

## Eleições 2022

Análise estatística para possíveis diferenças entre urnas modelo anterior a 2020 contra urnas 2020 em 6 estados da federação.

## Autor

Diante do momento de grande tumulto de informações e da produção de muitos resultados referentes ao momento políticos este autor buscou avaliar os resultados das eleições de 2022 à luz do conhecimento estatístico.

Oceanógrafo, com mestrado na área de ecologia e experiência em estatística, tem noção e concordância sobre certos resultados referentes a mudanças climáticas entre outras áreas ambientais, tão exaltadas pela esquerda.

Apesar de sua formação, este autor insiste em tentar ajudar com sua pequena contribuição avaliando os resultados expressos nos expostos por Fernando Cerimedo (lives censuradas, BrazilWasStolen/La Derecha Diário) e do vídeo editado e explicativo do Nataniel Antonioli (Desmistificando fraude nas eleições, canal Fábrica de Nerds).

Por medo da perseguição estatal e medo de prisão sem defesa, este autor opta por identidade anônima até que se faça necessário representar a favor dos resultados aqui apresentados com segurança.

Allejo, 2022.

## Apresentação

Este trabalho é uma investigação estatística que utiliza planilhas de dados referentes ao segundo turno das eleições brasileiras de 2022 (dados originais anteriores ao dia 4 de novembro), onde se comparam urnas novas de 2020 contra todas as outras consideradas de modelos mais antigos e similares.

Perante tudo que está ocorrendo no Brasil:

Os movimentos políticos orgânicos recentes;

Enfraquecimento das lideranças políticas na câmara e senado;

A falta de isonomia no processo eleitoral com apoio do TSE, prática do ativismo judicial recente, prática de censura, ao partidarismo declarado dos órgãos superiores do legislativo de forma unilateral;

Ao partidarismo da mídia bi-polar, possivelmente, pela geração da convulsão social, da polarização, consequentemente, aumento em sua audiência;

**Mas principalmente, devido ao clamor dos movimentos pacíficos sociais dos brasileiros a considerar o resultado conturbado e das últimas eleições.**

Apresento este manifesto de forma didática com finalidade de evitar a censura, auxiliar os setores técnicos que estão buscando avaliar os resultados das eleições e de reunir profissionais da área de análise de dados, estatística e tecnologia de informação para averiguar os resultados eleitorais **sem viés político partidário ou ideológico** (inclusive na busca de eliminar o viés do autor).

Busca-se, principalmente, sanar **ou** revelar a possibilidade de manipulação dos resultados destas eleições e de revelar também a possibilidade do ímpeto de manipulação da sociedade (originado de direita ou esquerda), haja visto, que todos os sistemas possuem vulnerabilidade e que o exercício da liberdade engloba o direito de se expressar e questionar o estado e as imposições político-ideológicas, afim de contribuir, incluindo concordantes e discordantes destes resultados.

Acrescento que,

Este documento busca principalmente a libertação de uma possível situação ilegítima, tal qual prejudicou, prejudica ou prejudicará a todos, independente de classes sociais, das diferenças individuais, das crenças, dos aspectos físicos ou pensamentos políticos bem ou mal intencionados, em outras palavras, **o objetivo deste trabalho é auxiliar na identificação, ou não, de possível ato corrompido que pode ter influenciado os resultados eleitorais para um projeto de poder pessoal em detrimento de um projeto de país.** Busca-se, também, não apresentar opinião individual.

Convido profissionais idôneos, livres de influência de partidos políticos, que não tenham relações próximas com personas políticas ou personalidades públicas, que tenham capacidade de deixar seus corações de lado e vistam-se de consciência plena pela razão para contribuir com opiniões sobre os resultados presentes neste documento.

Por fim, obrigado por **nos-ler** até aqui, agora, muito provavelmente você também faz parte deste manifesto.

*Enquanto houverem pessoas  
esclarecidos dialogando, a sociedade  
não se lançará a sorte.*

## **Índice**

### **O que é a estatística**

População e amostra: estatística descritiva vs. inferencial

Ciência determinística e probabilística

Medidas

Definição clássica de probabilidade de Laplace

Hipótese

Método

Poder de comparação de médias

Amostragens

Redução de variabilidade (retirada de capitais)

### **Metodologia específica**

Fonte de dados

Verificação posterior à eleição

### **Tratamentos**

Definição axiomática de Kolmogorov

Kolmogorov-Smirnov & Lilliefors

Gráfico de intervalos de confiança

Teste de Kruskal Wallis.

Regressão linear de Spearman

### **Resultados**

Gráficos de distribuição para modelos de urnas, Estados analisados

Gráfico de intervalos de confiança (95%) e teste de Kruskal-Wallis

Gráfico de nuvem de pontos e correlação de Spearman

### **Conclusões**

## O que é a estatística

A estatística é uma ciência matemática que estuda dados numéricos ou categóricos, sua organização, análise e comparação na busca de resultados que se aproximam, geralmente 95 ou 99%, da realidade da qual esses dados foram amostrados.

São dados numéricos: conjunto dos números naturais.

São dados categóricos, divididos em ordinais e nominais: conjunto de números naturais para fins classificatórios, itens em ordens por razão de amostra são dados categóricos ordinais; cor, nota musical e modelo de equipamento eletrônico são dados categóricos nominais.

Através da estatística podemos verificar características essenciais verdadeiras inerentes a grupos de dados provenientes do meio físico, digital, literário, ou qualquer universo que permita a realização de medidas, bem como, realizar projeções, extrair resultados, calcular probabilidades diversas, identificar influências nas amostras, etc, destes universos.

É necessária muita responsabilidade durante a aplicação dos testes estatísticos para que não se aceitem hipóteses falsas ou nulas pela má conduta durante a realização dos processos de preparação, organização, atendimento a premissas e da própria escolha dos testes estatísticos adequados. É também necessário muita responsabilidade e honestidade para desenvolver hipóteses que identifiquem um resultado sem manipulação de premissas argumentativas.

The Epistemic Integrity of Scientific Research, Jan De Winter & Laszlo Kosollosky, Science and Engineering Ethics volume 19, pages 757–774 (2013).

Correction of errors in scientific research, Russell M. Church, Jonathon D. Crystal & Charles E. Collyer, Behavior Research Methods, Instruments, & Computers volume 28, pages 305–310 (1996).

### População e amostra: estatística descritiva vs. inferencial

Uma distinção fundamental da estatística é aquela entre população e amostra. A população é o objeto integral de uma análise, a amostra é parte aleatória deste objeto. Dentro destas determinações, um estudo observacional pode querer descrever as preferências eleitorais (Amostra) de todos os eleitores de um país (objeto integral). Já um estudo experimental pode querer compreender se homens entre 35-50 anos (Amostra) são mais sujeitos aos efeitos colaterais de uma droga que mulheres (Amostra) nessa idade (seres humanos como objeto integral). Assim, como definição adicional, os eleitores e os homens e mulheres na referida faixa etária são a população de cada pesquisa.

É evidente que não é possível, ou ao menos muito caro e ineficiente, perguntar a preferência de todos os milhões de eleitores de um país, ou realizar experimento com todos os homens e mulheres daquela idade no globo para chegar às conclusões desejadas. Porém, o pesquisador pode desejar avaliar uma amostra, ou seja, uma parcela significativa daquela população que pode ser diretamente observada.

A estatística descritiva é ramo da estatística que permite uma descrição sistemática e precisa dos dados relativos primordialmente a uma população. Os conceitos utilizados na estatística descritiva servem também de base para a estatística inferencial.

A estatística inferencial é o ramo da estatística que descreve os procedimentos de pesquisa e as regras matemáticas que permitirão a consideração de uma amostra como seguramente representativa da população, assim como uma medida da robustez dessa segurança, com análise de conceitos como margem de erro e confiabilidade.

A estatística também pode ser determinística ou probabilística, e realiza a verificação de uma ou mais hipóteses pré-concebidas a depender de critérios amostrais como quantidades de amostras, características das amostragens e poder dos testes corretamente aplicados

## Ciência determinística e probabilística

A estatística determinística se propõe a avaliar dados totais de um dado conjunto amostral em um universo limitado e completo e apresenta resultados que podem ser replicados com precisão conhecida.

A teoria da probabilidade, se propõe a avaliar problemas associados aos fenômenos não determinísticos (aleatórios) para o desenvolvimento e compreensão dos métodos estatísticos, sobretudo no que se refere à Estatística inferencial. Isso acontece porque as conclusões obtidas nos processos inferenciais são baseadas em dados aleatoriamente escolhidos, consequentemente, sempre admitem determinada margem de incerteza (5 a 1%, aceitável ou até maior nível de precisão). Por isso, a teoria probabilística se constitui no base da estatística inferencial.

## Medidas

Todas as medidas dependem de determinados parâmetros, que são sempre constantes e estão relacionados a características de uma população representada por meio de uma variável aleatória. Esta, por sua vez, constitui suporte para os modelos teóricos que representam, com aceitável grau de aproximação, inúmeras situações de problemas reais, presentes no nosso cotidiano.

Além de médias, medianas, moda que são medidas de tendência central (estatística descritiva), parâmetros como amplitude, variância, desvio padrão coeficiente de variação, medidas de dispersão, (estatística inferencial) requerem tratamento das amostras analisando grupo similares entre eles, para que não haja distorções populacionais que influenciem possíveis resultados.

**No caso dos arquivos utilizados para verificar os resultados das eleições 2022, deve-se separar capitais de cidades do interior de cada estado, assim, diferencia-se a comparação de cidades do interior entre elas para cada estado das capitais que podem ser comparadas entre elas para o país inteiro.**

## Definição clássica de probabilidade de Laplace

Laplace define probabilidade considerando um espaço finito de possibilidades que pode ser chamado de  $\Omega$  onde existem  $n$  elementos (número de casos possíveis) e todos os resultados são equiprováveis (mesma probabilidade de ocorrência). Diante disso, um evento  $A$  qualquer ocorrendo dentro deste espaço finito chamado de  $\Omega$ , então, a probabilidade de  $A$  é definida pela razão entre o número de casos favoráveis à ocorrência de  $A$  dividido pelo número de casos possíveis.

$$P(A) = \frac{n^{\circ} \text{ de casos favoráveis à ocorrência de } A}{n^{\circ} \text{ de casos possíveis}} = \frac{\#(A)}{\#(\Omega)}$$

## Hipótese

A busca pelo conhecimento requer, em primeiro lugar, a formulação de ideias e a adequação de procedimentos para investigação de fatos mensuráveis. Essas ideias são conhecidas como hipóteses e necessitam de confirmação por métodos amostrais que sejam seguros e previamente conhecidos. Após a realização de amostragens, as informações são organizadas (tratadas) e submetidas a teste estatísticos para verificação de hipóteses propostas.

Além de testes de estimativas, existem testes de significância e de aderência que podem ser utilizados para verificar certas afirmações sobre determinado parâmetro desconhecido. No geral, as hipóteses podem ser duas ou mais, e se apresentam como nulas ( $H_0$ ) ou alternativas ( $H_1, H_2, \dots$ ). Os testes de significância têm como objetivo principal rejeitar essa hipótese (ideia) ou não rejeitar a hipótese inicial feita sobre certo parâmetro. Também se aplicam testes de hipóteses, com mais precisão, quando pretendemos verificar se os dados amostrais analisados seguem determinado modelo de distribuição, neste caso os testes são chamados de testes de aderência.

Para explicar melhor é necessário um exemplo prático de possíveis **urnas eletrônicas desonestas** que, tomando por base um resultado eleitoral geral 50%/50%, ao ser verificada 100000 urnas diferentes começa a apresentar um comportamento estranho repetindo muitas vezes um mesmo candidato por modelo fabricado específico analisado. A partir disto surge a dúvida e necessidade de verificação de que estas urnas podem estar viciadas, isto é a formulação da hipótese e da aplicação de um teste para verificar a possibilidade de as urnas estarem equilibradas ( $H_0$ ) ou viciadas ( $H_1$ ). Ou seja, é necessário aplicar um teste para comprovar se a proporção de “candidato X” é igual a de “Candidato Y”.

De forma a exemplificar uma situação em que a hipótese nula ( $H_0$ ) a ser testada é a afirmação de que a proporção de urnas em uma eleição 50%/50%  $p$ , é  $p = 0,5$ , equivalendo a 50% de probabilidade. Em contraponto, se ocorrem indícios que apontam para uma proporção amostral  $p \neq 0,5$ . Logo, estamos diante de um dos problemas de teste de hipóteses para  $p$ . Assim foram utilizados testes paramétricos para preparação dos dados e comparação de diferenças entre as medidas de tendência central para verificar possíveis influências externas no pleito eleitoral 2022.

No exemplo citado das urnas desonestas, a hipótese referente a  $p$  deve ser testada. Caso não haja problema então devesse verificar que  $p = 0,5$ , esta será referida como hipótese nula ( $H_0$ ) e será a hipótese testada, sendo assim  $\theta = \theta_0$ , ou seja,  $H_0 : p = 0,5$ , 50% de ocorrência como uma premissa provisória. No entanto, caso não haja confirmação dessa premissa chamamos de hipótese alternativa ( $H_1$ ),  $\theta \neq \theta_0$ , ou seja  $H_0 : p < 0,5$ . Assim, as possíveis formulações de  $H_0$  e  $H_1$  para os testes paramétricos de hipóteses são as seguintes:

- a)  $H_0 : \theta = \theta_0$   
 $H_1 : \theta \neq \theta_0$
- b)  $H_0 : \theta = \theta_0$   
 $H_1 : \theta < \theta_0$
- c)  $H_0 : \theta = \theta_0$   
 $H_1 : \theta > \theta_0$

Para este exemplo, todos os testes  $H_0$  é sempre  $H_0 : \theta = \theta_0$ , enquanto  $H_1$  é formulado com símbolos distintos:  $\neq$  ou  $<$  ou  $>$ . Isso acontece por que  $H_1$  representa uma negação de  $H_0$ , apoiada em algum indício ou alguma suspeita do pesquisador. Caso ocorra essas situações alternativas, é confirmada a que as urnas estão apresentando resultados desproporcionais e desonestos. Cada teste apresentará premissas e formas de resultados diferentes para  $p$ , bem como, para aceitar ou rejeitar sua  $H_0$ .

O nível de significância de um teste ( $\alpha$ ) está associado à possibilidade de rejeição de  $H_0$ , e é a partir dele que nós demarcamos, no gráfico de distribuição o que chamamos de região de rejeição de  $H_0$ , RR, ou região crítica, RC, ou verificamos na tabela para o valor resultante Z, são exemplos de métodos diferentes direcionados a cada tipo de teste, que variam bastante.

Também existem possibilidades aceitáveis em que se pode assumir um erro ao aceitar uma hipótese nula o que é designado por  $\beta$ . Geralmente, testes são realizados estabelecendo-se previamente o valor de  $\alpha$ , sem mencionar o valor de  $\beta$ . **Neste caso, o cálculo associado a probabilidade desse erro ocorre quando  $H_0$  é falsa, e existem infinitas possibilidades de  $H_0$  ser falsa (Urnas viciadas serem atestadas como legítimas) que é  $H_0 : \theta = \theta_0$ , enquanto há só uma possibilidade dela ser verdadeira (Urnas viciadas serem verificadas como viciadas).**

#### Método

Existem testes paramétricos (ANOVA) e não paramétricos (Kruskal-Wallis) em que as hipóteses são testadas de maneiras diferentes, porém com maior poder de inferência ou definição de influência de fatores chaves presentes nas amostragens. Assim, certo conjunto de dados analisados dependem de um pressuposto de distribuição normal, onde os dados podem ser considerados paramétricos e comparados em testes paramétricos, caso contrário, deve-se aplicar testes não-paramétricos.

Diante da necessidade de se verificar a legitimidade dos resultados de uma eleição buscou-se a realização de um teste de amostras que comparem os resultados de dois candidatos diante de variáveis como Estado, Cidade, Zona, Seção, Modelos de urnas, número de votos, a serem verificados para dois candidatos, buscou-se um teste paramétrico de comparação de medidas de tendência central, que compare diferenças e influências verificáveis.

#### Poder de comparação de médias

O teste de comparação de medidas de tendência central Kruskal-Wallis foi escolhido para identificar possíveis interferências no resultado das eleições a se comparar quantidades de votos por modelos de urnas, hipótese discutida pelo mundo político, mídia e população até o presente momento. Considerando que o teste ANOVA necessita de certas suposições para ser realizado, dentre essas a verificação de nível de significância.

Também foram aplicados testes de regressão linear de Spearman, que denota o grau de correlação entre duas variáveis pareadas, também considerado um teste não paramétrico, mas robusto o suficiente para expressar possíveis correlações.

#### Amostragens

Primeiro foi necessário realizar um número de observações grande o suficiente (disponível no próprio site do TSE), depois foi necessário tratar os dados e organizar para reduzir a variância amostral separando as capitais por número de eleitores aptos, depois foram retiradas seções com menos de 100 eleitores aptos e mais de 400 ou 450 dependendo de uma redução de perda de informações.

#### Redução de variabilidade (retirada de capitais)

Não há correlação verificável entre população/demografia e a diferença entre as médias comparáveis para dois grupos, em outras palavras, **diferença populacional amostral não modifica resultados finais de comparação de médias ou medianas, mas aumenta a variância de um grupo diante do outro**, por isso não é recomendável comparar populações tão diferentes em tamanhos pela ótica estatística. Resultados com grupos de n amostral diferentes levantam questionamentos estatísticos que são respondidos por essa mesma diferença de n amostral entre populações e precisam ser filtrados.

**Nenhum desses tratamentos afetam a expressão dos resultados finais**, somente a variabilidade das amostras quanto ao tamanho dos conjuntos amostrados. Os dados não são selecionados por critério de preferência de candidatos em nenhum tratamento, nem por zonas ou seções em regiões específicas, etnia, crenças, sexo, ou qualquer característica discriminatória de populações, mas sim pelos tamanhos dos conjuntos de amostras de forma neutra. Também não existem “questões de infraestrutura diversa” (explicação abstrata) para justificar a localização das urnas e sua distribuição em capitais e cidades do interior, como alegado por certos profissionais encontrados na internet.

#### Metodologia específica

##### Fonte de dados

Este trabalho utilizou uma amostra significativa (Votos das cidades exceto capitais) dentro da população integral (Estado) em um estudo observacional para 6 estados da federação (Acre, Bahia, Ceará, Goiás, Maranhão, Paraná, Santa Catarina) para realização de estatística inferencial probabilística (Testes de Kruskal-Wallis e Correlação de Spearman) identificando diferenças entre medidas de tendências centrais e possíveis correlações entre variáveis pareadas.



As amostras são planilhas referentes ao pleito eleitoral do ano de 2022 que foram disponibilizadas no próprio site do TSE a partir do dia 1/11 e baixadas pelo autor deste documento no dia 3/11. Acrescenta-se que na madrugada do dia 4, o site do TSE ficou fora do ar retornando com planilhas e arquivos modificados com data de criação para o dia 4/11.

As planilhas aqui utilizadas não se encontram mais no site oficial e devem ser conseguidas de forma alternativa na internet através do link:

<https://1drv.ms/u/s!Aist-PiC7g8JeZJm2MN-boHde0?e=GEpaJw>

Planilhas organizadas podem ser baixadas em:

[https://1drv.ms/u/s!Aist-PiC7g8JaU-F\\_991nuuZTPg?e=SbdaNa](https://1drv.ms/u/s!Aist-PiC7g8JaU-F_991nuuZTPg?e=SbdaNa)

Porém, é importante que o leitor se sinta convidado a realizar este mesmo trabalho com as planilhas disponibilizadas no próprio site oficial a fim de verificar os mesmos, ou diferentes resultados. Segue o link de acesso aos arquivos presentes no site:

<https://dadosabertos.tse.jus.br/dataset/resultados-2022-boletim-de-urna>

Também está disponível a planilha auxiliar que combina os códigos de modelo das urnas com seus anos de produção:

<https://dadosabertos.tse.jus.br/dataset/correspondencia-entre-numero-interno-e-modelo-da-urna-1>

Esta planilha foi alocada numa pasta destino que possa ser encontrada pela fórmula 1 aplicada na planilha do excel, e renomeada para “Organização de urnas do tse” como descrito na mesma fórmula.

As planilhas foram organizadas e colunas que não continham informações relevantes foram removidas somente Unidade Federativa, Cidade, Zona eleitoral, Seção, Número de eleitores aptos, Número de comparecimentos, Ano modelo de urna, Candidatos, Quantidade de votos por candidatos, Brancos, Nulos e Urnas organizadas por novas e antigas e Código gerado pela iniciação da urna.

No excel, foram utilizadas as seguintes fórmulas para auxiliar a organização dos votos em linhas e conversão dos códigos de urnas em Ano modelo de urna e por fim, Urnas organizadas por novas e antigas:

Fórmula 1:

=SE(E(K2>'D:\... \[Organização de urnas do tse.xlsx]Planilha1'!\$B\$2;K2<'D:\... \[Organização de urnas dotse.xlsx]Planilha1'!\$C\$2);2009;SE(E(K2>'D:\... \[Organizaçãodeurnasdotse.xlsx]Planilha1'!\$B\$3;K2<'D:\... \[Organização de urnas do tse.xlsx]Planilha1'!\$C\$3);2010;SE(E(K2>'D:\... \[Organização de urnas dotse.xlsx]Planilha1'!\$B\$4;K2<'D:\... \[Organizaçãodeurnasdotse.xlsx]Planilha1'!\$C\$4);2011;SE(E(K2>'D:\... \[Organizaçãodeurnasdotse.xlsx]Planilha1'!\$B\$5;K2<'D:\... \[Organizaçãodeurnasdotse.xlsx]Planilha1'!\$C\$5);2011;SE(E(K2>'D:\... \[Organização de urnas do tse.xlsx]Planilha1'!\$B\$6;K2<'D:\... \[Organização de urnas dotse.xlsx]Planilha1'!\$C\$6);2011;SE(E(K2>'D:\... \[Organizaçãodeurnasdotse.xlsx]Planilha1'!\$B\$7;K2<'D:\... \[Organizaçãodeurnasdotse.xlsx]Planilha1'!\$C\$7);2015;SE(E(K2>'D:\... \[Organizaçãodeurnasdots e.xlsx]Planilha1'!\$B\$8;K2<'D:\... \[Organização de urnas do tse.xlsx]Planilha1'!\$C\$8);2020;0))))))

Esta formula serve para localizar o arquivo de referência na pasta e identificar os anos de produção das urnas utilizadas com os códigos dos modelos que estão disponíveis no site do TSE.

Grupo de Fórmulas 2:

=SE(I7926="LULA";J7926;M7927)

=SE(I7926="JAIR BOLSONARO";J7926;N7927)

=SE(I7926="Branco";J7926;O7927)

=SE(I7926="Nulo";J7926;P7927)

Obs.: Este grupo é colocado para cada célula a começar da última célula preenchida e replicado para cima.

Fórmula 3:

=SE(L2<2020;"Urnas antigas";"Novo modelo 2020")

Essa fórmula separou urnas em novas e antigas a fim de atender aos testes de verificação categóricas relevantes às hipóteses para urnas honestas ou viciadas.

Ao final da organização de dados foi aplicada a função do excel “Remover duplicadas” para alinhar as quatro opções de votos na mesma linha, pareando dados em atendimento à suposição do teste Anova. Por fim, os dados foram transferidos para uma planilha em um software específico de estatística Statistica versão trial para windows, TIBCO, e pode ser baixado via link:

<https://www.tibco.com/resources/product-download/tibco-statistica-trial-download-for-windows#>

Vídeos quem explicam as rotinas estatísticas necessárias para a realização dos testes estatísticos no software acima podem ser encontrados em:

Teste de normalidade:

<https://www.youtube.com/watch?v=QCBVPZ07Evk>

Teste Kruskal-Wallis

<https://www.youtube.com/watch?v=VDLp9dJA5Ak>

Teste Spearman

<https://www.youtube.com/watch?v=nQLuyguT7DA>

Gráfico variados

[https://www.youtube.com/watch?v=q0Vbe\\_K0ALo](https://www.youtube.com/watch?v=q0Vbe_K0ALo)

Verificação posterior à eleição/Não auditoria

É importante ressaltar que as análises deste trabalho correspondem a inferência nos resultados posteriores à realização das eleições nacionais e que **não trata** de uma avaliação direta de hardware das urnas eleitorais, software, quanto ao sigilo mantido sobre seu código fonte a posteriori, influência de repositórios externos verificados pelo relatório das forças armadas, número de patrimônio repetido nas urnas antigas, mesmo número descodificável por criptografia hexadecimal, vazamentos de nomes e votos como falha de sigilo das urnas, defeitos frequentes nas urnas durante dia de votação, influência direta por algum grupo originário da Venezuela como apontado no documento da contratação, entre outras falhas de segurança apontadas até o presente momento.

Acrescento que as planilhas apontam a utilização de urnas modelo 2009 no pleito 2022, mesmo sob a determinação legal de que urnas com mais de 10 anos de funcionamento não podem ser utilizadas para uma eleição. Acrescento também que urnas 2009 foram movidas de lugar no dia 20 de novembro (domingo) sob alegação pelo próprio TSE de que seriam levadas à reciclagem, a considerar que estas não teriam sido utilizadas na eleição. Não atestado que sejam as mesmas urnas, porém a mesma data de fabricação.

## Tratamentos

### Definição axiomática de Kolmogorov

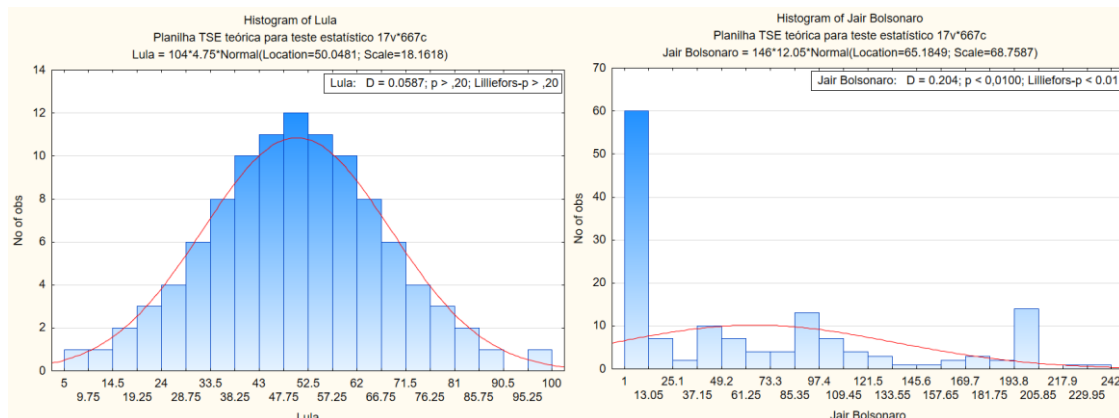
Após a realização do tratamento de dados as variáveis avaliadas foram testadas a distribuição normal incluindo análise do teste de Kolmogórov-Smirnov pressupondo que os dados devem ser paramétricos para a realização de teste ANOVA.

Kolmogorv-Smirnov é um teste de grau do ajuste, ou seja, serve para verificar se as pontuações que obtivemos da amostra seguem uma distribuição normal ou gaussiana permitindo medir o grau de concordância entre a distribuição de um conjunto de dados com uma distribuição teórica específica (parábola teórica). Seu objetivo é indicar se os dados provêm de uma população que possui a distribuição teórica especificada e se apresentam como dados paramétricos.

Neste teste, a hipótese nula ( $H_0$ ) estabelecerá que a distribuição empírica é semelhante à teórica, assim o resultado deve apresentar um valor  $p < 0,05$  para que a  $H_0$  seja rejeitada. A hipótese nula estabelecerá que a distribuição de frequência observada é consistente com a distribuição teórica representativa de dados paramétricos como a variável  $\alpha$  indicando o nível de significância do teste.

### Kolmogorov-Smirnov & Lilliefors

Os testes de Kolmogorov-Smirnov têm como função descritiva de amostras seu nível de aproximação com a distribuição normal, sendo esta, parábola descritiva de frequência de medidas baseadas na frequência de dados em torno de uma média considerando certo grau de dispersão pré-determinado. Também é realizado o teste de Lilliefors de forma complementar para avaliar se a discrepância máxima dos dados é grande o suficiente para ser estatisticamente significativa, sendo sua distribuição de probabilidade de forma estocástica menor que o teste Kolmogorov-Smirnov. Assim, considerando cuidadosamente a distribuição, bem como os parâmetros utilizados.



1 – À esquerda um exemplo hipotético de distribuição normal, à direita uma distribuição hipotética não normal.

Com base nessas premissas não foi verificada distribuição normal para dados paramétricos não permitindo a aplicação de teste ANOVA unidirecional também ditos teste paramétrico.

A variável D é resultante do teste de Kolmogorov-Smirnov & Lilliefors cuja significância p deve ser maior que 0,05, bem como p para Lilliefors também maior que 0,05 representando distribuição normal dos dados. **Se a estatística D é significativa** (possui p menor que 0,05), então **a hipótese ( $H_0$ ) de que a respectiva distribuição é normal deve ser rejeitada**. Para muitos programas de software, os valores de probabilidade que são relatados são baseados naqueles tabulados por Massey (1951); esses valores de probabilidade são válidos quando a média e o desvio padrão da distribuição normal são conhecidos a priori e não estimados a partir dos dados. No entanto, geralmente esses parâmetros são

calculados a partir dos dados reais. Nesse caso, o teste de normalidade envolve uma hipótese condicional complexa ("qual a probabilidade de obter uma estatística  $D$  dessa magnitude ou maior, contingente à média e ao desvio padrão computados a partir dos dados"), e é assim que as probabilidades de Kolmogorov-Smirnov e Lilliefors devem ser interpretadas (Lilliefors, 1967).

Parâmetro também importante de modelos estatísticos preditivos é a homocedasticidade que ocorre quando grupos de dados de uma ou mais observações, o padrão de variância (ou independente) em relação às variáveis explicativas, permanecem constantes. A homocedástica é importante por contrastar com uma desorganização aleatória suficientemente alta para caracterizar uma distribuição normal.

#### Gráfico de intervalos de confiança

O gráfico de médias com intervalo de confiança foi utilizado, o intervalo é representado por um limite representando a mínima de 95% para mais e 95% para menos de intervalo de confiança em torno da média.

#### Teste de Kruskal Wallis

O teste de Kruskal-Wallis é uma alternativa não paramétrica ao ANOVA unidirecional. Assim, foi utilizado para comparar os dois grupos de amostras (Urnas organizadas em novas 2020 e antigas anteriores à 2020), sendo considerada hipótese nula ( $H_0$ ) quando as diferentes amostras na comparação foram retiradas da mesma distribuição ou de distribuições com a mesma mediana. Assim, a interpretação do teste de Kruskal-Wallis é basicamente semelhante à da ANOVA unidirecional, exceto que é baseada em classificações e não em médias.

#### Regressão linear de Spearman

Para a verificação da correlação entre população e votos direcionados aos candidatos Lula e Bolsonaro foram aplicado o teste de correlação de Spearman para comparecimentos ao pleito eleitoral e votos nos dois candidatos disponíveis no segundo turno.

O teste Spearman  $R$  é compreendido como o coeficiente de correlação produto-momento de Pearson (Pearson  $r$ ); isto é, em termos da proporção de variabilidade contabilizada, exceto que Spearman  $R$  é calculado a partir de fileiras. Como mencionado, Spearman  $R$  assume que as variáveis em consideração foram medidas em pelo menos uma escala ordinal (ordem de classificação); ou seja, as observações individuais (casos) podem ser classificadas em duas séries ordenadas. Os parâmetros resultantes são  $r^2$  sendo grau de correlação (%) e  $p$  como grau de significância da amostra de duas variáveis pareadas, a depender do seu tamanho.

## Resultados

Gráficos de distribuição para modelos de urnas, Estados analisados, eleições 2022 2 turno.

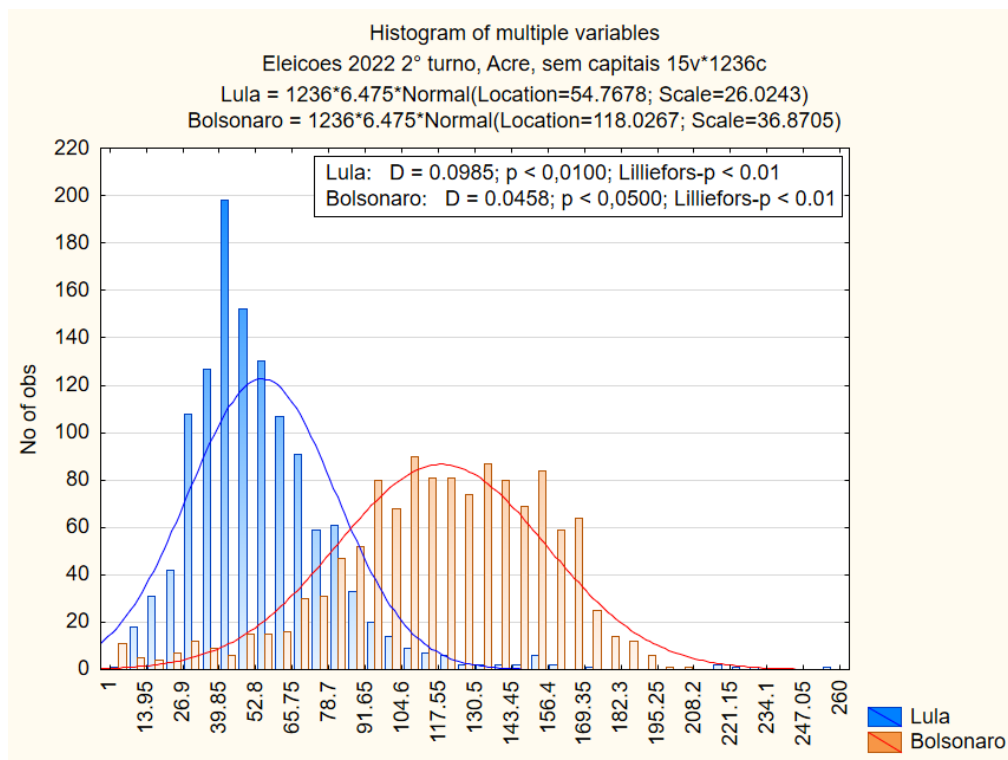


Figura 2 – Distribuição dos votos de Lula e Bolsonaro no Acre onde **Bolsonaro foi vitorioso**.

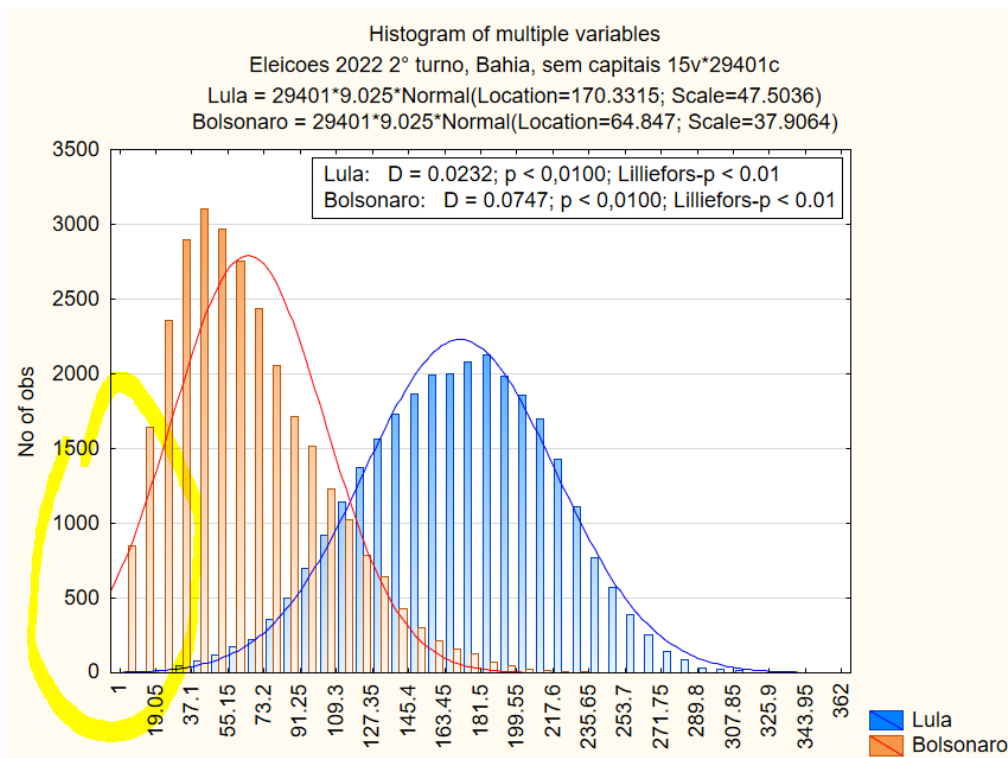


Figura 3 – Distribuição dos votos de Bolsonaro e Lula na Bahia onde **Lula foi vitorioso**, demarcação de irregularidades cedásticas nos dados.

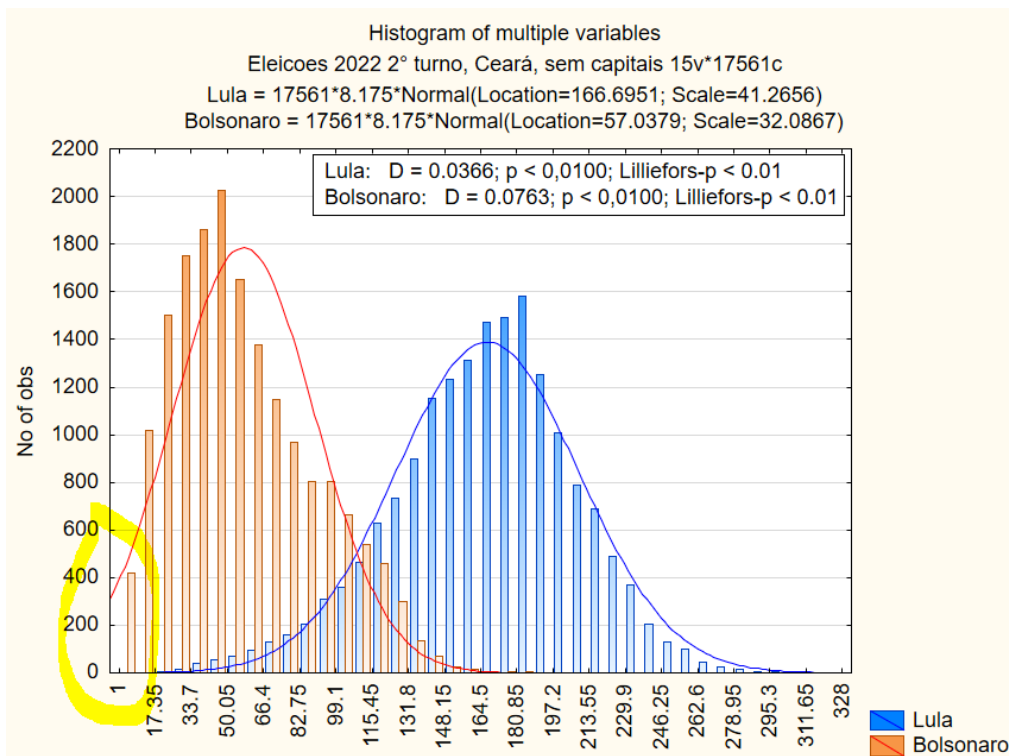


Figura 4 – Distribuição dos votos de Bolsonaro e Lula no Ceará onde **Lula foi vitorioso**, demarcação de irregularidades cedásticas nos dados.

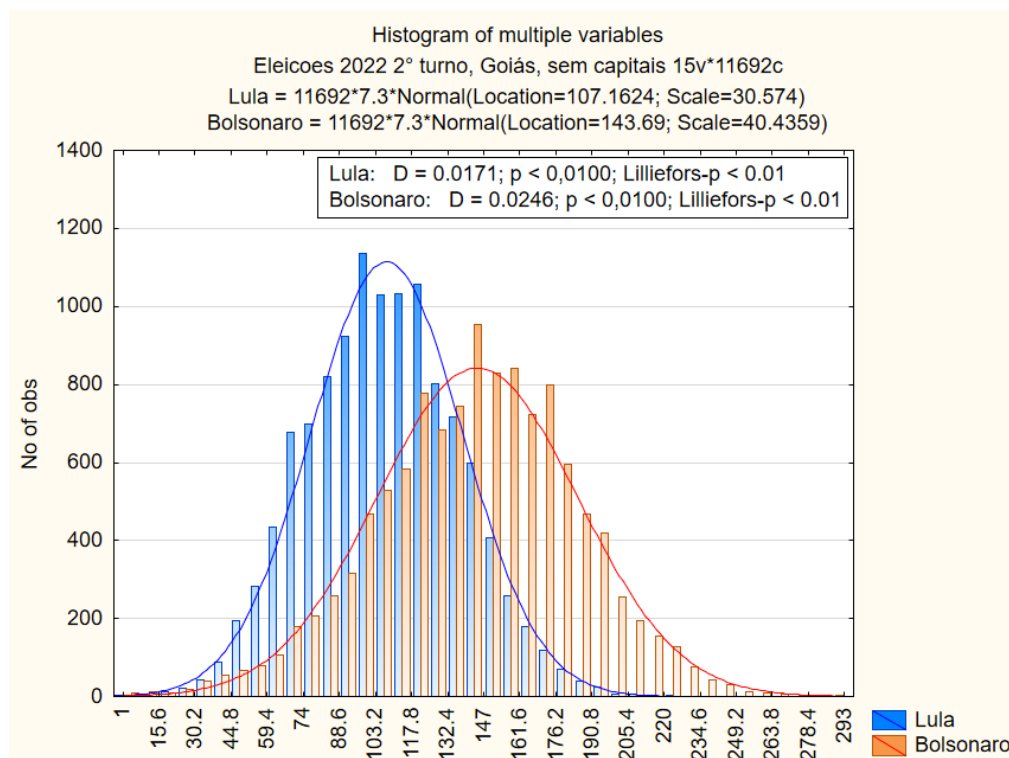


Figura 5 – Distribuição dos votos de Lula e Bolsonaro no Goiás onde **Bolsonaro foi vitorioso**.

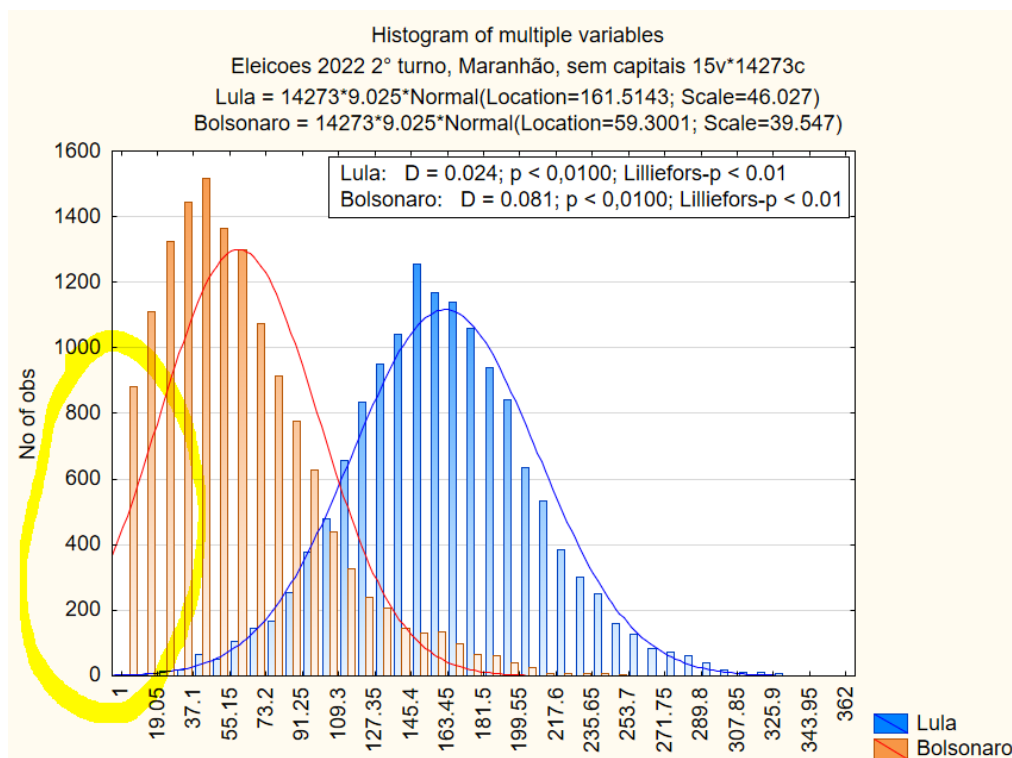


Figura 6 – Distribuição dos votos de Bolsonaro e Lula no Maranhão onde **Lula foi vitorioso**, demarcação de irregularidades cedásticas nos dados.

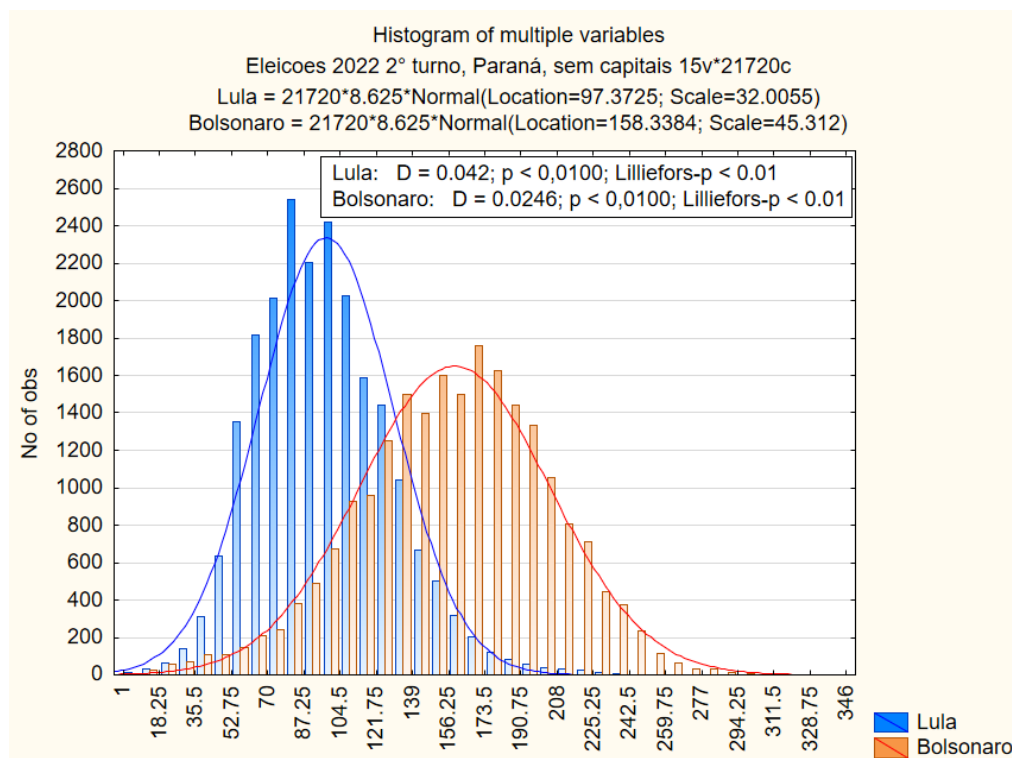


Figura 7 – Distribuição dos votos de Lula e Bolsonaro no Paraná onde **Bolsonaro foi vitorioso**.

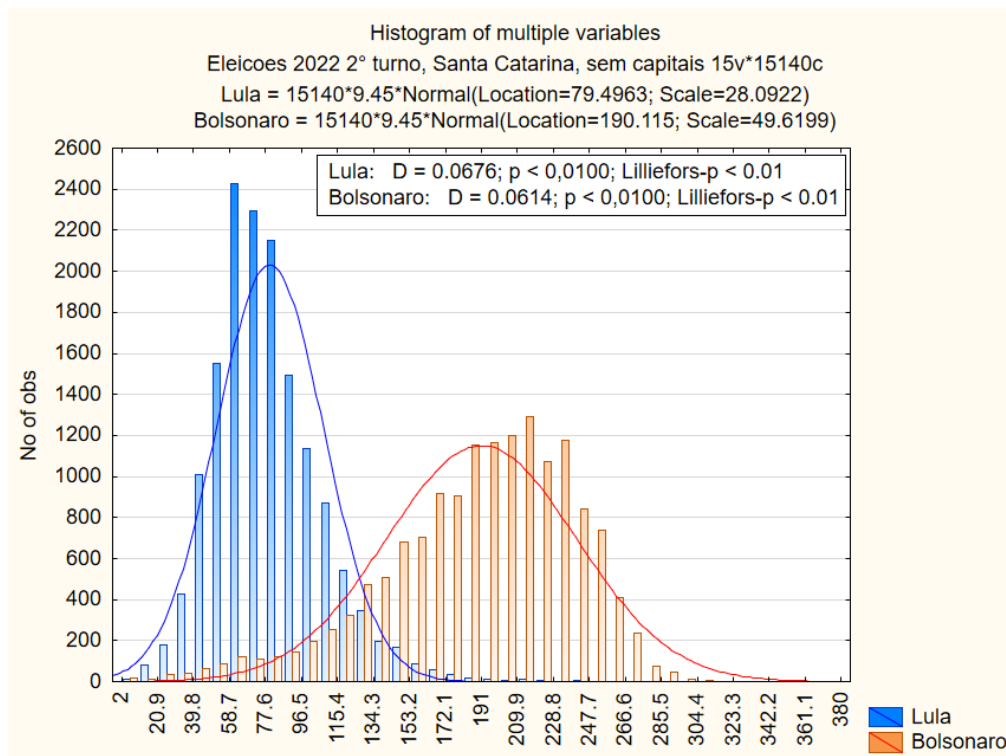


Figura 8 – Distribuição dos votos de Lula e Bolsonaro em Santa Catarina onde **Bolsonaro foi vitorioso**.

Os dados se apresentaram pareados, independentes e sem verificação de distribuição normal, assim foi selecionado o teste de comparação de medidas de tendência central Kruskal-Wallis. Os histogramas para Bolsonaro no Nordeste, apesar de inconclusivos, apresentaram uma aproximação incomum para as frequências 0 ~ 10, demonstrando tendência a zero para as frequências verificadas. Isso não foi observado nos estados em que Lula foi derrotado.



## Gráfico de intervalos de confiança (95%) e teste de Kruskal-Wallis

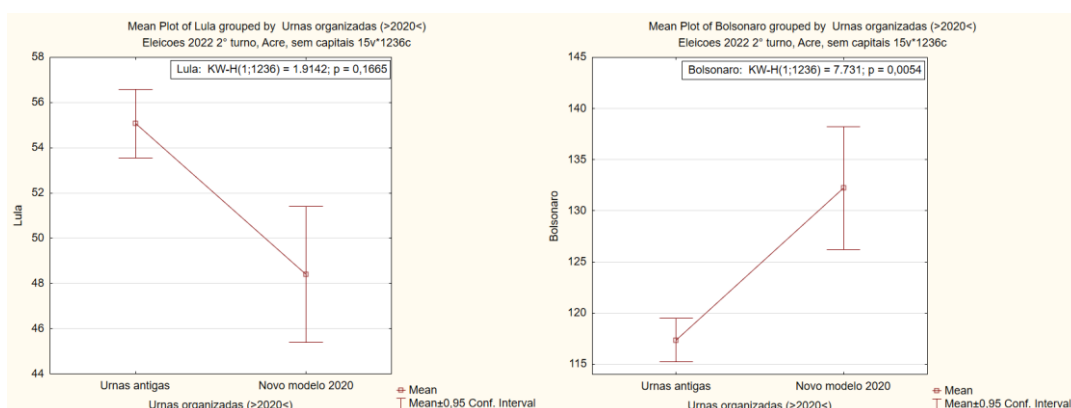


Figura 9 – Gráfico para o Acre, com médias (intervalos 95% para mais e menos) apontando **tendência de aumento de votos para Lula e redução de votos para Bolsonaro nas urnas antigas**. Não ocorreram diferenças estatísticas significantes para Lula entre urnas antigas e novas ( $z' = 1,38$ ;  $p = 0,16$ ,  $H_0$ ), porém ocorreram diferenças estatísticas significantes para Bolsonaro entre urnas antigas e novas ( $z' = 2,78$ ;  $p > 0,01$ ,  $H_1$ ).

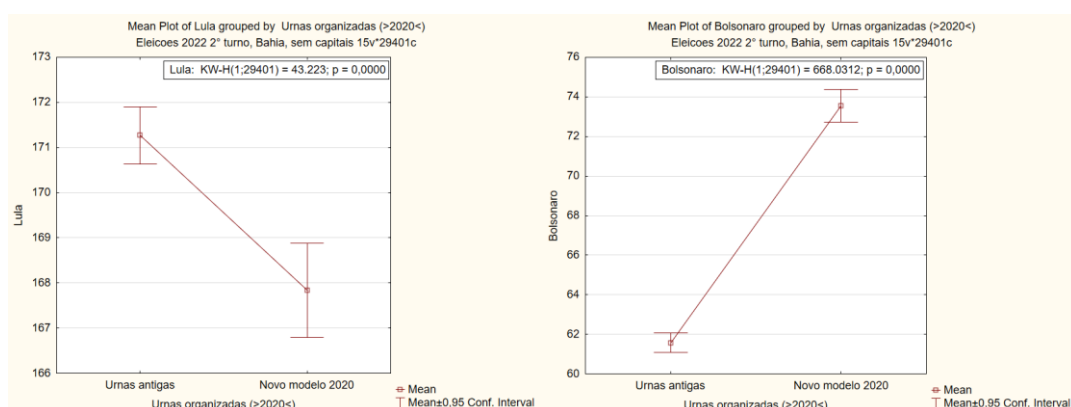


Figura 10 – Gráfico para o Bahia, com médias (intervalos 95% para mais e menos) apontando **tendência de aumento de votos para Lula e redução de votos para Bolsonaro nas urnas antigas**. Ocorreram diferenças estatísticas significantes para Lula entre urnas antigas e novas ( $z' = 6,57$ ;  $p < 0,01$ ,  $H_1$ ), bem como diferenças estatísticas significantes para Bolsonaro entre urnas antigas e novas ( $z' = 25,84$ ;  $p > 0,01$ ,  $H_1$ ).

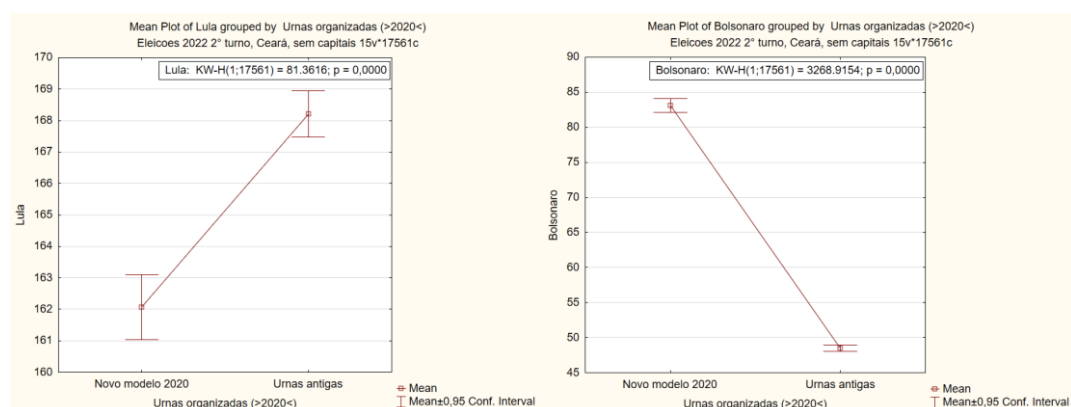


Figura 11 – Gráfico para o Ceará, com médias (intervalos 95% para mais e menos) apontando **tendência de aumento de votos para Lula e redução de votos para Bolsonaro nas urnas antigas**. Ocorreram diferenças estatísticas significantes para Lula entre urnas antigas e novas ( $z' = 9,02$ ;  $p < 0,01$ ,  $H_1$ ), bem como diferenças estatísticas significantes para Bolsonaro entre urnas antigas e novas ( $z' = 57,17$ ;  $p > 0,01$ ,  $H_1$ ).

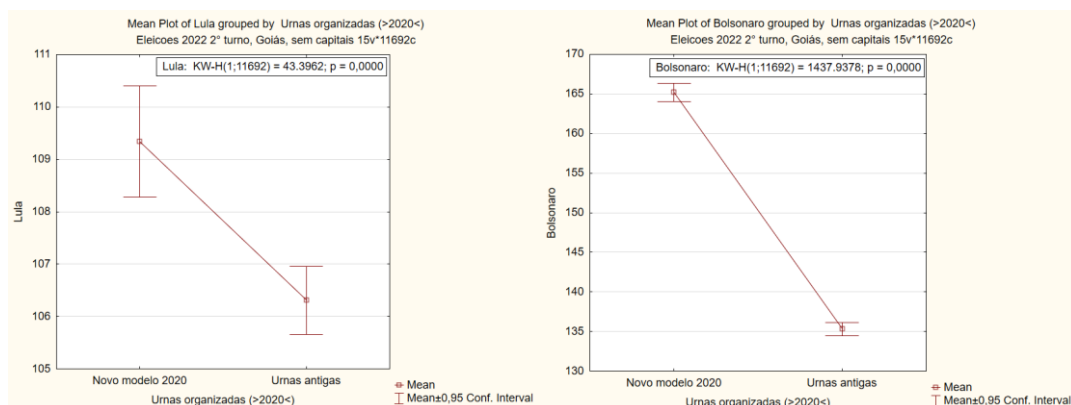


Figura 12 – Gráfico para o Goiás, com médias (intervalos 95% para mais e menos) apontando de **tendência de redução de votos para ambos os candidatos, Lula e Bolsonaro, nas urnas antigas**. Ocorreram diferenças estatísticas significantes para Lula entre urnas antigas e novas ( $z' = 6,58$ ;  $p < 0,01$ ,  $H_1$ ), bem como diferenças estatísticas significantes para Bolsonaro entre urnas antigas e novas ( $z' = 37,91$ ;  $p > 0,01$ ,  $H_1$ ).

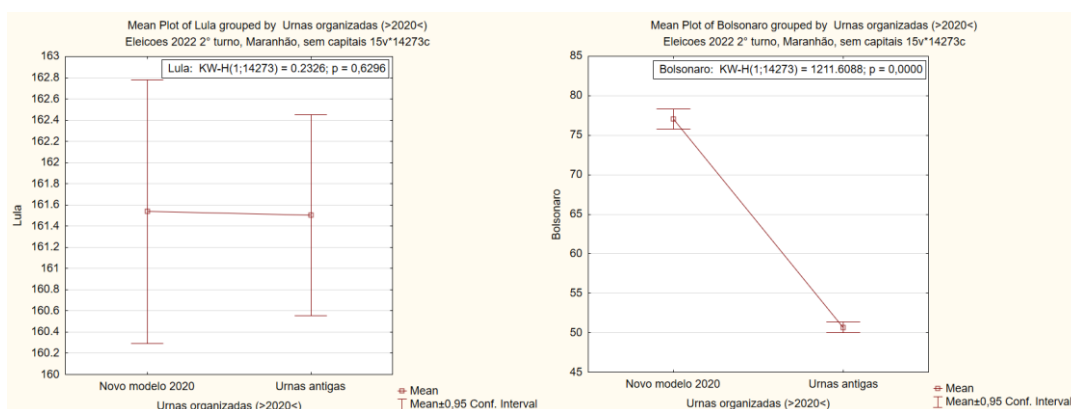


Figura 13 – Gráfico para o Maranhão, com médias (intervalos 95% para mais e menos) apontando verificação de **tendência de igualdades para Lula e redução para Bolsonaro nos votos observados entre urnas antigas e novas**. Não ocorreu diferenças estatísticas, nem significantes, para os votos do Lula entre urnas modelos antigos e novos ( $z' = 0,48$ ;  $p = 0,63$ ,  $H_0$ ), porém ocorreu diferenças estatísticas significantes para Bolsonaro ( $z' = 34,81$ ;  $p < 0,01$ ,  $H_1$ ).

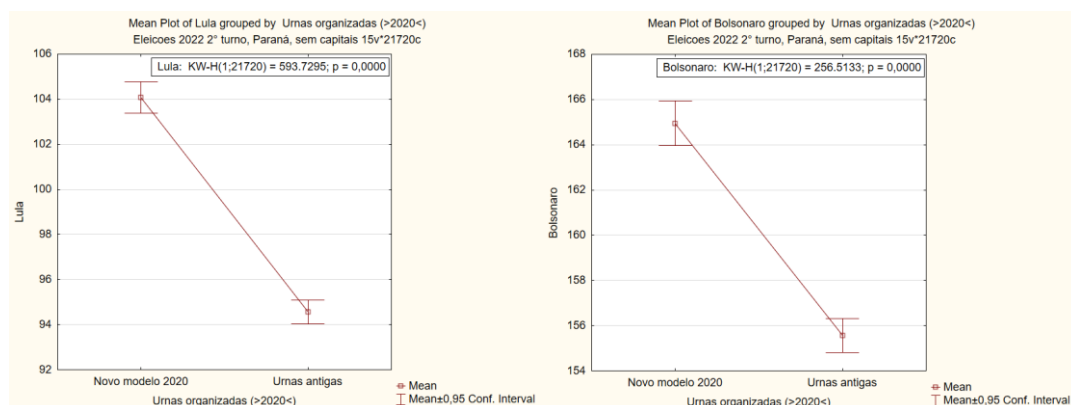


Figura 14 – Gráfico para o Paraná, com médias (intervalos 95% para mais e menos) apontando verificação de **tendência de redução de votos para ambos os candidatos, Lula e Bolsonaro, nas urnas antigas**. Ocorreram diferenças estatísticas significantes para Lula entre urnas antigas e novas ( $z' = 24,36$ ;  $p > 0,01$ ,  $H_1$ ), bem como diferenças estatísticas significantes para Bolsonaro entre urnas antigas e novas ( $z' = 16,02$ ;  $p < 0,01$ ,  $H_1$ ).

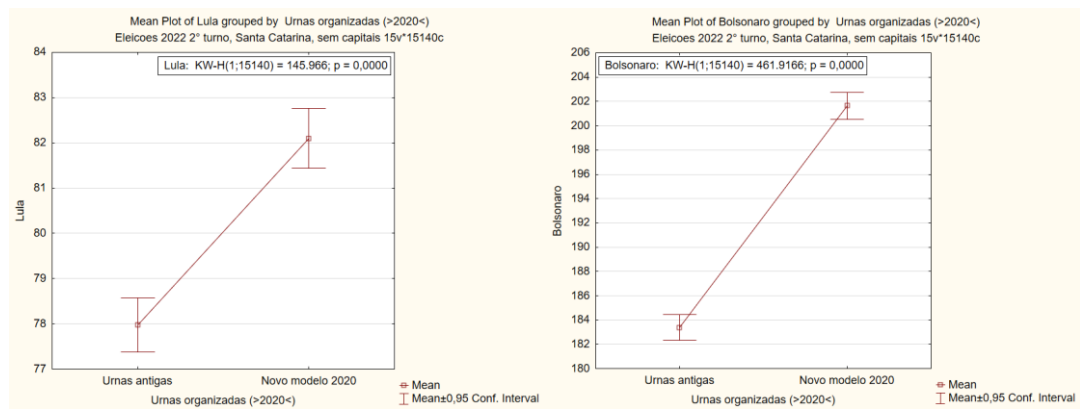


Figura 15 – Gráfico para o Santa Catarina, com médias (intervalos 95% para mais e menos) apontando verificação de **tendência de redução de votos para ambos os candidatos, Lula e Bolsonaro, nas urnas antigas**. Ocorreram diferenças estatísticas significantes para Lula entre urnas antigas e novas ( $z' = 12,08$ ;  $p > 0,01$ ,  $H_1$ ), bem como diferenças estatísticas significantes para Bolsonaro entre urnas antigas e novas ( $z' = 21,49$ ;  $p < 0,01$ ,  $H_1$ ).

## Gráfico de nuvem de pontos e correlação de Spearman

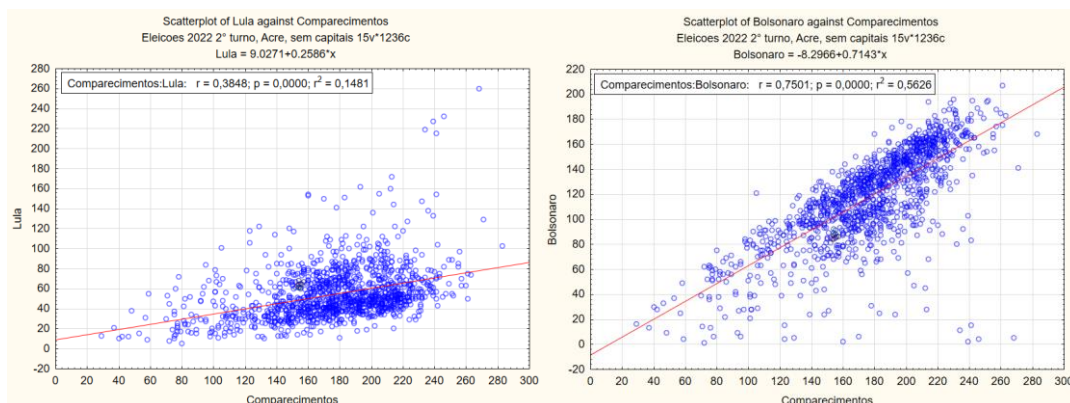


Figura 16 – Gráfico de nuvens de dispersão para o Acre sem forma aparente, com correlação (comparecimentos e votos no candidato) de 14% para Lula e 56% para Bolsonaro demonstrando possível engajamento maior de comparecimento e votos em Bolsonaro. O gráfico apresenta distribuição de pontos natural aleatório sem formas aparentes.

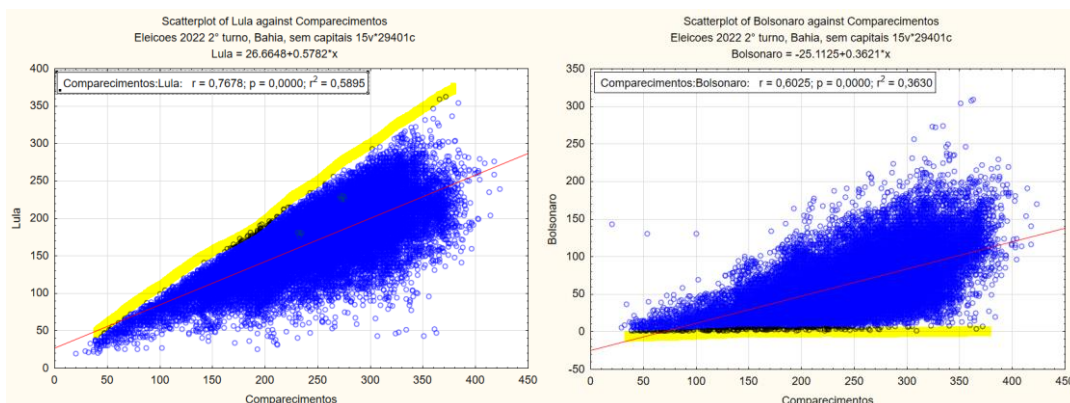


Figura 17 – Gráfico de nuvens de dispersão para a Bahia com correlação (comparecimentos e votos no candidato) de 58% para Lula e 36% para Bolsonaro, com formação geométrica de nuvem improvável para o tipo de situação (linha amarela).

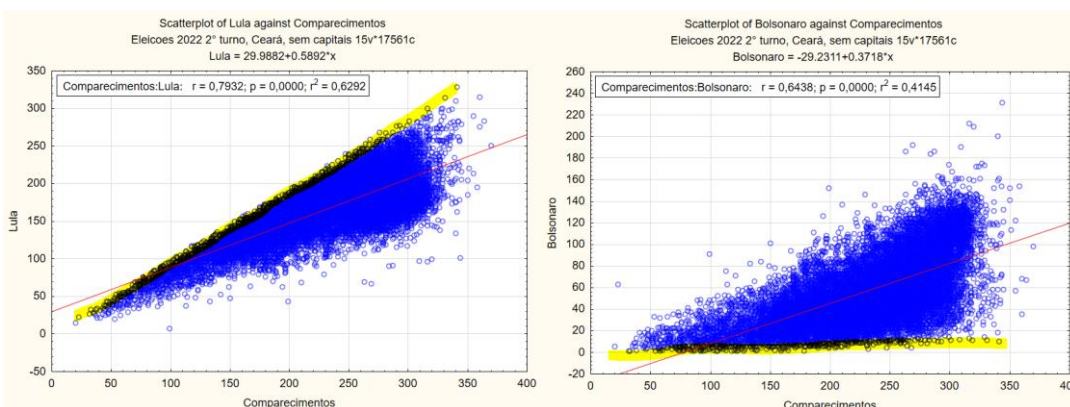


Figura 18 – Gráfico de nuvens de dispersão para a Ceará com correlação (comparecimentos e votos no candidato) de 63% para Lula e 41% para Bolsonaro, com formação geométrica de nuvem improvável para o tipo de situação (linha amarela).

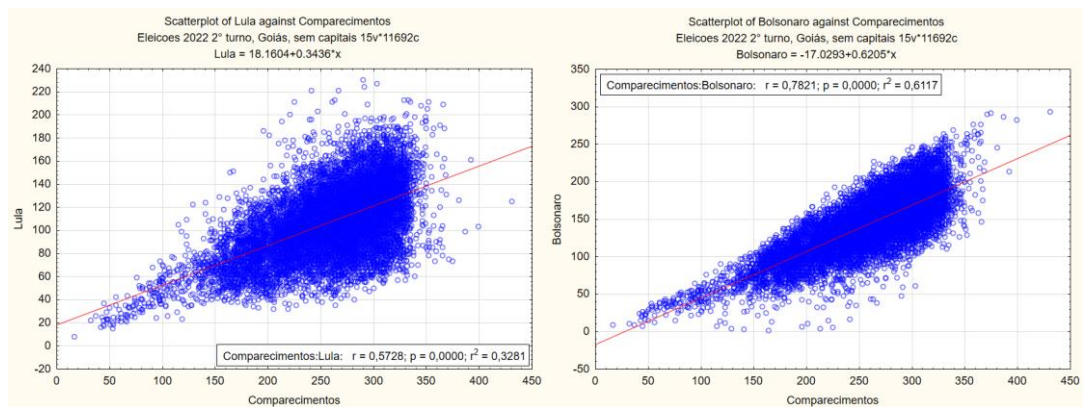


Figura 19 – Gráfico de nuvens de dispersão para o Goiás sem forma aparente, com correlação (comparecimentos e votos no candidato) de 33% para Lula e 61% para Bolsonaro demonstrando possível engajamento maior de comparecimento e votos em Bolsonaro.

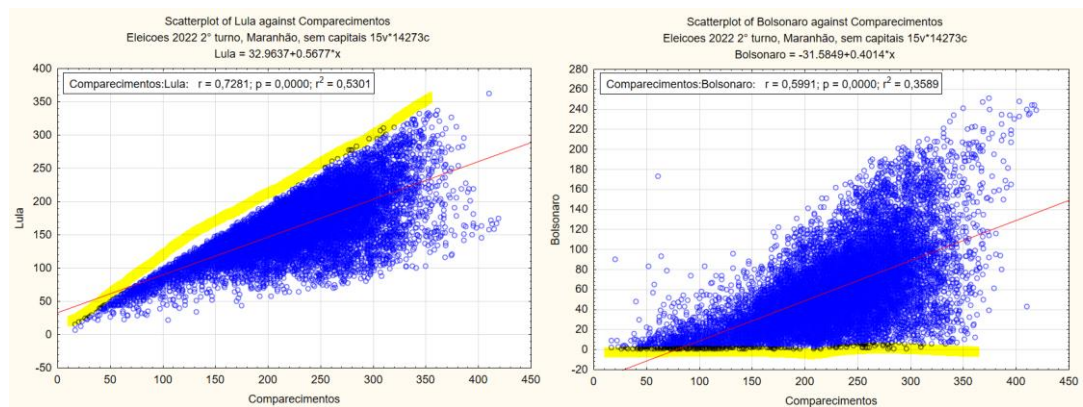


Figura 20 – Gráfico de nuvens de dispersão para a Maranhão com correlação (comparecimentos e votos no candidato) de 53% para Lula e 35% para Bolsonaro, com formação geométrica de nuvem improvável para o tipo de situação (linha amarela).

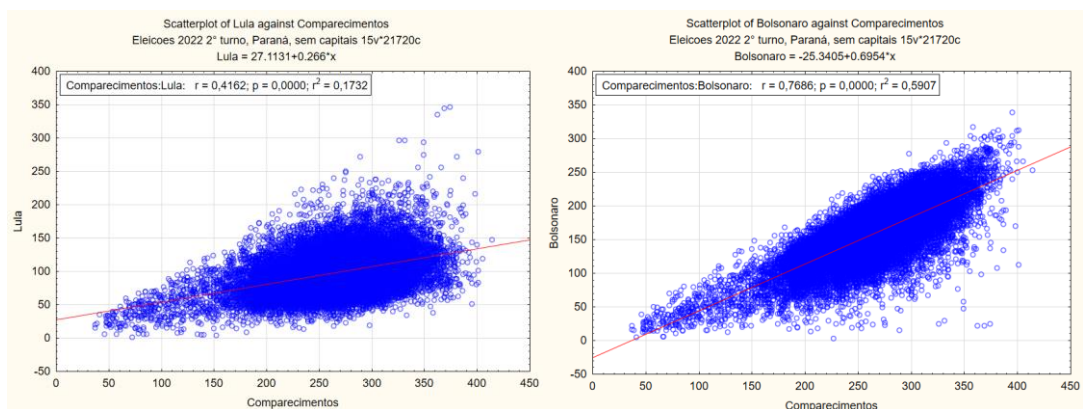


Figura 21 – Gráfico de nuvens de dispersão para o Paraná sem forma aparente, com correlação (comparecimentos e votos no candidato) de 17% para Lula e 59% para Bolsonaro demonstrando possível engajamento maior de comparecimento e votos em Bolsonaro.



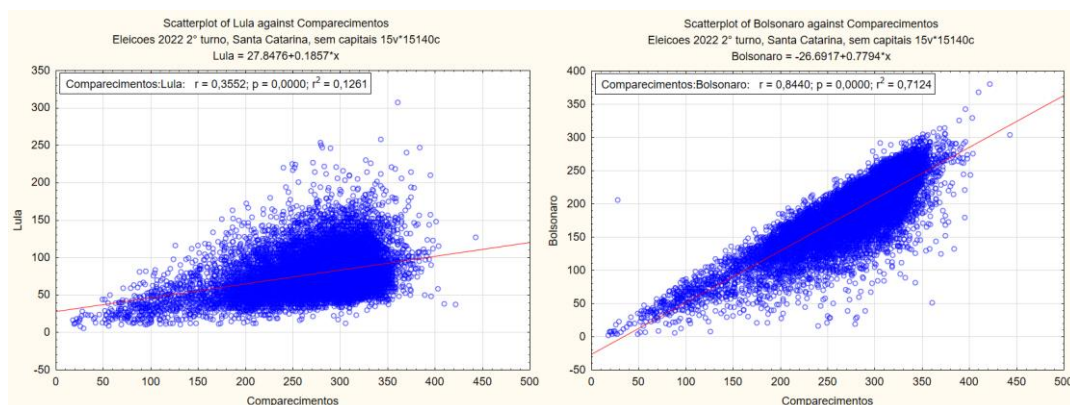


Figura 22 – Gráfico de nuvens de dispersão para o Santa Catarina sem forma aparente, com correlação (comparecimentos e votos no candidato) de 17% para Lula e 59% para Bolsonaro demonstrando possível engajamento maior de comparecimento e votos em Bolsonaro.

### Conclusões

Uma eleição pressupõe isonomia em todas as suas dimensões: legal, propaganda, de ocupação de espaços físicos, demanda temporal e capital, sendo todas essas dimensões discutíveis exceto uma delas, a dimensão da liberdade de voto. Por isso é de fundamental importância que urnas eletrônicas brasileiras e o processo que leva ao voto ocorra de forma isonômica.

Em teoria, ao considerar o tamanho (relativamente enorme) das amostras, os resultados eleitorais para votos entre urnas de modelos diferentes deveriam apresentar resultados estatisticamente similares entre eles, pois, uma urna não apresenta forma, cor, ou qualquer informação anexada que possa interferir na preferência eleitoral por candidato x ou y, muito menos pode ser distribuída por grupos eleitorais ou preferências por “currais” eleitorais, ou qualquer aspecto que diferencia uma escolha eleitoral perante o eleitor.

Este trabalho compara argumentos e análises apontados pelo técnico ligado à Eduardo Bolsonaro, Fernando Cerimedo (denunciante da fraude), quem apresenta uma consultoria internacional que verifica possíveis anomalias. Também tenta analisar os argumentos do Nataniel Antonioli, quem realiza uma abordagem de desmistificação da fraude nas urnas eleitorais na tentativa de reproduzir os processos analíticos e verificar resultados em meio ao tumulto de informações sobre o pleito eleitoral de 2022.

Diante do presente exposto este trabalho corrobora os resultados apresentados pela consultoria exposta por Cerimedo mas não consegue verificar os argumentos citados por Nataniel. Assim, foram observadas diferenças inexplicáveis por vias naturais entre urnas antigas e novas que diferem estatisticamente entre si, porém, coincidem entre diferentes estados para o Nordeste. Cerimedo afirma que existem anomalias estatísticas com diferenças entre urnas velhas e novas nas cidades do interior dos estados quando comparados grupos de dados de Lula e Bolsonaro o que se verifica no presente trabalho.

Por outro lado foi observado alguns argumentos apontados por Nataniel sobre as análises realizadas por Cerimedo. Dentre eles:

- 1- “Questões de infraestrutura” onde ele afirma que Cerimedo misturou dados das capitais com as cidades de interior. Neste caso, os resultados apresentaram diferenças de variância que denotava indução ao erro analítico. Sendo assim, as capitais foram separadas das análises aqui realizadas confirmando as afirmações na apresentação do argentino sobre seu tratamento dos dados.
- 2- “Urnas novas foram distribuídas nas capitais e urnas antigas foram distribuídas no interior do estado”, o que induziria alguma relação matemática entre densidade populacional e votos no candidato Lula serem mais frequentes em urnas antigas. Esta é uma afirmação falsa,

considerando que este trabalho verificou distribuição aleatória das urnas em todos os estados analisados, exceto, no Amazonas, em que realmente as urnas antigas foram distribuídas no interior e as novas na capital, por este motivo o estado do Amazonas e outros com problemas similares não foram analisados neste trabalho.

Para estes resultados de correlação foram utilizados gráficos de nuvem com que apresentaram forma extremamente improváveis para populacionais e votos em candidatos, com delineamentos geométricos de recorte.

O trabalho do Argentino foi transparente ao citar que contou com uma consultoria técnica internacional e foi publicado 5 dias após os resultados da eleição porém não apresentou as rotinas estatísticas nem como onde e como elas foram realizadas, já o trabalho do Nataniel que atuou para desqualificar a denúncia do grupo BrazilWasStolen/La Derecha Diário e tem amplo uso pela esquerda foi publicado 5 dias após ao exposto pelo Cerimedo, no dia 10 de novembro.

O autor deste trabalho (Allejo) iniciou no dia 10 após observar ambos com grande desconfiança, finalizando no dia 29, 19 dias no total, o que denota que ambos os casos citados, Nataniel e Cerimedo contaram com auxílio de equipes técnicas que os apoiaram. No vídeo do Nataniel, foi verificado a execução de comandos em python com grande repercussão para download e organização de tabelas com diversos cortes de edição. Ele também utilizou a densidade das cidades do Brasil inteiro via wikipedia, o que por si só, demanda um trabalho descomunal.

O vídeo de apresentação do Cerimedo ocorreu ao vivo e sem cortes, porém, não apresentou o processamento dos dados, sendo assim foi prioridade deste trabalho verificar a fidelidade desses processamentos de dados afim de oferecer alternativa a uma possível manipulação de informações. Assim, **Se este trabalho não verificasse os mesmos resultados, as mesmas anomalias do Cerimedo, seria igualmente imprescindível sua publicação.**

Por outro lado, o vídeo do Nataniel não explicou precisamente como executar os comandos em python para organizar as planilhas, o que, demanda muito tempo e energia de um analista que está começando na área até mesmo um mais experiente, limitando os próprios interessados a não executar os mesmos passos. Nataniel parte de pressupostos não testados e apresenta resultados com base teórica questionável.

Diante de presente exposto, este trabalho verificou modificações não naturais nos resultados eleitorais, bem como a tentativa de desqualificação das informações expostas pelo argentino nas lives BrazilWasStolen pelo canal Fábrica de Nerds do Nataniel Antonioli.