

Estado da publicação: O preprint não foi submetido para publicação

Influência do Contexto ou Atributos Individuais? Modelos Multiníveis Aplicados à Análise de Resultados Eleitorais

Thiago Cortez Costa

<https://doi.org/10.1590/SciELOPreprints.10161>

Submetido em: 2024-10-09

Postado em: 2024-10-09 (versão 1)

(AAAA-MM-DD)

INFLUÊNCIA DO CONTEXTO OU ATRIBUTOS INDIVIDUAIS? MODELOS MULTINÍVEIS APLICADOS À ANÁLISE DE RESULTADOS ELEITORAIS

THIAGO CORTEZ COSTA

<https://orcid.org/0000-0002-5365-4732>

[<thcortez@senado.leg.br>](mailto:thcortez@senado.leg.br)

Instituto de Ciência Política, Universidade de Brasília, DF, Brasil

RESUMO: Modelos de regressão convencionais têm entre seus pressupostos: 1) que a variável resposta é numérica e contínua (em jargão técnico, diz-se que tem distribuição de frequências ‘normal’); e 2) que as observações são independentes entre si. Contudo, para grande parte dos problemas de pesquisa em investigações sociais esse não é o caso. Frequentemente a variável resposta é qualitativa, por exemplo, em formato binário, categórico nominal ou ordinal. Outra dificuldade comum ocorre em relação à independência entre as observações, pois o resultado de uma observação pode estar correlacionado de diversas formas ao resultado das demais. Os modelos generalizados de efeito misto, também chamados de modelos hierárquicos generalizados, constituem uma solução para essas limitações. São uma técnica apropriada para o estudo de casos em que se acredita que as características e os comportamentos de uma determinada observação são influenciados pelo conglomerado ou agrupamento a que ela pertence. Desse modo, possibilita-se que o pesquisador considere, num mesmo modelo, dados de contexto, dados de indivíduos e a interação entre os dois níveis. O presente trabalho pretende difundir o uso da modelagem hierárquica como ferramenta de análise de dados organizados em níveis ou agrupamentos quando a variável resposta não é variável numérica contínua (isto é, não possui uma distribuição normal). Para tanto, apresenta-se a teoria estatística subjacente à técnica, acompanhada de uma análise aplicada aos resultados de eleições nacionais do Brasil.

Palavras-chave: modelos hierárquicos generalizados, efeitos contextuais, desempenho individual, análise de dados eleitorais, dados eleitorais.

CONTEXTUAL INFLUENCE OR INDIVIDUAL ATTRIBUTES? MULTILEVEL MODELLING APPLIED TO THE ANALYSIS OF ELECTORAL OUTCOMES

ABSTRACT: Traditional regression models typically rest on two key assumptions: 1) that the response variable is numerical and continuous (commonly assumed to follow a normal distribution), and 2) that the observations are independent of one another. However, in many social science research contexts, these assumptions often do not hold. Frequently, the response variable is qualitative—such as binary, nominal, or ordinal—and the independence of observations is often violated, as outcomes from one observation may be correlated with others. Generalized mixed-effects models, also known as hierarchical generalized models, address these limitations. These models are particularly well-suited for cases where the characteristics and behaviors of an observation are thought to be influenced by the cluster or group to which it belongs. This allows researchers to account for both individual-level and contextual-level data within a single model, as well as the interaction between these levels. This paper aims to promote the use of hierarchical modeling as a valuable tool for analyzing data that is structured in levels or clusters, particularly when the response variable is not continuous or normally distributed. To illustrate the technique, we present supporting statistical theory and apply it to analyze results of Brazilian national elections.

Keywords: Generalized Hierarchical Modeling, Political Behavior, Election Data Analysis, Contextual Effects, Individual Performance

INTRODUÇÃO

Uma grande parte dos fenômenos estudados na Ciência Política e nas Ciências Humanas em geral podem ser compreendidos como estruturas de agrupamentos. Em certos casos, a estrutura de agrupamento é simples, e cada unidade de observação pertence a um único grupo, havendo apenas um fator de agrupamento. Outros casos mais complexos possuem uma estrutura de níveis e subníveis hierarquizados (aninhados e contidos uns nos outros).

Por exemplo, um estudo longitudinal observa uma característica em um conjunto de indivíduos, fazendo medições repetidas ao longo do tempo, não necessariamente com a mesma periodicidade para todos os indivíduos. Essas observações podem ser agrupadas por indivíduo, já que as medidas observadas num mesmo indivíduo estão correlacionadas ao longo do tempo – um caso simples. Para complicar, no entanto, cada indivíduo pode estar agrupado dentro de outra unidade (ou outro nível), como um domicílio, uma família, uma escola, ou um local de trabalho. Tais unidades, por sua vez, poderão estar agrupadas dentro de localidades geográficas, como bairros, cidades, estados e países. Para tomar um exemplo nas ciências naturais, imaginemos um estudo que acompanhe o crescimento de uma dada espécie de planta, com espécimes agrupados por tipo de solo em que se encontram, e os tipos de solo agrupados por região climática. E em um exemplo de administração de empresas, as vendas de uma empresa podem estar agrupadas por vendedores, e os vendedores por lojas em que trabalham.

O ponto comum a todos esses casos é que as observações feitas não são independentes entre si. Há uma correlação entre as observações feitas dentro do mesmo grupo. Ocorrências de uma e de outra criança subnutrida numa família estão correlacionadas porque elas pertencem a uma mesma unidade domiciliar cuja renda per capita é baixa e, portanto, ambas estão igualmente sujeitas à insegurança alimentar. Constatações de crescimento acelerado de espécimes de uma planta em um tipo de solo estão correlacionadas porque aquele tipo de solo é favorável àquela espécie. Tais correlações podem existir e devem ser testadas. Modelos lineares e modelos lineares generalizados, que assumem completa independência entre observações, são inapropriados nesses casos. Em situações em que há correlação passível de agrupamento, a técnica mais adequada é a de modelos hierárquicos. Modelos hierárquicos levam em consideração a estrutura das observações ao estimar o modelo com uma equação para cada nível de agrupamento (ver RAUDENBUSH e BRYK, 2002 e PINHEIRO e BATES, 2004).

As eleições são eventos exemplares de como os dados podem estar agrupados em níveis hierárquicos. Com bastante frequência, sistemas políticos democráticos organizam as disputas pelo voto em urnas, zonas, domicílios e distritos eleitorais – os níveis menores, nesse caso as urnas para votação, estão aninhadas e contidas sucessivamente nas zonas eleitorais, as quais estão dispostas nos níveis mais

altos, domicílio eleitoral e depois distrito eleitoral. Eleitores e candidatos agem e interagem dentro de circunscrições ou domicílios eleitorais. Os partidos políticos, por sua vez, podem se estruturar em diretórios estaduais ou municipais – que equivalem aos distritos eleitorais nas eleições nacionais e municipais, respectivamente. No caso brasileiro, a disputa pelo voto e, depois do pleito, a própria representação política também segue a lógica de distritos. Fica difícil compreender o funcionamento das eleições, portanto, sem considerar tal estrutura multinível.

Nas eleições para a Câmara dos Deputados, conforme os dados que analisaremos, temos em um nível os candidatos e candidatas e num segundo nível as Unidades da Federação (UFs) por onde concorrem (BRASIL, 1988, arts. 45 e 46). Nossa análise pressupõe que os resultados obtidos por esses candidatos e candidatas são uma função do contexto estadual, das características pessoais daqueles que concorrem (do efeito geral que tais características podem ter, e do efeito específico que tais características podem ter em alguns contextos estaduais). A inclinação de um eleitor a votar em um candidato com certas características é afetada pelo distrito em que vota e, conseqüentemente, também são afetadas as chances de eleição daquele candidato com as tais características, naquele distrito. A modelagem hierárquica nos permitirá investigar se há evidências para crer que as características dos distritos eleitorais realmente interferem nos resultados eleitorais e na composição do grupo de eleitos ou se são somente as características individuais dos candidatos que interferem nos resultados eleitorais. Nosso foco neste estudo será investigar fatores que contribuíram para o sucesso ou fracasso eleitoral dos candidatos a deputados federais nas eleições de 2006. Optamos pela modelagem hierárquica porque queremos distinguir como as características individuais dos candidatos interferem em sua elegibilidade, separadamente da influência das características das unidades da federação pelas quais concorrem.

1. OBJETIVO DO PROJETO

A técnica de modelagem será aplicada para explicação dos resultados eleitorais de homens e mulheres que se candidataram à Câmara dos Deputados em 2006. No estudo desse ano eleitoral específico, pretendemos medir o quanto do desempenho eleitoral dos candidatos e candidatas se deve às características de sua UF de origem, ao partido pelo qual concorriam e às suas características pessoais. Nesse ínterim, algumas questões de pesquisa se levantam: indicadores de desenvolvimento socioeconômico (renda, educação e saúde) na UF estão positivamente associadas à probabilidade de mais mulheres serem eleitas? Ou independente do desenvolvimento, os homens têm mais chances de serem eleitos? Seria o nível educacional um fator importante para a eleição de homens e mulheres? Homens e mulheres mais jovens têm menos chance de serem eleitos? Tendo em vista as obrigações familiares e domésticas, será que candidatos e candidatas casados têm piores resultados eleitorais? Os partidos políticos têm peso nas chances de eleição, ou a política brasileira é de fato personalista? Há evidências de que os que concorrem pela reeleição estão em vantagem sobre os demais candidatos? Qual a relação entre

todas essas variáveis em conjunto e as chances de sucesso eleitoral de homens e mulheres? E as variáveis que têm peso nas chances individuais exercem a mesma influência no país como um todo, ou em algumas UFs mais que nas outras? São questões que nos propomos a investigar.

2. RECURSOS TECNOLÓGICOS UTILIZADOS NO PROJETO

O banco de dados de origem da modelagem foi construído e manuseado em MSOffice Excel, no qual foram feitas também as análises gráficas. Em seguida, o banco foi salvo em formato “.sav” e importado ao programa HLM 6.0, onde foi feito o ajuste do modelo hierárquico.

3. DESCRIÇÃO DA PESQUISA

- 1) Dimensão Temporal: a investigação empreendida neste trabalho enfocou um momento específico – a eleição para deputados federais no ano de 2006 – e por isso é do tipo *cross-section*.
- 2) Unidades de Observação e Análise: a pesquisa pretende chegar a conclusões acerca da influência dos contextos estaduais e das características individuais dos candidatos sobre os resultados eleitorais. Desse modo, a observação e a análise são feitas em dois níveis: são usadas informações das UFs e dos candidatos individualmente.
- 3) Técnica de coleta de dados: os microdados são secundários, obtidos junto às fontes primárias oficiais - o TSE, para dados eleitorais e de perfil socioeconômico de candidatas e candidatos, e o IBGE, para dados demográficos dos estados.
- 4) Técnica de análise: quantitativa, com o uso da modelagem hierárquica.

4. TIPOS DE DADOS COLETADOS

Recorremos à literatura sobre comportamento político, sistemas eleitorais e representação política para identificar variáveis que pudessem explicar os resultados de homens e mulheres na eleição de 2006 para deputado federal. As variáveis adotadas para o estudo foram: 1) no plano estadual, de caráter estrutural (socioeconômico, demográfico e geográfico): as UFs pelas quais os indivíduos concorreram à eleição (sobre estudos com variáveis geográficas, ver AVELAR, 2000; MIGUEL e QUEIROZ, 2006 e OLIVEIRA, 2004); as taxas de urbanização das UFs, densidade demográfica, tamanho da população residente, distribuição da população residente por sexo, tamanho do eleitorado, distribuição do eleitorado por sexo, PIB per capita, taxa de alfabetização de adultos e esperança de vida ao nascer (sobre estudos com variáveis demográficas e socioeconômicas, ver CASTRO, 1997; REIS, 1978; SOARES, 1973; SOARES, 2000; ALVES e ARAUJO, 2007; e NORRIS, 2004); 2) de caráter socioeconômico, no plano individual: sexo, nível educacional, idade, estado civil; 3) de caráter político, no nível individual: os partidos políticos pelos quais os candidatos concorreram nas eleições de 2006, e a reeleição – isto é, se

os candidatos concorriam ou não à reeleição naquele pleito; 4) de caráter político no nível estadual: a densidade da concorrência eleitoral (relação candidato/vaga) e a magnitude do distrito (número de vagas de deputados federal a que a UF tem direito na Câmara dos Deputados) (sobre o uso dessas variáveis em estudos de comportamento eleitoral, ver ALVES e ARAUJO, 2006; ARAUJO, 2005; HTUN, 2005; STEWART, 2001).

Os dados referentes a essas variáveis, aos nomes dos participantes na disputa e aos resultados eleitorais de todos os candidatos foram extraídos: 1) do banco de dados eleitorais do TSE referentes às eleições de 2006; 2) das bases de dados do IBGE (PNAD-2006; CPIS-DP; e CCN-DP).

O quadro 1 lista e define cada uma das variáveis, indicando ainda os códigos conforme inseridos nos modelos, e o nível em que aparecem.

		Código	Nome	Descrição
Variável Resposta		RESULTAD	Resultado Eleitoral	0, se não-eleito 1, se eleito
	Variáveis Explicativas	2º Nível - UFs	CANDPVAG	Densidade da Disputa Eleitoral
TAMDIST			Magnitude do Distrito	número de vagas para deputado na UF
POPURBAN			Taxa de Urbanização	Percentual da População na UF vivendo em zonas urbanas
DENSIDAD			Densidade Demográfica	número de habitantes por km ²
EXPECTAT			Esperança de Vida ao Nascer	Número de anos que se espera que um recém nascido viva em média, caso as taxas de mortalidade se mantivessem constantes
TAXAALFA			Taxa de Alfabetização de Adultos	Percentual de alfabetizados dentre os indivíduos com 15 anos ou mais
PIB_PER			PIB per Capita	PIB dividido pelo número de habitantes
ELEITORA			Fragmentação da Disputa pelo Voto Feminino	número de eleitoras por candidata
RZSEXELE			Razão de Sexo no Eleitorado	número de homens a cada 100 mulheres no eleitorado
ELEITPCA			Fragmentação da Disputa Eleitoral	número de eleitores por candidato (h e m)
1º Nível - Candidatos		SEXO1	Sexo	0, se mulher 1, se homem
		PRIMCOUM	Primeiro Grau Completo ou Mais	0, se não concluiu o Primeiro Grau 1, se possui o Primeiro Grau Completo ou mais
		SEGCOUM	Segundo Grau Completo ou Mais	0, se possui até o Primeiro Grau completo 1, se possui o Segundo Grau completo ou mais
		SUPCOMPL	Ensino Superior Completo	0, se não concluiu o Ensino Superior, 1, se concluiu
	CAS	Estado Civil	0, não-casado(a) 1, casado(a)	
	IDADE	Idade	número de anos de vida completos	

		REELEICA	Reeleição	0, se não concorre a reeleição 1, se concorre
--	--	----------	-----------	---

Quadro 1: Lista de Variáveis e Níveis de Análise usados no modelo

Fonte: elaborado pelo autor

5. FERRAMENTAS E MÉTODOS UTILIZADOS

Para proceder com a análise aqui proposta, foi usado o modelo hierárquico generalizado (um caso especial de modelo hierárquico). A escolha do tipo de modelo generalizado justifica-se pelo fato de a variável resposta ser binomial, conforme indicado no Quadro 1 (daí a necessidade de o modelo ser generalizado); e quanto ao 'hierárquico', deve-se ao fato de as variáveis explicativas estarem em dois níveis distintos: o primeiro nível são os indivíduos e o segundo nível, as UFs. O uso do modelo hierárquico tradicional não é apropriado por três razões:

- 1) Dada a restrição no valor da variável resposta, o efeito aleatório do nível 1 só pode assumir um dentre dois valores (0 ou 1, eleito ou não-eleito), e, portanto, não pode ser normalmente distribuído;
- 2) O efeito aleatório do nível 1 não tem variância homogênea. Ao invés disso, a variância deste efeito aleatório depende do valor esperado conforme explicado no item 3 a seguir;
- 3) Finalmente, no modelo tradicional não há restrições ao valor esperado da variável resposta: ele pode assumir qualquer valor real (distribuição normal, variável numérica contínua). Em contraste, o valor esperado de uma variável binomial Y , entendido como a probabilidade esperada de que $Y=1$, e não pode ser menor que 0 ou maior que 1. Assim, um modelo apropriado para prever Y deve restringir os valores esperados ao intervalo $[0,1]$. Sem esta restrição, os efeitos estimados pelo modelo, em geral, não podem ser interpretados (RAUDENBUSH e BRYK, 2002: 38, 274)

O programa HLM permite ajustar modelos de efeito misto para variáveis oriundas de uma distribuição binomial como uma extensão dos modelos lineares generalizados de McCullagh e Nelder (1989) ao caso de informações oriundas de dois ou mais níveis. Para uma descrição formal aprofundada dos modelos hierárquicos generalizados, consultar McCULLOCH e SEARLE (2001) e RAUDENBUSH e BRYK (2002).

6. PLANEJAMENTO E ETAPAS DA REALIZAÇÃO

Os modelos hierárquicos permitem que se considerem na análise a influência de variáveis oriundas de dois ou mais níveis de observações. No caso em questão, consideramos os indivíduos como um primeiro nível de observação e as UFs como um segundo. O motivo de incluir as UFs num modelo sobre os resultados eleitorais individuais se deve ao fato de que supomos que parte desses resultados

podem ser devidos ao contexto interno da UF pela qual um candidato concorre à Câmara. Isto é, algumas variáveis podem ter efeito diferenciado por UF sobre a elegibilidade de homens e mulheres, enquanto outras exercem o mesmo efeito sobre as chances de eleição de todos os indivíduos, independente da UF em questão. Num modelo hierárquico cada variável pode ter um efeito fixo, representado pelo efeito médio da variável sobre o maior nível de observação, e um efeito aleatório, representado pela variação do efeito dessa variável sobre as unidades dentro daquele nível – esse tipo de modelo é chamado de efeito misto justamente por considerar a existência de efeito fixo e de efeito aleatório.

Inicialmente pensamos em fazer o agrupamento das observações por UF usando os Partidos Políticos como um terceiro nível, dentro da UF. Entretanto, muitos partidos apresentaram somente um candidato em algumas UFs, o que inviabiliza a sua inclusão como um nível de análise à parte no método hierárquico. Por esse motivo, os partidos políticos entraram no modelo como co-variável e não como um nível de análise. Após a definição das variáveis de interesse, a execução da coleta dos dados e o estudo das variáveis, procedeu-se ao processo de ajuste do modelo – descrito adiante. Conforme já vimos, a variável resposta Y_{ij} é o resultado eleitoral Y do candidato i na UF j . Y_{ij} pode assumir dois valores:

$$Y_{ij} = \begin{cases} 0, & \text{se não eleito} \\ 1, & \text{se eleito} \end{cases} \quad (1)$$

Assim, a probabilidade de um candidato ser eleito é:

$$\Pr(Y_{ij} = 1) = \phi_{ij} \quad (2)$$

Aplicando a função de ligação logit à probabilidade de eleição ϕ_{ij} , temos o log das chances de eleição η_{ij} :

$$\eta_{ij} = \log \left(\frac{\phi_{ij}}{1 - \phi_{ij}} \right) \quad (3)$$

Tentamos inicialmente ajustar um modelo de η_{ij} que incluísse todas as variáveis explicativas que pretendíamos estudar, para em seguida ir retirando, uma a uma, as menos significativas estatisticamente. Contudo, o grande número de variáveis fez com que a matriz do modelo ficasse quase singular, provavelmente por multicolinearidade entre as variáveis. Optamos então por iniciar o ajuste sem os partidos, que sozinhos representavam 28 variáveis (uma variável binomial para cada partido, ficando de

fora o PV, usado como referência). O modelo ajustado inicialmente para η_{ij} pode ser descrito formalmente como:

$$\eta_{ij} = \beta_{0j} + \beta_{1j}(SEXO1)_{ij} + \beta_{2j}(PRIMCOUM)_{ij} + \beta_{3j}(SEGCOUMA)_{ij} + \beta_{4j}(SUPCOMPL)_{ij} + \beta_{5j}(CAS)_{ij} + \beta_{6j}(Idade)_{ij} + \beta_{7j}(REELEICA)_{ij}$$

$$\beta_{0j} = \gamma_0 + \alpha_1(CANDPVAG)_j + \alpha_2(TAMDIST)_j + \alpha_3(POPURBAN)_j + \alpha_4(DENSIDAD)_j + \alpha_5(EXPECT)_j + \alpha_6(TAXAALFA)_j + \alpha_7(PIB_PER)_j + \alpha_8(ELEITORA)_j + \alpha_9(RZSEXELE)_j + \alpha_{10}(ELEITPCA)_j + u_{0j}$$

$$\beta_{1j} = \gamma_1 + u_{1j} \quad (4)$$

$$\beta_{2j} = \gamma_2 + u_{2j}$$

$$\beta_{3j} = \gamma_3 + u_{3j}$$

$$\beta_{4j} = \gamma_4 + u_{4j}$$

$$\beta_{5j} = \gamma_5 + u_{5j}$$

$$\beta_{6j} = \gamma_6 + u_{6j}$$

$$\beta_{7j} = \gamma_7 + u_{7j}$$

No modelo expresso na equação 4, i é o índice dos candidatos $\{i \in \mathbf{N}/1 \leq i \leq 4946\}$, j o índice das UFs $\{j \in \mathbf{N}/1 \leq j \leq 27\}$. Sendo s o índice das variáveis do nível 2 (UFs) $\{s \in \mathbf{N}/1 \leq s \leq 10\}$ e q o índice das variáveis do nível 1 (candidatos) $\{q \in \mathbf{N}/1 \leq q \leq 7\}$, α_s representa o efeito fixo da variável s das UFs; o β_{aj} representa o efeito fixo γ_q da variável q dos candidatos, mais o efeito aleatório u_{aj} da variável q sobre a UF j .

Em seguida, retirou-se do modelo expresso em 4, uma a uma, as variáveis cujos parâmetros foram menos significativos estatisticamente (que são os que tinham maior p-valor, representados com fundo cinza no quadro 2). O quadro 2 resume as etapas de retirada dos parâmetros.

		Variáveis	Etapas						
			1	2	3	4	5	6	7
Efeito Fixo	2º Nível - UF	INTRCPT2	.222	.143	.145	.147	.178	.226	.087
		CANDPVAG	.051	.039	.039	.034	.042	.043	.046
		TAMDIST	.528	.471	.475	.429	.430	.415	.480
		POPURBAN	.841	.830	.652	.629	.584	.372	.354
		DENSIDAD	.198	.196	.197	.198	.221	.295	.338
		EXPECTAT	.882	-	-	-	-	-	-
		TAXAALFA	.864	.911	-	-	-	-	-
		PIB_PER	.702	.724	.611	.617	.607	-	-
		ELEITORA	.480	.487	.488	.477	.463	.428	.366
		RZSEXELE	.445	.428	.415	.427	.472	.575	-
		ELEITPCA	.853	.857	.855	-	-	-	-
	1º Nível - candidatos	SEXO1	.439	.438	.438	.440	.446	.444	.457
		PRIMCOUM	.229	.228	.229	.230	.273	.272	.276
		SEGCOUN	.031	.031	.031	.031	.035	.035	.035
		SUPCOMPL	.000	.000	.000	.000	.000	.000	.000
		CAS	.237	.237	.237	.236	.257	.267	.261
		IDADE	.636	.636	.635	.637	-	-	-
		REELEICA	.000	.000	.000	.000	.000	.000	.000
	Efeito Aleatório	1º Nível - candidatos	INTRCPT1	.075	.115	.169	.235	>.500	>.500
SEXO1			.002	.002	.002	.002	.013	.013	.013
PRIMCOUM			>.500	>.500	>.500	>.500	>.500	>.500	>.500
SEGCOUN			>.500	>.500	>.500	>.500	>.500	>.500	>.500
SUPCOMPL			.020	.020	.020	.020	.177	.178	.176

	CAS	.208	.209	.209	.207	.231	.230	.229
	IDADE	.144	.144	.145	.144	-	-	-
	REELEICA	.084	.084	.084	.085	.246	.244	.242

Quadro

2: p-Valor dos parâmetros dos modelos ajustados a cada etapa (etapas 1 a 7)

Fonte: Elaborado pelos autores

		Variáveis	Etapas						
			8	9	10	11	12	13	14
Efeito Fixo	2º Nível - UF	INTRCPT2	.012	.005	.002	.000	.000	.000	.000
		CANDPVAG	.031	.036	.025	.000	.000	.000	.000
		TAMDIST	-	-	-	-	-	-	-
		POPURBAN	.491	.495	.533	-	-	-	-
		DENSIDAD	.448	.559	-	-	-	-	-
		EXPECTAT	-	-	-	-	-	-	-
		TAXAALFA	-	-	-	-	-	-	-
		PIB_PER	-	-	-	-	-	-	-
		ELEITORA	.544	-	-	-	-	-	-
		RZSEXELE	-	-	-	-	-	-	-
	ELEITPCA	-	-	-	-	-	-	-	
	1º Nível - candid	SEXO1	.455	.479	.466	.468	.466	.477	.535
		PRIMCOUM	.278	.274	.275	.268	.276	.247	.236
		SEGCOUM	.035	.035	.035	.034	.034	.014	.014
SUPCOMPL		.000	.000	.000	.000	.000	.000	.000	

		CAS	.263	.275	.282	.272	.276	.210	.185
		IDADE	-	-	-	-	-	-	-
		REELEICA	.000	.000	.000	.000	.000	.000	.000
Efeito Aleatório	1º Nível - candidatos	INTRCPT1	>.500	>.500	>.500	>.500	.403	.002	.005
		SEXO1	.013	.013	.012	.012	.019	.018	.006
		PRIMCOUM	>.500	>.500	>.500	>.500	-	-	-
		SEGCOUN	>.500	>.500	>.500	>.500	>.500	-	-
		SUPCOMPL	.184	.185	.188	.185	.187	.108	.094
		CAS	.237	.227	.231	.223	.126	.108	.093
		IDADE	-	-	-	-	-	-	-
		REELEICA	.245	.244	.246	.248	.284	.239	-

Quadro 3: p-Valor dos parâmetros dos modelos ajustados a cada etapa (etapas 8 a 14)

Fonte: Elaborado pelos autores

Manteve-se a variável ‘Sexo’ ao longo do ajuste dos modelos por ser uma das variáveis de interesse da pesquisa, apesar do alto p-valor do efeito fixo encontrado em todas as etapas. Também justifica a manutenção de ‘Sexo’ no modelo, o fato de esta variável haver apresentado um efeito aleatório significativo (o p-valor do efeito aleatório é baixo). Isto indica que embora não haja evidência de efeito do Sexo sobre os resultados eleitorais quando consideramos o país inteiro, parece haver efeito do Sexo sobre os resultados por UF. O passo seguinte foi incluir os partidos políticos no modelo. Numa primeira tentativa, ao incluí-los todos separadamente gerou-se novamente o problema de singularidade da matriz do modelo, possivelmente pela existência de partidos demasiadamente pequenos – indicando que deveríamos agregar partidos e reduzir o número de categorias. Optamos, então, por agregar os partidos com menos candidatos em categorias segundo orientação ideológica, conforme o quadro 4.

	Tamanho	Categoria	Descrição	Partidos
1ª Agregação	De 1 a 100 candidatos	NANA	Partidos Trabalhistas e Comunistas	PRTB, PCdoB, PTN, PCO, PSTU, PCB
		NANB	Demais orientações ideológicas	PSL, PAN, PRB
	De 101 a 200 candidatos	MEDA	Partidos Trabalhistas e Comunistas	PSOL, PTdoB, PTC
		MEDB	Demais orientações ideológicas	PP, PMN, PHS, PSDC, PRONA, PRP
2ª Agregação	De 0 a 200 candidatos	MdNANA	Partidos Trabalhistas e Comunistas	NANA + MEDA
		MdNANB	Demais orientações ideológicas	NANB + MEDB

Quadro 4: Criação de Categorias para os Partidos com menos candidatos

Fonte: Elaborado pelos autores

Num primeiro ajuste, incluímos apenas as categorias NANA e NANB no modelo, mantendo os demais partidos desagregados. O problema de quase singularidade da matriz do modelo permaneceu. Tentou-se novamente o ajuste, dessa vez incluindo também as categorias MEDA e MEDB e o problema foi sanado, gerando o modelo da etapa 15 do quadro 4. Na etapa 17 executamos uma nova agregação, juntando todos os partidos com menos de 200 candidatos em duas categorias distintas entre si pela orientação ideológico-partidária.

	Variáveis	Etapas					
		15	16	17			
Efeito Fixo	2º Nível	INTRCPT2	.000	.000	.000		
		CANDPVAG	.005	.007	.007		
		...	-	-	-		
	1º Nível	SEXO1	.390	.400	.451		
		PRIMCOUM	.185	.205	.203		
		SEGCOUM	.012	.012	.014		
		SUPCOMPL	.000	.000	.000		
		CAS	.969	-	-		
		IDADE	-	-	-		
		REELEICA	.000	.000	.000		
		NANA	.338	.281		MdNANA	
		NANB	.399	.446	.839		
		MEDA	.293	.279		MdNANB	
		MEDB	.073	.070	.126		
		PMDB	.000	.000	.000		
		PDT	.087	.085	.091		
		PFL	.000	.000	.000		
		PL	.102	.113	.122		
		PPS	.043	.035	.040		
		PSB	.029	.027	.032		
		PSC	.443	.288	.327		
		PSDB	.000	.000	.000		
		PT	.000	.000	.000		
	PTB	.249	.224	.239			
	Efeito Aleatório	1º Nível	INTRCPT1	.036	.012	.013	
			SEXO1	.031	.061	.005	
			PRIMCOUM	-	-	-	
SEGCOUM			-	-	-		
SUPCOMPL			.064	.117	.151		
CAS			.418	-	-		
IDADE			-	-	-		

REELEICA	-	-	-	
NANA	.136	.044		MdNANA
NANB	>.500	>.500	.098	
MEDA	>.500	>.500		MdNANB
MEDB	.027	.006	.008	
PMDB	.006	.001	.002	
PDT	.277	.117	.130	
PFL	.362	.196	.170	
PL	.023	.005	.006	
PPS	.200	.052	.053	
PSB	.014	.003	.004	
PSC	.253	.088	.096	
PSDB	.083	.031	.037	
PT	.039	.010	.013	
PTB	.043	.011	.014	

Quadro 5: p-Valor dos parâmetros dos modelos ajustados a cada etapa (continuação do Quadro 02)

Fonte: Elaborado pelo autor a partir de dados do TSE e do IBGE

7. Benefícios do método, resultados e impactos potenciais ou efetivos

O modelo final ajustado na etapa 17 do quadro 4 ficou configurado de acordo com a equação 5:

$$\begin{aligned} \eta_{ij} = & \beta_{0j} + \beta_{1j}(SEX01)_{ij} + \beta_{2j}(PRIMCOUM)_{ij} + \beta_{3j}(SEGCOUMA)_{ij} + \beta_{4j}(SUPCOMPL)_{ij} \\ & + \beta_{5j}(REELEICA)_{ij} + \beta_{6j}(MdNANA)_{ij} + \beta_{7j}(MdNANB)_{ij} + \beta_{8j}(PMDB)_{ij} + \beta_{9j}(PDT) \\ & + \beta_{10j}(PFL) + \beta_{11j}(PL) + \beta_{12j}(PPS) + \beta_{13j}(PSB) + \beta_{14j}(PSC) + \beta_{15j}(PSDB) + \beta_{16j}(PT) \\ & + \beta_{17j}(PTB) \end{aligned}$$

(5)

$$\beta_{0j} = \gamma_0 + \alpha_1(CANDPVAG)_j + u_{0j}$$

$$\beta_{1j} = \gamma_1 + u_{1j}$$

$$\beta_{4j} = \gamma_4 + u_{4j}$$

$$\beta_{2j} = \gamma_2$$

$$\beta_{5j} = \gamma_5$$

$$\beta_{3j} = \gamma_3$$

$$\beta_{6j} = \gamma_6 + u_{6j}$$

$$\beta_{7j} = \gamma_7 + u_{7j}$$

$$\beta_{13j} = \gamma_{13} + u_{13j}$$

$$\beta_{8j} = \gamma_8 + u_{8j}$$

$$\beta_{14j} = \gamma_{14} + u_{14j}$$

$$\beta_{9j} = \gamma_9 + u_{9j}$$

$$\beta_{15j} = \gamma_{15} + u_{15j}$$

$$\beta_{10j} = \gamma_{10} + u_{10j}$$

$$\beta_{16j} = \gamma_{16} + u_{16j}$$

$$\beta_{11j} = \gamma_{11} + u_{11j}$$

$$\beta_{17j} = \gamma_{17} + u_{17j}$$

$$\beta_{12j} = \gamma_{12} + u_{12j}$$

Onde i é o índice dos candidatos $\{i \in \mathbf{N}/1 \geq i \geq 4946\}$, j o índice das UFs $\{j \in \mathbf{N}/1 \geq j \geq 27\}$.

Sendo s o índice das variáveis do nível 2 (UFs) $\{s \in \mathbf{N}/1 \geq s \geq 10\}$ e q o índice das variáveis do nível 1 (candidatos) $\{q \in \mathbf{N}/1 \geq q \geq 7\}$, α_s representa o efeito fixo da variável s das UFs; o β_{qj} representa o efeito fixo γ_q da variável q dos candidatos, mais o efeito aleatório u_{qj} da variável q sobre a UF j . Os quadros 5 e 6 informam os valores dos parâmetros calculados do modelo.

		Variáveis	Coeficiente		Desvio Padrão	Razão T	gl	p-valor	Razão de Chances	Intervalo de Confiança
Efeito Fixo	2º Nível	INTRCPT2	γ_0	-4.295276	0.513353	-8.367	25	.000	0.013633	(0.005,0.039)
		CANDPVAG	α_1	-0.062598	0.012007	-5.213	25	.007	0.939321	(0.916,0.963)
	1º Nível	SEXO1	γ_1	0.232661	0.289071	0.805	26	.451	1.261954	(0.697,2.284)
		PRIMCOUM	γ_2	-0.733953	0.433264	-1.694	4925	.203	0.480008	(0.205,1.122)
		SEGCOUN	γ_3	0.920006	0.321028	2.866	4925	.014	2.509304	(1.337,4.708)
		SUPCOMPL	γ_4	0.917475	0.169733	5.405	26	.000	2.502963	(1.767,3.546)
		REELEICA	γ_5	2.840033	0.128971	22.021	4925	.000	17.116323	(13.293,22.039)
		MdNANA	γ_6	0.092421	0.278208	0.332	26	.839	1.096826	(0.620,1.942)
		MdNANB	γ_7	0.718558	0.364061	1.974	26	.126	2.051472	(0.972,4.332)
		PMDB	γ_8	2.076502	0.344313	6.031	26	.000	7.976517	(3.934,16.174)
		PDT	γ_9	0.763123	0.269847	2.828	26	.091	2.144964	(1.233,3.733)
		PFL	γ_{10}	2.153541	0.361347	5.960	26	.000	8.615314	(4.103,18.090)
		PL	γ_{11}	0.828844	0.396475	2.091	26	.122	2.290669	(1.015,5.170)
		PPS	γ_{12}	0.950066	0.256856	3.699	26	.040	2.585880	(1.526,4.382)
		PSB	γ_{13}	1.154678	0.428311	2.696	26	.032	3.173001	(1.317,7.645)
PSC	γ_{14}	0.576240	0.409296	1.408	26	.327	1.779336	(0.768,4.123)		

	PSDB	γ_{15}	1.760130	0.219376	8.023	26	.000	5.813193	(3.705,9.120)
	PT	γ_{16}	1.837858	0.275605	6.668	26	.000	6.283066	(3.568,11.064)
	PTB	γ_{17}	0.660760	0.446948	1.478	26	.239	1.936262	(0.774,4.847)

Quadro 6: Parâmetros do Modelo Final Ajustado (Efeito Fixo)

Fonte: Elaborado pelo autor a partir de dados do TSE e do IBGE

		Variáveis	Componente da Variância	Desvio Padrão	gl	Qui-Quadrado	p-valor	
Efeito Aleatório	1º Nível	INTRCPT1	u_{0j}	2.10343	1.45032	18	33.96665	.013
		SEXO1	u_{1j}	1.20536	1.09789	19	38.89967	.005
		PRIMCOUM	-	-	-	-	-	-
		SEGCUM	-	-	-	-	-	-
		SUPCOMPL	u_{4j}	0.29083	0.53928	19	25.28365	.151
		REELEICA	-	-	-	-	-	-
		MdNANA	u_{6j}	0.61262	0.78270	19	27.29918	.098
		MdNANB	u_{7j}	1.28655	1.13426	19	37.14210	.008
		PMDB	u_{8j}	0.84125	0.91720	19	42.79269	.002
		PDT	u_{9j}	0.42259	0.65007	19	25.98924	.130

		PFL	u_{10j}	1.24919	1.11767	19	24.71585	.170
		PL	u_{11j}	1.42935	1.19555	19	37.94067	.006
		PPS	u_{12j}	0.37230	0.61017	19	29.86943	.053
		PSB	u_{13j}	1.73313	1.31649	19	39.86858	.004
		PSC	u_{14j}	1.02525	1.01255	19	27.38178	.096
		PSDB	u_{15j}	0.20132	0.44869	19	31.35612	.037
		PT	u_{16j}	0.74395	0.74395	19	35.36556	.013
		PTB	u_{17j}	2.11992	1.45600	19	34.99390	.014

Quadro 7: Parâmetros do Modelo Final Ajustado (Efeito Aleatório)

Fonte: Elaborado pelo autor a partir de dados do TSE e do IBGE

O intercepto γ_0 , do valor de -4,295276 (Quadro 5), é o log das chances de eleição de um tipo padrão de candidato, que possui todas as variáveis com valor 0 e $u_{0j} = 0$. Transformando o log das chances em probabilidade, calculamos a expressão:

$$\frac{1}{(1 + \exp\{4,295276\})} = 0,01345 \quad (6)$$

Em palavras, o modelo indica que, hipoteticamente, uma mulher que cursou até o Primeiro Grau incompleto, que não concorreu à reeleição e era candidata pelo Partido Verde sem concorrência numa UF padrão (em que $u_{0j} = 0$), teve 1,34% de probabilidade de ser eleita em 2006.

A densidade da concorrência (CANDPVAG) está associada a menores chances de vitória eleitoral ($\alpha_1 = -0,062598$, conforme Quadro 5). O coeficiente α_1 negativo indica que quando aumenta o número de candidatos por vaga, diminuem as chances de eleição. Assim, dentro de uma UF qualquer (isto é, mantendo-se constante a variabilidade aleatória u_{0j}), cada unidade acrescida na densidade da concorrência em 2006 provocou uma diminuição de 0,062 no log das chances de eleição. Ou seja, numa UF hipotética que tivesse 1 candidato concorrendo para cada vaga, a probabilidade de eleição de um candidato qualquer seria 1,26% menor do que se não houvesse nenhum candidato por vaga (conforme a equação 7).

$$\frac{1}{(1 + \exp\{4,295276 + 0,062598 \times 1\})} = 0,01264 \quad (7)$$

O sexo dos candidatos aparentemente não teve um efeito geral sobre a elegibilidade em 2006. O p-valor do efeito fixo do sexo foi alto, da ordem de 0,451, indicando não haver evidências de que na média nacional o sexo tenha interferido no log das chances de eleição. Contudo, houve uma variação do efeito do sexo entre as UFs – conforme indicado pelo Quadro 6, o componente da variância do sexo entre as UFs u_{1j} é igual a 1,20536 e possui p-valor baixo. Isso quer dizer que embora na média nacional o sexo não haja tido grande efeito sobre o log das chances, em algumas UFs ele

teve grande relevância, interferindo nas chances de eleição de homens e mulheres. Ainda assim, de acordo com os coeficientes calculados no modelo 13, candidatos do sexo masculino tiveram no geral maiores chances de sucesso eleitoral em 2006 ($\gamma_1 = 0,232661$). Mantendo constantes todas as outras variáveis do modelo, um candidato homem obteve 0,23 log de chances a mais do que se fosse mulher, ou 1,26 vezes as chances de uma mulher com as mesmas características (conforme a Razão de Chances do sexo no Quadro 5, pois $\exp(0,23)=1,26$). Portanto, o sexo masculino está associado a maiores chances de eleição. Nossa candidata de referência – que cursou até o Primeiro Grau incompleto, não concorre à reeleição, é membro do Partido Verde, numa UF padrão (em que $u_{0j} = 0$ e $u_{1j} = 0$), com uma densidade de concorrência igual a 9 candidatos por vaga (próxima à densidade média nacional) – teria uma probabilidade de 0,7% de ser eleita. Um candidato homem, na mesma situação, teria 0,9% de probabilidade de eleição.

Já com relação à escolaridade, os resultados do ajuste do modelo 5.13 mostraram que têm maiores chances de eleição aqueles com maiores níveis educacionais. Curiosamente, o coeficiente γ_2 da variável PRIMCOUM assumiu um valor negativo (-0,733953, conforme Quadro 5), indicando que os candidatos que possuíam o Primeiro Grau completo ou mais estavam em desvantagem em relação aos que sequer tinham completado o Primeiro Grau. Contudo, os coeficientes de SEGCOUM e SUPCOMPL assumiram valores positivos e maiores em módulo que o coeficiente de PRIMCOUM ($\gamma_3 = 0,920006$ e $\gamma_4 = 0,917475$, respectivamente, de acordo com o Quadro 5). Desse modo, somos levados a crer que indivíduos com o Primeiro Grau completo e Segundo Grau incompleto estavam em desvantagem eleitoral em relação aos que sequer completaram o Primeiro Grau. Estes por sua vez estavam em desvantagem em relação aos que completaram o Segundo Grau e aos que completaram o Nível Superior. Numa UF padrão, em que os componentes da variância fossem iguais a zero, possuir o Segundo Grau completo aumentaria o log das chances de eleição de um indivíduo em 0.92. Possuir nível superior aumentaria o log das chances ainda mais 0,917.

Ao longo das etapas do ajuste os componentes da variância de PRIMCOUM e de SEGCOUM foram retirados por apresentarem p-valor alto (ver Quadro 2). Isto indica que possuir Primeiro ou Segundo Grau completo teve um efeito geral para todo o país, mas não teve efeito diferenciado por UF. Possuir o Ensino Superior completo, por sua vez, teve efeito variado por UF, conforme evidenciado no Quadro 5.6 ($u_{4j} = 0,29$).

Concorrer à reeleição foi o atributo que teve o maior efeito fixo sobre a elegibilidade. Tudo o mais constante, em 2006, um candidato à reeleição teve 2.84 log de chances a mais que um candidato novato ($\gamma_5 = 2,8400$, conforme Quadro 5). Traduzindo em chances, um candidato à reeleição teve 17,11 vezes as chances de eleição de um novato com as mesmas características. O componente aleatório de REELEICA também apresentou p-valor grande ao longo das etapas de ajuste do modelo, sendo por esse motivo retirado do modelo. Cremos, por isso, que ‘concorrer à reeleição’ não teve efeito diferenciado por UF.

Na média nacional, os partidos que apresentaram 200 candidatos ou menos nas eleições de 2006 (representados no modelo pelas variáveis MdNANA e MdNANB), mais o PL, o PSC e o PTB não tiveram grande significância estatística (seu p-valor foi alto, conforme Quadro 5). Contudo, o p-valor da componente da variância desses partidos foi baixo (ver Quadro 6). Ou seja, embora quando considerados no país como um todo tais partidos não hajam tido efeito significativo, em algumas UFs, o peso desses partidos sobre os resultados eleitorais foi maior que em outras. Os demais partidos tiveram um importante efeito sobre os resultados eleitorais na média nacional. Tomando candidatos com as mesmas características, numa mesma UF, sairia em vantagem o candidato que fosse filiado ao PFL, com 2,15 log de chances a mais que um candidato pelo PV. Em seguida, viria um candidato pelo PMDB, com 2,07 log de chances de vantagem sobre o candidato do PV. Depois viria um do PT (1,83 log de chances a mais), um do PSDB (1,76 log de chances a mais), um do PSB (1,15 log de chances a mais), um do PPS (0,95 log de chances a mais), e um do PDT (0,76 log de chances a mais).

O PDT e o PFL apresentaram componentes da variância com p-valor alto (conforme Quadro 5.6), ou seja, esses dois partidos não tiveram comportamento diferenciado por UF, influenciando de igual modo a elegibilidade de seus candidatos por todo o país. A componente da variância dos demais partidos apresentou p-valor baixo (Quadro 6), indicando que a influência desses partidos sobre as chances de eleição variou de UF para UF.

8. CONTRIBUIÇÕES PARA A SISTEMATIZAÇÃO DO PROCESSO DE PESQUISA

Um resultado relevante do processo de ajuste do modelo foi o fato de que diversas variáveis apontadas pela literatura como importantes na explicação das eleições tiveram pouca significância estatística. Por apresentarem p-valor alto, ficaram de fora do modelo variáveis como: 1) no nível da UF: expectativa de vida ao nascer, taxa de alfabetização de adultos, PIB per capita, taxa de urbanização,

densidade demográfica, densidade da disputa, razão de sexo no eleitorado, magnitude do distrito eleitoral e número de eleitoras por candidata; e 2) no nível do candidato: idade e estado civil. Além disso, o efeito do sexo dos candidatos não ser estatisticamente significativo para explicar as chances de eleição, considerando o Brasil como um todo, é fato curioso, em face das discussões sobre políticas afirmativas de promoção da participação feminina. Esses achados sugerem a necessidade de uma reflexão sobre as medidas que têm sido utilizadas pela literatura (incluindo a que é utilizada nesse artigo) para medir o sucesso ou o insucesso da participação eleitoral de candidatos e candidatas. É preciso refletir sobre quem são os eleitos e as eleitas e em que condições obtiveram sucesso. Parte deste exercício precisará responder por quais motivos o sexo não foi relevante para explicar a chance de eleição e, ainda assim, o resultado das eleições naquele ano (e em todos os demais) obteve tão poucas mulheres eleitas.

Por hora, cabe-nos destacar o fato de que a metodologia permitiu que trabalhássemos com observações em dois níveis (UF e candidato), resolvendo um problema bastante frequente em investigação social e política: o de conciliar informações de diferentes estratos sem incorrer em conclusões que cometam falácias ecológicas. Também foi possível identificar para cada variável se há um efeito fixo geral sobre o resultado individual, se há um efeito variado de acordo com a UF em questão, e quais são esses efeitos. Desse exercício, vimos que ao considerar a estrutura de níveis associada à origem dos dados (isto é, a organização do Estado Brasileiro em unidades federativas, e do sistema político e eleitoral a partir delas), algumas variáveis aparentemente perderam importância como fator determinante de resultados eleitorais. Para confirmar isso, caberá testar se as conclusões se verificam em outras implementações de modelagem hierárquica em softwares mais recentes (como pacotes específicos do R, por exemplo), testar se o mesmo vale para outros anos eleitorais e realizar análises longitudinais, expandindo o período analisado para incluir resultados de outros pleitos.

CONSIDERAÇÕES FINAIS

O presente estudo destacou a relevância da modelagem hierárquica para a análise de dados estruturados em níveis, com foco específico no contexto eleitoral. Ao aplicar essa técnica à eleição para a Câmara dos Deputados de 2006 no Brasil, foi possível explorar tanto as características individuais dos candidatos quanto as peculiaridades das Unidades da Federação (UFs) às quais pertencem, evidenciando como esses diferentes níveis interagem para influenciar os resultados eleitorais.

Os resultados da modelagem revelaram que, embora variáveis como o sexo dos candidatos não apresentem grande significância estatística no nível nacional, o efeito do sexo variou entre as UFs, sugerindo a necessidade de um olhar mais atento ao contexto regional. Da mesma forma, o nível educacional mostrou um impacto positivo nas chances de eleição, principalmente para aqueles com ensino superior completo, enquanto a concorrência eleitoral (número de candidatos por vaga) demonstrou reduzir as chances de sucesso. Além disso, a reeleição emergiu como o fator de maior peso, oferecendo aos candidatos à reeleição uma vantagem substancial sobre os novatos.

A inclusão de variáveis partidárias também foi crucial, com partidos mais tradicionais e de maior expressão política mostrando um efeito significativo sobre as chances de eleição de seus candidatos. Contudo, esses efeitos também variaram conforme o contexto das UFs, refletindo a diversidade e complexidade do sistema eleitoral brasileiro.

Por fim, a pesquisa trouxe à tona alguns resultados inesperados, como a ausência de impacto de variáveis frequentemente destacadas pela literatura, como estado civil, idade dos candidatos e indicadores socioeconômicos de desenvolvimento das UFs. Esses achados sugerem a necessidade de uma revisão de conceitos pré-estabelecidos e indicam que futuros estudos poderiam beneficiar-se de uma abordagem mais granular, explorando variações regionais e outros fatores contextuais que possam influenciar o comportamento eleitoral.

Em suma, este trabalho demonstrou que a modelagem hierárquica generalizada oferece uma ferramenta poderosa para a análise de dados eleitorais, permitindo que se capturem nuances importantes que poderiam ser negligenciadas por abordagens tradicionais. A metodologia aqui aplicada não apenas reforça a importância das características individuais e contextuais nas eleições brasileiras, mas também demonstra sua aplicabilidade em contextos de pesquisa cujos dados têm ou podem ter uma estrutura multinível subjacente. Essa estrutura precisa ser levada em consideração pelo investigador para melhorar o processo de modelagem, fornecendo estimativas mais precisas e modelos mais eficientes - menores erros residuais e maior capacidade explicativa.

REFERÊNCIAS

ARAÚJO, Clara e ALVES, José Eustáquio D. Do lugar e das chances das mulheres nas eleições: efeitos e impactos do sistema eleitoral e de indicadores sociais sobre as cotas. ANPOCS, 30º encontro. Caxambu, MG, outubro de 2006.

ARAÚJO, Clara. “Potencialidades e Limites da Política de Cotas no Brasil”, in *Revista Estudos Feministas*, v.9 n.1, 2001, Florianópolis, p.231-252.

AVELAR, Lucia. *Desenvolvimento e Poder Político: O resgate de uma tese*. Projeto de Pesquisa. UnB, Brasília, 2000.

BRASIL. Congresso Nacional. *Constituição Federal de 1988*. Disponível em <<http://www.planalto.gov.br/>>, acesso em 30 de junho de 2008.

HTUN, Mala. “Women, Political Parties and Electoral Systems in Latin America”, in BALLINGTON, Julie e KARAM, Azza (eds.). *Women in Parliament: Beyond Numbers*. Stockolm: IDEA, 2005.

MCCULLOCH, Charles E. e SEARLE, Shayle R. *Generalized, Linear, and Mixed Models*. New York: John Wiley and Sons, 2001.

MIGUEL, Luis Felipe e QUEIROZ, Cristina Monteiro. Diferenças regionais e o êxito relativo de mulheres em eleições municipais no Brasil. In *Revista Estudos Feministas*, v.14 n.2, 2006, Florianópolis, pp. 363-385.

NORRIS, Pippa. *Electoral Engineering: voting rules and political behavior*. Cambridge: Cambridge University Press, 2004.

OLIVEIRA, Evelina A. Notas sobre a participação não-eleitoral no Nordeste, in *Revista Opinião Pública*, v.10 n.2, 2004, Campinas. Disponível em <www.scielo.br>, acesso em 01 de abril de 2008.

PINHEIRO, José C. e BATES, Douglas M. *Mixed-Effects Models in S and S-Plus*. New York: Springer, 2004.

RAUDENBUSH, S.W. e BRYK, A.S. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage Publications, 2002.

STEWART III, Charles. *Analyzing Congress*. New York: WW Norton Company, 2001.

DECLARAÇÃO DE DISPONIBILIDADE DE DADOS DA PESQUISA: O conjunto de dados de apoio aos resultados deste estudo ainda não está disponível ao público.

FINANCIAMENTO: O presente trabalho foi realizado com apoio do Senado Federal, órgão em que o autor trabalha e que o permitiu dedicar-se à elaboração.

CONTRIBUIÇÃO DAS/DOS AUTORES/AS: **Thiago Cortez Costa:** Conceitualização, Metodologia, Validação, Análise Formal, Investigação, Recursos, Curadoria de Dados, Redação - Original, Redação – Revisão e Edição, Visualização, Supervisão, Administração de Projeto, Aquisição de Financiamento.

DECLARAÇÃO DE CONFLITO DE INTERESSE: O autor declara não haver conflito de interesse.

MINIBIOGRAFIAS DOS/DAS AUTORAS DO PAPER:

Pesquisador do Instituto de Ciência Política, da UnB, e do Instituto de Pesquisa DataSenado, do Senado Federal. É cientista político pela UnB e Mestre pela Escola Nacional de Ciências Estatísticas (ENCE/IBGE).

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença [Creative Commons CC-BY](#).
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.