



UNIVERSIDADE FEDERAL DE PERNAMBUCO

CENTRO DE INFORMÁTICA

SISTEMAS DE INFORMAÇÃO

RICARDO FERREIRA DOS SANTOS JÚNIOR

**USO DA LEI DE NEWCOMB-BENFORD PARA DETECÇÃO DE ANOMALIAS EM
BASES DE DADOS: UMA ANÁLISE EXPERIMENTAL DAS DESPESAS PELA
COTA PARA EXERCÍCIO DA ATIVIDADE PARLAMENTAR**

Recife

2022

RICARDO FERREIRA DOS SANTOS JÚNIOR

**USO DA LEI DE NEWCOMB-BENFORD PARA DETECÇÃO DE ANOMALIAS EM
BASES DE DADOS: UMA ANÁLISE EXPERIMENTAL DAS DESPESAS PELA
COTA PARA EXERCÍCIO DA ATIVIDADE PARLAMENTAR**

Trabalho de Conclusão apresentado no curso de Sistemas de Informação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Sistemas de Informação.

Orientador: Prof. Vinicius Cardoso Garcia

Recife

2022

RICARDO FERREIRA DOS SANTOS JÚNIOR

**USO DA LEI DE NEWCOMB-BENFORD PARA DETECÇÃO DE ANOMALIAS EM
BASES DE DADOS: UMA ANÁLISE EXPERIMENTAL DAS DESPESAS PELA
COTA PARA EXERCÍCIO DA ATIVIDADE PARLAMENTAR**

Trabalho de Conclusão apresentado no curso de
Sistemas de Informação do Centro de Informática da
Universidade Federal de Pernambuco como requisito
parcial para obtenção do grau de Bacharel em Sistemas
de Informação.

Recife, 15 de Maio de 2022.

BANCA EXAMINADORA

Prof. Vinicius Cardoso Garcia (Orientador)

UNIVERSIDADE FEDERAL DE PERNAMBUCO

Prof. José Carlos Cavalcanti (2º membro da banca)

UNIVERSIDADE FEDERAL DE PERNAMBUCO

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

SANTOS JUNIOR, Ricardo Ferreira dos.

Uso da lei de Newcomb-Benford para detecção de anomalias em bases de dados: Uma análise experimental das despesas pela cota para exercício da atividade parlamentar / Ricardo Ferreira dos SANTOS JUNIOR. - Recife, 2022. 85 : il., tab.

Orientador(a): Vinicius Cardoso Garcia
(Graduação) - Universidade Federal de Pernambuco, Centro de Informática, , 2022.

Inclui referências, apêndices, anexos.

1. Computação. 2. Matemática. 3. Estatística . I. Garcia, Vinicius Cardoso. (Orientação). II. Título.

000 CDD (22.ed.)

Dedico este trabalho a minha avó, que sempre me apoiou
e acreditou em mim

AGRADECIMENTOS

Aos meus pais e avô que estiveram sempre comigo. As minhas irmãs, que me incentivaram e me ajudaram nos momentos difíceis. Aos amigos, por todo o apoio e pela ajuda que muito contribuíram para a realização deste trabalho. Ao meu orientador, Vinicius Cardoso Garcia, por toda paciência e apoio. Ao meu professor de estatística, Pedro Machado Manhães, que se dispôs a ajudar em um momento muito difícil.

“É o tempo da travessia: e, se não ousarmos fazê-la,
teremos ficado, para sempre, à margem de nós mesmos.”

(Fernando Teixeira de Andrade)

RESUMO

Em razão do crescimento exponencial de dados, emergiu a necessidade de mudanças nos métodos de verificação de sua autenticidade, com destaque nas relações de confiança nos dados que se encontram, principalmente, em bases abertas. O principal critério para propor a criação desta nova ferramenta se consolidou com a apropriação de uma técnica de distribuição de dígitos, principalmente, utilizada na área de detecção de fraudes contábeis, baseado, em suma, nas probabilidades de ocorrência de números iniciados por dígitos menores são superiores às probabilidades de ocorrência de números iniciados por dígitos mais elevados, fenômeno este denominado por Newcomb (1881) e Benford (1938), da Lei de Newcomb-Benford, dando início, portanto, a uma mudança nos novos paradigmas das relações contábeis. A realização deste trabalho emerge da necessidade de avaliação de possíveis divergências nas despesas atribuídas a cada parlamentar por meio de uma cota destinada a custear despesas típicas do exercício do seu mandato, despesas encontradas no portal de dados abertos da Câmara dos Deputados Federais, onde esperou-se verificar se existe ou não conformidade das bases com a Lei de Newcomb-Benford. A partir da análise do conjunto de dados disponibilizados pela Câmara dos Deputados, concluiu-se que nem todos os dados estavam em conformidade com a Lei de Newcomb-Benford, de acordo com a análise estatística do Desvio Absoluto Médio.

Palavras-chave: Lei de Newcomb-Benford, Desvio Absoluto Médio, Cota para Despesas Parlamentar.

ABSTRACT

Due to the exponential growth of data, there was a need for changes in the methods of verifying its authenticity with emphasis on trust relationships in data that are mostly found in open data. The main criteria about propounding the creation of this new tool was consolidated with the appropriation of a digit distribution technique, mainly used in the area of accounting fraud detection, based on the probabilities of numbers occurrence starting with smaller digits being higher than the probabilities of numbers starting with higher digits, a phenomenon named by Newcomb (1881) and Benford (1938), from the Newcomb-Benford Law, initiating a change on new models of accounting relationships. The accomplishment of this work comes from the need to estimate possible divergences in expenses assigned to each parliamentarian through a quota destined to defray typical costs to the exercise of its mandate, found in the open data portal of the Chamber of Federal Deputies, where it was expected to verify whether or not the bases matched the Newcomb-Benford Law. From the analysis of the data set made available by the Chamber of Deputies, it was concluded that not all data were in compliance with the Newcomb-Benford Law, according to the statistical analysis of the Mean Absolute Deviation.

Keywords: Newcomb-Benford Law, Mean Absolute Deviation, Quota for Parliamentary Expenses.

LISTA DE FIGURAS

Figura 1 – Frequências observadas e esperadas do 1º dígito para o número de mortes

Figura 2 - Funcionalidades

Figura 3 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2008

Figura 4 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2008

Figura 5 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2009

Figura 6 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2009

Figura 7 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2010

Figura 8 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2010

Figura 9 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2011

Figura 10 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2011

Figura 11 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2012

Figura 12 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2012

Figura 13 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2013

Figura 14 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2013

Figura 15 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2014

Figura 16 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2014

Figura 17 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2015

Figura 18 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2015

Figura 19 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2016

Figura 20 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2016

Figura 21 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2017

Figura 22 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2017

Figura 23 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2018

Figura 24 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2018

Figura 25 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2019

Figura 26 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2019

Figura 27 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2020

Figura 28 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2020

Figura 29 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2021

Figura 30 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2021

Figura 31 - Conclusão Geral dos Resultados

Figura 32 - Conformidade Geral dos Resultados

Figura 33 - Conclusão dos Valores dos Documentos

Figura 34 - Conformidade dos Resultados dos Documentos

Figura 35 - Conclusão dos Valores das Glosas

Figura 36 - Conformidade dos Resultados das Glosas

LISTA DE TABELAS

Tabela 1 - Probabilidade de cada dígito proposto por Benford

Tabela 2 - Percentagem de vezes que os números naturais de 1 a 9 são usados como primeiros dígitos em números, conforme 20.229 observações

Tabela 3 - Valores Críticos Desvio Absoluto Médio(DAM)

Tabela 4 – Ano 2008 do campo vlrDocumento, Teste Z, X^2 e DAM

Tabela 5 – Ano 2008 do campo vlrGlosa, Teste Z, X^2 e DAM

Tabela 6 – Ano 2009 do campo vlrDocumento, Teste Z, X^2 e DAM

Tabela 7 – Ano 2009 do campo vlrGlosa, Teste Z, X^2 e DAM

Tabela 8 – Ano 2010 do campo vlrDocumento, Teste Z, X^2 e DAM

Tabela 9 – Ano 2010 do campo vlrGlosa, Teste Z, X^2 e DAM

Tabela 10 – Ano 2011 do campo vlrDocumento, Teste Z, X^2 e DAM

Tabela 11 – Ano 2011 do campo vlrGlosa, Teste Z, X^2 e DAM

Tabela 12 – Ano 2012 do campo vlrDocumento, Teste Z, X^2 e DAM

Tabela 13 – Ano 2012 do campo vlrGlosa, Teste Z, X^2 e DAM

Tabela 14 – Ano 2013 do campo vlrDocumento, Teste Z, X^2 e DAM

Tabela 15 – Ano 2013 do campo vlrGlosa, Teste Z, X^2 e DAM

Tabela 16 – Ano 2014 do campo vlrDocumento, Teste Z, X^2 e DAM

Tabela 17 – Ano 2014 do campo vlrGlosa, Teste Z, X^2 e DAM

Tabela 18 – Ano 2015 do campo vlrDocumento, Teste Z, X^2 e DAM

Tabela 19 – Ano 2015 do campo vlrGlosa, Teste Z, X^2 e DAM

Tabela 20 – Ano 2016 do campo vlrDocumento, Teste Z, X^2 e DAM

Tabela 21 – Ano 2016 do campo vlrGlosa, Teste Z, X^2 e DAM

Tabela 22 – Ano 2017 do campo vlrDocumento, Teste Z, X^2 e DAM

Tabela 23 – Ano 2017 do campo vlrGlosa, Teste Z, X^2 e DAM

Tabela 24 – Ano 2018 do campo vlrDocumento, Teste Z, X^2 e DAM

Tabela 25 – Ano 2018 do campo vlrGlosa, Teste Z, X^2 e DAM

Tabela 26 – Ano 2019 do campo vlrDocumento, Teste Z, X^2 e DAM

Tabela 27 – Ano 2019 do campo vlrGlosa, Teste Z, X^2 e DAM

Tabela 28 – Ano 2020 do campo vlrDocumento, Teste Z, X^2 e DAM

Tabela 29 – Ano 2020 do campo vlrGlosa, Teste Z, X^2 e DAM

Tabela 30 – Ano 2021 do campo vlrDocumento, Teste Z, X^2 e DAM

Tabela 31 – Ano 2021 do campo vlrGlosa, Teste Z, X^2 e DAM

Tabela 32 – Conclusões de conformidade

LISTA DE EQUAÇÕES

$$(1) F_a = \log_{10}\left(\frac{a+1}{a}\right)$$

$$(2) Z = \frac{|AP - EP| - \left(\frac{1}{2N}\right)}{\sqrt{\frac{EP(1-EP)}{N}}}$$

$$(3) Qui - Quadrado = \sum_{i=1}^K \frac{(AC - EC)^2}{EC}$$

$$(4) DAM = \sum_{i=1}^K \frac{|AP - EP|}{K}$$

LISTA DE ABREVIATURAS E SIGLAS

LNB	Lei de Newcomb-Benford
DAM	Desvio Médio Absoluto (Mean Absolute Deviation)
X^2	Qui-Quadrado (Chi-Squared)
Z	Estatística Z (Z-Score)

SUMÁRIO

INTRODUÇÃO	16
MOTIVAÇÃO	17
OBJETIVO	18
PERGUNTA DE PESQUISA	18
HIPÓTESES	18
OBJETIVOS	18
Objetivo Geral	18
Objetivos Específicos	18
ESTRUTURA DE TRABALHO	19
REFERENCIAL TEÓRICO	20
LEI NEWCOMB-BENFORD	20
TESTES ESTATÍSTICOS	24
Estatística Z (Z-Score)	24
Qui-Quadrado (Chi-square)	25
Desvio Absoluto Médio(Mean Absolute Deviation)	26
METODOLOGIA	27
PROPOSTA DE FERRAMENTA	27
Descrição Geral	27
Visão Geral da Arquitetura	27
Requisitos da Aplicação	27
Versionamento da Ferramenta	28
Desenvolvimento dos códigos na aplicação	28
Funcionalidades	29
Verificação e Implantação da Ferramenta	30
PROPOSTA DO TCC	31
COLETA DOS DADOS E PREPARAÇÃO DOS DADOS PARA ANÁLISE	31
TRATAMENTO DOS DADOS	32
ANÁLISE ESTATÍSTICA	32
RESULTADOS	34
ANÁLISE DAS DESPESAS PELA CONTA PARA EXERCÍCIO DA ATIVIDADE PARLAMENTAR	34
Ano 2008	34
Ano 2009	36
Ano 2010	39
Ano 2011	42
Ano 2012	45

Ano 2013	47
Ano 2014	50
Ano 2015	52
Ano 2016	55
Ano 2017	58
Ano 2018	61
Ano 2019	63
Ano 2020	66
Ano 2021	69
CONCLUSÃO	76

1. INTRODUÇÃO

O presente estudo tem como razão precípua estimular o desenvolvimento, no meio acadêmico, sobre a discussão acerca de bases de dados a fim de garantir a confiabilidade. A adequação desses conjuntos de dados, no que diz respeito à qualidade, é de extrema importância para garantir que as bases são confiáveis para serem manipuladas e, por conseguinte, não acarretem problemas em possíveis resultados gerados a partir dessas bases (CRUSOE, J; MELIN, U., 2018).

A partir disso, na tentativa de contornar o problema foi identificada uma relação já existente entre os dados de caráter numérico e a essência peculiar dos números anômalos proposta inicialmente por Simon Newcomb. Foi observado por S. Newcomb, em 1881, que em uma tabela de logaritmos comuns, as primeiras páginas apresentavam um desgaste maior do que as últimas páginas, indicando que os números usados começam mais com o dígito 1 do que com o dígito 9 (BENFORD, 1938).

Benford, em 1937, propôs uma distribuição dos dígitos, baseada na descoberta de Newcomb, em uma tabela logarítmica, onde o numeral 1 encontra-se na primeira posição em 30.1% das vezes; o numeral 2 encontra-se na primeira posição em 17,6%, e assim por diante, onde os valores tendem a diminuir na medida que chegam ao último numeral, o 9 (TABELA 1).

Tabela 1 - Probabilidade de cada dígito proposto por Benford

Dígitos	Probabilidade
1	0.301
2	0.176
3	0.125
4	0.097
5	0.079
6	0.067
7	0.058
8	0.051
9	0.046

Fonte: Benford (1938).

A natureza peculiar da descoberta de S. Newcomb pode criar um cenário perfeito para utilização do que, futuramente, foi denominado de a “Lei de Newcomb-Benford”. A utilização prática da análise pela Lei de Newcomb-Benford está na detecção de anomalias, identificando números fictícios e possíveis fraudes, através de testes para a ocorrência ou não da distribuição previsível de dígitos em conjuntos de dados. Fato bastante relevante ao considerarmos que a maioria dos fraudadores não considera o padrão da Lei de Newcomb-Benford ao criar documentações, ou realizar transações falsas para encobrir seus rastros (ASSOCIATION OF CERTIFIED FRAUD EXAMINERS, 2020).

Sendo assim, a utilização da análise de Benford pode ser empregada como prova admissível de fraude em tribunais, pois é um método amplamente utilizado por especialistas. Além disso, muitos tipos de dados podem se encaixar nessas conjecturas, incluindo contagens de população, dados contábeis e tráfego de rede (KESSEL E., 2020).

A realização deste trabalho emerge da necessidade de identificação da conformidade de bases de dados de uma maneira simplificada e prática. Com isso, convergiu a ideia de criar uma ferramenta acessível e simples de usar que pudesse receber essas bases e aplicar a Lei de Newcomb-Benford. Devido a esse valor auditável, a ideia de utilizar uma base de dados abertos que, por sua vez, possuísse esse caráter contábil, direcionou sua utilização para uma base em específico como estudo de caso: a disponibilizada pela Câmara dos Deputados Federais do Brasil com as despesas pela cota para exercício da atividade parlamentar, onde uma análise experimental da ferramenta foi realizada.

1.1 MOTIVAÇÃO

A proposta de criação de ferramenta que utilize a Lei de Newcomb-Benford emerge da necessidade de investigar se os dados de despesas pela cota para exercício da atividade parlamentar, encontrados no portal de dados abertos da Câmara dos Deputados Federais do Brasil, são fidedignos.

1.2 OBJETIVO

1.2.1 PERGUNTA DE PESQUISA

Os dados existentes no portal de dados abertos da Câmara dos Deputados Federais, em sua totalidade, podem ser considerados confiáveis?

1.2.2 HIPÓTESES

No processo de coleta dos dados, algumas informações podem, de maneira errônea, possuir anomalias ou imprecisões quanto aos seus informações, e com isso, as resoluções obtidas a partir do uso desses dados podem gerar conclusões equivocadas sobre as despesas da Câmara dos Deputados.

1.2.3 OBJETIVOS

1.2.3.1 Objetivo Geral

O objetivo geral deste trabalho é investigar se os dados de despesas pela cota para exercício da atividade parlamentar, encontradas no portal de dados abertos da Câmara dos Deputados Federais do Brasil, estão em conformidade com a Lei de Newcomb-Benford.

1.2.3.2 Objetivos Específicos

- a) Criar uma ferramenta que possibilite a utilização da Lei de Newcomb-Benford em arquivos de planilhas;
- b) Identificar a conformidade das bases de dados com testes estatísticos em relação à Lei de Newcomb-Benford;
- c) Permitir a personalização dos dados de entrada.

1.3 ESTRUTURA DE TRABALHO

A organização do trabalho se dá pela seguinte forma:

- Capítulo 1 – Introdução
- Capítulo 2 – Referencial Teórico
- Capítulo 3 – Metodologia
- Capítulo 4 – Proposta do TG
- Capítulo 5 – Resultados
- Capítulo 6 – Conclusões

2 REFERENCIAL TEÓRICO

Para entender a relevância da utilização da Lei de Newcomb-Benford para a detecção de anomalias em dados abertos, é necessário primordialmente compreender os principais conceitos e implicações acerca da lei, com vistas à sua utilização para a melhora da confiabilidade de bases de dados.

2.1 LEI NEWCOMB-BENFORD

A adequação dos conjuntos de dados no que diz respeito à qualidade é de extrema importância para garantir que as bases são confiáveis para serem manipuladas, pois fatores como a integridade, a fragmentação, a precisão, e até o contexto que os dados foram captados, se menosprezados, podem causar problemas em possíveis resultados gerados a partir dessas bases (CRUSOE, J; MELIN, U., 2018).

Identificada a relação existente entre os dados de caráter numérico e a essência peculiar dos números anômalos foram propostos inicialmente por Simon Newcomb. Foi observado por S. Newcomb em 1881 que, em uma tabela de logaritmos comuns, as primeiras páginas apresentavam um desgaste maior do que as últimas páginas, indicando que os números usados começam mais com o dígito 1 do que com o dígito 9. Mostrou também que a ocorrência, em um conjunto de números, de acontecer é maior em números com o primeiro dígito baixo. Portanto, a Lei de Newcomb-Benford, ou também chamada de Lei dos Números Anômalos, pode ser considerada uma lei de probabilidade geral de ampla aplicação.

A partir de uma coleta e observação em conjuntos de dados (TABELA 2) das mais variadas fontes, Benford conseguiu provar que a lei estava presente em suas observações que incluíam desde endereços, taxas de mortes, comprimentos de rios, quadrado de números naturais, entre outros (BENFORD, 1938).

TABELA 2 - PERCENTAGEM DE VEZES QUE OS NÚMEROS NATURAIS DE 1 A 9 SÃO USADOS COMO PRIMEIROS DÍGITOS EM NÚMEROS, CONFORME 20.229 OBSERVAÇÕES

PERCENTAGE OF TIMES THE NATURAL NUMBERS 1 TO 9 ARE USED AS FIRST DIGITS IN NUMBERS, AS DETERMINED BY 20,229 OBSERVATIONS

Group	Title	First Digit									Count
		1	2	3	4	5	6	7	8	9	
A	Rivers, Area	31.0	16.4	10.7	11.3	7.2	8.6	5.5	4.2	5.1	335
B	Population	33.9	20.4	14.2	8.1	7.2	6.2	4.1	3.7	2.2	3259
C	Constants	41.3	14.4	4.8	8.6	10.6	5.8	1.0	2.9	10.6	104
D	Newspapers	30.0	18.0	12.0	10.0	8.0	6.0	6.0	5.0	5.0	100
E	Spec. Heat	24.0	18.4	16.2	14.6	10.6	4.1	3.2	4.8	4.1	1389
F	Pressure	29.6	18.3	12.8	9.8	8.3	6.4	5.7	4.4	4.7	703
G	H.P. Lost	30.0	18.4	11.9	10.8	8.1	7.0	5.1	5.1	3.6	690
H	Mol. Wgt.	26.7	25.2	15.4	10.8	6.7	5.1	4.1	2.8	3.2	1800
I	Drainage	27.1	23.9	13.8	12.6	8.2	5.0	5.0	2.5	1.9	159
J	Atomic Wgt.	47.2	18.7	5.5	4.4	6.6	4.4	3.3	4.4	5.5	91
K	n^{-1} , $\sqrt{n}$, ...	25.7	20.3	9.7	6.8	6.6	6.8	7.2	8.0	8.9	5000
L	Design	26.8	14.8	14.3	7.5	8.3	8.4	7.0	7.3	5.6	560
M	<i>Digest</i>	33.4	18.5	12.4	7.5	7.1	6.5	5.5	4.9	4.2	308
N	Cost Data	32.4	18.8	10.1	10.1	9.8	5.5	4.7	5.5	3.1	741
O	X-Ray Volts	27.9	17.5	14.4	9.0	8.1	7.4	5.1	5.8	4.8	707
P	Am. League	32.7	17.6	12.6	9.8	7.4	6.4	4.9	5.6	3.0	1458
Q	Black Body	31.0	17.3	14.1	8.7	6.6	7.0	5.2	4.7	5.4	1165
R	Addresses	28.9	19.2	12.6	8.8	8.5	6.4	5.6	5.0	5.0	342
S	$n!$, $n^2 \dots n!$	25.3	16.0	12.0	10.0	8.5	8.8	6.8	7.1	5.5	900
T	Death Rate	27.0	18.6	15.7	9.4	6.7	6.5	7.2	4.8	4.1	418
Average		30.6	18.5	12.4	9.4	8.0	6.4	5.1	4.9	4.7	1011
Probable Error		± 0.8	± 0.4	± 0.4	± 0.3	± 0.2	± 0.2	± 0.2	± 0.2	± 0.3	—

Fonte: BENFORD (1938).

Com isso, Benford, em 1937, aperfeiçoou a ideia de S. Newcomb, chegando a conclusão de que a frequência dos primeiros dígitos pode ser expressa por uma equação logarítmica (EQUAÇÃO 1).

$$F_a = \log_{10}\left(\frac{a+1}{a}\right) \quad (1)$$

A frequência do primeiro dígito segue de perto uma relação logarítmica, que no caso do dígito 1 ($F_a = 1$) que é de 0,306, sendo muito próximo do $\log 2$. Já no caso do dígito 2, era 0,185, muito próxima do $\log 3 - \log 2$, e assim por diante. Com isso, chegou-se à conclusão de que a frequência dos primeiros dígitos pode ser expressa por esta equação logarítmica, onde a é um dígito no intervalo fechado de 1 e 9.

Como consequência, a Lei de Newcomb-Benford começou a ser amplamente utilizada, como observa o Fraud Examiners Manual (2020):

O objetivo de uma análise da Lei de Benford é identificar números fictícios. Muitos fraudadores não consideram o padrão da Lei de Benford ao criar documentação ou transações falsas para encobrir seus rastros. Consequentemente, testar conjuntos de

dados para a ocorrência ou não da distribuição previsível de dígitos pode ajudar a identificar números incluídos que não são legítimos.

Diversos critérios têm sido avaliados ao utilizar a Lei de Newcomb-Benford para detecção de fraudes. Para Mark J. Nigrini (2012) é necessário que:

- a) O conjunto de dados deve descrever os tamanhos de fenômenos semelhantes (por exemplo, o valor monetário das transações de compra);
- b) Não deve haver valores mínimos ou máximos embutidos no conjunto de dados;
- c) O conjunto de dados não deve consistir em números atribuídos (por exemplo, números de conta, números de transação, números de ordem de compra ou números de cheque);
- d) O conjunto de dados deve ter mais itens pequenos do que grandes (por exemplo, uma empresa normalmente tem um número maior de transações pequenas do que grandes).

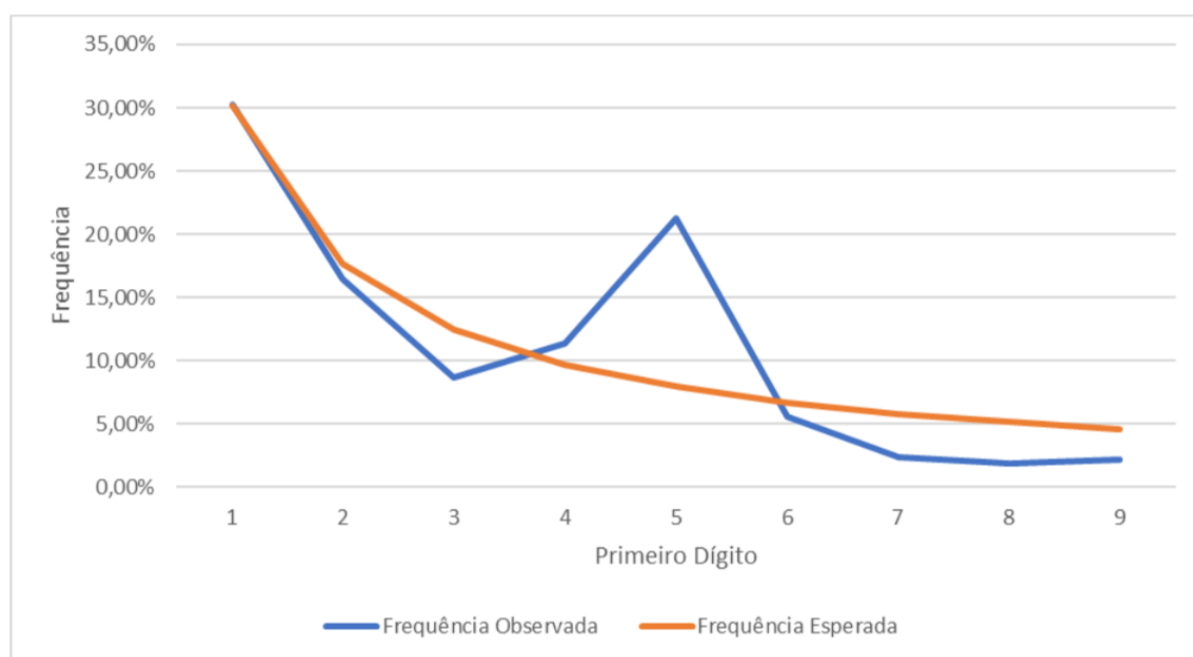
Enquanto isso, Kessel E. (2020) é mais direto, propondo a necessidade de que os dados precisam ser numéricos, grandes, gerados aleatoriamente e precisam representar magnitudes de eventos. Além de explicitar que o uso de dados criteriosos sendo analisados através da Lei de Newcomb-Benford pode gerar provas valiosas de fraudes.

Essa análise pode ser prova admissível de fraude em tribunais federais e estaduais. A comunidade de contabilidade forense geralmente aceita a metodologia, que é referenciada no Fraud Examiners Manual. Contadores forenses, examinadores de fraudes, contadores e auditores usam a lei de Benford para detectar anomalias que requerem investigação. A combinação do uso amplamente aceito do método, reputação acadêmica e ampla disponibilidade de especialistas tornam provável a admissibilidade das análises de Benford. (KESSEL E. 2020, tradução nossa)

Nigrini (2012) presumiu que dados contábeis sem manipulação acompanhavam de perto a lei de Newcomb-Benford, de tal forma que desvios consideráveis em relação à distribuição proposta por Benford poderiam insinuar uma possível fraude nos dados. Um exemplo do uso prático de tal técnica é mencionado por Nigrini (2020), que constatou anomalias em esquemas de empresas de fachada, onde o infrator possuía um fornecedor fictício que gerava valores a serem utilizados para as faturas fraudulentas apresentadas pelo suposto fornecedor.

Para exemplificar, durante a pandemia do Covid-19, um estudo foi realizado relacionando o número de casos confirmados e mortes no Brasil com a lei de Newcomb-Benford (TERRA, A. V. et al, 2021), onde foi analisada a conformidade com a lei e, segundo autor, revelou indícios de rejeição da hipótese nula do primeiro dígito, em outras palavras, foi identificada uma não conformidade nos números apresentados (FIGURA 1).

Figura 1 – Frequências observadas e esperadas do 1º dígito para o número de mortes



Fonte: TERRA, A. V. et al, (2021).

Em outro estudo realizado por MENEZES (2021), conclui-se que o número de casos confirmados da doença COVID-19 na China e em Nova Zelândia seguiu de perto a Lei de Benford, mas foi rejeitada em países como a Itália, Brasil e Estados Unidos.

A lei dos números anômalos foi utilizada em diversas áreas, desde sua utilização em dados contábeis (NIGRINI, 2012) até sua utilização no estudo em genomas de células eucariontes e procariontes (FRIAR, J. L., GOLDMAN T., PÉREZ-MERCADER J. 2012). Costa (2012) fez uma análise que identificou um total de 721 publicações nacionais e internacionais no período de 1881 a 201; além disso, foram encontradas 145 pesquisas relevantes produzidas com enfoque em auditoria contábil entre os anos de 1988 e 2011.

A partir disso, também é importante entender como ocorrerão os testes e assegurar que a análise permita a identificação da conformidade ou não dessas bases com a Lei de Newcomb-Benford.

2.2 TESTES ESTATÍSTICOS

Tendo em vista, a importância de avaliar a conformidade de um conjunto de dados com a Lei de Newcomb-Benford, Nigrini (2012) decidiu utilizar alguns testes descritos em livros de estatística, para obter esse resultado. São eles: Estatística Z (Z-Score), Teste do Qui-Quadrado (Chi-square) e o Desvio Absoluto Médio (Mean Absolute Deviation).

2.2.1 Estatística Z (Z-Score)

Nigrini (2012) demonstrou que o uso do teste Estatística Z, ou também chamado de teste Z, é adequado para verificar se a proporção real (AP) de um dígito se diferencia consideravelmente do que é proporção esperada (EP) pela Lei de Benford. Em sua pesquisa, a estatística Z funcionou bem para verificação não só do primeiro dígito, como também do segundo dígito.

A fórmula estatística Z (EQUAÇÃO 2) consiste em considerar a magnitude absoluta da diferença (a distância numérica do real ao esperado), o tamanho do conjunto de dados e a proporção esperada. Além de que, se rejeitada, a hipótese nula indica que há diferença estatisticamente relevante entre as duas distribuições apontadas, AP e EP, portanto, havendo uma não conformidade entre as distribuições (FONSECA; MARTINS, 1993).

$$Z = \frac{|AP - EP| - (\frac{1}{2N})}{\sqrt{\frac{EP(1-EP)}{N}}} \quad (2)$$

Fórmula do Z score – A fórmula adaptada de FONSECA e MARTINS (1993)

Onde AP é a proporção real, EP é a proporção esperada e N o número de registros. O autor utiliza para esse teste um nível de significância de 5%, além de considerar aceitável apenas cinco picos significativos, visto que a análise é feita para testar um dígito

individualmente, e não aponta uma conformidade ou uma não conformidade geral. Com isso, com a utilização de grande base de dados, a estatística Z tende aceitar desvios cada vez menores. Um acréscimo à estatística Z é um teste combinado para os dígitos usando o teste do Qui-Quadrado ou o teste de Kolmogorov-Smirnoff (NIGRINI, 2012).

2.2.2 Qui-Quadrado (Chi-square)

O teste Chi-square ou também chamado de Qui-Quadrado é dado por uma fórmula estatística (EQUAÇÃO 3) que compara um conjunto de resultados reais com os resultados esperados. A hipótese nula é que os dígitos da distribuição estejam em conformidade com a Lei de Newcomb-Benford. Isso nos permite comparar possíveis divergências na relação entre os resultados experimentais e a distribuição esperada do fenômeno.

$$Qui - Quadrado = \sum_{i=1}^K \frac{(AC - EC)^2}{EC} \quad (3)$$

Fórmula do Chi - Square – A fórmula adaptada de FONSECA e MARTINS (1993)

Onde AC é a contagem real, EC a contagem esperada, e K representa o grau de liberdade, que nesse caso é 8, pois o valor de K é 9 devido a contagem dos nove primeiros dígitos utilizados. No caso do teste Qui-Quadrado a hipótese nula é não indicar não haver discrepância significativa entre a contagem real e a contagem esperada (FONSECA; MARTINS, 1993).

Em relação ao teste de Estatística Z, pode-se concluir que valores do teste Qui-Quadrado altos para os vários dígitos relacionados a Estatísticas Z altas, e o inverso também ocorre com valores Qui-Quadrados baixos relacionados a Estatísticas Z baixas. Se o valor do Qui-Quadrado ultrapasse 15,51, a hipótese nula de conformidade vai ser rejeitada e seria deduzido que os valores não estão em conformidade com a Lei de Newcomb-Benford. Caso não ultrapasse, não há evidências suficientes para rejeitar a hipótese nula de que não há diferença significativa entre as proporções reais e as da Lei de Newcomb-Benford (NIGRINI, 2012).

O teste Qui-Quadrado também sofre com o mesmo problema da Estatística Z, uma vez que conjuntos de dados superiores a 5.000 registros já começam a apresentar essa sensibilidade a pequenos desvios. Nesse caso, o Qui-Quadrado quase sempre vai ser maior que o valor crítico.

2.2.3 Desvio Absoluto Médio(Mean Absolute Deviation)

Segundo Nigrini (2012), os testes anteriores levavam em consideração o número de registros nos cálculos, aumentando essa sensibilidade à medida em que esses números de registros cresciam, além disso, o autor explica o quão impraticável são esses testes nas bases reais que, geralmente, superam esses números. O desvio absoluto médio de um conjunto de dados é a média das distâncias entre cada conjunto de dados e a média (EQUAÇÃO 4). Isso nos permite entender a variabilidade no conjunto de dados e, conseqüentemente, identificar se essa variabilidade ultrapassa os valores críticos, demonstrando assim uma não conformidade entre a LNB.

$$DAM = \sum_{i=1}^K \frac{|AP - EP|}{K} \quad (4)$$

Fórmula do Desvio Absoluto Médio – A fórmula adaptada de FONSECA e MARTINS (1993)

Onde AP é a proporção real, EP é a proporção esperada, e K representa o número de iterações, que nesse caso é 8. O Teste do Desvio Absoluto Médio do primeiro dígito sinalizam conformidade bem próxima. Segundo Nigrini (2012), os valores críticos de Desvio Absoluto Médio para utilização em Newcomb-Benford foram propostos (TABELA 3).

TABELA 3 - Valores Críticos do Desvio Absoluto Médio(DAM)

Descrição	Valores
Close	0.000 a 0.006
Acceptable	0.006 a 0.012
Marginally Acceptable	0.012 a 0.015
Nonconformity	Acima 0.015

Fonte: Adaptada de Nigrini (2012)

3 METODOLOGIA

Neste capítulo é apresentado o processo de desenvolvimento da ferramenta. A proposta desse projeto é encontrar uma forma rápida e simples que auxilie a utilização da Lei de Newcomb-Benford em planilhas de dados.

3.1 PROPOSTA DE FERRAMENTA

3.1.1 Descrição Geral

Bendata é uma aplicação¹ que possibilita a utilização da Lei de Newcomb-Benford para demonstração de possíveis anomalias nos arquivos de planilhas, permitindo a personalização das entradas. Além de aplicar testes estatísticos para identificação da conformidade das bases em relação à Lei de Newcomb-Benford.

3.1.2 Visão Geral da Arquitetura

A aplicação Bendata foi desenvolvida para utilizar a Lei de Newcomb-Benford para, a partir dos dados, demonstrar possíveis anomalias em bases de dados. A ferramenta tem o planejamento de ser desenvolvida utilizando as seguintes linguagens, ferramentas e bibliotecas:

- Linguagem de Programação: Python;
- Deploy: Biblioteca Streamlit;
- Biblioteca de software para manipulação e análise de dados: Pandas;
- Biblioteca de visualização de dados para Python: Plotly;

3.1.3 Requisitos da Aplicação

Os requisitos funcionais levantados para criação da ferramenta proposta pelo autor, são elas:

- [UC01] Importar arquivos com dados na ferramenta.
 - [AC01 - 01] Adicionar arquivos em formato CSV;

¹ <https://share.streamlit.io/rfsj/benford/main/streamlit.py>

- [AC01 - 02] Personalizar a escolha do separador de arquivos de planilhas: vírgula(",") e ponto e vírgula(";")
- [UC02] Mecanismo que identificação de anomalias;
 - [AC02 - 01] Utilizar a Lei de Newcomb-Benford como técnica de detecção de anomalias;
 - [AC02 - 02] Utilizar testes estatísticos para identificar a conformidades das bases de entrada com a Lei de Newcomb-Benford;
- [UC03] Dashboard com todas as informações obtidas com o método;
 - [AC03 - 01] Visualizar Informações sobre os dados;
 - [AC03 - 02] Visualizar com gráficos o padrão da Lei de Newcomb-Benford;
 - [AC03 - 03] Visualizar Informações sobre os testes estatísticos;

3.1.4 Versionamento da Ferramenta

A lista das versões das linguagens de programação e bibliotecas utilizadas no projeto, são elas:

- Python 3.9.12
- Conda 4.12.0
- Streamlit 1.4.0
- Plotly 5.5.0
- Pandas 1.4.1
- Scipy 1.8.0
- Numpy 1.20.3

3.1.5 Desenvolvimento dos códigos na aplicação

Uma breve explicação dos arquivos que compõem a ferramenta Bendata.

- No arquivo `calculateBenford.py` é onde estão os principais cálculos da Lei de Newcomb-Benford (CODEDROME, 2019).
- No arquivo `stats.py` encontram-se os testes estáticos da ferramenta (FONSECA, MARTINS, 1993).
- Já no arquivo `loadData.py` encontram-se as funções de manipulação de dados da planilha com a biblioteca pandas.
- O arquivo `functionBenford.py` possui funções para manipulação dos dados gerados pelos cálculos da lei.

- É no arquivo *streamlit.py* que está toda visualização da aplicação (textos, gráficos e imagens).

3.1.6 Funcionalidades

Este é o pacote de funcionalidades existentes na ferramenta para personalizar o processo de tomada de decisão (FIGURA 2).

- Entrada de arquivos de planilhas: A ferramenta permite a adição de arquivos com formato *Comma-separated Values* (CSV);
- Escolha de separador: É muito comum que arquivos de planilhas tenham formas diferentes de separar suas lacunas, o que poderia ser um grande problema durante a leitura dos dados. Devido a esse problema, foi criado um método para escolher o separador entre vírgula(","), e ponto e vírgula(";");
- Escolha de coluna: Devido ao fato de que a Lei de Newcomb-Benford utilizar de um conjunto de números para realizar o processo, nem todas as colunas de um arquivo de planilhas são úteis para a utilização da lei. Com isso, a ferramenta possui uma opção de escolher qual coluna será usada para gerar os resultados.
- Limpeza dos Dados: Um dos principais problemas encontrados durante a limpeza dos dados é como a Lei de Newcomb-Benford lida com números negativos. Nigrini (2012) ignora os sinais de números negativos, obtendo o módulo do número em questão. Contudo, na tentativa de eliminar um possível viés causado pela manipulação desses números, foram criadas três maneiras de lidar com esse problema na ferramenta.
 - A mudança do sinal negativo linha por linha;
 - A remoção da linha por completo;
 - A separação do conjunto negativo para uma nova análise.

Figura 2 - Funcionalidades

Upload the results of your experiment to see the importance of using the Newcomb-Benford Law

Upload file

Drag and drop file here

Limit 400MB per file • CSV

Browse files

benford1.csv

5.7KB

You selected `UploadedFile(id=1, name='benford1.csv', type='text/csv', size=5811)`

What's your file separate?

☒ ;

☐ ,

Select column:

example

What will be the negative protocol?

☒ remove negative from each cell

☐ remove lines with negatives

☐ separate negative lines for analysis

☐ no conditions

Fonte: Elaborada pelo autor.

3.1.7 Verificação e Implantação da Ferramenta

Para verificação, a ferramenta demonstra três testes de análise estatística encontrados na ferramenta Bendata, são eles:

- A Estatística Z;
- Qui-Quadrado;
- Desvio Médio Absoluto.

Para implantação, com o intuito de acelerar o processo de implantação e disponibilização online, foi escolhida a biblioteca *streamlit* que permitiu a disponibilidade da aplicação em um domínio online², e o controle de versão é feito a partir do github dentro de um repositório³.

² <https://share.streamlit.io/rfsj/benford/main/streamlit.py>

³ <https://github.com/rfsj/benford/>

4 PROPOSTA DO TCC

Existe um duplo objetivo neste trabalho de graduação, onde o autor propõe a criação de uma ferramenta, o Bendata, abordada no capítulo anterior, que utilize da Lei de Newcomb-Benford para permitir a descoberta de possíveis anomalias nas bases de dados e a utilização dessa ferramenta na Cota para o Exercício da Atividade Parlamentar, onde espera-se que haja uma melhora na confiança dessa informação.

A questão que norteia este trabalho é a seguinte: Quais bases poderiam ser utilizadas para extrair um conhecimento útil e relevante nos últimos anos? Com esse intuito, foi proposto a utilização dessa ferramenta em cotas de gastos públicos.

Tendo isso em vista, os dados escolhidos são da Câmara dos Deputados Federais e de domínio público, as bases são referentes às despesas pela cota para exercício da atividade parlamentar devido ao seu caráter contábil, e são disponibilizadas no próprio site da Câmara dos Deputados. São analisados os dados das despesas pela cota para exercício da atividade parlamentar entre os anos de 2008 e 2021⁴.

4.1 COLETA DOS DADOS E PREPARAÇÃO DOS DADOS PARA ANÁLISE

Os dados foram coletados no site de dados abertos da Câmara dos Deputados Federais. Para aplicação do teste da Lei de Newcomb-Benford, os dados encontram-se dividido por anos, e contém 31 colunas de dados com a explicação, conforme está descrito no Anexo – A, de cada um dos campos⁵. A escolha das colunas se deu seguindo os críticos propostos por Kessel E. (2020), onde os dados precisam ser numéricos, grandes, gerados aleatoriamente e precisam representar magnitudes de eventos.

Levando em consideração os tipos das despesas, conforme está descrito no Anexo – B, os resultados obtidos foram gerados a partir dos campos `vlrDocumento` (Valor do documento) e `vlrGlosa` (Valor da Glosa) encontrados dentro dessas bases de dados, sendo elas:

- a) `vlrDocumento`: O seu conteúdo é o valor de face do documento fiscal ou o valor do documento que deu causa à despesa(Câmara dos deputados).

⁴<https://dadosabertos.camara.leg.br/swagger/api.html#staticfile>

⁵<https://www2.camara.leg.br/transparencia/cota-para-exercicio-da-atividade-parlamentar/explicacoes-so-bre-o-formato-dos-arquivos-xml>

- b) vlrGlosa: O seu conteúdo representa o valor da glosa do documento fiscal que incidirá sobre o Valor do Documento, ou o valor da glosa do documento que deu causa à despesa(Câmara dos deputados).

As colunas foram utilizadas em sua totalidade, não diferenciando os tipos de despesas, nem o autor das despesas. O intuito é aumentar a confiança na base por completo.

4.2 TRATAMENTO DOS DADOS

Em todos os anos foram adotados os seguintes procedimentos para análise dos dados:

- a) Localização da coluna numérica;
- b) Exclusão de dados em branco: No campo vlrDocumento não foram encontrados valores nulos. Já no campo vlrGlosa, devido ao caráter eletivo da glosa, existem campos com valores nulos;
- c) Exclusão de valores duplicados;
- d) Tratamento de números negativos: Segundo Nigrini (2012), a melhor escolha é ignorar os sinais de números negativos, obtendo o módulo do número em questão.

4.3 ANÁLISE ESTATÍSTICA

Os testes de conformidade demonstraram que levam em consideração o número de registros que funcionam bem bases pequenas, mas sofrem de um sério problema com grandes conjuntos de dados. Em seu livro, Nigrini (2012), mostrou simulações com bases de 1000 e 25000 registros retiradas de uma base de Newcomb-Benford; contudo, na prática, possui-se grandes conjuntos de dados e, na maioria dos casos, as estatísticas não indicam conformidade.

As estatísticas Z e Qui-Quadrado sofrem com o tamanho das bases, pois ambas levam em consideração o número de registros da base de dados. Já o teste de desvio absoluto médio não considera o número de registros em seus cálculos; logo, não se sujeita aos problemas com grandes bases de dados. Esse teste não possui valores críticos objetivos, mas o autor propõe uma tabela de valores criada a partir dos resultados de suas análises, podendo essa ser utilizada como parâmetro para comparações. A conclusão é que a melhor solução é o teste desvio absoluto médio, pois não considera o número de registros (NIGRINI, 2012).

5 RESULTADOS

5.1 ANÁLISE DAS DESPESAS PELA CONTA PARA EXERCÍCIO DA ATIVIDADE PARLAMENTAR

Essa análise é realizada a partir dos dados abertos obtidos no site oficial da Câmara dos Deputados.

5.1.1 Ano 2008

a) Campo vlrDocumento

Tabela 4 – Ano 2008 do campo vlrDocumento, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	105	0.23596	133.945	0.301	2.93971	6.2549	0.00723
2	69	0.15506	78.32	0.176	1.09791	1.10907	0.00233
3	46	0.10337	55.625	0.125	1.30796	1.66545	0.0024
4	51	0.11461	43.165	0.097	1.17487	1.42215	0.00196
5	53	0.1191	35.155	0.079	3.04825	9.05829	0.00446
6	27	0.06067	29.815	0.067	0.43893	0.26578	0.0007
7	35	0.07865	25.81	0.058	1.76238	3.27222	0.00229
8	29	0.06517	22.695	0.051	1.25085	1.75162	0.00157
9	28	0.06292	20.47	0.046	1.59082	2.76995	0.00188
Soma		1		1		27.56943	0.02482

Z crítico = 1.96 α = 0.05

Σ DAM

ΣX^2 = 15.507 α = 0.05

Perto

Aceitável

Marginalmente aceitável

Não conformidade

Grau de Liberdade = 8

0.000 a 0.006

0.006 a 0.012

0.012 a 0.015

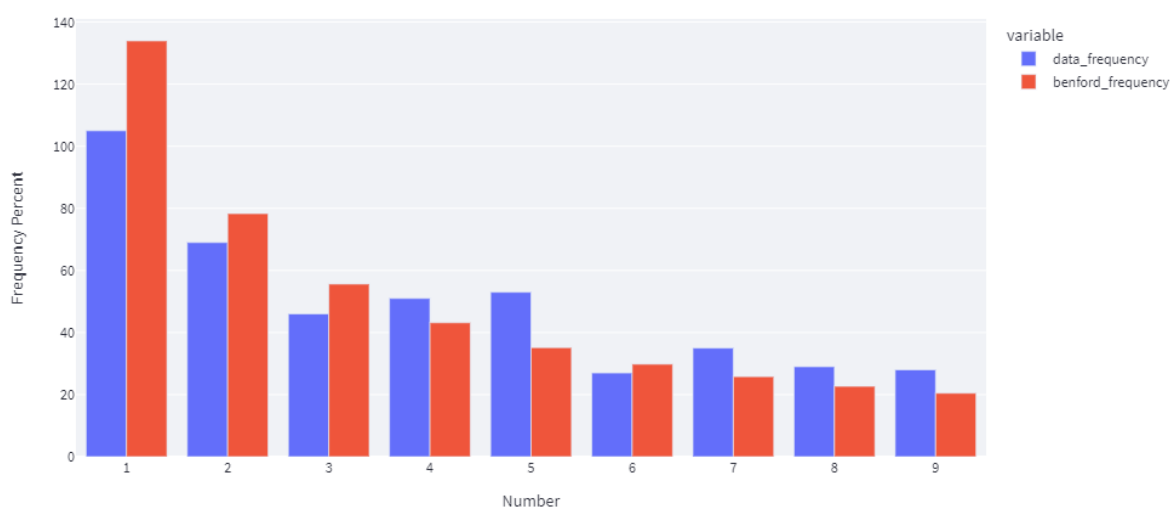
Acima 0.015

Fonte: Elaborada pelo autor.

A análise ocorreu em 445 registros. O teste Z demonstrou que para os dígitos 1 e 5, o valor ultrapassou significativamente o Z crítico, com isso, essa evidência pode rejeitar a hipótese nula do teste Z. No caso do teste X², demonstrou que é que o valor ultrapassou o

valor estipulado , podendo assim rejeitar a hipótese nula. Por fim, o DAM gerado foi de 0.02482, mostrando assim uma não conformidade com a Lei de Newcomb-Benford.

Figura 3 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2008



Fonte: Elaborada pelo autor.

b) Campo vlrGlosa

Tabela 5 – Ano 2008 do campo vlrGlosa, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	10	0.29412	10.234	0.301	0.08749	0.00535	0.00076
2	6	0.17647	5.984	0.176	0.00721	0.00004	0.00005
3	2	0.05882	4.25	0.125	0.90749	1.19118	0.00735
4	6	0.17647	3.298	0.097	1.27599	2.21371	0.00883
5	2	0.05882	2.686	0.079	0.11826	0.1752	0.00224
6	2	0.05882	2.278	0.067	0.19069	0.03393	0.00091
7	1	0.02941	1.972	0.058	0.34631	0.4791	0.00318
8	2	0.05882	1.734	0.051	0.20736	0.04081	0.00087

9	3	0.08824	1.564	0.046	0.76627	1.31848	0.00469
Soma		1		1		5.4578	0.02888

Z crítico = 1.96 α = 0.05

Σ DAM

ΣX^2 = 15.507 α = 0.05

Perto

Aceitável

Marginalmente aceitável

Não conformidade

Grau de Liberdade = 8

0.000 a 0.006

0.006 a 0.012

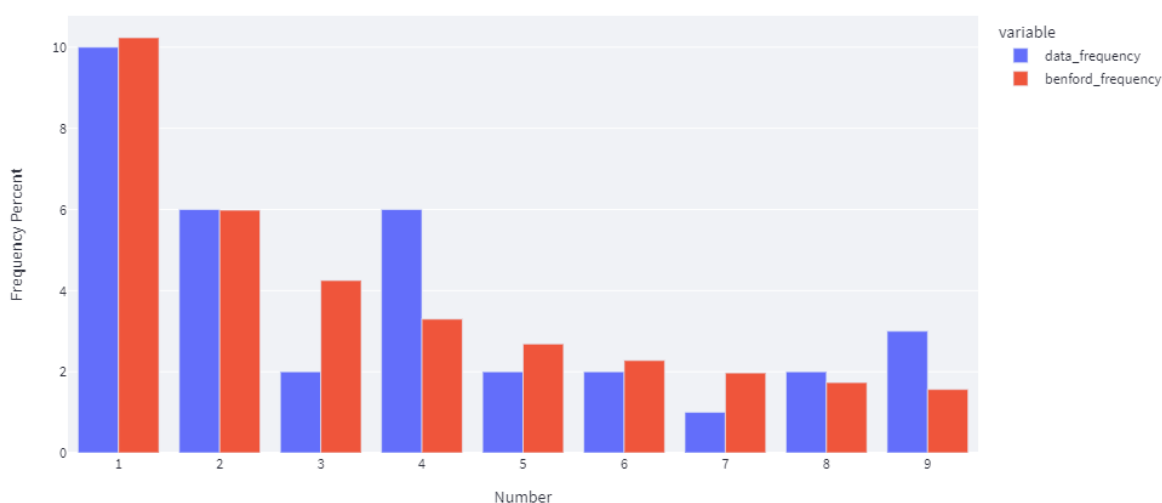
0.012 a 0.015

Acima 0.015

Fonte: Elaborada pelo autor.

A análise ocorreu em 34 registros. No teste Z, não houve valores que ultrapassassem significativamente o Z crítico, com isso, não é possível rejeitar a hipótese nula do teste Z. No caso do teste X^2 , o valor não ultrapassou o valor estipulado, sendo assim, não podendo rejeitar a hipótese nula. Por fim, o DAM gerado foi de 0.02888, mostrando assim uma não conformidade com a Lei de Newcomb-Benford.

Figura 4 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2008



Fonte: Elaborada pelo autor.

5.1.2 Ano 2009

a) Campo vlrDocumento

Tabela 6 – Ano 2009 do campo vlrDocumento, Teste Z, X^2 e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X^2	DAM
--------	-----------------	-----------------	----------------	----------------	---------	-------	-----

1	75141	0.28334	79824.297	0.301	19.82436	274.76936	0.00196
2	37645	0.14195	46674.672	0.176	46.04089	1746.8784	0.00378
3	37251	0.14047	33149.625	0.125	24.07871	507.43491	0.00172
4	25160	0.09487	25724.109	0.097	3.69797	12.37046	0.00024
5	29048	0.10953	20950.563	0.079	58.28987	3129.67656	0.00339
6	20817	0.0785	17768.199	0.067	23.67529	523.13617	0.00128
7	16458	0.06206	15381.426	0.058	8.93961	75.35137	0.00045
8	12555	0.04734	13525.047	0.051	8.55788	69.57397	0.00041
9	10892	0.04107	12199.062	0.046	12.11135	140.04446	0.00055
Soma		1		1		6479.23566	0.01378

Z crítico = 1.96 $\alpha = 0.05$

Σ DAM

$\Sigma X^2 = 15.507$ $\alpha = 0.05$

Perto

Aceitável

Marginalmente aceitável

Não conformidade

Grau de Liberdade = 8

0.000 a 0.006

0.006 a 0.012

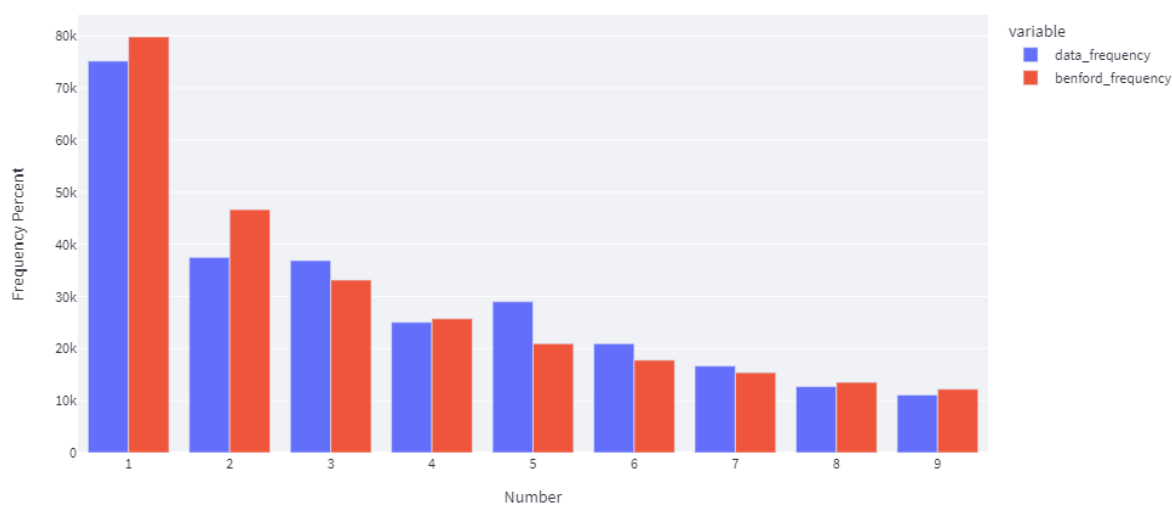
0.012 a 0.015

Acima 0.015

Fonte: Elaborada pelo autor.

A análise ocorreu em 265197 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.01378, sendo considerado marginalmente aceitável conforme a Lei de Newcomb-Benford.

Figura 5 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2009



Fonte: Elaborada pelo autor.

b) Campo vlrGlosa

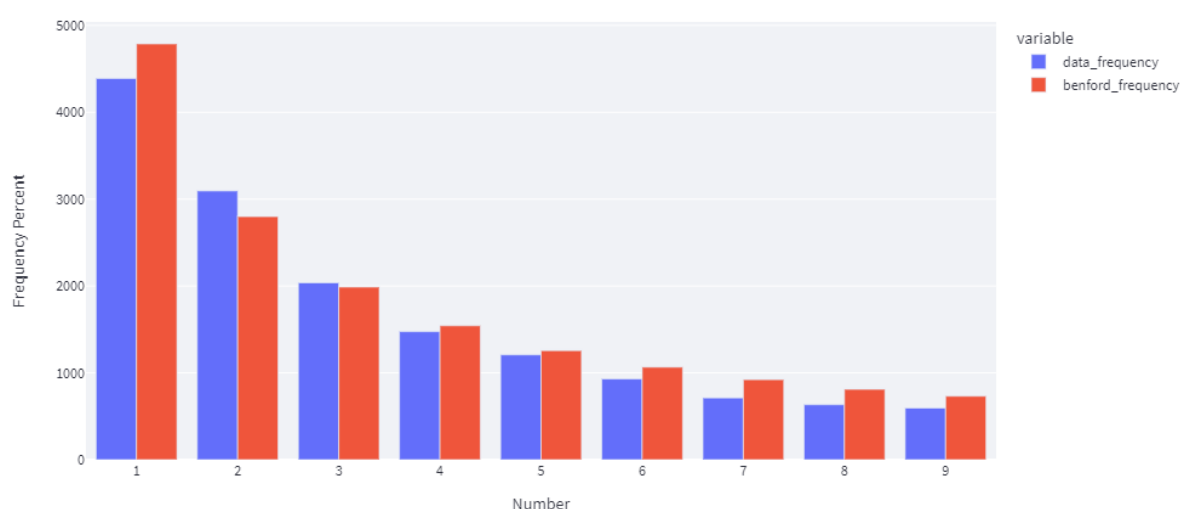
Tabela 7 – Ano 2009 do campo vlrGlosa, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	4390	0.27598	4788.007	0.301	6.87114	33.08466	0.00278
2	3095	0.19457	2799.632	0.176	6.13923	31.16204	0.00206
3	2037	0.12806	1988.375	0.125	1.15377	1.18911	0.00034
4	1475	0.09273	1542.979	0.097	1.80778	2.99495	0.00047
5	1208	0.07594	1256.653	0.079	1.41542	1.88367	0.00034
6	930	0.05846	1065.769	0.067	4.28969	17.2957	0.00095
7	712	0.04476	922.606	0.058	7.12697	48.07565	0.00147
8	634	0.03986	811.257	0.051	6.37036	38.73007	0.00124
9	595	0.0374	731.722	0.046	5.15585	25.54646	0.00096
Soma		1		1		199.96231	0.01061
Z crítico = 1.96 α = 0.05		Σ DAM					
Σ X² = 15.507 α = 0.05		Perto	Aceitável	Marginalmente aceitável		Não conformidade	
Grau de Liberdade = 8		0.000 a 0.006	0.006 a 0.012	0.012 a 0.015		Acima 0.015	

Fonte: Elaborada pelo autor.

A análise ocorreu em 15907 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.01061, sendo considerado aceitável conforme a Lei de Newcomb-Benford.

Figura 6 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2009



Fonte: Elaborada pelo autor.

5.1.3 Ano 2010

a) Campo vlrDocumento

Tabela 8 – Ano 2010 do campo vlrDocumento, Teste Z, X^2 e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X^2	DAM
1	97273	0.27853	105119.735	0.301	28.94547	585.72494	0.0025
2	51859	0.14849	61465.36	0.176	42.68325	1501.36845	0.00306
3	51226	0.14668	43654.375	0.125	38.73847	1313.2591	0.00241
4	35926	0.10287	33875.795	0.097	11.71933	124.08094	0.00065
5	37413	0.10713	27589.565	0.079	61.62246	3497.6947	0.00313
6	25186	0.07212	23398.745	0.067	12.09283	136.51503	0.00057

7	18198	0.05211	20255.63	0.058	14.89236	209.02047	0.00065
8	18004	0.05155	17810.985	0.051	1.48077	2.09167	6.00E-05
9	13942	0.03992	16064.81	0.046	17.1434	280.50891	0.00068
Soma		1		1		7650.26421	0.01371

Z crítico = 1.96 α = 0.05

Σ DAM

ΣX^2 = 15.507 α = 0.05

Perto

Aceitável

Marginalmente aceitável

Não conformidade

Grau de Liberdade = 8

0.000 a 0.006

0.006 a 0.012

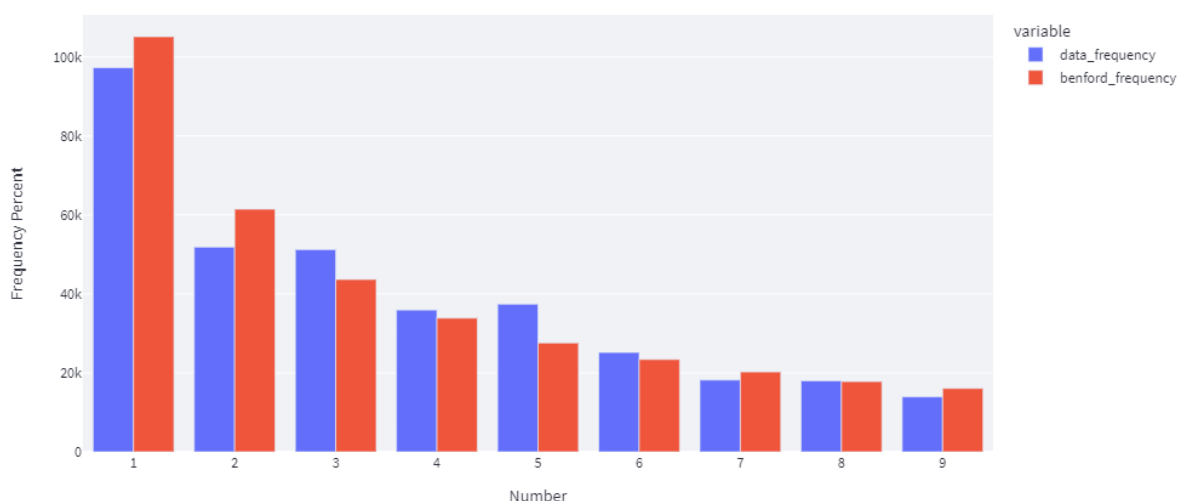
0.012 a 0.015

Acima 0.015

Fonte: Elaborada pelo autor.

A análise ocorreu em 349235 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.01371, sendo considerado marginalmente aceitável conforme a Lei de Newcomb-Benford.

Figura 7 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2010



Fonte: Elaborada pelo autor.

b) Campo vlrGlosa

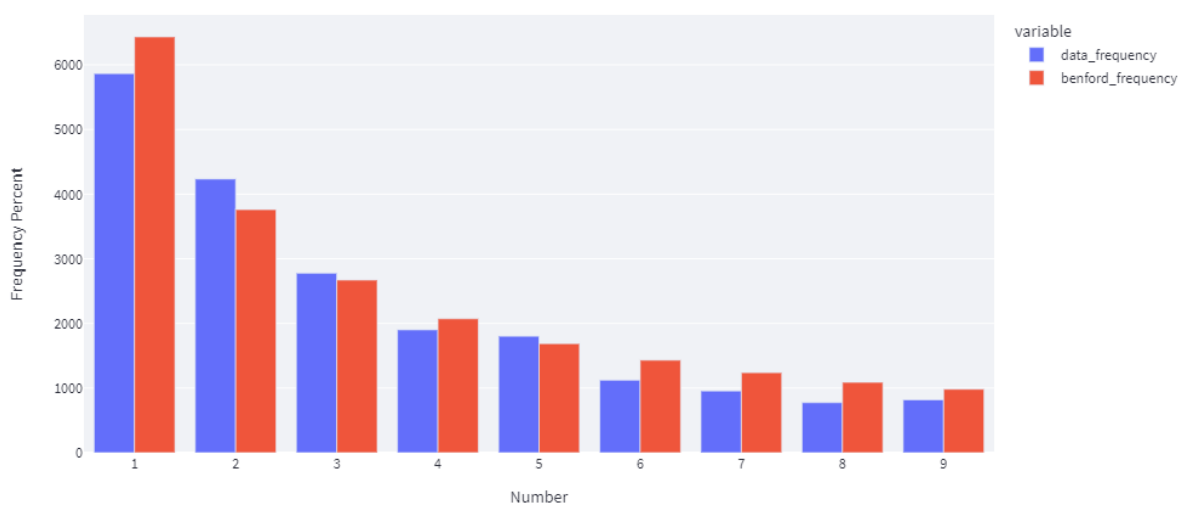
Tabela 9 – Ano 2010 do campo vlrGlosa, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	5865	0.27444	6432.671	0.301	8.45824	50.09589	0.00295
2	4234	0.19812	3761.296	0.176	8.48198	59.40747	0.00246
3	2778	0.12999	2671.375	0.125	2.19506	4.25582	0.00055
4	1901	0.08895	2072.987	0.097	3.96359	14.26904	0.00089
5	1803	0.08437	1688.309	0.079	2.89585	7.79124	0.0006
6	1123	0.05255	1431.857	0.067	8.43652	66.62163	0.00161
7	956	0.04473	1239.518	0.058	8.28251	64.84977	0.00147
8	775	0.03626	1089.921	0.051	9.77644	90.99305	0.00164
9	819	0.03832	983.066	0.046	5.34106	27.38133	0.00085
Soma		1		1		385.66524	0.01302
Z crítico = 1.96 α = 0.05		Σ DAM					
$\Sigma X^2 = 15.507$ α = 0.05		Perto	Aceitável	Marginalmente aceitável		Não conformidade	
Grau de Liberdade = 8		0.000 a 0.006	0.006 a 0.012	0.012 a 0.015		Acima 0.015	

Fonte: Elaborada pelo autor.

A análise ocorreu em 21371 registros. Tanto o teste Z quanto o teste X² possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.01302, sendo considerado marginalmente aceitável conforme a Lei de Newcomb-Benford.

Figura 8 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2010



Fonte: Elaborada pelo autor.

5.1.4 Ano 2011

a) Campo vlrDocumento

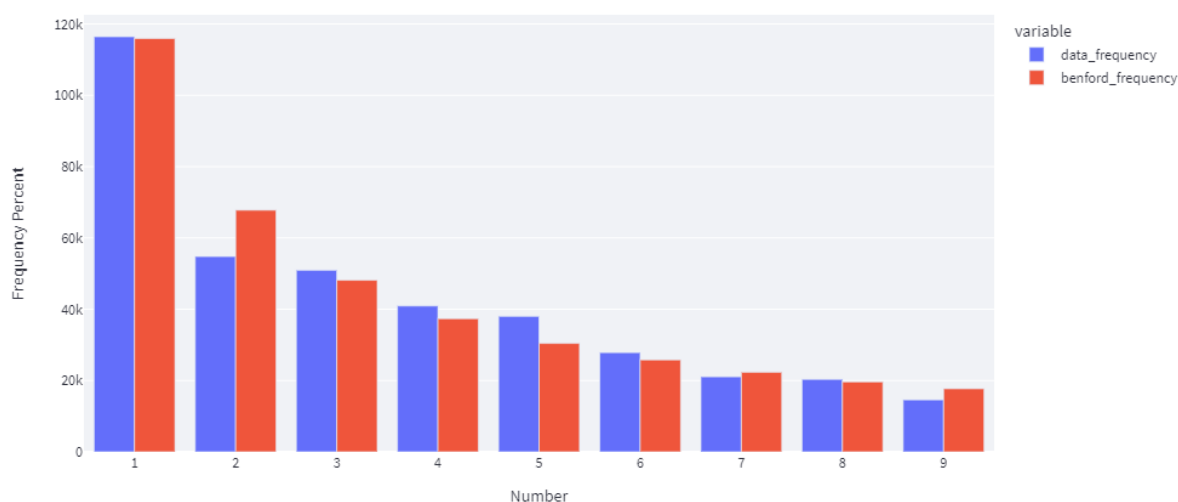
Tabela 10 – Ano 2011 do campo vlrDocumento, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	116430	0.30228	115935.869	0.301	1.73402	2.10604	0.00014
2	54796	0.14226	67789.744	0.176	54.97585	2490.60364	0.00375
3	50946	0.13227	48146.125	0.125	13.63881	162.82307	0.00081
4	40923	0.10625	37361.393	0.097	19.38785	339.52279	0.00103
5	38022	0.09872	30428.351	0.079	45.3579	1895.0585	0.00219
6	27855	0.07232	25806.323	0.067	13.19969	162.63756	0.00059
7	21074	0.05471	22339.802	0.058	8.72226	71.72197	0.00037
8	20352	0.05284	19643.619	0.051	5.18461	25.54538	0.0002
9	14595	0.03789	17717.774	0.046	24.01553	550.3918	0.0009
Soma		1		1		5700.41075	0.00998
Z crítico = 1.96 α = 0.05		Σ DAM					
Σ X² = 15.507 α = 0.05		Perto	Aceitável	Marginalmente aceitável		Não conformidade	
Grau de Liberdade = 8		0.000 a 0.006	0.006 a 0.012	0.012 a 0.015		Acima 0.015	

Fonte: Elaborada pelo autor.

A análise ocorreu em 385169 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.00998, sendo considerado aceitável conforme a Lei de Newcomb-Benford.

Figura 9 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2011



Fonte: Elaborada pelo autor.

b) Campo vlrGlosa

Tabela 11 – Ano 2011 do campo vlrGlosa, Teste Z, X^2 e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X^2	DAM
1	5308	0.27268	5859.266	0.301	8.60611	51.86558	0.00315
2	3567	0.18324	3426.016	0.176	2.64404	5.80163	0.0008
3	2582	0.13264	2433.25	0.125	3.2129	9.09342	0.00085
4	1784	0.09165	1888.202	0.097	2.51142	5.75047	0.00059
5	1605	0.08245	1537.814	0.079	1.77196	2.93531	0.00038

6	1101	0.05656	1304.222	0.067	5.81145	31.66576	0.00116
7	924	0.04747	1129.028	0.058	6.27155	37.23245	0.00117
8	800	0.0411	992.766	0.051	6.26392	37.4295	0.0011
9	731	0.03755	895.436	0.046	5.60897	30.19668	0.00094
Soma		1		1		211.9708	0.01014

Z crítico = 1.96 α = 0.05

Σ DAM

$\Sigma X^2 = 15.507$ α = 0.05

Perto

Aceitável

Marginalmente aceitável

Não conformidade

Grau de Liberdade = 8

0.000 a 0.006

0.006 a 0.012

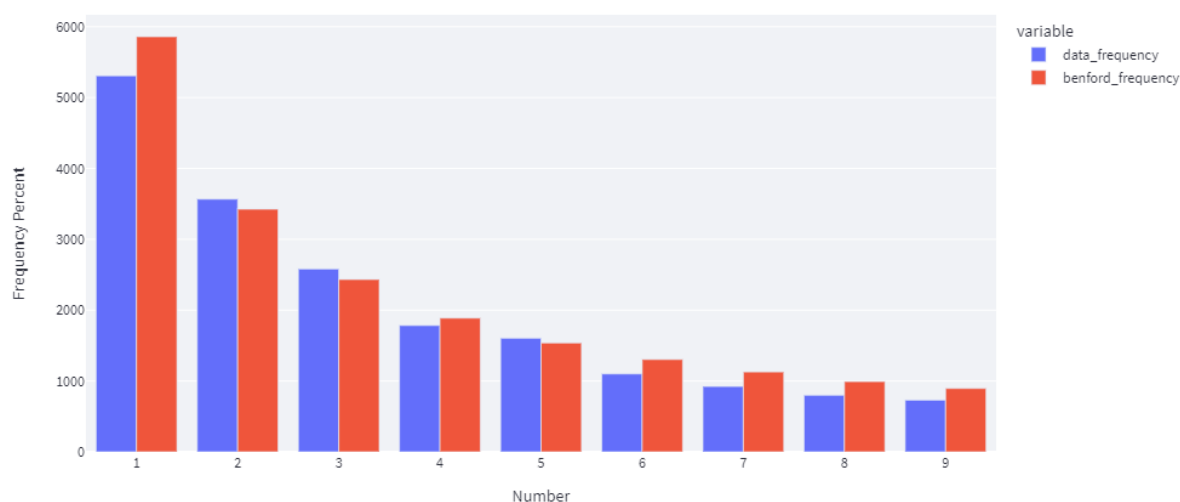
0.012 a 0.015

Acima 0.015

Fonte: Elaborada pelo autor.

A análise ocorreu em 19466 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.01014, sendo considerado aceitável conforme a Lei de Newcomb-Benford.

Figura 10 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2011



Fonte: Elaborada pelo autor.

5.1.5 Ano 2012

a) Campo vlrDocumento

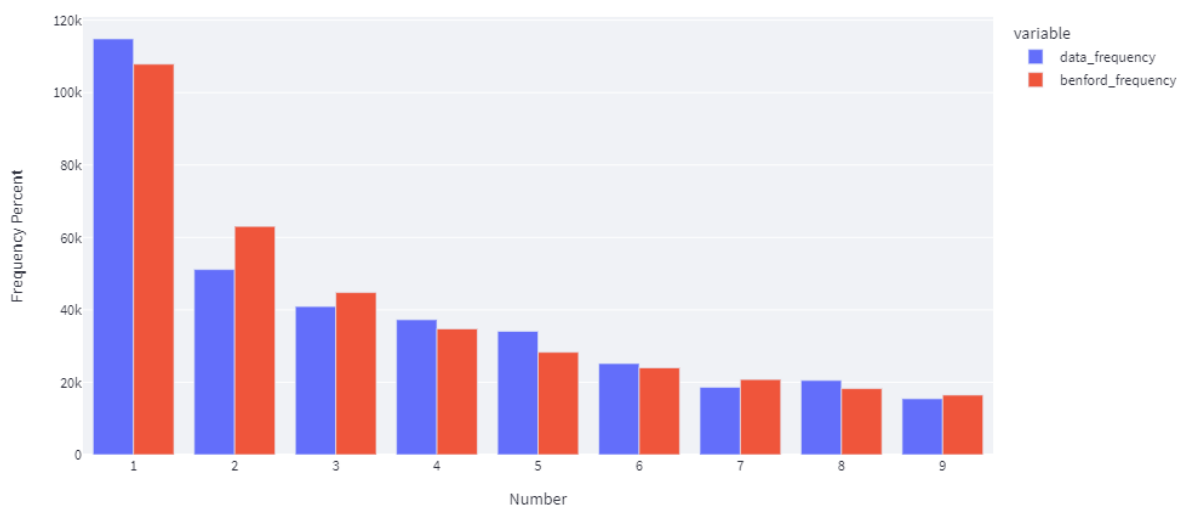
Tabela 12 – Ano 2012 do campo vlrDocumento, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	114844	0.32048	107861.845	0.301	25.42646	451.97158	0.00216
2	51143	0.14272	63068.72	0.176	52.31133	2255.04493	0.0037
3	40945	0.11426	44793.125	0.125	19.43493	330.58792	0.00119
4	37287	0.10405	34759.465	0.097	14.26366	183.78974	0.00078
5	34079	0.0951	28309.255	0.079	35.72934	1175.93901	0.00179
6	25184	0.07028	24009.115	0.067	7.8466	57.49295	0.00036
7	18616	0.05195	20784.01	0.058	15.4907	226.14824	0.00067
8	20526	0.05728	18275.595	0.051	17.08423	277.1085	0.0007
9	15466	0.04316	16483.87	0.046	8.11288	62.85292	0.00032
Soma		1		1		5020.93579	0.01167
Z crítico = 1.96 $\alpha = 0.05$		Σ DAM					
$\Sigma X^2 = 15.507 \alpha = 0.05$		Perto	Aceitável	Marginalmente aceitável		Não conformidade	
Grau de Liberdade = 8		0.000 a 0.006	0.006 a 0.012	0.012 a 0.015		Acima 0.015	

Fonte: Elaborada pelo autor.

A análise ocorreu em 358345 registros. Tanto o teste Z quanto o teste X² possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.01167, sendo considerado aceitável conforme a Lei de Newcomb-Benford.

Figura 11 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2012



Fonte: Elaborada pelo autor.

b) Campo vlrGlosa

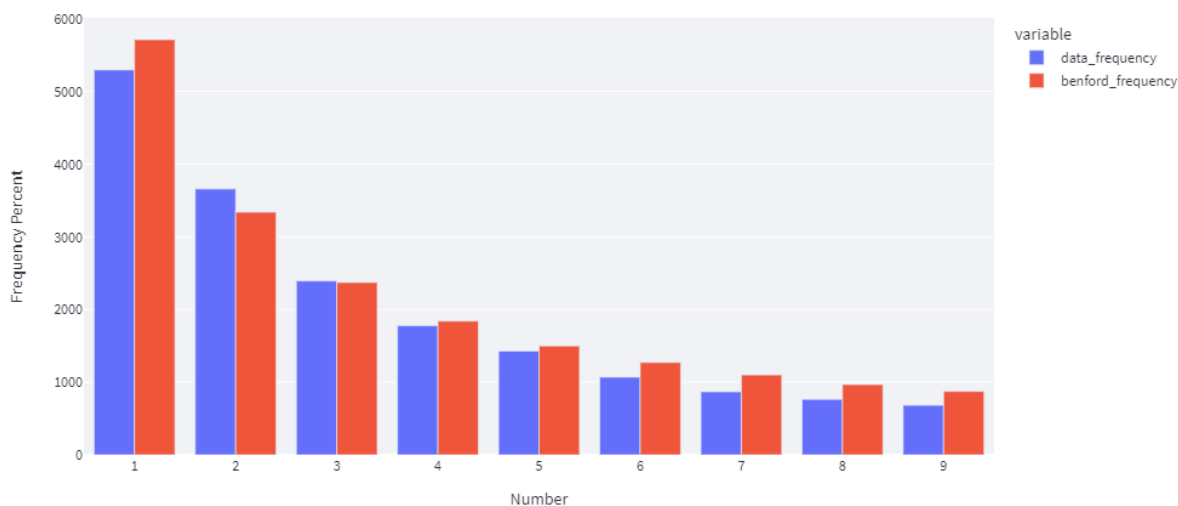
Tabela 13 – Ano 2012 do campo vlrGlosa, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	5294	0.27907	5709.97	0.301	6.57634	30.30332	0.00244
2	3658	0.19283	3338.72	0.176	6.07767	30.53257	0.00187
3	2392	0.12609	2371.25	0.125	0.44456	0.18158	0.00012
4	1777	0.09367	1840.09	0.097	1.53547	2.16313	0.00037
5	1428	0.07528	1498.63	0.079	1.88767	3.32877	0.00041
6	1069	0.05635	1270.99	0.067	5.85116	32.10093	0.00118
7	868	0.04576	1100.26	0.058	7.19889	49.02905	0.00136
8	764	0.04027	967.47	0.051	6.69854	42.79207	0.00119
9	683	0.036	872.62	0.046	6.55467	41.20436	0.00111
Soma		1		1		231.63578	0.01005
Z crítico = 1.96 α = 0.05		Σ DAM					
Σ X² = 15.507 α = 0.05		Perto	Aceitável	Marginalmente aceitável		Não conformidade	
Grau de Liberdade = 8		0.000 a 0.006	0.006 a 0.012	0.012 a 0.015		Acima 0.015	

Fonte: Elaborada pelo autor.

A análise ocorreu em 18970 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.01005, sendo considerado aceitável conforme a Lei de Newcomb-Benford.

Figura 12 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2012



Fonte: Elaborada pelo autor.

5.1.6 Ano 2013

a) Campo vlrDocumento

Tabela 14 – Ano 2013 do campo vlrDocumento, Teste Z, X^2 e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X^2	DAM
1	127728	0.32623	117849.928	0.301	34.41494	827.97086	0.0028
2	51068	0.13043	68908.928	0.176	74.86932	4619.12152	0.00506
3	44352	0.11328	48941	0.125	22.1733	430.292	0.0013
4	40448	0.10331	37978.216	0.097	13.33398	160.614	0.0007

5	35925	0.09176	30930.712	0.079	29.5873	806.41249	0.00142
6	24278	0.06201	26232.376	0.067	12.48928	145.60578	0.00055
7	23457	0.05991	22708.624	0.058	5.11339	24.66317	0.00021
8	24509	0.0626	19967.928	0.051	32.98458	1032.72282	0.00129
9	17706	0.04522	18010.288	0.046	2.31759	5.14102	9.00E-05
Soma		1		1		8052.54366	0.01342

Z crítico = 1.96 α = 0.05 Σ DAM ΣX^2 = 15.507 α = 0.05

Perto

Aceitável

Marginalmente aceitável

Não conformidade

Grau de Liberdade = 8

0.000 a 0.006

0.006 a 0.012

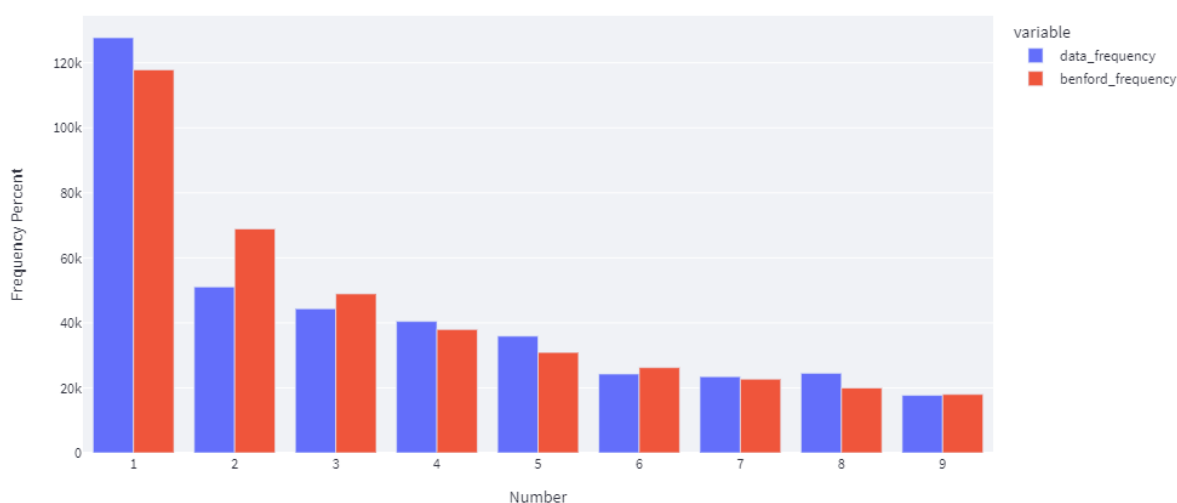
0.012 a 0.015

Acima 0.015

Fonte: Elaborada pelo autor.

A análise ocorreu em 391528 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.01342, sendo considerado marginalmente aceitável conforme a Lei de Newcomb-Benford.

Figura 13 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2013



Fonte: Elaborada pelo autor.

b) Campo vlrGlosa

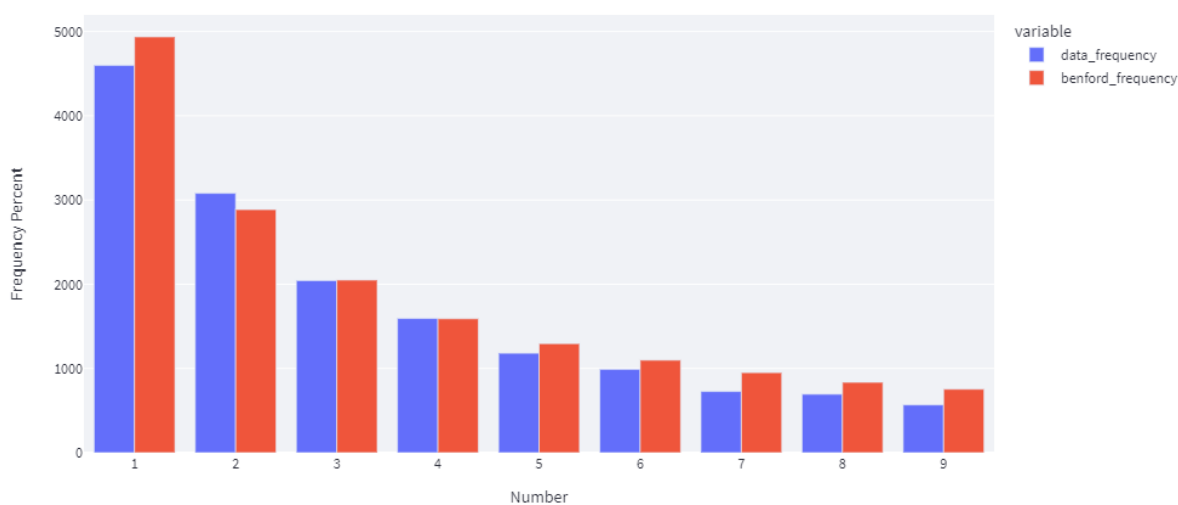
Tabela 15 – Ano 2013 do campo vlrGlosa, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	4599	0.28041	4936.701	0.301	5.74026	23.10085	0.00229
2	3080	0.18779	2886.576	0.176	3.95577	12.96098	0.00131
3	2044	0.12463	2050.125	0.125	0.13281	0.0183	4.00E-05
4	1594	0.09719	1590.897	0.097	0.06868	0.00605	2.00E-05
5	1182	0.07207	1295.679	0.079	3.27633	9.97386	0.00077
6	991	0.06042	1098.867	0.067	3.35318	10.58844	0.00073
7	727	0.04433	951.258	0.058	7.47488	52.86857	0.00152
8	695	0.04238	836.451	0.051	5.00282	23.92057	0.00096
9	568	0.03463	754.446	0.046	6.93104	46.07634	0.00126
Soma		1		1		179.51396	0.0089
Z crítico = 1.96 α = 0.05		Σ DAM					
Σ X² = 15.507 α = 0.05		Perto	Aceitável	Marginalmente aceitável		Não conformidade	
Grau de Liberdade = 8		0.000 a 0.006	0.006 a 0.012	0.012 a 0.015		Acima 0.015	

Fonte: Elaborada pelo autor.

A análise ocorreu em 16401 registros. Tanto o teste Z quanto o teste X² possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.0089, sendo considerado aceitável conforme a Lei de Newcomb-Benford.

Figura 14 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2013



Fonte: Elaborada pelo autor.

5.1.7 Ano 2014

a) Campo vlrDocumento

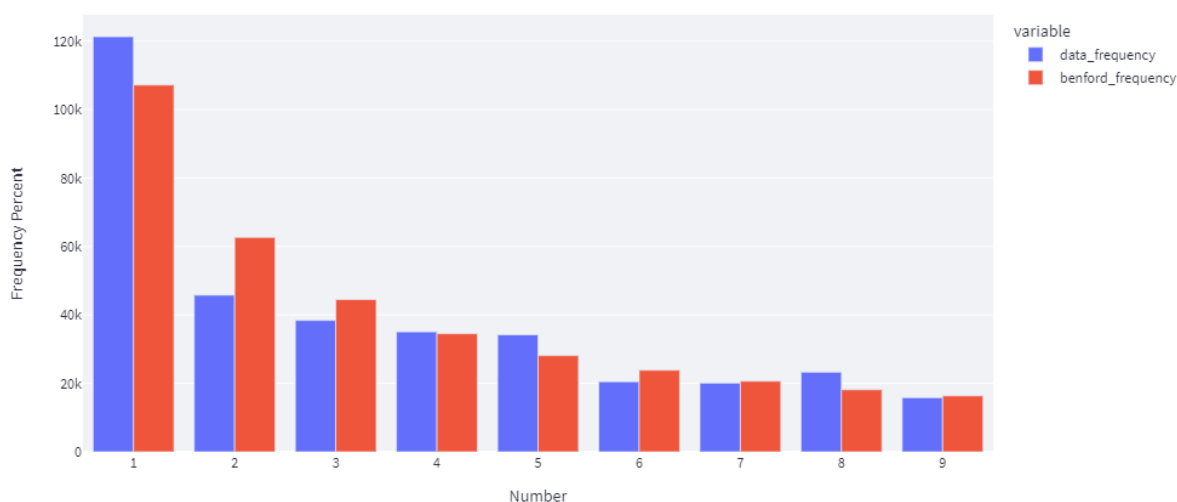
Tabela 16 – Ano 2014 do campo vlrDocumento, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	121300	0.34064	107183.692	0.301	51.57067	1859.14618	0.0044
2	45784	0.12857	62672.192	0.176	74.31376	4550.83858	0.00527
3	38412	0.10787	44511.5	0.125	30.90427	835.8267	0.0019
4	35074	0.0985	34540.924	0.097	3.01558	8.22705	0.00017
5	34204	0.09605	28131.268	0.079	37.72453	1310.92825	0.00189
6	20505	0.05758	23858.164	0.067	22.47143	471.27301	0.00105
7	20106	0.05646	20653.336	0.058	3.92046	14.505	0.00017
8	23294	0.06542	18160.692	0.051	39.09811	1450.98276	0.0016
9	15840	0.04448	16380.232	0.046	4.31762	17.81725	0.00017
Soma		1		1		10519.54478	0.01662
Z crítico = 1.96 $\alpha = 0.05$		Σ DAM					
Σ X² = 15.507 $\alpha = 0.05$		Perto	Aceitável	Marginalmente aceitável		Não conformidade	
Grau de Liberdade = 8		0.000 a 0.006	0.006 a 0.012	0.012 a 0.015		Acima 0.015	

Fonte: Elaborada pelo autor.

A análise ocorreu em 356092 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.01662, mostrando assim uma não conformidade com a Lei de Newcomb-Benford.

Figura 15 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2014



Fonte: Elaborada pelo autor.

b) Campo vlrGlosa

Tabela 17 – Ano 2014 do campo vlrGlosa, Teste Z, X^2 e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X^2	DAM
1	4124	0.26826	4627.273	0.301	8.84038	54.73714	0.00364
2	2846	0.18513	2705.648	0.176	2.96189	7.28058	0.00101
3	1978	0.12867	1921.625	0.125	1.36263	1.65388	0.00041
4	1485	0.0966	1491.181	0.097	0.15482	0.02562	0.00004
5	1139	0.07409	1214.467	0.079	2.24154	4.68952	0.00055

6	1009	0.06563	1029.991	0.067	0.66101	0.42779	0.00015
7	741	0.0482	891.634	0.058	5.18036	25.44834	0.00109
8	678	0.0441	784.023	0.051	3.86856	14.33743	0.00077
9	553	0.03597	707.158	0.046	5.91593	33.60591	0.00111
Soma		1		1		142.20621	0.00877

Z crítico = 1.96 $\alpha = 0.05$

Σ DAM

$\Sigma X^2 = 15.507$ $\alpha = 0.05$

Perto

Aceitável

Marginalmente aceitável

Não conformidade

Grau de Liberdade = 8

0.000 a 0.006

0.006 a 0.012

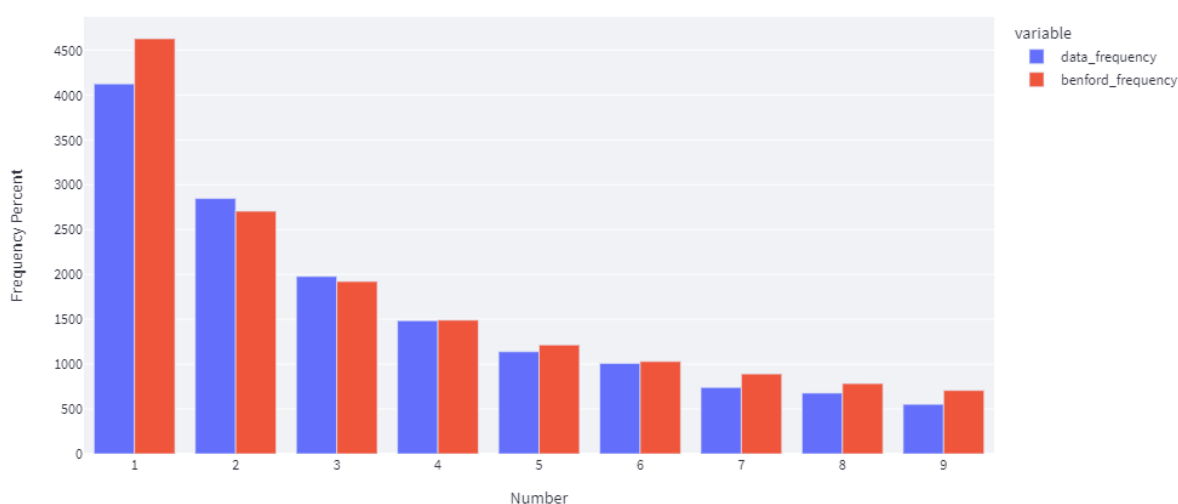
0.012 a 0.015

Acima 0.015

Fonte: Elaborada pelo autor.

A análise ocorreu em 15373 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.00877, sendo considerado aceitável conforme a Lei de Newcomb-Benford.

Figura 16 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2014



Fonte: Elaborada pelo autor.

5.1.8 Ano 2015

a) Campo vlrDocumento

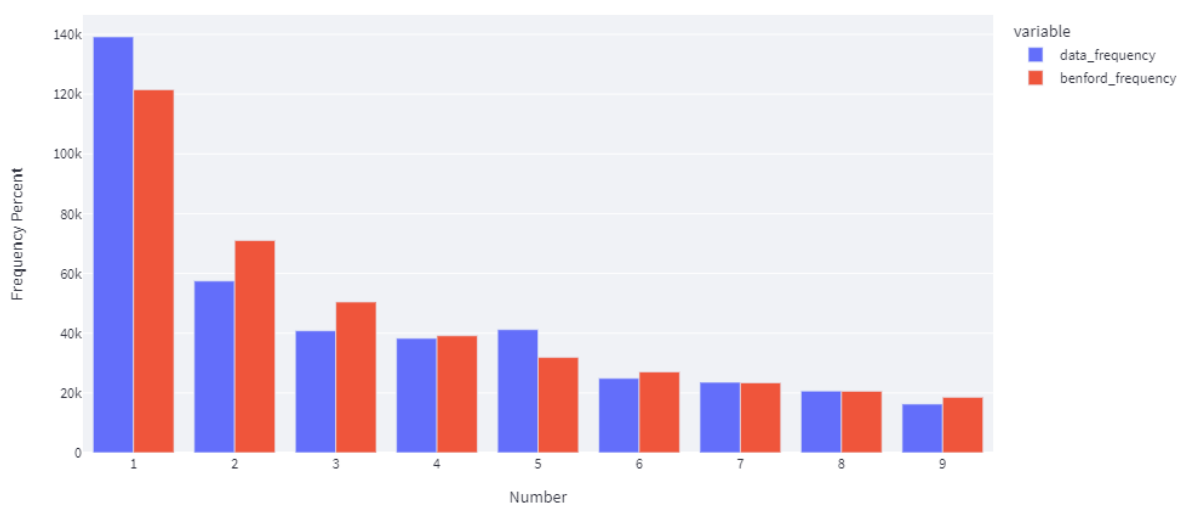
Tabela 18 – Ano 2015 do campo vlrDocumento, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	139167	0.34479	121492.63	0.301	60.64823	2571.21238	0.00487
2	57463	0.14237	71038.88	0.176	56.11001	2594.41756	0.00374
3	40816	0.10112	50453.75	0.125	45.86722	1841.01727	0.00265
4	38278	0.09483	39152.11	0.097	4.64618	19.51538	0.00024
5	41197	0.10207	31886.77	0.079	54.32528	2718.38078	0.00256
6	24897	0.06168	27043.21	0.067	13.50831	170.32806	0.00059
7	23588	0.05844	23410.54	0.058	1.19164	1.34521	0.00005
8	20660	0.05119	20585.13	0.051	0.53209	0.27231	0.00002
9	16299	0.04038	18566.98	0.046	17.03722	277.03661	0.00062
Soma		1		1		10193.52556	0.01534
Z crítico = 1.96 α = 0.05		Σ DAM					
$\Sigma X^2 = 15.507$ α = 0.05		Perto	Aceitável	Marginalmente aceitável		Não conformidade	
Grau de Liberdade = 8		0.000 a 0.006	0.006 a 0.012	0.012 a 0.015		Acima 0.015	

Fonte: Elaborada pelo autor.

A análise ocorreu em 403630 registros. Tanto o teste Z quanto o teste X² possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.01534, mostrando assim uma não conformidade com a Lei de Newcomb-Benford.

Figura 17 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2015



Fonte: Elaborada pelo autor.

b) Campo vlrGlosa

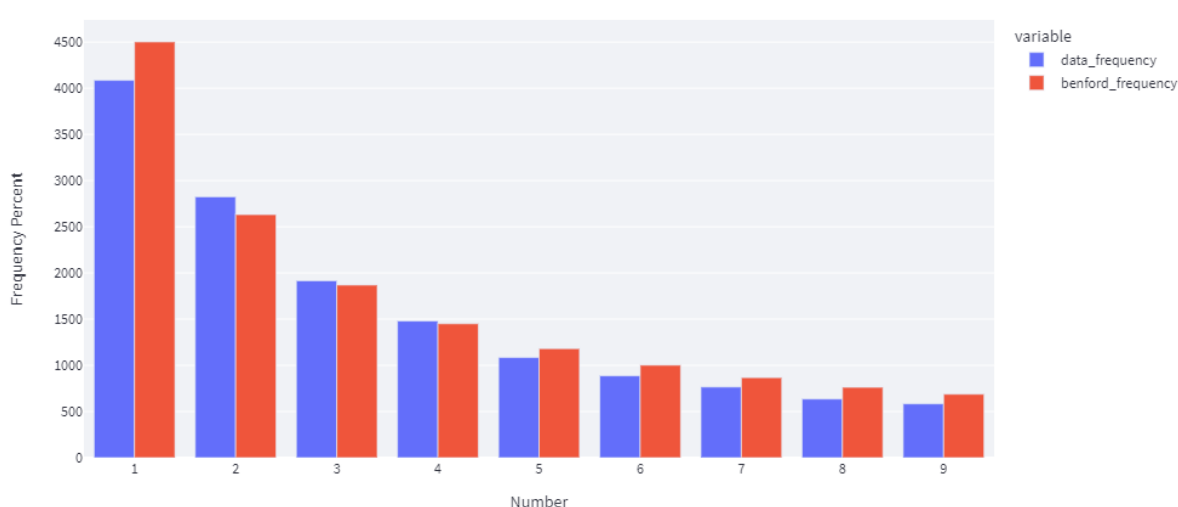
Tabela 19 – Ano 2015 do campo vlrGlosa, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	4086	0.27318	4502.057	0.301	7.40775	38.44985	0.00309
2	2825	0.18887	2632.432	0.176	4.12394	14.08676	0.00143
3	1916	0.1281	1869.625	0.125	1.13421	1.15031	0.00034
4	1480	0.09895	1450.829	0.097	0.79212	0.58652	0.00022
5	1085	0.07254	1181.603	0.079	2.91321	7.89786	0.00072
6	887	0.0593	1002.119	0.067	3.74849	13.22436	0.00086
7	766	0.05121	867.506	0.058	3.53334	11.87711	0.00075
8	636	0.04252	762.807	0.051	4.69447	21.08006	0.00094
9	584	0.03905	688.022	0.046	4.04071	15.72708	0.00077
Soma		1		1		124.07991	0.00912
Z crítico = 1.96 α = 0.05		Σ DAM					
Σ X² = 15.507 α = 0.05		Perto	Aceitável	Marginalmente aceitável		Não conformidade	
Grau de Liberdade = 8		0.000 a 0.006	0.006 a 0.012	0.012 a 0.015		Acima 0.015	

Fonte: Elaborada pelo autor.

A análise ocorreu em 14957 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.00912, sendo considerado aceitável conforme a Lei de Newcomb-Benford.

Figura 18 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2015



Fonte: Elaborada pelo autor.

5.1.9 Ano 2016

a) Campo vlrDocumento

Tabela 20 – Ano 2016 do campo vlrDocumento, Teste Z, X^2 e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X^2	DAM
1	131249	0.34249	115348.016	0.301	55.99725	2191.98649	0.00461
2	58100	0.15161	67446.016	0.176	39.64255	1295.08043	0.00271
3	39278	0.1025	47902	0.125	42.12139	1552.61526	0.0025
4	35156	0.09174	37171.952	0.097	11.00071	109.33142	0.00058
5	36897	0.09628	30274.064	0.079	39.65999	1448.87324	0.00192
6	23848	0.06223	25675.472	0.067	11.80407	130.07176	0.00053

7	20412	0.05326	22226.528	0.058	12.53669	148.13433	0.00053
8	20539	0.0536	19544.016	0.051	7.30227	50.65454	0.00029
9	15735	0.04106	17627.936	0.046	14.59306	203.26865	0.00055
Soma		1		1		7130.01612	0.01422

Z crítico = 1.96 α = 0.05

Σ DAM

$\Sigma X^2 = 15.507$ α = 0.05

Perto

Aceitável

Marginalmente aceitável

Não conformidade

Grau de Liberdade = 8

0.000 a 0.006

0.006 a 0.012

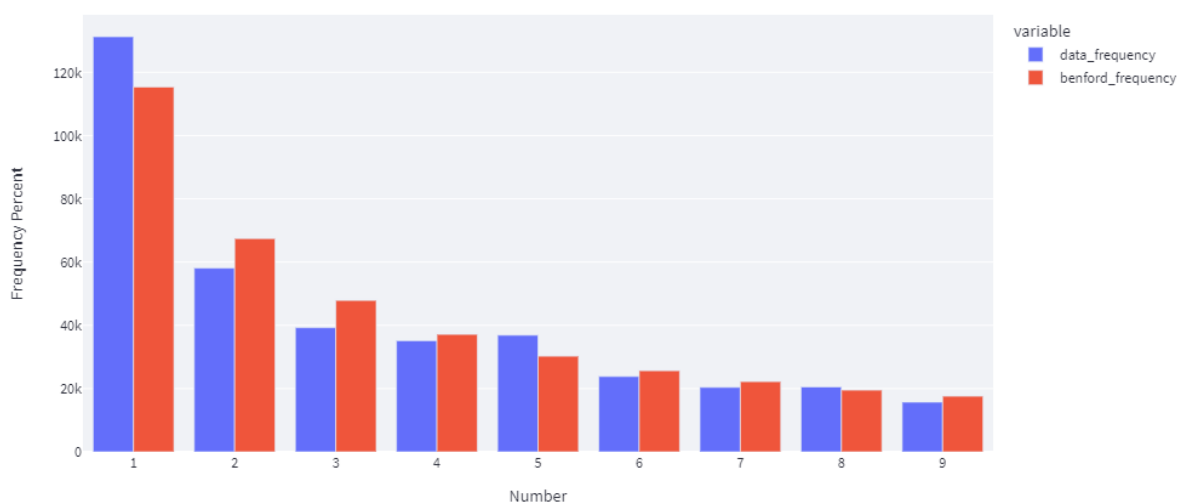
0.012 a 0.015

Acima 0.015

Fonte: Elaborada pelo autor.

A análise ocorreu em 383216 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.01422, sendo considerado marginalmente aceitável conforme a Lei de Newcomb-Benford.

Figura 19 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2016



Fonte: Elaborada pelo autor.

b) Campo vlrGlosa

Tabela 21 – Ano 2016 do campo vlrGlosa, Teste Z, X² e DAM

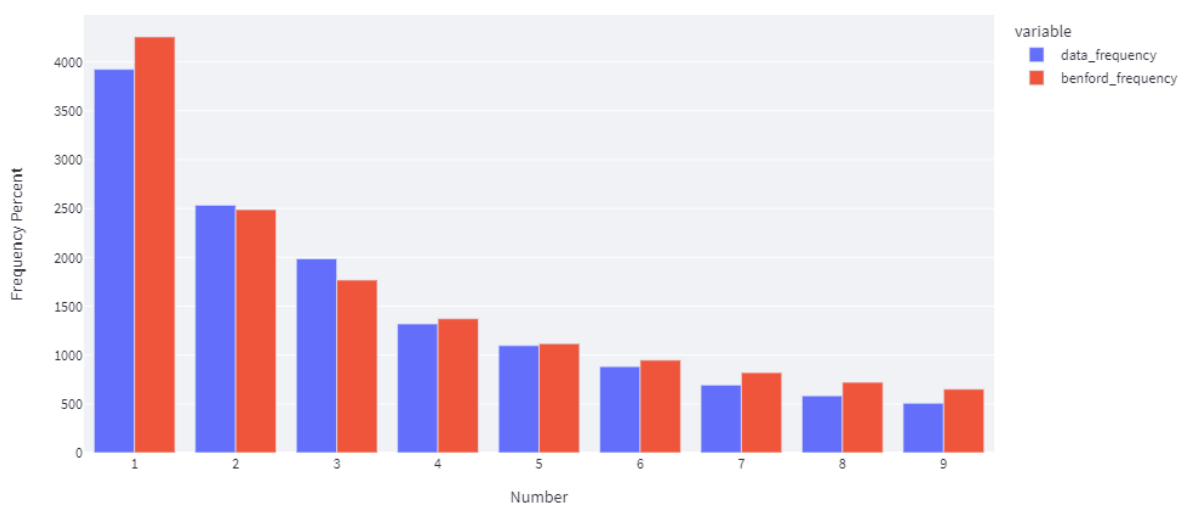
Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	3925	0.2775	4257.344	0.301	6.08311	25.944	0.00261
2	2534	0.17916	2489.344	0.176	0.97495	0.80108	0.00035
3	1986	0.14041	1768	0.125	5.52986	26.88009	0.00171
4	1319	0.09326	1371.968	0.097	1.49066	2.04495	0.00042
5	1097	0.07756	1117.376	0.079	0.61958	0.37157	0.00016
6	882	0.06236	947.648	0.067	2.19097	4.54774	0.00052
7	693	0.049	820.352	0.058	4.56322	19.77021	0.001
8	583	0.04122	721.344	0.051	5.26846	26.5325	0.00109
9	507	0.03585	650.624	0.046	5.74478	31.70472	0.00113
Soma		1		1		138.59686	0.00899

Z crítico = 1.96 α = 0.05	Σ DAM			
$\Sigma X^2 = 15.507$ α = 0.05	Perto	Aceitável	Marginalmente aceitável	Não conformidade
Grau de Liberdade = 8	0.000 a 0.006	0.006 a 0.012	0.012 a 0.015	Acima 0.015

Fonte: Elaborada pelo autor.

A análise ocorreu em 14144 registros. Tanto o teste Z quanto o teste X² possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.00899, sendo considerado aceitável conforme a Lei de Newcomb-Benford.

Figura 20 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2008



Fonte: Elaborada pelo autor.

5.1.10 Ano 2017

a) Campo vlrDocumento

Tabela 22 – Ano 2017 do campo vlrDocumento, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	125982	0.34244	110737.9	0.301	54.78994	2098.49189	0.0046
2	60157	0.16351	64750.4	0.176	19.88393	325.85627	0.00139
3	38718	0.10524	45987.5	0.125	36.23688	1149.13031	0.0022
4	32933	0.08952	35686.3	0.097	15.33487	212.42496	0.00083
5	33219	0.09029	29064.1	0.079	25.39218	593.96967	0.00125
6	23751	0.06456	24649.3	0.067	5.9202	32.73695	0.00027
7	18935	0.05147	21338.2	0.058	16.94709	270.65874	0.00073
8	17948	0.04878	18762.9	0.051	6.10316	35.39229	0.00025
9	13980	0.038	16923.4	0.046	23.16101	511.93044	0.00089
Soma		1		1		5230.59152	0.01241

Z crítico = 1.96 $\alpha = 0.05$

Σ DAM

$\Sigma X^2 = 15.507$ $\alpha = 0.05$

Perto

Aceitável

Marginalmente aceitável

Não conformidade

Grau de Liberdade = 8

0.000 a 0.006

0.006 a 0.012

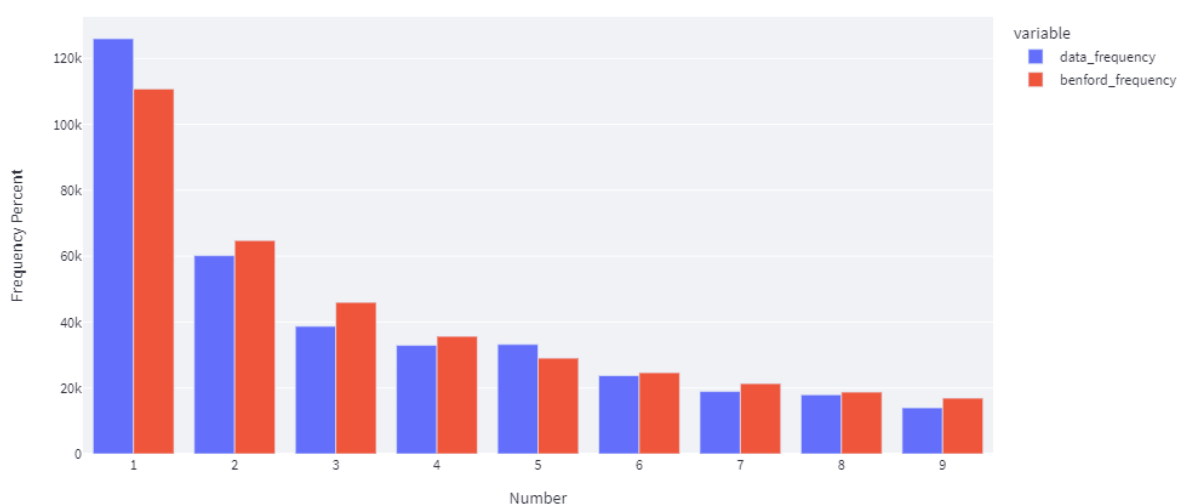
0.012 a 0.015

Acima 0.015

Fonte: Elaborada pelo autor.

A análise ocorreu em 367900 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.01241, sendo considerado marginalmente aceitável conforme a Lei de Newcomb-Benford.

Figura 21 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2017



Fonte: Elaborada pelo autor.

b) Campo vlrGlosa

Tabela 23 – Ano 2017 do campo vlrGlosa Teste Z, X^2 e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X^2	DAM
1	3689	0.28478	3899.154	0.301	4.01587	11.32674	0.0018
2	2290	0.17678	2279.904	0.176	0.2214	0.04471	0.00009
3	1774	0.13695	1619.25	0.125	4.09793	14.78929	0.00133

4	1218	0.09403	1256.538	0.097	1.12924	1.18196	0.00033
5	990	0.07642	1023.366	0.079	1.07054	1.08787	0.00029
6	746	0.05759	867.918	0.067	4.26681	17.12604	0.00105
7	663	0.05118	751.332	0.058	3.3015	10.38495	0.00076
8	539	0.04161	660.654	0.051	4.83858	22.40158	0.00104
9	546	0.04215	595.884	0.046	2.07124	4.176	0.00043
Soma		1		1		82.51914	0.00712

Z crítico = 1.96 $\alpha = 0.05$

Σ DAM

$\Sigma X^2 = 15.507$ $\alpha = 0.05$

Perto

Aceitável

Marginalmente aceitável

Não conformidade

Grau de Liberdade = 8

0.000 a 0.006

0.006 a 0.012

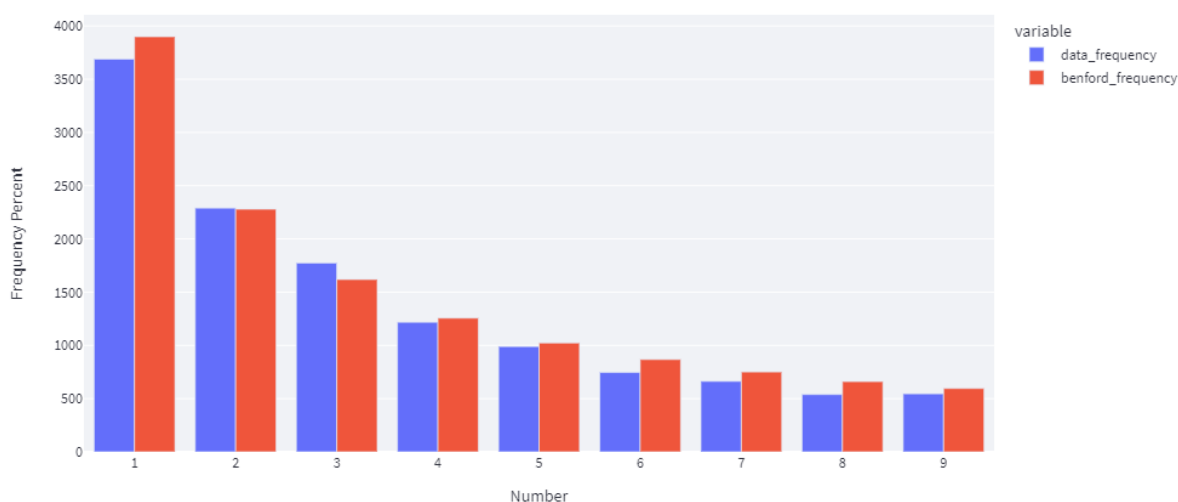
0.012 a 0.015

Acima 0.015

Fonte: Elaborada pelo autor.

A análise ocorreu em 12954 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.00712, sendo considerado aceitável conforme a Lei de Newcomb-Benford.

Figura 22 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2017



Fonte: Elaborada pelo autor.

5.1.11 Ano 2018

a) Campo vlrDocumento

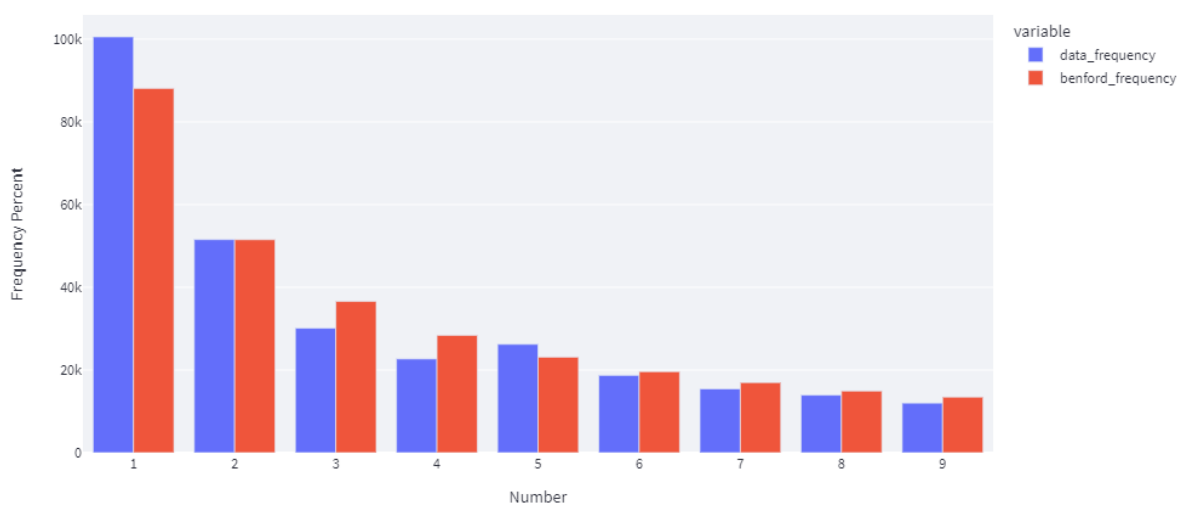
Tabela 24 – Ano 2018 do campo vlrDocumento, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	100562	0.34354	88108.419	0.301	50.17987	1760.23678	0.00473
2	51546	0.17609	51518.544	0.176	0.13083	0.01463	0.00001
3	30121	0.1029	36589.875	0.125	36.1502	1143.65911	0.00246
4	22691	0.07752	28393.743	0.097	35.61152	1145.36776	0.00216
5	26235	0.08963	23124.801	0.079	21.30834	418.3101	0.00118
6	18720	0.06395	19612.173	0.067	6.59177	40.58564	0.00034
7	15449	0.05278	16977.702	0.058	12.08414	137.647	0.00058
8	13939	0.04762	14928.669	0.051	8.31049	65.60831	0.00038
9	12011	0.04103	13465.074	0.046	12.82504	157.02336	0.00055
Soma		1		1		4868.45269	0.01239
Z crítico = 1.96 α = 0.05		Σ DAM					
$\Sigma X^2 = 15.507$ α = 0.05		Perto	Aceitável	Marginalmente aceitável		Não conformidade	
Grau de Liberdade = 8		0.000 a 0.006	0.006 a 0.012	0.012 a 0.015		Acima 0.015	

Fonte: Elaborada pelo autor.

A análise ocorreu em 292719 registros. Tanto o teste Z quanto o teste X² possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.01239, sendo considerado marginalmente aceitável conforme a Lei de Newcomb-Benford.

Figura 23 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2018



Fonte: Elaborada pelo autor.

b) Campo vlrGlosa

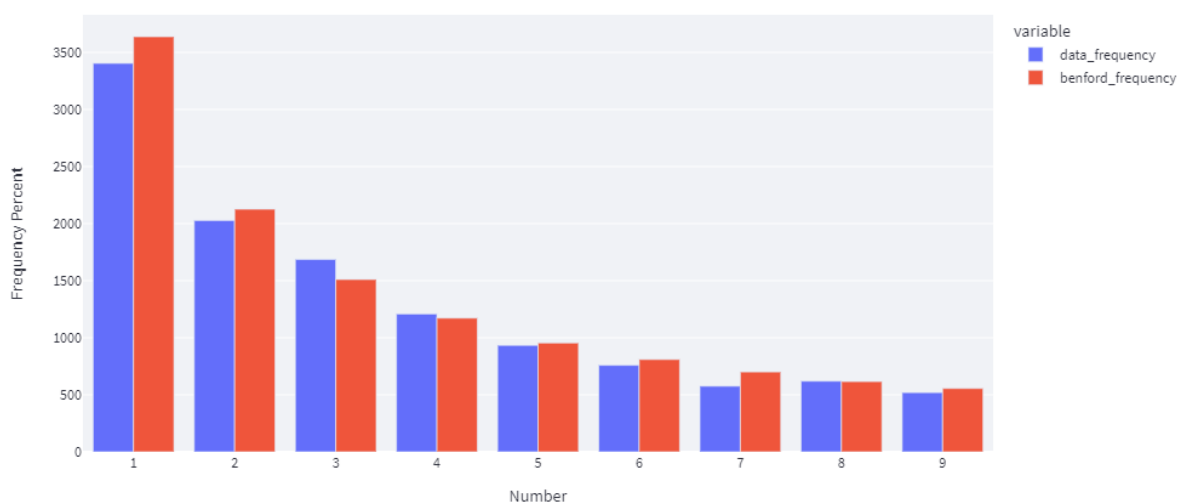
Tabela 25 – Ano 2018 do campo vlrGlosa, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	3405	0.28171	3638.187	0.301	4.61414	14.94595	0.00214
2	2028	0.16778	2127.312	0.176	2.3601	4.63631	0.00091
3	1687	0.13957	1510.875	0.125	4.83023	20.53116	0.00162
4	1209	0.10002	1172.439	0.097	1.10828	1.14011	0.00034
5	934	0.07727	954.873	0.079	0.68699	0.45627	0.00019
6	759	0.06279	809.829	0.067	1.83097	3.19029	0.00047
7	576	0.04765	701.046	0.058	4.84653	22.30453	0.00115
8	620	0.05129	616.437	0.051	0.12664	0.02059	3.00E-05
9	520	0.04302	556.002	0.046	1.54149	2.33119	0.00033
Soma		1		1		69.5564	0.00718
Z crítico = 1.96 α = 0.05							
Σ DAM							
Σ X² = 15.507 α = 0.05		Perto	Aceitável	Marginalmente aceitável		Não conformidade	
Grau de Liberdade = 8		0.000 a 0.006	0.006 a 0.012	0.012 a 0.015		Acima 0.015	

Fonte: Elaborada pelo autor.

A análise ocorreu em 12087 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.00718, sendo considerado aceitável conforme a Lei de Newcomb-Benford.

Figura 24 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2018



Fonte: Elaborada pelo autor.

5.1.12 Ano 2019

a) Campo vlrDocumento

Tabela 26 – Ano 2019 do campo vlrDocumento, Teste Z, X^2 e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X^2	DAM
1	103136	0.35639	87107.594	0.301	64.95464	2949.33871	0.00615
2	51899	0.17934	50933.344	0.176	4.71121	18.30808	0.00037
3	28310	0.09783	36174.25	0.125	44.20037	1709.681	0.00302

4	22276	0.07697	28071.218	0.097	36.39635	1196.40522	0.00223
5	23816	0.0823	22862.126	0.079	6.57015	39.79838	0.00037
6	18722	0.06469	19389.398	0.067	4.95834	22.97235	0.00026
7	15212	0.05257	16784.852	0.058	12.50448	147.38667	0.0006
8	13237	0.04574	14759.094	0.051	12.85689	156.97238	0.00058
9	12016	0.04152	13312.124	0.046	11.49691	126.19605	0.0005
Soma		1		1		6367.05884	0.01408

Z crítico = 1.96 $\alpha = 0.05$

Σ DAM

$\Sigma X^2 = 15.507$ $\alpha = 0.05$

Perto

Aceitável

Marginalmente aceitável

Não conformidade

Grau de Liberdade = 8

0.000 a 0.006

0.006 a 0.012

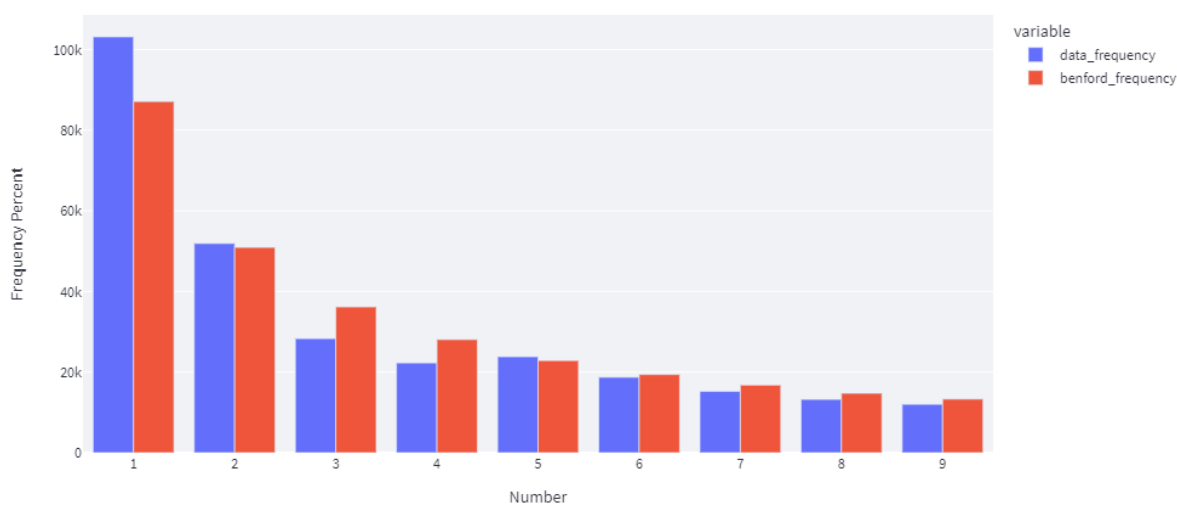
0.012 a 0.015

Acima 0.015

Fonte: Elaborada pelo autor.

A análise ocorreu em 289394 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.01408, sendo considerado marginalmente aceitável conforme a Lei de Newcomb-Benford.

Figura 25 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2019



Fonte: Elaborada pelo autor.

b) Campo vlrGlosa

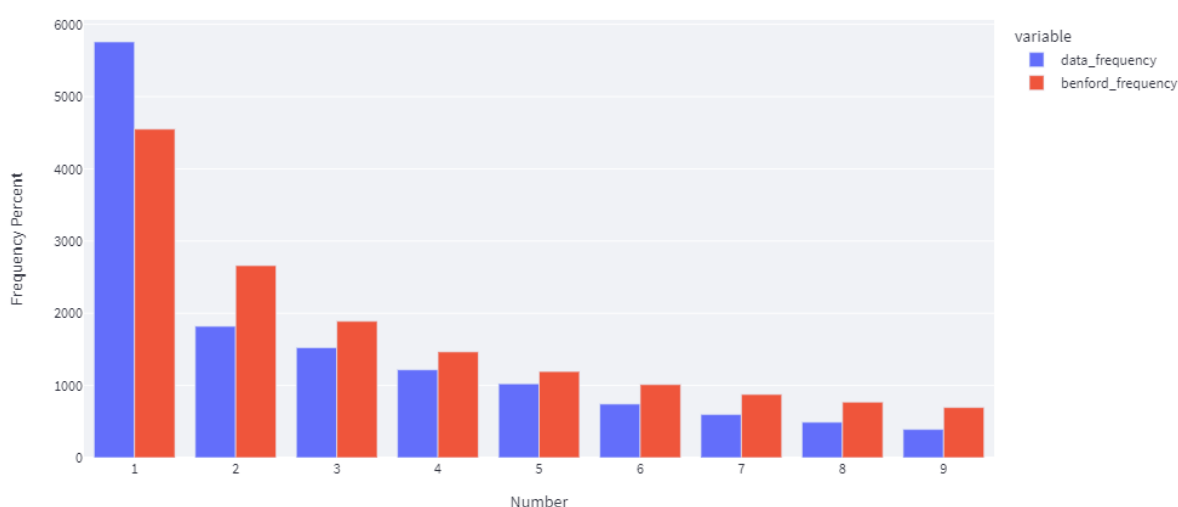
Tabela 27 – Ano 2019 do campo vlrGlosa, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	5758	0.3809	4550.217	0.301	21.40693	320.58686	0.00888
2	1820	0.12039	2660.592	0.176	17.94213	265.57808	0.00618
3	1525	0.10088	1889.625	0.125	8.95486	70.35861	0.00268
4	1218	0.08057	1466.349	0.097	6.81122	42.06176	0.00183
5	1024	0.06774	1194.243	0.079	5.11818	24.26866	0.00125
6	744	0.04922	1012.839	0.067	8.72917	71.35824	0.00198
7	600	0.03969	876.786	0.058	9.61361	87.3765	0.00203
8	492	0.03255	770.967	0.051	10.29491	100.94153	0.00205
9	392	0.02593	695.382	0.046	11.75946	132.35982	0.00223
Soma		1		1		1114.89006	0.02911
Z crítico = 1.96 α = 0.05		Σ DAM					
Σ X² = 15.507 α = 0.05		Perto	Aceitável	Marginalmente aceitável		Não conformidade	
Grau de Liberdade = 8		0.000 a 0.006	0.006 a 0.012	0.012 a 0.015		Acima 0.015	

Fonte: Elaborada pelo autor.

A análise ocorreu em 15117 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.02911, mostrando assim uma não conformidade com a Lei de Newcomb-Benford.

Figura 26 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2019



Fonte: Elaborada pelo autor.

5.1.13 Ano 2020

a) Campo vlrDocumento

Tabela 28 – Ano 2020 do campo vlrDocumento, Teste Z, X^2 e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X^2	DAM
1	57914	0.35025	49770.952	0.301	43.65498	1332.28777	0.00547
2	31957	0.19327	29101.952	0.176	18.43371	280.09458	0.00192
3	17216	0.10412	20669	0.125	25.6726	576.86434	0.00232
4	13310	0.08049	16039.144	0.097	22.67318	464.37808	0.00183
5	13717	0.08296	13062.808	0.079	5.95971	32.76226	0.00044
6	9828	0.05944	11078.584	0.067	12.29578	141.1697	0.00084

7	8157	0.04933	9590.416	0.058	15.07567	214.2432	0.00096
8	6482	0.0392	8432.952	0.051	21.8028	451.3501	0.00131
9	6023	0.03643	7606.192	0.046	18.57969	329.53374	0.00106
Soma		1		1		3822.68377	0.01615

Z crítico = 1.96 α = 0.05

Σ DAM

ΣX^2 = 15.507 α = 0.05

Perto

Aceitável

Marginalmente aceitável

Não conformidade

Grau de Liberdade = 8

0.000 a 0.006

0.006 a 0.012

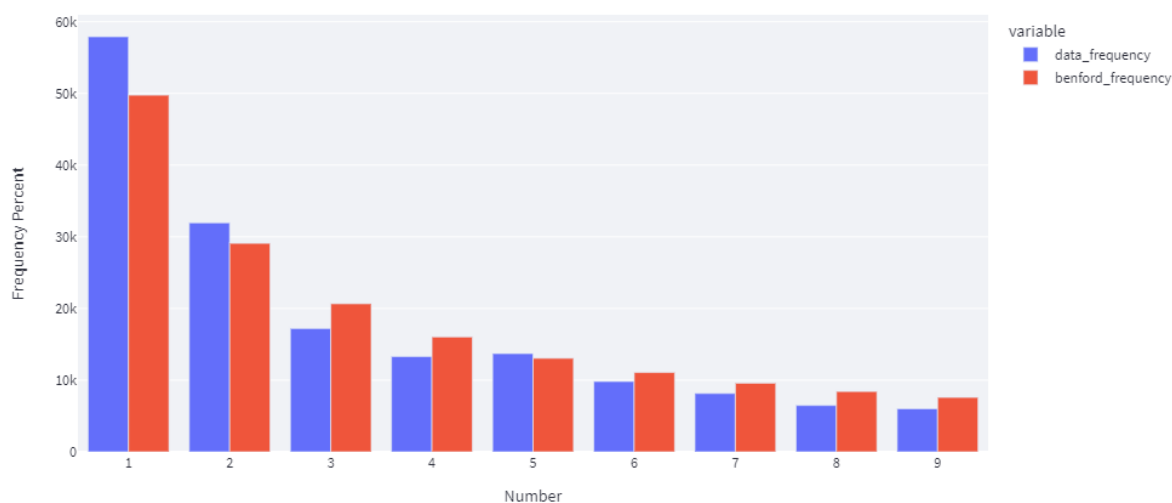
0.012 a 0.015

Acima 0.015

Fonte: Elaborada pelo autor.

A análise ocorreu em 165352 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.01615, mostrando assim uma não conformidade com a Lei de Newcomb-Benford.

Figura 27 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2020



Fonte: Elaborada pelo autor.

b) Campo vlrGlosa

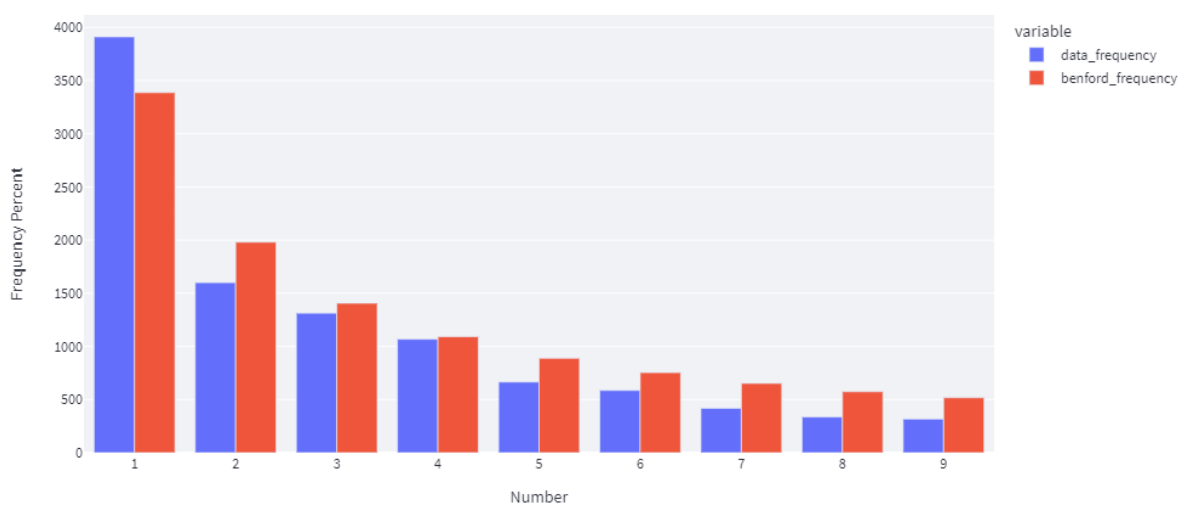
Tabela 29 – Ano 2020 do campo vlrGlosa, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	3911	0.34761	3386.551	0.301	10.76891	81.21737	0.00518
2	1600	0.14221	1980.176	0.176	9.39935	72.99038	0.00375
3	1313	0.1167	1406.375	0.125	2.64755	6.19955	0.00092
4	1068	0.09492	1091.347	0.097	0.72779	0.49946	0.00023
5	665	0.05911	888.829	0.079	7.8056	56.36565	0.00221
6	586	0.05208	753.817	0.067	6.30908	37.35992	0.00166
7	418	0.03715	652.558	0.058	9.44036	84.31044	0.00232
8	336	0.02986	573.801	0.051	10.16918	98.55214	0.00235
9	317	0.02818	517.546	0.046	9.00287	77.71038	0.00198
Soma		1		1		515.20529	0.0206
Z crítico = 1.96 α = 0.05		Σ DAM					
$\Sigma X^2 = 15.507$ α = 0.05		Perto	Aceitável	Marginalmente aceitável		Não conformidade	
Grau de Liberdade = 8		0.000 a 0.006	0.006 a 0.012	0.012 a 0.015		Acima 0.015	

Fonte: Elaborada pelo autor.

A análise ocorreu em 11251 registros. Tanto o teste Z quanto o teste X² possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (NIGRINI, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.0206, mostrando assim uma não conformidade com a Lei de Newcomb-Benford.

Figura 28 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2020



Fonte: Elaborada pelo autor.

5.1.14 Ano 2021

a) Campo vlrDocumento

Tabela 30 – Ano 2021 do campo vlrDocumento, Teste Z, X² e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X ²	DAM
1	70498	0.33652	63056.49	0.301	35.44283	878.1978	0.00395
2	44989	0.21475	36870.24	0.176	46.57591	1787.73623	0.00431
3	26608	0.12701	26186.25	0.125	2.78291	6.79261	0.00022
4	17657	0.08429	20320.53	0.097	19.65915	349.12436	0.00141
5	15068	0.07193	16549.71	0.079	11.99752	132.65879	0.00079
6	11424	0.05453	14035.83	0.067	22.81928	486.01728	0.00139
7	8394	0.04007	12150.42	0.058	35.10712	1161.33362	0.00199
8	7430	0.03547	10683.99	0.051	32.31094	991.05773	0.00173
9	6985	0.03334	9636.54	0.046	27.64913	729.58389	0.00141
Soma		1		1		6522.50231	0.0172

Z crítico = 1.96 $\alpha = 0.05$

Σ DAM

Σ X² = 15.507 $\alpha = 0.05$

Perto

Aceitável

Marginalmente aceitável

Não conformidade

Grau de Liberdade = 8

0.000 a 0.006

0.006 a 0.012

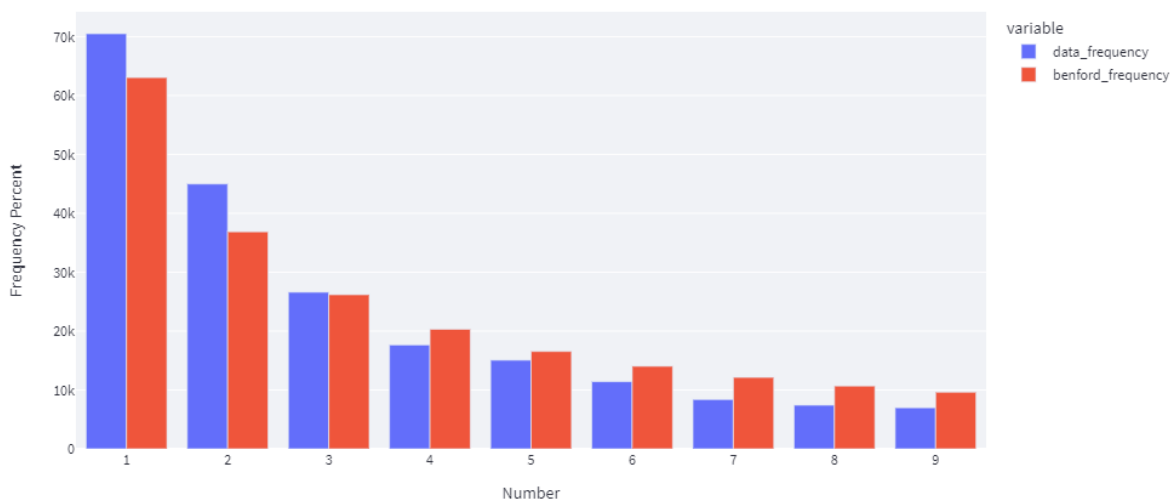
0.012 a 0.015

Acima 0.015

Fonte: Elaborada pelo autor.

A análise ocorreu em 209490 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (Nigrini, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.0172, mostrando assim uma não conformidade com a Lei de Newcomb-Benford.

Figura 29 - Gráfico de frequência de distribuição do campo vlrDocumento do ano de 2021



Fonte: Elaborada pelo autor.

b) Campo vlrGlosa

Tabela 31 – Ano 2021 do campo vlrGlosa, Teste Z, X^2 e DAM

Dígito	Freq. Observada	Prob. Observada	Freq. Esperada	Prob. Esperada	Teste Z	X^2	DAM
1	4720	0.38046	3734.206	0.301	19.28538	260.24001	0.00883
2	1927	0.15533	2183.456	0.176	6.03434	30.12183	0.0023
3	1317	0.10616	1550.75	0.125	6.33208	35.23396	0.00209
4	1133	0.09133	1203.382	0.097	2.11992	4.11642	0.00063

5	818	0.06594	980.074	0.079	5.37789	26.80204	0.00145
6	702	0.05659	831.202	0.067	4.62159	20.08315	0.00116
7	521	0.042	719.548	0.058	7.60703	54.78621	0.00178
8	447	0.03603	632.706	0.051	7.55825	54.5067	0.00166
9	412	0.03321	570.676	0.046	6.77909	44.11973	0.00142
Soma		1		1		530.01005	0.02132

Z crítico = 1.96 $\alpha = 0.05$

Σ DAM

$\Sigma X^2 = 15.507$ $\alpha = 0.05$

Perto

Aceitável

Marginalmente aceitável

Não conformidade

Grau de Liberdade = 8

0.000 a 0.006

0.006 a 0.012

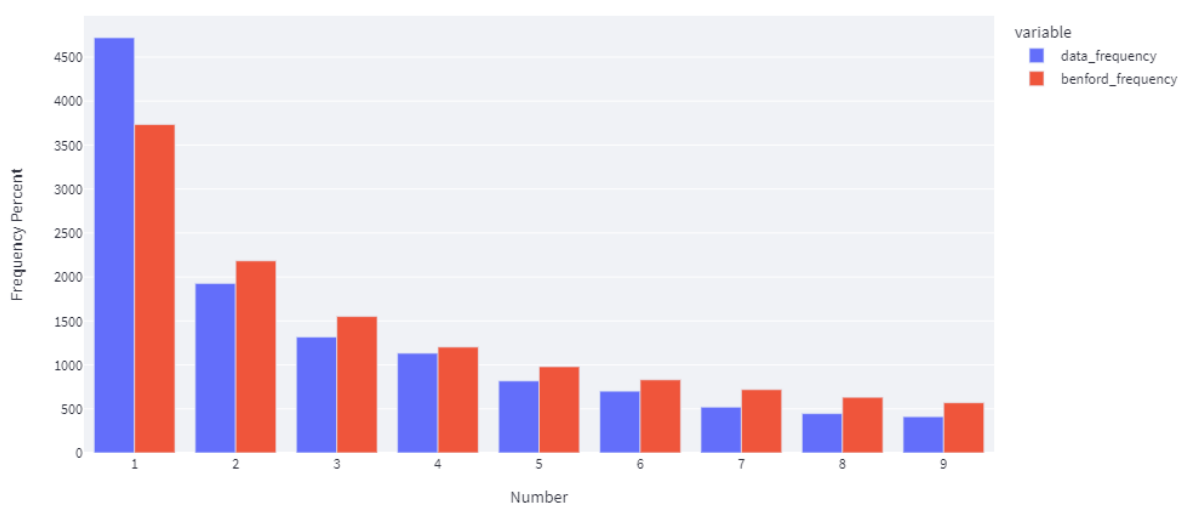
0.012 a 0.015

Acima 0.015

Fonte: Elaborada pelo autor.

A análise ocorreu em 12406 registros. Tanto o teste Z quanto o teste X^2 possuem problemas com números grandes bases de dados, levando considerarem o valor de registros em seus cálculos, sendo assim, impossibilitando seu uso para bases maiores que 5000 registros (Nigrini, 2008). Por fim, o DAM, por não utilizar o número de registros em seus cálculos, tende a ser mais assertivo em relação à Lei de Newcomb-Benford, e valor resultante foi de 0.02132, mostrando assim uma não conformidade com a Lei de Newcomb-Benford.

Figura 30 - Gráfico de frequência de distribuição do campo vlrGlosa do ano de 2021



Fonte: Elaborada pelo autor.

5.1.15 Resultado do Experimento

Esta tabela mostra a conclusão da conformidade do experimento iniciado no capítulo 5, e quais foram consideradas aceitáveis, marginalmente aceitáveis e em não conformidade.

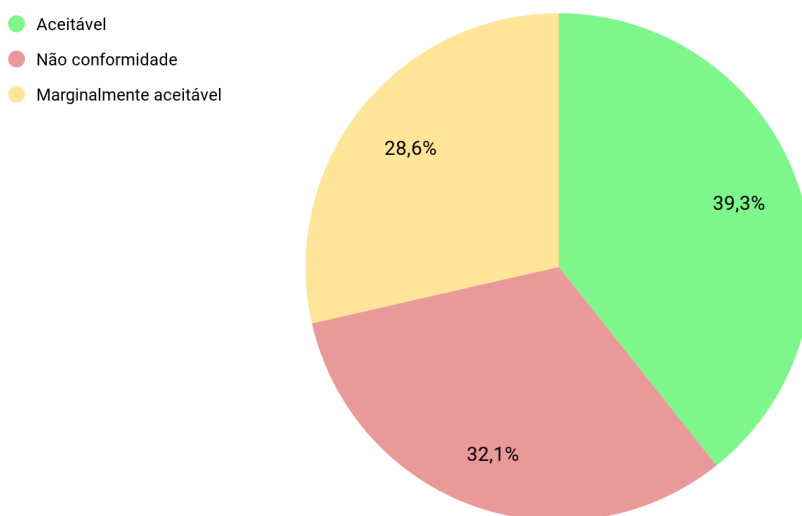
Tabela 32 – Conclusões de conformidade

Aceitável	Ano 2009 - vlrGlosa
	Ano 2011 - vlrDocument
	Ano 2011 - vlrGlosa
	Ano 2012 - vlrDocument
	Ano 2012 - vlrGlosa
	Ano 2013 - vlrGlosa
	Ano 2014 - vlrGlosa
	Ano 2015 - vlrGlosa
	Ano 2016 - vlrGlosa
	Ano 2017 - vlrGlosa
	Ano 2018 - vlrGlosa
Marginalmente aceitável	Ano 2009 - vlrDocument
	Ano 2010 - vlrDocument
	Ano 2010 - vlrGlosa
	Ano 2013 - vlrDocument
	Ano 2016 - vlrDocument
	Ano 2017 - vlrDocument
	Ano 2018 - vlrDocument
	Ano 2019 - vlrDocument
Não conformidade	Ano 2008 - vlrDocument
	Ano 2008 - vlrGlosa
	Ano 2014 - vlrDocument
	Ano 2015 - vlrDocument
	Ano 2019 - vlrGlosa
	Ano 2020 - vlrDocument
	Ano 2020 - vlrGlosa
	Ano 2021 - vlrDocument
	Ano 2021 - vlrGlosa

Fonte: Elaborada pelo autor.

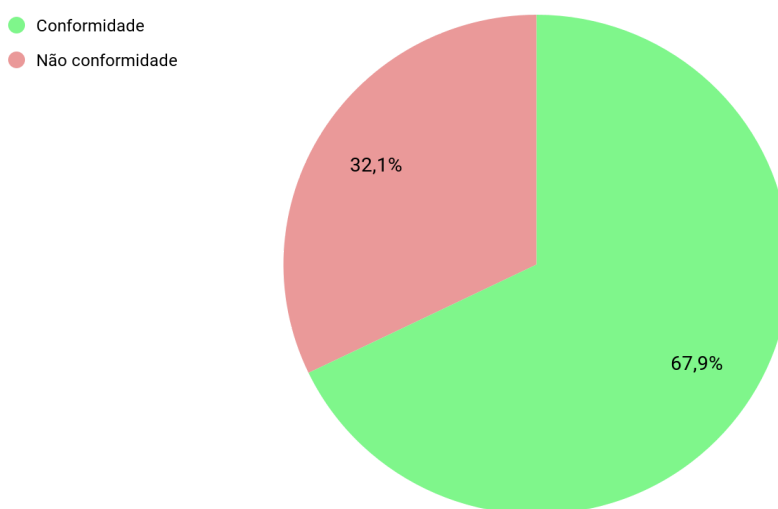
A partir da análise da Conclusões de conformidade (TABELA 32) é demonstrada a Conclusão Geral dos Resultados (FIGURA 31), onde é possível visualizar as porcentagens geral das conformidades. Além disso, também é demonstrada uma análise de Conformidade Geral dos Resultados (FIGURA 32), considerando se os conjuntos estão em conformidade ou não.

Figura 31 - Conclusão Geral dos Resultados



Fonte: Elaborada pelo autor.

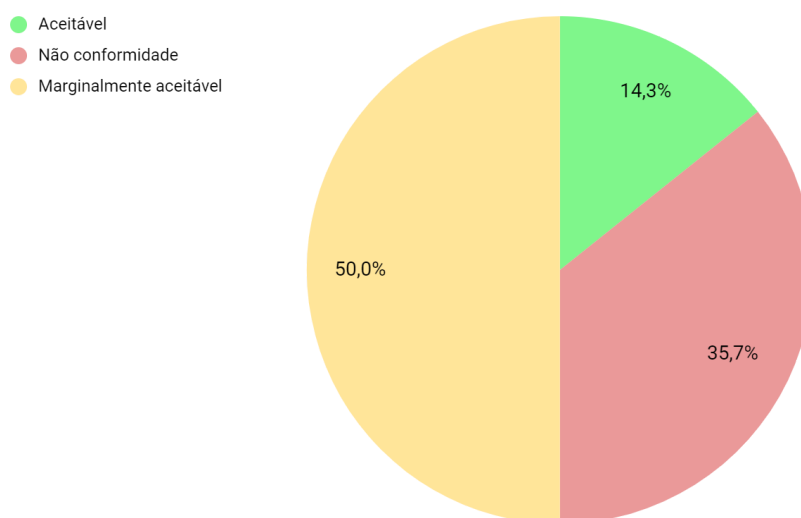
Figura 32 - Conformidade Geral dos Resultados



Fonte: Elaborada pelo autor.

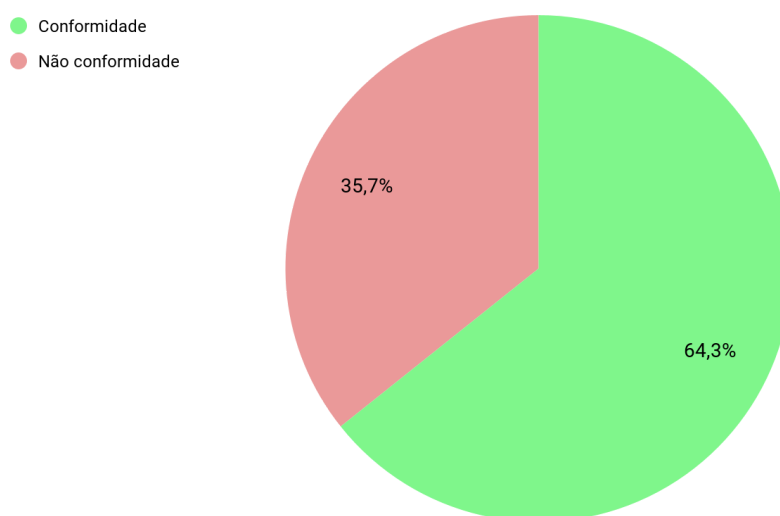
Também foi possível analisar separadamente os resultados obtidos da Conclusão dos Valores dos Documentos (FIGURA 33), onde é possível visualizar as porcentagens das conformidades dos documentos. Além disso, também é demonstrada uma análise de Conformidade dos Resultados dos Documentos (FIGURA 34), considerando se os conjuntos estão em conformidade ou não.

Figura 33 - Conclusão dos Valores dos Documentos



Fonte: Elaborada pelo autor.

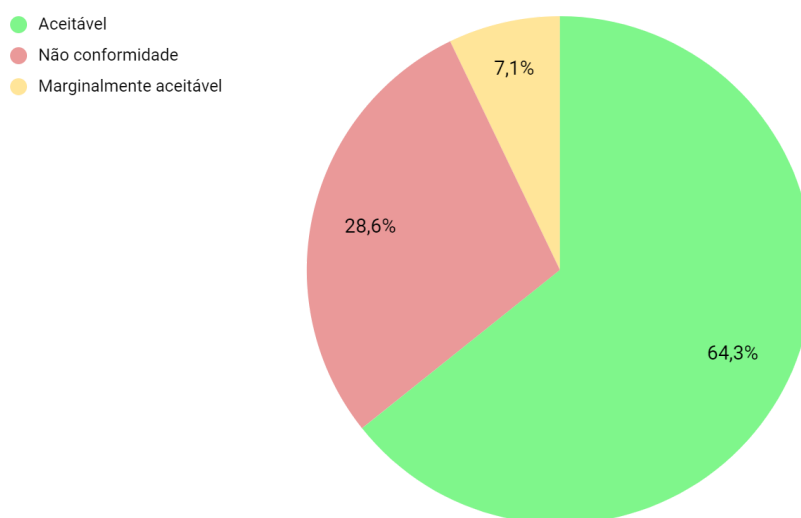
Figura 34 - Conformidade dos Resultados dos Documentos



Fonte: Elaborada pelo autor.

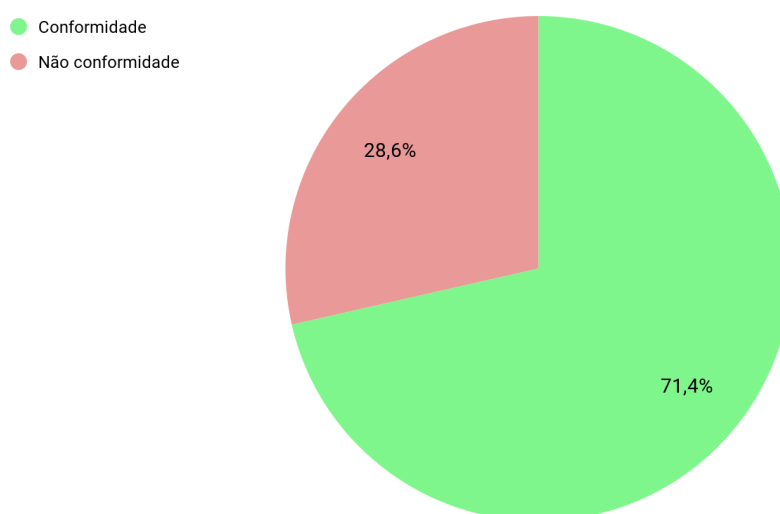
Também foi possível analisar separadamente os resultados obtidos da Conclusão dos Valores das Glosas (FIGURA 35), onde é possível visualizar as porcentagens das conformidades das glosas. Além disso, também é demonstrada uma análise de Conformidade dos Resultados das Glosas (FIGURA 36), considerando se os conjuntos estão em conformidade ou não.

Figura 35 - Conclusão dos Valores das Glosas



Fonte: Elaborada pelo autor.

Figura 36 - Conformidade dos Resultados das Glosas



Fonte: Elaborada pelo autor.

6 CONCLUSÃO

O intuito dessa pesquisa foi a criação de uma ferramenta que possibilite a utilização da Lei de Newcomb-Benford em bases de dados, e a utilização prática de uma base específica para ser analisada, sendo essa base, a de despesas pela cota para exercício da atividade parlamentar.

Durante a análise estatística, particularmente os Testes Z e X^2 , na maioria das vezes, rejeitavam a hipótese nula devido ao tamanho das amostras, apresentando uma leitura equivocada dos dados. Então, a solução foi a utilização do desvio absoluto médio para indicar a conformidade dos dados em relação à Lei de Newcomb-Benford.

Foi verificado que nem todos os conjuntos de dados disponibilizados pela Câmara dos Deputados Federais seguem a Lei de Newcomb-Benford utilizando a análise estatística do DAM. Além disso, foi observado que as glosas obtiveram um resultado de conformidade maior do que os documentos.

No que se refere a suposição de que as frequências das distribuições não estarem em conformidade com a Lei de Newcomb-Benford não significa que exista uma confirmação conclusiva de adulteração; além disso, que uma conformidade não assegura fidedignidade dos dados, e uma não conformidade indica que os necessitam de uma averiguação mais meticulosa (CECCATO, 2013).

A conclusão geral é que a Lei de Newcomb-Benford é um bom orientador para possíveis descobertas de inconsistências, onde ela pode identificar nessas planilhas alguma impropriedade (CECCATO, 2013).

Os objetivos geral e específico do trabalho foram atingidos ao criar uma ferramenta simples e acessível, o Bendata, para utilização que possua técnicas para possibilitar a descoberta de anomalias, e que também mostram a conformidade das bases com a Lei de Newcomb-Benford.

Sendo assim, sugere-se a utilização da lei de Newcomb-Berford em bases de dados abertas para garantir a qualidade dos dados disponibilizados.

Vale ressaltar que esse estudo, não é prova conclusiva de adulteração dos dados fornecidos pela Câmara dos Deputados, e seu único propósito para realização do estudo foi verificar a adequação dos dados à Lei de Newcomb-Benford. Para entender a causa da não

conformidade será necessária uma averiguação mais meticulosa de natureza diferente da aqui proposta.

Para trabalhos futuros, sugere-se a implementação de uma funcionalidade que facilite a utilização da ferramenta diretamente em bases de dados abertos, além de outros testes estáticos que não sofram com bases que possuem alto volume de registros.

REFERÊNCIAS

- ASSOCIATION OF CERTIFIED FRAUD EXAMINERS. Fraud Examiners Manual: 2020 International Edition. 2020.
- BENFORD, F. The Law of Anomalous Numbers. American Philosophical Society, [s. l.], v. 78, ed. 4, p. 551-572, 1938.
- CÂMARA DOS DEPUTADOS. Explicações sobre o formato dos arquivos XML. CÂMARA DOS DEPUTADOS, 2017. Disponível em: <https://www2.camara.leg.br/transparencia/cota-para-exercicio-da-atividade-parlamentar/explicacoes-sobre-o-formato-dos-arquivos-xml>. Acesso em: 22 fev. 2022.
- CÂMARA DOS DEPUTADOS. Despesas pela Cota para Exercício da Atividade Parlamentar. CÂMARA DOS DEPUTADOS, 2017. Disponível em: <https://dadosabertos.camara.leg.br/swagger/api.html>. Acesso em: 22 fev. 2022.
- CODEDROME. Benford's Law in Python. CODEDROME, 2019. Disponível em: <https://www.codedrome.com/benfords-law-in-python/> 27/11/2021. Acesso em: 11 fev. 2022.
- CRUSOE, J; MELIN, U. Investigating open government data barriers: A literature review and conceptualization. Electronic Government. p. 169-183. 2018.
- FONSECA, J. S.; MARTINS, G. A. Curso de estatística, Atlas, ed. 6. 1993.
- FRIAR, J. L.; GOLDMAN, T.; PÉREZ-MERCADER, J. Genome sizes and the Benford distribution. PLOS ONE. 2012.
- KESSEL, E. Benford's Law: Potential Applications for Insider Threat Detection. Carnegie Mellon University's Software Engineering Institute Blog. 2020
- MENEZES, R. O. Aplicação da Lei de Benford nos números de casos confirmados de COVID-19 em diferentes países. REMAT: Revista Eletrônica da Matemática. v.7. n.1. p.e 3005. 2021.

TERRA, A. V. et al. Avaliação dos dados reportados sobre a covid-19 no brasil: aplicação da lei de newcomb-benford ao 1º e 2º dígitos significativos do número de casos confirmados e de mortos pelo novo coronavírus. Revista SIMEP. Vol. 1 No. 2. 2021.

ANEXO A – COLUNAS DO CONJUNTO DE DADOS COTA PARA EXERCÍCIO DA ATIVIDADE PARLAMENTAR

Segue as explicações pertinentes a cada elemento de dado publicado como "Dados Abertos" acerca das despesas realizadas e publicadas na Transparência, quanto ao uso da CEAP (Cota para o Exercício da Atividade Parlamentar)(CÂMARA DOS DEPUTADOS, 2017).

Elemento de Dado	Nome do Dado	Definição do Dado
txNomeParlamentar	Nome Parlamentar	Nome adotado pelo Parlamentar ao tomar posse do seu mandato. Compõe-se de dois elementos: um prenome e o nome; dois nomes; ou dois prenomes, salvo, a juízo do Presidente da Casa legislativa, que poderá alterar essa regra para que não ocorram confusões.
ideCadastro	Identificador Único do Parlamentar	Número que identifica unicamente um deputado federal na CD.
nuCarteiraParlamentar	Número da Carteira Parlamentar	Documento usado para identificar um deputado federal na CD. Pode alterar a cada Legislatura nova.
nuLegislatura	Número da Legislatura	Legislatura: Período de quatro anos coincidente com o mandato parlamentar dos Deputados Federais. No contexto da cota CEAP, representa o ano base de início da legislatura e é utilizado para compor a Carteira Parlamentar, pois esta poderá ser alterada à medida que se muda de Legislatura.
sgUF	Sigla da UF	No contexto da cota CEAP, representa a unidade da federação pela qual o deputado foi eleito e é utilizada para definir o valor da cota a que o deputado tem.

sgPartido	Sigla do Partido	O seu conteúdo representa a sigla de um partido. Definição de partido: é uma organização formada por pessoas com interesse ou ideologia comuns, que se associam com o fim de assumir o poder para implantar um programa de governo. Tem personalidade jurídica de direito privado e goza de autonomia e liberdade no que diz respeito à criação, organização e funcionamento, observados os princípios e preceitos constitucionais.
codLegislatura	Código da Legislatura	Legislatura: Período de quatro anos coincidente com o mandato parlamentar dos Deputados Federais. No contexto da cota CEAP, o seu conteúdo representa o código identificador da Legislatura, que um número ordinal sequencial, alterado de um em um, a cada início de uma nova Legislatura (por exemplo, a Legislatura que iniciou em 2011 é a 54ª Legislatura).
numSubCota	Número da Subcota	No contexto da Cota CEAP, o conteúdo deste dado representa o código do Tipo de Despesa referente à despesa realizada pelo deputado e comprovada por meio da emissão de um documento fiscal, a qual é debitada na cota do deputado.
txtDescricao	Descrição da Subcota	O seu conteúdo é a descrição do Tipo de Despesa relativo à despesa em questão.
numEspecificacaoSubCota	Número da Especificação da Subcota	No contexto da Cota CEAP, há despesas cujo Tipo de Despesa necessita ter uma especificação mais detalhada (por exemplo, “Combustível”). O conteúdo deste dado representa o código desta especificação mais detalhada.
txtDescricaoEspecificacao	Descrição da Especificação da Subcota	Representa a descrição especificação mais detalhada de um referido Tipo de Despesa.
txtFornecedor	Fornecedor	O conteúdo deste dado representa o nome do fornecedor do produto ou serviço presente no documento fiscal
txtCNPJCPF	CNPJ/CPF	O conteúdo deste dado representa o CNPJ ou o CPF do emitente do documento fiscal, quando se tratar do uso da cota em razão do reembolso despesas comprovadas pela emissão de documentos fiscais.

txtNumero	Número do Documento	O conteúdo deste dado representa o número de face do documento fiscal emitido ou o número do documento que deu causa à despesa debitada na cota do deputado.
indTipoDocumento	Indicativo de Tipo de Documento Fiscal	Este dado representa o tipo de documento do fiscal – 0 (Zero), para Nota Fiscal; 1 (um), para Recibo; e 2, para Despesa no Exterior.
datEmissao	Data de Emissão	O conteúdo deste dado é a data de emissão do documento fiscal ou a data do documento que tenha dado causa à despesa.
vlrDocumento	Valor do Documento	O seu conteúdo é o valor de face do documento fiscal ou o valor do documento que deu causa à despesa. Quando se tratar de bilhete aéreo, esse valor poderá ser negativo, significando que o referido bilhete é um bilhete de compensação, pois compensa um outro bilhete emitido e não utilizado pelo deputado (idem para o dado vlrLiquido abaixo).
vlrGlosa	Valor da Glosa	O seu conteúdo representa o valor da glosa do documento fiscal que incidirá sobre o Valor do Documento, ou o valor da glosa do documento que deu causa à despesa.
vlrLiquido	Valor Líquido	O seu conteúdo representa o valor líquido do documento fiscal ou do documento que deu causa à despesa e será calculado pela diferença entre o Valor do Documento e o Valor da Glosa. É este valor que será debitado da cota do deputado. Caso o débito seja do Tipo Telefonia e o valor seja igual a zero, significa que a despesa foi franqueada.
numMes	Mês	O seu conteúdo representa o Mês da competência financeira do documento fiscal ou do documento que deu causa à despesa. É utilizado, junto com o ano, para determinar em que período o débito gerará efeito financeiro sobre a cota.
numAno	Ano	O seu conteúdo representa o Ano da competência financeira do documento fiscal ou do documento que deu causa à despesa. É utilizado, junto com o mês, para determinar em que período o débito gerará efeito financeiro sobre a cota.

numParcela	Número da Parcela	O seu conteúdo representa o número da parcela do documento fiscal. Ocorre quando o documento tem de ser reembolsado de forma parcelada.
txtPassageiro	Passageiro	O conteúdo deste dado representa o nome do passageiro, quando o documento que deu causa à despesa se tratar de emissão de bilhete aéreo.
txtTrecho	Trecho	O conteúdo deste dado representa o trecho da viagem, quando o documento que deu causa à despesa se tratar de emissão de bilhete aéreo.
numLote	Número do Lote	No contexto da Cota CEAP, o Número do Lote representa uma capa de lote que agrupa os documentos que serão entregues à Câmara para serem ressarcidos. Este dado, juntamente com o Número do Ressarcimento, auxilia a localização do documento no Arquivo da Casa.
numRessarcimento	Número do Ressarcimento	No contexto da Cota CEAP, o Número do Ressarcimento indica o ressarcimento do qual o documento fez parte por ocasião do processamento do seu reembolso. Este dado, juntamente com o Número do Ressarcimento, auxilia a localização do documento no Arquivo da Casa.
vlrRestituicao	Valor da Restituição	O seu conteúdo representa o valor restituído do documento fiscal que incidirá sobre o Valor do Documento.
nuDeputadoId	Identificador do Solicitante	Número que identifica um Parlamentar ou Liderança na Transparência da Cota para Exercício da Atividade Parlamentar.

ANEXO B – TIPOS DE DESPESAS

PASSAGEM AÉREA - RPA
MANUTENÇÃO DE ESCRITÓRIO DE APOIO À ATIVIDADE PARLAMENTAR
LOCOMOÇÃO, ALIMENTAÇÃO E HOSPEDAGEM
COMBUSTÍVEIS E LUBRIFICANTES.
CONSULTORIAS, PESQUISAS E TRABALHOS TÉCNICOS.
DIVULGAÇÃO DA ATIVIDADE PARLAMENTAR.
AQUISIÇÃO DE MATERIAL DE ESCRITÓRIO.
PASSAGEM AÉREA - REEMBOLSO
TELEFONIA
SERVIÇOS POSTAIS
FORNECIMENTO DE ALIMENTAÇÃO DO PARLAMENTAR
HOSPEDAGEM ,EXCETO DO PARLAMENTAR NO DISTRITO FEDERAL.
LOCAÇÃO DE VEÍCULOS AUTOMOTORES OU FRETAMENTO DE EMBARCAÇÕES
AQUISIÇÃO OU LOC. DE SOFTWARE; SERV. POSTAIS; ASS.
SERVIÇO DE SEGURANÇA PRESTADO POR EMPRESA ESPECIALIZADA.
ASSINATURA DE PUBLICAÇÕES
SERVIÇO DE TÁXI, PEDÁGIO E ESTACIONAMENTO
LOCAÇÃO OU FRETAMENTO DE VEÍCULOS AUTOMOTORES
LOCAÇÃO OU FRETAMENTO DE AERONAVES
LOCAÇÃO OU FRETAMENTO DE EMBARCAÇÕES
PASSAGENS TERRESTRES, MARÍTIMAS OU FLUVIAIS
PARTICIPAÇÃO EM CURSO, PALESTRA OU EVENTO SIMILAR
PASSAGEM AÉREA - SIGEPA