# O impacto de defeitos observados na biometria do mesário sobre o resultado do segundo turno das eleições 2022

A importância de se analisar os votos sem mesário, com falha dupla, é que estes votos são uma fraude largamente conhecida e de fácil implementação, conhecida como voto fantasma em que um terceiro vota por alguém, este voto em um sistema com controle por reconhecimento biométrico em que o mesário não quiser se comprometer, ou ainda em que se quiser medir a efetividade da fraude para algum outro processo do esquema, resultaria na mensagem de log que será descrita abaixo.

## Definição do defeito considerado, o voto sem mesário identificado.

Considerando o Voto Padrão nas urnas disponibilizadas em 2022 para as eleições ocorrem em alguns passos considerando os fluxos de exceção e a nomenclatura que serão dados neste documento.

### Os Passos de votação com os fluxos de exceção de interesse:

1. O “eleitor” entrega seu título de eleitor e um documento de identificação a um “primeiro” mesário.
2. O primeiro lê em voz alta o número do título de eleitor.
3. O “segundo” mesário digita o título de eleitor lido pelo primeiro, uma vez o título do eleitor digitado o nome do eleitor é exibido, o nome é lido em voz alta pelo mesmo.
4. O “primeiro” mesário confirma o nome e pede ao eleitor que ele se identifique biometricamente no terminal do mesário.
5. O “eleitor” oferece um de seus dedos que tiveram a biometria coletada anteriormente.
6. A biometria é avaliada pela máquina por similaridade em quatro tentativas.
   1. Caso a biometria não identifique o “eleitor” o “segundo” mesário exige um documento de identificação.
   2. O “segundo” mesário confirma a identificação do “eleitor” e oferece um de seus dedos para que o equipamento o reconheça.
      1. O mesário tem a sua identificação confirmada e a votação continua no passo 7.
      2. O mesário não tem sua identificação reconhecida e algumas logs são possivelmente informadas neste momento:
         1. Não encontrou digital coletada em nenhum dos arquivos, vai salvar a digital em novo arquivo.
         2. Digital coletada bate com uma digital gravada após o registro inicial de mesário. Não é possível associar a habilitação a um mesário.
         3. Biometria do mesário $ID encontrada em eleitores $ID2.
         4. Biometria do mesário $ID encontrada nos arquivos coletados $ID4.
7. O terminal do eleitor é habilitado.
8. O “eleitor” vota.
9. O voto do eleitor é computado.
10. Reinicia-se o processo.

O trabalho abaixo irá se concentrar nas seguintes falhas encadeadas:

Durante a identificação do eleitor:

1. Eleitor não identificado por biometria.
2. Mesário não identificado por biometria com as seguintes mensagens que avaliei como falha catastrófica sem possível controle imediato daqui em diante chamada de “dupla falha”:
   1. **Não encontrou digital coletada em nenhum dos arquivos, vai salvar a digital em novo arquivo.**
   2. **Digital coletada bate com uma digital gravada após o registro inicial de mesário. Não é possível associar a habilitação a um mesário.**

## Comparando a existência e a não existência de falhas:

### Estatística Descritiva básica:

Realizando um levantamento apenas comparando a existência e não existência dos votos a partir dos boletins de urna foram estratificados três grupos:

Chamaremos a amostra que contém o público completo de: **TODOS**

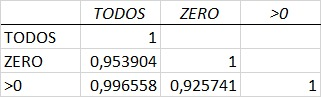
Chamaremos a amostra **que não contém** votos com “dupla falha” de: **ZERO**

Chamaremos a amostra **que contém** votos com “dupla falha” de: **> 0**

Agrupados em CD\_MUNICIPIO + NR\_VOTAVEL, resultando em 22773 categorias observável em Contagem abaixo:

## 

Considerando apenas a correlação entre estes três grupos agrupados por cidade e nr de voto e temos o seguinte resultado:



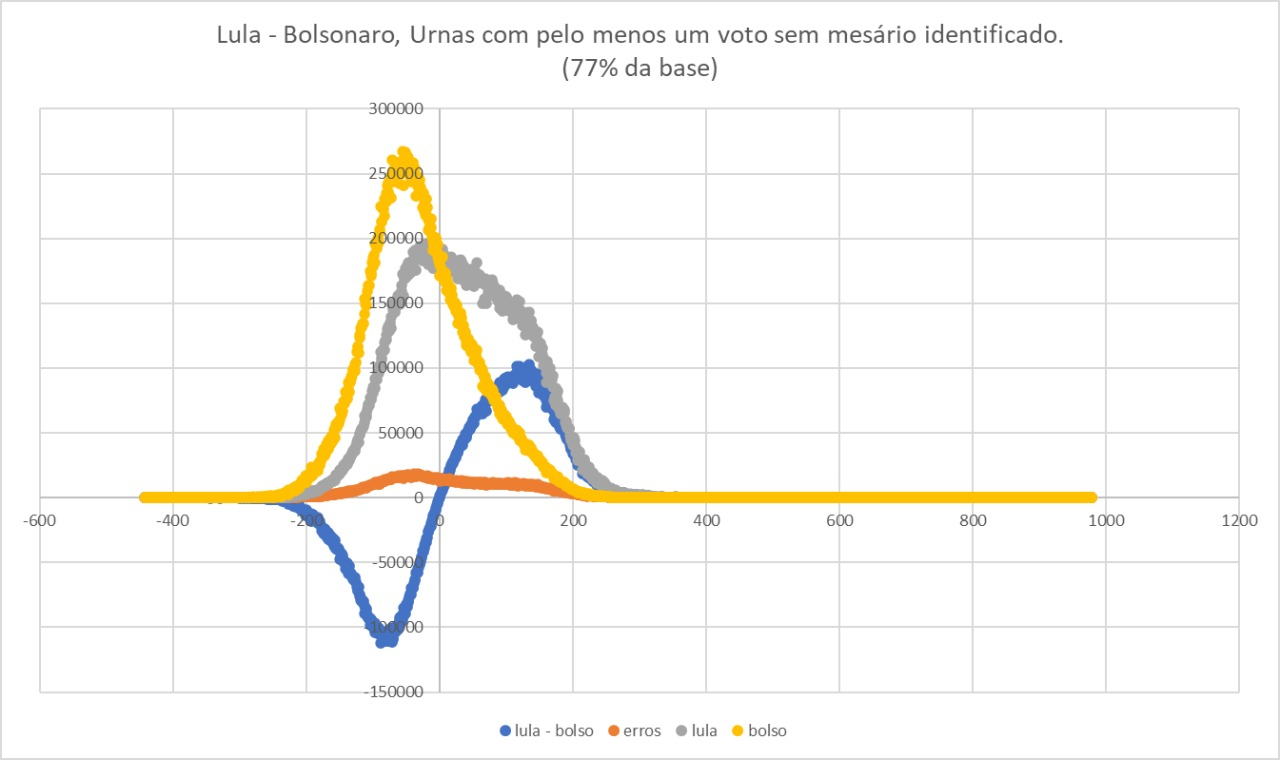
Onde podemos enxergar uma alta correlação (> 0,95) de **TODOS** com os outros 2 grupos e uma baixa correlação de **ZERO** com **>0**

Analisando ainda os tipos de urnas em cada grupo temos:

|  |  |
| --- | --- |
| Votos agrupados por urna | |
| **>0** | **ZERO** |
|  |  |

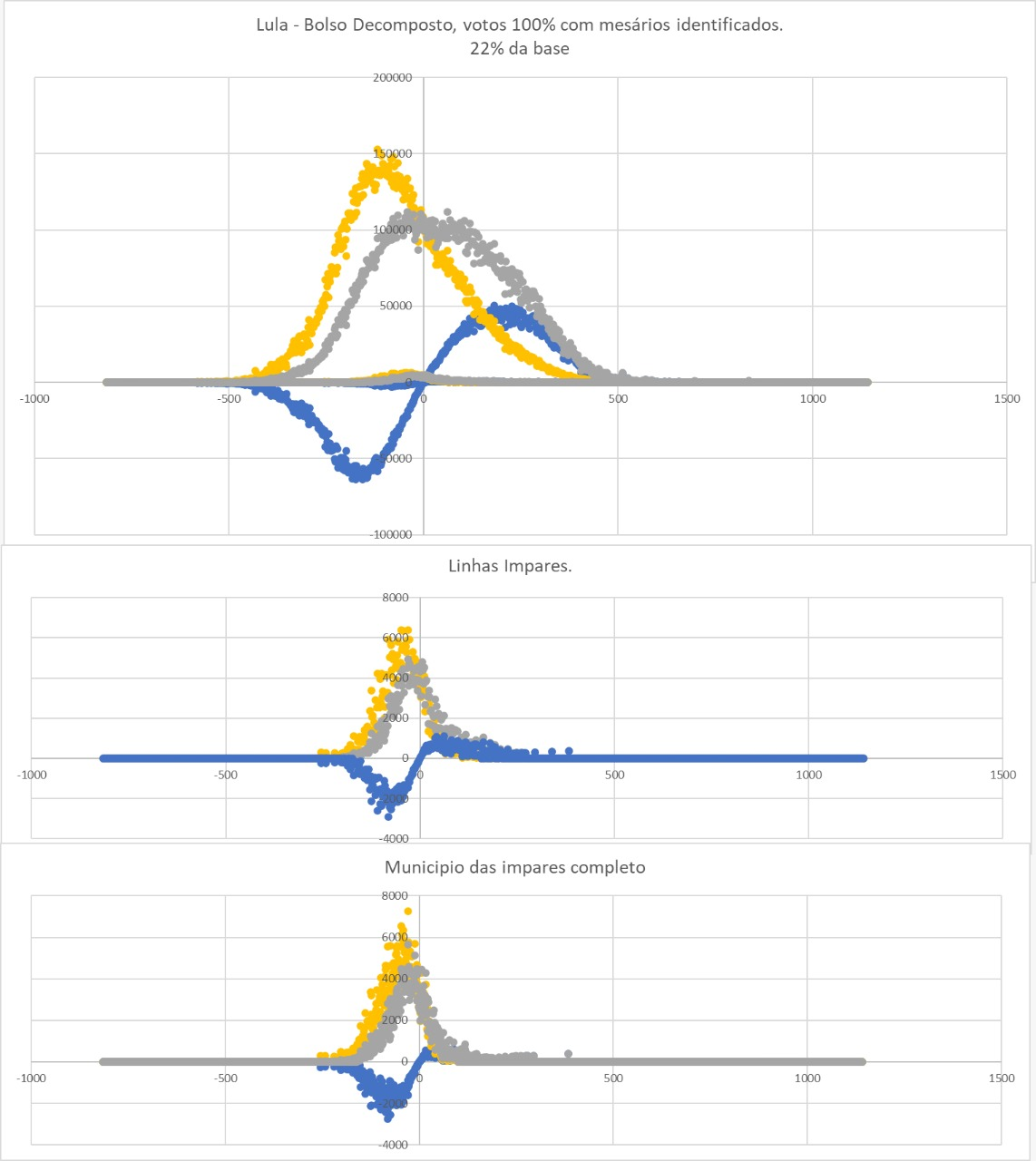
Podemos observar acima que por algum motivo urnas com a dupla falha são **96% maiores** do que urnas em que a falha não ocorreu.

Para aprofundar a análise sobre estes três grupos sobre a mesma base de boletins de urnas separados nos grupos acima em relação a uma curva normal os dados foram agrupados pelo resultado da subtração entre o voto 13 e o voto 22 como x e sumarizados os votos de Lula e Bolsonaro, as curvas resultantes são as seguintes:



Em laranja a quantidade de votos com dupla falha, podemos observar que ela não escalou na mesma proporção que os votos apurados nas urnas, o que por si só já demonstra uma distribuição exclusiva aparentemente mais relacionada a distribuição espacial do que a ocorrência de votos, possivelmente descartando o mesmo como falha, dado que como falha deveria ser mais frequente quanto mais um fenômeno é observado.

Ao obter a mesma curva para a base de dados de urnas **ZERO** foi possível observar:



No primeiro gráfico os votos de Lula e Bolsonaro aparentemente encaixam em normais, mas foi possível observar que os resultados pares e resultados ímpares tinham tendências diferentes perfazendo normais diferentes, a partir desta observação foi possível estratificar mais um público:

|  |
| --- |
| Exemplo de observações de dados ímpares e pares em tendências diferentes: |
|  |

Sumarizando a situação até o momento:

1 – As urnas foram separadas da ocorrência ou não de duas determinadas linhas no log.

2 – As urnas que não tinham esta linha no log ao sumarizar seus dados pelo RESULTADO de Votos 13 – Votos 22 apresentou DUAS NORMAIS DIFERENTES uma considerando os números PARES e outra considerando os números ímpares que daqui em diante chamaremos de **IMPAR.**

Então sem o encadeamento positivo das duas variáveis as seguintes observações não seriam possíveis:

Os votos das urnas ímpares cobrem todos os estados da federação e, totalizando todos os votos e agrupando as duas massas pela cidade e comparando o total de urnas temos que:

|  |
| --- |
|  |
| Urnas agrupadas por cidade com maior cobertura entre o público **TODOS vs ZERO** e **TODOS vs ÍMPAR** |

Podemos observar que as urnas ÍMPAR em uma grande frequência cobrem 50%+ das urnas de um determinado município e que com uma frequência muito grande as urnas **ZERO** cobrem 100% das urnas observadas.

Essa observação me levou a cogitar se não seriam propositalmente agrupadas para observar o processamento das mesmas, o que me levou a analisar o modelo das urnas conforme base urnas para as cidade sumarizadas por modelo:



O que nos levou a observar que os modelos de urnas para cada cidade são quase sempre exclusivos em comparação ao resto da massa em que foram empregadas um número misto de urnas.

Como engenheiro de sistemas me parece a implementação de uma massa de testes para um piloto controlado facilmente observável através de gráficos.

Surpreende, no entanto, a necessidade de manipular o resultado de mais de 101 mil urnas em que as mensagens de LOG objetivo do estudo pode potencialmente ter sido desativado propositalmente, assim como essa manipulação necessitar de código fonte específico que pode ser implementado como parâmetro ou integrado como um novo código fonte, o que faz mais sentido, dado que como estamos falando de hardware, uma massa de testes desse tipo deveria envolver inclusive o próprio código fonte.

A implementação disto aparentemente foi feita sem camadas de controle e até forjando dados, estes mesmos artifícios podem ter sido utilizados por um agente externo com um simples acesso ao código fonte e conhecimento dos procedimentos e das chaves, manipulando o resultado da própria eleição.

## Distribuição de falhas:

Voltando ao objetivo original do estudo, fizemos análise de coberturas por categoria, estruturei a nossa análise com o objetivo de comparar através da fórmula de correlação os diversos municípios do Brasil e, como fator de correlação o voto dado, desta forma os dados foram estratificados em cd\_municipio e nr\_votável do boletim de urna dos dados informados pelos dados abertos do TSE, resultando em 12986 categorias a serem comparadas nas análises seguintes.

### Cobertura de base.

A primeira análise é uma análise de cobertura das categorias das bases conforme evoluem as quantidades de “dupla falha” nas seções:



Nas colunas:

TODOS: todos os votos

ZERO: seções **sem** votos com “dupla falha”.

>N : seções **com** mais que N votos com “dupla falha”, estratificadas conforme uma sequência numérica Fibonacci a partir de 8 com o objetivo de obter números mais significativos.

Podemos observar que com **ZERO** e com **>13** votos com “dupla falha” temos **20%** ou mais dos votos o que já tornaria uma amostra válida pela regra de Paretto, considerando uma divisão aleatória, em ambos extratos temos mais de 92% das categorias preenchidas.

Correlação entre os públicos estratificado por quantidade de defeitos

Se fossemos considerar falha simples e aleatória de procedimento, considerando uma cobertura de municípios x voto de 92% em uma classificação de 12986 categorias para estratificar os dados para identificar se as bases não mudam a correlação com a ocorrência de “dupla falha” fizermos uma correlação entre grupos da seguinte forma:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Eleição Completa (TODOS) | vs | Fração da Eleição sem dupla falha (ZERO) | vs | Outras Frações com N votos com dupla falha ou mais (> N). |



Foi utilizada a fórmula de correlação CORREL(AmostraA, AmostraB) do excel, considerando as 12986 categorias da amostra TODOS preenchidas zeros os dados não encontrados em outras amostras.

Por padrões estatísticos vamos utilizar o grau de confiança de 95% como padrão para as afirmações do resultado das nossas estatísticas, desta forma podemos afirmar:

**1:** **pertencem** a um público compatível a amostra que contém **TODOS** os votos e **ZERO** votos com “dupla falha” (95,3 > 95 é verdadeiro).

**2:** **pertencem** a um público compatível a amostra que contém **TODOS** os votos e **>0** votos com “dupla falha” (99,6 > 95 é verdadeiro).

**3:** **não pertencem** a um público compatível amostras com **ZERO** votos com “dupla falha” e **>0** votos com “dupla falha” (92,5 > 95 é falso).

E considerando que o fato da cobertura de 20% ser um fator relevante na correlação, a mesma amostra com **>13** votos com dupla falha apresentam falsa correlação com o **TODOS** (79,4 > 95 é falso) os votos e com a amostra de **ZERO** (61,9 > 95 é falso) votos com “dupla falha”, assim:

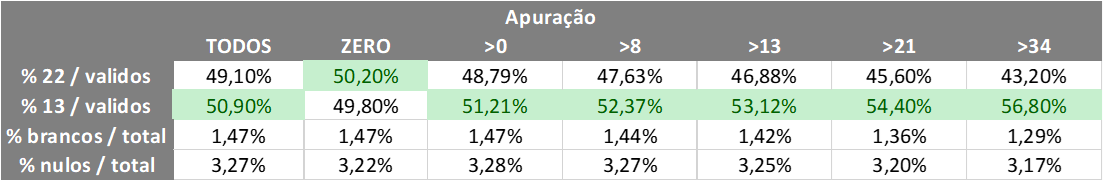
**4:** **não pertence** a um publico compatível a amostra com **>13** votos com dupla falha incompatível com a mostra que contém **TODOS** os votos e com a amostra que contem **ZERO** votos com dupla falha.

Desta a afirmativa de que as amostras que a “dupla falha” acima descrita tenha ocorrência puramente aleatória é falsa com grau de certeza de 95%.

### Resultado da apuração do segundo turno nas amostras:

Ok, a comparação entra amostras mostra que a dupla falha não é aleatória, e, como depende da operação de um usuário pode indicar ação intencional e não intencional, a ação não intencional tende a ser aleatória, já a ação intencional não é aleatória e visa atingir um determinado resultado.

Vamos ver qual é o impacto dos votos com “dupla falha” sobre os votos dos dois candidatos:



É possível ver que quanto maior a concentração de duplas falhas uma urna tem maior o resultado percentual para um dos candidatos, observamos também que os votos brancos e nulos têm uma distorção negativa, começando em um percentual que dá a vitória ao candidato com menos votos no resultado que considera o público como um todo.

O que nos leva a não descartar a intencionalidade da dupla falha.

### Resultado de uma análise de falha sigma sobre e o resultado numérica quantitativa sobre o resultado.

Vamos considerar a modelagem de defeito industrial como padronizado pela Fundação Seis Sigma para processos industriais no levantamento inicial de oportunidades do fluxo DMAIC:

Em uma abordagem pragmática considerando que a formula de:

|  |  |
| --- | --- |
| **Fórmula DPMO** | **Tabela DPMO x Sigma** |
|  | nivel sigma |

**Número de defeitos:** Cada defeito identificado na verdade consiste de dois defeitos: um defeito no reconhecimento da biometria do eleitor, e um segundo defeito na biometria do mesário.

**Número de oportunidades:** Duas oportunidades por peça.

**Número de peças:** Número Total de Votos.

Estou chamando de pragmático pois o número de defeitos poderia ser ainda maior se considerarmos a falha única que seria a falha apenas da biometria do eleitor, nestes casos uma vez que estamos considerando que ambas as observações estão sendo feitos para ambas as leituras (mesário + eleitor) o tempo todo o que não é um fato, mas que otimizaria o número apresentando um DPMO menor e consequentemente o sigma maior, mesmo assim observamos:



O padrão de um processo otimizado em sigma é de **3** para um processo industrial bem construído, entretanto podemos observar que considerando a amostra integral, o sigma está em 2,4 o que é abaixo do ideal (sigma, quanto maior melhor) e, se considerar apenas a massa em que ocorreu pelo menos uma vez por urna, **2,** o que é quase inaceitável em um processo industrial, especialmente se consideramos que o novo padrão da indústria tende a ser o 6.

### Efeito potencial sobre o resultado da eleição:



Montando apenas os dados quantitativos podemos observar que: o número da diferença entre os candidatos sempre é maior que o número de defeitos independente da quantidade de erros por urna, a diferença entre os votos apurados para o candidato 13 e candidato 22, é positiva para 22 com **ZERO** duplas falhas.

## Conclusão:

Em uma análise técnica podemos observar que considerando apenas as urnas que tiveram erros catastróficos, que seria o erro no reconhecimento biométrico simultâneo do eleitor e do mesário, sem a qual não é possível avaliar a lisura completa da votação, é muito abaixo de um padrão industrialmente aceito, calculado à 2.4 Sigma, considerando ser este um projeto moderno de processo está bastante fora do padrão, é possível afirmar ainda que o resultado da votação está sendo distorcido para um dos candidatos na proporção da existência da ocorrência do defeito na votação da urna, criando públicos entre urnas que não se correlacionam considerando a estratificação por quantidade de defeitos, reforçando a impressão de que uma mudança relevante no resultado teve a origem em um defeito na votação.

A análise quantitativa indica que, se descartadas as urnas com defeito catastrófico, o resultado final da eleição seria mudado drasticamente dando a vitória a outro candidato.