v7.n11.2026

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

# PREVISÃO PROBABILÍSTICA DO SEGUNDO TURNO PRESIDENCIAL BRASILEIRO DE 2026 POR INFERÊNCIA ECOLÓGICA E MATRIZES HISTÓRICAS DE TRANSFERÊNCIA DE VOTOS

## *PROBABILISTIC FORECASTING OF THE 2026 BRAZILIAN PRESIDENTIAL RUNOFF USING ECOLOGICAL INFERENCE AND HISTORICAL VOTE TRANSITION MATRICES*

## *PRONÓSTICO PROBABILÍSTICO DE LA SEGUNDA VUELTA PRESIDENCIAL BRASILEÑA DE 2026 MEDIANTE INFERENCIA ECOLÓGICA Y MATRICES HISTÓRICAS DE TRANSFERENCIA DE VOTOS*

Dâmocles Aurélio Nascimento da Silva Alves[1], Lianny Milenna de Sá Melo[2], Ana Paula Veras Sobral[3], Arundo Nunes da Silva Júnior[4], Maria Luciene Andrade Silva[5]

e7119424
https://doi.org/10.47820/recima21.v7i11.9424

PUBLICADO: 11/2026

**RESUMO**

Em eleições de dois turnos, o resultado final depende do destino dos votos dos candidatos derrotados, dos votos brancos e nulos e da abstenção. Este artigo estima matrizes de transferência por inferência ecológica, com dados por seção eleitoral das eleições presidenciais de 2014, 2018 e 2022, e as combina em uma previsão probabilística para o segundo turno de 2026, elaborada antes de sua realização. Em três exercícios retrospectivos, que envolvem apenas duas eleições-alvo distintas, o erro absoluto médio nacional do modelo básico foi de 1,11 ponto percentual, contra 3,30 pontos do modelo ingênuo. No modelo final, com encolhimento das estimativas estaduais, o erro nacional foi de 1,24 ponto e o erro médio por unidade da federação caiu de 2,20 para 1,72 pontos. A classificação dos candidatos derrotados acrescenta informação preditiva, sem identificar transferências individuais nem efeitos causais da ideologia. Sob a especificação probabilística adotada, a mediana simulada é de 47,9% dos votos válidos para Lula e de 52,1% para Flávio Bolsonaro, com intervalo condicional de 95% de 44,6% a 51,3% para Lula, que vence em 12,1% das vinte mil simulações. Essa frequência é condicional ao modelo e não constitui uma probabilidade eleitoral validada. Ela varia de 3,6% a 23,5% conforme a escala do choque estrutural e de 2,1% a 20,7% quando uma das eleições históricas é excluída, enquanto a mediana permanece entre 47,1% e 48,5%.

**PALAVRAS-CHAVE:** Previsão eleitoral. Inferência ecológica. Transferência de votos. Segundo turno.

***ABSTRACT***

*In two-round elections, the final outcome depends on where the votes of eliminated candidates, blank and null votes and abstention go. This article estimates vote transition matrices through ecological inference, using polling-station data from the 2014, 2018 and 2022 Brazilian presidential elections, and combines them into a probabilistic forecast of the 2026 runoff made before it took place.*

[1] Doutor em Biometria e Estatística Aplicada pela Universidade Federal Rural de Pernambuco (UFRPE) e professor adjunto da Universidade de Pernambuco (UPE), com atuação em Probabilidade e Estatística.
[2] Doutora em Psicologia Cognitiva pela Universidade Federal de Pernambuco (UFPE) e psicóloga da Universidade de Pernambuco (UPE), com atuação em Psicologia Cognitiva e Psicologia da Educação Matemática.
[3] Doutora em Odontologia pela Universidade de São Paulo (USP) e professora associada da Universidade de Pernambuco (UPE), com atuação em Patologia Bucal.
[4] Doutorando em Biometria e Estatística Aplicada pela Universidade Federal Rural de Pernambuco (UFRPE), com atuação em modelagem estatística de séries temporais.
[5] Mestra em Educação pela Universidade de Pernambuco e professora da Secretaria de Educação de Pernambuco, com atuação em Educação.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

*Across three retrospective exercises, which involve only two distinct target elections, the basic model has a national mean absolute error of 1.11 percentage points, compared with 3.30 points for the naive model. In the final model, which shrinks state estimates, the national error is 1.24 points and the mean state-level error falls from 2.20 to 1.72 points. The classification of eliminated candidates adds predictive information, without identifying individual transitions or causal effects of ideology. Under the adopted probabilistic specification, the simulated median is 47.9% of valid votes for Lula and 52.1% for Flávio Bolsonaro, with a conditional 95% interval of 44.6% to 51.3% for Lula, who wins in 12.1% of twenty thousand simulations. This frequency is conditional on the model and is not a validated electoral probability. It ranges from 3.6% to 23.5% depending on the scale of the structural shock and from 2.1% to 20.7% when one historical election is excluded, while the median stays between 47.1% and 48.5%.*

***KEYWORDS:*** *Election forecasting. Ecological inference. Vote transitions. Runoff elections.*

***RESUMEN***
*En elecciones a dos vueltas, el resultado final depende del destino de los votos de los candidatos derrotados, de los votos en blanco y nulos y de la abstención. Este artículo estima matrices de transferencia mediante inferencia ecológica, con datos por mesa electoral de las elecciones presidenciales de 2014, 2018 y 2022, y las combina en un pronóstico probabilístico de la segunda vuelta de 2026, elaborado antes de su realización. En tres ejercicios retrospectivos, que comprenden solo dos elecciones objetivo distintas, el error absoluto medio nacional del modelo básico fue de 1,11 puntos porcentuales, frente a 3,30 puntos del modelo ingenuo. En el modelo final, con contracción de las estimaciones estatales, el error nacional fue de 1,24 puntos y el error medio por unidad federativa bajó de 2,20 a 1,72 puntos. La clasificación de los candidatos derrotados aporta información predictiva, sin identificar transferencias individuales ni efectos causales de la ideología. Bajo la especificación probabilística adoptada, la mediana simulada es del 47,9% de los votos válidos para Lula y del 52,1% para Flávio Bolsonaro, con un intervalo condicional del 95% entre el 44,6% y el 51,3% para Lula, que gana en el 12,1% de veinte mil simulaciones. Esta frecuencia es condicional al modelo y no constituye una probabilidad electoral validada. Varía entre el 3,6% y el 23,5% según la escala del choque estructural y entre el 2,1% y el 20,7% cuando se excluye una de las elecciones históricas, mientras la mediana permanece entre el 47,1% y el 48,5%.*

***PALABRAS CLAVE:*** *Pronóstico electoral. Inferencia ecológica. Transferencia de votos. Segunda vuelta.*

## INTRODUÇÃO

No sistema presidencial brasileiro, o segundo turno é uma eleição nova, mas cercada de informação. Os dois finalistas são conhecidos, assim como o tamanho de cada eleitorado em disputa. O que não se sabe é o destino dos votos que não foram para eles. Em 2026, essa incerteza tem dimensão concreta. Flávio Bolsonaro, do Partido Liberal (PL), terminou o primeiro turno com 56,1 milhões de votos, e Luiz Inácio Lula da Silva (Lula), do Partido dos Trabalhadores (PT), com 53,9 milhões. A diferença de 2,2 milhões é menor que os 9,3 milhões de votos dados aos demais candidatos, os 6,0 milhões de votos brancos e nulos e os 33,5 milhões de abstenções (Brasil, 2026).



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

A literatura de previsão eleitoral trabalha com modelos de fundamentos, como a economia e a popularidade do governo, e com modelos que agregam pesquisas de intenção de voto. Turgeon e Rennó (2012) e Bertholini, Rennó e Turgeon (2022) representam a primeira abordagem, e Jackman (2005), Linzer (2013) e Bunker (2020), a segunda. Este estudo explora outra fonte de informação, o resultado do primeiro turno combinado com padrões históricos de redistribuição dos votos. A estratégia segue a tradição da inferência ecológica de Goodman (1953), King (1997) e Rosen et al. (2001), conjunto de métodos que estimam o comportamento de grupos a partir de dados agregados. Como o voto é secreto, não se sabe em quem votou no segundo turno cada eleitor de um candidato derrotado. Sabe-se apenas quantos votos cada opção recebeu em cada seção nos dois turnos, e é da comparação desses totais, em centenas de milhares de seções, que se estimam as taxas de transferência de cada grupo para cada destino. Essas taxas formam uma matriz de transferência.

O problema de pesquisa é saber se as transferências estimadas em eleições anteriores permitem prever o segundo turno presidencial brasileiro com erro menor que o de uma regra simples baseada apenas no primeiro turno. Essa regra, chamada aqui de modelo ingênuo, supõe que os votos dos derrotados, os brancos e nulos e as abstenções se dividam no segundo turno na mesma proporção observada entre os dois finalistas no primeiro. O objetivo geral é desenvolver e avaliar um método de previsão do segundo turno a partir dos resultados do primeiro turno e de padrões históricos de transferência. Os objetivos específicos são estimar as matrizes de transferência de 2014, 2018 e 2022 para cada unidade da federação (UF), avaliar seu desempenho em aplicações históricas posteriores, verificar se a classificação política dos candidatos derrotados acrescenta informação preditiva e produzir uma previsão probabilística para o segundo turno de 2026, elaborada antes do segundo turno de 25 de outubro de 2026.

A pesquisa se justifica pela escassez de previsões do segundo turno independentes das pesquisas de intenção de voto e pelo uso de dados públicos, que, associado à disponibilização dos procedimentos computacionais e das especificações do modelo, favorece a replicação dos resultados. O artigo não compara o método com agregadores de pesquisas, de modo que sua vantagem em relação a eles não é avaliada. Além desta introdução, o texto apresenta o referencial teórico, a metodologia, os resultados e a discussão e, por fim, as considerações finais.

## 1. REFERENCIAL TEÓRICO

### 1.1. Previsão eleitoral por fundamentos e por pesquisas

Os modelos de fundamentos partem da teoria do voto econômico retrospectivo, segundo a qual o eleitor recompensa ou pune o governo conforme o desempenho da economia e a



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

avaliação da gestão (Lewis-Beck; Stegmaier, 2000). No Brasil, Turgeon e Rennó (2012) enfrentaram o problema do pequeno número de eleições com dados subnacionais, e Bertholini, Rennó e Turgeon (2022) mostraram que modelos baseados na popularidade presidencial superaram os baseados em pesquisas e resistiram à eleição disruptiva de 2018. Esses modelos, porém, preveem o voto no candidato governista no agregado e não tratam a dinâmica específica entre o primeiro e o segundo turno.

A segunda família combina pesquisas ao longo da campanha em modelos bayesianos dinâmicos que estimam o viés de cada instituto, como propuseram Jackman (2005) e Linzer (2013). Em sistemas multipartidários, esses modelos foram estendidos para integrar fundamentos e pesquisas (Stoetzer et al., 2019) e adaptados a democracias recentes (Bunker, 2020, 2022). Seu ponto fraco é herdar os erros das pesquisas, que variam entre países e eleições, mas não diminuíram ao longo do tempo (Jennings; Wlezien, 2018). A literatura sobre combinação de previsões mostra que reunir informações obtidas por caminhos diferentes tende a reduzir o erro em relação a cada previsão isolada (Graefe et al., 2014), o que motiva o uso de uma fonte de informação distinta das pesquisas.

### 1.2. O voto no segundo turno

Nos sistemas de dois turnos, cada turno cria incentivos diferentes. No primeiro, a eleição pode ser decidida se um candidato obtiver a maioria absoluta dos votos válidos, e o eleitor decide se vota em sua opção preferida ou em um candidato com mais chances. No segundo, a escolha se reduz a dois nomes, e o voto passa a expressar sobretudo qual dos finalistas o eleitor considera menos indesejável (Cox, 1997). O eleitor cujo candidato foi eliminado escolhe entre duas alternativas que não eram sua primeira preferência ou deixa de votar validamente. No Brasil, Rennó e Hoepers (2010) mostraram que, em 2006, parte dos eleitores votou em outros candidatos no primeiro turno para punir Lula e voltou a ele no segundo, o que indica que o voto do primeiro turno nem sempre revela a preferência final e torna necessário estimar as transferências em vez de supô-las.

O caso francês ilustra até onde essa lógica pode chegar. Na eleição presidencial de 2017, Romero et al. (2020) estimaram por inferência ecológica as transferências entre os turnos e mostraram que os eleitores dos candidatos eliminados se dividiram de forma muito desigual entre os finalistas. Nas eleições legislativas de 2024, o bloco que liderou o primeiro turno terminou em terceiro lugar em cadeiras, depois que mais de duzentos candidatos desistiram nas disputas a três para concentrar votos contra ele (Mișcoiu; Gherghina, 2025; France, 2024). O episódio mostra que o segundo turno pode inverter a ordem do primeiro quando a rejeição a um finalista se concentra entre os eleitores dos demais. Neste estudo, a matriz estima os destinos



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

dos votos de cada grupo, sem medir rejeição, motivações individuais ou efeitos causais de apoios.

No Brasil recente, essa escolha é fortemente estruturada pela identificação com o PT e pela rejeição a ele. Samuels e Zucco (2018) mostram que o antipetismo é uma identidade política tão relevante quanto o petismo, e Fuks, Ribeiro e Borba (2021) argumentam que, em 2018, essa rejeição se ampliou para uma antipatia generalizada aos partidos, que favoreceu Jair Bolsonaro. Disso decorre a expectativa de que o destino dos votos dos candidatos derrotados dependa de sua posição em relação ao PT. Há também uma dimensão territorial. Desde 2006, a base eleitoral do PT deslocou-se para o Nordeste e para os eleitores de menor renda (Hunter; Power, 2007; Zucco, 2008), o que faz a forma como os votos se redistribuem variar entre regiões. Por isso, as transferências são estimadas separadamente para cada UF.

### 1.3. Inferência ecológica de transferências de votos

Como o voto é secreto, a transferência individual entre turnos não é observada. A regressão de Goodman (1953) supõe que a proporção de eleitores que passa de uma opção do primeiro turno para outra do segundo é a mesma em todas as seções de uma unidade e estima essas proporções por mínimos quadrados. Sem cuidados adicionais, o método pode gerar proporções negativas ou superiores a 100%, problema evitado ao se impor que cada proporção fique entre 0% e 100% e que as proporções de uma mesma origem somem 100%. King (1997) e Rosen et al. (2001) propuseram modelos hierárquicos que permitem variação das taxas entre unidades no caso com várias origens e vários destinos, e Pavía e Romero (2022) desenvolveram estimadores por programação linear.

A principal limitação é a falácia ecológica, isto é, inferir comportamento individual a partir de agregados. A hipótese de homogeneidade das taxas dentro de cada UF é violada quando características das seções, como renda, religião ou região, estão associadas às próprias propensões de transferência. Nesse caso, a regressão atribui a um grupo de origem um comportamento que reflete, em parte, o contexto das seções em que esse grupo é numeroso, o chamado viés de agregação (King, 1997; Greiner; Quinn, 2009). Usar seções eleitorais, com cerca de 330 eleitores em média, reduz a heterogeneidade dentro das unidades, mas não elimina esse viés, que depende da correlação entre composição e comportamento, e não do tamanho da unidade. As restrições de não negatividade e soma unitária garantem estimativas logicamente possíveis, mas não garantem que elas coincidam com os verdadeiros fluxos, que permanecem não observados. O problema é mais grave para grupos pequenos, cujas taxas são pouco restringidas pelos dados. Por isso, este artigo não pretende identificar comportamento individual.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Seu objetivo é preditivo, e o critério de desempenho é o erro nas aplicações históricas posteriores.

### 1.4. Hipóteses

A hipótese nula $H_0$ estabelece que, nas aplicações históricas, as diferenças pareadas entre os erros absolutos do modelo de transferência e do modelo ingênuo não tendem a favorecer o primeiro. A hipótese alternativa $H_1$ estabelece uma diferença favorável ao modelo de transferência. A comparação é feita por UF. Como as unidades compartilham choques nacionais e há apenas duas eleições-alvo distintas, os testes funcionam como diagnósticos desse conjunto de eleições, e não como prova de superioridade em eleições futuras. Com base na literatura sobre antipetismo, examina-se também se separar os candidatos derrotados por posição política melhora a previsão em relação a reuni-los em um único grupo, contraste que avalia informação preditiva, sem testar um efeito causal da ideologia.

## 2. METODOLOGIA

### 2.1. Fontes e preparação dos dados

Os dados vêm do Portal de Dados Abertos do Tribunal Superior Eleitoral (TSE), com dois arquivos para cada eleição presidencial de 2014, 2018 e 2022 (Brasil, 2014, 2018, 2022). O primeiro é a votação por seção eleitoral, que traz os votos de cada candidato e os votos brancos e nulos. O segundo é o detalhe da apuração por seção, que traz o número de eleitores aptos e as abstenções. Os registros foram filtrados para o cargo de presidente e agregados por turno, UF, município, zona e seção, chave usada para parear as seções entre os turnos.

O arquivo de detalhe contém as mesmas seções nos dois turnos, 428.894 em 2014, 454.490 em 2018 e 472.075 em 2022. Foram excluídas as seções sem registro de votação nominal no segundo turno, 42 em 2014, 42 em 2018 e 47 em 2022, com 748, 1.139 e 657 eleitores aptos, respectivamente, e comparecimento nulo no segundo turno. Restaram 428.852, 454.448 e 472.028 seções, incluindo as do exterior, tratadas como uma unidade à parte. As exclusões correspondem a menos de 0,001% do eleitorado e não alteram os totais nacionais na primeira casa decimal. O número de eleitores aptos não é idêntico em todas as seções nos dois turnos. Ele difere em 9.571 seções em 2014, 17.026 em 2018 e 49.710 em 2022, com soma das diferenças absolutas de 15.202, 24.103 e 72.306 eleitores, embora os totais nacionais coincidam. Por isso, as frações de origem são calculadas sobre os aptos do primeiro turno e as de destino sobre os aptos do segundo, de modo que cada linha some um em cada turno. Para 2026, os resultados do primeiro turno por UF foram obtidos no sistema de divulgação de resultados do



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

TSE, com todas as seções totalizadas, e os nomes e partidos dos candidatos foram conferidos com esse registro (Brasil, 2026).

Em cada seção, o eleitorado apto é dividido em sete categorias de origem no primeiro turno, que são o candidato do PT, o principal candidato de oposição ao PT, os demais candidatos de esquerda, os demais candidatos de centro, os demais candidatos de direita, os votos brancos e nulos e a abstenção. O principal candidato de oposição é, em cada eleição, o adversário do PT no segundo turno. O destino tem quatro categorias, que são PT, oposição, brancos e nulos e abstenção.

A classificação dos demais candidatos, apresentada no Quadro 1, é documentada por um protocolo em três etapas. Na primeira, cada candidato recebeu a posição de seu partido na escala ideológica de Bolognesi, Ribeiro e Codato (2023), obtida por consulta a especialistas em 2018, de 0 (extrema esquerda) a 10 (extrema direita), agrupada em esquerda (até 4,49), centro (de 4,50 a 7,00) e direita (acima de 7,00). Na segunda, partidos ausentes da escala, como UP, PPL, União Brasil, Missão e Democrata, foram classificados pela autodefinição programática. Na terceira, a classificação do candidato foi revista quando sua campanha se afastava da posição do partido, com base no programa registrado e no posicionamento público em relação ao PT e ao principal opositor. Dos 33 candidatos cujos partidos constam da escala, 27 mantiveram a classificação partidária, e seis foram revistos. Marina Silva, em 2014, do PSB (4,05), foi classificada como centro por ter concorrido como alternativa aos dois polos e apoiado o candidato do PSDB no segundo turno. Geraldo Alckmin (PSDB, 7,11), Henrique Meirelles (MDB, 7,01) e Álvaro Dias (Podemos, 7,24), em 2018, e Simone Tebet (MDB, 7,01), em 2022, foram classificados como centro porque seus partidos estão no limite entre centro e direita e eles se apresentaram como candidatos de centro. Padre Kelmon, em 2022, do PTB (6,10), foi classificado como direita pela campanha alinhada ao candidato do PL. Como a classificação influencia as matrizes, a seção de robustez testa uma classificação alternativa.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

**Quadro 1.** Classificação dos candidatos derrotados por eleição

| Eleição | PT | Principal opositor (finalista) | Outros de esquerda | Outros de centro | Outros de direita |
|---|---|---|---|---|---|
| 2014 | Dilma Rousseff | Aécio Neves (PSDB) | Luciana Genro (PSOL), Zé Maria (PSTU), Mauro Iasi (PCB), Rui Costa Pimenta (PCO) | Marina Silva (PSB), Eduardo Jorge (PV) | Pastor Everaldo (PSC), Levy Fidelix (PRTB), José Maria Eymael (DC) |
| 2018 | Fernando Haddad | Jair Bolsonaro (PSL) | Ciro Gomes (PDT), Guilherme Boulos (PSOL), Vera Lúcia (PSTU), João Goulart Filho (PPL) | Geraldo Alckmin (PSDB), Henrique Meirelles (MDB), Marina Silva (Rede), Álvaro Dias (Podemos) | João Amoêdo (Novo), Cabo Daciolo (Patriota), José Maria Eymael (DC) |
| 2022 | Luiz Inácio Lula da Silva | Jair Bolsonaro (PL) | Ciro Gomes (PDT), Vera Lúcia (PSTU), Sofia Manzano (PCB), Léo Péricles (UP) | Simone Tebet (MDB), Soraya Thronicke (União) | Felipe d'Ávila (Novo), Padre Kelmon (PTB), José Maria Eymael (DC) |
| 2026 | Luiz Inácio Lula da Silva | Flávio Bolsonaro (PL) | Samara (UP), Hertz Dias (PSTU), Edmilson Costa (PCB), Rui Costa Pimenta (PCO) | Augusto Cury (Avante) | Renan Santos (Missão), Ronaldo Caiado (PSD), Romeu Zema (Novo), Clariana Barão (DC), Wilson Grassi (Democrata) |

Fonte: Elaborado pelos autores (2026), com base em Brasil (2014, 2018, 2022, 2026) e Bolognesi, Ribeiro e Codato (2023).

### 2.2. Modelo de transferência

Seja $x_{ir}$ a fração do eleitorado apto da seção $i$ na origem $r$, com $r = 1, \ldots, 7$, e $y_{ik}$ a fração no destino $k$, com $k = 1, \ldots, 4$. Sob a hipótese de homogeneidade dentro de cada UF $u$, as frações se relacionam pela matriz de transferência $B^u$, cujos elementos $\beta_{rk}^u$ são as taxas de transferência da origem $r$ para o destino $k$, conforme a Equação 1, em que $\varepsilon_{ik}$ é um termo de erro de média zero.

$$y_{ik} = \sum_{r=1}^{7} x_{ir}\, \beta_{rk}^{u} + \varepsilon_{ik} \tag{1}$$

Como cada linha da matriz é uma distribuição de proporções, impõem-se as restrições da Equação 2.

$$\beta_{rk}^{u} \geq 0, \qquad \sum_{k=1}^{4} \beta_{rk}^{u} = 1 \tag{2}$$

A matriz de cada UF é estimada por mínimos quadrados ponderados, com peso $N_i$ igual ao número de eleitores aptos da seção, conforme a Equação 3, sujeita às restrições da Equação 2.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

$$\hat{B}^u = \arg\min_{B} \sum_{i \in u} N_i \sum_{k=1}^{4} \left( y_{ik} - \sum_{r=1}^{7} x_{ir}\, \beta_{rk} \right)^2 \tag{3}$$

O problema é quadrático com restrições lineares e foi resolvido pelo algoritmo SLSQP da biblioteca SciPy (Virtanen et al., 2020), com tolerância de $10^{-12}$, limite de 500 iterações e ponto de partida uniforme. São estimadas vinte e oito matrizes por eleição, incluindo o Distrito Federal e o exterior. As oitenta e quatro otimizações convergiram, e as somas das linhas diferem de um em, no máximo, $2{,}22 \times 10^{-16}$. Aplicadas ao próprio primeiro turno, as matrizes reproduzem os totais do segundo com erro relativo máximo, por UF e finalista, de 0,70% em 2014, 0,91% em 2018 e 0,50% em 2022. Esse ajuste aos agregados não demonstra que as taxas estimadas coincidam com os fluxos individuais.

### 2.3. Agregação e medida de desempenho

Seja $X_r^u$ o número de eleitores da UF *u* na origem *r*. O número previsto de votos no destino *k* resulta da soma, entre UFs e origens, dos eleitores de cada origem multiplicados pelas taxas estimadas, conforme a Equação 4.

$$\hat{V}_k = \sum_{u} \sum_{r=1}^{7} X_r^u\, \hat{\beta}_{rk}^u \tag{4}$$

A fatia prevista do PT nos votos válidos e o erro de previsão são dados pela Equação 5, em que PT e D indicam os dois finalistas e *s* é a fatia observada. O erro é calculado para o Brasil e para cada UF, e o modelo ingênuo atribui ao PT a proporção entre os finalistas observada no primeiro turno.

$$\hat{s} = \frac{\hat{V}_{\mathrm{PT}}}{\hat{V}_{\mathrm{PT}} + \hat{V}_{\mathrm{D}}}, \qquad e = \hat{s} - s \tag{5}$$

### 2.4. Avaliação retrospectiva

As matrizes de 2014 e 2018 são aplicadas aos primeiros turnos posteriores nos exercícios 2014→2018, 2018→2022 e 2014→2022. Os dois últimos têm o mesmo resultado observado, de modo que há apenas duas eleições-alvo distintas. Para cada UF, calcula-se a média, nos três exercícios, da diferença entre os erros absolutos do modelo de transferência e do modelo ingênuo, o que evita contar cada exercício como uma observação territorial independente. Essas 28 médias são avaliadas pelo teste unilateral dos postos sinalizados de



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Wilcoxon (1945), que pressupõe simetria das diferenças, e por *bootstrap* por UF (Efron; Tibshirani, 1993), com dez mil reamostragens e semente 20261011. A comparação entre a classificação política e o bloco único de derrotados usa o mesmo procedimento, com cinco mil reamostragens e semente 20261007. Esses procedimentos não eliminam os choques nacionais compartilhados pelas UFs nem a reutilização de eleições, e por isso a análise privilegia a magnitude dos erros.

### 2.5. Encolhimento das estimativas estaduais

Para estabilizar as linhas estimadas sobre grupos pequenos, cada linha da matriz estadual é combinada com a linha nacional correspondente, obtida pela média das linhas estaduais ponderada pelo número de eleitores de cada origem, conforme a Equação 6. O peso da linha estadual cresce com o número de eleitores $n_r^u$ da origem na UF.

$$\tilde{\beta}_r^u = w_r^u \, \hat{\beta}_r^u + \left(1 - w_r^u\right) \bar{\beta}_r, \qquad w_r^u = \frac{n_r^u}{n_r^u + \kappa} \tag{6}$$

O parâmetro $\kappa$ controla a intensidade do encolhimento. Com $\kappa$ igual a cinquenta mil, a linha estadual recebe metade do peso quando a origem reúne cinquenta mil eleitores na UF, o equivalente a cerca de 150 seções médias, e predomina acima disso. O valor foi definido na especificação original após o exame de uma grade de valores nos mesmos exercícios usados na avaliação, o que constitui uma limitação. Ele tem, porém, uma interpretação substantiva independente desses exercícios, a de que grupos com menos de cerca de 150 seções oferecem pouca variação entre seções para identificar suas taxas. O procedimento é inspirado na regularização bayesiana empírica (Gelman et al., 2013), mas não deriva de uma posterior especificada conjuntamente. O objetivo do encolhimento é a estabilidade das estimativas territoriais, que compõem a previsão por UF, e não a precisão nacional. Como diagnóstico, foram examinados sete valores de $\kappa$, de dez mil a um milhão. O erro territorial médio de $\kappa$ igual a cinquenta mil, 1,725 ponto, é próximo ao do melhor valor da grade, vinte e cinco mil, com 1,719 ponto, e a previsão nacional para 2026 varia apenas entre 47,8% e 47,9% em toda a grade.

O artigo usa, portanto, três modelos. O modelo ingênuo é a referência de comparação. O modelo básico usa as matrizes estaduais sem encolhimento e é a referência das variações da seção de robustez, pois mede o desempenho do modelo de transferência em si. O modelo final aplica o encolhimento com $\kappa$ igual a cinquenta mil e é o usado na previsão para 2026. Os resultados retrospectivos são apresentados separadamente para os modelos básico e final.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

### 2.6. Combinação probabilística das matrizes

A previsão para 2026 combina as matrizes do modelo final de 2014, 2018 e 2022 por calibração preditiva empírica. A distribuição simulada não é uma posterior obtida de uma verossimilhança e de uma distribuição a priori especificadas conjuntamente. Para cada origem *r*, sorteia-se um vetor nacional de transferências $\theta_r$ de uma distribuição de Dirichlet centrada na média $m_r$ das três matrizes nacionais do modelo básico, conforme a Equação 7. A distribuição de Dirichlet foi escolhida por gerar vetores de proporções positivas que somam um, como as linhas de uma matriz de transferência, e por ter um único parâmetro de concentração, que pode ser estimado com poucas observações (Gelman et al., 2013).

$$\theta_r \sim \mathrm{Dirichlet}\left(\alpha_r\, m_r\right), \qquad m_r = \frac{1}{3}\sum_t \bar{\beta}_r^{(t)} \tag{7}$$

A concentração $\alpha_r$ controla a dispersão dos sorteios e é estimada pelo método dos momentos da Equação 8, em que $S^2_{rk}$ é a variância entre as três matrizes nacionais. Origens cujas transferências mudaram pouco entre eleições recebem concentração alta. Os valores estimados são 742,8 para o PT, 420,3 para o principal opositor, 7,6 para a esquerda, 38,9 para o centro, 3,0 para a direita, 16,2 para brancos e nulos e 114,5 para a abstenção. Estimados com apenas três matrizes, esses valores são instáveis. Excluindo uma eleição de cada vez, a concentração da esquerda varia de 3,6 a 70,1, a da direita de 2,0 a 58,1 e a do centro de 29,1 a 57,1, e a influência de cada eleição sobre a previsão é avaliada na seção de robustez.

$$\hat{\alpha}_r = \frac{\sum_k m_{rk}\left(1 - m_{rk}\right)}{\sum_k S^2_{rk}} - 1 \tag{8}$$

Em vinte mil simulações, com semente 20261007, sorteia-se um vetor por origem, e seu desvio em relação à média é somado à média das matrizes estaduais do modelo final. O mesmo desvio atinge todas as UFs. Valores negativos são truncados em zero e cada linha é renormalizada, o que produz matrizes válidas, embora as linhas estaduais resultantes não tenham distribuição de Dirichlet exata. As matrizes simuladas são aplicadas ao primeiro turno de 2026, o que fornece, em cada UF *u*, a fatia $p_u$ de Lula e o total $V_u$ de votos válidos. O erro



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

estrutural, que representa fatores não captados pelas transferências históricas, como a campanha do segundo turno, é introduzido em escala logística, conforme a Equação 9, com um sorteio *z* comum a todas as UFs.

$$p_u^* = \mathrm{logit}^{-1}\big(\mathrm{logit}(p_u) + \tau z\big), \qquad z \sim \mathcal{N}(0,1) \tag{9}$$

O parâmetro *τ*, igual a 0,05566, é calibrado para que o desvio-padrão aproximado do choque na fatia nacional seja igual à raiz do erro quadrático médio do modelo final nos exercícios retrospectivos, 1,30 ponto percentual. A previsão nacional resulta da soma dos votos simulados em todas as UFs, o que mantém as fatias entre zero e um e a coerência entre os resultados estaduais e o nacional. Mediana, quantis e frequência de vitória são condicionais a essa especificação. Uma checagem retrospectiva usa apenas as matrizes de 2014 e 2018 para projetar 2022, com semente 20261011. Sem choque estrutural, o intervalo de 95% vai de 50,6% a 53,3%, com mediana de 52,0%, e contém o resultado observado, de 50,9%. Com o choque calibrado no único erro anterior disponível, 1,41 ponto, o intervalo vai de 49,0% a 55,1%. Com uma única eleição-alvo e parâmetros definidos retrospectivamente, essa checagem não estabelece a cobertura dos intervalos em eleições futuras.

O processamento foi feito em Python, com as bibliotecas pandas (McKinney, 2010), NumPy (Harris et al., 2020) e SciPy (Virtanen et al., 2020). Os scripts, as matrizes estimadas, os vetores de entrada de 2026, as sementes e os resultados das simulações compõem o material de replicação encaminhado à revista, e os dados de origem são públicos (Brasil, 2014, 2018, 2022, 2026).

## 3. RESULTADOS E DISCUSSÃO

### 3.1. Matrizes de transferência

A Tabela 1 apresenta as matrizes nacionais do modelo básico, obtidas pela soma ponderada das vinte e oito matrizes estaduais. Cada linha corresponde a uma categoria de origem no primeiro turno, e cada coluna, a um destino no segundo. Em 2018, por exemplo, estima-se que 78,8% dos votos da esquerda não petista tenham ido para Haddad, 9,5% para Bolsonaro e 11,7% para brancos, nulos e abstenção. Em cada ano, as três colunas de uma linha somam 100%.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

**Tabela 1.** Taxas estimadas de transferência de votos entre o primeiro e o segundo turno, segundo o modelo de inferência ecológica (% da categoria de origem)

| Categoria de origem no primeiro turno | 2014 (Dilma e Aécio) | | | 2018 (Haddad e Bolsonaro) | | | 2022 (Lula e Bolsonaro) | | |
|---|---|---|---|---|---|---|---|---|---|
| | PT | PSDB | Branco, nulo ou abstenção | PT | PSL | Branco, nulo ou abstenção | PT | PL | Branco, nulo ou abstenção |
| Votos no candidato do PT | 95,4 | 0,2 | 4,4 | 95,0 | 0,0 | 5,0 | 96,3 | 0,0 | 3,7 |
| Votos no principal opositor | 0,2 | 99,4 | 0,4 | 0,0 | 99,2 | 0,8 | 0,0 | 99,9 | 0,1 |
| Votos em outros de esquerda | 50,7 | 34,5 | 14,8 | 78,8 | 9,5 | 11,7 | 39,8 | 39,4 | 20,8 |
| Votos em outros de centro | 29,0 | 58,3 | 12,7 | 23,5 | 49,6 | 26,9 | 36,4 | 44,0 | 19,6 |
| Votos em outros de direita | 70,3 | 10,6 | 19,1 | 17,2 | 52,6 | 30,2 | 20,9 | 61,2 | 17,9 |
| Votos brancos e nulos | 43,2 | 10,8 | 46,0 | 31,0 | 9,6 | 59,4 | 16,1 | 27,8 | 56,1 |
| Abstenção | 0,3 | 3,6 | 96,1 | 1,3 | 1,2 | 97,5 | 2,3 | 4,0 | 93,7 |

Fonte: Elaborada pelos autores (2026), com base em Brasil (2014, 2018, 2022).

Nota: As linhas são categorias de origem definidas por totais de votos por seção, e não eleitores acompanhados individualmente entre os turnos. As taxas são estimativas do modelo básico sob a hipótese de homogeneidade dentro de cada UF.

As taxas dos dois finalistas e da abstenção são estáveis entre eleições, e a variação se concentra nos grupos de candidatos derrotados. A taxa estimada da esquerda não petista para o PT cai de 78,8% em 2018 para 39,8% em 2022. A parcela do centro atribuída ao adversário do PT supera a atribuída ao PT nas três eleições, sem maioria absoluta em 2018 e 2022. A taxa de 70,3% dos demais candidatos de direita para o PT em 2014 exige cautela. Esse grupo reuniu apenas 1,3 milhão de votos, cerca de 1% do eleitorado, e a estimativa permanece elevada em todos os procedimentos, variando de 66,7% com programação linear a 78,7% com zonas eleitorais e chegando a 71,7% com encolhimento. A persistência do valor não indica que ele seja correto. Ela é compatível com viés de agregação, pois esses candidatos tiveram mais votos em seções onde o PT cresceu entre os turnos por outras razões. O efeito sobre a previsão é relevante. Quando essa linha é substituída pela média de 2018 e 2022, a previsão para 2026 com o padrão de 2014 cai de 49,6% para 47,2%.

### 3.2. Testes retrospectivos

A Tabela 2 e a Figura 1 comparam os erros nacionais dos modelos. O erro é a diferença entre a previsão e a votação real do PT nos votos válidos, em pontos percentuais, e um valor positivo indica superestimação.



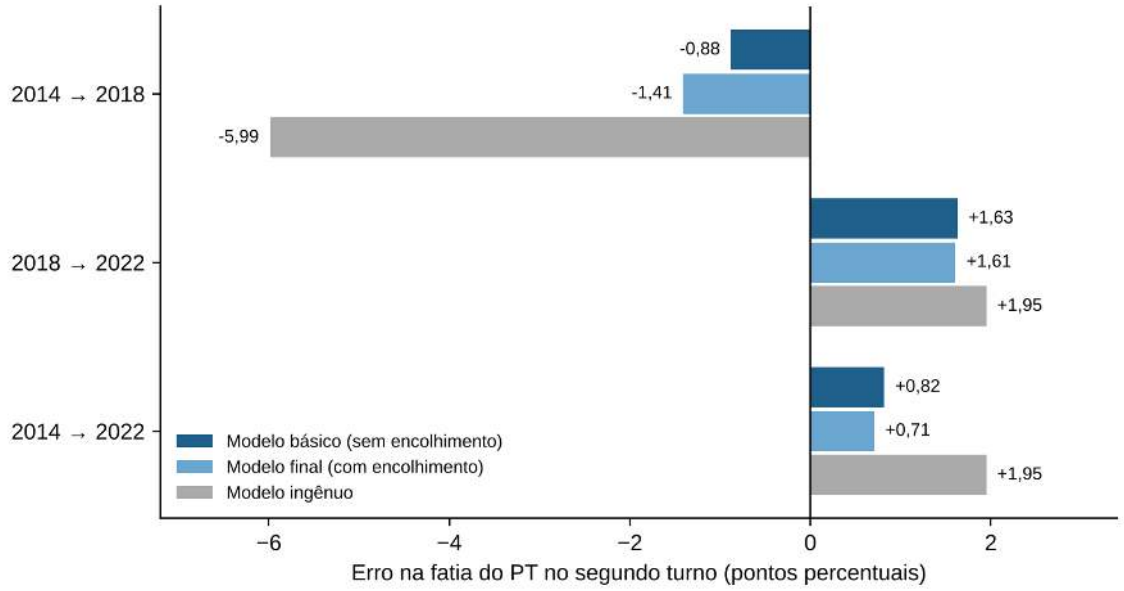
Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

**Tabela 2**. Erros nacionais e territoriais dos modelos básico, final e ingênuo na previsão da votação do PT no segundo turno

| Exercício | Votação real do PT (%) | Erro do modelo básico (p.p.) | Erro do modelo final (p.p.) | Erro do modelo ingênuo (p.p.) | Erro médio por UF, básico, final e ingênuo (p.p.) |
|---|---|---|---|---|---|
| 2014 → 2018 | 44,87 | −0,88 | −1,41 | −5,99 | 3,17, 2,01 e 5,36 |
| 2018 → 2022 | 50,90 | +1,63 | +1,61 | +1,95 | 1,81, 1,74 e 1,81 |
| 2014 → 2022 | 50,90 | +0,82 | +0,71 | +1,95 | 1,61, 1,42 e 1,81 |
| Média dos erros absolutos | | 1,11 | 1,24 | 3,30 | 2,20, 1,72 e 2,99 |

Fonte: Elaborada pelos autores (2026), com base em Brasil (2014, 2018, 2022).
Nota: p.p. indica pontos percentuais. Erro positivo indica superestimação da votação do PT.

**Figura 1.** Erro nacional na previsão da votação do PT nos exercícios retrospectivos, por modelo (pontos percentuais)




Fonte: Elaborada pelos autores (2026), com base em Brasil (2014, 2018, 2022).

O modelo básico tem erros nacionais de −0,88, +1,63 e +0,82 ponto, com erro absoluto médio de 1,11 e raiz do erro quadrático médio de 1,17 ponto, contra 3,30 pontos do modelo ingênuo. O modelo final tem erros de −1,41, +1,61 e +0,71 ponto, com erro absoluto médio de 1,24 e raiz do erro quadrático médio de 1,30 ponto. A vantagem não é uniforme. Ela é grande em 2014→2018, quando o modelo ingênuo não capta a migração do voto de Ciro Gomes para o PT, e pequena em 2018→2022, em que o erro territorial médio do modelo básico, 1,8139 ponto, é ligeiramente maior que o do modelo ingênuo, 1,8100 ponto. Na comparação por UF, o modelo básico erra menos que o ingênuo em 79% das unidades, com redução média de 0,80 ponto



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

(Wilcoxon, p = 0,002, intervalo *bootstrap* de 0,29 a 1,32 ponto de redução). No modelo final, a redução média é de 1,27 ponto, em 82% das unidades (p < 0,001, intervalo de 0,79 a 1,79 ponto). Esses resultados rejeitam $H_0$ nas eleições analisadas, mas a superioridade se refere à média de três exercícios com duas eleições-alvo, cujas UFs compartilham os mesmos choques nacionais, e não demonstra desempenho superior em eleições futuras.

O encolhimento reduz o erro territorial médio de 2,20 para 1,72 pontos, mas aumenta o erro nacional de 1,11 para 1,24 ponto. A escolha do modelo final prioriza, portanto, a estabilidade das previsões estaduais, com pequeno custo na precisão nacional. Com $\kappa$ escolhido em dois exercícios e avaliado no terceiro, o encolhimento reduziu o erro territorial nos três casos, de 3,17 para 2,70, de 1,81 para 1,74 e de 1,61 para 1,45 ponto, mas os exercícios compartilham eleições e essa verificação não é independente.

### 3.3. Previsão para o segundo turno de 2026

A Tabela 3 apresenta as previsões determinísticas obtidas com cada matriz histórica e a combinação probabilística. Cada cenário aplica aos grupos de 2026 as taxas do grupo equivalente na eleição indicada, sem supor que os candidatos sejam os mesmos.

**Tabela 3**. Previsões para o segundo turno de 2026 (% dos votos válidos)

| Cenário ou modelo | Lula (PT) | Flávio Bolsonaro (PL) |
|---|---|---|
| Taxas dos grupos equivalentes de 2014 (modelo final) | 49,6 | 50,4 |
| Taxas dos grupos equivalentes de 2018 (modelo final) | 47,5 | 52,5 |
| Taxas dos grupos equivalentes de 2022 (modelo final) | 46,6 | 53,4 |
| Combinação probabilística, mediana das simulações | 47,9 | 52,1 |
| Intervalo condicional de 95% | 44,6 a 51,3 | 48,7 a 55,4 |
| Frequência de vitória nas simulações | 12,1% | 87,9% |
| Modelo ingênuo | 49,0 | 51,0 |

Fonte: Elaborada pelos autores (2026), com base em Brasil (2014, 2018, 2022, 2026).

As previsões determinísticas variam de 46,6% a 49,6% para Lula, conforme a eleição de referência, dispersão que reflete as mudanças no perfil dos candidatos e na composição dos grupos entre eleições. A combinação probabilística representa essa variação pela dispersão da distribuição de Dirichlet, mas apenas na medida em que as três eleições históricas a capturam. Ficam fora do modelo a incerteza de estimação de cada matriz, a correspondência imperfeita entre os grupos de 2026 e os históricos, como a candidatura de Renan Santos, e os fatos da campanha do segundo turno, representados apenas de forma agregada pelo choque estrutural. Por isso, a mediana de 47,9% não sintetiza todas as fontes de incerteza.

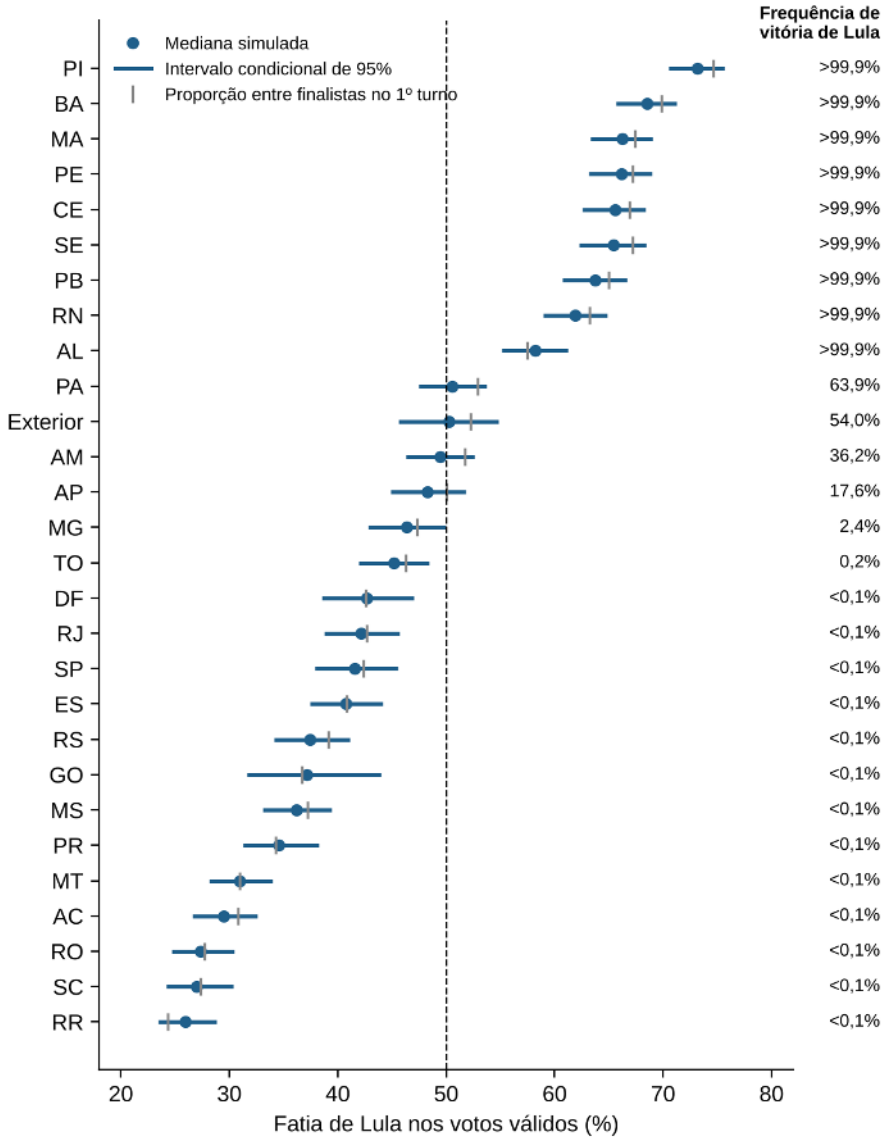
Sob a especificação principal, Lula vence em 2.429 das vinte mil simulações, ou 12,1%, com erro de Monte Carlo de 0,23 ponto percentual, que mede apenas a precisão computacional.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Essa frequência é condicional às matrizes, à classificação, ao encolhimento e à escala do choque e não constitui uma probabilidade eleitoral validada nem um resultado eleitoral. A Figura 2 apresenta a previsão por UF. As frequências de vitória de Lula são de 63,9% no Pará, 54,0% no exterior, 36,2% no Amazonas, 17,6% no Amapá e 2,4% em Minas Gerais, e os intervalos estaduais não são independentes entre si, pois compartilham o choque comum.

**Figura 2.** Previsão da votação de Lula no segundo turno de 2026, por unidade da federação, com mediana e intervalo condicional de 95% (% dos votos válidos)




Fonte: Elaborada pelos autores (2026), com base em Brasil (2014, 2018, 2022, 2026).



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Nota: O traço vertical indica a proporção de Lula entre os finalistas no primeiro turno. Os intervalos estaduais compartilham o choque comum e não são independentes.

### 3.4. Robustez e sensibilidade

A Tabela 4 apresenta variações do modelo básico em que uma escolha é alterada de cada vez. Um valor positivo indica superestimação da votação do PT, e a última coluna é a média das previsões determinísticas com as três matrizes.

**Tabela 4**. Erros nos exercícios retrospectivos e previsão determinística para 2026, por versão do modelo

| Versão do modelo | 2014 → 2018 (p.p.) | 2018 → 2022 (p.p.) | 2014 → 2022 (p.p.) | Erro absoluto médio (p.p.) | Lula em 2026 (%) |
|---|---|---|---|---|---|
| Modelo básico (seções e mínimos quadrados) | −0,88 | +1,63 | +0,82 | 1,11 | 47,8 |
| Variação 1. Zonas e mínimos quadrados | −1,40 | +1,65 | +0,54 | 1,20 | 47,7 |
| Variação 2. Zonas e programação linear | −2,41 | +1,57 | +0,44 | 1,47 | 47,7 |
| Variação 3. Classificação alternativa | −0,64 | +3,47 | +1,81 | 1,97 | 48,0 |
| Variação 4. Derrotados em um único grupo | −5,13 | +1,91 | +0,06 | 2,37 | 48,1 |
| Variação 5. Modelo final, com encolhimento | −1,41 | +1,61 | +0,71 | 1,24 | 47,9 |
| Modelo ingênuo | −5,99 | +1,95 | +1,95 | 3,30 | 49,0 |

Fonte: Elaborada pelos autores (2026), com base em Brasil (2014, 2018, 2022, 2026).

As variações 1 e 2 usam zonas eleitorais e diferem apenas no estimador, de modo que sua comparação isola o efeito da programação linear em relação aos mínimos quadrados. Com zonas, os mínimos quadrados têm erro absoluto médio de 1,20 ponto, e a programação linear, de 1,47, com previsão de 47,7% em ambos os casos. A comparação entre a variação 1 e o modelo básico isola o efeito da agregação. A classificação alternativa, que passa Marina Silva, em 2018, e Simone Tebet, em 2022, para a esquerda e Augusto Cury, em 2026, para a direita, eleva o erro do exercício 2018→2022 para 3,47 pontos. Reunir todos os derrotados em um único grupo eleva o erro de 2018 para 5,13 pontos, próximo ao do modelo ingênuo. Na comparação por UF, a separação por posição política reduz o erro em 64% das unidades, com redução média de 0,58 ponto (p = 0,015, intervalo *bootstrap* de 0,12 a 1,03 ponto). A classificação acrescenta informação preditiva nos exercícios analisados, sem demonstrar efeito causal da ideologia.

A Tabela 5 mostra a sensibilidade da previsão para 2026 a três escolhas da etapa probabilística, com os mesmos sorteios da especificação principal.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

**Tabela 5**. Sensibilidade da previsão para 2026 às escolhas da etapa probabilística

| Especificação | Mediana de Lula (%) | Intervalo condicional de 95% (%) | Frequência de vitória de Lula (%) |
|---|---|---|---|
| **Escala do choque comum (múltiplo de τ)** | | | |
| 0 (sem choque) | 47,9 | 45,8 a 50,2 | 3,6 |
| 0,5 | 47,9 | 45,4 a 50,5 | 6,3 |
| 1,0 (especificação principal) | 47,9 | 44,6 a 51,3 | 12,1 |
| 1,5 | 47,9 | 43,5 a 52,4 | 18,4 |
| 2,0 | 47,9 | 42,5 a 53,6 | 23,5 |
| **Eleições usadas na calibração** | | | |
| 2014 e 2018 (sem 2022) | 48,5 | 45,0 a 51,9 | 20,7 |
| 2014 e 2022 (sem 2018) | 48,1 | 44,4 a 51,7 | 16,1 |
| 2018 e 2022 (sem 2014) | 47,1 | 44,3 a 49,9 | 2,1 |
| **Choque estadual independente adicional** | | | |
| Desvio igual a τ | 47,9 | 44,5 a 51,4 | 12,8 |
| Desvio igual a 2τ | 47,9 | 44,2 a 51,7 | 14,2 |

Fonte: Elaborada pelos autores (2026), com base em Brasil (2014, 2018, 2022, 2026).
Nota: Vinte mil simulações em cada especificação. As frequências são condicionais ao modelo e não constituem probabilidades eleitorais validadas.

A mediana é estável em todas as especificações, entre 47,1% e 48,5%, mas a frequência de vitória de Lula não é. Ela varia de 3,6% a 23,5% conforme a escala do choque estrutural, cuja calibração depende de apenas três erros históricos. A eleição de 2014 tem a maior influência. Sem ela, a frequência cai para 2,1%, e sem 2018 ou 2022 sobe para 16,1% e 20,7%, o que reflete, em parte, a taxa de 2014 dos demais candidatos de direita discutida acima. Choques estaduais independentes, somados ao choque comum, quase não alteram o resultado nacional, mas ampliam a incerteza estadual, elevando a frequência de vitória de Lula em Minas Gerais de 2,4% para 5,8% e 13,7%. A previsão deve, portanto, ser lida como uma vantagem de Flávio Bolsonaro, cuja magnitude é estável, com probabilidade de inversão condicional e sensível às escolhas do modelo.

### 3.5. Discussão

Os exercícios retrospectivos são favoráveis ao modelo de transferência, mas com evidência limitada, pois envolvem duas eleições-alvo e a vantagem se concentra em 2014→2018. O contraste entre as matrizes indica que o destino estimado dos votos dos derrotados varia entre eleições. Na esquerda não petista, a parcela atribuída ao PT cai de 78,8% em 2018 para 39,8% em 2022, enquanto o eleitorado de Ciro Gomes, principal candidato do grupo, diminui de 13,3 para 3,6 milhões de votos. Uma interpretação possível é que parte dos eleitores próximos ao PT tenha votado em Lula já no primeiro turno de 2022, alterando a composição dos que permaneceram com Ciro, o que dialoga com Samuels e Zucco (2018). O



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

modelo não testa essa interpretação, pois os agregados não acompanham os mesmos indivíduos.

A parcela dos votos de centro atribuída ao adversário do PT supera a atribuída ao PT nas três eleições, inclusive em 2022, quando Simone Tebet declarou apoio a Lula. Isso não estima o efeito de apoios, que exigiria outro desenho de pesquisa. Até 11 de outubro de 2026, Ronaldo Caiado havia declarado apoio pessoal a Flávio Bolsonaro, embora o PSD tivesse liberado seus filiados (Gazeta do Povo, 2026), Augusto Cury havia anunciado apoio condicionado a Flávio (Poder360, 2026) e Renan Santos não havia declarado apoio a nenhum dos finalistas (O Povo, 2026). Essas informações jornalísticas contextualizam a disputa e não entram no modelo.

O modelo de transferência pode complementar previsões baseadas em pesquisas por usar uma fonte diferente de informação. Seu ganho sobre o modelo ingênuo é maior quando os votos dos derrotados têm destinos assimétricos, como em 2018. Os valores de p e os intervalos *bootstrap* descrevem a regularidade nas eleições analisadas e não substituem a avaliação prospectiva, que só será possível com novas eleições.

## 4. CONSIDERAÇÕES FINAIS

Nos exercícios históricos, o modelo básico apresentou erro absoluto médio nacional de 1,11 ponto percentual, contra 3,30 do modelo ingênuo, e o modelo final, usado na previsão, apresentou 1,24 ponto, com menor erro territorial. Esses resultados se referem a três exercícios com duas eleições-alvo e não asseguram desempenho equivalente em eleições futuras. Aplicada ao primeiro turno de 2026, a combinação probabilística produz mediana de 47,9% para Lula, intervalo condicional de 44,6% a 51,3% e frequência de vitória de 12,1%. A previsão é um exercício condicionado aos resultados do primeiro turno e à especificação estatística adotada, e a frequência de vitória não constitui uma probabilidade eleitoral validada nem um resultado eleitoral. Nas análises de sensibilidade, essa frequência varia de 2,1% a 23,5%, enquanto a vantagem de Flávio Bolsonaro na mediana permanece estável.

O estudo tem limitações. As taxas são supostas homogêneas dentro de cada UF e estão sujeitas a viés de agregação. A simulação não propaga toda a incerteza de estimação das matrizes, e a dispersão é calibrada com apenas três eleições. A classificação dos candidatos de 2026 não tem equivalência perfeita com a dos grupos históricos, e os fatos da campanha do segundo turno não entram no modelo. O código, as sementes e os resultados encaminhados à revista permitem avaliar a previsão após a eleição, com a especificação fixada antes de seu resultado.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Como agenda de pesquisa, propõe-se ampliar a série com eleições anteriores, estimar modelos hierárquicos que admitam heterogeneidade territorial e incorporem a incerteza das matrizes, avaliar a calibração das probabilidades com novas eleições e comparar o método com agregadores de pesquisas. Também são possíveis aplicações às eleições para governador e a outros sistemas de dois turnos.

## DECLARAÇÃO DE USO DE INTELIGÊNCIA ARTIFICIAL

Este trabalho utilizou duas ferramentas de inteligência artificial generativa. O Claude, da Anthropic, foi utilizado na implementação do código em Python de preparação dos dados do TSE, de estimação das matrizes de transferência e das simulações, na redação e na revisão do texto, na tradução dos títulos e resumos e na formatação. O ChatGPT, da OpenAI, foi utilizado no levantamento bibliográfico, na revisão textual e metodológica e em uma implementação independente dos cálculos e dos scripts de replicação. As ferramentas também sugeriram alternativas de método e de análise, avaliadas pelos autores antes de sua adoção. A validação seguiu três procedimentos. Os totais reconstruídos por seção foram conferidos com os resultados oficiais do TSE. Os erros retrospectivos, os testes estatísticos e a checagem de 2022 foram reproduzidos exatamente por duas implementações independentes do código, e a simulação principal, com diferença restrita ao erro de Monte Carlo. Os autores conferiram os scripts, os resultados numéricos e as interpretações e assumem integral responsabilidade pelo conteúdo.

## FINANCIAMENTO

Este estudo não recebeu financiamento específico de agências de fomento públicas, privadas ou sem fins lucrativos.

## REFERÊNCIAS

BERTHOLINI, F.; RENNÓ, L.; TURGEON, M. Against all odds: forecasting Brazilian presidential elections in times of political disruption. **Revista Latinoamericana de Opinión Pública**, Salamanca, v. 11, n. 1, 2022. Disponível em: https://revistas.usal.es/cuatro/index.php/1852-9003/article/view/25882. Acesso em: 6 out. 2026.

BOLOGNESI, B.; RIBEIRO, E.; CODATO, A. Uma nova classificação ideológica dos partidos políticos brasileiros. **Dados**: Revista de Ciências Sociais, Rio de Janeiro, v. 66, n. 2, e20210164, 2023. DOI: https://doi.org/10.1590/dados.2023.66.2.303.

BUNKER, K. A two-stage model to forecast elections in new democracies. **International Journal of Forecasting**, Amsterdam, v. 36, n. 4, p. 1407-1419, 2020.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

BUNKER, K. Forecasting two-horse races in new democracies: accuracy, precision and error. **Revista Latinoamericana de Opinión Pública**, Salamanca, v. 11, n. 1, 2022. Disponível em: https://revistas.usal.es/cuatro/index.php/1852-9003/article/view/25374. Acesso em: 6 out. 2026.

COX, G. W. **Making votes count**: strategic coordination in the world's electoral systems. Cambridge: Cambridge University Press, 1997.

EFRON, B.; TIBSHIRANI, R. J. **An introduction to the bootstrap**. New York: Chapman & Hall, 1993.

FUKS, M.; RIBEIRO, E.; BORBA, J. From antipetismo to generalized antipartisanship: the impact of rejection of political parties on the 2018 vote for Bolsonaro. **Brazilian Political Science Review**, São Paulo, v. 15, n. 1, 2021.

GELMAN, A. et al. **Bayesian data analysis**. 3. ed. Boca Raton: CRC Press, 2013.

GOODMAN, L. A. Ecological regressions and behavior of individuals. **American Sociological Review**, Washington, DC, v. 18, n. 6, p. 663-664, 1953.

GRAEFE, A. et al. Combining forecasts: an application to elections. **International Journal of Forecasting**, Amsterdam, v. 30, n. 1, p. 43-54, 2014.

GREINER, D. J.; QUINN, K. M. R × C ecological inference: bounds, correlations, flexibility and transparency of assumptions. **Journal of the Royal Statistical Society**: Series A, Oxford, v. 172, n. 1, p. 67-81, 2009.

HARRIS, C. R. et al. Array programming with NumPy. **Nature**, London, v. 585, p. 357-362, 2020. DOI: https://doi.org/10.1038/s41586-020-2649-2.

HUNTER, W.; POWER, T. J. Rewarding Lula: executive power, social policy, and the Brazilian elections of 2006. **Latin American Politics and Society**, Cambridge, v. 49, n. 1, p. 1-30, 2007.

JACKMAN, S. Pooling the polls over an election campaign. **Australian Journal of Political Science**, Abingdon, v. 40, n. 4, p. 499-517, 2005. DOI: https://doi.org/10.1080/10361140500302472.

JENNINGS, W.; WLEZIEN, C. Election polling errors across time and space. **Nature Human Behaviour**, London, v. 2, p. 276-283, 2018. DOI: https://doi.org/10.1038/s41562-018-0315-6.

KING, G. **A solution to the ecological inference problem**: reconstructing individual behavior from aggregate data. Princeton: Princeton University Press, 1997.

LEWIS-BECK, M. S.; STEGMAIER, M. Economic determinants of electoral outcomes. **Annual Review of Political Science**, Palo Alto, v. 3, p. 183-219, 2000. DOI: https://doi.org/10.1146/annurev.polisci.3.1.183.

LINZER, D. A. Dynamic Bayesian forecasting of presidential elections in the states. **Journal of the American Statistical Association**, Alexandria, v. 108, n. 501, p. 124-134, 2013.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

MCKINNEY, W. Data structures for statistical computing in Python. In: PYTHON IN SCIENCE CONFERENCE, 9., 2010, Austin. **Proceedings** [...]. Austin: SciPy, 2010. p. 56-61. DOI: https://doi.org/10.25080/Majora-92bf1922-00a.

MIȘCOIU, S.; GHERGHINA, S. Candidate withdrawal in the French 2024 national legislative elections. **French Politics**, London, v. 23, n. 1, p. 40-57, 2025. DOI: https://doi.org/10.1057/s41253-024-00272-1.

PAVÍA, J. M.; ROMERO, R. Improving estimates accuracy of voter transitions: two new algorithms for ecological inference based on linear programming. **Sociological Methods & Research**, Thousand Oaks, 2022. DOI: https://doi.org/10.1177/00491241221092725.

RENNÓ, L. R.; HOEPERS, B. Voto estratégico punitivo: transferência de votos nas eleições presidenciais de 2006. **Novos Estudos CEBRAP**, São Paulo, n. 86, p. 141-161, 2010. DOI: https://doi.org/10.1590/S0101-33002010000100008.

ROMERO, R. et al. Assessing uncertainty of voter transitions estimated from aggregated data: application to the 2017 French presidential election. **Journal of Applied Statistics**, Abingdon, v. 47, n. 13-15, p. 2711-2736, 2020. DOI: https://doi.org/10.1080/02664763.2020.1804842.

ROSEN, O. et al. Bayesian and frequentist inference for ecological inference: the R×C case. **Statistica Neerlandica**, Oxford, v. 55, n. 2, p. 134-156, 2001. DOI: https://doi.org/10.1111/1467-9574.00162.

SAMUELS, D.; ZUCCO, C. **Partisans, antipartisans, and nonpartisans**: voting behavior in Brazil. Cambridge: Cambridge University Press, 2018.

STOETZER, L. F. et al. Forecasting elections in multiparty systems: a Bayesian approach combining polls and fundamentals. **Political Analysis**, Cambridge, v. 27, n. 2, p. 255-262, 2019. DOI: https://doi.org/10.1017/pan.2018.49.

TURGEON, M.; RENNÓ, L. Forecasting Brazilian presidential elections: solving the N problem. **International Journal of Forecasting**, Amsterdam, v. 28, n. 4, p. 804-812, 2012.

VIRTANEN, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. **Nature Methods**, New York, v. 17, p. 261-272, 2020. DOI: https://doi.org/10.1038/s41592-019-0686-6.

WILCOXON, F. Individual comparisons by ranking methods. **Biometrics Bulletin**, Washington, DC, v. 1, n. 6, p. 80-83, 1945.

ZUCCO, C. The president's "new" constituency: Lula and the pragmatic vote in Brazil's 2006 presidential elections. **Journal of Latin American Studies**, Cambridge, v. 40, n. 1, p. 29-49, 2008.

**Dados e documentos oficiais**

BRASIL. Tribunal Superior Eleitoral. **Resultados das eleições 2014**: votação por seção eleitoral (votacao_secao_2014_BR.zip) e detalhe da apuração por seção eleitoral (detalhe_votacao_secao_2014.zip). Brasília, DF: TSE, 2014. Conjuntos de dados. Disponível em: https://dadosabertos.tse.jus.br/dataset/resultados-2014. Acesso em: 6 out. 2026.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

BRASIL. Tribunal Superior Eleitoral. **Resultados das eleições 2018**: votação por seção eleitoral (votacao_secao_2018_BR.zip) e detalhe da apuração por seção eleitoral (detalhe_votacao_secao_2018.zip). Brasília, DF: TSE, 2018. Conjuntos de dados. Disponível em: https://dadosabertos.tse.jus.br/dataset/resultados-2018. Acesso em: 6 out. 2026.

BRASIL. Tribunal Superior Eleitoral. **Resultados das eleições 2022**: votação por seção eleitoral (votacao_secao_2022_BR.zip) e detalhe da apuração por seção eleitoral (detalhe_votacao_secao_2022.zip). Brasília, DF: TSE, 2022. Conjuntos de dados. Disponível em: https://dadosabertos.tse.jus.br/dataset/resultados-2022. Acesso em: 6 out. 2026.

BRASIL. Tribunal Superior Eleitoral. **Resultados**: eleição geral ordinária de 2026, 1º turno, presidente, arquivos de resultados por unidade da federação ({uf}-c0001-e006257-u.jws), eleição 6257. Brasília, DF: TSE, 2026. Disponível em: https://resultados.tse.jus.br/oficial/ele2026/6257/dados/. Acesso em: 6 out. 2026.

FRANCE. Ministère de l'Intérieur. **Résultats des élections législatives 2024**. Paris: Ministère de l'Intérieur, 2024. Disponível em: https://www.resultats-elections.interieur.gouv.fr. Acesso em: 6 out. 2026.

**Fontes jornalísticas**

GAZETA DO POVO. Caiado oficializa apoio a Flávio Bolsonaro no segundo turno. **Gazeta do Povo**, Curitiba, 6 out. 2026. Disponível em: https://www.gazetadopovo.com.br/eleicoes/2026/caiado-oficializa-apoio-a-flavio-bolsonaro-no-segundo-turno/. Acesso em: 11 out. 2026.

O POVO. Renan Santos declarou apoio a Lula ou a Flávio Bolsonaro? Veja. **O Povo**, Fortaleza, 9 out. 2026. Disponível em: https://www.opovo.com.br/noticias/politica/eleicoes/2026/10/09/renan-santos-declarou-apoio-a-lula-ou-a-flavio-bolsonaro-veja.html. Acesso em: 11 out. 2026.

PODER360. Augusto Cury declara apoio a Flávio Bolsonaro no 2º turno. **Poder360**, Brasília, 8 out. 2026. Disponível em: https://www.poder360.com.br/poder-eleicoes-2026/augusto-cury-declara-apoio-a-flavio-bolsonaro-no-2o-turno/. Acesso em: 11 out. 2026.



Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.