

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

LEANDRO MAGNO CORREA DA SILVA

**ANÁLISE ESTATÍSTICA PARA DETECÇÃO DE REGULARIDADES EM ELEIÇÕES
BRASILEIRAS**

DISSERTAÇÃO

CAMPO MOURÃO

2020

LEANDRO MAGNO CORREA DA SILVA

**ANÁLISE ESTATÍSTICA PARA DETECÇÃO DE REGULARIDADES EM ELEIÇÕES
BRASILEIRAS**

Dissertação apresentada ao Programa de Pós Graduação em Inovações Tecnológicas, da Universidade Tecnológica Federal do Paraná, como requisito parcial para a obtenção do grau de Mestre.
Orientadora: Prof^a. Dr^a. Magda Cardoso Mantovani.

**CAMPO MOURÃO
2020**

Dados Internacionais de Catalogação na Publicação

Silva, Leandro Magno Correa da
Análise estatística para detecção de regularidades em eleições brasileiras/ Leandro Magno Correa da Silva. – Campo Mourão, 2020.
1 arquivo de texto (106 f.) : PDF ; 6,51MB

Orientador: Magda Cardoso Mantovani
Dissertação (Mestrado) – Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Inovações Tecnológicas, Campo Mourão, 2020.
Inclui bibliografia: f. 95-102

1.Estatística. 2. Eleições. 3. Crime eleitoral 4. Inovações tecnológicas – Dissertações.
I. Mantovani, Magda Cardoso, orient. II. Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Inovações Tecnológicas. IV. Título.

Biblioteca da UTFPR - Câmpus Campo Mourão

Bibliotecária/Documentalista:
Lígia Patrícia Torino – CRB-9/1278

FOLHA DE APROVAÇÃO

LEANDRO MAGNO CORREA DA SILVA

ANÁLISE ESTATÍSTICA PARA DETECÇÃO DE REGULARIDADES EM ELEIÇÕES BRASILEIRAS

Dissertação apresentada ao Programa de Pós Graduação em Inovações Tecnológicas, da Universidade Tecnológica Federal do Paraná, como requisito parcial para a obtenção do grau de Mestre.

Data de aprovação: 18 de fevereiro de 2020.

Prof^ª. Dr^ª. Magda Cardoso Mantovani
Universidade Tecnológica Federal do Paraná

Prof. Dr. Leandro de Santana Costa
Universidade Estadual de Londrina

Prof. Dr. Diogo Heron Macowski
Universidade Tecnológica Federal do Paraná

Prof. Dr. Wyrllen Everson de Souza
Universidade Tecnológica Federal do Paraná

A folha de aprovação assinada encontra-se na coordenação do curso ou programa.

AGRADECIMENTOS

A Deus por permitir a conclusão desse trabalho me dando saúde, força e me capacitando, e principalmente guiando a cada passo.

Agradeço à minha orientadora Professora Doutora Magda Cardoso Mantovani, pelo empenho, sabedoria e paciência com que me guiou nesta trajetória.

Deixo registrado também, o meu reconhecimento à minha família: à minha esposa Rejane Bortolozzo e às minhas filhas Laura e Heloisa Bortolozzo, pois sem o apoio delas seria impossível vencer esse desafio.

Agradeço ao apoio incondicional e infraestrutura necessária em Campo Mourão fornecidos pelos meus amigos e compadres Giulio Wagner Volante e Edivana de Albuquerque Canto Volante.

Agradeço aos meus amigos de trabalho no IFPR que contribuíram incentivando e ajudando a propiciar condições para o desenvolvimento deste trabalho.

Agradeço também aos meus colegas do programa de mestrado.

Enfim, a todos que diretamente ou indiretamente fizeram parte dessa conquista, o meu muito obrigado.

*“Bem-aventurados os que têm um coração de pobre, porque deles é o Reino dos Céus!
Bem-aventurados os que choram, porque serão consolados!
Bem-aventurados os mansos, porque possuirão a terra!
Bem-aventurados os que têm fome e sede de justiça, porque serão saciados!
Bem-aventurados os misericordiosos, porque alcançarão misericórdia!
Bem-aventurados os puros de coração, porque verão Deus!
Bem-aventurados os pacíficos, porque serão chamados filhos de Deus!”
(Evangelho Segundo São Mateus 5, 3-10, A Bíblia Sagrada.)*

RESUMO

Usando conceitos e ferramentas estatísticas, este trabalho foi direcionado a compreender como se deu a participação do eleitorado nas urnas nas 5 últimas eleições presidenciais brasileiras, comparando-as e obtendo regularidades e anomalias nas distribuições de votos e comparecimento. Verificou-se a ausência de indicativos de mecanismos de fraude, reforçando a fragilidade da tese de fraude em eleições brasileiras. Observou-se uma regularidade clara entre as eleições que elegem candidatos do partido do governo e um comportamento mais irregular na eleição de 2018.

Palavras-chaves: Estatística. Eleições. Democracia. Urna Eletrônica. Fraude.

ABSTRACT

Using statistical concepts and tools, this work was aimed at understanding how the electorate participated in the polling stations in the last 5 Brazilian presidential elections, comparing them and obtaining regularities and anomalies in the distribution of votes and attendance. There were no indications of fraud mechanisms, reinforcing the fragility of the fraud thesis in Brazilian elections. Clear regularity was observed between elections electing candidates from the ruling party and more irregular behavior in the 2018 election.

Key-words: Statistics. Elections. Democracy. Voting Machine. Fraud.

SUMÁRIO

1	INTRODUÇÃO	10
2	ELEIÇÕES NO BRASIL E A URNA ELETRÔNICA	14
2.1	A História das Eleições no Brasil	15
2.1.1	Brasil Colônia (1500-1822)	15
2.1.2	Império Independente (1822-1889)	16
2.1.3	Primeira República (1889-1930)	16
2.1.4	Refundação da República (1930-1937)	18
2.1.5	República Democrática(1945-1964)	18
2.1.6	Regime Militar (1964-1985)	19
2.1.7	Nova República (1985 até hoje)	19
2.1.8	A Urna Eletrônica: Hardware e Software	20
2.2	O Processo de Votação Eletrônico	22
2.3	A Segurança da Urna Eletrônica	25
3	DINÂMICAS DE VOTAÇÃO NAS URNAS	30
3.1	Correlações Espaciais nas Estatísticas Eleitorais na França	31
3.2	Detecção Estatística de Irregularidades Eleitorais Sistemáticas	35
3.3	Envolvimento Cívico do Eleitorado: Medida da Desordem nas Urnas	38
3.4	Anomalias Estatísticas nas Eleições Russas ne 2011–2012 Reveladas pela Análise de Correlação 2D	42
3.5	Testes para Manipulação de Eleitores em Pequenas Seções de Votação	49
3.6	O estudo de possíveis fraudes eleitorais pode ser muito mais frequente do que se possa imaginar	59
3.6.1	Detectando fraudes eleitorais por irregularidades nas distribuições de votos	59
3.7	Próximos passos	60
4	FERRAMENTAS ESTATÍSTICAS	61
4.1	Distribuição Gaussiana e Distribuição Lognormal	61
4.2	Lei de Potência ou Escala	62
4.3	Teste Kolmogorov-Smirnov	62
4.4	Distribuição Beta	64
4.5	Distribuição Weibull	65
4.6	Assimetria e Curtose	66

4.6.1	Assimetria	66
4.6.2	Curtose	67
4.7	Entropia Shannon	68
4.8	Conclusão do capítulo	69
5	ENVOLVIMENTO CÍVICO EM ELEIÇÕES PRESIDENCIAIS BRASILEIRAS	70
5.1	Apresentação dos Dados	71
5.2	Análise dos Dados	72
5.3	Resultados	90
5.4	Discussão	92
5.5	Conclusões	94
	REFERÊNCIAS	96
6	APÊNDICES	104
6.1	Procedimentos ao Processar Dados do TSE	104
6.2	Principais Trechos de Código Fonte em R	105

1 INTRODUÇÃO

“A democracia surgiu quando, devido ao fato de que todos são iguais em certo sentido, acreditou-se que todos fossem absolutamente iguais entre si” – Aristóteles (384-322 a.C)

Desde seu surgimento, a palavra “democracia” tem sido crescentemente usada, assim como muitos esforços para implantá-la, como IDEA (2017), DI(2018) e IDW(2018) com o objetivo de dar a todos os cidadãos condições igualitárias de uma vida justa. Quando Aristóteles, vendo a recém criada democracia já se mostrava incrédulo, é razoável que Churchill, mais de 2260 anos depois, ainda lamentasse não haver nada melhor.

“A democracia é a pior forma de governo, exceto todas as outras formas que foram tentadas de tempos em tempos” - CHURCHILL(1947).

O Brasil vive hoje em uma democracia em maturação, com seus avanços e fragilidades ainda por sanar, principalmente no tocante à estabilidade dos partidos (DARAUJO, 2015). No entanto, mais do que nunca a democracia está enraizada culturalmente em muito do que se faz no país. Ao mesmo tempo, graças às recentes instabilidades no governo (BRASIL ID, 2018), é possível notar que as circunstâncias levam a população a retomar a atenção ao que acontece no cenário político, levando a repensar o sentido de cada fato ou ação que remete ao processo democrático.

A democracia surgiu em Atenas, na Grécia, no período histórico intitulado “clássico” pela história, entre os anos de 508 a 322 antes de Cristo. Neste período o voto, instrumento de participação democrática, era proferido pessoalmente numa assembleia, inicialmente, diretamente pelo cidadão participante e posteriormente aperfeiçoado com práticas usadas até hoje, como voto escrito, conselhos, representantes e corpo eleitoral (TIERNO, 2018), o que constitui um processo progressivamente inovador.

As tantas inovações que surgiram ao longo da história transformaram a democracia e seus processos de tal forma que todos as opiniões tornaram-se opções discretas e conseqüentemente os fatos viraram números, criando um cenário complexo e desconfiável por natureza. Não era de se esperar algo diferente quando se trata de normatizar processos de escolha de governantes, e estes por sua vez podem mudar de forma tão profunda a vida de tantas pessoas.

Desde a primeira eleição no Brasil, em 1821, ainda para eleger deputados da corte de Lisboa, passando pela primeira eleição direta em 1881 o sistema eleitoral brasileiro tem se desenvolvido de forma bastante irregular, numa instabilidade constante,

em que ora se dava por processos duvidosos, e ora por processos que tendem a legitimidade da eleição. (BRASIL, 2014)

Uma verdadeira “montanha russa” de sobe e desce no nível de democracia no Brasil foi descrita por BIZZARRO (2017), por vezes descrito como uma espiral onde aconteceram grandes mudanças; e a cada mudança grande parte do processo anterior é substituído. Este cenário instável é demonstrado ainda pior ao se olhar para outros estudos como ARANHA (2018 e 2019) e JUSBRASIL (2017).

Nos últimos anos, pesquisadores tem utilizado conceitos e ferramentas da estatística na tentativa de entender fenômenos coletivos emergentes das interações individuais (como unidades elementares) em estruturas sociais. São exemplos de estudos de fenômenos eleitorais: hierarquia no processo de candidatura (MANTOVANI et al., 2011), análises de resultados de eleições (MANTOVANI, 2013; COSTA FILHO, 1999; COSTA FILHO, 2002; LYRA, 2003; ARARIPE, 2006; ARAÚJO, 2010; TRAVIESO, 2006; FORTUNATO, 2006; ARARIPE, 2009; CHATTERJEE, 2013), e diferença de gênero em processo de candidatura (CARDOSO, 2020).

Este contexto intrincado remete ao ponto de vista dos sistemas complexos, que são difíceis de modelar usando a abordagem de sucessivos refinamentos à procura de elementos fundamentais. A abordagem de um sistema com o entendimento apropriado que é um sistema complexo permite observar propriedades quando olhado para o todo o que não é possível se olhado por partes. Assim, os estudos de objetos com características similares a estes sistemas permitem definir padrões e propriedades gerais presentes em classes de diferentes sistemas (MACAU, 2002). Neste caso, o comportamento dos eleitores nas urnas frente a modelos matemáticos e estatísticos de propriedades conhecidas que indiquem possíveis conclusões acerca dos pleitos eleitorais, seja por regularidades ou irregularidades.

A busca constante pela eleição sem fraudes persiste, não só no Brasil, mas em todo mundo. Em alguns países, pesquisadores (KLIMEK, 2012, BORGUESI, 2012, JIMENEZ, 2017 e KOBAK 2012) propuseram e utilizaram alguns métodos estatísticos para verificar se há indicadores de fraudes em pleitos eleitorais; ou até mesmo concluir que os padrões são bastante adversos. Em alguns casos os indicativos foram contundentes.

No entanto, olhando a história das eleições do Brasil (BRASIL, 2014), nota-se que, conforme se tentava implantar um processo onde o cidadão realmente faria valer seu voto, surgiam várias formas de fraudes que comprometiam parcialmente ou completamente a eleição. Algumas formas de fraude duraram décadas.

A urna eletrônica (BRASIL, VA, 2017), cujas primeiras previsões datam de 1932, e que foi utilizada somente em 1996 numa eleição municipal, trouxe a necessidade

de haver uma base de dados unificada, que por sua vez se propôs a ser fidedigna com a votação. Mas para tanto, foi necessário o cadastramento de candidatos e (em 1985) dos quase 70 milhões de eleitores. Assim, o Brasil dá um passo relevante no sentido de expurgar de vez as fraudes (tanto por parte dos eleitores como por parte dos escrutinadores): a informatização do processo. A informatização foi um feito que possibilita o funcionamento da urna eletrônica brasileira e também evita fraudes como votar em mais de uma seção. Também é o que nos permite fazer esta pesquisa sobre esta grande base de dados.

Optamos por estudar as últimas eleições do Brasil porque os dados disponíveis no Repositório de Dados do Tribunal Superior Eleitoral (TSE) (BRASIL, BD, 2019) encontram-se completos e organizados (consistentes). Assim, a proposta é estudar uma polêmica que existe diante do nosso sistema eleitoral, e investigar o comportamento dos eleitores nas urnas através dos dados de votação.

De que adianta o processo eleitoral brasileiro usar de estratégias, sistemas e equipamentos *sui generis* (inclusive exportados para vários países do mundo), se ao término ainda pode-se observar indicativos de fraude? A que servem grandes discursos ideológicos dos candidatos e representantes do TSE, se por fim, prevalecem no poder políticos que usam de métodos fortuitos para corromper o sistema eleitoral que tanto zela pela verdade das urnas?

O objetivo geral desta pesquisa é, usando ferramentas matemáticas, estatísticas e computacionais, verificar possibilidade da existência regularidades (e/ou irregularidades, como indicativos de fraudes) no processo eleitoral brasileiro, durante a história recente (a partir de 1998, quando passou-se a usar as urnas eletrônicas), mais especificamente nas eleições presidenciais brasileiras.

Foram usadas variáveis como: número de eleitores aptos a votar, números de votos brancos e nulos, número de comparecimento e de abstenções, e totais de votação por candidato para analisar a existência de padrões de comportamento e desvios neste padrão no processo eleitoral.

Os objetivos específicos deste trabalho são:

- Entender o comportamento e os relacionamentos das variáveis brancos e nulos, votos no primeiro e segundo colocados, e abstenções;
- Fazer um apanhado na literatura de ferramentas estatísticas que outros pesquisadores têm usado para produzir trabalhos similares usando base de dados eleitorais de outros países;
- Obter e ajustar os dados eleitorais brasileiros fornecidos pelo Tribunal Superior Eleitoral (TSE) de forma a serem usáveis por ferramentas estatísticas, como a

plataforma da linguagem R;

- Aplicar as ferramentas estatísticas na base de dados eleitoral brasileira ajustada;
- Identificar comportamentos anômalos do eleitorado em uma ou mais eleições, em relação às demais;
- Propor um modelo de potência utilizando os votos brancos e nulos.

O capítulo 2 mostra um histórico do sistema eleitoral brasileiro, incluindo o que é possível saber sobre a urna eletrônica e sua segurança, com ênfase no sistema atual e na polêmica em relação a fraudes.

O capítulo 3 traz uma revisão bibliográfica do que tem sido pesquisado sobre eleições recentemente ao redor do mundo, através de trabalhos de análise e de dinâmicas nas urnas do ponto de vista dos sistemas complexos.

O capítulo 4 mostra algumas ferramentas estatísticas utilizadas no produzir os resultados do próximo capítulo, como as distribuições de probabilidade Gaussiana, Log-normal, Beta e Weibull; também leis de potência, teste Kolmogorov-Smirnov, Entropia de Shannon, e Assimetria e Curtose em distribuições.

O capítulo 5 trás o panorama dos dados resultantes das eleições presidenciais no Brasil nas últimas décadas, métodos e resultados obtidos pela análise de irregularidades em eleições presidenciais brasileiras. Trás também as consequentes conclusões sobre estes resultados, dentre elas a de que não há indicativos de fraude nas eleições presidenciais brasileiras, ao menos não há indicativos usando os métodos propostos pelos pesquisadores KLIMEK (2012), BORGUESI (2012), JIMENEZ (2017) e KOBAC (2012). No entanto há alguns indícios nos resultados de que há um comportamento diferente dos eleitores que elegem um candidato da oposição ao seu governante anterior, ou seja, o eleito não tinha a “máquina” governamental na mão, bem como a indicação de que nas últimas eleições presidenciais no Brasil houve a presença de fatores diferentes que mostra uma distribuição dos dados estatisticamente mais distante da normalidade aparente os pleitos anteriores.

O que nos leva a levantar questões sobre o grau de dificuldade de um candidato se eleger presidente no Brasil quando já está ocupando o cargo de presidente, portanto “tem a máquina na mão”, e questionar os fatores que podem ter contribuído com o vencedor em 2018, causando uma votação representativa mesmo nas urnas em que não foi vencedor, o que não parece ter ocorrido nas eleições anteriores, pontos de partida para investigações futuras, bem como a possibilidade de se quantificar numericamente a dificuldade real de um candidato a presidente se eleger.

2 ELEIÇÕES NO BRASIL E A URNA ELETRÔNICA

Seguindo a natureza complexa ao mesmo tempo desconfiável como já citamos no capítulo 1, não era de se esperar menos do que o aumento da desconfiança do eleitorado em geral ao se proclamar que a votação passaria a ser através de um computador, a despeito de chamá-la de “urna eletrônica” (BRASIL, UE, 2019) para manter a maior familiaridade com as urnas tradicionais, modelo usado até o início da década de 1990.

O Tribunal Superior Eleitoral (TSE) disponibilizou em seu sítio oficial na internet em uma seção intitulada “Perguntas Frequentes sobre as Eleições” (BRASIL, PF, 2018), com perguntas como: “Por que outros países não usam urna eletrônica?” e “A urna pode mostrar em quem o eleitor votou?”. Quase às vésperas da eleição de 2018, o TSE (BRASIL NV, 2018) publica nota de esclarecimento sobre a segurança das urnas eletrônicas, mencionando um vídeo (sem mencionar o autor) que questiona a idoneidade da urna eletrônica.

A partir das reações oficiais do TSE (BRASIL, PF, 2019; BRASIL NV, 2018; BERNARDINETTI, 2019) acerca da desconfiança da população e/ou de tentativas de desqualificar todo o processo eleitoral brasileiro pode-se observar o panorama geral e a que ponto chega a crença ou não dos eleitores na urna eletrônica (BRASIL, PS, 2010; TECMUNDO, 2018; ARANHA, 2019; DEL PICCIA, 2017), mesmo com um parecer de uma comissão bastante qualificada de pesquisadores da Unicamp que tem respaldado a equipe de desenvolvimento da urna eletrônica do TSE. (BRASIL, RF, 2002)

Esta desconfiança dá-se primordialmente como resultado atual da história eleitoral brasileira, que equaciona alternados momentos de aumento e diminuição da legitimidade democrática no processo eleitoral, decorrente diretamente da história política. A seguir, uma breve história das eleições no Brasil, conforme, primordialmente BRASIL, 2014), desde as primeiras eleições até hoje, incluindo as dinâmicas do processo de votação, observando o funcionamento e segurança (onde residem a maior parte das dúvidas).

A inovação é, possivelmente, a única constante no processo democrático brasileiro (assim como em outros países), tanto por parte dos órgãos promotores como por eventuais transgressores, ou mesmo, em alguns momentos históricos, como uma das poucas opções para tentar fazer do pleito um processo minimamente democrático.

2.1 A História das Eleições no Brasil

Brasil (2014) trás a história oficial das eleições no Brasil. Nogueira (2012), BRASIL (2009) e Kleina (2019) acrescentam alguns pormenores e que ilustram a história.

2.1.1 Brasil Colônia (1500-1822)

A primeira eleição no Brasil ocorreu em 23 de janeiro de 1532 e definiu os membros do Conselho Municipal da Vila de São Vicente (atual São Paulo). Sem nenhum pudor quanto à exclusão, os eleitores eram apenas os “homens bons”: nobres de linhagem, senhores de engenho, membros da alta burocracia militar e os “homens novos” (burgueses enriquecidos pelo comércio). Estes foram os critérios utilizados para selecionar cidadãos aptos a votar no Brasil. Como a fonte de dados era o recenseamento populacional, constituiu-se a definição do “voto censitário”.

Durante o Brasil colônia ocorreram as eleições para os cargos de vereador, juiz ordinário, procurador e outros oficiais. O processo seguiu tal como era na corte portuguesa, de acordo com as Ordenações do Reino. Esta foi a primeira fase de nossa democracia.

O processo eleitoral do Brasil colônia era constituído pelas etapas (Brasil, 2014): escolha dos eleitores, formação de duplas, elaboração de nove listas, redução de nove listas para três, pelouros e sorteio. Curiosamente, pelouro era uma bola para guardar a lista dos candidatos que seria sorteada de uma arca por um menino de até 7 anos.

Nota-se já aí um tênue limite entre a (pseudo) democracia e o autoritarismo, sugerida por existência de sorteio dentre uma casta seleta da qual a população em geral não participava por não se qualificar dentre os “homens bons”. Este sistema persistiu por aproximadamente três séculos, enraizando (inclusive de forma cultural) a tradição de exclusão.

A partir de 1799 o contexto histórico de Portugal tornou-se instável, o que causou a vinda da família real para o Brasil e o posterior retorno em 1821. Ao término desta “estadia”, Dom João VI convocou os brasileiros para eleger os deputados para as cortes de Lisboa. Este pleito se realizou em 4 graus (níveis): os cidadãos das freguesias nomeavam compromissários, estes escolhiam os eleitores de paróquia, estes designavam os eleitores de comarca, e estes por sua vez elegiam os deputados.

2.1.2 Império Independente (1822-1889)

A segunda fase foi a do Império independente, a partir da declaração da independência do Brasil, em 1822. Dom Pedro I convocou eleições para a Assembleia Geral Constituinte e Legislativa. (BRASIL, 2014)

Seguindo o legado do sistema de níveis português o sistema de votação foi adaptado em apenas dois graus: quem recebia salários e soldos não podia votar em primeiro grau, e para votar em segundo grau era necessário possuir decente subsistência por emprego, indústria ou bens.

Este sistema considerava os eleitores e candidatos como personagens, a saber: eleitor de primeiro grau, eleitor de primeiro grau liberto, eleitor de primeiro grau criminoso, eleitor de primeiro e segundo graus estrangeiro. Já para ser deputado era necessário possuir renda mínima de 400 mil réis e ser católico (professar a religião do estado). Para ser senador era necessário ter renda mínima de 800 mil réis anuais, ser cidadão brasileiro, ter 40 anos ou mais de idade (exceção aos príncipes da Casa Imperial, 25 anos), ser “pessoa de saber, capacidade e virtudes, com preferência de ter feitos serviços à Pátria” - Constituição de 1824, art. 45 (Nogueira, 2012).

Neste período houveram muitas inovações importantes que alteraram a legislação que regulamentava as eleições de tal forma a garantir a maioria ao partido da situação, diferentes sistemas de listas e votos distritais, dentre outras. Em 1875, com a introdução da Lei Saraiva, o voto foi proibido aos analfabetos. (Brasil, 2014)

2.1.3 Primeira República (1889-1930)

A constituição de 1891 definiu como cidadãos os brasileiros natos ou naturalizados. Poderiam votar apenas os cidadãos maiores de 21 anos alistados na forma da lei. Na primeira eleição para presidente da República, em 1894, para possuir direito ao voto, era necessário ser homem e alfabetizado. Com isto, apenas 2,2% da população conseguiu votar neste pleito.

Neste período, houve algum progresso na legislação eleitoral, sem com isto permitir que analfabetos e mulheres pudessem votar, mas trouxe o fim do voto censitário¹ e a permanência do voto direto. Ironicamente nenhuma lei conseguiu extirpar de vez as fraudes eleitorais.

¹ Sistema em que o critério para ter direito de ser eleitor ou candidato é baseado em uma informação tipicamente obtida em um senso demográfico, como, por exemplo: sexo, religião, etnia, etc. No Brasil estavam aptos a votar e concorrer às eleições era restrito aos chamados “homens bons”: indivíduos descendentes de famílias abastadas, com títulos nobiliárquicos ou donos de muitas propriedades; excluindo todo o restante da população, principalmente os pobres e desvalidos

Através do decreto nº200-A, de 1890, foi definido o regulamento para a eleição do Congresso Nacional Constituinte, mantido o voto direto e incluído todos os eleitores já qualificados pela Lei Saraiva. O maior problema era a falta de transparência do processo, já que a qualificação dos eleitores e todo o processo de alistamento ficava completamente nas mãos das autoridades locais.

Em 23 de junho do mesmo ano foi promulgado o Regulamento Alvim (decreto nº511) que regulamentou a eleição para o Congresso Nacional Constituinte. Dentre as novidades, determinou as condições de elegibilidade para cada cargo: para ser eleito deputado era necessário ter mais de 7 anos de cidadania brasileira e para ser eleito senador ter mais que 35 anos de idade. Trouxe também de volta a lista completa por estado e a revogação do voto distrital e a legado à mesa eleitoral receber e apurar os votos. Assim, o Governo Provisório garantiu a maioria no Congresso Nacional.

A Constituição de 1891 instituiu o sistema presidencialista. O presidente e seu vice deveriam ser eleitos pela maioria absoluta dos votos, em sufrágio direto. O Congresso Nacional passou a ser responsável pelo processo eleitoral para os cargos federais e os estados pelos processos eleitorais nos cargos estaduais e municipais. O que instituiu uma descentralização do processo eleitoral como um todo e permitiu uma diversidade de modelos de processos.

Em 7 de dezembro de 1896 foi promulgada a lei nº426 que passou a permitir o voto a descoberto, e institucionalizou o que depois passou a ser chamado de voto cabresto, já que permitia que o eleitor levasse uma segunda via de seu voto e o proclamasse publicamente no momento de votação, permitindo um total controle da votação por quem quer que pressionasse o eleitor.

A lei Rosa e Silva (1.269 de 15 de novembro de 1904) revogou toda a legislação eleitoral anterior, acrescentou o voto cumulativo mantendo o voto distrital (com cinco deputados por distrito) e o limitado, e ainda possibilitando o voto a descoberto (tal qual a Lei nº426 de 1896). A formação das comissões de alistamento passaram a ser formadas por: um juiz de direito, dois dos maiores contribuintes de imposto predial, dois dos maiores contribuintes sobre propriedade rural e três cidadãos eleitos pelo município. Estes deveriam saber ler e escrever, provando isto em livro específico as suas informações pessoais: nome, filiação, idade, profissão e residência.

Com a lei nº 3.139 de 2 de agosto de 1916, somente o Poder Judiciário passou a ser o responsável pelo alistamento eleitoral. Passaram a ser obrigatória as comprovações de idade, capacidade de subsistência, residência no município por mais de 2 meses e prova de saber ler e escrever.

O período que precede 1930 foi apelidado de idade das trevas eleitorais no Brasil, caracterizado pelo revezamento de elites tradicionais no poder, quando as

eleições eram mera legitimação popular de alguns dos mesmos políticos, e era comum o uso de práticas fraudulentas (como o voto cabresto, onde jagunços de coronéis conduziam eleitores à força às seções de votação).

2.1.4 Refundação da República (1930-1937)

Uma das bandeiras do movimento de 1930 foi moralizar as eleições, porquanto tão público e notório eram as fraudes desde o império). Assim, em 1932 o poder judiciário centralizou as atividades eleitorais, sendo este o único responsável por todo o processo eleitoral, ato prático que criou a justiça eleitoral, instituída em definitivo com a constituição de 1934.

Neste mesmo ano, foi criado o primeiro código eleitoral brasileiro, que trouxe inovações vitais para a democracia, como o voto secreto e o voto para as mulheres, mesmo que ainda facultativo. Instituiu também o sistema representativo proporcional, e regulou em todo o país as eleições federais, estaduais e municipais; obrigando o pré registro de todas as candidaturas, mas ainda permitindo a inscrição de candidatos sem partido. Infelizmente ainda não permitiu que analfabetos votassem, mas trouxe uma inovação que foi o primeiro passo para a urna eletrônica: a autorização para se usar uma máquina de votar.

Houve uma reforma no Código Eleitoral de 1935 (lei Nº 48 de 4 de maio de 1935) que trouxe mais algumas inovações, dentre elas, rebaixou a idade mínima do voto (de 21) para 18 anos, tornou obrigatório o voto para mulheres em função pública remunerada e limitou a candidatura avulsa. Este código reformado nunca entrou em ação já que no período de 1937 a 1945, Getúlio Vargas comandou o Estado Novo, que fez extinguir a Justiça Eleitoral, abolir dos partidos políticos, instituir de eleições indiretas para presidente com mandato de seis anos, além de dissolver as casas legislativas e nomear interventores nos estados.

2.1.5 República Democrática(1945-1964)

Em 1945, com novo golpe, os militares destituíram Getúlio e sua constituição, reinstalando a Justiça Eleitoral, desta vez em definitivo. O ato se fez através da promulgação de um Código Eleitoral (lei nº 7.586 - Lei Agamenon) e a principal inovação foi a obrigatoriedade dos candidatos a filiação partidária e penalidades para o eleitor que tentasse se inscrever mais de uma vez ou mesmo votar se passando por outro eleitor; também manteve a inscrição do eleitor *ex officio*².

² Realizado por imperativo legal ou em razão do cargo ou da função (diz-se de ato). Ação legal deliberativa que seja atribuição do cargo ou função de quem o exerce

Outras inovações surgiram em 1950 (Lei nº 1.165) quando foram definidos os critérios para a padronização de cédulas e o fim do alistamento *ex officio*. Mais inovações foram instituídas em 1955 (Lei nº 2.250) com a folha individual de votação, quando se fixou o eleitor na mesma seção eleitoral. Outra lei (nº 2.582) criou a cédula de votação oficial.

No período de 1945 a 1964 passaram pelo cargo máximo do executivo brasileiro nove presidentes e houveram tentativas de golpe militar em 1951, 1954 e 1961. Os esforços do governo, dos partidos e dos candidatos se concentraram em coagir o cidadão a acreditar que faz parte de seu dever-poder unicamente comparecer às urnas, deixando de lado outras práticas políticas como manifestações populares e atividades afins.

As inovações seguem: em 1955 surgiu a primeira cédula eleitoral oficial, com a eleição de Juscelino Kubistcheck. A cédula permitia votar para presidente e vice presidente de chapas distintas. Já em 1958, Sócrates Ricardo Puntel criou uma “urna” mecânica, com teclas e duas réguas para mostrar os cargos. A invenção foi reconhecida mas não chegou a ser adotada pela Justiça Eleitoral.

2.1.6 Regime Militar (1964-1985)

De 1964 a 1985, conhecido como período militar, a legislação imposta através dos atos institucionais (1, 2, 3, 4, 5 e 14) permitiu uma série de retrocessos democráticos, desde instituir as eleições indiretas, passando pelo fechamento do congresso à vontade do presidente, cassação dos direitos políticos do cidadão, proibição de *habeas corpus* para presos políticos estabelecendo uma verdadeira perseguição, até possibilitar a pena de morte para casos de guerra externa, psicológica ou adversa ou revolucionária ou subversiva. (Brasil, 2014)

Em 15 de julho de 1965 é instituído o Código Eleitoral Brasileiro, a partir do Ato Institucional de 9 de abril de 1964. (BRASIL, 2018)

2.1.7 Nova República (1985 até hoje)

A partir de 1974, o Presidente Geisel iniciou uma abertura lenta e gradual, iniciando com o fim do Ato Institucional nº 5, o fim da censura prévia, a liberação do *habeas corpus*, dentre outras ações. Este mesmo Código Eleitoral de 1964 passou por várias modificações até o texto vigente hoje. (BRASIL, 2018)

Em 1986 foi dado o primeiro passo em direção à informatização: iniciou-se o cadastramento de 70 milhões de eleitores, o que permitiria posteriormente o funcionamento da urna eletrônica. Todo o esforço não foi em tempo suficiente para

evitar que as eleições de 1988 ainda fosse manual, com uma escrutinação demorada e cheia de oportunidades de fraude. Apesar disso, a novidade foi o retorno do voto para os analfabetos e o voto direto para algumas categorias municipais.

Antes do surgimento da urna eletrônica, na cidade de Brusque, em Santa Catarina, o juiz Carlos Prudência pretendia fazer a apuração de forma mais rápida, usando um computador. Embora a iniciativa tenha sido bem intencionada, não foi autorizada pelo TRE local. No entanto, no primeiro turno eleitoral de 1989, Brusque virou destaque nacional por ter registrado sua apuração antes de todo o resto do Brasil. O sistema não era oficial e os votos foram contados também manualmente, o que mostrou o mesmo resultado. (KLEINA, 2019)

O ministro Sepúlveda Pertence, em 1994, inaugurou o processamento eletrônico dos resultados com a Justiça Eleitoral. Que consiste de um sistema de computadores e sistemas centralizados e uma rede de recepção de dados. Atualmente este é o recurso que possibilita a divulgação rápida dos resultados.

Em 1995, assume o projeto no TSE uma equipe de especialistas formada por Paula Nakaya, Psvaldo Imamura, Mauro Hashioka, Antônio Marcondes e Giuseppe Janino. Três deles eram técnicos do INPE (Instituto Nacional de Pesquisas Espaciais), um analista do CTA (Centro Técnico da Aeronáutica) e um servidor do TSE Janino, hoje atual secretário em tecnologia de eleições. Esta equipe criou o Coletor Eletrônico de Voto (CEV) que mais tarde foi batizado de “Urna Eletrônica”.

A primeira oportunidade que a urna eletrônica foi usada ocorreu em 1996 em apenas 57 cidades abrangendo um total aproximado de duzentos mil eleitores. (KLEINA, 2019)

2.1.8 A Urna Eletrônica: Hardware e Software

O que é conhecido hoje como Urna Eletrônica é um dispositivo eletro eletrônico composto de uma tela, teclado e CPU (unidade central de processamento) em uma só máquina. Quando de sua concepção, dentro os requisitos para sua criação estavam a simplicidade de uso. Assim, para a votação foi escolhido um modelo de teclado numérico como o de uma calculadora. (KLEINA, 2019)

Nos primeiros modelos o dispositivo já contava com uma tecla votar em branco, mas não tinha a tecla para votar nulo. Assim, e eleitor precisava digitar um número de candidato inválido se assim o desejasse fazer.

A impressão do voto, questão de bastante controversa (RAMALHO, 2018; ARANHA, 2018; SHALDERS, 2018; e SIMPLICIO, 2018), foi implementada na primeira versão da urna de forma a manter o legado de poder escrutinar os votos em caso

houver algum questionamento sobre o processo eleitoral. Quem defende que o voto não deve ser impresso acredita que a impressão seria um retrocesso que mais traria novas oportunidades de antigas fraudes do que mais segurança à urna.

Embora a muitos defendam que o voto precisa ser impresso em paralelo para haver algo palpável para uma escrutinação manual, como a organização não governamental IDEA (*Institute for Democracy and Electoral Assistance*), inacreditavelmente o plenário do Superior Tribunal de Justiça do Brasil (STF) aprovou, por 8 votos a 2, a suspensão da impressão dos votos nas eleições de 2018, com a chancela de Procuradoria Geral da República (PGR) alegando que a impressão do voto comprometeria seu sigilo. (BRASIL, VI, 2018 e RAMALHO, 2018).

Por fim, até o momento, evidenciando o caráter altamente dinâmico da legislação acerca do tema, o projeto de emenda constitucional (PEC 135/19) favorável à impressão dos votos recentemente obteve a aprovação de legalidade constitucional na Comissão de Constituição e Justiça da Câmara dos Deputados (BRASIL, PC, 2019), seguindo à frente em seus trâmites legais.

No primeiro modelo da urna havia um dispositivo para imprimir o voto e o depositar em uma urna plástica. Isto foi abolido em 1998, retornou em 2002, e foi substituído em 2004 pelo registro digital.

A partir das eleições municipais de 2000 todos os eleitores passaram, a usar a urna eletrônica, que possuíam fones de ouvido para deficientes auditivos.

Ao longo dos anos, a cada pleito, o TSE apresentou novos modelos de urnas, que usaram sistemas operacionais diversos, são eles:

ANO	MODELO	SISTEMA OPERACIONAL (SO)
1998	UE98	VirtuOS
2000	UE2000	VirtuOS
2002	UE2002	Windows CE
2004	UE2004	Windows CE
2006	UE2006	Windows CE
2008	UE2006	Linux próprio
2010	UE2006	Linux próprio
2014	UE2006	Linux próprio
2018	UE2006	Linux próprio

Tabela 1 – Modelos de urnas eletrônicas e SO. (BRASIL, MU, 2019 e BRASIL, TT, 2018).

Apesar desta informação oficial, nos códigos dos programas de assinaturas de chaves criptográficas (BRASIL, MU, 2019) constam os seguintes modelos: ue2006, ue2008, ue2009, ue2010, ue2011, ue2013, ue2015.

VirtuOS é um sistema operacional desenvolvido pela MICROBASE (2019) com suporte no Brasil desde 1983. Este sistema é multiusuário e multitarefa, projetado exclusivamente para micro computadores de 16 e 32 bits compatíveis com o padrão PC AT.

Windows CE é um sistema operacional produzido pela MICROSOFT (2019) aberto, escalável e de 32 bits para dispositivos portáteis, tablets e sistemas embarcados, compatível com os processadores Intel x86 e compatíveis, MIPxS, ARM, e SuperH Hitachi.

O Linux é um sistema operacional de código aberto, o que permitiu que a equipe do TSE desenvolvesse sua própria versão específica para atingir todos os requisitos, principalmente de segurança, que a utilização necessita. (BRASIL, MU, 2019)

As próximas inovações na urna eletrônica foram: a biometria, capacidade para o voto em trânsito, e a possibilidade de uso de um aplicativo do TSE em substituição ao título de eleitor, em 2008, 2010 e 2018, respectivamente. (KLEINA, 2019)

O TSE publicou (Brasil, EU, 2019) o nome de alguns dos sistemas que fazem a urna eletrônica funcionar, dentre eles: Sistema VOTA encarretado de coletar e apurar os votos de uma seção eleitoral; Sistema SCUE, cuja função é permitir a carga da urna responsável e executar a preparação e a instalação do sistema operacional, software e dados de eleição nas urnas; o Sistema ATUE que possibilita o (auto) teste e validação do funcionamento das urnas antes do dia da eleição; e o Sistema GEDAI-UE que é um gerenciador de dados, aplicativos e interface com a urna, que permite que as equipes dos cartórios eleitorais e dos Tribunais Regionais Eleitorais (TREs) façam a carga de sistemas nas urnas eletrônicas.

2.2 O Processo de Votação Eletrônico

Com início da informatização eleitoral no Brasil passou a ser possível o entendimento do cidadão brasileiro de como o processo é extenso e complexo, e exige que cada participante desempenhe seu papel. Ao eleitor coube o cadastramento digital, e mais recentemente o cadastramento biométrico. Aos milhares de voluntários que trabalham nas eleições, coube aprender e executar todos os passos com exatidão. A equipe de desenvolvimento dos sistemas foi quem idealizou, planejou e programou cada um dos sistemas, tanto o de votação, trânsito de dados e totalização, como os de infraestrutura, bem como sua implantação e funcionamento.

“A partir do cadastro único de eleitores e do registro de candidaturas, os dados são preparados. As mídias são geradas e as urnas

eletrônicas carregadas. São preparados três tipos distintos de urnas: urna de votação, urna de contingência e mesa receptora de justificativa eleitoral. A identificação de todas as urnas preparadas é armazenada na tabela de correspondência. Após a votação, o boletim de urna é transmitido para o sistema de totalização e os resultados consolidados são divulgados. As mídias são geradas e as urnas eletrônicas carregadas. São preparados três tipos distintos de urnas: urna de votação, urna de contingência e mesa receptora de justificativa eleitoral. A identificação de todas as urnas preparadas é armazenada na tabela de correspondência. Após a votação, o boletim de urna é transmitido para o sistema de totalização e os resultados consolidados são divulgados.” (BRASIL, PE, 2019).

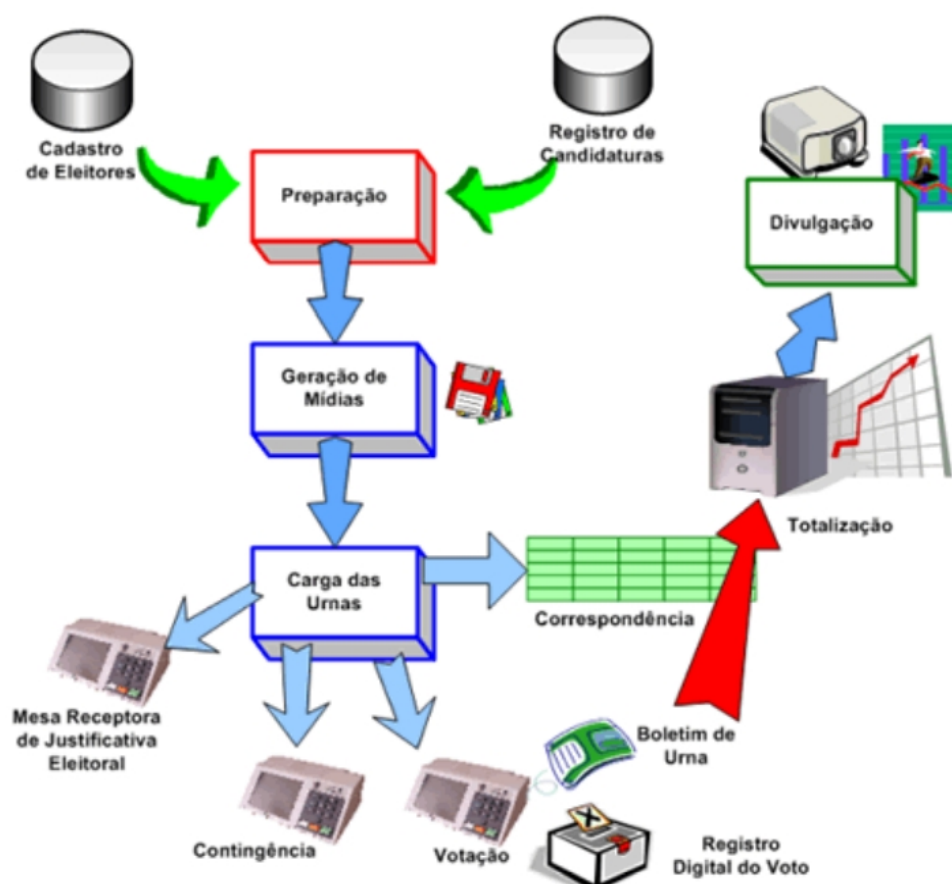


Figura 2.2.1 – Processo Eleitoral Eletrônico. Fonte: BRASIL, PE, 2019.

Dentre todas as fases que compõem o processo, uma das mais passíveis de questionamento é a do registro do voto, por ser o ponto a atividade fim da urna. Segundo o TSE (BRASIL, RD, 2019), o registro digital do voto (RDV), criado em 2003, acontece em duas tabelas de dados, uma com o RDV propriamente dito e outra com a totalização, que ao encerramento da sessão dará origem ao boletim de urna.

Na tabela do RDV, cada linha (ou registro) guarda as informações do evento de votação de um eleitor, com as colunas (ou campos) do número do candidato votado, ou um indicador de branco ou nulo. Por exemplo, no pleito de 2018 as colunas foram: “deputado federal”, “deputado estadual”, “senador”, “governador” e “presidente”. Esta

tabela é que permite a recontagem dos votos de uma urna e ela é o único local onde esta informação fica armazenada. Não há uma coluna com qualquer identificação do eleitor.

Na tabela de totalização há uma linha (ou registro) para cada candidato e um para brancos e nulos de cada cargo, contendo o número do candidato (ou uma identificação de branco ou nulo) e a quantidade de votos na urna.

Note que fazer uma recontagem de votos digitais implica somente em retotalizar os votos da tabela do RDV e gerar novamente a tabela de totalização, que boletim de urna. Como não há voto impresso, não há outra fonte de verificação da votação a não ser a tabela do RDV, e este é um dos principais questionamento de quem se opõe a urna eletrônica, como ARANHA (2018).

Estas duas tabelas são encriptadas com a tecnologia de chaves assimétricas, onde ao menos uma das chaves é pública para permitir a assinatura de autenticidade do voto. (BRASIL, FA, 2018)

Ao término da sessão, o boletim de urna (BU) é submetido ao processamento que levará à totalização dos votos de cada urna.

O sistema que totaliza os votos passa pelos seguintes passos de verificação conforme figura 2.2.2:

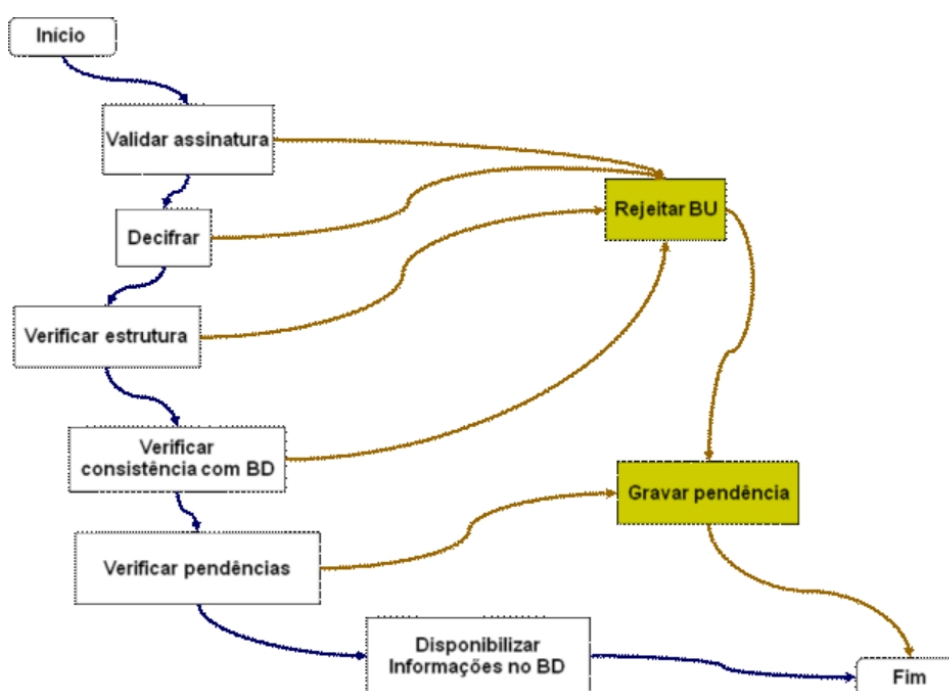


Figura 2.2.2 – Processamento do Boletim de Urna. Fonte: BRASIL, BU, 2019.

O boletim de urna é processado conforme os passos:

- A validação da compatibilidade da chave pública de assinatura digital do boletim de urna

- com a chave privada do Totalizador;
- A decriptografia do boletim de urna de forma segmentada;
- A leitura do boletim de urna decriptografado;
- O armazenamento do boletim de urna criptografado e decriptografado.

(Fonte: BRASIL, BU, 2019)

O juiz eleitoral é quem dá o tratamento apropriado caso de houver alguma pendência, ou mesmo inconformidade nos dados. Havendo conformidade em todas as verificações, os dados são armazenados no banco de dados e processados pelo sistema. (BRASIL, BU, 2019)

A urna gera também uma tabela de *log* onde fica registrado cada evento que muda o estado da urna a saber: data, hora, sistema que gerou o registro, e descrição do evento, sem a identificação do eleitor ou mesmo dos números votados. Por exemplo:

```
05/10/2008 07:05:40 - GAP - iniciado: Oficial
05/10/2008 07:05:40 - GAP - versão: 1.12.20.0
05/10/2008 07:05:45 - GAP - Urna operando em turno 1 Turno
05/10/2008 07:05:45 - GAP - Número Interno da UE 00546735
05/10/2008 07:05:47 - GAP - Modelo da Urna: UE2000
05/10/2008 08:09:14 - Vota - Aguardando digitação do título
05/10/2008 08:09:39 - Vota - Título digitado pelo mesário
05/10/2008 08:09:40 - Vota - Título digitado pelo mesário é inválido
05/10/2008 08:09:48 - Vota - Aguardando digitação do título
05/10/2008 08:10:59 - Vota - Título digitado pelo mesário
05/10/2008 08:11:05 - Vota - Eleitor habilitado
05/10/2008 08:11:13 - Vota - Voto confirmado para Vereador
05/10/2008 08:11:19 - Vota - Voto confirmado para Prefeito
05/10/2008 08:11:21 - Vota - Voto computado
05/10/2008 08:11:21 - Vota - Aguardando digitação do título
```

Figura 2.2.3 – Exemplo de arquivo de *log* de Urna. Fonte: BRASIL, FL, 2018.

2.3 A Segurança da Urna Eletrônica

Segurança é um termo bastante abrangente que inclui uma dinâmica constante para mantê-la. De forma geral, algo está seguro se o proprietário está ao menos um passo à frente do criminoso. É importante ter em mente que a segurança de recursos de informação tem 3 componentes:

- Confidencialidade: proteção contra acesso não autorizado;
- Integridade: proteção contra estragos;
- Disponibilidade: proteção contra interferências no seu funcionamento.

COULOURIS (2007)

Assim, para que se possa dizer que a urna eletrônica é segura ou mesmo o sistema eleitoral o seja, é necessário que: não haja acesso não autorizado; não seja possível estragar nenhuma parte do sistema (seu *software* ou *hardware*); e que as informações estejam disponíveis a quem deve, e as operações estejam acessíveis e

funcionais conforme planejado. O que torna a tarefa de segurança do Tribunal Superior Eleitoral (TSE) bem mais complexa.

Ciente da “corrida de gatos e ratos” que são as incumbências de segurança, o TSE está constantemente realizando ações para estar sempre à frente dos ataques no sentido de mitigar toda e qualquer possibilidade de fraude na urna eletrônica, tentando sempre provar que é muito segura. (BRASIL, US, 2019). Dentre estas ações, inúmeras características foram adicionadas tanto à urna quanto como a todo o processo eleitoral, compondo várias camadas de segurança tanto tecnológicas como sócio-administrativas; listadas por um dos mentores da urna, e hoje secretário de tecnologia da informação do TSE, Giuseppe Janino. (BRASIL, MV, 2019):

- A urna eletrônica possui lacre físico em todas as áreas de acesso ao seu interior, para garantir a inviolabilidade física, além de mais de 30 barreiras de segurança, algumas descritas a seguir;
- Sistema de controle de versões: todo o software da urna é mantido em ferramenta de controle de versões, permitindo saber quem realizou modificações, quais foram elas e quando foram realizadas;
- Testes de software por várias equipes: todo o software da urna é testado, antes e depois da lacração, por várias equipes: pela própria equipe de desenvolvimento, por equipe dedicada no TSE e pelos TREs;
- Seis meses de abertura do código-fonte: pelo período de seis meses que antecedem a Cerimônia de Lacração, várias entidades podem auditar todo o código-fonte dos sistemas eleitorais;
- Teste Público de Segurança: o processo eletrônico é aberto à comunidade em geral para inspeção de código e exercício do hardware e do software da urna em busca de falhas;
- Cerimônia de Lacração e Assinatura Digital: em cerimônia pública no TSE, os sistemas eleitorais são compilados e assinados digitalmente;
- Cerimônia de Geração de Mídias, Carga e Lacre das Urnas: em cerimônia pública nos TREs ou cartórios eleitorais, todas as mídias da urna são geradas, o software com os dados de eleitores e candidatos é carregado na urna, que, por fim, recebe lacres físicos numerados;
- Tabela de correspondência: ao final do processo de carga, a urna gera um código único (código da carga) que faz a correspondência entre determinada urna e a sua preparação para eleição. Somente urnas com código de carga válido poderão ter resultados totalizados;
- Cadeia de segurança em hardware: a urna eletrônica possui um dispositivo de segurança em hardware que valida todo o software da cadeia de inicialização da urna (BIOS, bootloader e kernel), o que garante que somente o software gerado durante a Cerimônia de Lacração e Assinatura Digital pode ser executado na urna;
- Processo de fabricação seguro: as urnas saem da fábrica não operacionais, sendo necessário passar por processo de certificação digital nos TREs ou no TSE para que entrem em funcionamento;
- Projeto de hardware e software dedicados somente à eleição: a urna só possui mecanismos necessários para a votação e apuração de resultados, tanto em hardware quanto em software (não há suporte à comunicação via rede, por exemplo);
- Verificação de assinatura dos aplicativos da urna: todos os aplicativos da UE possuem assinatura digital embutida, verificada pelo kernel antes de sua execução (somente aplicativos gerados durante a Cerimônia de Lacração e Assinatura Digital podem ser executados na urna eletrônica);
- Verificação de assinatura dos dados de eleitores e candidatos: todos os dados que alimentam a urna são protegidos por assinatura digital, gerada pelo software responsável pela guarda original, garantindo que informações de eleitores e candidatos se mantenham íntegras e autênticas;

- Criptografia da biometria do eleitor: os dados biométricos do eleitor são criptografados no software responsável pela sua guarda e são decifrados somente na urna;
- Criptografia da imagem do kernel do Linux: protege o kernel do Linux usado na urna contra engenharia reversa e execução fora da UE;
- Criptografia do sistema de arquivos da urna: protege vários arquivos da UE contra engenharia reversa e cópia indevida;
- Criptografia de chaves da urna: protege as chaves utilizadas pela UE, para que sejam usadas somente por ela;
- Criptografia do Registro Digital do Voto (RDV): protege o arquivo de RDV contra leituras sucessivas ao longo da votação;
- Derivação de chaves na urna: calcula as chaves de criptografia do kernel, do sistema de arquivos, das outras chaves e do RDV, de modo que não é possível obtê-las por engenharia reversa do software ou por análise das mídias;
- Embaralhamento no RDV: os votos são gravados no arquivo de RDV tal como digitados pelo eleitor e embaralhados cargo a cargo entre todos os eleitores, de modo a impedir que se conheça a ordem em que ocorreu a votação e a quebra do sigilo do voto;
- Boletim de urna impresso: antes da gravação de qualquer arquivo de resultado, o Boletim de Urna (BU), documento com resultado da apuração da seção eleitoral, é impresso e torna-se documento público;
- Assinatura de software dos arquivos de resultado: todos os resultados produzidos pela urna são assinados digitalmente, utilizando uma biblioteca desenvolvida pelo Cepesc/Abin;
- Assinatura de hardware dos arquivos de resultado: todos os resultados produzidos pela urna são assinados digitalmente, utilizando também o dispositivo de segurança em hardware (cada urna possui uma chave de assinatura diferente);
- Criptografia do BU: a porção do arquivo de BU que possui os resultados apurados da seção é criptografada na urna e só pode ser decifrada pelo Sistema de Totalização;
- QR Code no BU: no BU impresso, há um QR Code que possui assinatura digital própria, o que permite a obtenção de cópia digital do documento, a oportunidade de comparação com o resultado recebido pelo TSE e até mesmo a totalização paralela dos votos;
- Código verificador do BU: código numérico com token de autenticação que permite a digitação somente de Bus válidos no Sistema de Apuração (procedimento de recuperação de dados realizado em cerimônia pública);
- Votação paralela: após sorteio realizado na véspera da eleição, urnas que seriam utilizadas na eleição são separadas para teste de votação monitorada, para fins de conferência da fidedignidade do resultado apurado pela UE (sorteio e votação realizados em cerimônias públicas);
- Conferência de hash e assinatura digital: em diversos momentos (sobretudo nas cerimônias públicas) é possível fazer a verificação de integridade e autenticidade do software da urna, tanto com o auxílio de software desenvolvido pelo TSE quanto por software desenvolvido por outras entidades;
- Conferência de hash de assinatura digital no dia da eleição: a verificação dos hashes pode ser realizada em urnas instaladas nas seções eleitorais antes do início da votação (UEs sorteadas na véspera);
- Log da urna: todas as operações realizadas pelo software da UE são registradas em arquivo de log, que pode ser entregue aos partidos políticos para averiguação;
- Entrega do RDV: os partidos políticos podem solicitar cópia de todos os arquivos de RDV de todas as urnas, para validação dos resultados totalizados.

Seguindo a própria cartilha de segurança, podemos especular que estas medidas devem ser apenas algumas das publicáveis – para não perder o fator surpresa – dentre outras que só virão à tona caso algum evento de fraude seja descoberto.

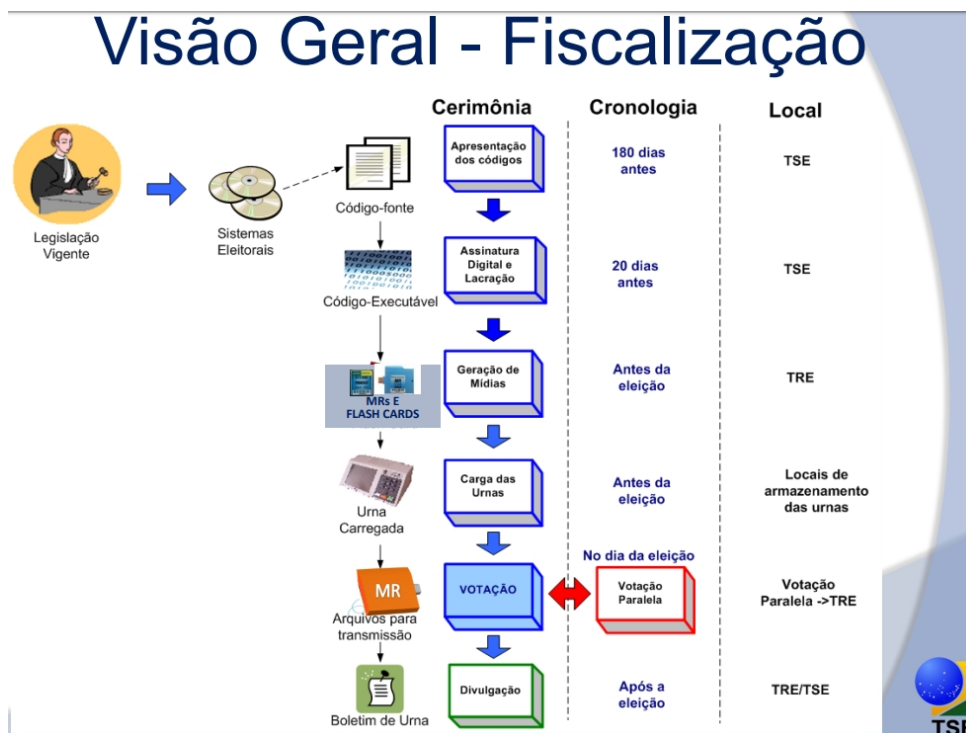


Figura 2.3.1 – Fiscalização do Processo Eleitoral. Fonte: BRASIL, GJ, 2015.

Em 17 de julho de 2015, o poder legislativo federal do Brasil, através da Câmara dos Deputados, compôs uma comissão parlamentar de inquérito sobre crimes cibernéticos. (BRASIL, CC, 2019). Dentre tantas reuniões da comissão, numa delas (a de 03 de novembro de 2015) contou com a participação do secretário de tecnologia da informação do TSE, Giuseppe Janino, para prestar esclarecimentos sobre uma indicação no arquivo de *LOG* de tentativa de executar um programa que não compõe o Ecosistema da Urna Eletrônica (assim chamado pelo TSE), portanto não autorizado, no dia 27 de agosto de 2012, embora tivesse o nome de um programa do Ecosistema, conforme relatou Maria Aparecida da Rocha Cortiz – Advogada, membro do Conselho Multidisciplinar Independente (CMind), que também contribuiu com a comissão nesta ocasião. (BRASIL, NT, 2015).

Podemos observar este incidente de duas formas: a primeira é constatar que a urna tem uma segurança frágil e que se fatos como este descrito pela advogada forem comprovados, então é sim possível fraudar a urna eletrônica com relativa facilidade; a segunda forma é atentar-se para o fato de que o sistema de *LOG* é bastante eficiente e que, graças ao registro por ele gerado, foi possível observar uma tentativa de fraude, o resultado da atuação de uma possível praga virtual, ou, quiçá, uma titubeada de algum operador voluntário.

Mas a advogada trouxe outra informação contundente: além da primeira tentativa de execução de um programa não autorizado na urna eletrônica, houveram outras duas sem sucesso em 27 e 28 de agosto de 2012, por conta de versão incompatível.

Mas no dia 21 de setembro o programa foi rodado com sucesso. No dia 24 de setembro foram geradas as mídias de carga para as demais urnas e este programa rodou no Município de Londrina, no Paraná, nas eleições de 2012. (BRASIL, CD, 2015)

A advogada descobriu, na eleição de 2014, com o ajuda do Professor Diego Aranha (autor de ARANHA, 2018) que a empresa que trabalhou neste programa era a SmartMatic, numa contratação questionável através de um aditivo de contrato. O que ela encontrou, disse à comissão, foi um programa executável inserido no sistema de segurança da urna, como um subsistema do Sistema SIS, o primeiro a ser carregado na urna. Neste programa havia um outro programa chamado “*Inserator*”, capaz de assinar digitalmente programas não oficiais.

Como descrito pelo TSE, somente ele é quem assina os programas, então este código não poderia, de forma alguma, ir para as urnas, mas sim ficar restrito à equipe de programação do TSE. O TSE, na pessoa do Sr. Giuseppe Janino alegou que este programa não foi mais usado depois de 2006. Portanto não foi para as urnas. E acrescentou que este programa não poderia ter fraudado as eleições, porque ele tinha a capacidade somente de inserir assinaturas digitais numa tabela onde ficam as autorizações de execução oficial dos programas autorizados pelo TSE. Esclareceu também que este programa substituíra de forma “doméstica” uma autoridade certificadora externa, que foi contratada em 2006, ressaltando que o programa nunca foi para uma urna, pois faz parte do conjunto de programas que só ficam no TSE.

Por fim, este episódio nem consta no relatório final da CPI (BRASIL, CC, 2019), que não menciona a urna eletrônica nem o dito programa *Inserator*.

Tendo observado história política eleitoral no Brasil e o que foi possível do funcionamento e história da urna eletrônica, pode-se observar as eleições pelo mundo e ferramentas estatísticas usadas para verificar indícios de fraudes.

3 DINÂMICAS DE VOTAÇÃO NAS URNAS

Processos eleitorais compreendem desde o cadastro dos eleitores, candidaturas, votação, totalização, divulgação dos resultados, prestação de contas, logística eleitoral, até a diplomação dos eleitos, com o objetivo de democratizar as escolhas dos representantes de uma sociedade através do voto. Investigar a dinâmica desses processos e detectar regularidades e irregularidades pode contribuir substancialmente visto que o resultado de uma eleição define o destino de uma nação. Nesse sentido, análises estatísticas têm contribuído como uma possibilidade para quantificar e qualificar o nível de confiabilidade dos resultados de eleições.

Decisões individuais são tomadas em termos de um campo “cultural” subjacente, que enviesa localmente a decisão da população de uma determinada região, com correlações de curto alcance. BORGUESI (2010) fez um estudo estatístico das eleições francesas no período de 1992 a 2010, encontrando a distribuição das taxas de comparecimento às urnas estáveis através dos anos bem como uma correlação das taxas de comparecimento ou fração de votos no vencedor decaindo logaritmicamente com a distância entre as cidades.

Pesquisadores mostram um resultado estatisticamente anômalo em duas eleições Russas de 2011 e 2012, que favoreceram o presidente eleito Vladimir Putin. KOBAC, 2012. O mesmo resultado é obtido por KLIMEK (2012) analisando distribuições de votos para eleições de 11 países.

Outro trabalho desse contexto, (BORGUESI, 2012), traz a correlação entre abstenções, votos brancos e nulos em países com votação não obrigatória, determinando um comportamento padrão existente e propondo uma normatização, a qual sugere como regra de comportamento uma “lei fraca” para o processo.

Pesquisadores analisaram 21 eleições em 10 países e encontraram anomalias estatísticas significativas compatível com fraude de eleitores na Rússia de 2007 a 2011, na Venezuela de 2006 a 2013 e em Uganda em 2011, especialmente na Venezuela em 2013, onde as menores seções de votação foram decisivas para o resultado do Eleições presidenciais. Para esta análise, observaram os dados de pequenas seções de votação, e desenvolveram um método estatístico para testar e quantificar até que ponto os resultados de uma determinada eleição exhibe traços de manipulação do eleitor. (JIMENEZ, 2017).

Uma apresentação mais detalhada dessas investigações será abordada nesse capítulo com o intuito de esclarecer os direcionamentos e objetivos que levaram a propor o trabalho apresentado no capítulo 5.

3.1 Correlações Espaciais nas Estatísticas Eleitorais na França

BORGUESI (2010) propôs uma observação simplificada (binária: “SIM” ou “NÃO”) das decisões que os eleitores podem tomar no momento do voto considerando que quase todas as decisões individuais, de mais variadas naturezas, são influenciadas por seus grupos familiares ou de amizades, ou ainda pelas comunidades que pertencem, criando um processo de imitação coletiva que leva a fenômenos de massa (bons ou ruins) como: tendências, modas e bolhas, cujos motivos podem ser de difícil identificação na análise dos resultados.

Neste estudo, para mapear as imitações, BORGUESI (2010), levou em conta tanto a proximidade entre cidades como os eventos dentro de cada cidade, encontrando o que chamou de “campo cultural”, que permite estimar quanta influência a proximidade ou não gera entre seções próximas. Este campo cultural resulta em intenções, convicções ou propensões, seja para votar ou não, ou mesmo votar alinhado com ideologias mais à esquerda ou mais à direita. Foi feita uma análise das correlações espaciais dos padrões de votação, analisando a taxa de comparecimento das eleições nacionais francesas de 1992 a 2010 num total de 13 eventos, sendo 6 eleições com opções binárias, 3 referendos e 3 segundo turnos de eleições presidenciais.

Para cada eleitor i , $S_i = 1$ representa a participação do eleitor e $S_i = 0$ à não participação. Para uma cidade α , o número total de eleitores é N_α ; o número total de votantes é $V_\alpha = \sum_{i=1}^{N_\alpha} S_i$ a taxa de comparecimento é $p_\alpha = V_\alpha/N_\alpha$ e o número total de votos vencedores é W_α . Por conveniência, trabalhou-se com taxas em logaritmos para a participação ou votos vencedores: $\tau_\alpha = \ln(V_\alpha/(N_\alpha - V_\alpha)) = \ln(p_\alpha/(1 - p_\alpha))$ e $\rho = \ln(W_\alpha/(V_\alpha - W_\alpha))$. Cada cidade é caracterizada pelas suas coordenadas espaciais de sua prefeitura, e a distância entre duas cidades (α e β) é definida por $r(\alpha\beta)$, $R_\alpha \cdot r_{\alpha\beta} = |R_\alpha - R_\beta|$.

A figura 3.2.1 mostra a distribuições colapsadas da variável para toda a base de dados:

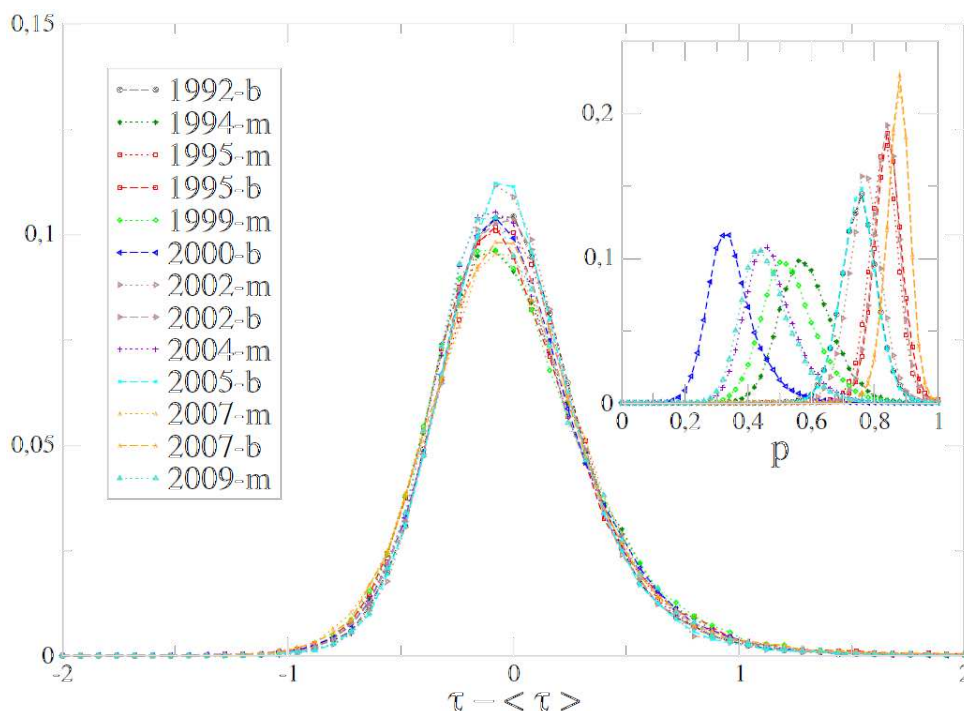


Figura 3.1.1 – Distribuição sobre os municípios de $P(\tau - \langle \tau \rangle)$ para cada eleição, onde τ é a taxa de comparecimento logarítmica e $\langle \tau \rangle$ é a média sobre todos o municípios. A inserção mostra as distribuições da participação direta taxas, p , para cada eleição. Fonte: BORGHESI, 2010.

Note que as distribuições são estáveis, ou seja, quando reescaladas ocorre o colapso de todas elas.

Sendo τ a taxa logarítmica de comparecimento em diferentes cidades, e $P(\tau)d\tau$ a probabilidade de uma dada cidade, independente do seu tamanho ter uma taxa de comparecimento τ em $d\tau$. Assim, a taxa de comparecimento média $m = \langle \tau \rangle$ varia substancialmente entre as eleições. O aspecto da distribuição $\hat{\tau} = \tau - \langle \tau \rangle$ é notavelmente constante, mesmo reescalando para a raiz quadrada de σ . Os primeiros 3 acumulados de $P(\tau)$ são os mesmos para todas as eleições.

O teste de Kolmogorov-Smirnov, não rejeita a hipótese das distribuições nos anos $P(\hat{\tau}/\sigma)$ serem a mesma.

Observou-se também que as distribuições τ que aparecem na figura 3.1.1 mostram curvas Gaussianas, com diferentes assimetrias e curtoses. Verificou também que a estatística τ depende do tamanho da cidade; que tanto a média como a variância da distribuição condicional $P(\tau|N)$ decai com N ; e que cidades menores têm taxas médias de comparecimento visivelmente altas, (Figura 3.1.2).

O que é esperado por um processo simples binomial, onde se prevê que $\sigma_{\tau}^2(N) = 1/Np(1-p) + \sum^2$, onde p é a (N -dependente) taxa de comparecimento, enquanto \sum^2 é a “verdadeira” variância da taxa de comparecimento.

Apenas assumindo-se que \sum^2 é N -independente, em cada caso onde a variância varia significativamente mais rápido (em fator 2 ou mais) que uma predição binomial simples que pressupõe eleitores independentes.

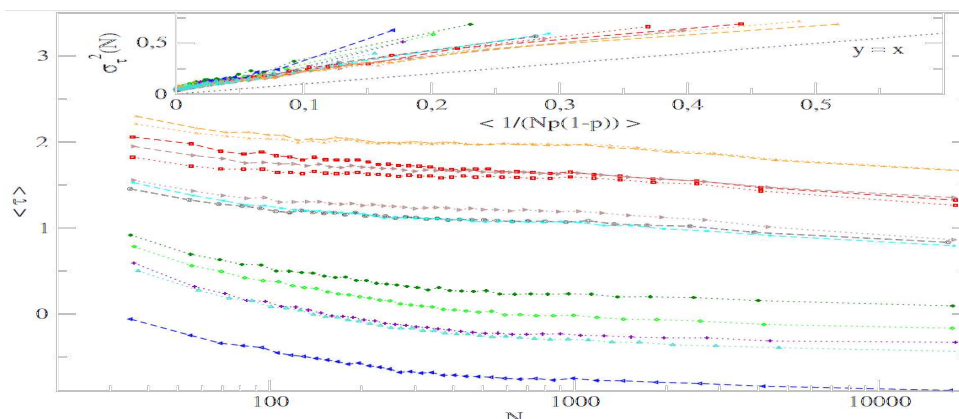


Figura 3.1.2 – Média (figura principal) e variância (figura menor) da distribuição condicional $P(\tau|N)$. Estas quantidades foram obtidas como médias das porções de dados com o tamanho de um milhar de cidades. Fonte: BORGHESI, 2010.

A possível interpretação é que os votos de diferentes indivíduos não são independentes dentro da mesma cidade, conduzindo a um valor binomial efetivamente pequeno de N . BORGHESI (2010) sugere o seguinte: consideremos que dentro de uma família, ou grupo de amigos próximos a decisão de votar ou não é exatamente a mesma. Se o tamanho da distribuição destes grupos de “clones” é $Q(s)$, então é fácil de mostrar que para um N suficientemente grande, essas correlações na decisão equivalem a substituir N por $N\langle s \rangle_Q / \langle s^2 \rangle \leq N$.

Assim, BORGHESI (2010) destaca num primeiro momento que as taxas de participação nas eleições francesas são estáveis ao longo do tempo, uma vez que a taxa de participação média é levada em consideração. A dependência do tamanho da variação da taxa de participação pode ter sugerido alguma influência, descrito como efeito de “pastoreio” intra-cidade. Mesmo assim, BORGHESI (2010), acredita que os dados são mais consistentes com pequenas cidades com maior dispersão em vieses idiossincráticos locais, sem ainda vislumbrar um argumento convincente para o porquê disso acontecer. Uma explicação poderia ser a diferença sistemática no contexto sociocultural da população das grandes cidades em comparação com o das pequenas cidades.

Foi verificado se há correlação espacial entre as taxas de comparecimento usando a função de correlação

$$C_\tau(r) = \frac{\langle (\tau_\alpha - m_\alpha) \rangle |_{r_{\alpha\beta}=r}}{\langle (\tau_\alpha - m_\alpha)^2 \rangle} \quad (3.1)$$

onde m_α é a média de τ sobre todas as cidades com tamanho na mesma janela (*bin*) que o de α para remover correlações espaciais sistemáticas entre os tamanhos das cidades, o que pode ser visto na Figura 3.1.3.

Note que para todo o conjunto de eleições (França 1992 a 2010), a correlação espacial das taxas de participação ou da fração dos votos vencedores decai logaritmicamente com a distância entre as cidades. (BORGUESI, 2010)

O resultado central desse trabalho é que para todas as eleições, $C_\tau(r)$ é correlacionado a longa distância entre as cidades, com decaimento logarítmico para $0 < r < L$, $\tau(r) = -\lambda^2 \ln(r/L)$ e $C_\tau(r > L) \approx 0$.

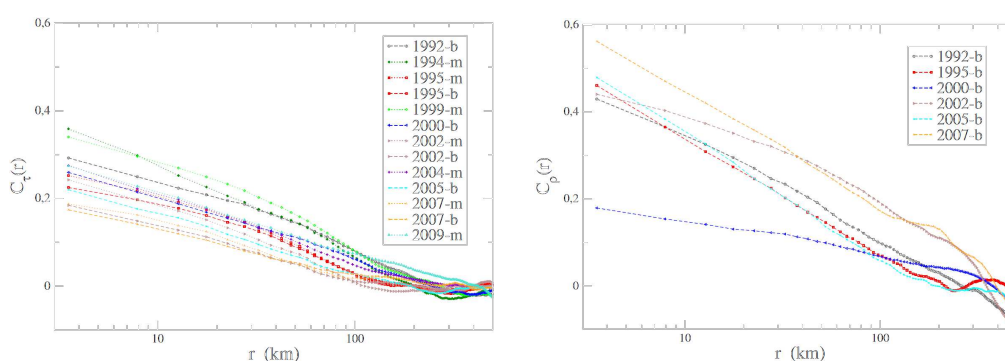


Figura 3.1.3 – Correlações espaciais $C_\tau(r)$ das taxas de participação logarítmica (esquerda) e $C_\rho(r)$ dos votos vencedores (direita), mostrando uma dependência logarítmica em r com uma distância de corte $L \approx 300$ km. A inclinação logarítmica média λ^2 é $\approx 0,065 \pm 0,01$ para taxas de participação e $0,11 \pm 0,02$ para votos vencedores. Fonte: BORGUESI, 2010.

O lento decaimento das correlações é característico de um campo aleatório difusivo em duas dimensões.

Com base nessas observações empíricas e na analogia com a equação de difusão aleatória bidimensional, BORGUESI (2010) propôs que decisões individuais podem ser racionalizadas em termos de um campo “cultural” subjacente que influencia localmente a decisão da população de uma determinada região, além de um campo idiossincrático, dependente da cidade, com correlações de curto alcance.

BORGUESI (2010) propõe o modelo dado pela equação na classe de universalidade da equação de difusão aleatória:

$$\frac{\partial \phi(R, t)}{\partial t} = \nabla \cdot [D(R) \nabla \phi(R, t)] + \eta(R, t) + \nu(R) F(t) \quad (3.2)$$

sabendo que a solução é dada por:

$$C_\phi(r) = \frac{\langle \phi(r) \phi(0) \rangle}{\langle \phi(0)^2 \rangle} \approx -A^2 \ln \frac{r}{L}, \ell \ll r \ll L \quad (3.3)$$

onde ℓ_ϕ é um ponto de corte em pequena escala e L é o tamanho linear do sistema como obtido para os dados empíricos.

3.2 Detecção Estatística de Irregularidades Eleitorais Sistemáticas

Em tempos de eleição, é natural que o eleitor esteja atento aos percentuais em que este ou aquele candidato aparecem à frente ou atrás nas pesquisas eleitorais, ou seja, observe apenas os votos válidos. Mas apenas uma pequena parcela dos eleitores dá importância aos votos não válidos: os votos nulos, os votos em brancos e as abstenções.

KLIMEK (2012) observou os dados dos resultados das eleições em 11 países (eleições parlamentares na Áustria, Canadá, República Checa, Finlândia, Rússia 2011, Suíça e Espanha, eleições parlamentares europeias na Polônia, e eleições presidenciais na França, Romênia, Rússia 2012 e Uganda), comparando o número de eleitores que compareceram às urnas (t) com o número de votos para o candidato vencedor (v). Em todos os países e pleitos observados, o comparecimento às urnas é facultativo e sem ônus.

Foram construídos histogramas bidimensionais com cada uma das eleições da base de dados, chamados de “impressão digital padronizada” da eleição. A intensidade da cor indica o número de urnas com aquela quantidade de comparecimento e votos para o vencedor. O resultado é mostrado na Figura 3.2.1

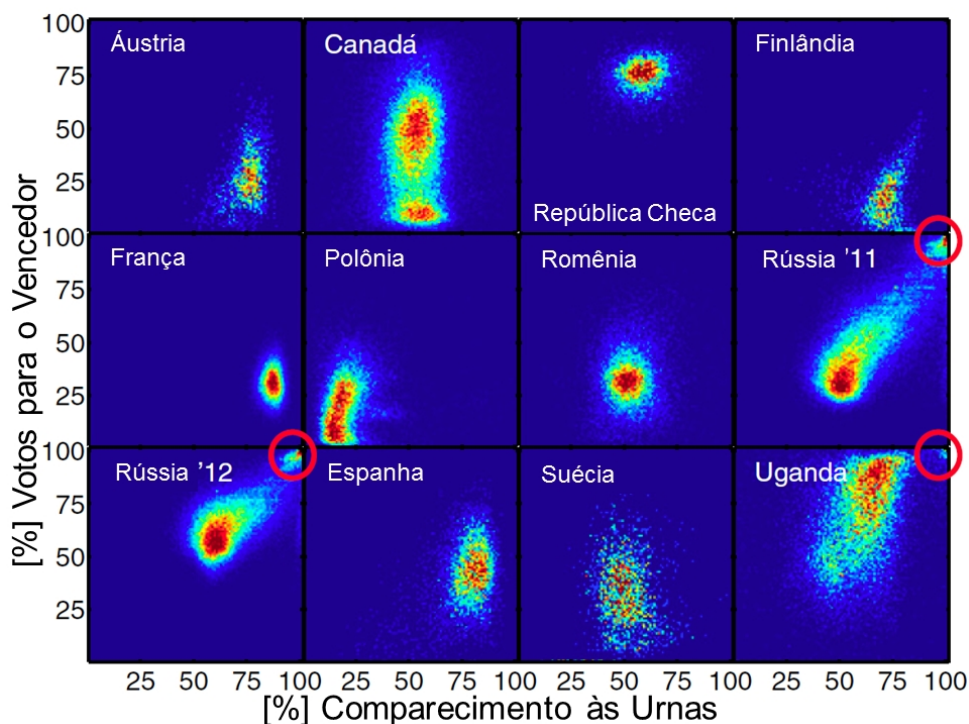


Figura 3.2.1 – Impressões Digitais das Eleições: histogramas onde o eixo x traz o comparecimento às urnas e no eixo y o percentual de votos para o partido (ou candidato) vencedor em diferentes eleições recentes em diferentes países: Áustria, Canadá, República Checa, Finlândia, França, Polônia, Romênia, Rússia 2011, Rússia 2012, Espanha, Suíça e Uganda. As cores mais avermelhadas mostram maior número de ocorrências. Fonte: KLIMEK, (2012).

Para criar a figura 3.2.1, os dados foram utilizados da seguinte forma: seja N_i o total de votantes em cada sessão i , e W_i o número de votos para o vencedor. Optou-se por calcular uma taxa de votação (ν_i) dada pelo logaritmo da razão entre todos os votos que não foram para o vencedor, pelo número de votos para o vencedor, para cada i .

$$\nu_i = \log \frac{N_i - W_i}{W_i} \quad (3.4)$$

Na figura 3.2.1, podemos notar que nas duas eleições na Rússia e na eleição em Uganda que a distribuição de votos e da taxa de comparecimento são bimodais, em ambos, votos e comparecimento. Um aglomerado maior ocorre para valores intermediários e um outro ocorre próximo de 100% de comparecimento e 100% votos para o partido vencedor. Esse pico sugere dois mecanismos de fraude: (i) incremental, quando com uma dada quantidade de votos para um partido são adicionados a urna e/ou votos do outro partido são retirados, (ii) extrema, corresponde a uma participação completa e quase todos os votos para um único partido. Essa fraude forma o aglomerado no canto superior direito.

Usando um escala apropriada, é possível observar que os dados da distribuição dos votos e do comparecimento às urnas seguem aproximadamente uma distribuição

Gaussiana. Observou-se também que a curtose de tal distribuição apresenta um achatamento superior em relação a eleições normais, conforme Figura 3.2.2.

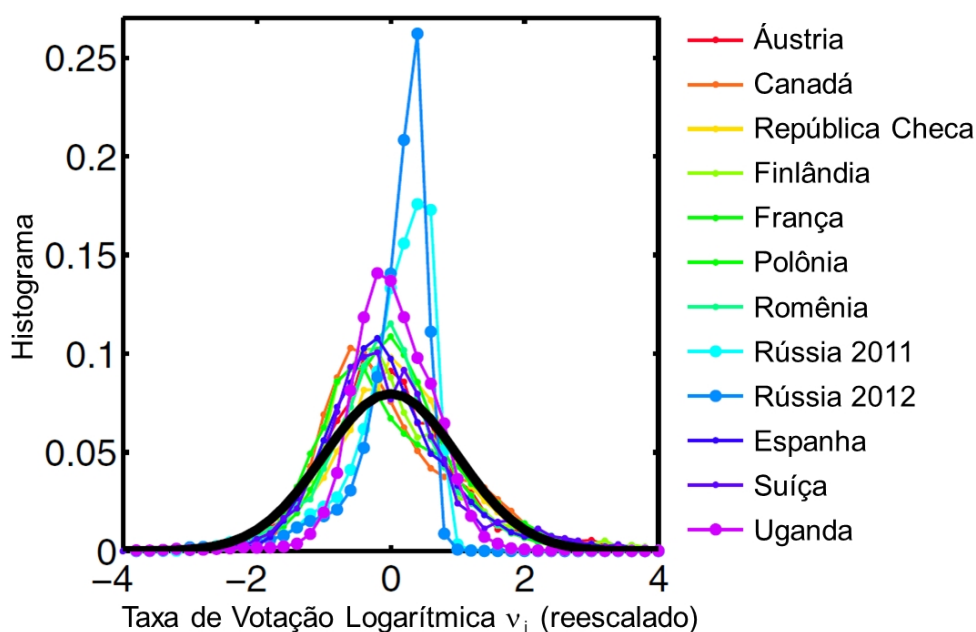


Figura 3.2.2 – Comparação da curtose entre as eleições: dados de diferentes eleições em diferentes países em um nível de agregação similar: distribuições das taxas logarítmicas de votos v_i das partes vencedoras como distribuições reescaladas com média zero e variância unitária. Fica bem visível as diferenças nas linhas de Uganda e Rússia. Fonte: KLIMEK, 2012.

Generalizando o método aplicado às eleições Russas, KLIMEK, (2012) propôs então um modelo paramétrico que identifica a presença e quantifica a intensidade de tais mecanismos fraudulentos, no sentido de reproduzir os padrões visíveis na figura 3.2.1. (KLIMEK, 2012). Em um segundo momento, propôs que, dentro de cada seção eleitoral, a taxa de comparecimento e votos para o candidato ou partido vencedor possuem distribuição Gaussiana, com média e desvio padrão tomados como os valores que maximizam os dados empíricos, visíveis na 3.2.2.

Na simulação do modelo, o primeiro passo foi encontrar as distribuições empíricas V_i/N_i e W_i/N_i para estimar os parâmetros necessários. Os mecanismos de fraude são inseridos no modelo, como segue: fraude incremental - com probabilidade f_i , cédulas são retiradas de ambos, não votantes e oposição, e são adicionadas ao partido vencedor, que é deslocada para a parte vencedora e foi estimada a partir dos dados reais da eleição (número de cédulas transferidas para o partido vencedor também foi estimada dos dados empíricos); e fraude extrema - com probabilidade f_e , quase todos os votos dos não votantes e da oposição são contabilizados para o partido vencedor.

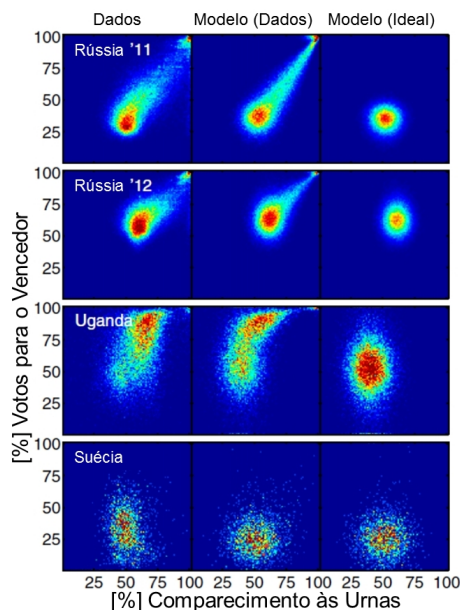


Figura 3.2.3 – Comparação do modelo com os resultados empíricos observados para Rússia 2011 e 2012, Uganda e Suíça. Do lado esquerdo, estão os dados reais, no centro, a simulação do modelo considerando os mecanismos de fraude e, a direita, o modelo considerando eleições justas (isto é, a ausência do mecanismo de fraude). Fonte: KLIMEK, 2012.

A figura 3.2.3 mostra a comparação do modelo com os resultados reais, simulações do modelo com e sem os mecanismos de fraude.

Para descrever o aglomerado principal, observado para Rússia e Uganda, foi necessário usar $f_i = 0,64$ para Rússia 2011 e $f_i = 0,39$ para Rússia Unida 2012, no mecanismo de fraude incremental. Isso significa uma fraude em torno de 64% das seções eleitorais em 2011 e de 39% em 2012. O segundo pico no canto superior direito, próximo a 100% de comparecimento onde existe aproximadamente 3.000 seções eleitorais com 100% de votos para Rússia Unida nos dados de 2011, representa um eleitorado de mais de 2 milhões de pessoas. Para o melhor ajuste do modelo, considerou-se $f_e = 0,033$ para 2011 e $f_e = 0,021$ para 2012, isto é, 2 a 3% das seções eleitorais tiveram fraude extrema.

3.3 Envolvimento Cívico do Eleitorado: Medida da Desordem nas Urnas

Cada eleição produz uma proporção variável de cidadãos que não participa da votação. A proporção da população não envolvida, seja por não-registro, abstenção ou votação em branco ou nulo, tem sido muito menos estudada do que o próprio voto. Atualmente, tais comportamentos estão aumentando entre as democracias mais antigas e seu significado pode estar mudando.

Além da abstenção passiva (devido ao descuido ou indiferença), uma recusa ativa do voto - possivelmente com uma mensagem política - está aumentando entre as categorias populacionais que geralmente participam das eleições.

Usando o conceito de entropia da mecânica estatística, BORGUESI (2012) investigou o envolvimento do eleitorado de acordo com a totalização dos votos nas urnas. Os dados investigados foram agregados em três escalas diferentes:

(1) Em escala municipal. Nessa escala, os fenômenos estudados são com respeito ao tamanho da população. As 76 eleições estudadas a nível municipal são recentes, depois de 1990, e em 11 países diferentes (Alemanha, Áustria, Canadá, República Checa, França, Itália, México, Polônia, Romênia, Espanha e Suíça).

(2) Em grande escala, por exemplo, nacional, provincial etc. Aqui, a análise é focada sobre a evolução temporal. Os países são estudados por seus aspectos históricos e são os mesmos da escala municipal. O estudo começa no início dos chamados regimes democráticos, após a Segunda Guerra Mundial, e mesmo antes, para alguns casos.

(3) Em nível de urnas, sobre as 100 cidades mais populosas, sempre que possível (ou seja, para o Canadá, França, México, Polônia e Romênia).

Considerou-se, para cada eleição e cada escala específica (município, província, país, etc.):

$$N = N_a + N_{bn} + N_c, \quad (3.5)$$

onde: N é o número total de votos registrados; N_a o número de abstenções, os eleitores que não compareceram para votar; N_{bn} votos brancos ou nulos; N_c votos em favor da lista de escolhas, votos válidos. O envolvimento cívico dos eleitores é medido por meio das três razões: $pa = N_a/N$, $pc = N_c/N$ e $pbn = N_{bn}/N$.

$$P_a + P_c + P_{bn} = 1 \quad (3.6)$$

As razões, relacionam-se numa quantidade onipresente em física estatística; a entropia. Nesse contexto, ela é definida por

$$S(P_a, P_c, P_{bn}) = -P_a \cdot \log(P_a) - P_c \cdot \log(P_c) - P_{bn} \cdot \log(P_{bn}) \quad (3.7)$$

em que \log é logaritmo na base 2. S é chamada entropia do envolvimento cívico ou envolvimento entrópico, e considerada como uma medida da desordem do envolvimento cívico em um nível coletivo. A ordem máxima, $S = 0$, é obtida quando uma das razões é igual a 1 (e então as outras duas são zero). Em contraste, a desordem máxima corresponde a $P_a = P_c = P_{bn} = 1/3$ e a entropia máxima, $S = \log(3) \approx 1,58$.

Assim como na física, o sistema entrópico conserva o mesmo total (100%) de energia, ou seja: 1. Da mesma forma que a soma das probabilidades de votos em cada

uma das possibilidades (um candidato, outro candidato, branco, nulo e abstenções) totaliza o número de eleitores inscritos para votar em cada urna.

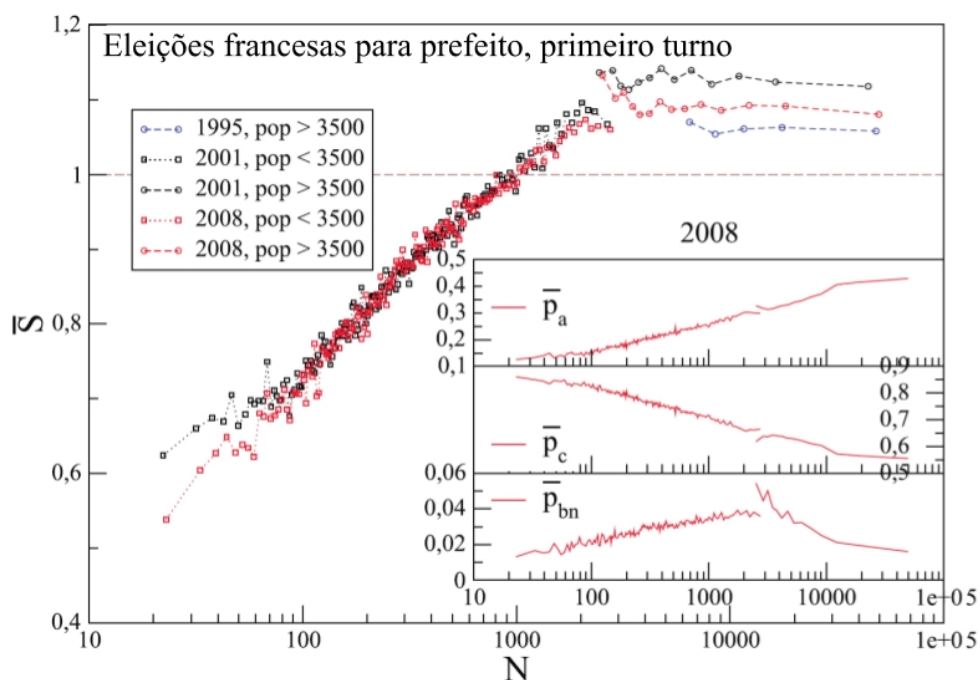


Figura 3.3.1 – Valores médios do envolvimento entrópico para França: S é tomado como uma função do número de eleitores N , para o primeiro turno de eleições para prefeito na França. Existem dois tipos de regra de votação, as quais dependem do tamanho da população, maior ou menor do que 3.500 habitantes. A inserção mostra os valores médios de P_a , P_c e P_{bn} como funções de N para eleições municipais em 2008 (que leva, para municípios altamente populosos, a um platô de S , apesar das variações em P_a , P_c e P_{bn}). Para cada N , os valores médios \bar{S} , \bar{P}_a , \bar{P}_c e \bar{P}_{bn} , são calculados sobre ≈ 200 municípios de tamanho $\approx N$. Fonte: BORGUESI, 2012.

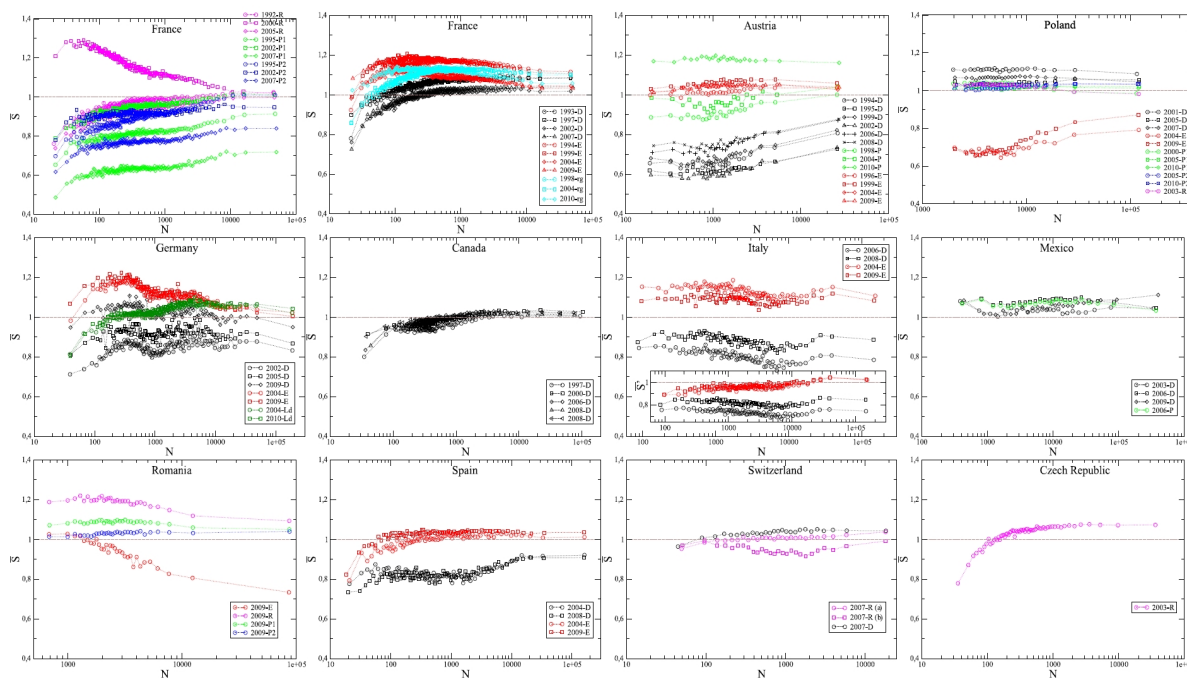


Figura 3.3.2 – Valores médios \bar{S} do envolvimento entrópico para municípios como uma função do número de eleitores registrados N . Cada ponto resulta da média sobre uma amostra de ≈ 100 (20 para França) municípios de tamanho $\approx N$. No gráfico da Itália, a inserção mostra uma variante de S em que votos brancos são agrupados a votos válidos. Fonte: BORGUESI (2012).

A razão P_a de abstenções é proporcional ao tamanho do município, N . Espera-se então um crescimento do envolvimento entrópico com o tamanho do município. Isso foi observado para eleições de primeiro turno em 2001 e 2008 na França, como pode ser visto na figura 3.3.1. Essa figura identifica também um valor limiar do número de eleitores de uma cidade, 3.500, a partir do qual o envolvimento entrópico torna-se constante. A entropia do envolvimento cívico também foi calculada para outros países, como pode ser visto na Figura 3.3.2.

Uma outra análise é realizada agregando-se países da base de dados em dois grupos: um com eleições mais antigas e outro com eleições mais recentes, tendo um total de 321 eleições divididas entre os dois grupos com aproximadamente o mesmo número de eleições em cada um. O histograma do envolvimento entrópico S são comparados para estes dois grupos na Figura 3.3.3. Os histogramas para o grupo das eleições mais recentes mostra um acentuado pico em ≈ 1 , enquanto o grupo de eleições mais antigas tem uma ampla distribuição. Em eleições atuais, existem poucas eleições com envolvimento entrópico pequeno (digamos $S \leq 0,8$, porém a muitas eleições com $S \approx 1$.

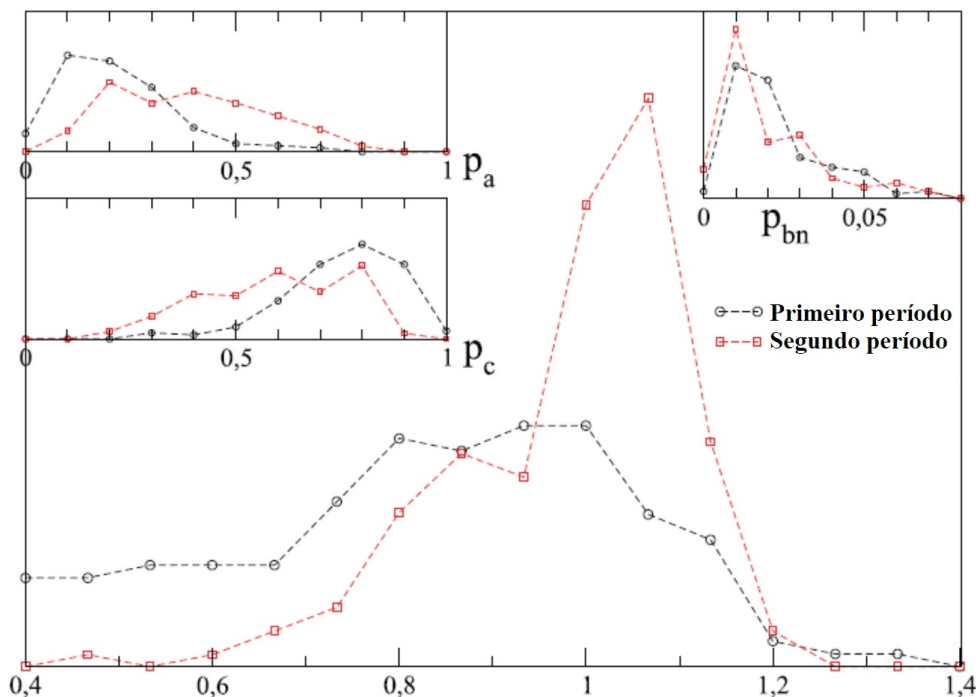


Figura 3.3.3 – Evolução temporal do envolvimento entrópico S em grande escala (nacional, provincial etc.) de 321 eleições, além de referendos suos.

Para cada país, os resultados eleitorais são divididos em dois grupos: os que ocorreram em um primeiro período de tempo e os ocorreram num segundo período. Os histogramas de S (de P_a , P_c e P_{bn} nas inserções) mostram o envolvimento entrópico do primeiro e do segundo grupo. (BORGHESI, 2012).

3.4 Anomalias Estatísticas nas Eleições Russas em 2011–2012 Reveladas pela Análise de Correlação 2D

Já sabemos que as variáveis comparecimento (ou não) e os votos inválidos (brancos e nulos) compõem um pano de fundo que, embora não contribua diretamente para o resultado do vencedor, constitui parte do cenário descrito pelo resultado, e que podem indicar irregularidades de tal monta, somente possíveis com uma intervenção desleal no processo eleitoral; chegando ao ponto de alterar o resultado das eleições.

Observando as eleições realizadas em legislativas em 4 de dezembro de 2011 (para o parlamento russo) e a eleição presidencial em 4 de março de 2012, pesquisadores usaram métodos estatísticos para fazer comparações de dados para trazer à tona algumas anomalias que podem ter comprometido sistemática e persistentemente os resultados, favorecendo significativamente o partido pró-governo Rússia Unida e seu líder Vladimir Putin. (KOBÁK, 2012)

A partir do resultado das eleições legislativas, a população russa, acreditando que o pleito foi manipulado promoveu protestos públicos em larga escala, nunca

vistos na Rússia desde o início dos anos 90. Mesmo assim nenhuma das alegadas manipulações foi oficialmente reconhecida.

Após as eleições parlamentares, e com um resultado entendido como absurdo por boa parte da população, iniciaram-se um gama de atividades antifalsificação sem precedentes nas eleições presidenciais, como transmissão ao vivo da maioria das seções eleitorais e intenso controle público por parte de observadores voluntários. Atividades estas que inspiraram os pesquisadores a utilizar métodos de espectroscopia de correlação bidimensional (2D).

Para encontrar irregularidades em dados eleitorais a estatística é uma ferramenta muito eficiente. Também identifica este par de eleições como uma oportunidade ímpar para comparar seus dados, já que a maioria dos líderes do partido depois se candidatou à presidência. (KOBAK, 2012)

Entre os 3 meses que se passaram entre uma eleição e outra, a distribuição sociogeográfica não poderia ter se alterado de forma relevante, assim os conjuntos de dados deveriam mostrar padrões semelhantes.

Foram analisados os dados oficiais das duas eleições disponibilizados no sítio online do Comitê Central Eleitoral da Rússia, detalhados por urnas. Os partidos que concorreram nas eleições parlamentares foram sete, e quatro deles não passaram de um limiar de 7% dos votos válidos. Cinco candidatos concorreram à presidência. Na Rússia existem 95 mil urnas. A estatística eleitoral alcançou mais de 109 milhões de eleitores aptos a votar, com 65,7 milhões de votos nas eleições legislativas e 71,7 milhões de votos nas eleições presidenciais. Rússia Unida (UR) venceu as eleições parlamentares com um resultado de 49,31%, e Vladimir Putin derrotou seus adversários com 63,60% dos votos válidos, estabelecendo um recorde.

As principais irregularidades detectadas foram as seguintes: (i) correlação notavelmente alta entre resultados de comparecimento às urnas e de votação no candidato ou partido vencedor; (ii) um grande número de seções de votação onde os resultados do partido Rússia Unida (RU) e de Putin são dados por um número percentual redondo; (iii) distritos eleitorais mostrando uma dispersão anormalmente baixa ou (iv) anormalmente alta nos resultados seções de votação; (v) diferença substancial entre os resultados nas seções de votação em papel e eletrônicas. Estas anomalias dificilmente se ajustam aos pressupostos de uma eleição livre e justa.

As figuras 3.4.1A e 3.4.1E mostram histogramas 2D do número de votos a favor do partido Rússia Unida (RU) e [Vladimir] Putin como função do comparecimento e respectiva cada seção de votação. Além dos principais aglomerados em $\approx 52\%$ de participação e $\approx 30\%$ dos votos para RU e $\approx 60\%/55\%$ para Putin, existem duas características proeminentes em ambas as parcelas que os distinguem claramente

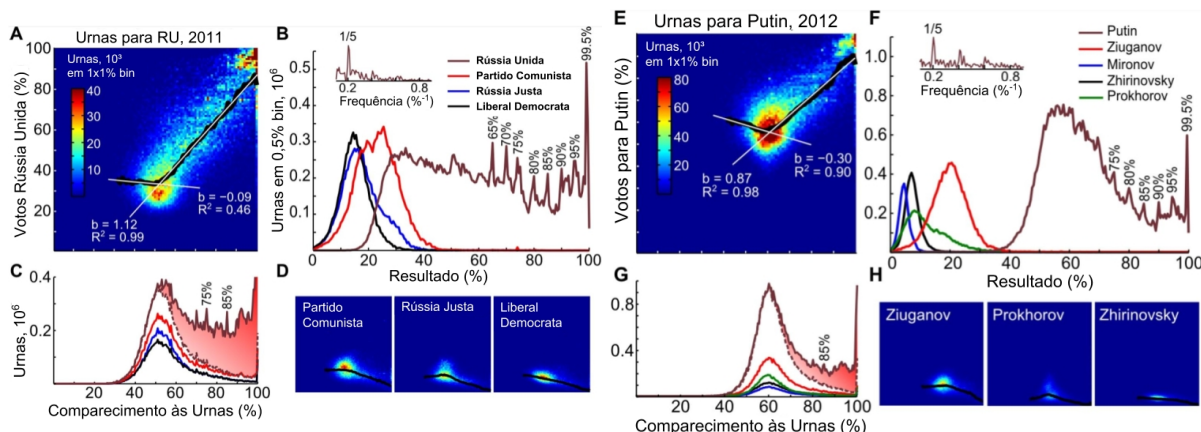


Figura 3.4.1 – Resumo dos resultados da Rússia Unida: (A) Cédulas obtidas nas seções de votação que mostram uma certa participação e resultado da Rússia Unida (em 1% - 1% das posições). Número de cédulas é codificado por cores; o *cluster* no canto superior direito está muito saturado para permitir que outros dados fiquem visíveis. A curva preta mostra um resultado geral para cada urna. As linhas brancas mostram ajustes lineares à curva preta antes e depois dos 50% de participação; o valor de R^2 e o coeficiente de regressão são representados ao lado de cada ajuste. (B) Número total de boletins de voto para cada partido, dependendo do resultado na estação de voto (em urnas de 0,5%). Inserção mostra o espectro de energia de Fourier do traço da Rússia Unida. (C) Número de boletins dependendo do comparecimento às urnas (0,5% de caixas). O código de cores é o mesmo que em (B). A linha tracejada mostra a parte do traço UR proporcional à soma de todas as outras partes; sombreamento vermelho mostra a diferença. O traço UR é truncado em 100% de participação por motivos de clareza; o valor máximo é $0,98 \cdot 10^6$. (D) Histogramas bidimensionais para três outros partidos eleitos. A escala de cores é a mesma que em (A). (E-H) Parcelas semelhantes para as eleições presidenciais. Fonte: KOBAK, 2012.

dos histogramas de outros participantes (Figuras 3.4.1D e 3.4.1H): (i) um conjunto incomum de votos na vizinhança de 95% de participação, e (ii) uma longa cauda de votos começando no pico central, que mostra uma alta correlação dos resultados com o comparecimento (marcado por curvas, linha central inclinação na espectroscopia 2D).

Os aglomerados com 90 a 100% de rendimento na participação de $\approx 3,5$ milhões de votos para os vencedores nas duas eleições e pode ser rastreada até seis repúblicas de Distrito Federal do Norte do Cáucaso e Repúblicas da Mordóvia, Bascortostão e Tartaristão. Em cada uma dessas nove regiões, há um número de círculos eleitorais que exibem resultados de votação com dispersão extremamente baixa entre as seções de votação, significativamente menor que o valor de dispersão imposto por modelo binomial (por exemplo, 25 círculos eleitorais com p-valor $< 0,0001$ para parlamentares e 9 círculos eleitorais para as eleições presidenciais (conforme os dados de KOBAK, 2012). Isso sugere que os resultados nesses grupos constituintes

foram artificialmente fixados em determinada porcentagem valores.

É importante observar uma projeção dos histogramas no eixo vertical, o que dá uma distribuição do número de boletins de voto para RU e Putin dependendo de seus resultados em cada votação estação (Figuras 3.4.1B e 3.4.1F). A característica única destes histogramas são picos pontiagudos localizados em números “redondos” de 65%, 70%, 75% etc. O caráter periódico destes picos é evidente a partir dos espectros de Fourier que mostram harmônicos proeminentes em $1/5\%-1$ (inserções). De longe, o pico mais alto em ambos os casos está localizado em 99,5% e é originário apenas de uma única região de República Chechena. Outros picos também podem ser rastreados de volta a determinadas circunscrições, mas geralmente não confinado a uma única região. Esses picos, que são altamente significativos estatisticamente (dados de KOBAK, 2012), compreendem 1.4 milhões de votos para UR e 1,3 milhão de votos para Putin.

O caráter sobrenatural geral dos picos sugere fortemente que os votos para os vencedores foram manipulados a posteriori para fixar (de forma inescrupulosa e ingênua) os votos em valores redondos. A segunda característica que se destaca do histograma 2D nas Figuras 3.4.1A e 3.4.1E é uma correlação notável entre a participação e o resultado da RU (correlação coeficiente de 0,68) e Putin (0,53). Observe que em índices de comparecimentos mais baixos ambas as correlações são negativas, tornando-se positivo apenas em retornos superiores nas posições dos principais aglomerados. Os histogramas para outros concorrentes mostram comportamento exatamente oposto: correlação baixa ou mesmo positiva em taxas de comparecimento mais baixos e correlação negativa mais adiante (Figuras 3.4.1D e 3.4.1H).

A correlação entre participação e os resultados da votação é um fenômeno bem conhecido, observado em muitos países. (BORGUESI, 2012 e KLIMEK, 2012). No entanto, dependências como forte como encontrado aqui são difíceis de explicar sem uma suposição de pressão administrativa e/ou manipulação de voto (KOBAK, 2012).

Com base nos dados apresentados por KOBAK (2012), a correlação entre resultados de participação e votação na escala nacional poderia ter surgido devido a agregação de dispersão amplamente dispersa, mas não correlacionada resultados de diferentes territórios, dado grandes diferenças culturais e socioeconômicas entre regiões da Rússia, bem como entre as áreas urbanas e rurais. Para resolver esta questão, os dados apresentados nas Figuras 3.4.1A e 3.4.1E foram decompostos em três partes: áreas urbanas, áreas rurais, e as nove repúblicas acima mencionadas. Ambas as áreas urbanas e rurais exibem separadamente correlações; um maior detalhamento ao nível da região mostra que a alta correlação não é característica cada região, mas está confinado a apenas algumas regiões da Rússia. Além disso, em regiões que demonstram altas correlações semelhantes já são observadas ao nível dos círculos eleitorais individuais. Isto mostra que as correlações observadas não são um artefato

de agregação, mas uma característica interna de fatores específicos.

Um dos exemplos mais marcantes de tais correlações é dado pela cidade de Moscou, onde a eleição parlamentar resultou em um nível extremamente alto de correlação entre o comparecimento às urnas e os votos para o RU (partido Rússia Unida) (Figura 3.4.2A). A situação foi totalmente invertida nas eleições para presidente, onde o resultado de Putin foi fortemente anticorrelacionado com o comparecimento (Figura 3.4.2B). Além disso, as projeções horizontais dos histogramas 2D (que mostra o número de cédulas em função de afluência) adquiriu formas semelhantes para todos os candidatos.

Outra particularidade observada é a diferença do desvio padrão dos votos obtidos pelo partido Rússia Unida e por Putin nos respectivos pleitos, apenas em Moscou. A possibilidade mais aceita é que, após o pleito para o Parlamento Russo, com os eleitores bem mais atentos às possibilidades de fraudes e a implantação de mais e mais rígidos mecanismos de controle, o desvio padrão se concentrou acerca de 5%, principalmente em Moscou, onde o candidato obteve um percentual de votação entre 40 e 50% aproximadamente.

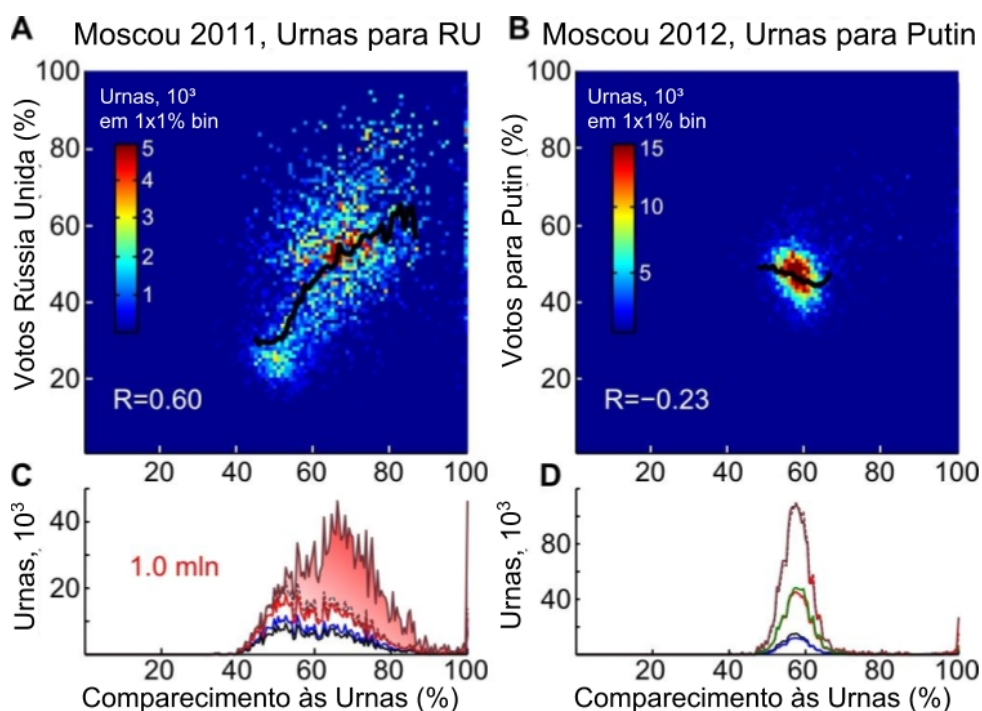


Figura 3.4.2 – Resultado das Votações na Cidade de Moscou: Número de votos para RU (A) e Putin (B) nas seções de votação que mostram uma certa participação e resultam em eleições parlamentares e presidenciais, respectivamente (em 1% X 1% de bins). R representa o coeficiente de correlação (excluindo 5% dos votos no mais alto e 5% nos comparecimentos mais baixos). Note-se que dois clusters distintos de cédulas em 50% e 70% de participação e uma alta correlação positiva entre o comparecimento às urnas e resultado de RU em (A) se transformou em um único aglomerado bem confinado e correlação negativa entre resultado e Putin em (B). (C, D) Projeções horizontais de (A) e (B), juntamente com os histogramas de outros participantes. As cores são as mesmas das Figura 3.4.1 e 3.4.2. O número vermelho em (C) mostra a área do sombreamento vermelho, semelhante à Figura 3.4.1C. Fonte: KOBAK, 2012.

Por fim, em ambas as eleições, algumas seções de votação ($\approx 5,5\%$ em todo o país) foram equipadas com urnas eletrônicas para digitalizar as cédulas e contar automaticamente os votos, reduzindo assim a possibilidade de interferência humana, e (muitos acreditam) reduzindo também a oportunidade de fraudes. Na análise mostrada na Figura 3.4.3 que dentro dos mesmos distritos eleitorais o resultado da RU nas seções de votação eletrônicas era em média 7,1% menor do que nos tradicionais baseados em papel (diferença significativa com $p = 10^{-51}$) e o resultado de Putin foi 4,7% menor $p = 10^{-35}$).

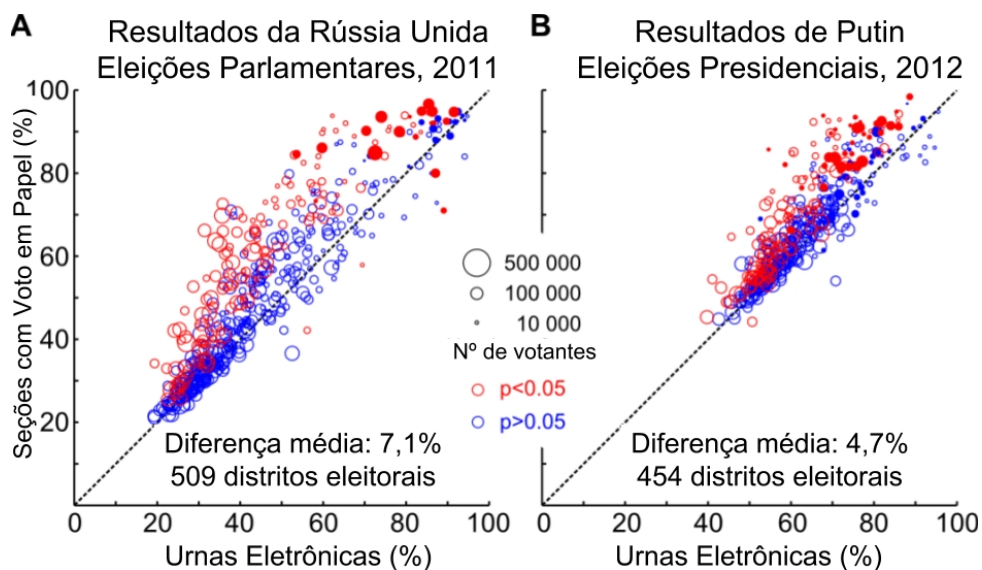


Figura 3.4.3 – Correlação entre os resultados dos vencedores nas seções de voto eletrônico: as áreas circulares são proporcionais ao número de eleitores registrados em um distrito. Círculos preenchidos mostram círculos eleitorais localizados nas nove repúblicas. Círculos vermelhos mostram distritos eleitorais onde os resultados de UR / Putin em seções de votação eletrônicas e em papel são significativamente diferentes com $p < 0,005$; círculos azuis mostram todos os restantes círculos eleitorais. KOBÁK, 2012

Apesar de haver a incerteza de que as seções de votação eletrônicas constituem um subconjunto representativo, estas diferenças são consistentes com as estimativas apresentadas. (KOBÁK, 2012).

KOBÁK (2012) partiu com dos dados oficiais das duas eleições, obtidos a partir do sítio online do Comitê Central Eleitoral Russo (izbirkom.ru). As correlações obtidas a partir destes dados foram obtidas usando correção de Spearman (ao invés de Pearson) para garantir uma análise mais robusta quando se trata de dados discrepantes, como por exemplo de instalações militares e hospitais, onde o comparecimento à votação típica se aproxima de 100%.

Foi utilizada análise de correlação 2D para identificar com eficiência várias anomalias nas recentes eleições russas. Embora em todas as métricas discutidas as eleições presidenciais pareçam mais justas que as parlamentares, várias anomalias ainda chegam aos milhões de votos. A análise estatística por si só não serve (e não consegue servir) como uma prova conclusiva de qualquer fraude possível, mas ela destaca claramente as impressões digitais alarmantes nos resultados da votação.

As abordagens e métodos propostos por KOBÁK (2012) podem ser estendidas para quantificar as impressões digitais da fraude eleitoral em qualquer outra eleição, deixando as ferramentas prontas para aplicar aos dados de qualquer outro país, inclusive os do Brasil, seja os das últimas eleições de 2018 como todos os outros países

advindos ou não de pleitos com uso de urnas eletrônicas.

3.5 Testes para Manipulação de Eleitores em Pequenas Seções de Votação

Após vermos lado a lado o caso de duas eleições russas com forte indicação de fraude, vejamos uma outra pesquisa (JIMENEZ, 2017) que abrangeu também a Rússia (em 2011) e, dentre outros países, também a Venezuela. Neste caso em específico a votação obtida em seções eleitorais pequenas somadas fizeram a diferença no resultado do pleito como um todo.

A principal hipótese baseia-se em alguns aspectos sociais, dentre eles, que pequenas seções eleitorais são facilmente suscetíveis à coerção dos eleitores porque é mais fácil identificar indivíduos opostos, há menos testemunhas oculares e as partes interessadas esperam razoavelmente menos visitas de observadores eleitorais.

Alguns cenários de fraude numa grande variedade de modalidade, que vão desde candidatos “terroristas” que obrigam os eleitores ao votar em si, sob pena de morte (literal e ilegalmente, através de milícias), até políticos que usam de clientelismo - modalidade em que o eleitor voluntariamente vota sempre em seu candidato porque este o beneficia direta ou indiretamente assim que ascende ao governo.

O clientelismo consiste em uma prática que não chega a ser ilegal, pois o eleitor entende que o candidato, uma vez eleito, o representa; mas sim imoral, porque este o representa privilegiando seus eleitores em detrimento dos demais. É uma prática tão sutil que ocorre mesmo nas democracias entendidas como mais desenvolvidas, como Estados Unidos e Austrália, dentre outras. Já em países como o Zimbábue, a fraude não é tão discreta, chegando à imprensa oficial rumores de transporte e intimidação dos eleitores.

Outros casos com ações não tão extremas mas bastante explícitos surgiram, não só em rumores, como em estudos dos dados em vários países, como: Venezuela, Rússia, e etc.

A hipótese de JIMENEZ (2017) é que em seções de votação pequenas, existe uma facilidade em manipular o eleitor para que vote em quem o opressor desejar, tomando como base alguns fatores: pequenos centros podem ser particularmente suscetíveis à fraude eleitoral porque (i) é mais fácil identificar indivíduos opostos, (ii) esses centros quase sempre estão em áreas supostamente pró-governamentais, (iii) há menos testemunhas oculares e (iv) os centros são visitados com menos frequência por observadores eleitorais.

Para analisar estas seções eleitorais, JIMENEZ (2017) desenvolveu um teste estatístico para quantificar até que ponto os resultados de uma determinada eleição

exibir traços de manipulação do eleitor, mostrando discrepâncias entre as pequenas seções de votação e suas seções vizinhas.

O método aplicado aos resultados de 21 eleições em 10 diferentes países. Em particular, JIMENEZ (2017) desenvolveu um teste de significância estatística que nos permite investigar se o comportamento de voto em pequenas seções de votação é substancialmente diferente das seções de grande dimensão.

A possibilidade destas irregularidades em pequenas estações é firmemente rejeitada pelo método proposto por JIMENEZ (2017) em várias eleições em oito países, incluindo Venezuela, antes que o atual partido governante controlasse o poder eleitoral. Observou-se um ponto de virada no tamanho das irregularidades eleitorais entre 2004 e 2006 na Venezuela, culminando em um período de efeito da distorção do eleitor - como distorções nos resultados das eleições presidenciais de 2013.

Tal fato pode levar a crer que o principal fator que impacta neste estudo, é de fato se o partido ou candidato vencedor já tinha o poder nas mãos, com todas as vantagens e recursos de se possuir a “máquina governamental” em seu favor. O que sugere uma observação adicional ao tratar os dados do Brasil, como o faremos nos capítulos em diante.

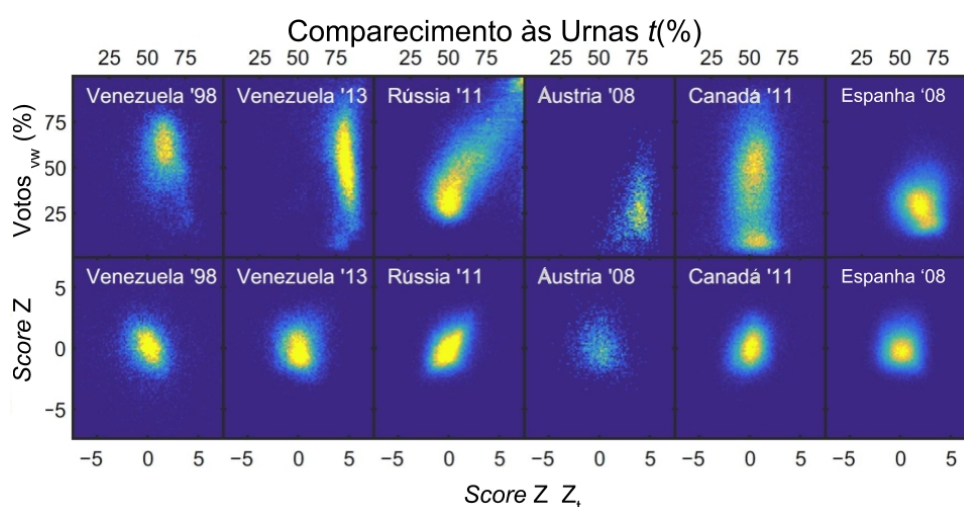


Figura 3.5.1 – Comparação das impressões digitais eleitorais e sua variante padronizada para seis eleições diferentes: os IDEPs (Impressão Digital Eleitoral Personalizada) (linha inferior) são comparados com as impressões digitais da eleição (linha superior) como introduzidas por KLIMEK, 2012 para um conjunto selecionado de eleições, nomeadamente (da esquerda para a direita), Venezuela 1998 e 2013, Rússia 2011, Áustria 2008, Canadá 2011 e Espanha 2008. Fonte: JIMENEZ, 2017.

A figura 3.5.1 mostra as impressões digitais da eleição, conforme introduzidas por KLIMEK, 2012, e as impressões digitais eleitorais padronizadas (IDEPs) de várias eleições (Venezuela 1998 e 2013, Rússia 2011, Áustria 2008, Canadá 2011 e Espanha 2008).

Observe que a padronização estratificada nos IDEPs corrige a multimodalidade (Canadá) e mobilização heterogênea de eleitores (Áustria) que foi observado em versões anteriores de impressões digitais de eleições.

No entanto, não há razão para acreditar que os IDEPs (Impressões Digitais Eleitorais Padronizadas) devam seguir o modelo normal bivariado não correlacionado, tal como foi sugerido para as impressões digitais. O voto estratégico pode introduzir alguma correlação entre o comparecimento padronizado, Z_t , e o número padronizado do eleitor para o partido vencedor, Z_{vw} , porque mais comparecimentos podem ser associados à mobilização a favor ou contra o vencedor.

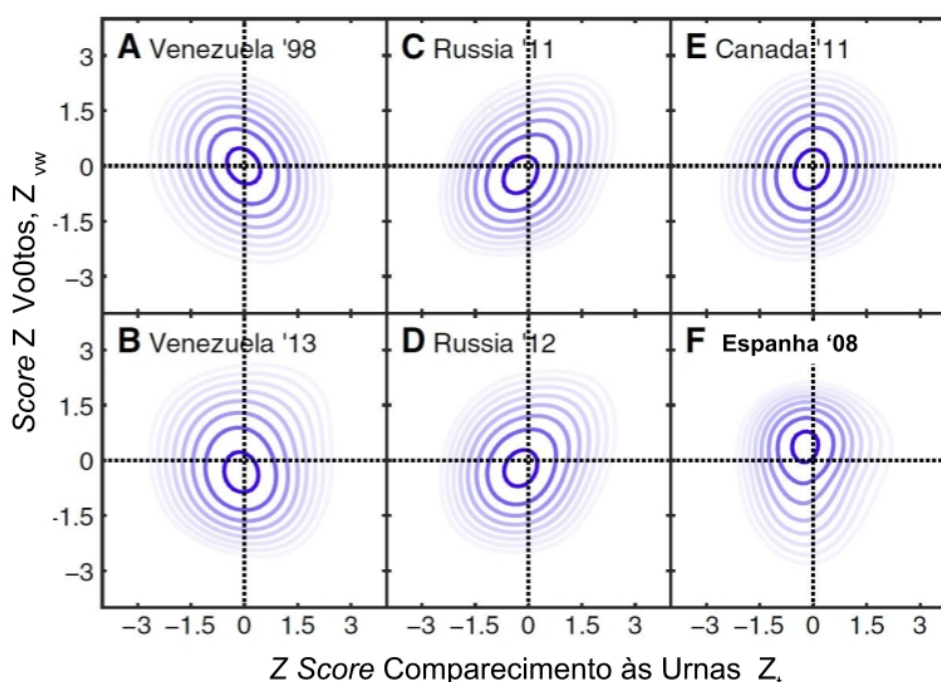


Figura 3.5.2 – Simetria dos IDEPs: os IDEPs de (A) Venezuela 1998 e (B) 2013, (C) Rússia 2011, (D) Rússia 2012, (E) Canadá 2011 e (F) Espanha 2011 são mostrados usando representações de contorno para as densidades de pontos de dados. Os IDEPs parecem ser axialmente simétricos. Eles estão próximos da simetria elíptica. Fonte: JIMENEZ, 2017.

Se a vizinhança eleitoral é homogênea o suficiente, argumentos assintóticos podem ser invocados para alegar que as distribuições marginais empíricas dos escores Z podem ser aproximadamente uma distribuição normal padrão, embora com valores mais extremos do que os esperados para uma amostra gaussiana devido à complexidade dos processos eleitorais. Portanto, a única alegação possível é que a distribuição conjunta deve ser unimodal, centrada na origem, e aproximadamente suportada em uma área normal de alta confiança. Além disso, espera-se uma simetria particular para a IDEP de uma eleição em que certos tipos de fraude podem ser descartados. Isso se torna aparente em uma visualização de contorno de histogramas bidimensionais (2D) de todas as seções eleitorais para várias eleições. (Figura 3.5.3). Os IDEPs dessas

eleições parecem ser axialmente simétricas, quase elíptica simétrica em vários casos.

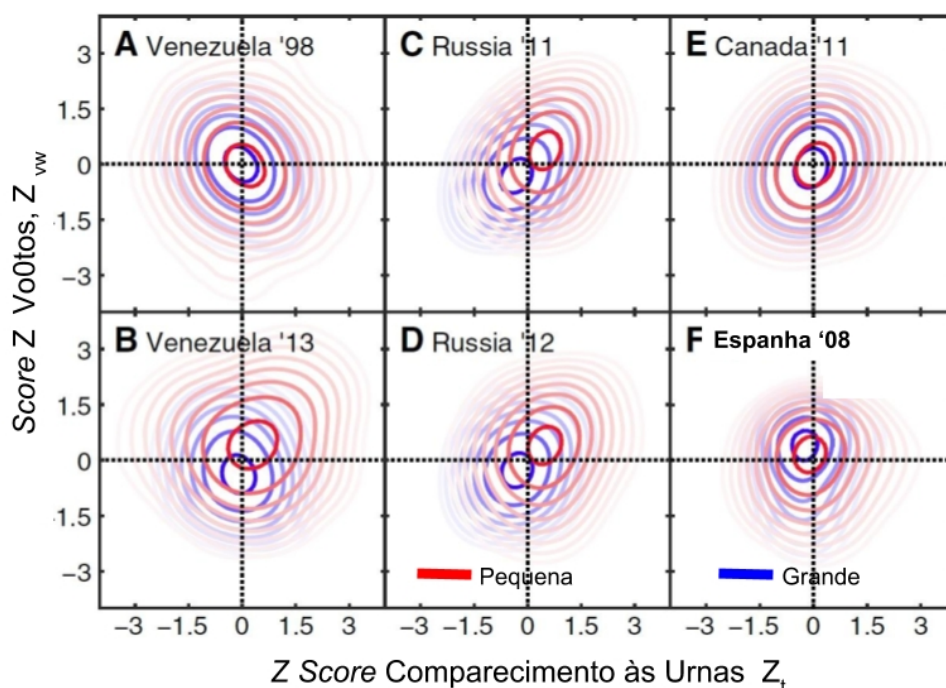


Figura 3.5.3 – Comparação dos IDEPs de seções eleitorais grandes (azuis) e pequenas (vermelhas) para (A) Venezuela 1998, (B) Venezuela 2013, (C) Rússia 2011, (D) Rússia 2012, (E) Canadá 2011 e (F) Espanha 2011. Os centros das distribuições de pequenos e grandes centros coincidem para Venezuela 1998, Canadá 2011 e Espanha 2011. No entanto, há uma clara discrepância nos IDEPs de pequenos e grandes centros na Venezuela 2013 e Rússia 2011 e 2012. O comparecimento reescalado e os votos para o vencedor são substancialmente maiores em pequenos centros para estas eleições. Esta é uma evidência clara de que os resultados das eleições nesses pequenos centros mostram distorções sistemáticas. Fonte: JIMENEZ, 2017.

A figura 3.5.3 mostra as visualizações dos IDEPs para pequenas e grandes seções de votação usando linhas de contorno (por equidensidade). Nesta figura, pequena seções eleitorais são aquelas que têm um tamanho de eleitorado abaixo percentil $p = 20$. Resultados semelhantes podem ser observados para todas as outras escolhas de p . Para descartar que os resultados não são impulsionados por poucas e pequenas seções eleitorais com contagem atípica de votos, foi feita uma remoção simples de dados discrepantes de escore Z baseado na simetria elíptica. Para Venezuela 1998, Canadá 2011 e Espanha 2011, os centros dos IDEPs para as pequenas e grandes seções eleitorais coincidem, e as formas das distribuições são difíceis de distinguir.

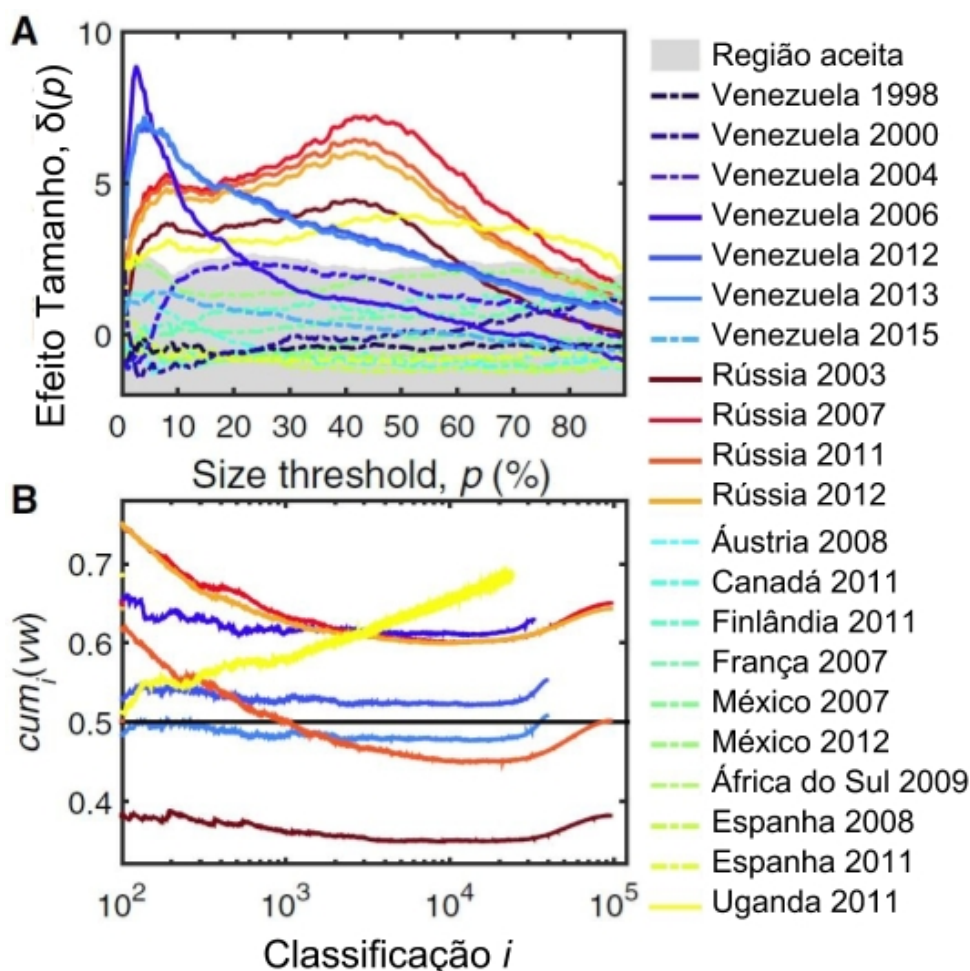


Figura 3.5.4 – (A) Os resultados do teste estatístico para manipulação do eleitor, quantificados pelo tamanho do efeito (p), são mostrados para 21 eleições diferentes. A região cinzenta contém as eleições sem diferenças significativas entre seções eleitorais pequenas e grandes. Essas eleições também são mostradas como linhas tracejadas. Linhas sólidas mostram eleições que são compatíveis com a hipótese dos fraudadores nos pequenos centros. Estas incluem as eleições na Rússia, Uganda e Venezuela de 2006 a 2013, onde os efeitos mais fortes são observados. (B) Diferentes visualizações dos resultados atípicos observados da Rússia, Uganda e Venezuela. Nestas visualizações, as seções eleitorais são ordenadas de forma decrescente de acordo com o seu número de eleitores, e a percentagem de votos para o vencedor é calculada usando apenas seções eleitorais até o *ranking* determinado no eixo x (em escala logarítmica). Para a Venezuela 2013, é apenas a adição de pequenas seções eleitorais que empurra os resultados para determinar o vencedor. Fonte: JIMENEZ, 2017.

Isto não acontece para Venezuela 2013, Rússia 2011 e Rússia 2012. Os escores Z destas eleições são substancialmente deslocados para cima e para direita do gráfico para pequenas seções eleitorais. Isso significa comparecimento mais alto do eleitorado e mais votos para o vencedor em pequenas seções eleitorais, conforme sua vizinhança direta, uma evidência clara de uma distorção sistemática dos resultados

das eleições nestas seções, que é consistente com os efeitos da coação de eleitores nesses locais.

Para quantificar sistematicamente estas distorções, considera-se a distância padronizada $\delta(p)$ como o efeito tamanho, que dá o número de desvios padrão pelos quais centros pequenos e grandes estão em média deslocadas umas das outras, conforme medido sobre um conjunto de eleições não discrepantes. A figura 3.5.4A mostra resultados para $\delta(p)$ para as 21 eleições em estudo. A região cinza da figura 3.5.4A, a “região aceita”, contém os resultados para todas as eleições que não são classificados como discrepantes de acordo com a maioria dos resultados do teste modificado de Thompson tau¹. Eleições que estão fora desta região mostram resultados que são compatíveis com a suposição de manipulação generalizada de eleitores em pequenas seções eleitorais.

Foram observados os efeitos mais fortes na Venezuela entre 2006 e 2013, com valores de $\delta(p)$ que quase atingem um fator de 10 para limites de tamanho em torno do 5º percentil. No entanto, os valores de $\delta(p)$ estão claramente fora da região aceita para uma ampla gama de opções de p . Tais desvios fortes estão totalmente ausentes dos dados venezuelanos em eleições, bem como na última eleição parlamentar venezuelana (isto é, 2015). As eleições russas entre 2007 e 2012 também mostram desvios significativos, com $\delta(p)$ entre 4 e 7.

Os resultados também são encontrados para a Rússia 2003 e Uganda, mas com substancialmente tamanhos de efeito menores com valores entre 3 e 4. Certamente, estes resultados anômalos para a Rússia, Uganda e Venezuela podem ser traços de fraude eleitoral em pequenas seções eleitorais e levantam sérias dúvidas a integridade das eleições relacionadas.

A região aceita na figura 3.5.4A encapsula eleições que mostram desvios abaixo de 3 desvios-padrão, em consistência com um nível de confiança de 95%. Os desvios-padrão nas eleições na Rússia de 2007 a 2012 e Venezuela de 2006 a 2013, portanto, indicam eventos verdadeiramente extremos. Para avaliar a robustez dos resultados acima, foi feita uma validação cruzada.

Foi encontrada somente a eleição na Venezuela em 2006, que caiu nas regiões aceitas para uma ampla gama de escolhas de p na figura 3.6.1A, torna-se insignificante em 9 das 21 validações cruzadas de exclusão única. Em cada um desses nove casos, foi removido um conjunto de dados de eleição do conjunto de referência de eleições “confiáveis”. Todas as outras remoções não produziram resultados significativos. Nas eleições da Rússia de 2007 a 2012 e Venezuela de 2012 e 2013 permanecem

¹ O teste Thompson Tau modificado é uma maneira de encontrar *outliers* em um conjunto de dados. Um potencial *outlier* é testado por vez, usando uma versão do teste t. O teste de Tau elimina valores discrepantes a mais de dois desvios-padrão da média.

significativos em todas as 21 validações cruzadas.

Foi avaliada a robustez dos resultados por três testes randômicos. Primeiro, para cada eleição com desvios significativos $\delta(p)$, foram permutados aleatoriamente os tamanhos do eleitorado nas seções eleitorais, mantendo as porcentagens vw e t fixas. Assim, cada seção é aleatoriamente designada para o grupo de seções pequenas ou grandes. Isso remove completamente qualquer correlação entre o tamanho do eleitorado e os percentuais vw e t , e não eram esperados resultados significativos sob esta randomização.

Depois de isso, observa-se que para cada uma dessas eleições, $\delta(p)$ se enquadra região de aceitação mostrada na figura 3.5.4. Para o segundo teste foi permutada aleatoriamente o vw , enquanto no terceiro teste o t foi permutado aleatoriamente, mantendo todas as outras variáveis fixas em cada teste. Ao permutar vw , foi obtido o mesmo conjunto de eleições com desvios significativos como nos dados originais.

Ao permutar t , apenas os resultados da Rússia de 2003 e da Uganda de 2011 se tornam insignificante. Note-se que no segundo teste, a correlação entre tamanho e participação é preservada pela permuta aleatória das porcentagens vw .

De maneira semelhante, a correlação entre n e vw é preservada aleatoriamente permutando t . Estas correlações residuais são na maioria dos casos suficiente para o método produzir resultados consistentes. Uma visualização diferente dos resultados atípicos relatados para a Rússia, Uganda e Venezuela de 2006 a 2013 são fornecidos na figura 3.6.1B.

Aqui, as seções eleitorais são classificadas de acordo com o número de eleitores de maneira descendente. Em seguida, foram calculadas as porcentagens de votos para os vencedores sobre todas as seções até a classificação dada i (isto é, acima do número correspondente de eleitores). Este número é denotado $cum_i(vw)$.

Se o comportamento de voto nas seções é independente do tamanho da seção, é esperada uma inclinação de zero para $cum_i(vw)$ para fileiras altas i . Isso significa que a adição de seções eleitorais cada vez menores não muda os resultados globais das eleições.

Para Uganda, foi observado um aumento logarítmico de $cum_i(vw)$ de seções eleitorais relativamente grandes. Foi notado um padrão diferente para a Rússia e Venezuela, onde $cum_i(vw)$ aumenta claramente nas seções menores. Portanto, a adição de essas pequenas seções têm um impacto substancial nos resultados das eleições.

Para a Rússia 2011 e Venezuela 2013, as contribuições das seções eleitorais muito pequenas que empurram o número total de votos para a barreira de 50%. No

caso venezuelano, em que o atual presidente (Nicolás Maduro) foi eleito por uma pluralidade de sistemas, os resultados irregulares das menores seções eleitorais foram determinantes no resultado.

O método desenvolvido por JIMENEZ (2017) para testar anomalias estatísticas de eleições resulta em pequenas seções de votação passíveis de manipulação dos eleitores. Praticamente foi elaborada uma ferramenta comparativa baseada em IDEPs de diferentes países.

A análise levou em conta 21 eleições nacionais em 10 países mostra irregularidades estatísticas na Venezuela 2006–2013, Rússia 2007–2012 e, em menor escala, Rússia 2003 e Uganda 2011. A manipulação eleitoral pode ser a constatação a que leva os resultados nas eleições presidenciais venezuelanas de 2013, em particular.

A integridade do sistema eleitoral venezuelano tem sido questionada desde a realização do referendo revogatório presidencial em 2004.

Particularmente nas últimas eleições, algumas táticas para influenciar e/ou pressionar o eleitor foram amplamente divulgadas. A primeira dessas táticas é através votação assistida, que pode estar associada à coerção dos eleitores. Isso foi detectado em 6,3% das seções de votação e observadas em 2012, 4,7% em 2013, e até 6% em 2015, principalmente em cidadãos que foram pressionados a votar para candidatos do partido no poder.

Em segundo lugar está o crescimento de pequenos centros na última década. O organismo eleitoral Conselho Nacional Eleitoral (CNE) venezuelano justificou esta política com base na necessidade de descentralizar grandes centros e aumentar o número de pequenos centros em áreas diversas. No entanto, estes centros são mais propensos a irregularidades e atos de intimidação e/ou violência nos dias de eleição. Nestes pequenos centros, os votos são principalmente de cidadãos que dependem dos programas do governo social, o que os torna vulneráveis ao *modus operandi* do mecanismo oficial. Em eleições extremamente competitivas, como a de 2013, uma manipulação nesses centros pode ter sido crítica para o vencedor ganhar a maioria dos votos.

Já nas eleições parlamentares de 2015, os tipos de irregularidades relatadas 2 anos antes tiveram menor impacto. Além da menor mobilização de chavistas, em grande parte devido à crise, dois fatores parecem ter desempenhado um papel importante: a implantação pelas forças da oposição de ativistas e estudantes, particularmente em áreas controladas pelo chavismo e, portanto, mais vulneráveis a possíveis fraudes; e o papel institucional desempenhado pelas forças armadas numa eleição mais afoita por um monitoramento mais forte da comunidade.

Na Rússia, por outro lado, as eleições não atingiram os padrões mínimos

a serem considerados democráticos. A pressão sobre os eleitores é escancarada. Nas eleições parlamentares de 2003, há evidência de pressão sobre vários milhares de trabalhadores que foram instruídos por seus empregadores a solicitar cédulas de ausentes e foram posteriormente transportados no dia da eleição para que seu voto pudesse ser monitorado em centros previamente designados sob ameaça (até) de perda de emprego.

Nas eleições parlamentares de 2007, houveram denúncias de ameaças contra eleitores, de uso indevido de cédulas de votação de eleitores ausentes e de eleitores sendo encaminhados para centros designados. Nas eleições parlamentares de 2011, os funcionários públicos foram convidados a assinar cartas de apoio ao partido no poder. Proprietários de grandes empresas, seguindo as instruções das autoridades locais, também pressionaram seus funcionários para votar na Rússia Unida. (KOBAK, 2012)

Semelhante às eleições de 2011 em que a fraude eleitoral levou a manifestações generalizadas, as eleições presidenciais de 2012 foram marcadas por um grande número de irregularidades, embora com menor impacto nos resultados eleitorais. Comparado com eleições anteriores, foi relatado que, em 2012, houve uma grande tentativa de controlar o voto por tais práticas como voto maciço usando voto de ausente ou exigindo que os funcionários votem em seus locais de trabalho.

Foram observados maiores traços estatísticos de pressão sobre eleitores nas seções de votação nas eleições parlamentares de 2011 em que vários tipos de fraude foram analisados. (KOBAK, 2012). Estimou-se que o tamanho da fraude nas seções eleitorais que favoreceram partido no poder, Rússia Unida, é de 11%. Pesquisa mais recente (COLLIER, 2012) mostrou que a mistura de fraudes eleitorais de manipulação, enchimento de urnas e pressão dos eleitores.

As eleições de 2011 no Uganda exigem uma avaliação totalmente diferente. Este país tem desfrutado de mais um cenário competitivo desde a reintrodução das eleições multipartidárias em 2005, mas manteve as características do autoritarismo eleitoral, a chave para a manutenção do Movimento Nacional de Resistência (GRN) no poder.

Em 2011, três práticas comuns em eleições passadas, também predominantes na África Subsaariana, são observadas: intimidação, compra de votos e manipulação de votos. No contexto da chamada monetização das eleições, a pressão sobre alguns eleitores foi exercida através de suborno (com a ameaça das consequências caso eles não votassem para o candidato que tinha comprado o seu voto; e também através do desenvolvimento programas de governo (ameaçando os eleitores com a perda de benefícios se eles não votassem em quem fora determinado). Neste caso, os traços

estatísticos de suborno assemelham-se aos da manipulação de eleitores nos outros dois países.

É possível notar uma tendência em que quanto menor o centro, maior o partido que compra votos é (Figura 14B). No entanto, o método desenvolvido por JIMENEZ (2017) não contempla a diferença entre a mera compra de votos e a combinação de compra e coerção. Tal método seria igualmente eficaz no diagnóstico irregularidades que são consistentes com outras formas de manipulação supracitadas. Por exemplo, ameaçando e intimidando minorias étnicas ou raciais para ficar ausente das urnas levaria para uma diminuição da participação que coincide com o aumento dos limites das pequenas seções eleitorais.

Essa forma de “negação de voto” deslocaria os IDEPs de pequenas seções para o quadrante superior esquerdo na figura 3.5.4. Este efeito seria captado pela distância padronizada $\delta(p)$ também, como qualquer outra forma de deslocamento sistemático entre os IDEPs de pequenas e grandes seções eleitorais.

No entanto, também se pode imaginar formas de manipulação do eleitor que não serão detectadas pelo método proposto por JIMENEZ (2007). Este é particularmente o caso se os eleitores visados forem “bem misturados” na população, isto é, não se agrupa em pequenas ou grandes seções eleitorais. Essas formas de manipulação dos eleitores não levariam a um deslocamento de os IDEPs correspondentes na figura 3.2.3 e, portanto, não seriam detectados pelo teste desenvolvido por JIMENEZ (2017).

Outra limitação deste método é que, para o teste de significância funcionar, precisa-se definir uma referência conjunto de eleições “confiáveis”. Não é claro a priori se uma determinado eleição pode ser considerada confiável, porque é justamente isso que o teste deve decidir.

No entanto, contorna-se essa questão aqui construindo o conjunto de referência a partir de resultados eleitorais de democracias ocidentais que não apresentaram irregularidades sistemáticas na grande maioria da literatura forense eleitoral. Assim, assumimos que não haja irregularidades nestas eleições, ao invés de observarmos que seu impacto combinado no deslocamento $\delta(p)$ é tipicamente de magnitude menor do que o seu impacto nas eleições, marcado como irregular pelo teste proposto.

3.6 O estudo de possíveis fraudes eleitorais pode ser muito mais frequente do que se possa imaginar

3.6.1 Detectando fraudes eleitorais por irregularidades nas distribuições de votos

ROZENAS (2017) desenvolveu um novo método para detectar fraudes eleitorais de padrões irregulares na distribuição de votos, observando algumas eleições alvo de alegações de fraude. Em muitas seções de votação os resultados foram percentuais arredondados (por exemplo, 0,50, 0,60, 0,75) para o partido no poder, o que parece estranha e altamente implausível em grandes eleições. Usando resultados analíticos e simulações, mostrou-se que a frequência desses resultados é totalmente plausível devido a leis numéricas simples e, por si só, não constitui evidência de fraude. Para evitar erros falsos positivos na detecção de fraude, ROZENAS (2017) propôs um método de densidade de kernel reamostrado (RKD) para medir se resultados estranhos ocorrem com muita frequência para levantar uma suspeita estatisticamente qualificada de fraude. Foram analisados os dados eleitorais da Rússia e do Canadá, bem como dados simulados. Por fim, ainda disponibilizou um pacote de software para uma fácil implementação do método, através da publicação de uma biblioteca (pacote) em linguagem R.

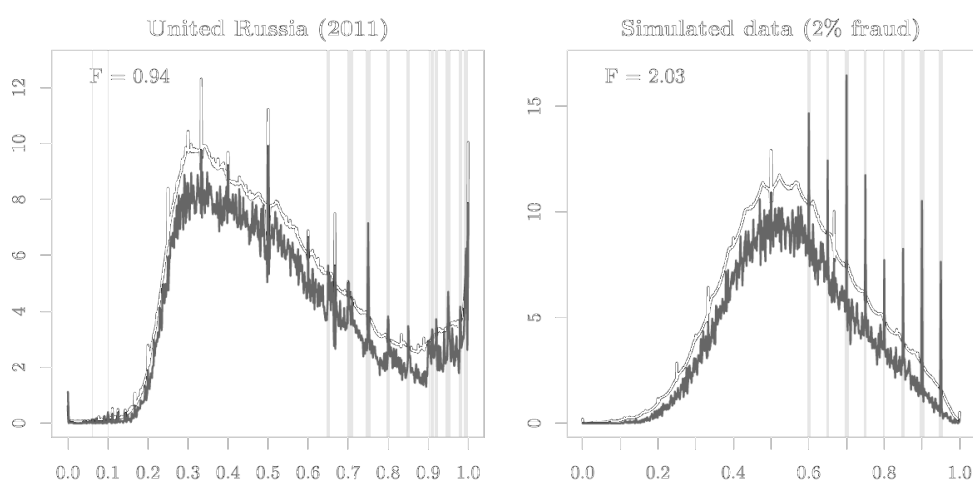


Figura 3.6.1 – Resultados comparativos entre simulação e dados reais da Rússia: densidades de kernel das ações de voto observadas (curva inferior) e seus envelopes superiores reamostrados (curva superior). Barras cinzas verticais indicam resultados potencialmente falsificados. Fonte: ROZENAS, 2017.

A figura 3.6.1 ilustra a mecânica do método RKD. As figuras mostram as densidades de kernel das ações de voto observadas e o envelope superior das densidades de kernel dos dados reamostrados. As barras verticais cinza representam os pontos nos quais a densidade de kernel dos dados observados excede o envelope superior

da densidade de kernel dos dados reamostrados. Por exemplo, uma barra cinza larga localizada em 0,75 indica que muitos distritos em que o partido recebeu 75% dos votos são suspeitos.

O painel esquerdo da figura 3.6.1 mostra os resultados do partido Rússia Unida nas eleições de 2011. A densidade do kernel dos dados observados (curva cinza inferior) excede o envelope superior da densidade do kernel dos dados reamostrados (curva preta superior) em vários pontos, especialmente nos distritos onde o partido Rússia Unida recebeu uma alta proporção de votos (em 0,75, 0,85 e em vários pontos acima de 0,9). Dado que os agentes locais do governo cometeriam fraudes inflacionando o número de votos do partido no poder para um grande número, esse padrão parece plausível. Também é consistente com pesquisas anteriores sobre essas eleições, que concluíram que o governo russo estava cometendo fraude não em distritos marginais, mas em distritos onde tinha forte apoio.

O painel direito da figura 3.6.1 mostra como a medida de fraude é executada em um conjunto de dados simulado com dois por cento dos distritos contaminados por fraude. Primeiro, o método identifica corretamente o conjunto de seções eleitorais com resultados fraudulentos - as barras cinzas verticais estão localizadas precisamente nos pontos em que os dados das eleições foram contaminados na simulação. Segundo, a quantidade estimada de fraude está muito próxima da verdade básica. Terceiro, alguns picos de massa de probabilidade nesses dados (por exemplo, 1/2 ou 2/3) são corretamente identificados como não indicativos de fraude.

Uma das contribuições de ROZENAS (2017) é questionar o quanto podem ou não ser verídicas as suspeitas de fraude eleitoral. Considerou a hipótese nula H_0) a não existência de fraude, deixando para H_1 em diante para outras hipóteses de fraudes, tentando estimar a proporção em que as ações ilícitas influenciaram ou não no resultado final.

3.7 Próximos passos

Após observar este ferramental estatístico matemático, é inevitável não ansiar por aplicá-lo aos dados eleitorais brasileiros, disponíveis no sítio online do TSE - Tribunal Superior Eleitoral. É o que será feito nos capítulos adiante.

Mas não antes de observar atentamente um pouco da história das eleições no Brasil, mas principalmente a urna eletrônica, que causa tantas percepções diversas (e até extremas) no eleitorado deste país.

4 FERRAMENTAS ESTATÍSTICAS

Neste capítulo será trazido à tona alguns conceitos e ferramentas estatísticas que são usadas no desenvolvimento do trabalho.

4.1 Distribuição Gaussiana e Distribuição Lognormal

Distribuição Gaussiana ou Normal é um modelo que descreve o comportamento de vários fenômenos aleatórios, como fenômenos naturais. A importância dessa distribuição se dá por conta do Teorema do Limite Central.

O teorema pode ser enunciado da seguinte maneira: Seja x_1, \dots, x_n , n variáveis aleatórias independentes e identicamente distribuídas com média μ e desvio padrão σ . A distribuição da soma das n variáveis tende a apresentar um comportamento normal, como

$$x_1 + x_2 + \dots + x_n = \sum x_i \sim normal. \quad (4.1)$$

Ou seja, as médias de amostras retiradas de distribuição qualquer tendem a convergir para uma distribuição normal quando o número de observações (tamanho da amostra) aumenta (BITTENCOURT, 2006).

A função densidade dessa distribuição é denotada como

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad (4.2)$$

onde $\mu \in \mathfrak{R}$ é a média, $\sigma > 0$ o desvio padrão e $x \in \mathfrak{R}$ (BITTENCOURT, 2006). Uma distribuição é gaussiana quando se apresenta na forma de um “sino”.

Uma variável X tem distribuição lognormal quando o seu logaritmo $Y = \log(X)$ tem distribuição normal. A função de densidade desta distribuição é denotada da seguinte maneira:

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right], \quad (4.3)$$

com $x > 0$ e 0 caso contrário.

A distribuição lognormal tem a sua relevância devido a um resultado análogo do teorema central do limite. Enquanto a distribuição gaussiana ocorre quando são somadas várias variáveis independentes, a distribuição lognormal ocorre como um produto de várias variáveis independentes (sempre positivas).

4.2 Lei de Potência ou Escala

Para sistemas complexos não há uma definição precisa, porém, em geral, eles apresentam propriedades emergentes que decorrem em grande parte da interação não-linear entre suas partes. Assim, suas propriedades apenas se tornam notáveis quando vistos como um todo. Uma das propriedades marcantes de muitos desses sistemas é a presença de Leis de Escala ou Leis de Potência.

Uma lei é dita Lei de Escala ou Lei de Potência se entre dois escalares Y e N a relação é escrita da seguinte maneira:

$$Y = Y_0 N^\beta, \quad (4.4)$$

onde Y_0 (constante de proporcionalidade) e β (expoente) são constantes.

A Lei de Escala é expressa por uma reta em um gráfico log-log, pois a Equação 5.4 pode ser escrita como:

$$\log(Y) = \log(Y_0) + \beta \log(N) \quad (4.5)$$

que é a mesma forma da equação de uma reta.

Outro conceito a se observar é o de **intercepto**. Intercepto é o ponto em que uma linha *plotada* pela função descrita pela equação que relaciona os dois escalares se encontra com o eixo das ordenadas (y), ou seja, a resultante para $x = 0$. O intercepto é resultado direto do ângulo de inclinação da reta e da constante que (se houver) a define como mais distante do eixo das abcissas (x). A reta resultante da relação escalar entre duas variáveis, em nosso caso, é originada de uma regressão linear entre as variáveis originadas na base de dados das seções eleitorais, como: votos brancos e nulos, votos no vencedor, taxa de comparecimento, dentre outras, conforme descrito no próximo capítulo.

4.3 Teste Kolmogorov-Smirnov

Criado pelos russos Andrei Kolmogorov e Nikolai Smirnov, o teste de Kolmogorov-Smirnov (Teste KS ou Teste K-S) é um teste estatístico não paramétrico sobre a igualdade (ou diferença) de distribuições de probabilidade contínuas e unidimensionais (uma sequência de amostras com apenas uma variável) aplicáveis na comparação de uma amostra com uma distribuição de probabilidade usada como referência (conhecida) ou com outra amostra (o que faz com que o teste possa ser biamostrais). Este teste permite avaliar quantitativamente a distância entre (as curvas geradas por) duas funções. (FELLER, 1948)

O teste consiste se confirmar a hipótese 0 (zero) – quando os dados se assemelham aos dados da distribuição de referência – ou a hipótese 1 (um – quando os dados não se assemelham aos dados da distribuição de referência. Para tanto, faz-se necessário somar as distâncias entre cada respectivo ponto da imagem de cada função, da seguinte forma:

$$D = \sup_x |F_0(x) - F_{data}(x)| \quad (4.6)$$

O D resultante desta fórmula é um conjunto de elementos (unidimensional) onde cada elemento é a distância Y entre um mesmo X em cada curva. O somatório de todos os valores de D quando se aumenta a precisão de X tende à integral que é a área entre as curvas. (SMIRNOV, 1939)

Figura 4.3.1 – Aplicação da Kolmogorov-Smirnov

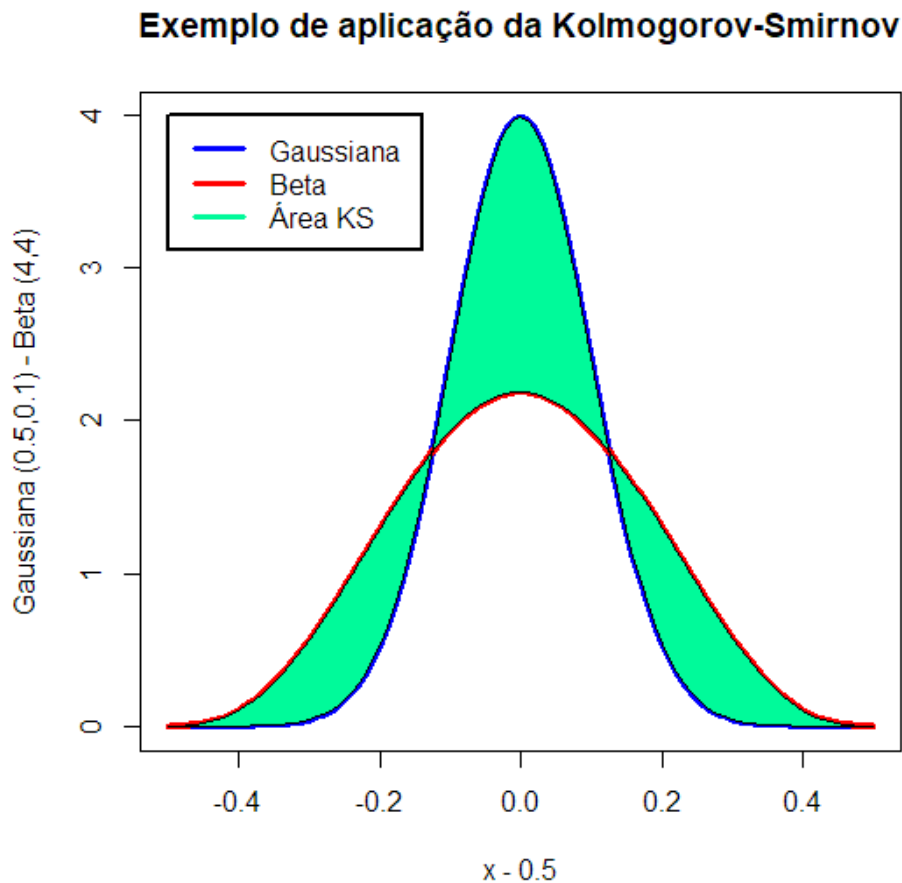


Figura 4.3.1 – Exemplo de utilização (usando as distribuições Beta e Normal Gaussiana) calculando de área entre entre curvas geradas proposto por Komogorov-Smirmov. Fonte: Própria.

4.4 Distribuição Beta

Distribuição Beta é um conjunto de distribuições de probabilidade contínuas definidas no intervalo fechado ou aberto de zero a um, e possuem dois parâmetros positivos (α e β) presentes como expoentes da variável aleatória, e controlam o formato da distribuição. A função pode ser, por inferência bayesiana, ser a conjugada de outras distribuições como Bernoulli, binomial, binomial negativa e geométrica.

A aplicabilidade desta distribuição tem sido vastamente explorada em uma grande quantidade de áreas da ciência, mais especificamente para modelar o comportamento de variáveis aleatórias limitadas a intervalos de tamanho finito.

A função densidade de probabilidade para a distribuição beta é a seguinte:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha + \beta)} x^{\alpha-1} (1-x)^{\beta-1} \quad (4.7)$$

Figura 4.4.1 – Distribuição Beta

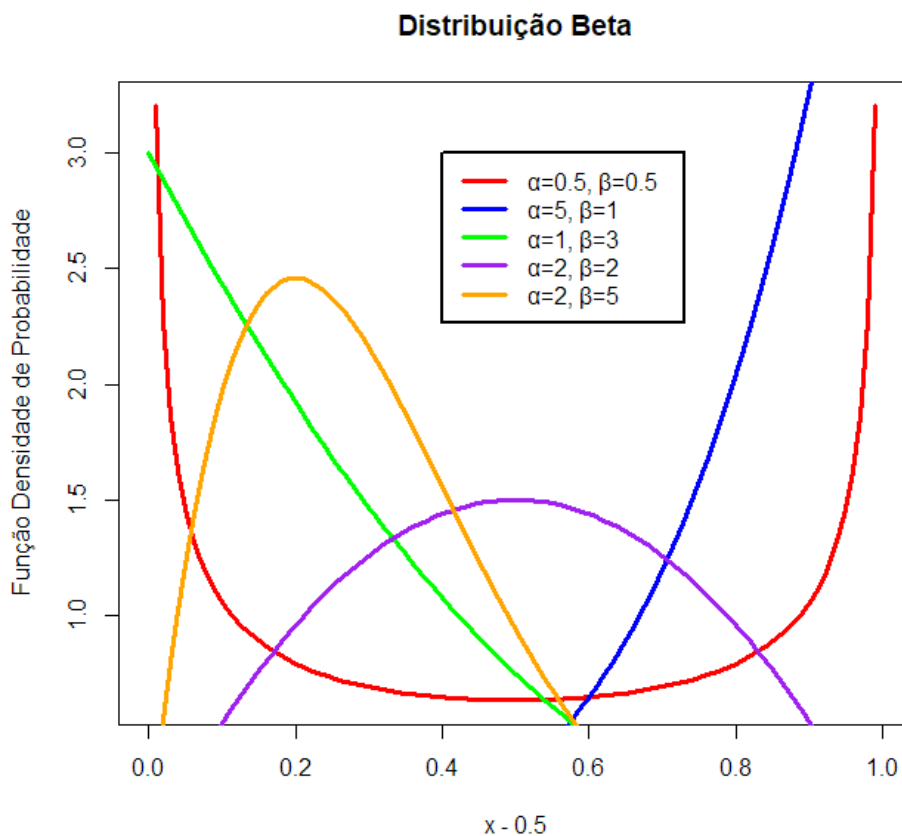


Figura 4.4.1 – Plotagem Distribuição Beta com diferentes parâmetros α e β . Fonte: Própria.

4.5 Distribuição Weibull

A distribuição de Weibull é uma distribuição de probabilidade contínua, criada em 1951 por Waloddi Weibull, utilizada em uma vasta gama de aplicações das mais diversas áreas da ciência. Esta distribuição é bastante usada em engenharia de confiabilidade, determinação de tempo de vida médio e taxa de falhas em função do tempo da população analisada. Dentre várias propriedades, uma das mais buscadas é a capacidade de fazer previsões de acurácia razoável mesmo com quantidade pequena de dados disponíveis.

A sua função de densidade de probabilidade é dada por:

$$f(x; \lambda; k) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0, \\ 0 & x < 0. \end{cases} \quad (4.8)$$

Figura 4.5.1 – Distribuição Weibull

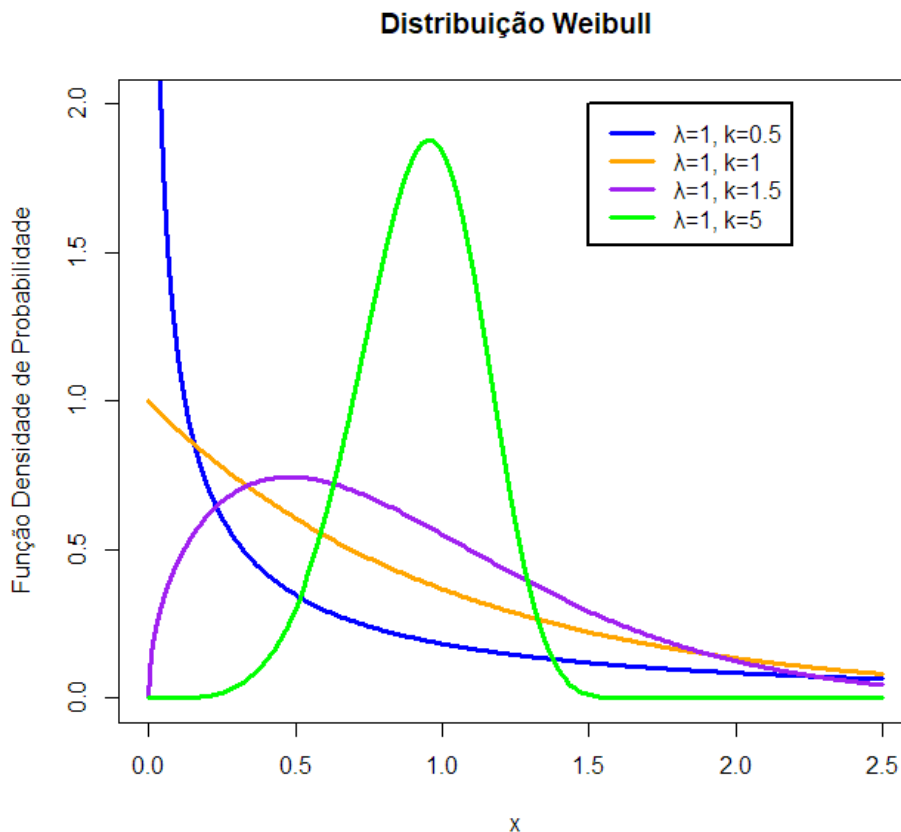


Figura 4.5.1 – Plotagem Distribuição Weibull com diferentes parâmetros λ e k . Fonte: Própria.

O parâmetro λ é medido na mesma unidade de x . Quando aplicado a áreas que envolvem tempo, λ representa este tempo característico da aplicação. Estatisticamente,

λ é determinado parâmetro de escala, pois com k constante, a variação de λ causa a compressão ou expansão do gráfico.

O parâmetro k é adimensional e é chamado de declividade, já que determina a declividade da função de distribuição acumulada plotada. Estatisticamente, k é o parâmetro de forma, uma vez que k altera drasticamente o comportamento da distribuição.

Ambos os parâmetros k e λ são definidos para o intervalo de 0 a $+\infty$.

4.6 Assimetria e Curtose

Assimetria e Curtose permitem, em conjunto com as medidas de posição e dispersão, descrever e compreender completamente as distribuições de frequências. CRESPO (2002). As distribuições de frequência são diferentes entre si quanto ao valor numérico, à variabilidade e também quanto à sua forma. Assimetria e curtose são maneiras de olhar numericamente para a forma da distribuição.

4.6.1 Assimetria

Uma distribuição é simétrica quando os valores da média, mediana e moda são iguais. $\bar{x} = Md = Mo$ ou $\bar{x} = \hat{x} = \tilde{x}$

Figura 4.6.1 – Média, Mediana e Moda

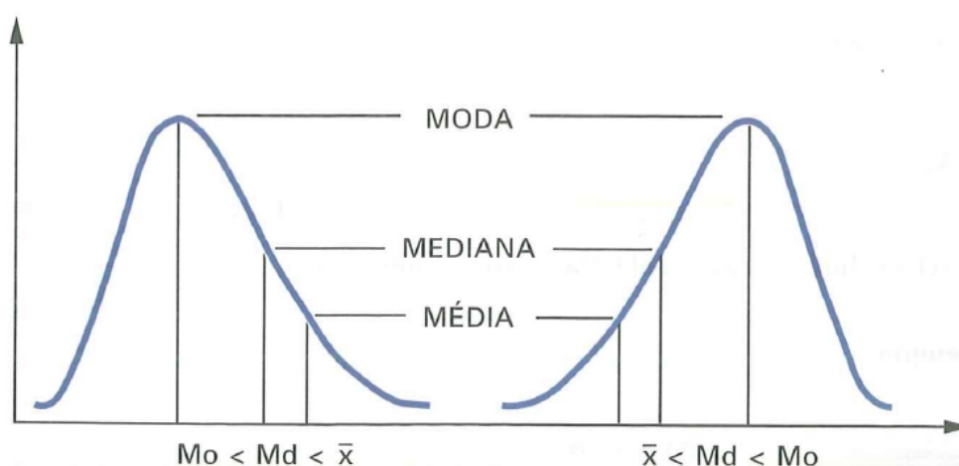


Figura 4.6.1 – Curvas normais assimétricas mostrando a média (\bar{x}), mediana (M_d) e moda (M_o). Fonte: CRESPO, 2002.

Para calcular a média (\bar{x}), mediana (M_d) e moda (M_o), considera-se uma sequência com a distribuição onde cada elemento é a quantidade de ocorrências nesta distribuição. A média é calculada somando-se todos os elementos e dividindo-se esta soma pela quantidade de elementos (PLACKETT, 1958), seja de uma mostra ou de

uma população. A mediana é o elemento central desta sequência (WRIGHT, 1599). A moda é o valor que mais ocorre na sequência. (PEARSON, 1895).

Com a média, mediana e moda, defini-se a simetria ou não, e o tipo de curva quanto a assimetria:

- Simétrica quando $\bar{x} = M_d = M_o$;
- Assimétrica positiva quando $\bar{x} > M_d > M_o$;
- Assimétrica negativa quando $\bar{x} < M_d < M_o$.

Para ser possível comparar as assimetrias, PEARSON (1895) propôs um coeficiente de assimetria:

$$A_s = \frac{3.(\bar{x} - M_d)}{S} \quad (4.9)$$

, onde A_s é o coeficiente de simetria, S é o desvio padrão, \bar{X} é a média e M_d é a mediana. Se o coeficiente de assimetria (A_s) estiver entre 0,15 e 1 entende-se que é uma assimetria moderada; e se o coeficiente de assimetria (A_s) for maior que 1 entende-se que é uma assimetria forte.

4.6.2 Curtose

Curtose é o grau de achatamento de uma distribuição em relação à distribuição normal padrão (CRESPO, 2002), ou, uma medida de forma que dá característica o achatamento da curva da função de distribuição de probabilidade (CASELLA, 2010). É definida por:

$$\frac{m_4(\mu)}{\sigma^4} + (-3) \quad (4.10)$$

onde $m_4(\mu)$ é o quarto momento central e σ é desvio-padrão.

Classificação da distribuição de dados:

- Leptocúrtica: quando o valor é maior que zero, a distribuição é mais alta e estreita, mais concentrada que a distribuição normal;
- Mesocúrtica: quando o valor é zero, tem o mesmo achatamento que a distribuição normal;
- Platicúrtica: quando o valor é menor que zero, sendo mais "achatada"(baixa) que a distribuição normal.

4.7 Entropia Shannon

Entropia é um termo originalmente advindo da física, mais especificamente da termodinâmica, que mede o grau de liberdade molecular de um sistema, e está associado ao seu número de possibilidades de distribuição de partículas nos mais diversos aspectos observáveis. Por esta definição, entropia também é relacionada a aleatoriedade, dispersão de matéria e energia, e "desordem" de um sistema (termodinâmico), ou seja, é o grau de irreversibilidade de um determinado sistema. Então, quanto menor a chance do sistema voltar ao seu estado original, maior será o grau de entropia.

Na estatística, a entropia foi utilizada inicialmente por Ludwig BOLTZMANN (1877) que visualizou um método probabilístico para medir a entropia de um determinado número de partículas de um gás ideal, na qual ele definiu entropia como proporcional ao logaritmo neperiano do número de microestados que um gás pode ocupar, da seguinte forma:

$$S = k \cdot \ln(\omega) \quad (4.11)$$

Onde S é a entropia, k é a constante de Boltzmann e ω é o número de microestados possíveis para o sistema.

Claude SHANNON (1948) apresentou uma teoria cujo conceito central é a entropia de uma fonte de informação, determinando o que hoje é conhecido como Teoria da Informação, assim o conceito de entropia ganhou nova aplicação definindo como armazenar e transmitir informação de maneira mais econômica. Tal aplicação de entropia não deixou mais dúvida de que entropia é muito útil no contexto de probabilidades e nos sistemas de informação, e não necessariamente em teorias físicas como Termodinâmica ou Mecânica Estatística. (LIMA, 2004).

Esta publicação de Shannon, que inicialmente era chamada de teoria da comunicação, pelo contexto e proposta, veio a ser um marco da revolução digital desenvolvida nas décadas subsequentes, na medida que todo dispositivo contém microprocessador(es) e/ou microcontrolador(es), conceitualmente um resultado direto descendente desta teoria. Shannon esteve em contato com Alan Turing quando este esteve nos Estados Unidos em 1943 compartilhando os feitos em criptoanálise que levou os aliados à vitória e ao fim da 2ª guerra mundial. Na oportunidade, Shannon conheceu o artigo que Turing estava produzindo que tornou-se, posteriormente, conhecido como Máquina Conceitual de Turing, mecanismo que embasou a Teoria da Computação tal como conhecemos hoje.

4.8 Conclusão do capítulo

Tendo visto rapidamente estas ferramentas estatísticas, com seu funcionamento (e em algumas um pouco de sua história para que se possa dimensionar o quanto consagradas são tendo transcorrido tanto tempo de uso) é possível observar sua aplicação no capítulo adiante que trás o desenvolvimento do trabalho.

5 ENVOLVIMENTO CÍVICO EM ELEIÇÕES PRESIDENCIAIS BRASILEIRAS

Os conteúdos vistos anteriormente conduziram e possibilitaram o estudo sobre eleições que será apresentado no presente capítulo. O capítulo 2 apresenta um levantamento histórico do sistema eleitoral brasileiro, o capítulo 3 contém uma revisão de alguns artigos analisando o comportamento do eleitorado em eleições de vários países do mundo, e o capítulo 4 mostra ferramentas estatísticas utilizadas nas análises apresentadas a seguir.

Eleições representam fenômenos sociais de grande interesse para todas as pessoas em todos os países do mundo, pois o destino de uma sociedade depende do resultado delas. Em uma sociedade democrática a escolha de seus representantes é realizada de forma livre e justa através de eleições em que o voto de cada cidadão deve contar igualmente. Porém existem países que embora adotem o sistema democrático a elite política dominante direciona o resultado da eleição de acordo com seus interesses de forma subjacente. A possibilidade de coerção e mobilização de eleitores, de forma tácita, por parte dessas elites direciona o interesse de pesquisadores para a investigação de resultados de eleições.

Além disso, eleições nacionais representam experiências sociais em grande escala, onde as pessoas são agrupadas por uma divisão eleitoral constituindo um grande número de pequenos grupos de eleitores para votarem de acordo com suas preferências. Tais resultados formam grandes bases de dados que analisadas como um todo pode apresentar comportamentos sistemáticos, isto é, uma análise desse fenômeno visto como um sistema complexo pode apontar regularidades ou irregularidades, caso existam.

O sistema eleitoral brasileiro, como visto no capítulo 2, conta com praticamente 100% de urnas eletrônicas, o que torna os processos de votação e apuração muito ágil, porém eleitores e políticos, em especial na última eleição presidencial, manifestaram desconfiança em relação aos resultados de votação em urnas eletrônicas, apesar da ênfase que o Tribunal Eleitoral do Brasil – TSE (BRASIL, MV, 2019), dá a confiabilidade do uso desses equipamentos.

Levando em conta tal polêmica e estudando trabalhos que realizaram investigações nesse contexto (BORGHESI, 2010; BORGHESI, 2012; KOBAK, 2012; KLIMEK, 2017; JIMENEZ, 2017; ROZENAS, 2017), a pesquisa realizada nesse trabalho foi direcionada a analisar o envolvimento do eleitorado em eleições presidenciais brasileiras no segundo turno de votação para as últimas 5 eleições que ocorreram a cada quatro anos de 2002 a 2018, para as quais os dados brutos são disponibilizados no site do TSE

(BRASIL, BD, 2019). Os procedimentos usados aqui foram iniciados por uma análise da fração de votos do primeiro colocado e da fração de eleitores que compareceram para votar em cada urna do Brasil, com o intuito de observar indícios de fraude, caso existissem, no entanto os resultados não apresentaram indicativo de mecanismos de fraude como encontrado em eleições da Rússia e da Uganda por (KLIMEK, 2017).

Contudo o procedimento apresentou, de forma subjacente, algumas regularidades e irregularidades no comportamento do eleitorado nas urnas. Na direção de fazer uma investigação mais detalhada, foram realizadas análises das distribuições de algumas variáveis relacionadas as proporções de votos, de comparecimento e de votos brancos e nulos, apresentando comportamentos estáveis com comportamento anômalo para a eleição de 2018 e uma maior participação do eleitorado em 2002. Foi também realizada uma análise da entropia para as variáveis conservativas: fração de votos válidos, fração de comparecimento e fração de brancos e nulos, obtendo uma medida da “desordem” na distribuição dessas variáveis. Esse procedimento permitiu observar e comparar o envolvimento cívico do eleitorado em cada uma das cinco eleições presidenciais analisadas indicando um comportamento diferenciado entre pequenos e grande eleitorados.

Para finalizar as investigações aqui, foi usado um procedimento para relacionar o número de votos do primeiro e do segundo colocados com o número de votos brancos e nulos, através do comportamento médio por faixas de votos dos candidatos, apresentando um modelo analítico lei de potência. Os procedimentos realizados nesse estudo apresentaram resultados que reforçam a fragilidade da tese de fraude nas urnas em eleições no Brasil e ajudam a compreender o envolvimento cívico do eleitorado em cada ano de eleição.

5.1 Apresentação dos Dados

A base de dados foi obtida do repositório de dados eleitorais, uma compilação bruta dos dados referentes as eleições brasileiras, disponibilizado no site do Supremo Tribunal Eleitoral (BRASIL, BD, 2019).

Os dados filtrados constam das variáveis: número de eleitores cadastrados (aptos), comparecimento, votos no primeiro colocado, votos no segundo colocado, brancos e nulos, por urna, nos segundos turnos de eleições presidenciais do Brasil nos anos de 2002, 2006, 2010, 2014, 2018. Às análises que serão apresentadas nesse trabalho se restringem a esses anos devido as eleições para cargos políticos no Brasil ocorrerem a cada quatro anos, anos pares, e a completude dos dados disponíveis, apenas a partir de 2002 os dados disponíveis estão completos. A Tabela 2 apresenta os dados totais para cada ano analisado.

Ano	Qt urnas	Qt Aptos	% C	% VP_A	% VP_C	% VS_A	% VS_C	% V_BN_A	% V_BN_C
2002	320.122	115.174.456	78,8	56,4	57,6	37,4	36,4	6,0	6,2
2006	360.875	125.795.016	80,7	58,0	57,1	35,9	36,8	6,0	6,0
2010	398.986	135.493.224	78,4	53,3	52,3	40,1	40,1	6,7	6,5
2014	427.734	142.445.761	78,9	49,7	48,4	44,1	45,2	6,3	6,1
2018	453.372	146.754.108	78,8	48,8	49,8	41,8	40,6	9,6	9,4

Tabela 2 – Valores totais dos dados: As colunas apresentam os valores totais por ano das variáveis quantidades de urnas representada por **Qt urnas**, quantidade aptos (eleitores cadastrados na urna) por **Qt Aptos**, porcentagem de eleitores que compareceram para votar por **%C**, porcentagem de votos para o primeiro colocado em relação aos aptos por **%VP_A** e em relação ao comparecimento por **%VP_C**, porcentagem de votos para o segundo colocado em relação aos aptos por **%VS_A** e em relação ao comparecimento por **%VS_C**, porcentagem de votos brancos mais nulos em relação aos aptos por **%V_BN_A** e em relação ao comparecimento por **%V_BN_C**.

5.2 Análise dos Dados

A investigação do comportamento do eleitorado nas urnas nas 5 eleições presidenciais brasileiras da base de dados, foi iniciada desenvolvendo um procedimento usado por KLIMEK (2017) para investigar indicativo de fraude nas urnas em eleições presidenciais de 11 países. Foram construídas as densidades bidimensionais da variável $f_c = C/A$ fração de comparecimento (C é o número de eleitores que compareceram para votar e A o número de eleitores aptos a votar em uma dada urna), versus $f_p = V_p/C$, fração de votos do vencedor (V_p é o número de votos para o vencedor na urna), onde a intensidade da cor representa o número de urnas com dadas frações f_c e f_p . Na análise realizada em KLIMEK (2017) os histogramas para duas eleições da Rússia e uma da Uganda apresentaram um aglomerado de urnas no canto superior direito, o que significa um número considerável de urnas contendo aproximadamente 100% de comparecimento e 100% de votos para o vencedor. Os autores sugerem dois mecanismos de fraude: fraude incremental e extrema, como já mencionado no capítulo 3. Fraude incremental significa que, com uma determinada taxa, votos para um partido são adicionados à urna e/ou votos para outros partidos são retirados. Fraude extrema corresponde a detectar uma participação completa e quase todos os votos para um único partido.

No entanto, como pode ser observado na Figura 5.2.1, para as eleições presidenciais no Brasil a taxa de comparecimento se concentra em torno de 80% e a taxa de votos para o vencedor se apresenta concentrada abaixo de 90%, não apresentando aglomerado nos extremos das variáveis para nenhum ano de eleição, exibindo um comportamento esperado para as variáveis, e conseqüentemente mostrando que não

há indicativo dos mecanismos de fraude detectados para eleições da Rússia e da Uganda. No entanto é possível notar uma distribuição similar nos anos 2006, 2010 e 2014, enquanto os anos 2002 e 2018 mostram uma configuração diferenciada. Para comparar os intervalos de concentração dessas frações e também das frações de votos brancos e nulos e de votos para o segundo colocado são mostradas algumas medidas de posição na Tabela 3.

Observe que a média de f_C é aproximadamente 78% sobre os 5 anos de eleições, sendo em 2006 sensivelmente acima, próximo a 80%, o que era esperado, pois o sistema eleitoral brasileiro determina que o voto é obrigatório para todos os brasileiros, não analfabeto, entre 18 e 70 anos, apesar de não haver muita dificuldade em justificar a ausência para votar. Note também que 2006 tem a maior taxa média de comparecimento e menor taxa média de votos brancos e nulos, 2006 tem a eleição com maior participação e menor rejeição aos candidatos entre todas as eleições analisadas. Pode ser visto ainda que 99% das frações f_C e f_p estão abaixo de 93% e 90%, respectivamente, para todas as eleições confirmando que não há uma quantidade significativa de urnas com uma fração “massiva” (muito próximo a 100%) de comparecimento e votos para o vencedor, mostrando a ausência do mecanismos de fraude extrema (nos mesmos padrões dos pleitos na Rússia e Uganda) para eleições presidenciais brasileiras analisadas. Porém, observando um comportamento aparentemente “irregular” para as eleições de 2002 e 2018, parece interessante investigar melhor a relação entre comparecimento e votos para o primeiro colocado nas urnas, já que parece haver um padrão entre as três eleições intermediárias, aspecto observado na figura 5.2.1.

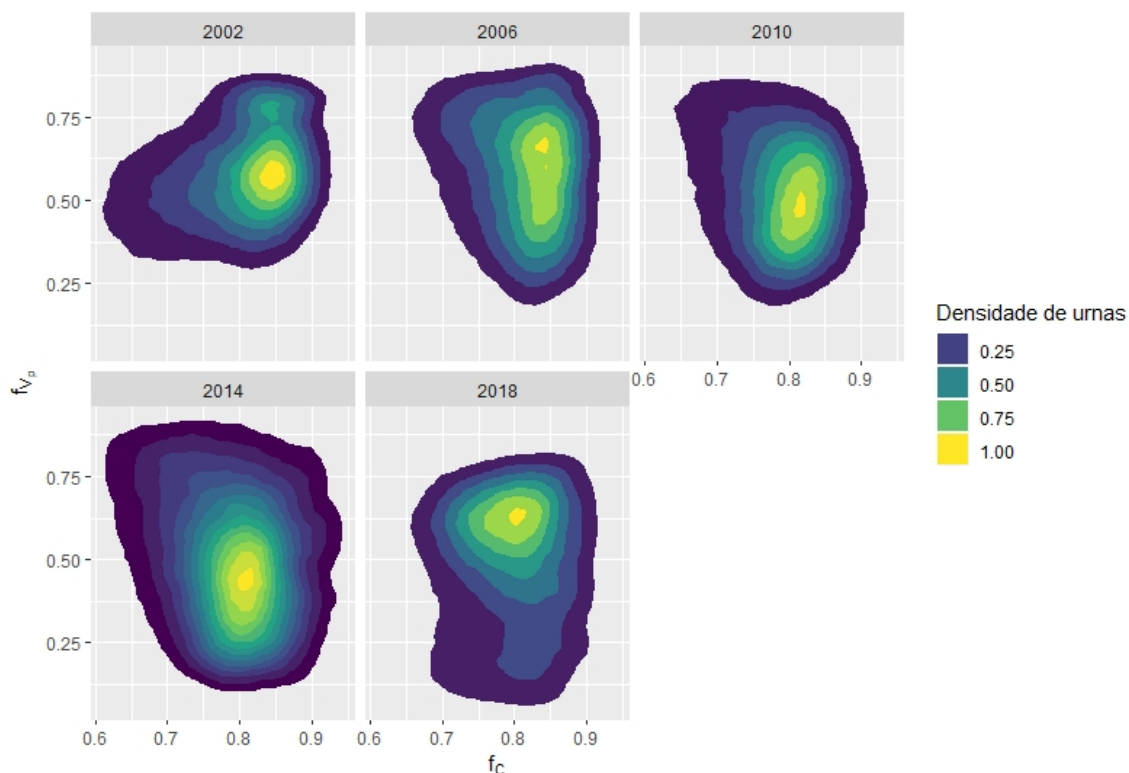


Figura 5.2.1 – Densidade bidimensional da taxa de comparecimento versus taxa de votos do vencedor: O eixo-x representa os valores da variável f_c , número de eleitores que compareceram para votar sobre o número de eleitores aptos na urna, e o eixo-y representa os valores da variável f_p , número de votos para o primeiro colocado sobre número de eleitores que compareceram para votar na urna, a intensidade da cor, aumentando do azul para o amarelo representa o número de urnas com essas dadas frações.

Para isso construímos a distribuições de probabilidade normalizadas da variável

$$\nu = \log_{10}\left(\frac{C - V_p}{V_p}\right) \quad (5.1)$$

para todos os anos de eleições da nossa base de dados, juntamente com a distribuição normal padrão, como pode ser visto na figura 5.2.2(A). Uma variável similar a essa foi utilizada por KLIMEK (2017), que mostrou que a distribuição de probabilidades para países que apresentaram indícios de fraude possuem curtose bem superior a dos outros países. Observe que a variável ν tende a $-\infty$ quando V_p tende a C e tende a ∞ quando V_p tende a 0, omitimos os dados de um número pequeno de urnas para as quais $C = V_p$ e $V_p = 0$, para as quais ν não está definida. Para os anos 2006, 2010 e 2014 as distribuições colapsam e são bem ajustadas pela distribuição normal, o que não ocorre para os anos de 2002 e 2018, o ano de 2018 é o que mais se afasta do padrão normal como pode ser observado nas distribuições acumuladas mostradas na figura 5.2.2(B), juntamente com a distribuição acumulada da normal padrão, (distribuição normal de

Variáveis	Ano	1º percentil	1º quartil	2º quartil = Mediana	Média	3º quartil	99º percentil
f_c	2002	0,500	0,740	0,811	0,788	0,853	0,930
	2006	0,601	0,740	0,819	0,807	0,855	0,929
	2010	0,574	0,747	0,794	0,784	0,831	0,917
	2014	0,588	0,748	0,795	0,789	0,837	0,930
	2018	0,588	0,744	0,795	0,788	0,840	0,922
f_{bn}	2002	0,014	0,042	0,056	0,061	0,074	0,159
	2006	0,008	0,034	0,051	0,060	0,078	0,169
	2010	0,011	0,043	0,061	0,064	0,082	0,152
	2014	0,009	0,038	0,055	0,061	0,076	0,174
	2018	0,019	0,063	0,088	0,093	0,120	0,204
f_p	2002	0,184	0,462	0,564	0,564	0,671	0,866
	2006	0,192	0,449	0,593	0,580	0,717	0,904
	2010	0,188	0,408	0,528	0,533	0,653	0,895
	2014	0,138	0,350	0,487	0,496	0,635	0,903
	2018	0,048	0,345	0,528	0,488	0,643	0,817
f_s	2002	0,084	0,266	0,374	0,374	0,472	0,752
	2006	0,046	0,210	0,345	0,359	0,494	0,765
	2010	0,059	0,272	0,403	0,401	0,527	0,762
	2014	0,056	0,292	0,442	0,441	0,590	0,827
	2018	0,118	0,255	0,358	0,418	0,559	0,911

Tabela 3 – Quartis e médias das variáveis f_c , f_{bn} , f_p e f_s em cada ano de eleição da base de dados.

média $\mu = 0$ e desvio padrão $\sigma = 1$). Com o intuito de confirmar a qualidade dos ajustes foi aplicado o teste de Komolgorov-Smirnov, apresentado no capítulo 4. No entanto não é possível usar todos os dados por exceder a capacidade para a função `ks.test` do software R no cálculo do p – valor exato, por isso o procedimento foi realizado com 20 amostras de tamanho 1000, e para tornar a amostra mais representativa, a amostragem foi realizada por conglomerado (por estado), sorteando um número proporcional ao respectivo número de urnas em cada estado de maneira a obter aproximadamente 1000 urnas, tomando as médias das estatísticas (D e p – valor) do teste Kolmogorov-Smirnov. Pode-se considerar o ajuste aceitável com 95% de confiança se $D \leq 0,043$ e $p \geq 0,05$, os valores obtidos para a distribuição de em cada ano estão na tabela 4. Isso reforça o aspecto visual das distribuições permitindo concluir que, exceto na eleição de 2018, que teve um comportamento anômalo comparado as outras eleições, para as outras quatro eleições da base de dados a variável v tem distribuição aproximadamente normal.

Ano de Eleição	2002	2006	2010	2014	2018
Estatística D	0,03	0,02	0,02	0,02	0,08
Estatística p -valor	0,17	0,40	0,28	0,32	$2,28 \times 10^{-6}$

Tabela 4 – Estatísticas do teste de Komolgorov-Smirnov para o ajuste pela distribuição normal da variável v para amostras de tamanho 1000 por conglomerados (por estados federativos). Os valores das estatísticas representam as médias para 20 amostras, sendo que para os 4 primeiros anos $D < 0,04$ e p – valor $> 0,05$, enquanto em 2018 ocorre o contrário ($D > 0,04$ e p – valor $< 0,05$), isso para todas as amostras.

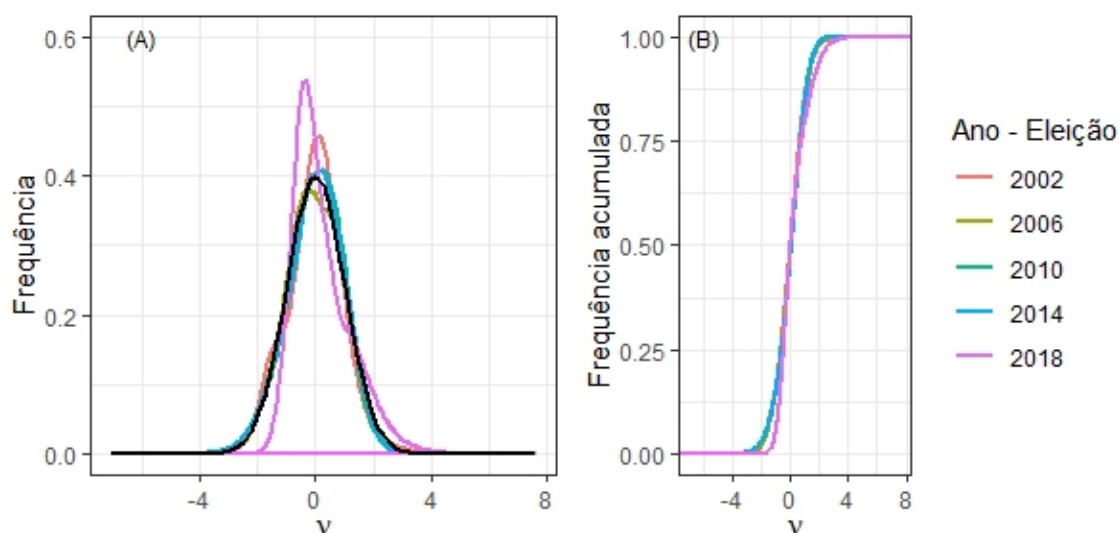


Figura 5.2.2 – Distribuição da variável v normalizada para os 5 anos de eleições da base de dados: (A) Distribuição da variável definida pela Equação 5.1. (B) Distribuição acumulada da variável v . Onde a curva para cada ano, nas duas figuras, tem a cor disponibilizada na legenda a direita das figuras e a curva em preto representa a distribuição normal padrão em (A) e a acumulada da normal padrão em (B).

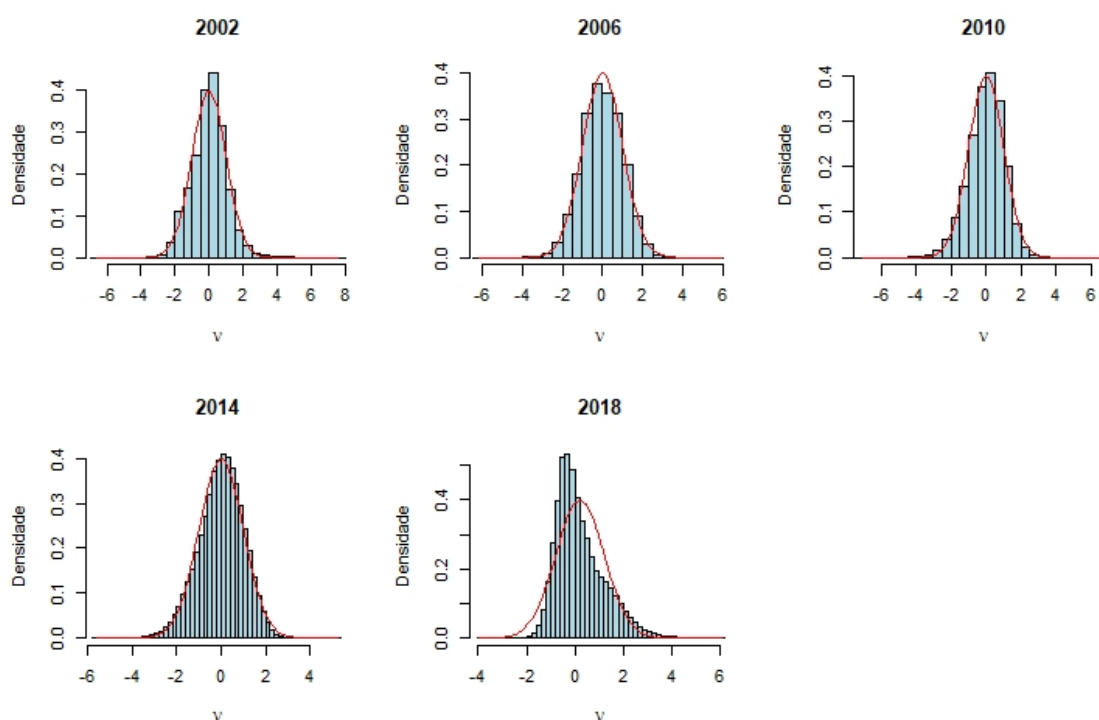


Figura 5.2.3 – Histograma da variável v ajustados pela distribuição normal para cada ano de eleição da base de dados: Cada gráfico em azul representa o histograma para 40 classes e a curva em vermelho a distribuição normal padrão em cada um dos 5 anos de eleição analisados.

Foi calculada também a curtose e a assimetria das curvas, figura 5.2.4. Note que os anos 2002, 2006, 2010 e 2014 apresentam uma assimetria entre 0 e -0,3 enquanto 2018 tem uma assimetria positiva acentuada, próxima a 1. A curtose no entanto se encontra entre 3 e 3,4 para os anos 2006, 2010 e 2014 enquanto 3,8 para 2002 e acima de 1 para 2018. No que 2002 tem a assimetria muito próxima a zero porém uma curtose consideravelmente alta o que mostra uma grande concentração de urnas com v próxima da média, no caso de 2018 que apresentou assimetria e curtose positivas indicando valores mais frequente abaixo da média o que mostra uma grande concentração de urnas com V_P acima de 50%. O que pode ser notado facilmente observando a figura 5.2.1.

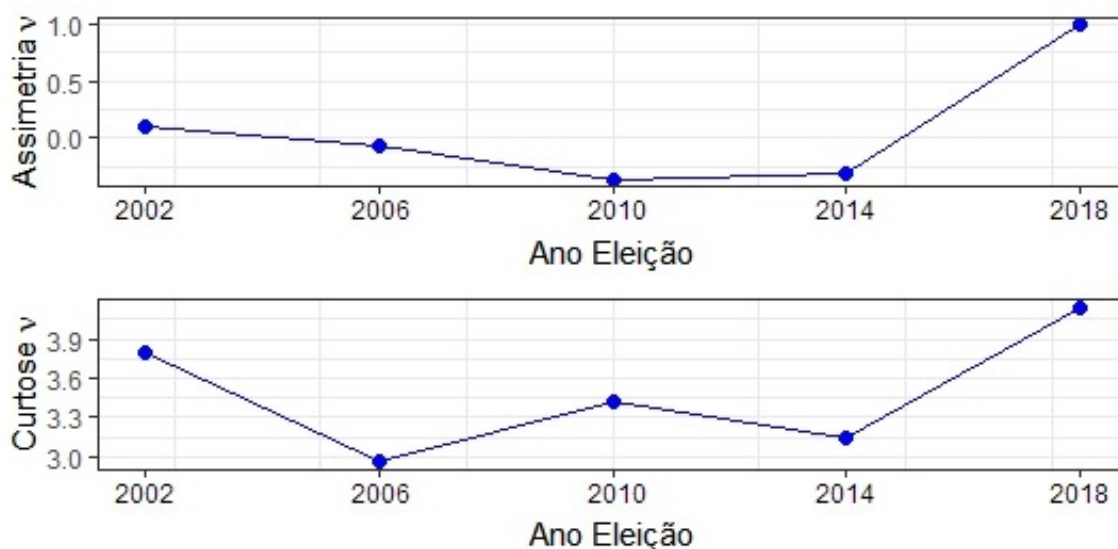


Figura 5.2.4 – Assimetria e Curtose da distribuição da variável v : O gráfico superior mostra os valores da assimetria e o gráfico inferior os valores da curtose das distribuições da variável v para cada ano da base de dados.

Os procedimentos mostrados até aqui apresentam um padrão de regularidade na distribuição de votos para às eleições dos anos 2006, 2010 e 2014 e comportamentos irregulares nos anos 2002 e 2018. Para aprofundar ainda mais a compreensão de como se dá a distribuição dos votos na urna, lembramos que no segundo turno de eleições brasileiras o comparecimento se distribui em três tipos de votos, votos para o primeiro colocado, votos para o segundo colocado e votos brancos e nulos. Exceto em alguns casos de erros do eleitor, brancos e nulos representam eleitores que consideram que os dois candidatos classificados para o segundo turno os não servem para o representar, o que é entendido como voto de “protesto”.

Portanto entender como acontece a distribuições do comparecimento e dos votos brancos e nulos nas urnas pode ajudar a compreender e comparar como se deu a formação de opinião nas eleições da base de dados. A Figura 5.2.5(A) mostra a distribuição e a média da variável f_C nos 5 anos de eleições analisadas e Figura 5.2.5(B) suas distribuições acumuladas, (BORGHESI, 2010) calculou a distribuição da variável comparecimento reescalada onde obtiveram o padrão normal estável para várias eleições francesas. Note que a distribuições são similares com médias muito próximas a 80%, o que era esperado, como observado anteriormente, devido ao sistema de voto obrigatório adotado no Brasil. A Tabela 3 mostra o valor das médias que mais precisamente é 78% para todos os anos exceto para 2006 que é sensivelmente acima, 80%, na Figura 5.2.5(B) as distribuições acumuladas confirmam a similaridade entre as distribuições para todos os anos.

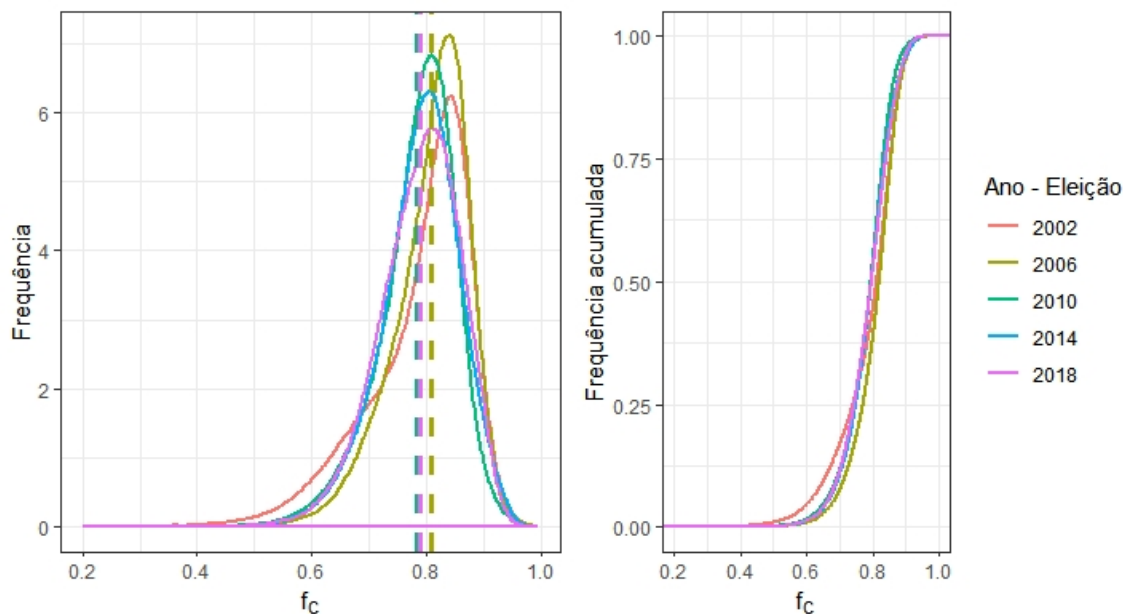


Figura 5.2.5 – Distribuição da variável f_c —para todas as eleições da base de dados: (A) Distribuição da variável f_c , número de eleitores que compareceram pra votar sobre o número de eleitores aptos, por urna. (B) Distribuição acumulada da variável f_c . Onde a curva para cada ano, em ambos os gráficos, tem a cor disponibilizada na legenda a direita das figuras.

As distribuições da variável f_c foram bem ajustadas pela distribuição Weibull, isto é, pela função densidade de probabilidade dada por

$$f(x) = \begin{cases} \frac{\alpha_i}{\beta_i^{\alpha_i}} x^{\alpha_i} \exp\left[-\left(\frac{x}{\beta_i}\right)^{\alpha_i}\right] & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (5.2)$$

em que α_i e β_i com $i = 1, \dots, 5$ representam os parâmetros para os ajustes de cada ano das 5 eleições da base de dados e $x = f_c$. O que mostra que o a fração de comparecimento é estável sobre todos os anos. Os histogramas para $x = f_c$ juntamente com os ajustes do modelo (5.2) e para cada ano são mostrados na Figura 5.2.6. Como pode ser visto todos os anos são bem ajustados pela distribuição Weibull, apenas o ano de 2002 não fica tão bem ajustado. Os parâmetros para os ajustes em cada ano de eleição pode ser visto na tabela 5.

Para verificar a qualidade do ajuste foi aplicado o teste de Komolgorov-Smirnov com o mesmo procedimento usado para a variável v , obtendo as estatística D e p -valor para cada ano apresentadas na tabela 5. Com esses valores pode-se considerar com, 95% de confiança, que os ajustes pela distribuição Weibull são aceitáveis para a variável f_c sobre todos os anos de eleições da base de dados, pois apesar de 2002 não ficar tão bem ajustado, o valor médio de D é menor que 0.04 e o p -valor não é muito preciso para amostras grandes como as utilizadas aqui. A Figura 5.2.7 mostra as assimetrias e curtoses para as distribuições em cada ano. Todos os anos apresentam

Ano de Eleição	2002	2006	2010	2014	2018
Parâmetro α	11,31	14,71	13,63	13,00	13,01
Parâmetro β	0,82	0,83	0,81	0,82	0,82
Estatística D	0,04	0,02	0,02	0,03	0,02
Estatística p -valor	0,02	0,40	0,42	0,18	0,55

Tabela 5 – Parâmetros α e β dos ajustes pela distribuição Weibull ajustada e as estatísticas do teste de Komolgorov-Smirnov para esses ajustes f_c para amostras de tamanho 1000 por conglomerados (por estados federativos), para cada ano de eleição da base de dados. Os valores das estatísticas representam as médias para 20 amostras, sendo que para os 4 últimos anos $D < 0,04$ e p -valor $> 0,05$, para todas as amostras, enquanto em 2002 a média de D é aceitável mas o p -valor fica abaixo de 0.5

assimetria negativa e crescente com o passar dos anos, a maior concentração da variável se encontra acima da média que é aproximadamente 78%, o que é possível observar também na Tabela 3 notando que a mediana é maior do que a média para todos os anos. A curtose é positiva para todos os anos, próxima a 5 para os quatro primeiros anos e a 4 para 2018. Os mesmos procedimentos foram realizados para a variável f_{bn} , soma de votos brancos e nulos sobre o comparecimento por urna.

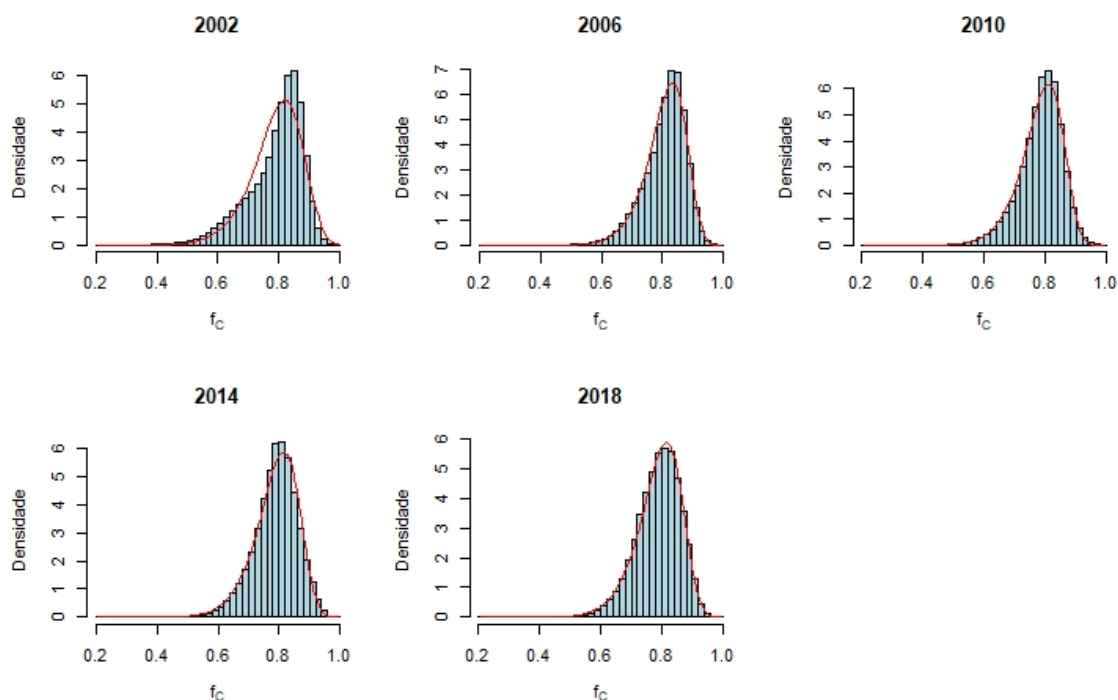


Figura 5.2.6 – Histograma da variável f_c ajustada pela Distribuição Weibull para cada ano de eleição da base de dados: Cada gráfico em azul representa o histograma para 40 classes e a curva em vermelho a distribuição de Weibull, modelo (2) com os parâmetros obtidos dos respectivos dados empíricos em cada um dos 5 anos de eleição analisados.

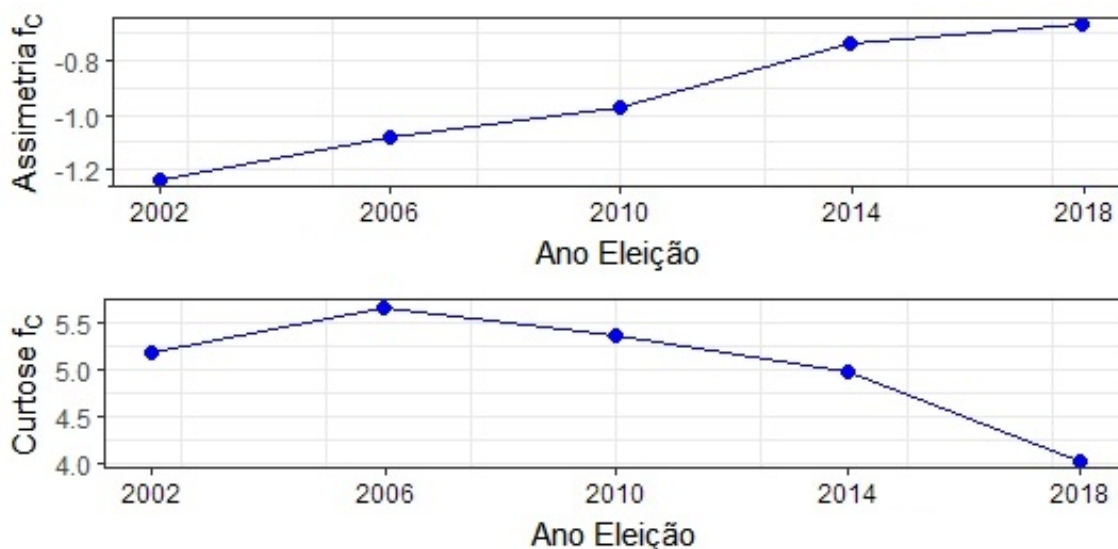


Figura 5.2.7 – Assimetria e Curtose da distribuição da variável f_c : O gráfico superior mostra os valores da assimetria e o gráfico inferior os valores da curtose das distribuições da variável f_c para cada ano da base de dados.

Os mesmos procedimentos foram realizados para a variável f_{bn} , soma de votos brancos e nulos sobre o comparecimento por urna. A Figura 5.2.8(A) disponibiliza a distribuição de f_{bn} para todos os anos de eleições da base de dados, observe que as médias sobre todos anos exceto 2018 tem média próxima à 6% enquanto 2018 tem média próxima à 9%, o que também pode ser visto na Tabela 3.

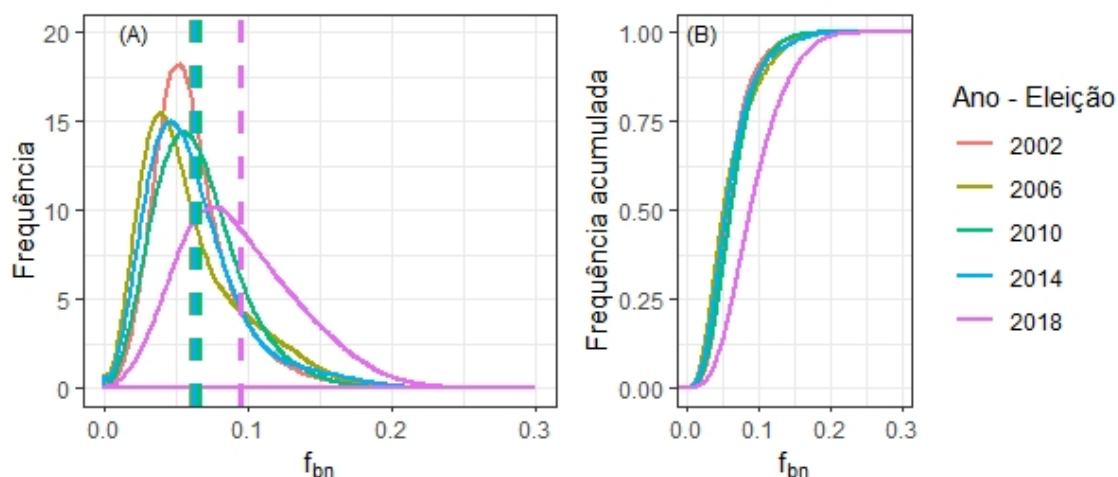


Figura 5.2.8 – Distribuição da variável f_{bn} para todas as eleições da base de dados: (A) Distribuição da variável f_{bn} , número de votos brancos mais nulos sobre o comparecimento, por urna. (B) Distribuição acumulada da variável f_{bn} . A curva para cada ano, em ambos os gráficos, tem a cor disponibilizada na legenda a direita das figuras.

Testadas algumas distribuições como ajuste para variável, a distribuição Beta foi a que melhor se ajustou sobre todos os anos de eleição, assim o ajuste para f_{bn}

mostrados na Figura 5.2.9 é dado pela densidade de probabilidade

$$f(x) = \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} x^{\alpha_i-1} (1-x)^{\beta_i-1}, \quad (5.3)$$

sendo α_i e β_i parâmetros calculados para os dados empíricos com $i = 1, \dots, 5$ representando cada ano das 5 eleições da base de dados e $\Gamma(t) = \int_0^\infty y^{t-1} e^{-y} dy$ a função γ , tomando $x = f_{bn}$.

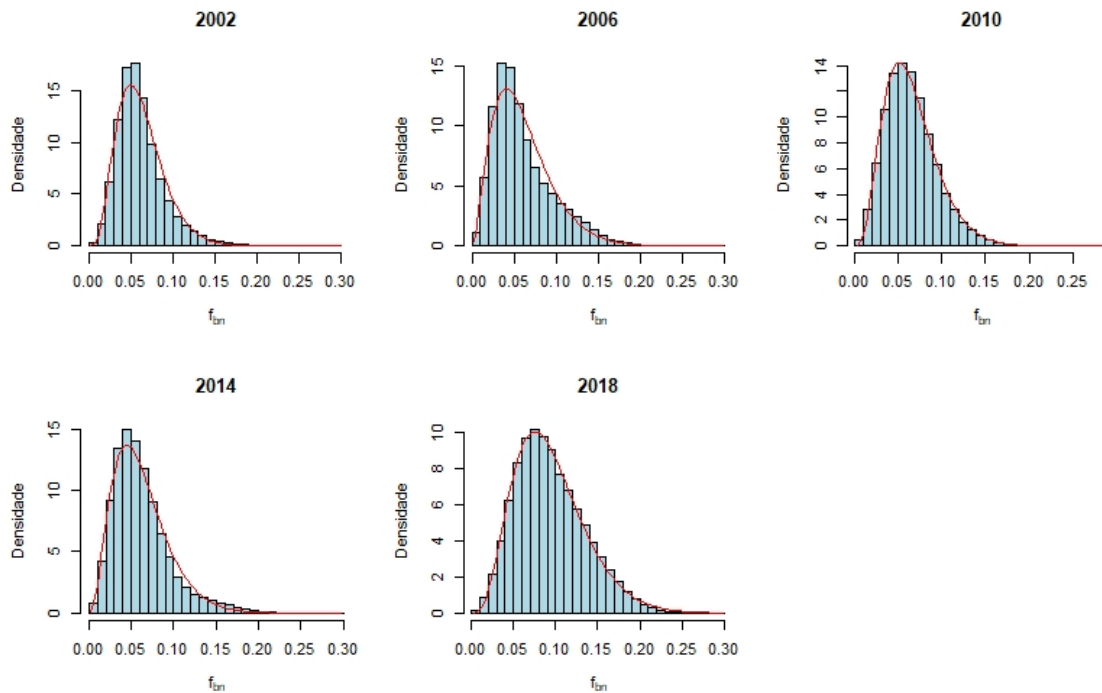


Figura 5.2.9 – Histograma da variável f_{bn} ajustados pela distribuição Beta para cada ano de eleição da base de dados: Cada gráfico em azul representa o histograma para 40 classes e a curva em vermelho a **distribuição de Beta** com os parâmetros obtidos dos respectivos dados empíricos em cada um dos 5 anos de eleição analisados.

Assim como nas análises das variáveis v e f_c foi aplicado o teste de Kolmogorov-Smirnov para verificar a qualidade do ajuste das distribuições de f_{bn} pelo modelo (5.3), as estatísticas obtidas pelo teste estão apresentadas na tabela 6. Como pode ser observado, tem-se $D \leq 0,04$ e $p \geq 0,05$ para todos os anos e os ajustes são aceitos com 95% de confiança. Isso mostra que a forma da distribuição é estável sobre os anos de eleição, no entanto os parâmetros são próximos para os anos de 2006, 2010 e 2014, e substancialmente diferentes para 2002 e 2018, como pode ser confirmado com os valores para as assimetrias e curtoses sobre os anos de eleição disponibilizados na Figura 5.2.10. Assimetrias e curtose próximas para os três anos intermediários e maiores para 2002 e menores para 2018. Portanto os anos intermediários preservam forma e parâmetros.

Ano de Eleição	2002	2006	2010	2014	2018
Parâmetro α	4,63	2,79	4,25	3,33	4,37
Parâmetro β	70,13	43,22	60,97	50,59	42,18
Estatística D	0,03	0,03	0,02	0,03	0,02
Estatística p -valor	0,17	0,09	0,46	0,14	0,60

Tabela 6 – Parâmetros α e β dos ajustes pela distribuição Beta ajustada e as estatísticas do teste de Komolgorov-Smirnov para esses ajustes f_{bn} para amostras de tamanho 1000 por conglomerados (por estados federativos), para cada ano de eleição da base de dados. Os valores das estatísticas representam as médias para 20 amostras, para todas as amostras ficaram muito próximos dos valores médios.

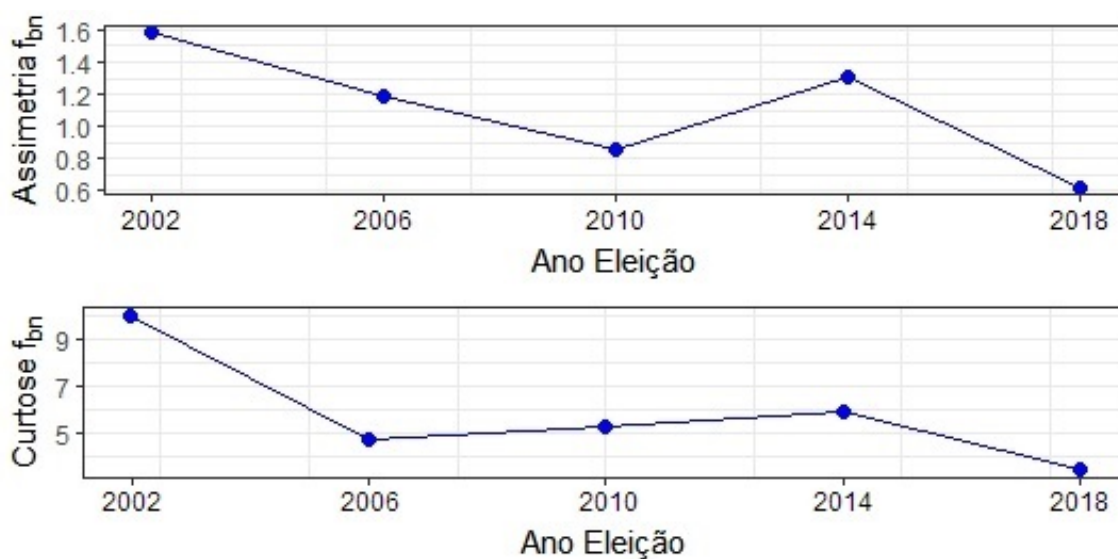


Figura 5.2.10 – Assimetria e Curtose da distribuição da variável f_{bn} : O gráfico superior mostra os valores da assimetria e o gráfico inferior os valores da curtose das distribuições da variável f_{bn} para cada ano da base de dados

As análises realizadas até aqui apresentaram algumas regularidades, em especial para os três anos intermediários de eleições da base de dados, mas os anos 2002 e 2018 mostraram um comportamento um tanto diferenciado, 2018 ainda mais anômalo quando comparados aos outros anos, pois no caso da variável ν não preserva a mesma forma de distribuição que os outros anos. Uma análise feita por BORGHESI (2012) investigou o envolvimento cívico do eleitorado para eleições municipais de 11 países sem voto compulsório utilizando o cálculo da entropia de Shannon para as frações de votos brancos e nulos, votos válidos e abstenções obtendo um comportamento médio em relação ao tamanho do eleitorado aproximadamente logarítmico.

Considerando a definição de entropia já mencionada¹, na direção de investigar o envolvimento do eleitorado nas eleições presidenciais brasileiras, e que a entropia aqui mede o grau de desordem ou quanto o envolvimento do eleitorado distancia

¹ Para a área de Teoria da Informação, a entropia é definida como uma forma de medir o grau médio de incerteza de fontes de informação, na física estatística mede a chance de um sistema de partícula voltar ao seu estado original quando perturbado. (LIMA, 2004)

o comportamento esperado, foram calculadas as entropias S para as proporções: $f_V = V_v/A_p$ ($V_v = V_p + V_s$) é o número de votos válidos, votos do primeiro mais votos do segundo colocados na urna, e A_p número de aptos (eleitores cadastrados na urna), $f_A = A_a/A_p$ (A_a é o número de abstenções na urna), $f_{BN} = V_{bn}/A_p$ (V_{bn} é o número de votos brancos mais nulos na urna), o cálculo da entropia de Shannon.

$$S = -f_V \log_{10}(f_V) - f_A \log(f_A) - f_{BN} \log_{10}(f_{BN}) \quad (5.4)$$

As variáveis devem satisfazer $f_V + f_A + f_{BN} = 1$, conservação da probabilidade. Maximizando a entropia S obtém-se que o ponto de máximo ocorre para $f_V = f_A = f_{BN} = 1/3$ e o valor máximo é $S = \log_{10}(3) = 0,477$, no qual consideramos aqui a "desordem" máxima possível para o envolvimento do eleitorado. O resultado para S em cada ano de eleição foram agrupados por cidade e depois por janela da quantidade de aptos igualmente espaçadas em logaritmo na base 10 e são mostrado no gráfico da Figura 5.2.11. Em termos simples, consideramos nessa análise que a entropia mede o quão incerto é o resultado, quanto maior o valor da entropia mais irregular o resultado, neste sentido pode-se observar que 2018 apresenta maior irregularidade comparado aos outros e 2002, para eleitorados menores do que 30.000 apresenta-se mais irregular do que para os anos intermediários, porém fica mais ordenado para eleitorados maiores.

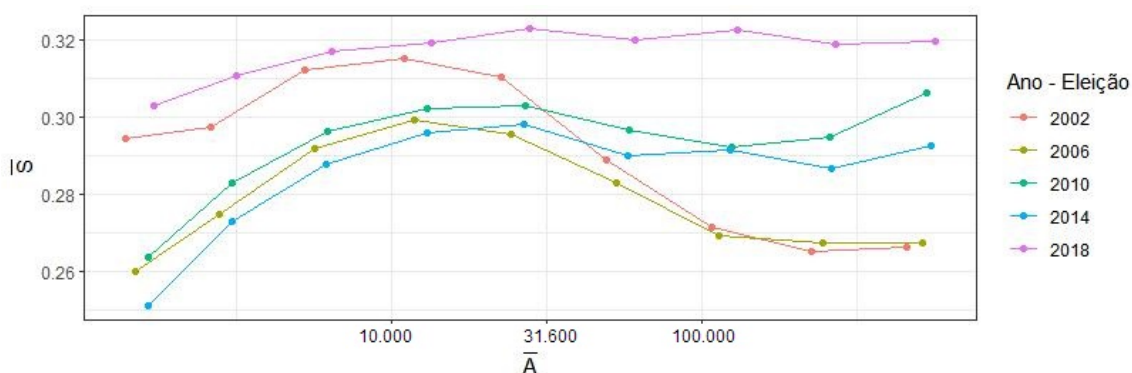


Figura 5.2.11 – Entropia média versus tamanho do eleitorado: \bar{S} representa a entropia média por cidade e por janela ω igualmente espaçada do logaritmo na base 10 do tamanho do eleitorado nas cidades e \bar{A} a média do eleitorado por janela para cada ano de eleição da base de dados.

A desordem aparente nas eleições de 2002 e 2018 parece ser devido a distribuição dos votos brancos e nulos, votos que devem medir o grau de insatisfação do eleitorado com ambos os candidatos, com intuito de investigar melhor esse "grau de insatisfação" foram construídos os gráficos da relação entre votos para o primeiro colocado, V_p , versus votos brancos mais nulos, V_{bn} , por cidade, mostrado na Figura 5.2.12 e o gráfico de V_s , votos para o segundo colocado, versus Figura 5.2.13. É perceptível em ambos os gráficos de dispersão a tendência, subjacente às flutuações,

de crescimento linear da variável V_{bn} em relação a V_p quanto em relação a V_s . Para mostrar de forma mais robusta esse resultado, amenizando o efeito das flutuações, foram tomadas as médias em janelas igualmente espaçadas da variável $\log_{10}(V_p)$, e dos respectivos valores de $\log_{10}(V_{bn})$, a relação entre os valores médios são mostrados juntamente com o respectivo ajuste linear na Figura 5.2.14. O mesmo procedimento é realizado para a relação V_s versus V_{bn} e é mostrado na Figura 5.2.15.

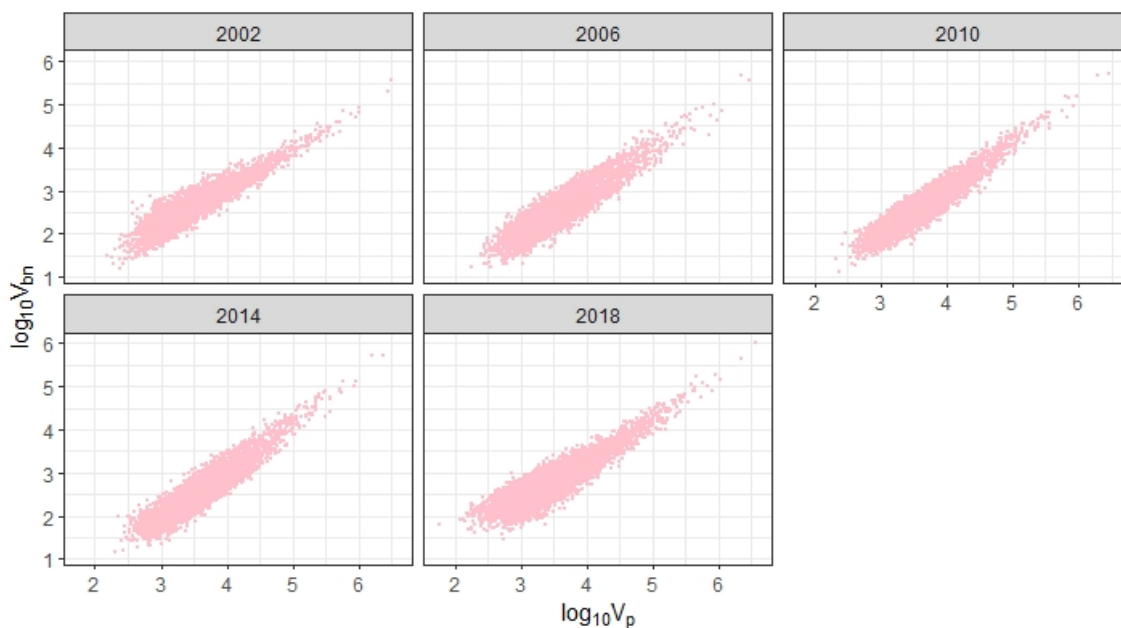


Figura 5.2.12 – Gráfico de dispersão para o primeiro colocado: gráfico de $\log_{10}(V_p)$ versus $\log_{10}(V_{bn})$, em que V_p representa os votos para o primeiro colocado e V_{bn} os votos brancos e nulos por cidade para os 5 anos da base de dados.

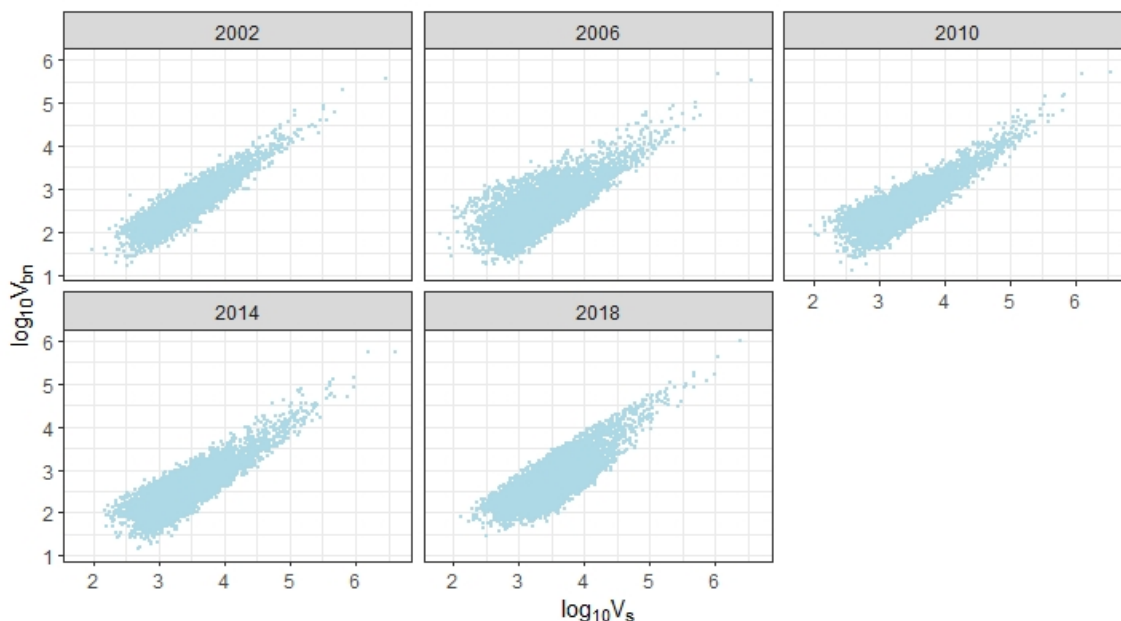


Figura 5.2.13 – Gráfico de dispersão para o segundo colocado: gráfico de $\log_{10}(V_s)$ versus $\log_{10}(V_{bn})$, em que V_s representa os votos para o segundo colocado e V_{bn} os votos brancos e nulos por cidade para os 5 anos da base de dados.

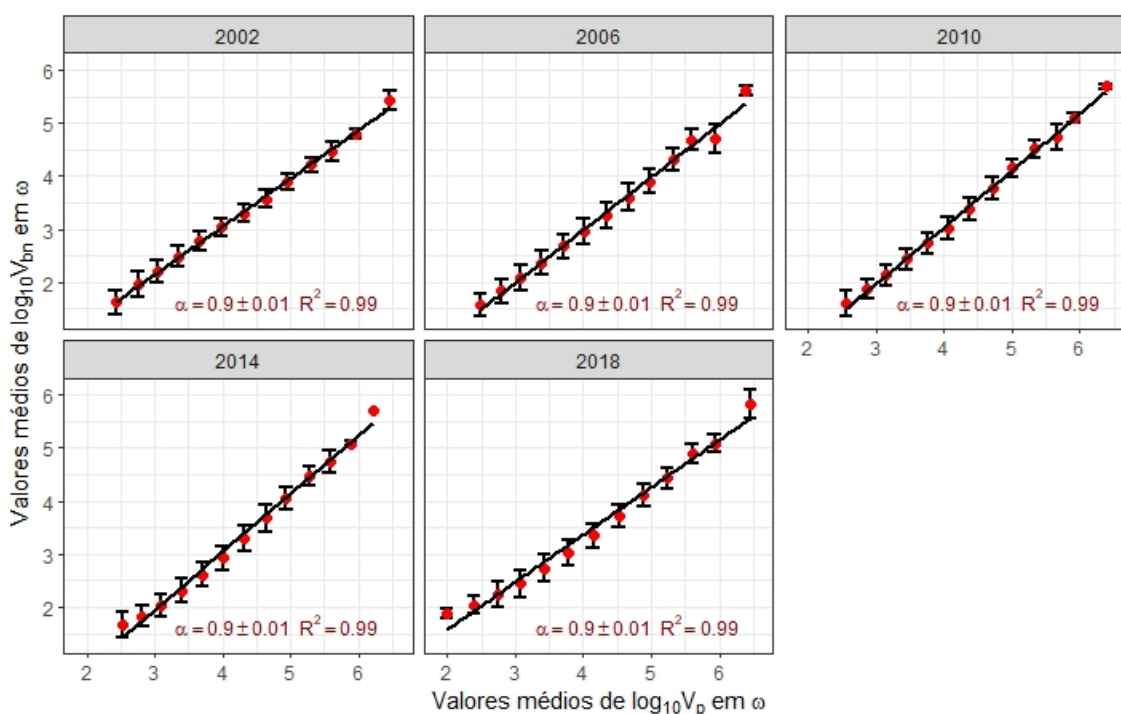


Figura 5.2.14 – Ajustes lineares dos valores médios para o primeiro colocado: Os pontos em vermelho representam os valores médios da variável $\log_{10}(V_p)$ versus os valores médios correspondentes da variável $\log_{10}(V_{bn})$ em janelas ω igualmente espaçadas em $\log_{10}(V_p)$, para 13 janelas. As retas em preto representam os ajustes lineares, via mínimos quadrado e as barras de erro os desvios padrão na janela ω . A inserção em cada figura mostra o coeficiente de determinação do ajuste, R^2

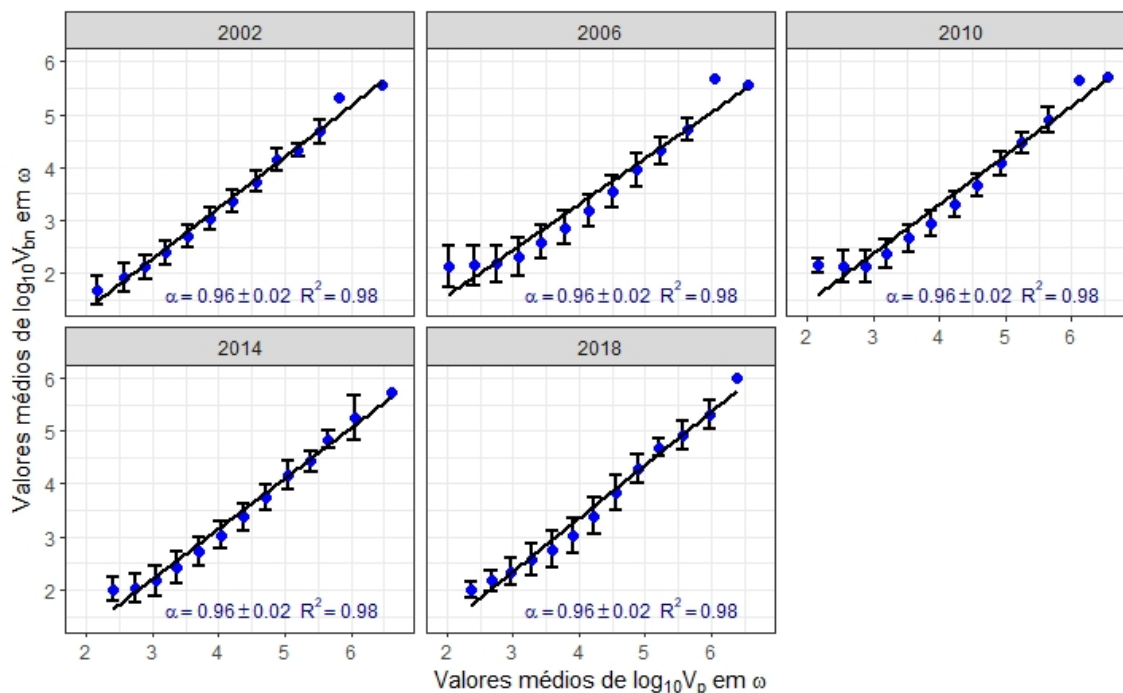


Figura 5.2.15 – Ajustes lineares dos valores médios para o segundo colocado: Os pontos em azul escuro representam os valores médios da variável $\log_{10}(V_s)$ versus os valores médios correspondentes da variável $\log_{10}(V_{bn})$ em janelas ω igualmente espaçadas em $\log_{10}(V_p)$, para 13 janelas. As retas em preto representam os ajustes lineares, via mínimos quadrado e as barras de erro os desvios padrão na janela ω . A inserção em cada figura mostra o coeficiente de determinação do ajuste, R^2

Os coeficientes de determinação mostrados nas inserções dos gráficos são muito próximos a 1, o que mostra que os ajustes para os valores médios são muito consistentes. E para analisar a qualidade da representação dos dados pelos valores médios nas janelas ω foram calculadas as variâncias da variável $\log_{10}(V_{bn})$ em cada janela para todas as eleições e mostradas na Figura 5.2.16 para o primeiro colocado e na Figura 5.2.17 para o segundo colocado. Ambos os casos apresentam pouca variação o que mostra ser consistente representar os dados das cidades pelas médias por janelas.

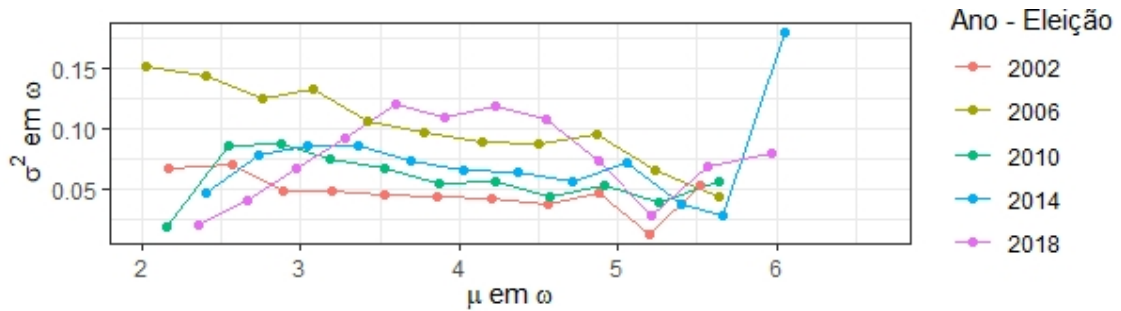


Figura 5.2.16 – Variância nas janelas ω para o primeiro colocado: Valores médios da variável $\log_{10}(V_p)$ versus variância σ^2 da variável $\log_{10}(V_{bn})$ para as 13 janelas ω usadas na obtenção dos valores médios mostrados na Figura 5.2.14.

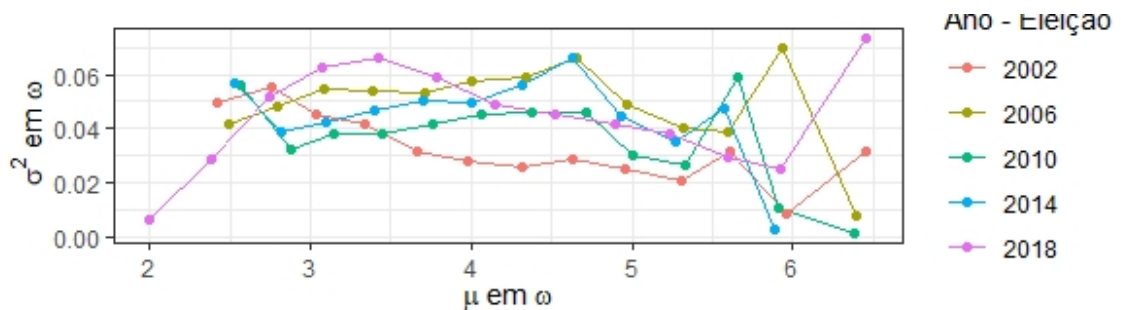


Figura 5.2.17 – Variância nas janelas ω para o segundo colocado: Valores médios da variável $\log_{10}(V_s)$ versus variância σ^2 da variável $\log_{10}(V_{bn})$ para as 13 janelas ω usadas na obtenção dos valores médios mostrados na Figura 5.2.15.

Os resultados apresentados possibilitam escrever para os valores médios o modelo

$$\langle \log_{10}(\mathbf{V}_{bn}) \rangle = \mathbf{A}_{ki} + \alpha_{ki} \langle \log_{10}(\mathbf{V}_k) \rangle \quad (5.5)$$

em que A_{ki} e α_{ki} para $i = 1, \dots, 5$, representando os parâmetros do ajuste para os 5 anos de eleições da base de dados e $k = p, s$ representando primeiro e segundo colocados. Ou ainda, observado que as variâncias por janela tem pouca variação é consistente considerar, equivalentemente, o modelo (5.5)

$$\log_{10}(\mathbf{V}_{bn}) = \mathbf{A}_{ki} + \alpha_{ki} \log_{10}(\mathbf{V}_k) + \varphi(\mathbf{V}_k) \quad (5.6)$$

com os parâmetros mencionados acima e φ um ruído estocástico oriundo das (poucas) variações das variâncias e supondo a opção de voto de um eleitor é independente da opção do outro, partir do modelo linear acima pode-se considerar o modelo lei de potência para os dados.

$$\mathbf{V}_{bn} = \mathbf{a}_{ki} \mathbf{V}_k^{(\alpha_{ki})} \xi(\mathbf{V}_k) \quad (5.7)$$

Para $A_{ki} = \log_{10}(a_{ki})$ e $\varphi = \log_{10}(\xi)$ para $i = 1, \dots, 5$ e $k = p, s$, ou seja, o modelo é válido para todos os anos de eleição da base de dados e para o primeiro e segundo

colocado. O resultado obtido aqui é que o número de votos brancos e nulos, aglomerado por cidade, cresce com o número de votos para o primeiro colocado e também com número de votos para o segundo colocado através do modelo lei de potência. Os expoentes α_{ki} e os interceptos A_{ki} são mostrados para os cinco anos de eleições e para os dois candidatos na Figura 5.2.18.

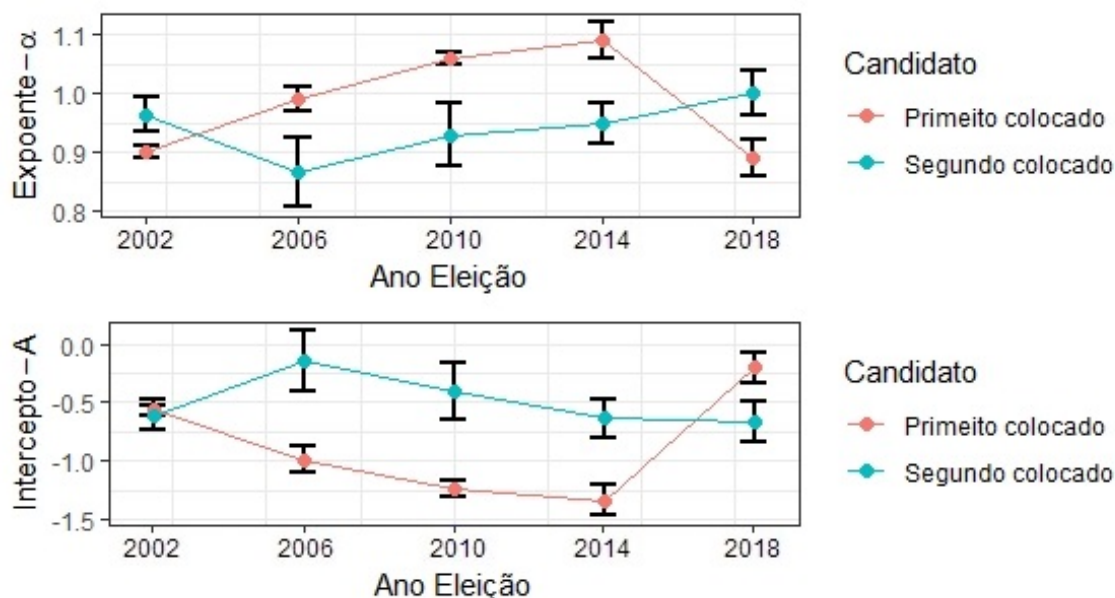


Figura 5.2.18 – Expoentes e interceptos do modelo lei de potência para os dois candidatos: o gráfico na parte superior da figura apresenta os expoentes e o gráfico na parte inferior os interceptos do modelo lei de potência obtido dos ajustes mostrados nas Figuras 5.2.14 e 5.2.15. As barras de erro representam o erro padrão para os respectivos parâmetros do modelo.

Nos anos de 2002, 2006 e 2018 os expoentes do primeiro colocado são menores do que 1, ou seja, os votos brancos e nulos crescem em uma menor proporção do que os votos do primeiro colocado.

Em 2010 e 2014 os expoentes para o primeiro colocado são maiores do que 1, portanto nesses anos os brancos e nulos crescem numa proporção maior do que os votos do vencedor. No caso do segundo colocado, todos os expoentes são menores ou igual a 1, assim brancos e nulos crescem sempre numa proporção menor do que os votos do segundo colocado. Outra observação é a inversão dos expoentes, em 2002 e 2018, os expoentes dos primeiros colocados ficam abaixo dos expoentes dos segundos colocados, o que significa que brancos e nulos crescem como uma proporção maior dos votos do segundo colocado do que do primeiro colocado, enquanto para os anos intermediários em que ocorre o fenômeno de "reeleição" (2006, 2010 e 2014), ocorre a inversão, nesse caso brancos e nulos crescem como uma proporção maior dos votos do primeiro do que do segundo. No caso dos interceptos também ocorre o fenômeno contrário em relação nos mesmos anos de eleição, porém o fenômeno é invertido,

em 2002 e 2018 os interceptos do primeiro colocado ficam acima dos interceptos dos segundo colocados, sendo em 2002 muito próximo os expoentes do primeiro e segundo colocados, ocorrendo também a inversão nos anos intermediários.

5.3 Resultados

Com o intuito de verificar a possibilidade de indicativo de fraude em eleições brasileiras, e utilizando um procedimento usado em KLIMEK(2017), que constatou a presença de dois mecanismos de fraude (incremental e extrema) em eleições da Rússia e da Uganda, foram construídos as densidades bidimensional da variável f_C , taxa de comparecimento, versus a variável f_P , taxa de votos para o primeiro colocado, na urna, para 5 eleições presidenciais nos anos 2002, 2006, 2010, 2014 e 2018. Esses mecanismos se apresentam através de um fenômeno, muito pouco provável, um aglomerado de urnas com 100% de comparecimento e 100% de votos para o vencedor, a partir da tabela 3 pode se concluir que para as eleições investigadas aqui esse fenômeno não ocorre.

No entanto é interessante observar algumas diferenças entre as eleições. Em 2002, 2006 e 2010 não há quantidade de urnas significativa com a taxa de votos para o vencedor abaixo de 20%, em especial em 2002 que fica acima de 25%, enquanto para as duas últimas eleições tem-se uma massa considerável de urnas com baixa votação para o vencedor, o que pode ser confirmado pelo primeiro percentil mostrado na Tabela 2, de forma substancial em 2018 que 10% das urnas apresentam a taxa abaixo de 4,8%. É notável também na Figura 5.2.1 uma distribuição parecida para os anos intermediários e distribuições “diferenciadas” para os anos de 2002 e 2018.

A densidade da variável ν definida em (5.1), mostrada para ν normalizada na Figura 5.2.2(A), apresenta densidades muito próximas a densidade de probabilidade normal padrão para os anos 2002, 2006 e 2010, enquanto para os anos 2002 e 2018 não parecem seguir esse padrão. A Figura 5.2.2(B) mostra a distribuição acumulada e a curva em rosa é a que mais se distancia da normal. Os histogramas com o respectivo ajuste pela distribuição normal é mostrado para cada ano, separadamente, na Figura 5.2.3 para uma melhor visualização dos ajustes.

Aplicado o teste de Kolmogorov-Smirnov para amostras de tamanho 1000, obtidas aglomerando por estado, de cada ano de eleição, pode-se aceitar com 95% de confiança o ajuste da distribuição de ν pela distribuição normal para as 4 primeiras eleições da base de dados, mas o teste rejeita o ajuste para a eleição de 2018. A Figura 5.2.4 mostra as assimetrias próximas a zero para os 4 primeiros anos de eleição e curtose próxima a três, valor da distribuição normal, para os 3 anos intermediários e assimetria próxima a 1 e curtose próxima a 4,2 para 2018.

No segundo turno o eleitor tem três possibilidades de escolha: um dos dois candidatos ou anular seu voto (branco e nulo), o que representa, a menos de um erro do eleitor, uma rejeição aos dois candidatos, mas essa rejeição poderia ser manifestada também através da abstenção do eleitor, porém no Brasil o voto é compulsório e isso é pouco provável. A Figura 5.2.5(A) mostra as densidades para as frações de comparecimento, f_C , observe que as médias são muito próximas a 78%, exceto 2006 que está sensivelmente acima, em 80%, a forma das distribuições são consideravelmente próximas, as distribuições acumuladas na Figura 5.2.5(B) reforça a similaridade entre as distribuições.

A Figura 5.2.6 mostra os histogramas de f_C ajustado pela distribuição Weibull para cada ano de eleição da base de dados separadamente. Aplicado o teste de Kolmogorov-Smirnov para amostras de tamanho 1000, o teste confirma o ajuste para todos os anos com 95% de confiança. As assimetrias mostradas são negativas e crescentes sobre os anos, as curtoses decrescem a partir de 2006 como pode ser visto na Figura 5.2.7. Os mesmos procedimentos realizados para a variável fração de brancos e nulos f_{bn} , mostram distribuições de mesma forma, Figura 5.2.8(A), porém de média sensivelmente deslocada para a direita para o ano de 2018, em relação aos outros anos, para os quatro primeiros anos da base de dados a médias são muito próximas a 6% enquanto para 2018 ela se aproxima de 9,3%.

Este fenômeno já era esperado, como visto na tabela 3, que traz as taxas de votos brancos e nulos tanto pelos aptos quanto pelo comparecimento em 9,6% e 9,4% respectivamente. Notoriamente os votos brancos e nulos foram díspares das demais eleições, o que provavelmente trata-se de um aumento nos votos de “protesto”, como já discutido.

O deslocamento apresentado na distribuição de 2018 pode ser visto na distribuição acumulada na Figura 5.2.8(B). Os histogramas ajustados pela distribuição beta são mostrados na Figura 5.2.9.

Os resultados do teste de Kolmogorov-Smirnov para amostras de tamanho 1000, confirmam com 95% de confiança o ajuste pela distribuição Beta para todas as eleições. Assimetrias decrescente para os anos 2002, 2006, 2010 e 2018 e curtose próxima a 9 para 2002, próximas a 5 para 2006, 2010 e 2014, e próxima a 3 para 2018. Menor curtose e menor assimetria para 2018.

Para finalizar as análises por urna foram calculadas as entropias, definida em (5.2), para as frações f_V (fração de votos válidos), f_A (fração de abstenções), f_{BN} (frações de brancos e nulos) em relação aos aptos na urna. A Figura 5.2.11 mostra as médias por janelas do eleitorado por cidade, o comportamento é muito próximo para os 3 anos intermediários e para eleitorado de até 30.000 eleitores e acima desse

tamanho as médias se aproximam para os anos 2002 e 2006 ficando abaixo de 0,29 e se aproximam também para os anos 2010 e 2014 ficando acima de 0,28. O ano de 2018 apresenta entropia média notavelmente acima dos outros anos em todas as janelas.

Observando que votos brancos e nulos representam a rejeição do eleitorado pelos dois candidatos, para finalizar as análises foram construídos os gráficos dos logaritmos das variáveis votos para o primeiro colocado versus brancos e nulos por cidade, Figura 5.2.12 e o mesmo procedimento para os votos do segundo colocado Figura 5.2.13. Uma tendência linear emerge em meio as flutuações dos dados. Para contornar as flutuações e confirmar a tendência linear, as Figuras 5.2.13 e 5.2.14, apresentam as médias tomadas por janelas igualmente espaçadas na variável logaritmo dos votos do primeiro e do segundo colocados e seus respectivos ajustes lineares, respectivamente.

O coeficientes de determinação estão inseridos nas figuras com os respectivos gráficos, todos com valores muito próximos a 1 mostram que os ajustes são muito consistentes. Para verificar a representatividade pelos valores médios na janela foram calculadas as variâncias para o primeiro colocado na Figura 5.2.16 e para o segundo colocado na Figura 5.2.17, nos dois casos as variâncias são aproximadamente constantes para todos os anos e com uma variação um pouco maior para 2018, esse resultado mostra que é consistente representar os dados pelos seus valores médios. Diante desses resultados obtém-se o modelo linear (5.5) e conseqüentemente o modelo lei de potência (5.7) com os parâmetros de ajustes mostrados na Figura 5.2.18.

Para todos os anos e para os dois candidatos os expoentes são menores do que 1, escala sublinear, com exceção de 2010 e 2014 para o primeiro colocado que apresentam expoente maiores do que 1, escala superlinear, sendo os expoentes dos segundos colocados em 2002 e 2018 menores que o do primeiro, os interceptos apresentam uma relação inversa a dos expoentes.

5.4 Discussão

Os resultados das investigações mostraram um comportamento esperado para as distribuições das proporções de comparecimento versus proporções de votos do primeiro colocado, apresentando picos apenas em torno de 80% de comparecimento e 50% de votos para o primeiro colocado, mostrando uma normalidade e extinguindo a hipótese dos mecanismos de fraude incremental e extrema introduzidos por KLI-MEK(2012) nas eleições presidenciais brasileiras analisadas aqui. Porém foi observado um comportamento similar para os anos intermediários, 2006 (reeleição de Luiz Inácio Lula da Silva), 2010 (eleita Dilma Russel) e 2014 (reeleita Dilma Russel), isso sugere

um padrão para “reeleição” dado que foram eleitos candidatos que estavam com a “máquina na mão”, ou seja, candidatos do partido do governo.

As distribuições da variável ν para esses três anos intermediários foram muito bem ajustadas pela distribuição normal, resultado confirmado pelo teste de Kolmogorov-Smirnov e pode ser observado pelas curtoses e assimetrias mostradas na Figura 5.2.4, isso sugere que o padrão de reeleição é do tipo normal. No entanto a eleição de 2002 também teve o ajuste normal aceitável pelo teste de Kolmogorov-Smirnov, mas como apresentado na Figura 5.2.4 a curtose nesse ano fica sensivelmente acima da distribuição normal. Pode-se então considerar que o padrão normal para a variável ν se mantém nos 4 primeiros anos de eleições da base de dados, porém com melhor ajuste nos anos intermediários. Visivelmente, a distribuição de ν para eleição de 2018 não apresenta algum tipo de padrão regular conhecida, veja as Figuras 5.2.2 e 5.2.3, aplicado o teste de Kolmogorov-Smirnov as estatísticas do teste rejeitam a hipótese de ajuste normal.

Assim a relação entre votos do primeiro e os demais votos levada em conta na definição de ν se apresenta irregular em 2018. Foi investigado também a rejeição aos dois candidatos, que pode ser analisada através do comparecimento e dos votos brancos e nulos. As distribuições da taxa de comparecimento às urnas, f_C , mostradas na Figura 5.2.5, apresentam um padrão para todos os anos de eleições analisadas, com uma média de 78% de comparecimento, exceto para 2018 que ficou sensivelmente acima, 80%, o que pode ser visto também nas Tabelas 2 e 3. O padrão Weibull foi verificado para todas as eleições e assim a distribuição da f_C nas urnas é estável, resultado esperado já que no Brasil o voto é compulsório. A distribuição da fração de brancos mais nulos nas urnas, f_{bn} , mostrada na Figura 5.2.8, apresenta uma forma regular assimétrica positiva para todos os anos com médias próximas a 6% sobre todos os anos exceto para 2018 que ficou acima de 9%, pode também ser observado nas Tabelas 2 e 3. O padrão obtido encontrado para as distribuições de brancos e nulos é a distribuição Beta, assim essa variável também apresentam uma forma estável sobre os anos, com parâmetros bem diferenciados no ano de 2018.

As análises dessas variáveis indicam um padrão para o envolvimento do eleitorado sobre todas as eleições analisadas, com um “deslocamento” marcante em 2018, mostrando uma maior e acentuada rejeição aos dois candidatos. Enquanto 2002 apresenta uma maior participação e maior aceitação dos candidatos pelo eleitorado, o comparecimento apresentou 2% acima e menor proporção de brancos e nulos em relação a todos outros anos, como pode ser constatado nas Tabelas 2 e 3.

O envolvimento cívico definido por BORGUESI (2012), que relaciona as taxas de abstenções, brancos e nulos e votos válidos através da Entropia de β , foi aplicado nesse trabalho para investigar como se deu o envolvimento do eleitorado nas 5 eleições

presidenciais brasileiras que compõem a base de dados. As entropias calculadas aqui representam a medida do quanto o comportamento do eleitorado se afasta do esperado, ou ainda da desordem na distribuição dessas razões. Para isso foram calculadas as entropias, definidas pelo modelo (5.4). A Figura 5.2.11 mostra que para eleitorados pequenos, menores do que aproximadamente 30.000, os três anos intermediários, anos que apresentaram o fenômeno de "reeleição" apresentam entropias muito parecidas e crescentes.

No caso de grandes eleitorados, maiores do que 30.000, tem-se para as eleições de 2002 e 2006 (eleito Luiz Inácio Lula da Silva em ambas) o envolvimento cívico tendendo a uma organização estável, dado que as entropias se estabilizam próximas a 0.27, enquanto para as eleições de 2010 e 2014 (eleita Dilma Rousseff em ambas) há uma inversão e as entropias decaem para eleitorados intermediários, entre 50.000 e 500.000, e voltam a crescer para os grandes aglomerados de eleitores, indicando uma instabilidade no envolvimento cívico. As eleições de 2018 (eleito Jair Bolsonaro) a entropia é aproximadamente constante para eleitorados maiores que 10.000 e bem acima de todas as outras sobre todo o eleitorado. Indicando um comportamento de consideravelmente fora do padrão esperado.

Na última análise realizada aqui, foi obtido um modelo lei de potência (5.7) que relaciona as variáveis votos do candidato com os votos brancos e nulos, para o primeiro e o segundo colocados em todos os anos de eleições analisados. Os expoentes e interceptos dos modelos são mostrados na Figura 5.2.16. Os expoentes do ajuste para o primeiro colocado nos três anos intermediários são maiores ou iguais a 1 ($\alpha_{pi} \geq 1$ para $i = 2, 3, 4$), mostrando que, nesses anos, conforme crescem os votos do primeiro colocado, crescem numa proporção maior os votos brancos e nulos, que tem uma proporção crescente sobre os três anos, visto que os expoentes crescem a cada ano. Para os anos 2002 e 2018 os votos brancos e nulos crescem numa proporção menor do que os votos do primeiro colocado, assim como acontece para a relação de votos no segundo colocado, e brancos e nulos em todos os anos de eleição da base de dados.

5.5 Conclusões

As análises apresentadas aqui mostraram que não há indicativo de mecanismos de fraude como os encontrados por KLIMEK (2017) em eleições da Rússia e da Uganda, usando os mesmos métodos, o que reforça a segurança em nosso sistema eleitoral como é enfatizado pelo TSE. No entanto foram obtidos modelos estáveis para as distribuições das variáveis taxa de comparecimento e brancos e nulos sobre todos os anos de eleições. E modelo normal para distribuição da variável ν nos quatro primeiros anos de eleições, mostrando que a eleição de 2018 foi anômala comparada as outras,

sem um padrão regular. Os cálculos das entropias permitiram observar que eleitorados com tamanho até 30.000 se comportaram como se as três eleições intermediárias fossem reeleição, enquanto eleitorados maiores parece ter distinguido os dois candidatos, Dilma e Lula como duas candidaturas distintas. Isso possibilitou compreender uma diferença considerável na dinâmica do envolvimento cívico relacionada ao tamanho do eleitorado. E também mostraram que 2018 teve o envolvimento cívico mais irregular entre os anos analisados. Para a análise por cidade foi obtido um modelo lei de potência relacionando votos dos dois candidatos com os votos brancos e nulos.

Os resultados apresentados aqui contribuíram para a compreensão de como se deu o envolvimento cívico nas 5 últimas eleições presidenciais brasileiras. Levando em conta que os votos para o primeiro colocado, votos para o segundo colocado e votos brancos e nulos representam os três estados de opinião do eleitorado (voto no vencedor, voto no perdedor, e o “voto protesto”), assim pode ser interessante uma análise mais aprofundada para obtenção de um modelo relacionando as três variáveis.

REFERÊNCIAS

- ARANHA, Diego F.; Graaf, Jeroen. The Goog, the Bad, and the Ugly: Two Decades of E-Voting in Brazil. *Security and Privacy in Brazil*. IEEE Security and Privacy. p.22-30, 2018.
- ARANHA, Diego F.; Barbosa, Pedro Y.S.; Cardoso, Thiago N.C.; Araújo, Caio Lüders; Matias, Paulo. - The return of software vulnerabilities in the Brazilian voting machine. Elsevier. *Computers & Security*. v.86, p.335–349, 2019.
- ARARIPE, L. E.; Costa Filho, R. N.; Herrmann, H. J.; Andrade, J. S. Plurality voting: the statistical laws of democracy in Brazil, *Int. J. Mod. Phys. C* 17, 1809-1813, 2006.
- ARARIPE, L. E.; Costa Filho, R. N. Role of parties in the vote distribution of proportional elections. **Physica A**, v. 378, p.4167, 2009.
- ARAÚJO, N. A. M.; Andrade, J. S.; Herrmann, H. J. Tactical voting in plurality elections, *PLoS ONE* 5, e12446, 2010.
- BERNARDINETTI, Sergio. Ex-juiz eleitoral escreve artigo sobre as urnas eletrônicas. Disponível em: <<http://www.tre-pr.jus.br/imprensa/noticias-tre-pr/2018/Junho/ex-juiz-escreve-artigo-sobre-as-urnas-eletronicas>>. Acesso em 02 jun. 2019.
- BITTENCOURT, H.R.; Viali, L. Contribuições Para o Ensino da Distribuição Normal ou Curva de Gauss em Cursos de Graduação. *In: III SEMINÁRIO INTERNACIONAL DE PESQUISA EM EDUCAÇÃO MATEMÁTICA*, 10, Águas de Lindóia, 2006.
- BIZZARRO, Fernando, Coppedge, Michael. *Variedades da Democracia no Brasil*. Opinião Pública vol.23 no.1 Campinas Apr. 2017.
- BOLTZMANN, Ludwig. Sobre a relação entre a segunda lei da teoria do calor mecânico e o cálculo da probabilidade ou as frases sobre balanço de calor. *Relatórios de reunião da Academia Imperial de Ciências de Viena (II 76): 428*. Reimpressão em: *Tratados Científicos de Ludwig Boltzmann*. II Viena p. 164-223, 1877.
- BORGUESI, C.; Bouchaud, J.-P. Spatial correlations in vote statistics: a diffusive field model for decision-making. *The European Physical Journal B*. EDP Sciences, Societ'a Italiana di Fisica, Springer-Verlag, 2010.
- BORGHESI, Christian; Chiche, Jean; Nadal, Jean-Pierre. Between order and disorder: a 'weak law' on recent electoral behavior among urban voters? **PLoS ONE**, v. 7 n. 7,

2012.

BRASIL. Tribunal Superior Eleitoral. O sistema eleitoral brasileiro. Brasília: TSE, 2009.

BRASIL. Tribunal Superior Eleitoral. Eleições no Brasil: uma história de 500 anos / Ane Ferrari Ramos Cajado, Thiago Dornelles, Amanda Camylla Pereira. – Brasília: Tribunal Superior Eleitoral, 2014.

BRASIL. Tribunal Superior Eleitoral. Urna eletrônica: 20 anos a favor da democracia. – Brasília: Tribunal Superior Eleitoral, 2016.

BRASIL. Tribunal Superior Eleitoral. Código eleitoral anotado e legislação complementar / Tribunal Superior Eleitoral. – 13. ed. – Brasília: Tribunal Superior Eleitoral, 2018.

BRASIL, BD. Repositório de Dados Eleitorais do Tribunal Superior Eleitoral. Disponível em: <<http://www.tse.jus.br/eleicoes/estatisticas/repositorio-de-dados-eleitorais>>. Acesso em: 24 mar. 2019.

BRASIL, BU. Justiça Eleitoral. Processamento do Boletim de Urna. Disponível em: <<http://www.justicaeleitoral.jus.br/arquivos/tse-processamento-do-boletim-de-urna>>. Acessado em: 29 set. 2019.

BRASIL, CC. Câmara dos Deputados. Comissão Parlamentar de Inquérito - Crimes Cibernéticos. Disponível em: <<https://www2.camara.leg.br/atividade-legislativa/comissoes/comissoes-temporarias/parlamentar-de-inquerito/55a-legislatura/cpi-crimes-ciberneticos>>. Acessado em: 28 set. 2019.

BRASIL, CD. Câmara dos Deputados. Debatedores divergem sobre segurança de urna eletrônica. Disponível em: <<https://www.camara.leg.br/noticias/474592-debatedores-divergem-sobre-seguranca-de-urna-eletronica/>>. Acessado em: 27 set. 2019. Datado de: 03 nov. 2015.

BRASIL, EU. Justiça Eleitoral. Ecosistema da urna. Disponível em: <<http://www.tse.jus.br/eleicoes/processo-eleitoral-brasileiro/logistica-e-preparacao/ecossistema-da-urna>>. Acessado em: 29 dez. 2019.

BRASIL, FA. Justiça Eleitoral. Sistema eletrônico de votação. Formatos dos arquivos de assinatura. Disponível em: <<http://www.justicaeleitoral.jus.br/arquivos/tse-documentacao-tecnica-sobre-a-urna-formato-assinatura>>. Acessado em: 28 set. 2019. Datado de out. 2018.

BRASIL, FL. Justiça Eleitoral. Sistema eletrônico de votação. Formato dos arquivos de log. Disponível em: <<http://www.justicaeleitoral.jus.br/arquivos/tse-documentacao->

tecnica-sobre-a-urna-formato-log>. Acessado em: 28 set. 2019. Datado de out. 2018.

BRASIL, GJ. Câmara dos Deputados. Comissão Parlamentar de Inquérito de Crimes Cibernéticos. Audiência Pública. Apresentação de Giuseppe Janino Secretário de Tecnologia da Informação do Tribunal Superior Eleitoral. Disponível em: <<https://www2.camara.leg.br/atividade-legislativa/comissoes/comissoes-temporarias/parlamentar-de-inquerito/55a-legislatura/cpi-crimes-ciberneticos/documentos/audiencias-publicas/audiencia-publica-dia-03-11.15/apresentacao-giuseppe-janino-secretario-de-tecnologia-da-informacao-do-tribunal-superior-eleitoral-tse>>. Acessado em: 28 set. 2019. Datado de: 03 nov. 2015.

BRASIL, ID. Veja os principais documentos do processo de impeachment de Dilma Rousseff. Agência Senado. Disponível em: <<https://www12.senado.leg.br/noticias/materias/2016/08/22/veja-os-principais-documentos-do-processo-de-impeachment-de-dilma-rousseff>>. Acesso em 01 set. 2018.

BRASIL, MU. Tribunal Superior Eleitoral. Modelos de urna eletrônica. Disponível em: <<http://www.tse.jus.br/eleicoes/urna-eletronica/seguranca-da-urna/modelos-de-urna-eletronica>>. Acessado em 20 set. 2019.

BRASIL, MV. Justiça Eleitoral. Segurança do Processo Eleitoral. Mitos e Verdades. Disponível em: <<http://www.justicaeleitoral.jus.br/spe/#mitos-verdades>>. Acessado em: 27 set. 2019.

BRASIL, NT. Câmara dos Deputados. Comissão Parlamentar de Inquérito - Crimes Cibernéticos. Audiência Pública. Reunião nº 2282/15. Notas taquigráficas. Disponível em: <<http://www2.camara.leg.br/atividade-legislativa/comissoes/comissoes-temporarias/parlamentar-de-inquerito/55a-legislatura/cpi-crimes-ciberneticos/documentos/notas-taquigraficas/nt031115-crc-sem-revisao>>. Acessado em: 28 set. 2019. Datado de 03 nov. 2015.

BRASIL, NV. Tribunal Superior Eleitoral. Nota de esclarecimento: utilização das urnas eletrônicas nas eleições de 2014. Disponível em: <<http://www.tse.jus.br/imprensa/noticias-tse/2018/Outubro/nota-de-esclarecimento-utilizacao-das-urnas-eletronicas-nas-eleicoes-de-2014>>. Acessado em 27 set. 2019. Datado de: 05 out. 2018.

BRASIL, PC. Câmara dos Deputados. CCJ aprova admissibilidade de PEC que torna obrigatório voto impresso. Disponível em: <<https://www.camara.leg.br/noticias/627201-CCJ-APROVA-ADMISSIBILIDADE-DE->

PEC-QUE-TORNA-OBRIGATORIO-VOTO-IMPRESSO>. Acessado em 02 jan. 2020.
Datado de: 17 dez. 2019.

BRASIL, PE. Tribunal Superior Eleitoral. Processo eleitoral no Brasil. Disponível em:
<<http://www.tse.jus.br/eleicoes/processo-eleitoral-brasileiro/funcionamento-do-processo-eleitoral-no-brasil>>. Acesso em 29 dez.
2019.

BRASIL, PF. Tribunal Superior Eleitoral. Sistema eletrônico de votação: perguntas mais frequentes / Tribunal Superior Eleitoral. – 3ª ed. – Brasília: Tribunal Superior Eleitoral, 2018.

BRASIL, PS. Justiça Eleitoral do Brasil. Disponível em:
<<http://www.justicaeleitoral.jus.br/arquivos/pesquisa-instituto-sensus-sobre-a-justica-eleitoral>>. Acesso em 29 set. 2017. Datado de: nov.
2010.

BRASIL, RD. Tribunal Superior Eleitoral. Registro digital do voto. Disponível em:
<<http://www.tse.jus.br/eleicoes/urna-eletronica/seguranca-da-urna/registro-digital-do-voto>>. Acessado em: 28 set.
2019.

BRASIL, RF. Justiça Eleitoral. Relatório Final de Avaliação do Sistema Informatizado das Eleições. Disponível em: <<http://www.justicaeleitoral.jus.br/arquivos/relatorio-final-de-avaliacao-do-sistema-informatizado-das-eleicoes>>. Acessado em: 27 set. 2019.
Disponível também em:
<http://www.ic.unicamp.br/~tomasz/misc/rel_final_site_TSE.pdf>. Acessado em: 29 dez. 2019. Datado de: maio 2002.

BRASIL, TT. Tribunal Superior Eleitoral. Urna eletrônica usa tecnologia desenvolvida por técnicos do próprio TSE. Disponível em:
<<http://www.tse.jus.br/imprensa/noticias-tse/2018/Setembro/urna-eletronica-usa-tecnologia-desenvolvida-por-tecnicos-do-proprio-tse>>. Acessado em 20 set. 2019.
Datado de: 15 set. 2018.

BRASIL, UE. Tribunal Superior Eleitoral. A Urna Eletrônica. Disponível em:
<<http://www.tse.jus.br/eleicoes/urna-eletronica>>. Acesso em 03 jun. 2019.

BRASIL, US. Tribunal Superior Eleitoral. A Urna Eletrônica. Disponível em:
<<http://www.tse.jus.br/o-tse/escola-judiciaria-eleitoral/publicacoes/revistas-da-eje/artigos/revista-eletronica-eje-n.-6-ano-4/por-que-a-urna-eletronica-e-segura>>. Acesso em 30 dez. 2019.

BRASIL, VA. Urna eletrônica completa 20 anos de uso no Brasil. Disponível em:
<<http://www.tse.jus.br/o-tse/cultura-e-historia/museu-do-voto/exposicoes-e>

programas-educativos/20-anos-da-urna-eletronica>. Acesso em 26 jun. 2017.

BRASIL, VI. Liminar suspende regra da minirreforma eleitoral que prevê voto impresso. Disponível em:

<<http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=380571>>.

Acesso em 02 jan. 2020. Datada de 06 Jun. 2018. CARDOSO, M.; Mendes, R.S.; Souza, J.T.G.; Ribeiro, H.V. Gender difference in candidature processes for Brazilian elections, *Physica A* 537 122525, 2020.

CASELLA, George, e Berger, Roger L. Inferência estatística. 2. ed. Centage Learning, 2010.

CHATTERJEE, A.; Mitrovic, M. Fortunato, S. Universality in voting behavior: an empirical analysis, *Sci. Rep.* 3 1049, 2013.

CHURCHILL, Winston S. "Democracy is the worst form of Government except all those other forms that have been tried from time to time". House of Commons, 11 nov. 1947.

COLLIER, P.; Vicente, P. C. , Violence, bribery, and fraud: The political economy of elections in Sub-Saharan Africa. *Public Choice* 153, p. 117–147. 2012.

COULOURIS, George; Dollimore, Jean; Kindberg, Tim. Sistemas Distribuídos: Conceitos e Projetos. 4.ed. Bookman Editora. 2007.

COSTA FILHO, R.N.; Almeida, J.S.; Andrade, JR.; Moreira, J.E. Brazilian elections: voting for a scaling democracy. **Physica A**, v. 322, p. 698, 2002.

COSTA FILHO, R. N.; Almeida, M. P.; Andrade, J. S.; Moreira, J. E. Scaling behavior in a proportional voting process, *Phys. Rev. E* 60 1067-1068, 1999.

CRESPO, Antônio Arnot. Estatística Fácil. 17.ed. São Paulo. Editora Saraiva. 2002.

DARAUJO, Maria Celina;Ribeiro, Guilherme Leite. Perspectivas sobre desenvolvimento democrático e qualidade da democracia: Brasil e América Latina. *Soc. e Cult.*, Goiânia, v. 18, n. 2, p. 61-77, jul./dez. 2015.

DEL PICCIA, Walter. As urnas brasileiras são vulneráveis. Disponível em: <<https://jornal.usp.br/artigos/as-urnas-brasileiras-sao-vulneraveis>>. Acesso em 26 jun. 2017.

DI. Democracy International. Disponível em: <<http://democracyinternational.com/>>. Acesso em 03 nov. 2018.

FELLER, W. "On the Kolmogorov-Smirnov limit theorems for empirical distribution *Annals of Math. Stat.*, Vol. 1,pp. 177-188. 1948.

FORTUNATO, Santo; Castellano, Claudio. Scaling and Universality in Proportional

Elections. PRL 99, Physical Review Letters Week Ending. Set. 2007.

IDEA. The International Institute for Democracy and Electoral Assistance. Disponível em: <<https://www.idea.int>>. Acesso em 26 jun. 2017.

IDW. International Democracy Watch. Centro Studi Sul Federalismo. Disponível em: <<http://www.internationaldemocracywatch.org/>>. Acesso em 03 nov. 2018.

IHAKA, Ross, Gentleman, Robert. Linguagem de Programação R. Departamento de Estatística da Universidade de Auckland, Nova Zelândia, 2018.

JIMENEZ, Raul; Hidalgo, Manuel; Klimek, Peter. Testing for voter rigging in small polling stations. Science Advances. v.3 n.6, 2017. Disponível em: <<http://advances.sciencemag.org/content/3/6/e1602363>>. Acesso em: 5 jul, 2017.

JUSBRASIL, Folha Política. A urna eletrônica brasileira é a mais ultrapassada do mundo, diz especialista que violou o sistema. 2014. Disponível em: <<https://folhapolitica.jusbrasil.com.br/noticias/112550488/a-urna-eletronica-brasileira-e-a-mais-ultrapassada-do-mundo-diz-especialista-que-violou-o-sistema>>. Acesso em 26 jun. 2017.

KLEINA, Nilton, TecMundo "A história da urna eletrônica - TecMundo". 2018. Disponível em: <https://www.youtube.com/watch?v=WQM_sEEJi14>. Acesso em 09 set. 2019.

KLIMEK, Peter; Yegorov, Yuri; Hanel, Rudolf; Thurner, Stefan. Statistical detection of systematic election irregularities. **PNAS**, v. 109, n. 41, p. 16469-16473, out. 2012. Disponível em: <<https://doi.org/10.1073/pnas.1210722109>>. Acesso em: 28 jun. 2017.

KOBAK, Dmitry; Shpilkin, Sergey Pshenichnikov, Maxim S. Statistical anomalies in 2011–2012 Russian elections revealed by 2D correlation analysis. **Physics and Society**. 2012. Disponível em: <<https://arxiv.org/pdf/1205.0741.pdf>>. Acesso em: 23 jun. 2017.

LIMA, Bernardo N. B.; CIOLETTI, Leandro M.; CUNHA, Marcelo O. T.; BRAGA, Gastão A. Entropia: Introdução à Teoria Matemática da (des)Informação. Departamento de Matemática. UFMG, 2004.

LYRA, M. L.; Costa, U. M. S.; Costa Filho, R. N. Andrade, J. S. Generalized Zipf's law in proportional voting processes, Europhys. Lett. 62 131-137, 2003.

MACAU, E.E.N. Sistemas Complexos. Instituto Nacional de Pesquisas Espaciais - INPE, 2002.

MANTOVANI, M.C.; RIBEIRO, H.V.; MORO, M.V.; PICOLI JUNIOR, S.; MENDES, R.S. Scaling laws and universality in the choice of election candidates. **A Letters Journal Exploring The Frontiers of Physics**, v. 96, p. 48001-p1, 2011.

MANTOVANI, M.C.; RIBEIRO, H.V.; LENZI, E.K.; MENDES, R.S. Engagement in the

electoral processes: scaling laws and the role of the political position. **Phys. Rev. E**, v. 88, p. 1, 2013.

MICROBASE, Sistema Operacional VirtuOS. Disponível em:

<www.microbase.com.br/produtos/produtos_virtuos.htm>. Acesso em 27 set. 2019.

MICROSOFT, Sistema Operacional Windows CE. Disponível em:

<<https://docs.microsoft.com/en-us/previous-versions>>. Acesso em 27 set. 2019.

NOGUEIRA, Octaciano. Constituições Brasileiras. Vol. 1, 1824. - 3. ed. - Brasília: Senado Federal. Subsecretaria de Edições Técnicas. 2012.

PEARSON, Karl. Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material, Philosophical Transactions of the Royal Society of London, Ser. A, 186, 343-414. 1895.

PLACKETT, R. L. Studies in the History of Probability and Statistics: VII. The Principle of the Arithmetic Mean. Biometrika. Vol. 45, No. 1/2 (Jun., 1958), pp. 130-135. Oxford University Press. Jun. 1958.

RAMALHO, Renan; OLIVEIRA, Mariana. G1 e TV Globo. Por 8 a 2, STF derruba voto impresso nas eleições de 2018. Disponível em:

<<https://g1.globo.com/politica/noticia/relator-no-stf-admite-impresao-de-votos-mas-diz-que-implantacao-pelo-tse-pode-ser-gradual.ghtml>>. Acessado em 24 set. 2019. Datado de: 06 jul. 2018.

ROZENAS, Arturas. Detecting Election Fraud from Irregularities in Vote-Share Distributions. Wilf Family Department of Politics, New York University. Cambridge University Press. 2017.

SHALDERS, André. Eleições 2018: como funciona a contagem de votos com a urna eletrônica. Disponível em: <<https://www.bbc.com/portuguese/brasil-45079504>>. Acesso em 02 jun. 2019. Datado de 08 ago. 2018.

SHANNON, C. E. A Mathematical Theory of Communication. Bell System Technical Journal n. 27, p. 379-423 e p. 623-656, 1948.

SIMPLICIO, Marcos A. Votação Eletrônica: Desafios no Brasil e no mundo. Ou: porque falar de insegurança da urna brasileira é "pura paranóia". Universidade de São Paulo, 2018

SMIRNOV, N. "On the estimation of the discrepancy between empirical curves of distribution for two independent samples," Bulletin Mathématique de l'Université de Moscou, Vol. 2, fasc. 2. 1939.

TECMUNDO. 92% dos brasileiros não confiam na urna eletrônica. Disponível em:

<<https://www.tecmundo.com.br/seguranca/133524-92-brasileiros-nao-confiam-urna->

eletronica.htm>. Acessado em: 27 set. 2019. Datado de: 23 ago. 2018.

TIERNO, P. Formação da pólis e surgimento da democracia na Grécia antiga: história e consciência da Atenas clássica. **HOLOGRAMATICA – Facultad de Ciencias Sociales UNLZ**. Año XIV, v.21, n. 21, p. 99-119, 2018.

TRAVIESO, G.; Costa, L. F. Spread of opinions and proportional voting, Phys. Rev. E 74 036112, 2006.

WRIGHT, Edward. Certain Errors in Navigation. Londres: Ed. Agas. 1599.

6 APÊNDICES

6.1 Procedimentos ao Processar Dados do TSE

O Tribunal Superior Eleitoral do Brasil (BRASIL, BD, 2019) disponibiliza os dados dos eleitores e eleições, as estatísticas e os resultados de todos os pleitos eleitorais onde houve a utilização do CEV (Coletor Eletrônico de Votos) popularmente conhecido como “urna eletrônica”, em seu repositório de dados eleitorais, no sítio oficial via internet.

Deste repositório, foram usados apenas os dados de Votação por Seção Eleitoral (tipo 1) e o Detalhe da Apuração por Seção Eleitoral (tipo 2), nos arquivos nomeados de “votacao_secao_AAAA_BR.txt” e “detalhe_votacao_secao_AAAA_BR.txt”, respectivamente, onde “AAAA” é o ano da eleição, dos anos que houveram eleições gerais federais presidenciais.

Ambos os arquivos trazem os dados utilizados nos boletins de urna, publicado nas seções no momento do encerramento. No conjunto de arquivos 1 há os dados os dados totais de comparecimento em cada seção. Os arquivos do tipo 1 possuem a quantidade de votos por candidato (nulos e brancos são registrados como candidatos) e algumas poucas informações da seção. Os arquivos do tipo 2 possuem a totalização de votos válidos, inválidos e totais da urna.

Foi utilizada a linguagem de programação R (IHAKA, 2018), por ser apropriada para em manipulação, análise e visualização gráfica de dados, para importar e rearranjar os dez arquivos (dos pleitos presidenciais brasileiros de 2002, 2006, 2010, 2014 e 2018) de forma a criar um único arquivo onde cada registro (ou linha de texto) contenha todos os dados úteis de uma seção, incluindo os votos no candidato vencedor e os votos no segundo mais votado, no turno que foi decisivo.

A este arquivo foi acrescentado o campo (variável ou coluna) região (do país) observando-se a unidade da federação onde a seção de votação ocorreu para possibilitar análises considerando-se a região. Outra variável acrescentada foram quatro entropias diferentes.

A primeira entropia foi calculada a partir das frações de voto no primeiro e segundo (V_1 e V_2) colocado pelos votos válidos (V_V):

$$E_1 = -\frac{V_1}{V_V} \cdot \ln\left(\frac{V_1}{V_V}\right) - \frac{V_2}{V_V} \cdot \ln\left(\frac{V_2}{V_V}\right)$$

A segunda entropia foi calculada a partir das frações de voto no primeiro e segundo colocados (V_1 e V_2) e da soma de brancos com nulos (V_{bn}), pelo número de

comparecimento (V_C) às urnas:

$$E_2 = -\frac{V_1}{V_C} \cdot \ln\left(\frac{V_1}{V_C}\right) - \frac{V_2}{V_C} \cdot \ln\left(\frac{V_2}{V_C}\right) - \frac{V_{bn}}{V_C} \cdot \ln\left(\frac{V_{bn}}{V_C}\right)$$

A terceira entropia foi calculada a partir das frações de votos no primeiro e no segundo colocados (V_1 e V_2), a soma de brancos e nulos (V_{bn}), e o número de abstenções (V_{Abs}) pelo eleitorado cadastrado, os aptos (V_A) para votar na seção:

$$E_3 = -\frac{V_1}{V_A} \cdot \ln\left(\frac{V_1}{V_A}\right) - \frac{V_2}{V_A} \cdot \ln\left(\frac{V_2}{V_A}\right) - \frac{V_{bn}}{V_A} \cdot \ln\left(\frac{V_{bn}}{V_A}\right) - \frac{V_{Abs}}{V_A} \cdot \ln\left(\frac{V_{Abs}}{V_A}\right)$$

A quarta entropia foi calculada a partir das frações de votos válidos (V_v), a soma de brancos e nulos (V_{bn}) e o número de abstenções (V_{Abs}) pelo eleitorado cadastrado para votar, os aptos (V_A) na seção:

$$E_4 = -\frac{V_v}{V_A} \cdot \ln\left(\frac{V_v}{V_A}\right) - \frac{V_{bn}}{V_A} \cdot \ln\left(\frac{V_{bn}}{V_A}\right) - \frac{V_{Abs}}{V_A} \cdot \ln\left(\frac{V_{Abs}}{V_A}\right)$$

Com esta base de dados criada em um único arquivo, o passo seguinte foi gerar uma outra base de dados, com um registro por cidade, com as médias de cada variável (coluna ou campo) para cada cidade, possibilitando gerar or gráficos para as análises mostradas no capítulo 5.

6.2 Principais Trechos de Código Fonte em R

Script (programa) em R que gera uma base da dados com um registro por seção de votação:

```

1 iPleito <- 0
2 for (iPleito in 1:7)
3 {
4   esteano<-resultpleitos[iPleito , ]
5   votc1 <- vot.det %>%
6     filter (ANO_ELEICAO==esteano$ANO, NR_TURNO==esteano$TURNO, NR_VOTAVEL==esteano$
7       PRIMEIRO) %>%
8     group_by (ANO_ELEICAO,CD_MUNICIPIO,NR_ZONA,NR_SECAO)
9   votc2 <- vot.det %>%
10     filter (ANO_ELEICAO==esteano$ANO, NR_TURNO==esteano$TURNO, NR_VOTAVEL==esteano$
11       SEGUNDO) %>%
12     group_by (ANO_ELEICAO,CD_MUNICIPIO,NR_ZONA,NR_SECAO)
13   votc1$QT_VOTO1<-votc1$QT_VOTOS
14   votc1$QT_VOTO2<-0
15   votc2$QT_VOTO1<-0
16   votc2$QT_VOTO2<-votc2$QT_VOTOS
17   votacao <- rbind (votc2 , votc1)
18   sqlQuery <- paste0 (
19     "SELECT ",
20     "ANO_ELEICAO,NR_TURNO,SG_UF,CD_MUNICIPIO,NM_MUNICIPIO,NR_ZONA, ",
21     "NR_SECAO,QT_APTOS,QT_COMPARECIMENTO,QT_ABSTENCOES, ",
22     "QT_VOTOS_NOMINAIS,QT_VOTOS_BRANCOS,QT_VOTOS_NULOS, ",
23     "sum(QT_VOTO1),sum(QT_VOTO2) ",
24     "from votacao ",
25     "where ANO_ELEICAO = ",esteano$ANO," ",
26     "group by ANO_ELEICAO, CD_MUNICIPIO, NM_MUNICIPIO, NR_ZONA, NR_SECAO")
27   secao <- sqlDf (sqlQuery)

```

```

26 names(secao)<-c(
27   "ANO_ELEICAO", "NR_TURNO", "SG_UF",
28   "CD_MUNICIPIO", "NM_MUNICIPIO", "NR_ZONA",
29   "NR_SECAO", "QT_APTOS", "QT_COMPARECIMENTO",
30   "QT_ABSTENCOES", "QT_VOTOS_NOMINAIS", "QT_VOTOS_BRANCOS",
31   "QT_VOTOS_NULOS", "QT_VOTO1", "QT_VOTO2")
32 secao$QT_VOTOS_VALIDOS<-secao$QT_VOTOS_NOMINAIS
33 secao$VL_LN_VOTOS_VALIDOS<-ln (secao$QT_VOTOS_VALIDOS)
34 secao$VL_LN_COMPARECIMENTO<-ln (secao$QT_COMPARECIMENTO)
35 secao$VL_LN_APTOS<-ln (secao$QT_APTOS)
36 secao$VL_E1_FV1<-secao$QT_VOTO1/secao$QT_VOTOS_VALIDOS
37 secao$VL_E1_FV2<-secao$QT_VOTO2/secao$QT_VOTOS_VALIDOS
38 secao$VL_ENTROPIA1<-
39   (
40     -secao$VL_E1_FV1*ln (secao$VL_E1_FV1)
41     -secao$VL_E1_FV2*ln (secao$VL_E1_FV2)
42   )
43 secao$VL_E2_FV1<-secao$QT_VOTO1/secao$QT_COMPARECIMENTO
44 secao$VL_E2_FV2<-secao$QT_VOTO2/secao$QT_COMPARECIMENTO
45 secao$VL_E2_FVNB<-(secao$QT_VOTOS_BRANCOS + secao$QT_VOTOS_NULOS) / secao$QT_
    COMPARECIMENTO
46 secao$VL_ENTROPIA2<-
47   (
48     -secao$VL_E2_FV1*ln (secao$VL_E2_FV1)
49     -secao$VL_E2_FV2*ln (secao$VL_E2_FV2)
50     -secao$VL_E2_FVNB*ln (secao$VL_E2_FVNB)
51   )
52 secao$VL_E3_FV1<-secao$QT_VOTO1/secao$QT_APTOS
53 secao$VL_E3_FV2<-secao$QT_VOTO2/secao$QT_APTOS
54 secao$VL_E3_FVNB<-(secao$QT_VOTOS_BRANCOS+secao$QT_VOTOS_NULOS) / secao$QT_APTOS
55 secao$VL_E3_FABS<-secao$QT_ABSTENCOES / secao$QT_APTOS
56 secao$VL_ENTROPIA3<-
57   (
58     -secao$VL_E3_FV1*ln (secao$VL_E3_FV1)
59     -secao$VL_E3_FV2*ln (secao$VL_E3_FV2)
60     -secao$VL_E3_FVNB*ln (secao$VL_E3_FVNB)
61     -secao$VL_E3_FABS*ln (secao$VL_E3_FABS)
62   )
63 secao$VL_E4_FVV<-secao$QT_VOTOS_VALIDOS / secao$QT_APTOS
64 secao$VL_ENTROPIA4<-
65   (
66     -secao$VL_E4_FVV*ln (secao$VL_E4_FVV)
67     -secao$VL_E3_FVNB*ln (secao$VL_E3_FVNB)
68     -secao$VL_E3_FABS*ln (secao$VL_E3_FABS)
69   )
70 todassecoes <-rbind (todassecoes, secao)
71 }
72 write.table (todassecoes, "EntropiaPorSecao-TodosAnos.csv", append = FALSE, sep = ";", ,row.names
    = FALSE, col.names = TRUE)

```

Script (programa) em R que gera uma base de dados com um registro por município, com a soma dos números das seções e com as médias das entropias:

```

1 qtAnos<-nrow (resultpleitos)
2 qtAnos<- strtoi (qtAnos)
3 for (iAno in 1:qtAnos)
4 {
5   regano<-resultpleitos [iAno,]
6   anocorrente<-regano$ANO

```

```

7 | votacaoano<-votacao[votacao$ANO_ELEICAO==anocorrente ,]
8 | municipios <- distinct(votacaoano,SG_UF,CD_MUNICIPIO,NM_MUNICIPIO)
9 | qtMunicipios <- nrow(municipios)
10 | qtMunicipios <- strtoi(qtMunicipios)
11 | for (i in 1:qtMunicipios)
12 | {
13 |   ummunicipio <- municipios[i, ]
14 |   secoesmunicipio <- votacaoano[votacaoano$CD_MUNICIPIO==ummunicipio$CD_MUNICIPIO, ]
15 |   ummunicipio$ANO<-anocorrente
16 |   ummunicipio$VL_MI1<-mean(secoesmunicipio$VL_ENTROPIA1, na.rm=TRUE)
17 |   ummunicipio$VL_MI2<-mean(secoesmunicipio$VL_ENTROPIA2, na.rm=TRUE)
18 |   ummunicipio$VL_MI3<-mean(secoesmunicipio$VL_ENTROPIA3, na.rm=TRUE)
19 |   ummunicipio$VL_MI4<-mean(secoesmunicipio$VL_ENTROPIA4, na.rm=TRUE)
20 |   ummunicipio$QT_VOTOS_VALIDOS<-sum(secoesmunicipio$QT_VOTOS_VALIDOS, na.rm=TRUE)
21 |   ummunicipio$QT_COMPARECIMENTO<-sum(secoesmunicipio$QT_COMPARECIMENTO, na.rm=TRUE)
22 |   ummunicipio$QT_APTOS<-sum(secoesmunicipio$QT_APTOS, na.rm=TRUE)
23 |   ummunicipio$TT_VOTO1<-sum(secoesmunicipio$QT_VOTO1, na.rm=TRUE)
24 |   ummunicipio$TT_VOTO2<-sum(secoesmunicipio$QT_VOTO2, na.rm=TRUE)
25 |   ummunicipio$VL_VOTO1<-mean(secoesmunicipio$QT_VOTO1, na.rm=TRUE)
26 |   ummunicipio$VL_VOTO2<-mean(secoesmunicipio$QT_VOTO2, na.rm=TRUE)
27 |   municipioentropia<-rbind(municipioentropia , ummunicipio)
28 | }
29 | }
30 | write.table(municipioentropia , "Municipio-Entropia-TodosAnos.csv" , append = FALSE, sep = ";" , ,
   | row.names = FALSE, col.names = TRUE)

```