

Quem quer ser um deputado?

Uma análise multinível das eleições para a Câmara dos Deputados em 2022

Felipe Marques Esteves Lamarca^{*,†}

Resumo

O que aumenta as chances de um candidato se eleger? A literatura especializada da Ciência Política procurou responder esta questão através de diversas abordagens ao longo do tempo. Este trabalho faz uma análise estatística dos dados eleitorais de 2022 para avaliar o que impacta nas chances de candidatos a deputado federal serem eleitos no Brasil. São utilizadas técnicas de modelagem estatística, como a aplicação de modelos de regressão logística multinível e métricas de avaliação da capacidade explicativa e preditiva de modelos. Os resultados sugerem que as despesas de campanha, com variações entre os partidos, explicam boa parte dos resultados nas urnas.

Palavras-chave: Eleições Legislativas; Regressão Logística; Modelos Multinível; Avaliação de Modelos.

Introdução

Uma das questões mais fundamentais que se colocam sobre sistemas políticos é o que leva um candidato a ser vitorioso nas eleições em detrimento de outros. Entender as razões do sucesso eleitoral é importante para os *players* em disputa: conforme a Escola da

^{*}Escola de Matemática Aplicada, FGV EMAp.

[†]Escola Superior de Ciências Sociais, FGV CPDOC.


Teoria da Escolha Racional (DOWNS, 2013), os atores agem racionalmente e, portanto, os partidos políticos e candidatos moldam a forma como se apresentam ao eleitorado na medida em que conhecem os fatores que conduzem à eleição.

Além disso, a resposta para essa pergunta é importante para a compreensão de questões mais gerais dos regimes políticos — se, por exemplo, a ideologia dos candidatos é o fator determinante para que eles recebam ou percam votos, então essa é uma característica importante do regime como um todo, porque é aquilo que o eleitorado confere mais importância ao computar seu voto.

Naturalmente, é improvável que uma única variável explique o sucesso eleitoral dos candidatos. Conforme sugere a literatura, é de se esperar que o sucesso eleitoral se deva a uma complexa conjunção de uma série de variáveis de naturezas distintas: socioeconômicas, regionais, conjunturais etc. Em particular, a literatura é enfática ao afirmar que as despesas de campanha influenciam bastante em um bom desempenho nas urnas, apesar de não garantirem vitória.

Ainda assim, as eleições de 2022 foram diferentes das anteriores em vários sentidos. Cabe destacar, principalmente, que o alto grau de polarização do eleitorado já vinha alterando a dinâmica das disputas para cargos eletivos desde 2018. Na ocasião, candidatos a cargos legislativos conseguiram angariar grandes quantidades de votos mesmo dispondo de baixo efetivo financeiro para gastos de campanha, devendo seu sucesso eleitoral a outros fatores — por exemplo, o apoio a um ou outro candidato ao Executivo. O cenário em 2022, no que diz respeito à polarização, talvez tenha sido ainda mais crítico.

Em vista dessa questão, este trabalho propõe uma análise das eleições para deputado federal em 2022 através do ajuste de modelos de regressão logística multinível. Trata-se de uma tentativa de elencar quais covariáveis explicam melhor o sucesso eleitoral dos candidatos à Câmara dos Deputados nas últimas eleições do país utilizando uma abordagem capaz de captar o quadro geral dos dados, sem ignorar informações mais específicas — por estado ou partido, por exemplo.

Para isso, serão utilizados os dados eleitorais disponibilizados pelo Tribunal Superior Eleitoral (TSE). O acesso foi mediado pela Base dos Dados, que disponibiliza as informações de forma estruturada e sob fácil acesso através de uma API. O *dataset* final é resultado da composição de quatro tabelas — `candidatos`, `resultados_candidato`, `bens_candidato` e `despesas_candidato`. Assim, foi possível compor um conjunto de informações socioeconômicas de cada candidato, bem como os respectivos resultados nas urnas. Os códigos em R e Python utilizados para extrair os dados, tratá-los e ajustar modelos a partir deles pode ser encontrado neste repositório do GitHub ¹.

Este trabalho está subdividido em mais quatro seções, além desta introdução. Na (i) primeira, faço uma breve revisão da literatura que mapeia, a partir de diferentes abordagens, os determinantes para o sucesso eleitoral para diferentes cargos eletivos. Em

¹ <https://github.com/felipelmc/ElectoralDynamics>

seguida, (ii) apresento a abordagem utilizada nesta empreitada e discuto em que medida ela contribui para a literatura disponível. Na (iii) terceira seção, discuto os resultados obtidos e termino na (iv) quarta seção retomando os principais achados do trabalho, suas limitações e sugerindo possíveis caminhos para pesquisas futuras.

1 Revisão de literatura

A introdução deste trabalho enuncia a questão que se pretende discutir ao longo deste relatório: quais fatores mais influenciam no sucesso de um candidato na disputa para a Câmara dos Deputados? A literatura da Ciência Política vem discutindo o tema exaustivamente, justamente pela importância dessa resposta para a realização de diagnósticos sobre os sistemas representativos de forma geral.

No contexto de países nos quais as eleições legislativas são realizadas utilizando o modelo de lista aberta, reconhecido por incentivar a competição intrapartidária, bem como a personalização do voto e das campanhas eleitorais (NICOLAU, 2006), passa a ser importante determinar quais fatores convergem para o sucesso eleitoral de determinados candidatos em detrimento de outros.

As várias pesquisas nacionais que se debruçaram sobre este assunto frequentemente consideram que variáveis econômicas, em particular as despesas de campanha, têm efeito positivo para explicar os votos nos candidatos (FILHO, 2005; MANCUSO, 2015). O artigo pioneiro que procurou estimar o efeito das despesas de campanha no voto foi Samuels (2001), utilizando o Método dos Mínimos Quadrados Ordinários. No entanto, publicações mais recentes avaliam que o método utilizado por Samuels não era ideal por assumir que o número de votos de um candidato é de natureza contínua, e não discreta, como é o caso. Essas críticas levaram à utilização de outros modelos para estimar o impacto que as covariáveis escolhidas surtem no número de votos, como Poisson ou Binomial Negativo (KENNEDY, 2008; ARRAES; NETO; SIMONASSI, 2017).

A despeito das diferenças relativas aos desenhos de pesquisa — ajustes de modelos lineares ou lineares generalizados; análise de diferentes cargos eletivos; ou opções por diferentes níveis de agregação dos dados —, variáveis econômicas parecem ter grande potencial para explicar resultados eleitorais. Trata-se de uma análise que se pretende aplicar aos dados eleitorais de 2022, a fim de verificar se as especificidades conjunturais das últimas eleições alteraram de alguma forma esse diagnóstico.

Além disso, parte da literatura também investiga as potenciais diferenças de efeito das despesas de campanha entre candidatos incumbentes, isto é, que já possuem mandato, e desafiantes (MENEZES, 2010; SPECK; MANCUSO, 2013). Na etapa de manipulação dos dados de 2022, criou-se uma nova coluna do *dataset* para possibilitar a realização desse mesmo teste em dados mais recentes. A seguir, detalho a metodologia utilizada no desenvolvimento das análises.

2 Desenho de pesquisa e metodologia

Ao analisar as eleições de 2010 para deputados estaduais no Ceará, Arraes, Neto e Simonassi (2017) justificam o limitado escopo de análise para “[...] maximizar a homogeneidade da amostra.” (ARRAES; NETO; SIMONASSI, 2017, p. 1061). Segundo os autores, as diferenças nas dinâmicas eleitorais entre os estados impactariam negativamente a acurácia das estimativas obtidas, o que justificaria analisar apenas um estado em apenas um ano eleitoral. No entanto, para compreender a dinâmica nacional das eleições de deputados federais, a abordagem mais adequada seria a utilização de modelos multinível. Notadamente, esse tipo de modelo torna possível obter estimativas gerais, ainda levando em consideração as diferenças regionais, caso elas existam.

Em linha com a literatura que procura medir o impacto de covariáveis socioeconômicas e políticas no resultado eleitoral, este artigo faz uma análise dos dados das eleições para deputados federais ao nível nacional, ou seja, considerando todas as unidades federativas. Para isso, serão ajustados modelos multinível de regressão logística, o que pressupõe que a variável resposta seja modelada de forma binária, indicando se o candidato foi eleito ou não. Os modelos multinível serão ajustados considerando diferentes componentes: agrupamentos por estado e partido, com interceptos e/ou coeficientes angulares variáveis. Essa abordagem é importante para avaliar de que forma as informações disponíveis sobre os candidatos interferem na capacidade de cada um deles se eleger.

No que diz respeito à abordagem utilizada para realizar inferências estatísticas sobre os dados, optou-se pelo ajuste de modelos frequentistas. Essa escolha se justifica por alguns motivos: (i) a simplicidade de se ajustar os modelos no R e a eficiência computacional; (ii) o uso frequente dessa abordagem na literatura, o que possibilita replicações e comparações mais diretas de resultado; e (iii) a maior familiaridade do autor com o ajuste de modelos frequentistas usando a linguagem R.

Para avaliar a qualidade e as capacidades explicativa e preditiva dos modelos, serão utilizadas algumas métricas clássicas como o AIC, R^2 , AUC e acurácia. O AIC, ou Akaike Information Criterion, é uma forma simples de calcular a qualidade do modelo, e é dado por $2k - 2\ln(\hat{L})$, sendo k o número de parâmetros a serem estimados no modelo e \hat{L} o valor que maximiza a verossimilhança do modelo. O valor do AIC por si só não é muito informativo, ao passo que é utilizado mais frequentemente para que seja possível comparar modelos — dadas algumas opções de modelos para os dados, prefere-se o modelo de menor AIC.

O cálculo do R^2 para modelos multinível não é uma tarefa direta, e por isso foi utilizado o pacote do R, `MuMIn`, mais especificamente a função `r.squaredGLMM()`. Trata-se de uma medida de bondade de ajuste que indica a proporção da variabilidade da variável dependente que pode ser explicada pelas covariáveis do modelo. Por fim, o AUC e a acurácia são calculados separando os dados em conjuntos de treinamento e teste na proporção 70%/30%. Após ajustar os modelos aos dados de treinamento, calcula-se a

acurácia a partir dos acertos das classes no conjunto de teste. O AUC, baseado na ROC curve, identifica a área sob a curva e resume o quão bem o modelo consegue distinguir duas classes.

3 Resultados

3.1 Covariáveis e Variável Resposta

Antes de detalhar os modelos ajustados neste artigo e seus resultados, é importante fazer uma análise geral das covariáveis e identificar aquelas que parecem estar relacionadas à variável resposta. A tabela abaixo faz um detalhamento mais geral das variáveis do *dataset* utilizado:

Variável	Especificação	Domínio
ano	Ano da disputa eleitoral	[2022]
sigla_uf	Sigla da unidade federativa	Estados brasileiros
id_candidato_bd	Identificador único do candidato na Base dos Dados	—
nome_urna	Nome do candidato na urna	—
sigla_partido	Sigla do partido do candidato	—
cargo	Cargo disputado	“deputado federal”
genero	Sexo biológico do candidato	[“masculino”, “feminino”]
raca	Raça do candidato	[“branca”, “parda”, “preta”, “indígena”, “amarela”, “nao informado”]
resultado	Resultado final do candidato	0 = não eleito; 1 = eleito
votos	Número de votos no estado	[0, 1.492.047]
bens_declarados	Bens declarados (R\$)	[0, 158.184.458,79]
despesas_campanha	Despesas de campanha (R\$)	[1,83, 3.633.254,13]
concorreu_18	Indica se o candidato concorreu para deputado federal em 2018	0 = não; 1 = sim
tentando_reeleicao	Indica se o candidato está tentando reeleição para deputado federal	0 = não; 1 = sim

Tabela 1 – Especificações das variáveis disponíveis para a análise

A análise exploratória dos dados mostrou que algumas covariáveis estão fortemente

correlacionadas à variável resposta (no caso, **resultado**). Na tabela abaixo, reporto as variáveis que apresentaram os maiores Índices de Correlação de Pearson. Elas serão utilizadas nos ajustes dos modelos:

Variável	Correlação com a variável resposta
despesas_campanha	0.596
tentanto_reeleicao	0.586
sigla_partido	0.262

Tabela 2 – Tabela de correlação entre algumas covariáveis e a variável resposta **resultado**

É importante observar que, apesar de a correlação entre o resultado e a sigla do partido ser relativamente baixa, é razoável imaginar que essa variável esteja relacionada à variável resposta de alguma forma. Em particular, é fato incontestado que a filiação a um ou outro partido interfere nas chances de um candidato conseguir se eleger. Observe o gráfico abaixo, por exemplo, que informa a média de gastos de campanha dos candidatos de cada partido político:

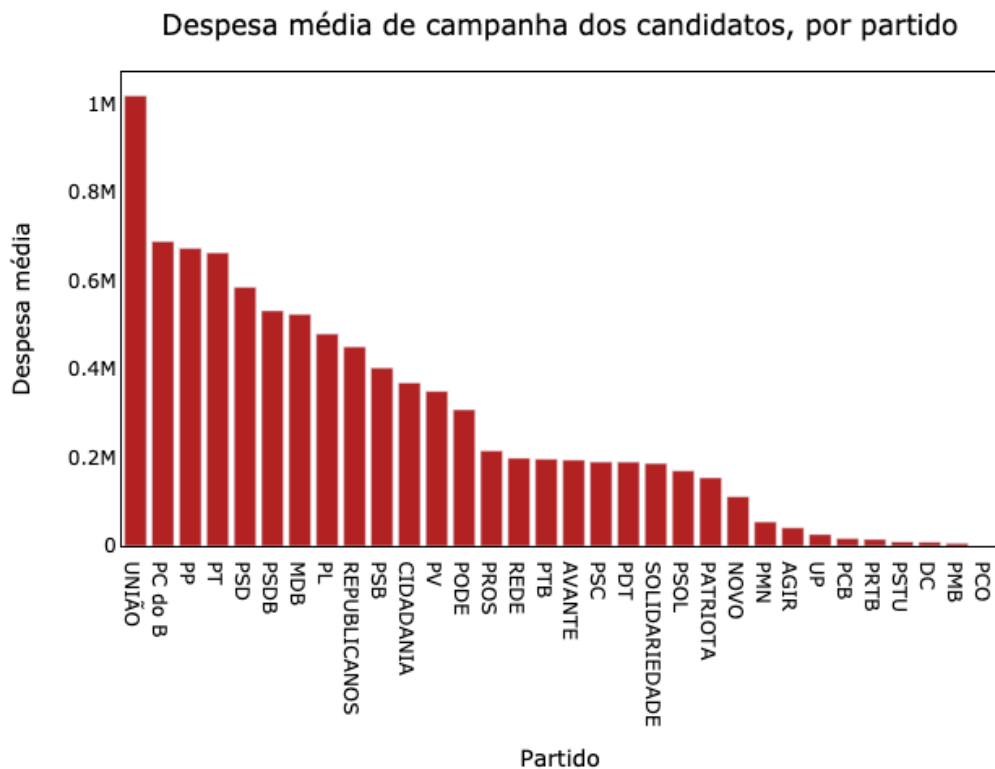


Figura 1 – Despesa média em campanha dos candidatos por partido

Com exceção do PC do B e PSD, foram justamente os partidos que apresentaram os maiores gastos médios de campanha que obtiveram mais cadeiras na Câmara dos

Deputados. Vale verificar também como se distribui a quantidade de votos, variável discreta, de acordo com os gastos de campanha².

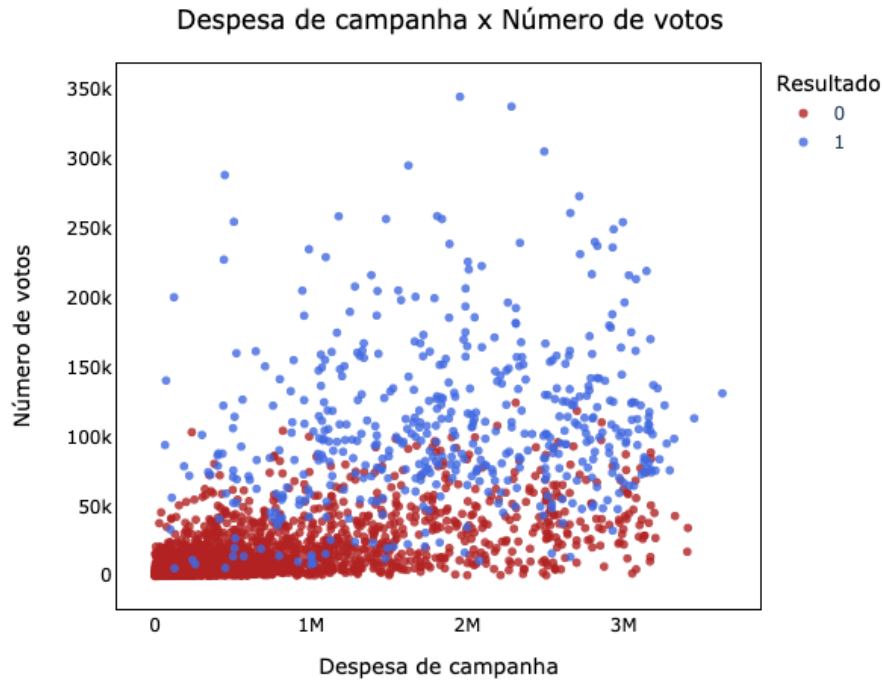


Figura 2 – Despesas de campanha *versus* Número de votos

O gráfico mostra um cenário interessante: o resultado para os candidatos que gastam muito pouco durante a campanha é praticamente certo — a derrota nas urnas. A recíproca, no entanto, não é verdadeira. Mesmo que utilize recursos financeiros em abundância durante a campanha, não é incomum que o candidato não obtenha retorno desses gastos em votos e, no final, não consiga se eleger. Isso sugere que outras variáveis também impactam no desempenho eleitoral dos candidatos. Na próxima seção, comento em detalhes os modelos ajustados para compreender a dinâmica das eleições para deputado federal em 2022.

3.2 Modelos ajustados

Os modelos ajustados pretendiam responder a algumas questões prévias levantadas a partir da literatura. Os dois primeiros modelos pretendem testar se o efeito das despesas de campanha variam de acordo com o estado. Por isso, o **modelo 1** é ajustado com a variável resposta sendo o **resultado** e uma única covariável, a despesa de campanha. Em seguida, ajustamos o **modelo 2**, semelhante ao primeiro, mas que permite que o efeito dos gastos de campanha varie entre os estados. É importante observar que a variável

² A quantidade de votos é um *proxy* natural para resultado eleitoral, embora essa associação não seja direta no caso de sistemas de representação proporcional.

`despesas_campanha` foi padronizada, medida importante para garantir a convergência dos modelos multinível ajustados:

Modelo		Estimação	Erro padrão	AIC
Modelo 1	Intercepto	-3.73883	0.08169	2188
	<code>despesas_campanha</code>	1.31954	0.03888	
Modelo 2 (multinível)	Intercepto	-3.74614	0.08249	2183.2
	<code>despesas_campanha</code>	1.33276	0.05544	

Tabela 3 – Resultados dos modelos 1 e 2

Observe que os resultados reportados são bastante semelhantes, e todos são estatisticamente significativos. No entanto, o modelo 2 reportou poucas variações entre as unidades federativas. Em outras palavras, parece que não houve ganhos em modelar os dados considerando os agrupamentos por estado em relação ao modelo mais simples, embora o AIC tenha sido ligeiramente menor. Esse resultado sugere que as despesas de campanha têm efeito perene por todo o país, favorecendo candidatos que gastam mais.

	Variância	Erro padrão
Intercepto	0	0
<code>despesas_campanha</code>	0.03113	0.1764

Tabela 4 – Incerteza sob os coeficientes de efeito aleatório do modelo 2

Mas, se a regionalidade não parece alterar a dinâmica das disputas eleitorais, ainda faz sentido verificar se alguma outra covariável o faz. Em particular, observamos na análise exploratória que o partido no qual o candidato concorre à eleição parece fazer diferença no resultado. Para testar essa hipótese, ajustou-se um modelo multinível que varia o intercepto por partido e permite que o efeito dos gastos de campanha varie entre eles. A seguir estão reportados os resultados.

Modelo		Estimação	Erro padrão	AIC
Modelo 3 (multinível)	Intercepto	-4.58019	0.24825	1908.5
	<code>despesas_campanha</code>	1.54133	0.09606	

Tabela 5 – Resultados do modelo 3

Apesar de introduzir erro à estimação dos coeficientes em relação aos outros modelos, o modelo 3 apresentou um resultado melhor em relação aos dois primeiros. Além disso, os resultados do modelo multinível sugerem que há variabilidade no efeito dos gastos de campanha por partido:

	Variância	Erro padrão
Intercepto	0.9455	0.9724
despesas_campanha	0.1002	0.3165

Tabela 6 – Incerteza sob os coeficientes de efeito aleatório do modelo 3

Finalmente, o último modelo ajustado adiciona a covariável `tentando_reeleicao` ao modelo para avaliar de que forma ela afeta os resultados. A implementação do modelo permite (i) que os interceptos variem de acordo com o partido e (ii) que o efeito de `tentando_reeleicao` varie por partido. O fato de o candidato estar tentando a reeleição parece surtir um efeito positivo importante nas suas chances de ser eleito:

Modelo		Estimação	Erro padrão	AIC
Modelo 4 (multinível)	Intercepto	-4.71683	0.24871	1827.6
	despesas_campanha	1.27714	0.05654	
	tentando_reeleicao	1.53558	0.26056	

Tabela 7 – Resultados do modelo 4

	Variância	Erro padrão
Intercepto	1.0229	1.0114
tentando_reeleicao	0.6078	0.7796

Tabela 8 – Incerteza sob os coeficientes de efeito aleatório do modelo 4

Dentre todos os modelos ajustados, o modelo 4 é o que possui maior capacidade preditiva, e relativa capacidade explicativa. Conforme abordado na análise exploratória, observamos que gastar pouco em campanha acarreta, de forma geral, derrota nas urnas. Ao mesmo tempo, altos gastos de campanha não necessariamente resultam em vitória. Observe a tabela abaixo, que reporta algumas métricas de avaliação dos modelos³:

Modelo	AUC	R^2	Acc geral	Acc resultado == 1	Acc resultado == 0
Modelo 1	95.65%	34.80%	94.5%	39.49%	98.28%
Modelo 3	96.52%	50.97%	95.4%	52.22%	98.37%
Modelo 4	96.85%	50.34%	95.9%	57.32%	98.59%

Tabela 9 – Métricas de avaliação dos modelos ajustados: AUC, R^2 e Acurácia

³ O modelo 2 foi desconsiderado na medida em que observamos que ele não adicionava mudanças relevantes ao modelo 1.

Os modelos reportam alta acurácia na previsão dos candidatos que não serão eleitos, mas baixa acurácia na previsão dos vitoriosos, refletindo o que foi observado na análise exploratória. Ainda assim, observe que a inclusão da variável `tentando_reeleicao` levou a um aumento razoável na acurácia das previsões dos candidatos eleitos. A capacidade explicativa dos modelos é razoável, principalmente considerando o fato de que o contexto eleitoral é impactado por uma série de outras variáveis que não estão no escopo dos modelos ajustados.

4 Considerações finais

Este relatório traz algumas contribuições para a comunidade da Ciência Política que analisa o impacto de variáveis políticas e econômicas nas eleições brasileiras. De modo geral, a literatura discute a importância dos gastos em campanha para o sucesso eleitoral, também considerando outras covariáveis importantes. Neste trabalho, aplicamos algumas abordagens da literatura aos dados mais recentes disponíveis sobre as eleições para a Câmara dos Deputados. Em particular, utilizamos modelos simples e modelos multinível de regressão logística para tentar determinar de que forma a despesa de campanha impacta no sucesso de um candidato nas urnas.

Os resultados indicam que, mesmo em um cenário de disputa eleitoral extremamente polarizado — no qual surge a hipótese de que alguns candidatos poderiam ser eleitos mesmo com baixos gastos de campanha, garantindo apoio por questões ideológicas —, a quantia disponível para gastar em campanha ainda é fundamental para a vitória. Mais especificamente, o fato de um candidato gastar pouco em campanha é um indicativo claro de que não será eleito; gastar muito, no entanto, não garante sucesso.

Além disso, o fato de o candidato estar tentando a reeleição também surte efeito positivo na sua chance de se eleger. O efeito dos gastos de campanha e de tentativa de reeleição variam por partido, mas parecerem ser relativamente perenes em todo o país, sem alterações regionais significativas.

Uma limitação metodológica deste trabalho é o fato de que a variável resposta é modelada de forma binária. Outras abordagens poderiam utilizar a natureza discreta do número de votos para utilizar outras modelagens e conseguir extrair outras informações dos dados. Por exemplo, seríamos capazes de saber quantos votos a mais um candidato conseguiria obter caso houvesse aumento de uma unidade nas despesas de campanha, *ceteris paribus* (ARRAES; NETO; SIMONASSI, 2017). Esse tipo de análise permitiria diagnósticos mais precisos sobre o efeito das despesas de campanha.

Além disso, ao focar em variáveis econômicas e políticas, este artigo não dá conta de observar o efeito de covariáveis sociais na dinâmica da disputa. É verdade que esse é o caminho adotado pela literatura especializada de forma geral, mas trabalhos mais recentes vêm identificando o efeito de variáveis sociais nas eleições brasileiras, principalmente na

relação entre essas variáveis e as despesas de campanha (CAMPOS; MACHADO, 2020).

Trabalhos futuros poderiam explorar não apenas essas abordagens, mas também avaliar outros níveis de agregação dos dados — por exemplo, outros cargos eletivos, tanto do Poder Executivo quanto Legislativo. É possível que os modelos que utilizem gastos de campanha como covariável sejam capazes de explicar melhor os resultados eleitorais para cargos majoritários, situação na qual número de votos indica precisamente se o candidato será eleito ou não. Um desdobramento possível é o ajuste de modelos multinível por tipo de candidatura, sob a hipótese de que há diferenças no efeito dos gastos de campanha entre os grupos.

Referências

- ARRAES, R.; NETO, O. A.; SIMONASSI, A. Despesas de Campanha e Sucesso Eleitoral nos Pleitos Legislativos Brasileiros. *Dados - Revista de Ciências Sociais*, v. 60, n. 4, p. 1059–1093, 2017.
- CAMPOS, L. A.; MACHADO, C. *Raça e Eleições no Brasil*. Porto Alegre, RS: Zouk, 2020.
- DOWNS, A. Introdução e Capítulo 8 — A estática e a dinâmica das ideologias partidárias. In: _____. *Uma teoria econômica da democracia*. São Paulo: Editora da Universidade de São Paulo, 2013.
- FILHO, D. B. F. Gastos Eleitorais: Os Determinantes das Eleições? Estimando a Influência dos Gastos de Campanha nas Eleições de 2002. *Revista Urutágua*, v. 8, p. 1–10, 2005.
- KENNEDY, P. *A Guide to Econometrics*. 6. ed. Oxford: Blackwell Publishing, 2008.
- MANCUSO, W. P. Investimento Eleitoral no Brasil: Balanço da Literatura (2001-2012). *Revista de Sociologia e Política*, v. 23, n. 54, p. 155–183, 2015.
- MENEZES, A. M. S. *Eleições para Senador no Brasil e Gastos de Campanha*. Dissertação (Mestrado em Economia) — Universidade Federal do Ceará, Fortaleza, 2010.
- NICOLAU, J. O Sistema Eleitoral de Lista Aberta no Brasil. *DADOS - Revista de Ciências Sociais*, Rio de Janeiro, v. 49, n. 4, p. 689–720, 2006.
- SAMUELS, D. Incumbents and Challengers on a Level Playing Field: Assessing the Impact of Campaign Finance in Brazil. *Journal of Politics*, v. 63, p. 569–584, 2001.
- SPECK, B. W.; MANCUSO, W. P. O que Faz a Diferença? Gastos de Campanha, Capital Político, Sexo e Contexto Municipal nas Eleições para Prefeito em 2012. *Cadernos Adenauer*, XIV, p. 247–270, 2013.