii) SAR e (iii) GWR, especificados na metodologia deste artigo. As Tabelas 1 e 2 mostram alguma significância da presença de diretórios em maior ganho de votos no município. Porém, o R-Quadrado desses modelos demonstra pouco poder de explicação entre as duas variações. Nos modelos referentes às Tabelas 3 e 4, incluiu-se a matriz de vizinhança nos modelos anteriores. O ganho em poder de explicação é bastante expressivo. O R-Quadrado sobe de 2,67% para 19,14% no caso do PT e de 3,05% para 15,30% no caso do PSDB. Nestes dois últimos casos, a presença de um diretório instalado no município continua estatisticamente significativa com relação à presença de um maior nível do *share* de votos.

**Tabela 1** – Estimadores do Modelo Clássico (i) – PT

Variáveis	Estimador	Erro Padrão	Est. Z	P-Valor	
$\beta_0$	0,1081699	0,005159	20,96	0,00	***
$\beta_1$	0,02633798	0,006096	4,32	0,00	***
R-Quadrado-Ajustado:	0,028210				

\*\*\* Significante a 1% de erro

(Nenhum estimador é estatisticamente significativo a 10% de erro)

**Tabela 2** – Estimadores do Modelo Clássico (i) – PSDB

Variáveis	Estimador	Erro Padrão	Est. Z	P-Valor	
$\beta_0$	0,1559693	0,006710	23,24	0,00	***
$\beta_1$	0,0373700	0,008106	4,60	0,00	***
R-Quadrado-Ajustado:	0,030486				

\*\*\* Significante a 1% de erro

(Nenhum estimador é estatisticamente significativo a 10% de erro)

Uma observação importante é que, realizando o teste de Breusch-Pagan, verifica-se um Valor-P (significância observada) de 0,19 para o caso do partido PT (Tabela 3) e um Valor-P de 0,12 para o caso do partido PSDB (Tabela 4), não é possível rejeitar a hipótese de homocedasticidade para ambos os modelos. Portanto, para os dois casos, os estimadores são robustos quanto à variância do erro. O mesmo, entretanto, não pode ser dito para os resultados da Tabela 5, quando se inclui a variável IPVS nos modelos SAR.

**Tabela 3** – Estimadores do Modelo SAR (ii) – PT

Variáveis	Estimador	Erro Padrão	Est. Z	P-Valor	
$\beta_0$	0,0516503	0,007513	6,87	0,00	***
$\beta_1$	0,0373700	0,005547	3,87	0,00	***
$\rho$	0,4829759	0,046009	10,49	0,00	***
R-Quadrado:	0,191395				

\*\*\* Significante a 1% de erro

(Nenhum estimador é estatisticamente significativo a 10% de erro)

**Tabela 4** – Estimadores do Modelo SAR (ii) – PSDB

Variáveis	Estimador	Erro Padrão	Est. Z	P-Valor	
$\beta_0$	0,0516503	0,007513	6,87	0,00	***
$\beta_1$	0,0215348	0,005547	3,87	0,00	***
$\rho$	0,4829759	0,046009	10,49	0,00	***
R-Quadrado:	0,153016				

\*\*\* Significante a 1% de erro

(Nenhum estimador é estatisticamente significativo a 10% de erro)

**Tabela 5** – Estimadores do Modelo SAR com o IPVS (iii) – PT

Variáveis	Estimador	Erro Padrão	Est. Z	P-Valor	
$\beta_0$	0,0185795	0,01051	7,95	0,00	***
$\beta_1$	0,0192943	0,005525	4,36	0,00	***
$\rho$	0,4637989	0,046162	8,31	0,00	***
$\beta_3$	0,0135470	0,003297	4,10	0,00	***
R-Quadrado:	0,209177				

\*\*\* Significante a 1% de erro

(Nenhum estimador é estatisticamente significativo a 10% de erro)

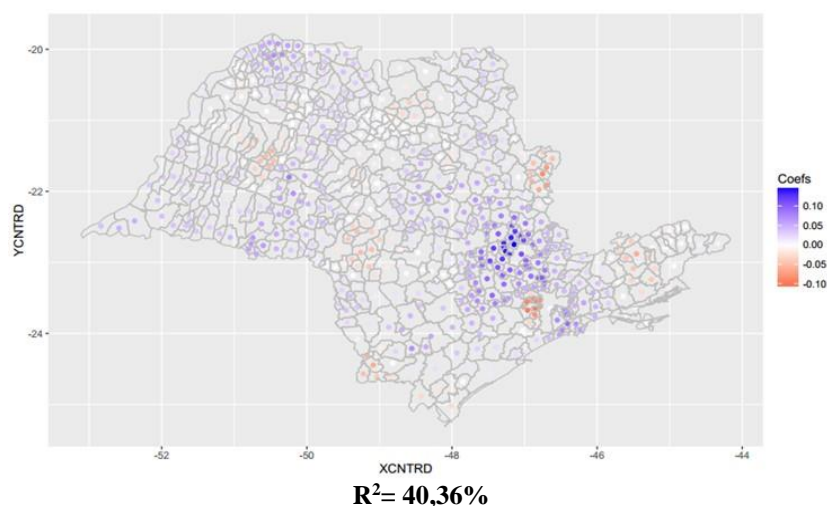
**Tabela 6** – Estimadores do Modelo SAR com o IPVS (iii) – PSDB

Variáveis	Estimador	Erro Padrão	Est. Z	P-Valor	
$\beta_0$	0,0993961	0,016456	6,03	0,00	***
$\beta_1$	0,0335725	0,007583	4,42	0,00	***
$\rho$	0,4071884	0,0492275	8,27	0,00	***
R-Quadrado:	0,154182				

\*\*\* Significante a 1% de erro

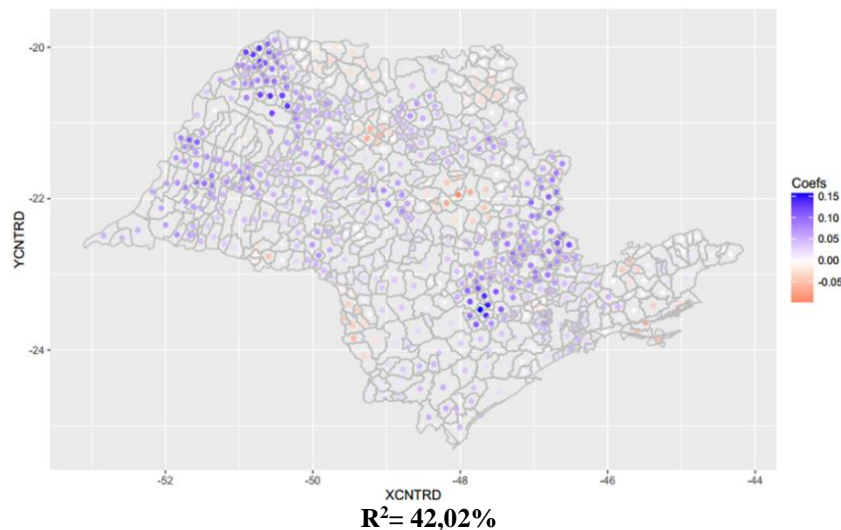
(Nenhum estimador é estatisticamente significativo a 10% de erro)

Um aspecto interessante nas Tabelas 5 e 6, é que a variável não é estatisticamente significativa para explicar o comportamento do partido PSDB (por este motivo é expurgada da Tabela 6), mas é significativa para o partido PT (Tabela 5). Todavia, não é possível assegurar robustez para este resultado, tendo em vista que não se consegue manter a hipótese de homocedasticidade para o caso do PT. Assim sendo, não é possível concluir diferença da importância da variável entre os dois partidos. As Figuras 7 e 8 destacam alguns resultados da aplicação dos modelos GWR.



**Figura 7** – Distribuição espacial do parâmetro “dummy da presença de diretórios” da regressão GWR do desempenho eleitoral para Deputado Federal em 2010 para o PT





**Figura 8** – Distribuição espacial do parâmetro “dummy da presença de diretórios” da regressão GWR do desempenho eleitoral para Deputado Federal em 2010 para o PSDB

Os resultados foram particularmente relevantes para a regressão realizada por meio da técnica GWR, mediante o emprego do *software* R 3.1.3 e o pacote *spgwr*. Nos dois mapas que foram criados utilizando essa metodologia, percebe-se a formação de *clusters* em diferentes regiões para cada um dos partidos. Os modelos GWR elevam os R-Quadrados para cada modelo para um patamar de 0,40 (PT) e 0,42 (PSDB), mais do que o dobro quando comparados com o modelo SAR nos dois primeiros casos. O resultado parece refletir o que já foi observado nas análises de autocorrelação, conforme Figuras 5 e 6: não existe um padrão global de votação onde mais diretórios respondam, necessariamente a mais votos. O que existe, de fato, são *clusters* onde a área de influência dos diretórios é maior, e algumas áreas onde, de forma surpreendente, em algumas regiões, a presença de diretórios acarreta em menor desempenho eleitoral.

## 8 CONSIDERAÇÕES FINAIS

Para que o estimador calculado realmente represente o efeito isolado do desempenho do partido, ou seja, se for mesmo necessário simular o efeito *ceteris paribus* da organização partidária sobre o desempenho do partido nas eleições, deve-se certificar que determinadas hipóteses estejam garantidas. Caso contrário, ter-se-á apenas um indicador de estatística descritiva e não um preditor estatisticamente confiável. Dentre os pressupostos exigidos, talvez o mais complicado seja garantir que o modelo não possua viés sobre suas variáveis. Ou seja, que todo o efeito captado pelo parâmetro seja dado apenas por suas variáveis independentes. Este problema tem repercutido sobre estudos recentes que, ao invés de tentar controlar todas as independentes possivelmente relacionadas ao modelo, o que também traz um efeito indesejado ao aumentar a variância do estimador, tentam trabalhar o desenho a que a pesquisa está relacionada.

Uma maneira de tentar “controlar” o viés de variável omitida e que vem se mostrando bastante eficiente, como tentou-se demonstrar aqui, tem sido considerar o aspecto espacial nos modelos. Tal inclusão somente é possível com o desenvolvimento tecnológico observado, desde o final do século passado, em instrumentos informacionais. A relevância informação espacial vai além da própria informação espacial que, na maioria das análises, é omitida, pois consegue trazer informações quase intangíveis, tais como a “cultura” local, ou as “preferências dos consumidores”. Ora, se tais informações são relevantes nas explicações das relações sociais e estão correlacionadas ao locus geográfico, trazer a variável geográfica minimiza o viés contido no termo de erro desses modelos, reduzindo a incerteza e, portanto, melhorando a explicação causal e predição dos modelos, objetivos primordiais dos modelos aplicados. Os resultados demonstram uma alternativa válida e ainda a ser bastante utilizada para melhorar a capacidade de interpretação sobre uma realidade complexa.

Quando se reflete sobre a organização partidária, muitos aspectos tendem a ficar de fora nesse modelo. Em outras palavras, muitos são os vieses de omissão nos modelos. Como os partidos se adaptam para conseguir os votos de seus eleitores? Esta “conquista dos votos” ou formação de *constituencies*, se dá, de forma geral, em nível local. A partir do momento em que se inclui a variável espacial, também se está incluindo parte dos aspectos locais de cada eleitorado. Este é o motivo da melhora nesses modelos.

Neste estudo, buscou-se incluir o aspecto geográfico na explicação do desempenho eleitoral dos partidos PT e PSDB, em conjunção com sua distribuição de estruturas partidárias. Foram encontrados indícios de que a organização partidária organizada como diretório pode ser um bom previsor para desempenho eleitoral, mas sobretudo em algumas regiões específicas do estado de São Paulo. As organizações partidárias denominadas Comissões Provisórias não foram observadas como significativas, muito embora não houvesse, para essa variável, graus de liberdade suficientes para uma análise mais robusta. Como lição fundamental, vislumbra-se o potencial que a dimensão espacial confere a análises estatísticas. Partiu-se de um modelo com capacidade explicativa muito baixa, em direção ao modelo GWR, em que se concluiu que não se trata de um fenômeno aleatório no espaço, mas fato comprovável, dado o aumento significativo do R-Quadrado e a configuração em *clusters* da distribuição espacial dos desempenhos eleitorais dos dois partidos analisados. Conclui-se que a introdução da dimensão espacial aprimora, em muito, a qualidade dos modelos de explicação do desempenho eleitoral.

## REFERÊNCIAS

- Centro de Política e Economia do Setor Público da Fundação Getulio Vargas. (2010). *Resultados eleitorais agregados por município*. Recuperado de <http://cepesp.fgv.br/pt-br/nossos-bancos>
- Francisco, E. de R. (2010). *Indicadores de renda baseados em consumo de energia elétrica: Abordagens domiciliar e regional na perspectiva da estatística espacial* (Tese de Doutorado em Administração de Empresas). Escola de Administração de Empresas de São Paulo da Fundação Getulio Vargas, São Paulo, SP, Brasil.
- Francisco, E. de R., Fagundes, E. B., Ponchio, M. C., & Zambaldi, F. (2010). Development of an indicator of propensity to energy commercial losses using geospatial statistical techniques and socio-economic data: the case of AES Eletropaulo. *RAM. Revista de Administração Mackenzie*, 11(4), 178-197.
- Guarnieri, F. H. E. (2009). *A força dos “partidos fracos”- Um estudo sobre a organização dos partidos brasileiros e seu impacto na coordenação eleitoral* (Tese de Doutorado em Ciência Política). Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, São Paulo, SP, Brasil.
- Mittal, V., Kamakura, W. A., & Govind, R. (2004). Geographic patterns in customer service and satisfaction: An empirical investigation. *Journal of Marketing*, 68(3), 48-62.
- Morandi, E., Ribeiro, R., Hernandez, E., Camara, B., Spinola, L., & Francisco, E. de R. (2016). *Análise Geoespacial da relação entre transporte público sobre trilhos, renda e tempo médio de deslocamento*. XL Encontro da EnANPAD.
- Ribeiro, P. F. (2013). Organização e poder nos partidos brasileiros: Uma análise dos estatutos. *Revista Brasileira de Ciência Política*, n. 10, 225–265.
- Sousa Braga, M. do S., & Pimentel, J., Jr. (2013). Estrutura e organização partidária municipal nas eleições de 2012. *Cadernos ADENAUER*, 2, 13–36.

Sousa Braga, M. do S., Rodrigues-Silveira, R., & Borges, T. (2012). Organización, territorio y sistema partidario: Difusión territorial de la organización de los partidos y sus potenciales impactos sobre la estructura del sistema partidario en Brasil. *América Latina Hoy*, 62(18).

Tribunal Superior Eleitoral. (2010). *Dados sobre organização partidária*. Recuperado de <http://www.tse.jus.br/>