Varredura de fiscalização por local de votação — método, dados e resultados

Documento técnico para avaliação. Gerado automaticamente dos resultados atuais em 24/10/2026; código no commit c7c47d0 2026-10-24. Revisado em 24/10/2026 após parecer externo (ver seção 11).

Pergunta. Existem locais de votação onde seções com (a) votação depois do horário e/ou (b) habilitação manual de eleitores pelo mesário acima do esperado apresentam resultado diferente do das seções vizinhas, de forma recorrente entre eleições? E o padrão difere entre eleições de prefeito e de presidente? O objetivo é ordenar locais para observação presencial. Os dados não observam compra de voto, coação ou intenção, nem em quem votaram os eleitores tardios ou habilitados manualmente; observam horários, habilitações, encerramento e votos por seção.

Sinalização não é acusação. Toda diferença medida aqui pode ter explicação legítima (fila, falha de biometria, perfil do eleitorado da rua atendida pela seção). A pontuação ordena prioridades de observação; não é probabilidade nem estimativa de irregularidade.

1. Dados usados

Todas as fontes são públicas, do Portal de Dados Abertos do TSE (dadosabertos.tse.jus.br). A aquisição lê os membros dos arquivos ZIP por HTTP Range, guarda apenas as colunas e linhas necessárias em Parquet comprimido, e registra URL, hash SHA-256 e recibo de cada captura. Nenhum arquivo nacional inteiro é armazenado.

Conjunto (TSE)	O que é	O que se usa aqui
Boletins de urna por seção (bweb_<turno>t_<UF>)	Uma linha por seção × candidatura, com aptos, comparecimento, abstenções, votos, horários de abertura/encerramento, urna efetivada, eleitores habilitados manualmente (campo do BU) e seções agregadas.	Aptos, comparecimento, votos válidos e por candidato (lado A/B), horário de encerramento, agregação. Base de todas as seções.
Logs das urnas (arqurnatot … <ano>_<turno>t_<UF>.zip; logjez/logd.dat)	Registro de eventos de cada urna com horário: habilitações, tipo de habilitação, voto computado, cancelamentos, mensagens de operação.	Contagem de votos computados por horário (após 17h, após 19h), sequência de habilitação manual, cancelamentos, último voto, intervalos entre votos tardios, reversões de relógio.
Correspondências efetivadas (ceft_<turno>t_<UF>)	Para cada seção, a urna esperada e a urna efetivamente usada, cargas e origem do voto (RED = recuperação de dados).	Troca de urna (esperada ≠ efetivada) e contingência; usada para não contar encerramento tardio quando houve troca.
Perfil do eleitorado por seção (perfil_eleitor_secao_2022_<UF>)	Eleitores por seção com faixa etária, escolaridade e biometria cadastrada; nome do local de votação.	Controles da regressão de habilitação manual (% com biometria, % 60+, % baixa escolaridade), nome dos locais. Perfil de 2022 para todas as eleições.
Logs de urnas de contingência	Logs cujo nome de arquivo identifica a seção de destino quando a urna original foi substituída.	Re-identificados pelo nome do arquivo (identity_status PARSED_FILENAME) e incluídos; marcados como contingência.

Recortes: 11 estados (Bahia, São Paulo, Minas Gerais, Rio de Janeiro, Paraná, Rio Grande do Sul, Pernambuco, Ceará, Pará, Santa Catarina, Maranhão). Eleições com logs de urna disponíveis e processados: presidente 2022 (1º e 2º turnos) e 2026 (1º turno); prefeito 2024 (1º e 2º turnos). Eleições anteriores (2012–2020) foram capturadas em boletins e correspondências para estudos de persistência, mas não entram nesta varredura por não terem logs processados.

1.1 Cobertura observada

Eleição	Dia	Seções (BU)	Com log	Log completo	Contingência / recarregadas	Votos após 17h (todos os logs)	… em contingência	Votos com sequência manual	Manuais após 17h
Pres. 2022 T1	02/10/2022	366.736	366.602	361.481	8.064	3.293.439	179.772	8.390.817	293.041
Pres. 2022 T2	30/10/2022	366.736	365.307	362.728	12.782	30.707	2.149	7.729.414	2.896
Pref. 2024 T1	06/10/2024	376.597	376.565	373.869	4.659	127.472	21.906	6.194.585	11.501
Pref. 2024 T2	27/10/2024	77.186	77.183	77.074	891	9.386	121	1.051.582	625
Pres. 2026 T1	04/10/2026	384.987	384.971	383.273	2.381	2.744.574	40.049	6.483.425	145.412

“Log completo” = o número de votos computados no log é igual ao comparecimento do boletim da seção. Só seções com log completo entram nos indicadores de horário e habilitação; comparecimento e encerramento usam todas as seções elegíveis (≥ 50 aptos).

2. Como os logs são lidos

Cada log é lido linha a linha, somente no dia da eleição e no módulo de votação (VOTA). Uma máquina de estados por urna acompanha a habilitação de cada eleitor: **solicitação de dado pessoal para habilitação manual** → **captura da digital do mesário** → **eleitor habilitado** → **voto computado**. Só quando a sequência completa termina em voto computado o voto conta como “sequência manual”. Mensagens de tipo de habilitação biométrica, “eleitor não possui biometria”, cancelamentos e suspensões reiniciam ou classificam a sequência; um voto com rótulos conflitantes é contado como conflito, não como manual. Horários são os do relógio de cada urna; reversões de relógio são contadas e interrompem a sequência. A taxonomia foi conferida contra o formato oficial dos logs (TSE, 2022) e contra amostras reais antes de ser adotada; a contagem de votos computados de cada log foi reconciliada com o comparecimento do boletim.

Validação: no piloto Bahia 2022 T2, 33.082 de 33.084 logs com identidade inequívoca tiveram contagem de votos igual ao comparecimento presidencial do BU. Nada disso valida o relógio das urnas nem prova que toda habilitação manual foi registrada; o que se mede é o que o log registra.

3. Indicadores por seção

Seção elegível: pelo menos 50 aptos e votos válidos > 0. Limiares P99 são calculados por estado, só entre seções elegíveis com log completo, dentro de cada eleição.

Indicador	Definição
late	Votos após 17h: fração acima do P99 do estado e pelo menos 20 votos após as 17h
manual	Habilitação manual pelo mesário acima do esperado para a seção (resíduo da regressão > P99 do estado)
manual_late	Pelo menos 10 votos com sequência manual após as 17h
fast	Ritmo acelerado após 17h: mediana entre votos abaixo de 40 s com pelo menos 20 votos tardios
last19	Último voto computado às 19h ou depois
turnout	Comparecimento muito acima: z > 3 frente à mediana estadual (MAD) e ≥ 10 p.p. acima da mediana municipal
close	Encerramento da urna mais de 60 min após as 17h sem troca de urna registrada
cancel	Cancelamentos/suspensões de habilitação pelo mesário acima do P99 do estado e pelo menos 5

Habilitação manual esperada. Para cada estado e eleição, regressão linear (mínimos quadrados) da fração de votos com sequência manual sobre: % de eleitores com biometria cadastrada, % com 60 anos ou mais e % de baixa escolaridade da seção (perfil 2022), estimada nas seções com log completo (mínimo de 200 seções). O resíduo (observado − esperado) é o que se compara ao P99. Isso ajusta por eleitorado idoso ou sem biometria; não separa causas.

Comportamento da seção (usado na varredura): TARDIO = late ou last19 ou close; MANUAL = manual ou manual_late. Comportamento = TARDIO ou MANUAL. Os demais indicadores (fast, turnout, cancel) entram no score do estudo de inconsistências e nos testes de direção, mas não na definição de comportamento da varredura.

4. Comparação com as seções vizinhas e “diferença acima do limiar”

Para cada seção com comportamento, os **pares** são as outras seções do mesmo local de votação (mesmo município, zona e código de local) que não têm comportamento na mesma eleição. Mede-se a diferença, em pontos percentuais, entre a fração de votos válidos do lado A na seção e a média ponderada dessa fração nos pares. Lado A = PT nas presidenciais; nas eleições de prefeito, o candidato líder do município (lado B = PL ou o 2º colocado).

z = diferença ÷ √(variância da fração do lado A entre os pares + erro binomial da própria seção, $100^2 \cdot p(1-p)/$votos válidos). **Diferença acima do limiar** = |diferença| ≥ 5 p.p. e |z| ≥ 2 (nas colunas dos arquivos chama-se “real”, nome herdado da primeira versão). **Diferença alta** = |diferença| ≥ 10 p.p. (descritiva).

Níveis de comparação. “Local”: 3 ou mais pares no mesmo prédio — é o resultado principal. “Local (poucos pares)”: 1 ou 2 pares; a média de referência é do prédio, mas a dispersão do z vem das seções comuns do município. “Município”: nenhum par no prédio — seção única no local ou prédio em que todas as demais seções também têm comportamento — e a comparação é com as seções comuns do município. Em 2026 T1, das 10.346 seções com comportamento no nível município, 3.289 são seção única e 7.057 estão em prédios sem par comum.

Taxa-base. O mesmo teste é aplicado a cada seção comum contra as demais seções comuns da sua referência, deixando-a de fora (leave-one-out), no mesmo nível. A fração de seções comuns que passa no teste é o que se espera por acaso e por heterogeneidade normal. A taxa-base é muito maior no nível município (eleitorados diferentes) do que no nível local; por isso os níveis não são misturados e o resultado principal usa só o nível local.

4.1 Taxa de diferença acima do limiar e direção, por nível (2026 T1 e 2022 T1)

Eleição	Nível	Com comp.: n	acima do limiar	%	favor A	favor B	% A	Comuns: n	acima do limiar	%	favor A	favor B	% A
Pres. 2026 T1	local	15.593	664	4,3	506	158	76,2	293.231	7.175	2,4	3.901	3.274	54,4
Pres. 2026 T1	local (poucos pares)	10.136	96	0,9	70	26	72,9	37.787	472	1,2	246	226	52,1
Pres. 2026 T1	município	10.346	809	7,8	717	92	88,6	17.701	2.332	13,2	1.839	493	78,9
Pres. 2022 T1	local	16.013	549	3,4	346	203	63,0	273.894	6.994	2,6	4.014	2.980	57,4
Pres. 2022 T1	local (poucos pares)	11.147	87	0,8	54	33	62,1	35.340	330	0,9	179	151	54,2
Pres. 2022 T1	município	11.313	899	7,9	739	160	82,2	18.894	2.441	12,9	1.905	536	78,0
Pref. 2024 T1	local	2.692	76	2,8	45	31	59,2	313.846	6.344	2,0	3.239	3.105	51,1
Pref. 2024 T1	local (poucos pares)	836	11	1,3	5	6	45,5	39.149	393	1,0	206	187	52,4
Pref. 2024 T1	município	905	152	16,8	69	83	45,4	19.019	2.751	14,5	1.358	1.393	49,4

Também se calcula a diferença de comparecimento frente aos pares (p.p.) e, quando metade ou mais das seções do local têm comportamento (“local inteiro”), a diferença do local inteiro frente às seções comuns do município. A razão de chances de Mantel–Haenszel (favor A × grupo, estratos = nível) resume a direção controlando a composição por nível.

5. Teste de direção por permutação

Pergunta: as seções sinalizadas por um indicador votam, em média, mais para um lado do que as demais seções do mesmo local? Estatística: diferença (A − B) na seção menos a média ponderada do seu local, ponderada por votos válidos, entre sinalizadas e não sinalizadas. Distribuição nula: a sinalização é embaralhada 500 vezes (semente 20261016) em dois esquemas — dentro do município e, como sensibilidade mais fiel à pergunta do mesmo prédio, dentro do próprio local; p = fração de embaralhamentos com diferença igual ou maior em módulo (bilateral); q = correção de Benjamini–Hochberg. A faixa do acaso nos gráficos é o intervalo entre os percentis 2,5 e 97,5 do nulo (centrada no nulo, não no efeito). Resultados completos em relatorios/fiscalizacao/testes_permutacao.csv; resumo:

Eleição	Indicador	Sinalizadas	Lado A % (sinal.)	Lado A % (demais)	Dif. no local (p.p.)	Faixa 95% do acaso (município)	p (município)	q	p (local)
Pres. 2022 T1	late	3.615	55,5	48,2	0,27	[-0,20; 0,18]	0,002	0,005	0,004
Pres. 2022 T1	manual	3.617	50,3	48,3	-0,10	[-0,20; 0,20]	0,315	0,545	0,323
Pres. 2022 T1	manual_late	5.783	67,8	48,0	0,20	[-0,15; 0,13]	0,006	0,014	0,002
Pres. 2022 T1	last19	7.510	68,2	47,8	0,42	[-0,11; 0,12]	0,002	0,005	0,002
Pres. 2022 T1	close	35.337	59,8	46,9	0,18	[-0,06; 0,05]	0,002	0,005	0,002
Pres. 2022 T1	score2	11.417	64,1	47,7	0,29	[-0,10; 0,09]	0,002	0,005	0,002
Pres. 2026 T1	late	3.780	52,9	45,0	1,18	[-0,20; 0,19]	0,002	0,005	0,002
Pres. 2026 T1	manual	3.644	47,6	45,0	0,15	[-0,22; 0,18]	0,168	0,335	0,158
Pres. 2026 T1	manual_late	859	70,4	45,0	0,86	[-0,42; 0,37]	0,002	0,005	0,002
Pres. 2026 T1	last19	5.561	70,0	44,6	0,88	[-0,14; 0,14]	0,002	0,005	0,002
Pres. 2026 T1	close	31.937	61,7	43,3	0,68	[-0,06; 0,06]	0,002	0,005	0,002
Pres. 2026 T1	score2	7.654	64,8	44,6	1,00	[-0,12; 0,13]	0,002	0,005	0,002
Pref. 2024 T1	late	855	57,7	55,3	-0,16	[-0,35; 0,37]	0,411	0,569	0,293
Pref. 2024 T1	manual	3.617	56,6	55,3	0,11	[-0,20; 0,20]	0,281	0,533	0,256
Pref. 2024 T1	manual_late	72	59,6	55,3	-0,25	[-1,41; 1,31]	0,697	0,764	0,715
Pref. 2024 T1	last19	62	54,5	55,3	-0,37	[-1,41; 1,58]	0,649	0,764	0,615
Pref. 2024 T1	close	424	57,0	55,3	-0,11	[-0,50; 0,57]	0,701	0,764	0,603
Pref. 2024 T1	score2	586	57,5	55,3	-0,09	[-0,45; 0,48]	0,689	0,764	0,651

Associação não é causa nem importância prática: significância diz apenas que a diferença é maior do que o acaso produz; a direção pode refletir o perfil de quem enfrenta fila.

6. Prefeito × presidente, nas mesmas seções

Seções presentes nas três eleições de primeiro turno (presidente 2022, prefeito 2024, presidente 2026). A contagem bruta de seções com desvio “só nas presidenciais” é muito maior que “só nas municipais” porque há muito mais seções com comportamento nos 1º turnos presidenciais (fila); a comparação correta usa taxas com denominadores e, de forma estrita, as mesmas seções com comportamento nos dois cargos. Lado A tem significado distinto por cargo (PT × líder do município); por isso direção e pontos de direção usam só as presidenciais.

Eleição	Recorte	Presentes	Com comp.	% com comp.	Com comp. e acima do limiar	% entre com comp.	Comuns acima do limiar	% entre comuns
Pres 2022 T1	todas	356.214	37.286	10,47	1.470	3,94	9.400	2,95
Pres 2022 T1	local	283.049	15.692	5,54	534	3,40	6.848	2,56
Pref 2024 T1	todas	356.214	4.159	1,17	220	5,29	8.294	2,36
Pref 2024 T1	local	302.729	2.585	0,85	72	2,79	5.595	1,86
Pres 2026 T1	todas	356.214	33.907	9,52	1.399	4,13	8.448	2,62
Pres 2026 T1	local	289.929	14.819	5,11	578	3,90	6.136	2,23

Par	Recorte	Seções c/ comp. nos dois cargos	Acima do limiar na presidencial	na municipal	% pres.	% mun.	Só pres.	Só mun.	p (McNemar)
Pres 2022 T1 × Pref 2024 T1	todas	1.086	56	59	5,16	5,43	45	48	0,836
Pres 2022 T1 × Pref 2024 T1	local	353	14	11	3,97	3,12	12	9	0,664
Pres 2026 T1 × Pref 2024 T1	todas	1.445	88	93	6,09	6,44	72	77	0,743
Pres 2026 T1 × Pref 2024 T1	local	522	21	19	4,02	3,64	17	15	0,860

Comparação estrita: nas mesmas seções com comportamento nos dois cargos, a propensão a ultrapassar o limiar é parecida (p alto no teste de McNemar): não há evidência de assimetria entre cargos.

7. Pontuação do local e prioridade de observação

A pontuação é uma soma transparente, pensada para ser lida e discutida, não um modelo ajustado. Componentes, máximos e tipo (operacional = fila/habilitação manual e volume; resultado = diferenças de votação):

Componente	Regra	Máx.	Tipo
Recorrência	3 pontos por seção com comportamento em 2+ eleições (até 5) + 2 por seção em 3+ eleições (até 3)	21	operacional
Desvio nas presidenciais	5 pontos por seção com comportamento e diferença acima do limiar em alguma presidencial (até 5)	25	resultado
Direção única	8 pontos se há 2+ diferenças acima do limiar nas presidenciais e todas para o mesmo lado	8	resultado
Comportamento em 2026 T1	3 por seção com comportamento em 2026 T1 (até 4)	12	operacional
Diferença em 2026 T1	3 por seção com diferença acima do limiar em 2026 T1 (até 3)	9	resultado
Prédio inteiro atípico	8 se ≥50% das seções têm comportamento em 2026 T1 e o prédio difere do município com $\|z\| \geq 2$; 4 se só a primeira condição	8	resultado (vs município) / operacional
Volume	votos após 17h no local: ≥100 → 3, ≥500 → 6; habilitações manuais após 17h: ≥20 → 3, ≥50 → 6	12	operacional
2+ seções com desvio só nas presidenciais	5 se 2+ seções do local têm diferença acima do limiar só nas presidenciais (redundante com o desvio presidencial; mantido com registro)	5	resultado

Níveis de **prioridade de observação**: alta ≥ 45 pontos; média 25–44; baixa < 25. É ordem de prioridade, não probabilidade. Dos 451 locais de prioridade alta, 120 não têm nenhuma seção com diferença acima do limiar frente à própria referência (93 têm o prédio inteiro divergindo do município) e 28 não têm nenhum ponto de resultado: entram por fila/habilitação manual recorrentes, volume e comportamento em 2026. Na prioridade média, 1.825 de 2.387.

Prioridade para o 2º turno de 2026: locais com seções presentes em 2026 e comportamento em 2026 T1 ou recorrente, ordenados por pontos. No mapa de fiscalização entram os níveis alta e média; “seções a checar” são as seções do local com comportamento em 2026 T1 ou recorrentes (2+ eleições) existentes em 2026; fiscais sugeridos = 1 por local de prioridade alta + 1 a cada 2 de média, arredondado para cima em cada zona (ponto de partida para alocação).

8. Resultados principais

8.1 Contagens, com o nível “local” (mesmo prédio, ≥3 pares) como resultado principal

Eleição	Seções	Com comp.	Tardia	Manual	Com comp. no nível local	acima do limiar (local)	% (com comp., local)	% (comuns, local)	acima do limiar (município)	Locais inteiros
Pres. 2022 T1	366.601	38.473	35.628	8.635	16.013	549	3,4	2,6	899	6.690
Pres. 2022 T2	366.600	3.691	70	3.629	2.346	78	3,3	2,6	86	387
Pref. 2024 T1	376.447	4.433	887	3.641	2.692	76	2,8	2,0	152	527
Pref. 2024 T2	77.157	741	3	738	681	21	3,1	1,8	9	15
Pres. 2026 T1	384.794	36.075	32.961	4.275	15.593	664	4,3	2,4	809	6.354

8.2 Direção (nível local) e resumo por estratos

Eleição	Favor A / B (com comp., local)	% A	Favor A / B (comuns, local)	% A	OR Mantel–Haenszel (estratos = nível)	Comparec. dif. p.p.	Excesso A além dos votos após 17h: % (nível local)	% (todos os níveis)
Pres. 2022 T1	346 / 203	63,0	4.014 / 2.980	57,4	1,29	-0,13	61,3	86,5
Pres. 2022 T2	41 / 37	52,6	4.573 / 3.298	58,1	0,64	-0,08	100,0	100,0
Pref. 2024 T1	45 / 31	59,2	3.239 / 3.105	51,1	0,99	-0,23	97,8	98,3
Pref. 2024 T2	10 / 11	47,6	695 / 639	52,1	1,07	-1,02	100,0	100,0
Pres. 2026 T1	506 / 158	76,2	3.901 / 3.274	54,4	2,45	-0,21	70,8	87,1

• No mesmo prédio, seções com comportamento ultrapassam o limiar em 2,8% a 4,3% dos casos, contra 1,8% a 2,6% nas seções comuns (em 2026, razão de 1,7): a maioria vota como o resto do prédio.

• Direção: em 2026 T1, no mesmo prédio, 76% das diferenças das seções com comportamento pendem para o lado A (PT), contra 54% nas comuns (OR de Mantel–Haenszel 2,45). Em 2022 T1 a distância é pequena (63% contra 57%; OR 1,29). Nas eleições de prefeito não se observa assimetria. A direção é robusta só em 2026 T1.

• Misturando os níveis de comparação, 2026 T1 teria 1.569 diferenças acima do limiar, 809 delas no nível município, onde ambos os grupos pendem para A: por isso o resultado principal é o do nível local.

• Locais com ao menos uma seção com comportamento: 29.032; prioridade alta 451, média 2.387, baixa 26.194. Entre os locais com alguma diferença nas presidenciais, direção única para A em 1.880, para B em 587, mista em 17.

• Prioridade 2026 T2: 18.689 locais, 451 de prioridade alta e 2.355 média. Mapa de fiscalização: 2.806 locais, 14.107 seções a checar, 1.826 fiscais sugeridos.

9. Limitações e o que não se pode concluir

• Os dados não observam compra de voto, coação, transporte de eleitores ou intenção. Observam horários, habilitações, encerramento e votos por seção.

• Não se sabe em quem votaram os eleitores habilitados manualmente ou após as 17h; compara-se o resultado da seção inteira. Em 2026 T1, no nível local, mesmo que todos os votos após as 17h fossem do lado A, em 71% das 506 seções com comportamento que destoaram a favor de A o excesso não seria explicado por eles (o contrafactual cobre só os votos após as 17h, não os habilitados manualmente): a diferença é da seção, não só dos eleitores tardios.

• Votar após as 17h é legal para quem estava na fila; habilitação manual é o procedimento previsto quando a biometria falha ou o eleitor não a tem cadastrada. Ambos medem fila e falha, não irregularidade.

• A diferença frente aos vizinhos reflete também composição do eleitorado: seções do mesmo prédio atendem ruas e perfis distintos. O teste z e a taxa-base reduzem, mas não eliminam, esse fator.

• Seções sem par comum no prédio (seção única ou prédio inteiro com comportamento) são comparadas ao município, referência fraca com taxa-base alta nos dois grupos; estão separadas do resultado principal e não devem ser lidas como diferença frente ao prédio.

• A direção favorável ao lado A nas seções com comportamento é robusta só em 2026 T1; em 2022 é pequena e, no agregado, majoritariamente efeito de composição por nível. É compatível com várias explicações (perfil

de quem enfrenta fila; densidade; operação irregular); os dados não distinguem entre elas.

• Horários são do relógio de cada urna, não certificados. Reversões são contadas; desvios constantes de relógio não são detectáveis. O perfil do eleitorado é de 2022 e é usado para todas as eleições.

• A pontuação é uma regra transparente escolhida pelos autores, não um modelo validado contra casos confirmados; ordena prioridades de observação. O componente “2+ seções com desvio só nas presidenciais” é redundante com o desvio presidencial e foi mantido com registro; mudanças de pesos mudam a lista.

• 2º turnos (2022 e prefeito 2024) têm poucas seções com comportamento e baixo poder; prefeito 2024 T2 só existe em municípios com segundo turno.

10. Reprodutibilidade

Ambientes: Python 3.12; pacote auditabr 0.2.0 (requirements.lock) para aquisição e leitura de logs; .venv-report (requirements-report.lock) para análises e PDFs. Sequência de scripts, todos em scripts/:

Etapa	Script	Saída
Catálogo CKAN e aquisição de BU/logs por UF	uf_pipeline.py, mun_logs_pipeline.py, bu_capture_all.py	research/estados/<UF>/…, research/municipais/<UF>/…, research/bu/
Correspondências de urnas e perfil do eleitorado	ceft_capture.py, profile_capture.py	research/ceft/, research/perfil/2022/
Leitura dos logs (máquina de estados)	analyze_pilot_operations.py (motor), process_bahia_document_logs.py, extract_contingency_logs.py, reidentify_logs.py	operations.jsonl por eleição/UF
Análise por seção e cargo	analyze_office.py, reanalyze_with_logs.py	sections.parquet, candidate_sections.parquet, summary.json
Indicadores, testes de permutação, ferramenta	make_fiscalizacao_study.py	relatorios/fiscalizacao/ (secoes_indicadores_*.parquet, testes_permutacao.csv, PDF)
Varredura por local	make_varredura_locais.py	relatorios/fiscalizacao/varredura/ (locais_risco.csv, comparacao_por_nivel.csv, pres_x_pref_*.csv, VARREDURA_LOCAIS.xlsx, PDF)
Mapa de fiscalização 2026 T2	make_mapa_fiscalizacao_2026t2.py	relatorios/fiscalizacao/mapa_2026_T2/ (XLSX e PDF por UF)
Estes documentos	make_documentos_metodo.py	relatorios/fiscalizacao/documentos/

Parâmetros fixos: elegibilidade ≥ 50 aptos; P99 por estado; diferença acima do limiar ≥ 5 p.p. e |z| ≥ 2; pares mínimos 3; permutação 500 réplicas, semente 20261016, dois esquemas; cortes de prioridade 25 e 45 pontos. Cada captura tem recibo com URL, SHA-256 e data; cada saída tem data de geração. O repositório git do projeto registra o código; os dados brutos ficam fora do versionamento por tamanho.

11. Registro de alterações (24/10/2026, após parecer externo)

• Resultado principal restrito ao nível “local” (≥3 pares no mesmo prédio); nível “município” reportado em separado, com origem (seção única / prédio sem par) e taxa-base própria; tabela por nível publicada (comparacao_por_nivel.csv).

• “Diferença real” passou a “diferença acima do limiar” na prosa; os nomes de colunas (“real”) foram mantidos para não quebrar planilhas e scripts.

• Direção (favor A/B) e pontos de direção passaram a usar só as presidenciais, onde lado A tem significado estável; prefeito reportado à parte. “Só municipais” passou a exigir presença nos dados presidenciais, espelhando “só presidenciais”.

• Cruzamento prefeito × presidente substituído pela comparação pareada nas mesmas seções, com denominadores, e pela comparação estrita (seções com comportamento nos dois cargos, McNemar).

• “Risco” passou a “prioridade de observação”; pontos separados em operacionais e de resultado; coluna sinal_resultado; componente “só presidenciais” renomeado para “2+ seções com desvio só nas presidenciais”, peso mantido e redundância registrada.

• Teste de permutação: desvio frente à média ponderada do local (nulo centrado), faixa do acaso pelos quantis do nulo, embaralhamento adicional dentro do local; gráfico redesenhado (faixa cinza = acaso; ponto = observado).

• Textos: retiradas atribuições causais (“muda o resultado”, “não é acaso”, “o que importa”); acrescentada a ressalva de que não se sabe em quem votaram os eleitores tardios; tabelas redimensionadas para a largura do texto.

• Maranhão incluído em toda a cadeia (11 estados).

12. Glossário

Termo	Significado
Seção	Menor unidade de votação (uma urna); identificada por município, zona e número.
Local de votação	Prédio (escola etc.) que reúne várias seções; identificado por município, zona e código do local.
Zona eleitoral	Divisão administrativa que agrupa seções; pode abranger vários municípios ou parte de um.
BU / boletim de urna	Resultado oficial impresso e transmitido por cada urna ao fim da votação.
Log de urna	Arquivo com os eventos registrados pela urna, com horário.
Habilitação manual	Liberação do eleitor pelo mesário sem reconhecimento biométrico, com registro da digital do mesário.
Lado A / lado B	PT / PL nas presidenciais; líder / 2º colocado do município nas eleições de prefeito.
p.p.	Pontos percentuais (diferença entre duas porcentagens).
P99	Valor acima do qual está apenas 1% das seções do estado.
z	Quantos desvios-padrão a diferença da seção está da variação normal entre seções da sua referência.
Diferença acima do limiar	≥5 p.p. e $\|z\| \geq 2$; chamada “real” nas colunas dos arquivos.
Nível de comparação	local (≥3 pares no prédio), local (poucos pares), município (sem par no prédio).
Permutação	Teste que embaralha os rótulos muitas vezes para ver o que aconteceria por acaso.
q (BH)	p corrigido para muitos testes simultâneos (Benjamini–Hochberg).
OR de Mantel–Haenszel	Razão de chances combinada entre estratos; 1 = sem diferença de direção entre grupos.
McNemar	Teste pareado para duas medidas na mesma unidade (aqui, a mesma seção em dois cargos).
Contingência	Urna substituta usada quando a original falha.
CEFT	Correspondência entre a urna prevista e a efetivamente usada em cada seção.