WHITEPAPER · BRASIL 2026

# Entre os turnos

Estimando como os eleitores brasileiros se moveram do primeiro para o segundo turno da eleição presidencial de 2026.

Rennan Chagas · Documento de trabalho v1.0 · 30 de outubro de 2026 · Resultados exploratórios · Versão curta (artigo)


## RESUMO

O sigilo do voto no Brasil faz com que nenhuma base de dados registre como cada eleitor mudou de ideia entre os turnos. Reconstruímos esses movimentos para a eleição presidencial de outubro de 2026 a partir dos resultados oficiais de cerca de 499 mil seções eleitorais nos dois turnos, usando uma técnica estatística chamada inferência ecológica, com pesquisas de opinião como conferência.

Estimamos que 97% dos eleitores de Lula e 95% dos de Flávio Bolsonaro no primeiro turno ficaram com eles; que os 9,3 milhões de eleitores dos candidatos eliminados se dividiram quase igualmente entre os finalistas; e que a margem de Lula, de cerca de 4,4 milhões de votos, veio principalmente de quem não votara no primeiro turno e de eleitores de branco e nulo. Testes em eleições simuladas mostram que o método é preciso quando suas hipóteses valem (erro típico de 1–3% dos eleitores), mas bem pior (cerca de 11%) quando o hábito de voto varia com a composição local, e que suas faixas de incerteza são estreitas demais. Por isso apresentamos os resultados como **exploratórios**.


**97,3% · 95,1%**
dos eleitores de Lula e de Flávio no 1º turno votaram neles de novo

**3,81 mi vs 3,82 mi**
eleitores dos eliminados que foram para Lula e para Flávio: divisão quase igual

**+4,8 milhões**
de vantagem líquida de Lula entre quem não votou e eleitores de branco/nulo do 1º turno: mais que toda a margem de 4,4 milhões

**1,3% → 11,5%**
parcela típica de eleitores colocados no lugar errado em eleições simuladas, do cenário mais favorável ao mais difícil

# 1. Introdução

As eleições presidenciais brasileiras são decididas em dois turnos. Em 2026, o primeiro turno foi em 4 de outubro e o segundo em 25 de outubro. Flávio Bolsonaro liderou o primeiro turno com 56,1 milhões de votos (47,0% dos válidos), à frente de Lula, com 53,9 milhões (45,2%). Três outros candidatos, Augusto Cury, Renan Santos e Ronaldo Caiado, receberam 2,9%, 2,2% e 2,2%. No segundo turno, Lula venceu com 64,3 milhões de votos (51,8%) contra 59,8 milhões.

Uma pergunta natural é como o eleitorado se moveu nesse intervalo. Os apoiadores dos candidatos eliminados escolheram um finalista em vez do outro? Quem se absteve no primeiro turno voltou? Quantos eleitores trocaram de lado? A resposta importa para entender o que moldou o resultado, mas os dados para respondê-la diretamente não existem: o voto é secreto, e nenhum registro oficial liga o voto de uma pessoa no primeiro turno ao do segundo.

O que existe são os resultados de cada seção eleitoral nos dois turnos. Este trabalho os usa, junto com um modelo estatístico, para estimar a tabela completa de transições: para cada grupo de eleitores do primeiro turno, que parcela acabou em cada grupo do segundo. Também dizemos com clareza até onde essas estimativas merecem confiança.

**O que este trabalho não faz.** Não identifica como qualquer indivíduo votou, não faz previsão e não pretende encerrar debates sobre por que os eleitores escolheram como escolheram. É uma reconstrução retrospectiva e exploratória.

# 2. Dados

## 2.1 Resultados oficiais por seção

O Tribunal Superior Eleitoral (TSE) publica um boletim de urna (BU) para cada urna, ou seção eleitoral. Cada uma atende, em média, cerca de 318 eleitores aptos. Para cada seção e turno, o boletim informa o número de aptos, quantos compareceram, quantos se abstiveram e os votos de cada candidato, além de brancos e nulos.

Baixamos os 56 arquivos oficiais (26 estados, o Distrito Federal e o exterior, em dois turnos) e conferimos cada um com os checksums publicados pelo TSE. Somando nossos dados processados, reproduzimos **exatamente** os totais nacionais oficiais de cada candidato, branco, nulo e abstenção, nos dois turnos (diferença zero).

## 2.2 Pareando seções entre os turnos

A análise precisa que a mesma seção seja observada nos dois turnos. No fim, pareamos 499.207 unidades que existem nos dois turnos, cobrindo todos os 158.745.079 eleitores aptos das seções que funcionaram. (Outros 423 eleitores aptos estavam em 41 seções que nunca funcionaram; ficam de fora.)

## 2.3 Nove tipos de eleitor do 1º turno, cinco escolhas no 2º

Os eleitores do primeiro turno são agrupados em: Lula, Flávio Bolsonaro, Augusto Cury, Renan Santos, Ronaldo Caiado, outros candidatos, branco, nulo e quem não votou. “Outros candidatos” reúne sete candidaturas pequenas, cada uma abaixo de 0,5% dos votos válidos (inclusive Romeu Zema, 0,33 milhão); isso evita que o modelo persiga grupos pequenos demais para estimar. No segundo turno as escolhas são Lula, Flávio, branco, nulo ou não votar.

## 2.4 Eleitores que mudaram de seção

Em 55.722 das unidades pareadas, cobrindo 18,6 milhões de eleitores aptos, o número de aptos foi diferente nos dois turnos. Uma primeira versão da nossa análise simplesmente descartava essas unidades. Uma auditoria dos registros do TSE mostrou o motivo: o eleitor pode pedir para votar em outro lugar no segundo turno (a *transferência temporária*), e o saldo dessas transferências explica a diferença em cada uma das unidades, sem resíduo.

Isso não significa que 18,6 milhões de pessoas se mudaram. O movimento líquido é de apenas cerca de 37 mil eleitores entrando e 37 mil saindo, no país. Mas um saldo líquido esconde trocas, então consideramos também um número maior e mais bem sustentado, de cerca de 65 mil entradas e saídas, que é o nosso caso-base. No modelo, cada seção ganha então uma pequena conta de “entrada técnica” e “saída técnica” para as contas fecharem. Testamos a sensibilidade a essa escolha com alternativas de até cerca de 1 milhão de mudanças (Seção 5.4). Os votos de quem se mudou são desconhecidos.

## 2.5 Pesquisas de opinião

Usamos três pesquisas publicadas como conferência, nunca como fonte principal: Quaest e Nexus/BTG (ambas publicadas em 24 de outubro) e Datafolha (relatório de 24 de outubro, mais uma divulgação anterior com um parágrafo sobre os eleitores de quatro candidatos eliminados). As três estão registradas no TSE (BR-09158/2026, BR-08234/2026 e BR-09540/2026). Em geral, perguntam em quem a pessoa votou no primeiro turno e em quem pretende votar no segundo. O relatório final do Datafolha não traz essa tabela, então usamos só o parágrafo anterior. Nenhuma das pesquisas publica detalhe suficiente para calcular tamanhos efetivos de amostra por subgrupo.

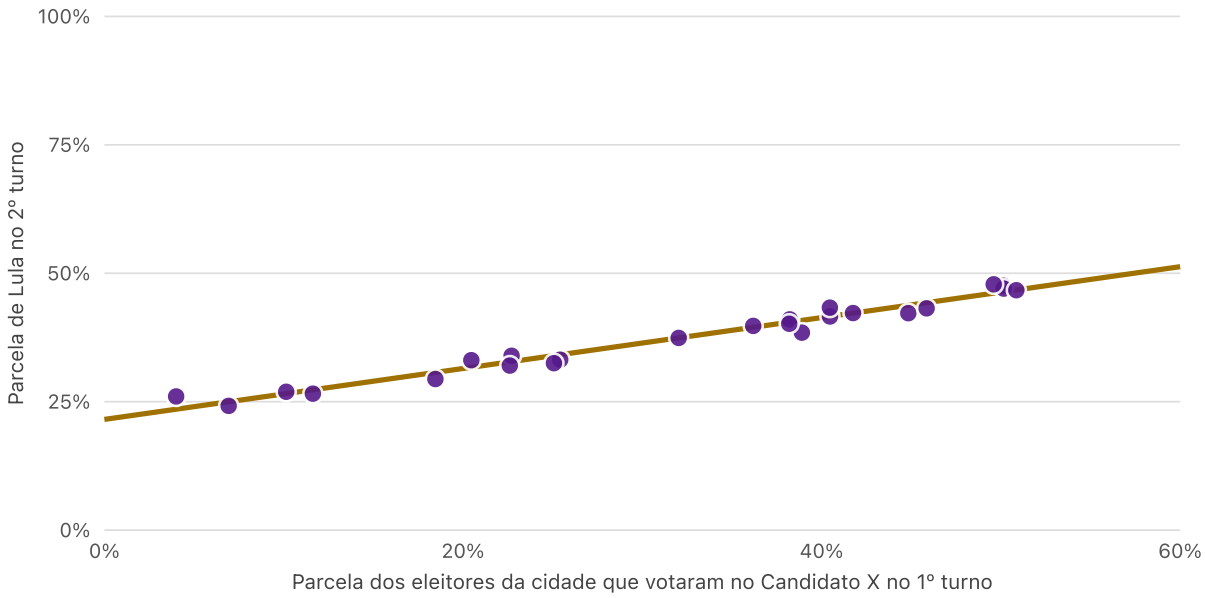
# 3. Método, em linguagem simples

## 3.1 A ideia central

Suponha que saibamos, para muitas cidades, a parcela de eleitores que votou no Candidato X no primeiro turno e a parcela que votou em Lula no segundo. Se as cidades com mais eleitores de X deram mais votos a Lula no segundo turno, isso sugere que os eleitores de X foram para Lula. Uma reta pelas cidades revela a taxa: onde a reta encontra a borda direita (uma cidade só de eleitores de X) está a estimativa de como os eleitores de X se comportaram. Esse tipo de raciocínio se chama *inferência ecológica*: inferir o comportamento de indivíduos a partir de dados de grupos.

**Figura 1 · A ideia central, e onde ela quebra**

Cada ponto é uma cidade imaginária; a reta é o melhor resumo em linha reta.




| Parcela real dos eleitores de X que escolheram Lula | O que a reta diz | Erro |
|---|---|---|
| **70%** | **71%** | **1 p.p.** |

Dados inventados. Neste mundo, 70% dos eleitores do Candidato X realmente escolheram Lula, e 22% de todos os outros também.

A fraqueza é a que o controle revela. Se os eleitores de X se comportam diferente onde são muitos do que onde são poucos, um "efeito de vizinhança", a reta deixa de refletir o que qualquer eleitor de X fez. Estatísticos chamam isso de *viés ecológico*. Nada nos dados das urnas pode descartá-lo por completo, e é por isso que a análise exige um programa cuidadoso de testes (Seção 5).

## 3.2 O modelo

Adaptamos o modelo Brown–Payne modificado (Brown e Payne, 1986; Forcina, Gnaldi e Bracalente, 2012), usando a aproximação normal na forma apresentada por Forcina e Pavía (2026, Seção 2.1, Equação 1). Em palavras:

- Dividimos o país em **183 áreas** de seções (cerca de 2.500 unidades pareadas cada, dentro de um mesmo estado, formadas por municípios e zonas eleitorais consecutivos; não são vizinhanças geográficas exatas). Dentro de uma área supomos que um dado tipo de eleitor do primeiro turno tem a *mesma chance* de acabar em cada escolha do segundo turno, em todas as seções.
- Dadas essas chances, o modelo prevê quantos votos cada escolha deve ter em cada seção (a soma, sobre os grupos do primeiro turno, do tamanho do grupo vezes a chance). Em seguida escolhe as chances que fazem as previsões se aproximarem ao máximo do que foi de fato contado, em todas as seções da área.

- Como as seções reais são mais variáveis do que um sorteio de moeda seria, o modelo admite aleatoriedade extra, estimada a partir dos dados (*sobredispersão*).
- Áreas diferentes recebem chances diferentes, de modo que o modelo capta, por exemplo, que eleitores do Nordeste se comportam diferente dos do Sul.

## 3.3 Fazendo as contas fecharem

As chances ajustadas são aproximadas. Para transformá-las numa tabela de contagens de eleitores, usamos o *ajuste proporcional iterativo*, uma técnica clássica que ajusta a tabela até que as linhas somem exatamente o resultado do primeiro turno e as colunas o do segundo turno de cada seção. Somamos então as seções para obter tabelas do país, de uma região ou de um estado. As tabelas são contagens esperadas, não uma lista de indivíduos.

Forcina e Pavía (2026) comparam formas de fazer esta etapa e concluem que um método de mínimo qui-quadrado é um pouco melhor que o ajuste proporcional iterativo quando ambos são aplicados aos totais de uma área. Mantemos o ajuste proporcional iterativo porque ele pode ser aplicado seção a seção, respeitando os totais de cada seção; o mínimo qui-quadrado deles trabalha com os totais da área. Na base estrita, trocar um método pelo outro no nível da área move cerca de 0,1% dos eleitores; reconciliar seção a seção, em vez de por área, move cerca de 0,5%. Em nossas simulações, os dois métodos diferem em menos de 0,02 ponto percentual dos eleitores em todos os cenários, sem vencedor consistente.

## 3.4 Quem mudou de seção

Para os eleitores que entraram ou saíram de uma seção, o modelo acrescenta uma conta técnica de "entrada" e "saída" (Seção 2.4). As saídas são supostas um sorteio ao acaso entre as pessoas da seção; os que entram têm um padrão próprio de escolhas, estimado. A passagem direta de entrada para saída é descartada. Essas contas reúnem cerca de 0,04% dos eleitores no caso-base.

## 3.5 Quantificando a incerteza

Apresentamos dois tipos de faixa para cada célula da tabela.

- **Limites rígidos.** Dados os totais de uma seção, não podem ir mais eleitores do grupo A para o grupo B do que o menor dos dois grupos, e um mínimo pode ser forçado. Somar esses limites em todas as seções dá uma faixa sempre verdadeira, mas em geral larga.
- **Faixa de 95% por reamostragem.** Repetimos a análise em 1.000 reamostras das áreas (técnica chamada *bootstrap*) e ficamos com os 95% centrais dos resultados. Isso capta o quanto os resultados dependem de quais áreas estão nos dados, mas não a incerteza sobre as hipóteses do modelo. Mostramos na Seção 5 que essas faixas são estreitas demais.

## 3.6 Misturando pesquisas (análise de sensibilidade)

Como teste de estresse, e no espírito de Klima et al. (2017) e Greiner e Quinn (2010), que combinam dados agregados com pesquisas, também ajustamos versões que misturam a informação das urnas com as pesquisas. Uma ("confiando nas pesquisas") trata as respostas das pesquisas como informativas, com uma força que variamos entre três valores plausíveis. A outra ("admitindo viés

das pesquisas") dá a cada instituto seu próprio viés e admite respostas esporadicamente muito fora do esperado. São aproximações nacionais, não reanálises completas; servem só para comparação (Figura 8).

### Para os curiosos: o modelo em poucas linhas

Na seção $s$, seja $x_{si}$ o número de eleitores do grupo $i$ no primeiro turno e $y_{sj}$ o número que escolheu $j$ no segundo. Com $p_{ij}$ a chance de o grupo $i$ acabar em $j$, o modelo usa uma aproximação normal com

$$\text{média: } \mu_s = \Sigma_i \, x_{si} \, p_i$$
$$\text{covariância: } V_s = \varphi \cdot \Sigma_i \, x_{si} \, [\, \mathrm{diag}(p_i) - p_i p_i^{\top} \,], \; \varphi \geq 1$$

e maximiza a verossimilhança resultante sobre os $p_i$ em cada área (otimizador L-BFGS-B, gradientes analíticos, φ concentrado). A extensão para mudanças de cadastro troca a média por $f \cdot x_s P + e \cdot q$, em que $f$ é a parcela dos eleitores da seção que permanece, $e$ o número de entrantes e $q$ suas probabilidades de destino, e acrescenta termos de variância hipergeométrica para quem sai. A etapa de reconciliação é o ajuste proporcional iterativo em cada seção com zeros estruturais (entrada → saída é impossível). A especificação completa está no repositório (`docs/population/model_specification.md`).

É uma adaptação, não uma réplica dos métodos publicados: usa sobredispersão comum por área, sem covariáveis, logits limitados e agrupa candidatos muito pequenos.

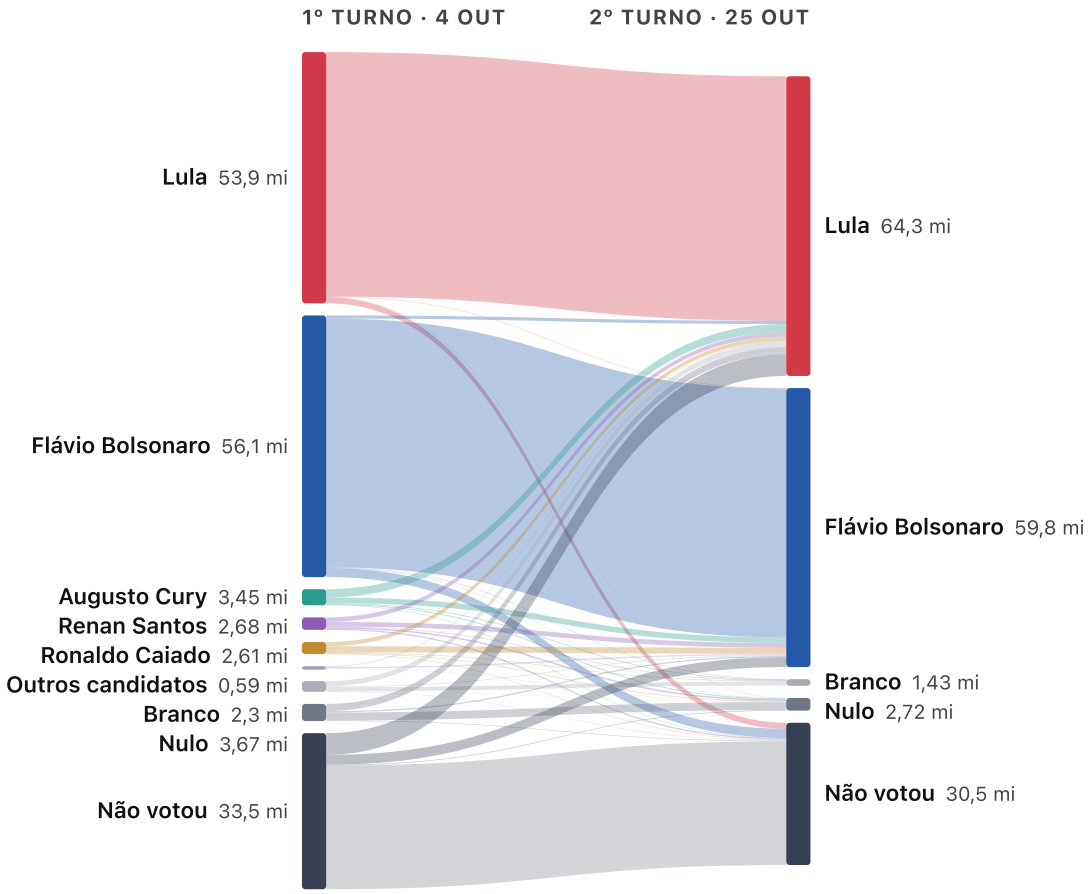
# 4. Resultados

Todas as figuras abaixo são interativas. O filtro *Onde* muda a geografia de todos os gráficos de uma vez. O filtro *Hipótese sobre mudanças de cadastro* alterna entre seis hipóteses sobre mudanças de cadastro e tamanho das áreas do modelo; elas mudam muito pouco (Seção 5.4).

## 4.1 O panorama

**Figura 2 · Onde cada grupo de eleitores foi parar**

Esquerda: escolha no 1º turno (4 out). Direita: escolha no 2º turno (25 out). Largura da faixa = número estimado de eleitores.




Brasil (país inteiro): 158.745.079 eleitores aptos; o modelo foi ajustado em 183 áreas (1 sem convergência plena).

Passe o mouse sobre uma barra para seguir um grupo; sobre uma faixa para ver tamanho e incerteza. “Não votou” inclui eleitores aptos que ficaram em casa. Cerca de 65 mil eleitores cujo cadastro mudou de seção não aparecem no desenho (0,04%).

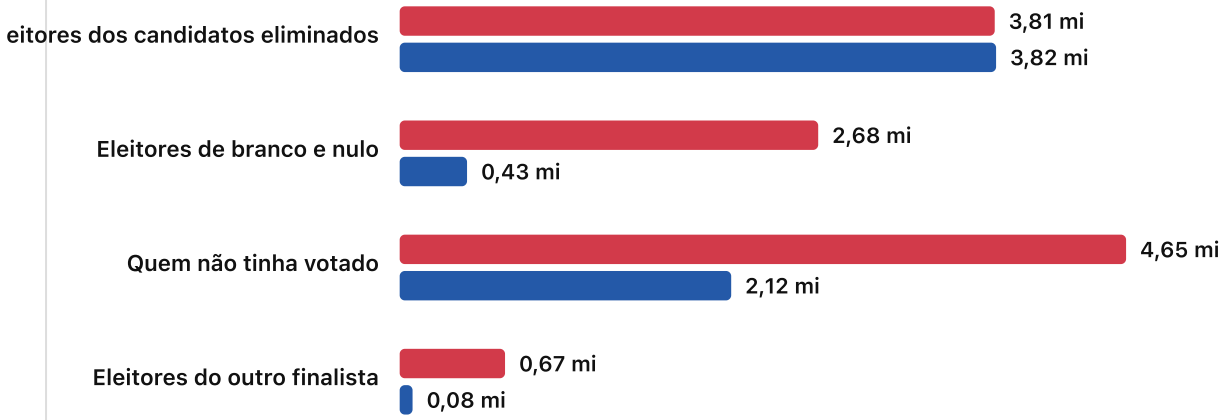
A maioria dos eleitores não mudou. Estimamos que 97,3% dos eleitores de Lula no primeiro turno e 95,1% dos de Flávio votaram neles de novo. O principal “vazamento” de cada finalista foi o não comparecimento (2,4% para Lula, 3,5% para Flávio). A troca direta foi pequena e num só sentido: cerca de 0,67 milhão de eleitores de Flávio foram para Lula (1,2%), contra cerca de 82 mil no caminho inverso (0,15%).

## 4.2 O que decidiu o segundo turno

Lula ganhou 10,4 milhões de votos entre os turnos e Flávio 3,7 milhões. A figura abaixo mostra de onde vieram os votos novos.

**Figura 3 · De onde vieram os votos de cada finalista no segundo turno**

Por tipo de eleitor do 1º turno. Troque para "Vantagem líquida" para ver como se forma a margem final.




Brasil (país inteiro): Lula manteve 97,3% dos seus eleitores do 1º turno; Flávio manteve 95,1%.

As quatro barras líquidas somam a margem final entre os finalistas.

- **Eleitores dos candidatos eliminados** (9,3 milhões). Dividiram-se quase igualmente: 3,81 milhões para Lula, 3,82 milhões para Flávio. Outros 18% não votaram ou votaram em branco ou nulo.
- **Eleitores de branco e nulo** (6,0 milhões). 2,68 milhões escolheram Lula e 0,43 milhão escolheram Flávio.
- **Quem não votou no 1º turno** (33,5 milhões). 79% se abstiveram de novo. Entre os que voltaram, 4,65 milhões escolheram Lula e 2,12 milhões Flávio, uma divisão de cerca de 69% para 31%.

Em termos líquidos, a margem de Lula, de cerca de 4,4 milhões, é a soma de 2,5 milhões vindos de quem não votou, 2,2 milhões de eleitores de branco e nulo, praticamente zero dos eleitores dos eliminados e 0,3 milhão *contra* ele vindos dos eleitores dos próprios finalistas no primeiro turno (Flávio partiu de uma base maior). A "divisão igual" dos eleitores dos eliminados é onde nos sentimos menos à vontade, já que as pesquisas discordam (Seção 5.3).

## 4.3 A tabela completa

**Figura 4 · Matriz de transição**

Cada linha é um grupo do 1º turno; cada coluna, uma escolha no 2º turno.

| | Lula | Flávio Bolsonaro | Branco | Nulo | Não votou | Total |
|---|---|---|---|---|---|---|
| Lula | 97,3%<br>52.417.073 | 0,15%<br>82.468 | 0,011%<br>5.695 | 0,12%<br>64.903 | 2,4%<br>1.285.687 | 53.879.538 |
| Flávio Bolsonaro | 1,2%<br>674.805 | 95,1%<br>53.338.561 | 0,14%<br>77.424 | 0,097%<br>54.533 | 3,5%<br>1.935.879 | 56.104.503 |
| Augusto Cury | 51,6%<br>1.778.107 | 34,8%<br>1.200.720 | 5,0%<br>170.747 | 5,8%<br>199.696 | 2,8%<br>97.882 | 3.448.569 |
| Renan Santos | 35,8%<br>957.595 | 37,9%<br>1.012.524 | 4,5%<br>119.938 | 12,0%<br>320.504 | 9,9%<br>263.879 | 2.675.887 |
| Ronaldo Caiado | 35,2%<br>916.811 | 50,3%<br>1.310.017 | 5,0%<br>129.627 | 5,4%<br>139.449 | 4,2%<br>108.138 | 2.605.148 |
| Outros candidatos | 27,1%<br>158.787 | 50,6%<br>296.729 | 4,2%<br>24.873 | 11,5%<br>67.227 | 6,7%<br>39.149 | 587.143 |
| Branco | 49,5%<br>1.138.219 | 7,3%<br>168.829 | 36,6%<br>841.053 | 2,1%<br>47.792 | 4,5%<br>104.212 | 2.300.798 |
| Nulo | 42,0%<br>1.542.399 | 7,2%<br>262.642 | 0,84%<br>30.683 | 46,3%<br>1.702.413 | 3,7%<br>135.039 | 3.674.249 |
| Não votou | 13,9%<br>4.652.654 | 6,3%<br>2.123.311 | 0,085%<br>28.526 | 0,37%<br>124.046 | 79,3%<br>26.528.768 | 33.469.244 |
| Total | 64.256.232 | 59.826.041 | 1.430.496 | 2.723.305 | 30.509.005 | 158.745.07 |

Brasil (país inteiro): 158.745.079 eleitores aptos; o modelo foi ajustado em 183 áreas (1 sem convergência plena). As linhas somam 100% (ignorando cerca de 0,04% de eleitores que mudaram de cadastro). Passe o mouse ou toque numa célula para ver a faixa de 95%.

A cor da célula é proporcional à parcela da linha. Passe o mouse ou toque numa célula para ver a faixa de 95% e os limites rígidos.

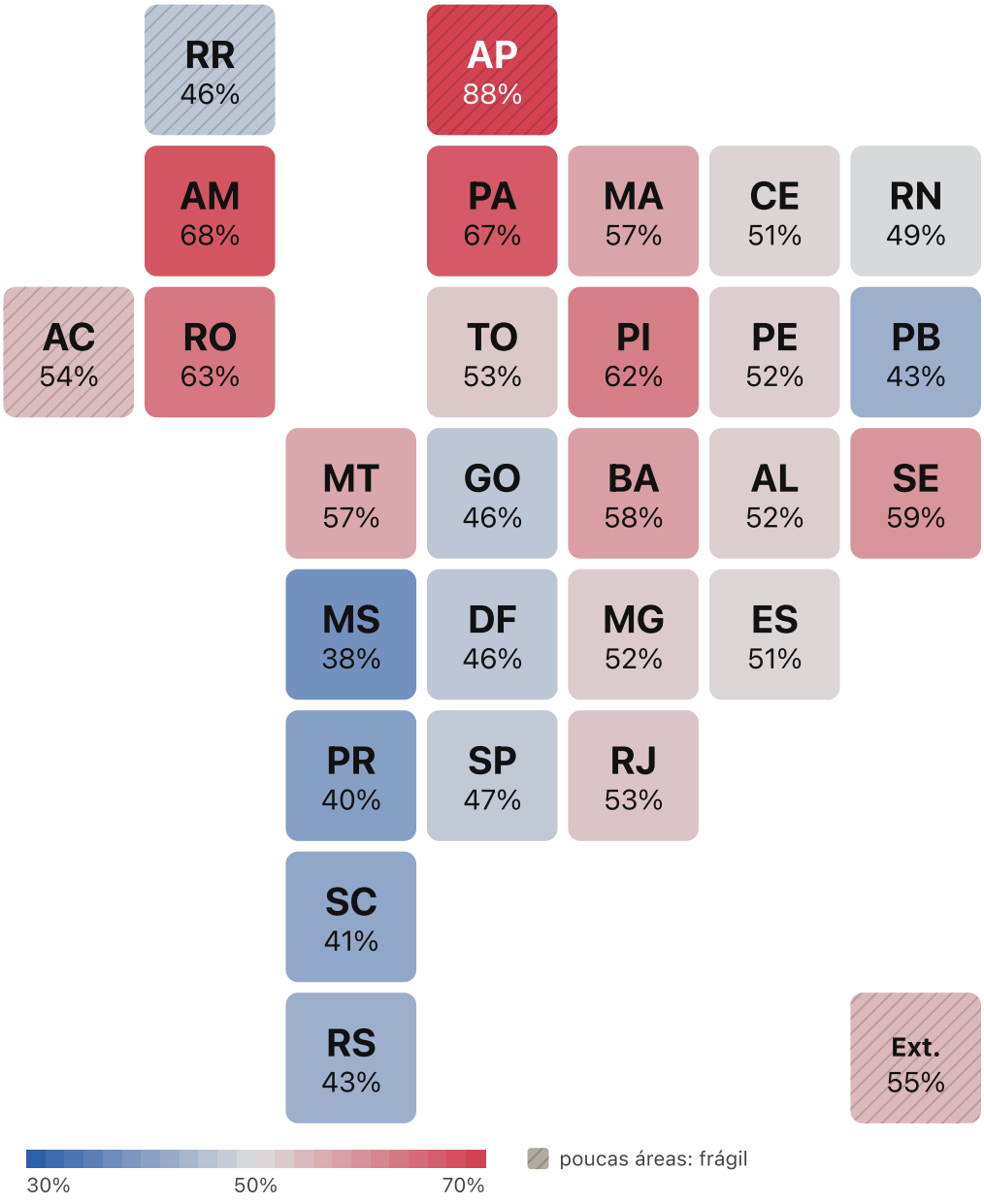
As diferenças entre os candidatos eliminados são visíveis. Os eleitores de Cury pendem para Lula (52% contra 35% para Flávio). Os de Caiado pendem para Flávio (50% contra 35%). Os de Renan Santos se dividem (36% Lula, 38% Flávio), com um 12% incomumente alto de nulos. Os eleitores

de branco foram uma mistura: 49% foram para Lula e 37% votaram em branco de novo; entre os de nulo, 42% foram para Lula e 46% anularam de novo.

## 4.4 Geografia

**Figura 5 · Geografia das transições**

Mapa em blocos dos 26 estados, do Distrito Federal e do voto no exterior. Clique num bloco para selecioná-lo em todos os gráficos.




Vermelho: eleitores dos eliminados inclinados a Lula; azul: inclinados a Flávio. Percentual entre os que escolheram um dos finalistas.

Blocos hachurados dependem de pouquíssimas áreas do modelo e devem ser vistos como frágeis.

Os contrastes regionais são claros. Entre os eleitores dos eliminados que escolheram um finalista, 54% escolheram Lula no Nordeste, 65% no Norte e 41% no Sul. O retorno de quem não votara no

primeiro turno variou de 17% (Norte) a 26% (Nordeste). Os resultados por estado são mais ruidosos que os nacionais; Acre, Amapá, Roraima e o exterior se baseiam em uma única área do modelo cada.

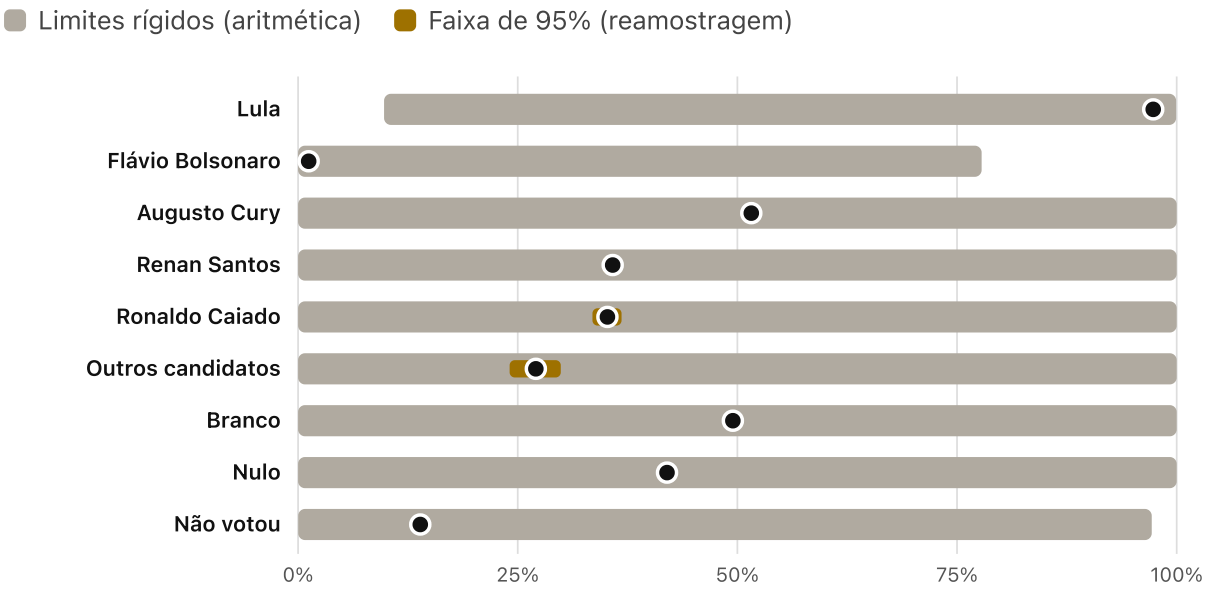
# 5. Quão confiáveis são essas estimativas?

## 5.1 As faixas

A Figura 6 compara a melhor estimativa com os limites rígidos e com a faixa de 95% por reamostragem. Os limites rígidos, que usam só aritmética, são largos para a maioria das células: os totais oficiais sozinhos dizem pouco sobre fluxos individuais. As faixas de reamostragem são bem mais estreitas, porque o modelo supõe uma estrutura; se a estrutura está certa é o assunto das duas próximas seções.

**Figura 6 · Estimativas, faixas de 95% e limites rígidos**

Parcela de cada grupo do 1º turno que foi para o destino escolhido.




Brasil (país inteiro): 158.745.079 eleitores aptos; o modelo foi ajustado em 183 áreas (1 sem convergência plena).

Cinza: limites rígidos só com os totais oficiais. Dourado: faixa de 95% reamostrando as áreas (1.000 sorteios). Ponto: melhor estimativa. As faixas existem apenas no caso-base.

## 5.2 Testes em eleições simuladas

Geramos 1.000 eleições artificiais usando tamanhos reais de seções de dois lugares (Acre e o exterior), cinco grupos de eleitores inventados e seis tipos de mundo:

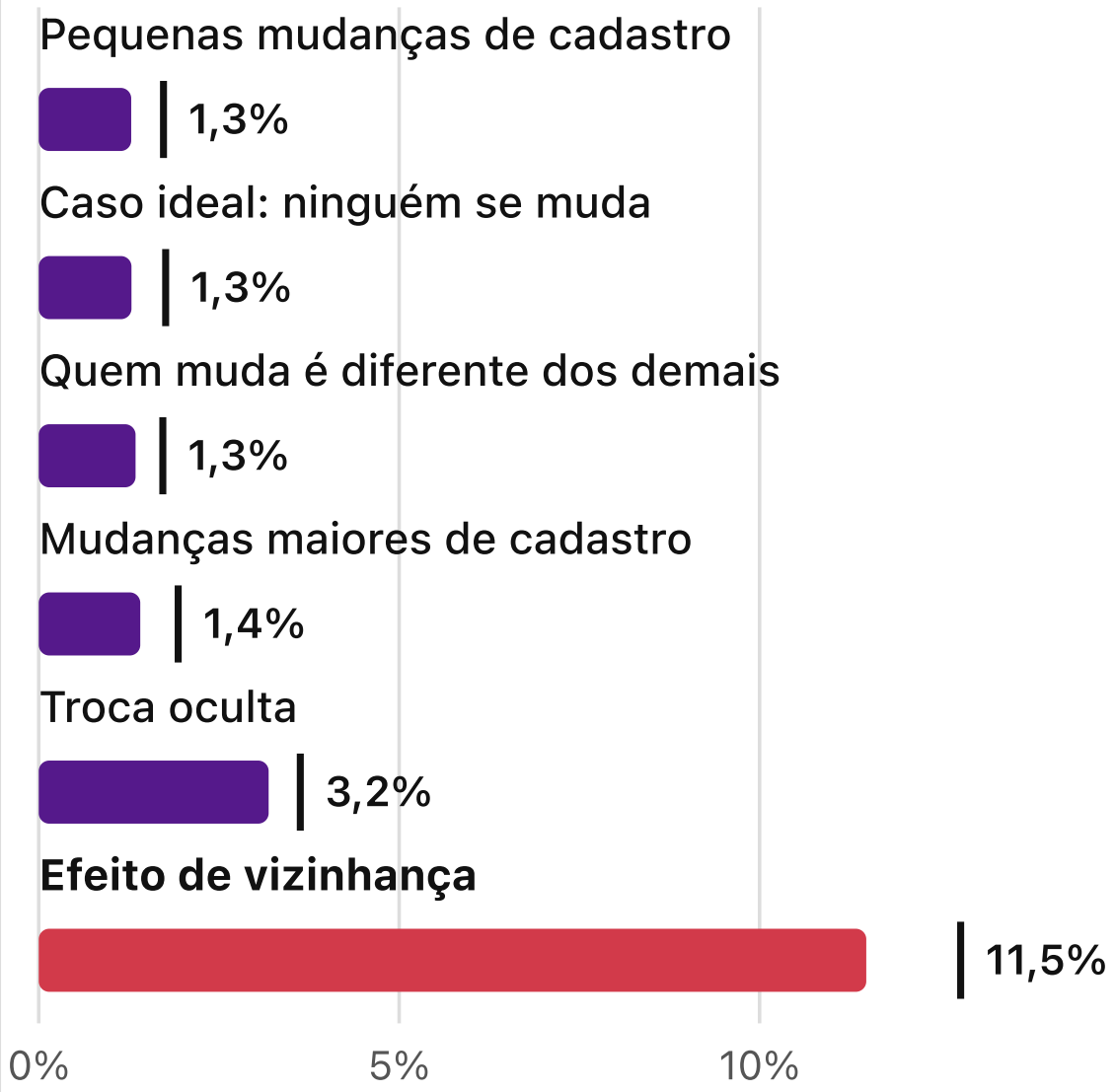
- **Caso ideal:** ninguém muda de seção e todos se comportam como o modelo supõe.
- **Mudanças de cadastro (pequenas, maiores, seletivas):** alguns eleitores entram ou saem das seções, ao acaso ou não.

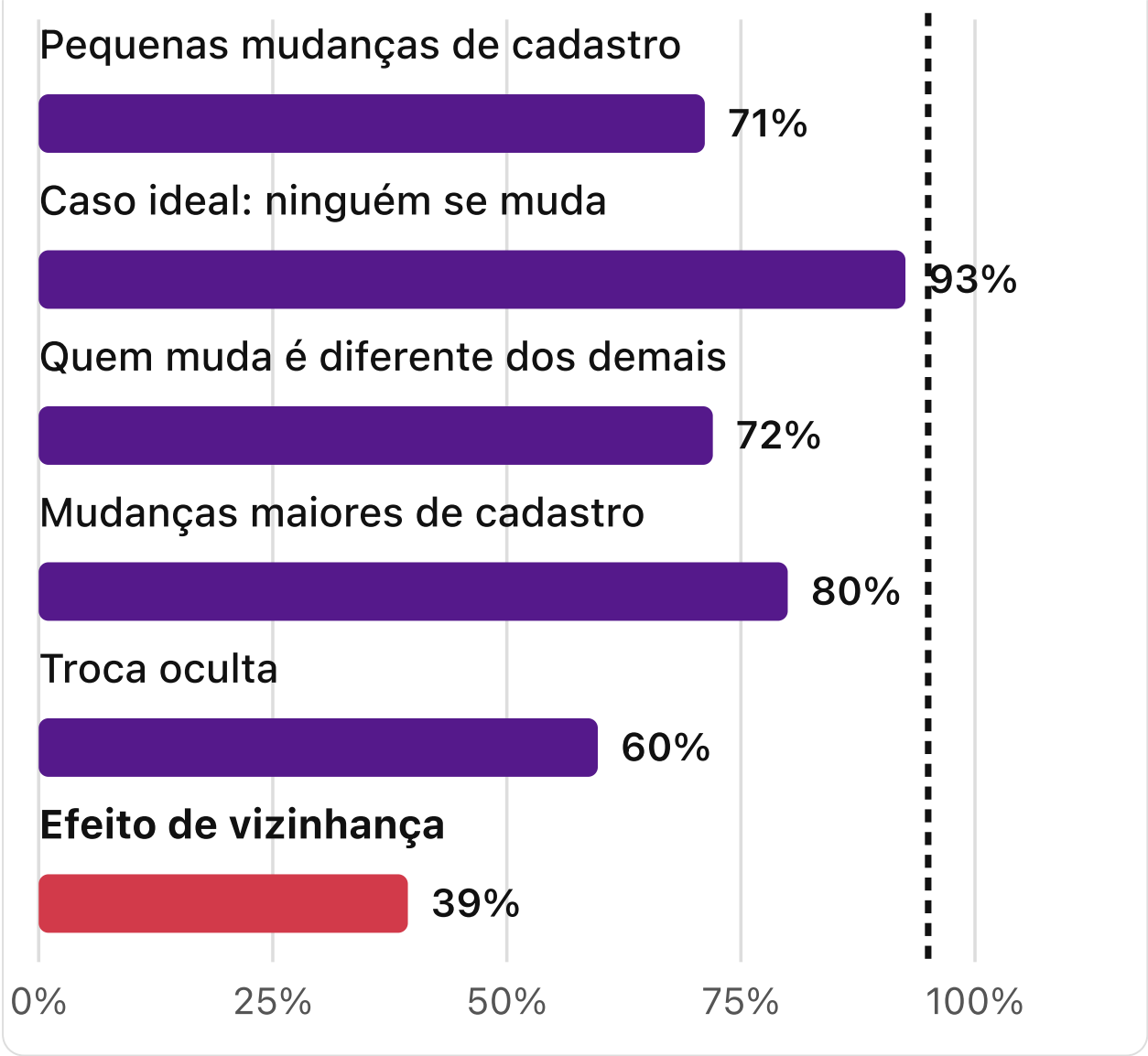
- **Troca oculta:** chegam tantos quanto saem, então os totais não revelam.
- **Efeito de vizinhança:** o comportamento depende de quem está ao redor, o caso do exemplo ilustrativo.

Em cada eleição artificial sabíamos a verdade, rodamos o método e medimos o erro como a parcela de eleitores atribuída a uma célula errada da tabela. Em dez eleições artificiais por cenário (60 no total) conferimos também as faixas de 95%, com 80 reamostras cada; são poucas, então as taxas de acerto abaixo são aproximadas.

**Figura 7 · Acurácia em eleições simuladas**

Erro médio e frequência com que a faixa de 95% continha a verdade.






A barra vermelha marca o cenário em que o hábito de voto varia com a composição local. Os traços pretos no painel esquerdo marcam o percentil 95 do erro.

| Cenário | Erro típico | Erro em dia ruim (5% piores) | Erro típico se as seções com cadastro alterado fossem descartadas | Parcela das "faixas de 95%" que acertaram | Ajustes que convergiram plenamente |
|---|---|---|---|---|---|
| Pequenas mudanças de cadastro | 1,3% | 1,7% | 1,5% | 71% | 63% |
| Caso ideal: ninguém se muda | 1,3% | 1,8% | 1,3% | 93% | 30% |
| Quem muda é diferente dos | 1,3% | 1,7% | 1,6% | 72% | 63% |

| Cenário | Erro típico | Erro em dia ruim (5% piores) | Erro típico se as seções com cadastro alterado fossem descartadas | Parcela das "faixas de 95%" que acertaram | Ajustes que convergiram plenamente |
|---|---|---|---|---|---|
| demais | | | | | |
| Mudanças maiores de cadastro | 1,4% | 1,9% | 6,7% | 80% | 95% |
| Troca oculta | 3,2% | 3,6% | 1,9% | 60% | 56% |
| Efeito de vizinhança | 11,5% | 12,8% | 11,5% | 39% | 36% |

"Erro típico" é a média sobre as eleições simuladas; "dia ruim" é o percentil 95. A quarta coluna é o erro da nossa primeira versão, que descartava as seções com cadastro alterado; mostra por que a extensão importa quando os cadastros mudam muito (mudanças maiores: 6,7% contra 1,4%). No cenário de troca oculta a versão antiga parece melhor (1,9% contra 3,2%), mas só porque avalia apenas as seções que manteve. "Ajustes que convergiram plenamente" é a parcela de ajustes que atendeu a um critério numérico estrito; um experimento separado mostrou que mais esforço do otimizador melhorou a convergência mas não eliminou o viés.

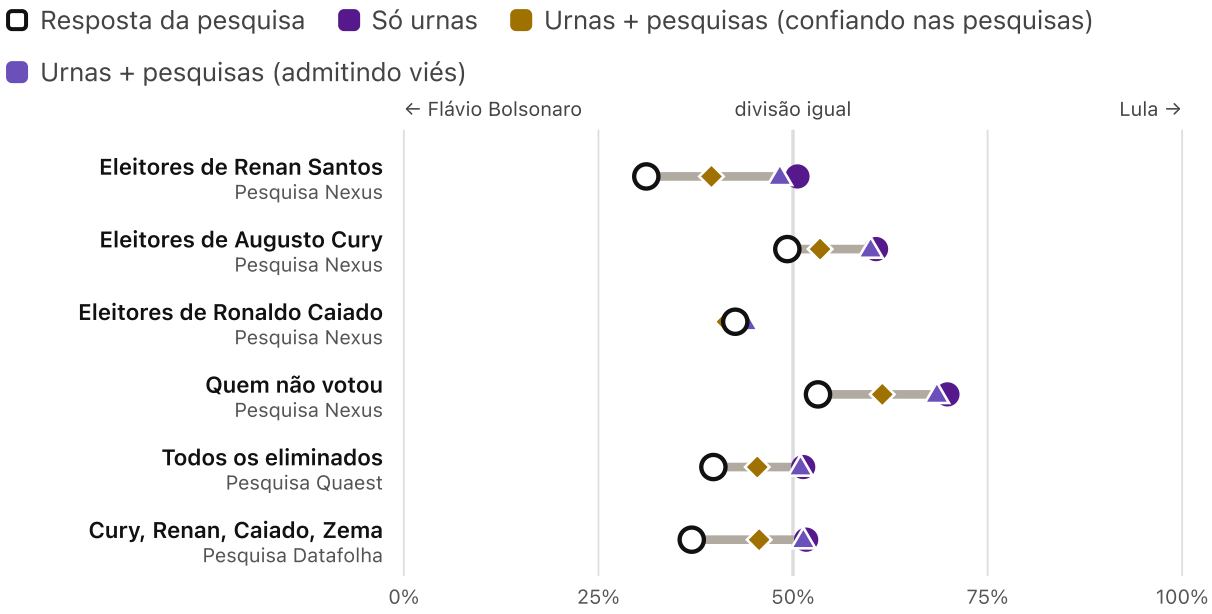
> **O QUE OS TESTES DIZEM**
>
> Quando as hipóteses do modelo valem, ele recupera bem a verdade (cerca de 1,3% dos eleitores no lugar errado) e as faixas de 95% ficam perto do correto (93%). Com troca oculta, o erro sobe para cerca de 3%, e com efeito de vizinhança para cerca de 11%, com faixas de 95% que só acertam cerca de 39% das vezes. Nossas próprias metas de acurácia e cobertura, definidas antes, não foram plenamente atingidas, e mantivemos as falhas visíveis.

## 5.3 Pesquisas

A figura abaixo coloca as estimativas pelas urnas ao lado do que responderam os entrevistados. As estimativas pelas urnas concordam bem com a Nexus para os eleitores de Caiado, ficam a poucos pontos para os de Cury e divergem para os de Renan Santos, para quem não votou e para o conjunto dos eliminados na Quaest e no Datafolha. Parte da diferença é esperada: as pesquisas medem intenção antes do voto, com amostras pequenas, erro de memória e não resposta; e as respostas das pesquisas dizem respeito à parcela que escolheu Lula *entre quem escolheu um finalista*, o que não se compara diretamente com uma parcela bruta. Mas a distância também nos diz que a estimativa pelas urnas para esses grupos não está acima de dúvida.

**Figura 8 · Urnas versus pesquisas**

Entre quem escolheu Lula ou Flávio, a parcela que escolheu Lula, por grupo do 1º turno.




As pesquisas medem intenção de voto pouco antes da eleição, com amostras pequenas; o método das urnas mede o resultado real com hipóteses de modelo. Nenhum dos dois é o gabarito.

Misturar pesquisas ao modelo puxa as estimativas na direção das pesquisas quando se confia nelas, e as deixa perto da resposta das urnas quando se admite viés das pesquisas. A mistura que confia nas pesquisas leva a parcela de Lula entre os eleitores dos eliminados a cerca de 45%; a que admite viés, a cerca de 51%.

## 5.4 Sensibilidade às hipóteses

**Figura 9 · Os números principais dependem das nossas hipóteses?**

Estimativas para o Brasil sob as seis hipóteses alternativas sobre mudanças de cadastro e tamanho das áreas do modelo.

| | Caso-base | Base, áreas maiores | Menos mudanças | Menos, áreas maiores | Mais mudanças | Muitas mudanças |
|---|---|---|---|---|---|---|
| Eleitores de Lula que ficaram com Lula | 97,3% | 97,4% | 97,3% | 97,4% | 97,7% | 98,0% |
| Eleitores de Flávio que ficaram com Flávio | 95,1% | 95,3% | 95,1% | 95,3% | 95,6% | 96,0% |
| Eleitores dos eliminados que escolheram Lula* | 49,9% | 50,4% | 49,9% | 50,4% | 50,5% | 51,2% |
| Eleitores de branco/nulo que escolheram Lula* | 86,1% | 86,1% | 86,2% | 86,2% | 85,8% | 85,5% |
| Quem não votou no 1º turno e votou no 2º | 20,7% | 20,6% | 20,7% | 20,6% | 19,9% | 19,2% |
| Dos que voltaram a votar, quem escolheu Lula* | 68,7% | 69,2% | 68,7% | 69,2% | 69,2% | 69,7% |

*Entre quem escolheu Lula ou Flávio. Os números quase não se mexem: essas hipóteses não são o que limita a análise.

As estimativas principais quase não se mexem entre as seis hipóteses (mudanças de cadastro de cerca de 37 mil a cerca de 1 milhão; áreas do modelo de 2.500 ou 5.000 unidades). Isso nos diz que a análise não é frágil a *essas* escolhas, mas nada diz sobre a incógnita maior, o efeito de vizinhança,

que não pode ser testado só com os dados reais.

**Áreas menores mudam a resposta.** Forcina e Pavía (2026) ajustam o modelo em áreas muito menores (cerca de 50 a 90 seções), argumentando que ali os eleitores são mais parecidos. Reajustamos a base estrita com áreas de cerca de 500 e 250 unidades (769 e 1.370 áreas). Entre os eleitores de candidatos eliminados que escolheram Lula ou Flávio, a parcela de Lula cai de 50,6%, com áreas de 2.500 unidades, para 48,9% e 48,6%; cerca de 0,7% e 1,0% dos eleitores mudam de célula. O deslocamento tem a mesma direção nas cinco regiões e no exterior (maior no Norte, cerca de 6 pontos, e menor no Sul); o Rio de Janeiro é o principal estado que se move no sentido oposto. Ele se mantém após uma segunda passagem de otimização (depois da qual 41 e 83 ajustes de áreas pequenas, cerca de 6% dos eleitores, ainda não atingem nosso critério de convergência). Áreas menores tornam a hipótese de chances iguais mais plausível, mas dão menos informação a cada ajuste, e os dados reais não dizem qual efeito predomina. Por isso lemos o resultado principal como incerto em pelo menos cerca de 2 pontos nessa direção, além das faixas da Seção 3.5.

## 5.5 Saúde numérica

No caso-base, 182 dos 183 ajustes de área atenderam aos nossos critérios numéricos de convergência; um não atendeu e está sinalizado. Nas versões com pesquisas, as cinco variantes convergiram sem transições divergentes (R-hat máximo 1.006).

# 6. Limitações

1. **Sem dados individuais.** Cada número aqui é uma estimativa para um grupo. O sigilo do voto impede a verificação contra indivíduos.
2. **O efeito de vizinhança.** O método supõe a mesma chance de troca dentro de uma área. Se isso estiver errado de formas que acompanham a composição local, as estimativas podem errar em torno de 11% dos eleitores, como na nossa pior simulação.
3. **Faixas estreitas demais.** Nossas faixas de 95% continham a verdade entre 39% e 93% das vezes nos testes. Leia-as como incerteza mínima.
4. **Mudanças de seção.** Sabemos quantos eleitores mudaram de seção (ao menos 37 mil, provavelmente cerca de 65 mil), não como votaram. Hipóteses alternativas quase não alteram os resultados, mas continuam sendo hipóteses.
5. **Lugares pequenos.** Acre, Amapá, Roraima e o exterior dependem de uma única área cada; seus resultados e faixas não são confiáveis.
6. **Candidatos pequenos.** Candidatos abaixo de 0,5% dos votos válidos são agrupados; nada dizemos sobre eles individualmente.
7. **Pesquisas.** Pesquisas medem intenção e recordação. Usamo-las só como conferência, não como gabarito, e os modelos com pesquisas são aproximações nacionais.
8. **Não é uma réplica.** O método adapta trabalhos publicados; não é uma réplica exata. O artigo completo de Forcina e Pavía (2026), fornecido como PDF local, e seu suplemento oficial foram consultados, e a primeira área do suplemento foi reproduzida. As áreas do modelo deles são muito menores que as nossas (cerca de 50 a 90 seções); a Seção 5.4 mostra como nossos resultados mudam com áreas menores. Forcina, Gnaldi e Bracalente (2012) é citado pela

descrição deles. Para Klima et al. (2017), o acesso continua limitado aos metadados e ao software publicados.

9. **Sem revisão por pares.** Este é um documento de trabalho. Algumas das nossas metas de acurácia e cobertura, fixadas antes, não foram atingidas.

# 7. Glossário

**Seção eleitoral (urna)**
A menor unidade para a qual há resultados publicados; cerca de 300 eleitores.

**Voto em branco / nulo**
Votar em ninguém (branco) ou uma cédula inválida (nulo). Nenhum conta como voto em candidato.

**Abstenção / "não votou"**
Eleitor apto que não compareceu.

**Inferência ecológica**
Inferir como indivíduos se comportam a partir de dados de grupos.

**Viés ecológico / efeito de vizinhança**
Erro que surge quando indivíduos se comportam diferente conforme a composição do seu grupo.

**Bootstrap / "faixa de 95% por reamostragem"**
Repetir a análise em muitas versões reamostradas dos dados para ver quanto os resultados se mexem.

**Limites rígidos**
O intervalo de valores de uma célula da tabela que é compatível com os totais oficiais, qualquer que seja o comportamento.

**Transferência temporária**
Pedido do eleitor para votar em outro local em determinado turno.

**Ajuste proporcional iterativo**
Técnica que ajusta uma tabela até que suas linhas e colunas somem totais dados.

# 8. Referências e fontes


1. Brown, P. J., & Payne, C. D. (1986). Aggregate data, ecological regression, and voting transitions. *Journal of the American Statistical Association*, 81(394), 452–460.

2. Forcina, A., & Pavía, J. M. (2026). Estimating Vote Transition Counts in Ecological Inference. *Statistics and Public Policy, 13*(1), Article 2654396. https://doi.org/10.1080/2330443X.2026.2654396.

3. Forcina, A., Gnaldi, M., & Bracalente, B. (2012). A revised Brown and Payne model of voting behaviour applied to the 2009 elections in Italy. *Statistical Methods & Applications,*

*21*(1), 109–119. doi:10.1007/s10260-011-0184-x

4. Klima, A., Schlesinger, T., Thurner, P. W., & Küchenhoff, H. (2017). Combining aggregate data and exit polls for the estimation of voter transitions. *Sociological Methods & Research*. doi:10.1177/0049124117701477

5. Greiner, D. J., & Quinn, K. M. (2010). Exit polling and racial bloc voting: Combining individual-level and R×C ecological data. *The Annals of Applied Statistics*, 4(4), 1774–1796. doi:10.1214/10-AOAS353

6. Tribunal Superior Eleitoral. *Dados Abertos: Resultados 2026; Eleitorado 2026*. dadosabertos.tse.jus.br

7. Pesquisas presidenciais Quaest (BR-09158/2026), Nexus/BTG (BR-08234/2026) e Datafolha (BR-09540/2026), outubro de 2026, conforme registradas no TSE.