

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE MATEMÁTICA
BACHARELADO EM ESTATÍSTICA

MARIANA DE FREITAS PINTAR

ELEIÇÕES GERAIS 2014: FRAUDE OU NÃO?

UBERLÂNDIA
JUNHO DE 2016

MARIANA DE FREITAS PINTAR

ELEIÇÕES GERAIS 2014: FRAUDE OU NÃO?

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação de Bacharelado em Estatística da Faculdade de Matemática da UFU como requisito parcial para obtenção do título de bacharel em Estatística.

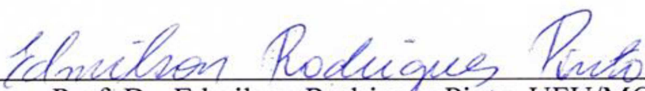
Orientadora: Prof.^a Dr.^a Priscila Neves Faria

UBERLÂNDIA
JUNHO DE 2016

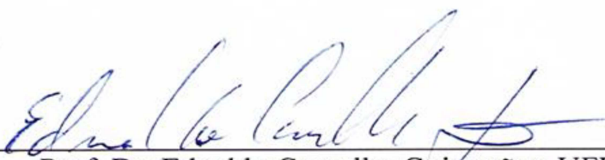
ELEIÇÕES GERAIS 2014: FRAUDE OU NÃO?

Monografia aprovada para a obtenção do título de
Bacharel no Curso de Graduação em Estatística
da Universidade Federal de Uberlândia (MG)
pela banca examinadora formada por:

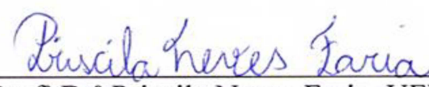
Uberlândia, 28 de junho de 2016



Prof. Dr. Edmilson Rodrigues Pinto, UFU/MG
Membro



Prof. Dr. Ednaldo Carvalho Guimarães, UFU/MG
Membro



Profª Drª Priscila Neves Faria, UFU/MG
Orientador

AGRADECIMENTOS

À minha orientadora, Doutora Priscila Neves Faria, por toda compreensão, tempo, dedicação, encorajamento e apoio que me possibilitou concluir não só este trabalho, mas muitos outros que realizamos.

À minha família, que me proporcionaram crescimento e equilíbrio profissional, intelectual e espiritual. Por tudo o que me ensinaram e continuam ensinando até hoje.

Ao meu companheiro, Katon Oliveira, que esteve presente em todos os momentos de minha graduação alegrando e me incentivando sempre quando necessário.

Aos meus amigos, que tornaram a graduação uma aprendizagem prazerosa e balanceada.

Aos meus professores, que me instruíram e se dedicaram pelo ensino abrindo portas para meu futuro profissional.

SUMÁRIO

1. Introdução.....	8
2. Objetivos	11
2.1. Geral.....	11
2.2. Específicos	11
3. Referencial Teórico	12
3.1. A lei de Newcomb-Benford	12
3.2. O Uso da Lei NB em Eleições	14
3.3. As Eleições Brasileiras	16
3.4. Testes Estatísticos	19
4. Metodologia da Pesquisa.....	21
5. Resultados e Discussões.....	31
6. Conclusões	38
REFERÊNCIAS	39
APÊNDICE	44

RESUMO

Frequentes movimentos e discussões questionam a eficiência do processo de apuração eleitoral e os resultados do segundo turno das eleições brasileiras de 2014, principalmente tratando-se da escolha para presidente. O presente trabalho pretende avaliar a confiabilidade dos resultados eleitorais para presidência do Brasil pautada em um método estatístico chamado Lei de Newcomb-Benford, ou Lei do Primeiro Dígito, que modela uma probabilidade para os dígitos de uma amostra. Dada essa peculiaridade da distribuição dos dígitos de dados agregados, a Lei é amplamente aplicada para detectar indícios de fraude onde se possa haver manipulação de números, podendo ser notas fiscais, números apresentados em relatórios financeiros ou, como proposto, resultados eleitorais. Foi possível concluir que o método é um ferramenta eficaz de apoio a auditoria e controle a adulteração. Pela análise estatística dos dados constatou-se a adequação dos números de votos para os candidatos Dilma Roussef e Aécio Neves e para votos nulos, entretanto, para os votos em branco, os dados não se aderiram à lei.

Palavras-chave: Eleições brasileiras de 2014, Lei de Newcomb-Benford, Lei contra Fraude.

ABSTRACT

Frequent movements and discussions question the efficiency of the electoral calculation's process and the second round's results of Brazilian elections in 2014. Mostly when it is about the presidential election. This study intends to evaluate the reliability of the election results for presidency through a statistic method called the Newcomb-Benford law or the First Digit law, which delineates a probability for the digits of a sample. Taking into consideration the peculiarity of the distribution of digits of aggregate data, this law is widely applied to detect evidences of fraud where there may be manipulation of numbers, such as in invoices, numbers presented in financial reports or, as proposed, election results. It was possible to conclude that the method is an effective support tool to auditing and controlling the adulteration. From the statistic analysis of the data, it was verified the suitability of the numbers of votes for the candidates Dilma Rousseff and Aécio Neves and also for null votes. However, in relation to blank votes, the data did not adhere to the law.

Key-words: Brazilian election 2014, Newcomb-Benford's Law, law against fraud.

1. Introdução

É imprescindível evidenciar que para delimitar uma amostragem, o auditor deve utilizar os conhecimentos advindos das Ciências Matemáticas e Estatísticas, com a finalidade de que os riscos de controle, de testes de detalhamentos e incertezas sejam controlados e minimizados (DINIZ et al., 2006).

Certamente, no atual ambiente de variações políticas, crises econômicas, escândalos e as incertezas que a administração pública brasileira vem enfrentando, ferramentas e estudos estatísticos podem oferecer uma metodologia científica indispensável no processo decisório e transparência fidedigna dos resultados gerais das eleições presidenciais brasileiras (NIGRINI, 2000).

Em todo o mundo, sejam nas novas ou velhas democracias, as alegações de fraudes eleitorais sucedem com ampla frequência (LEHOUCQ, 2003). A eleição presidencial de 1976 nos Estados Unidos é um exemplo, quando o republicano Rutherford Hayes foi empossado no cargo, embora o governador Samuel Tilden houvesse apresentado a maioria dos votos populares. Outro exemplo ocorreu na Rússia, nas eleições presidenciais de 2008, quando Dmitry Medvedev, com apoio de Vladimir Putin, tomou posse da presidência mesmo na afluência de crescentes tendências autoritárias, o que gerou grandes ondas de protestos e denúncias de fraude.

Não diferentemente dos outros países, em 1982 ocorreu uma tentativa de fraude nas eleições para governador do estado do Rio de Janeiro no Brasil, evento conhecido como "Caso Proconsult", cuja empresa contratada para realizar a apuração dos resultados transferia votos do candidato Leonel Brizola para o candidato Moreira Franco, apoiado pelos militares.

O Brasil foi pioneiro no uso de urnas eletrônicas nos anos 2000, contudo, até hoje, mantém o mesmo tipo de equipamento, que segundo o site do Tribunal Superior Eleitoral, foi implantado a fim de minimizar os riscos à proteção e garantir a integridade do conteúdo criptografado.

Entretanto, especialistas em segurança de dados garantem que o sistema nacional é ultrapassado e não confiável. O modelo utilizado no país, chamado primeira geração, é a primeira versão de três. Na versão adotada, a urna conta os votos eletronicamente, o que torna a apuração rápida, porém, não permite a verificação pelo eleitor, nem recontagem dos sufrágios por não haver documento impresso do voto.

Após as eleições, houve uma série de denúncias de caráter geral com evidências aparentes (antes de uma auditoria específica), que chegaram à coordenação do partido efetivo.

Uma delas referia-se ao desvio de votos nas urnas: o modelo das urnas brasileiras permitiria inserir softwares maliciosos para desviar votos durante seu registro ou apuração (HURSTI, 2006, FELDMAN et al. 2007). Outra denúncia foi o desvio de votos na transmissão e na totalização dos votos: a percepção de fraude foi agravada por ter se tornado público que um grupo pequeno de pessoas sob coordenação da STI/TSE (Secretaria de Tecnologia da Informação do Tribunal Superior Eleitoral) teve acesso em ambiente fechado aos dados parciais da totalização antes da divulgação oficial (FILHO et al, 2014).

Atualmente, o Brasil é o único país a usar este modelo de urna, países como Alemanha, Irlanda, Inglaterra e Paraguai testaram e abandonaram o sistema por falta de transparência ou falta de confiabilidade, e com isso, a maioria dos países utiliza o modelo de segunda geração que permite ao eleitor checar os candidatos escolhidos além de emitir um comprovante em papel com o voto.

As últimas eleições gerais brasileiras foram realizadas nos dias 5 e 26 de outubro de 2014 referentes ao primeiro e segundo turno, respectivamente. Na ocasião, os eleitores votaram em presidente da República, governador, senador e deputados federal e estadual.

Ao longo dos anos de 2014 e 2015 houve uma série de protestos encabeçados pelos movimentos "Vem pra rua", "Movimento Brasil Livre", etc que se caracterizavam pela pauta de anticorrupção e democracia e espelharam a insatisfação dos cidadãos quanto ao cenário nacional e seus representantes políticos.

Por haver muita discussão sobre os resultados dos votos, é de grande importância, mediante um procedimento, verificar algum tipo de fraude, intencional ou não, na coleta de dados. No processo de apuração dos resultados, as informações são criptografadas e transmitidas computacionalmente, tal situação aumenta a chance de haver alterações nos valores devido a erros ocasionados durante o procedimento.

O Tribunal Superior Eleitoral (TSE) possui um portal oficial online e dispõe informações referentes às quantidades de votos do primeiro e segundo turno das eleições de cada município do Brasil, acrescido dos votos em trânsito e de outros países.

De acordo com Mebane (2006), o método ideal para análise eleitoral seria aquele que não dependeria nem de suposições especiais sobre as configurações políticas específicas que disputam a eleição, nem em qualquer teoria particular sobre como a eleição foi conduzida.

A Lei de Newcomb-Benford ou "Lei do Primeiro Dígito" mostra através de observações que a frequência de números naturais em uma série histórica apresenta uma probabilidade natural, em particular, no primeiro dígito (HOESCHL; BUENO, 2016). Dados de população, áreas de rios e estatísticas de beisebol apresentam conformidade com a lei, com

a mesma natureza, dados de contagem de votos, possuem as mesmas propriedades e também deveriam seguir esta lei.

À vista disso, o presente trabalho propõe um estudo para avaliar a confiabilidade das contagens dos votos utilizando o método estatístico baseado na Lei de Newcomb-Benford (LNB), com intuito de testar a aderência dos dados e assim, esclarecer possíveis irregularidades nos resultados das eleições.

2. Objetivos

2.1. Geral

Haja vista o problema de pesquisa sugerido pretende-se perscrutar, como objetivo geral deste trabalho, a veracidade, consistência e compreensão dos resultados à presidência do segundo turno das Eleições Gerais de 2014 verificando observações discrepantes baseando-se no padrão determinado pela Lei de Newcomb-Benford.

2.2. Específicos

Para atender o propósito desta pesquisa, seguem os seguintes objetivos específicos:

- 1) Verificar a aplicabilidade e metodologia da Lei de Newcomb-Benford aos dados eleitorais das eleições presidenciais brasileiras de 2014;
- 2) Estipular as hipóteses estatísticas, pressupostos e testes adequados baseados na LNB a dados de eleições;
- 3) Avaliar os resultados identificando possíveis valores incoerentes ou anômalos nas eleições presidenciais brasileiras de 2014.

3. Referencial Teórico

3.1. A lei de Newcomb-Benford

A lei de Newcomb-Benford explica uma distribuição anômala dos números inteiros de 1 a 9. O fenômeno foi observado pela primeira vez em 1881 pelo matemático e astrônomo americano Simon Newcomb, que publicou um artigo no *American Journal of Mathematics* descrevendo suas observações de que as primeiras páginas das tábuas de logaritmos eram mais manuseadas e desgastadas, enquanto que as últimas páginas apresentavam-se menos desgastadas. Isto é, as pessoas iam muito mais à busca dos valores dos logaritmos que começavam por 1 do que aqueles que começavam com o dígito 9 (HILL, 1998, 1999).

Porém, foi somente em 1938, que o físico Frank Benford chegou à mesma conclusão e publicou um artigo descrevendo o mesmo acontecimento que observou em mais de 20 mil dados analisados em diferentes amostras de diversas fontes, como áreas de rios, estatísticas de beisebol, números de endereços, número de artigos de uma revista, entre outros (HILL, 1996, 1998, 1999).

A distribuição de Benford se dá nos ditos dígitos significativos, isto é, nos dígitos à extrema esquerda dos valores, com exceção do zero, independentemente do número de algarismos de cada valor da amostra. Portanto, considerando o número 52.168, o número 5 é o primeiro dígito, o 2 é o segundo dígito, e assim sucessivamente até o número 8 que é o quinto dígito.

O professor Theodore P. Hill usualmente aplica um experimento para seus alunos de matemática. Parte deles deve lançar uma moeda 200 vezes e registrar fielmente o resultado, enquanto a outra parte deve apenas simular a jogada e inventar os 200 arremessos. Posteriormente, simplesmente por olhar os resultados, o professor consegue apontar quais foram os que fraudaram os lançamentos. Hill explica que a maioria das pessoas é incapaz de definir quais são as reais probabilidades de um evento como este e, portanto, não consegue inventar dados convincentes (BROWNE, 1998).

Benford constatou que a probabilidade dos números naturais de 1 a 9 nem sempre era de 11,11% ($1/9$), como esperado intuitivamente, mas que o número 1 tinha 30% de probabilidade de ocorrer, o número 2 tinha 17,06%, e em contraste, o número 9 aparecia 4,6% das vezes, formando uma curva logarítmica decrescente. (NIGRINI, 1996; HILL, 1998, 1999).

Benford (1938) conta em sua publicação "*The Law of Anomalous Numbers*" que o método consiste em selecionar qualquer tabulação de dados que não seja muito restrita no

intervalo numérico, ou condicionada de algum modo, e fazer uma contagem das vezes que os números naturais 1, 2, 3, ..., 9 ocorrem como primeiros algarismos, e que, se o número zero ocorrer antes do primeiro número natural, ele deve ser ignorado.

Embora muitos dados sigam a distribuição de Benford, existem também muitos exemplos em que a lei não pode ser atribuída. Listas telefônicas de uma região, por exemplo, geralmente começam sempre com um mesmo conjunto de números, e assim, não se aderem à lei (HILL, 1995). Teixeira e Kira (2016) ilustram que outro conjunto de dados que não segue a lei para o primeiro dígito é a sequência de números primos, cuja frequência dos algarismos na primeira posição é praticamente a mesma. Outro exemplo é para avaliar números de CPF ou RG, pois não são números gerados de forma natural.

Por outro lado, a gama de casos e ocorrências em que é possível aplicar a lei é vasta: populações e áreas de países, comprimentos de rios, notas fiscais, etc (ROUSSEAU, 2012). Um exemplo bem comum é o número de páginas dos livros de uma biblioteca. Considerando que todas as páginas estejam enumeradas, ao escolher qualquer livro e abrir uma página aleatória, há 30% de probabilidade (em vez de 11,11%) de que o primeiro algarismo da página escolhida seja 1, 17% de probabilidade de que o primeiro algarismo da página seja 2 e que os algarismos sucessivos apresentem probabilidade decrescente (MELÃO, 2014).

Pesquisas no campo da teoria das probabilidades, tais como Hill (1998), Pinkham (1961) e Raimi (1969), mostram que a Lei de Newcomb-Benford aplica-se ao conjunto de dados que tem as seguintes propriedades:

- a) É invariante escalar, ou seja, multiplicando-se os valores de uma amostra por uma constante, a distribuição não se altera;
- b) Advém de uma escolha a partir de uma variedade de diferentes fontes.

Hill (1998) constata ainda que a distribuição que se baseia na Lei de Newcomb-Benford possui a característica de base invariante, ou seja, a mudança de base da função logarítmica não afeta a distribuição dos dígitos.

Este resultado é obtido de uma análise mais rigorosa da Teoria do Limite Central na forma de teoremas para a mantissa de variáveis aleatória sobre o efeito da multiplicação. Neste sentido, quando o número de variáveis cresce, a função densidade tende a distribuição logarítmica. Hill (1996) demonstrou rigorosamente que a distribuição da distribuição obtida a partir de amostras aleatórias advindas de uma variedade de diferentes distribuições é a distribuição de Newcomb Benford.

O resultado investigado por Benford não define apenas uma distribuição para os primeiros dígitos, mas uma distribuição para todos os dígitos significativos de um número. Mais formalmente, um conjunto de números que siga a Lei de Benford teria a mantissa de seus logaritmos uniformemente distribuída. Um teste para verificar tal ocorrência, pode também ser usado para testar a conformidade dos dados com a Lei de NB, como proposto por Nigrini (2012).

Raimi (1969) aponta que os dados precisam somente se aproximar de uma sequência geométrica. Nigrini (2012) concorda e ainda propõe o uso de uma regressão linear para testar se as mantissas ordenadas formam uma linha reta.

Alexander (2009) foi além e propôs um rigoroso teste de conformidade com a Lei de Benford, chamado Mantissa Arc Test (MA), que consiste em verificar se as mantissas se distribuem uniformemente ao longo do intervalo $[0,1]$. Um ajuste para este teste, quando tratado para altas populações, é sugerido por Nigrini (2012), pois o autor diz que quando analisados um grande conjunto de dados, o teste sinaliza conformidade facilmente.

Mebane (2006) afirma que o teste dos segundos dígitos é um método ideal para ser aplicado em dados eleitorais como base para auditorias rotineiras, ressaltando que o teste não é suscetível no cenário de urnas, mas que funciona muito bem em zonas eleitorais ou municípios. Mebane (2006) ilustrou que o teste é sensível a muitos padrões de manipulação como, por exemplo, a falsificação de cédulas, e que o método, unido a uma ferramenta de recontagem, seria um recurso de excelente capacidade para sinalizar fraudes.

Hill (1995) declara que é uma surpresa que os dígitos não sejam igualmente prováveis, mas reivindicar o estudo desta lei e descrever sua distribuição é de fato impressionante. Nigrini (1996) menciona que existem inúmeras aplicações da Lei de Benford que vão desde dados de hidrologia, preços das ações, ciência da computação na análise de erros de arredondamento até contabilidade para detectar a fraude fiscal.

3.2. O Uso da Lei NB em Eleições

Walter Mebane, um dos autores que mais estudaram a aplicabilidade da lei de Newcomb-Benford em eleições, analisou dados eleitorais de vários países, incluindo os Estados Unidos, Rússia e México. Em 2009, o cientista avaliou cada uma das urnas da eleição presidencial do Irã e constatou que os resultados haviam sido fraudados a favor de um dos candidatos à presidência (BELLOS, 2015).

Para fraudar uma eleição não é mais necessário lançar milhares de cédulas nos rios ou contratar pessoas para forjar repetidos votos ilícitos. Em vez disso, nos dias de hoje, apenas é necessário ter acesso a uma porta de entrada e algumas linhas de código de computador. Detectar tais manipulações é um difícil e urgente problema (MEBANE, 2006).

Pericchi e Torres (2004) propuseram usar a distribuição do segundo dígito da Lei de Benford para testar votos fraudulentos. Mebane (2006) diz que a lei de Benford é mais bem entendida como um método que indica para onde investigações poderiam ser direcionadas.

O resultado de Hill (1995) declara que a lei é obtida quando a variável medida é o resultado de uma "mistura imparcial de distribuições", e, portanto, foi proposto que teste de ajuste desta lei é realizado para detectar a possibilidade de alteração de números. Ressalta-se que, isto não é uma prova de que os dados foram adulterados, mas é um indício de alteração.

Quando se trata de dados de eleições, é preferível utilizar o segundo dígito significativo, pois as urnas têm números pré-especificados e projetados de eleitores que podem afetar a análise do primeiro dígito, pois já se presumiria as quantidades e os valores máximos para cada urna (PERICCHI e TORRES, 2004). Mebane (2006) ainda destaca que dados de contagem de votos é um processo que se assemelha fortemente à distribuição da Lei de Benford quando estudados os segundos dígitos.

Pericchi e Torres (2004) em seu trabalho sobre um referendo revogatório do mandato do presidente Hugo Chávez, na Venezuela, mostraram que os votos eletrônicos brutos não aderiram à lei, e após a apuração e correção computacional, os votos aderiram à lei, o que levou ao indício de fraude nas eleições.

Nos Estados Unidos, várias grandes empresas e órgãos Federais utilizam softwares para detecção de fraudes que possuem um sistema baseado na Lei de Newcomb-Benford como instrumento de apoio à auditoria (BROWNE, 1998).

Para saber se as contagens de votos satisfazem a lei é usado o teste qui-quadrado de aderência para verificar se a frequência observada dos segundos dígitos dos dados coincidem com as proporções dos segundos dígitos da lei de Benford. Se os valores observados do teste são maiores que os valores críticos específicos identificados pela teoria estatística, a conclusão é que a distribuição dos segundos dígitos não caracteriza as contagens referentes aos votos (MEBANE, 2007).

3.3. As Eleições Brasileiras

A insatisfação dos cidadãos brasileiros com a política e o descaso por parte da administração pública é um cenário que se transcreve há mais de 100 anos. O eleitorado do ano de 2014 não foi o primeiro que surgiram dúvidas e incertezas quanto aos resultados e eficiência do processo eleitoral eletrônico.

A era republicana no Brasil teve início em 1889, com a proclamação da República pelo Marechal Deodoro da Fonseca, e vigora até os dias de hoje. Nesses anos, o país passou por importantes mudanças de governo, inclusive um período de ditadura militar.

A primeira inovação eleitoral trazida pela República foi a eliminação do "voto censitário" que definia renda mínima para qualificar o eleitor. Essa inovação foi promulgada por Deodoro, sendo considerada a primeira Lei Eleitoral da República que tratava da qualificação dos eleitores. A Constituição de 1891 implantou o voto universal para os cidadãos, instituiu o presidencialismo e o voto aberto para eleger o presidente e vice-presidente por maioria absoluta dos votos, entretanto, mulheres, analfabetos e militares de baixa patente ficavam de fora do sufrágio.

Após a Proclamação da República, a política no Brasil foi marcada pelo governo de presidentes civis ligados ao setor agrário. De 1894 a 1930 o poder era alternado entre políticos dos dois estados mais ricos da época, Minas Gerais e São Paulo, e, por serem fazendeiros de elite, favoreciam sempre o setor agrícola.

Os coronéis, indivíduos de grande poder econômico, para garantir a eleição dos candidatos que apoiavam, ofereciam regalias aos eleitores e verificavam o preenchimento das cédulas, já que o voto não era secreto. Roberto Catelli Jr. (2005), em seu artigo *A República do Voto*, menciona fraudes na República da Oligarquia, quando, na apuração dos votos, cédulas e atas eram adulteradas além de intervenções quando o candidato de oposição era eleito.

A quebra desta troca de governo provocou a Revolução de 1930 e através de um golpe de estado, Getúlio Vargas assumiu a presidência do Brasil introduzindo a Era Vargas de 1930 a 1945. Foi somente no início desse período, com a reforma eleitoral, que o direito ao voto foi concedido às mulheres e a idade mínima do eleitor abaixou para 18 anos de idade.

Após a Era Vargas, a Assembleia Constituinte de 1946 criou a quinta constituição brasileira, que estabeleceu os poderes Executivo, Legislativo e Judiciário. Outra singularidade foi que se deliberou que os eleitores seriam todos os brasileiros alfabetizados

de ambos os sexos acima de 18 anos e o voto deveria ser secreto. A autonomia de cada estado foi ampliada, sendo os governadores e o presidente do país eleitos por voto direto.

Em 1964, com o golpe militar, outorgou-se o direito de cassar o mandato de deputados e senadores e suspenderam-se as garantias constitucionais quando julgassem necessário. A Constituição foi modificada, partidos políticos foram dissolvidos e, em 1967, a nova Carta da Constituição estabeleceu que as eleições para presidência da República fossem indiretas, e assim, a população não possuía o direito ao voto e os representantes eram escolhidos por uma assembleia fechada.

Nas eleições gerais de 1982 ocorreu uma tentativa de fraude na apuração dos resultados para governador do estado do Rio de Janeiro no Brasil, o evento ficou conhecido como "Caso Proconsult", empresa contratada pelo TSE para realizar a apuração dos votos. O programador do sistema e ex-oficial do exército transferia votos de Leonel Brizola para Moreira Franco, candidato apoiado pelos militares. Entretanto, Brizola, ao descobrir todo o facete, montou um esquema de apuração e acompanhamento e, juntamente com seu partido, comunicou a imprensa o acontecido provando que os resultados haviam sido manipulados. Com isso, Brizola conseguiu se eleger governador em 1982 e se reelegeu em 1990.

O ano de 1988 foi de extrema importância para a história da cidadania brasileira, quando a Constituição Federal de 1988, também conhecida como a Constituição Cidadã, foi promulgada em 5 de outubro deste ano e é considerada a mais completa, principalmente quando se refere aos direitos à cidadania dos brasileiros.

Foi somente em 1989, com o fim absoluto do militarismo, que todos os brasileiros com idade acima de 16 anos, independente de sua renda ou escolaridade, puderam votar e eleger o presidente da República. Foi neste mandato que os brasileiros presenciaram outro marco para sua história política: Fernando Collor de Melo, primeiro presidente eleito diretamente por voto popular após o militarismo, foi também o primeiro a ser julgado e condenado por Crime de Responsabilidade pelos escândalos e envolvimento na corrupção, foi, portanto, o primeiro Presidente da República a sofrer o processo de *impeachment*. O processo determinou que o ex-presidente fosse suspenso de direitos e funções públicas durante o período de oito anos.

Foi neste contexto que surgiram os "Caras-pintadas", um movimento promovido por organizações estudantis com um objetivo principal: remover o presidente do poder. Este episódio demonstrou a insatisfação do povo com a política e economia brasileira.

Os dois pleitos subsequentes foram notáveis pelo lançamento do Plano Real e pela estabilização da economia, o que levou a primeira reeleição presidencial da história política brasileira, quando o povo nomeou Fernando Henrique Cardoso presidente em 1994 e 1998.

Nesse ínterim surge o sistema eletrônico de votação, implantado a partir da eleição municipal de 1996, representou um marco histórico na evolução e tecnologia do modelo de apuração e totalização dos votos no país. Na ocasião, os municípios e capitais com mais de 200 mil eleitores receberam as primeiras urnas eletrônicas que representavam um terço das seções eleitorais. Já na eleição seguinte, em 1998, dois terços das seções eleitorais encontravam-se equipadas com essas máquinas, representando um avanço no sistema eleitoral nacional (TSE, 2014).

Este sistema veio com a intenção de conferir com maior rapidez e eliminar a interferência humana no sistema eleitoral, a automação permitiu que todas as fases do processo pudessem ser acompanhadas e auditadas por representantes de partidos, instituições e por qualquer cidadão interessado.

Oliveira (1999) relata que o voto obrigatório foi implantado no Brasil em 1932 e se transformou em uma norma constitucional em 1934 com a justificativa de ser uma necessidade garantir a presença de todos os eleitores nas eleições. Apesar de cerca de 85% dos eleitores comparecerem às urnas, o estudo de Oliveira (1999) mostra que independente da renda ou escolaridade, grande parcela dos eleitores estão desmotivados pela política. Uma maneira de medir o grau dessa "alienação política" foi proposta por Santos (1987), ele expõe que este índice pode ser entendido como a soma das abstenções com os votos nulos e brancos.

No segundo turno das eleições de 2014, embora as quantidades de votos nulos e brancos não tenham sido altas, 3,7% e 1,3%, respectivamente, o percentual de abstenções chegou a 21% dos eleitores aptos a votar, demonstrando o alto índice de desmotivação à participação dos eleitores (TSE, 2015).

De acordo com o Glossário Eleitoral do Tribunal Superior Eleitoral (TSE, 2016), o índice de abstenção eleitoral é calculado como o percentual de eleitores que, tendo direito, não se apresentam às urnas, diferentemente dos casos em que o eleitor, apresentando-se, vota em branco ou nulo. O voto em branco é aquele em que o eleitor não manifesta preferência por nenhum dos candidatos. Antigamente como o voto branco era considerado válido (isto é, era contabilizado e dado para o candidato vencedor), ele era tido como um voto de conformismo, na qual o eleitor se mostrava satisfeito com o candidato que vencesse as eleições.

Entretanto, atualmente, vigora no pleito eleitoral o princípio da maioria absoluta de votos válidos. Este princípio considera como votos válidos os que são nominais e de legenda, para os cálculos eleitorais, desconsiderando os votos em branco e os nulos.

3.4. Testes Estatísticos

Em muitas situações tem-se interesse em tomar decisões de aceitar ou rejeitar determinadas afirmações, fundamentando-se em um conjunto de evidências (BOLFARINE; SANDOVAL, 2000). Estas afirmações ou hipóteses são respostas supostas e prováveis para um problema, cuja adequação e comprovação, serão verificadas através de pesquisas e testes (LAKATOS; MARCONI, 2007).

O teste de hipótese é um instrumento poderoso para o avanço da ciência, pois auxilia o pesquisador a investigar e comprovar sua teoria a fim de expandir seus dados para amplas generalizações, isentos de valores e opiniões de indivíduos (KERLINGER, 1980).

Ao desenvolver um teste de hipóteses é necessário estabelecer duas hipóteses antagônicas: uma hipótese nula, também designada por H_0 , é a suposição de interesse tida como verdadeira e a hipótese alternativa ou H_1 , que é oposta à nula (BOLFARINE; SANDOVAL, 2000). Nestas condições, o objetivo do teste estatístico de hipóteses é fornecer uma metodologia que permita verificar se os dados amostrais trazem evidências que apoiem ou não as hipóteses formuladas (BUSSAB; MORETTIN, 2004).

Nigrini (2000) introduziu a aplicação da Lei de Newcomb-Benford fundamentada na relação com os testes de hipóteses Z-teste e Qui-quadrado de aderência, para assegurar uma melhor detecção dos desvios padrões e análise do nível e significância.

4.4.1. Teste Z

Para comparar as frequências observadas da amostra com as frequências esperadas pela distribuição, Nigrini (2000) propõe o uso do teste Z a fim de medir o grau de significância dos desvios encontrados de cada um dos dígitos analisados. Stevenson (2001) relata que os testes utilizados para mais de 20 observações, tanto para médias quanto para proporções, medem o desvio de uma amostra em relação a um valor teórico, em outras palavras, o teste evidencia se um determinado dígito aparece com maior ou menor frequência numa determinada posição em comparação com a distribuição de Newcomb-Benford. O teste Z é utilizado para duas amostras de populações com variâncias conhecidas e que convergem assintoticamente para a distribuição normal.

As hipóteses H_0 – Não há diferenças estatisticamente significativas entre as proporções observadas e esperadas, e H_1 – Há diferenças estatisticamente significativas entre as proporções, são testadas para um valor crítico de Z com um nível de significância de 5%, mais comumente utilizado (STEVENSON, 2001).

A estatística calculada usa o tamanho da amostra (n), quando a amostra é grande, a estatística Z indicará uma diferença significativa, onde, em termos práticos, a diferença é pouco significativa. A estatística é calculada para cada dígito ou combinação de dígitos, ou seja, para o teste dos dois primeiros dígitos existiria 90 estatísticas (NIGRINI, 2000).

4.4.2. Teste Qui-Quadrado (χ^2)

Referindo-se ao teste Qui-Quadrado ou χ^2 , geralmente é aplicado para testar a hipótese de uma amostra ser proveniente de uma distribuição de probabilidade que é definida em H_0 , ou seja, a distribuição de probabilidade proposta pela Lei de Newcomb-Benford (DOWNING et al., 2005).

O teste Qui-quadrado é um teste global, ou seja, analisa a amostra na sua globalidade e não individualmente, no caso, para cada dígito como a estatística Z , assim, ele mede o grau de conformidade da distribuição de probabilidade observada com a esperada segundo a lei de Newcomb-Benford (SANTOS, et al. 2003).

É observado como resultado que a hipótese nula estabelece que as populações não difiram relativamente da frequência esperada com o que realmente ocorre. Por outro lado, a hipótese alternativa estabelece que as diferenças amostrais reflitam diferenças significativas com relação ao que era esperado pela LNB (LEVIN, 1987).

4. Metodologia da Pesquisa

Neste estudo foi utilizado o banco de dados disponibilizado pelo portal oficial do Tribunal Superior Eleitoral, com os boletins de urna do segundo turno de cada estado com eleitores aptos a votar, que contou com todos os estados brasileiros e alguns estados do exterior. Os boletins de urna são arquivos das seções eleitorais que constam as informações assim como saíram da urna eletrônica, ou seja, sem o processamento do sistema de totalização.

Utilizou-se o software SPSS apenas para organização e contabilização dos dados que resultou em um banco de dados com as variáveis agregadas por zona eleitoral: UF, município, quantidade de eleitores aptos a votar, quantidade de eleitores presentes, tipo de urna (as urnas foram classificadas como "apurada" ou "não instalada"), partido, número votável (número que representava o candidato ou partido, sendo que para votos brancos e nulos, os números eram "95" e "96", respectivamente), nome do candidato e quantidade de votos por município.

Em estatística, a distribuição de probabilidade descreve a chance de uma variável assumir cada valor do domínio. A distribuição mais conhecida é a Distribuição Normal ou Gaussiana. Em um conjunto de dados distribuído normalmente, espera-se que 68% dos valores estejam a um desvio padrão da média, 95% dos valores estejam até dois desvios padrão da média e 99,7% dos valores estejam em até três desvios padrão da média. Neste aspecto, a distribuição de Newcomb-Benford permite avaliar o comportamento dos dígitos de números produzidos de forma natural e auxiliar na tomada de decisão.

A distribuição de Newcomb-Benford se dá nos ditos dígitos significativos, isto é, nos dígitos à extrema esquerda dos valores, com exceção do zero, independentemente do número de algarismos de cada valor da amostra. Portanto, considerando o número 52.168, o número 5 é o primeiro dígito, o 2 é o segundo dígito, e assim sucessivamente até o número 8 que é o quinto dígito.

Considerando-se a sucessão dos números Fibonacci, que é uma sequência numérica iniciada em 0 ou 1, na qual, cada termo subsequente corresponde à soma dos dois números anteriores, para o cálculo das frequências de cada dígito dos 7 primeiros números Fibonacci: 1, 2, 3, 5, 8, 13, 21, a frequência do algarismo 1 como primeiro dígito é 2, pois ele aparece nos números "1" e "13". A frequência para o algarismo 2 também é 2, pois aparece nos números "2" e "21", já para os demais algarismos 3, 5 e 8, a frequência para cada é 1, pois aparecem apenas uma vez como primeiro dígito (TEIXEIRA; KIRA, 2016).

A tabela que atua de suporte às disposições do estudo foi publicada por Newcomb (1881) no *American Journal of Mathematics* composta pelas probabilidades de ocorrência do

primeiro e segundo dígito significativo que pertence ao conjunto dos números inteiros de 0 até 9 (Tabela 1):

Tabela 1 - Frequências dos dígitos da Lei de Benford.

Número	1º Dígito	2º Dígito
0	-	0,11968
1	0,30103	0,11389
2	0,17609	0,10882
3	0,12494	0,10433
4	0,09691	0,10030
5	0,07918	0,09667
6	0,06695	0,09337
7	0,05799	0,09035
8	0,05115	0,08757
9	0,04576	0,08499

Como proposto por Newcomb, os primeiros dígitos dos dados, expostos na Tabela (1), seguem uma distribuição logarítmica. Durtschi, et al. (2004) relatam que Newcomb calculou que a probabilidade do primeiro dígito de um número diferente de zero conforme a Equação (1):

$$P(d) = \text{Log}_{10} \left(1 + \frac{1}{d} \right) \quad [1]$$

Onde d é o valor do primeiro dígito significativo, inteiro e não nulo, e $P(d)$ é a probabilidade de ocorrência. A Lei de Newcomb-Benford corresponde graficamente para o primeiro e segundo dígito, respectivamente nas Figuras (1) e (2):

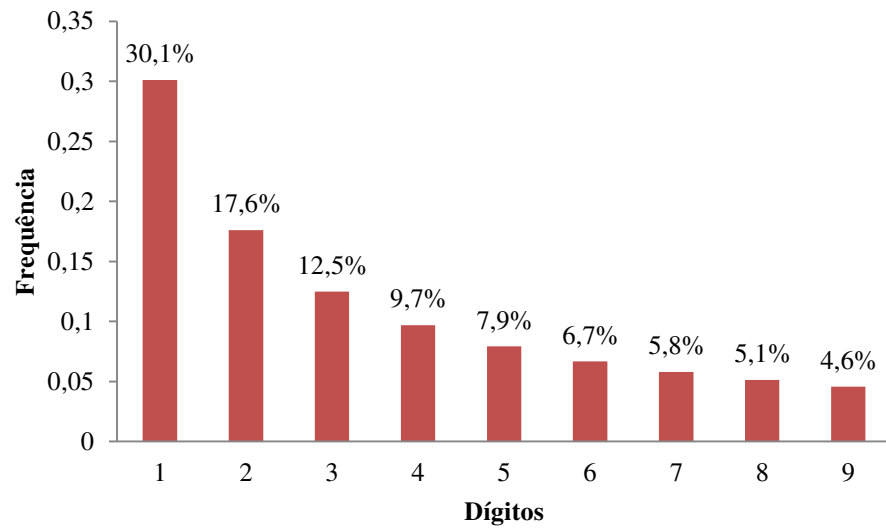


Figura 1 - Distribuição do primeiro dígito de Newcomb-Benford.

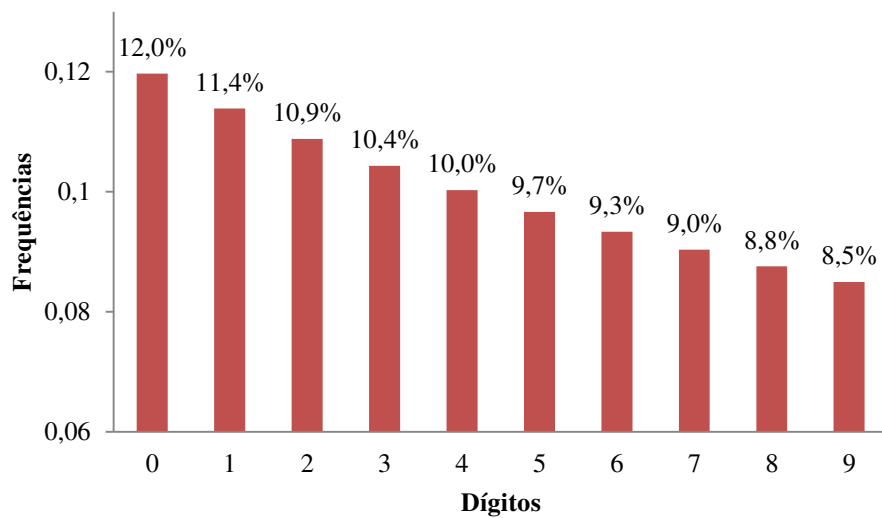


Figura 2 - Distribuição do segundo dígito de Newcomb-Benford.

Santos, et al. (2003) demonstraram em seu trabalho a distribuição originária da LNB. Suponha-se que existe uma distribuição de probabilidade universal com função densidade de probabilidade dada por $p(x)$ no conjunto dos elementos com as características de invariância escalar e origem de uma variedade de diferentes fontes, então p é invariante escalar para uma função real λ de variável k , isto é:

$$p(kx) = \lambda(k)p(x)$$

[2]

Pode-se supor então que, como $p(x)$ é função densidade de probabilidade, então:

$$\int_{-\infty}^{\infty} p(x) dx = 1 \quad [3]$$

Fazendo uma mudança de variável $kx = u$ e $dx = du/k$ na primeira integral, então:

$$\int_{-\infty}^{\infty} p(kx) dx = \int_{-\infty}^{\infty} \lambda(k) p(x) dx \quad [4]$$

$$\int_{-\infty}^{\infty} p(u) \frac{du}{k} = \int_{-\infty}^{\infty} \lambda(k) p(x) dx \quad [5]$$

$$\frac{1}{k} \int_{-\infty}^{\infty} p(u) du = \lambda(k) \int_{-\infty}^{\infty} p(x) dx \quad [6]$$

E assim, pode-se dizer que $\lambda(k) = \frac{1}{k}$.

Derivando e integrando a equação (4) com relação à k , então $\lambda'(k) = -1/k^2$ tem se:

$$[p(kx)]' = [\lambda(k)p(x)]' \quad [7]$$

$$p'(kx)x = \lambda'(k)p(x)$$

$$p'(kx)x = -\frac{1}{k^2}p(x)$$

E para $k = 1$ resulta a equação diferencial:

$$xp'(x) = -p(x) \quad [8]$$

$$\frac{p'(x)}{p(x)} = -\frac{1}{x}, \forall x \quad [9]$$

Resolvendo:

$$\int_1^x \frac{p'(t)}{p(t)} dt = - \int_1^x \frac{1}{t} dt$$

$$\ln[p(x)] = -\ln(x)$$

$$\ln[p(x)] = \ln\left(\frac{1}{x}\right)$$

Portanto,

$$p(x) = \frac{1}{x} \quad [10]$$

É solução de (8).

A função densidade de probabilidade $p(x) = 1/x$ não representa uma distribuição de probabilidade própria, mas como os fenômenos reais impõem um ponto de corte na distribuição pode-se, assim, ser considerada esta distribuição para efeito dos cálculos das probabilidades para a ocorrência do primeiro dígito. Então:

$$P(\text{primeiro dígito significativo} = d) = \frac{\int_d^{d+1} p(x) dx}{\int_1^{10} p(x) dx} = \frac{\ln(d+1) - \ln(d)}{\ln(10) - \ln(1)}$$

Para $d \in \{1, 2, \dots, 9\}$.

Limites de intenção:

$$\frac{\ln\left(\frac{d+1}{d}\right)}{\ln(10)} = \log_{10}\left(1 + \frac{1}{d}\right)$$

Berger e Hill (2014) estenderam a teoria, e, de maneira geral, encontraram a distribuição para todos os dígitos significativos. A função dessa generalização, também logarítmica, é apresentada na Equação (11):

$$P(D_i \dots D_k) = \log_{10}[1 + 1/(D_i \dots D_k)] \quad [11]$$

Foi possível encontrar também, a função de probabilidade das distribuições do segundo dígito significativo, como mostra a Equação (12) (BERGER; HILL, 2014):

$$P(X = D_{2_i}) = \sum_{k=1}^9 \text{Log}_{10}[1 + 1/(10D_{1_k} + D_{2_i})] \quad [12]$$

Em que $1 \leq D_{1_k} \leq 9; 0 \leq D_{2_i} \leq 9$;

Generalizando, para todos os inteiros positivos k , todo $d_1 \in \{1, 2, \dots, 9\}$ e todo $d_j \in \{1, 2, \dots, 9\}, j = 2, \dots, k$, a probabilidade fica como na Equação (13) (HILL, 1995).

$$P(D_1 = d_1, \dots, D_k = d_k) = \log_{10} \left[1 + \left(\sum_{i=1}^k d_i \times 10^{k-i} \right)^{-1} \right] \quad [13]$$

A aplicação da Lei de Benford é realizada através da comparação dos valores observados no estudo para o dígito em análise com relação aos valores esperados para esse mesmo dígito da distribuição de Benford. Esta comparação se faz por meio de um teste de aderência.

A exigência de que as mantissas sejam distribuídas uniformemente é a base matemática da Lei de Benford. Nigrini (2012) exemplifica a mantissa do logaritmo de 200 como nas Equações (14) e (15):

$$\log_{10}(200) = 2,30103 \quad [14]$$

$$\text{mantissa do } [\log_{10}(200)] = 0,30103 \quad [15]$$

Se os dados apresentarem uma conformidade com a lei de Benford, suas mantissas serão uniformes no intervalo de $[0,1)$. Um teste para verificar se as mantissas são uniformemente distribuídas ao longo desse intervalo pode ser usado para atestar a conformidade com a LNB (NIGRINI, 2012).

Duas condições necessárias, mas não suficientes para que as mantissas sejam uniformemente distribuídas no intervalo $[0,1)$ é que a média seja 0,50 e a variância seja $1/12$. Além disso, as mantissas ordenadas devem formar uma linha reta que vai de $-1/n$ à $(n-1)/n$, com uma inclinação de $1/n$ (CECATO, 2013).

Uma reta de regressão pode ser usada para testar se as mantissas ordenadas formam uma linha reta, onde y é o conjunto das mantissas ordenadas e x é o ranque das mantissas que vai de 1 a n . Para isso, pode-se testar o intercepto, a inclinação, e o coeficiente de determinação R^2 que devem se aproximar, respectivamente, de zero, $1/n$ e 1, para melhor conformidade (NIGRINI, 2012).

O teste proposto por Alexander (2009), Mantissa Arc Test ou teste MA, propõe que cada valor da mantissa seja plotado em uma circunferência de raio 1 e com centro de origem em (0,0), e assim, calcula-se se o centro de origem dos dados diferem significativamente do ponto (0,0). Caso seja encontrada uma diferença significativa, então os pontos não são uniformemente distribuídos no círculo e os dados não se assemelham a uma distribuição de Benford.

O teste MA procede do fato de uma circunferência com perímetro $2\pi r$. Uma vez que o raio é de tamanho 1, como indicado por Alexander (2009), o comprimento da circunferência será apenas 2π .

Para se calcular o centro de gravidade, cada número precisa ser convertido em um ponto no círculo unitário. As coordenadas x e y dos pontos são calculadas como mostram as Equações (16) e (17):

$$\text{coordenada } x = \cos(2\pi(\log(x_i) \bmod 1)) \quad [16]$$

$$\text{coordenada } y = \sin(2\pi(\log(x_i) \bmod 1)) \quad [17]$$

Onde x_i são os valores dos votos em cada zona eleitoral, $\cos$ e $\sin$ se referem às funções trigonométricas cosseno e seno. O $\log$ é calculado na base 10 e $\bmod 1$ é calculado para tomar apenas a parte fracional do $\log$ (ALEXANDER, 2009).

O centro de gravidade e suas coordenadas x e y são calculados da seguinte forma (Equações 18 e 19):

$$\text{Coordenada } x \text{ do C.G.} = \frac{\sum_{i=1}^N \cos(2\pi(\log(x_i) \bmod 1))}{N} \quad [18]$$

$$\text{Coordenada } y \text{ do C.G.} = \frac{\sum_{i=1}^N \sin(2\pi(\log(x_i) \bmod 1))}{N} \quad [19]$$

O tamanho do vetor de centro de gravidade, L^2 , e seu respectivo *valor p* são calculados conforme as equações 19 e 20:

$$L^2 = (\text{coordenada } x \text{ do CG})^2 + (\text{coordenada } y \text{ do CG})^2 \quad [20]$$

$$\text{valor } p = 1 - e^{-L^2 \times N} \quad [21]$$

Alexander (2009) comenta que a Equação (20) mostra que ao passo que o centro de gravidade se distancia do ponto (0,0), o comprimento do vetor aumenta. A Equação (21) mostra que à medida que o vetor L^2 aumenta e N tende ao infinito, o *valor p* tende a 1. As hipóteses a serem testadas são H_0 : As mantissas dos dados são uniformemente distribuídas contra H_1 : As mantissas dos dados não são uniformemente distribuídas. O *valor p* é significativo quando é menor que o valor do vetor L^2 , assim a hipótese nula é rejeitada, do contrário, um *valor p* maior que L^2 é não significativo e a hipótese nula não pode ser rejeitada. O *valor p* se baseia na distribuição qui-quadrado com 2 graus de liberdade (ALEXANDER, 2009).

Nigrini (2012) comenta que o problema do Teste MA é que basta o comprimento do vetor L^2 ser pequeno para que o teste sinalize conformidade, com um *valor p* menor do que 0,05. O autor sugere, que para altas populações, o uso de $\sqrt{N}$ ou $\sqrt[3]{N}$ no cálculo do *valor p*, para considerar o comportamento de dados reais, em que N é o tamanho da amostra utilizada.

O teste de hipótese é o instrumental metodológico que permite ao pesquisador, por exemplo, apreciar sobre a validade de expandir seus dados para amplas generalizações ou, ao contrário, verificar se esses são extremamente valiosos por diferirem do que se conhece até então a esse respeito (LAKATOS; MARCONI, 2003).

De acordo com Santos et al. (2003), dois testes estatísticos são interessantes para avaliar a conformidade de um conjunto de dados a Lei de Benford. O *teste - z* avalia sob um determinado nível de significância as diferenças entre a probabilidade observada (p_o) e a probabilidade esperada (p_e) no tocante a ocorrência dos dígitos numéricos (NASCIMENTO, et al, 2014).

A estatística Z é utilizada para medir o grau de significância entre as diferenças de probabilidade observada e esperada, associadas a cada um dos dígitos analisados. Por outras palavras, este teste evidencia se um determinado dígito aparece com maior ou menor

frequência numa determinada posição em comparação com a distribuição de Benford (Durtschi, et al., 2004). A estatística de Teste Z é calculada como se segue na Equação (22):

$$Z = \frac{|p_o - p_e|}{(p_e * (1 - p_e)/n)^{1/2}} \approx N(0,1) \quad [22]$$

Onde p_e indica a proporção esperada tendo em conta a distribuição de Benford, p_o indica a proporção observada tendo em conta a distribuição da amostra analisada, e n o número de observações.

O valor crítico de Z para um nível de significância de 5% é de 1,96, ou seja, rejeitamos H_0 sempre que o valor de Z obtido for superior a 1,96. A estatística Z apresenta duas características. A estatística calculada usa o tamanho da amostra (n) sendo que, para amostras grandes, as estatísticas Z calculadas serão mais elevadas. Para grandes amostras a estatística Z indicará uma diferença significativa, onde, em termos práticos, a diferença é pouco significativa. A estatística Z é calculada para cada dígito ou combinação de dígitos, ou seja, para o teste dos dois primeiros dígitos existiria 90 estatísticas (NIGRINI, 2000). De maneira similar, outro teste importante é o teste qui-quadrado de aderência que possui a mesma função que o teste z , porém em um âmbito global dos dígitos, ou seja, para toda a distribuição.

O teste Qui-quadrado é utilizado para estimar a probabilidade da obtenção de um valor relevante por mero acaso e não pelo fato de que existem diferenças reais entre as variáveis analisadas em comparações entre as frequências e não entre valores médios. É observado que como resultado a hipótese nula estabelece que as populações não difiram relativamente à frequência com que ocorre refere a uma característica particular. Por outro lado, a hipótese experimental estabelece que as diferenças amostrais reflitam diferenças reais na população matriz a partir da frequência relativa (LEVIN, 1987).

Recomendado para grandes amostras, o teste Qui-quadrado de aderência é útil para verificar se a distribuição das frequências observadas dos dados se ajusta a um modelo teórico pré-determinado. Caso as frequências esperadas não distingam estatisticamente das frequências observadas, pode-se inferir que a característica em estudo da população tem a distribuição pré-estabelecida, entretanto, quando já é conhecido que o fenómeno se adere à lei de Benford, se a distribuição de algum dígito não siga a lei, conclui-se que algum número teve maior ou menor frequência do que o esperado pela lei.

O teste é realizado utilizando a estatística apresentada em (23) que converge para uma distribuição Qui-quadrado com $k - 1$ graus de liberdade:

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \approx \chi_{k-1}^2 \quad [23]$$

Em que O_i é a frequência de votos observada, E_i é a frequência esperada pela distribuição de Benford e k é a quantidade total de dígitos (9 para o primeiro e 10 para os demais dígitos significativos). O valor obtido é comparado ao valor crítico da distribuição qui-quadrado, em um nível de significância estabelecido com $(k - 1)$ graus de liberdade (SIEGEL; CASTELLAN Jr, 2008). Neste estudo adotou-se o nível de significância α de 5%.

As hipóteses a serem testadas são H_0 : As frequências observadas dos votos não diferem das frequências esperadas em relação à distribuição de Benford, ou seja, a quantidade de votos tem distribuição de Benford contra H_1 : As frequências observadas diferem das frequências esperadas em relação à distribuição de Benford, ou seja, a quantidade de votos não tem distribuição de Benford. No caso de o valor calculado para a estatística qui-quadrado ser maior que o valor crítico, ao nível de 5% de significância, a hipótese nula (H_0) é rejeitada e as frequências dos votos não seguem distribuição de Benford.

Quando $k \geq 2$, o teste qui-quadrado não deverá ser utilizado se houver qualquer frequência esperada inferior a 1 ou se mais de 20% das frequências esperadas forem menores que 5 (SIEGUEL; CASTELLAN Jr, 2008).

Como ferramenta, foi utilizado o software estatístico R que é gratuito e possui inúmeras técnicas estatísticas para análise dos dados. Além disso, neste trabalho, fez-se uso de um pacote fundamental para aplicação da lei de Newcomb-Benford, o *benford.analysis* (CINELLI, 2014).

5. Resultados e Discussões

O segundo turno das Eleições Presidenciais foi realizado no dia 26 de outubro de 2014, e a apuração dos resultados, exposta no site do Tribunal Superior Eleitoral, totalizava 105.542.273 votos, sendo que 54.501.118 foram para a candidata eleita e 51.041.155 para o candidato não eleito. Para fins de transparência e investigação, foram coletados os dados dos boletins de urna de cada estado que, posteriormente, foram agregados por zona eleitoral.

Após as análises iniciais, foram descartados valores faltantes, bem como as urnas não apuradas, eliminando-se da análise 41 observações, destarte, procedeu-se a análise que resultou em 112.289.506 votos.

Utilizou-se a regressão linear para testar a uniformidade da distribuição das mantissas dos logaritmos. O método confirmou que as mantissas formavam uma linha quase reta ao longo do intervalo $[0, 1)$ quando ordenadas, confirmando essa propriedade da Lei de Benford (Figura 3).

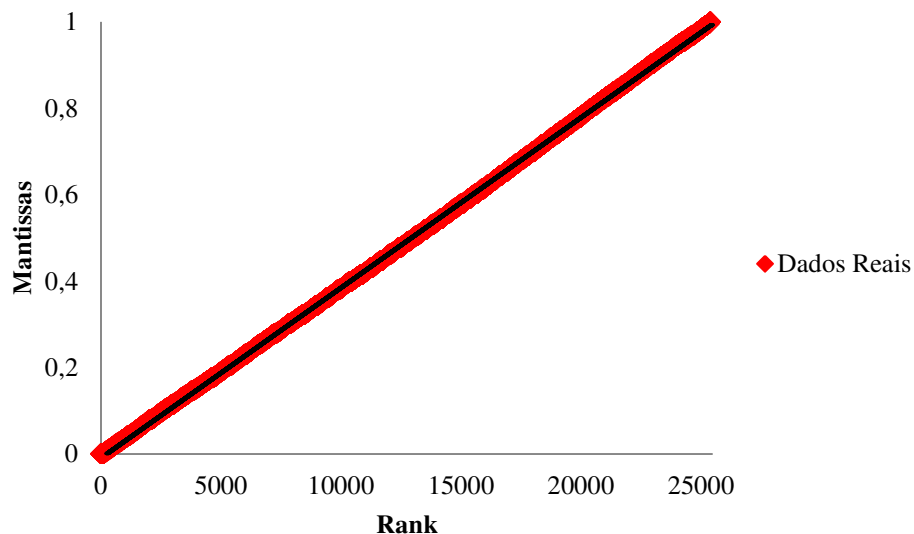


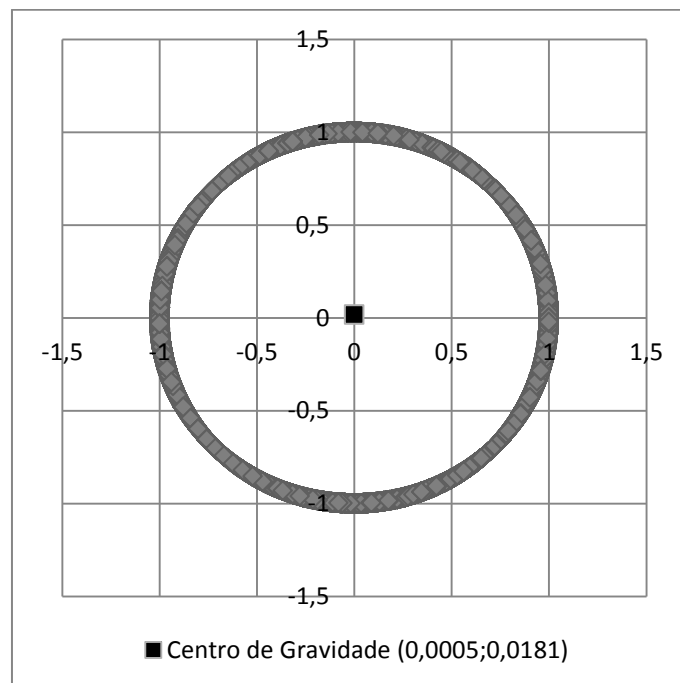
Figura 3 – Distribuição das mantissas dos logaritmos dos votos.

Além disso, testou-se o intercepto, a inclinação e o coeficiente de determinação. A variável y é o conjunto das mantissas ordenadas e a variável x é o ranque que vai de 1 até N . O *valor p* exposto na Tabela (2) indica que tanto o intercepto quanto a inclinação foram significativos para o modelo. O coeficiente de determinação $R^2 = 0,9998$ foi próximo de 1 o que comprova o bom ajuste da reta de regressão.

Tabela 2 - Regressão linear das mantissas.

	Coeficientes	Erro padrão	Estatística t	Valor p
Intercepto	-0,0083	$5,29 \times 10^{-5}$	-156,154	0
Inclinação	$3,94 \times 10^{-5}$	$3,6 \times 10^{-9}$	10928,31	0

O Mantissa Arc Test confirmou a uniformidade da distribuição das mantissas dos valores. O *valor p* do teste foi de 0,00956 e menor que 5% de significância que constata que as mantissas são uniformes e a distribuição global dos votos se aproxima da distribuição de Benford. A Figura (4) exibe as mantissas transformadas em coordenadas e plotadas no plano cartesiano.

**Figura 4** – Mantissa Arc Test.

Das quantidades de votos do segundo turno das eleições presidenciais brasileiras, foram avaliados quatro subconjuntos separadamente tal como número de votos brancos, número de votos nulos, números de votos para a candidata Dilma e números de votos para o candidato Aécio. Mediante testes estatísticos, foi possível verificar a conformidade dos registros das quantidades dos votos com a lei de Benford, exceto para os votos em branco, ilustrando a aplicabilidade desta lei como ferramenta de auditoria da qualidade e confiabilidade dos resultados.

Para a quantidade de votos brancos, a soma total foi de 1.918.339, com mínimo de 1 voto por zona, máximo de 5.799 votos e a média foi de 303 votos por zona eleitoral. No teste do segundo dígito, calculou-se o valor da estatística Qui-quadrado, para $k - 1 = 9$ graus de liberdade, o *valor p* foi significativo, rejeitando a hipótese nula de aderência à Lei de Benford a um nível α de 5% de significância. A Estatística Z calculada para cada algarismo aponta distorção nos números 0 e 8 que podem ser confirmados na Tabela (3).

Tabela 3 – Resultados das estatísticas para o teste do 2º dígito para votos em branco.

Algarismo	Frequência Observada	Frequência de NB	Teste Z
0	0,135	0,120	3,661*
1	0,115	0,114	0,288
2	0,111	0,109	0,501
3	0,101	0,104	0,910
4	0,098	0,100	0,569
5	0,098	0,097	0,442
6	0,090	0,093	0,893
7	0,088	0,090	0,550
8	0,080	0,088	2,263*
9	0,084	0,085	0,206
Soma	1	1	

O gráfico das frequências de votos em branco pela frequência esperada da LNB é exibido na Figura (5):

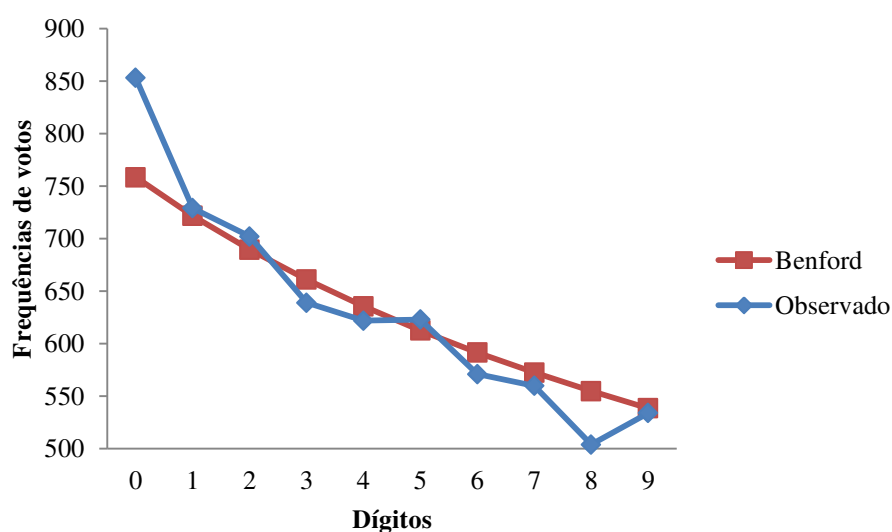


Figura 5 - Frequências esperadas da lei de Benford pelas frequências observadas da quantidade de votos brancos.

Para a quantidade de votos nulos, o limite inferior foi de 1 voto, o limite superior foi de 17.745 votos, com um total de 5.211.553 votos e média de 822 por zona eleitoral. O teste de aderência para o segundo dígito apontou evidências para não rejeição da hipótese nula (*valor* $p > 0,05$), atestando que não houve sinais de distorção nas quantidades de votos nulos. Para o teste Z apenas o algarismo 0 se destoa da distribuição (Tabela 4). A Figura (6) expõe as frequências observadas pelas esperadas da LNB.

Tabela 4 - Resultados das estatísticas para o teste do 2º dígito para votos nulos.

Algarismo	Frequência Observada	Frequência de NB	Teste Z
0	0,128	0,120	2,01*
1	0,112	0,114	0,45
2	0,106	0,109	0,66
3	0,101	0,104	0,81
4	0,102	0,100	0,41
5	0,101	0,097	1,27
6	0,090	0,093	0,96
7	0,089	0,090	0,49
8	0,090	0,088	0,78
9	0,081	0,085	1,26
Soma	1	1	

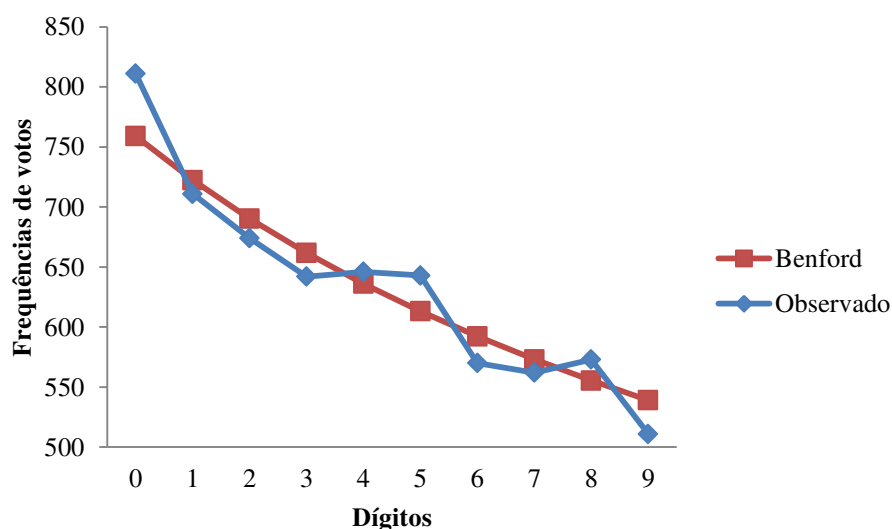


Figura 6 - Frequências esperadas da lei de Benford pelas frequências observadas da quantidade de votos nulos.

Para a quantidade de votos para a candidata Dilma, o limite inferior foi de 1 voto, o limite superior foi de 102.732 votos, a média de votos foi de 8.541 e o total de votos foi de 54.362.136 por zona eleitoral. Os resultados do teste de aderência evidenciaram que os votos à candidata correspondem à distribuição NB (*valor p* > 0,05). Pelo teste Z, também não há evidências para alegar distorções nos valores dos dígitos, veja na Tabela (5). A Figura (7) expõe as frequências observadas pelas esperadas da LNB.

Tabela 5 - Resultados das estatísticas para o teste do 2º dígito para votos para a candidata Dilma Rousseff.

Algarismo	Frequência Observada	Frequência de NB	Teste Z
0	0,120	0,120	0,164
1	0,108	0,114	1,456
2	0,110	0,109	0,216
3	0,105	0,104	0,162
4	0,104	0,100	0,859
5	0,101	0,097	1,132
6	0,091	0,093	0,616
7	0,088	0,090	0,703
8	0,087	0,088	0,194
9	0,087	0,085	0,496
Soma	1	1	

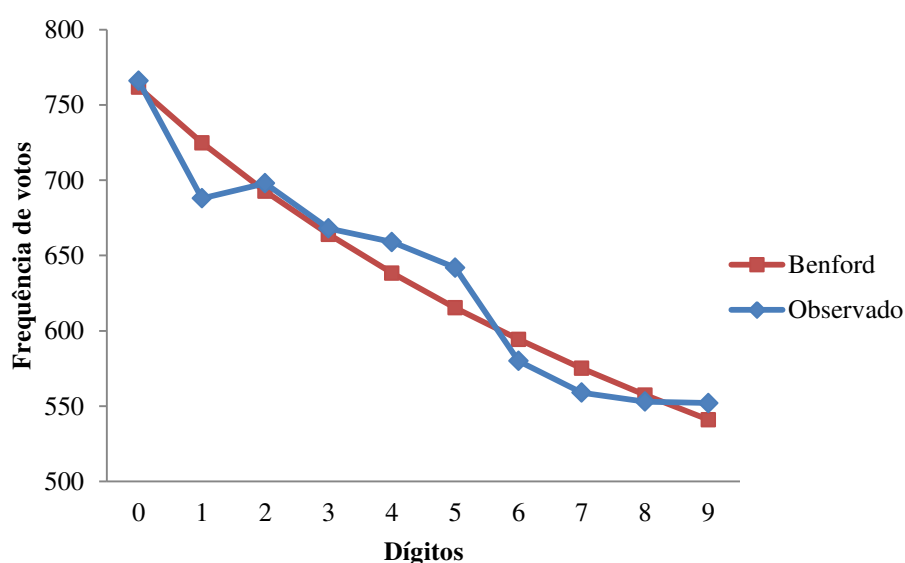


Figura 7 - Frequências esperadas da lei de Benford pelas frequências observadas da quantidade de votos para Dilma Rousseff.

Para a quantidade de votos para o candidato Aécio, o limite inferior foi de 5 votos, o limite superior foi de 131.292 votos e a média foi de 7.980 votos por zona eleitoral. O número total de votos para este candidato foi de 50.797.478,00 votos. Embasando-se no teste Qui-quadrado, comparou-se a distribuição de probabilidades descrita pela lei NB com os valores dos votos para o candidato, os resultados apontaram que não houve indícios de distorções significativas (*valor p* > 0,05). Observando a Tabela (6), nota-se que nenhum dos valores da estatística Z foi significativo, não evidenciando possíveis manipulações. A Figura (8) expõe as frequências observadas pelas esperadas de Benford.

Tabela 6 - Resultados das estatísticas para o teste do 2º dígito para votos para o candidato Aécio Neves.

Algarismo	Frequência Observada	Frequência de NB	Teste Z
0	0,123	0,120	0,931
1	0,109	0,114	1,145
2	0,106	0,109	0,714
3	0,107	0,104	0,731
4	0,107	0,100	1,731
5	0,094	0,097	0,696
6	0,093	0,093	0,146
7	0,089	0,090	0,445
8	0,090	0,088	0,733
9	0,081	0,085	1,036
Soma	1	1	

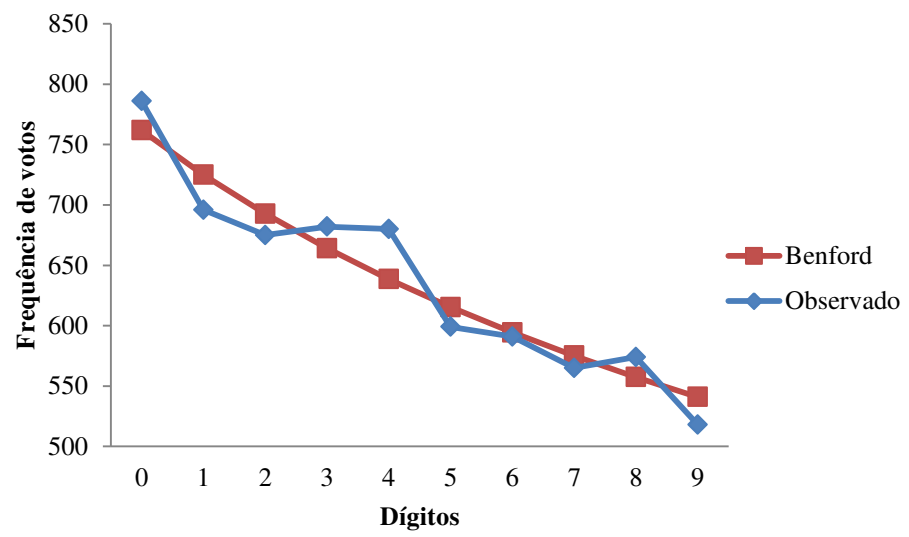


Figura 1 - Frequências esperadas da lei de Benford pelas frequências observadas da quantidade de votos para Aécio Neves.

6. Conclusões

O teste dos segundos dígitos vem como um instrumento de alerta que indica para onde é necessário dispor maior atenção. De fato, para o uso desta lei, é necessário que se tenha cautela, pois muitos dados podem não provir de natureza aleatória ou simplesmente apresentar um feitiço que não se adere à lei.

Ante todo o exposto, constatou-se que as quantidades de votos nulos e para os candidatos apresentaram uma conformidade aceitável para os segundos dígitos, o que sugere que não houve algum tipo de interferência nos dados. Já as quantidades de votos brancos apontaram uma desconformidade dos segundos dígitos analisados em conjunto e requerem certa verificação.

Desta forma, embora os votos em branco não tenham se adequadado à lei, eles não alteram de maneira efetiva os resultados das eleições. Porém, ainda assim, é uma ocasião que demanda uma melhor atenção. A análise dos dígitos se mostra como uma ferramenta poderosa no aprimoramento e controle não intuitivo de processos eleitorais.

REFERÊNCIAS

ALEXANDER, J. C. *Remarks on the use of Benford's Law*. Novembro de 2009.

BELLOS, A. **Alex através do espelho: Como a vida reflete os números e como os números refletem a vida**. Cap. 2. Maio de 2015. Disponível em: https://books.google.com.br/books?id=u7tUCgAAQBAJ&pg=PT47&lpg=PT47&dq=walter+mebane+e+a+lei+de+Benford&source=bl&ots=e7Bb5naXcS&sig=uTbJ_RLYPOekL4Eej9fs0irzOis&hl=pt-BR&sa=X&ved=0ahUKEwj364CDjKrNAhVGgJAKHUnZCWMQ6AEILjAD#v=onepage&q=walter%20mebane%20e%20a%20lei%20de%20Benford&f=false. Acesso em: maio de 2016.

BENFORD, F. *The Law of Anomalous Numbers*. American Philosophical Society. Vol. 78, pag. 551-572, Março, 1938.

BERGER, A. HILL. T. P. *An Introduction to Benford's Law*. Cap. 1. 2014. Disponível em: <<http://press.princeton.edu/chapters/s10526.pdf>>. Acesso em: Outubro de 2015.

BOLFARINE, H.; SANDOVAL, M. C. **Introdução à Inferência Estatística**. São Paulo. Setembro de 2000.

BROWNE, M. W. *Following Benford's Law, or Looking Out for No. 1*. The New York Times, agosto de 1998. Disponível em: <<http://www.rexswain.com/benford.html>>. Acesso em: maio de 2016.

BUSSAB, W. O. MORETTIN, P. A. **Estatística básica**. 4. ed. São Paulo, 2004. 321 p.

CATELLI, R. J. **A República do Voto**. 19 de novembro de 2005. Disponível em: <<http://mlb-historia.blogspot.com.br/2005/09/repblica-do-voto.html>>. Acesso em: maio de 2016.

CECCATO, F. **Aplicações da Lei Newcomb-Benford à Auditoria de Obras Públicas**. Dezembro de 2013. 485 f. Dissertação de Mestrado – Universidade de Brasília. Brasília, 2013.

CINELLI, C. (2014). *benford.analysis*: Benford Analysis for data validation and forensic analytics. R package version 0.1.1. <http://CRAN.R-project.org/package=benford.analysis>.

CINELLI, C. **Indício de Fraude nas Eleições? Usando a Lei de Benford**. Análise Real. 2 de novembro de 2014. Disponível em: <https://analisereal.com/2014/11/02/indicio-de-fraude-nas-eleicoes-usando-a-lei-de-benford/>. Acesso em: agosto de 2015.

DINIZ, J. A. SANTOS, J. DIENG, M. DINIZ, M. A. A. **Comprovação de eficácia da aplicação de modelos contabilométricos no campo da auditoria digital das contas públicas municipais: caso de um tribunal de contas de um estado brasileiro**. In: congresso USP de controladoria e contabilidade. São Paulo, 2006.

DOWNING, D. CLARK, J. **Estatística Aplicada**. São Paulo, 2005.

DURTSCHI, C. HILLISON, W. PACINI, C. *The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data*. Journal of Forensic Accounting. Vol. 5, pag. 17-34, 2004.

FELDMAN, A. HALDERMAN, J. A. FELTEN, E. W. *Security analysis of the diebold Accuvote-ts voting machine*. In Proc. of the USENIX Workshop on Accurate Electronic Voting Technology (EVT'07), pages 2–2. USENIX Association. 2007.

FILHO, A. B. CARVALHO, M. A. M. TEIXEIRA, M. C. SIMPLICIO Jr, M. A. FERNANDES, C.T. **Auditoria Especial no Sistema Eleitoral 2014**. Disponível em: <<http://docplayer.com.br/11789476-Auditoria-especial-no-sistema-eleitoral-2014.html>>. Acesso em: março de 2016.

HILL, T. P. *A Statistical Derivation of the Significant-Digit Law*. Statistical Science. Vol. 10, pag. 354-363, 1995 e 1996.

_____. *The Difficulty of Faking Data*. Chance, v. 26, p. 8-13, 1999.

_____. *The First Digit Phenomenon*. American Scientist, Dezembro, 1998.

HOESCHL, H. C.; BUENO, T. C. D'A. **Eleições Eletrônicas 2014 no Brasil**. Relatório. 2016. Disponível em: <<http://institutoi3g.org.br/blog/lei-de-newcomb-benford-nas-ultimas-eleicoes-presidenciais-no-brasil>>. Acesso em: abril de 2016.

HURSTI, H. Diebold TSx evaluation. *Technical report, Black Box Voting, Inc.* 2006. Disponível em: <www.blackboxvoting.org/BBVtsxstudy.pdf>. Acesso em: 06 maio 2016.

IBM Corp. Released 2013. *IBM SPSS Statistics for Windows*, Version 22.0. Armonk, NY: IBM Corp.

KERLINGER, F. N. **Metodologia da pesquisa em ciências sociais: um tratamento conceitual**. São Paulo, 1980.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos da Metodologia Científica**. 5ª ed. São Paulo, 2003. Disponível em: <https://docente.ifrn.edu.br/olivianeta/disciplinas/copy_of_historia-i/historia-ii/china-e-india> Acesso em: março de 2016.

LEHOUCQ, F. 2003. *Electoral Fraud: Causes, Types, and Consequences*. Annual Review of Political Science, Vol. 6, 2003.

LEVIN, J. **Estatística Aplicada a Ciências Humanas**. 2ª ed. São Paulo: Harbra, 1987.

MEBANE, W. R. *Election Forensics: Statistics, Recounts and Fraud*. Annual Meeting of the Midwest Political Science Association. Chicago. Abril de 2007.

_____. *Election Forensics: Vote Counts and Benford's Law*. Summer Meeting of the Political Methodology Society, UC-Davis, July 20-22, 2006.

MELÃO, H. Jr. **Outro indício de fraude nas eleições**. Saturno V. 2014. Disponível em: <<http://www.saturnov.com/artigos/331-outro-indicio-de-fraude-nas-eleicoes>>. Acesso em: maio de 2016.

NASCIMENTO, T. D. FILHO, E. M. S. BUSCACIO, L. **Detecção De Fraudes: O Uso Da Lei De Benford Para Avaliar Dados Educacionais E Financeiros**. XLVI Simpósio Brasileiro de Pesquisa Operacional. Salvador-BA, 2014.

NEWCOMB, S. *Note on the Frequency of Use of the Different Digits in Natural Numbers*. American Journal of Mathematics. Vol. 4, pag 39-40, 1881.

NIGRINI, M J. *A taxpayer compliance application of Benford's Law*. Journal American Taxation Association. Pag, 72-91, 1996.

_____. *Benford's Law: Application for Forensic Accounting, Auditing and Fraud Detection*. Wiley and Sons: New Jersey, 2012.

_____. *Digital Analysis Using Benford's Law: Tests Statistics for Auditors*. Global Audit Publication. Canadá, 2000.

OLIVEIRA, L. H. H. **Voto obrigatório e equidade um estudo de caso**. São Paulo Perspec., São Paulo , v. 13, n. 4, p. 144-152, Dec. 1999 . Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-88391999000400016&lng=en&nrm=iso. Acesso em: maio de 2016.

PERICCHI, L. R.; TORRES, D. *La Ley de Newcomb-Benford y sus Aplicaciones al Referendum Revocatorio en Venezuela*. Reporte Técnico no-definitivo. 2. ed. Outubro, 2004.

PINKHAM, R. S. *On the Distribution of First Significant Digits*. The Annals of Mathematical Statistics. Vol. 32, pag. 1223-1230, dezembro, 1961.

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

RAIMI, R. *The Peculiar Distribution of First Significant Digits*. Scientific American, 221 (6), 109-120 (1969).

ROUSSEAU, C. A Lei de Benford: aprendendo a fazer ou a detectar fraudes? 10 setembro 2012. Disponível em: <<http://blog.kleinproject.org/?p=1446&lang=pt-br>>. Acesso em: 19 setembro 2015.

SANTOS, J. TENÓRIO, J. N. B. SILVA, L. G. C. **Uma Aplicação da Teoria das Probabilidades na Contabilometria: A lei de Newcomb-Benford como medida para análise de dados no campo da auditoria contábil**. Unb. Vol. 6. Brasília, 2003.

SANTOS, W.G. dos. **Crise e castigo: partidos e gerais na política brasileira**. São Paulo, Vértice, 1987, capítulo III.

SIEGEL, S.; CASTELLAN JR., N. J. **Estatística Não-paramétrica para Ciências do Comportamento**. Métodos de pesquisa. 2. ed. Porto Alegre, 2008.

STEVENSON, Willian J. **Estatística Aplicada à Administração**. Ed. Harbra, São Paulo. 2001

TEIXEIRA, A.C. KIRA, E. **Lei de Benford e Aplicações**. Instituto de Matemática e Estatística da USP. 2016.

TRIBUNAL SUPERIOR ELEITORAL-TSE. Eleitor. Glossário. **Abstenção Eleitoral**. Disponível em: <http://www.tse.jus.br/eleitor/glossario/termos-iniciados-com-a-letra-a#abstencao-eleitoral>. Acesso em: Junho de 2016.

TRIBUNAL SUPERIOR ELEITORAL-TSE. Imprensa. Notícias. **Conheça a história da urna eletrônica brasileira, que completa 18 anos**. Junho de 2014. Disponível em: <http://www.tse.jus.br/imprensa/noticias-tse/2014/Junho/conheca-a-historia-da-urna-eletronica-brasileira-que-completa-18-anos>. Acesso em: Maio de 2016.

TRIBUNAL SUPERIOR ELEITORAL-TSE. Repositório de Dados Eleitorais. **Resultados**. Apresenta resultados das eleições gerais brasileiras. Disponível em: <<http://www.tse.jus.br/eleicoes/estatisticas/repositorio-de-dados-eleitorais>>. Acesso em: 26 abril 2015.

APÊNDICE

APÊNDICE A: Rotina utilizada no software R.

#Os dados foram salvos em um arquivo de Excel e, portanto, é necessário carregar o pacote
#*XLConnect* que conecta diretamente o R ao Excel.

#Os dados quando importados vêm em formato *data.frame* e precisam ser transformados em
#tipo vetor para uso da função *benford.analysis*.

#*VOTOx10* é a variável das quantidades dos votos por zona eleitoral multiplicada por 10
(mudança de escala).

#Carregando o conector Excel com R#
library(XLConnect)

#Carregando pacote Benford
library(benford.analysis)

VOTOS AÉCIO #
#####

#Lendo dados do Excel#
dados_aecio<- readWorksheet(loadWorkbook("Aecio.xlsx"),sheet=1)
names(dados_aecio)

#Verificando se é data frame#
is.data.frame(dados_aecio)

#Armazenando valores em tipo vetor#
dados <- dados_aecio\$dados_aecio

votos<-c(dados_aecio\$VOTOx10)

#Verifica se foi transformado em vetor.
is.vector(votos)

#Função que calcula as probbibilidades e compara com as probabilidades esperadas pela
#distribuição da LNB. Calcula o valor da estatística Qui-quadrado e do Mantissa Arc Test.
bfd_votos <- benford(dados_aecio\$VOTOx10, number.of.digits=2)

plot(bfd_votos)
print(bfd_votos)

separando os dados
split_votos_uf <- split(dados_aecio, dados_aecio\$NOME_CAND)

```

bfd_votos_uf <- lapply(split_votos_uf, function(x) benford(x$VOTOx10,
number.of.digits=2))
#calcula a estatística Qui-quadrado separadamente.
chi_votos_uf <- sapply(bfd_votos_uf, function(x) chisq(x)$stat)
chi_votos_uf

```

```

#####
# VOTOS EM BRANCO #
#####

```

```

#Lendo dados do Excel#
dados_branco<- readWorksheet(loadWorkbook("Branco.xlsx"),sheet=1)
names(dados_branco)

```

```

#Verificando se é data frame#
is.data.frame(dados_branco)

```

```

#Armazenando valores em tipo vetor#
dados <- dados_branco$dados_branco

```

```

votos<-c(dados_branco$VOTOx10)

```

```

#Verifica se foi transformado em vetor.
is.vector(votos)

```

```

#Função que calcula as probbibilidades e compara com as probabilidades esperadas pela
#distribuição da LNB. Calcula o valor da estatística Qui-quadrado e do Mantissa Arc Test.

```

```

bfd_votos <- benford(dados_branco$VOTOx10, number.of.digits=2)

```

```

plot(bfd_votos)
print(bfd_votos)

```

```

# separando os dados
split_votos_uf <- split(dados_branco, dados_branco$NOME_CAND)

```

```

bfd_votos_uf <- lapply(split_votos_uf, function(x) benford(x$VOTOx10,
number.of.digits=2))

```

```

#calcula a estatística Qui-quadrado separadamente.
chi_votos_uf <- sapply(bfd_votos_uf, function(x) chisq(x)$stat)
chi_votos_uf

```

```

#####
# VOTOS NULOS #
#####

```

```

#Lendo dados do Excel#
dados_nulo<- readWorksheet(loadWorkbook("Nulo.xlsx"),sheet=1)
names(dados_nulo)

#Verificando se é data frame#
is.data.frame(dados_nulo)

#Armazenando valores em tipo vetor#
dados <- dados_nulo$dados_nulo

votos<-c(dados_nulo$QTD_VOTOS)

#Verifica se foi transformado em vetor.
is.vector(votos)

#Função que calcula as probbibilidades e compara com as probabilidades esperadas pela
#distribuição da LNB. Calcula o valor da estatística Qui-quadrado e do Mantissa Arc Test.

bfd_votos <- benford(dados_nulo$ VOTOx10, number.of.digits=2)

plot(bfd_votos)
print(bfd_votos)

# separando os dados
split_votos_uf <- split(dados_nulo, dados_nulo$UF)

bfd_votos_uf <- lapply(split_votos_uf, function(x) benford(x$ VOTOx10,
number.of.digits=2))

#calcula a estatística Qui-quadrado separadamente.
chi_votos_uf <- sapply(bfd_votos_uf, function(x) chisq(x)$stat)
chi_votos_uf

#####
# VOTOS DILMA #
#####

#Lendo dados do Excel#
dados_dilma<- readWorksheet(loadWorkbook("DILMA.xlsx"),sheet=1)
names(dados_dilma)

#Verificando se é data frame#
is.data.frame(dados_dilma)

#Armazenando valores em tipo vetor#
dados <- dados_dilma$dados_dilma

```

```
votos<-c(dados_dilma$ VOTOx10)
```

```
#Verifica se foi transformado em vetor.  
is.vector(votos)
```

```
#Função que calcula as probbibilidades e compara com as probabilidades esperadas pela  
#distribuição da LNB. Calcula o valor da estatística Qui-quadrado e do Mantissa Arc Test.
```

```
bfd_votos <- benford(dados_dilma$ VOTOx10_sum, number.of.digits=2)  
plot(bfd_votos)  
print(bfd_votos)
```

```
# separando os dados  
split_votos_uf <- split(dados_dilma, dados_dilma$UF)
```

```
bfd_votos_uf <- lapply(split_votos_uf, function(x) benford(x$ VOTOx10,  
number.of.digits=2))
```

```
#calcula a estatística Qui-quadrado separadamente.  
chi_votos_uf <- sapply(bfd_votos_uf, function(x) chisq(x)$stat)  
chi_votos_uf
```