



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA
CURSO DE GRADUAÇÃO EM FÍSICA

ALEF THIAGO FARIAS GUIMARÃES

BUSCA POR PADRÕES NOS RESULTADOS DO SEGUNDO TURNO DAS ELEIÇÕES
PRESIDENCIAIS BRASILEIRAS DE 2022

FORTALEZA

2023

ALEF THIAGO FARIAS GUIMARÃES

BUSCA POR PADRÕES NOS RESULTADOS DO SEGUNDO TURNO DAS ELEIÇÕES
PRESIDENCIAIS BRASILEIRAS DE 2022

Trabalho de Conclusão de Curso apresentado
ao Curso de Graduação em Física do Centro
de Ciências da Universidade Federal do Ceará,
como requisito parcial à obtenção do grau de
bacharel em Física.

Orientador: Prof. Dr. Diego Rabelo da
Costa

Coorientador: Prof. Dr. Raimundo No-
gueira da Costa Filho

FORTALEZA

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

G976b Guimarães, Alef Thiago Farias.
 Busca por padrões nos resultados do segundo turno das eleições presidenciais brasileiras de 2022 / Alef
 Thiago Farias Guimarães. – 2023.
 39 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, 3, Fortaleza, 2023.
Orientação: Prof. Diego Rabelo da Costa.
Coorientação: Prof. Dr. Raimundo Nogueira da Costa Filho.

1. Eleições. 2. Voto. 3. Data analysis. I. Título.

CDD

ALEF THIAGO FARIAS GUIMARÃES

BUSCA POR PADRÕES NOS RESULTADOS DO SEGUNDO TURNO DAS ELEIÇÕES
PRESIDENCIAIS BRASILEIRAS DE 2022

Trabalho de Conclusão de Curso apresentado
ao Curso de Graduação em Física do Centro
de Ciências da Universidade Federal do Ceará,
como requisito parcial à obtenção do grau de
bacharel em Física.

Aprovada em: 13/12/2023

BANCA EXAMINADORA

Prof. Dr. Diego Rabelo da Costa (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Raimundo Nogueira da Costa
Filho (Coorientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Saulo Davi Soares e Reis
Universidade Federal do Ceará (UFC)

Prof. Dr. Jorge Luiz Coelho Domingos
Universidade Federal do Ceará (UFC)

A minha família

AGRADECIMENTOS

Dedico este trabalho aos meus pais, que apesar dos percalços, nunca desistiram e sempre me apoiaram.

À minha cadela, Nicky, que sempre me acompanha nos meus estudos e nas minhas rotinas matinais.

Ao meu padrinho, que acreditou em mim e contribuiu muito para minha formação.

Aos meus tios, Ranilson e Germana, que sempre apoiaram meu desenvolvimento intelectual.

À minha tia Rita, que me deu meu primeiro livro.

Ao meu tio Roberto, pelas discussões que me ajudaram no desenvolvimento deste trabalho.

Ao meu amigo Luis, por sempre me lembrar de que mudanças nem sempre são ruins.

À minha namorada Lucilia, por estar sempre ao meu lado.

Aos meus amigos, Mariana, Seza e Jussara, que fizeram os tempos entre aulas passarem mais rápido.

Ao professor Ramos, por ter uma didática incomparável e ser uma fonte de inspiração.

Aos professores Raimundo e Diego por terem me guiado na execução deste trabalho.

Todos aqui citados contribuíram para que eu possa ser quem sou hoje, e por isso, vocês têm a minha profunda gratidão.

“There is no reason to be alive if you can’t do
deadlift”

(Jon Pall Sigmarsson)

RESUMO

Desde o segundo turno das eleições brasileiras presidenciais de 2014, quando o resultado foi questionado pelo partido perdedor, tem sido levantado um debate em contrassenso das literaturas predominantes, isto é, de que as urnas são seguras e o processo é eficiente. Apesar da falta de evidências concretas sobre violações das urnas, é interessante buscarmos mecanismos que demonstrem adulterações nas urnas que estão para além da inspeção física dos dispositivos e da análise dos *softwares* neles empregados. Para isso, recorreremos à estatística, buscando identificar padrões que emergem ao analisarmos as distribuições dos votos, pois desvios dessa norma podem ser indícios de fraudes sistemáticas nas urnas. Neste contexto, realizamos uma análise estatística dos dados retirados dos boletins de urnas disponíveis no portal dados abertos do tribunal superior eleitoral dos anos de 2014, 2018 e 2022, tomando o censo populacional dos municípios do Instituto Brasileiro de Geografia e Estatística. Resultados das distribuições de $\tau = \log(V/(N - V))$, onde V e N correspondem a quantidade de votos registrados na urna e o número total de votos aptos nas urnas, respectivamente, foram discutidos para as capitais, estados e regiões, e comparados com as eleições presidenciais dos anos de 2014, 2018 e 2022.

Palavras-chave: Eleições. Votos. Análise de dados.

ABSTRACT

Since runoff of the Brazilian presidential elections of 2014, when the outcome was contested by the losing party, there has been an ongoing debate that goes against the prevailing literature, that is, the voting machines are safe and the process is efficient. Besides the lack of evidence about the violation of the ballots, it's interesting to explore mechanisms that demonstrate adulterations with the ballots beyond the physical inspections of the devices and the analyses made on their softwares employed in them. To address that, we resort to statistics, aiming to identify patterns that emerges when analyzing the vote distribution, as deviations from the norm maybe evidence of systematic ballots fraud. In this context, we conducted a statistical analysis of the data extracted from the ballots reports available on the *Tribunal Superior Eleitoral* website for the years of 2014, 2018 and 2022, and the population census of the counties in the *Instituto Brasileiro de Geografia e Estatística*. Results of the distributions $\tau = \log(V/(N - V))$, where V and N correspond to the total votes cast in a given ballot and the total votes that could be cast in the ballot, respectively, were discussed for capital, states and regions, and compared with presidential elections on the years of 2014, 2018 and 2022.

Keywords: Elections. Votes. Data analysis.

LISTA DE FIGURAS

- Figura 1 – A figura da esquerda mostra a distribuição de τ para as *communes* francesas nos anos de análise de 1992 à 2009. Como a média muda de ano para ano, a gaussiana é centralizada para facilitar a comparação. O *inset* no canto superior mostra a distribuição da taxa de comparecimento p , que é a quantidade de votantes dividida pela quantidade de votantes aptos. No painel da direita temos a correlação espacial da variável τ 14
- Figura 2 – A figura à esquerda mostra como as distribuições de tau, que aqui ele chama de *nu*, aparecem em diferentes votações. Já a figura da direita mostra a relação entre comparecimento nas urnas e votos para o vencedor. É de se esperar que o número de urnas em que 100% dos votos foram para o vencedor, ou seja pontos no canto superior direito, seja reduzido. 15
- Figura 3 – Essa foi a cédula de votação que confundiu eleitores americanos na votação de 2000. Apesar do campo do Al Gore conter uma seta com o número cinco indicando o local do furo do voto, muitos confundiram que a indicação do local do voto era a linha que demarca o espaço do Bush com o de Al Gore e por isso acabaram depositando seu voto no 4 furo. 17
- Figura 4 – Histograma para 100 números aleatórios entre 0 e 1. 19
- Figura 5 – Histogramas para 100 amostras aleatórias de uma distribuição gaussiana com variância de 0.6. Na imagem da esquerda temos 2 *bins*, na central 8, e na direita 32 *bins*. 20
- Figura 6 – Histogramas para 100 amostras aleatórias de uma distribuição gaussiana com variância de 0.6. Na imagem à esquerda temos um histograma comum, no centro temos a representação do histograma por pontos e na direita temos um histograma normalizado. 20
- Figura 7 – Histograma dos coeficientes de urna τ da cidade de São Paulo. 25
- Figura 8 – Resultado da distribuição dos votos contabilizados no segundo turno das eleições presidenciais de 2022 de todas as urnas do Brasil. Aqui foram utilizados 100 *bins*, pois a regra de Sturges resulta em poucos pontos. No eixo horizontal temos o valor de τ das urnas e no eixo vertical a probabilidade de uma urna ter esse valor. 27

Figura 9 – Para esse gráfico ordenamos o τ de todas as zonas eleitorais das eleições presidenciais de 2022 em uma lista de ordem crescente e separando elas em vinte grupos de mesmo tamanho e então calculamos a média de tau e do número de votantes na zona. Cada ponto desse representa as médias dos grupos. O resultado diverge do encontrado na eleições francesas onde a curva é sempre decrescente	28
Figura 10 – Distribuição de τ por região apresentada de forma simétrica e obtida ao se subtrair as distribuições ao seu valor médio, facilitando assim a comparação por terem todas um centro de distribuição comum localizado no 0.	28
Figura 11 – Distribuições dos coeficientes τ associadas a votação do segundo turno presidencial para as urnas dos 26 estados mais o distrito federal do território brasileiro no ano de 2022.	30
Figura 12 – Distribuições dos coeficientes τ associadas a votação do segundo turno presidencial para as urnas das 27 capitais brasileiras no ano de 2022.	31
Figura 13 – Relações entre os vários dados extraídos dos vários municípios brasileiros: densidade populacional, média da gaussiana (x_0), variância do gaussiana (s) e o erro quadrático médio (RMSE).	33
Figura 14 – Imagem tirada pelo celular de Marília de Alencar, então diretora de Inteligência do Ministério da Justiça momentos após o começo da reunião com o ex-ministro Anderson Torres (SERRA, 2023). Na imagem podemos ver pelo título que há um filtro para as cidades com mais de 75% de votantes no Lula, com a lista ordenada pelas cidades com a maior quantidade de votantes absolutos no Lula.	34
Figura 15 – Mapa da variação do τ médio do segundo para o primeiro turno nas eleições federais de 2014, 2018 e 2022. Note que cada mapa tem um intervalo diferente que foi definido pelo maior valor dos 10% menores valores e pelo menor valor dos 10% maiores.	34

Figura 16 – Variação de τ do segundo para o primeiro turno por município. No eixo horizontal temos a diferença percentual dos votos no primeiro turno do candidato vencedor no segundo turno para o perdedor do segundo turno nos municípios. As cores indicam essa diferença para o segundo turno. Para minimizar o erro os gráficos contêm apenas as cidades com população maior que 10 mil habitantes. É bem evidente que existem uma relação linear entre as variáveis, por isso a regressão linear desses pontos está presente nesses gráficos. Note que quanto mais uma cidade vota no candidato orientado à esquerda menor é o índice τ , o que indica um menor engajamento dos habitantes da cidade, e quanto mais a cidade vota no candidato à direita do espectro político maior tende a ser o índice τ 35

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Física Social	13
1.2	Eleições	13
1.3	Problematização	16
1.4	Escopo do trabalho	17
2	TEORIA	19
2.1	Histogramas	19
2.2	Erro quadrático médio	21
2.3	Distribuição Gaussiana	21
2.4	Outros conceitos importantes	22
2.4.1	<i>Normalização de distribuições</i>	<i>22</i>
2.4.2	<i>Média (μ)</i>	<i>22</i>
2.4.3	<i>Variância (σ^2)</i>	<i>23</i>
2.4.4	<i>Assimetria</i>	<i>23</i>
2.4.5	<i>Curtose</i>	<i>23</i>
3	METODOLOGIA	24
3.1	Coeficiente de Urna	24
3.2	Tratamento de Dados	25
3.3	Distribuição	25
4	RESULTADOS	27
4.1	Resultados das distribuições no brasil e nas regiões	27
4.2	Resultados por estado	29
4.3	Resultados por cidades	29
4.4	Correlações de dados	32
4.5	Interferência da PRF nas eleições de 2022	32
5	CONCLUSÕES E TRABALHOS FUTUROS	36
	REFERÊNCIAS	37

1 INTRODUÇÃO

1.1 Física Social

Tentar compreender o comportamento humano através de leis matemáticas é uma prática antiga. Uma das primeiras tentativas de realizar isso foi feita por Thomas Hobbes em seu livro “Leviatã”, onde ele, inspirado por Euclides e Galileu, tenta buscar axiomas e derivar postulados sobre o comportamento humano (BALL, 2004). Em sua concepção T. Hobbes imaginou o homem como uma máquina complexa, e portanto deveria seguir leis como os objetos obedecem a lei da inércia. Essa visão positivista e idealizada sobre o comportamento humano não foi muito longe por falhar em fazer previsões, já que seus axiomas faziam suposições muito fortes sobre o comportamento humano.

Cerca de 11 anos após a publicação do “Leviatã”, surgiu o livro “Observações sobre os Censos de Mortalidade” (do inglês, *Observations on the Bills of Mortality*) de John Graunt. Esse livro faz uma análise dos censos de mortalidade publicados na Inglaterra e com isso semeia as primeiras ideias da estatística. Esse ramo da matemática é no fim do século XIX aplicado à física por James Clerk Maxwell e Ludwig Boltzmann que descrevem toda a termodinâmica a partir da teoria atômica da matéria. Acredita-se que eles foram, pelo menos parcialmente, inspirados a abandonar o conceito de descrição de trajetórias por um modelo estatístico por conta das diversas leis quantitativas que surgiam do comportamento de grupo das pessoas (CASTELLANO *et al.*, 2009).

A física estatística surge como uma consequência dos estudos de dinâmica dos gases (FLAMM, 1998). Mais tarde, foi notado que ela também poderia ser utilizada para descrever vários outros sistemas, dentre eles estavam as dinâmicas sociais. Contudo, o acesso à informação para fazer as análises foi um empecilho para o desenvolvimento desse ramo no começo do século XX, o que mudou com a chegada da internet. Agora, com um acesso à grandes quantias de dados esse ramo da física aplicado à análise da sociedade tem crescido de forma bastante expressiva.

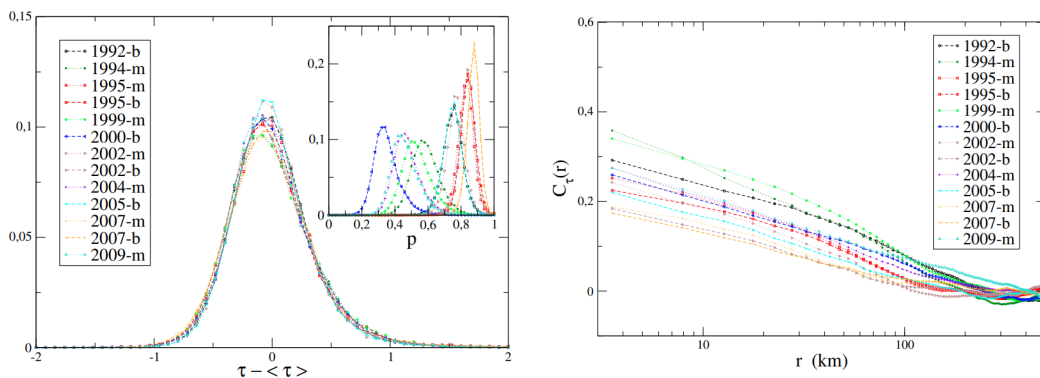
1.2 Eleições

Entender os processos eleitorais e seus resultados é um assunto de altíssima relevância, pois esses sistemas impactam profundamente a vida das pessoas. Dado a sua relevância, não é de se espantar que haja desconfiança nesse processo, por isso, são buscados mecanismos de

segurança para aumentar a confiabilidade do processo.

Nos anos recentes, a crescente aplicação da física estatística como ferramenta de análise social tem mostrado diversos padrões emergentes que caracterizam as votações (CASTELLANO *et al.*, 2009). O artigo (BORGHESI; BOUCHAUD, 2010) intitulado *Spatial correlations in vote statistics: a diffusive field model for decision-making* é um exemplo desse tipo de estudo das votações no caso específico da França. Para ilustrar a análise feita no artigo (BORGHESI; BOUCHAUD, 2010), apresentamos na Fig. 1 dois resultados discutidos para a votação da França associados ao parâmetro τ , que mede o comparecimento nas urnas e será o mesmo abordado posteriormente no presente trabalho. Na Fig. 1 à esquerda, nota-se que as distribuições dos τ 's seguem distribuições gaussianas que são levemente assimétricas e possuem curtose baixas. No painel da direita, é mostrado que há uma correlação espacial positiva para esse parâmetro.

Figura 1 – A figura da esquerda mostra a distribuição de τ para as *communes* francesas nos anos de análise de 1992 à 2009. Como a média muda de ano para ano, a gaussiana é centralizada para facilitar a comparação. O *inset* no canto superior mostra a distribuição da taxa de comparecimento p , que é a quantidade de votantes dividida pela quantidade de votantes aptos. No painel da direita temos a correlação espacial da variável τ .

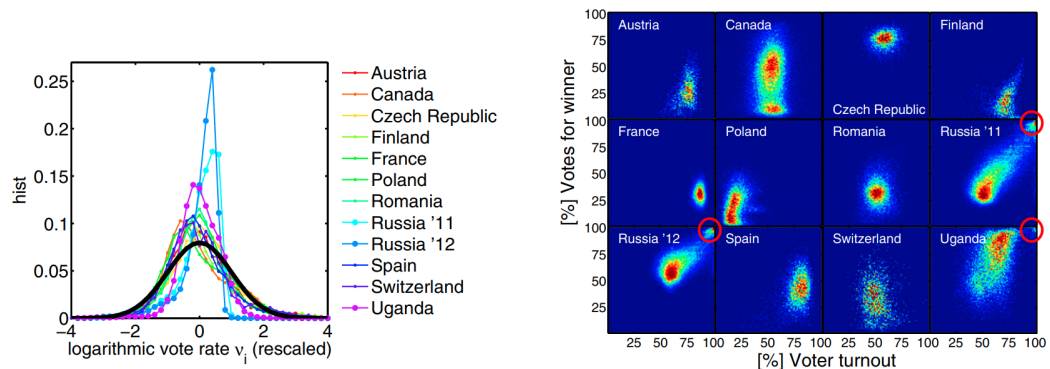


Fonte: (BORGHESI; BOUCHAUD, 2010)

Desvios desses padrões podem indicar irregularidades sistemáticas nas votações como foi reportado no artigo *Statistical detection of systematic election irregularities* (KLIMEK *et al.*, 2012), onde as eleições da Rússia e de Uganda se destacam por terem padrões bem distintos dos demais como podemos ver na Fig. 2.

Ao fazer esse tipo de análise, ganhamos além de um entendimento mais profundo do processo eleitoral, uma ferramenta que contribui para o fortalecimento da confiabilidade no processo. O primeiro e um dos exemplos mais conhecidos de aplicação da física estatística para estudar sistemas eleitorais foi o utilizado para descrever uma lei de potência encontrada nas frações de votos das eleições para deputado estadual de 1998, reportado na Ref. (FILHO *et*

Figura 2 – A figura à esquerda mostra como as distribuições de tau, que aqui ele chama de ν , aparecem em diferentes votações. Já a figura da direita mostra a relação entre comparecimento nas urnas e votos para o vencedor. É de se esperar que o número de urnas em que 100% dos votos foram para o vencedor, ou seja pontos no canto superior direito, seja reduzido.



Fonte: (KLIMEK *et al.*, 2012)

al., 1999). Quando o estudo foi replicado para eleições em outros países, descobriu-se que o coeficiente da potência encontrado não era universal, apenas a lei de potência em si. Mais tarde, esse comportamento foi descrito utilizando o modelo de Sznajd, que é uma extensão do modelo de Ising da física de estado sólido.

Para entender como funcionam as eleições devemos entender como ocorre o processo de escolha dos vencedores. Para deputado estadual, deputado federal e vereador, temos um sistema de escolha proporcional. Nele, todos os votos do estado são somados e divididos pelo número de cadeiras que o estado possui, o valor obtido é o quociente eleitoral do estado. Então soma-se a quantidade de votos que um partido obteve e divide esse valor pelo quociente eleitoral. Esse valor arredondado para baixo dirá quantas vagas o partido terá direito. As n vagas conquistadas pelo partido serão ocupadas pelos n candidatos mais votados do partido (TRIBUNAL SUPERIOR ELEITORAL, 2022a).

Diferente desse sistema, as eleições para prefeito, governador, senador e presidente são majoritárias. Esse sistema é mais direto, ganha quem tiver mais votos absolutos, ou seja, quem tiver mais que 50% dos votos. Caso a maioria não seja alcançada, é feito um segundo turno com os dois candidatos mais votados, nesse caso basta a maioria para vencer (TRIBUNAL SUPERIOR ELEITORAL, 2022a). Durante esse trabalho analisarei o padrão emergente nas proporções entre as pessoas que foram votar e as que se abstiveram no segundo turno das eleições de 2022.

Uma característica importante do nosso sistema eleitoral é o voto obrigatório. Poucas nações no mundo adotaram esse sistema quando se tornaram democracias e muitas delas

abandonaram a prática com o tempo. Alguns países que ainda adotam essa regra são Uruguai, Bélgica, Argentina e Austrália (TRIBUNAL SUPERIOR ELEITORAL, 2022b).

Outro diferencial das nossas eleições são as urnas eletrônicas para todas as eleições. Não são incomuns países que adotam ela para eleições menores como a de prefeito, mas são poucos que adotam para eleições grandes como são as presidenciais no nosso país (SE, 2022).

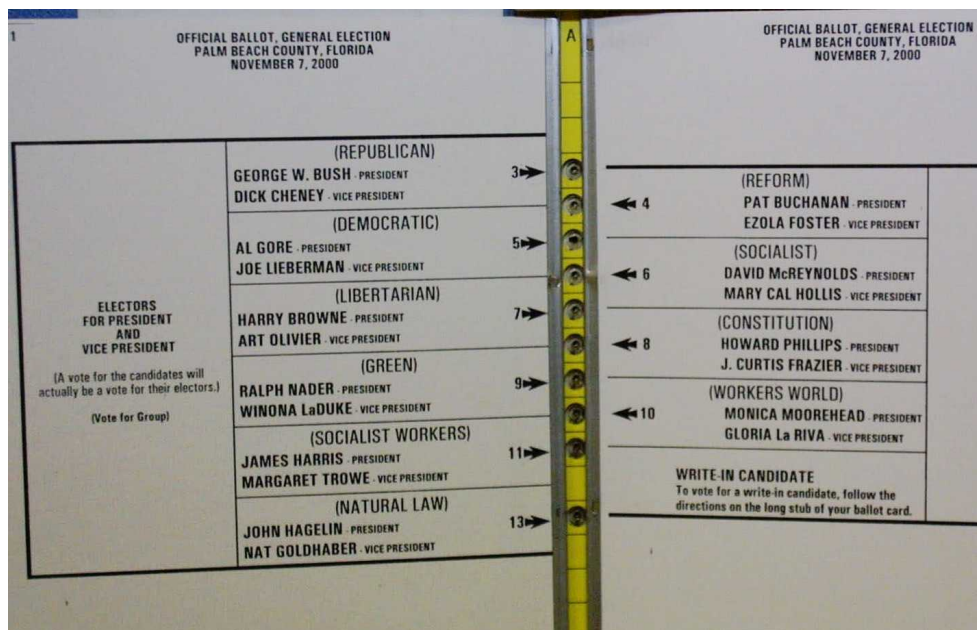
1.3 Problemática

A democracia, apesar de ser um ótimo sistema político, tem suas falhas. Ainda no berço da democracia, Atenas, Platão apontava que a natureza da democracia de levar à governança os mais queridos à população e não os mais qualificados faziam dela um péssimo sistema (BBC, 2021). Existem também problemas com fraudes eleitorais. No Brasil por exemplo, o código eleitoral prevê punições para algumas, como: tentar votar duas vezes, tentar votar em nome de uma pessoa falecida ou não (ANDRADE, 2023)(TRIBUNAL SUPERIOR ELEITORAL, 2023). Além disso o voto de cabresto é um fenômeno que caracteriza fortemente eleições do começo do século passado mas não se limita à esse período. Além das inúmeras falhas no próprio sistema de votação como os problemas com os votos borboleta nos anos 2000 nos estados unidos que gerou uma recontagem de votos (SANDOVAL, 2020).

O problema exposto por Sócrates é sistemático e geralmente leva debates políticos e campanhas políticas para rumos que não são construtivos. Um exemplo clássico na política brasileira são os políticos comendo pastel durante a campanha. Essa ação por si não mostra a capacidade de governar de um individuo, afinal, qualquer um pode ir em uma feira comer pastel. Contudo, esse ato aproxima o político dos eleitores, fazendo que eles recebam mais votos [ver reportagem na Ref. (KLISIEWICZ, 2022)].

Sempre são testadas novas formas de realizar a votação e com essas novas formas virão problemas e benefícios. O caso americano citado acima ocorreu pois estavam testando uma forma de computar os votos em um cartão chamado cédula borboleta, como mostrado na Fig. 3. Nele o eleitor tinha que furar um cartão em um local pré-determinado para indicar o voto no candidato desejado. Começaram a desconfiar que algo não estava certo quando as contagens de votos indicavam altos números de votantes em um candidato acusado de ser antisemita e racista em bairros de maioria judeu e negra, e após uma inspeção foi determinado que votantes do candidato Al Gore estavam se confundindo com o sistema e votando no candidato errado. Por conta desse erro foi tentada uma recontagem, mas no final quem decidiu o resultado foi a

Figura 3 – Essa foi a cédula de votação que confundiu eleitores americanos na votação de 2000. Apesar do campo do Al Gore conter uma seta com o número cinco indicando o local do furo do voto, muitos confundiram que a indicação do local do voto era a linha que demarca o espaço do Bush com o de Al Gore e por isso acabaram depositando seu voto no 4 furo.



Fonte: (PREVIDELLI, 2020)

suprema corte americana (SANDOVAL, 2020; PREVIDELLI, 2020).

Quando o Brasil inovou em 1996 com a urna eletrônica, muita suspeita foi levantada sobre esse sistema. Por um lado, dificultou-se a prática do voto de cabresto e de muitas fraudes como as citadas acima (ANDRADE, 2023), mas por outro a falta de um sistema físico como um papel fez as pessoas desconfiarem do processo.

Especialmente nos últimos anos têm sido levantado dúvidas sobre a confiabilidade no sistema, então caracterizar os padrões emergentes das eleições pode nos ajudar a evidenciar as peculiaridades e as similaridades de cada eleição. Isso pode ser uma forma de evidenciar as possíveis fraudes e falhas no sistema em um dado ano. Tendo isso em mente, esse trabalho busca alguns desses padrões com o fim de encontrar se existem diferenciações nos padrões esperados.

1.4 Escopo do trabalho

Esse trabalho começa com uma apresentação das ferramentas estatísticas básicas utilizadas para a realização da análise que se segue, onde iremos observar distribuições características que emergem dos resultados das eleições de 2022. Tais conceitos são apresentados no Capítulo 2. A metodologia utilizada na análise, coleta e tratamento dos dados é apresentada no

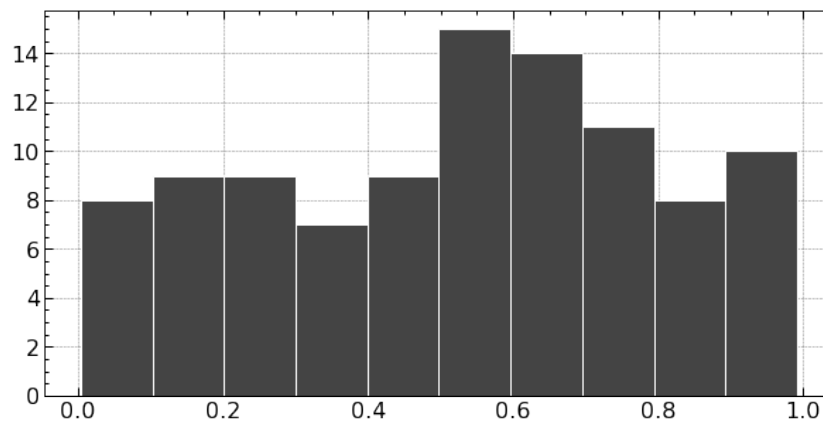
Capítulo 3. Os resultados são discutidos no Capítulo 4. O foco da análise será no entorno do parâmetro que está relacionado aos índices de comparecimento das urnas, o mesmo discutido na Ref. (CASTELLANO *et al.*, 2009) e mostrado na Fig. 2. Com a análise similar, agora para o segundo turno das eleições presidenciais brasileiras de 2022, tentaremos identificar alguma relação das tendências estatísticas obtidas com o impacto das *blitze* das polícias federais realizadas no segundo turno da eleição de 2022 sobre a região Nordeste.

2 TEORIA

2.1 Histogramas

Histogramas são ferramentas muito comuns quando analisamos dados, pois eles expõem a frequência com que uma variável aparece em um dado intervalo. Para compreendermos bem sua funcionalidade e como eles vão aparecer nesse trabalho vamos entender primeiro como eles são montados. Começamos obtendo o valor máximo e mínimo de um conjunto de dados, então dividimos o espaço entre eles em intervalos que chamamos de barras ou do inglês *bins*. Normalmente esses intervalos têm o mesmo comprimento porém isso não é obrigatório. Definidos os *bins*, contamos quantas medidas estão contidas dentro de cada intervalo e com isso desenhamos um retângulo sobre cada *bin* que tem a altura dessa quantidade. Podemos ver isso em prática na Fig. 4.

Figura 4 – Histograma para 100 números aleatórios entre 0 e 1.



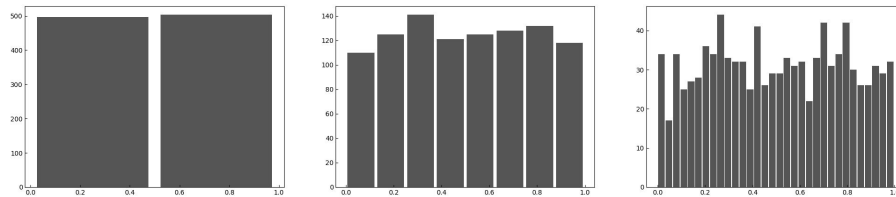
Fonte: Elaborado pelo autor.

Definimos *bins* matematicamente como $[a, b)$, ou seja, inclui a até b , mas não contém b , com exceção do último *bin* que deve conter o último valor do histograma. Quando olhamos para o exemplo da Fig. 4 não vemos um padrão claro, contudo em análise de dados de uma certa amostra, procura-se por algum tipo de padrão que o histograma deve nos ajudar a ilustrar. Para isso devemos saber escolher a quantidade correta de *bins*. Quando escolhemos poucos intervalos o padrão que buscamos pode não aparecer, já quando escolhemos *bins* em excesso o ruído nos dados irá atrapalhar na visualização, como podemos ver na Fig. 5.

Existem vários meios para determinar o número ideal de intervalos a serem utilizados, um dos mais simples é $k = \lceil \sqrt{n} \rceil$, onde k é o número de *bins* e n é o número de amostras. Porém,

para amostras grandes essa expressão entrega intervalos em excesso. Também é frequente o uso da fórmula de Sturges (STURGES, 1926): $k = \lceil \log_2(n) + 1 \rceil$. Nela os valores de k para grandes n são menores e no geral são mais satisfatórios, portanto esse foi o critério utilizado nos histogramas do presente trabalho. Contudo, é bom lembrar que apesar da fórmula entregar boas estimativas outros valores também foram testados e o mais adequado a cada situação foi escolhido.

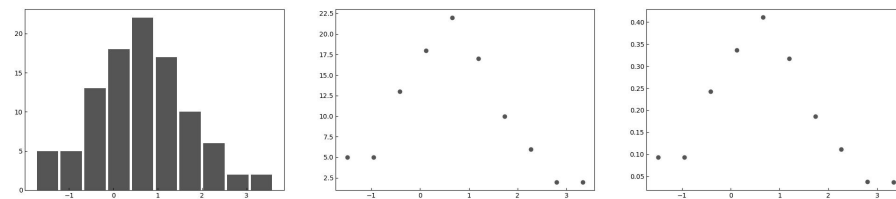
Figura 5 – Histogramas para 100 amostras aleatórias de uma distribuição gaussiana com variância de 0.6. Na imagem da esquerda temos 2 *bins*, na central 8, e na direita 32 *bins*.



Fonte: Elaborado pelo autor.

Em alguns casos ocorre que o excesso de barras pode ser um poluente gráfico visual, que pode impactar negativamente na compreensão e interpretação do histograma. Por isso não é raro representarmos a barra por um ponto localizado no meio do *bin* e com a mesma altura da barra, como podemos ver na Fig. 6.

Figura 6 – Histogramas para 100 amostras aleatórias de uma distribuição gaussiana com variância de 0.6. Na imagem à esquerda temos um histograma comum, no centro temos a representação do histograma por pontos e na direita temos um histograma normalizado.



Fonte: Elaborado pelo autor.

Uma última característica que é preciso ressaltar é o fato que quando os espaçamentos dos *bins* é constante, o histograma mostra uma densidade de dados que podemos transformar isso em uma probabilidade de encontrar uma nova medida em um *bin*, simplesmente normalizando o histograma, isto é, multiplicar a altura de todos os retângulos por uma constante de tal forma que soma da área de todos eles passe a ser 1. Essa constante é a inversa da área total da curva, já que queremos $1 = A \sum y_i \Delta x$, então $A = \frac{1}{\sum y_i \Delta x}$. O exemplo de um histograma normalizado pode ser visto na Fig. 6.

2.2 Erro quadrático médio

Há varias maneiras de quantificar o quanto uma curva teórica descreve um conjunto de pontos. Para calcular uma delas, comecemos com os resíduos:

$$r_i = y_i - \hat{y}_i, \quad (2.1)$$

onde y_i é o valor medido do i -ésimo ponto da amostra e \hat{y}_i é o i -ésimo valor predito pelo modelo. O resíduo vai nos dar a distância entre o ponto teórico e o experimental, assim quanto menor o seu valor mais justa é a predição do modelo. O próximo passo seria tirar a média, porém se somarmos todos os resíduos diretamente o resultado não seria útil, pois poderíamos cair em casos onde os resíduos estão distribuídos como no conjunto $r = \{-1, 2, 1, 0, -2\}$ que teria o valor médio 0 e $r = \{0, 0, 0, 0\}$ que também tem a média igual a zero. Claramente o segundo conjunto é mais adequado, então para evitar esses casos somamos os valores dos resíduos ao quadrado. No exemplo, tem-se que o primeiro conjunto ficará então com a média quadrada de 2 enquanto que o segundo terá valor 0. Assim, tem-se um parâmetro que quanto mais próximo de 0 é o seu valor mais adequada é a curva para o conjunto de dados. O que obtemos nesse processo é o chamado erro quadrático médio que por si só poderia ser o parâmetro que analisaria a adequação no modelo, contudo é comum usarmos a raiz quadrada desse valor pois a unidade dele é a mesma da variável y . A raiz do erro quadrático médio (da sigla em inglês *RMSE* para *Root Mean-Square Error*) é o parâmetro de adequação ideal para utilizar e sua fórmula completa é:

$$RMSE = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{N}}. \quad (2.2)$$

2.3 Distribuição Gaussiana

Um dos mais importantes teoremas da estatística é o teorema central do limite (SANDERSON, 2023). De maneira simplificada, ele diz que se sortearmos n conjuntos X_i de uma distribuição $P(x)$ e então calcular as médias deles, e com isso fizermos um histograma, a distribuição que vai surgir é uma gaussiana, independente de $P(x)$. Por conta desse teorema essa distribuição é bastante comum e aparece em diversas áreas da ciência. Vamos então entender um pouco mais sobre ela.

A distribuição gaussiana já normalizada é dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad (2.3)$$

em que μ corresponde ao valor esperado do conjunto de dados e σ ao desvio padrão que é raiz quadrado da variância. A curva da Eq. (2.3) tem um formato de sino e é simétrica. A média, a moda e a mediana coincidem no pico da distribuição em $x = \mu$, enquanto σ dita o quão dispersa é a distribuição. Graficamente, um valor baixo de σ significa que a curva tem um pico alto por volta de μ e que decai rapidamente para qualquer direção, enquanto um alto valor de σ denota uma curva mais larga. Com o desvio padrão é possível calcular o intervalo onde uma porcentagem x das medidas tem chance de estar, por exemplo: é esperado que 68.3% dos valores vão estar há 1σ para mais ou para menos de μ enquanto 95% dos valores vão estar há 2σ de distância.

2.4 Outros conceitos importantes

2.4.1 Normalização de distribuições

Toda distribuição $P(x)$ deve passar por um processo de normalização para que a área da curva sobre o intervalo (a, b) seja idêntica a probabilidade de encontrar um novo valor dentre desse intervalo. Para garantir isso devemos assegurar que a probabilidade de encontrar qualquer valor sob todo o domínio de $P(x)$ deve ser de 100% e para isso multiplicamos $P(x)$ por uma constante A . Sabendo disso, temos que a constante de normalização A no caso discreto pode ser escrita como

$$1 = A \sum P(x_i) \Rightarrow A = \frac{1}{\sum P(x_i)}, \quad (2.4)$$

e no caso contínuo, tal como

$$1 = A \int P(x) dx \Rightarrow A = \frac{1}{\int P(x) dx}. \quad (2.5)$$

2.4.2 Média (μ)

Uma das características mais básicas quando falamos da descrição de distribuições estatísticas é a medida da média. Para calculá-la somamos todos os possíveis valores da distribuição multiplicados pela probabilidade deles acontecerem, assim temos para o caso discreto que:

$$\mu = E[X] = \sum_{i=0}^n x_i P(x_i), \quad (2.6)$$

e no caso contínuo ela se torna

$$\mu = E[X] = \int xP(x)dx. \quad (2.7)$$

O valor esperado é uma generalização do conceito de média aritmética. Podemos ver isso no caso em que qualquer elemento tem igual probabilidade de ser escolhido, implicando em $P(x) = 1/N$. Assim, tem-se que

$$E[X] = \sum_{i=0}^n x_i P(x_i) = P(x) \sum_{i=0}^n x_i = \frac{1}{N} \sum_{i=0}^n x_i, \quad (2.8)$$

correspondendo a própria definição de média aritmética.

2.4.3 Variância (σ^2)

Outro parâmetro importante para estudar distribuições é a variância. Esta quantidade nos fornece uma noção de quanto os valores se afastam do valor esperado e portanto o quão dispersa é a distribuição. O variância é dada por

$$\sigma^2 = \text{var}(X) = E[(X - E[X])^2]. \quad (2.9)$$

2.4.4 Assimetria

Para quantificar a assimetria de uma curva utilizamos o parâmetro

$$\gamma = E \left[\left(\frac{X - E[X]}{\sigma} \right)^3 \right]. \quad (2.10)$$

De acordo com a Eq. (2.10), tem-se que quanto mais positivo (negativo) é o valor de γ , mais o centro de massa da distribuição vai estar para à direita (esquerda). Logicamente, quando esse valor for 0, a curva é simétrica.

2.4.5 Curtose

A curtose, dada por

$$\text{Kurt}[X] = E \left[\left(\frac{X - E[X]}{\sigma} \right)^4 \right], \quad (2.11)$$

é um parâmetro que quantifica o quão achatada é uma curva. Valores negativos indicam uma curva achatada. Valores positivos indicam uma curva mais pontiaguda. Já o 0 identifica uma gaussiana.

3 METODOLOGIA

3.1 Coeficiente de Urna

Para realizar a análise estatística dos votos, foram testados vários parâmetros e relações entre eles. Começando a apresentação aqui deles pelo mais simples, verificou-se diretamente se os votos totais (V), os votos totais para o candidato vencedor (W) e para o perdedor (L) apresentavam alguma distribuição. Entretanto, nenhum padrão ficou evidente nesta primeira observação dos dados.

Em seguida, buscou-se relações entre os parâmetros: V , W e L . Analisamos distribuições de $q = W - L$, $q = W/N$, $q = (W - L)/N$ e $q = (W - L)/V$, onde N denota o total de votos aptos nas urnas. As três últimas expressões geraram distribuições semelhantes a gaussianas, porém a última expressão para q gerou menores RMSE. Com base nesse resultado, a priori essa análise de q apresentava-se ser a coeficiente da análise mais apropriado. Contudo, revisando a literatura da área, em especial o artigo (BORGHESI; BOUCHAUD, 2010) que fez uma análise parecida com a que eu estamos propondo, observou-se que um coeficiente mais apropriado seria o seguinte:

$$\tau = \log \left(\frac{V}{N - V} \right). \quad (3.1)$$

Avaliando o emprego do coeficiente da Eq. (3.1), vimos que os RMSE's apresentaram resultados mais baixos. Dessa maneira, decidimos adotá-lo. Para entender esse coeficiente devemos notar que ele faz uma fração entre as pessoas que votaram (V) e as que se ausentaram ($N - V$). Portanto, quanto maior o número de pessoas que compareceram no dia da votação maior é o valor de τ . Vale notar que quando todos comparecem, tem-se que $\lim_{N \rightarrow V} \tau = \infty$, o que leva a um problema na análise. Para contornar esse problema, removeremos os pontos associados a tais casos. Da Eq. (3.1), teremos o valor de 0 quando aproximadamente metade das pessoas comparecerem e valores menores que 0 quando tivermos menos da metade. Podemos ver então outro problema, quando todos se ausentam: $\lim_{N \rightarrow 0} \tau = -\infty$. Assim, para resolver esse problema, também removemos esses pontos.

Baseada na breve discussão da presente seção, nota-se que uma análise utilizando-se o parâmetro τ se mostra apropriada. Espera-se uma leve assimetria negativa nas curvas de distribuições τ , enquanto que sistemas com possíveis irregularidades o que se observa é uma assimetria positiva e uma curtose elevada (KLIMEK *et al.*, 2012).

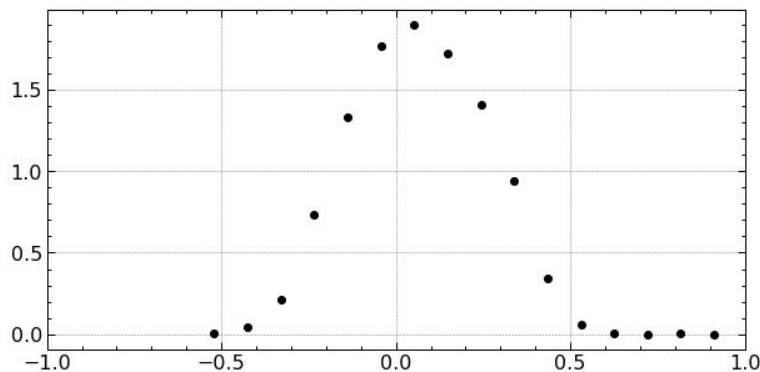
3.2 Tratamento de Dados

Os dados aqui trabalhados foram retirados dos boletins de urna disponíveis no site do portal dados abertos do Tribunal Superior Eleitoral (TSE) (TRIBUNAL SUPERIOR ELEITORAL, 2022c). Através da linguagem de programação Python, das bibliotecas Pandas e Numpy e do Jupyter notebook, extraímos de cada boletim de urna os dados de interesse: número da zona eleitoral, número da seção, código do município, nome do município, sigla da unidade federal, quantidade de votantes aptos, quantidade de comparecimentos e quantidade de abstenções. Para complementar os dados, extraímos do censo do Instituto Brasileiro de Geografia e Estatística (IBGE) de 2022 (IBGE, 2022) a população e a área dos municípios. Toda manipulação e representação gráfica dos dados também foi feita em Python.

3.3 Distribuição

Para ilustrar a análise dos dados, como discutiremos no próximo capítulo para as diferentes capitais, estados e regiões, apresentamos na Fig. 7 o histograma da distribuição do parâmetro τ para a cidade mais populosa do país, isto é São Paulo, pois devido ao seu espaço amostral grande qualquer distribuição deve ficar evidente. É fácil perceber pela Fig. 7 que a distribuição se assemelha à uma gaussiana, e de fato ao utilizar ela para modelar as curvas obtemos baixos RMSE's que diminuem quanto maior é o N amostral, como veremos no próximo capítulo na Fig. 13.

Figura 7 – Histograma dos coeficientes de urna τ da cidade de São Paulo.



Fonte: Elaborado pelo autor.

Porém, apesar da semelhança com a curva gaussiana, é possível encontrar na literatura que essas distribuições tenham uma assimetria positiva, ou seja, são levemente inclinadas

para a esquerda. Esse comportamento descrito no artigo (BORGHESI; BOUCHAUD, 2010) é justificado por que nele temos uma análise sobre o sistema eleitoral francês onde os coeficientes τ são calculados sobre as *communes*, que são divisões administrativas que não possuem equivalente nas nossas divisões, e eles demonstram uma relação inversa entre tamanho da *commune* e o valor médio de τ para elas.

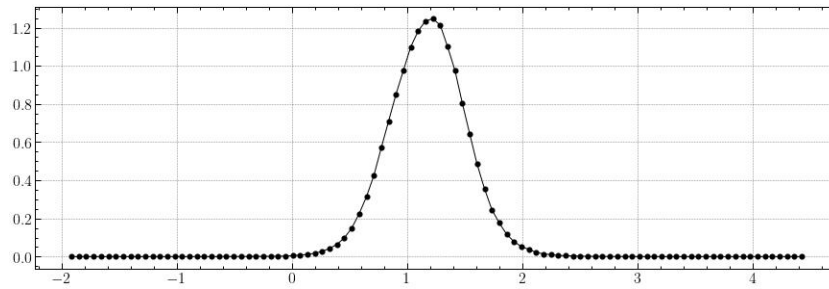
Em resumo, eram esperadas curvas gaussianas levemente assimétricas para a esquerda. De qualquer forma, modelamos as distribuições com gaussianas. Para entender o limite deste modelo, calculamos as curvas para todas as cidades, estados, regiões e para o país utilizando o método de ajuste por mínimos quadrados não-lineares, que consiste em minimizar o RMSE variando os parâmetros da curva.

4 RESULTADOS

4.1 Resultados das distribuições no brasil e nas regiões

A Fig. 8 apresenta o histograma com os votos contabilizados no segundo turno das eleições presidenciais de 2022 de todas as urnas do Brasil. Nota-se uma gaussiana quase que perfeita, com valor médio $\langle x \rangle \approx 1.17$, variância $\sigma^2 \approx 0.11$, assimetria ≈ -0.0064 , e curtose ≈ 0.87 . Lembrando que para ser uma gaussiana perfeita, ela deveria possuir 0 de assimetria e de curtose.

Figura 8 – Resultado da distribuição dos votos contabilizados no segundo turno das eleições presidenciais de 2022 de todas as urnas do Brasil. Aqui foram utilizados 100 *bins*, pois a regra de Sturges resulta em poucos pontos. No eixo horizontal temos o valor de τ das urnas e no eixo vertical a probabilidade de uma urna ter esse valor.

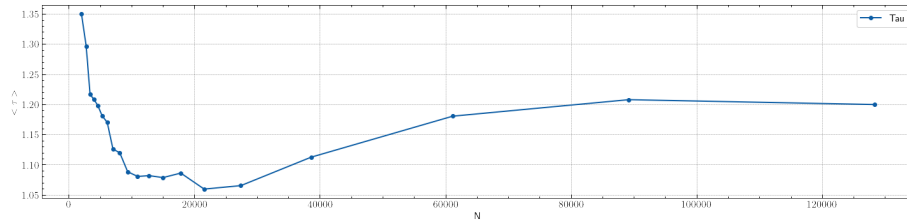


Fonte: Elaborado pelo autor.

Para entendermos o porquê desse resultado ser diferente do encontrado na França (BORGHESI; BOUCHAUD, 2010) (ver painel da esquerda da Fig. 1), associamos as *communes* com as zonas eleitorais, pois elas possuem ordens de grandeza similares. Para tal, reproduzimos o gráfico dos votos aptos nas urnas N versus $\langle \tau \rangle$, como mostra a Fig. 9, em analogia ao painel da direita da Fig. 1 no Capítulo 1.

Pela Fig. 9, percebe-se que a relação entre N e $\langle \tau \rangle$ é diferente para as nossas eleições quando comparadas as eleições francesas (BORGHESI; BOUCHAUD, 2010), pois temos um alto valor de τ para zonas com baixo número de pessoas, e esse valor diminui até alcançar um mínimo por volta de $N \approx 20000$ e depois sobe novamente. No caso francês (BORGHESI; BOUCHAUD, 2010) podemos concluir que quanto maior é a *commune* menos as pessoas têm engajamento para votar, enquanto que no presente caso das eleições brasileiras, temos que o crescimento das zonas até certo ponto também reflete em uma baixo engajamento, com o porém de que após um limiar o engajamento volta a crescer.

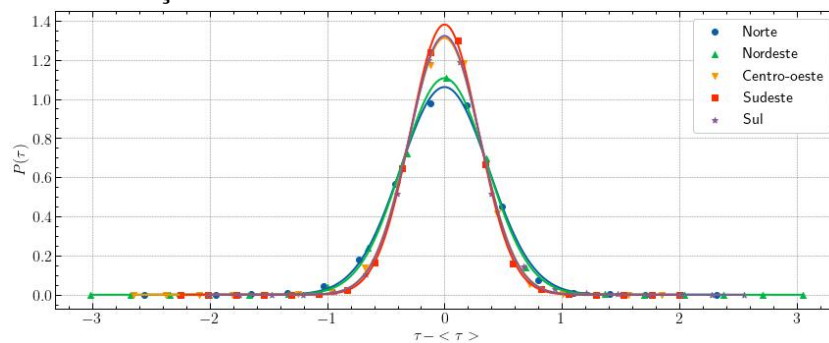
Figura 9 – Para esse gráfico ordenamos o τ de todas as zonas eleitorais das eleições presidenciais de 2022 em uma lista de ordem crescente e separando elas em vinte grupos de mesmo tamanho e então calculamos a média de tau e do número de votantes na zona. Cada ponto desse representa as médias dos grupos. O resultado diverge do encontrado na eleições francesas onde a curva é sempre decrescente



Fonte: Elaborado pelo autor.

Baseado na comparação até aqui apresentada entre os resultados das eleições brasileiras e francesas (BORGHESI; BOUCHAUD, 2010), e percebendo que elas apresentam formas diferentes de como que $\langle \tau \rangle$ varia com N , sendo as distribuições brasileiras mais normais possivelmente por estar calculando τ por urna, somos então motivados a verificar como deveria ser a distribuição de τ por zona. Essa análise está apresentada na Fig. 10. Percebe-se que as distribuições resultantes tiveram curtoses e assimetrias menores do que aquelas para as eleições francesas. É importante notar que as eleições na França não são obrigatórias, por isso a relação entre $\langle \tau \rangle$ e N encontrada na Fig. 10 seja diferente.

Figura 10 – Distribuição de τ por região apresentada de forma simétrica e obtida ao se subtrair as distribuições ao seu valor médio, facilitando assim a comparação por terem todas um centro de distribuição comum localizado no 0.



Fonte: Elaborado pelo autor.

Os valores dos RMSE, assimetrias, curtose, médias, e variâncias das distribuições são apresentados na Tabela 1 por região do território brasileiro. Pela Fig. 10 e Tabela 1, nota-se que as distribuições regionais também apresentam distribuições gaussianas. O interessante é que a variância dessas distribuições é diretamente proporcional as áreas das regiões. O único ponto em exceção é a região centro-oeste, que tem uma variância baixa apesar de ter uma área

um pouco maior do que a região nordeste.

Tabela 1 – Valores dos RMSE, assimetrias, curtose, médias, e variâncias das distribuições para as diferentes regiões brasileiras.

	RMSE	Assimetria	Curtose	Média	Variância
Norte	0.016	-0.34	1.2	1.1	0.38
Nordeste	0.013	-0.10	0.50	1.2	0.36
Centro-oeste	0.020	-0.17	0.69	1.2	0.30
Sudeste	0.0081	-0.021	0.63	1.1	0.29
Sul	0.013	0.30	1.1	1.3	0.30

4.2 Resultados por estado

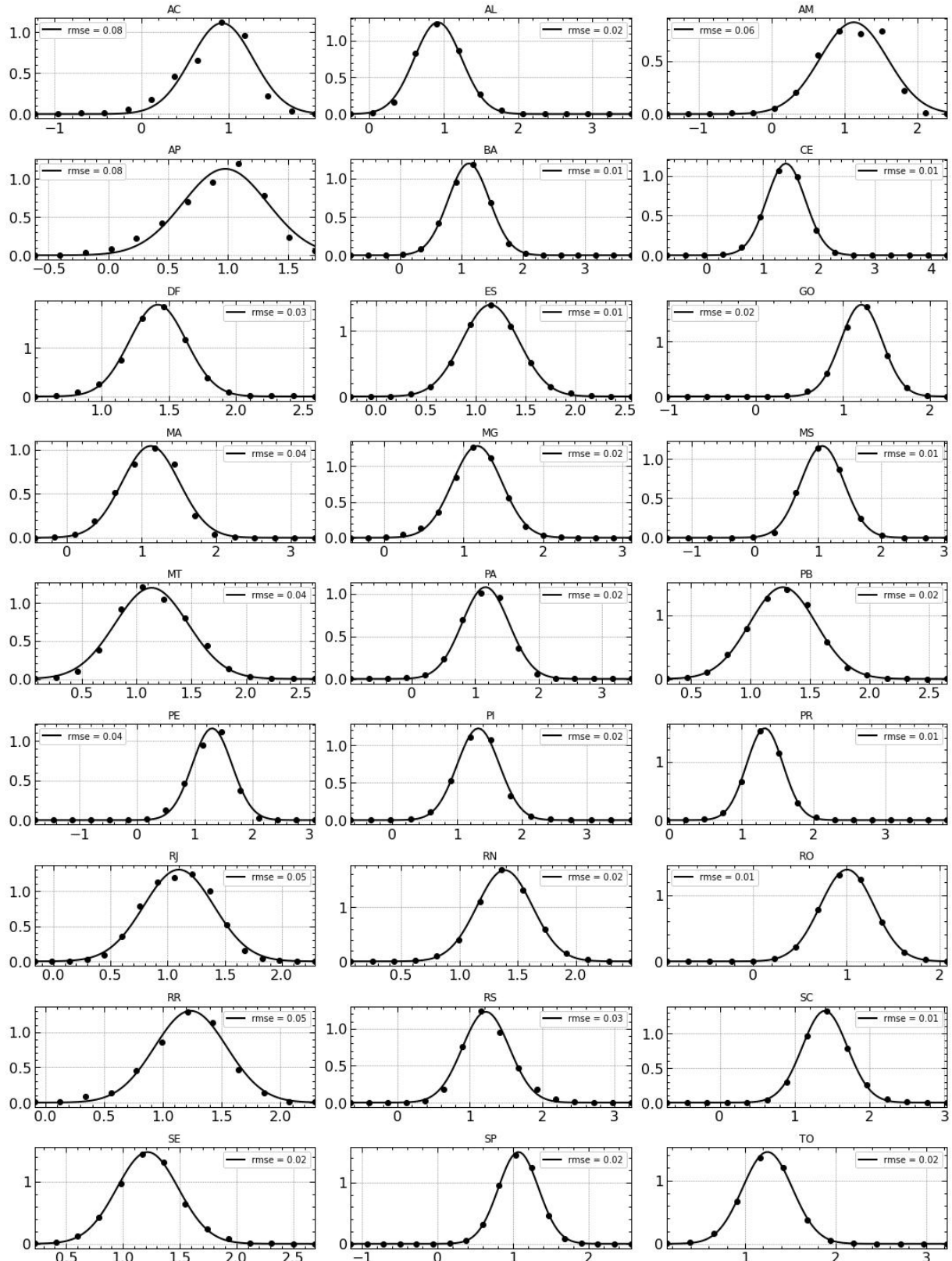
Na Fig. 11 apresentamos os resultados das distribuições por estado do coeficiente τ relativas a votação do segundo turno para o cargo de presidente no ano de 2022. Da Fig. 11 é possível ver que três estados destoam dos demais por terem o erro quadrático médio mais elevado, são eles: Acre, Amazonas e Amapá. Isto é, analisando as curvas de ajustes gaussianas e os pontos dos dados extraídos, percebe-se que os pontos dos dados para esses três estados não são bem *fitados* pela curva gaussiana. Essa diferença para os demais estados pode estar relacionada com o fato dessas regiões estarem no *rank* dos 5 estados com menor densidade demográfica.

4.3 Resultados por cidades

Resultados análogos ao da Seção 4.2 mostrados na Fig. 11 para as distribuições do parâmetro τ para os estados são agora discutidos para as 27 capitais brasileiras na Fig. 12. É importante ter em mente que as capitais não necessariamente são as 27 cidades mais populosas do país, entretanto, visando uma distribuição uniforme das várias regiões no país e sabendo que todas elas são suficientemente grandes para apresentarem claramente distribuições bem expressivas, decidimos fazer tal análise por capital. Caso escolhêssemos as cidades mais populosas, teríamos uma aglomeração de 7 cidades no estado de São Paulo, o que poderia levar a um comportamento não-universal do país inteiro.

Da Fig. 12, verifica-se que as curvas de ajuste gaussianas para as cidades do Rio de Janeiro e Porto Velho são as que mais se distanciam dos pontos extraídos das urnas, apresentando valores de RMSE's uma ordem de grandeza maior do que as das outras capitais. As curvas de Porto Velho e do Macapá têm uma assimetria leve à direita. No artigo (KLIMEK *et al.*, 2012),

Figura 11 – Distribuições dos coeficientes τ associadas a votação do segundo turno presidencial para as urnas dos 26 estados mais o distrito federal do território brasileiro no ano de 2022.

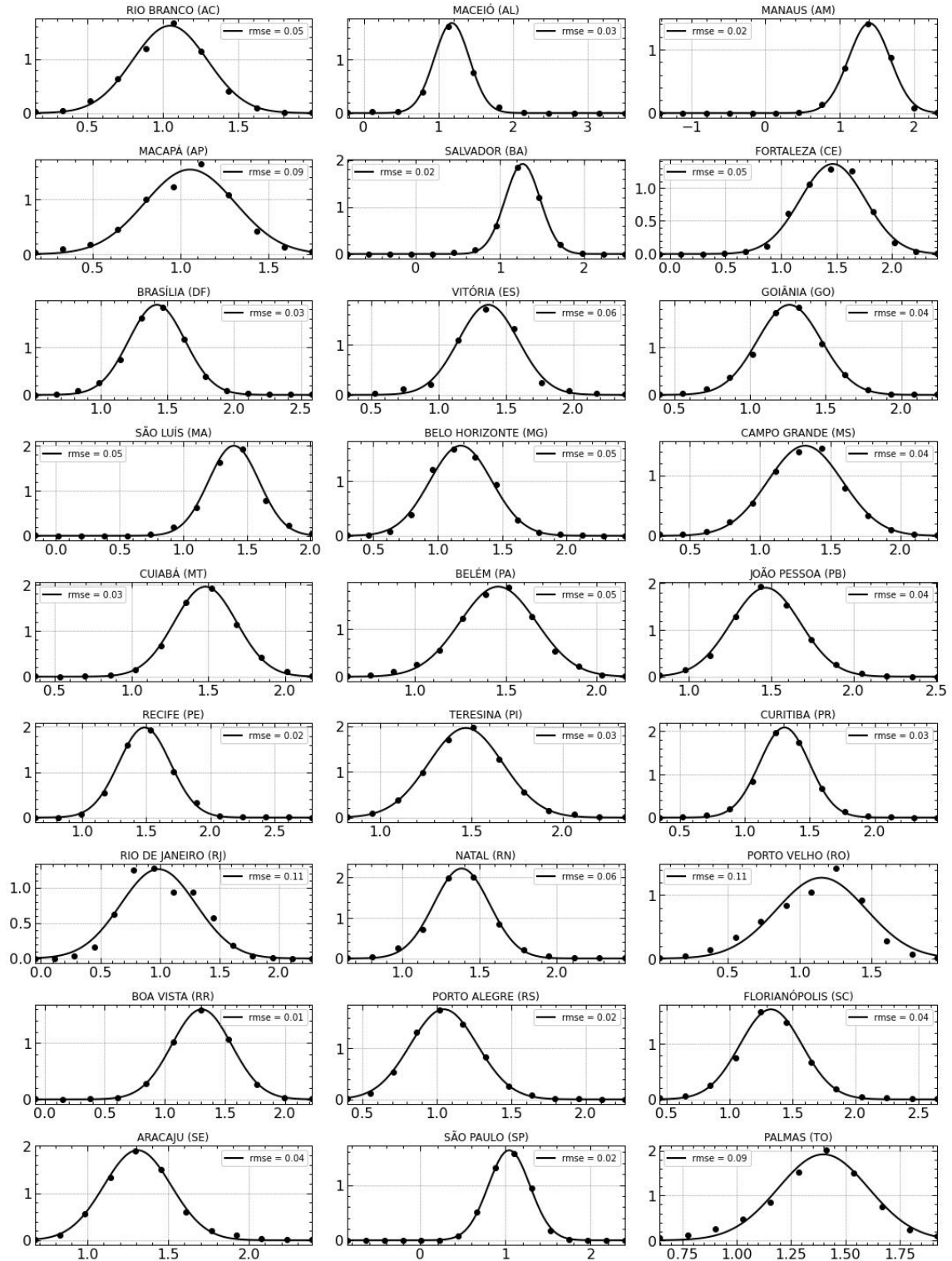


Fonte: Elaborado pelo autor.

é mostrado que eleições com falhas tem assimetrias similares, porém a ordem de grandeza da assimetria dessas curvas não é comparável às encontradas nesses estados. Provavelmente essa

assimetria é apenas fruto do acaso.

Figura 12 – Distribuições dos coeficientes τ associadas a votação do segundo turno presidencial para as urnas das 27 capitais brasileiras no ano de 2022.



Fonte: Elaborado pelo autor.

4.4 Correlações de dados

Para entender quais fatores podem influenciar no parâmetro τ e na distribuição que ele gera, representamos graficamente o valor de τ em função da população dos municípios e das densidades populacionais e calculamos as médias, variâncias e RMSE's. Esses gráficos são apresentados na Fig. 13. A maioria desses dados cruzados parece não ter nenhuma correlação a priori, com exceção da densidade contra população que aparenta ter uma relação linear em um *plote* $\log \times \log$ e do RMSE que decresce quanto maior é a população, mostrando que de fato quanto maior o N amostral mais a curva se ajusta à curva teórica.

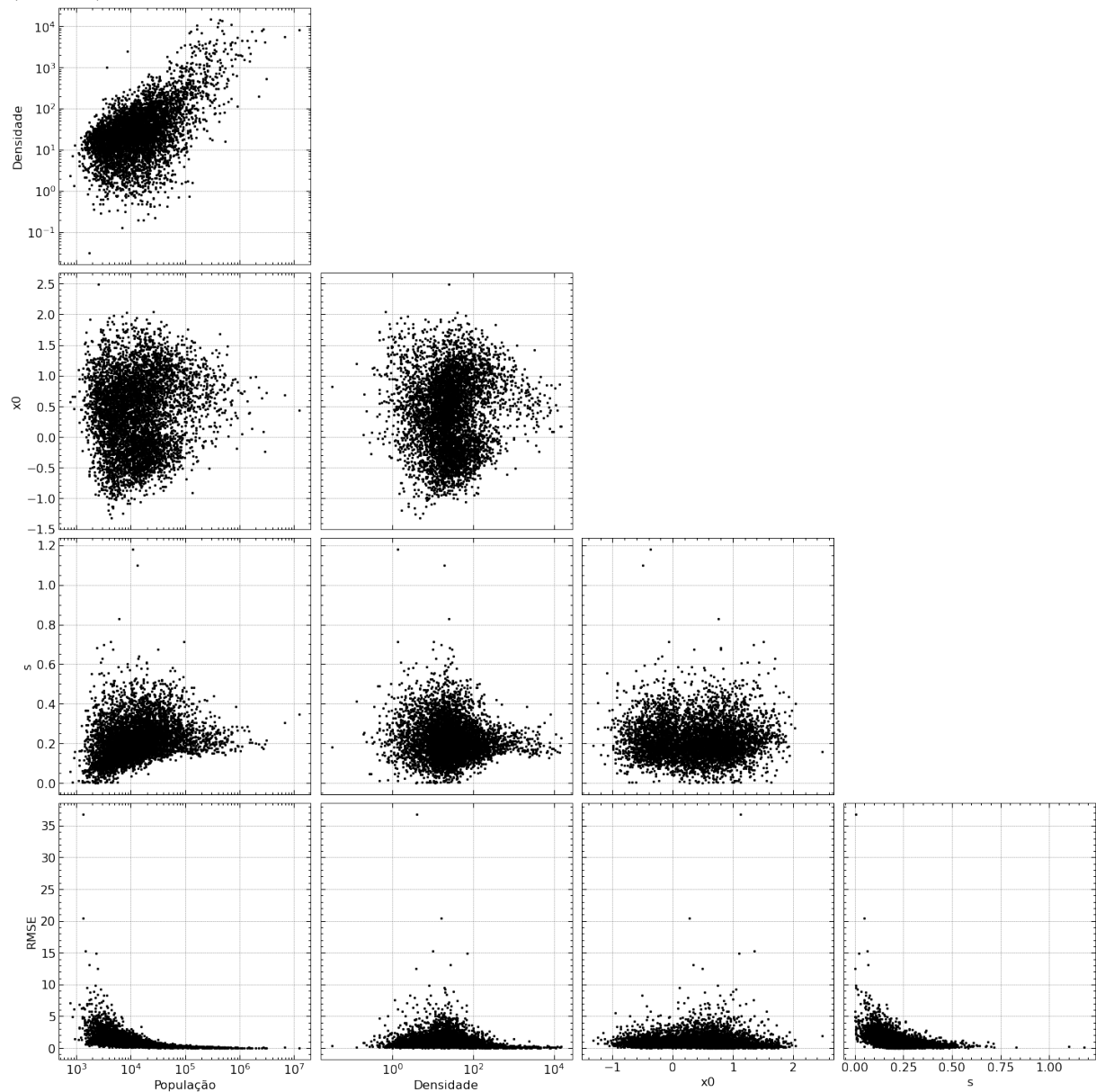
4.5 Interferência da PRF nas eleições de 2022

No dia 9 de agosto de 2023 foi decretado pelo Supremo Tribunal Federal (STF) a prisão do ex-diretor da polícia rodoviária federal (PRF), Silvinei Vasques, por possíveis tentativas de dificultar a locomoção por estradas nas cidades onde o candidato a presidência Luiz Inácio Lula da Silva teve mais votantes no primeiro turno (JORDÃO, 2022). A investigação que levou a essa prisão partiu de um servidor público que trabalhava na Diretoria de Inteligência do Ministério Justiça ao receber um pedido não usual onde foi requisitado a ele um relatório onde os candidatos do segundo turno tiveram mais que 75% dos votos no primeiro 14. Segundo a polícia federal, que investiga o caso, uma das imagens desse relatório foi provavelmente apresentada ao então ministro da Justiça, Anderson Torres, que juntamente com Silvinei utilizou a listagem contida nas imagens para basear um novo plano de trabalho da PRF, que sobrepunha o antigo lançado antes do primeiro turno. Ver imagem na Fig. 14 com o mapa das cidades. (SERRA, 2023).

No dia anterior as eleições um dos ministros do STF, Alexandre de Moraes, coibiu as ações em ordem judicial devido ao número anormal de *blitzes* que estavam planejadas para acontecer no nordeste (SERAPIÃO, 2023). No dia seguinte houve descumprimento da ordem, contudo à tarde as *blitze* cessaram e o ministro depois afirmou que nenhum eleitor foi impedido de votar. Mas será que a medida foi de fato eficaz? Podemos analisar a situação utilizando o coeficiente τ para avaliar se de fato as *blitze* impactaram negativamente ou não no comparecimento dos eleitores do nordeste.

Na Fig. 15 podemos ver como a média dos τ 's por município variou do segundo para o primeiro turno no nordeste nas eleições presidências de 2012, 2014 e 2022. Nessa figura

Figura 13 – Relações entre os vários dados extraídos dos vários municípios brasileiros: densidade populacional, média da gaussiana (x_0), variância do gaussiana (s) e o erro quadrático médio (RMSE).



Fonte: Elaborado pelo autor.

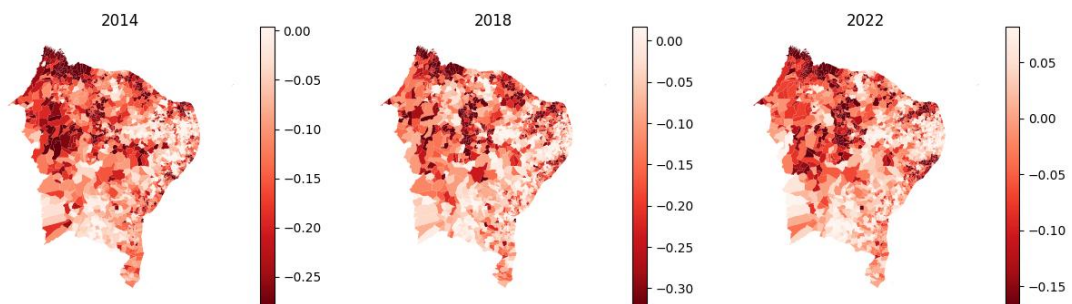
observamos que os mapas contém variações, mas a aparência deles não muda significativamente, as regiões em que o número de votantes cai bastante continuaram sendo as mesmas e as regiões em que o número pouco variou também são equivalentes. É importante observar que no último ano (2022) a variação do τ médio foi positiva no nordeste, por isso o intervalo dos τ no mapa mudou para esse ano, o que implica no fato das pessoas estarem mais engajadas para votar no segundo turno de 2022 do que elas normalmente engajam. Mas como é possível verificar se os valores de τ não poderiam ser ainda maiores ou mantidos abaixo do que poderiam ser pelas *blitze*?

Figura 14 – Imagem tirada pelo celular de Marília de Alencar, então diretora de Inteligência do Ministério da Justiça momentos após o começo da reunião com o ex-ministro Anderson Torres (SERRA, 2023). Na imagem podemos ver pelo título que há um filtro para as cidades com mais de 75% de votantes no Lula, com a lista ordenada pelas cidades com a maior quantidade de votantes absolutos no Lula.



Fonte: (CAPITAL, 2023)

Figura 15 – Mapa da variação do τ médio do segundo para o primeiro turno nas eleições federais de 2014, 2018 e 2022. Note que cada mapa tem um intervalo diferente que foi definido pelo maior valor dos 10% menores valores e pelo menor valor dos 10% maiores.

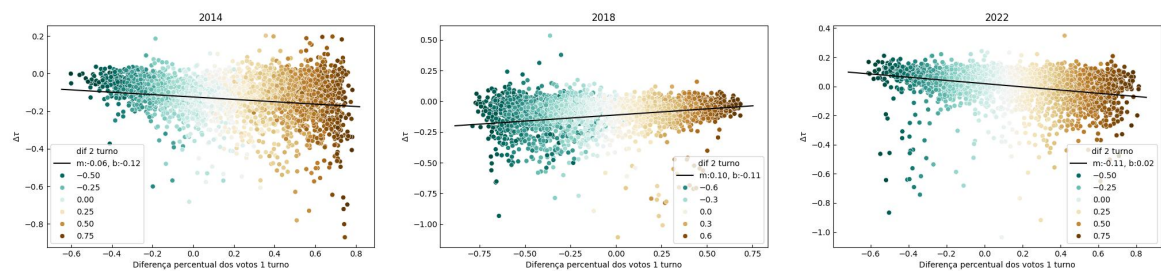


Fonte: Elaborado pelo autor.

Para fazer a verificação da possibilidade do τ ser maior, podemos observar o comportamento do τ médio dos municípios para os anos disponíveis em busca de um padrão. O resultado pode ser visto na Fig. 16. É encontrada uma relação linear entre a variação do τ e a diferença do percentual de votos no primeiro turno dos candidatos que participaram do segundo turno. Podemos ver que o padrão se mantém constante para todos os anos, levando em conta que em 2018 o candidato vencedor era orientado à direita no espectro político. Isso nos leva a concluir que as cidades que votam mais alinhadas com os candidatos de direita perdem menos

votantes no segundo turno do que os candidatos alinhados à esquerda. O ano de 2022 não foi exceção. Caso a tentativa de impedir os eleitores de chegarem nos locais de votação fossem sucedidas seria possível ver uma queda nos pontos mais à esquerda. Como isso não ocorreu, é plausível concluir que de fato as *blitzes* não afetaram significativamente no comparecimento das pessoas nas urnas.

Figura 16 – Variação de τ do segundo para o primeiro turno por município. No eixo horizontal temos a diferença percentual dos votos no primeiro turno do candidato vencedor no segundo turno para o perdedor do segundo turno nos municípios. As cores indicam essa diferença para o segundo turno. Para minimizar o erro os gráficos contêm apenas as cidades com população maior que 10 mil habitantes. É bem evidente que existem uma relação linear entre as variáveis, por isso a regressão linear desses pontos está presente nesses gráficos. Note que quanto mais uma cidade vota no candidato orientado à esquerda menor é o índice τ , o que indica um menor engajamento dos habitantes da cidade, e quanto mais a cidade vota no candidato à direita do espectro político maior tende a ser o índice τ .



Fonte: Elaborado pelo autor.

5 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho entendemos um pouco mais sobre a física social, uma ciência antiga mas que está ganhando destaque em anos recente, que aplica os métodos de análise da física à sistemas sociais. Comentamos um pouco da sua história e discutimos como esses modelos podem mostrar como a nossa sociedade funciona. Também discutimos sobre o sistema eleitoral brasileiro e como descrever os resultados através de modelos físico-estatísticos pode ser útil.

Em seguida, vimos um pouco das ferramentas de análise estatística que permitem o estudo dos sistemas sociais e discutimos como foi feita a extração de dados.

Uma análise sobre o coeficiente τ foi realizada para distintas divisões administrativas no nosso país e mostramos que, consistentemente, as distribuições desse coeficiente resultam em gaussianas. Pesquisas anteriores mostram que a variância dessas distribuições são independentes do ano de eleição, diferente do valor médio. Então, seria interessante comparar os valores aqui encontrados com os valores da eleições passadas em trabalhos posteriores.

Como mostrado no artigo (KLIMEK *et al.*, 2012), um forte indício de fraude seria um alto desvio do formato de gaussianas. Aqui encontramos o contrário, as curvas se ajustam a gaussianas melhor do que as curvas encontradas para outros países. Possíveis motivos para que a nossa curva seja diferente são as correlações do número de aptos a votar N e τ que temos, e o fato do voto ser obrigatório no Brasil.

Além disso uma breve análise sobre as variações dos coeficientes mostrou que as *blitze* realizadas no dia do segundo das eleições presidenciais de 2022 não impactaram significativamente no comparecimento dos votantes.

REFERÊNCIAS

- ANDRADE, A. L. M. S. de. **Fraude eleitoral**. 2023. Disponível em: <https://www.infoescola.com/direito/fraude-eleitoral/>. Acesso em: dez 2023.
- BALL, P. **Critical Mass**. [S. l.]: Arrow Books, 2004.
- BBC. **Qual era a melhor forma de governo para Platão, que fazia duras críticas à democracia**. 2021. Disponível em: <https://www.bbc.com/portuguese/internacional-56004586>. Acesso em: dez 2023.
- BORGHESI, C.; BOUCHAUD, J.-P. Spatial correlations in vote statistics: a diffusive field model for decision-making. **The European Physical Journal B**, Springer, v. 75, p. 395–404, 2010.
- CAPITAL, C. **Reunião sobre blitze da PRF no 2º turno usou mapa com votos de Lula**. 2023. Disponível em: <https://www.cartacapital.com.br/politica/reuniao-sobre-blitze-da-prf-no-2o-turno-usou-mapa-com-votos-de-lula-veja-imagens/>. Acesso em: dez 2023.
- CASTELLANO, C.; FORTUNATO, S.; LORETO, V. Statistical physics of social dynamics. **Reviews of modern physics**, APS, v. 81, n. 2, p. 591, 2009.
- FILHO, R. C.; ALMEIDA, M.; ANDRADE, J.; MOREIRA, J. *et al.* Scaling behavior in a proportional voting process. **Physical Review E**, APS, v. 60, n. 1, p. 1067, 1999.
- FLAMM, D. History and outlook of statistical physics. **arXiv preprint physics/9803005**, 1998.
- IBGE. **Censo 2022**. 2022. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=resultados>. Acesso em: jul 2023.
- JORDÃO, P. **Entenda a prisão do Silvinei Vasques, ex-diretor-geral da PRF**. 2022. Disponível em: <https://www.cnnbrasil.com.br/politica/entenda-a-prisao-do-silvinei-vasques-ex-diretor-geral-da-prf/>. Acesso em: nov 2023.
- KLIMEK, P.; YEGOROV, Y.; HANEL, R.; THURNER, S. Statistical detection of systematic election irregularities. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 109, n. 41, p. 16469–16473, 2012.
- KLISIEWICZ, E. L. **Simpático ou oportunista? De onde vem a fixação dos políticos pelo pastel de feira?** 2022. Disponível em: <https://www.tribunapr.com.br/eleicoes/2022/simpatico-ou-oportunista-de-onde-vem-a-fixacao-dos-politicos-pelo-pastel-de-feira/>. Acesso em: dez 2023.
- PREVIDELLI, F. **BUSH X AL GORE: EM 2000, ELEIÇÕES DOS ESTADOS UNIDOS TIVERAM RECONTAGEM DOS VOTOS**. 2020. Disponível em: <https://aventurasnahistoria.uol.com.br/noticias/reportagem/bush-x-al-gore-uma-das-eleicoes-norte-americanas-mais-controversias-da-historia.phtml>. Acesso em: dez 2023.
- SANDERSON, G. **But what is the Central Limit Theorem?** 2023. Disponível em: <https://www.youtube.com/watch?v=zeJD6dqJ5lo>. Acesso em: dez 2023.

SANDOVAL, P. X. D. **“Parem de contar os votos!”: os ecos de Bush x Gore e o desastre eleitoral da Flórida em 2000.** 2020.

Disponível em: <https://brasil.elpais.com/internacional/2020-11-06/parem-de-contar-os-votos-os-ecos-de-bush-x-gore-e-o-desastre-eleitoral-da-florida-em-2000.html>. Acesso em: dez 2023.

SE Tribunal Regional Eleitoral. **Além do Brasil, outros 46 países utilizam urnas eletrônicas nas eleições.** 2022. Disponível em: <https://www.tre-se.jus.br/comunicacao/noticias/2021/Maio/alem-do-brasil-outros-46-paises-utilizam-urnas-eletronicas-nas-eleicoes#:~:text=eletr%C3%B4nicas%20nas%20elei%C3%A7%C3%B5es-,Publica%C3%A7%C3%A3o%20que%20vem%20circulando%20nas%20redes%20sociais%20afirma%20que%20somente,A%20informa%C3%A7%C3%A3o%20%C3%A9%20falsa!> Acesso em: dez 2023.

SERAPIÃO, F. **Moraes proíbe operações da PRF que envolvam transporte público no dia da eleição.** 2023. Disponível em: <https://www1.folha.uol.com.br/poder/2022/10/moraes-proibe-operacoes-da-pf-e-prf-que-envolvam-transporte-publico-no-dia-da-eleicao.shtml>. Acesso em: dez 2023.

SERRA, M. M. P. **Mapa com concentração de eleitores de Lula foi utilizado em reunião sobre blitzes da PRF.** 2023. Disponível em: <https://oglobo.globo.com/politica/noticia/2023/08/09/mapa-com-concentracao-de-eleitores-de-lula-foi-utilizada-em-reuniao-sobre-blitze-da-prf-veja.ghml>. Acesso em: dez 2023.

STURGES, H. A. The choice of a class interval. **Journal of the american statistical association**, New York, v. 21, n. 153, p. 65–66, 1926.

TRIBUNAL SUPERIOR ELEITORAL. **Eleições majoritárias e proporcionais.** 2022. Disponível em: <https://www.tre-sc.jus.br/eleicoes/tire-suas-duvidas/eleicoes-majoritarias-e-proporcionais>. Acesso em: dez 2023.

TRIBUNAL SUPERIOR ELEITORAL. **Eleições pelo Mundo: obrigatoriedade do voto e uso de urnas eletrônicas.** 2022. Disponível em: <https://www.tse.jus.br/comunicacao/noticias/2021/Setembro/serie-201celeicoes-pelo-mundo201d-traz-historico-sobre-obrigatoriedade-do-voto-e-uso-de-urnas-eletronica>. Acesso em: dez 2023.

TRIBUNAL SUPERIOR ELEITORAL. **Portal dados abertos.** 2022. Disponível em: <https://dadosabertos.tse.jus.br/dataset>. Acesso em: dez 2022.

TRIBUNAL SUPERIOR ELEITORAL. **As fraudes no processo de votação manual.** 2023. Disponível em: <https://www.justicaeleitoral.jus.br/urna-eletronica/historico-das-fraudes-nas-eleicoes.html>. Acesso em: dez 2023.