



Universidade do Estado do Rio de Janeiro

Centro de Ciências Sociais

Faculdade de Administração e Finanças

Leonardo Botelho Ferreira da Silva

A lei Newcomb-Benford e as notas de empenhos da Marinha do Brasil

Rio de Janeiro

2021

Leonardo Botelho Ferreira da Silva

A lei Newcomb-Benford e as notas de empenhos da Marinha do Brasil



Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Ciências Contábeis, da Universidade do Estado do Rio de Janeiro. Área de concentração: Controle de Gestão.

Orientador: Prof. Dr. Francisco José dos Santos Alves

Rio de Janeiro

2021

CATALOGAÇÃO NA FONTE
UERJ/REDE SIRIUS/BIBLIOTECA CCS/B

S586 Silva, Leonardo Botelho Ferreira da.
A lei Newcomb-Benford e as notas de empenhos da Marinha do Brasil / Leonardo Botelho Ferreira da Silva. – 2021.
177 f.

Orientador: Francisco José dos Santos Alves.
Dissertação (mestrado) – Universidade do Estado do Rio de Janeiro, Faculdade de Administração e Finanças.
Bibliografia: f. 100-111.

1. Auditoria – Brasil – Teses. 2. Brasil. Marinha – Teses. 3. Fraude – Prevenção – Teses. 4. Contabilidade – Brasil – Teses. I. Alves, Francisco José dos Santos. II. Universidade do Estado do Rio de Janeiro. Faculdade de Administração e Finanças. III. Título.

CDU 657.6(81)

Bibliotecária: Regina Souza do Patrocínio CRB7/4954

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte.

Assinatura

Data

Leonardo Botelho Ferreira da Silva

A lei Newcomb-Benford e as notas de empenhos da Marinha do Brasil

Dissertação apresentada, como requisito parcial para obtenção do título de Mestre, ao Programa de Pós-Graduação em Ciências Contábeis, da Universidade do Estado do Rio de Janeiro. Área de concentração: Controle de Gestão.

Aprovada em 03 de março de 2021.

Banca Examinadora:

Prof. Dr. Francisco José dos Santos Alves (Orientador)
Faculdade de Administração e Finanças - UERJ

Prof. Dr. José Francisco Moreira Pessanha
Faculdade de Administração e Finanças - UERJ

Prof. Dr. Frederico Antonio Azevedo de Carvalho
Universidade Federal do Rio de Janeiro

Rio de Janeiro
2021

DEDICATÓRIA

Dedico este estudo aos meus queridos pais, Vera e Maurillo, por toda a educação que me proporcionaram e pela formação do meu caráter. Ao meu irmão Bernardo.
A minha esposa Caroline, por todo apoio, compreensão e parceria.

AGRADECIMENTOS

O meu primeiro agradecimento é aos meus pais Vera Lúcia e Maurillo Euclides. Sem a educação proporcionada por vocês, sendo pagando colégios e cursos ou fazendo o papel de professor, me ensinando o conteúdo, não conseguiria chegar aonde cheguei. Vocês construíram a base do conhecimento que possuo hoje para conseguir galgar sucesso acadêmico, profissional e na vida. Agradeço também por moldarem meu caráter me tornando a pessoa que sou hoje. Vocês são meus exemplos.

Agradeço ao meu irmão Bernardo Botelho, por sempre estar presente nas minhas conquistas.

A minha esposa Caroline Lima, por ser meu oráculo acadêmico, disponível para tirar minhas dúvidas e me ouvir a qualquer hora do dia. Obrigado por estar ao meu lado em qualquer situação. Muito obrigado.

Ao meu orientador, prof. Dr. Francisco José dos Santos Alves, agradeço a sua disponibilidade e orientação diuturna, mesmo em época de pandemia, se fez muito presente da forma EAD. Agradeço pelo apoio e por meu fazer buscar novos conhecimentos para a confecção desta dissertação. Foi um prazer ter recebido sua orientação.

Ao prof. Dr. José Francisco Moreira Pessanha, por todos os ensinamentos da metodologia quantitativa e por ter me apresentado o *software* R, utilizado nesta dissertação. Como o senhor, me tornei um entusiasta deste *software*. Foi um prazer ter conhecido o senhor e ser seu aluno nas disciplinas MQA1 e MQA2, onde só os fortes sobrevivem.

Agradeço a professora Dra. Andréa Paula Osório Duque e professora Dra. Tânia Maria de Oliveira Almeida Gouveia, pelos ensinamentos na área de metodologia da pesquisa, pela atenção e paciência de responderem todas minhas dúvidas durante o mestrado.

Ao Prof. Dr. Frederico Antonio Azevedo de Carvalho, pelas diversas contribuições e apontamentos que ajudaram muito esta pesquisa. Aos demais docentes do Programa, por todo o aprendizado adquirido nas atividades acadêmicas realizadas.

Agradeço também a minha companheira de turma Monica Sobreira por ser minha dupla em todas as disciplinas do mestrado e ter tido paciência para fazer todos os trabalhos e dois artigos comigo. Obrigado.

A toda equipe da secretaria do Programa de Pós-Graduação em Ciências Contábeis (PPGCC), pela atenção e apoio de sempre.

Agradeço também a Marinha do Brasil, pela oportunidade de crescer profissionalmente com este mestrado. Ao Almirante de Esquadra Celso Luiz Nazareth, ao Almirante de Esquadra Renato Rodrigues de Aguiar Freire e ao Vice-Almirante (IM) Hugo Cavalcante Nogueira por terem me dado a oportunidade impar de participar do processo seletivo do mestrado.

Assim sendo, muito obrigado a todos, sem vocês a conclusão desta etapa não seria possível.

Numbers don't lie, sir. Politics, poetry, promises - those are lies!
Numbers are as close as we get to the handwriting of God.

Dr. Hermann Gottlieb

RESUMO

SILVA, Leonardo Botelho Ferreira da. *A lei Newcomb-Benford e as notas de empenhos da Marinha do Brasil*, 2021. 177 f. Dissertação (Mestrado em Ciências Contábeis) - Faculdade de Administração e Finanças, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2021.

O Centro de Controle Interno da Marinha (CCIMAR) é órgão responsável pela auditoria no âmbito da Marinha do Brasil (MB), atuando no combate à fraude. Com o crescimento do volume de dados gerados, é indispensável ao auditor a utilização de novas ferramentas para apoiar seu trabalho. Assim, a lei Newcomb-Benford (LNB) vem ganhando espaço como ferramenta capaz de auxiliar o auditor na sua tomada de decisão. Esta lei é considerada uma ferramenta útil na fase de planejamento da auditoria para seleção da amostra auditada. Ela é um fenômeno baseado na quantidade de vezes que um determinado algarismo deve aparecer em uma posição específica de um determinado número. Deste modo este estudo tem como objetivo geral avaliar se as notas de empenhos emitidas em 2019 por organizações militares da Marinha do Brasil estão em conformidade com a lei Newcomb-Benford. A metodologia de pesquisa é caracterizada pelo método hipotético dedutivo, com abordagem quantitativa, concepção descritiva e utilizou com critério de seleção da amostra a amostragem não probabilística por julgamento. Como instrumento de pesquisa foram analisadas 46.194 notas de empenhos emitidas no ano de 2019 por 27 organizações militares, que movimentaram um total de R\$ 6.721.861.945,92. Para isto, foram utilizadas as análises pré-Benford, análises primárias, análises avançadas, análises associadas e análises de conformidade sugeridas por Nigrini (2012) e Nigrini (2020) com o auxílio do *software* R e testes estatísticos. O estudo revelou que seis organizações militares não apresentaram conformidade com a lei Newcomb-Benford em alguma das análises realizadas. Concluiu-se assim que de forma geral as notas de empenhos emitidas pelas organizações militares estudadas seguem a lei Newcomb-Benford e esta pode ser utilizada como ferramenta na fase de planejamento da auditoria.

Palavras-chave: Lei Newcomb-Benford. Análise de dados. Auditoria. Notas de empenhos.
Marinha do Brasil

ABSTRACT

SILVA, Leonardo Botelho Ferreira da. *The Newcomb-Benford's law and the Brazilian Navy purchases orders*, 2021.177 f. Dissertação (Mestrado em Ciências Contábeis) - Faculdade de Administração e Finanças, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2021.

The Navy Internal Control Center (CCIMAR) is responsible for auditing within the Brazilian Navy (MB), working to combat fraud. With the growth in the volume of data generated, it is essential for the auditor to use new tools to support his work. Thus, the Newcomb-Benford's law (LNB) has been gaining ground as a tool capable of assisting the auditor in his decision making. This law is considered a useful tool in the audit planning phase for selecting the audited sample. It is a phenomenon based on the number of times that a given number must appear in a specific position of a given number. Thus, this study has the general objective of assessing whether the purchases orders issued in 2019 by military organizations of the Brazilian Navy are in compliance with the Newcomb-Benford law. The research methodology is characterized by the hypothetical deductive method, with a quantitative approach, descriptive design and using non-probabilistic sampling by judgment as a criterion for sample selection. As a research instrument, 46.194 purchases orders issued in 2019 by 27 military organizations that handled a total of R\$ 6.721.861.945,92 were analyzed. For this, pre-Benford analyzes, primary analyzes, advanced analyzes, associated analyzes and conformity analyzes suggested by Nigrini (2012) and Nigrini (2020) were used with the aid of the R software and statistical tests. The study found that six military organizations failed to comply with the Newcomb-Benford's law in some form of analysis. It was concluded that, in general, the purchases orders issued by the studied military organizations follow the Newcomb-Benford's law and this can be used as a tool in the planning stage of the audit.

Keywords: Newcomb-Benford's law. Data analysis. Audit. Purchases orders. Brazilian Navy.

LISTA DE FIGURAS

Figura 1 -	Ciclo da fraude.....	26
Figura 2 -	Probabilidades esperadas pela LNB para o primeiro e segundo dígito....	54
Figura 3 -	Probabilidades esperadas pela LNB para os dois primeiros dígitos.....	55
Figura 4 -	Consulta Tesouro Gerencial.....	78
Figura 5 -	Importação de dados xlsx.....	112
Figura 6 -	Substituição dos <i>not available</i>	112
Figura 7 -	Aplicação do perfil de dados.....	112
Figura 8 -	Criação de histogramas.....	112
Figura 9 -	Programação do gráfico temporal.....	113
Figura 10 -	Inicialização dos pacotes.....	113
Figura 11 -	Análises globais para os primeiros dígitos.....	113
Figura 12 -	Análise local para os primeiros dígitos.....	113
Figura 13 -	Análises globais para os segundos dígitos.....	114
Figura 14 -	Análise local para os segundos dígitos.....	114
Figura 15 -	Análises globais para os dois primeiros dígitos.....	114
Figura 16 -	Análise local para os dois primeiros dígitos.....	114
Figura 17 -	MAD para a análise de segunda ordem dos primeiros dígitos.....	114
Figura 18 -	MAD para a análise de segunda ordem dos dois primeiros dígitos.....	114
Figura 19 -	Análise da soma para os primeiros dígitos.....	115
Figura 20 -	Análise da soma para os dois primeiros dígitos.....	115
Figura 21 -	Análise dos números duplicados.....	115
Figura 22 -	Modelo fator de distorção.....	115
Figura 23 -	Histogramas dos CeIM1 e CeIM2.....	117
Figura 24 -	Histogramas dos CeIM3 e CeIM4.....	117
Figura 25 -	Histogramas dos CeIM5 , CeIM6 e CeIM7.....	118
Figura 26 -	Histogramas das EAM1 e EAM2.....	118
Figura 27 -	Histogramas das EAM3 e EAM4.....	119
Figura 28 -	Histogramas das BE1 e BE2.....	119
Figura 29 -	Histogramas das BE3 e BE4.....	120
Figura 30 -	Histogramas das BE5 e BE6.....	120

Figura 31 -	Histogramas das BE7 e BE8.....	121
Figura 32 -	Histogramas dos HN1 e HN2.....	121
Figura 33 -	Histogramas dos HN3 e HN4.....	122
Figura 34 -	Histogramas dos HN5 e HN6.....	122
Figura 35 -	Histogramas dos HN7 e HN8.....	123
Figura 36 -	Gráfico valor empenho mensal dos CeIM1 e CeIM2.....	124
Figura 37 -	Gráfico valor empenho mensal dos CeIM3 e CeIM4.....	124
Figura 38 -	Gráfico valor empenho mensal dos CeIM5, CeIM7 e CeIM8.....	125
Figura 39 -	Gráfico valor empenho mensal das EAM1 e EAM2.....	125
Figura 40 -	Gráfico valor empenho mensal das EAM3 e EAM4.....	126
Figura 41 -	Gráfico valor empenho mensal das BE1 e BE2.....	126
Figura 42 -	Gráfico valor empenho mensal das BE3 e BE4.....	127
Figura 43 -	Gráfico valor empenho mensal das BE5 e BE6.....	127
Figura 44 -	Gráfico valor empenho mensal das BE7 e BE8.....	128
Figura 45 -	Gráfico valor empenho mensal dos HN1 e HN2.....	128
Figura 46 -	Gráfico valor empenho mensal dos HN3 e HN4.....	129
Figura 47 -	Gráfico valor empenho mensal dos HN5 e HN6.....	129
Figura 48 -	Gráfico valor empenho mensal dos HN7 e HN8.....	130
Figura 49 -	Gráfico temporal dos CeIM.....	130
Figura 50 -	Gráfico temporal das EAM.....	131
Figura 51 -	Gráfico temporal das BE.....	131
Figura 52 -	Gráfico temporal dos HN.....	132
Figura 53 -	Distribuição dos dois primeiros dígitos dos CeIM1 e CeIM2.....	137
Figura 54 -	Distribuição dos dois primeiros dígitos dos CeIM3 e CeIM4.....	138
Figura 55 -	Distribuição dos dois primeiros dígitos dos CeIM5, CeIM6 e CeIM7....	138
Figura 56 -	Distribuição dos dois primeiros dígitos das EAM1 e EAM2.....	138
Figura 57 -	Distribuição dos dois primeiros dígitos das EAM3 e EAM4.....	139
Figura 58 -	Distribuição dos dois primeiros dígitos das BE1 e BE2.....	139
Figura 59 -	Distribuição dos dois primeiros dígitos das BE3 e BE4.....	139
Figura 60 -	Distribuição dos dois primeiros dígitos das BE5 e BE6.....	140
Figura 61 -	Distribuição dos dois primeiros dígitos das BE7 e BE8.....	140
Figura 62 -	Distribuição dos dois primeiros dígitos dos HN1 e HN2.....	140
Figura 63 -	Distribuição dos dois primeiros dígitos dos HN3 e HN4.....	141

Figura 64 -	Distribuição dos dois primeiros dígitos dos HN5 e HN6.....	141
Figura 65 -	Distribuição dos dois primeiros dígitos dos HN7 e HN8.....	141
Figura 66 -	Análise da soma para os primeiros dígitos do CeIM1.....	144
Figura 67 -	Análise da soma para os primeiros dígitos do CeIM2.....	144
Figura 68 -	Análise da soma para os primeiros dígitos do CeIM3.....	145
Figura 69 -	Análise da soma para os primeiros dígitos do CeIM4.....	145
Figura 70 -	Análise da soma para os primeiros dígitos do CeIM5.....	146
Figura 71 -	Análise da soma para os primeiros dígitos do CeIM6.....	146
Figura 72 -	Análise da soma para os primeiros dígitos do CeIM7.....	147
Figura 73 -	Análise da soma para os primeiros dígitos da EAM1.....	147
Figura 74 -	Análise da soma para os primeiros dígitos da EAM2.....	148
Figura 75 -	Análise da soma para os primeiros dígitos da EAM3.....	148
Figura 76 -	Análise da soma para os primeiros dígitos da EAM4.....	149
Figura 77 -	Análise da soma para os primeiros dígitos da BE1.....	149
Figura 78 -	Análise da soma para os primeiros dígitos da BE2.....	150
Figura 79 -	Análise da soma para os primeiros dígitos da BE3.....	150
Figura 80 -	Análise da soma para os primeiros dígitos da BE4.....	151
Figura 81 -	Análise da soma para os primeiros dígitos da BE5.....	151
Figura 82 -	Análise da soma para os primeiros dígitos da BE6.....	152
Figura 83 -	Análise da soma para os primeiros dígitos da BE7.....	152
Figura 84 -	Análise da soma para os primeiros dígitos da BE8.....	153
Figura 85 -	Análise da soma para os primeiros dígitos do HN1.....	153
Figura 86 -	Análise da soma para os primeiros dígitos do HN2.....	154
Figura 87 -	Análise da soma para os primeiros dígitos do HN3.....	154
Figura 88 -	Análise da soma para os primeiros dígitos do HN4.....	155
Figura 89 -	Análise da soma para os primeiros dígitos do HN5.....	155
Figura 90 -	Análise da soma para os primeiros dígitos do HN6.....	156
Figura 91 -	Análise da soma para os primeiros dígitos do HN7.....	156
Figura 92 -	Análise da soma para os primeiros dígitos do HN8.....	157
Figura 93 -	Análise da soma para os dois primeiros dígitos do CeIM1.....	157
Figura 94 -	Análise da soma para os dois primeiros dígitos do CeIM2.....	158
Figura 95 -	Análise da soma para os dois primeiros dígitos do CeIM3.....	158
Figura 96 -	Análise da soma para os dois primeiros dígitos do CeIM4.....	159

Figura 97 -	Análise da soma para os dois primeiros dígitos do CeIM5.....	159
Figura 98 -	Análise da soma para os dois primeiros dígitos do CeIM6.....	160
Figura 99 -	Análise da soma para os dois primeiros dígitos do CeIM7.....	160
Figura 100 -	Análise da soma para os dois primeiros dígitos da EAM1.....	161
Figura 101 -	Análise da soma para os dois primeiros dígitos da EAM2.....	161
Figura 102 -	Análise da soma para os dois primeiros dígitos da EAM3.....	162
Figura 103 -	Análise da soma para os dois primeiros dígitos da EAM4.....	162
Figura 104 -	Análise da soma para os dois primeiros dígitos da BE1.....	163
Figura 105 -	Análise da soma para os dois primeiros dígitos da BE2.....	163
Figura 106 -	Análise da soma para os dois primeiros dígitos da BE3.....	164
Figura 107 -	Análise da soma para os dois primeiros dígitos da BE4.....	164
Figura 108 -	Análise da soma para os dois primeiros dígitos da BE5.....	165
Figura 109 -	Análise da soma para os dois primeiros dígitos da BE6.....	165
Figura 110 -	Análise da soma para os dois primeiros dígitos da BE7.....	166
Figura 111 -	Análise da soma para os dois primeiros dígitos da BE8.....	166
Figura 112 -	Análise da soma para os dois primeiros dígitos do HN1.....	167
Figura 113 -	Análise da soma para os dois primeiros dígitos do HN2.....	167
Figura 114 -	Análise da soma para os dois primeiros dígitos do HN3.....	168
Figura 115 -	Análise da soma para os dois primeiros dígitos do HN4.....	168
Figura 116 -	Análise da soma para os dois primeiros dígitos do HN5.....	169
Figura 117 -	Análise da soma para os dois primeiros dígitos do HN6.....	169
Figura 118 -	Análise da soma para os dois primeiros dígitos do HN7.....	170
Figura 119 -	Análise da soma para os dois primeiros dígitos do HN8.....	170

LISTA DE TABELAS

Tabela 1 -	Publicações sobre a LNB.....	23
Tabela 2 -	Dados analisados por Benford (1938).....	32
Tabela 3 -	Proporção esperada pela lei Newcomb-Benford.....	33
Tabela 4 -	Proporção esperada para os dois primeiros dígitos.....	34
Tabela 5 -	Frequências esperadas pela lei Newcomb-Benford.....	54
Tabela 6 -	Faixa de valores do MAD.....	64
Tabela 7 -	D^2 crítico.....	66
Tabela 8 -	Valor $K-S_{\text{crítico}}$	72
Tabela 9 -	Grupos estudados e respectivas OMs.....	76
Tabela 10 -	Código de cada OM utilizada no Tesouro Gerencial.....	78
Tabela 11 -	Perfil de dados dos CeIM.....	116
Tabela 12 -	Perfil de dados das EAM.....	116
Tabela 13 -	Perfil de dados das BE.....	116
Tabela 14 -	Perfil de dados dos HN.....	116
Tabela 15 -	Estatística descritiva dos saldos das NE em reais.....	123
Tabela 16 -	Análise global para os primeiros dígitos da LNB.....	133
Tabela 17 -	P-valor do teste Z por primeiro dígito dos CeIM.....	133
Tabela 18 -	P-valor do teste Z por primeiro dígito das EAM.....	134
Tabela 19 -	P-valor do teste Z por primeiro dígito das BE.....	134
Tabela 20 -	P-valor do teste Z por primeiro dígito dos HN.....	134
Tabela 21 -	Análise global para os segundos dígitos da LNB.....	135
Tabela 22 -	Teste Z para os segundos dígitos dos CeIM.....	135
Tabela 23 -	Teste Z para os segundos dígitos das EAM.....	136
Tabela 24 -	Teste Z para os segundos dígitos das BE.....	136
Tabela 25 -	Teste Z para os segundos dígitos dos HN.....	136
Tabela 26 -	Análise global para os dois primeiros dígitos da LNB.....	137
Tabela 27 -	Dois primeiros dígitos em não conformidade com a LNB.....	142
Tabela 28 -	Análise de segunda ordem.....	143
Tabela 29 -	Maiores desvios da análise da soma para os primeiros dígitos.....	171
Tabela 30 -	Maiores desvios da análise da soma para os dois primeiros dígitos.....	172

Tabela 31 -	Valores com maior incidência nos CeIM.....	173
Tabela 32 -	Valores com maior incidência nas EAM.....	174
Tabela 33 -	Valores com maior incidência nas BE.....	175
Tabela 34 -	Valores com maior incidência nos HN.....	176
Tabela 35 -	Resultados do modelo fator de distorção.....	177

LISTA DE ABREVIATURAS E SIGLAS

BAeNSPa	Base Aérea Naval de São Pedro da Aldeia
BE	Base e Estação Naval
BNA	Base Naval de Aratu
BFLa	Base Fluvial de Ladário
BNN	Base Naval de Natal
BNRG	Base Naval do Rio Grande
ENRN	Estação Naval do Rio Negro
BNRJ	Base Naval do Rio de Janeiro
BNVC	Base Naval de Val-de-Cães
CCIMAR	Centro de Controle Interno da Marinha
CeIM	Centro de Intendência da Marinha
CeIMBe	Centro de Intendência da Marinha em Belém
CeIMLa	Centro de Intendência da Marinha em Ladário
CeIMMa	Centro de Intendência da Marinha em Manaus
CeIMNa	Centro de Intendência da Marinha em Natal
CeIMRG	Centro de Intendência da Marinha em Rio Grande
CeIMSa	Centro de Intendência da Marinha em Salvador
CeIMSPA	Centro de Intendência da Marinha em São Pedro da Aldeia
EAM	Escolas de Aprendizes-Marinheiros
EAMCE	Escola de Aprendizes-Marinheiros do Ceará
EAMES	Escola de Aprendizes-Marinheiros do Espírito Santo
EAMPE	Escola de Aprendizes-Marinheiros de Pernambuco
EAMSC	Escola de Aprendizes-Marinheiros de Santa Catarina
FD	Fator de Distorção
H ₀	Hipótese Nula

H ₁	Hipótese Alternativa
HCM	Hospital Central da Marinha
HN	Hospital Naval
HNBe	Hospital Naval de Belém
HNBra	Hospital Naval de Brasília
HNLa	Hospital Naval de Ladário
HNMD	Hospital Naval Marcílio Dias
HNNa	Hospital Naval de Natal
HNRe	Hospital Naval de Recife
HNSa	Hospital Naval de Salvador
IBRACON	Instituto dos Auditores Independentes do Brasil
INEP	Instituto Nacional de Pesquisas e Estudos Educacionais
ISS	Imposto Sobre Serviço de Qualquer Natureza
K-S	Kolmogorov-Smirnov
MA	Média Atual
MAD	Desvio Absoluto Médio
MB	Marinha do Brasil
ME	Média Esperada
MFD	Modelo Fator de Distorção
MPDFT	Ministério Público do Distrito Federal e Territórios
MPM	Ministério Público Militar
NA	<i>Not Available</i>
OMs	Organizações Militares
PIB	Produto Interno Bruto
SEFAZ-RJ	Secretaria da Fazenda do Estado do Rio de Janeiro
UG	Unidade Gestora
UGE	Unidade Gestora Executora
χ^2	Qui Quadrado

LISTA DE SÍMBOLOS

a	Primeiro dígito de um número
ab	Dois primeiros dígitos de um número
b	Segundo dígito de um número
B_{12j}	Frequência esperada do j -ésimo dígito como os dois primeiros dígitos
B_{1j}	Frequência esperada do j -ésimo dígito como o primeiro dígito
B_{2j}	Frequência esperada do j -ésimo dígito como o segundo dígito
D_n	Diferença de dois termos subsequentes
f_{12j}	Frequência observada dos dois primeiros dígitos do j -ésimo dígito
f_{1j}	Frequência observada no primeiro dígito do j -ésimo dígito
f_{2j}	Frequência observada no segundo dígito do j -ésimo dígito
ft	Curva qualquer de uma combinação de dois dígitos
int	Utilizar a parte inteira do resultado
k	Número de linhas do teste estatístico
K	Quantidade de algarismos analisados
\log	Logaritmo na base 10
$\text{mod}1$	Utilizar a parte decimal do resultado
n	Número de elementos da amostra
pe	Proporção esperada
PE	Proporção absoluta esperada da amostra
po	Proporção observada
PO	Proporção absoluta de observações da amostra
r	Razão de uma progressão geométrica
T_n	n -ésimo termo de uma progressão geométrica
w	Primeiro termo de uma progressão geométrica
X_1, X_2, \dots, X_n	n dados da amostra
α	Nível de significância

SUMÁRIO

	INTRODUÇÃO	20
1	REFERENCIAL TEÓRICO	25
1.1	A fraude e a análise de dados	25
1.2	A lei Newcomb-Benford	30
1.2.1	<u>Um Breve Histórico sobre a LNB</u>	31
1.2.2	<u>Estudos anteriores que utilizaram a LNB para detecção de fraudes</u>	39
1.3	Procedimentos não estatísticos e testes estatísticos utilizados para verificação de conformidade com a lei Newcomb-Benford	52
1.3.1	<u>Procedimentos não estatísticos</u>	52
1.3.1.1	Análises pré-Benford.....	52
1.3.1.2	Análises primárias.....	54
1.3.1.3	Análises avançadas.....	56
1.3.1.3.1	Análise de segunda ordem.....	57
1.3.1.3.2	Análise da soma.....	58
1.3.1.4	Análises associadas.....	59
1.3.1.4.1	Análise dos números duplicados.....	59
1.3.1.4.2	Análise dos últimos dois dígitos.....	60
1.3.1.4.3	Modelo fator de distorção.....	60
1.3.1.5	Análises de conformidade.....	62
1.3.1.5.1	Z_{teste} - modelo fator de distorção.....	63
1.3.1.5.2	Desvio absoluto médio (Mean Absolute Deviation).....	63
1.3.1.5.3	<i>Mantissa arc test</i>	64
1.3.2	<u>Testes estatísticos</u>	66
1.3.2.1	Teste Z.....	68
1.3.2.2	Teste qui quadrado (χ^2).....	70
1.3.2.3	Teste Kolmogorov-Smirnov (K-S).....	71
2	METODOLOGIA	73
2.1	Método da pesquisa	73
2.2	Seleção da amostra	74
2.3	Coleta de dados	77

2.4	Tratamento de dados	79
2.4.1	<u>Importação e edição dos dados</u>	80
2.4.2	<u>Aplicação das análises pré-Benford</u>	81
2.4.3	<u>Aplicação das análises primárias</u>	82
2.4.4	<u>Aplicação das análises avançadas</u>	84
2.4.5	<u>Aplicação das análises associadas</u>	85
2.5	Delimitação da pesquisa	85
3	ANÁLISE E DISCUSSÃO DOS RESULTADOS	87
3.1	Análises pré-Benford dos dados	87
3.2	Análises primárias dos dados	89
3.3	Análises avançadas dos dados	93
3.4	Análises associadas dos dados	94
	CONSIDERAÇÕES FINAIS	96
	REFERÊNCIAS	100
	APÊNDICE A - Linguagem de programação empregada.....	112
	APÊNDICE B - Resultados das análises pré-Benford.....	116
	APÊNDICE C - Resultados das análises primárias.....	133
	APÊNDICE D - Resultados das análises avançadas.....	143
	APÊNDICE E - Resultados das análises associadas.....	173

INTRODUÇÃO

A fraude é um fenômeno social em que os benefícios dos fraudadores são provenientes de prejuízos das vítimas. Essas vítimas podem ser pessoas físicas, empresas, o governo ou até a sociedade como um todo (BAESENS; VAN VLASSELAER; VERBEKE, 2015).

Nos Estados Unidos, uma empresa perde por ano, em média, 5% do seu lucro, devido às fraudes. No tocante ao seguro saúde, é estimada uma perda maior que 40 bilhões de dólares por ano. No Reino Unido, o custo anual da fraude é de 73 bilhões de libras. Isso demonstra a importância das organizações e do governo lutarem para a mitigação da fraude, investindo em mecanismos para sua detecção (BAESENS; VAN VLASSELAER; VERBEKE, 2015).

A *Association of Certified Fraud Examiners* (2012) emitiu um relatório dividindo a fraude em três principais categorias: apropriação indevida de ativos, fraude nas demonstrações financeiras e corrupção. Este relatório concluiu que a apropriação indevida de ativos é o tipo de fraude mais comum nos Estados Unidos, representando 87% dos casos, sendo, entretanto, o tipo que tem menor média de perda, por volta de 120 mil dólares por fraude. As fraudes nas demonstrações contábeis, que representam apenas 8% dos casos, têm, por outro lado, a maior perda média, cerca de 1 milhão de dólares por fraude. Por último, a corrupção representa um terço dos casos e causa uma perda média de 250 mil dólares por fraude.

Quando o assunto é fraude, a situação no Brasil foi ilustrada por Kanitz (2012), ao afirmar que “o Brasil não é um país corrupto, é apenas um país pouco auditado”. Isto está em linha com o que foi relatado pelo IBRACON (2013), que apresentou, em diversos países, a quantidade de habitantes para cada auditor. No Brasil, esta proporção é de 20 mil habitantes por auditor, enquanto em países como Estados Unidos, Inglaterra e Holanda está razão é de, respectivamente, 2.100, 1.100 e 899 habitantes por auditor. O IBRACON (2016) alerta, também, que existe um descaso com as atividades de auditoria interna, auditoria externa e controle que, só ganham destaque no Brasil, quando escândalos de corrupção ocupam espaço na mídia.

Um relatório, emitido pela Federação de Indústria do Estado de São Paulo (2010), frisa que a fraude é responsável por prejuízos importantes para o Brasil, onde, em média, 1,84% do Produto Interno Bruto é perdido por ano, devido a corrupção. Em 2008, aproximadamente 55,3 bilhões de reais foram desviados no Brasil por este motivo. O relatório

frisa, também, que um menor controle da corrupção implica em um aumento dos recursos desviados, ressaltando, assim, a importância da auditoria no setor público e privado.

Em 2019, o Brasil obteve a 106ª posição entre 180 países no relatório emitido pela Transparência Internacional, que divulga o *ranking* do Índice de Percepção de Corrupção. Esta foi sua pior posição desde 2012, quando os dados passaram a ser comparáveis ano a ano. Esta classificação significou a quarta queda anual consecutiva. Este relatório considera a corrupção como principal barreira para a prosperidade econômica e social do Brasil (TRANSPARÊNCIA INTERNACIONAL, 2020).

No Brasil, o combate à fraude e à corrupção se dá pela atuação de diversos órgãos, cada um em sua esfera e escopo de atuação, sendo que a intenção é um elemento importante para diferenciar a fraude do erro (BRASIL, 2018). Sá (2002) considera que a diferenciação entre a fraude e o erro é dada por uma linha tênue. A fraude é cometida por uma pessoa que, através de um conjunto de ações premeditadas, tem com o intuito conseguir algum tipo de benefício ilegal. Por outro lado, o erro é proveniente de uma ação involuntária, que pode ou não causar dano a uma pessoa ou empresa.

No âmbito da Marinha do Brasil (MB), o órgão responsável pela auditoria, atuando no combate à fraude, é o Centro de Controle Interno da Marinha (CCIMAR), que tem como missão “planejar, dirigir, coordenar e executar as atividades de controle interno, por meio da realização de auditoria, análise e apresentação de contas no âmbito do Comando da Marinha” (BRASIL, 2012).

O principal papel dos auditores é analisar procedimentos, verificar sua veracidade e conformidade com as legislações vigentes, devendo, neste contexto, reportar a alta administração qualquer suspeita de fraude (CHURCH; MCMILLAN; SCHNEIDER, 2001). O *Institute of Internal Auditors* (2012) salienta que o fato dos auditores analisarem as contas do setor público amplia a convicção da população com relação ao comprometimento e probidade das instituições públicas.

Com o crescimento do volume de dados gerados, é indispensável ao auditor a utilização de novas ferramentas para apoiar seu trabalho. Neste contexto, a lei Newcomb-Benford (LNB) ganha espaço como ferramenta capaz de auxiliar o auditor na sua tomada de decisão (RIBEIRO; MONSUETO, 2015). A efetividade, facilidade e simplicidade advindas das aplicações da LNB mostra sua viabilidade na fase de planejamento da auditoria e seleção da amostra auditada, configurando-a como ferramenta útil para o auditor (COSTA; SANTOS; TRAVASSOS, 2012).

Lagioia *et al.* (2011) consideram a lei Newcomb-Benford um instrumento válido nas auditorias, por facilitar o seu planejamento e a escolha da amostra que será averiguada mais a fundo. Nesta lei, a probabilidade de um número se iniciar com os algarismos “1”, “2” e “3” é maior do que com os algarismos restantes - aproximadamente 60%.

A utilização da análise de dados, com apoio de *softwares*, é de grande valia para os auditores, pois torna seu trabalho mais eficaz e simples, contribuindo para detecção de fraudes ou anomalias de modo mais rápido e com menor custo (DURTSCHI; HILLISON; PACINI, 2004). Na mesma linha, Santos, Diniz e Ribeiro Filho (2003) consideram a utilização da análise de dados, em conjunto com a lei Newcomb-Benford, uma ferramenta poderosa na etapa de planejamento do trabalho do auditor.

Silva (2009) trata a LNB como uma abordagem útil utilizada internacionalmente na detecção de fraudes e na escolha de amostras de auditoria. Esta lei foi utilizada por diversos autores para averiguar o viés de fraude em notas de empenhos (NE) como Santos, Diniz e Corrar (2005), Costa, Santos e Travassos (2012), Freitas Costa *et al.* (2012), Costa (2012), Nascimento, Filho e Buscacio (2014), Ganassin, Costa e Wilbert (2016), Silva, Travassos e Costa (2017), Cella e Zanolla (2018), De Oliveira *et al.* (2018) e Macieira *et al.* (2019). Assim, este estudo parte da seguinte indagação: As notas de empenhos emitidas, no ano de 2019, por organizações militares (OM) da Marinha do Brasil seguem a lei Newcomb-Benford?

Considerando este problema de pesquisa, a dissertação tem como objetivo geral avaliar se as notas de empenhos emitidas em 2019 por organizações militares da Marinha do Brasil estão em conformidade com a lei Newcomb-Benford.

Para que o objetivo geral deste trabalho seja alcançado, será necessário o cumprimento dos seguintes objetivos específicos:

- aquilatar a utilização da LNB na detecção de fraudes;
- identificar os testes estatísticos utilizados para verificar a conformidade com lei Newcomb-Benford;
- identificar os procedimentos não estatísticos, propostos por Nigrini (2012) e Nigrini (2020), para verificar a conformidade com lei Newcomb-Benford; e
- analisar a distribuição das notas de empenhos emitidas em 2019 por organizações militares da Marinha do Brasil e sua conformidade com LNB.

O interesse em pesquisar sobre a LNB tem aumentado pelo mundo no decorrer dos anos. Kossovsky (2014) observou um crescimento na quantidade de artigos publicados sobre este tema entre os anos de 1940 e 2009. O autor dividiu os artigos em três classes: trabalhos

teóricos, preocupados com a fundamentação matemática da lei, trabalhos expositivos, que explicam as propriedades da lei e trabalhos que aplicam a lei e tiram conclusões a partir da realização de testes. Esta pesquisa terá como foco a última classe. A Tabela 1 apresenta o crescimento, no decorrer dos anos, das publicações de artigos sobre a lei Newcomb-Benford:

Tabela 1 – Publicações sobre a LNB

Ano	Número de publicações
1940 - 49	5
1950 - 59	5
1960 - 69	18
1970 - 79	40
1980 - 89	72
1990 - 99	101
2000 - 09	416

Fonte: KOSSOVSKY, 2014, p. vii.

No cenário global, os Estados Unidos estão na vanguarda das publicações que relacionam a lei Newcomb-Benford e a auditoria contábil, com a produção de 45,52% do total de artigos. O Brasil, por sua vez, contribui com apenas 8,28% das publicações, que estão concentradas, principalmente, em três autores: Josenildo dos Santos, professor da Universidade Federal de Pernambuco; Josedilton Alves Diniz, professor da Universidade Federal da Paraíba; e Luiz João Corrar, professor da Universidade de São Paulo (COSTA, 2012).

Santos, Diniz e Corrar (2005), Costa, Santos e Travassos (2012), Freitas Costa *et al.* (2012), Costa (2012), Nascimento, Filho e Buscacio (2014), Ganassin, Costa e Wilbert (2016), Silva, Travassos e Costa (2017), Cella e Zanolla (2018), De Oliveira *et al.* (2018) e Macieira *et al.* (2019), avaliaram a conformidade da distribuição das notas de empenhos emitidas em períodos anteriores a 2019 com a LNB. Neste período a despesa pública era realizada sob os preceitos da Lei nº 8.666, de 21 de junho de 1993, que permitia a aquisição de bens e contratação de serviços (exceto os de engenharia) sem licitação, até o limite de R\$ 8.000,00. O ano de 2019 foi o primeiro ano em que a despesa pública foi realizada à luz do Decreto nº 9.412, de 18 de junho de 2018, que atualizou os valores das aquisições de bens e contratações de serviços (exceto os de engenharia), sem licitação, para R\$ 17.600,00.

Pelos motivos expostos, esta pesquisa torna-se relevante na medida em que busca contribuir para o aumento do conhecimento sobre aplicações da lei Newcomb-Benford na área da auditoria contábil, além de apoiar a difusão do conhecimento da própria LNB.

Este propósito é ratificado por Silva (2009), ele afirma que embora a lei Newcomb-Benford seja um instrumento para mensurar a qualidade das informações, com grande aplicação internacional para detecção de fraudes, no Brasil a lei é pouco utilizada.

Vale ressaltar que a conformidade com a LNB não implica que não exista a chance de haver fraude ou erro nas informações examinadas. Da mesma forma, potenciais resultados de não conformidade não significa obrigatoriamente a existência de fraudes ou erros nos dados analisados. A certeza de fraudes nos dados só existirá após o auditor realizar uma auditoria qualitativa, o que poderá confirmar ou não os indícios de fraudes (MILANI FILHO, 2013).

Este trabalho se limita a estudar as notas de empenhos emitidas por 27 Unidades Gestoras Executoras (UGE) da Marinha no Brasil, no ano de 2019. O fato de serem estudadas notas de empenhos de UGE da MB está alinhado com o objetivo do trabalho de avaliar se as notas de empenhos emitidas em 2019 por organizações militares da Marinha do Brasil estão em conformidade com a lei Newcomb-Benford. Foram selecionadas apenas as Unidades Gestoras Executoras, em razão destas centralizarem a execução financeira de diversas Unidades Gestoras Responsáveis, permitindo, assim, um aumento na quantidade de dados analisados. Já para escolha das organizações militares para a amostra, foi utilizada a seleção não probabilística por julgamento, permitindo assim a seleção de organizações militares semelhantes quanto à missão para permitir a comparação dos resultados encontrados.

Esta dissertação é constituída de seis seções, iniciando com essa introdução. Na próxima seção foi apresentado o referencial teórico utilizado como base para o trabalho. Na terceira seção foi apresentada a metodologia aplicada. Na quarta seção foi feita a análise e interpretação dos dados, e, finalmente, na quinta e última seção foram apresentadas as considerações finais e sugestão para trabalhos futuros. Os apêndices apresentam a linguagem de programação utilizada nos procedimentos empregados nesta dissertação, os resultados da análise pré-Benford, os resultados da análise primária, os resultados da análise avançada e os resultados da análise associada.

1 REFERENCIAL TEÓRICO

Esta seção apresenta o arcabouço conceitual utilizado como base para alcançar o objetivo geral proposto nesta dissertação, que é avaliar se as notas de empenhos emitidas em 2019 pelas organizações militares da Marinha do Brasil estão em conformidade com a lei Newcomb-Benford. No referencial teórico, abordam-se os seguintes tópicos: a fraude e a análise de dados, a lei Newcomb-Benford e os procedimentos não estatísticos e testes estatísticos utilizados de forma complementar para verificação de conformidade com a lei Newcomb-Benford.

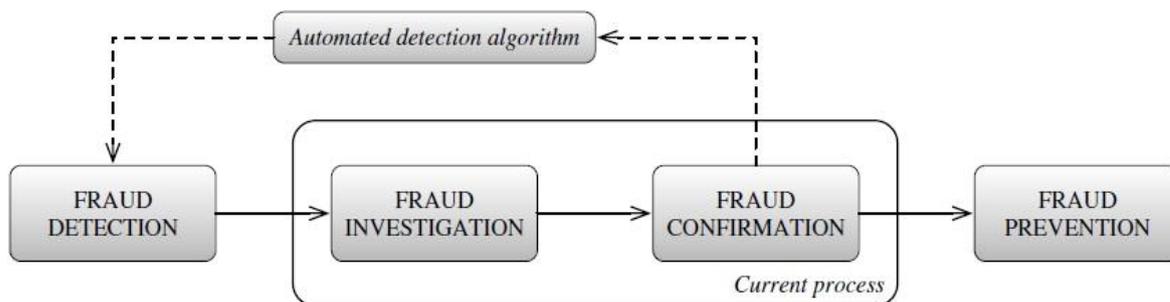
1.1 A fraude e a análise de dados

Para estudar a fraude, deve-se, primeiramente, saber o que a caracteriza. Para esta ser efetivada, devem estar presentes, obrigatoriamente, os seguintes elementos: haver uma afirmação falsa, o indivíduo deve estar ciente que a afirmação que fez é falsa, ter a intenção de enganar a vítima, a vítima considerou a declaração como verdadeira e a vítima foi prejudicada financeiramente ou de outra maneira qualquer (GEE, 2015).

Baesens, Van Vlasselaer e Verbeke (2015) definiram fraude como um crime incomum, bem planejado e imperceptivelmente oculto, que evolui com o tempo, que, muitas vezes, é cuidadosamente organizado e aparece em muitas formas. Eles abordam que o ciclo da fraude possui quatro atividades essenciais:

- Detecção da fraude, a análise de dados detecta uma possível fraude;
- Investigação da fraude, a expertise humana é necessária para avaliar se a detecção da etapa anterior configura ou não uma fraude;
- Confirmação da fraude, etapa na qual é confirmada a fraude. Caso seja confirmado um novo tipo de fraude, o algoritmo da análise de dados é retroalimentado com essa nova informação; e
- Prevenção de fraude, que se baseia na ideia de sempre estar um passo à frente do fraudador, prevenindo das fraudes antes que elas sejam cometidas.

Figura 1 – Ciclo da fraude



Fonte: BAESENS; VAN VLASSELAER; VERBEKE, 2015, p. 23.

Umar, Samsudin e Mohamed (2015) afirmaram que, as funções dos auditores internos e externos são ferramentas importantes na prevenção de fraudes. Na mesma linha, Albrecht *et al.* (2019) mencionam a relevância dos auditores para a detecção de fraudes e a utilização, por estes, da análise de dados.

Baesens, Van Vlasselaer, e Verbeke (2015) consideram que a análise de dados fornece ferramentas poderosas aos auditores, aumentando, assim, sua capacidade em analisar uma grande quantidade de dados, com a finalidade de detectar fraudes. Balaniuk (2010), entretanto, alerta que a disponibilidade de grande quantidade de dados não garante, necessariamente, que o auditor consiga extrair deles alguma evidência ou confirmação de fraude. É necessário que ele, para uma melhor análise, saiba utilizar ferramentas estatísticas e conheça as normas da administração pública.

Albrecht *et al.* (2019) consideram que o processo de análise de dados se assemelha ao de procurar uma agulha no palheiro, visto que as transações fraudulentas nos registros eletrônicos estão em quantidades muito menores do que o total de transações realizadas. Assim, Gee (2015) salienta que entender o negócio que está sendo avaliado, suas práticas e seus procedimentos ajuda a identificar e diferenciar as anomalias que não são fraudes das que realmente são.

Na mesma linha de Albrecht *et al.* (2019), Kovalerchuk, Vityaev e Holtfreter (2007) frisam que, no cotidiano, a porcentagem transações fraudulentas é muito pequena quando comparada com o total de transações efetuadas no dia a dia.

Gee (2015) divide a análise de dados em três tipos: análise de dados exploratória, confirmatória e qualitativa. A análise exploratória consiste no estágio inicial, quando pouco é sabido sobre os dados e as hipóteses são formadas, na maior parte das vezes, utilizando técnicas visuais e gráficas. A análise de dados confirmatória ocorre quando as hipóteses

levantadas na etapa inicial são aceitas ou rejeitadas. Já a análise qualitativa é utilizada para evidenciar as conclusões de dados não numéricos e é largamente utilizada nas ciências sociais.

O ciclo da análise de dados é, também, composto por três estágios. O primeiro é a avaliação e análise, quando o cenário que será investigado é inteiramente compreendido, permitindo a formação de um arcabouço para escolher uma linha base para a realização de comparações, com a finalidade de identificar vieses de fraudes. O segundo estágio é a tecnologia e *software*, quando, já identificada a linha base, são utilizados *softwares* como ferramentas para identificar os dados que se distanciaram da sua linha base definida na primeira etapa. O estágio final é composto por auditoria e investigação, no qual são considerados os dados identificados na segunda fase para uma análise mais criteriosa, pois nenhum tipo de *software* ou tecnologia pode confirmar uma fraude na etapa anterior (GEE, 2015).

Baesens, Van Vlasselaer, e Verbeke (2015) afirmam que são três os principais motivos para as organizações utilizarem a análise de dados para identificação de possíveis fraudes. O primeiro motivo é que modelos estatísticos utilizados na análise de dados são mais precisos, quando comparados com os métodos tradicionais, em virtude de terem um maior poder de detecção. O segundo motivo é devido à análise de dados possuir uma maior eficiência operacional, tendo em vista que consegue avaliar toda a população analisada com muita rapidez. O último motivo é o baixo custo da implementação da análise de dados nas organizações, quando comparado com o impacto que a não detecção da fraude acarreta.

Legitimando a afirmação de Baesens, Van Vlasselaer, e Verbeke (2015), Bierstaker, Brody e Pacini (2006), ao realizar uma pesquisa com 86 contadores que trabalhavam como examinadores de fraudes certificados na auditoria interna e tinham, em média, 13,94 anos de experiência, concluíram que, embora pouco usadas, as técnicas de análise de dados, mineração de dados e contabilidade forense sejam mais eficazes. Eles perceberam, também, que esta pouca utilização é motivada pelas preocupações com o custo, mesmo que uma ocorrência de fraude possa ser mais danosa para a empresa do que o custo de implementação das citadas técnicas.

Segundo Gee (2015), uma vantagem possibilitada pela análise de dados é permitir ao auditor examinar 100% dos dados e fornecer uma lista de dados anômalos. Isto, porém, não significa que todos os dados anômalos são fraudulentos. O auditor necessitará aplicar seu julgamento profissional, habilidades analíticas e intuição para, depois, certificar se o dado é uma fraude ou não.

Na mesma linha de Gee (2015), Özer e Babacan (2013) alertaram que, embora a análise de dados revele anomalias dos dados tendo como referência uma linha base, isto não caracteriza uma fraude ou manipulação, apenas um viés. Assim, a análise de dados funciona como uma ferramenta para o auditor, pois sinaliza onde estão os dados que não estão em conformidade com a linha base. Não obstante, apenas através dela não é possível afirmar se os dados são fraudulentos ou não, somente são levantados indícios para uma posterior análise qualitativa do auditor.

A auditoria é regularmente auxiliada pela análise de dados, permitindo ao auditor, encontrar comportamentos anormais, como duplicações, arredondamentos e tendências para algum dígito específico (MILANI FILHO, 2013). Nigrini e Mittermaier (1997) salientaram que a análise de dados é adequada para casos em que as transações são armazenadas em formato eletrônico e os auditores encontrem grandes conjuntos de dados, nos quais a auditoria da totalidade das transações se torna algo impraticável e economicamente inviável.

Assim, em situações que demandem a necessidade da escolha de uma amostra de auditoria para uma verificação mais qualitativa, a análise de dados, em conjunto com ferramentas estatísticas, auxilia o auditor nesta escolha. Esta metodologia torna possível o auditor inspecionar todos os dados existentes e contribuir para sua tomada de decisão, indicando quais amostras ele deve escolher, diminuindo, assim, os custos e o tempo gastos na auditoria. (COSTA; DUTRA, 2014).

Diniz et al. (2006) alertam que a delimitação da amostra de auditoria não deve ser feita de maneira aleatória e sim utilizando conhecimento científico de matemática e estatística, controlando tecnicamente, assim, os riscos.

O processo de utilizar a análise de dados em conjunto com testes estatísticos auxilia o auditor na detecção de possíveis fraudes, permitindo a análise de 100% da população por um baixo custo. (BAESENS, 2014). Na mesma linha de Baensens (2014), Gee (2015) salienta que, com os avanços da tecnologia, a produção de uma grande quantidade de dados por dia pode esconder algumas transações maliciosas, tornando-as muito difíceis de serem detectadas pelos métodos tradicionais. Ele considera, também, a análise de dados uma ferramenta útil para detecção de potenciais sinais de fraudes, com o auxílio de *software*, pois este método tem a capacidade de analisar grande quantidade de dados e destacar as anomalias.

A lei Newcomb-Benford pode ser empregada na análise de dados. Sua utilização permite ao auditor analisar toda a população de forma rápida e escolher para a amostra de auditoria os dados que não estão em conformidade com esta lei. Ela é aplicada para identificar anomalias em vários tipos de registros contábeis, tais como diários de desembolso, cheques,

depósitos, transferências e outros dados financeiros, verificando, assim, a possibilidade de atividade fraudulenta nos dados avaliados (MANTONE, 2013).

Coglitore e Berryman (1988) atestam que os procedimentos da análise dados tendo como base a LNB são úteis para detectar declarações de receitas subestimadas, vendas fictícias, inventário fictício e sobrevalorizado, dívidas incobráveis subestimadas e subsídios para contas duvidosas, passivos de compra não registados, despesas registradas a menor e capitalização de despesa inadequada.

Nigrini (1999) afirma que nem todos os tipos de dados seguem a lei Newcomb-Benford e descreve quais as características estes devem conter para terem maiores chances de seguir esta lei. As características são as seguintes: os números devem descrever os tamanhos de fenômenos semelhantes, os números não devem ter máximos ou mínimos embutidos e os dados não devem ser números que sejam utilizados para identificar alguma coisa, como por exemplo, número de conta corrente, número de voo ou número de identificação de item no estoque.

Para os dados terem maior chance de aderir à lei Newcomb-Benford, Wallace (2002) propõe que eles devem cumprir dois critérios principais: a média do conjunto de dados deve ser maior que a mediana e o conjunto de dados deve ser caracterizado por um valor de assimetria positivo.

A lei Newcomb-Benford pode ser usada como uma técnica de triagem para detecção de fraudes, porém, como todas as ferramentas de análise de dados, estar de acordo com a lei não significa que a amostra não contenha dados fraudulentos e, da mesma forma, o fato dos dados da amostra não seguirem a lei não indica necessariamente a presença de dados fraudulentos na amostra (JAMAIN, 2001).

Na mesma linha de Jamain (2001), Nigrini (2012) exemplifica como utilizar a análise de dados em consonância com a LNB. Caso o auditor tenha quatro amostras e após a análise de dados verificar que três delas seguem a lei Newcomb-Benford, ele deve adotar a estratégia de focar na amostra que não seguiu a lei, porque ela tem mais probabilidade de conter fraude ou erro.

Ettredge e Srivastava (1999) consideram a análise de dados um método de avaliar os padrões em uma amostra de números e determinar se esta é equivalente a uma população de números. Eles afirmam que a distribuição explicitada pela lei Newcomb-Benford é usada como *benchmark* para testar a validade dos dados e o desvio do *benchmark* pode configurar fraude.

Bhattacharya, Xu e Kumar (2011) afirmam que a utilização da lei Newcomb-Benford como uma ferramenta na detecção de fraudes em dados financeiros é motivada pelo fato dela não poder ser facilmente reproduzida manualmente, ou seja, mesmo que o fraudador conheça a lei Newcomb-Benford dificilmente conseguirá reproduzi-la. Assim, quando na amostra existirem dados inventados ou fraudulentos, ela tem maior chance de não estar em conformidade com a LNB, facilitando, deste modo, a identificação de fraudes, por meio de análises baseadas nas proporções observadas e esperadas.

O relatório da ACFE, AICPA e IIA (2008) afirma que o uso da análise de dados, baseada na LNB, no gerenciamento da auditoria permite, mais rapidamente, a identificação de atividades fraudulentas. Adicionalmente, considera esta análise útil para examinar relatórios de despesas, contas contábeis e transações financeiras, para a escolha dos dados que passarão por análises adicionais.

Gee (2015) reitera que a lei Newcomb-Benford é uma ferramenta que faz parte de muitos planos de auditoria, pelo motivo de fornecer um ponto de partida para o trabalho do auditor. Na mesma direção, Nigrini (1996) considera esta lei como promissora e útil na área de auditoria contábil. Os auditores que avaliam conjuntos de dados precisarão avaliar se as frequências observadas são válidas para os dados em consideração, podendo assim detectar viés de fraude ao avaliar toda a amostra.

Entretanto, é sempre importante lembrar que resultados de não conformidade, na utilização da análise de dados com a lei Newcomb-Benford, não caracteriza que os dados apurados contenham fraude obrigatoriamente, isto apenas pode ser certificado por técnicas minuciosas de auditoria. A não conformidade apenas fornece ao auditor indícios e não a certeza de fraudes. Da mesma forma, a conformidade com a LNB não expressa a certeza da não existência de fraudes nos dados averiguados e sim a falta de viés para fraude, obtendo, assim, uma credibilidade maior nos dados que estão em conformidade com relação aos que não possuem conformidade (MILANI FILHO et. al., 2016).

1.2 A lei Newcomb-Benford

A lei Newcomb-Benford é definida como um fenômeno baseado na quantidade de vezes que um determinado algarismo deve aparecer em uma posição específica de um

determinado número. A análise de dados, pautada na lei, faz a busca de transações fraudulentas, com a premissa que os dígitos de cada número tendem a aparecer nas proporções esperadas. Neste contexto, pode-se observar que a lei Newcomb-Benford não é uma técnica infalível para capturar fraudes, mas sim uma ferramenta útil para ajudar na identificação de dados maliciosos para futuros testes adicionais e, portanto, auxiliar auditores na sua busca de detectar fraudes (DURTSCHI; HILLISON; PACINI, 2004).

Assim, a lei Newcomb-Benford é uma ferramenta que usada em conjunto com a análise de dados, permite ao auditor verificar toda a amostra em busca de anomalias. Quando esta ferramenta aponta um conjunto de dados anômalos, isto não implica que este conjunto de dados seja fraudulento, mas sim que eles precisam ser analisados por um auditor, que aplicará seu julgamento profissional, suas habilidades analíticas e sua intuição para, só assim, determinar se os dados são fraudulentos ou não. É importante salientar que é natural a ocorrência de falsos positivos, ou seja, dados que para a LNB são considerados fraudulentos e que, após uma análise mais criteriosa do auditor, são classificados, como dados honestos (GEE, 2015).

Nigrini (1999) deu uma explicação intuitiva para a lei Newcomb-Benford, considerando que uma aplicação financeira renda 10% ao ano. Caso nesta aplicação tenha \$100 milhões, seu primeiro dígito é “1” e só sofrerá alteração quando o valor atingir \$ 200 milhões. Com a taxa de juros de 10%, isso ocorrerá após 7,3 anos. Já para a aplicação passar de \$500 milhões para \$600 milhões e assim trocar seu primeiro algarismo, isto demorará, aproximadamente, 1,9 anos, tempo muito menor do que aquele para alterar seu primeiro dígito de “1” para “2”. Fazendo isso para todos os primeiros algarismos, nota-se que a aplicação passará mais tempo com os primeiros dígitos “1”, “2” e “3” do que com outros dígitos. Assim, quando maior o dígito menor o tempo que ele permanecerá como primeiro dígito da aplicação. Isso irá ocorrer até a aplicação chegar a \$1 bilhão e o ciclo, novamente, se repetirá.

1.2.1 Um Breve Histórico sobre a LNB

O primeiro a escrever sobre o fenômeno que depois seria conhecido como a lei Newcomb-Benford foi Simon Newcomb. Ele notou que as primeiras páginas das tabelas logarítmicas, utilizadas na época para realizar multiplicações de forma mais rápida e prática,

se encontravam mais desgastadas, devido ao uso, em relação às outras. Pautado nesta observação, deduziu que os números das primeiras folhas seriam mais utilizados que os das últimas, ou seja, números começando com os algarismos mais baixos seriam mais procurados nas tabelas logarítmicas do que com algarismos mais altos. Assim, obteve a conclusão, de que os primeiros dígitos significativos dos dados provenientes da natureza não seriam distribuídos uniformemente conforme previsto, mas seguiam, sim, uma lei logarítmica (NEWCOMB, 1881).

Muitos anos depois de Newcomb (1881), Benford (1938) escreveu sobre o tema e, com o mesmo *insight* de Newcomb, notou que as primeiras páginas de uma tabela logarítmica mostravam mais desgaste do que as últimas, indicando que os números mais usados começam mais com o dígito “1” do que com o dígito “9”, concluindo, assim, que os números mais utilizados têm como primeiro dígito um algarismo de menor valor. Ele coletou 20.229 números retirados das mais diferentes fontes, que não tinham correlação entre si, para serem submetidos à análise de seus primeiros algarismos. Na Tabela 2, pode-se ver como ficou a distribuição dos dados coletados.

Tabela 2 – Dados analisados por Benford (1938)

Grupo	Descrição	Total	Primeiro Dígito								
			1	2	3	4	5	6	7	8	9
A	Rios	335	31,0	16,4	10,7	11,3	7,2	8,6	5,5	4,2	5,1
B	População	3259	33,9	20,4	14,2	8,1	7,2	6,2	4,1	3,7	2,2
C	Constantes	104	41,3	14,4	4,8	8,6	10,6	5,8	1,0	2,9	10,6
D	Jornais	100	30,0	18,0	12,0	10,0	8,0	6,0	6,0	5,0	5,0
E	Calor específico	1389	24,0	18,4	16,2	14,6	10,6	4,1	3,2	4,8	4,1
F	Pressão	703	29,6	18,3	12,8	9,8	8,3	6,4	5,7	4,4	4,7
G	H.P perdido	690	30,0	18,4	11,9	10,8	8,1	7,0	5,1	5,1	3,6
H	Peso molecular	1800	26,7	25,2	15,4	10,8	6,7	5,1	4,1	2,8	3,2
I	Drenagem	159	27,1	23,9	13,8	12,6	8,2	5,0	5,0	2,5	1,9
J	Peso atômico	91	47,2	18,7	5,5	4,4	6,6	4,4	3,3	4,4	5,5
K	n^{-1}, \sqrt{n}, \dots	5000	25,7	20,3	9,7	6,8	6,6	6,8	7,2	8,0	8,9
L	Designer	560	26,8	14,8	14,3	7,5	8,3	8,4	7,0	7,3	5,6
M	<i>Readers' Digest</i>	308	33,4	18,5	12,4	7,5	7,1	6,5	5,5	4,9	4,2
N	Custo	741	32,4	18,8	10,1	10,1	9,8	5,5	4,7	5,5	3,1
O	Raio X	707	27,9	17,5	14,4	9,0	8,1	7,4	5,1	5,8	4,8
P	Liga Americana	1458	32,7	17,6	12,6	9,8	7,4	6,4	4,9	5,6	3,0
Q	Radiação	1165	31,0	17,3	14,1	8,7	6,6	7,0	5,2	4,7	5,4
R	Endereço	342	28,9	19,2	12,6	8,8	8,5	6,4	5,6	5,0	5,0
S	$n^1, n^2, \dots, n!$	900	25,3	16,0	12,0	10,0	8,5	8,8	6,8	7,1	5,5
T	Taxa de mortalidade	418	27,0	18,6	15,7	9,4	6,7	6,5	7,2	4,8	4,1
	Média	1011	30,6	18,5	12,4	9,4	8,0	6,4	5,1	4,9	4,7
	Erro		±0,8	±0,4	±0,3	±0,2	±0,2	±0,2	±0,2	±0,2	±0,3

Fonte: BENFORD, 1938, p. 3.

Através da Tabela 2, Benford (1938) concluiu que a frequência dos números com o primeiro algarismo “a”, segundo algarismo “b” e os dois primeiros algarismos “ab” podem ser expressas pelas Equações 1, 2 e 3 respectivamente:

$$F(a) = \log\left(\frac{a+1}{a}\right) \quad a \in \{1,2, \dots,9\} \quad (1)$$

$$F(b) = \sum_{a=1}^9 \log\left(1 + \frac{1}{ab}\right) \quad b \in \{0,1, \dots,9\} \quad (2)$$

$$F(ab) = \log\left(1 + \frac{1}{ab}\right) \quad ab \in \{10,11, \dots,99\} \quad (3)$$

Hill (1995a) demonstrou a fórmula geral para a probabilidade de um determinado dígito ocorrer devido à lei Newcomb-Benford é calculada através da Equação 4.

$$Prob(D_1 = d_1, \dots, D_k = d_k) = \log\left[1 + \left(\frac{1}{\sum_{i=1}^k d_i \times 10^{k-i}}\right)\right] \quad (4)$$

Com as Equações 1 e 2, que depois seriam conhecidas como lei Newcomb-Benford, Benford (1938) chegou à frequência esperada dos dígitos da primeira e segunda posição, explicitadas na Tabela 3, e com a Equação 3 chegou a frequência esperada dos dois primeiros dígitos, explicitada na Tabela 4.

Tabela 3 – Proporção esperada pela lei Newcomb-Benford

Dígito	Primeira posição	Segunda posição
0	0,000	0,120
1	0,301	0,114
2	0,176	0,108
3	0,125	0,104
4	0,097	0,100
5	0,079	0,097
6	0,067	0,093
7	0,058	0,090
8	0,051	0,088
9	0,046	0,085

Fonte: BENFORD, 1938, p. 6.

Tabela 4 – Proporção esperada para os dois primeiros dígitos

10	4,14	40	1,07	70	0,62
11	3,78	41	1,05	71	0,61
12	3,48	42	1,02	72	0,6
13	3,22	43	1	73	0,59
14	3	44	0,98	74	0,58
15	2,8	45	0,95	75	0,58
16	2,63	46	0,93	76	0,57
17	2,48	47	0,91	77	0,56
18	2,35	48	0,9	78	0,55
19	2,23	49	0,88	79	0,55
20	2,12	50	0,86	80	0,54
21	2,02	51	0,84	81	0,53
22	1,93	52	0,83	82	0,53
23	1,85	53	0,81	83	0,52
24	1,77	54	0,80	84	0,51
25	1,70	55	0,78	85	0,51
26	1,64	56	0,77	86	0,50
27	1,58	57	0,76	87	0,50
28	1,52	58	0,74	88	0,49
29	1,47	59	0,73	89	0,49
30	1,42	60	0,72	90	0,48
31	1,38	61	0,71	91	0,47
32	1,34	62	0,69	92	0,47
33	1,30	63	0,68	93	0,46
34	1,26	64	0,67	94	0,46
35	1,22	65	0,66	95	0,45
36	1,19	66	0,65	96	0,45
37	1,16	67	0,64	97	0,45
38	1,13	68	0,63	98	0,44
39	1,10	69	0,62	99	0,44

Fonte: O autor, 2021.

Diaconis e Freeman (1979) criticaram os dados da Tabela 2 presentes no artigo de Benford (1938), alertando que a soma das porcentagens dos dígitos de todos os grupos, ou seja, de cada linha tem como resultado 100%. Eles afirmam que, levando em consideração o *broken-stick model*, a chance de isso ocorrer é muito pequena. Neste modelo, a probabilidade de uma linha somar 100% é igual a $(6/8\pi)^{1/2}$, ou seja, a probabilidade das 20 linhas da Tabela 2 somarem 100% é de $(6/8\pi)^{10}$, aproximadamente 0,00006%. Eles também verificaram que no grupo “A” da Tabela 2, onde foram coletados 335 dados de rios, caso dezoito deles iniciassem com o algarismo “7” a porcentagem observada seria de 5,4% ($18/335 = 0,0537$) e se fossem dezenove dados a porcentagem seria de 5,7% ($19/335 = 0,0567$). O valor presente na Tabela

2, entretanto, é 5,5%, que é um valor improvável para tal situação. Outro ponto por eles salientado foi que a média das observações que se iniciam com os algarismos “3” e “9” na Tabela 2 estavam presentes, respectivamente, com os valores 12,4 e 4,7, quando na verdade seriam 12,26 e 4,775. Pelos motivos expostos, eles concluíram que, em todos os casos, o arredondamento foi utilizado para aproximar os valores da proporção esperada pela lei Newcomb-Benford, existindo, desta maneira, uma suspeita de que os dados da tabela foram manipulados para possibilitar que todas as linhas somassem 100%.

Os primeiros artigos sobre a lei Newcomb-Benford foram publicados por Goudsmit e Furry (1944) e Furry e Hurwitz (1945).

Goudsmit e Furry (1944) sugeriram os dados da Tabela 2 e só seguiram a LNB devido ao modo que eles foram distribuídos na tabela. Através da função $\log_A b/a$, sendo a e b dois números consecutivos, eles criaram uma série de números para A igual a 10 ou 100, e realizaram o teste de conformidade com a lei. Para “A” igual a 10, eles observaram que sua amostra seguiu a LNB, e para “A” igual a 100, viram que não houve apresentação de conformidade, concluindo, assim, que a lei Newcomb-Benford só é válida para “A” pequenos.

Furry e Hurwitz (1945) discutiram, apenas, a matemática por trás das fórmulas de Benford (1938).

O terceiro artigo a tratar do assunto foi publicado por Hsu (1948). Ele solicitou a 1044 estudantes universitários que inventassem números de quatro dígitos. A única regra era que estes números deveriam ser originais e não poderiam representar um evento ou data. Após a coleta de todos os números, ele percebeu que os mesmos não seguiam a LNB e concluiu que a lei não é válida para números inventados mentalmente. Nigrini (2020) considera este estudo importante, pois indica que a não conformidade com a lei Newcomb-Benford nem sempre indica fraude ou erro, em alguns casos os dados apenas não seguem a lei.

Quatro décadas depois, Hill (1988), por sua vez, fez um trabalho nos mesmos moldes de Hsu (1948). Ele explorou a possível conexão entre o fenômeno dos primeiros dígitos, descoberto por Benford (1938), e a distribuição dos primeiros dígitos de números gerados, aleatoriamente, por pessoas. Ao solicitar que 742 de estudantes de graduação em matemática escrevessem um número aleatório de seis dígitos e analisou os dois primeiros dígitos destes números, com o teste qui quadrado (χ^2) e o teste de Kolmogorov-Smirnov (K-S). Ele chegou à conclusão de que os números em questão não seguiam a LNB, seguindo neste caso uma distribuição uniforme, mesma conclusão constatada por Hsu (1948).

Outro artigo que tratou de escolha de números aleatórios foi o elaborado por Mosimann, Wiseman e Edelman (1995). Eles solicitaram que pessoas produzissem,

aleatoriamente, números. Os indivíduos foram orientados para tentarem produzir dígitos aleatórios em três locais, a fim de fabricar uma série de três números. Foi observado que as pessoas não selecionaram todos os dígitos com a mesma frequência. Dos 8.280 dígitos escolhidos neste estudo, os dígitos mais escolhidos foram “1”, “2”, “3”, “6”, “4”, “9”, “7”, “0”, “8” e “5”, nesta ordem. O artigo concluiu que, mesmo que uma pessoa estivesse ciente de que os dígitos dos números gerados seriam examinados quanto à uniformidade, o sucesso na construção de distribuições uniformes não é garantido.

Na mesma linha, Morais (2013) aplicou a lei Newcomb-Benford para detecção da inclusão de números inventados em uma base de dados hipotética, que seguia a lei. Ele solicitou a cinco grupos de cinco estudantes de ciências contábeis da Universidade de Brasília escrevessem aleatoriamente 2, 10, 19, 38 e 47 números, o que correspondia respectivamente a 1%, 5%, 10%, 20% e 25% da amostra. Dentre os 25 alunos entrevistados, apenas um conhecia a LNB, demonstrando, assim, a falta de conhecimento da lei no meio contábil. Os testes qui quadrado e Z apenas obtiveram sucesso em detectar os dados inseridos na amostra, quando a quantidade de dados inseridos foi de 20% e 25%. Assim, o experimento concluiu que a aplicação da lei, em conjunto com testes estatísticos, é uma ferramenta útil para determinação de amostra de auditoria, porém não é uma técnica infalível.

Uma propriedade da lei Newcomb-Benford foi provada por Pinkham (1961), quando ele demonstrou matematicamente a propriedade de invariância de escala da lei pela multiplicação. Esta propriedade significa que se multiplicarmos toda uma amostra por uma constante diferente de zero a nova amostra também seguirá a LNB. Esta propriedade é útil, pois é possível, por exemplo, medir áreas das ilhas em metros quadrados ou milhas quadradas ou até valores financeiros em dólar, euro ou real, que, em todos os casos, os dados continuarão a seguir a lei Newcomb-Benford em qualquer unidade de medida.

Um artigo que explicita a propriedade de invariância de escala é o de Burke e Kincanon (1991). Eles escolheram 20 constantes da física (o que eles mesmos consideraram uma amostra pequena) para testar sua conformidade com a lei Newcomb-Benford. Esse teste foi realizado com as constantes no sistema métrico americano e no sistema métrico mks (também utilizado no Brasil). Em ambos os sistemas, os dados seguiram a distribuição preconizada na LNB, conforme preconizado pela propriedade de invariância de escala demonstrada por Pinkham (1961).

Outra propriedade da LNB foi manifestada no trabalho de Hill (1995). Ele mostrou que ao retirar amostras aleatórias de populações aleatórias escolhidas sem nenhum viés, os dígitos resultantes da amostra iriam convergir para a distribuição prevista por Newcomb

(1881) e Benford (1938). Para Nigrini (2020), esta propriedade ajuda a explicar a utilização da lei Newcomb-Benford nas áreas de computação, modelagem matemática e detecção de fraudes contábeis.

A primeira pessoa a, formalmente, utilizar a lei Newcomb-Benford foi Good (1965). Naquela ocasião, ele montou uma tabela com números aleatórios, formados pelos três dígitos do meio das áreas das paróquias inglesas. Ele percebeu que isso não produz uma tabela de números aleatórios, pois a LNB afirma que os dígitos não são igualmente distribuídos. Assim, ele concluiu que os números escolhidos na sua tabela tinham um viés para os dígitos menos significativos, o que ratifica a descoberta de Newcomb (1881) e Benford (1938).

Em 1966, Flehinger (1966) desenvolveu uma prova matemática para a lei. Esta demonstração matemática, porém, foi duramente criticada. Nigrini (2020) salientou que as críticas foram motivadas pelo fato de terem sido usados métodos especiais de somatória e de média, e, também, argumentos matemáticos, que usavam situações particulares e resultavam na lei Newcomb-Benford, não constituindo, assim, uma prova matemática robusta.

No final dos anos 60, muitos trabalhos discutindo o tema foram publicados na Ásia. O primeiro foi realizado por Adhikari e Sarkar (1968) e demonstrava que em uma distribuição uniforme, na qual os números variavam entre 0 e 1, eles convergiam, após algumas operações matemáticas, para a distribuição prevista pela LNB.

Raimi (1969) provou, matematicamente, a ocorrência da lei Newcomb-Benford, através da utilização do teorema de Banach.

Raimi (1969a) fez uma revisão, não matemática, sobre a lei Newcomb-Benford, com explicações intuitivas, que, depois, passaram a ser chamadas, em diversos artigos, de “O fenômeno do primeiro dígito”. Para Nigrini (2020), foi neste artigo, pela primeira vez, que o tema circulou pelo mundo, em uma revista científica de respeito.

Nos anos 70, vários artigos utilizaram a lei Newcomb-Benford para avaliar a sequência de Fibonacci e a sequência de Luccas. O primeiro artigo a tratar destas séries foi o de Wlodarski (1971), que mostrou que os primeiros dígitos dos 100 primeiros números da sequência de Fibonacci e da sequência de Luccas seguem a LNB. Sentance (1973) propôs o mesmo estudo, analisando, porém, os 1000 primeiros números de ambas as sequências e concluiu que os primeiros dígitos das sequências estavam em conformidade com a lei. Brandy (1978) ampliou, ainda mais, as pesquisas anteriores, analisando os 2000 primeiros dígitos de ambas as sequências e, novamente, os primeiros dígitos seguiram o preconizado por Newcomb (1881) e Benford (1938).

Ainda no início da década de 70, alguns pesquisadores se perguntavam se a lei Newcomb-Benford poderia ser utilizada para um problema prático. Varian (1972), então, sugeriu que a lei poderia ser empregada como forma de avaliar a confiabilidade de previsão de dados em um contexto das ciências sociais. Ele tabelou dados demográficos e verificou que os mesmos estavam em conformidade com a LNB. Após isso, Varian (1972) fez previsões demográficas com os dados iniciais e estas estimativas também seguiram a lei Newcomb-Benford. Assim, concluiu que a lei poderia ser utilizada como um teste da razoabilidade de previsão. Ele afirmou, ainda, que a conformidade dos dados analisados com a LNB não é certeza que os dados não tenham sido manipulados. No entanto, caso os dados não seguirem a LNB é suficiente para levantar suspeita que os dados possam ter sofrido alguma espécie de manipulação.

Outro artigo que ratifica a ideia de Varian (1972) é o de Nigrini e Wood (1996), no qual analisaram as frequências dos dígitos do censo dos 3.141 condados dos Estados Unidos, no ano de 1990. O resultado da análise de dados foi que os valores estavam em conformidade com a lei Newcomb-Benford, em consonância com o resultado do teste qui quadrado. Os autores analisaram, também, as projeções oficiais da população para 1991 e 1992. As previsões da população do condado de 1991 e 1992 tinham, de maneira análoga, uma conformidade com a LNB, conformidade esta obtida mediante o teste qui quadrado.

Na área da computação, Tsao (1974) utilizou a lei para estudar os erros de arredondamento e a distribuição cumulativa de um número de ponto flutuante, concluindo que a distribuição dos primeiros dígitos significativos de um inteiro digital está em conformidade com a lei.

Raimi (1976) fez uma revisão de literatura com os 37 artigos mais relevantes sobre o tema e 15 outros que só fizeram um breve comentário. Ele fez críticas a alguns trabalhos, os chamando de “um pouco filosóficos” devido à abordagem utilizada para provar, matematicamente, a lei Newcomb-Benford. Sua crítica foi motivada pelo fato de que métodos que têm como resultado as mesmas probabilidades encontradas por Newcomb (1881) e Benford (1938) não provam, necessária e matematicamente, a lei. Neste artigo, ele, adicionalmente, enalteceu a ideia de Varian (1972) em usar a lei como teste de validação de dados científicos e frisou que cientistas sociais necessitam de todas as ferramentas disponíveis para descobrir fraudes em dados, aplicando, inclusive, o teste para os primeiros dígitos do número.

Nos anos 80, trabalhos que utilizaram a lei Newcomb-Benford para algum uso prático começaram a ser publicados.

Becker (1982) observou a parte decimal das taxas de falha e os tempos médios de falha de componentes eletrônicos seguiam, com relação aos primeiros dígitos, a lei Newcomb-Benford. Esta descoberta o fez sugerir a aplicação da lei como *checklist* destes dois atributos em erros sistêmicos.

Nelson (1984), por sua vez, avaliou a perda de precisão devido ao arredondamento para dois dígitos significativos, utilizando a lei Newcomb-Benford para calcular a perda máxima média de precisão.

A primeira aplicação contábil para a lei Newcomb-Benford foi feita por Carslaw (1988), quando avaliou os primeiros e segundos dígitos do lucro de 220 empresas listadas na bolsa de valores da Nova Zelândia, entre o dia primeiro de janeiro de 1981 e 31 de dezembro de 1985. Concluiu que, para um nível de significância de 5%, os primeiros dígitos da amostra estavam em conformidade com a lei, porém, quando analisados, os segundos dígitos da amostra eles não estavam em conformidade. Já o teste Z, para o segundo dígito, acusou desvios significativos para os algarismos “0” e “9”, mostrando que a amostra tinha poucos números com o segundo algarismo “9” e muitos com o segundo algarismo “0”, apontando indícios de que a amostra possuía arredondamentos para cima, como por exemplo: números como 890 milhões e 29,9 mil estariam arredondados, respectivamente, para 900 milhões e 30 mil.

1.2.2 Estudos anteriores que utilizaram a LNB para detecção de fraudes

Nesta seção foram abordados alguns estudos que utilizaram a lei Newcomb-Benford para identificar viés de fraude em áreas atinentes à contabilidade.

Thomas (1989) fez um trabalho nos mesmos moldes de Carslaw (1988), porém avaliou os segundos dígitos dos lucros (valores positivos) e prejuízos (valores negativos) auferidos por firmas dos Estados Unidos, listadas na COMPUSTAT. Os dados analisados totalizaram 80.000 resultados de empresas, sendo 69.000 com lucros e 11.000 com prejuízos. Foram observados mais zeros e menos nove do que o esperado nos algarismos no segundo dígito dos lucros avaliados. Por outro lado, quando verificado o segundo dígitos dos prejuízos, as empresas apresentaram padrões opostos (menos zeros e mais noves), levando à constatação de que as empresas dos Estados Unidos, assim como as da Nova Zelândia avaliadas pro Carslaw (1988), tendem a arredondar para cima seus lucros, para melhorar os números do

balanço alcançando assim, melhor *valuation*. Quando os dados eram referentes a prejuízos, entretanto, não existia este arredondamento para cima e sim uma subestimação do valor. Também foi observado que, por ocasião do exame do lucro por ação quanto aos segundos dígitos, um comportamento de arredondamento bem mais forte nestes valores do que na análise dos lucros.

Em ocasião posterior, Craig (1992) analisou o lucro por ação de maneira semelhante a Thomas (1989). Ele teve como objetivo verificar se os gerentes, de forma geral, arredondam os cálculos de lucro por ação de maneira tendenciosa. No trabalho foram utilizados dados financeiros de um período de 20 anos, de aproximadamente 2500 empresas listadas na Bolsa de Valores de Nova York ou negociadas no mercado de balcão, e verificada a conformidade dos dados com a lei Newcomb-Benford. Como resultado, o estudo obteve evidência de que os gerentes de empresas estatais têm maior probabilidade de arredondar o último decimal do lucro por ação pra cima do que para baixo. Um fator percebido para explicar este arredondamento foi que os gestores recebiam benefícios pautados no lucro por ação da empresa, sendo, portanto, a razão principal de inflarem este indicador.

Das e Zhang (2003) trataram, também, do tema lucro por ação, juntamente com a lei Newcomb-Benford. Os dados utilizados no trabalho deles foram obtidos no banco de dados da *Standard & Poors' Quarterly Compustat* e abrangem o período dos exercícios fiscais de 1989 a 1998, totalizando 103.944 observações trimestrais. O número de observações varia de 6.663 para o ano 1989 a 16.447 para o ano de 1997. O artigo investigou o dígito imediatamente à direita do decimal do número calculado de lucro por ação, expresso em centavos, com o auxílio da lei Newcomb-Benford. Como resultado, eles concluem que 54,5% dos balanços trimestrais das empresas arredondam seus lucros por ação. Eles verificam, também, que os gerentes realizam o arredondamento por três principais motivos: relatar lucros positivos, sustentar o desempenho recente e atender às previsões dos analistas.

Outro trabalho que teve como tema arredondamento de dados contábeis foi o de Nigrini (2005), que analisou os relatórios de ganhos divulgados em 2001 e 2002 pela empresa Enron. Ele usou a LNB como referência e verificou que os números das receitas trimestrais mostravam um excesso de zeros como segundo dígito, sugerindo que os números da receita eram arredondados para cima. A incidência de zeros como segundo dígito dos números de receita aumentou em 2002 (13% contra 12,8% em 2001). Ao analisar o lucro por ação, Nigrini (2005) notou que os dados mostravam uma distribuição acentuada em torno de zero, mostrando, assim, uma tendência a números redondos de lucros por ação, fato este que já tinha sido relatado por Thomas (1989).

Nigrini (1994) utilizou a LNB para detectar fraudes na folha de pagamento, comparando as frequências dos dois primeiros dígitos da amostra com a frequência esperada para os dois primeiros dígitos prevista por Benford (1931). Foram avaliados os dados por um período de dez anos e percebido um desvio significativo em relação à lei Newcomb-Benford. Os desvios se tornavam maiores nos últimos cinco anos estudados e eram ocasionados por números usados repetidamente. Com isso, Nigrini concluiu que o fraudador seguiu uma rotina nos últimos cinco anos, e nem se deu o trabalho de inventar números, apenas repetindo os já utilizados. Neste período, foi detectado um montante de US\$ 500.000,00 em fraudes.

Nigrini (1999) explicita que a lei Newcomb-Benford é uma ferramenta que pode ter aplicação na auditoria, visto que no caso de Wayne James Nelson, um gerente no escritório do Tesouro do Estado do Arizona que fraudou o Estado em, aproximadamente, dois milhões de dólares, o golpe seria detectado caso os cheques fossem auditados tomando como base a lei Newcomb-Benford. Alguns sinais de fraude, que a LNB conseguiria detectar, são:

- Wayne começou pequeno e, em seguida, aumentou os montantes em dólares.
- A maioria dos montantes foi pouco abaixo de US \$100.000, ou seja, manter os valores abaixo de um limite de controle adicional.
- Os números parecem ter sido escolhidos para dar a aparência de aleatoriedade. Nenhum dos valores de cheque foi duplicado, não havia números redondos e todos os valores incluíam centavos. O gerente, no entanto, repetiu alguns dígitos e combinações de dígitos e caracterizou uma tendência para os dígitos mais altos. As contagens para os dez dígitos de “0” a “9” foram 7, 19, 16, 14, 12, 5, 17, 22, 22 e 26, respectivamente, notando-se, então, que “7” a “9” foram os dígitos mais utilizados, em contraste com a LNB.

Silva, Korzenowski e Vaccaro (2014) analisaram duas ferramentas estatísticas e seu comportamento no reconhecimento de amostras que contivessem indícios de lavagem de dinheiro, tendo como base a lei Newcomb-Benford. Foram utilizados os testes qui quadrado e Kolmogorov-Smirnov. Os dados avaliados foram os lançamentos de oito contas corrente, sendo que apenas em três havia suspeita de lavagem de dinheiro, conforme critério estabelecido pelo Banco Central de Brasil. Eles concluíram que para amostras pequenas o teste qui quadrado é menos sensível que o K-S, pois não rejeita a hipótese nula se a amostra apresentar um leve desvio da LNB. No caso da amostra ser grande, o comportamento de ambos é semelhante.

Nigrini e Mittermaier (1997) estudaram ferramentas de análise de dados em conjunto com a lei Newcomb-Benford e sua utilização na contabilidade. As ferramentas de análise de

dados discutidas no artigo foram a análise dos primeiros dígitos, dos segundos dígitos, dos dois primeiros dígitos, dos últimos dois dígitos e da duplicação de números. Eles utilizaram o teste Z e o desvio absoluto médio (*MAD-Mean Absolute Deviation*) em uma companhia de petróleo listada na *New York Stock Exchange* e em uma companhia do setor elétrico, analisando respectivamente 72 mil e 30 mil faturas. Ambas as amostras seguiram a LNB para o primeiro, segundo e dois primeiros dígitos. Eles concluíram que os auditores externos e internos podem utilizar a LNB juntamente com estes procedimentos (teste Z e MAD) para testar a existência de fraude em dados, e determinar as amostras de auditoria.

Milani Filho (2013) estudou os dados financeiros de hospitais filantrópicos canadenses, localizados nas províncias de Ontário e Quebec, e sua conformidade com a lei Newcomb-Benford. Os dados financeiros explorados foram as receitas e despesas dos hospitais, que resultaram em 1334 observações extraídas entre os anos de 2009 e 2010. Para a análise de dados, foram utilizados o teste Z e o qui quadrado, ambos com um nível de significância de 5%. O trabalho concluiu que as receitas e despesas dos hospitais filantrópicos das cidades estudadas seguiram a LNB.

Cella e Rech (2017) verificaram a aplicabilidade da lei Newcomb-Benford para detectar fraudes nos dados financeiros da Petrobrás. Foram analisados os primeiros dígitos de diversas contas contábeis entre o primeiro trimestre de 2004 e o segundo trimestre de 2012. As contas analisadas foram Fornecedores, Imobilizado, Estoques, Receita, Custo dos Produtos Vendidos, Caixa e Lucro antes dos Impostos, que totalizaram 238 observações. Para haver comparação de desvios das empresas com relação à lei, foi feita a mesma análise em outras 40 empresas listadas na bolsa de valores no mesmo período. Na análise de conformidade, através do teste qui quadrado, a amostra contendo os dados da Petrobras e as amostras contendo todos os dados das outras empresas não apresentaram conformidade com a lei. Assim, o artigo concluiu que a adulteração dos dados contábeis da Petrobras não teria sido previamente descoberta com sucesso, apenas com base na utilização da LNB, para a análise dos primeiros dígitos das contas verificadas. Pois, se nenhuma empresa analisada apresentou conformidade com a LNB, a não conformidade da Petrobrás não levantaria nenhuma espécie de suspeita de fraude.

Oliveira *et al.* (2018), da mesma forma que Cella e Rech (2017), estudaram o caso Petrobras, aplicando a lei Newcomb-Benford do primeiro ao quarto dígito dos saldos dos balanços patrimoniais da empresa, entre o primeiro trimestre de 2008 e o segundo trimestre de 2015. Quando realizado o teste qui quadrado nos 5.544 dados, eles não seguiram a LNB em testes realizados do primeiro ao quarto dígito. Assim, eles concluíram que os resultados

encontrados evidenciavam a necessidade de uma auditoria mais rigorosa nestes dados que, posteriormente, foram revistos pela operação Lava Jato.

Ribeiro e Monsueto (2015) estudaram os pagamentos realizados por uma empresa da cidade de Goiânia, no período de setembro de 2010 até setembro de 2011. A análise de conformidade com a lei Newcomb-Benford para os primeiros dígitos, utilizando o teste qui quadrado, revelou impropriedades nos meses de setembro, novembro e dezembro de 2010, e nos meses de janeiro, maio, junho, julho e agosto de 2011. Já o teste Z revelou inconformidade nos dígitos “3”, “5” e “6”. Eles verificaram que a impropriedade no dígito “3” ocorreu devido ao pagamento de taxas do Conselho Regional de Engenharia e Arquitetura no valor de R\$ 33,00. O desvio significativo do dígito “5” foi explicado pela junção de diversos pagamentos em um único lançamento. Já com relação ao dígito “6”, houve excesso na observação devido às cobranças de taxas bancárias serem no valor fixo de R\$ 6,50. Assim ficou constatada, a utilidade da LNB para controle de dados financeiros.

Forster (2006) analisou, com o auxílio de testes estatísticos, as contas de 159 instituições sem fins lucrativos do Distrito Federal, entre os anos de 2002 e 2003, e sua conformidade com a lei Newcomb-Benford. As contas analisadas foram: Caixa, Banco, Despesa e Receita. Após a aplicação dos testes Z e qui quadrado para analisar os primeiros dígitos das contas, o autor concluiu que apenas a conta da despesa no ano de 2002 não estava em conformidade com a LNB, pois existiam desvios significativos nos dígitos “5”, “6” e “9”, apontando que essa conta necessitava de uma análise *in loco* da Promotoria de Justiça de Tutela de Fundações e Entidades de Interesse, órgão do Ministério Público do Distrito Federal e Territórios responsável pela análise dos demonstrativos contábeis do terceiro setor.

Drake e Nigrini (2000) analisaram 38.176 faturas de uma empresa, utilizando como processo para mensurar o nível de conformidade dos primeiros, segundos e dois primeiros dígitos com a LNB o desvio absoluto médio (MAD). Este processo obteve como resultado para o primeiro, segundo e dois primeiros dígitos os valores de 0,0076, 0,0078 e 0,001236, respectivamente. Concluindo, assim, que as faturas estavam em conformidade com a lei Newcomb-Benford nas três ocasiões estudadas, o que indicava que os dados não foram manipulados.

Özer e Babacan (2013) verificaram a conformidade com a lei Newcomb-Benford das transações realizadas pelos bancos turcos, entre 1990 e 2010, utilizando o teste qui quadrado. A análise de dados apontou que, apenas no ano de 1999, os bancos realizaram, nas suas contas extrapatrimoniais e nas movimentações realizadas em moeda estrangeira, transações fraudulentas. Assim, concluíram que os resultados obtidos identificaram, corretamente, as

fraudes realizadas em 1999. Por outro lado, não identificaram as fraudes realizadas no período de 1990 a 1998, período marcado por fraudes e manipulações bancárias na Turquia. Com isso, os autores consideraram a LNB uma ferramenta efetiva para detecção de fraudes, alertando, entretanto, que, mesmo se um conjunto de dados estiverem em conformidade com a lei, eles podem conter impropriedades.

Milani Filho et al. (2016) investigaram o comportamento da variável lucro antes dos juros, impostos, depreciação e amortização (*Earnings Before Interest, Taxes, Depreciation and Amortization* – Ebitda), nas empresas listadas nas seguintes bolsas de valores: BM&FBovespa (Brasil), *Moscow Exchange* (Rússia), *Bombay Stock Exchange* (Índia), *Hong Kong Stock Exchange* (China), *Shanghai Stock Exchange* (China), *Shenzhen Stock Exchange* (China) e *Johannesburg Stock Exchange Limited* (África do Sul). Eles observaram 2.193 empresas, que totalizaram 10.965 observações do Ebitda, no período de 2008 a 2012, com o objetivo de verificar se esta variável seguiu a distribuição prevista por Newcomb (1881) e Benford (1931) para os primeiros dígitos. Através do teste Z e qui quadrado, verificaram que todas as bolsas de valores seguiram a lei Newcomb-Benford entre os anos analisados, exceto a *Shenzhen Stock Exchange*. As empresas listadas nesta bolsa de valores não apresentaram conformidade com a LNB entre os anos de 2008 a 2011, indicando, como necessária, uma averiguação mais detalhada, para constatar os fatores que induziram tal cenário.

Santos et al. (2009) verificaram que a utilização da lei Newcomb-Benford, seria útil na auditoria tributária do imposto sobre serviço de qualquer natureza (ISS). Esta verificação foi feita através da comparação entre conclusões obtidas através da conformidade dos dados com a LNB e as obtidas pela auditoria contábil-fiscal realizada pelos auditores fiscais municipais, entre os anos de 2002 e 2004. Foram analisadas 1.958 notas fiscais de uma empresa de publicidade e foi observado um desvio positivo no primeiro dígito "2" e desvios negativos nos primeiros dígitos "7" e "8". A auditoria confirmou irregularidades nas notas fiscais que estavam nas amostras que não apresentaram conformidade com a LNB, concluindo assim a utilização desta lei foi útil na auditoria tributária.

Também utilizando a LNB na auditoria tributário do ISS, Lagioia et al. (2011) analisaram quatro empresas, que emitiram em conjunto 6.511 notas fiscais. Estas empresas pertenciam aos segmentos de contabilidade e auditoria, turismo, publicidade e laboratórios de análises clínicas. O trabalho comparou o resultado da conformidade da distribuição das notas fiscais com lei com os resultados do relatório dos auditores fiscais municipais. Concluíram que os resultados da fiscalização dos auditores confirmaram os indícios levantados pela não conformidade com a lei Newcomb-Benford. Deste modo, os autores salientaram que a

aderência à lei se torna uma ferramenta para o auditor na fase de planejamento da auditoria, dando preferência na escolha para amostra de auditoria as notas fiscais que não estão em conformidade com a lei.

Santos, Tenório e Silva (2003) analisaram 8.000 notas fiscais de venda emitidas, entre 1998 e 2001, que foram verificadas pela Secretaria da Fazenda de Pernambuco. Após a análise das amostras dos quatro anos, foi percebido que as notas fiscais emitidas em 1998 e 1999 seguiam a LNB. Já as emitidas em 2000 e 2001 não seguiam. Concluiu-se, assim, que estes dois últimos anos tem que receber maior atenção da auditoria, para que possa ser respondido o porquê do seu desvio com relação à lei.

Almeida (2011) avaliou a conformidade do primeiro ao quarto dígito dos dados do Produto Interno Bruto (PIB) dos municípios da Região Norte, entre 2004 a 2008, o consumo de energia elétrica por município do Amapá, entre 1997 e 2001, e o demonstrativo de arrecadação mensal de tributos do Estado de Macapá, no período de 1999 a 2010, com relação à LNB. No resultado obtido, mediante o teste qui quadrado, foi percebido que o PIB seguia a lei para os quatro dígitos avaliados, o consumo de energia seguiu a distribuição apenas para os primeiros dígitos e a arrecadação dos tributos aderiu à lei para os segundos, terceiros e quartos dígitos. A autora concluiu, portanto, que a conformidade do PIB e da arrecadação dos tributos com a lei dá confiabilidade aos dados. A não aderência do consumo de energia sugere a aplicação de técnicas de arredondamento na leitura dos dados.

Mir, Ausloos e Cerqueti (2014) analisaram, primeiramente, a conformidade, com a lei Newcomb-Benford, dos dados da receita tributária de alguns municípios italianos, entre os anos de 2007 e 2011, com o objetivo de encontrar algum indício de evasão fiscal. Estes dados foram obtidos no Centro de Pesquisa do Ministério das Finanças e Economia da Itália e, com a utilização do teste qui quadrado, sua conformidade com a lei foi testada para os primeiros dígitos. Excetuando-se os anos de 2008 e 2009, quando o material analisado apresentou grandes desvios da LNB, os autores constataram que os dados fiscais de todas as cidades italianas estavam em conformidade com a lei.

Após estes procedimentos, os mesmos autores examinaram as receitas tributárias das regiões localizadas ao sul da Itália (Calabria, Sicília e Campania). A escolha destas regiões se deu pelo motivo das mesmas serem conhecidas como zona de mercado negro. Concluíram que apenas a região da Campania, nos anos de 2008 e 2009, não demonstrava conformidade com a LNB, o que está em linha com o fato de esta região ser bastante conhecida por ter um baixo nível de governança, uma baixa prestação de serviços e uma presença substancial da máfia e do crime organizado, o que indica uma alta correlação com a evasão fiscal.

Adicionalmente, o trabalho alertou que a análise de todas as regiões juntas apontou evasão fiscal nos anos de 2008 e 2009, enquanto que a análise das regiões do sul, individualmente, demonstrou indícios de evasão fiscal, nos anos de 2007 e 2008, na região da Campânia. Isso alerta para o fato de que um tipo de análise pode indicar uma amostra como fraudulenta, enquanto outro tipo de análise pode indicar outra amostra como ilegítima.

Ainda verificando a receita tributária das regiões italiana, Ausloos, Cerqueti e Mir (2017) analisaram a receita de todas as 20 regiões da Itália, entre os anos de 2007 e 2011, e sua conformidade com a lei Newcomb-Benford. Após utilizar o teste qui quadrado para os primeiros dígitos, os autores verificaram que os dados da receita tributária das seguintes três regiões não seguiam a LNB: Campânia, em 2007 e 2008, reafirmando Mir, Ausloos e Cerqueti (2014); Sardenha, nos anos de 2007 e 2011, e Ligúria, no ano de 2008. Assim, concluíram que as três regiões apresentavam indícios de evasão fiscal. Eles alegaram que o motivo da evasão fiscal da região de Sardenha está ligada ao fato desta não possuir uma estrutura industrial altamente desenvolvida e as suas trocas comerciais serem frequentemente baseadas em *commodities*. No caso da região de Ligúria, a causa dos indícios de evasão fiscal se deve ao fato da penetração da economia paralela. A influência do crime organizado no sistema econômico foi o motivo exposto para a Campânia figurar entre as regiões com possível evasão fiscal, o mesmo motivo apresentado por Mir, Ausloos e Cerqueti (2014).

Christian e Gupta (1993) analisaram os dados do imposto de renda dos Estados Unidos, entre os anos de 1979 a 1986, investigando sinais de evasão fiscal secundária¹. Eles analisaram os dois últimos dígitos das receitas tributáveis declaradas por cada contribuinte, que, segundo a lei Newcomb-Benford, deveriam estar uniformemente distribuídos, fato não observado na amostra estudada. Assim, concluíram que havia indícios de redução da receita tributável, com o intuito de evasão fiscal secundária.

Com relação à evasão fiscal, Nigrini (1996) aplicou a lei Newcomb-Benford para analisar este tipo de fraude e foi o primeiro a analisar uma grande quantidade de dados. Foram analisados os dados dos contribuintes dos Estados Unidos, nos anos de 1985 e 1988, o que resultou num montante de 55.000 e 91.000 dados, respectivamente. Seu estudo detectou evasão fiscal em ambos os anos, observando que os itens tributáveis foram subestimados e os dedutíveis foram superestimados. Foi verificado, também, que os contribuintes de baixa renda não apresentavam distorções significativas, quando comparados aos de alta renda.

¹ A evasão fiscal secundária ocorre quando o contribuinte, deliberadamente, reduz sua renda tributável, com a finalidade de alcançar patamares inferiores da tabela do imposto de renda.

Ainda sobre o tema evasão fiscal, Macías e Igua (2018) aplicaram a lei Newcomb-Benford para averiguar as declarações de imposto de renda de pessoa física e jurídica referentes ao ano de 2014 no Equador, com o objetivo de procurar indícios de atos de evasão fiscal ou declaração incorreta. Estas informações foram extraídas do banco de dados do Centro de Estudos Fiscais da Receita Federal. Foram analisados o formulário 101, referente às pessoas jurídicas, o formulário 102, concernente às pessoas físicas, e o formulário 107, atinente às informações dos dependentes das pessoas jurídicas, estes formulários continham, respectivamente, 116.502, 720.387 e 3.481.873 dados. Com relação à conformidade com o primeiro dígito da lei, foi observado que os dados do formulário 101 seguiram a LNB, com exceção do patrimônio líquido. O fato que ocasionou o desvio foi que este dado tinha um limitante inferior de 800 dólares, o que prejudicou a utilização da lei. Os formulários 102 e 107, entretanto, apresentaram grandes desvios com relação à lei. A justificativa foi que esta espécie de informação era fornecida pelo contribuinte, estando, por isso, mais sujeita a fraudes.

Cella e Zanolla (2018), na área de contabilidade pública, avaliaram a conformidade, das despesas empenhadas, com a lei Newcomb-Benford, durante o primeiro quadrimestre de 2016, de dois municípios de Goiás, um com alto nível de transparência e outro com baixo nível. Foram analisadas 3068 notas de empenhos, atinentes a um montante de 39 milhões de reais, sendo 1.453 notas de empenhos do município de maior transparência e 1.615 do outro município. Na análise de dados foram aplicados os testes Z e qui quadrado com nível de significância de 5% nos primeiros e segundos dígitos dos valores das notas de empenhos. A conclusão da avaliação foi que as despesas de nenhum dos dois municípios estavam em conformidade com a lei Newcomb-Benford, sendo que, entretanto, o município com menor grau de transparência obteve maiores desvios com relação à lei do que o município mais transparente.

Costa, Santos e Travassos (2012) analisaram 134.281 notas de empenhos emitidas por 20 Unidades Gestoras (UG) de dois estados brasileiros. Foram utilizados o teste Z e qui quadrado com nível de significância de 5%, para mensurar a conformidade dos primeiros e segundos dígitos com a lei Newcomb-Benford. Como resultado do teste qui quadrado, foi observada a não conformidade com a lei para os primeiros e segundos dígitos em todas as 20 UG. Com relação à análise por dígito, os autores concluíram que, quanto à avaliação do primeiro dígito, ocorreram desvios relevantes entre a proporção observada e esperada nos algarismos “7” e “8” (excesso) e nos algarismos “9” e “6” (escassez). Eles perceberam, também, que este comportamento configura uma propensão de fuga à realização dos

processos licitatórios. Ao analisar o segundo dígito, foi percebido um excesso de incidência dos algarismos “0” e “5”, o que evidenciou o emprego de arredondamentos nos valores dos empenhos.

De Oliveira et al. (2018), analisaram os gastos do Gabinete da Presidência da República do Brasil e seus 24 Ministérios, no período de janeiro de 2011 a agosto de 2015, quanto à conformidade do primeiro ao quarto dígito com a LNB. Para isto, utilizaram os testes Z e qui quadrado, ambos com nível de significância de 5%. Na análise de dados, foi apurado que para nenhum dos quatro dígitos houve conformidade com a lei e o ministério que mais obteve distorção foi o de Ciências, Tecnologia e Informação, seguido pelo da Educação. Concluíram, assim, que em razão da não conformidade apresentada os gastos estudados deveriam ser auditados pelos respectivos órgãos de fiscalização, salientando que os ministérios de Ciências, Tecnologia e Informação e o da Educação exibiram relevantes dificuldades de gerenciamento durante o período estudado.

Macieira *et. al.*(2019) investigaram a conformidade da distribuição dos dois primeiros dígitos das notas de empenhos das Prefeituras Municipais do Estado de São Paulo, com a lei Newcomb-Benford. Após a verificação de conformidade, verificaram se grandes desvios da LNB tinha correlação com os pareceres desfavoráveis emitidos de 2012 a 2014. Para análise global dos primeiros e segundos dígitos da amostra de cada prefeitura foram utilizados os testes qui quadrado e o desvio absoluto médio (MAD). Foi percebida uma não conformidade em 68% na análise dos primeiros e segundo dígitos das notas de empenhos, já com relação à análise de contas, 35% das prefeituras municipais tiveram suas contas rejeitadas. Os autores do estudo concluíram que a utilização do MAD foi o mais apropriado para verificação de conformidade, pois não considera o tamanho da amostra. Eles também constataram que existe baixa correlação entre a não conformidade com a LNB e o parecer desfavorável do Tribunal de Contas do Estado de São Paulo.

Costa (2012) analisou 335.830 notas de empenhos emitidas em 2010 por 60 unidades gestoras distribuídas por três estados brasileiros. Foram realizados quatro testes estatísticos (Teste de Invariância Escalar, Semidesvio Absoluto, Discrepância Relativa e Fator de Detecção de Ruído) para verificar a conformidade das NE de cada UG com a LNB, totalizando 240 testes estatísticos. De todos os testes realizados em apenas cinco ocasiões foram verificadas conformidades com a lei Newcomb-Benford. O excesso de não conformidade levantou indícios de erros, manipulação de dados e fracionamento de despesas. O autor, porém, alertou que são apenas indícios e que seria necessária a análise qualitativa de um auditor para confirmar, realmente, a existência de fraudes.

Com o intuito de reconhecer desvios no tempo, Silva, Travassos e Costa (2017) estudaram o uso da LNB em um ambiente de auditoria contínua. Para isso, eles analisaram os primeiros dígitos de 210.899 notas de empenhos emitidas, em 2010, por 60 unidades gestoras localizadas em dois estados do nordeste brasileiro. Eles realizaram uma análise gráfica, com a finalidade de acompanhar, a discrepância relativa com a lei e padrões típicos plausíveis de fraudes ou fracionamento de despesas. Assim, concluíram que uma análise temporal de compatibilidade com a lei Newcomb-Benford é um mecanismo útil na auditoria contínua.

Santos, Diniz e Corrar (2005) utilizaram a lei Newcomb-Benford para analisar os primeiros dígitos das notas de empenhos de 20 municípios do estado da Paraíba localizados nas regiões do litoral e do sertão, totalizando, assim, 104.104 notas de empenhos. Para analisar os dados, foram utilizados o teste Z e o teste qui quadrado, ambos com o nível de significância de 5%. Foi percebido um desvio significativo nas notas de empenhos que se iniciavam com os algarismos “4”, “5”, “8” e “9”. Com isso, eles concluíram que os municípios apresentavam indícios de fracionamento de despesas, devido ao excesso de primeiros dígitos “4”, “5” e “8”, e, também, sinais de superfaturamento, devido à grande quantidade de primeiros dígitos “9”.

Diniz et al.(2006) utilizaram a lei Newcomb-Benford, para o planejamento da auditoria das contas públicas, com o objetivo de mensurar a correlação dos resultados apresentados nos testes de conformidade com a lei e o produto do relatório de auditoria emitido pelos tribunais de contas. Eles utilizaram a mesma amostra de Santos, Diniz e Corrar (2005), porém, na análise de dados, utilizaram a estatística de Pearson, para comparar o valor do teste qui quadrado com a quantidade de irregularidades encontradas e o parecer do Tribunal de Contas. O índice de correlação entre os valores da estatística do teste qui quadrado e o número de irregularidades apresentadas no relatório de auditoria foi de aproximadamente 66%, já o índice de correlação entre os valores da estatística do teste qui quadrado e o parecer do Tribunal de Contas (favorável ou desfavorável) foi de aproximadamente 47%. Embora a metodologia empregada não consiga mensurar variáveis importantes, tais como uniformização dos procedimentos de auditoria, a magnitude das irregularidades expostas e certos aspectos inerentes a cada auditor, os autores, mesmo assim, concluíram que, o modelo utilizado obteve um índice de correlação satisfatório em relação à quantidade de irregularidades.

Ganassin, Costa e Wilbert (2016) analisaram 7.037 notas de empenhos do Ministério Público Militar (MPM) e do Ministério Público do Distrito Federal e Territórios (MPDFT), entre 2008 e 2012, verificando sua conformidade com a LNB. Eles analisaram as notas de

empenhos agrupadas por modalidade, ou seja, as provenientes de licitação e as sem licitação. Como resultado, foi percebido que apenas as notas de empenhos emitidas sem licitação do MPM, nos anos de 2008 e 2010, e do MPDFT, nos anos de 2009, 2010 e 2012, não seguiram a lei Newcomb-Benford. Como resultado da análise, concluíram que os desvios mais significantes aconteceram em ano eleitoral, podendo dizer que as amostras das NE que não seguiram a LNB possuem indícios de execução orçamentária atípica.

Cunha (2013) utilizou a LNB para analisar as planilhas orçamentárias de obras públicas realizadas para a Copa do Mundo de 2014, com o intuito de identificar sobrepreços. Foram examinados os dados da restauração do estádio Maracanã, da criação da Arena Amazônia e a reestruturação do Aeroporto Internacional de Confins, nos quais foram observadas inconformidades com a lei, o que levantou indícios de sobrepreços. Estes indícios foram confirmados pela auditoria do Tribunal de Contas da União, que identificou um sobrepreço de, em média, 80%. A análise realizada concluiu, portanto, que a utilização da lei Newcomb-Benford é uma forma eficiente de avaliar a existência de sobrepreços em planilhas orçamentárias de obras públicas.

Ainda sobre obras públicas, Café (2015) aplicou a lei Newcomb-Benford para analisar a planilha de custos da obra do Veículo Leve sobre Pneus, em Brasília. Foram percebidos indícios de superfaturamento devido a não conformidade com a LNB. Em auditoria realizada pelo Tribunal de Contas do Distrito Federal foi averiguado que, para cada real que a lei indicou como suspeito de fraude, 41 centavos foram confirmados sobre preço. Concluiu, então, que o uso da lei Newcomb-Benford é útil para identificar fraudes em planilhas de obras públicas.

Nascimento, Filho e Buscacio (2014) analisaram dados retirados dos sites do Instituto Nacional de Pesquisas e Estudos Educacionais (INEP) e da Secretaria da Fazenda do Estado do Rio de Janeiro (SEFAZ-RJ). Os dados do INEP eram referentes ao número total de alunos matriculados nas escolas da rede pública (municipal e estadual), em cada município brasileiro. Os dados da SEFAZ-RJ, por sua vez, continham 200.545 notas de empenhos emitidas pelo Estado do Rio de Janeiro, entre 2008 e 2012. Foi verificada a conformidade dos dados do INEP com a LNB através do teste qui quadrado para os primeiros e segundos dígitos. Os dados no INEP apresentaram conformidade, porém, os dados da SEFAZ-RJ não estavam em conformidade com a LNB e o algarismo com maior desvio era o “8”. Assim, concluíram que as notas de empenhos, por terem um grande desvio entre a frequência observada e esperada, têm indícios de existência de irregularidades e fracionamento de despesas para fugir do limite de licitação de R\$8.000,00.

Freitas Costa et al. (2012) estudaram técnicas que são aplicadas em auditorias contábeis, para verificação de conformidade dos dados com a lei Newcomb-Benford. Eles analisaram os testes qui quadrado, teste Z, teste Kolmogorov-Smirnov, modelo fator de distorção, desvio absoluto médio, discrepância relativa e desvio da média da convolução. Estas técnicas foram utilizadas, simultaneamente, para avaliar a conformidade de 210.899 notas de empenhos emitidas em 2010, com a LNB. Concluíram que elas são ferramentas úteis para o auditor, utilizando sempre que possível mais de uma delas para tirar suas conclusões. Afirmaram, também, que técnicas que avaliam a amostra como um todo devem ser utilizadas para levantar indícios necessários para tomada de decisão, já as que analisam de forma local (dígito por dígito) são indicadas para a etapa de preparação da auditoria.

Sallaberry et al. (2019) analisaram a conformidade da lei Newcomb-Benford com as informações financeiras declaradas por candidatos a cargos políticos ao Tribunal Superior Eleitoral, no ano de 2014. Estas informações totalizaram 26.252 pedidos de candidaturas e 83.050 declarações bens. Os dados não seguiram a LNB para os testes qui quadrado e Z, resultando em grandes desvios para os algarismos “3”, “4”, “5”, “6”, “7” e “9” da primeira posição. Os autores também aplicaram o fator de distorção, que sinalizou os algarismo “3” e “5” como discrepantes, e a média dos desvios absolutos, que obteve o algarismo “5” como discrepante. Com isso, concluíram que deveriam ser verificadas as declarações de bens que se iniciavam com os dígitos “3” e “5”, o que representava 22,13% do total.

Graham, Hasseldine e Paton (2009) utilizam a lei Newcomb-Benford para testar a confiabilidade dos dados das zonas de pesca no Canadá. Os dados analisados foram obtidos no Departamento de Pescas e Oceanos e faziam referência à quantidade diária pescada, em kg, por pescador na zona cinza (um ambiente de pesca altamente regulamentado), na área de pesca do caranguejo da neve (ambiente regulado) e nas áreas de pesca da lagosta 33 e 34 (ambos ambientes pouco reguladas). Como resultado obtido a um nível de significância de 1%, foi observado que os dados da zona cinzenta estavam em conformidade com a LNB. Entretanto, as observações das áreas de pesca do caranguejo da neve e das duas áreas de pesca de lagosta não estavam em conformidade com a lei. O estudo concluiu que as zonas da pesca da lagosta, por serem pouco reguladas, estavam propensas a apresentar irregularidades. Já na área da pesca do caranguejo da neve, os frequentadores criaram uma maneira de contornar a regulamentação, mostrando que a regulamentação não estava sendo efetiva, produzindo, portanto, dados sem confiabilidade.

1.3 Procedimentos não estatísticos e testes estatísticos utilizados para verificação de conformidade com a lei Newcomb-Benford

Nesta seção serão abordados os procedimentos não estatísticos abordados por Nigrini (2012) e Nigrini (2020) e também os testes estatísticos utilizados para verificação dos dados com a LNB.

1.3.1 Procedimentos não estatísticos

Nigrini (2012) e Nigrini (2020) dividem os procedimentos utilizados por eles para analisar a conformidade dos dados com a lei Newcomb-Benford em cinco tipos de análises²:

- Análises pré-Benford;
- Análises primárias;
- Análises avançadas;
- Análises associadas; e
- Análises de conformidade.

Estas análises foram abordadas nas seções subseqüentes.

1.3.1.1 Análises pré-Benford

Esta análise se inicia com uma visão geral de alto nível nos dados, a qual abrange a análise do perfil de dados, o histograma dos dados e análise periódica dos dados. A visão geral de alto nível permite ao auditor entender com que espécie de dados ele está tratando, pois, através dos testes de alto nível, ele saberá quantos dados existem e sua distribuição em relação à ordem de grandeza e ao tempo (NIGRINI, 2020).

² Embora Nigrini (2012) e Nigrini (2020) chamem seus procedimentos de teste, eles não são testes estatísticos no sentido stricto sensu. Assim está dissertação substituiu o termo “teste” pelo termo “análise” para manter o rigor estatístico e acadêmico.

As análises pré-Benford contém três procedimentos que auxiliam o auditor a entender melhor os dados. Estes procedimentos fornecem diversas informações ao auditor, tais como a quantidade de dados da amostra, a distribuição dos dados em relação ao montante e a sua organização no tempo (NIGRINI, 2012).

A primeira análise pré-Benford é denominado perfil de dados e deve ser a primeira a ser realizada, pois, conforme Nigrini (2020), neste procedimento é possível detectar sérios problemas nos dados, que impossibilitariam a continuidade da análise. Este procedimento consiste em dividir os dados em cinco categorias: grandes e positivos, pequenos e positivos, zeros, pequenos e negativos, e grandes e negativos.

O perfil de dados alerta o auditor para algumas situações, tais como integridade dados, alta proporção de valores pequenos, valores nulos, presença de *outliers* e presença de valores negativos. Nigrini (2020) frisa, também, que a uma proporção aceitável de valores pequenos positivos em uma base de dados é de até 15% do total, caso contrário a amostra pode apresentar viés de não conformidade com a LNB.

Outra análise pré-Benford proposta por Nigrini (2012) e Nigrini (2020) é o histograma de dados, também chamado de estatística descritiva. Ele mostra ao auditor a forma da distribuição dos dados. Nigrini (2020) sugere que, para melhor visualização do histograma, ele deve conter 10 intervalos. Para amostras com mais de 5000 dados, deve possuir, entretanto, entre 14 a 20 intervalos. Para dados financeiros, ele também salienta que 70% dos dados costumam estar no primeiro intervalo e 20 % no segundo intervalo. Isto é devido a esta espécie de dados tenderem a ter vários números pequenos e poucos números grandes.

O terceiro procedimento é denominado gráfico periódico, que divide os dados de acordo com decorrer do tempo, ilustrando os meses no eixo horizontal e o valor total dos dados no eixo vertical. Para Nigrini (2020), este procedimento é útil para o auditor entender melhor seus dados e também detectar grandes anomalias. Através do gráfico periódico, o auditor é capaz de notar gastos mensais fora do normal, o que pode sinalizar alguma espécie de fraude.

1.3.1.2 Análises primárias

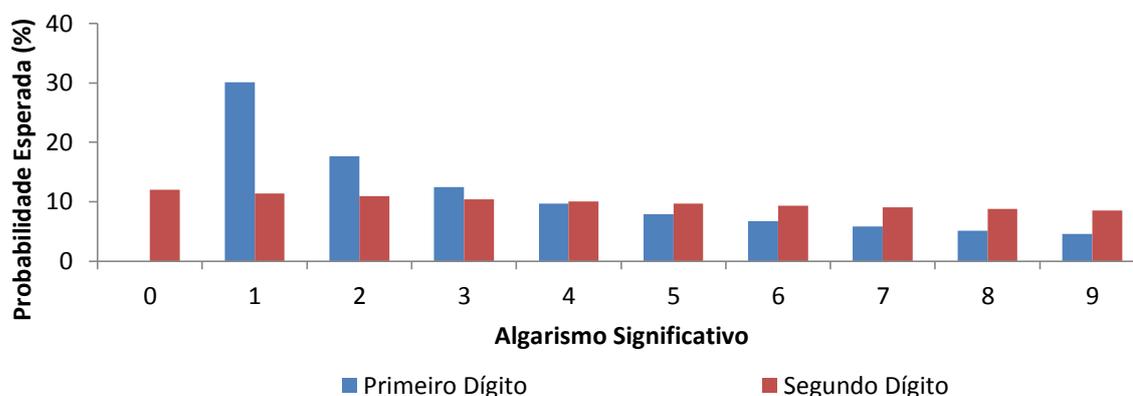
As análises primárias são divididas em três avaliações: avaliação dos primeiros dígitos, avaliação dos segundos dígitos e avaliação dos dois primeiros dígitos. Neles há a comparação da frequência dos dígitos da amostra analisada com a frequência esperada pela lei Newcomb-Benford. As frequências esperadas para os primeiros, segundos e dois primeiros dígitos estão representados na Tabela 5 e nas Figuras 2 e 3.

Tabela 5 – Frequências esperadas pela lei Newcomb-Benford

		Segundo algarismo significativo										Σ
		0	1	2	3	4	5	6	7	8	9	
Primeiro algarismo significativo	1	4,14	3,78	3,48	3,22	3,00	2,80	2,63	2,48	2,35	2,23	30,10
	2	2,12	2,02	1,93	1,85	1,77	1,70	1,64	1,58	1,52	1,47	17,61
	3	1,42	1,38	1,34	1,30	1,26	1,22	1,19	1,16	1,13	1,10	12,49
	4	1,07	1,05	1,02	1,00	0,98	0,95	0,93	0,91	0,90	0,88	9,69
	5	0,86	0,84	0,83	0,81	0,80	0,78	0,77	0,76	0,74	0,73	7,92
	6	0,72	0,71	0,69	0,68	0,67	0,66	0,65	0,64	0,63	0,62	6,69
	7	0,62	0,61	0,60	0,59	0,58	0,58	0,57	0,56	0,55	0,55	5,8
	8	0,54	0,53	0,53	0,52	0,51	0,51	0,50	0,50	0,49	0,49	5,12
	9	0,48	0,47	0,47	0,46	0,46	0,45	0,45	0,45	0,44	0,44	4,58
	Σ	11,97	11,39	10,89	10,43	10,03	9,65	9,33	9,04	8,75	8,51	≈100

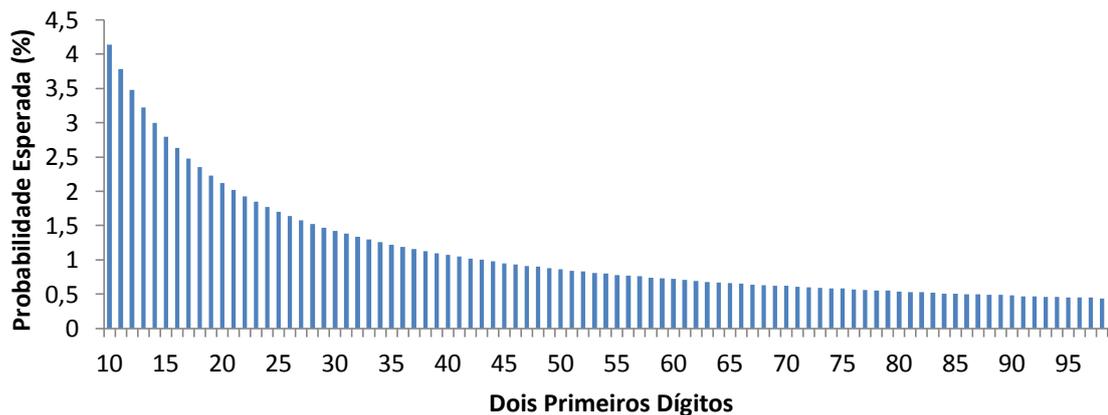
Fonte: O autor, 2021.

Figura 2 – Probabilidades esperadas pela LNB para o primeiro e segundo dígito



Fonte: O autor, 2021.

Figura 3 – Probabilidades esperadas pela LNB para os dois primeiros dígitos



Fonte: O autor, 2021.

Avaliando as análises primárias, Mantone (2013) considera como melhor método para investigação de dados financeiros é a avaliação dos dois primeiros dígitos, pois ele refina a captação da variação entre a frequência esperada e a observada, obtida nas avaliações dos primeiros e segundos dígitos, reduzindo, assim, a incidência de falsos positivos. Ele ressalta que a avaliação dos dois primeiros dígitos é um procedimento mais refinado para achar anomalias nos dados, que podem não aparecer quando realizados as análises dos primeiros e dos segundos dígitos, separadamente. Ratificando Mantone (2013), Kossovsky (2014) considera que a avaliação do primeiro dígito de forma isolada não é um procedimento muito efetivo ou uma maneira eficiente em detectar fraudes, pois uma não conformidade dos dados iniciados pelo dígito “8” obrigaria o auditor a verificar milhares de registros. Já a avaliação dos dois primeiros dígitos acusaria a não conformidade nos dados que se iniciam com os algarismos 85, facilitando, assim, o trabalho do auditor, tornando a tarefa mais rápida e menos onerosa.

Nigrini (2012) afirmou que suas pesquisas demonstraram que, para as análises primárias, amostras com números de quatro algarismos tendem a ter melhor conformidade com a lei Newcomb-Benford. Neste contexto, se a amostra possuir uma grande quantidade de números com menos de quatro algarismos, ela poderá mostrar um viés de não conformidade com a LNB. Ele também sugere não utilizar as análises primárias para uma amostra com menos de 300 dados e afirma que uma amostra para ter uma boa conformidade com a LNB deve ter no mínimo 1.000 dados. Procedimentos não estatísticos para verificação de conformidade com a LNB são mencionados no item 1.3.1.5 desta dissertação. Já os testes

estatísticos utilizados para avaliar a conformidade com a LNB são mencionados no item 1.3.2 desta dissertação.

A avaliação do segundo dígito foi utilizado por Nigrini (2005) para detectar o comportamento de arredondamento dos lucros divulgados pelas empresas. Foi detectado que o *controller* da organização tinha o hábito de arredondar os dados de venda e lucro líquido para cima. Esta constatação foi possível devido a análise dos segundos dígitos explicitar a excesso do algarismo “0” e poucos algarismos “9” na segunda posição dos dados. Este procedimento é considerado por Nigrini (2012) como uma ferramenta útil para detecção de problemas em dados eleitorais, valores de inventários, leituras de odômetros e números de vendas diárias. Ele ressalta que a conformidade na análise dos primeiros dígitos não implica em conformidade na análise dos segundos dígitos, elas são análises independentes e devem ser aplicadas e interpretadas separadamente.

Diekmann (2007) coletou os dados das tabelas publicadas em dois volumes do *American Journal of Sociology* (Volumes 101 e 102) e realizou a avaliação dos primeiros e segundos dígitos com o auxílio de testes estatísticos, a fim de verificar conformidade com a lei Newcomb-Benford. Ambas as avaliações apresentaram conformidade com a LNB. Após estas avaliações, foram coletados dados fabricados e realizados os mesmos procedimentos. Como resultado, os dados fabricados obtiveram, quando analisado os primeiros dígitos, conformidade com a lei, porém a conformidade não ocorreu quando realizada a avaliação para os segundos dígitos. Com isso, ele concluiu que, para melhor detecção de anomalias nos dados, não se deve usar apenas a avaliação dos primeiros dígitos da LNB e sim usá-lo juntamente com a avaliação dos segundos dígitos.

1.3.1.3 Análises avançadas

As análises avançadas compreendem duas análises, a análise de segunda ordem e a análise da soma.

1.3.1.3.1 Análise de segunda ordem

A análise de segunda ordem começou quando Raimi (1976) relaxou a restrição de que apenas uma sequência geométrica teria conformidade com a lei Newcomb-Benford. Ele notou que uma mistura intercalada de uma sequência geométrica também produziria uma distribuição igual àquela proposta por Newcomb (1881) e Benford (1938). Considerando T_n o n ésimo termo de uma progressão geométrica de primeiro termo “ w ” e razão “ r ”, e D_n a diferença de dois termos subsequentes pode-se escrever as Equações 5 e 6.

$$T_n = wr^{n-1} \quad (5)$$

$$D_n = wr^n - wr^{n-1} = w(r - 1)r^{n-1} \quad (6)$$

Com isso, uma amostra de “ n ” elementos se transforma em uma amostra de “ $n-1$ ” elementos. Esta amostra também é uma progressão geométrica com primeiro elemento “ $w(r-1)$ ” e razão “ r ”, conforme explicitado na Equação 6.

Primeiramente, Miller e Nigrini (2008) provaram matematicamente que distribuições normais, uniformes, triangulares e gammas, após cumprir o rito da análise de segunda ordem, fornecem uma nova sequência, que está em conformidade com a LNB. Os casos de não conformidade são extremamente raros, funcionando como um indicador de que existe alguma espécie de anomalia ou erro nos dados.

Esta análise foi utilizada pela primeira vez por Nigrini e Miller (2009), em quatro estudos de casos. No primeiro caso, analisaram 38.176 transações realizadas no contas a pagar da mesma empresa estudada por Drake e Nigrini (2000). No segundo estudo de caso, eles investigaram 154.935 lançamentos contábeis do sistema contábil de uma empresa, no ano de 2005. No terceiro, eles utilizaram a análise de segunda ordem para investigar os dados contábeis de uma grande cadeia de restaurantes. No quarto e último caso, eles usaram os dados de receitas de restaurantes e os custos com alimentos, no ano de 2003.

Para iniciar a análise de segunda ordem, o auditor, primeiramente, deve ordenar seus dados de forma crescente. Após isso, inicia a montagem da nova série, subtraindo o último dado do penúltimo, depois subtrai o penúltimo do antepenúltimo e assim, sucessivamente, até subtrair o segundo termo do primeiro, formando, desta maneira, uma amostra de $n-1$ elementos. Caso a amostra contenha números iguais, o resultado da subtração dará zero, este

elemento será descartado da amostra. Deste modo, o auditor compara a distribuição dos primeiros e dois primeiros dígitos da nova amostra com a LNB. (NIGRINI, 2020)

Com a relação à conformidade, Nigrini (2020) sugere que existem duas maneiras de avaliar se a análise de segunda ordem está aderente com a lei Newcomb-Benford. A primeira maneira é examinar os dados e fazer um julgamento gráfico e visual. A segunda é a utilização do procedimento do desvio absoluto médio (MAD), que é tratado, com mais detalhes, no item 1.3.1.5.2. Nigrini (2020) não considera os testes Z, qui quadrado e Kolmogorov–Smirnov como apropriados para avaliar a conformidade da análise de segunda ordem com a LNB, em razão de eles serem testes mais sensíveis e o fato de que os novos dados obtidos nunca serão exatamente uma distribuição de Newcomb-Benford.

1.3.1.3.2 Análise da soma

Esta análise, desenvolvida por Nigrini (1992), é baseada em somas dos números pertencentes a cada grupo de primeiros dígitos (1, 2, 3,..., 9) ou dois primeiros dígitos (10, 11, 12,..., 99) e não na simples contagem dos números pertencentes ao grupo, como nas análises primárias. Ele notou que as somas dos números pertencentes aos grupos eram aproximadamente iguais, ou seja, representavam 1/90 do total, quando analisados os dois primeiros dígitos e 1/9 do total quando analisado os primeiros dígitos. Isto é demonstrado pela Equação 7.

$$\begin{aligned}
 & \int_{\log(ft)*N}^{\log(ft+1)*N} ar^{n-1} dn = \\
 & = \frac{a}{r} * \left. \frac{r^x}{\ln r} \right|_{\log(ft)*N}^{\log(ft+1)*N} = \\
 & = \frac{a}{r \ln r} * (r^{\log(ft+1)*N} - r^{\log(ft)*N}) = \\
 & = \frac{a}{r \ln r}
 \end{aligned} \tag{7}$$

Na primeira linha, Nigrini (1992) calculou a área embaixo de uma curva qualquer de uma combinação de dois dígitos, chamada ft. Concluiu na última linha, que a área embaixo da curva não depende da função ft e isso significa que as áreas devem ser iguais.

A análise da soma complementa as análises primárias, já que estas analisam apenas a quantidade dos números, mas não sua ordem de grandeza. Por exemplo, a análise dos dois primeiros dígitos explicita a quantidade de dados iniciados pelos algarismos 30, porém ele não faz a diferenciação se o valor é R\$ 30,00 ou R\$ 300.000.000,00. A função da análise da soma é detectar a presença de números muito grandes comparados com o resto da amostra nos grupos dos primeiros dígitos e dois primeiros dígitos (NIGRINI, 2012).

1.3.1.4 Análises associadas

As análises associadas são compostas por outras três análises, são elas: a análise dos números duplicados, a análise dos dois últimos dígitos e o modelo fator de distorção.

1.3.1.4.1 Análise dos números duplicados

Esta análise identifica os números que causam picos nas análises primárias e na análise da soma. Ele foi proposto pela primeira vez por Nigrini (1992) e deve ser aplicado em todos os números, sejam eles positivos ou negativos. O procedimento consiste em identificar quais números estão repetindo, verificar a quantidade de repetições e ranquear os números, para analisar se eles causaram anomalias na análise de primeira ordem ou análise da soma. Kossovsky (2014) considera este teste útil, pois pode apontar para aqueles números falsos que o fraudador repetiu inconscientemente.

Este procedimento pode ser utilizado para qualquer tipo de dado e não apenas em dados financeiros. Ele foi utilizado por auditores de uma companhia aérea norte-americana para observar os depósitos de milhas aéreas realizados pelos seus clientes. Chegaram à conclusão que os números que mais se repetiam eram os de 500 e 802 milhas. O motivo da repetição desses valores era porque 500 milhas correspondiam à quantidade mínima de transferência e 802 milhas era a quantidade de milhas necessárias para aquisição de um trecho. Outro exemplo de utilização deste procedimento foi em uma empresa localizada no Tennessee, que utilizou a análise de dados para encontrar funcionários fantasmas. A aplicação deste procedimento foi realizado no número da conta em que a empresa fazia o depósito dos

salários. Foi observado que algumas contas se repetiam duas ou até três vezes. O motivo foi que casais que trabalhavam na empresa usavam a mesma conta para depósito e também funcionários mais jovens utilizavam a conta do amigo com que ele dividia apartamento, pois não tinha conta em banco (NIGRINI, 2020).

1.3.1.4.2 Análise dos últimos dois dígitos

A análise dos dois últimos dígitos é considerada, por Kossovsky (2014), uma ferramenta poderosa para detecção de números inventados, porém este tipo de análise não é apropriado para dados que ilustram pagamentos ou despesas. O motivo é que nesta espécie de dados os últimos dois dígitos representam os centavos, que na grande maioria das vezes estarão como números 00 e o procedimento dará a impressão que os dados estão sendo arredondados de maneira proposital. Por este motivo esta análise não será utilizada nesta dissertação para avaliar os valores das notas de empenhos.

Embora não seja salutar a utilização desta análise em dados financeiros, Nigrini (2012) enumera diversos tipos de dados em que este procedimento é satisfatório na detecção de anomalias, são eles: números de censos populacionais, resultados eleitorais, inventários, leitura de odômetros, pesos de peixes pescados, números de deduções fiscais e estatísticas de sites.

1.3.1.4.3 Modelo fator de distorção

O modelo fator de distorção (MFD) é um procedimento não estatístico proposto por Nigrini (2012) e Nigrini (2020). Ele foi introduzido na literatura científica por Nigrini (1992), em sua tese de doutorado, porém foi publicado pela primeira vez, por Nigrini (1996). Este procedimento detecta distorções que as análises primárias não são capazes de identificar. Este procedimento considera que o excesso de dígitos menores significa que os valores estão subestimados, já quando existe excesso de dígitos maiores significa que os números foram superestimados.

Este procedimento analisa se os dados foram superestimados ou subestimados, e mensura o tamanho deste desvio. Primeiramente, são excluídos da amostra os números inferiores a 10. Após este processo, todos os números, exceto os compreendidos entre 10,00 e 99,99, são transformados em números de apenas dois algarismos significativos, conforme Equação 8 (considerando $\log a$, logaritmo na base 10 de a).

$$a_{\text{modificado}} = \frac{10 a}{10^{\text{int}(\log a)}} \quad (8)$$

Assim, por exemplo, o número 8379 modificado está expresso na Equação 9.

$$\begin{aligned} a_{\text{modificado}} &= \frac{10 \times 8379}{10^{\text{int}(\log 8379)}} = \\ &= \frac{83790}{10^{\text{int}(\log 8379)}} = \\ &= \frac{83790}{10^3} = \\ a_{\text{modificado}} &= 83,79 \end{aligned} \quad (9)$$

Após este passo, calcula-se a média atual (MA) da nova amostra com “n” elementos, através da equação 10.

$$MA = \frac{\sum a_{\text{modificado}}}{n} \quad (10)$$

Para haver comparação das amostras estudadas, calcula-se a soma esperada (SE) por uma amostra que segue a LNB e, também, sua média esperada (ME), considerando “w” o primeiro termo, “r” a razão e “n” o número de elementos, pelas equações 11 e 12, respectivamente.

$$SE = \frac{w \times (r^n - 1)}{(r - 1)} \quad (11)$$

$$ME = \frac{90}{(n \times (10^{1/n} - 1))} \quad (12)$$

Assim, a média esperada para uma amostra de 123.456 itens está representada na Equação 13.

$$ME = \frac{90}{(123456 \times (10^{1/123456} - 1))} =$$

$$ME = \frac{90}{2,302606} = 39,08613 \quad (13)$$

Como este procedimento é para analisar amostras grandes Nigrini (1992) considerou que “n” seja um número muito grande e tenda ao infinito, temos que ME tende a 39,0865, como explicitado na Equação 14.

$$ME = \frac{90}{\lim_{n \rightarrow \infty} (n \times (10^{1/n} - 1))} =$$

$$\frac{90}{\ln 10} = 39,0865 \quad (14)$$

Assim, Nigrini (2012) sugere que para amostras muito grandes deve-se utilizar a média esperada com valor de 39,0865, como demonstrado na Equação 14.

O primeiro passo para aplicar o modelo fator de distorção é excluir da amostra todos os números menores que 10,00. Após isso, deve-se encontrar para todos os dados restantes seus $a_{\text{modificado}}$ e calcular sua média atual (MA). O próximo passo é calcular o fator de distorção (FD), por intermédio da Equação 15.

$$FD = \frac{MA - ME}{ME} \quad (15)$$

Este procedimento foi utilizado por Nigrini (1996) como ferramenta para análise de conformidade de dados com a lei Newcomb-Benford, com a finalidade de analisar a evasão fiscal norte americana. Ele também, utilizado por Dumas e Devine (2000) para verificar a veracidade dos dados de emissão de poluentes informados pelos países, e, por Diniz, Corrar e Slomsky (2010) para investigar as despesas municipais.

1.3.1.5 Análises de conformidade

Nesta seção foram abordados os procedimentos adotados por Nigrini (2012) e Nigrini (2020) para verificar se a amostra avaliada está em conformidade com a lei Newcomb-

Benford. Foram abordados os seguintes procedimentos: Z_{teste} para o modelo fator de distorção, desvio absoluto médio e *mantissa arc test*.

1.3.1.5.1 Z_{teste} - modelo fator de distorção

Este procedimento é utilizado para analisar o valor do fator de distorção (FD) obtido na Equação 15. O valor do Z_{teste} foi proposto por Nigrini (1992) e utilizado, pela primeira vez, com o intuito de investigar os dados da receita federal norte-americana por Nigrini (1996). O Z_{teste} é calculado conforme Equação 16, sendo FD o fator de distorção encontrado na Equação 15 para uma amostra com “n” dados.

$$Z_{\text{teste}} = \frac{FD}{\frac{0,638253}{\sqrt{n}}} \quad (16)$$

O valor do Z_{teste} é comparado com o valor do $Z_{\text{crítico}}$. Para este procedimento, Nigrini (1992) adotou as seguintes hipóteses:

H_0 : O fator de distorção da amostra é estatisticamente igual a zero.

H_1 : O fator de distorção da amostra é estatisticamente diferente de zero.

Caso $|Z_{\text{teste}}| < |Z_{\text{crítico}}|$ a hipótese nula (H_0) é aceita, já se $|Z_{\text{teste}}| > |Z_{\text{crítico}}|$ a hipótese nula é rejeitada e a hipótese alternativa (H_1) aceita.

1.3.1.5.2 Desvio absoluto médio (Mean Absolute Deviation)

Outro modo de analisar a conformidade da amostra como um todo é através do procedimento denominado *Mean Absolute Deviation* ou MAD. Este procedimento foi proposto por Drake e Nigrini (2000), Nigrini (2012) e Nigrini (2020) em substituição do teste Kolmogorov-Smirnov, haja vista que o MAD faz uma análise global da amostra, sem levar em consideração o tamanho da mesma e sim o número de algarismos analisados.

Drake e Nigrini (2000) consideram este procedimento válido para verificar a conformidade dos dados com a lei Newcomb-Benford, analisando 100% da amostra. Esta

ferramenta soma os desvios absolutos de cada algarismo e depois divide este valor pela quantidade de algarismos analisados - nove se for aplicada com a análise dos primeiros dígitos, dez se for aplicada com a análise dos segundos dígitos e 90 se aplicada com a análise dos dois primeiros dígitos.

Considerando a proporção observada como “po”, a proporção esperada como “pe” e a quantidade de algarismos analisados como “K”, o valor MAD_{teste} é calculado através da Equação 17.

$$MAD_{teste} = \frac{\sum_{i=1}^k |po-pe|}{K} \quad (17)$$

Para ser concluído qual o nível de conformidade obtido através do MAD_{teste} era utilizada a tabela de Drake e Nigrini (2000), porém esta tabela foi atualizada por Nigrini (2012). A faixa de valores do MAD adotado, por ambos os artigos, estão expressos na Tabela 6. Este trabalho, entretanto, adota, como base, os valores propostos por Nigrini (2012). As cores adotadas no campo conclusão da Tabela 6 também foram utilizadas na seção 3 (análise e discussão dos resultados), para realçar a conclusão do resultado do MAD.

Tabela 6– Faixa de valores do MAD

Teste	Drake e Nigrini(2000)	Nigrini(2012)	Conclusão
Primeiros dígitos	0,000 até 0,004	0,000 até 0,006	Alta conformidade
	0,004 até 0,008	0,006 até 0,012	Conformidade aceitável
	0,008 até 0,012	0,012 até 0,015	Conformidade marginalmente aceitável
	Acima de 0,012	Acima de 0,015	Não conformidade
Segundos dígitos	0,000 até 0,006	0,000 até 0,008	Alta conformidade
	0,006 até 0,012	0,008 até 0,010	Conformidade aceitável
	0,012 até 0,018	0,010 até 0,012	Conformidade marginalmente aceitável
	Acima de 0,018	Acima de 0,012	Não conformidade
Dois primeiros dígitos	0,0000 até 0,0006	0,0000 até 0,0012	Alta conformidade
	0,0006 até 0,0012	0,0012 até 0,0018	Conformidade aceitável
	0,0012 até 0,0018	0,0018 até 0,0022	Conformidade marginalmente aceitável
	Acima de 0,0018	Acima de 0,0022	Não conformidade

Fonte: O autor, 2021.

O MAD foi adotado nesta dissertação para verificar a conformidade com a lei Newcomb-Benford nas análises dos primeiros dígitos, dos segundos dígitos, dos dois primeiros dígitos e de segunda ordem.

1.3.1.5.3 *Mantissa arc test*

Mantissa arc test é um procedimento não estatístico para análise global da amostra e verificação de conformidade dos dois primeiros dígitos da amostra com a lei Newcomb-Benford. Este método foi utilizado pela primeira vez por Alexander (2009) e sugerido por Nigrini (2012) e Nigrini (2020) para análise global da amostra. Este procedimento consiste em comparar o centro de gravidade de um círculo de raio um e centro no ponto (0,0) (figura que uma amostra que segue perfeitamente a LNB forma pelo procedimento de Alexander (2009)) com o centro de gravidade da figura gerada por todos os dados da amostra. As coordenadas no eixo x e y de cada dado são calculadas pelas equações 18 e 19, sendo que $X_1, X_2, X_3, \dots, X_n$ são os “n” dados da amostra e “mod1” significa utilizar apenas a parte decimal do resultado de $\log x_1$.

$$\text{coordenada } X = \cos(2\pi (\log(x_1) \text{ mod } 1)) \quad (18)$$

$$\text{coordenada } Y = \sin(2\pi (\log(x_1) \text{ mod } 1)) \quad (19)$$

Este procedimento consiste na obtenção do centro de gravidade do círculo formado pelos dados da amostra, obtidos pelas Equações 20 e 21. Já para o cálculo da distância do centro de massa do círculo de origem (0,0) e raio um para o centro de massa da figura formada pelos dados da amostra utiliza-se a Equação 22.

$$X_{centro} = \frac{\sum_{i=1}^N \cos(2\pi (\log(x_i) \text{ mod } 1))}{N} \quad (20)$$

$$Y_{centro} = \frac{\sum_{i=1}^N \sin(2\pi (\log(x_i) \text{ mod } 1))}{N} \quad (21)$$

$$D_{centros}^2 = X_{centro}^2 + Y_{centro}^2 \quad (22)$$

Alexander (2009) adotou para o *Mantissa Arc Test* as seguintes hipóteses:

H_0 : A distribuição dos dados na mantissa é uniforme

H_1 : A distribuição dos dados na mantissa não é uniforme.

Para verificar se o procedimento aceita ou rejeita a hipótese nula, é comparado o seu nível de significância com o seu p-valor, que é calculado conforme Equação 23, proposta por Alexander (2009).

$$p \text{ valor} = 1 - \frac{1}{e^{D_{centro}^2 * n}} \quad (23)$$

No *Mantissa Arc Test*, caso p-valor < nível de significância aceita-se a hipótese nula e considera-se que os dados se distribuem uniformemente, seguindo, assim, a LNB. Por outro lado, caso p-valor > nível de significância rejeita-se a hipótese nula e considera-se que os dados não se distribuem uniformemente, não seguindo, assim, a LNB.

Outra forma de avaliar a conformidade, além da comparação do p-valor com o nível de significância, foi proposta por Alexander (2009). Ele sugere para cálculo do valor crítico de D^2 em uma amostra de “n” dados a utilização da Tabela 7, para “n” maior que 45 elementos.

Intervalo de confiança	D^2
90%	2,30/n
95%	2,99/n
99%	4,60/n

Fonte: ALEXANDER, 2009, p. 12.

Nigrini (2012) recomenda cautela na aplicação deste procedimento, pois ele fornece muitos falsos positivos ou erros tipo I, assunto que é tratado mais a frente nesta dissertação.

1.3.2 Testes estatísticos

Nigrini e Mittermaier (1997) afirmaram que a LNB, juntamente com a utilização de testes estatísticos, pode ser utilizada uma ferramenta útil pelos auditores na fase de planejamento para escolha de amostras de auditoria. Para isto, entretanto, os auditores necessitariam testar a semelhança da distribuição dos dígitos analisados com a LNB, comparando as frequências observadas com as esperadas, através de testes de conformidade.

Desta forma, os resultados encontrados poderiam ajudar os auditores a determinar as amostras de auditoria selecionada para procedimentos adicionais.

A escolha de amostras de auditoria é motivada pela escassez de recursos e tempo para uma análise minuciosa de todos os dados. Porém, isto acarreta riscos. Existem duas espécies de riscos que podem ocorrer, eles são os erros tipo I e tipo II. O erro tipo I ocorre quando o auditor, de maneira errônea, rejeita a hipótese nula do teste estatístico ao invés de aceitá-la, concluindo pela indicação indevida de fraude. O erro tipo II acontece quando o auditor aceita a hipótese nula, quando deveria rejeitá-la, concluindo pela indicação indevida de ausência de fraude. Os auditores devem se preocupar mais com o erro tipo II, tentando mitigá-lo, ao máximo, pois ele deixa passar despercebidos dados fraudulentos, que podem acarretar perdas financeiras significativas. Com a mitigação do erro tipo II, ocorre um aumento no erro tipo I, fazendo com que o auditor considere como fraudulentos os dados honestos, nos quais, mesmo com uma análise mais profunda, não encontre fraudes. O erro tipo I é menos nocivo ao auditor (GEE,2015).

Bhattacharya, Xu e Kumar (2011) também discutem sobre o tema e afirmam que os testes estatísticos de conformidade estão propensos a desacertos como, por exemplo, os erros tipo I e tipo II, que são denominados, também, de erro do falso alarme e erro da falha de alarme, respectivamente. Ambos são considerados erros de classificação, o erro tipo I ou alarme falso ocorre quando o teste estatístico considera uma amostra honesta como uma amostra que tenha viés fraude. Já o erro tipo II ou alarme com falha acontece quando a amostra está adulterada, porém o teste estatístico a considera como uma amostra honesta. Este erro é crítico para o auditor e deve ser o seu principal foco de mitigação.

Cleary e Thibodeau (2005) chamam a atenção que, embora a aplicação da análise de dados, usando a lei Newcomb-Benford, seja uma ferramenta para detecção de fraudes, os auditores que a utilizam devem estar cientes dos custos dos possíveis erros do Tipo I, passíveis de ocorrerem durante a análise de dados. Seu trabalho concluiu que existem maiores chances de este erro ocorrer quando realizada uma análise dígito por dígito da amostra do que quando a análise da LNB for realizada de forma global. Eles concluíram que o uso de uma abordagem dígito por dígito aumenta a chance de um erro Tipo I, embora aumente, também, as chances de encontrar fraudes. Assim, sugerem que a abordagem mais prudente seria iniciar a fase de análise com uma análise global, usando um teste qui quadrado, e depois seguir com uma análise local (dígito por dígito) somente se houver uma indicação de possível fraude na análise global.

Nesta seção, serão abordados os testes estatísticos que verificam se a amostra analisada está em conformidade com a lei Newcomb-Benford de forma global ou local. Serão abordados os seguintes testes estatísticos: teste Z, teste qui quadrado (χ^2) e teste Kolmogorov-Smirnov (K-S).

1.3.2.1 Teste Z

O teste Z é utilizado para avaliar se a distribuição observada de cada dígito da amostra está em conformidade com a distribuição esperada por Newcomb (1881) e Benford (1938). Esta ferramenta é considerada um teste local, pois analisa cada dígito separadamente e é utilizada para as análises dos primeiros dígitos, segundos dígitos e dois primeiros dígitos (KOSSOVKY, 2014).

Para o cálculo do Z_{teste} , Fleiss (1981) sugere a Equação 24, sendo “po” a proporção observada da amostra, “pe” a proporção esperada caso a amostra seguisse a lei Newcomb-Benford e “n” o número de elementos da amostra observada.

$$Z_{teste} = \frac{|po-pe| - \frac{1}{2n}}{\sqrt{\frac{pe(1-pe)}{n}}} \quad (24)$$

Ao invés de utilizar o Z_{teste} , pode-se calcular o limite do intervalo de confiança que a proporção de cada algarismo pode alcançar para estar em conformidade com a LNB. Assim, têm-se as Equações 25 e 26, utilizadas por Nigrini (2012), para analisar, respectivamente, o limite inferior e superior para o nível de significância α .

$$Limite Superior = pe + \left(Z_{\alpha/2} \sqrt{\frac{pe(1-pe)}{n}} \right) + \frac{1}{2n} \quad (25)$$

$$Limite Inferior = pe - \left(Z_{\alpha/2} \sqrt{\frac{pe(1-pe)}{n}} \right) - \frac{1}{2n} \quad (26)$$

O termo $1/2n$, conhecido como fator de correção contínuo, é apenas utilizado quando este valor é inferior ao primeiro termo do numerador ($1/2n < |po - pe|$) ou quando é

desejável realizar um teste mais conservador (NIGRINI, 1996; SHESKIN, 2007; KOSSOVSKY, 2014).

Porém, Levine *et. al.* (2008) afirmam que caso “n” seja um número muito grande, pode-se desconsiderar o fator de correção de continuidade $1/2n$ e retirar o módulo no cálculo do desvio da Equação 24. Assim, pode-se reescrever a Equação 24 conforme Fleiss (1981) da seguinte forma.

$$Z_{teste} = \frac{PO-PE}{\sqrt{PE(1-pe)}} \quad (27)$$

Em que PO é a quantidade absoluta de observações da amostra para o primeiro dígito analisado e PE a quantidade absoluta esperada para o dígito da posição, conforme a lei Newcomb-Benford. Já as Equações 28 e 29, de Nigrini (2012), são utilizadas no lugar das Equações 25 e 26 quando é desconsiderado o fator de correção de continuidade, ficam da seguinte forma:

$$Limite Superior = pe + \left(Z_{\alpha/2} \sqrt{\frac{pe(1-pe)}{n}} \right) \quad (28)$$

$$Limite Inferior = pe - \left(Z_{\alpha/2} \sqrt{\frac{pe(1-pe)}{n}} \right) \quad (29)$$

Este teste foi utilizado pela primeira vez no campo da auditoria contábil por Carlsaw (1988), com objetivo de mensurar a conformidade entre as frequências observadas e as esperadas para os dígitos da primeira e segunda posição, separadamente. Este é um teste local, no qual cada dígito é analisado separadamente.

No teste Z, adota-se as seguintes hipóteses:

H_0 : O dígito analisado está em conformidade com a LNB.

H_1 : O dígito analisado não está em conformidade com a LNB.

O teste Z é aplicado de forma local para cada dígito da amostra. Nele é decidido se a hipótese nula (H_0) é aceita através da comparação do Z_{teste} com o $Z_{crítico}$ ou comparando a proporção observada com o limite superior e inferior (Equações 20 e 21) ou através da comparação entre do p-valor do teste e seu nível de significância (α).

Se $|Z_{teste}| < |Z_{crítico}|$ ou p-valor $>$ nível de significância, aceita-se H_0 , já se $|Z_{teste}| > |Z_{crítico}|$ ou p-valor $<$ nível de significância, rejeita-se H_0 e aceita-se a hipótese alternativa

(H_1). Com relação aos limites superior e inferior, se o valor da proporção observada estiver compreendida entre o intervalo do limite inferior e do limite superior aceita-se H_0 , caso contrário rejeita-se H_0 e aceita-se H_1 .

O teste Z tem uma peculiaridade para amostras muito grandes, ele fica mais propício a erros tipo I ou falso positivos à medida que o número de elementos da amostra cresce (KRAKAR; ZGELA, 2009, NIGRINI; MITTERMAIER, 1997).

1.3.2.2 Teste qui quadrado (χ^2)

O teste qui quadrado é um teste estatístico utilizado para analisar a amostra de maneira global, ou seja, como um todo, e diferentemente do teste Z, que analisa a proporção observada de cada algarismo separadamente. O teste qui quadrado foi utilizado na presente dissertação para a análise dos primeiros dígitos, segundos dígitos e dos dois primeiros dígitos. O teste χ^2 tem a mesma peculiaridade do teste Z, à medida que o tamanho da amostra testada fica maior mais rigoroso o teste fica, ou seja, maior chance de ocorrer o erro tipo I. (NIGRINI, 2020; NIGRINI, 2012; KRAKAR; ZGELA, 2009, NIGRINI; MITTERMAIER, 1997).

Calcula-se o χ^2_{teste} através da Equação 30, conforme Kossovsky (2014), Nigrini (2012), Gee (2015), e Baesens, Van Vlasselaer e Verbeke (2015), Nigrini (2020), considerando “PO” a proporção absoluta observada, “PE” a proporção absoluta esperada e “k” o número de linhas do teste estatístico, nove para o teste do primeiro dígito, dez para o teste do segundo dígito e noventa para o teste dos dois primeiros dígitos.

$$\chi^2_{teste} = \sum_{i=1}^k \frac{(PE-PO)^2}{PE} \quad (30)$$

Este teste adota as seguintes hipóteses:

H_0 : A amostra segue a LNB.

H_1 : A amostra não segue a LNB.

O valor do χ^2_{teste} é comparado com o $\chi^2_{crítico}$, caso $\chi^2_{teste} < \chi^2_{crítico}$ aceita-se H_0 , já se $\chi^2_{teste} > \chi^2_{crítico}$ rejeita-se H_0 e aceita-se H_1 . O valor $\chi^2_{crítico}$ leva em consideração o nível de significância e o grau de liberdade da amostra. O grau de liberdade é igual ao número de linhas da tabela analisada (k) menos uma unidade, ou seja, k-1.

1.3.2.3 Teste Kolmogorov-Smirnov (K-S)

O teste Kolmogorov-Smirnov é um teste não paramétrico de aderência, baseado na função densidade acumulada. Ele compara a distribuição cumulativa de uma variável com uma distribuição específica, analisando a amostra como um todo. Sua aplicação com a lei Newcomb-Benford é feita através da comparação da frequência acumulada dos primeiros, segundos e dois primeiros dígitos dos dados pesquisados com a frequência acumulada proposta por Newcomb (1881) e Benford (1938).

Este teste foi utilizado, juntamente com a LNB, por Tam Cho e Gaines (2007) e Bhattacharya, Xu e Kumar (2011). No presente trabalho, o cálculo da estatística K-S foi feito de acordo com Bhattacharya, Xu e Kumar (2011), e explicitado nas Equações 31 (para o primeiro dígito), 32 (para o segundo dígito) e 33 (para os dois primeiros dígitos).

Considerando f_{1j} é a frequência observada no primeiro dígito do j -ésimo dígito e B_{1j} a frequência esperada do j -ésimo dígito como o primeiro dígito, conforme a LNB, tem-se a Equação 31:

$$K_1 = \max_i |\sum (f_{1j} - B_{1j})| \quad i, j = 1, 2, \dots, 9 \quad (31)$$

Considerando f_{2j} é a frequência observada no segundo dígito do j -ésimo dígito e B_{2j} a frequência esperada do j -ésimo dígito como o segundo dígito, conforme LNB, tem-se a Equação 32:

$$K_2 = \max_i |\sum (f_{2j} - B_{2j})| \quad i, j = 0, 1, 2, \dots, 9 \quad (32)$$

Considerando f_{12j} é a frequência observada dos dois primeiro dígito do j -ésimo dígito e B_{12j} a frequência esperada do j -ésimo dígito como os dois primeiro dígito, conforme a LNB, tem-se a Equação 33:

$$K_{12} = \max_i |\sum (f_{12j} - B_{12j})| \quad i, j = 10, 11, 12, \dots, 99 \quad (33)$$

O valor do $K-S_{\text{crítico}}$ varia de acordo com nível de significância adotado " α " e o número de dados da amostra " n ". Estes valores estão representados na Tabela 8.

Tabela 8 – Valor $K-S_{\text{crítico}}$

α	10%	5%	1%
$K-S_{\text{crítico}}$	$\frac{1,22}{\sqrt{n}}$	$\frac{1,36}{\sqrt{n}}$	$\frac{1,63}{\sqrt{n}}$

Fonte: O autor, 2021.

Este teste estatístico adota as seguintes hipóteses:

H_0 : A amostra segue a lei Newcomb-Benford.

H_1 : A amostra não segue a lei Newcomb-Benford.

Caso o valor do $K-S_{\text{teste}} < K-S_{\text{crítico}}$ aceita-se H_0 , já se $K-S_{\text{teste}} > K-S_{\text{crítico}}$ rejeita-se H_0 e aceita-se H_1 .

Nigrini (2012) e Nigrini (2020) explicam que ao aplicar este teste deve-se ter cautela, devido ao tamanho da amostra “n” para o cálculo do $K-S_{\text{crítico}}$. Isto faz com que ele tenha peculiaridade de proporcionar muitos erros tipo I ou falsos positivos, caso a amostra apreciada possua muitos elementos, fazendo com que a inclusão de “n” no cálculo do $K-S_{\text{crítico}}$ torne este teste não usual para algumas situações cotidianas. Silva, Korzenowski e Vaccaro (2014), quando utilizaram o teste qui quadrado e o teste K-S, concluíram que o teste K-S foi mais conservador para avaliar a suspeita de lavagem de dinheiro utilizando a LNB, ou seja, na avaliação do teste K-S mais amostras não seguiram a LNB.

2 METODOLOGIA

Nesta seção foram descritos os procedimentos metodológicos para realização da dissertação. Esta seção é dividida em cinco subseções: método da pesquisa, seleção da amostra, coleta dos dados, tratamento dos dados e delimitação da pesquisa.

2.1 Método da pesquisa

Quanto ao método que proporciona a base lógica de investigação, esta dissertação é caracterizada pelo método hipotético dedutivo, pois através de testes e formulação de hipóteses procura refutar axiomas que regem eventos - lei Newcomb-Benford (MARKONI; LAKATOS, 2007).

Para Malhotra (2012) quanto à abordagem do problema, a pesquisa é considerada quantitativa, pois procura quantificar os dados com auxílio da análise estatística ou outros procedimentos chegando à resposta de um problema. Isto se deve ao fato da pesquisa empregar os procedimentos estatísticos e procedimentos não estatísticos propostos por Nigrini (2012) e Nigrini (2020) para avaliar a conformidade das notas de empenhos emitidas com a lei Newcomb-Benford.

Já com relação à concepção da pesquisa, para Malhotra (2012) esta dissertação pode ser considerada como descritiva, por expor características de uma determinada população, que no presente trabalho são os empenhos emitidos por quatro tipos de Organizações Militares (OMs) da Marinha do Brasil (MB).

A seguir são apresentados os critérios para seleção da amostra e as OMs selecionadas para a pesquisa.

2.2 Seleção da amostra

Os tipos de amostragem podem ser divididos em dois grupos, a amostragem probabilística e a não probabilística. Esta pesquisa adotou a amostragem não probabilística do tipo por julgamento. Este tipo de amostragem é caracterizado pela seleção amostral ficar a cargo do julgamento pesquisador (MALHOTRA, 2012).

A população considerada nesta dissertação foram as 156 Unidades Gestoras Executoras (UGE) da Marinha do Brasil, que são responsáveis pela execução financeira de 444 organizações militares. A escolha pela população de UGE se deu pelo fato de elas centralizarem a execução financeira das Organizações Militares apoiadas, aumentando, assim, a quantidade de dados analisados. Esta escolha está em linha com o entendimento de Nigrini (2012), Nigrini (2020) e Kossovsky (2014) que, como já salientado, frisam que quanto maior a quantidade de dados analisados na amostra, maior será a confiabilidade dos testes estatísticos e procedimentos não estatísticos.

Foi escolhido analisar as organizações militares em grupos, de acordo com seu tipo. Isto possibilita a comparação dos resultados de organizações militares semelhantes, pois, conforme Kossovsky (2014), um dos motivos da não conformidade com a lei Newcomb-Benford é que o órgão segue uma configuração de negócio particular, que causa o desvio da lei. Assim, caso apenas uma organização do grupo apresente desvios da lei Newcomb-Benford, conforme Nigrini (2012), seus dados têm mais probabilidade de conter alguma fraude ou erro.

O ano de 2019 foi selecionado por ser o primeiro exercício financeiro completo no qual a despesa pública fora realizada à luz do Decreto nº 9.412, de 18 de junho de 2018, que atualizou os valores das modalidades de licitação. Este decreto estabeleceu a possibilidade de aquisição de bens e contratação de serviços (exceto os de engenharia), sem licitação, até o limite de R\$ 17.600,00.

Neste contexto, foram adotados, para esta pesquisa, quatro grupos, para analisar a conformidade da distribuição das notas de empenhos, emitidas em 2019, com a distribuição proposta por Newcomb (1881) e Benford (1938). Os grupos são compostos pelos seguintes tipos de organização militares: Centros de Intendência da Marinha (CeIM), Escolas de Aprendizes-Marinheiros (EAM), Bases e Estações Navais (BE) e Hospitais Navais (HN).

Os Centros de Intendência da Marinha são organizações militares da Marinha do Brasil, que possuem a missão de desempenhar as funções de coordenação do abastecimento e

atuar como órgão central na execução financeira, no pagamento e na obtenção das organizações militares por ele apoiadas, a fim de colaborar para a prestimosidade dos meios navais, aeronavais e de fuzileiros navais, sediados ou em trânsito na sua área de jurisdição, bem como dos estabelecimentos de terra por ele apoiados. Este grupo conta com os seguintes Centros de Intendências da Marinha (CeIM): Centro de Intendência da Marinha em Ladário (CeIMLa), Centro de Intendência da Marinha em Manaus (CeIMMa), Centro de Intendência da Marinha em Natal (CeIMNa), Centro de Intendência da Marinha em Rio Grande (CeIMRG), Centro de Intendência da Marinha em São Pedro da Aldeia (CeIMSPA), Centro de Intendência da Marinha em Salvador (CeIMSa) e Centro de Intendência da Marinha em Belém (CeIMBe). Este grupo movimentou no ano de 2019 o valor de R\$ 959.503.711,57 através de 15.945 notas de empenhos.

As Escolas de Aprendizes-Marinheiros são organizações militares de ensino, que têm como missão garantir ao discente o preparo mental, corpóreo, intelectual, moral e militar-naval, a fim de diplomar Marinheiros para o Corpo de Praças da Armada. Este grupo é composto pela Escola de Aprendizes-Marinheiros de Santa Catarina (EAMSC), Escola de Aprendizes-Marinheiros de Pernambuco (EAMPE), Escola de Aprendizes-Marinheiros do Ceará (EAMCE) e Escola de Aprendizes-Marinheiros do Espírito Santo (EAMES). Este grupo movimentou no ano de 2019 o valor de R\$ 594.604.805,02 através de 4.973 notas de empenhos.

As Bases e Estações Navais são organizações militares que tem a missão de prestar apoio logístico fixo às Forças e unidades navais da Marinha do Brasil que operem nas águas marítima, fluvial ou ribeirinha, sob sua área de jurisdição, a fim de contribuir para a defesa da navegação de interesse nacional e para o controle dessas áreas. Este grupo conta com as seguintes Bases e Estações Navais: Base Naval de Aratu (BNA), Base Naval de Natal (BNN), Base Naval de Val-de-Cães (BNVC), Base Naval do Rio Grande (BNRG), Base Fluvial de Ladário (BFLa), Estação Naval do Rio Negro (ENRN), Base Aérea Naval de São Pedro da Aldeia (BAeNSPA) e Base Naval do Rio de Janeiro (BNRJ). Este grupo movimentou no ano de 2019 o valor de R\$ 1.079.389.500,42 através de 7.831 notas de empenhos.

Os Hospitais Navais são organizações militares que tem a missão de prover assistência médico-hospitalar e odontológica, aos usuários do Sistema de Saúde da Marinha, além de executar medicina operativa e parcela da medicina pericial na sua área de jurisdição. Este grupo conta com os seguintes Hospitais Navais: Hospital Central da Marinha (HCM), Hospital Naval Marcílio Dias (HNMD), Hospital Naval de Salvador (HNSa), Hospital Naval

de Natal (HNNa), Hospital Naval de Recife (HNRe), Hospital Naval de Belém (HNBe), Hospital Naval de Ladário (HNLa) e Hospital Naval de Brasília (HNBra). Este grupo movimentou no ano de 2019 o valor de R\$ 4.088.363.928,91 através de 17.445 notas de empenhos.

De maneira resumida, a Tabela 9 explicita os subgrupos estudados nesta dissertação, bem como as organizações militares que compõem a amostra dos respectivos grupos.

Tabela 9 – Grupos estudados e respectivas OMs

Centros de Intendência da Marinha (CeIM)	Centro de Intendência da Marinha em Ladário
	Centro de Intendência da Marinha em Manaus
	Centro de Intendência da Marinha em Natal
	Centro de Intendência da Marinha em Rio Grande
	Centro de Intendência da Marinha em São Pedro da Aldeia
	Centro de Intendência da Marinha em Salvador
	Centro de Intendência da Marinha em Belém
Escolas de Aprendizes- Marinheiros (EAM)	Escola de Aprendizes-Marinheiros de Santa Catarina
	Escola de Aprendizes-Marinheiros de Pernambuco
	Escola de Aprendizes-Marinheiros do Ceará
	Escola de Aprendizes-Marinheiros do Espírito Santo
Bases e Estações Navais (BE)	Base Naval de Aratu
	Base Naval de Natal
	Base Naval de Val-de-Cães
	Base Naval do Rio Grande
	Base Naval de Ladário
	Estação Naval do Rio Negro
	Base Aérea Naval de São Pedro da Aldeia
	Base Naval do Rio de Janeiro
Hospitais Navais (HN)	Hospital Central da Marinha
	Hospital Naval Marcílio Dias
	Hospital Naval de Salvador
	Hospital Naval de Natal
	Hospital Naval de Recife
	Hospital Naval de Belém
	Hospital Naval de Ladário
	Hospital Naval de Brasília

Fonte: O autor, 2021.

No total as 27 organizações militares estudadas emitiram 46.194 notas de empenhos e movimentaram um volume financeiro de R\$ 6.721.861.945,92

2.3 Coleta de dados

Como fonte de pesquisa, o presente trabalho analisou as notas de empenhos emitidas por 27 UGE da MB, divididas em quatro grupos, no ano de 2019. As notas de empenhos, emitidas por cada UGE, foram obtidas através do uso do Tesouro Gerencial - ferramenta disponibilizada pela Secretaria do Tesouro Nacional, que possibilita a consulta de informações orçamentárias e financeiras do SIAFI (Sistema Integrado de Administração Financeira).

A consulta, no Tesouro Gerencial, foi efetuada da seguinte forma:

- No filtro do relatório, foram inseridos os filtros item de informação, UG executora, ano de lançamento e despesa empenhada;
- No campo item de informação, foi escolhida a função Despesa Empenhada (Controle de Empenho);
- No campo UG executora foi inserido os números das UGE (presentes na Tabela 10) escolhidas para análise de cada grupo;
- No campo ano de lançamento foi introduzido o ano de interesse que no presente trabalho é 2019;
- No campo filtro de exibição, foi selecionado o saldo da nota de empenho maior que zero; e
- A paginação escolhida foi por UG Executora, para cada aba do arquivo exportado em Excel representar os empenhos emitidos por uma UGE.

Os dados escolhidos para estar representados nas colunas do relatório emitido pelo Tesouro Gerencial foram: data de emissão, UG responsável, nota de empenho, modalidade de licitação, tipo da nota de empenho, favorecido, ND detalhada e saldo.

Na presente pesquisa dar-se-á maior enfoque aos campos data de emissão e saldo.

Para melhor ilustrar a pesquisa realizada no Tesouro Gerencial, a consulta realizada para obter o relatório, utilizado para analisar as notas de empenhos dos Centros de Intendência da Marinha, está representada na Figura 4.

Figura 4 – Consulta Tesouro Gerencial

FILTRO DO RELATÓRIO X Limpar ? x

Item Informação Na Lista (DESPESAS EMPENHADAS (CONTROLE EMPENHO)) ↓
 E ↕

UG Executora Na Lista (784810:CENTRO DE INTENDENCIA DA MARINHA EM BELEM, 788820:CENTRO DE INTENDENCIA DA MARINHA EM MANAUS, 791181:CENTRO DE INTEND. DA MARINHA EM S.PEDRO DA AL., 785810:CENTRO DE INTENDENCIA DA MARINHA EM RIO GRAND, 783810:CENTRO DE INTENDENCIA DA MARINHA EM NATAL, 782802:CENTRO DE INTENDENCIA DA MARINHA EM SALVADOR, 786810:CENTRO DE INTENDENCIA DA MARINHA EM LADARIO) ↑ ↓
 E ↕

Ano Lançamento Na Lista (2019) ↑

FILTRO DE EXIBIÇÃO ? x

Saldo - R\$ (Item Informação) Maior do que 0

PAGINAR POR: Q ?

Dia Emissão NE CCor	UG Responsável	Nota Empenho CCor	Modalidade Licitação NE CCor	NE - Tipo	Favorecido NE CCor	Natureza Despesa Detalhada	Item Informação	<Item Informação>
<Dia Emissão NE CCor>	<UG Responsável>	<Nota Empenho CCor>	<Modalidade Licitação NE CCor>	<NE - Tipo>	<Favorecido NE CCor>	<Natureza Despesa Detalhada>		<Saldo - R\$ (Item Informação)>

Fonte: O autor, 2021.

O código de cada organização militar presente na amostra estudada se encontra na Tabela 10.

Tabela 10 – Código de cada OM utilizada no Tesouro Gerencial

CeIM		EAM	
CeIMLa	786810	EAMSC	785600
CeIMMa	788820		
CeIMNa	783810	EAMPE	783602
CeIMRG	785810		
CeIMSPA	791181	EAMCE	783601
CeIMSa	782802		
CeIMBe	784810	EAMES	781600
BE		HN	
BNA	782801	HCM	765701
BNN	783800	HNMD	765720
BNVC	784800	HNSa	782700
BNRG	785800	HNNa	783701
BFLa	786800	HNRe	783702
ENRN	788810	HNBe	784700
BAeNSPA	791180	HNLa	786700
BNRJ	791800	HNBra	787700

Fonte: O autor, 2021.

As consultas realizadas e utilizadas para esta dissertação podem ser consultadas no Tesouro Gerencial através do seguinte caminho: Relatórios compartilhados → Consultas gerenciais → Relatórios de bancada dos órgãos superiores → 52000 Ministério da Defesa → 52131 Comando da Marinha → UERJ dissertação.

Na fase de tratamento dos dados, preferiu-se omitir os nomes das unidades gestoras executoras, tendo em vista que a identificação do nome ao lado da análise de dados pode ser interpretada como uma injusta acusação de fraude. Vale salientar que o fato de não apresentar conformidade com a lei Newcomb-Benford, segundo Kossovky (2014), pode ser devido a quatro motivos e não necessariamente é um sinal incontestável de fraude, Os motivos estão descritos a seguir.

- Falso positivo, uma amostra honesta que, por uma rara chance aleatória estatística, desvia da lei;
- Os tipos de dados que por uma maneira particular não seguem a lei, porém, são honestos;
- A organização de que os dados foram retirados segue uma configuração de negócio particular, que causa o desvio da lei; e
- Os dados foram fraudados e a organização é desonesta.

2.4 Tratamento de dados

Nesta seção estão presentes todos os procedimentos adotados no Tesouro Gerencial para importação dos dados, bem como a programação envolvida no *software* R (R CORE TEAM, 2019) para o tratamento dos dados e a realização das análises pré-Benford, análises primárias, análises avançadas, análises associadas, análises de conformidade e também dos testes estatísticos.

A linguagem de programação evidenciada nas figuras desta seção tem como base os dados de um Centro de Intendência da Marinha que foi chamado de “CeIM1”. Esta conduta foi assim escolhida, em vista da linguagem de programação ser a mesma para os quatro grupos, - a única diferença são os nomes das variáveis escolhidas, dos arquivos Excel importados e das suas abas.

2.4.1 Importação e edição dos dados

O primeiro passo para tratar os dados é exportá-los do Tesouro Gerencial, em formato Excel, para posterior análise no *software* R. Para isto, é necessário clicar no botão “exportar” e escolher a opção “Excel com formatação”. Após o *download* ser concluído, os quatro arquivos foram salvos na pasta “Dados”, localizada no *desktop* do computador, com os seguintes nomes: CeIM (grupo dos Centros de Intendência da Marinha), EAM (grupo das Escolas de Aprendizes-Marinheiros), BE (grupo das Bases e Estações Navais) e HN (grupo dos Hospitais Navais).

Em cada arquivo, para sua melhor manipulação, foi alterado o nome de cada aba, dos quatro arquivos Excel utilizados, de “Saldo - R\$ (Item Informação)_#” para a sigla do grupo contida na Tabela 9 e um numeral (para os CeIM varia de 1 a 7, para as EAM varia de 1 a 4 e para as BE e HN varia de 1 a 8) – no caso da primeira aba da tabela dos Centros de Intendência da Marinha “CeIM1”.

Outro procedimento adotado, antes de iniciar a análise de dados em *software* estatístico, foi a exclusão de todos os cabeçalhos, deixando apenas a tabela foco do estudo.

Por último, foram alterados os itens do cabeçalho da tabela extraída do Tesouro Gerencial pelos seguintes itens: Data, Ano, Codigo UGR, Nome UGR, Numero NE, Codigo Modalidade Licitação, Modalidade Licitação, Codigo Tipo de Empenho, Tipo de Empenho, CNPJ Favorecido, Nome Favorecido, Natureza de Despesa, Descrição Natureza de Despesa e Saldo. A acentuação, bem como o “ç”, não foram utilizados no cabeçalho, para melhor leitura da tabela pelo *software* estatístico utilizado.

O *software* escolhido para realizar o tratamento de dados foi o R versão 3.6.1, em razão, primeiramente, de se tratar de um *software* gratuito, que qualquer organização militar ou empresa pode utilizar sem acarretar custos adicionais. Sua linguagem de programação é simples e ainda possui compatibilidade com diferentes linguagens e sistemas operacionais. Além de ser estável e com suporte contínuo com lançamentos de diversos novos pacotes, outro ponto importante é o R ser uma plataforma amplamente utilizada, no Brasil e no mundo, por pesquisadores e pela indústria (R CORE TEAM, 2019). Alguns desses pacotes tratam especificamente da lei Newcomb-Benford e foram utilizados no tratamento de dados desta dissertação.

Para iniciar o tratamento dos dados, foi necessário importar a base de dados extraída do Tesouro Gerencial para o R. Já que a base de dados está em formato Excel, foi necessário

utilizar do pacote “readxl”. A linguagem de programação utilizada para realizar a importação está descrita na Figura 5 no Apêndice A.

A primeira linha se refere à instalação do pacote do R, e a segunda à inicialização do pacote. A última linha se refere à importação da primeira aba do arquivo Excel “CeIM” para sua respectiva variável CeIM1.

Devido à mesclagem das células data, algumas linhas das variáveis se encontravam com a notação NA, que significa *not available* ou dados omissos. Para reparar este problema, foi utilizada a linguagem de programação presente na Figura 6 no Apêndice A, em todas as variáveis, para substituir os valores “NA” pela data correspondente.

A próxima seção aborda como foi realizada a programação para realização da análise pré-Benford no *software* R.

2.4.2 Aplicação das análises pré-Benford

Nesta seção, se encontra a programação utilizada no *software* R para aplicar os procedimentos das análises pré-Benford denominados perfil de dados, histograma de dados e gráficos periódicos, já descritos na seção 1.3.1.1.

Conforme já salientado a linguagem de programação evidenciada nesta seção tem como base os dados de um Centro de Intendência da Marinha que foi chamado de “CeIM1”.

De acordo com Nigrini (2020), o primeiro procedimento é a realização da análise pré-Benford. O primeiro procedimento utilizado desta análise foi o perfil de dados, que consiste em dividir os dados nas seguintes cinco categorias: grandes e positivos, pequenos e positivos, zeros, pequenos e negativos, e grandes e negativos. Como os saldos das notas de empenhos são sempre maiores que zero, foi estabelecida, nesta dissertação, a seguinte divisão de categorias: números com um algarismo à esquerda da vírgula (a), números com dois algarismos à esquerda da vírgula (b), números com três algarismos à esquerda da vírgula (c), números com quatro algarismos à esquerda da vírgula (d) e números com cinco ou mais algarismos à esquerda da vírgula (e). Para realização deste procedimento, no *software* R, foram utilizados os comandos contidos na Figura 7 do Apêndice A.

O segundo procedimento da análise pré-Benford é a criação de histogramas. Como as amostras possuem entre 1.000 e 5.000 dados, Nigrini (2020) recomenda que os histogramas sejam construídos com 10 intervalos. Ele reitera que, para dados financeiros, 70% dos dados

costumam estar primeiro no primeiro intervalo e 20% no segundo intervalo, pois eles habitualmente possuem diversos valores pequenos e poucos grandes. Para obter os histogramas dos saldos das notas de empenhos, foram utilizados os comandos evidenciados na Figura 8 do Apêndice A.

O terceiro procedimento da análise pré-Benford consiste em ilustrar a quantidade empenhada, no decorrer do tempo, por cada CeIM. Nigrini (2020) frisa que este procedimento é profícuo para detectar anomalias, através de despesas fora do normal. Para este fim, foram gerados diversos gráficos com o auxílio do *software* R, tendo como eixo das abscissas os meses do ano e como eixo das ordenadas o total empenhado no respectivo mês. Com esta finalidade foram utilizados os comandos presentes na Figura 9 do Apêndice A.

Após a realização das análises pré-Benford, Nigrini (2012) e Nigrini (2020) sugerem a aplicação das análises primárias.

2.4.3 Aplicação das análises primárias

Nesta seção, estão os comandos utilizados no *software* R para aplicação dos procedimentos das análises primárias, já descritos na seção 1.3.1.2. As análises realizadas foram a dos primeiros dígitos, a dos segundos dígitos e a dos dois primeiros dígitos. A linguagem de programação evidenciada nas figuras desta seção tem como base os dados de um Centro de Intendência da Marinha que foi chamado de “CeIM1”.

Para aplicar as análises primárias nos saldos das notas de empenhos emitidas, no ano de 2019, pelos Centros de Intendência da Marinha, com auxílio do *software* R, o primeiro passo é instalar e iniciar os pacotes *benford.analysis* e *BenfordTests*, através dos comandos encontrados na Figura 10 do Apêndice A.

Após a instalação e inicialização dos pacotes anteriormente mencionados, realizou-se, primeiramente, a verificação de conformidade dos primeiros dígitos. Para esta verificação foram utilizados os testes qui quadrado (χ^2), Kolmogorov-Smirnov (K-S) e o procedimento do desvio absoluto médio (*Mean Absolute Deviation* – MAD). Todos eles têm, como hipóteses nula e alternativa, os seguintes pressupostos:

H_0 : A amostra segue a lei Newcomb-Benford.

H_1 : A amostra não segue a lei Newcomb-Benford.

Obtiveram-se os valores das estatísticas teste para os testes χ^2 , K-S e o procedimento MAD, através do *software* R, por meio do seguinte programa explicitado na Figura 11 do Apêndice A.

Já para uma análise local, foi aplicado o teste Z com um nível de significância de 5% para cada primeiro dígito da amostra, com a finalidade de investigar o desvio da probabilidade observada das amostras de cada algarismo, em relação a probabilidade esperada por Newcomb (1881) e Benford (1938). Através do código descrito na Figura 12 do Apêndice A, foi obtido o p-valor do teste Z para cada primeiro dígito ao nível de significância de 5%.

Já para análise dos segundos dígitos, foi percebido que o *software* R não contém pacotes, nem funções específicas para este fim. Assim, em virtude disto, este trabalho fez esta análise global através dos testes qui quadrado, Kolmogorov-Smirnov e do procedimento MAD.

Para obter os valores do χ^2_{teste} , conforme Equação 30, do $K-S_{\text{teste}}$, conforme Equação 32, do $K-S_{\text{crítico}}$ para o nível de significância de 5%, conforme Tabela 8, e do MAD_{teste} , conforme Equação 17, foi utilizado o seguinte código no R presente na Figura 13 do Apêndice A.

Já para o teste local dos segundos dígitos das notas de empenhos, foi utilizado o teste Z ao nível de significância de 5%. Os valores do Z_{teste} para cada segundo dígito, conforme Equação 27, foram obtidos através da aplicação da linha de comando inserida no *software* R e descrita na Figura 14 do Apêndice A.

Para a análise dos dois primeiros dígitos foram utilizados os seguintes procedimentos para avaliar a conformidade com a LNB: teste qui quadrado, teste Kolmogorov-Smirnov, o desvio absoluto médio e a *Mantissa Arc Test*.

Obteve-se a estatística teste e p-valor dos testes para os dois primeiros dígitos dos testes χ^2 , K-S, MAD e *Mantissa Arc Test*, por meio do *software* R, mediante o programa presente na Figura 15 do Apêndice A.

Para verificar a conformidade dos dois primeiros dígitos com a lei Newcomb- Benford de forma local foi aplicado o teste Z. Para aplicar este teste estatístico com o auxílio do R foi utilizado o comando explicitado na Figura 16 do Apêndice A.

Na próxima seção foi abordada a metodologia utilizada para aplicação das análises avançadas proposta por Nigrini (2012) e Nigrini (2020) com o auxílio do *software* R.

2.4.4 Aplicação das análises avançadas

Nesta seção, apresenta-se a linguagem de programação utilizada no *software* R para utilização das análises avançadas proposta por Nigrini (2012) e Nigrini (2020), já descrita na seção 1.3.1.3. Os procedimentos utilizados foram a análise de segunda ordem presente na seção 1.3.1.3.1 desta dissertação e a análise da soma presente na seção 1.3.1.3.2. A linguagem de programação evidenciada nas figuras desta seção tem como base os dados de um Centro de Intendência da Marinha que foi chamado de “CeIM1”.

Foi utilizado como procedimento de verificação de conformidade para a análise de segunda ordem o MAD, pois, como Nigrini (2020) alerta, este deve ser o único procedimento utilizado para verificar conformidade com a LNB neste caso.

Para fazer realizar a análise de segunda ordem e verificar sua conformidade em relação aos primeiros dígitos com a lei Newcomb-Benford através do desvio absoluto médio foi utilizada a linguagem de programação contida na Figura 17 do Apêndice A.

Já para realizar a análise de segunda ordem nos dois primeiros dígitos da amostra e avaliar sua conformidade, através do desvio absoluto médio, utilizou-se os seguintes comandos presentes na Figura 18 do Apêndice A.

Para a análise da soma, foi levado em consideração os primeiros e os dois primeiros dígitos. Como explicitado na seção 1.3.1.3.2 e explicitado pela Equação 7, os valores das somas dos números do primeiro dígito ou dois primeiros dígitos se distribuem uniformemente. Assim, pode-se afirmar que o valor esperado pela soma dos números que possuem como primeiro algarismo o número 1 tende a ser $1/9$ da soma de todos os valores, da mesma forma, o valor esperado pela soma dos números que possuem como primeiros algarismos o número 51 tende a ser $1/90$ da soma de todos os valores.

Para realizar esta análise no R e obter as informações do teste da soma para os primeiros dígitos, foram utilizados os comandos explicitados na Figura 19 do Apêndice A.

Seguindo a mesma linha, a análise da soma, englobando os dois primeiros dígitos, foi realizada mediante a utilização a linguagem de programação exposta na Figura 20 do Apêndice A.

Na próxima seção é abordada a metodologia utilizada nas análises associadas, propostas por Nigrini (2012) e Nigrini (2020), com o auxílio do *software* R.

2.4.5 Aplicação das análises associadas

Nesta seção, são aplicados os procedimentos das análises associadas, já descritos na seção 1.3.1.4, com exceção da análise dos dois últimos dígitos, pois, conforme Kossovsky (2014), esta análise não é apropriada para analisar despesas. Assim, foram realizadas a análise dos números duplicados e o modelo fator de distorção. A linguagem de programação evidenciada nas figuras desta seção tem como base os dados de um Centro de Intendência da Marinha que foi chamado de “CeIM1”.

Para a realização do teste dos números duplicados, foi aplicada a linguagem de programação no *software* R contida na Figura 21 do Apêndice A.

O outro procedimento da análise associada utilizado para análise das notas de empenhos emitidas foi o modelo fator de distorção, abordado nesta dissertação na seção 1.3.1.4.3. Para obter os valores do FD (Equação 15) e do Z_{teste} (Equação 16), foi utilizada a linguagem de programação demonstrada na Figura 22 do Apêndice A.

O *software* R, através do pacote *benford.analysis*, permite encontrar o valor do fator de distorção, por meio da função *dfactor()*, porém, não fornece o valor do Z_{teste} e nem o número de dados analisados “n”, que é necessário para o cálculo do Z_{teste} (Equação 16). Devido a estes motivos, foi utilizada a linguagem de programação já explicitada ao invés da função *dfactor()*.

2.5 **Delimitação da pesquisa**

Esta dissertação aplicou a análise de dados à luz da lei Newcomb-Benford, com o auxílio do *software* estatístico R, para verificar sua conformidade com as notas de empenhos emitidas em 2019 por 27 organizações militares.

Ressalta-se que os resultados encontrados neste estudo limitam-se às 27 organizações estudadas, no ano de 2019, e não podem ser generalizados para outras organizações militares da Marinha do Brasil. Outro fato a ser evidenciado é que esta dissertação apenas levantou se a amostra segue a lei Newcomb-Benford, o viés de fraude ou anomalias, só poderá ser confirmado após auditoria “*in loco*” pelo Centro de Controle Interno da Marinha.

Este trabalho teve como intuito avaliar se a distribuição das notas de empenhos emitidas por organizações militares da Marinha do Brasil segue a LNB. Conforme já mencionado a não conformidade pode indicar um viés de fraude, porém, apenas os resultados de uma auditoria podem ratificar este viés de fraude.

Foi escolhido o ano de 2019 para análise, pois durante a confecção desta dissertação era o único exercício financeiro completo em que os dados estavam disponíveis para análise. O ano de 2019 foi o primeiro em que as despesas públicas foram realizada à luz do Decreto nº 9.412, de 18 de junho de 2018, que atualizou os valores das aquisições de bens e contratações de serviços (exceto os de engenharia), sem licitação, para R\$ 17.600,00. Nos anos anteriores o limite era de R\$8.000,00.

A próxima seção apresenta a análise e discussão dos resultados obtidos através dos procedimentos explicitados na seção 2.4. Estes procedimentos serão aplicados nas notas de empenhos emitidas pelos Centros de Intendência da Marinha, pelas Escolas de Aprendizes-Marinheiros, pelas Bases e Estações Navais e pelos Hospitais Navais no ano de 2019.

3 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Nesta seção, serão descritos os resultados obtidos após a preparação dos dados e a realização das análises pré-Benford, primárias, avançadas e associadas, descritas na seção 2.4 desta dissertação. Esta seção é dividida em quatro subseções: A primeira seção contém as análises pré-Benford. Na segunda seção são executadas as análises primárias. Na seção subsequente contém as análises avançadas dos dados. Na última seção são realizadas as análises associadas. Em todas as seções são analisados os quatro grupos (CeIM, EAM, BE e HN).

3.1 Análises pré-Benford dos dados

Primeiramente foi aplicada a análise pré-Benford denominada perfil de dados nas notas de empenhos emitidas pelos Centros de Intendência da Marinha, Escolas de Aprendizizes-Marinheiros, Bases e Estações Navais e Hospitais Navais no ano de 2019, através da linguagem de programação explicitada na Figura 7 do Apêndice A. Os resultados obtidos para o perfil de dados estão presentes nas Tabelas 11, 12, 13 e 14 presentes no Apêndice B.

A quantidade de dados examinados é importante, pois, conforme explicitado por Nigrini (2020) e Nigrini (2012), quanto maior a quantidade de dados, maior será a confiabilidade dos procedimentos estatísticos e não estatísticos adotados, no que tange à conformidade com a lei Newcomb-Benford. Outro fato salientado por ele é que as amostras, para serem submetidas às análises que avaliam sua conformidade com a LNB, devem possuir no mínimo 1.000 elementos e nunca menos de 300 dados. Assim, analisando as Tabelas 11, 12, 13 e 14 do Apêndice B foi decidido, manter na amostra as BE2, BE4, BE5, BE6 e BE7 que emitiram respectivamente 874, 817, 753, 369 e 889 notas de empenhos, mesmo sendo amostras com menos de 1000 dados, que segundo Nigrini (2012) e Nigrini (2020) podem apresentar viés de não conformidade com a LNB.

Também foi decidido manter na amostra o HN1 mesmo possuindo apenas 257 notas de empenhos emitidas no ano de 2019, número inferior a 300, fato que Nigrini (2012) e Nigrini (2020) alertam que amostras pequenas não devem ser analisadas levando em consideração a LNB.

Outra análise realizada foi o perfil de dados. Nela foi avaliada a quantidade de valores pequenos que a amostra possui, pois, para Nigrini (2020), uma quantidade de dados pequenos superior a 15% pode dar um viés de não conformidade com a LNB. Na mesma linha, Nigrini (2012) reitera que suas pesquisas apontaram que, para a obtenção de melhor conformidade, com a LNB, os dados devem possuir quatro algarismos ou mais. Assim, nesta pesquisa, os empenhos com saldo inferior a R\$1.000,00 foram considerados dados pequenos. Neste contexto, as amostras CeIM3, CeIM5, CeIM6, CeIM7, EAM3, BE2, BE3 e BE4 podem apresentar viés de não conformidade, por possuírem, respectivamente, 22%, 17,5%, 18,2%, 17,7%, 28,14%, 16,13%, 21,02% e 16,77% de valores considerados pequenos – menores do que R\$1.000,00.

Com relação ao histograma de dados foi utilizada a linguagem de programação presente na Figura 8 do Apêndice A e através dela foram obtidas as Figuras 23, 24 e 25 (Histograma dos CeIM), Figuras 26 e 27 (Histograma das EAM), Figuras 28, 29, 30 e 31 (Histograma das BE) e as Figuras 32, 33, 34 e 35 (Histograma dos HN) todas presentes no Apêndice B.

Todos os histogramas das 27 OMs, representados nas Figuras 23 a 35 presentes no Apêndice B, explicitam que quase a totalidade dos dados se encontra no primeiro intervalo. Isto ocorreu devido à amplitude dos saldos das notas de empenhos variarem de valores pequenos até valores muito grandes. Esta afirmativa foi confirmada através da utilização do comando `summary(CeIM1$Saldo)` no *software* R. Seus resultados estão relatados na Tabela 15 do Apêndice B.

A Tabela 15 do Apêndice B deixa evidente a presença de *outliers* em todas as 27 organizações militares estudadas, visto que a distância entre o 3º Quartil e o máximo de cada amostra é maior que 1,5 vezes a distância interquartilica (distância entre o 1º Quartil e o 3º Quartil).

Este fato ratifica os resultados obtidos pelos histogramas explicitados pelas Figuras 23 a 35, nas quais se observa que quase a totalidade dos dados se localizou na primeira classe do seu respectivo histograma. Com isso, pode-se dizer que as 27 amostras das notas de empenhos seguiram um padrão no segundo procedimento da análise pré-Benford. Isto vai de encontro com o abordado por Nigrini (2012) e Nigrini (2020) que afirmam que dados financeiros costumam se concentrar no primeiro intervalo do histograma de dados.

Já para a confecção do gráfico temporal foi utilizada o programa representado na Figura 9 do Apêndice A. Após a aplicação do programa nos dados das 27 organizações militares, obteve-se como resultado os gráficos temporais representados pelas Figuras 36, 37 e

38 (CeIM), Figuras 39 e 40 (EAM), Figuras 41, 42, 43 e 44 (BE) e Figuras 45, 46, 47 e 48 (BE) que representam o total empenhado mensalmente, em milhares de reais, por cada organização militar. Já nas Figuras 49, 50, 51 e 52 estão os gráficos temporais das organizações militares individualmente e também divididas por grupo.

Através da análise das Figuras 36 a 52, observou-se uma espécie de padrão em todos os gráficos. Foi detectado que fevereiro, julho e dezembro são os meses com maiores volumes empenhados no ano em todas as amostras e os CeIM2, CeIM4, CeIM7, EAM2, BE8 e HN2 são as amostras que movimentam maiores volumes financeiros dentro de seus grupos.

Após a realização das análises pré-Benford foi decidido manter todas as amostras para as análises subsequentes, porém, levantados alguns motivos que podem explicar uma possível não conformidade com a lei Newcomb-Benford.

3.2 Análises primárias dos dados

Após a realização das análises pré-Benford, foram realizados os procedimentos das análises primárias. O primeiro procedimento realizado foi a verificação da conformidade das amostras, de forma global, com a lei Newcomb-Benford levando em consideração apenas os primeiros dígitos. Para isso foi utilizado o teste qui quadrado, o teste K-S e o procedimento MAD. Estes procedimentos foram realizados no R com os comandos evidenciados na Figura 11 do Apêndice A e seus resultados presentes na Tabela 16 do Apêndice C.

Para analisar a conformidade da amostra utilizando o teste qui quadrado, é necessário obter o valor do $\chi^2_{\text{crítico}}$ para um nível de significância de 5% e grau de liberdade 8. Este valor foi obtido através do comando `qchisq(0.95,8)` e o resultado obtido foi 15,507. Assim, analisando a Tabela 16 do Apêndice C conclui-se que as NE das EAM3, EAM4, BE4, BE6, HN1 e HN2 não estão em conformidade com a LNB com relação aos primeiros dígitos. Isto porque nestas ocasiões $\chi^2_{\text{teste}} > \chi^2_{\text{crítico}}$, rejeitando assim H_0 e aceitando H_1 . Já para as demais amostras foi obtido na Tabela 16 do Apêndice C, $\chi^2_{\text{teste}} < \chi^2_{\text{crítico}}$, aceitando assim a H_0 e concluindo que estas NE seguem a lei Newcomb-Benford, quando analisados os primeiros dígitos. Na Tabela 16 do Apêndice C as amostras que seguem a LNB para o teste qui quadrado estão com o fundo verde, já as que não seguem estão com o fundo em vermelho.

Com relação aos valores do MAD presentes na Tabela 16 do Apêndice C, foi comparado este valor com os explicitados por Nigrini (2012), na Tabela 6 (utilizando as

mesmas cores da conclusão desta Tabela) para interpretar os resultados. Assim, conclui-se, que apenas as notas de empenhos emitidas pelas EAM3, BE6 e HN1 não possuem conformidade com a lei Newcomb-Benford, de forma global, quando analisados os primeiros dígitos.

Para o teste Kolmogorov-Smirnov, o *software* R considera para cálculo do $K-S_{\text{teste}}$ a Equação 34, sendo “n” o número de dados da amostra testada.

$$K_1 = \max_i |\sum (f_{1j} - B_{1j})| \sqrt{n} \quad i, j = 1, 2, \dots, 9 \quad (34)$$

Destarte, ao invés de utilizar como $K-S_{\text{crítico}}$ para um nível de significância de 5% o valor contido na Tabela 8 de $1,36/\sqrt{n}$, utilizou-se o valor de 1,36 como $K-S_{\text{crítico}}$.

Diante do exposto, foi verificado que, para o teste Kolmogorov-Smirnov, todas as amostras obtiveram $K-S_{\text{teste}} < K-S_{\text{crítico}}$ com exceção da EAM3 em que $K-S_{\text{teste}} > K-S_{\text{crítico}}$. De tal modo, conclui-se que para este teste todas as amostras com exceção da EAM3 seguem a lei Newcomb-Benford, de maneira global, quando analisados os primeiros dígitos.

Já para análise dos primeiros dígitos de maneira local foi utilizado o teste Z ao nível de significância de 5% através do código descrito na Figura 12 do Apêndice A. Através deste código foi obtido o p-valor do teste Z de cada primeiro dígito. Estes valores estão evidenciados nas Tabelas 17, 18, 19 e 20 do Apêndice C representando respectivamente os CeIM, EAM, BE e HN..

Após analisar as Tabela 17, 18, 19 e 20, notou-se que os algarismos com o fundo em vermelho, obtiveram no teste Z p-valor < nível de significância. Assim, para esses algarismos, rejeita-se H_0 concluindo assim que a distribuição observada desses dígitos não está em conformidade com a LNB.

Já para analisar a distribuição dos segundos dígitos das notas de empenhos foi utilizada a linguagem de programação presente na Figura 13 do Apêndice A e os resultados dos testes qui quadrado e K-S, bem como o valor do K-S crítico e do MAD estão presentes na Tabela 21 do Apêndice C.

O valor do $\chi^2_{\text{crítico}}$ para o nível de significância de 5% e grau de liberdade 9 foi obtido pelo comando `qchisq(0.95,9)` no *software* R e retornou como valor 16,92. Assim os CeIM1, CeIM2 e CeIM5, as EAM1, EAM2 e EAM3, a BE1 e os HN6, HN7 e HN8 possuem $\chi^2_{\text{teste}} > \chi^2_{\text{crítico}}$, rejeitando desta forma H_0 e concluindo que as notas de empenhos destas amostras não estão em conformidade, de forma global, com a lei Newcomb-Benford, quando analisados os

segundos dígitos. Por outro lado, as outras amostras têm $\chi^2_{\text{teste}} < \chi^2_{\text{crítico}}$, aceitando assim H_0 e deduzindo que as NE destas amostras estão em conformidade, de forma global, com a LNB, quando analisados os segundos dígitos. Ou seja, as organizações militares em que as NE seguiram a LNB para os segundos dígitos estão com fundo verde, já as que não seguiram estão com o fundo vermelho na Tabela 21 do Apêndice C.

Para o MAD, utilizando como base os dados de Nigrini (2012) presentes na Tabela 6, conclui-se que apenas as NE do CeIM5 e das EAM2 e EAM3 não apresentam conformidade com a LNB. As cores utilizadas para colorir o fundo do resultado do $\text{MAD}_{\text{teste}}$ está em linha com as cores utilizadas na coluna “conclusão” da Tabela 6.

Já o teste K-S denota que apenas as notas de empenhos emitidas pelos CeIM5, EAM3, BE1 e HN6, no ano de 2019, não apresentam conformidade com a distribuição de Newcomb (1881) e Benford (1938) quando analisados os segundos dígitos, pois $K\text{-Steste} > K\text{-Scrítico}$. As demais organizações militares apresentaram $K\text{-Steste} < K\text{-Scrítico}$ denotando assim, que a distribuição dos segundos dígitos das notas de empenhos emitidas em 2019, está em conformidade com a lei Newcomb-Benford. Assim os resultados em que o teste K-S explicita conformidade com a LNB estão com fundo verde e os que não explicitam conformidade estão com o fundo vermelho.

Já para análise dos segundos dígitos das notas de empenhos de forma local foi utilizada a programação denotada na Figura 14. Através dela foi aplicado o teste Z com nível de significância de 5% e obtidos como resultado os valores Z_{teste} para cada segundo dígito das 27 amostras estudadas. Estes valores estão presentes nas Tabelas 22, 23, 24 e 25 do Apêndice C.

Com relação o teste Z, calcula-se o valor do $Z_{\text{crítico}}$ para o nível de significância de 5% através do comando `qnorm(0.975)` e obtém-se o valor de 1,96. Após análise das Tabelas 22, 23, 24 e 25 do Apêndice C, concluiu-se que os valores com o fundo vermelho apresentaram $|Z_{\text{teste}}| > Z_{\text{crítico}}$, rejeita-se assim H_0 e aceita-se H_1 . Com isso, pode-se afirmar que, para um nível de significância de 5%, estes dígitos não estão em conformidade com a LNB. Por outro lado, os valores com o fundo verde apresentaram $|Z_{\text{teste}}| < Z_{\text{crítico}}$, aceitando-se assim H_0 e concluindo que estes dígitos estão em conformidade com a LNB.

Para a verificação da conformidade dos dois primeiros dígitos com a lei Newcomb-Benford primeiramente foi aplicado o teste global através da programação descrita na Figura 15 do Apêndice A.

Com os resultados explicitados pelo *software* R, foi obtida a Tabela 26 do Apêndice C. Nela estão evidenciados os valores do χ^2_{teste} , do MAD, do K-S_{teste} e, também, o p-valor do *Mantissa Arc Test*.

Para analisar a conformidade das notas de empenhos e utilizar o teste qui quadrado, foi necessário obter o valor do $\chi^2_{\text{crítico}}$, para um nível de significância de 5% e grau de liberdade 89. Este valor foi auferido pelo comando `qchisq(0.95,89)` e obteve-se como resultado o valor 112,022. Deste modo, 12 amostras (fundo verde) apresentaram $\chi^2_{\text{teste}} < \chi^2_{\text{crítico}}$, aceita-se H_0 e conclui-se que, para o teste qui quadrado, estas amostras seguem a lei Newcomb-Benford, quando analisados os dois primeiros dígitos. Por outro lado, 15 organizações militares (fundo vermelho) apresentaram $\chi^2_{\text{teste}} > \chi^2_{\text{crítico}}$, rejeitando-se H_0 . Portanto, chega-se a conclusão que as notas de empenhos destas amostras, não seguem a LNB, quando analisados os dois primeiros dígitos.

Com relação ao MAD e comparando o valores da Tabela 6 explicitados por Nigrini (2012), conclui-se que cinco amostras apresentaram conformidade aceitável (fundo verde claro) e seis amostras apresentaram conformidade marginalmente aceitável (fundo amarelo) com a lei Newcomb-Benford, quando analisados os dois primeiros dígitos. No entanto, 16 organizações militares não apresentaram conformidade com a LNB, para os dois primeiros dígitos, quando analisados à luz do MAD.

Ao aplicar o teste Kolmogorov-Smirnov, o pacote utilizado para determinar a estatística teste utiliza a Equação 35 para calcular o K-S_{teste}.

$$K_{12} = \max_i |\sum (f_{12j} - B_{12j})| \sqrt{n} \quad i, j = 10, 11, 12, \dots, 99 \quad (35)$$

Assim, para um nível de significância de 5%, conforme Tabela 8, adota-se o valor de 1,36 como K-S_{crítico}. Com isso, percebe-se que 23 amostras (fundo verde), apresentaram K-S_{teste} < K-S_{crítico}, denotando que elas apresentaram conformidade, com a lei Newcomb-Benford, quando analisados, de forma global, os dois primeiros dígitos. Por outro lado, quatro amostras (fundo vermelho) apresentaram K-S_{teste} > K-S_{crítico}, rejeitando assim H_0 e concluindo que as notas de empenhos destas organizações militares não segue a lei Newcomb-Benford, quando analisados, de forma global, os dois primeiros dígitos.

O último item a ser analisado na Tabela 26 do Apêndice C é o resultado obtido no *Mantissa Arc Test*. Apenas cinco amostras (fundo verde) apresentaram p-valor < 0,05, concluindo assim que estas amostras seguiram a LNB. Já 22 amostras (fundo vermelho)

apresentaram $p\text{-valor} > 0,05$ denotando assim que elas não seguem a LNB para os dois primeiros dígitos de forma global, quando analisados à luz do *Mantissa Arc Test*.

Já para o estudo dos dois primeiros dígitos de forma local foi aplicado o teste Z com o nível de significância de 5% através do comando explicitado na Figura 16.

Com os dados alcançados através do comando anteriormente citado, foram obtidas as Figuras 53, 54 e 55 (CeIM), Figuras 56 e 57 (EAM), Figuras 58, 59, 60 e 61 (BE) e Figuras 62, 63, 64 e 65 (HN), todas no Apêndice C., que explicitam as frequências observadas, bem como os limites superior e inferior, calculados conforme as Equações 28 e 29 ao nível de significância de 5%.

Aferiu-se, portanto, que alguns dois primeiros dígitos não estão em conformidade com a LNB, por sua frequência observada estar fora da faixa delimitada pelo limite superior e pelo limite inferior do teste Z, para um nível de significância de 5%. Os dois primeiros dígitos que não apresentaram conformidade com a LNB, encontram-se evidenciados na Tabela 27 do Apêndice C.

3.3 Análises avançadas dos dados

Com relação às análises avançadas iniciou-se com a realização da análise de segunda ordem para os primeiros dígitos e dois primeiros dígitos utilizando para verificação de conformidade o MAD. Os procedimentos adotados no *software* R foram os descritos na Figura 17 do Apêndice A (análise de segunda ordem para o primeiro dígito) e Figura 18 do Apêndice A (análise de segunda ordem para os dois primeiros dígitos) e através deles foram obtidos os resultados descritos na Tabela 28 do Apêndice C.

Percebe-se, assim, que utilizando os valores propostos por Nigrini (2012) presentes na Tabela 6, foi percebido que na Tabela 28 do Apêndice C apenas uma organização militar não possui conformidade com a LNB para análise de segunda ordem quando analisados os primeiros dígitos.

Na mesma linha, porém com relação aos dois primeiros dígitos, apenas os HN3 e HN4 apresentaram uma conformidade marginalmente aceitável com a LNB, para a análise de segunda ordem, quando analisados com o MAD. As outras 25 amostras não apresentaram conformidade com a lei Newcomb-Benford para a análise de segunda ordem quando analisados os dois primeiros dígitos

Outra análise avançada aplicada foi a análise da soma. Primeiramente foi aplicada a análise da soma com relação aos primeiros dígitos conforme código presente na Figura 19 do Apêndice A. Com base neste código foram geradas as Figuras 66 a 72 (CeIM), Figuras 73 a 76 (EAM), Figuras 77 a 84 (BE) e Figuras 85 a 92 (HN) do Apêndice D, cada uma representando os resultados obtidos na análise da soma para os primeiros dígitos de cada organização militar.

As Figuras 66 a 92 explicitam a existência de diferença entre as ordens de grandeza dos números iniciados por cada primeiro dígito. Caso a soma dos números iniciados por um determinado algarismo possua um desvio positivo em relação à média das somas, significa que seus números apresentam uma ordem de grandeza grande quando comparado com o restante da amostra. Por outro lado, no caso dos números que possuam desvios negativos, significa que a ordem de grandeza deles é pequena quando comparada com os outros da amostra. Na Tabela 29 do Apêndice C são apresentados os primeiros dígitos com maiores desvios (positivos e negativos), em ordem decrescente, de cada organização militar.

Já para aplicar a análise da soma levando em consideração os dois primeiros dígitos foi utilizada a linguagem de programação contida na Figura 20 do Apêndice A. Assim, obtiveram-se como resultados do *software* R as Figuras 93 a 99 (CeIM), Figuras 100 a 103 (EAM), Figuras 104 a 111 (BE) e Figuras 112 a 119 (HN) do Apêndice D.

Na mesma linha da análise da soma para o primeiro dígito, as Figuras 93 a 119 explicitam a existência de diferenças entre as ordens de grandeza dos números iniciados por cada dois primeiros dígitos. Os dois primeiros dígitos com maiores desvios (positivos e negativos) em ordem decrescente, de cada organização militar, estão apresentados na Tabela 30 do Apêndice D.

3.4 Análises associadas dos dados

Com relação às análises associadas dos dados foram aplicadas a análise dos números duplicados e o modelo fator de distorção. Nesta dissertação primeiramente foi aplicada a análise dos números duplicados com auxílio do *software* R. Para isto foi utilizada a linguagem de programação contida na Figura 21 do Apêndice A. Deste procedimento foram extraídas as Tabelas 31, 32, 33 e 34 do Apêndice E onde estão ilustrados os valores das notas de empenhos, que mais se repetiram nos CeIM, EAM, BE e HN respectivamente.

As notas de empenhos que possuem como saldo os valores presentes nas Tabelas 31, 32, 33 e 34 do Apêndice E devido sua alta incidência, deveriam ser alvo de auditoria conforme Nigrini (2020). Por outro lado, os valores que possuem fundo azul são referentes aos empenhos de diárias, bolsas de estudos e gratificações por encargos de curso e concurso, valores estes que são tabelados, o que explicaria sua repetição excessiva. Para Nigrini (2012), este fato explicaria a excessiva repetição dos valores com fundo azul e excluiriam estas notas de empenhos da amostra de auditoria.

Para o modelo fator de distorção, foi utilizada a linguagem de programação contida na Figura 22 do Apêndice A no *software* R. Com os dados auferidos pelo *software*, foi obtida a Tabela 35 do Apêndice E.

Nesta dissertação foi considerado para o MFD um nível de significância de 5%, obtendo-se, assim, um $Z_{\text{crítico}}$ igual a 1,96. Assim sendo, os resultados da Tabela 34 do Apêndice E apontam que a EAM2, EAM3 e a BE5 foram as únicas amostras que apresentaram $|Z_{\text{teste}}| > |Z_{\text{crítico}}|$. Assim sendo, é rejeitada a hipótese nula e aceita a hipótese alternativa inferindo assim que o fator de distorção da amostra é estatisticamente diferente de zero. Conclui-se assim que segundo o modelo fator de distorção as notas de empenhos emitidas pela EAM2 foram superestimadas em 3,38%, já as notas de empenhos da EAM3 e BE5 foram subestimadas em 5,32% e 4,81% respectivamente cabendo assim segundo Nigrini (2012) e Nigrini (2020) uma análise qualitativa do auditor nestas organizações militares.

Por outro lado, as demais amostras apresentaram $|Z_{\text{teste}}| < |Z_{\text{crítico}}|$. Com isso, se aceita a hipótese nula, que infere que o fator de distorção das demais amostras é estatisticamente igual a zero. Em face do exposto, conclui-se que as notas de empenhos das demais amostras não estão subestimadas e nem tampouco superestimadas, já que possuem o fator de distorção estatisticamente igual a zero.

CONSIDERAÇÕES FINAIS

Este estudo teve como objetivo geral avaliar se as notas de empenhos emitidas em 2019 pelas organizações militares da Marinha do Brasil estão em conformidade com a lei Newcomb-Benford. Porém antes de cumprir este objetivo, esta dissertação teve como objetivo específico aquilatar a utilização da lei Newcomb-Benford na detecção de fraudes. Este objetivo foi alcançado na seção 1.2.2 ao explicitar a emprego da LNB em diversos temas como:

Pagamentos e transações - Nigrini e Mittermaier (1997), Nigrini (1999), Drake e Nigrini (2000), Santos, Tenório e Silva (2003), Özer e Babacan (2013), Silva, Korzenowski e Vaccaro (2014) e Ribeiro e Monsueto (2015);

Dados Financeiros - Forster (2006), Graham, Hasseldine e Paton (2009), Almeida (2011), Cella e Rech (2017), Oliveira *et al.* (2018) e Sallaberry *et al.* (2019);

Evasão fiscal - Christian e Gupta (1993), Nigrini (1996), Santos *et al.* (2009), Lagioia *et al.* (2011), Mir, Ausloos e Cerqueti (2014), Mir, Ausloos e Cerqueti (2014) e Macías e Igua (2018);

Notas de empenhos - Santos, Diniz e Corrar (2005), Diniz *et al.* (2006), Costa (2012), Costa, Santos e Travassos (2012), Freitas Costa *et al.* (2012), Nascimento, Filho e Buscacio (2014), Ganassin, Costa e Wilbert (2016), Silva, Travassos e Costa (2017), Cella e Zanolla (2018), De Oliveira *et al.* (2018) e Macieira *et al.* (2019);

Planilhas orçamentárias de obras públicas - Cunha (2013) e Café (2015); e

Folha de pagamento - Nigrini (1994).

O segundo objetivo específico deste trabalho foi identificar os testes estatísticos utilizados para verificar a conformidade com lei Newcomb-Benford. Este objetivo foi atingido pela seção 1.3.2 que identificou como testes estatísticos para verificação de conformidade com a LNB os testes qui quadrado, teste Z e teste Kolmogorov-Smirnov.

O terceiro objetivo específico deste trabalho foi identificar os procedimentos não estatísticos, propostos por Nigrini (2012) e Nigrini (2020), para verificar a conformidade com lei Newcomb-Benford. Este objetivo foi atingido pela seção 1.3.2 que identificou os procedimentos denominados análises pré-Benford, análises primárias, análises avançadas, análises associadas e análises de conformidade.

Para alcançar o objetivo geral foi escolhida uma amostra, por julgamento, de 27 organizações militares, que emitiram em conjunto 46.194 notas de empenhos e

movimentaram um montante de R\$ 6.721.861.945,92 no ano de 2019. Assim foi analisada a distribuição dos dígitos das 46.194 NE e comparada com a LNB, cumprindo assim, o quarto objetivo específico desta dissertação que foi analisar a distribuição das notas de empenhos emitidas em 2019 por organizações militares da Marinha do Brasil e sua conformidade com LNB.

Para cumprir o objetivo geral desta pesquisa que foi avaliar se as notas de empenhos emitidas em 2019 por organizações militares da Marinha do Brasil estão em conformidade com a lei Newcomb-Benford, foram realizadas as análises pré-Benford, primárias, avançadas e associadas conforme preconizado por Nigrini (2012) e Nigrini (2020) e os resultados estão explicitados, de forma resumida, a seguir.

Na análise primária dos primeiros dígitos, foram analisadas as notas de empenhos emitidas por cada OM de forma global, apenas a EAM3 não apresentou conformidade com a LNB em nenhum dos procedimentos estatístico e não estatísticos adotados. Já na análise local das NE emitidas pela EAM3 iniciadas pelos algarismos 1, 2, 3 e 4 apresentaram desvios significativos com a LNB.

Na análise primária dos segundos dígitos as NE da EAM3 e do CeIM5 não apresentaram conformidade com a LNB de forma global em nenhum dos procedimentos estatístico e não estatísticos adotados. Já quando realizada análise local, através do teste Z, foi explicitada a não conformidades com a LNB das NE iniciadas pelos dígitos 1, 2, 3, 4, 6, 7 e 9 para a EAM3 e 3, 4 e 8 para o CeIM5.

Na análise primária dos dois primeiros dígitos, as únicas amostras em que as NE emitidas não apresentaram conformidade com a LNB em nenhuma ocasião foram a BE7 e o CeIM5. Quando realizada a análise local foi percebida a não conformidades das NE iniciadas pelos dígitos 16, 20, 23, 26, 28, 33, 47, 50, 59, 70, 73 e 85 para a BE7 e 12, 14, 17, 25, 37, 39, 44, 47, 65, 78 e 89 para o CeIM5.

Na análise avançada de segunda ordem, quando analisados os primeiros dígitos apenas o HN1 não obteve conformidade para os dígitos 1 e 2. Já quando analisados os dois primeiros dígitos apenas os HN3 e HN4 obtiveram conformidade com a LNB. Assim, como preconizado por Kossovsky (2014) este procedimento (dos dois primeiros dígitos) não será levado em consideração para escolha de amostra de auditoria já que 25 das 27 organizações militares não apresentaram conformidade com a LNB.

Na análise da soma, foram percebidos diversos picos nas Figuras 66 a 119 e os algarismos presentes nas Tabelas 29 (primeiros dígitos) e 30 (dois primeiros dígitos) devem

passar por uma análise mais criteriosa conforme preconizado por Nigrini (2012) e Nigrini (2020).

No modelo fator de distorção, as notas de empenhos da EAM2 apresentaram indícios de estarem superestimadas, por outro lado, as NE das EAM3 e BE5 apresentaram indícios de estarem subestimadas, fato que, para Nigrini (2012) e Nigrini (2020) é motivo de uma análise mais rigorosa das NE.

Após todas as análises, as amostras que o estudo sugere um viés de fraude levando em consideração a não conformidade com lei de Newcomb-Benford, devendo fazer parte da amostra de auditoria e serem alvo de uma análise qualitativa de um auditor são as amostras CeIM5, EAM2, EAM3, BE5, BE7 e HN1. Assim dentre as 27 amostras estudadas apenas seis não seguiram a LNB em uma ou mais análises. Porém algumas dessas não conformidades podem ser explicadas por resultados obtidos nas análises pré-Benford e nas análises associadas.

O HN1 e as BE5 e BE7 possuem respectivamente 257, 753 e 889 notas de empenhos respectivamente fato que para Nigrini (2012) pode ocasionar não conformidade com a LNB tendo em vista que estas OMs possuem menos de 1000 notas de empenhos e em especial o HN1 que possui menos de 300 NE.

Outro fato que cabe salientar é que o CeIM5 e a EAM2 apresentaram respectivamente 17,5% e 28,17% de valores das notas de empenhos considerados pequenos por Nigrini(2012), ou seja inferiores a R\$1.000,00. Fato que ele alerta poder ocasionar viés de não conformidade com a LNB.

Já o CeIM5 e as EAM2 e EAM3 na análise dos números duplicados apresentaram diversos valores de NE repetidos o que pode ter ocasionado o desvio com a lei Newcomb-Benford. É possível verificar nas Tabelas 31 e 32 do Apêndice E que o CeIM5 e as EAM2 e EAM3 possuíram respectivamente 238, 188 e 167 notas de empenhos com valores que se repetem por mais de cinco vezes.

Assim de forma geral, conclui-se que as notas de empenhos da Marinha do Brasil seguem a lei Newcomb-Benford e está lei pode ser aplicada junto com a análise de dados para auxiliar na fase de planejamento e escolha de amostras de auditoria. Por outro lado, recomenda-se a verificação das CeIM5, EAM3 e BE7 devido as análises primárias, a verificação das NE das EAM2, BE5 e novamente da EAM3 devido aos resultados do modelo fator de distorção e a verificação das NE do HN1 devido a análise de segunda ordem dos primeiros dígitos.

É importante salientar que a não conformidade não configura fraude, apenas sugere uma análise *in loco* do auditor.

Por fim, sugere-se para pesquisas futuras, a análise das notas de empenhos emitidas após o ano de 2019 e verificar sua conformidade com a lei Newcomb-Benford. Além disso, sugere-se aplicar os procedimentos explicitados nesta dissertação para análise das notas de empenhos de outras organizações militares da MB no ano de 2019.

REFERÊNCIAS

ASSOCIATION OF CERTIFIED FRAUD EXAMINERS. *Report to the Nations on occupation fraud and abuse*, 2012. Disponível em:

<https://www.acfe.com/uploadedFiles/ACFE_Website/Content/rtnn/2012-report-to-nations.pdf> Acesso em: 14 abr.2020.

ASSOCIATION OF CERTIFIED FRAUD EXAMINERS, AMERICAN INSTITUTE OF CERTIFIED PUBLIC ACCOUNTANTS E INSTITUTE OF INTERNAL AUDITORS - ACFE; AICPA; IIA. *Managing the Business Risk of Fraud: A Practical Guide*, 2008.

Disponível em:

<https://www.acfe.com/uploadedfiles/acfe_website/content/documents/managing-business-risk.pdf> Acesso em: 14 abr.2020.

ADHIKARI, A. K.; SARKAR, B. P. Distribution of most significant digit in certain functions whose arguments are random variables. *The Indian Journal of Statistics*, Sankhyā, v. 30, n. 1/2, p. 47-58, June. 1968. Disponível em: < <https://www.jstor.org/stable/25051623>>. Acesso em: 14 abr. 2020.

ALBRECHT W. S.; ALBRECHT C. O.; ALBRECHT C. C.; ZIMBELMAN M. F. *Fraud examination*. 6. ed. Boston, Massachusetts, USA: Cengage Learning, 2019.

ALEXANDER; J. C. Remarks on the use of Benford's Law. *Social Science Research Network*, [S.l.], v.43, n. 11, p 31-49, Dec. 2009.

ALMEIDA,D. *Distribuição de Newcomb-Benford: teoria e aplicações no PIB da região Norte do Brasil, tributos e consumo de energia no estado do Amapá*.2011. 112 f. Trabalho de Conclusão de Curso (Licenciatura Plena em Matemática) – Faculdade de Matemática, Universidade Federal de Macapá, Amapá, 2011. Disponível em: < https://www2.unifap.br/matematica/files/2017/07/TCC-2011-Distribuicao.NB_.Daianne.pdf>. Acesso em: 20 maio. 2020.

AUSLOOS, M.; CERQUETI; R.; MIR; T. A. Data science for assessing possible tax income manipulation: the case of Italy. *Chaos, Solitons & Fractals*, USA, v. 104, n. 1/2, p. 238- 256, Nov. 2017. Disponível em: <

<https://www.sciencedirect.com/science/article/pii/S0960077917303314>>. Acesso em: 14 abr. 2020.

BAESENS, B. *Analytics in a big data world: the essential guide to data science and its applications*. 1. ed. Hoboken, New Jersey, USA : John Wiley & Sons, 2014.

_____.; VAN VLASSELAER, V.; VERBEKE, W. *Fraud analytics using descriptive, predictive, and social network techniques: a guide to data science for fraud detection*. 1. ed. Hoboken, New Jersey, USA: John Wiley & Sons, 2015.

BALANIUK, R. A Mineração de dados como apoio ao controle externo. *Revista do TCU*, Brasília, v. 42, n. 117, p. 79- 96, jan./abr. 2010. Disponível em: < <https://revista.tcu.gov.br/ojs/index.php/RTCU/article/view/282>>. Acesso em: 14 abr. 2020.

BECKER, P. W. Patterns in listings of failure-rate & MTTF values and listings of other data. *IEEE Transactions on Reliability*, USA, v. 32, n. 2, p. 132- 134, June 1982. Disponível em: < <https://ieeexplore.ieee.org/abstract/document/5221273>>. Acesso em: 14 abr. 2020.

BENFORD, F. The law of anomalous numbers. *Proceedings of the American philosophical society*, USA, v. 78, n. 4, p. 551- 572, Mar. 1938. Disponível em: < <https://www.jstor.org/stable/984802>>. Acesso em: 14 abr. 2020.

BHATTACHARYA, S; XU, D.; KUMAR, K. An ANN-based auditor decision support system using Benford's law. *Decision support systems*, USA, v. 50, n. 3, p. 576- 584, Feb. 2011. Disponível em: < <https://www.sciencedirect.com/science/article/abs/pii/S0167923610001351>>. Acesso em: 14 abr. 2020.

BIERSTAKER, J. L.; BRODY, R.; PACINI, C. Accountants perceptions regarding fraud detection and prevention methods. *Managerial Auditing Journal*, USA, v. 21, n. 5, p. 520- 535, June 2006. Disponível em: < <https://doi.org/10.1108/02686900610667283>>. Acesso em: 14 abr. 2020.

BOLTON, R. J.; HAND, D. J. Statistical fraud detection: A review. *Statistical science*, USA, v. 17, n. 3, p. 235- 255, Aug. 2002. Disponível em: <<https://projecteuclid.org/journals/statistical-science/volume-17/issue-3/Statistical-Fraud-Detection-A-Review/10.1214/ss/1042727940.full>>. Acesso em: 14 abr. 2020.

BRADY, W.G. More on Benford's law. *The Fibonacci Quarterly*, USA, v. 9, n. 1, p. 87- 88, Feb. 1971. Disponível em: < <https://www.mathstat.dal.ca/FQ/Scanned/16-1/brady.pdf>>. Acesso em: 14 abr. 2020.

BRASIL. Decreto 7.809, de 20 de setembro de 2012. Altera os Decretos nº 5.417, de 13 de abril de 2005, nº 5.751, de 12 de abril de 2006, e nº 6.834, de 30 de abril de 2009, que aprovam as estruturas regimentais e os quadros demonstrativos dos cargos em comissão e das funções gratificadas dos Comandos da Marinha, do Exército e da Aeronáutica, do Ministério da Defesa. *Diário Oficial da União*, Brasília, 20 de set. de 2012; Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2012/Decreto/D7809.htm> Acesso em: 10 mai. 2020.

_____. Tribunal de Contas da União. Referencial de combate a fraude e corrupção: aplicável a órgãos e entidades da Administração Pública. Brasília: TCU, Coordenação-Geral de Controle Externo dos Serviços Essenciais ao Estado e das Regiões Sul e Centro-Oeste, Secretaria de Métodos e Suporte ao Controle Externo (Semec), 2. ed, Brasília, 2018. Disponível em: < https://portal.tcu.gov.br/data/files/A0/E0/EA/C7/21A1F6107AD96FE6F18818A8/Referencial_combate_fraude_corrupcao_2_edicao.pdf>. Acesso em: 10 maio 2020.

BURKE, J.; KINCANON, E. Benford's law and physical constants: the distribution of initial digits. *American Journal of Physics*, USA, v. 59, n. 10, p. 952, Jan. 1991. Disponível em: < <http://dx.doi.org/10.1119/1.16838>>. Acesso em: 14 abr. 2020.

CAFÉ, R. M. *O uso da Lei de Benford na auditoria de obras públicas: o caso do VLP*. 2015. 54 f. Trabalho de Conclusão de Curso (Monografia Bacharelado em Economia) – Faculdade de Administração, Contabilidade e Economia, Universidade de Brasília, Brasília, 2015. Disponível em: <https://bdm.unb.br/bitstream/10483/11809/1/2015_RenataMottaCafe.pdf>. Acesso em: 20 mai. 2020.

CARSLAW, C. A. Anomalies in income numbers: Evidence of goal oriented behavior. *Accounting Review*, USA, v. 63, n. 2, p. 321- 327, Apr. 1988. Disponível em: <<https://www.jstor.org/stable/248109>>. Acesso em: 14 abr. 2020.

CELLA, R. S.; RECH, I. J. Caso Petrobras: a lei de Benford poderia detectar a fraude? *Revista de Gestão, Finanças e Contabilidade*, Salvador, v. 7, n. 3, p. 86 - 104, set./dez. 2017. Disponível em: <<https://www.revistas.uneb.br/index.php/financ/article/view/3665>>. Acesso em: 14 abr. 2020.

_____.; ZANOLLA, E. A Lei de Benford e a transparência: uma análise das despesas públicas municipais. *BBR. Brazilian Business Review*, Vitória, v. 15, n. 4, p. 331 - 347, jul/ago. 2018. Disponível em: <<https://doi.org/10.15728/bbr.2018.15.4.2>>. Acesso em: 14 abr. 2020.

CHRISTIAN, C. W.; GUPTA, S. New evidence on " secondary evasion ". *The Journal of the American Taxation Association*, Sarasota , v. 15, n. 1, p. 72 - 87, Mar./June 1993. Disponível em: <<https://www.proquest.com/scholarly-journals/new-evidence-on-secondary-evasion/docview/211084984/se-2?accountid=201395>>. Acesso em: 14 abr. 2020.

CHURCH, B. K.; MCMILLAN, J. J.; SCHNEIDER, A. Factors affecting internal auditors' consideration of fraudulent financial reporting during analytical procedures. *A Journal of Practice & Theory*, USA , v. 20, n. 1, p. 65 - 80, Mar. 2001. Disponível em: <<https://doi.org/10.2308/aud.2001.20.1.65>>. Acesso em: 14 abr. 2020.

CLEARY, R.; THIBODEAU, J. C. Applying digital analysis using Benford's law to detect fraud: the dangers of type I errors. *Auditing: A Journal of Practice & Theory*, USA , v. 24, n. 1, p. 77 - 81, May. 2005. Disponível em: <<https://doi.org/10.2308/aud.2005.24.1.77>>. Acesso em: 14 abr. 2020.

COGLITORE, F.; BERRYMAN, R. Analytical procedures: a defensive necessity. *Auditing: A Journal of Practice & Theory*, USA, v. 7, n. 2, p. 150 - 163, Mar./June 1988. Disponível em: <<https://doi.org/10.2307/2490618>>. Acesso em: 14 abr. 2020.

COSTA, G. P. C.; DUTRA, T. A. D. G. L. Auditoria financeira na era do Big Data: novas possibilidades para avaliação e resposta a riscos em demonstrações financeiras do Governo Federal. *Revista do TCU*, Brasília, v. 46, n. 131, p. 54 - 61, jan. 2014. Disponível em: <<https://doi.org/10.2307/2490618>>. Acesso em: 14 abr. 2020.

COSTA, J. I. F. *Desenvolvimento de metodologias contabilométricas aplicadas à auditoria contábil digital: uma proposta de análise da lei de Newcomb-Benford para os Tribunais de Contas*. 212. 448 f. Dissertação (Mestrado em Contabilidade) – Centro de Ciências Sociais Aplicadas, Universidade Federal de Pernambuco, Recife, 2012. Disponível em: <<https://repositorio.ufpe.br/handle/123456789/10416>>. Acesso em: 20 mai. 2020.

_____.; SANTOS, J.; TRAVASSOS, S. K. M. Análise de conformidade nos gastos públicos dos entes federativos: aplicação da Lei de Newcomb-Benford para o primeiro e segundo dígitos dos gastos em dois estados brasileiros. *Revista Contabilidade & Finanças*, São Paulo, v. 23, n. 60, p. 187 - 198, dez. 2012. Disponível em: < <https://doi.org/10.1590/S1519-70772012000300004>>. Acesso em: 14 abr. 2020.

CRAIG, T. R. Round-off bias in Earnings-Per-Share calculations. *Journal of Applied Business Research*, USA , v. 8, n. 4, p. 106 - 113, Sept/Dec. 1992. Disponível em: < www.proquest.com/scholarly-journals/round-off-bias-earnings-per-share-calculations/docview/227624747/se-2?accountid=201395 >. Acesso em: 14 abr. 2020.

CUNHA, F. C. R. *Aplicações da Lei Newcomb-Benford à auditoria de obras públicas*. 2013. 486 f. Dissertação (Mestrado Profissional em Regulação e Gestão de Negócios) - Faculdade de Administração, Contabilidade e Economia, Universidade de Brasília, Brasília, 2013. Disponível em: <<https://portal.tcu.gov.br/biblioteca-digital/aplicacoes-da-lei-newcomb-benford-a-auditoria-de-obras-publicas.htm>>. Acesso em: 20 mai. 2020.

DAS, S.; ZHANG, H. Rounding-up in reported EPS, behavioral thresholds, and earnings management. *Journal of Accounting and Economics*, USA , v. 35, n. 1, p. 31 - 50, Apr. 2003. Disponível em: < [https://doi.org/10.1016/S0165-4101\(02\)00096-4](https://doi.org/10.1016/S0165-4101(02)00096-4)>. Acesso em: 14 abr. 2020.

DIACONIS, P; FREEDMAN, D. On rounding percentages. *Journal of the American Statistical Association*, USA , v. 74, n. 366, p. 359 - 364, June 1979. Disponível em: < <https://www.jstor.org/stable/2286335>>. Acesso em: 14 abr. 2020.

DIEKMANN; A. Not the First Digit! Using Benford's Law to Detect Fraudulent Scientific Data. *Journal of Applied Statistics*, USA , v. 34, n. 3, p. 321 - 329, May 2007. Disponível em: < <https://doi.org/10.1080/02664760601004940>>. Acesso em: 14 abr. 2020.

DINIZ, J. A.; SANTOS, J. D.; DIENG, M.; DINIZ, M. A. A. Comprovação de eficácia da aplicação de modelos contabilométricos no campo da auditoria digital das contas públicas municipais: caso de um tribunal de contas de um estado brasileiro. In: CONGRESSO USP DE CONTROLADORIA E CONTABILIDADE, 6, 2006. São Paulo. *Anais eletrônicos...* Disponível em:<<https://congressosp.fipecafi.org/anais/artigos62006/261.pdf>>. Acesso em: 10 jun. 2020.

_____.; CORRAR, L. J; SLOMSKI, V. Análise digital: uma abordagem cognitiva na detecção de não conformidade em prestações de contas municipais. In: CONGRESSO USP CONTROLADORIA E CONTABILIDADE, 10, 2010, São Paulo. *Anais eletrônicos...* Disponível em:<https://congressosp.fipecafi.org/anais/artigos102010/an_resumo.asp?con=1&cod_trabalho=474&titulo=An%20E1lise+digital%3A+uma+abordagem+cognitiva+na+detec%20E3o+de+n%20E3o+conformidade+em+presta%20E7%F5es+de+contas+mun>. Acesso em: 10 jun. 2020.

DRAKE, P.D.; NIGRINI, M.J. Computer assisted analytical procedures using Benford's Law. *Journal of Accounting Education USA* , v. 18, n. 2, p. 127 - 146, Spring 2000. Disponível em: < [https://doi.org/10.1016/S0748-5751\(00\)00008-7](https://doi.org/10.1016/S0748-5751(00)00008-7)>. Acesso em: 14 abr. 2020.

DUFFIELD, G.; GRABOSKY, P. The psychology of fraud. *Trends and Issues in Crime and Criminal Justice*, Sydney , v. 54, n. 199, p. 1 - 6, Mar. 2001. Disponível em: < <https://search.informit.org/doi/abs/10.3316/agispt.20013757>>. Acesso em: 14 abr. 2020.

DUMAS, C. F.; DEVINE, J. H. Detecting Evidence of Non-Compliance in Self-Reported Pollution Emissions Data: an application of Benford'S Law. *American Agricultural Economics Association Annual Meeting*. USA , v. 76, n. 372, p. 195 - 239, July 2000. Disponível em: < <https://ageconsearch.umn.edu/record/21740/>>. Acesso em: 14 abr. 2020.

DURTSCHI, C.; HILLISON, W.; PACINI, C. The effective use of Benford's law in detecting fraud in accounting data , *Journal of Forensic Accounting*, USA , v. 5, n. 1, p. 17 - 34, Oct. 2004. Disponível em: < https://www.researchgate.net/profile/Cindy-Durtschi-2/publication/241401706_The_Effective_Use_of_Benford%27s_Law_to_Assist_in_Detecting_Fraud_in_Accounting_Data/links/54982f4a0cf2c5a7e342a59e/The-Effective-Use-of-Benford's-Law-to-Assist-in-Detecting-Fraud-in-Accounting-Data.pdf>. Acesso em: 14 abr. 2020.

ETTREDGE, M. L.; SRIVASTAVA, R. P. Using digital analysis to enhance data integrity. *Issues in Accounting Education*, USA , v. 14, n. 4, p. 675 - 690, Nov. 1999. Disponível em: < <https://doi.org/10.2308/iace.1999.14.4.675>>. Acesso em: 14 abr. 2020.

FIESP. Federação das Indústrias do Estado de São Paulo. *Corrupção: custos econômicos e propostas de combate*. 3. ed, São Paulo, 2010. Disponível em:< <https://www.fiesp.com.br/indices-pesquisas-e-publicacoes/indice-de-percepcao-da-corrupcao-2010/>> Acesso em: 06 mai. 2020.

FLEHINGER, B. J. On the probability that a random integer has initial digit A. *The American Mathematical Monthly*, USA , v. 73, n. 10, p. 1056 - 1061, Dec. 1966. Disponível em: < <https://www.jstor.org/stable/2314636>>. Acesso em: 14 abr. 2020.

FLEISS, J. L. *Statistical methods for rates and proportions*. 1 ed. Hoboken, New Jersey, USA: John Wiley & Sons, 1981.

FORSTER, R. P. *Auditoria contábil em entidades do terceiro setor: uma aplicação da lei Newcomb-Benford*. 2006. 66 f. Dissertação (Mestrado em Ciências Contábeis) – Faculdade de Administração, Contabilidade e Economia, Universidade de Brasília, Brasília, 2006. Disponível em: <<https://repositorio.unb.br/handle/10482/5956?mode=full>>. Acesso em: 20 mai. 2020.

FREITAS COSTA, J. I., HENRIQUES, D. B. B., BARROS MELO, S.; SANTOS, J. Análise de métodos contabilométricos para determinação de conformidade da Lei de Newcomb-Benford aplicados à Auditoria Contábil. *Gestão Pública: Práticas e Desafios*, Recife , v. 3, n. 6, p. 292 - 314, dez. 2012. Disponível em: <<https://periodicos.ufpe.br/revistas/index.php/gestaopublica/article/view/1139>>. Acesso em: 14 abr. 2020.

FURRY, W. H.; HURWITZ, H. Distribution of numbers and distribution of significant figures. *Nature*, USA , v. 155, n. 3924, p. 52 - 53, Jan. 1945. Disponível em: < <https://periodicos.ufpe.br/revistas/index.php/gestaopublica/article/view/1139>>. Acesso em: 14 abr. 2020.

GANASSIN, E. J. F.; COSTA, A. J. B.; WILBERT, M. D. Aplicação de Modelo Contabilométrico baseado na Lei de Newcomb-Benford no Controle de Contas Públicas. *Revista de Estudos Contábeis*, Londrina , v. 7, n. 12, p. 39 - 56, jan/jun. 2016. Disponível em: < <http://www.uel.br/revistas/uel/index.php/rec/article/view/20812/22798>>. Acesso em: 14 abr. 2020.

GEE, S. *Fraud and fraud detection: a data analytics approach*. 1 ed. Hoboken, New Jersey, USA: John Wiley & Sons, 2015.

GOOD, I. Letter to the editor. *The American Statistician*, USA , v. 5, n. 19, p. 41 - 43, June 1965. Disponível em: < <https://doi.org/10.1080/00031305.1965.10479764>>. Acesso em: 14 abr. 2020.

GOUDSMIT, S. A.; FURRY, W. H. Significant figures of numbers in statistical tables. *Nature*, USA , v. 154, n. 3921, p. 800 - 801, dez. 1944. Disponível em: < <https://doi.org/10.1038/154800a0>>. Acesso em: 14 abr. 2020.

GRAHAM S. D. J.; HASSELDINE J.; PATON D. Statistical fraud detection in a commercial lobster fishery, *New Zealand Journal of Marine and Freshwater Research*, Wellington , v. 43, n. 1, p. 457 - 463, jun. 2009. Disponível em: < <https://doi.org/10.1080/00288330909510014>>. Acesso em: 14 abr. 2020.

HILL, T. P. A statistical derivation of the significant-digit law. *Statistical science*, USA , v. 10, n. 4, p. 354 - 363, Nov. 1995. Disponível em: < <https://www.jstor.org/stable/2246134> >. Acesso em: 14 abr. 2020.

_____. Random-number guessing and the first digit phenomenon. *Psychological Reports*, USA , v. 62, n. 3, p. 967 - 97. June 1988. Disponível em: < <https://doi.org/10.2466/pr0.1988.62.3.967>>. Acesso em: 14 abr. 2020.

_____. The significant-digit phenomenon. *The American Mathematical Monthly*, USA , v. 102, n. 4, p. 322 - 327, Apr. 1995a. Disponível em: < <https://doi.org/10.1080/00029890.1995.11990578>>. Acesso em: 14 abr. 2020.

HSU, E. H. An experimental study on “mental numbers” and a new application. *The Journal of general psychology*, USA , v. 38, n. 1, p. 57 - 67, Jan. 1948. Disponível em: < <https://doi.org/10.1080/00221309.1948.9711768>>. Acesso em: 14 abr. 2020.

INSTITUTO DOS AUDITORES INDEPENDENTES DO BRASIL. *Brasil amarga o título de país pouco auditado*. 2016 Disponível em: <<http://www.ibracon.com.br/ibracon/Portugues/detNoticia.php?cod=3336>>. Acesso em: 5 maio 2020.

_____. *Interesse das auditorias no mercado é crescente*. 2013 Disponível em: <<http://www.ibracon.com.br/ibracon/Portugues/detNoticia.php?cod=1453>>. Acesso em: 5 maio 2020.

INSTITUTE OF INTERNAL AUDITORS. *Supplemental Guidance: The Role of Auditing in Public Sector Governance* 2012. Disponível em: <<http://docplayer.net/1724809->

Supplemental-guidance-the-role-of-auditing-in-public-sector-governance-2nd-edition.html>
Acesso em: 3 maio 2020.

JAMAIN, A. *Benford's Law*. 2001. 69 f. Dissertação (Mestrado em Matemática) - Faculdade de Matemática, Imperial College London, Londres, 2001. Disponível em: <<https://www.ma.imperial.ac.uk/~nadams/classificationgroup/Benfords-Law.pdf>>. Acesso em: 20 maio 2020.

_____. *Forensic analytics: methods and techniques for forensic accounting investigations*. 2. ed. Hoboken, New Jersey, USA: John Wiley & Sons, 2020.

SÁ, A. L. *Curso de auditoria*. 9. ed. São Paulo: Atlas, 2002.

KANITZ, S. *Blog para se pensar*. 2012. Disponível em: <<https://blog.kanitz.com.br/corruptao/>>. Acesso em: 14 abr. 2020.

KOSSOVSKY, A. E. *Benford's Law: Theory, the general law of relative quantities, and forensic fraud detection applications*. 1 ed. Hackensack, New Jersey, USA : World Scientific, 2014.

KOVALERCHUK, B; VITYAEV, E.; HOLTFRETER, R. Correlation of complex evidence in forensic accounting using data mining. *Journal of Forensic Accounting*, USA , v. 8, n. 1, p. 53 - 88, June 2007. Disponível em: < <https://silo.tips/download/correlation-of-complex-evidence-in-forensic-accounting-using-data-mining-2> >. Acesso em: 14 abr. 2020.

KRAKAR, Z.; ŽGELA, M. Application of Benford 's Law in Information Systems Auditing. *Journal of Information and Organizational Sciences*, USA , v. 33, n. 1, p. 60 - 78, July. 2009. Disponível em: <<http://archive.ceciis.foi.hr/app/index.php/ceciis/2008/paper/view/71/34>>. Acesso em: 14 abr. 2020.

LAGIOIA, U. C. T., ARAÚJO, I. J. C. D., ALVES FILHO, B. D. F., BARROS, M. A. B.; NASCIMENTO, S. G. O. D. A. Aplicabilidade da Lei de Newcomb-Benford nas fiscalizações do imposto sobre serviços-ISS. *Revista Contabilidade & Finanças*, São Paulo , v. 22, n. 56, p. 203 - 224, ago. 2011. Disponível em: < <https://doi.org/10.1590/S1519-70772011000200006>>. Acesso em: 14 abr. 2020.

LEVINE, D. M., STEPHAN, D. F., KREHBIEL, T. C., BERENSON, Mark L. *Estatística: teoria e aplicações*. 5. ed. Rio de Janeiro: LTC, 2008.

OLIVEIRA, J. A. J. de; MILANI FILHO, M. A. F.; FRANCISCHETTI, C. E.; OLIVEIRA JR, S. V. de; PADOVEZE, C. L. Uma análise nos gastos da união: aplicabilidade da lei de Newcomb-Benford. *Contabilometria- Brazilian Journal of Quantitative Methods Applied to Accounting*, Monte Carmelo, v. 5, n. 2, p. 123 - 141, jul/dez. 2018. Disponível em: < <http://www.fucamp.edu.br/editora/index.php/contabilometria/article/view/1112>>. Acesso em: 14 abr. 2020.

MACÍAS; A. L. O; IGUA; S. T. Encontrando datos anómalos en la tributación. Aplicación de la Ley de Benford en el Impuesto a la Renta en Ecuador. *SaberEs*, Quito , v. 10, n. 2, p. 203 - 224, dic. 2018. Disponível em: <<https://doi.org/10.35305/s.v10i2.191>>. Acesso em: 14 abr. 2020.

MACIEIRA, F.; NASCIMENTO, J.; MILLER, F.; ZITTEI, M. Aplicabilidade da lei de newcomb-benford à auditoria governamental: uma análise dos pareceres prévios do Tribunal de Contas do Estado de São Paulo. In: CONGRESSO INTERNACIONAL DE DESEMPENHO DO SETOR PÚBLICO, 3, 2019. Florianópolis. *Anais eletrônicos...* Disponível em: <<http://www.cidesp.com.br/index.php/Icidesp/3cidesp/paper/view/657>>. Acesso em: 10 jun. 2020.

MALHOTRA, N. K. *Pesquisa de Marketing: uma orientação aplicada*. 6. ed. Porto Alegre: Bookman, 2012.

MANTONE, P. S. *Using Analytics to Detect Possible Fraud: Tools and Techniques*. 1 ed. Hoboken, New Jersey, USA :John Wiley & Sons, 2013

MARKONI, M. A.; LAKATOS, E. M. *Fundamentos de metodologia científica*. São Paulo: Atlas, 2007.

MILANI FILHO, M. A. F. A Confiabilidade dos dados financeiros de hospitais filantrópicos canadenses: Um estudo empírico baseado na Lei de Benford. *Sociedade, Contabilidade e Gestão*, Rio de Janeiro , v. 8, n. 2, p. 48 - 63, ago. 2013. Disponível em: <<http://www.atena.org.br/revista/ojs-2.2.3-08/index.php/ufrj/article/viewArticle/1873>>. Acesso em: 14 abr. 2020.

_____. POKER J. H.; BELLI M. M.; SEGURA L. C. Bolsas de Valores dos BRICS: uma análise comparativa das informações financeiras baseada na Lei de Benford. *Contabilidade Vista & Revista*, Belo Horizonte, v. 27, n. 2, p. 64 - 87, ago. 2016. Disponível em: <<https://revistas.face.ufmg.br/index.php/contabilidadevistaerevista/article/view/2833>>. Acesso em: 14 abr. 2020.

MILLER, S. J.; NIGRINI, M. J. Order statistics and Benford's law. *International Journal of Mathematics and Mathematical Sciences*, USA , v. 40, n. 3, p. 30 - 48, set. 2008. Disponível em: <<https://doi.org/10.1155/2008/382948>>. Acesso em: 14 abr. 2020.

MIR; T. A. AUSLOOS; M. CERQUETI, R. Benford's law predicted digit distribution of aggregated income taxes: the surprising conformity of Italian cities and regions. *The European Physical Journal B*, USA , v. 87, n. 113, p. 1 - 8, Nov. 2014. Disponível em: <<https://doi.org/10.1140/epjb/e2014-50525-2>>. Acesso em: 14 abr. 2020.

MORAIS, E. S. *Lei Newcomb-Benford: uma análise da capacidade de detecção de fraude*. 2013. 35 f. Trabalho de Conclusão de Curso (Bacharel em Ciências Contábeis) - Faculdade de Economia, Administração e Contabilidade, Universidade de Brasília, Brasília, 2013. Disponível em: <<https://bdm.unb.br/handle/10483/11935>>. Acesso em: 20 maio 2020.

MOSIMANN, J. E.; WISEMAN, C. V.; EDELMAN, R. E. Data fabrication: can people generate random digits? *Accountability in research*, USA , v. 4, n. 1, p. 31 - 55, Mar. 1995. Disponível em: <<https://doi.org/10.1080/08989629508573866>>. Acesso em: 14 abr. 2020.

NASCIMENTO, T. D. FILHO, E. M. S. BUSCACIO, L. Detecção de Fraudes: o uso da Lei de Benford para avaliar dados educacionais e financeiros. SIMPÓSIO BRASILEIRO DE

PESQUISA OPERACIONAL, 46, 2014. Salvador-BA, *Anais eletrônicos...* Disponível em: <<http://www.din.uem.br/sbpo/sbpo2014/pdf/arq0211.pdf>>. Acesso em: 10 jun. 2020.

NELSON, L. Display Tables and Significant Digits. *Journal of Quality Technology USA*, v. 16, n. 3, p. 175 - 176, Sept. 1984. Disponível em: <<https://doi.org/10.1080/00224065.1984.11978911>>. Acesso em: 14 abr. 2020.

NEWCOMB, S. Note on the frequency of use of the different digits in natural numbers. *American Journal of mathematics*, USA , v. 4, n. 1, p. 39 - 40, Jan. 1881. Disponível em: <<http://gerry.lamost.org/upload/newcomb.pdf>>. Acesso em: 14 abr. 2020.

NIGRINI M.J. A taxpayer compliance application of Benford's Law. *Journal of the American Taxation Association USA* , v. 18, n. 1, p. 72 - 91, Mar. 1996. Disponível em: <<https://www.proquest.com/scholarly-journals/taxpayer-compliance-application-benfords-law/docview/211023799/se-2?accountid=201395>>. Acesso em: 14 abr. 2020.

_____. *Benford's Law: Applications for forensic accounting, auditing, and fraud detection*. 1.ed. Hoboken, New Jersey, USA: John Wiley & Sons, 2012.

_____. *Forensic analytics: methods and techniques for forensic accounting investigations*. 2.ed. Hoboken, New Jersey, USA: John Wiley & Sons, 2020.

_____. An assessment of the change in the incidence of earnings management around the Enron-Andersen episode. *Review of Accounting and Finance*, USA , v. 4, n. 1, p. 92 - 110, Mar. 2005. Disponível em: <<https://doi.org/10.1108/eb043420>>. Acesso em: 14 abr. 2020.

_____. I've got your number. *Journal of accountancy*, USA , v. 187, n. 5, p. 79 - 83, May 1999. Disponível em: <https://web.williams.edu/Mathematics/sjmiller/public_html/BrownClasses/197/benford/Nigrini_IveGotYourNumber.pdf>. Acesso em: 14 abr. 2020.

_____. MITTERMAIER, L. J. The use of Benford's law as an aid in analytical procedures. *Auditing: A journal of practice & theory*. USA, v. 16, n. 2, p. 52 - 67, June 1997. Disponível em: <<https://www.proquest.com/scholarly-journals/use-benfords-law-as-aid-analytical-procedures/docview/216734639/se-2?accountid=201395>>. Acesso em: 14 abr. 2020.

_____. MILLER, S. J. Data diagnostics using second-order tests of Benford's law. *Auditing: A Journal of Practice & Theory*, USA , v. 28, n. 2, p. 305 - 324, Nov. 2009. Disponível em: <<https://doi.org/10.2308/aud.2009.28.2.305>>. Acesso em: 14 abr. 2020.

_____. *The detection of income tax evasion through an analysis of digital distributions*. 1992. 309 f. Tese (Doutorado em Contabilidade) – Faculdade de Contabilidade, Universidade de Cincinnati, Cincinnati, 1992. Disponível em: <<https://www.proquest.com/docview/304052348?pq-origsite=gscholar&fromopenview=true>>. Acesso em: 20 maio 2020.

_____. Using digital frequencies to detect fraud. *The white paper*, USA, v. 8, n. 2, p. 3 - 6, maio 1994. Disponível em: <<https://doi.org/10.1111/j.1467-646X.2004.00108.x>>. Acesso em: 14 abr. 2020.

NIGRINI, M.; WOOD, W. Assessing the integrity of tabulated demographic data. *Preprint, Univ. Cincinnati and St. Mary's Univ, USA*, v. 15, n. 1, p. 23 - 36, Sept. 1995. Disponível em: <<https://dx.doi.org/doi:10.21220/s2-j7gh-w740>>. Acesso em: 14 abr. 2020.

OLIVEIRA, J. A. J.; GALEANO, R.; FRANCISCHETTI, C. E.; PADOVEZE, C. L.; MILANI FILHO, M. A. F. Aplicações da Lei de Newcomb-Benford nas demonstrações financeiras da Petrobrás. *Enfoque: Reflexão Contábil*, Maringá, v. 37, n. 1, p. 21 - 38, jan. 2018. Disponível em: <<https://doi.org/10.4025/enfoque.v37i1.35608>>. Acesso em: 14 abr. 2020.

OLIVEIRA, J. A. J. de; MILANI FILHO, M. A. F.; FRANCISCHETTI, C. E.; OLIVEIRA JR, S. V. de; PADOVEZE, C. L. Uma análise nos gastos da união: aplicabilidade da lei de Newcomb-Benford. *Contabilometria- Brazilian Journal of Quantitative Methods Applied to Accounting*, Monte Carmelo, v. 5, n. 2, p. 123 - 141, jul/dez. 2018. Disponível em: <<http://www.fucamp.edu.br/editora/index.php/contabilometria/article/view/1112>>. Acesso em: 14 abr. 2020.

ÖZER, G.; BABACAN, B. Benford's Law and Digital Analysis: Application on Turkish Banking Sector. *Business and Economics Research Journal*, USA, v. 4, n. 1, p. 29 - 41, jan. 2013. Disponível em: <<https://ssrn.com/abstract=2156485>>. Acesso em: 14 abr. 2020.

PINKHAM, R. S. On the distribution of first significant digits. *The Annals of Mathematical Statistics*, USA, v. 32, n. 41, p. 1223 - 1230, maio 1961. Disponível em: <<https://www.jstor.org/stable/2237922>>. Acesso em: 14 abr. 2020.

R CORE TEAM. R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. Disponível em: <<https://www.R-project.org/>>. Acesso em: 10 mar. 2020.

RAIMI, R. A. On the distribution of first significant figures. *The American Mathematical Monthly*, USA, v. 76, n. 4, p. 352 - 348, Sept. 1969. Disponível em: <<https://doi.org/10.1080/00029890.1969.120002152>>. Acesso em: 14 abr. 2020.

_____. The first digit problem. *The American Mathematical Monthly*, USA, v. 83, n. 7, p. 521 - 538, June 1976. Disponível em: <<https://doi.org/10.1080/00029890.1976.11994162>>. Acesso em: 14 abr. 2020.

_____. The peculiar distribution of first digits. *Scientific American*, USA, v. 221, n. 6, p. 109 - 121, Dec. 1969a. Disponível em: <<https://www.jstor.org/stable/24964397>>. Acesso em: 14 abr. 2020.

RIBEIRO, D. H.; MONSUETO, S. E. Lei de Newcomb-Benford aplicada no controle interno nas empresas: um estudo de caso no controle de lançamentos financeiros. *Revista de Administração Unimep*, Piracicaba, v. 13, n. 1, p. 1 - 18, abr. 2015. Disponível em: <<https://www.redalyc.org/pdf/2737/273738309001.pdf>>. Acesso em: 14 abr. 2020.

SÁ, A. L. *Curso de auditoria*. 9. ed. São Paulo: Atlas, 2002.

SALLABERRY, J. D.; FLACH, L.; CODESSO, M. M.; RODRIGUES, L.F. Sinalização de inconsistências a partir do patrimônio declarado de políticos no Brasil: aplicação da Lei Newcomb-Benford. *Revista Evidenciação Contábil & Finanças*, João Pessoa, v. 7, n. 3, p. 39

- 59, set. 2019. Disponível em: <<https://doi.org/10.22478/ufpb.2318-1001.2019v7n3.41903>>. Acesso em: 14 abr. 2020.

SANTOS, J., TENÓRIO, J. N. B.; SILVA, L. G. C. Uma aplicação da teoria das probabilidades na contabilometria: A Lei de Newcomb-Benford como medida para análise de dados no campo da auditoria contábil. *Revista Contabilidade, Gestão e Governança*, Brasília, v. 6, n. 1, p. 35 - 54, set. 2003. Disponível em: <https://www.revistacgg.org/contabil/article/view/192/pdf_61>. Acesso em: 14 abr. 2020.

_____.; DINIZ, A. J.; CORRAR, L. J. O Foco é a Teoria Amostral nos Campos da Auditoria Contábil Tradicional e da Auditoria Digital: testando a Lei de Newcomb- Benford para o primeiro dígito nas contas públicas. *BBR - Brazilian Business Review*, Vitória, v. 2, n. 1, p. 71 - 89, jan. 2005. Disponível em: <<https://www.redalyc.org/pdf/1230/123016184005.pdf>>. Acesso em: 14 abr. 2020.

_____.; _____.; RIBEIRO FILHO, J. F. A Lei de Newcomb-Benford: uma aplicação para determinar o DNA-equivalente das despesas no setor público. In: CONGRESSO DA USP DE CONTABILIDADE E CONTROLADORIA, 3, 2003. São Paulo. *Anais eletrônicos...* Disponível em: <<https://congressusp.fipecafi.org/anais/artigos32003/default.asp?con=1>>. Acesso em: 10 jun. 2020.

_____.; RIBEIRO FILHO, J. F.; LAGIOLIA, U.; ALVES FILHO, B. F.; ARAÚJO, I. J.C. Aplicações da lei de Newcomb-Benford na auditoria tributária do imposto sobre serviços de qualquer natureza (ISS). *Revista Contabilidade & Finanças*, São Paulo, v. 20, n. 49, p. 79 - 94, jan. 2009. Disponível em: <<https://doi.org/10.1590/S1519-70772009000100006>>. Acesso em: 14 abr. 2020.

SENTANCE, W.A. A further analysis of Benford's law. *Fibonacci Quarterly*, USA, v. 11, n. 1, p. 490 - 494, Jan. 1973. Disponível em: <<https://www.mathstat.dal.ca/FQ/Scanned/11-5/sentance.pdf>>. Acesso em: 14 abr. 2020.

SHESKIN, D. J. *Handbook of parametric and nonparametric statistical procedures*. 4 ed. Florida, USA: Chapman & Hall, 2007.

SILVA, Â. H. L. O Uso da Lei Newcomb-Benford na contabilidade e auditoria. *Revista do TCU*, Brasília, v. 115, n. 1, p. 19 - 25, jan. 2009. Disponível em: <<https://revista.tcu.gov.br/ojs/index.php/RTCU/article/view/314>>. Acesso em: 14 abr. 2020.

_____.; KORZENOWSKI, A. L.; VACCARO, G. L. R.. Uma aplicação da lei de Benford na identificação de padrões estatisticamente assinaláveis de suspeitas de fraude por lavagem de dinheiro. *Revista Espacios*, Caracas, v. 35, n. 7, p. 20 - 34, jun. 2014. Disponível em: <<https://www.revistaespacios.com/a14v35n07/14350720.html>>. Acesso em: 14 abr. 2020.

SILVA, W. B. S.; TRAVASSOS, S. K. M.; COSTA, J. I. F. Utilização da Lei de Newcomb-Benford como método identificador de desvios em ambientes de auditoria contínua: uma proposta de identificação de desvios no tempo. *Revista Contabilidade & Finanças*, São Paulo, v. 28, n. 73, p. 11-26, jan./abr. 2017. Disponível em: <http://www.scielo.br/pdf/rcf/v28n73/pt_1808-057X-rcf-28-73-00011.pdf>. Acesso em: 2 maio. 2020.

TAM CHO, W. K.; GAINES, B. J. Breaking the (Benford) law: statistical fraud detection in campaign finance. *The american statistician*, USA , v. 61, n. 3, p. 218 - 223, Aug. 2007. Disponível em: <<https://doi.org/10.1198/000313007X2234961>>. Acesso em: 14 abr. 2020.

THOMAS, J. K. Unusual patterns in reported earnings. *Accounting Review*, USA , v. 64, n. 4, p. 773 - 787, Oct. 1989. Disponível em: <<https://www.jstor.org/stable/247861>>. Acesso em: 14 abr. 2020.

TRANSPARÊNCIA INTERNACIONAL *Índice de Percepção da Corrupção* 2.ed, São Paulo, 2019. Disponível em: <<https://comunidade.transparenciainternacional.org.br/asset/67:indice-de-percepcao-da-corrupcao-2019?stream=1>>. Acesso em: 10 maio 2020.

TSAO, N. K. On the distributions of significant digits and round off errors. *Communications of the ACM*, USA , v. 17, n. 5, p. 269 - 271, maio 1974. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/360980.360998?casa_token=qZIoQ0jVW54AAAAA%3AnBxuLNTiL4AhCLu56sNe4x0iUSnlNy2mYAVD_ON4MUJ-JT5pup9EZnThPX5hZtkKJT77wWG39sop>. Acesso em: 14 abr. 2020.

UMAR, I.; SAMSUDIN, R. S.; MOHAMED, M. The types, costs, prevention and detection of occupational fraud: The ACFE perspective. In: INTERNATIONAL CONFERENCE ON ACCOUNTING STUDIES, 1, 2015, Johor, Malaysia. *Anais eletrônicos...* Disponível em: <http://repo.uum.edu.my/17605/1/501_ICAS2015%20501-505.pdf>. Acesso em: 10 jun. 2020.

VAN VLASSELAER, V.; ELIASSI-RAD, T.; AKOGLU, L.; SNOECK, M.; BAESENS, B. Gotcha! Network-based fraud detection for social security fraud. *Management Science*, USA , v. 63, n. 9, p. 3090 - 3110, Sept. 2017. Disponível em: <<https://doi.org/10.1287/mnsc.2016.2489>>. Acesso em: 14 abr. 2020.

VARIAN, H.R. Benford's law. *The American Statistician*, USA , v. 26, n. 9, p. 62 - 66, June 1972. Disponível em: <<https://www.jstor.org/stable/2682871>>. Acesso em: 14 abr. 2020.

WALLACE, W.A. Assessing the quality of data used for benchmarking and decision-making. *The Journal of Government Financial Management*, USA , v. 51, n. 3, p. 16 - 22, Apr. 2002. Disponível em: <<https://www.proquest.com/scholarly-journals/assessing-quality-data-used-benchmarking-decision/docview/222377882/se-2?accountid=195669>>. Acesso em: 14 abr. 2020.

WLODARSKI, J. Fibonacci and Lucas Numbers tend to obey Benford's law. *Fibonacci Quarterly* , USA , v. 9, n. 1, p. 87 - 88, Jan. 1971. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.389.8710>>. Acesso em: 14 abr. 2020.

APÊNDICE A – Linguagem de programação empregada

Figura 5 – Importação de dados xlsx

```
install.packages("readxl")
library(readxl)
CeIM1=read_excel("C:/Users/Leonardo Botelho/Desktop/Dados/CeIM.xlsx",sheet="CeIM1")
```

Fonte: O autor, 2021.

Figura 6 – Substituição dos *not available*

```
for (i in 1:length(CeIM1$Data)) {
  if( is.na(CeIM1$Data[i])==TRUE){
    CeIM1$Data[i]=CeIM1$Data[i-1]}}
```

Fonte: O autor, 2021.

Figura 7 – Aplicação do perfil de dados

```
a1=b1=c1=d1=e1=0
for (i in 1:length(CeIM1$Saldo)) {
  if (CeIM1$Saldo[i]<10) a1=a1+1 else
  if (CeIM1$Saldo[i]>=10 & CeIM1$Saldo[i]<100) b1=b1+1 else
  if (CeIM1$Saldo[i]>=100 & CeIM1$Saldo[i]<1000) c1=c1+1 else
  if (CeIM1$Saldo[i]>=1000 & CeIM1$Saldo[i]<10000) d1=d1+1 else
  if (CeIM1$Saldo[i]>=10000) e1=e1+1}
print(c(a1,b1,c1,d1,e1))
```

Fonte: O autor, 2021.

Figura 8 – Criação de histogramas

```
par(mfrow=c(1,2))
hist(CeIM1$Saldo,nc=10,xlab="valores em reais",ylab="Frequência",right=F,main= "Histograma CeIM1",col=5)
hist(CeIM2$Saldo,nc=10,xlab="valores em reais",ylab="Frequência",right=F,main= "Histograma CeIM2",col=5)
```

Fonte: O autor, 2021.

Figura 9 – Programação do gráfico temporal

```

Data1=as.Date(CeIM1$Data, format = "%d/%m/%Y")
mes1=months(Data1)
aux1=data.frame(mes1,CeIM1$Saldo)
colnames(aux1)=c("mes", "saldo")
m1=m2=m3=m4=m5=m6=m7=m8=m9=m10=m11=m12=0
for (i in 1:length(aux1$mes)) {
  if (aux1$mes[i]=="janeiro") m1=m1+aux1$saldo[i] else
  if (aux1$mes[i]=="fevereiro") m2=m2+aux1$saldo[i] else
  if (aux1$mes[i]=="marco") m3=m3+aux1$saldo[i] else
  if (aux1$mes[i]=="abril") m4=m4+aux1$saldo[i] else
  if (aux1$mes[i]=="maio") m5=m5+aux1$saldo[i] else
  if (aux1$mes[i]=="junho") m6=m6+aux1$saldo[i] else
  if (aux1$mes[i]=="julho") m7=m7+aux1$saldo[i] else
  if (aux1$mes[i]=="agosto") m8=m8+aux1$saldo[i] else
  if (aux1$mes[i]=="setembro") m9=m9+aux1$saldo[i] else
  if (aux1$mes[i]=="outubro") m10=m10+aux1$saldo[i] else
  if (aux1$mes[i]=="novembro") m11=m11+aux1$saldo[i] else
  if (aux1$mes[i]=="dezembro") m12=m12+aux1$saldo[i]}
mesesn=seq(1:12)
salDOS=c(m1,m2,m3,m4,m5,m6,m7,m8,m9,m10,m11,m12)
tabela1=data.frame(mesesn,salDOS)
plot(tabela1$meses,tabela1$salDOS/1000,type="b",pch=16,col="blue",xlab="Meses", ylab="Total empenhados em milhares de reais",main="CeIM1")

```

Fonte: O autor, 2021.

Figura 10 – Inicialização dos pacotes

```

install.packages("benford.analysis")
library(benford.analysis)
install.packages("BenfordTests")
library(BenfordTests)

```

Fonte: O autor, 2021.

Figura 11 – Análises globais para os primeiros dígitos

```

bCeIM1=benford(CeIM1$Saldo, number.of.digits = 1, sign = "positive", discrete=TRUE, round=3)
chisq(bCeIM1)
MAD(bCeIM1)
ks.benftest(CeIM1$Saldo,digits=1, pvalmethod="simulate", pvalsims=10000)

```

Fonte: O autor, 2021.

Figura 12 – Análise local para os primeiros dígitos

```

signifd.analysis(CeIM1$Saldo, digits=1, graphical_analysis=FALSE, freq=TRUE,ci_lines=c(0.05))

```

Fonte: O autor, 2021.

Figura 13 – Análises globais para os segundos dígitos

```

F1=signifd(CeIM1$Saldo,digits=1)
FT1=signifd(CeIM1$Saldo,digits=2)
S1=FT1-10*F1
PO1=table(S1)
PE1=c()
for (i in 1:10) {PE1[i]=p.this.digit.at.n(i-1,2)*sum(PO1)}
Tabela171=data.frame(PO1,PE1)
for (i in 1:10) {Tabela171$X2[i]=(((PO1[i]-PE1[i])^2)/PE1[i])}
sum(Tabela171$X2)
po1=c()
for (i in 1:10) {po1[i]=PO1[i]/sum(PO1)}
pe=c()
for (i in 1:10) {pe[i]=p.this.digit.at.n(i-1,2)}
MAD1=sum(abs(po1-pe))/10
pea=c()
pea[1]=pe[1]
for (i in 2:10) {
pea[i]=pea[i-1]+pe[i]}
po1a=c()
po1a[1]=po1[1]
for (i in 2:10) {
po1a[i]=po1a[i-1]+po1[i]}
KS1=max(abs(po1a-pea))
KS1c=1.36/sqrt(sum(PO1))

```

Fonte: O autor, 2021.

Figura 14 – Análise local para os segundos dígitos

```

for (i in 1:10) {Tabela171$Z[i]=((PO1[i]-PE1[i])/sqrt(PE1[i]*(1-pe[i])))}

```

Fonte: O autor, 2021.

Figura 15 – Análises globais para os dois primeiros dígitos

```

b2CeIM1=benford(CeIM1$Saldo, number.of.digits = 2, sign = "positive", discrete=TRUE, round=3)
chisq(b2CeIM1)
MAD(b2CeIM1)
marc(b2CeIM1)
ks.benftest(CeIM1$Saldo,digits=2, pvalmethod="simulate", pvalsims=10000)

```

Fonte: O autor, 2021.

Figura 16 – Análise local para os dois primeiros dígitos

```

signifd.analysis(CeIM1$Saldo, digits=2, graphical_analysis=TRUE, freq=FALSE, a1phas=20, tick_col="red",ci_col="green",ci_lines=c(0.05))
title("CeIM1", ylab = "Frequência Observada")

```

Fonte: O autor, 2021.

Figura 17 –MAD para a análise de segunda ordem dos primeiros dígitos

```

bCeIM1=benford(CeIM1$Saldo, number.of.digits = 1, sign = "positive", discrete=TRUE, round=3)
mad1=(sum(abs(bCeIM1$bfd[,5]-bCeIM1$bfd[,8])/sum(bCeIM1$bfd[,8])))9

```

Fonte: O autor, 2021.

Figura 18 –MAD para a análise de segunda ordem dos dois primeiros dígitos

```

b2CeIM1=benford(CeIM1$Saldo, number.of.digits = 2, sign = "positive", discrete=TRUE, round=3)
mad12=(sum(abs(b2CeIM1$bfd[,5]-b2CeIM1$bfd[,8])/sum(b2CeIM1$bfd[,8])))90

```

Fonte: O autor, 2021.

Figura 19 – Análise da soma para os primeiros dígitos

```
bCeIM1=benford(CeIM1$Saldo, number.of.digits = 1, sign = "positive", discrete=TRUE, round=3)
bCeIM1$bfd[,c(1,9,10)]
plot(bCeIM1,except=c("digits","second order","mantissa","chi squared","abs diff"))
suspectsTable(bCeIM1,by="abs.excess.summation")
```

Fonte: O autor, 2021.

Figura 20 – Análise da soma para os dois primeiros dígitos

```
b2CeIM1=benford(CeIM1$Saldo, number.of.digits = 2, sign = "positive", discrete=TRUE, round=3)
b2CeIM1$bfd[,c(1,9,10)]
plot(b2CeIM1,except=c("digits","second order","mantissa","chi squared","abs diff"))
suspectsTable(b2CeIM1,by="abs.excess.summation")
```

Fonte: O autor, 2021.

Figura 21 – Teste dos números duplicados

```
bCeIM1=benford(CeIM1$Saldo, number.of.digits = 1, sign = "positive", discrete=TRUE, round=3)
duplicatesTable(bCeIM1)
getDuplicates(bCeIM1,CeIM1,how.many=2)
```

Fonte: O autor, 2021.

Figura 22 – Teste modelo fator de distorção

```
aux=which(CeIM1$Saldo<10)
if (length(aux)==0) FDCEIM1=CeIM1$Saldo else FDCEIM1=CeIM1$Saldo[-aux]
collapsed1=c()
for (i in 1:length(FDCEIM1))
collapsed1[i]=(10*FDCEIM1[i])/(10^(trunc(log10(FDCEIM1[i]))))
AM1=sum(collapsed1)/length(collapsed1)
DF1=(AM1-39.0865)/39.0865
Zteste1=DF1/(0.638253/sqrt(length(collapsed1)))
```

Fonte: O autor, 2021.

APÊNDICE B – Resultados das análises pré-Benford

Tabela 11 – Perfil de dados dos CeIM

	CeIM1	CeIM2	CeIM3	CeIM4	CeIM5	CeIM6	CeIM7
a	3	2	2	8	0	2	3
b	6	5	62	25	7	33	49
c	110	81	600	306	233	362	553
d	534	468	1384	1164	643	1012	1382
e	955	997	970	1328	491	772	1423
Total	1608	1553	3018	2831	1374	2181	3407

Fonte: O autor, 2021.

Tabela 12– Perfil de dados das EAM

	EAM1	EAM2	EAM3	EAM4
a	4	1	0	0
b	7	2	13	6
c	46	135	265	97
d	385	582	363	390
e	856	756	347	718
Total	1298	1476	988	1211

Fonte: O autor, 2021.

Tabela 13 – Perfil de dados das BE

	BE1	BE2	BE3	BE4	BE5	BE6	BE7	BE8
a	2	0	0	0	1	0	0	5
b	11	12	44	14	1	0	10	11
c	84	129	216	123	26	4	84	68
d	386	381	473	387	254	103	320	309
e	710	352	504	293	471	262	475	1306
Total	1193	874	1237	817	753	369	889	1699

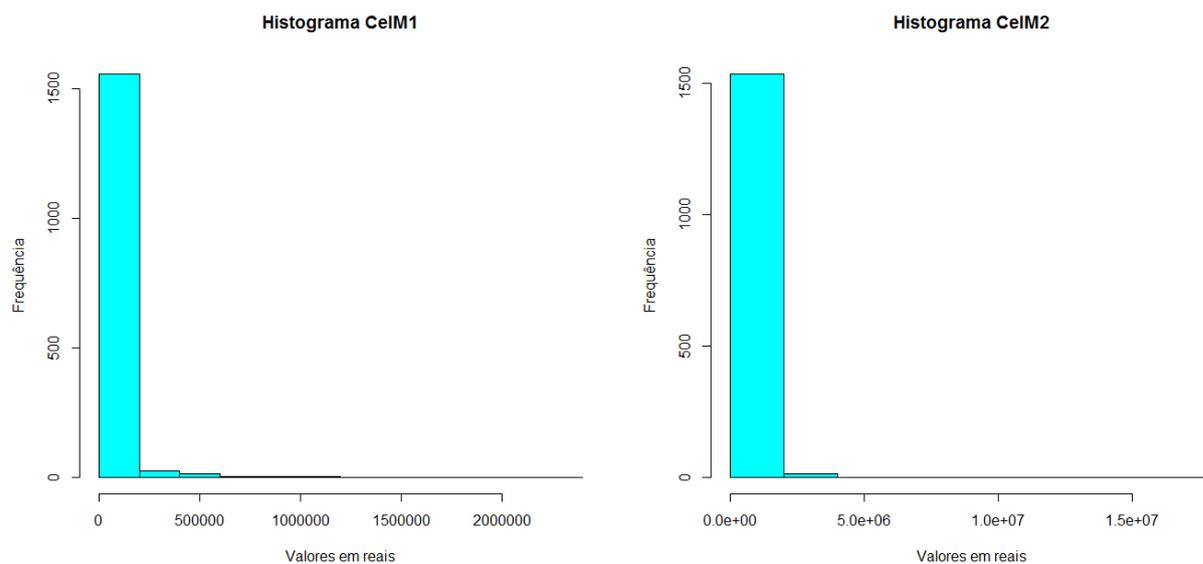
Fonte: O autor, 2021.

Tabela 14 – Perfil de dados dos HN

	HN1	HN2	HN3	HN4	HN5	HN6	HN7	HN8
a	0	8	2	0	0	0	10	1
b	0	19	3	4	2	2	0	9
c	22	163	54	121	72	104	59	103
d	108	1004	410	795	499	381	456	581
e	127	5185	1059	1557	1007	1317	974	1227
Total	257	6379	1528	2477	1580	1804	1499	1921

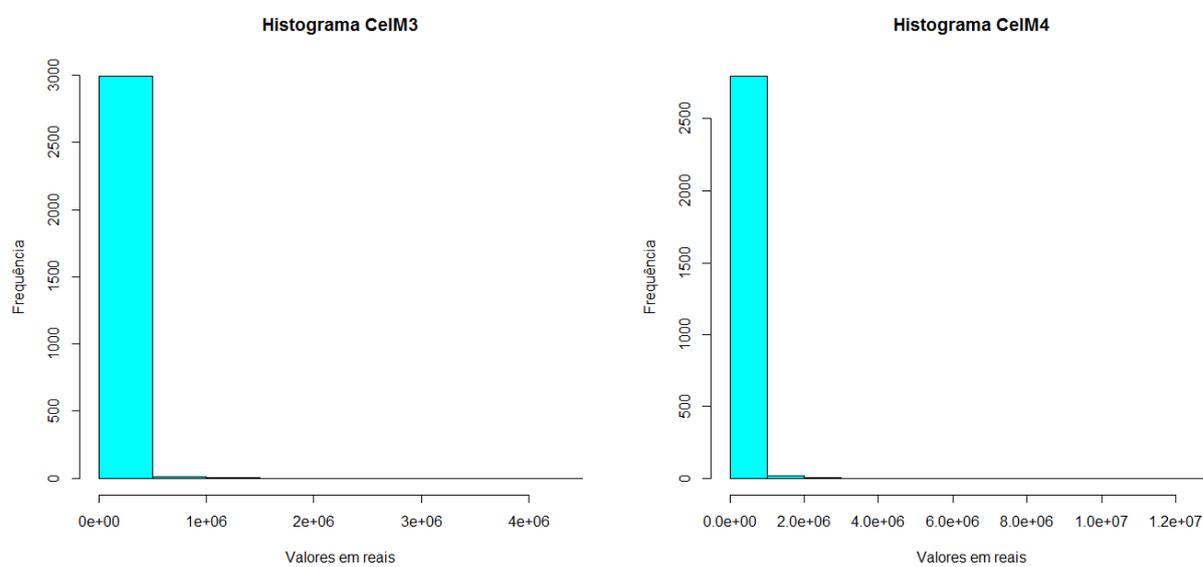
Fonte: O autor, 2021.

Figura 23 – Histogramas dos CeIM1 e CeIM2



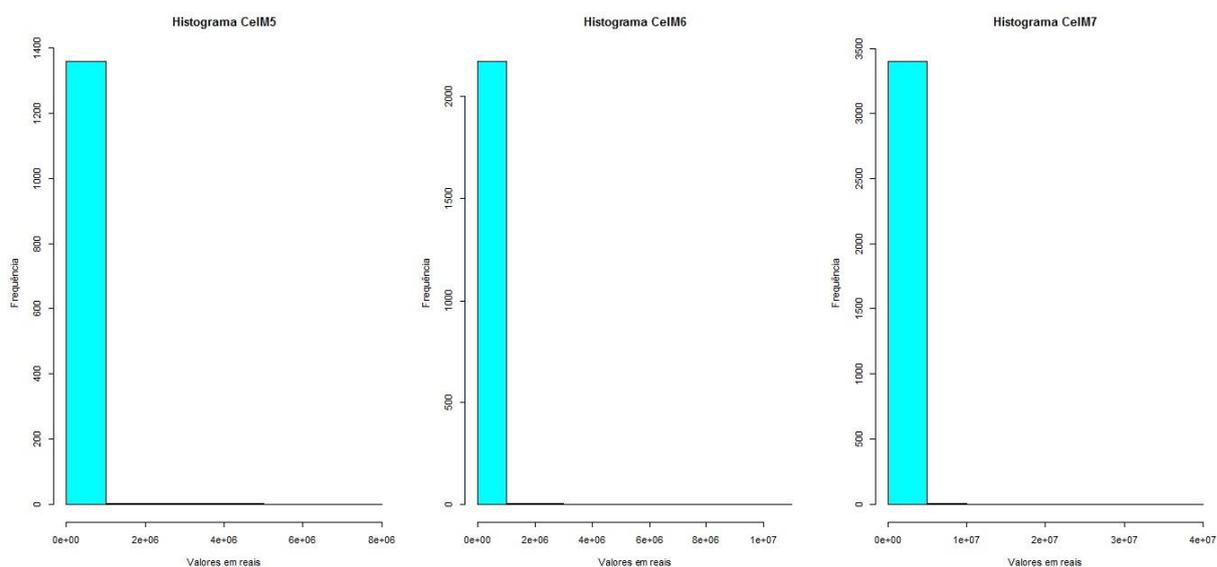
Fonte: O autor, 2021.

Figura 24 – Histogramas dos CeIM3 e CeIM4



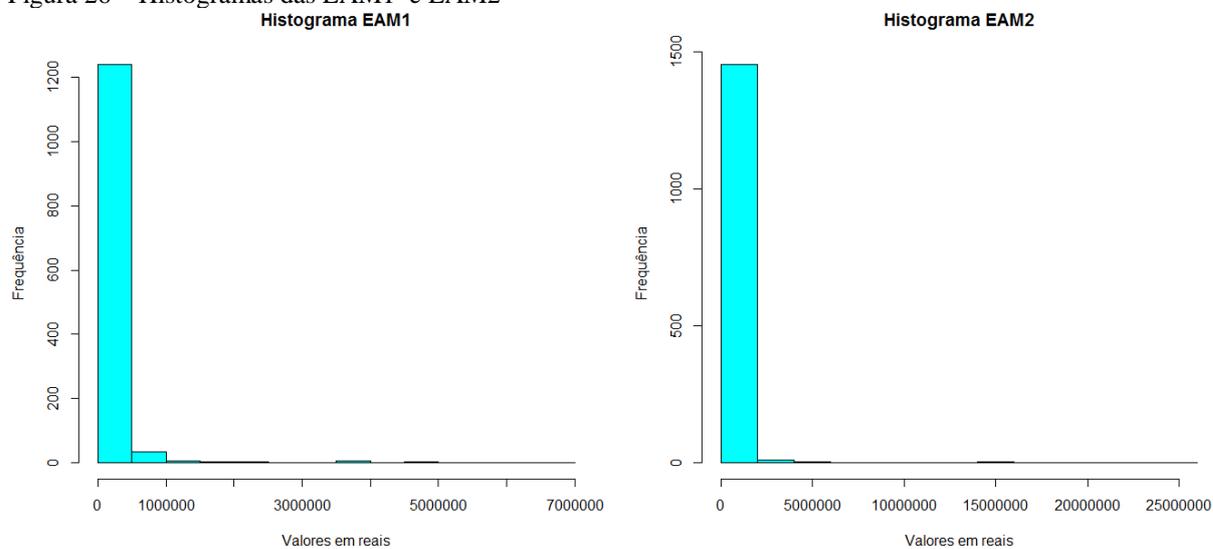
Fonte: O autor, 2021.

Figura 25 – Histogramas dos CeIM5 , CeIM6 e CeIM7



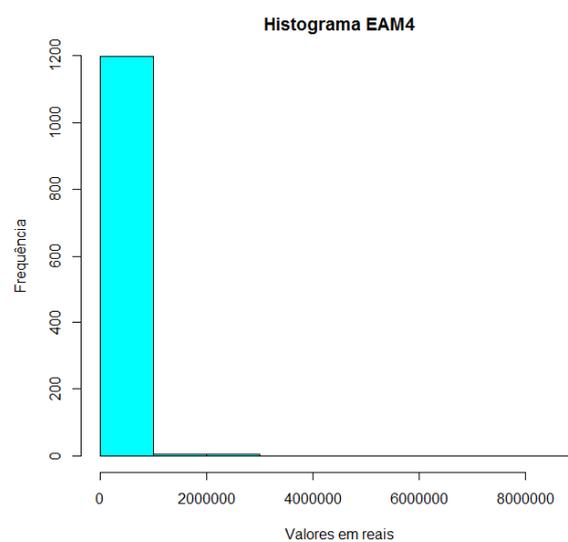
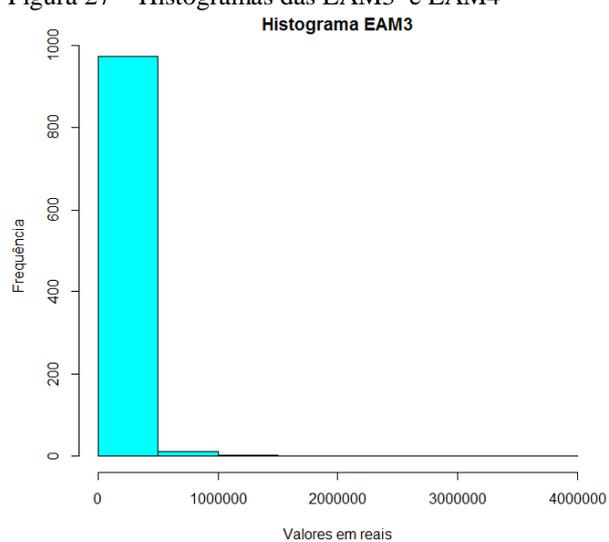
Fonte: O autor, 2021.

Figura 26 – Histogramas das EAM1 e EAM2



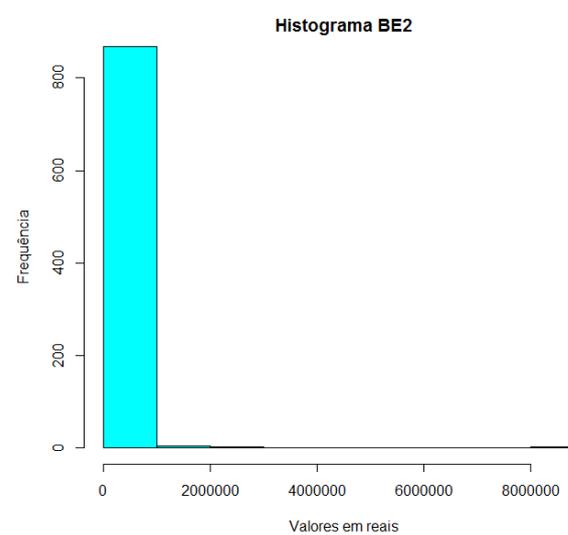
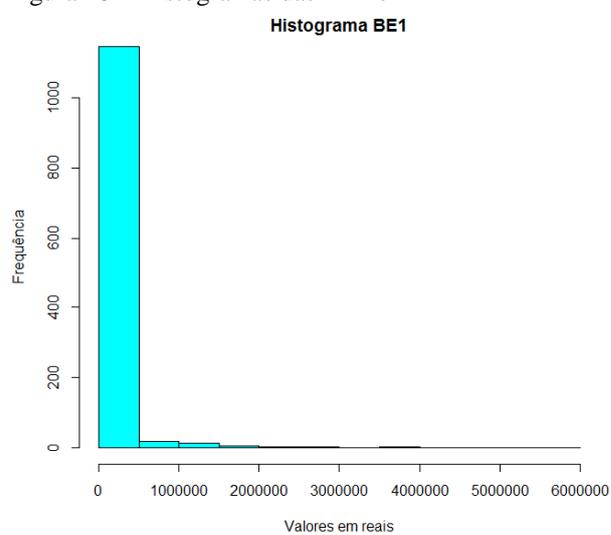
Fonte: O autor, 2021.

Figura 27 – Histogramas das EAM3 e EAM4



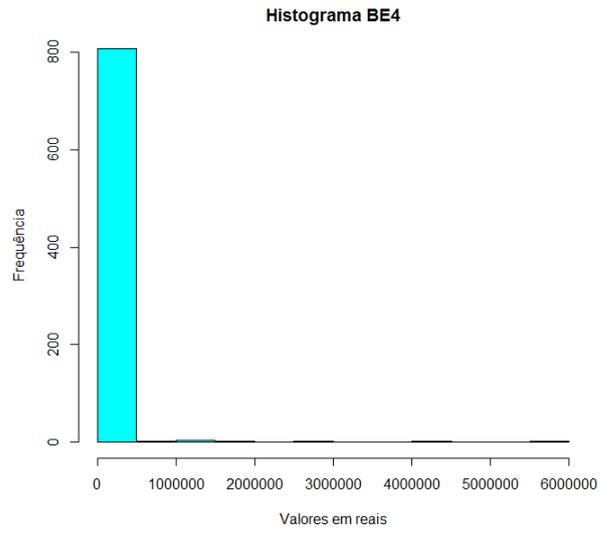
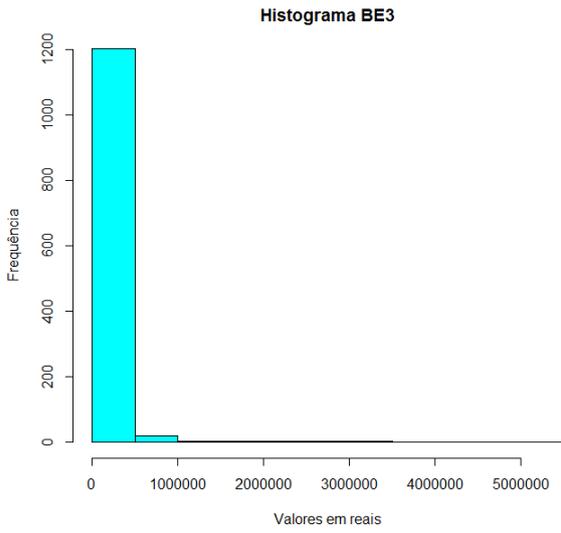
Fonte: O autor, 2021.

Figura 28 – Histogramas das BE1 e BE2



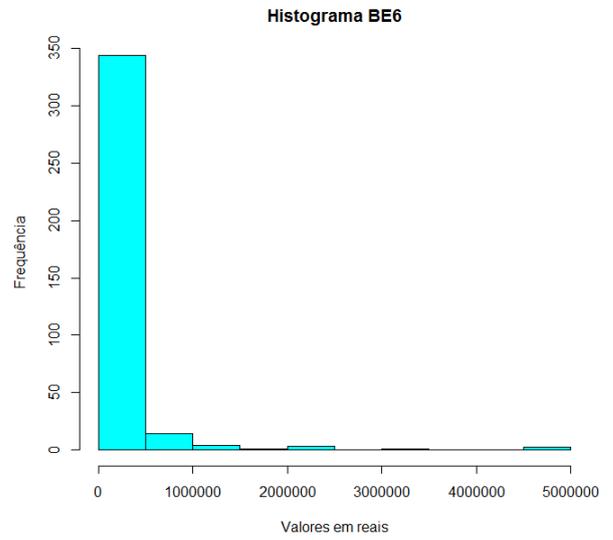
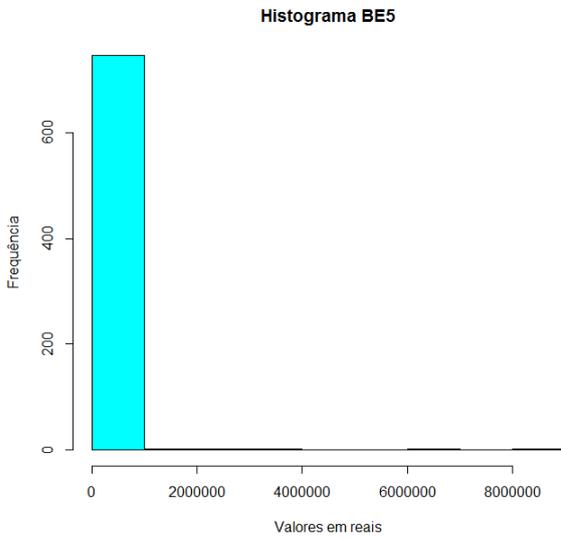
Fonte: O autor, 2021.

Figura 29 – Histogramas das BE3 e BE4



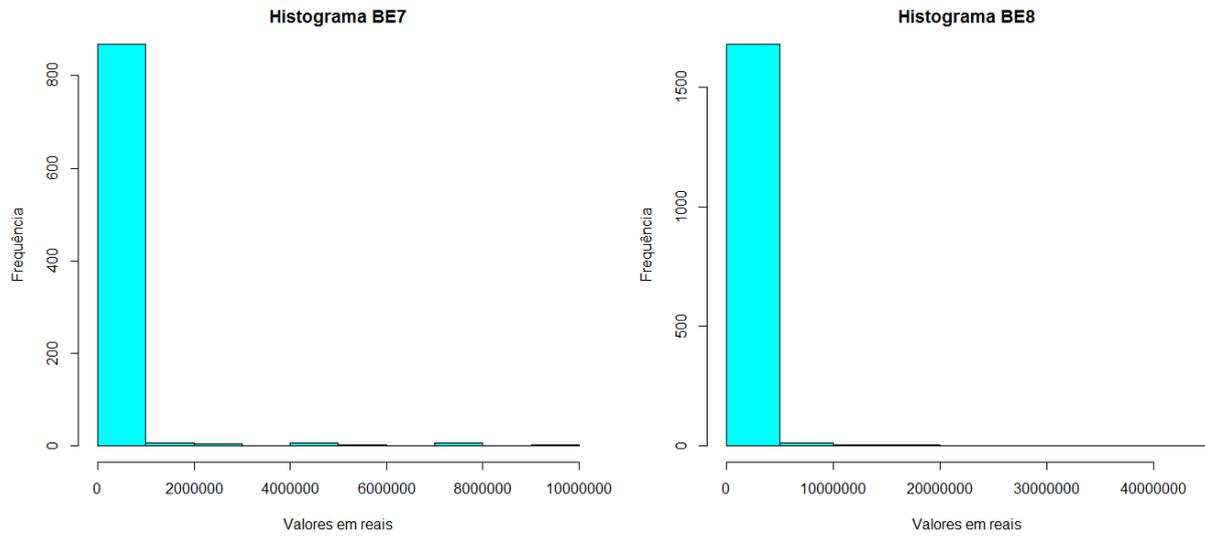
Fonte: O autor, 2021.

Figura 30 – Histogramas das BE5 e BE6



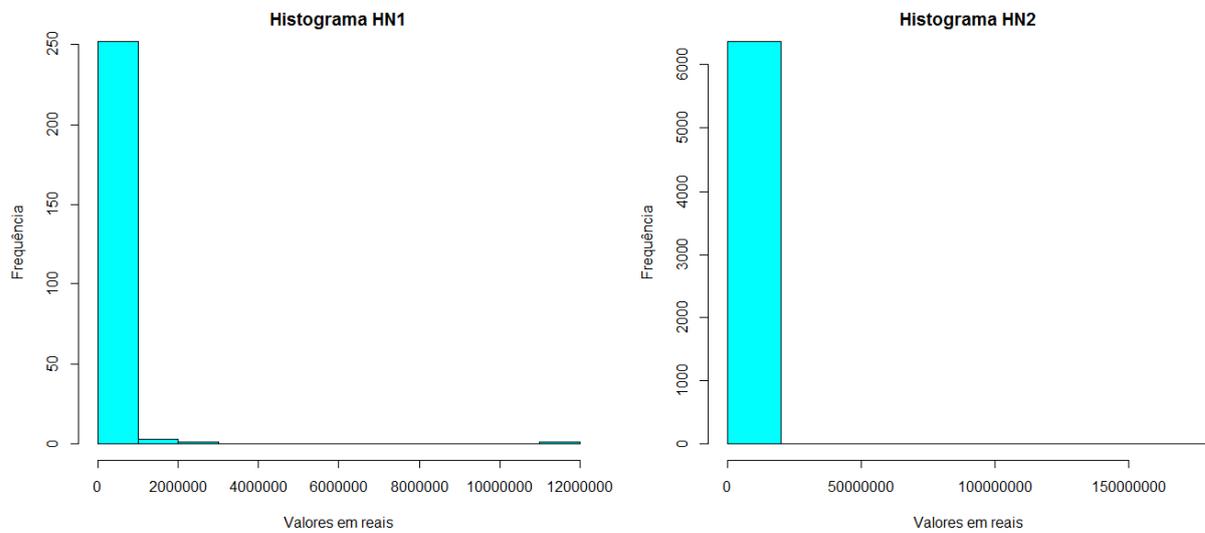
Fonte: O autor, 2021.

Figura 31 – Histogramas das BE7 e BE8



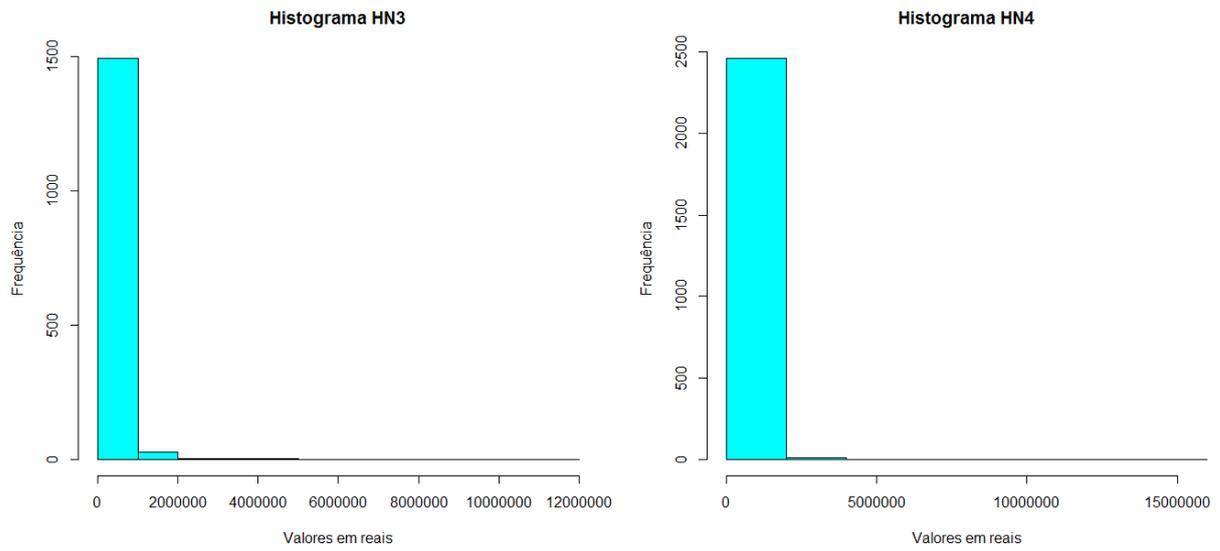
Fonte: O autor, 2021.

Figura 32 – Histogramas dos HN1 e HN2



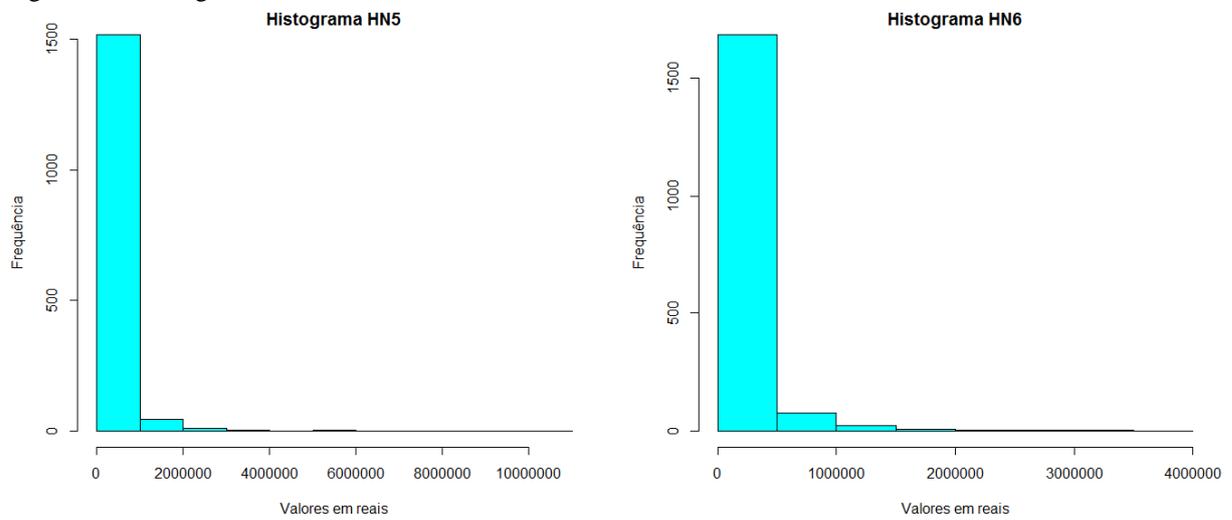
Fonte: O autor, 2021.

Figura 33 – Histogramas dos HN3 e HN4



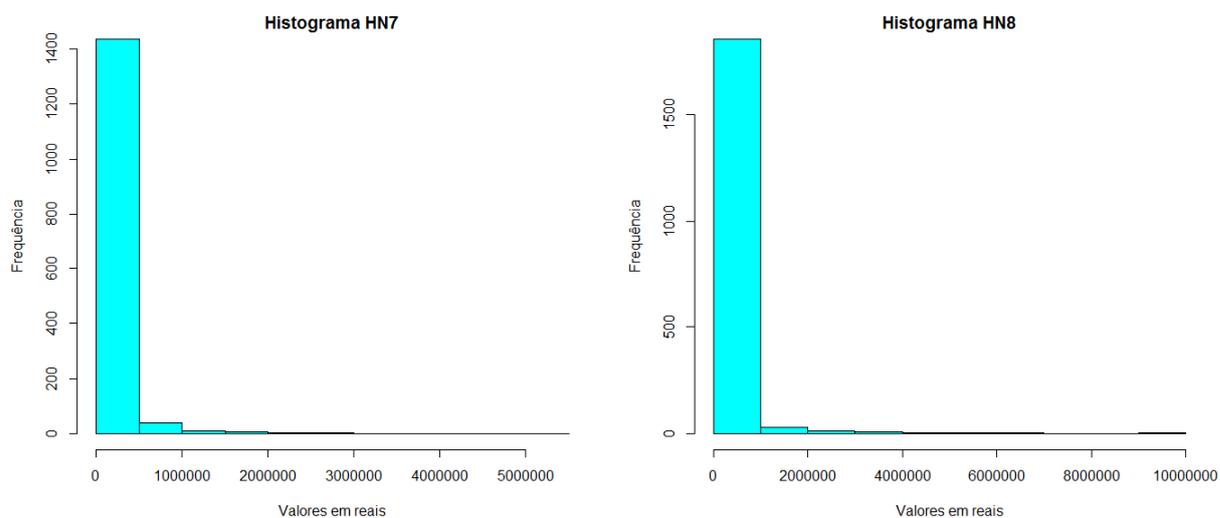
Fonte: O autor, 2021.

Figura 34 – Histogramas dos HN5 e HN6



Fonte: O autor, 2021.

Figura 35 – Histogramas dos HN7 e HN8



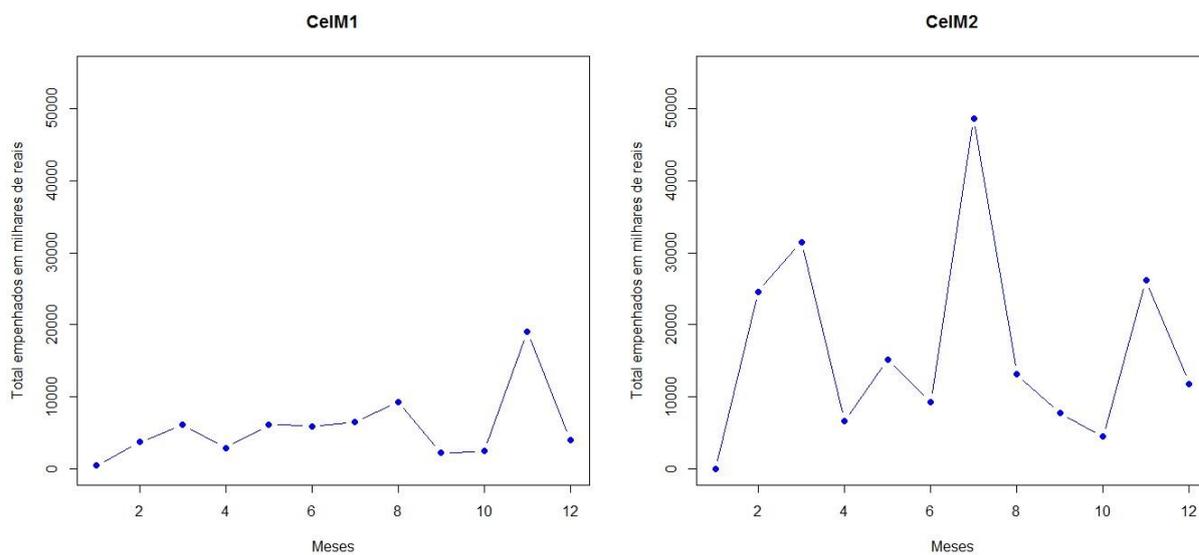
Fonte: O autor, 2021.

Tabela 15 – Estatística descritiva dos saldos das notas de empenho em reais

	Mínimo	1º Quartil	Média	Mediana	3º Quartil	Máximo
CeIM1	0,07	4.312,50	14.406,00	42.598,89	35.737,50	2.278.421,92
CeIM2	8,32	5.202,20	20.000,00	128.222,89	70.399,36	16.949.296,00
CeIM3	0,02	1.171,20	3.985,40	31.648,80	15.342,29	4.364.628,10
CeIM4	0,04	2.411,82	8.843,40	70.229,65	30.179,36	12.821.500,00
CeIM5	39,90	1.379,93	5.030,31	62.242,25	17.355,58	8.000.000,00
CeIM6	3,00	1.596,00	5.065,28	36.827,10	16.106,72	10.340.890,71
CeIM7	0,02	1.641,82	6.470,75	68.006,34	23.915,32	39.078.143,06
EAM1	1,68	6.152,86	21.487,96	126.483,46	79.133,13	6.633.895,32
EAM2	3,00	3.200,00	10.508,10	189.142,08	52.087,90	25.214.863,18
EAM3	38,24	716,48	4.101,00	39.542,63	18.590,50	3.914.807,88
EAM4	12,80	3.810,11	15.345,00	92.640,34	69.089,91	8.645.416,65
BE1	2,00	4.189,74	15.672,15	97.590,42	55.000,00	5.684.944,96
BE2	35,54	1.794,58	5.522,80	49.632,77	21.712,86	8.634.313,44
BE3	12,00	1.318,44	5.873,40	70.535,51	26.968,95	5.360.696,46
BE4	13,20	1.659,00	5.808,00	45.224,30	18.900,00	5.747.691,40
BE5	1,50	5.728,38	15.294,28	73.692,96	41.394,50	8.758.987,40
BE6	278,32	7.740,00	39.000,00	160.275,23	95.751,75	4.915.520,09
BE7	40,00	3.279,50	11.760,00	151.900,29	49.114,38	9.838.316,60
BE8	0,03	11.165,43	40.500,00	321.196,41	141.304,20	44.435.265,98
HN1	210,00	3.840,00	9.705,92	115.984,83	36.900,00	11.183.616,94
HN2	0,12	14.923,73	44.640,00	401.939,29	160.882,66	169.754.424,91
HN3	6,26	7.598,79	28.534,01	130.631,47	102.643,96	11.891.915,83
HN4	29,61	5.160,00	19.200,00	119.463,98	68.651,60	14.222.192,50
HN5	28,60	5.599,13	20.565,30	170.163,42	102.243,59	10.464.579,84
HN6	60,00	8.694,02	32.372,72	134.131,97	108.692,40	3.694.458,24
HN7	0,01	5.035,30	25.150,00	102.728,50	83.877,75	5.330.508,00
HN8	1,20	5.457,60	19.860,12	173.995,65	78.240,00	9.961.970,31

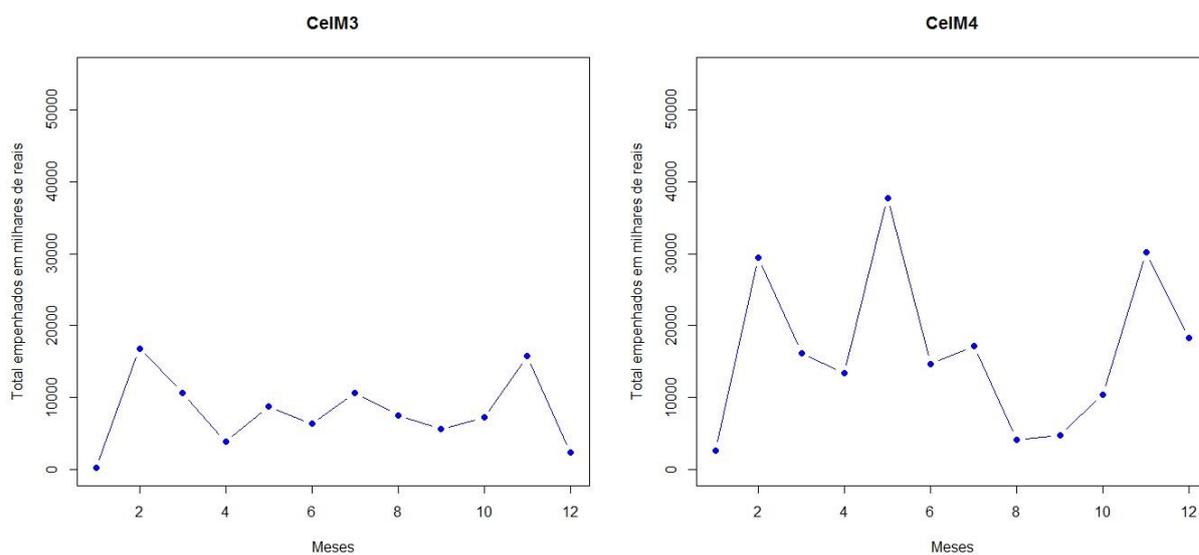
Fonte: O autor, 2021.

Figura 36 – Gráfico valor empenho mensal dos CeIM1 e CeIM2



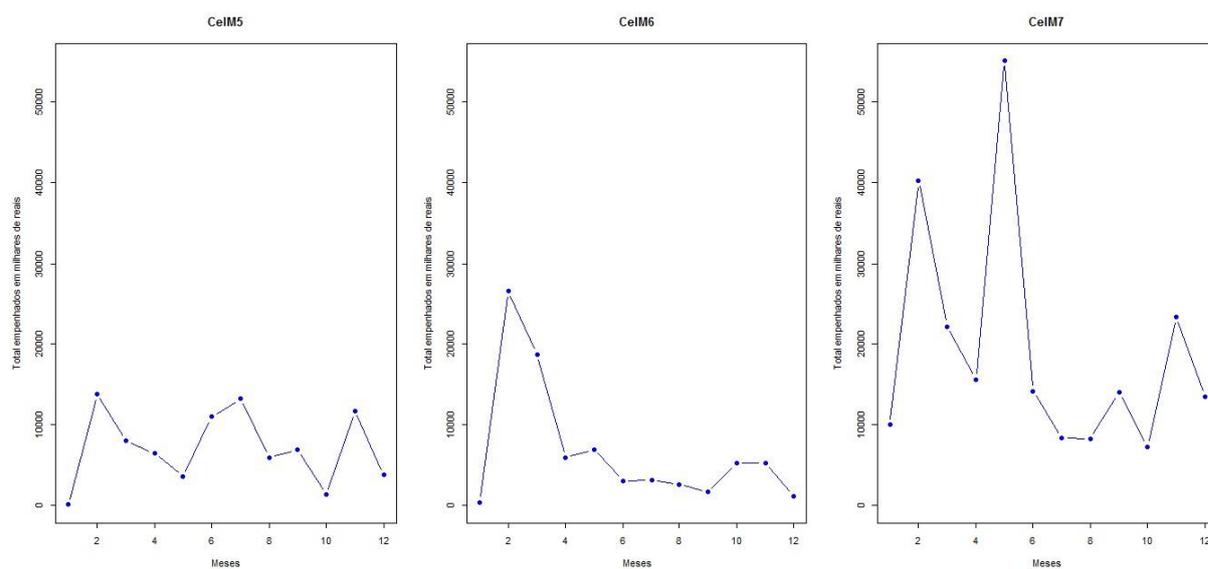
Fonte: O autor, 2021.

Figura 37 – Gráfico valor empenho mensal dos CeIM3 e CeIM4



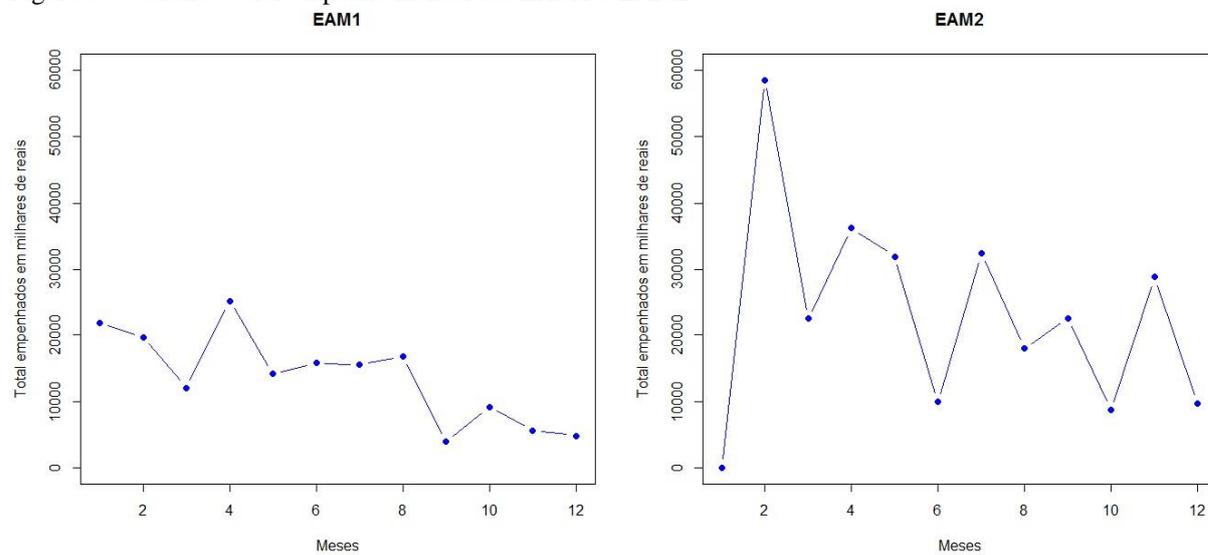
Fonte: O autor, 2021.

Figura 38 – Gráfico valor empenho mensal dos CeIM5, CeIM7 e CeIM8



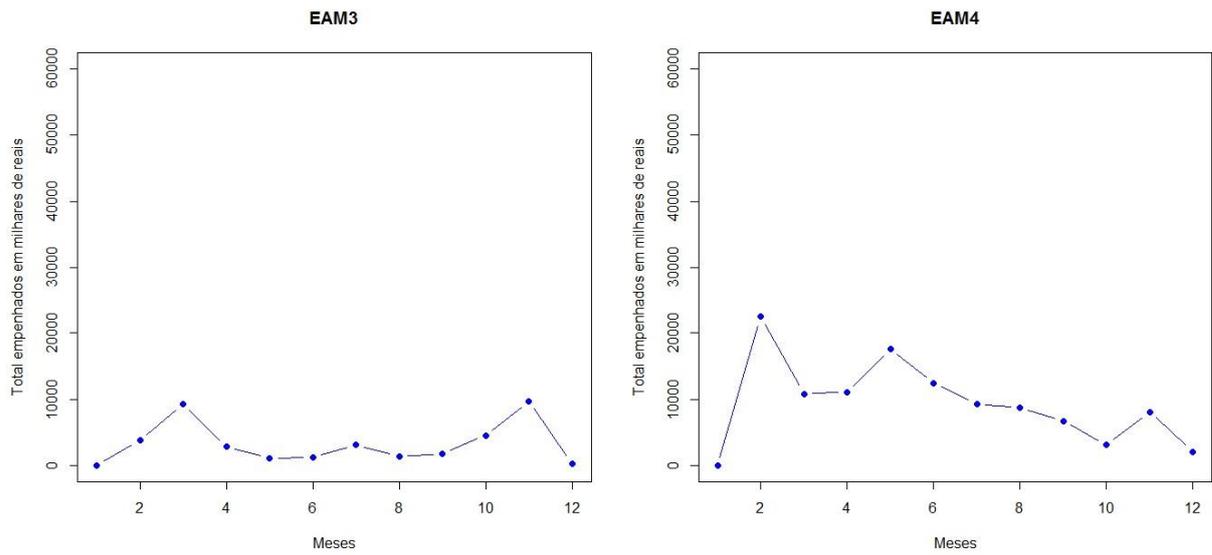
Fonte: O autor, 2021.

Figura 39 – Gráfico valor empenho mensal das EAM1 e EAM2



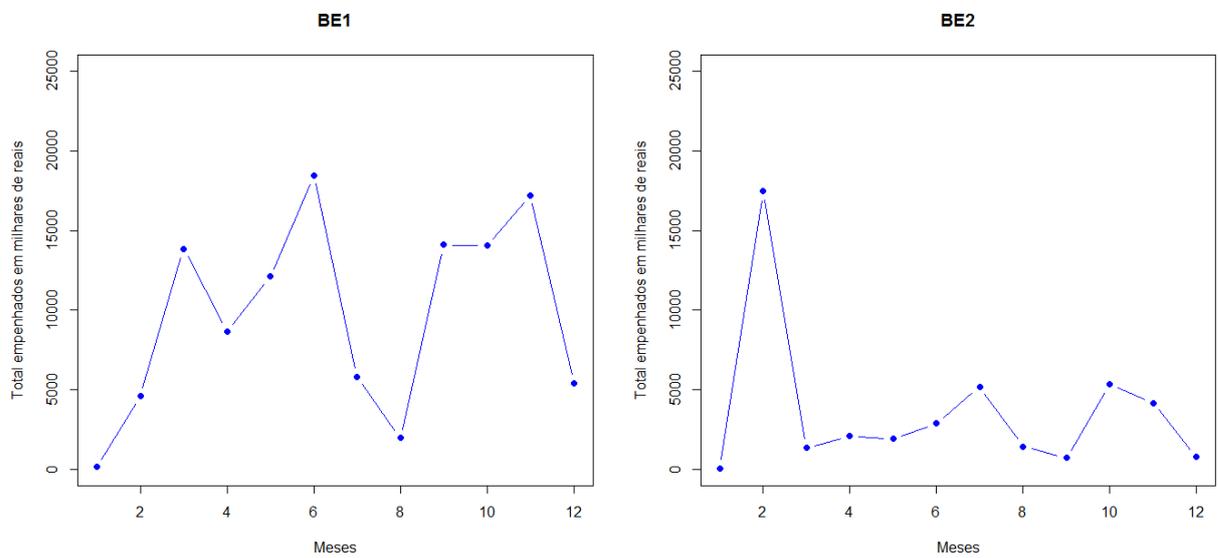
Fonte: O autor, 2021.

Figura 40 – Gráfico valor empenho mensal das EAM3 e EAM4



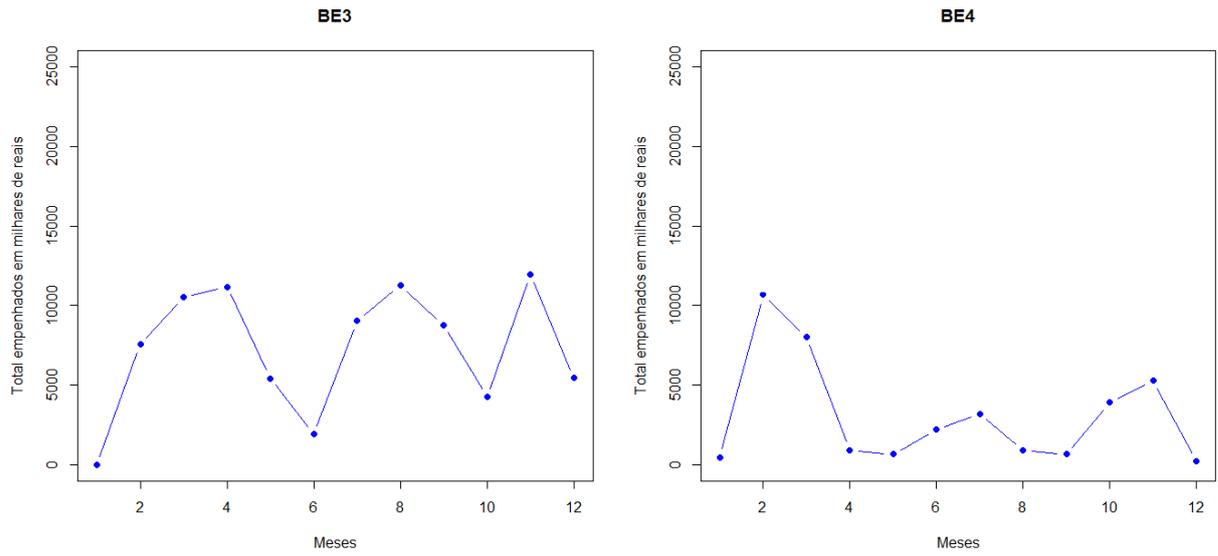
Fonte: O autor, 2021.

Figura 41 – Gráfico valor empenho mensal das BE1 e BE2



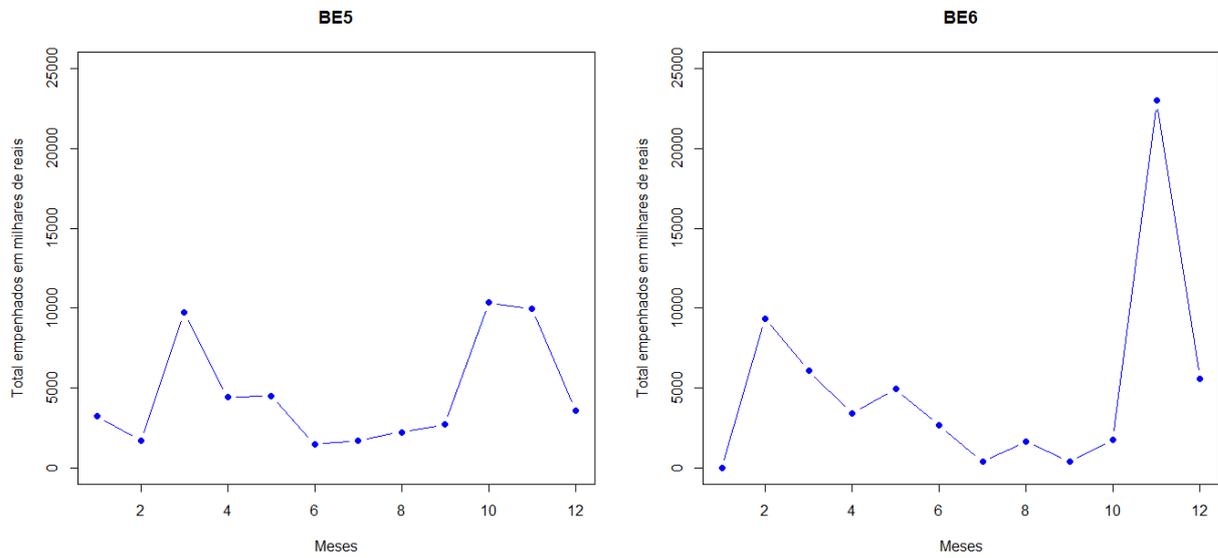
Fonte: O autor, 2021.

Figura 42 – Gráfico valor empenho mensal das BE3 e BE4



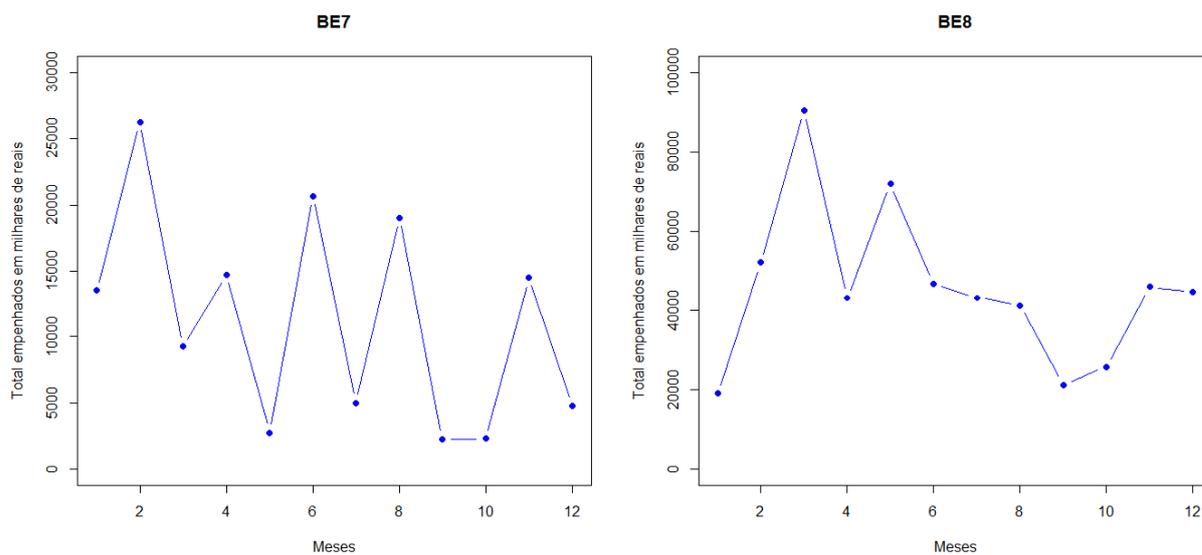
Fonte: O autor, 2021.

Figura 43 – Gráfico valor empenho mensal das BE5 e BE6



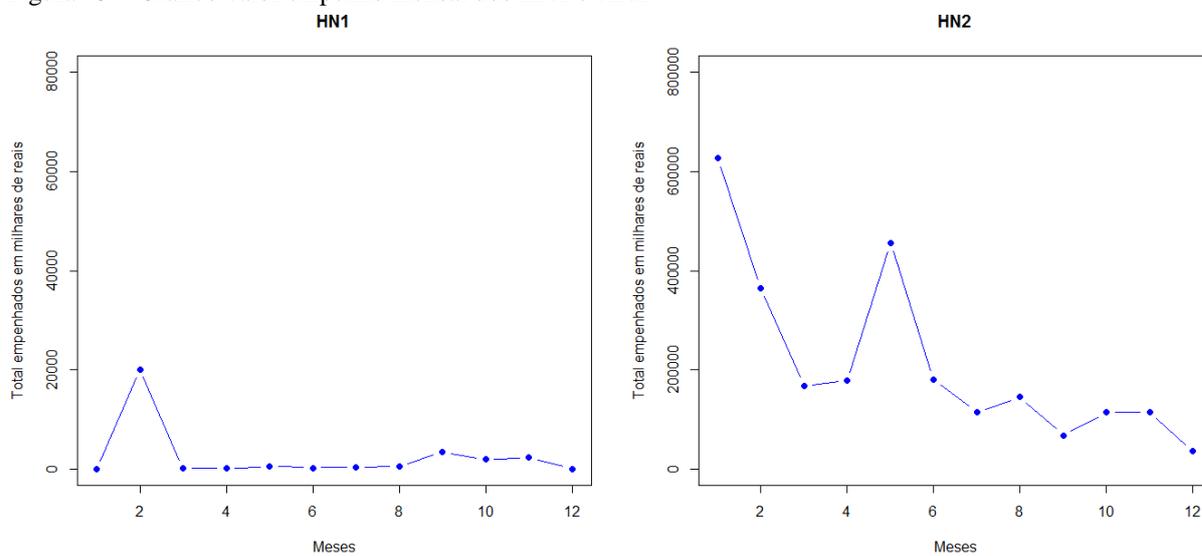
Fonte: O autor, 2021.

Figura 44 – Gráfico valor empenho mensal das BE7 e BE8



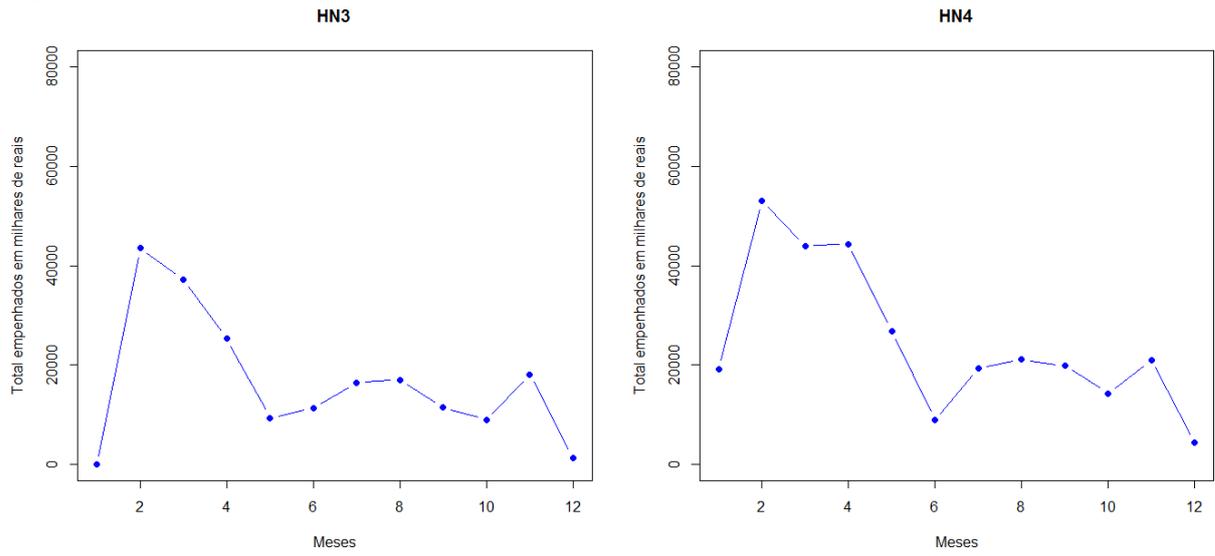
Fonte: O autor, 2021.

Figura 45 – Gráfico valor empenho mensal dos HN1 e HN2



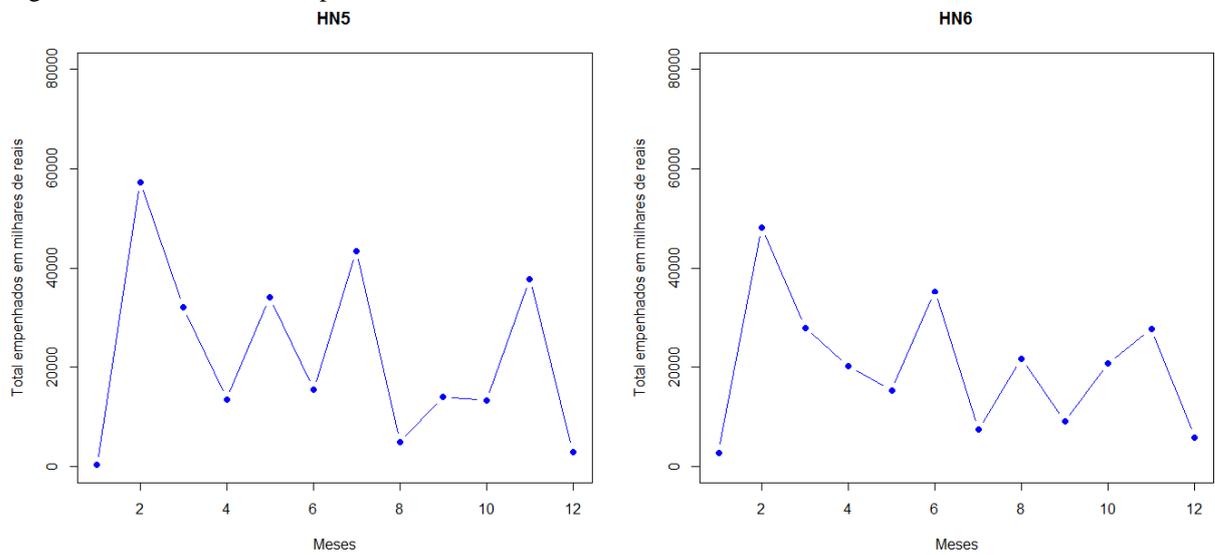
Fonte: O autor, 2021.

Figura 46 – Gráfico valor empenho mensal dos HN3 e HN4



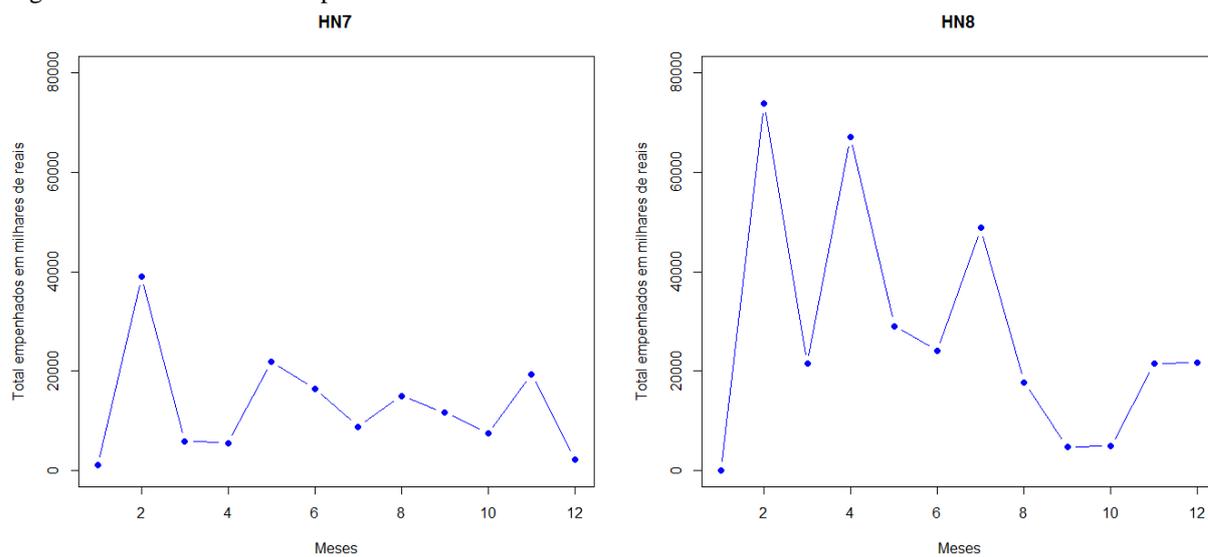
Fonte: O autor, 2021.

Figura 47 – Gráfico valor empenho mensal dos HN5 e HN6



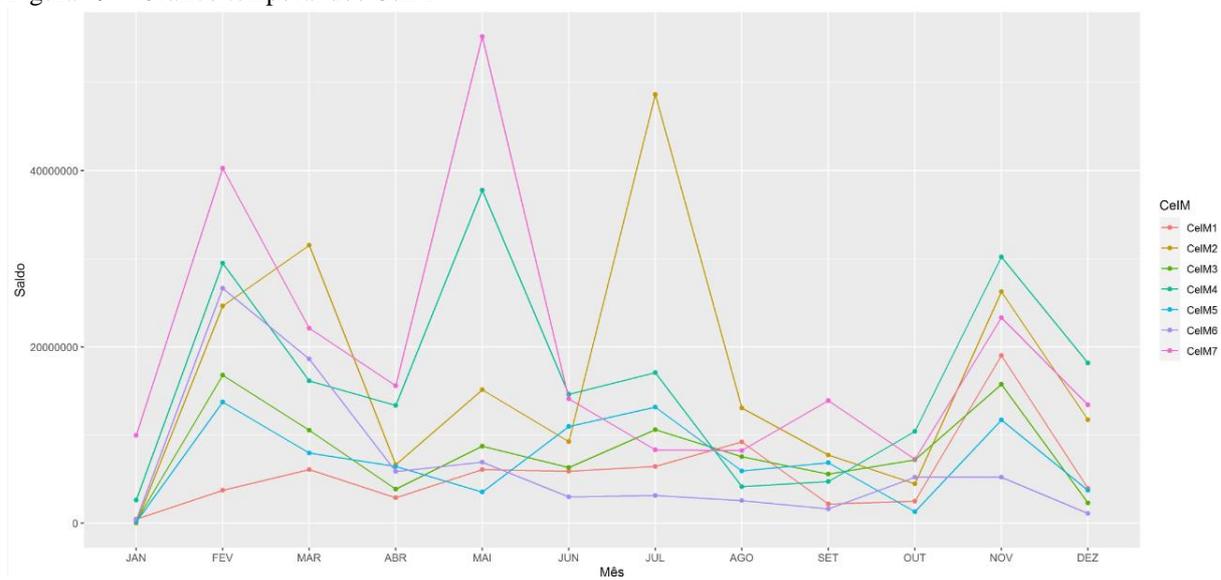
Fonte: O autor, 2021.

Figura 48 – Gráfico valor empenho mensal dos HN7 e HN8



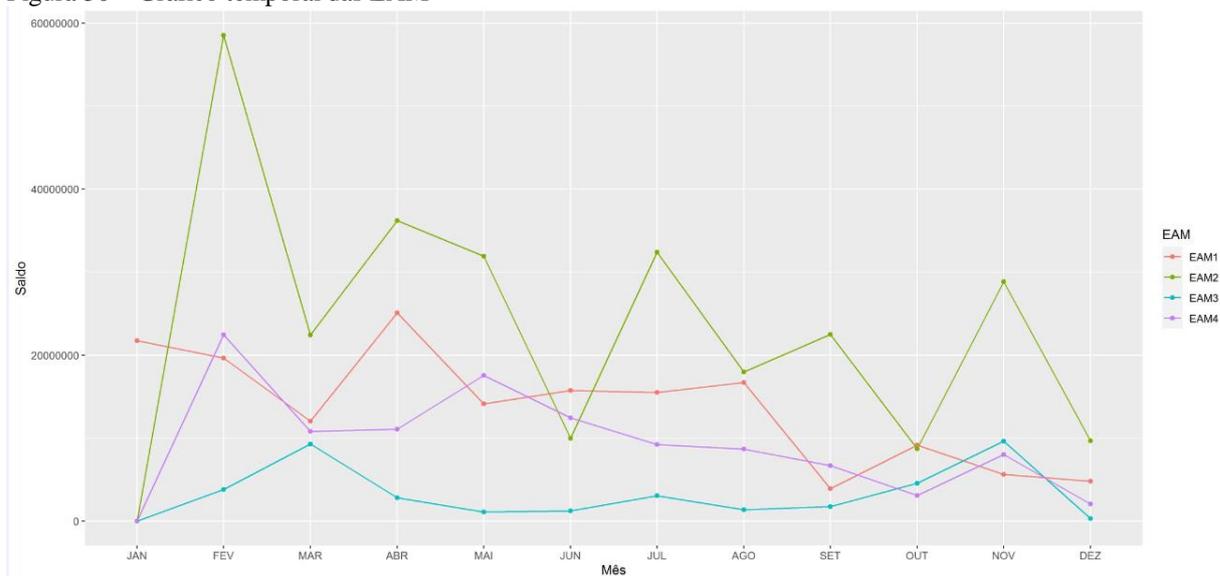
Fonte: O autor, 2021.

Figura 49 – Gráfico temporal dos CeIM



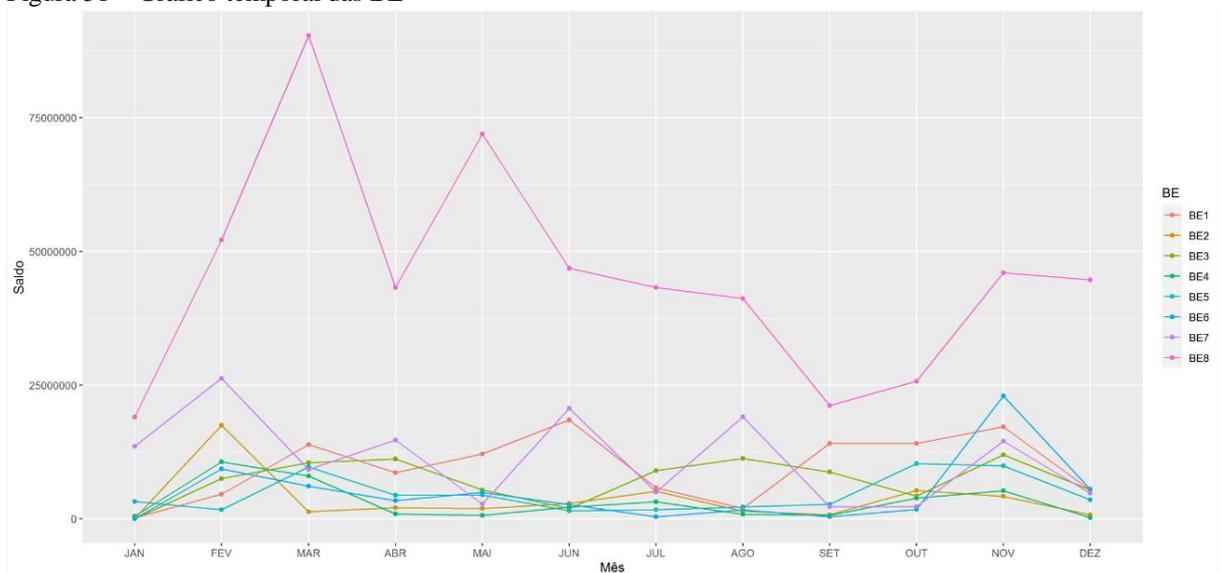
Fonte: O autor, 2021.

Figura 50 – Gráfico temporal das EAM



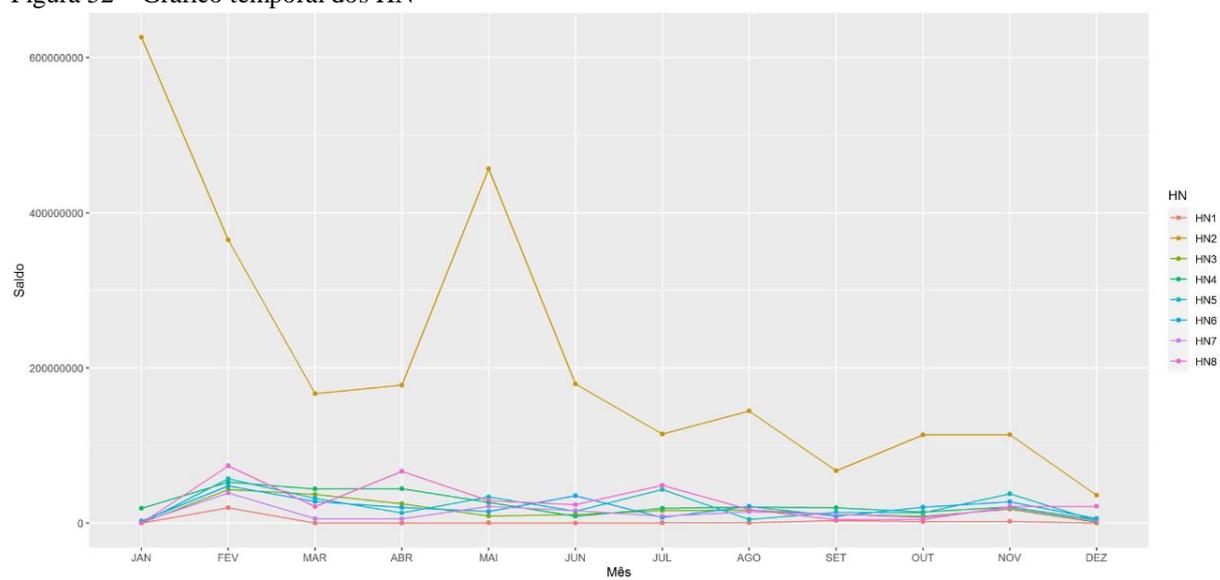
Fonte: O autor, 2021.

Figura 51 – Gráfico temporal das BE



Fonte: O autor, 2021.

Figura 52 – Gráfico temporal dos HN



Fonte: O autor, 2021.

APÊNDICE C – Resultados das análises primárias

Tabela 16 – Análise global para os primeiros dígitos da LNB

	χ^2	MAD	K-S
CeIM1	10.58	0.0081	0.895
CeIM2	13.10	0.0089	1.002
CeIM3	14.29	0.0053	0.920
CeIM4	9.09	0.0047	0.394
CeIM5	15.31	0.0100	0.604
CeIM6	6.42	0.0039	0.417
CeIM7	12.24	0.0052	0.949
EAM1	8,72	0,0066	0,478
EAM2	14,71	0,0081	1,264
EAM3	63,41	0,0259	2,073
EAM4	18,14	0,0121	0,961
BE1	4,95	0,0068	0,691
BE2	8,66	0,0079	1,042
BE3	6,31	0,0069	0,444
BE4	18,45	0,0107	0,661
BE5	10,29	0,0097	1,008
BE6	16,76	0,0159	0,889
BE7	8,19	0,0075	0,986
BE8	6,45	0,0056	0,538
HN1	15,64	0,0208	0,896
HN2	30,82	0,0049	1,191
HN3	4,18	0,0043	0,334
HN4	13,30	0,0064	0,838
HN5	3,92	0,0040	0,572
HN6	9,43	0,0062	0,571
HN7	5,74	0,0050	0,736
HN8	8,24	0,0060	0,600

Fonte: O autor, 2021.

Tabela 17 – P-valor do teste Z por primeiro dígito dos CeIM

Primeiro dígito	CeIM1	CeIM2	CeIM3	CeIM4	CeIM5	CeIM6	CeIM7
1	0,175	0,029	0,416	0,646	0,188	0,872	0,039
2	0,209	0,005	0,517	0,748	0,020	0,864	0,138
3	0,023	0,166	0,953	0,135	0,242	0,821	0,931
4	0,279	0,764	0,002	0,831	0,726	0,236	0,809
5	0,498	0,853	0,588	0,069	0,328	0,077	0,010
6	0,738	0,475	0,660	0,476	0,281	0,391	0,636
7	0,650	0,669	0,139	0,430	0,014	0,820	0,859
8	0,202	0,384	0,306	0,274	0,600	0,227	0,832
9	0,669	0,909	0,254	0,193	0,529	0,774	0,321

Fonte: O autor, 2021.

Tabela 18 – P-valor do teste Z por primeiro dígito das EAM

Primeiro dígito	EAM1	EAM2	EAM3	EAM4
1	0,774	0,558	0,002	0,036
2	0,118	0,342	0,000	0,318
3	0,989	0,369	0,000	0,184
4	0,941	0,657	0,034	0,014
5	0,157	0,448	0,150	0,449
6	0,152	0,046	0,072	0,297
7	0,456	0,068	0,522	0,029
8	0,226	0,516	0,070	0,701
9	0,312	0,021	0,088	0,297

Fonte: O autor, 2021.

Tabela 19 – P-valor do teste Z por primeiro dígito das BE

Primeiro dígito	BE1	BE2	BE3	BE4	BE5	BE6	BE7	BE8
1	0,132	0,994	0,520	0,254	0,125	0,053	0,084	0,353
2	0,320	0,472	0,325	0,704	0,954	0,492	0,968	0,334
3	0,433	0,020	0,573	0,393	0,325	0,987	0,834	0,899
4	0,724	0,320	0,254	0,162	0,714	0,693	0,638	0,396
5	0,792	0,598	0,113	0,319	0,244	0,466	0,021	0,619
6	0,497	0,155	0,438	0,966	0,503	0,632	0,948	0,198
7	0,636	0,698	0,604	0,722	0,015	0,755	0,621	0,714
8	0,356	0,793	0,460	0,000	0,561	0,002	0,400	0,126
9	0,525	0,628	0,723	0,285	0,800	0,049	0,832	0,582

Fonte: O autor, 2021.

Tabela 20 – P-valor do teste Z por primeiro dígito dos HN

Primeiro dígito	HN1	HN2	HN3	HN4	HN5	HN6	HN7	HN8
1	0,051	0,708	0,825	0,439	0,885	0,258	0,833	0,191
2	0,775	0,811	0,594	0,604	0,297	0,004	0,377	0,777
3	0,721	0,037	0,482	0,031	0,465	0,550	0,360	0,053
4	0,001	0,826	0,223	0,343	0,123	0,948	0,156	0,310
5	0,217	0,328	0,449	0,367	0,918	0,941	0,977	0,793
6	0,582	0,003	0,861	0,328	0,561	0,794	0,631	0,688
7	0,770	0,005	0,860	0,754	0,798	0,443	0,918	0,876
8	0,543	0,002	0,208	0,059	0,925	0,691	0,137	0,187
9	0,206	0,432	0,814	0,095	0,933	0,395	0,657	0,331

Fonte: O autor, 2021.

Tabela 21 – Análise global para os segundos dígitos da LNB

	χ^2	MAD _{teste}	K-S _{teste}	K-S _{crítico}
CeIM1	30,76	0,0107	0,029	0,034
CeIM2	20,28	0,0082	0,034	0,035
CeIM3	3,04	0,0028	0,009	0,025
CeIM4	16,30	0,0062	0,016	0,026
CeIM5	40,31	0,0130	0,038	0,037
CeIM6	9,49	0,0061	0,014	0,029
CeIM7	10,26	0,0046	0,009	0,023
EAM1	18,99	0,0091	0,014	0,038
EAM2	68,79	0,0190	0,034	0,035
EAM3	72,56	0,0243	0,068	0,043
EAM4	2,90	0,0042	0,009	0,039
BE1	24,03	0,0111	0,044	0,039
BE2	6,96	0,0074	0,017	0,046
BE3	9,99	0,0075	0,016	0,039
BE4	15,36	0,0096	0,038	0,048
BE5	6,90	0,0084	0,031	0,050
BE6	3,78	0,0088	0,016	0,071
BE7	6,95	0,0069	0,019	0,046
BE8	11,49	0,0064	0,020	0,033
HN1	3,45	0,0088	0,027	0,085
HN2	11,55	0,0031	0,007	0,017
HN3	10,55	0,0063	0,027	0,035
HN4	7,57	0,0047	0,023	0,027
HN5	6,11	0,0052	0,018	0,034
HN6	64,16	0,0119	0,059	0,032
HN7	17,00	0,0085	0,031	0,035
HN8	16,99	0,0076	0,013	0,031

Fonte: O autor, 2021.

Tabela 22 – Teste Z para os segundos dígitos dos CeIM

Segundo dígito	CeIM1	CeIM2	CeIM3	CeIM4	CeIM5	CeIM6	CeIM7
0	1,81	4,08	0,38	2,62	-1,62	1,25	1,60
1	-1,89	-0,39	-0,61	-0,26	-1,15	-1,04	-1,62
2	-0,96	-1,14	0,56	-0,73	0,39	1,49	0,23
3	-0,23	-0,25	-0,23	-1,07	-2,06	0,31	-1,76
4	3,96	0,44	-0,71	0,31	5,04	-0,48	1,10
5	1,23	-1,13	-0,73	1,48	-1,26	-0,93	0,04
6	-0,10	-0,70	-0,30	-2,22	-1,33	1,20	0,11
7	-1,33	-0,29	0,34	-1,30	-0,39	-1,50	-1,01
8	-2,72	0,54	1,08	0,14	2,74	-0,83	0,77
9	0,03	-1,55	0,29	0,83	-0,08	0,35	0,58

Fonte: O autor, 2021.

Tabela 23 – Teste Z para os segundos dígitos das EAM

Segundo dígito	EAM1	EAM2	EAM3	EAM4
0	1,25	2,03	1,64	-0,17
1	-2,52	-4,27	-2,56	0,55
2	0,07	2,20	2,09	-0,72
3	0,05	-2,64	-2,40	-0,22
4	0,72	5,02	4,86	0,15
5	-1,27	-0,50	-0,49	-0,59
6	0,36	2,70	4,02	0,88
7	3,07	-2,30	-3,58	-0,34
8	-0,65	-0,30	-1,41	-0,41
9	-1,03	-2,10	-2,51	0,93

Fonte: O autor, 2021.

Tabela 24 – Teste Z para os segundos dígitos das BE

Segundo dígito	BE1	BE2	BE3	BE4	BE5	BE6	BE7	BE8
0	3,68	0,98	0,52	0,02	1,22	-0,51	1,20	-0,10
1	-0,54	-0,06	-0,08	1,21	0,72	0,98	-0,55	-2,48
2	1,60	0,10	0,31	-0,10	0,71	-0,69	-0,51	0,24
3	-0,61	-1,02	-1,68	-0,83	-1,50	-0,77	0,25	-0,10
4	-2,38	-1,65	0,94	3,27	0,42	0,00	-0,47	0,61
5	-0,23	0,63	-1,88	-1,77	-0,47	0,06	0,01	1,37
6	0,46	0,16	1,42	-0,64	-0,54	0,28	-1,85	1,20
7	0,32	-0,94	-0,77	-0,71	-1,40	0,85	-0,04	-1,23
8	-0,97	1,01	0,77	-0,31	0,14	0,68	0,85	-0,41
9	-1,81	0,81	0,50	-0,31	0,52	-0,81	1,13	1,10

Fonte: O autor, 2021.

Tabela 25 – Teste Z para os segundos dígitos dos HN

Segundo dígito	HN1	HN2	HN3	HN4	HN5	HN6	HN7	HN8
0	1,01	-0,25	2,53	0,59	-0,39	7,70	1,40	1,20
1	0,34	-0,81	-0,16	1,70	1,11	-0,92	0,43	-2,93
2	-0,59	0,48	0,88	0,42	0,49	0,13	1,90	0,73
3	0,45	-0,84	-1,37	0,17	-0,15	-2,40	-2,91	-0,85
4	-1,41	0,13	-0,79	0,84	1,30	-1,17	-0,72	2,38
5	0,24	0,99	-1,27	-1,46	-1,43	-0,03	0,27	-0,36
6	0,00	0,06	0,03	-0,57	0,13	-1,49	0,54	-0,19
7	0,17	-1,50	-0,54	-1,11	-1,03	-0,08	-0,13	1,15
8	-0,33	2,67	0,47	-0,14	0,32	-1,41	0,80	-0,50
9	0,03	-0,82	0,01	-0,76	-0,48	-1,04	-1,80	-0,60

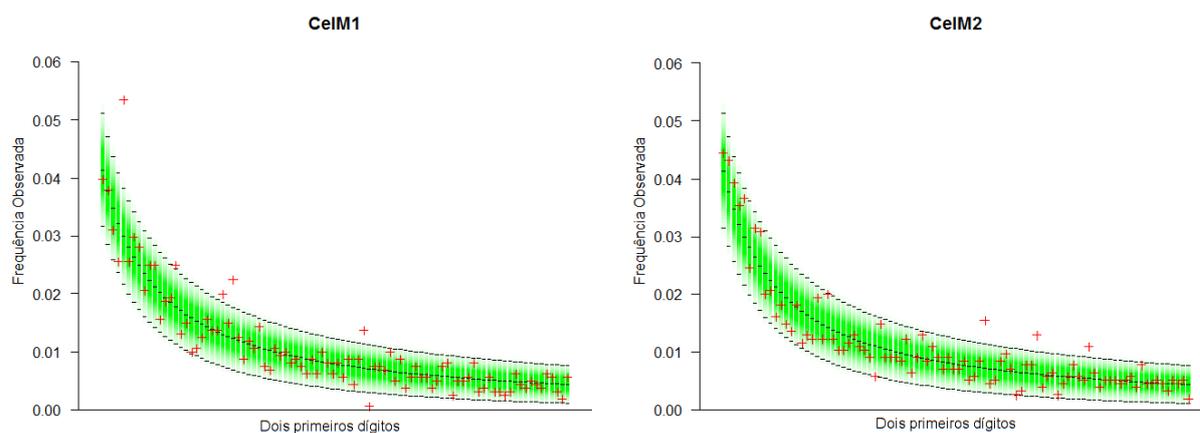
Fonte: O autor, 2021.

Tabela 26 – Análise global para os dois primeiros dígitos da LNB

	χ^2	MAD	K-S	<i>Mantissa Arc Test</i>
CeIM1	134.76	0,0023	1.039	0.339
CeIM2	108.10	0,0022	1.207	0.016
CeIM3	148.61	0,0017	1.028	0.039
CeIM4	118.46	0,0017	0.473	0.936
CeIM5	435.78	0,0037	1.432	0.111
CeIM6	116.44	0,0020	0.639	0.743
CeIM7	117.47	0,0015	0.964	0.100
EAM1	106,71	0,0024	0,605	0,425
EAM2	295,02	0,0036	1,406	0,026
EAM3	748,00	0,0050	2,374	0,000
EAM4	71,74	0,0021	0,961	0,025
BE1	120,00	0,0027	1,010	0,710
BE2	103,00	0,0030	1,040	0,086
BE3	110,00	0,0025	0,562	0,920
BE4	252,00	0,0032	0,738	0,490
BE5	117,00	0,0036	1,160	0,410
BE6	266,00	0,0057	1,040	0,420
BE7	146,00	0,0035	1,430	0,099
BE8	96,50	0,0021	0,546	0,990
HN1	109,10	0,0051	0,936	0,196
HN2	163,00	0,0013	1,214	0,170
HN3	85,99	0,0019	0,460	0,817
HN4	88,67	0,0015	1,116	0,052
HN5	86,24	0,0020	0,572	0,860
HN6	225,00	0,0028	0,838	0,108
HN7	101,70	0,0022	0,912	0,356
HN8	106,60	0,0020	0,771	0,679

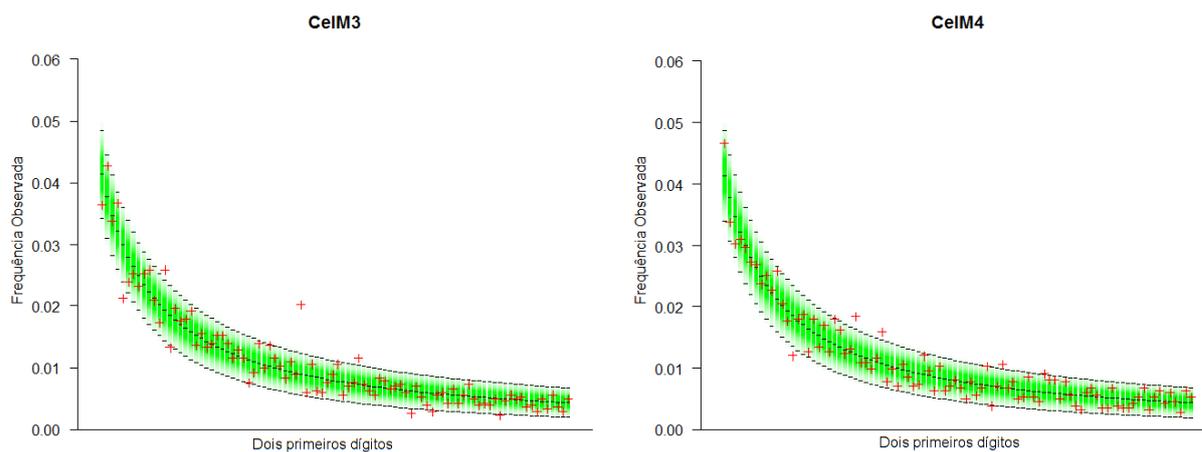
Fonte: O autor, 2021.

Figura 53 – Distribuição dos dois primeiros dígitos dos CeIM1 e CeIM2



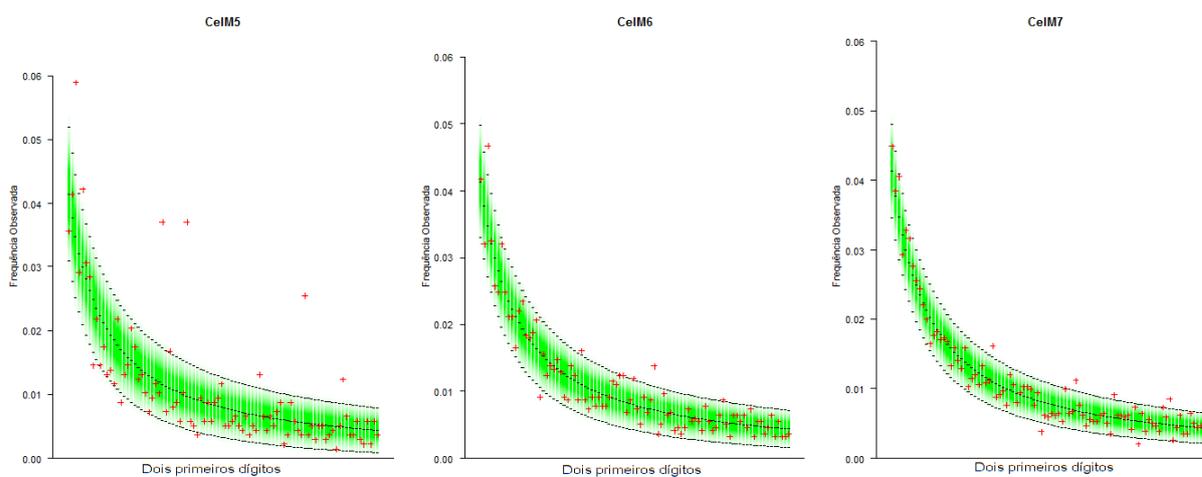
Fonte: O autor, 2021.

Figura 54 – Distribuição dos dois primeiros dígitos dos CeIM3 e CeIM4



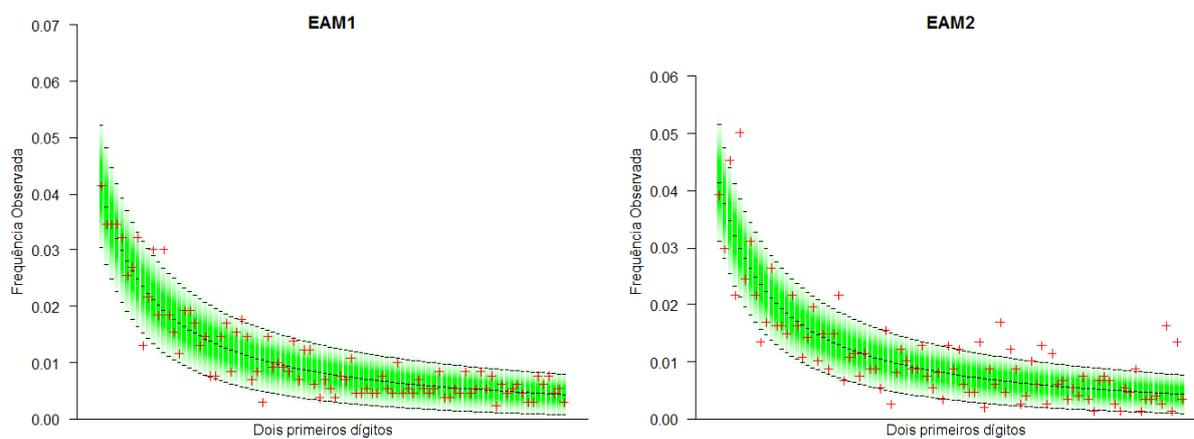
Fonte: O autor, 2021.

Figura 55 – Distribuição dos dois primeiros dígitos dos CeIM5, CeIM6 e CeIM7



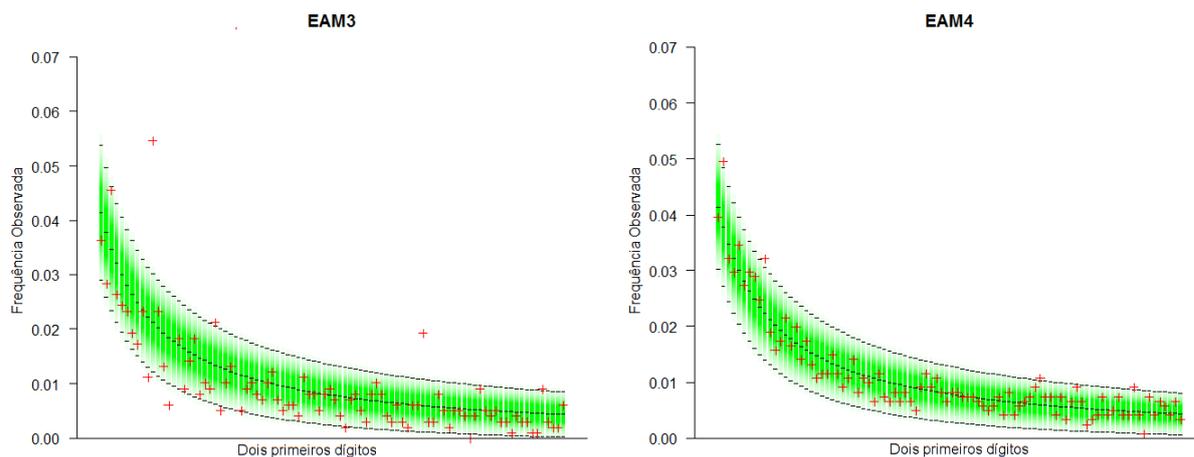
Fonte: O autor, 2021.

Figura 56 – Distribuição dos dois primeiros dígitos das EAM1 e EAM2



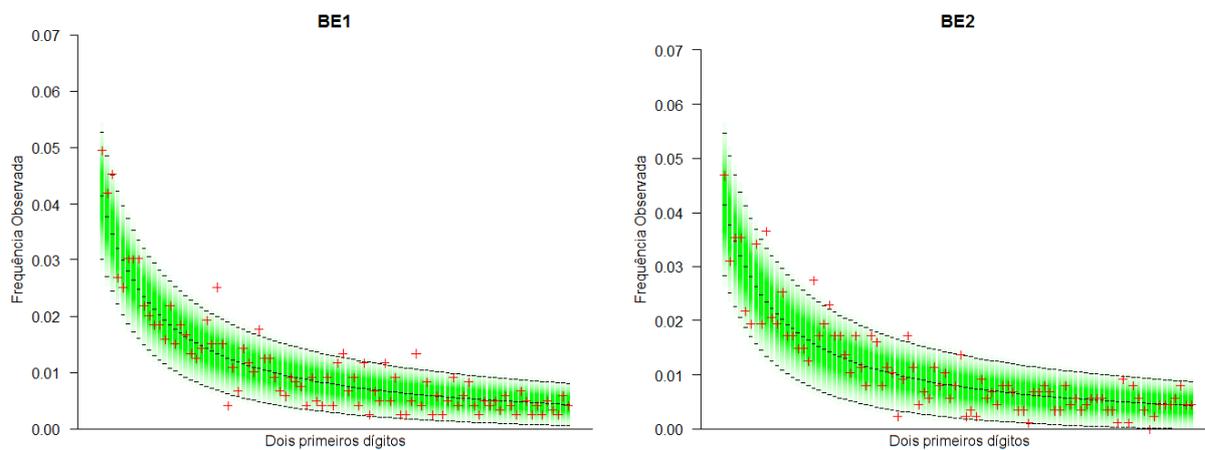
Fonte: O autor, 2021.

Figura 57 – Distribuição dos dois primeiros dígitos das EAM3 e EAM4



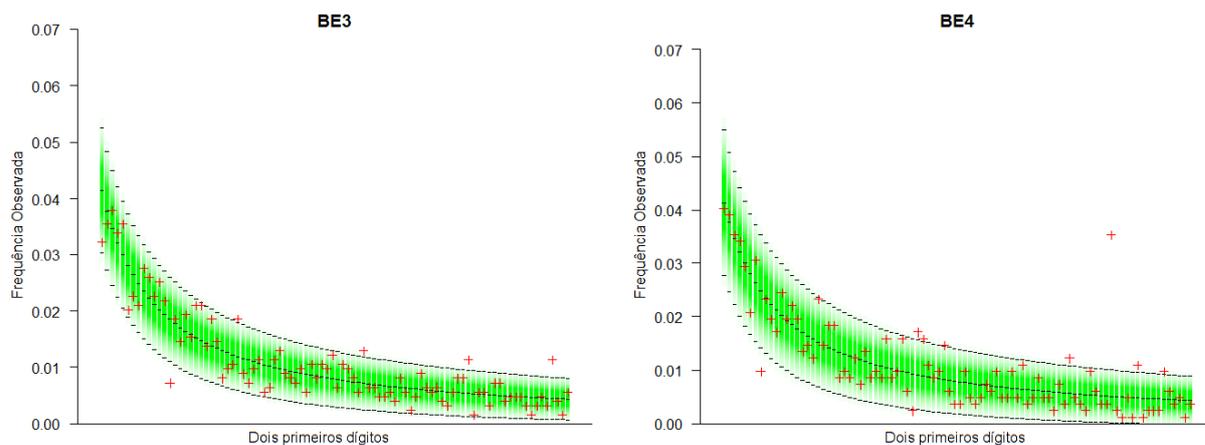
Fonte: O autor, 2021.

Figura 58 – Distribuição dos dois primeiros dígitos das BE1 e BE2



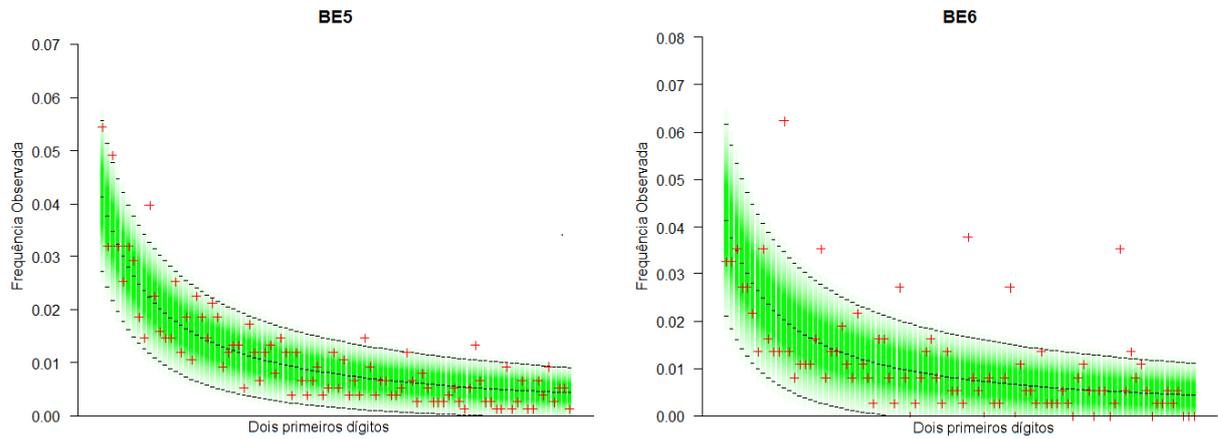
Fonte: O autor, 2021.

Figura 59 – Distribuição dos dois primeiros dígitos das BE3 e BE4



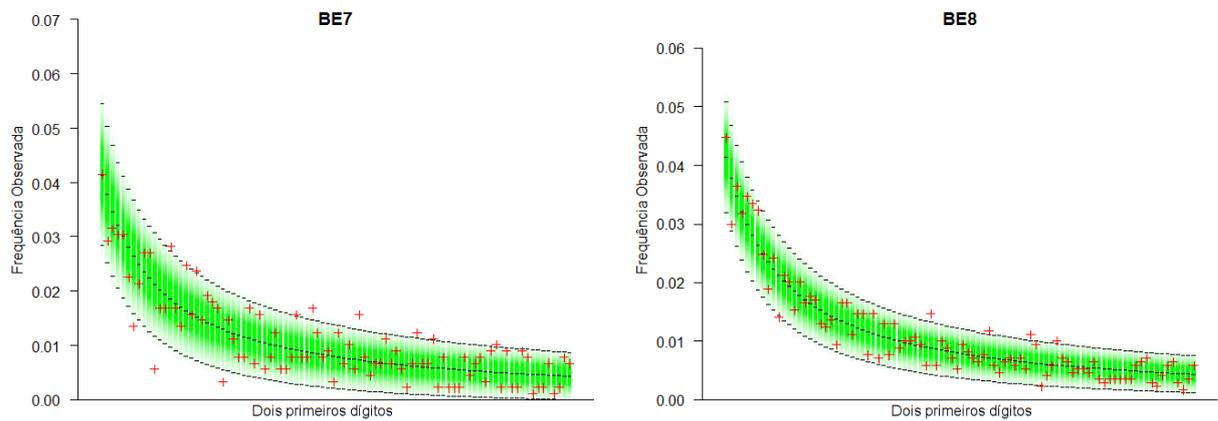
Fonte: O autor, 2021.

Figura 60 – Distribuição dos dois primeiros dígitos das BE5 e BE6



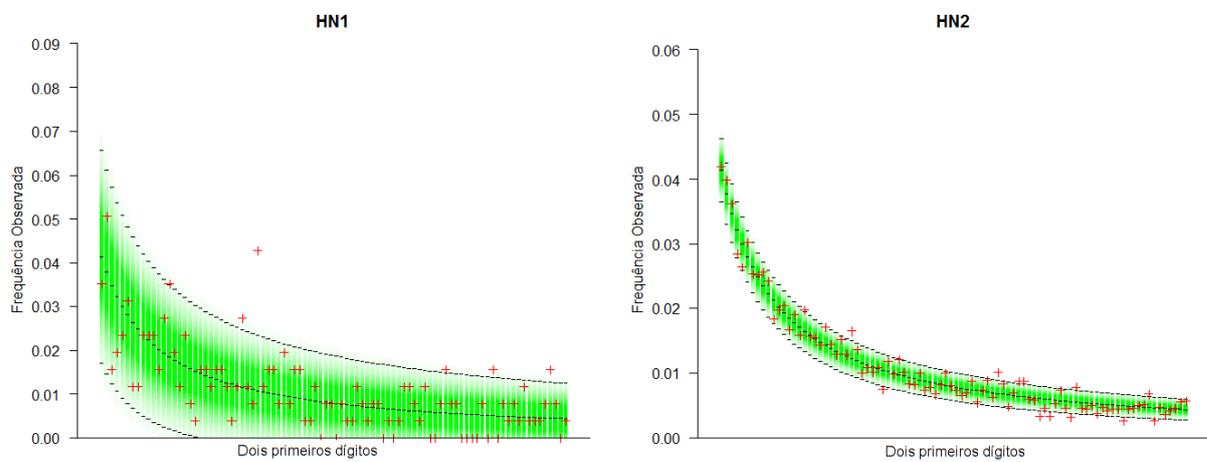
Fonte: O autor, 2021.

Figura 61 – Distribuição dos dois primeiros dígitos das BE7 e BE8



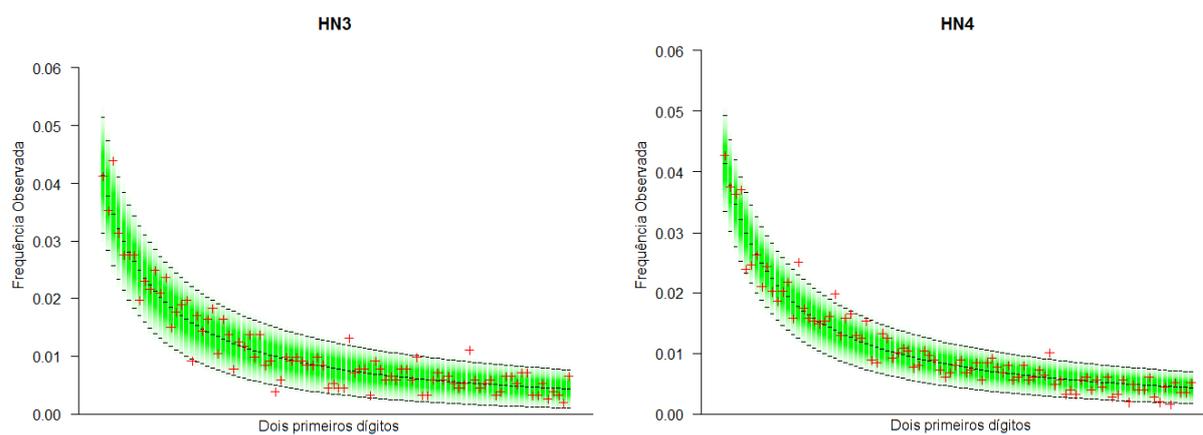
Fonte: O autor, 2021.

Figura 62 – Distribuição dos dois primeiros dígitos dos HN1 e HN2



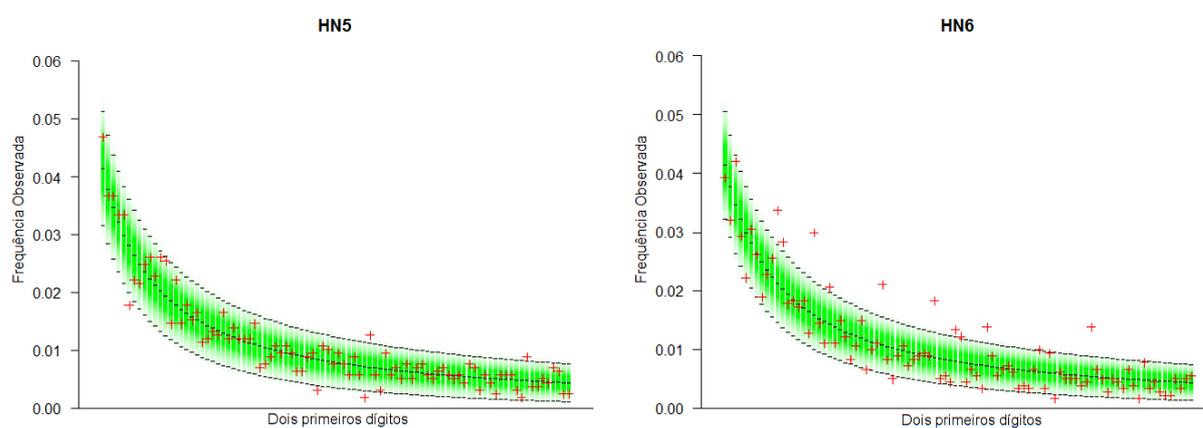
Fonte: O autor, 2021.

Figura 63 – Distribuição dos dois primeiros dígitos dos HN3 e HN4



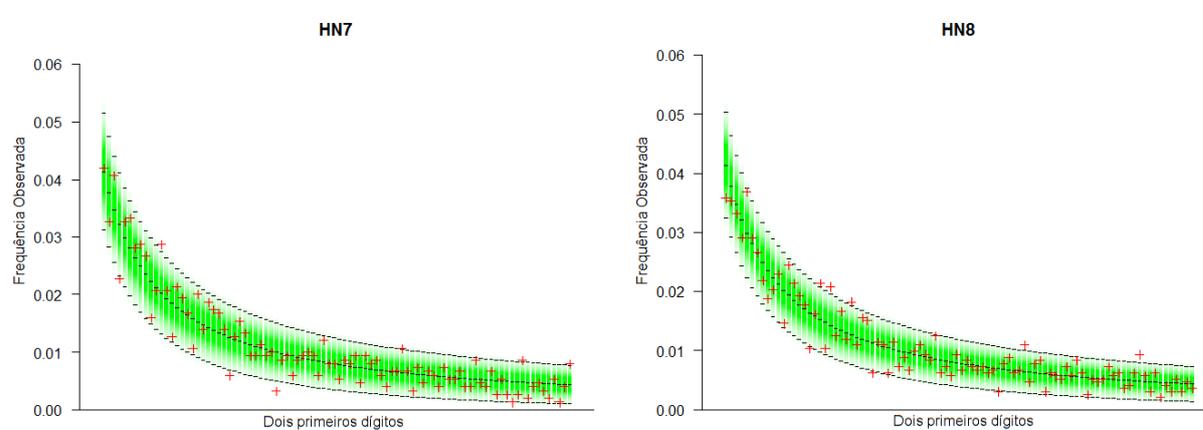
Fonte: O autor, 2021.

Figura 64 – Distribuição dos dois primeiros dígitos dos HN5 e HN6



Fonte: O autor, 2021.

Figura 65 – Distribuição dos dois primeiros dígitos dos HN7 e HN8



Fonte: O autor, 2021.

Tabela 27 – Dois primeiros dígitos em não conformidade com a LNB

OM	Dígitos	OM	Dígitos	OM	Dígitos
CeIM1	14, 24, 33, 35, 60 e 61	EAM3	19, 20, 23, 24, 32, 33, 36, 57, 72, 81 e 95	BE8	20, 49, 60, 68, 70 e 73
CeIM2	39, 60, 70 e 80	EAM4	11, 19, 72, 90 e 92	HN1	23, 37, 40, 76, 85 e 96
CeIM3	14, 22, 23, 48, 59, 69, 73 e 86	BE1	12, 32, 34, 40, 56, 64 e 70	HN2	26, 35, 41, 63, 67, 68, 71, 73, 77, 78, 87, 92 e 93
CeIM4	23, 35, 40, 61, 63 e 71	BE2	18, 27, 30, 43, 45, 55 e 91	HN3	27, 43, 57 e 80
CeIM5	12, 14, 17, 25, 37, 39, 44, 47, 65, 78 e 89	BE3	23, 36, 60, 80 e 96	HN4	24, 31, 72, 87 e 95
CeIM6	12, 27, 39, 51, 54, 60 e 80	BE4	17, 46, 47, 48, 52, 76, 84, 89 e 94	HN5	15, 51, 60, 61 e 91
CeIM7	39, 53, 63, 74, 81 e 88	BE5	12, 19, 60 e 81	HN6	20, 21, 27, 30, 40, 42, 50, 54, 55, 59, 60, 70, 73 e 80
EAM1	18, 20, 22, 37 e 41	BE6	21, 28, 43, 56, 64, 85 e 87	HN7	13, 21, 34, 43, 67, 88, 90 e 99
EAM2	12, 13, 14, 18, 33, 34, 42, 43, 53, 54, 56, 60, 61, 64, 66, 70, 72, 74, 82, 87, 90, 96 e 98	BE7	16, 20, 23, 26, 28, 33, 47, 50, 59, 70, 73 e 85	HN8	26, 28, 30, 34, 38, 62, 67 e 89

Fonte: O autor, 2021.

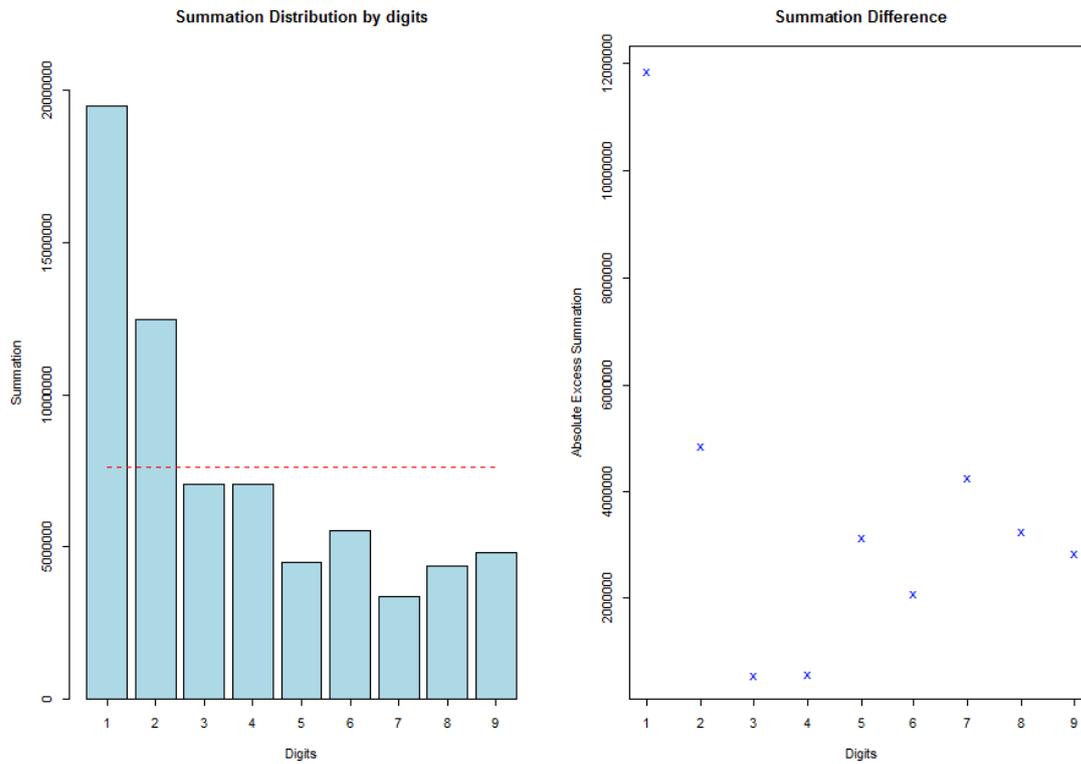
APÊNDICE D – Resultados das análises avançada

Tabela 28 – Análise de segunda ordem

	MAD	
	Primeiros dígitos	Dois primeiros dígitos
CeIM1	0,0059	0,0033
CeIM2	0,0056	0,0027
CeIM3	0,0079	0,0033
CeIM4	0,0048	0,0029
CeIM5	0,0072	0,0029
CeIM6	0,0068	0,0040
CeIM7	0,0064	0,0039
EAM1	0,0068	0,0024
EAM2	0,0062	0,0031
EAM3	0,0100	0,0035
EAM4	0,0043	0,0024
BE1	0,0087	0,0031
BE2	0,0087	0,0032
BE3	0,0092	0,0039
BE4	0,0145	0,0035
BE5	0,0078	0,0036
BE6	0,0117	0,0048
BE7	0,0106	0,0030
BE8	0,0053	0,0022
HN1	0,0205	0,0057
HN2	0,0037	0,0027
HN3	0,0045	0,0021
HN4	0,0055	0,0021
HN5	0,0090	0,0022
HN6	0,0041	0,0023
HN7	0,0057	0,0026
HN8	0,0031	0,0023

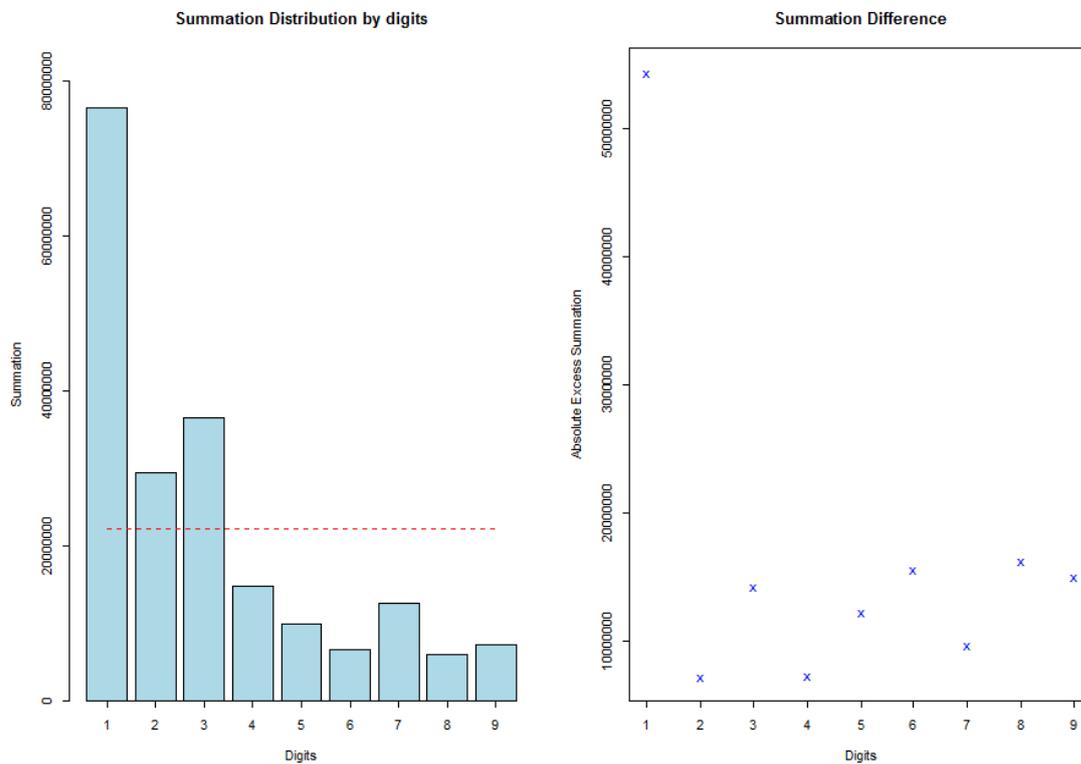
Fonte: O autor, 2021.

Figura 66 – Análise da soma para os primeiros dígitos do CeIM1



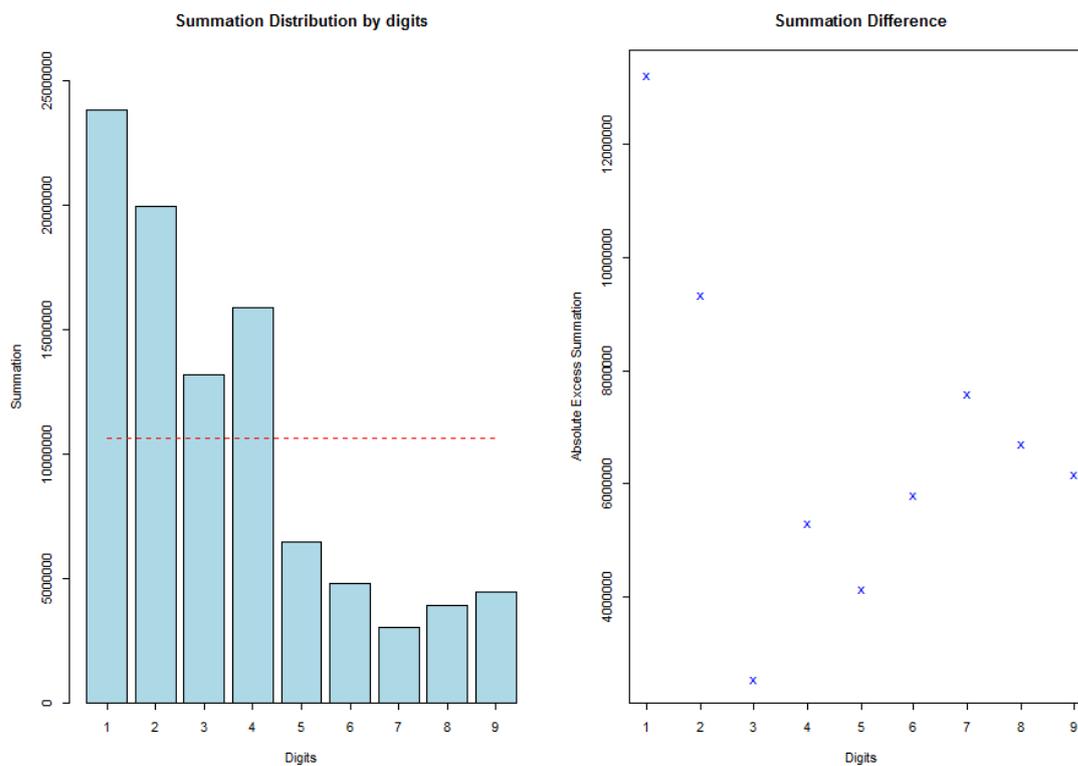
Fonte: O autor, 2021.

Figura 67 – Análise da soma para os primeiros dígitos do CeIM2



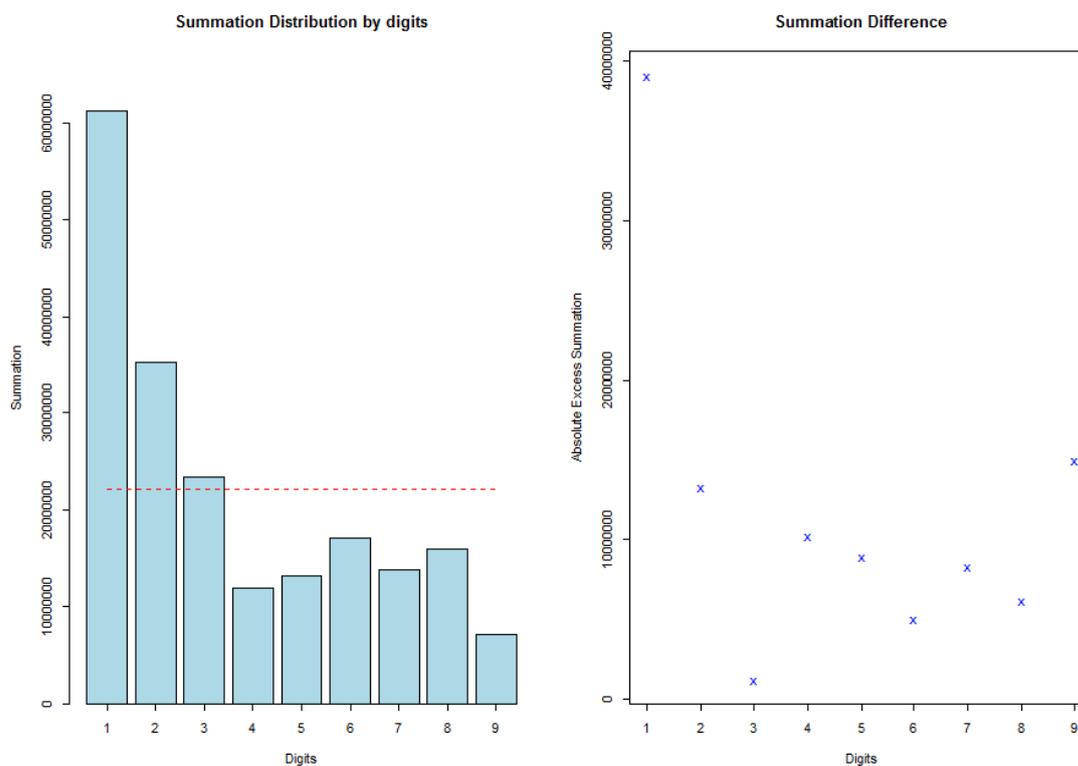
Fonte: O autor, 2021.

Figura 68 – Análise da soma para os primeiros dígitos do CeIM3



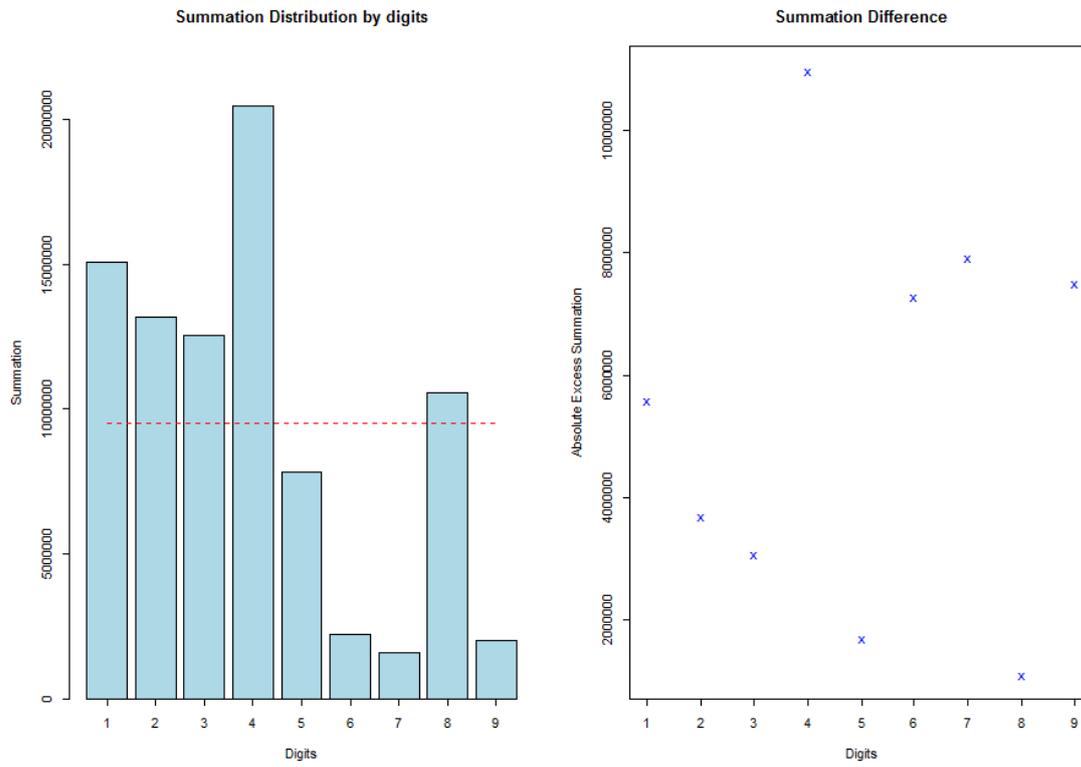
Fonte: O autor, 2021.

Figura 69 – Análise da soma para os primeiros dígitos do CeIM4



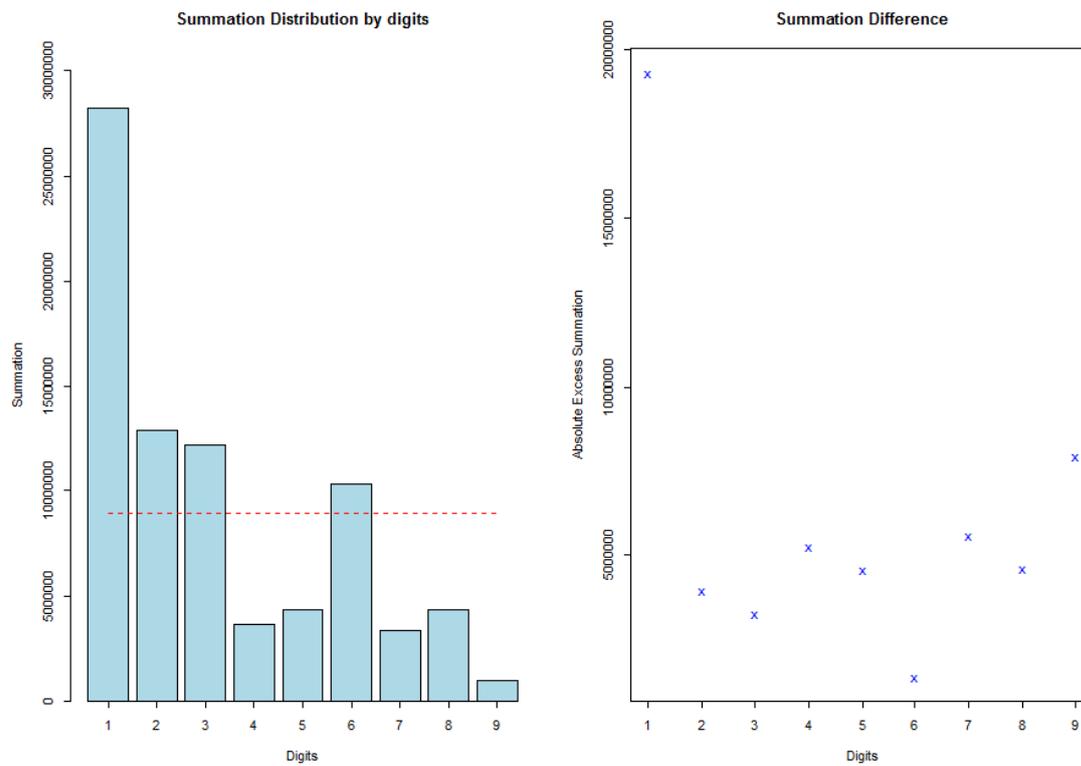
Fonte: O autor, 2021.

Figura 70 – Análise da soma para os primeiros dígitos do CeIM5



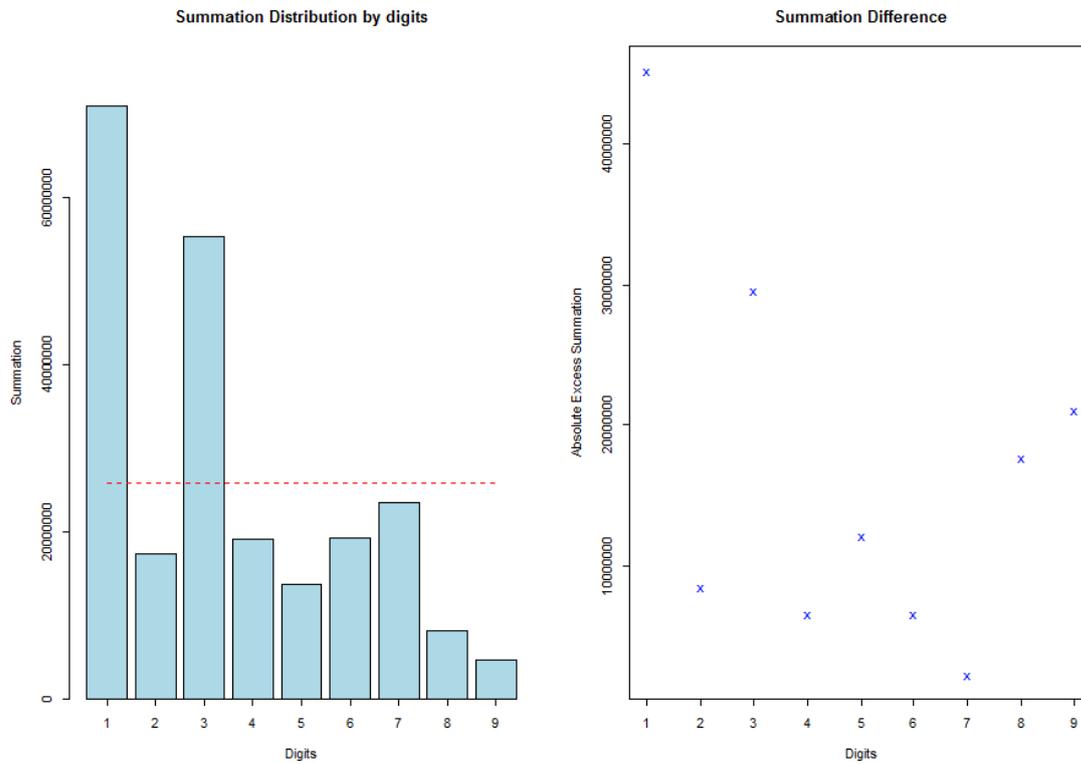
Fonte: O autor, 2021.

Figura 71 – Análise da soma para os primeiros dígitos do CeIM6



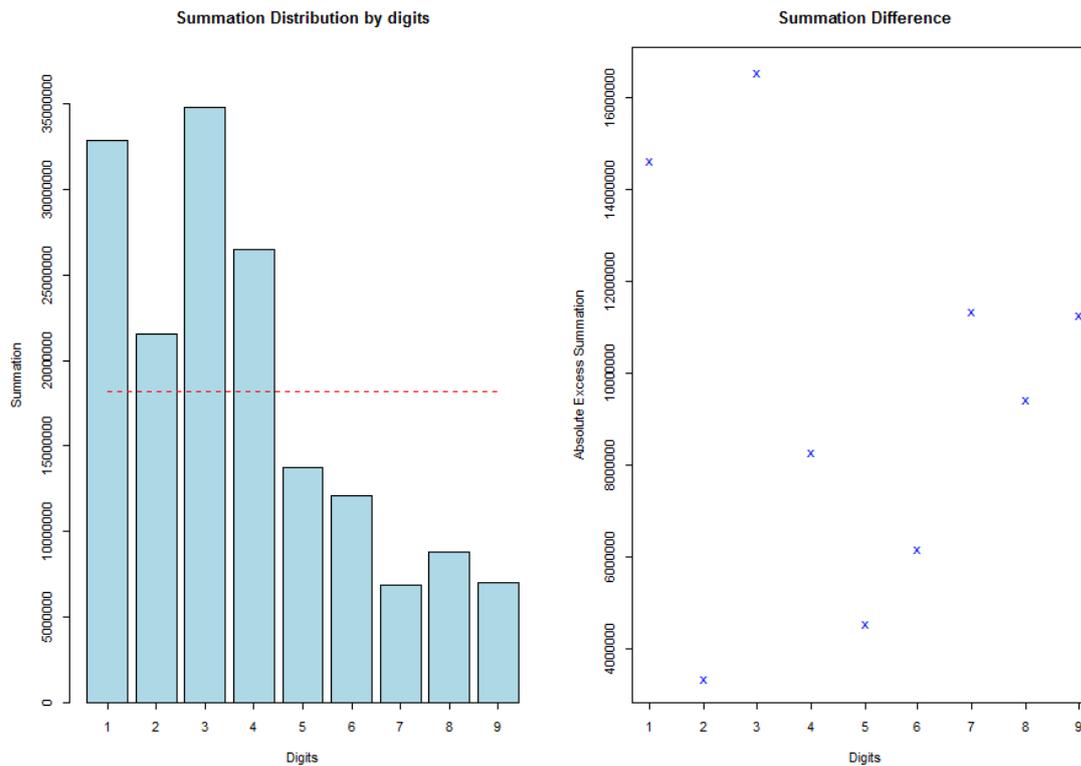
Fonte: O autor, 2021.

Figura 72 – Análise da soma para os primeiros dígitos do CeIM7



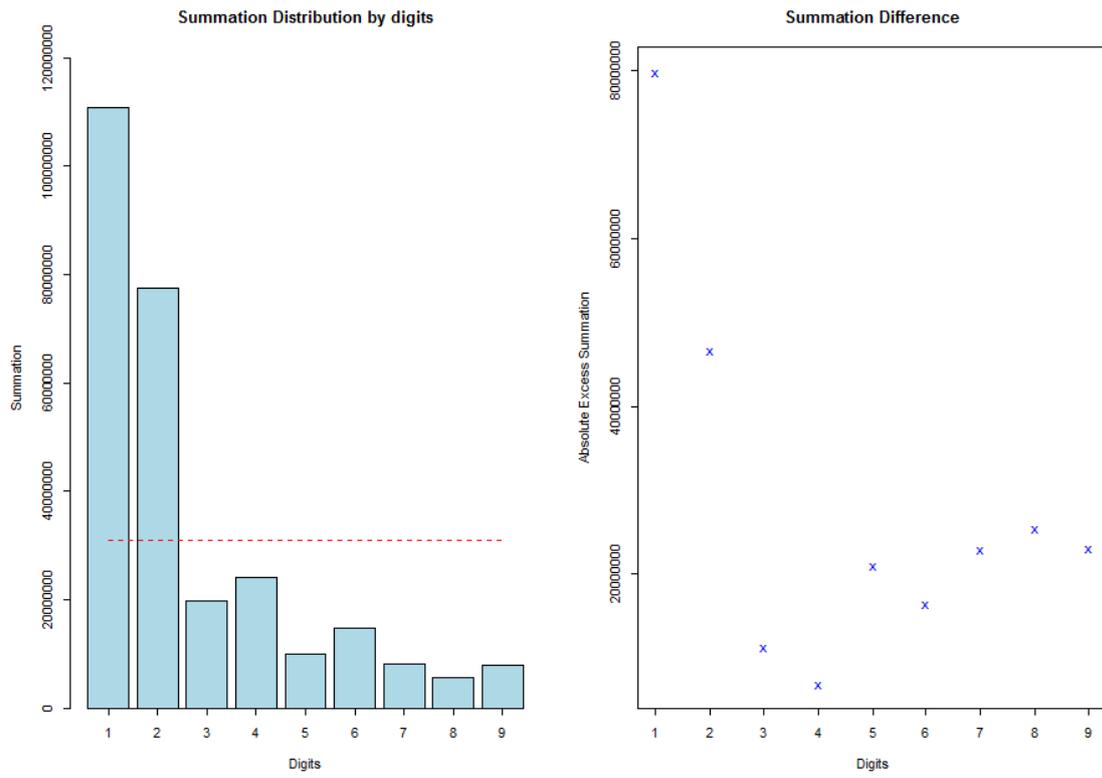
Fonte: O autor, 2021.

Figura 73 – Análise da soma para os primeiros dígitos da EAM1



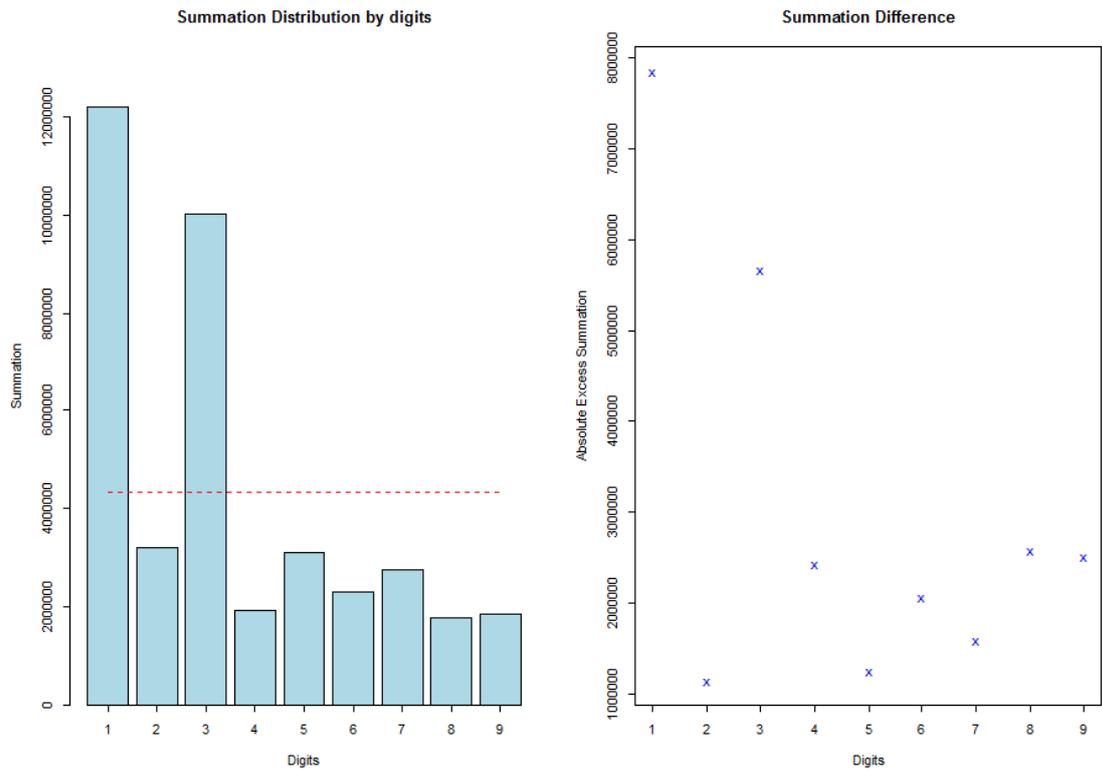
Fonte: O autor, 2021.

Figura 74 – Análise da soma para os primeiros dígitos da EAM2



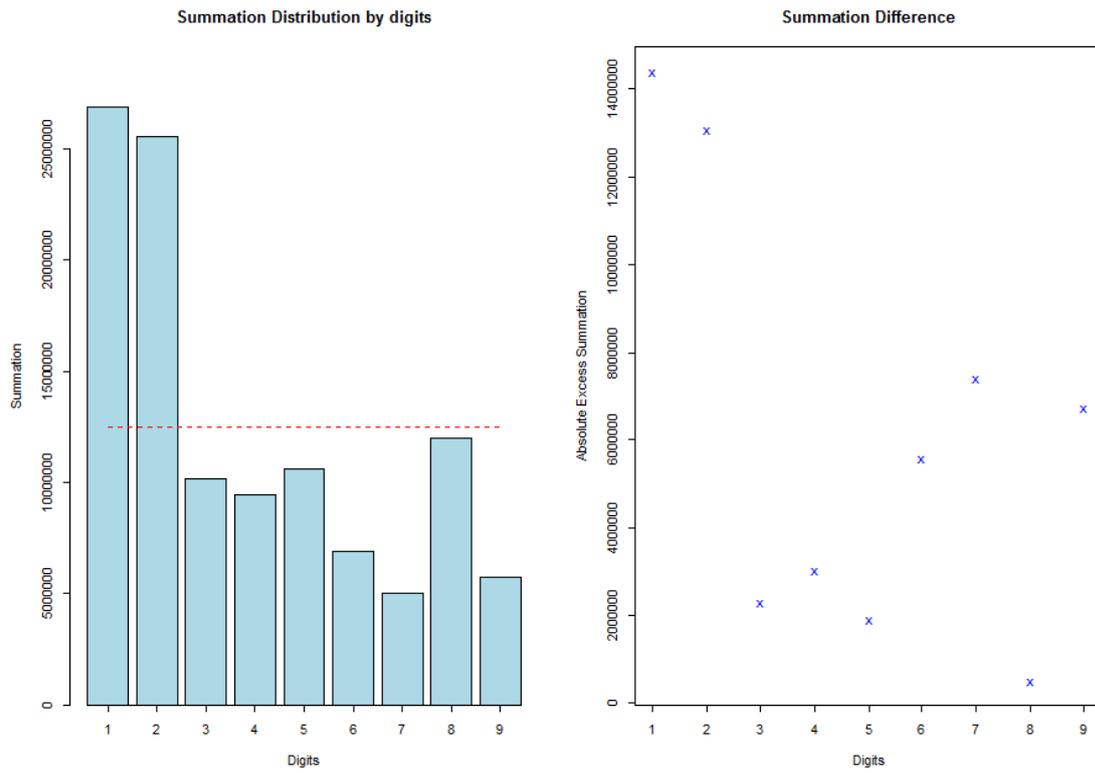
Fonte: O autor, 2021.

Figura 75 – Análise da soma para os primeiros dígitos da EAM3



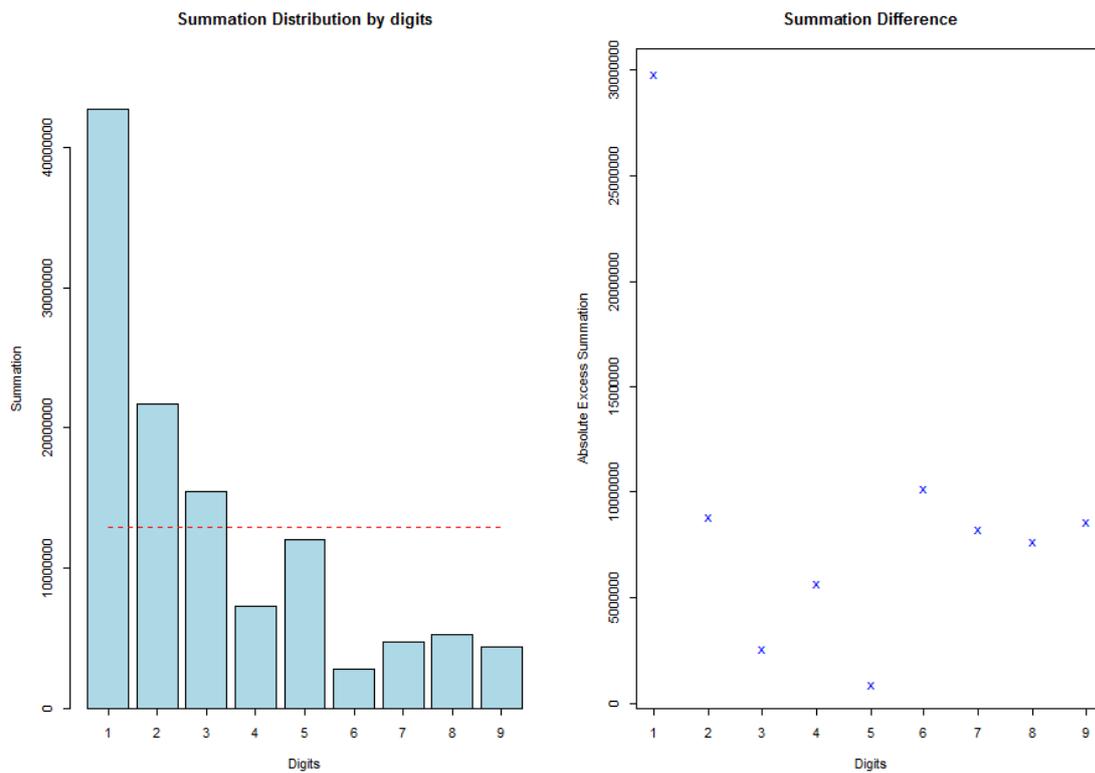
Fonte: O autor, 2021.

Figura 76 – Análise da soma para os primeiros dígitos da EAM4



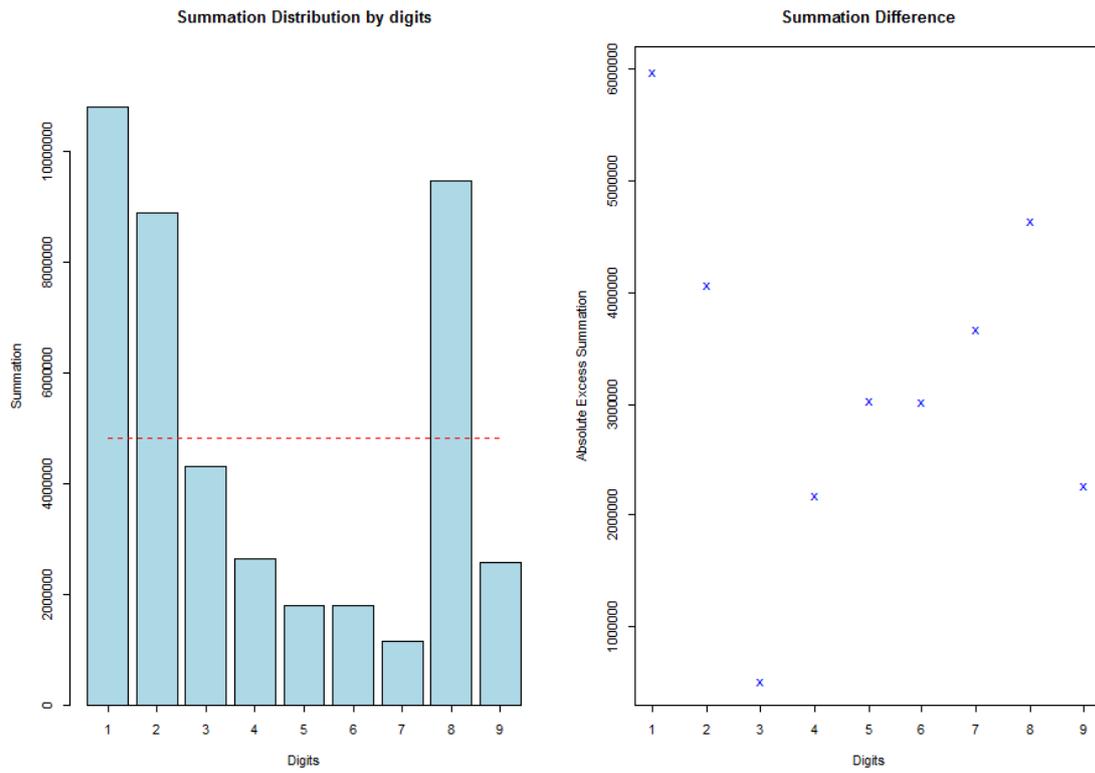
Fonte: O autor, 2021.

Figura 77 – Análise da soma para os primeiros dígitos da BE1



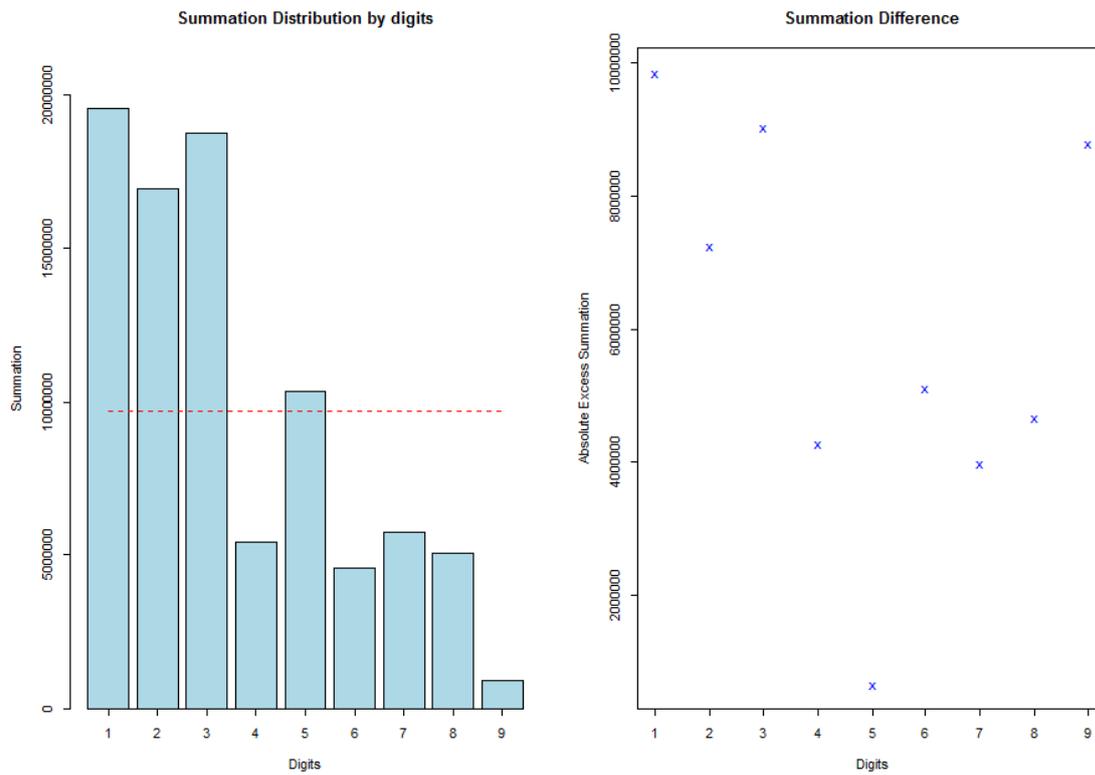
Fonte: O autor, 2021.

Figura 78 – Análise da soma para os primeiros dígitos da BE2



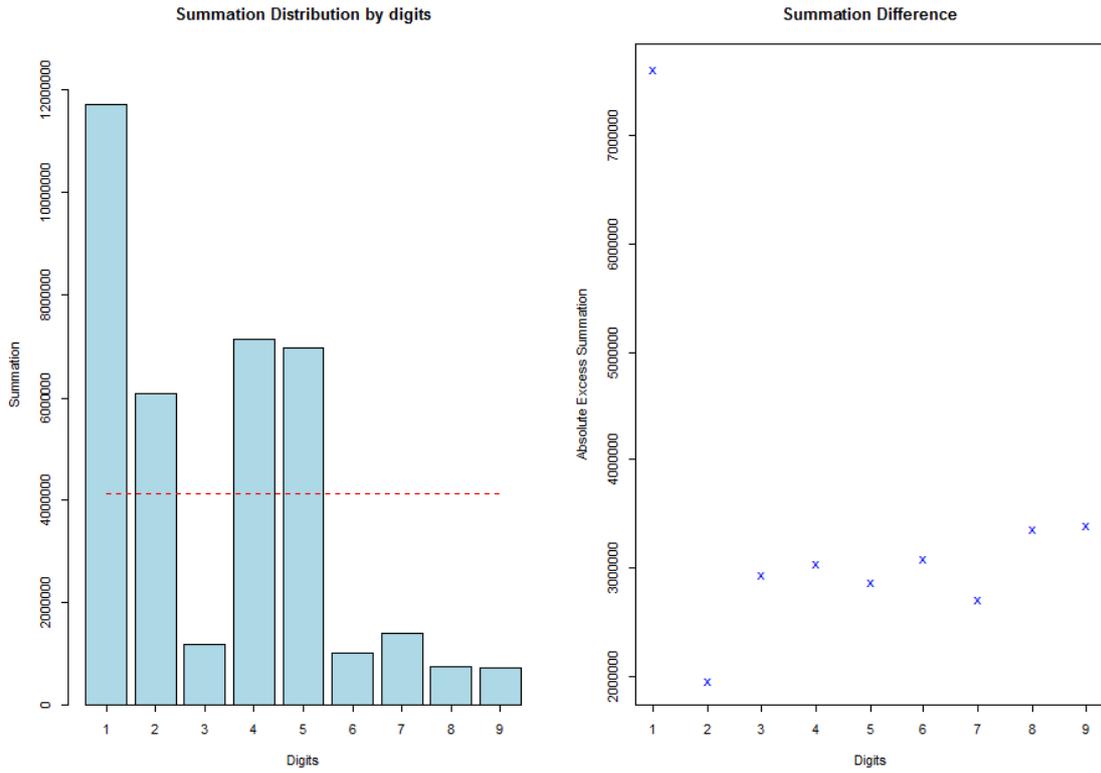
Fonte: O autor, 2021.

Figura 79 – Análise da soma para os primeiros dígitos da BE3



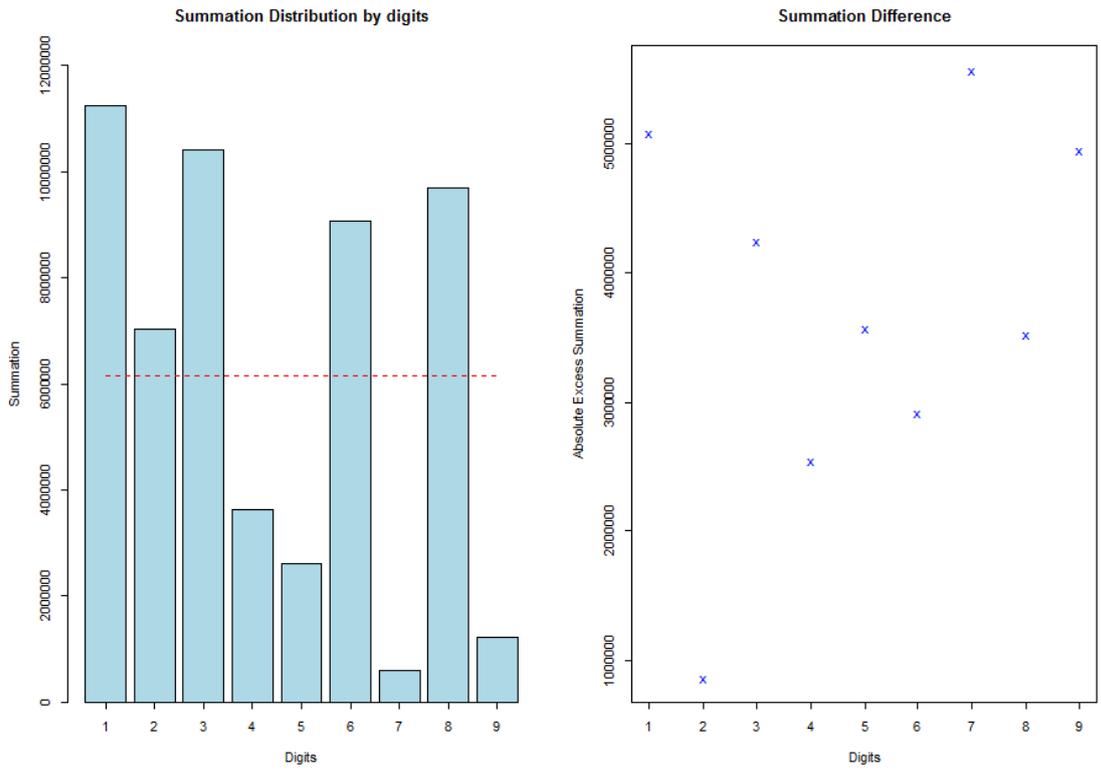
Fonte: O autor, 2021.

Figura 80 – Análise da soma para os primeiros dígitos da BE4



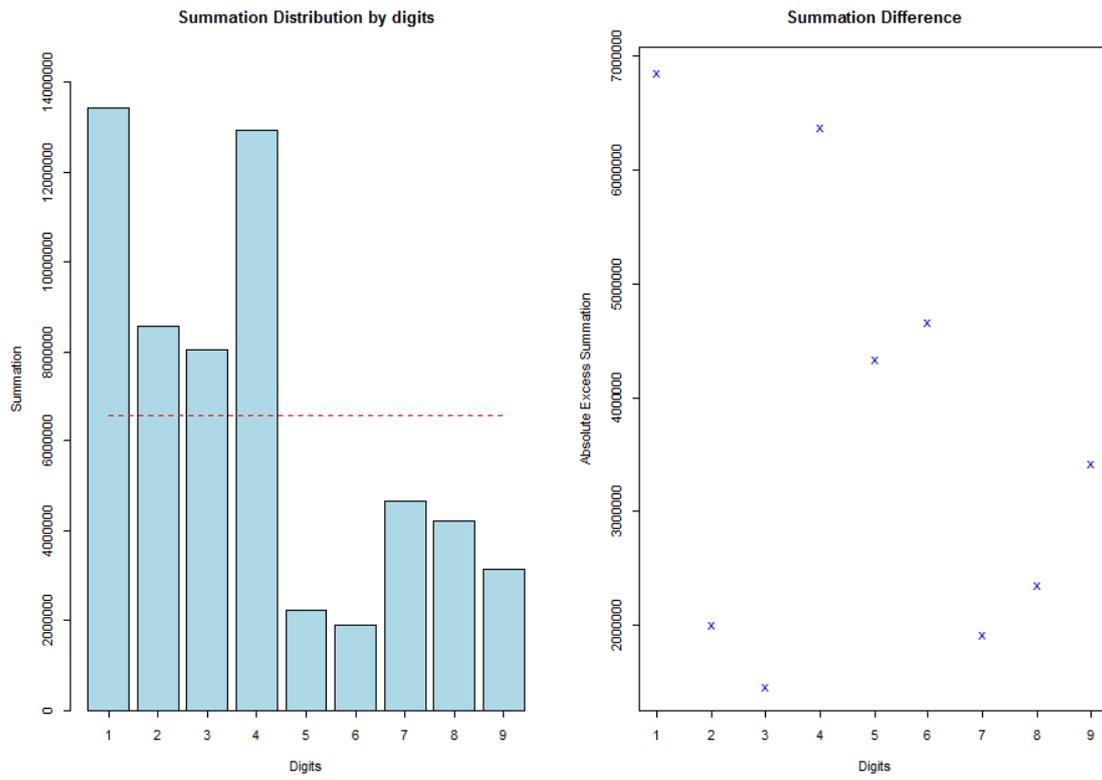
Fonte: O autor, 2021.

Figura 81 – Análise da soma para os primeiros dígitos da BE5



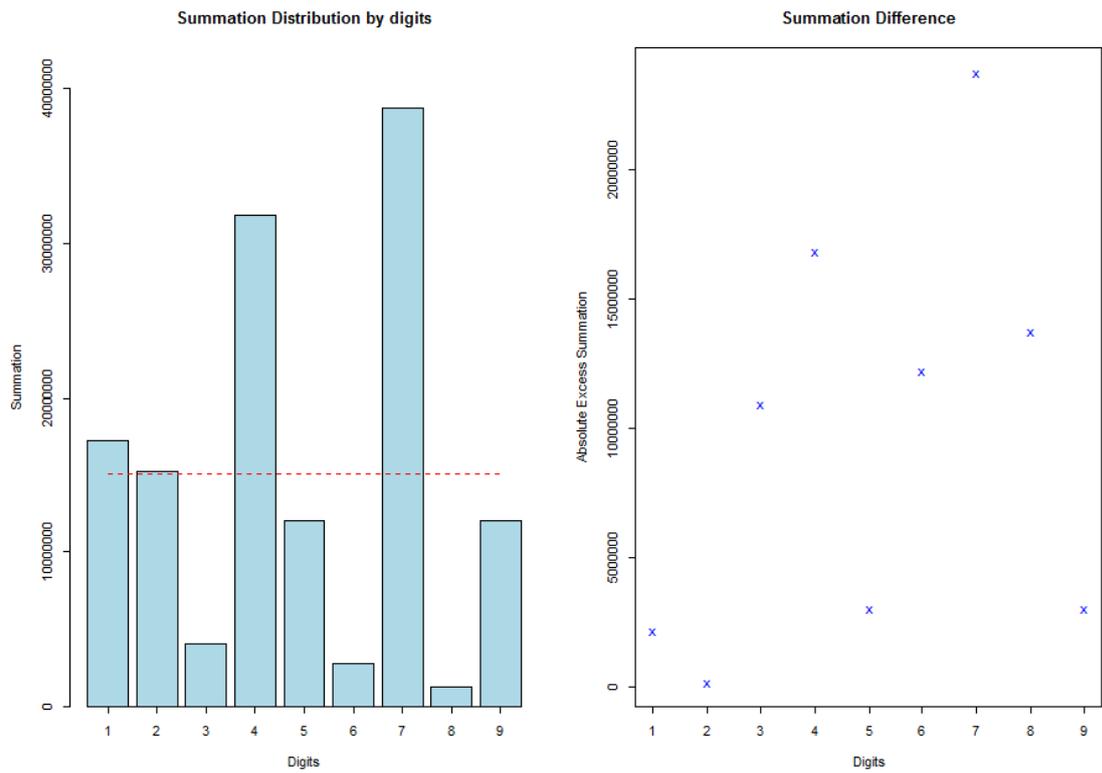
Fonte: O autor, 2021.

Figura 82 – Análise da soma para os primeiros dígitos da BE6



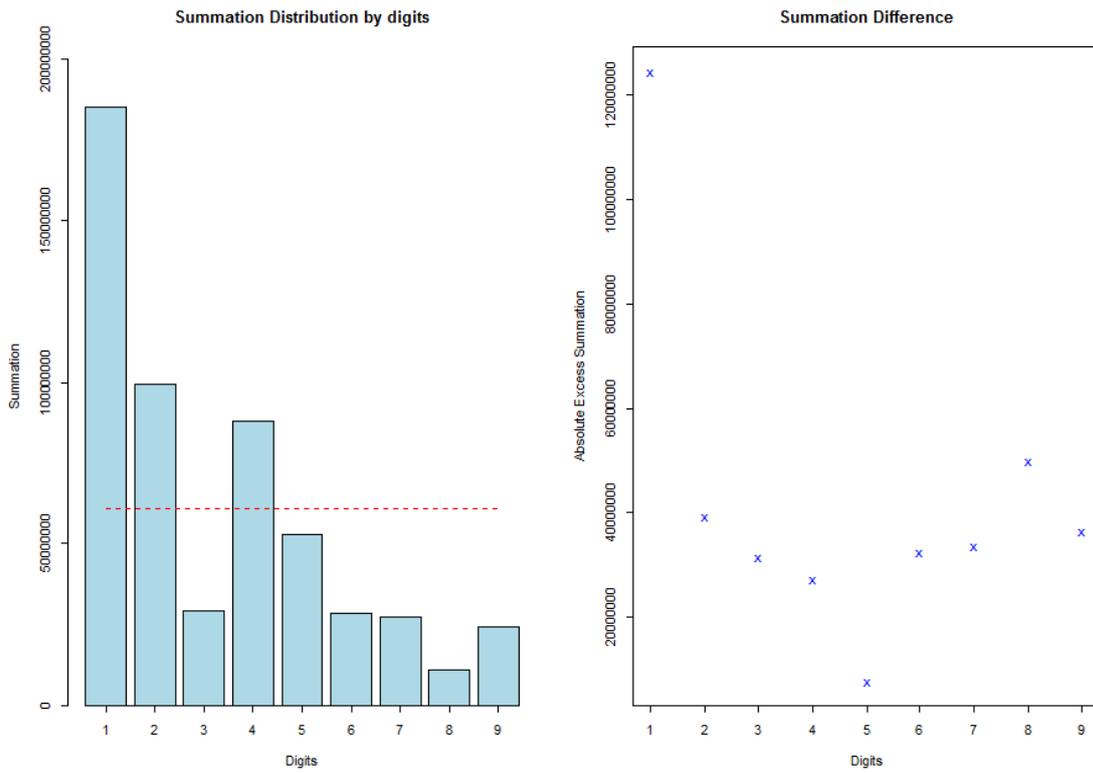
Fonte: O autor, 2021.

Figura 83 – Análise da soma para os primeiros dígitos da BE7



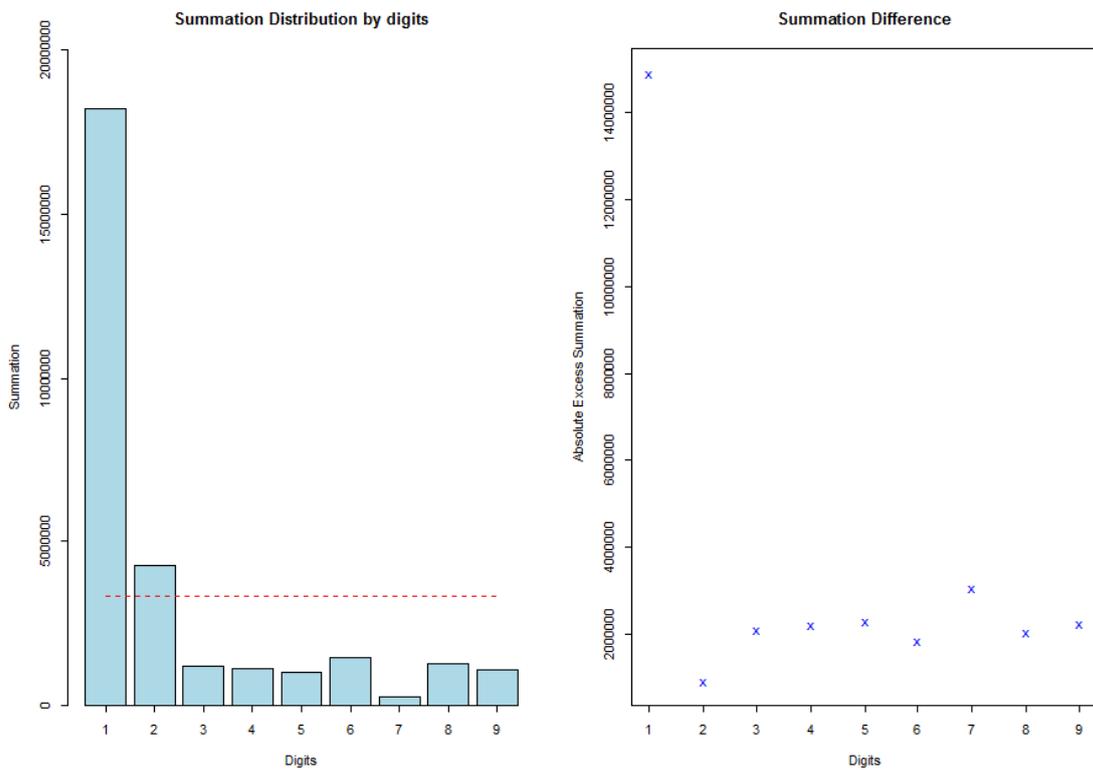
Fonte: O autor, 2021.

Figura 84 – Análise da soma para os primeiros dígitos da BE8



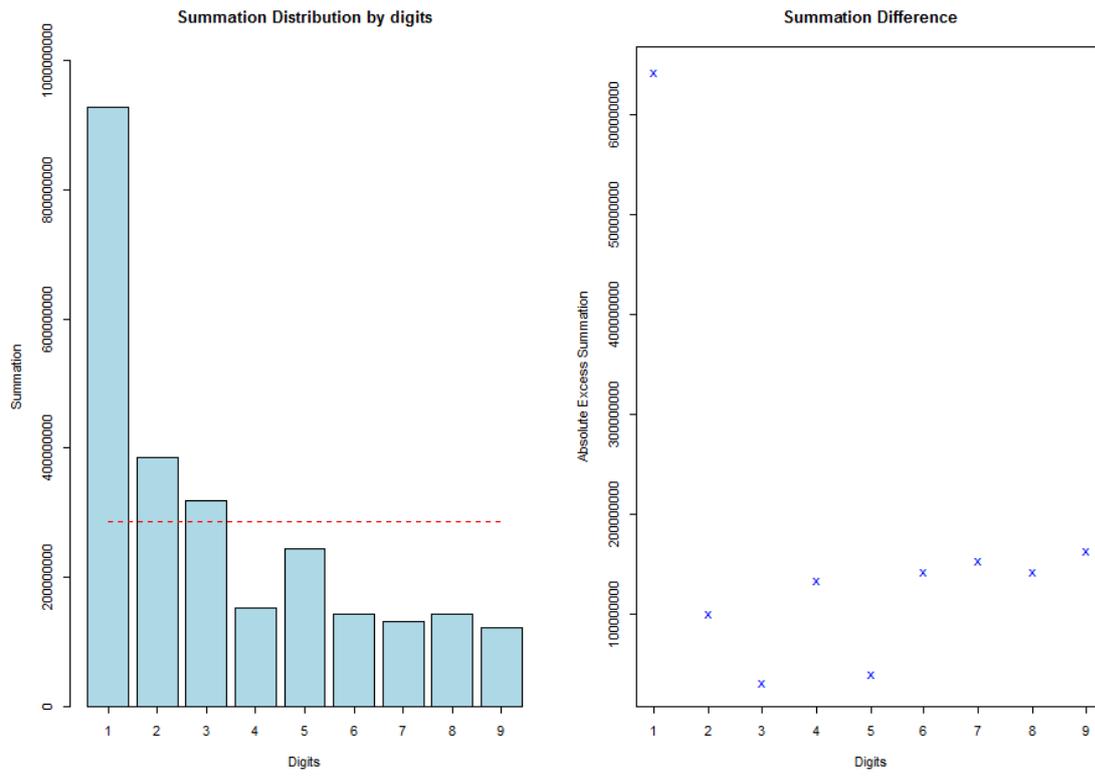
Fonte: O autor, 2021.

Figura 85 – Análise da soma para os primeiros dígitos do HN1



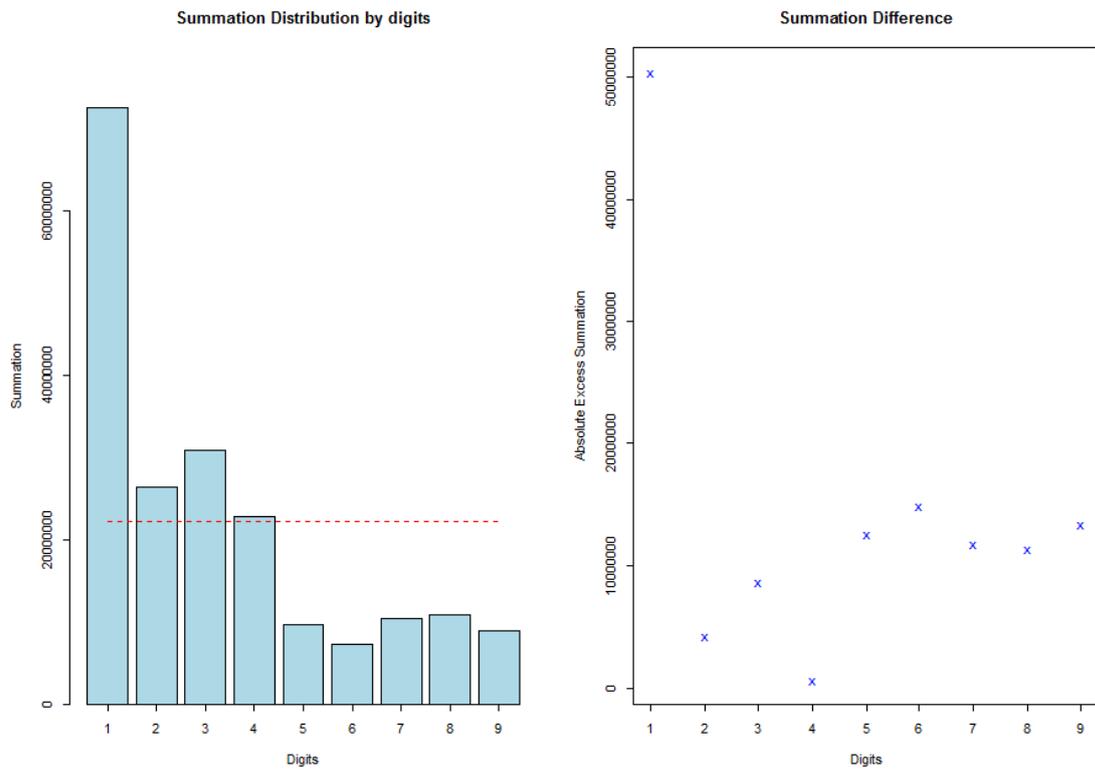
Fonte: O autor, 2021.

Figura 86 – Análise da soma para os primeiros dígitos do HN2



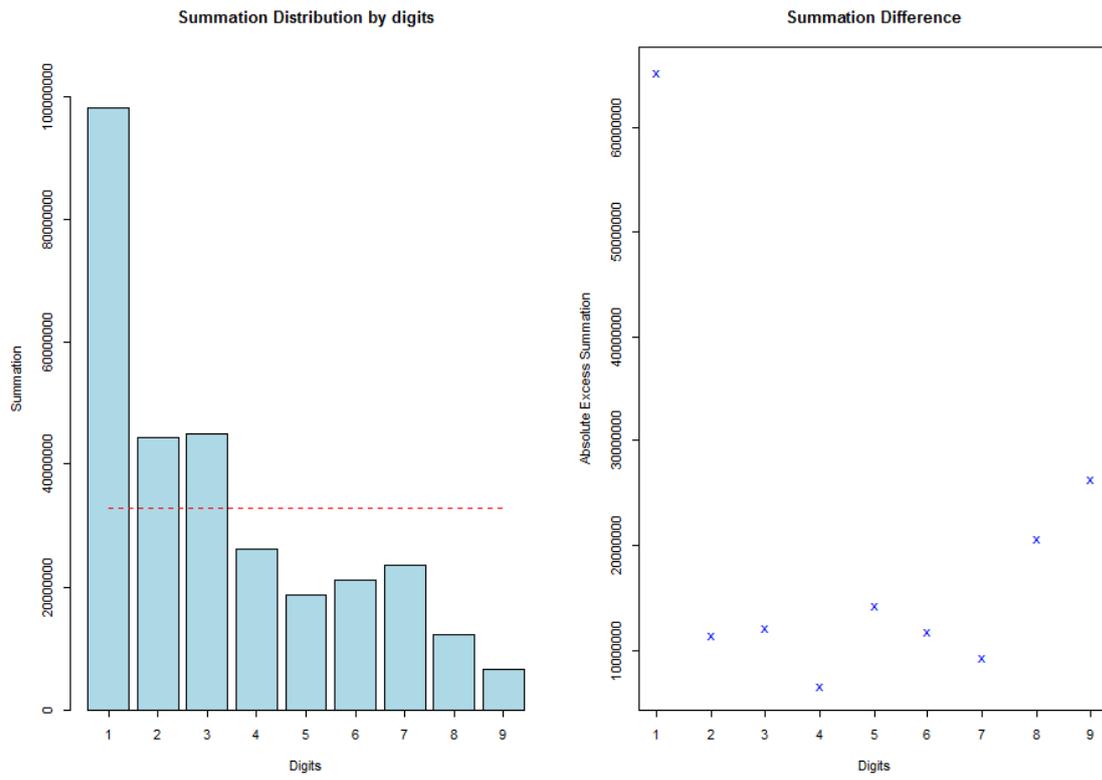
Fonte: O autor, 2021.

Figura 87 – Análise da soma para os primeiros dígitos do HN3



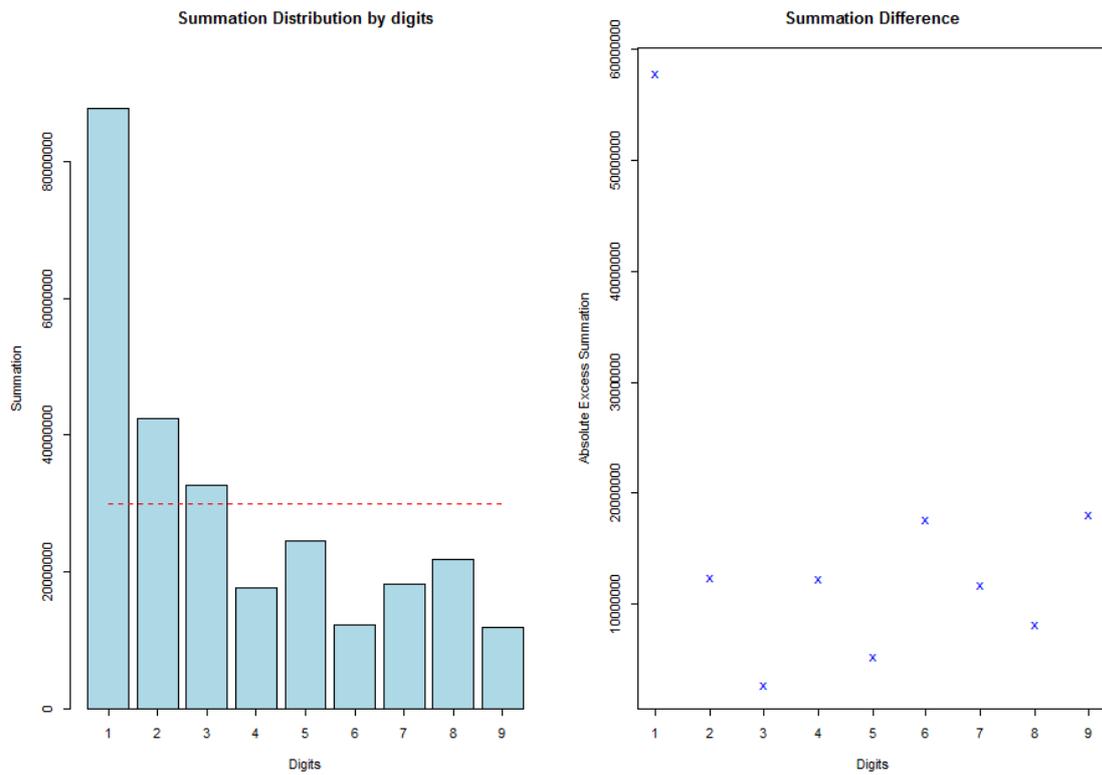
Fonte: O autor, 2021.

Figura 88 – Análise da soma para os primeiros dígitos do HN4



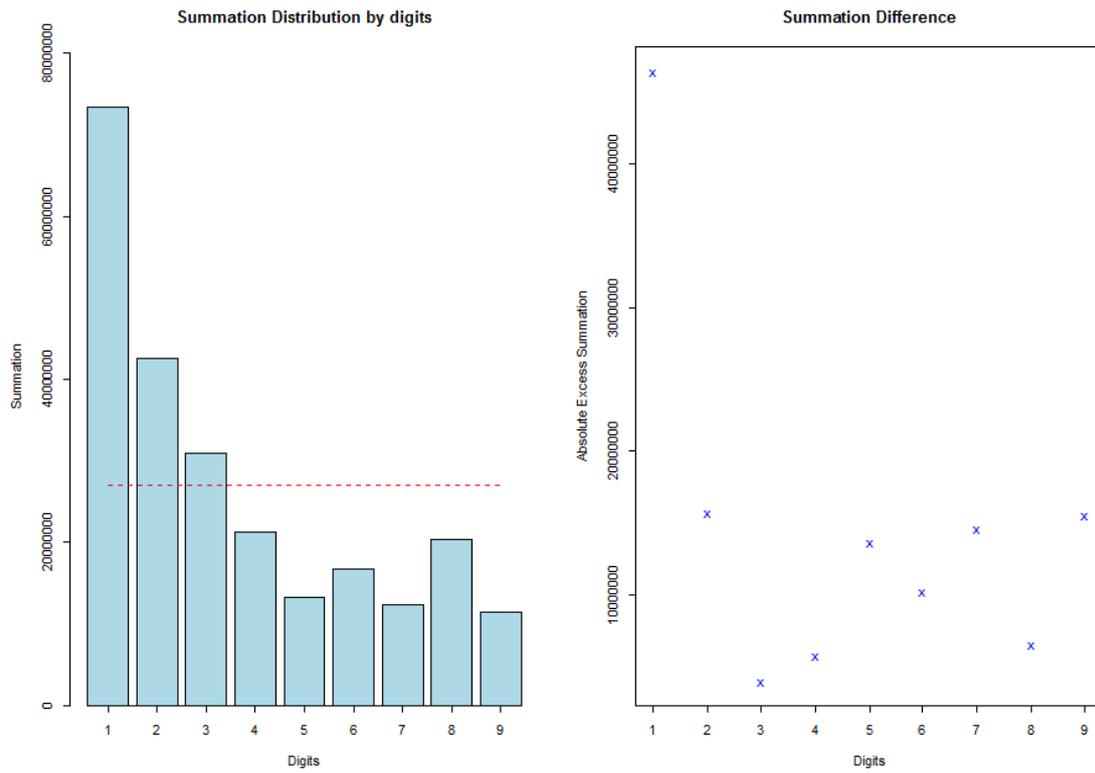
Fonte: O autor, 2021.

Figura 89 – Análise da soma para os primeiros dígitos do HN5



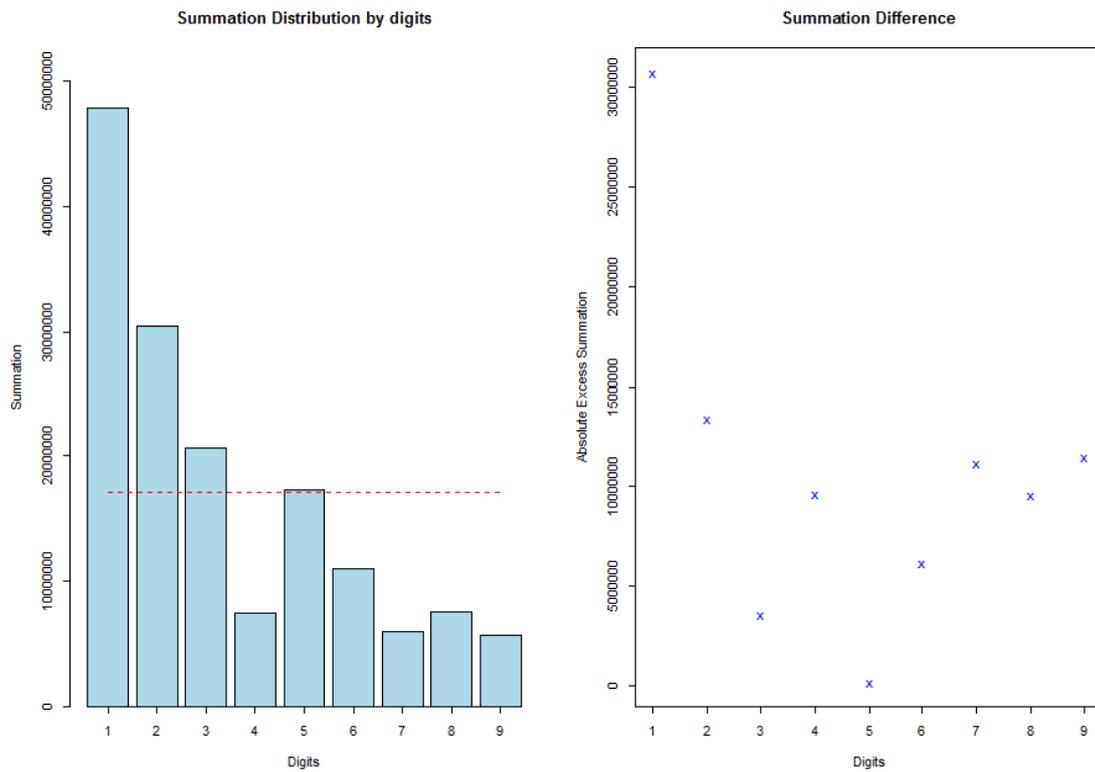
Fonte: O autor, 2021.

Figura 90 – Análise da soma para os primeiros dígitos do HN6



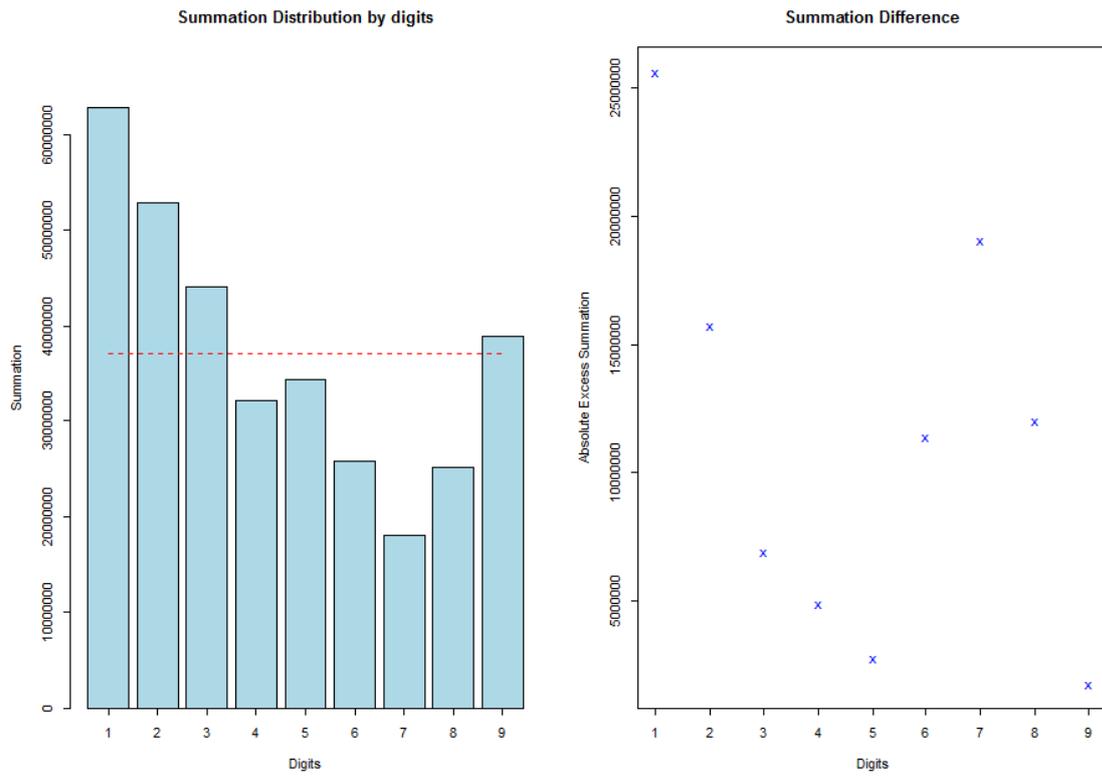
Fonte: O autor, 2021.

Figura 91 – Análise da soma para os primeiros dígitos do HN7



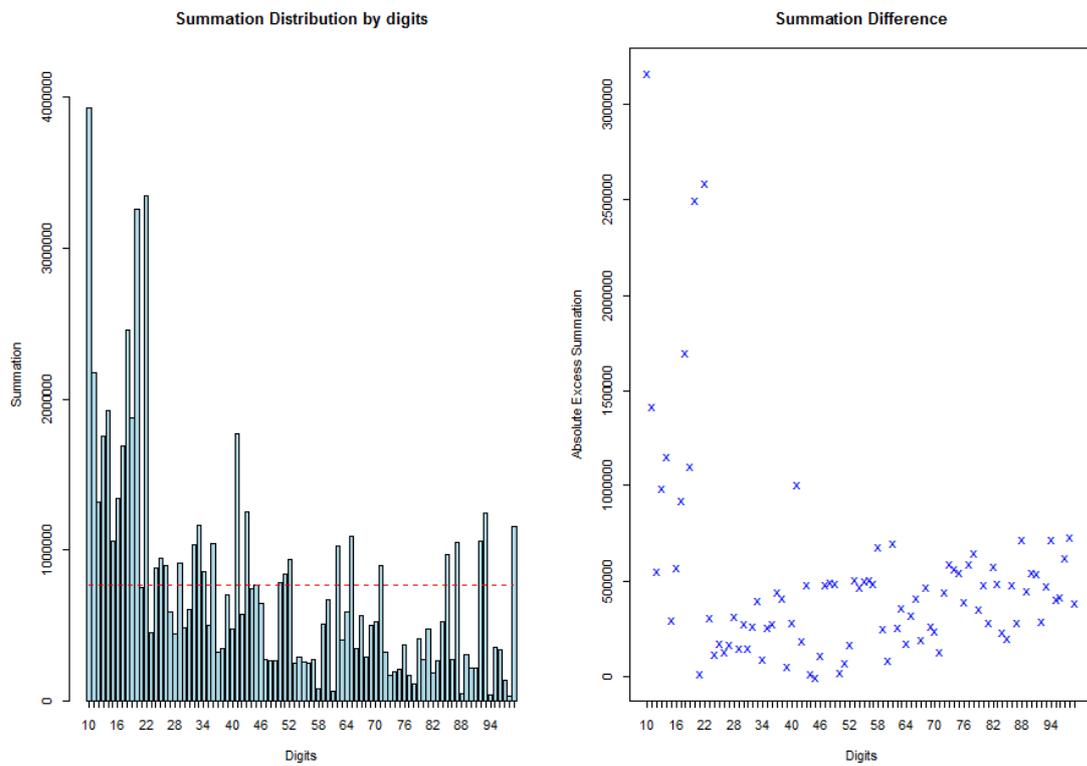
Fonte: O autor, 2021.

Figura 92 – Análise da soma para os primeiros dígitos do HN8



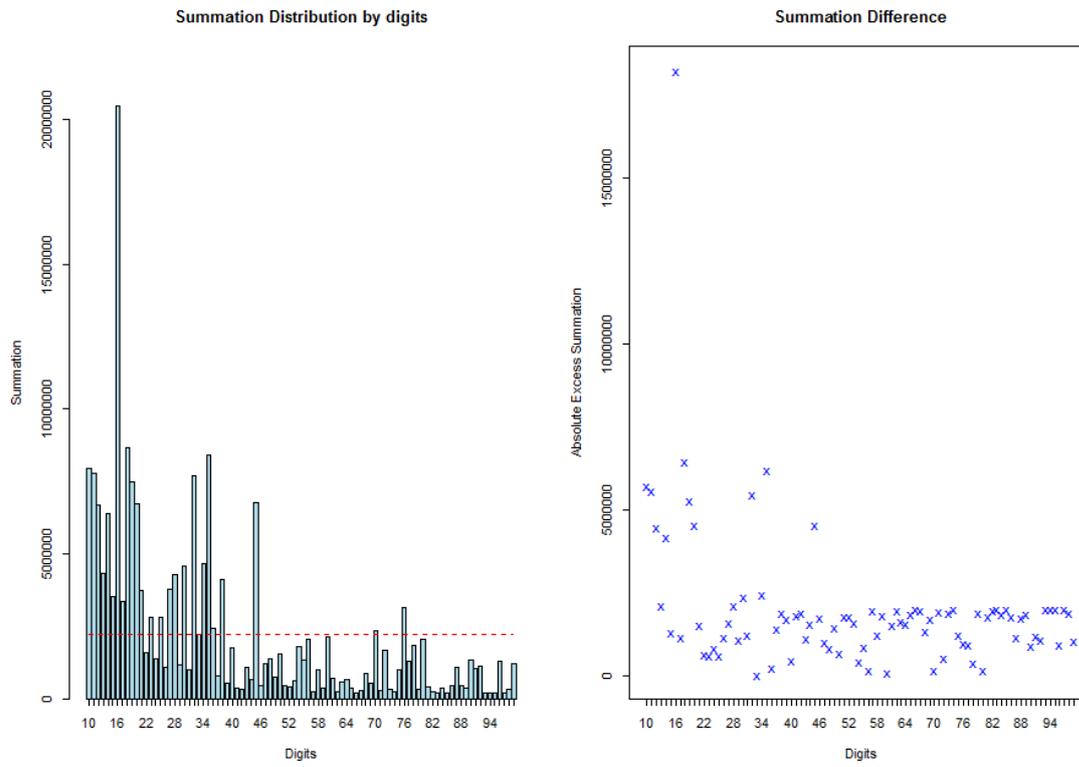
Fonte: O autor, 2021.

Figura 93 – Análise da soma para os dois primeiros dígitos do CeIM1



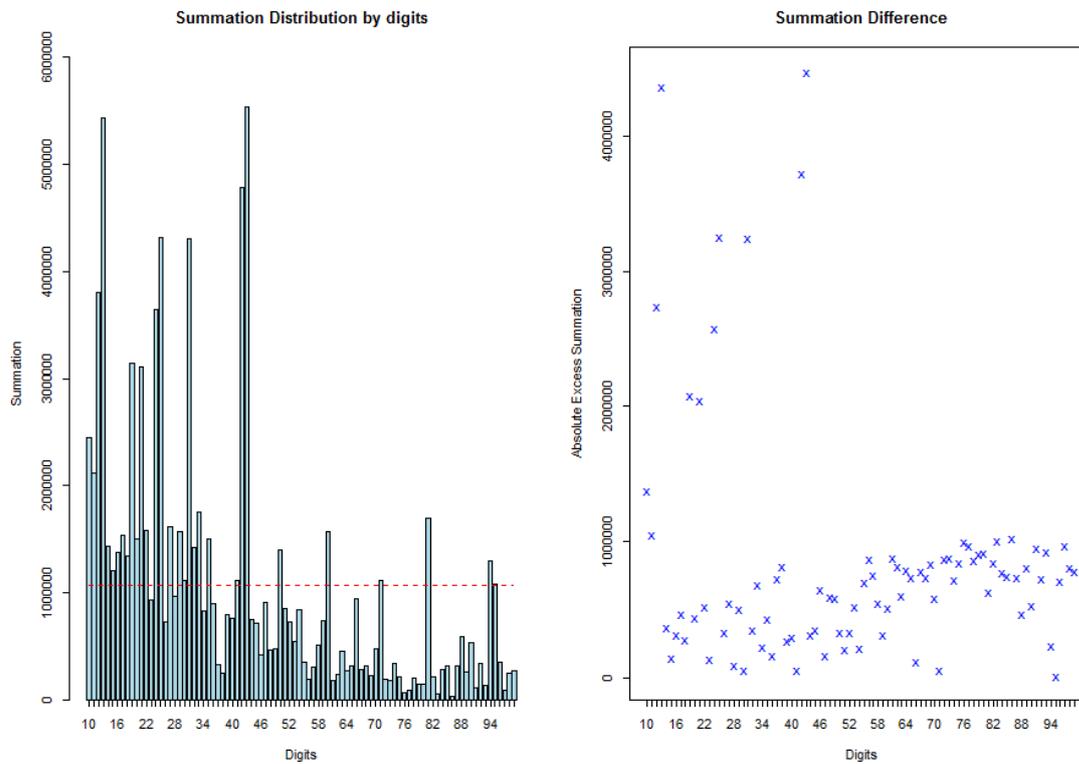
Fonte: O autor, 2021.

Figura 94 – Análise da soma para os dois primeiros dígitos do CeIM2



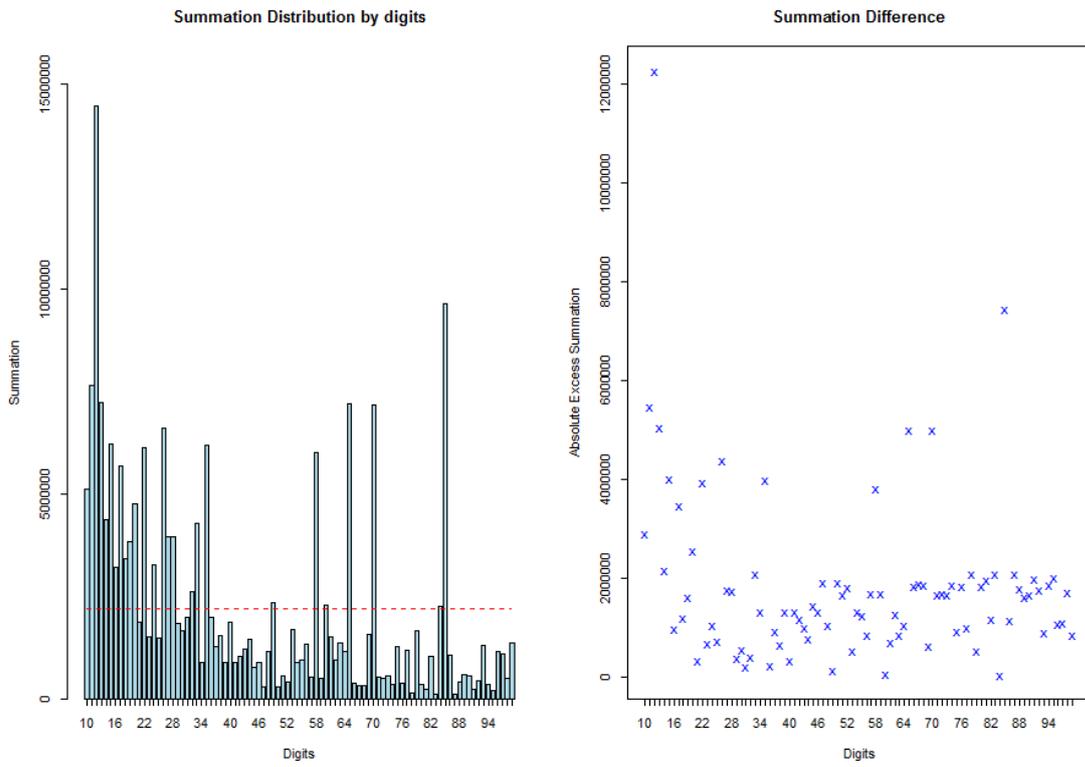
Fonte: O autor, 2021.

Figura 95 – Análise da soma para os dois primeiros dígitos do CeIM3



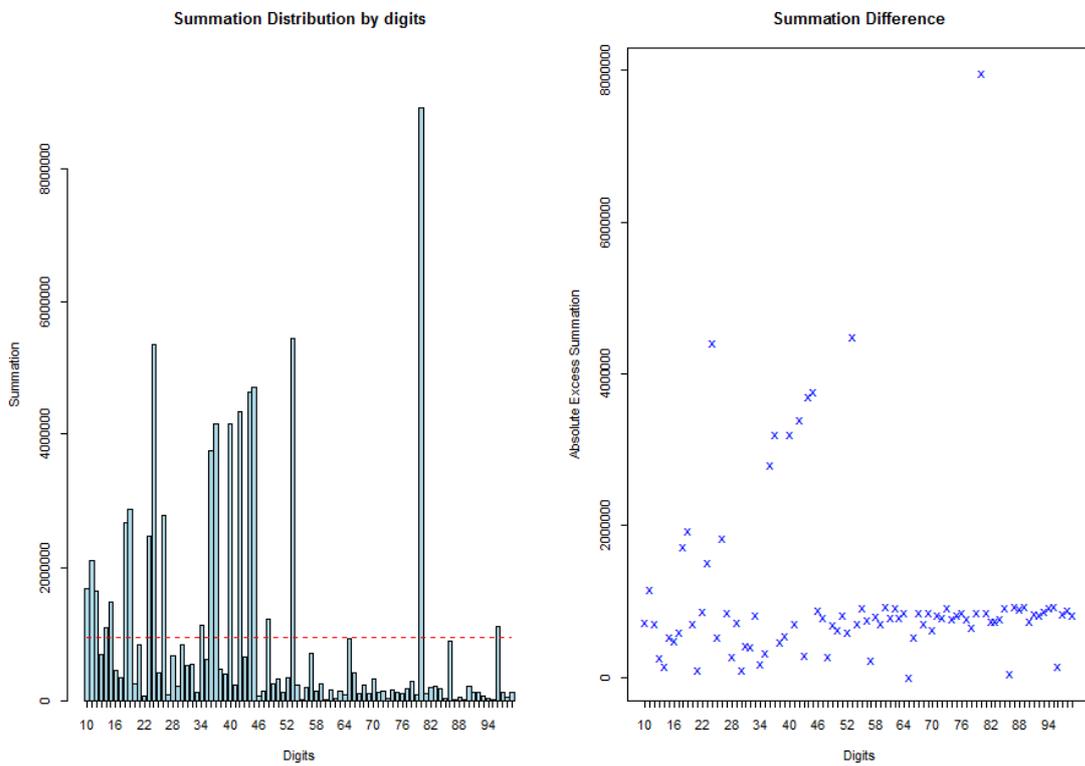
Fonte: O autor, 2021.

Figura 96 – Análise da soma para os dois primeiros dígitos do CeIM4



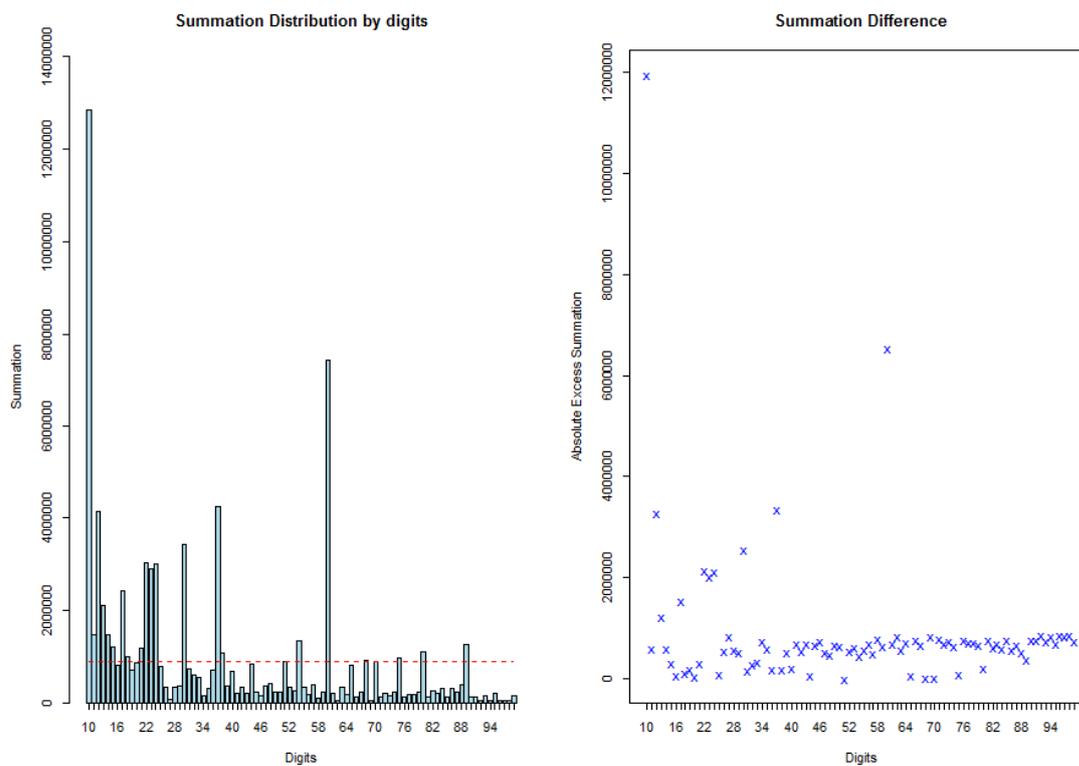
Fonte: O autor, 2021.

Figura 97 – Análise da soma para os dois primeiros dígitos do CeIM5



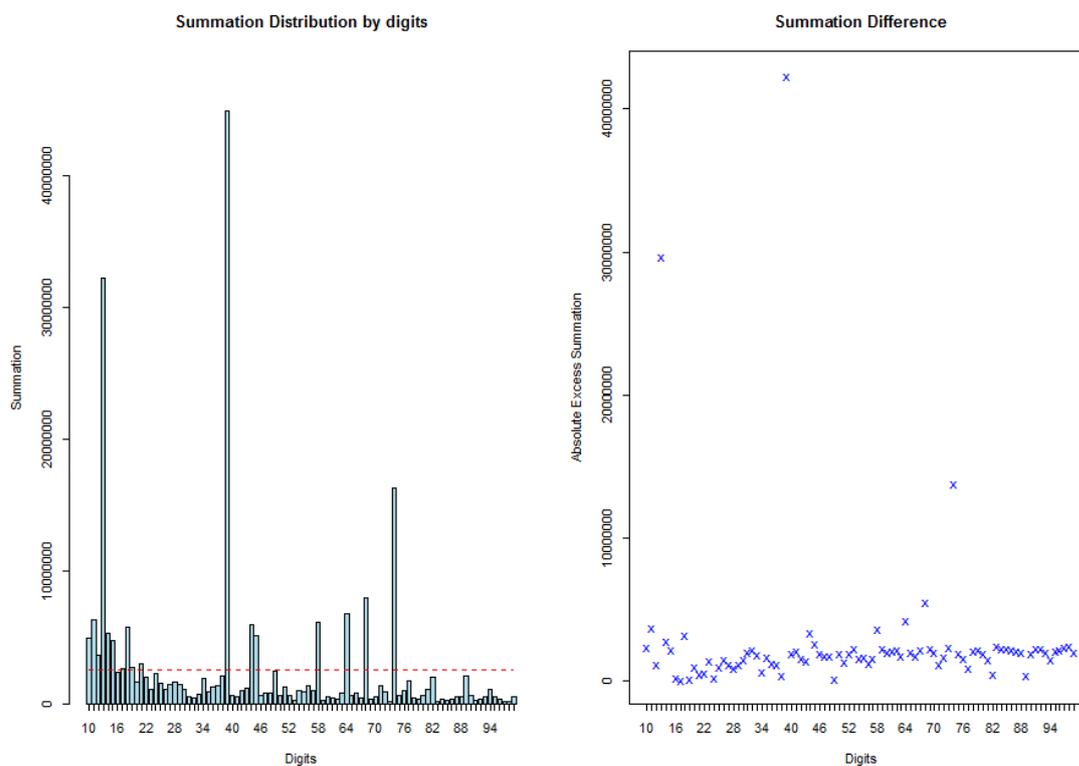
Fonte: O autor, 2021.

Figura 98 – Análise da soma para os dois primeiros dígitos do CeIM6



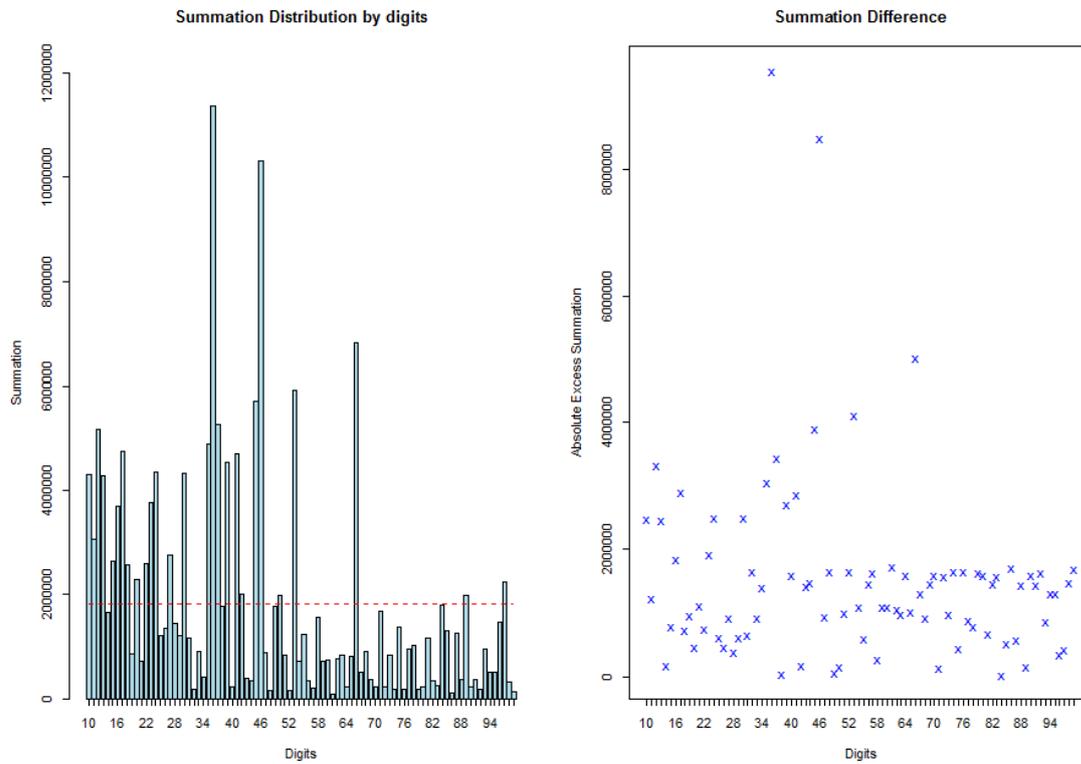
Fonte: O autor, 2021.

Figura 99 – Análise da soma para os dois primeiros dígitos do CeIM7



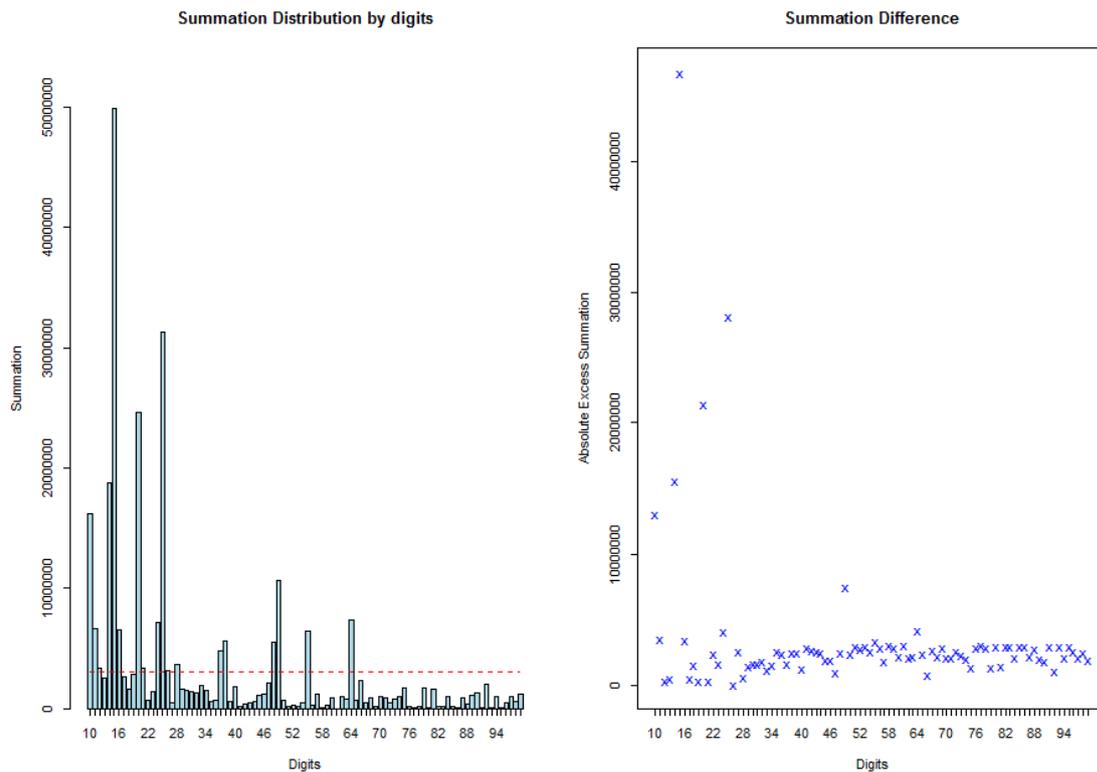
Fonte: O autor, 2021.

Figura 100 – Análise da soma para os dois primeiros dígitos da EAM1



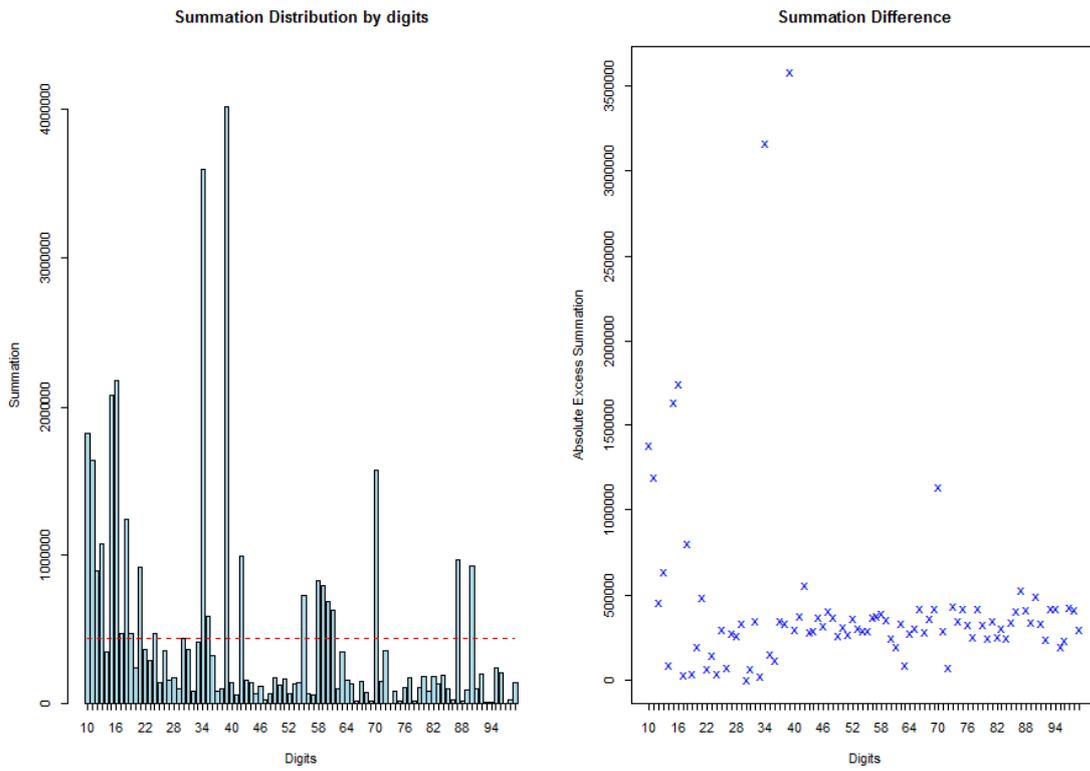
Fonte: O autor, 2021.

Figura 101 – Análise da soma para os dois primeiros dígitos da EAM2



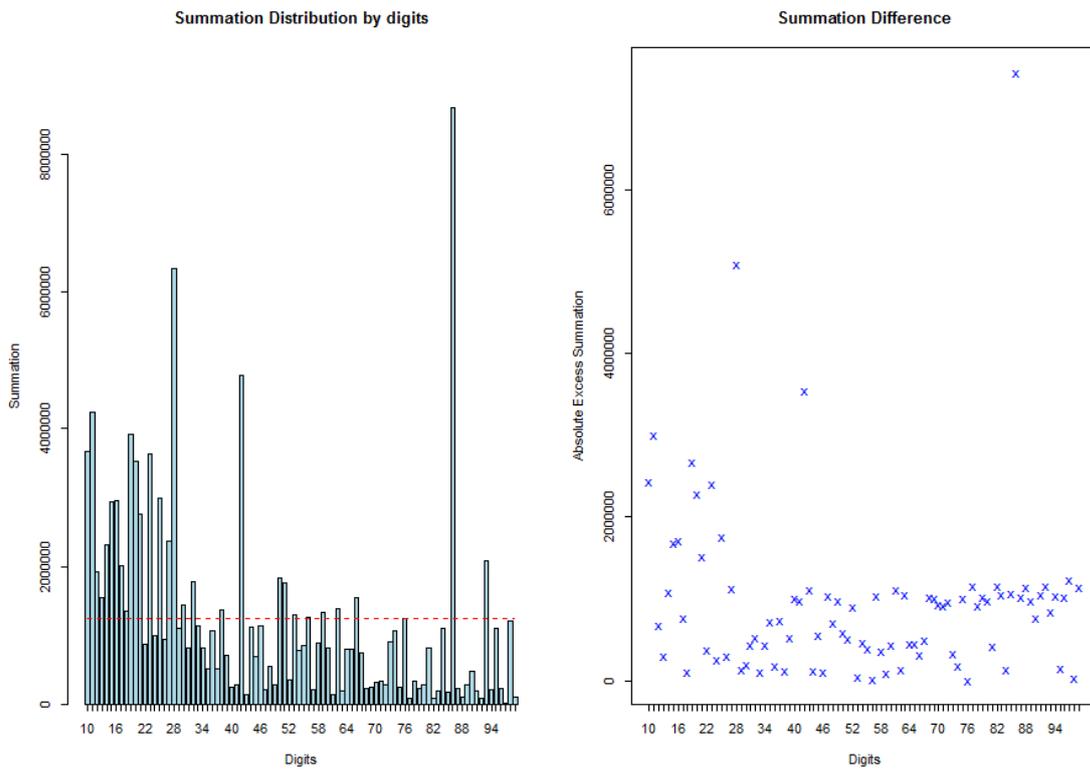
Fonte: O autor, 2021.

Figura 102 – Análise da soma para os dois primeiros dígitos da EAM3



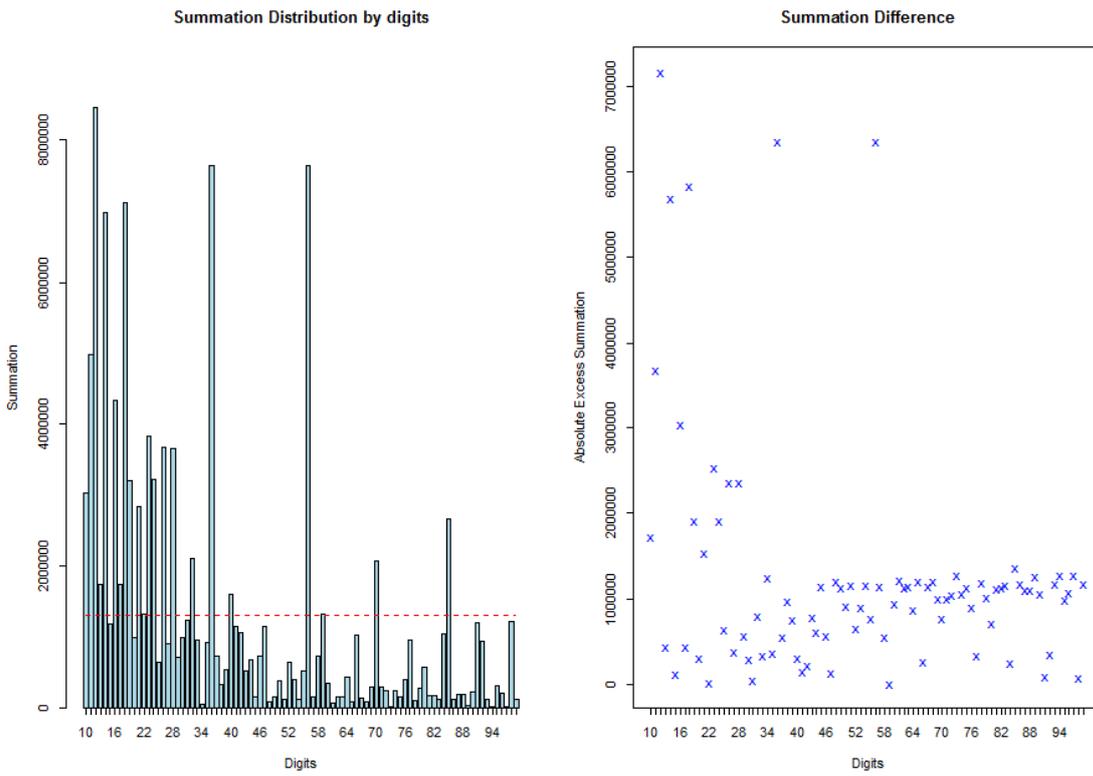
Fonte: O autor, 2021.

Figura 103 – Análise da soma para os dois primeiros dígitos da EAM4



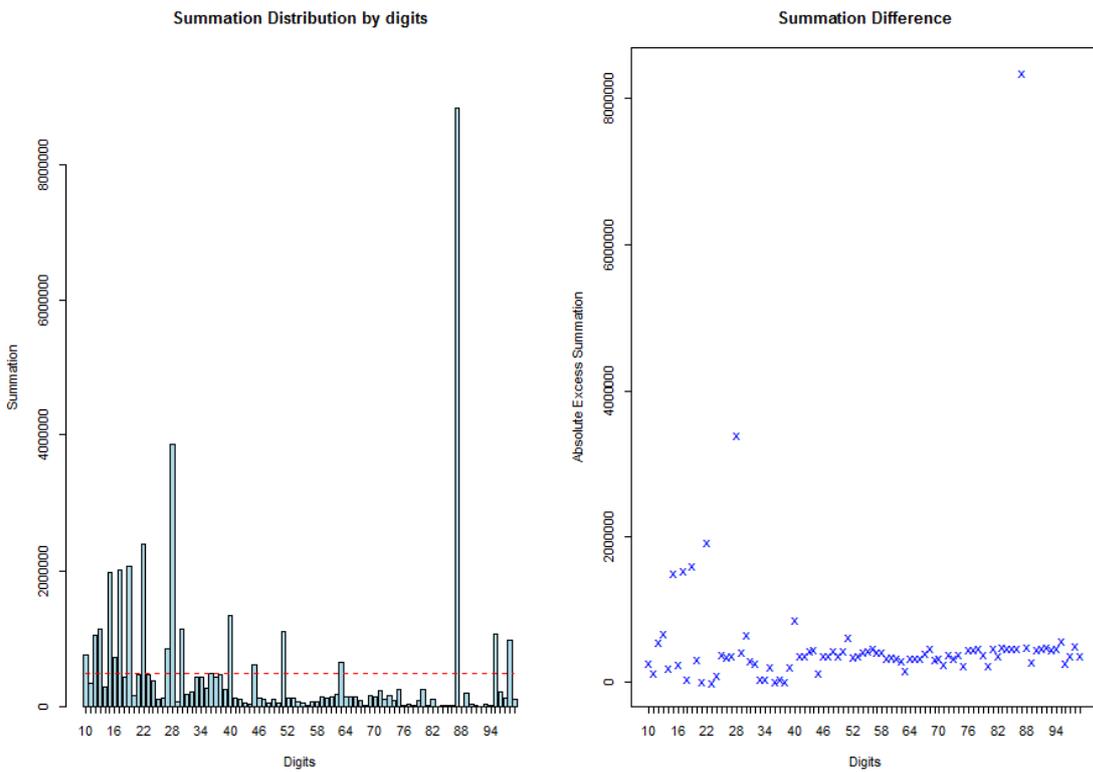
Fonte: O autor, 2021.

Figura 104 – Análise da soma para os dois primeiros dígitos da BE1



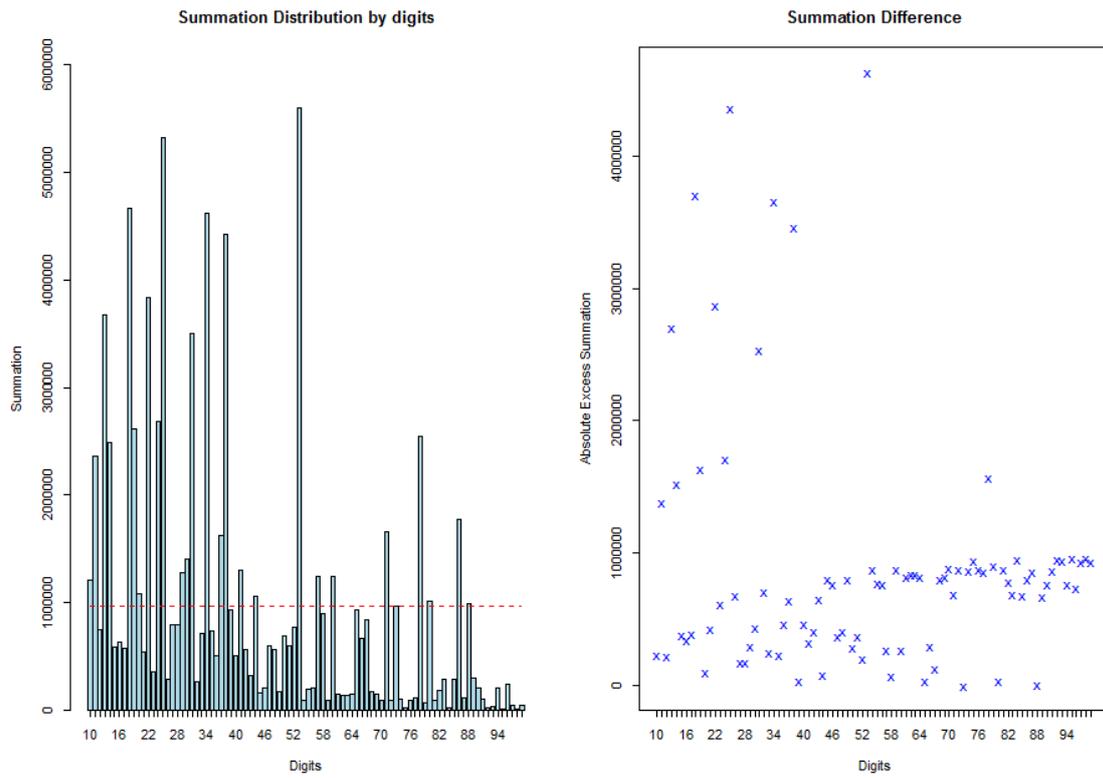
Fonte: O autor, 2021.

Figura 105 – Análise da soma para os dois primeiros dígitos da BE2



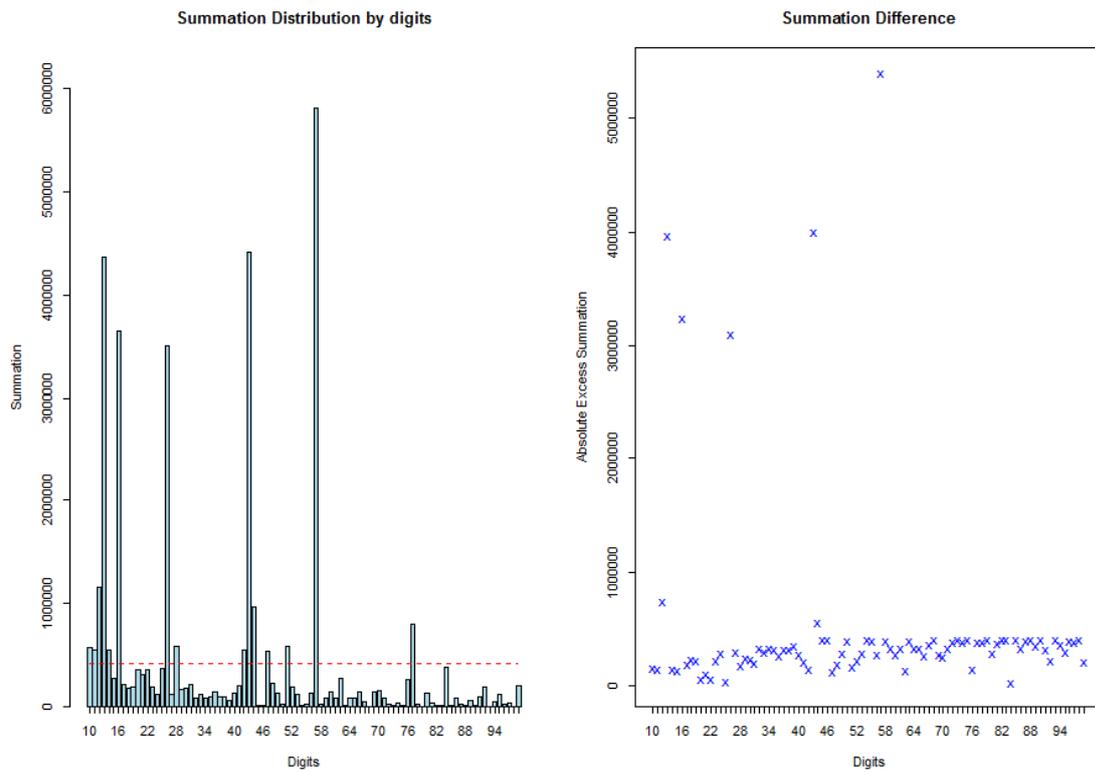
Fonte: O autor, 2021.

Figura 106 – Análise da soma para os dois primeiros dígitos da BE3



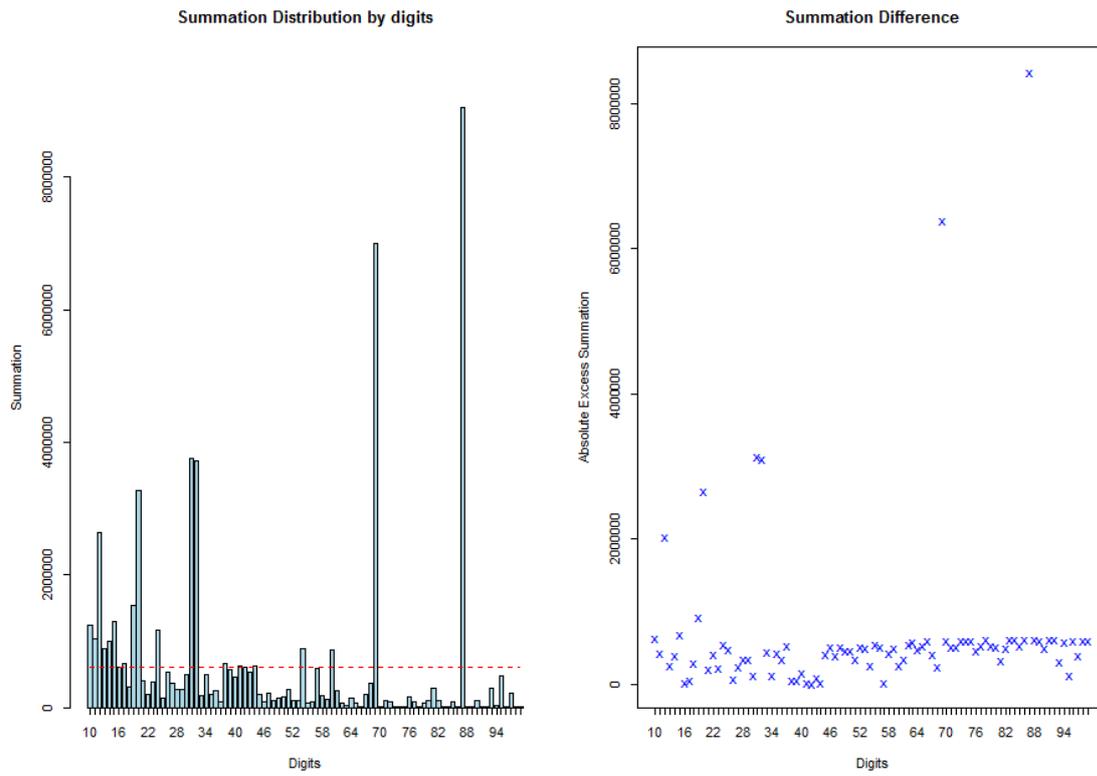
Fonte: O autor, 2021.

Figura 107 – Análise da soma para os dois primeiros dígitos da BE4



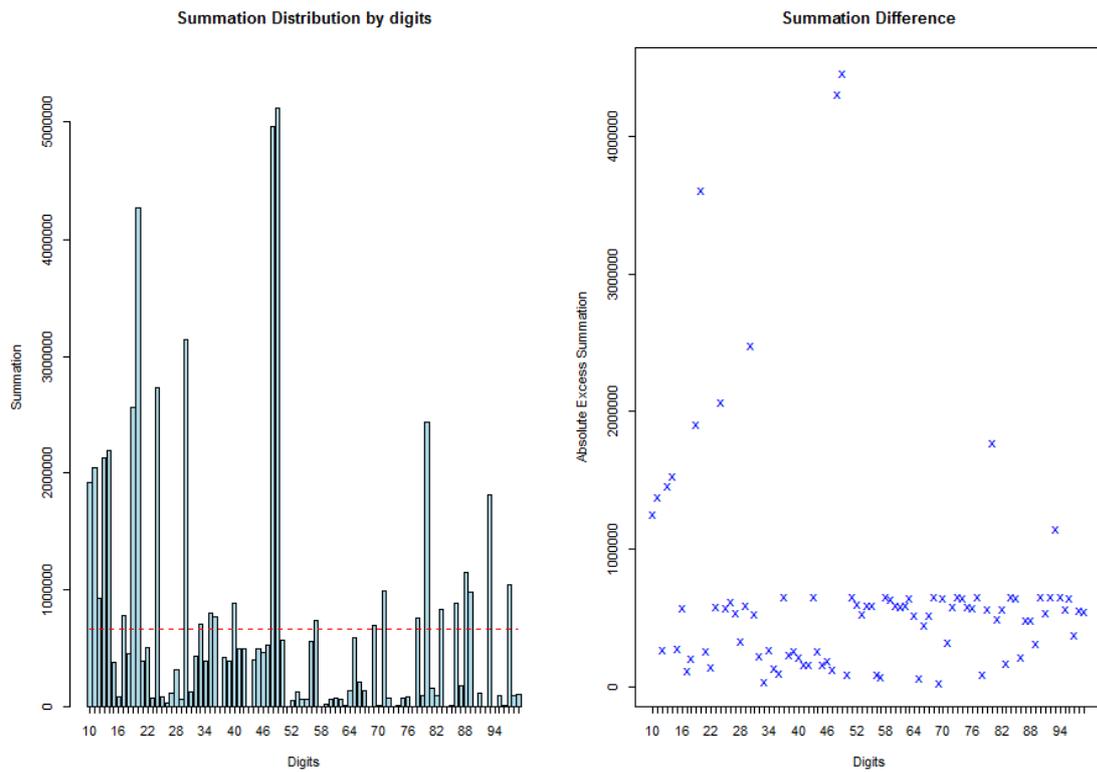
Fonte: O autor, 2021.

Figura 108 – Análise da soma para os dois primeiros dígitos da BE5



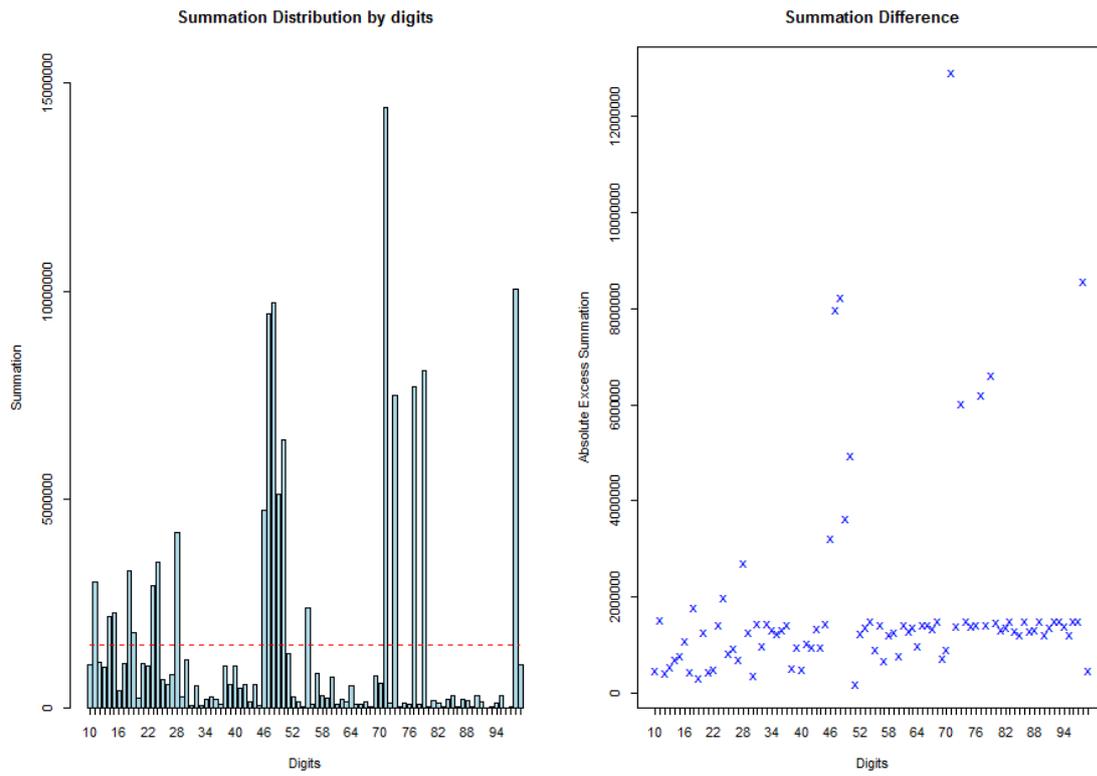
Fonte: O autor, 2021.

Figura 109 – Análise da soma para os dois primeiros dígitos da BE6



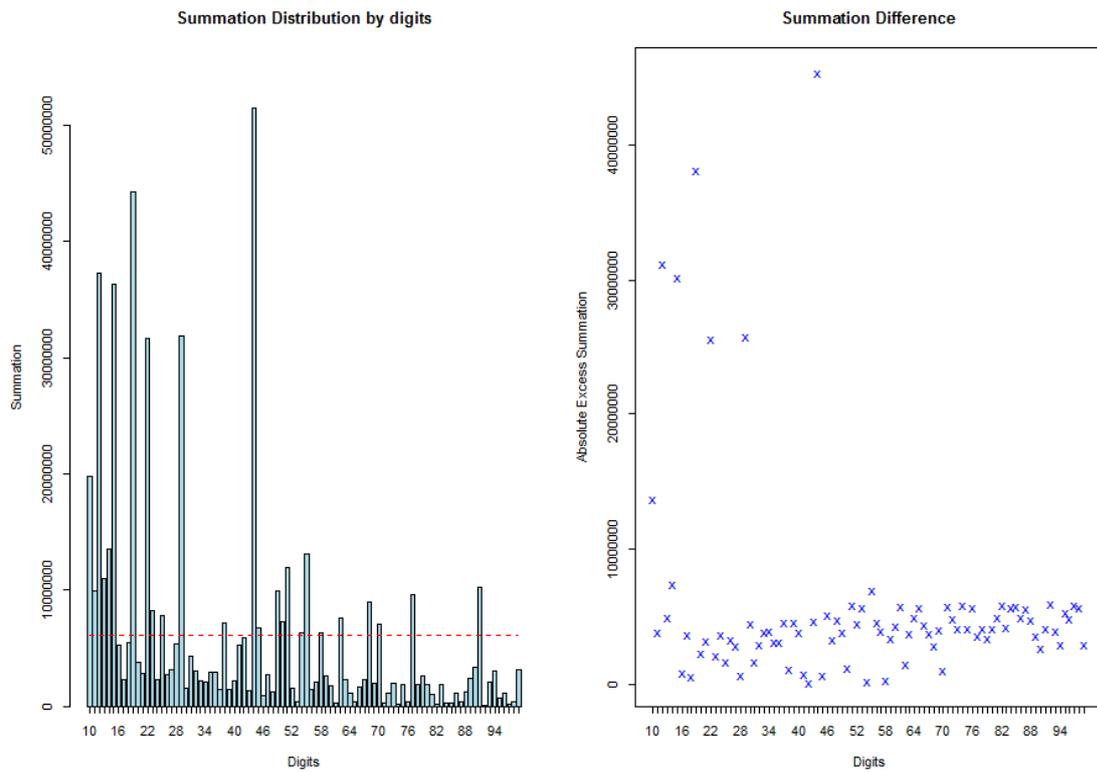
Fonte: O autor, 2021.

Figura 110 – Análise da soma para os dois primeiros dígitos da BE7



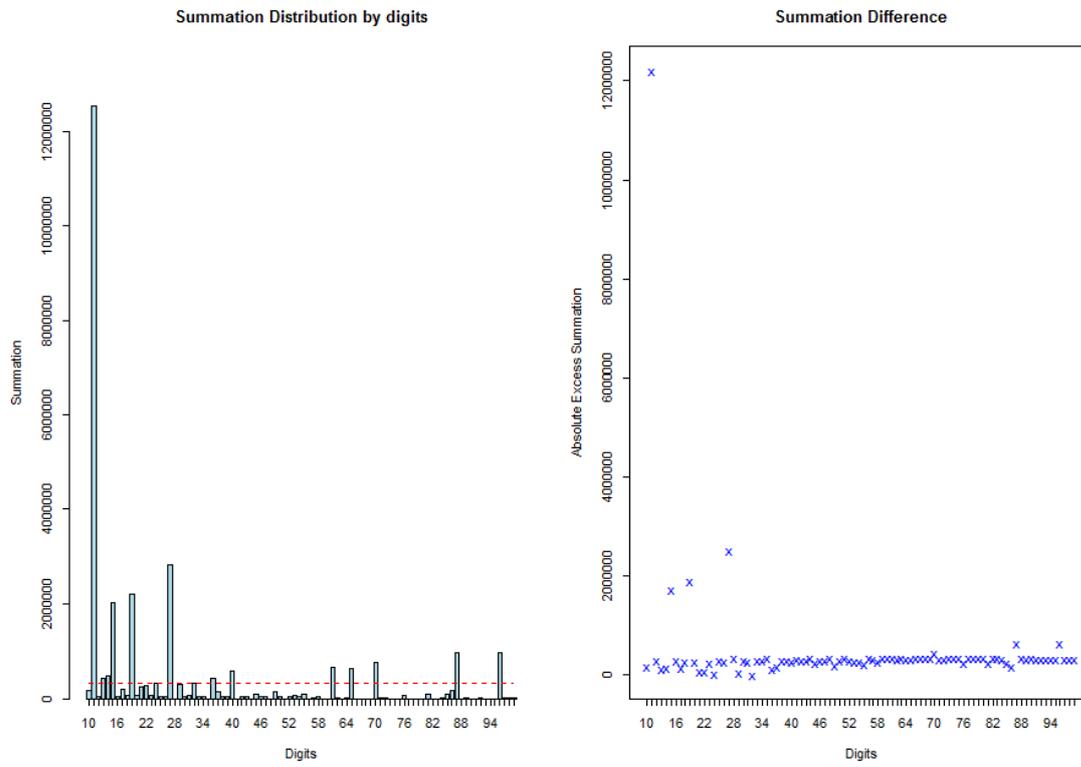
Fonte: O autor, 2021.

Figura 111 – Análise da soma para os dois primeiros dígitos da BE8



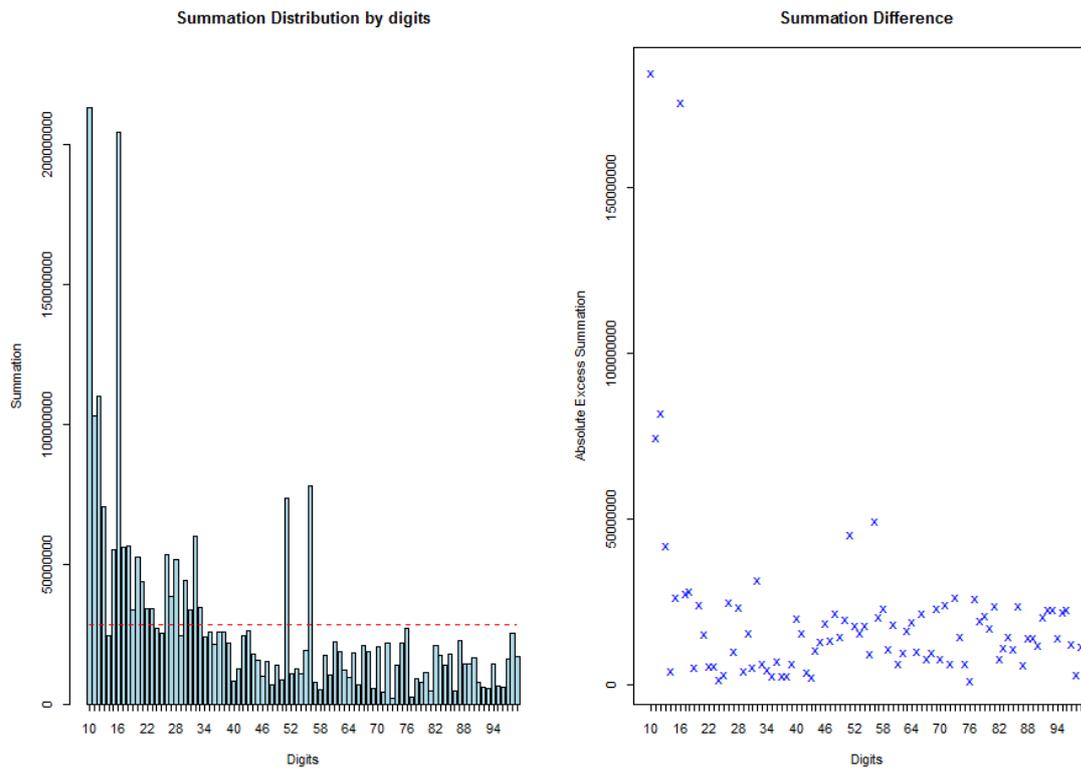
Fonte: O autor, 2021.

Figura 112 – Análise da soma para os dois primeiros dígitos do HN1



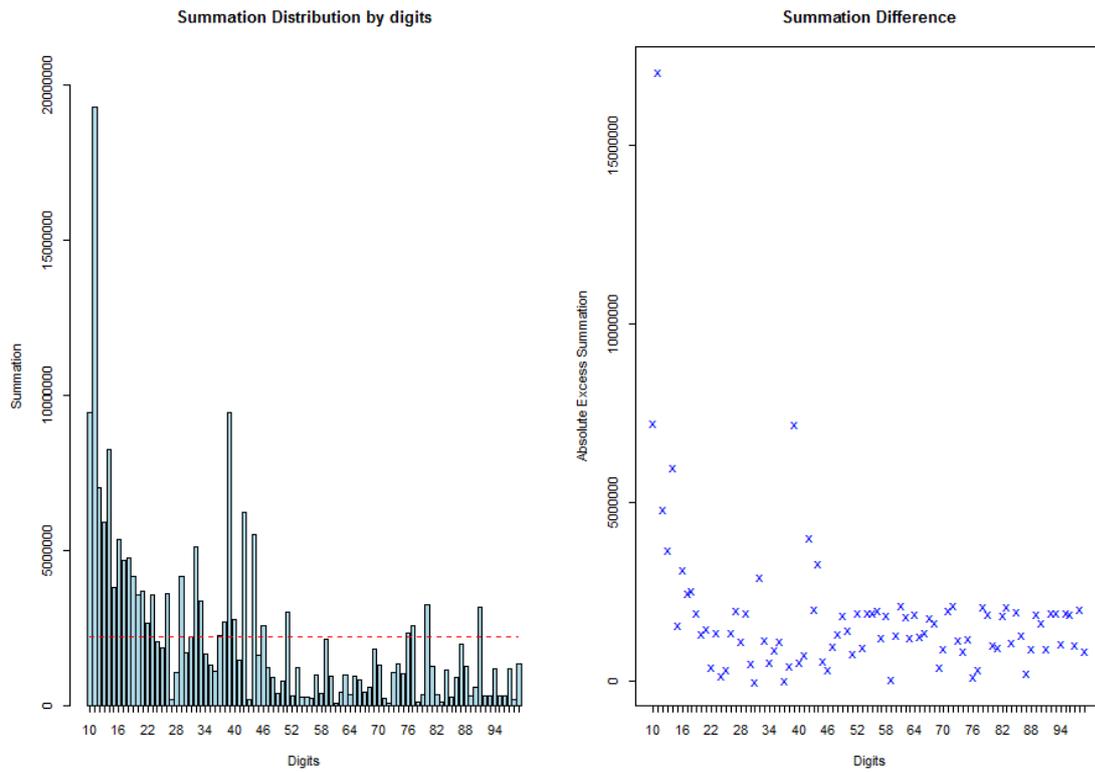
Fonte: O autor, 2021.

Figura 113 – Análise da soma para os dois primeiros dígitos do HN2



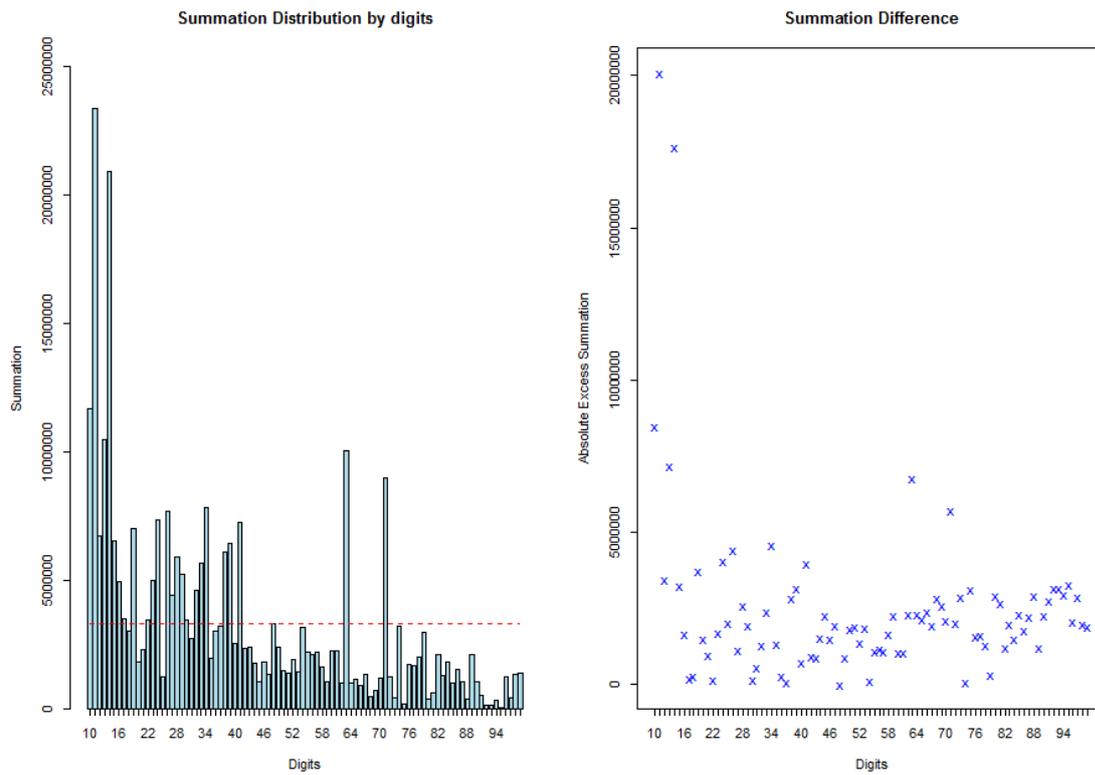
Fonte: O autor, 2021.

Figura 114 – Análise da soma para os dois primeiros dígitos do HN3



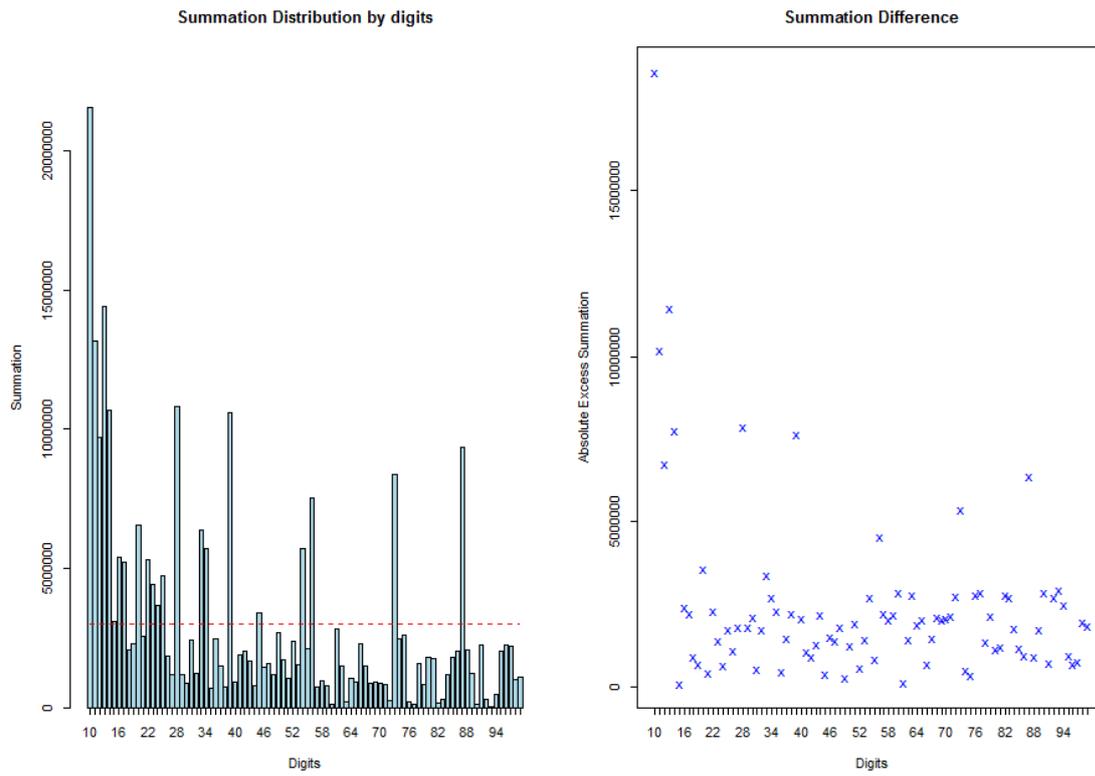
Fonte: O autor, 2021.

Figura 115 – Análise da soma para os dois primeiros dígitos do HN4



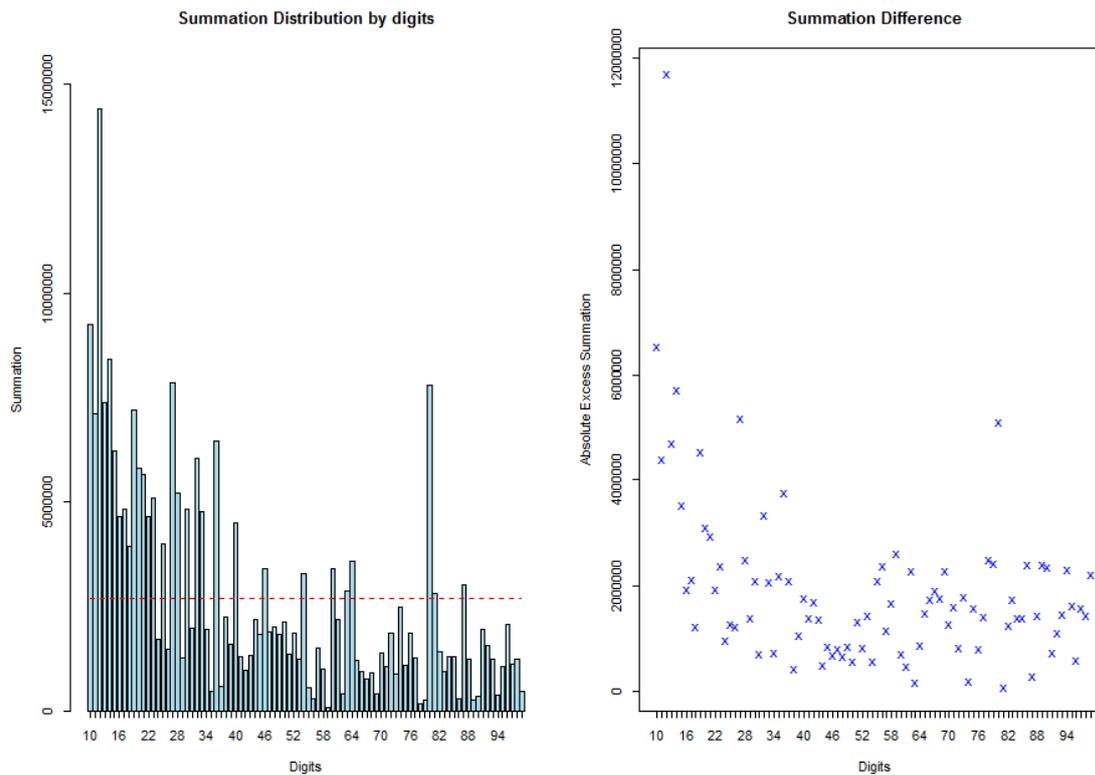
Fonte: O autor, 2021.

Figura 116 – Análise da soma para os dois primeiros dígitos do HN5



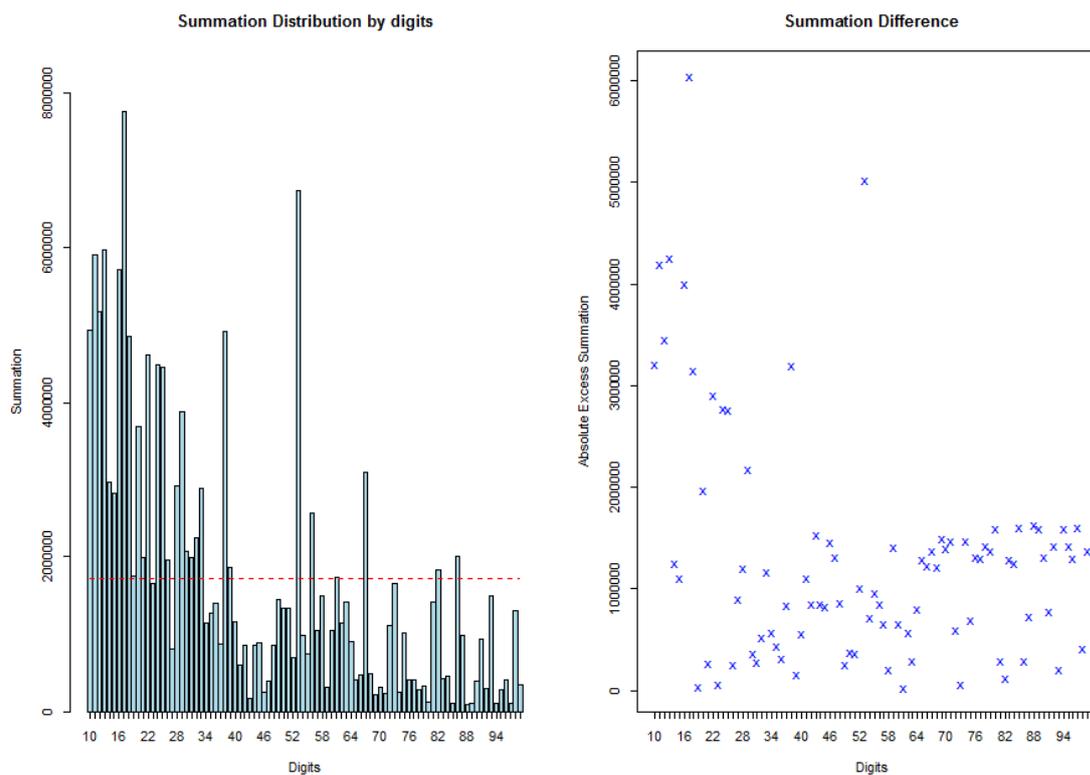
Fonte: O autor, 2021.

Figura 117 Análise da soma para os dois primeiros dígitos do HN6



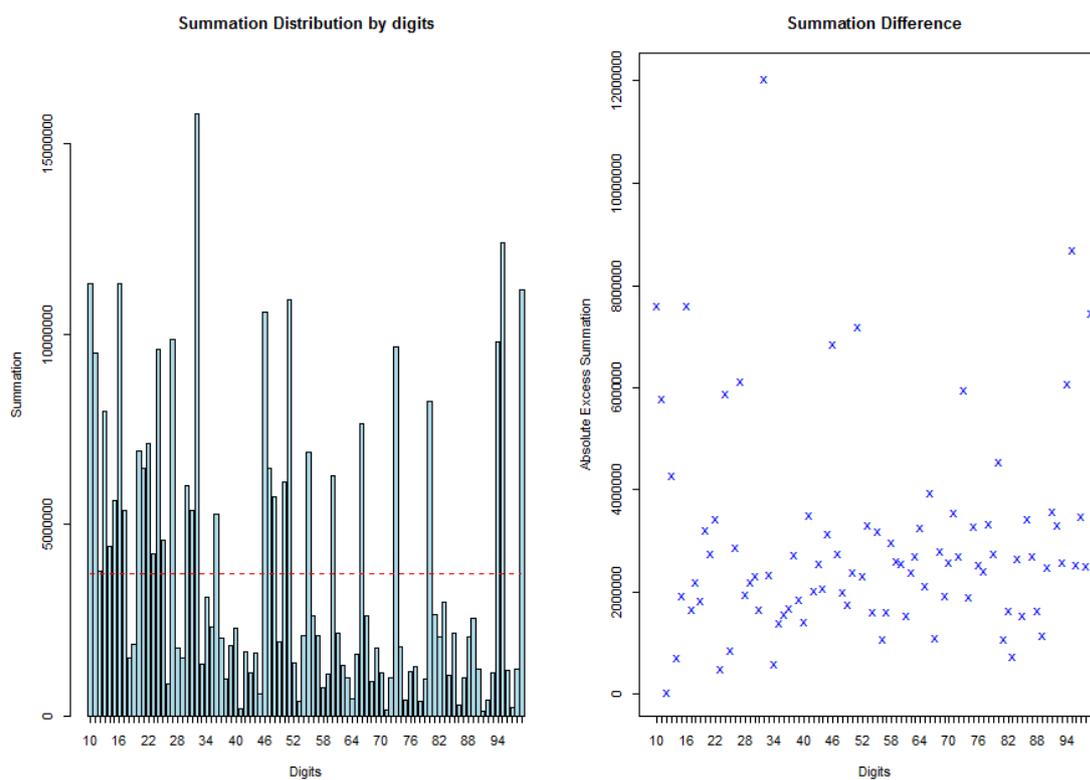
Fonte: O autor, 2021.

Figura 118 – Análise da soma para os dois primeiros dígitos do HN7



Fonte: O autor, 2021.

Figura 119 – Análise da soma para os dois primeiros dígitos do HN8



Fonte: O autor, 2021.

Tabela 29 – Maiores desvios da análise da soma para os primeiros dígitos

	Desvios positivos	Desvios negativos
CeIM1	1 e 2	7, 8 e 5
CeIM2	1 e 3	8, 6 e 9
CeIM3	1 e 2	7, 8 e 9
CeIM4	1 e 2	9, 4 e 5
CeIM5	4 e 1	7, 9 e 6
CeIM6	1 e 2	9, 7 e 4
CeIM7	1 e 3	9, 8 e 5
EAM1	3 e 1	7, 9 e 8
EAM2	1 e 2	8, 9 e 7
EAM3	1 e 3	8, 9 e 4
BE1	1 e 2	6, 9 e 7
BE2	1 e 8	7, 5 e 6
BE3	1 e 3	9, 6 e 8
BE4	1 e 4	9, 8 e 6
BE5	1 e 3	7, 9 e 5
BE6	1 e 4	6, 5 e 9
BE7	7 e 4	8, 6 e 3
BE8	1 e 2	8, 9 e 7
HN1	1 e 2	7, 5 e 9
HN2	1 e 2	9, 7 e 8
HN3	1 e 3	6, 9 e 5
HN4	1 e 3	9, 8 e 5
HN5	1 e 2	9, 6 e 4
HN6	1 e 2	9, 7 e 5
HN7	1 e 2	9, 7 e 4
HN8	1 e 2	7, 8 e 6

Fonte: O autor, 2021.

Tabela 30 – Maiores desvios da análise da soma para os dois primeiros dígitos

	Desvios positivos	Desvios negativos
CeIM1	10, 22 e 20	98, 94 e 88
CeIM2	16, 18 e 35	20, 66 e 85
CeIM3	43, 13 e 42	83, 76 e 97
CeIM4	12, 85 e 11	87, 83 e 78
CeIM5	80,53 e 24	95, 60 e 87
CeIM6	10, 60 e 37	96, 92 e 98
CeIM7	39, 13 e 74	83, 98 e 73
EAM1	36, 46 e 66	61, 86 e 99
EAM2	15, 25 e 20	61, 77 e 58
EAM3	39, 34 e 16	73, 97 e 93
EAM4	86, 28 e 42	97, 77 e 82
BE1	12, 56 e 36	97, 94 e 73
BE2	87, 28 e 22	83, 88 e 92
BE3	53, 25 e 18	95, 98 e 92
BE4	57, 43 e 13	79, 98 e 93
BE5	87, 69 e 31	86, 78 e 88
BE6	49, 48 e 20	37, 51 e 58
BE7	71, 98 e 48	96, 92 e 54
BE8	44, 19 e 12	92, 82 e 74
HN1	11, 27 e 19	44, 48 e 56
HN2	10, 16 e 12	73,77 e 71
HN3	11, 10 e 39	72, 61 e 78
HN4	11, 14 e 10	95, 93 e 92
HN5	10, 13 e 11	93,90 e 60
HN6	12, 10 e 14	59, 78 e 79
HN7	17,53 e 13	88, 97 e 85
HN8	32, 95 e 10	91, 71 e 41

Fonte: O autor,2021.

APÊNDICE E – Resultados das análises associadas

Tabela 31 – Valores com maior incidência nos CeIM

CeIM	Valor	Quantidade	CeIM	Valor	Quantidade
CeIM1	R\$ 14.406,00	48	CeIM5	R\$ 1.569,40	10
	R\$ 3.510,50	11		R\$ 1.210,30	9
	R\$ 17.346,00	8		R\$ 1.024,10	9
	R\$ 15.000,00	7		R\$ 1.120,00	9
	R\$ 20.038,80	7		R\$ 14.358,40	8
	R\$ 24.000,00	6		R\$ 3.699,30	7
	R\$ 67.200,00	6		R\$ 744,80	7
CeIM2	R\$ 1.770,00	8	CeIM6	R\$ 9.313,80	6
	R\$ 60.000,00	7		R\$ 2.640,00	12
	R\$ 30.000,00	7		R\$ 24.000,00	8
	R\$ 2.858,00	7		R\$ 12.000,00	8
	R\$ 10.000,00	6		R\$ 134.492,60	7
CeIM3	R\$ 480,00	34	R\$ 3.600,00	7	
	R\$ 8.000,00	12	R\$ 6.000,00	7	
	R\$ 12.000,00	8	R\$ 167,10	7	
	R\$ 40.000,00	8	R\$ 15.000,00	7	
	R\$ 1.604,80	7	R\$ 880,00	6	
	R\$ 378,00	7	R\$ 1.080,00	6	
	R\$ 1.220,00	6	R\$ 945,00	6	
	R\$ 132.300,00	6	R\$ 4.250,88	6	
	R\$ 26.460,00	6	R\$ 960,00	6	
	R\$ 13.845,00	6	R\$ 9.027,00	11	
CeIM4	R\$ 457,50	6	R\$ 6.000,00	11	
	R\$ 424,80	6	R\$ 2.832,00	10	
	R\$ 1.654,40	7	R\$ 3.600,00	9	
CeIM5	R\$ 1.000,00	6	R\$ 7.497,00	8	
	R\$ 448,40	41	R\$ 9.600,00	7	
	R\$ 372,40	34	R\$ 11.570,00	7	
	R\$ 784,70	26	R\$ 3.186,00	7	
	R\$ 1.457,30	18	R\$ 8.000,00	6	
	R\$ 1.233,10	17	R\$ 1.800,00	6	
	R\$ 896,80	15	R\$ 3.802,50	6	
	R\$ 651,70	11	R\$ 1.448,55	6	
R\$ 1.303,40	11	R\$ 4.457,50	6		

Fonte: O autor, 2021.

Tabela 32 – Valores com maior incidência nas EAM

EAM	Valor	Quantidade	EAM	Valor	Quantidade
EAM1	R\$ 4.248,00	4	EAM2	R\$ 6.000,00	6
	R\$ 8.000,00	4		R\$ 12.000,00	6
EAM2	R\$ 1.200,00	22		R\$ 2.700,00	6
	R\$ 1.440,00	16		R\$ 8.000,00	6
	R\$ 960,00	14		R\$ 3.360,00	6
	R\$ 640,00	11		R\$ 1.499,20	6
	R\$ 3.300,00	9		R\$ 9.800,00	6
	R\$ 2.800,00	9	EAM3	R\$ 240,00	62
	R\$ 16.800,00	9		R\$ 360,00	55
	R\$ 5.600,00	9		R\$ 200,00	31
	R\$ 7.496,00	9		R\$ 320,00	12
	R\$ 2.100,00	7		R\$ 3.600,00	7
	R\$ 5.400,00	7	EAM4	R\$ 1.800,00	5
	R\$ 10.000,00	6		R\$ 2.400,00	4
	R\$ 6.600,00	6		R\$ 973,50	4
R\$ 19.650,90	6	R\$ 7.134,00		4	
R\$ 20.880,00	6				

Fonte: O autor, 2021.

Tabela 33 – Valores com maior incidência nas BE

BE	Valor	Quantidade	BE	Valor	Quantidade
BE1	R\$ 30.000,00	7	BE5	R\$ 13.759,60	5
	R\$ 6.000,00	6		R\$ 26.425,20	4
	R\$ 40.000,00	5	BE6	R\$ 2.150,00	12
	R\$ 80.000,00	5		R\$ 2.858,00	12
	R\$ 12.968,19	5		R\$ 43.222,50	8
	R\$ 1.725,20	5		R\$ 56.189,43	7
	R\$ 37.491,65	5	BE7	R\$ 7.350,00	6
	R\$ 32.000,00	5		R\$ 1.470,00	6
	R\$ 53.475,80	5		R\$ 50.400,00	6
R\$ 1.058,40	6	R\$ 10.080,00		6	
R\$ 2.910,60	4	R\$ 18.185,70		5	
R\$ 1.323,00	4	R\$ 3.186,00		5	
BE2	R\$ 4.560,00	3	BE8	R\$ 64.285,56	5
	R\$ 10.054,80	3		R\$ 19.981,86	4
	R\$ 529,20	3		R\$ 77.000,00	4
	R\$ 360,00	8		R\$ 69.024,69	4
	R\$ 540,00	6		R\$ 22.220,56	3
	R\$ 1.920,00	5		R\$ 4.686,80	3
BE3	R\$ 1.120,00	5	R\$ 14.400,00	3	
	R\$ 240,00	5	R\$ 4.200,00	3	
	R\$ 8.400,00	23	R\$ 3.469,44	3	
	R\$ 28.320,00	7	R\$ 64.807,05	3	
	R\$ 121.000,00	5	R\$ 3.097,50	3	
	R\$ 480,00	3	R\$ 18.000,00	3	
BE4	R\$ 4.425,00	3	R\$ 35.000,00	3	
	R\$ 4.720,00	3	R\$ 16.800,00	3	
	R\$ 9.440,00	3	R\$ 1.609,12	3	
	R\$ 2.400,00	5	R\$ 52.000,00	3	
BE5	R\$ 2.400,00	5			

Fonte: O autor, 2021.

Tabela 34 – Valores com maior incidência nos HN

HN	Valor	Quantidade	HN	Valor	Quantidade
HN1	R\$ 7.600,00	2	HN4	R\$ 1.920,00	5
	R\$ 480,00	2		R\$ 1.600,00	5
	R\$ 7.200,00	2	HN5	R\$ 5.400,00	4
	R\$ 3.200,00	2		R\$ 18.000,00	4
	R\$ 5.106,50	2		R\$ 6.417,60	4
	R\$ 450,00	2		R\$ 7.000,00	4
	R\$ 9.600,00	2		R\$ 3.240,00	4
	R\$ 4.000,00	2	HN6	R\$ 40.000,00	12
	R\$ 15.200,00	2		R\$ 2.750,00	10
R\$ 18.000,00	27	R\$ 550,00		10	
R\$ 15.400,00	24	R\$ 2.500,00		10	
HN2	R\$ 33.600,00	23	HN6	R\$ 500,00	10
	R\$ 53.900,00	21		HN7	R\$ 1.200,00
	R\$ 9.200,00	21	R\$ 2.100,00		14
R\$ 5.760,00	8	R\$ 3.000,00	12		
HN3	R\$ 12.000,00	8	R\$ 3.600,00		8
	R\$ 28.800,00	6	R\$ 2.400,00		8
	R\$ 13.134,00	4	R\$ 2.000,00	8	
	R\$ 80.000,00	4	R\$ 2.800,00	8	
	R\$ 2.400,00	7	HN8	R\$ 67.200,00	6
R\$ 640,00	6	R\$ 22.992,00		5	
R\$ 6.000,00	5	R\$ 3.530,43		4	
R\$ 7.295,00	5	R\$ 4.200,00		4	
				R\$ 1.569,08	4

Fonte: O autor, 2021.

Tabela 35 – Resultados do modelo fator de distorção

	Fator de Distorção	Z_{teste}
CeIM1	-2,72%	-1,71
CeIM2	0,01%	0,01
CeIM3	-1,18%	-1,02
CeIM4	0,26%	0,21
CeIM5	-0,18%	-0,10
CeIM6	0,07%	0,05
CeIM7	-0,45%	-0,42
EAM1	-0,15%	-0,08
EAM2	3,38%	2,03
EAM3	-5,33%	-2,62
EAM4	0,25%	0,14
BE1	-3,04%	-1,645
BE2	2,77%	-1,281
BE3	0,64%	0,353
BE4	2,47%	1,105
BE5	-4,81%	-2,065
BE6	4,10%	1,234
BE7	3,78%	1,766
BE8	-1,17%	-0,754
HN1	4,01%	1,01
HN2	-1,23%	-1,54
HN3	-0,20%	-0,12
HN4	-2,15%	-1,67
HN5	-0,85%	-0,53
HN6	-1,86%	-1,24
HN7	-1,99%	-1,20
HN8	0,77%	0,53

Fonte: O autor, 2021.