

BENFORD COMO FERRAMENTA DE DETECÇÃO DE FRAUDES PELO MÉTODO DATA MINING, UM ESTUDO DOCUMENTAL.

Leonardo Soares Francisco de Almeida¹

Roni Cleber Bonizio²

Cesar Frederico dos Santos von Dollinger³

Bruna Barreto Santos Alves⁴

RESUMO

A crescente ansiedade das empresas por informações patrimoniais requer ações mitigadoras, de cunho objetivo instrucional. Esta pesquisa propõe realizar a avaliação do método de *Data Mining* aplicado a Lei de *Newcomb-Benford*, seleção amostral de dados, que apresentaram indícios de fraudes em conformidade nas organizações. Quanto a metodologia, documental, exploratória e explicativa. Os objetivos, apontam para realização de auditoria em conformidade, aplicando método *Data Mining* para seleção amostral, Lei de *Newcomb-Benford*, inferindo a busca de dados que vislumbrem indícios de manipulação. Foi realizado estudo de caso da empresa *Big-Benford*, pela análise das despesas de um centro de custo de tecnologia, nos anos de 2023 e 2024. Constatou-se fraudes por parte de colaboradores e terceiros. O achado da pesquisa demonstrou o benefício da importância da Lei de *Newcomb-Benford* como ferramenta de seleção amostral e instrumento de avaliação da integridade das informações financeiras, utilizadas pelos *stakeholders* na tomada de decisões.

PALAVRAS-CHAVE: Lei de *Newcomb-Benford*, fraude, auditoria de conformidade, seleção amostral, *Compliance*, *Data Mining*.

ABSTRACT

The growing concern of companies with asset-related information demands mitigating actions with an objective and instructional approach. This research proposes to evaluate the application of the Data Mining method using the Newcomb-Benford Law for sample data selection aimed at identifying signs of fraud in organizational compliance. The methodology is documentary, exploratory, and explanatory. The objective is to conduct a compliance audit by applying Data Mining based on the Newcomb-Benford Law to detect patterns that may indicate data

¹ Universidade Federal Fluminense – UFF- RJ/Brasil. Professor convidado do programa de MBA UFF Contabilidade. e-mail: leonardosoares@id.uff.br. Autor Correspondente.

² Universidade de São Paulo - USP. Professor doutor da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo - FEA-RP/USP. e-mail: rbonizio@usp.br. Autor

³ Universidade Federal Fluminense – UFF- RJ/Brasil. Professor Adjunto, Departamento de Ciências Contábeis.

⁴ Universidade Federal Fluminense – RJ/Brasil. Especialista do programa de MBA UFF Contabilidade. e-mail: brunabarreto@id.uff.br. Autor

manipulation. A case study was carried out on the company Big-Benford, analyzing the expenses of a technology cost center during the years 2023 and 2024. Fraud involving both employees and third parties was identified. The research findings demonstrated the usefulness of the Newcomb-Benford Law as a sample selection tool and an instrument to assess the integrity of financial information used by stakeholders in decision-making.

KEYWORDS: *Newcomb-Benford Law, fraud, compliance audit, sample selection, Compliance, Data Mining.*

1. INTRODUÇÃO

Num ambiente organizacional cada vez mais globalizado, restrito, tecnológico e em constante mudança, percebe-se o aumento da ocorrência de fraudes e desvios.

A fraude representa um risco eminente aos negócios, devido a diversos fatores, como a ausência da cultura de riscos, problemas relacionados à cultura e estrutura organizacional, governança corporativa das empresas e/ ou aos órgãos independentes, como Controles Internos, Auditoria Interna e *Compliance*.

De acordo com Mills (1994), a auditoria de conformidade consiste na realização de avaliação com a finalidade de assegurar que o sistema, programa, produto, serviço e/ ou processo aplicáveis apresentem todas características, critérios e parâmetros demandados.

Por outro lado, o avanço tecnológico também contribui para a informatização dos processos, o que viabiliza uma análise de conformidade para detecção das fraudes mais tempestiva e efetiva.

Dentre os testes estatísticos mais utilizados na análise de conformidade de um conjunto de dados, encontra-se a Lei de *Newcomb-Benford*, baseada na obra de *Newcomb* (1881) e *Benford* (1938).

De posse destes dados chegou-se ao problema da pesquisa, que se propõe em verificar e responder a seguinte pergunta: **A Lei de *Newcomb-Benford* representa uma ferramenta de controle e apoio à realização de auditoria para gestão financeira do negócio? Quais suas vantagens e desvantagens?**

2. FUNDAMENTAÇÃO TEÓRICA

2.1. FRAUDE

Um dos trabalhos mais antigos sobre o assunto foi realizado por Cressey (1953), que verificou o comportamento de fraudadores, descobrindo três fatores comuns associados ao ato de cometê-las: a existência de uma motivação (pressão), a presença de oportunidades (percepção sobre a vulnerabilidade do que se pretende fraudar) e a percepção moral do indivíduo sobre o certo e errado (racionalização). Dessa forma, surge a teoria do Triângulo da Fraude de Cressey:

Elementos da Teoria do Triângulo da Fraude

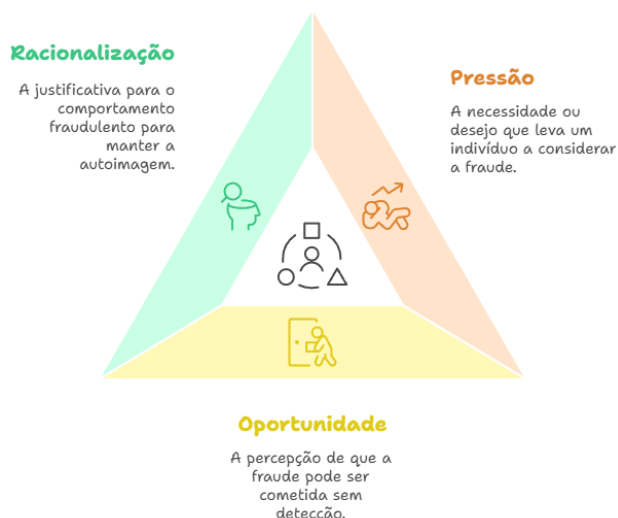


Figura 1 – Triângulo das fraudes

Fonte: Albrecht (2003)

O tema “fraude” ganhou destaque no cenário corporativo dada a sua recorrência e sofisticação ao longo dos últimos anos. Casos como Petrobras, Siemens, Banco Panamericano e Enrol coloca o assunto cada vez mais em voga apresentando ocorrência a aproximadamente duas décadas num contexto nacional e internacional.

2.2. DATA MINING

Num mundo onde grandes quantidades de dados são coletadas diariamente, empresas que geram enormes conjuntos de dados, incluindo transações de vendas, registros financeiros, descrições de produtos, promoções, perfis de empresas e desempenho e comentários dos clientes

devem observar e acompanhar sua origem e veracidade. Logo, descobrir informações valiosas e transformá-las em conhecimento organizado levou ao nascimento do *Data Mining*, ou, mineração de dados (HAM & KAMBER & PEI, 2012).

2.3. A LEI DE NEWCOMB-BEEENFORD

Em um contexto geral, a probabilidade de frequência do dígito 1 ao 9 cai exponencialmente, conforme mostra gráfico 01 em anexo. Esse evento denomina-se em termos estatísticos como “anomalia” e sua distribuição se aplica em análise dos dígitos em um amplo conjunto de dados numéricos, desde números contidos em páginas de jornais a números referentes a pesos atômicos (NIGRINI, 2000).

A Lei de *Newcomb-Benford*, também chamada de lei dos primeiros dígitos, determina que em uma sequência de valores, a distribuição esperada dos dígitos inferiores, ou seja, os dígitos 1 e 2, por exemplo, possui maior frequência do que os dígitos superiores, como 8 e 9 (HILL, 1995).

2.3.1. Histórico

Segundo Hill (1995), a Lei de *Newcomb-Benford* surgiu em 1938, quando o físico da *General Electric Company*, Frank Benford, publicou o artigo *The law of anomalous numbers* na revista *Proceedings*, da Sociedade Filosófica Americana.

Sua pesquisa se baseou no estudo de 20.229 conjuntos de números de diversas fontes, como distância de rios, contagens da população, estatísticas da liga de beisebol, números de endereços, tabelas com pesos atômicos, taxa de mortalidade, entre outros (HILL, 1995).

Benford (2020), ao lado de autores como Hill, Schneider, Rives e Tarter, oferecem uma base teórica robusta para a aplicação da Lei de *Benford*, explorando desde fundamentos matemáticos até aplicações em auditoria, segurança e modelagens estatísticas. Essa visão integrada, reforçada por estudos de Knorr, Ng, Moreno, Pedrosa, Kraus, Torres-Zúñiga, Tran e Tassinari, ampliam a aplicabilidade da lei para os domínios ambiental, contábil e computacional, descrevendo sua importância (BENFORD, Frank, 2020; HILL, Theodore P.; SCHNEIDER, Lenny, 2022; RIVES, Julio, 2024; TARTER, Timothy, 2025; KNORR, Edwin M.; NG, Raymond T., 2000; MORENO, David; VALLEJOS, Rocío; DI RIENZO, Julio, 2021; PEDROSA, Isabel; COSTA, Carlos J.; LAUREANO, Raul M. S., 2015; KRAUS, Peter; VALVERDE, Carmen,

2014; TORRES-ZÚÑIGA, Vicente, 2024; TRAN, Quoc T.; HUYNH, Tuan V., 2021; TASSINARI, Francisco, 2010).

As evidências indicaram que, ao verificar que os seus livros de logaritmos estavam mais gastos nas páginas que se iniciavam com os primeiros dígitos do que as que se iniciavam com os últimos, *Benford* (1938) realizou uma análise do estudo do astrônomo americano Simon Newcomb (1881), *Note on the frequency of use of the different digits in natural numbers*, na década de 20. Newcomb (1881) havia calculado a probabilidade de ocorrência dos números de 1 a 9 no primeiro dígito.

Frank Benford (2020) estabeleceu as bases do conceito de variáveis aleatórias com aderência condicional à Lei de *Benford*, alertando que sua obediência depende fortemente da base numérica usada.

A análise do espectro benfordiano mostra que variáveis que seguem a distribuição logarítmica em base 10 nem sempre mantêm esse padrão em outras bases, como 2 ou 16, o que exige cuidado na aplicação estatística (Benford, Frank, 2020). As probabilidades de ocorrência de distribuição do primeiro e do segundo dígito segundo a Lei de *Newcomb-Benford*, podem ser verificadas conforme tabela 01 em anexo.

2.3.2. Testes Estatísticos na Aplicação da Lei de *Newcomb-Benford*

Testes estatísticos como Teste-Z, X^2 (quiquadrado) e Média dos Desvios Absolutos são comumente utilizados para determinar se a diferença decorrente da comparação dos dados analisados em relação ao padrão da Lei de *Newcomb-Benford* representa uma não conformidade (NIGRINI, 1996). Segundo Nigrini (1996), o objetivo dos testes estatísticos tem como premissa provar uma das seguintes hipóteses:

- Não há diferença significativa, em termos estatísticos, entre as distribuições analisadas e a distribuição segundo a Lei de *Newcomb-Benford*; e
- Há significativa diferença, em termos estatísticos, entre as distribuições analisadas e a distribuição segundo a Lei de *Newcomb-Benford*.

Vale ressaltar que segundo Bugarin & Cunha (2015), importante confrontar ao menos dois testes, sendo estes:

Teste do Primeiro Dígito: segundo Bugarin & Cunha (2015), esse teste analisa as frequências de repetição dos números 1 a 9 nos primeiros dígitos dos valores de uma determinada

base de dados, dividindo, portanto, a amostra em 9 grandes grupos para proporcionar um panorama geral. A seguinte expressão traduz a distribuição esperada de um primeiro dígito:

$$\text{Prob}(D_1 = d_1) = \log_{10} \left(1 + \frac{1}{d_1} \right)$$

Onde, D_1 = primeiro dígito e $d_1 \in \{1, 2, \dots, 9\}$.

Como pode ser visto, *Prob* indica a probabilidade de observar o evento (em parênteses). Em seguida, log refere-se ao log da base 10. Nesse caso, a probabilidade esperada do primeiro dígito ser o número 1 consistem em $\log(1 + 1/1)$, igual a 0,3010 (conforme apresentado na tabela 01).

Teste do Segundo Dígito: segundo Bugarin & Cunha (2015), esse teste analisa as frequências de repetição dos números 0 a 9 nos segundos dígitos dos valores de uma determinada base de dados. Dividindo as amostras em 10 grupos, também apresenta um panorama geral, mas pode ser útil na detecção de desvios nos dados. A seguinte expressão traduz a distribuição esperada de um segundo dígito:

$$\text{Prob}(D_2 = d_2) = \sum_{d_1=1}^9 \log \left(1 + \frac{1}{d_1 d_2} \right);$$

Onde D_2 = segundo dígito e $d_2 \in \{0, 1, \dots, 9\}$.

Teste dos Dois Primeiros Dígitos: segundo Bugarin & Cunha (2015), esse teste, diferentemente dos demais apresentados, divide as amostras em 90 grupos (10 a 99), apresentando maior precisão nos resultados. A seguinte expressão traduz a distribuição esperada dos dois primeiros dígitos:

$$\text{Prob}(D_1 D_2 = d_1 d_2) = \log \left(1 + \frac{1}{d_1 d_2} \right);$$

Onde $D_1 D_2$ = Dois primeiros dígitos e $d_1 d_2 \in \{10, 11, \dots, 99\}$

Teste da Soma: segundo Nigrini & Wells (2012), o teste da soma norteia o auditor na análise dos dados, indicando números demasiadamente grandes ou excessivas repetições e alertando quanto à necessidade de maior cuidado na avaliação das informações.

Teste Z: teste comumente utilizado para verificar se a proporção real de um dígito varia de maneira significativa em relação à proporção esperada segundo a Lei de *Newcomb-Benford*. A estatística Z está em função do desvio absoluto entre as proporções esperadas e verificadas e do tamanho da amostra. A seguinte expressão traduz o Teste Z:

$$Z = \frac{|PR - PE| - \left(\frac{1}{2n}\right)}{\sqrt{\frac{PE(1-PE)}{n}}}$$

PR representa a proporção real, **PE** a proporção esperada, “**n**” o tamanho da base de dados e $1/2n$ representa um fator de correção usado somente quando inferior ao primeiro termo do numerador (devido ao desvio absoluto entre as proporções).

Segundo Nigrini & Wells (2012), o teste Z efetua a análise dos dígitos de forma individual, não indicando a aderência dos dados de forma integrada, podendo haver aproximadamente cinco picos significativos no teste dos dois primeiros dígitos.

X² (quiquadrado): segundo Bugarin & Cunha (2015), a estatística Qui-Quadrado de Pearson ou Chi-Square (CS) realiza comparação entre a frequência absoluta e a frequência esperada dos conjuntos de dados para verificar se os dígitos da distribuição (como um todo) são aderentes à Lei de *Newcomb-Benford*. A seguinte expressão traduz a estatística Qui-Quadrado:

$$\chi^2 = \sum_{i=1}^K \frac{(CR - CE)^2}{CE}$$

K representa as categorias das séries dos dígitos ($K=9$, $K=10$, $K=90$), **CR** representa a contagem real dos números da categoria, **CE** representa a contagem esperada para a categoria conforme a Lei de *Newcomb-Benford* e $(K-1)$ representa o número dos graus de liberdade (ou, a quantidade de dígitos analisados menos 1). Se o resultado for menor que o limite tabulado, tende-se à hipótese de aderência à distribuição conforme a Lei de *Newcomb-Benford*.

Cada dígito está relacionado a um valor qui-quadrado e, dessa forma, a soma de todos leva ao resultado do teste, comparado a um valor crítico. No caso do teste dos dois primeiros dígitos, consideram-se 89 graus de liberdade, que a um nível de significância de 5%, o valor crítico resulta em 112,022 e, caso contrário, a hipótese de aderência “pode” ser rejeitada.

Média dos Desvios Absolutos: segundo Nigrini & Wells (2012), a Média dos Desvios Absolutos refere-se à estatística calculada independentemente do tamanho da amostra, apresentando-se, dessa forma, como o teste estatístico mais apropriado para bancos de dados com maiores números de registros. Calculada a partir da média das frequências relativas e esperada de cada dígito, a estatística MAD se apresenta pela seguinte expressão:

$$MDA = \sum_{i=1}^k \frac{|PR - PE|}{k}$$

PR diz respeito à frequência relativa identificada para o primeiro dígito *i*, **PE** representa a frequência relativa esperada para *i* segundo a Lei de *Newcomb-Benford* e **K** refere-se às categorias das séries dos dígitos ($K=9$, $K=10$, $K=90$). Os 90 desvios absolutos calculados devem ser somados e, posteriormente, divididos pelo número de realizações possíveis (90).

Ao contrário dos demais testes, a **estatística MAD** não possui valores críticos objetivos. Nigrini (2000) oferece algumas diretrizes baseadas em sua experiência pessoal com banco de dados que foram testados à luz da Lei de *Benford*. Os seguintes valores críticos devem ser observados, conforme tabela 06 em anexo.

3. METODOLOGIA

A linha metodológica da pesquisa tem por base um “estudo de caso”, de uma empresa brasileira fictícia, da indústria de tecnologia que passou por problemas de fraudes financeiras em 2023 e 2024. Pretende-se demonstrar a aplicação da Lei de *Newcomb-Benford* como método de seleção amostral para direcionamento de análises de conformidade.

A partir da fundamentação teórica, avaliar o método de *Data Mining* denominado Lei de *Newcomb-Benford*, e sua efetividade como instrumento de seleção amostral aplicados a auditoria de conformidade, na identificação de fraudes financeiras *intramuros*.

Mediante análises do estudo de caso, verificou-se a efetividade da Lei de *Newcomb-Benford* como ferramenta de seleção amostral, limitados à uma base de notas fiscais de empresa **BIG-BENFORD**, que apresentou problemas de desvios financeiros.

4. ESTUDO DE CASO – **BIG-BENFORD**

4.1. OBJETIVO

Aplicou a Lei de *Newcomb-Benford* às despesas de um centro de custo da empresa de tecnologia fictícia nacional nos anos de 2023 e 2024. Identificou-se fraudes financeiras por parte de colaboradores e fornecedores.

A empresa analisada caracteriza-se em nacional do ramo de tecnologia de grande porte, com lucro de aproximadamente R\$ 2 bilhões e faturamento de R\$ 15 bilhões por ano, com cerca de 11 mil colaboradores. Possui uma estrutura de Controles Internos, uma área de Conformidade e auditoria contábil independente, que não apresentou reservas ou ênfases no relatório de auditoria dos anos de 2023 e 2024.

Quanto aos dados de despesas foram extraídos de base de notas fiscais.

4.2. ANÁLISES

Segundo Nigrini & Wells (2012), ao aplicar a Lei de *Newcomb-Benford* em uma base de dados, esta não deve sofrer nenhum tipo de tratamento anterior, como agrupamentos realizados na Curva ABC, por exemplo. A base de notas fiscais do centro de custo em questão fornecidas refere-se aos anos de 2023 e 2024.

Foram identificadas fraudes nas despesas nos anos de 2023 e 2024, base de dados deste estudo de caso que possui 5.574 registros, quantidade adequada para análise segundo os estudos de (Nigrini & Wells 2012).

Vale ressaltar que a base em questão se alinha com os pressupostos de aderência à Lei de *Newcomb-Benford*, visto que possui uma quantidade suficiente de itens (5.574 registros), composta por números gerados de forma natural (sem a atribuição de limites máximos e/ ou mínimos).

Complementarmente, aplicou-se o Teste do Primeiro Dígito em uma fração da base de dados referente a um fornecedor que efetuou fraudes (362 registros) e um segundo que não efetuou fraudes (453 registros), verificados por auditorias de conformidade realizadas.

Os resultados dos testes do fornecedor onde ocorreu fraudes mostram uma desconformidade das frequências de distribuição dos dígitos em relação à Lei de *Newcomb-Benford*, enquanto que os resultados do fornecedor que não efetuou demonstraram conformidade das distribuições dos dígitos em relação à Lei de *Newcomb-Benford*.

Esse fato demonstra provável adequação aderente dos dados à Lei de *Benford*, visto que a base de dados em questão sofreu manipulação humana.

Para avaliar se as diferenças identificadas entre as frequências de dígitos observadas e as frequências de dígitos da Lei de *Newcomb-Benford* são estatisticamente significativas, aplicou-se os testes estatísticos: Teste-Z, Qui-Quadrado e Média dos Desvios Absolutos sobre os resultados dos testes do Primeiro Dígito, dos Dois Primeiros Dígitos e da Soma. Sendo o objetivo testar as duas hipóteses a seguir:

- H0 (nula): Não há diferença significativa, em termos estatísticos, entre as distribuições analisadas e a distribuição segundo a Lei de *Newcomb-Benford*; e;
- H1: Há significativa diferença, em termos estatísticos, entre as distribuições analisadas e a distribuição segundo a Lei de *Newcomb-Benford*.

4.2.1. Teste do Primeiro Dígito

Extraíndo os valores contidos na base de notas fiscais referentes a 2023 e 2024 do referido centro de custo da organização alvo deste estudo, foram obtidos um total de 5.574 registros de despesas de diversos fornecedores. A coluna “Valor da Nota Fiscal” foi utilizada para verificação do primeiro dígito.

Foram contabilizadas as quantidades de itens por primeiro dígito (1 a 9) e calculadas as frequências observadas. Em seguida, as frequências foram comparadas aos percentuais de distribuição dos primeiros dígitos conforme a Lei de *Newcomb-Benford*.

O resultado das análises do teste do Primeiro Dígito para a base total ($n = 5.574$) deste estudo pode ser observado no gráfico 02 e tabela 03 em anexo.

Considerando um nível de significância de 5%, pode-se rejeitar a hipótese H_0 para a base total de despesas analisada, ou seja, exclui-se a hipótese de os dados terem distribuição conforme a Lei de *Newcomb-Benford*.

Em contrapartida aos dados apresentados, o resultado das análises do teste do Primeiro Dígito para a base de notas fiscais emitidas pelo **Fornecedor 1, (que não efetuou fraude)**, pode ser observado no gráfico 03 e 04 e tabela 04 em anexo. Os valores de despesas da base de notas fiscais do Fornecedor 1, apresentam uma tendência semelhante à distribuição conforme a Lei de *Newcomb-Benford*, ou seja, observa-se maiores ocorrências dos dígitos menores e menores ocorrências nos dígitos maiores.

Os resultados dos testes mostram que os valores de despesas da base do Fornecedor 1, não apresentam sinais de manipulação humana. Essa constatação está em linha com os resultados identificados pela auditoria de conformidade efetuada sobre as notas fiscais desse fornecedor.

Em complemento às análises apresentadas, o resultado do teste do Primeiro Dígito para a base de notas fiscais emitidas pelo **Fornecedor 2, (que efetuou fraude)**, pode ser observado no gráfico 05 e 06 e tabela 05 em anexo.

Os valores de despesas da base de notas fiscais do Fornecedor 2, apresentam tendência diferente da distribuição conforme a Lei de *Newcomb-Benford*, podem ser observados picos de ocorrências dos dígitos maiores em relação aos menores, comprovando a conformidade estatística com a distribuição de *Benford*. Foram aplicados aos resultados os testes estatísticos Teste-Z, Qui-Quadrado e Média dos Desvios Absoluto. A tabela 05 em anexo, apresenta os resultados obtidos.

Para os testes estatísticos *Teste Z* e *Qui-Quadrado* sobre a base de notas fiscais emitidas pelo Fornecedor 2, foi levado em consideração o nível de significância 5%, conforme aplicado por Nigrini & Wells (2012) em seus estudos e limiares de 1,96 (teste Z) e 15,51 (teste Qui-Quadrado). Qualquer resultado acima desses limiares, mostra uma desconformidade com a Lei de *Newcomb-Benford*.

4.2.2. Teste dos Dois Primeiros Dígitos

O teste dos Dois Primeiros Dígitos compara as frequências dos dois primeiros dígitos das sequências observadas com a distribuição segundo a Lei de *Newcomb-Benford*. Este teste mostra-se mais específico que o teste do Primeiro Dígito, já que a amostra se divide em 90 grupos (dígitos de 10 a 99). Valores inferiores a 10 são desconsiderados da análise.

Nigrini & Wells (2012) considera aceitável quatro ou cinco picos significativos no Teste dos Dois Primeiros Dígitos, uma vez que o Teste Z analisa cada dígito individualmente e não sinaliza uma conformidade dos dados de forma conjunta.

Neste estudo, foram contabilizadas as quantidades de itens referentes a cada dois primeiros dígitos (1 a 9) e calculadas as frequências observadas. Em seguida, as frequências foram comparadas aos percentuais de distribuição dos primeiros dígitos conforme a Lei de *Newcomb-Benford*. O resultado das análises do teste dos Dois Primeiros Dígitos para a base total deste estudo pode ser observado no gráfico 07 e tabela 05 em anexo.

Os valores de despesas da base total de fornecedores apresentam tendência diferente da distribuição da Lei de *Newcomb-Benford*, observam-se picos de ocorrências dos dígitos maiores em relação aos menores. Foram aplicados aos resultados os testes estatísticos Teste-Z, Qui-Quadrado e Média dos Desvios Absoluto para validar a ocorrência. A tabela a seguir apresenta os resultados obtidos:

Nesse sentido, considerando um nível de significância de 5%, não pode-se rejeitar a hipótese de que os dados estão em desconformidade com a Lei de Benford, dado o resultado dos testes estatísticos no geral. Em contrapartida aos dados apresentados, o resultado das análises do teste dos Dois Primeiros Dígitos para a base de notas fiscais emitidas pelo Fornecedor 1, pode ser observado no gráfico 08 e **Tabela 06**, em anexo.

Conforme apresentado no Gráfico 7, os valores de despesas da base de notas fiscais do Fornecedor 1, apresentam uma tendência semelhante à distribuição conforme a Lei de *Newcomb-RAGC*, v. 19, ago.; p. 105 – 121 / 2025

Benford, ou seja, podem ser observadas maiores ocorrências dos dígitos menores e menores ocorrências nos dígitos maiores. Foram aplicados aos resultados os testes estatísticos Teste-Z, Qui-Quadrado e Média dos Desvios Absoluto. A **Tabela 06** em anexo apresenta os resultados:

Nesse teste foi utilizada uma significância de 5%, conforme aplicado por Nigrini & Wells (2012) em seus estudos, e limiares de aceitação de 1,96 (para o teste Z) e 112,02 (para o teste Qui-Quadrado). Dessa forma, qualquer resultado acima do limiar, mostra uma desconformidade com a Lei de *Newcomb-Benford*.

Como pode ser observado na **Tabela 06**, todos os primeiros dígitos analisados possuem estatística individual em conformidade com a distribuição da Lei de *Newcomb-Benford*, pois apresentaram valor z baixo do limiar de 1,96. Observando, ainda, os resultados dos testes Qui-Quadrado e Média dos Desvios Absolutos, da mesma forma, ambos se apresentaram em conformidade com a Lei de *Newcomb-Benford*, pois o resultado Qui-Quadrado (108,16) se encontra abaixo do limiar de 112,02 e o resultado do MDA enquadra-se na categoria “Conformidade aproximada”, conforme Tabela 02.

Considerando um nível de significância de 5%, configura como possível rejeitar a hipótese H1 para a base de notas fiscais do Fornecedor 1, ou seja, exclui-se a hipótese de os dados terem distribuição em não conformidade com a Lei de *Newcomb-Benford*. Cabe ressaltar que aceitar a hipótese nula (conformidade com a Lei de *Newcomb-Benford*) não significa que a hipóteses nula seja verdadeira, mas apenas que não existe evidência para rejeitá-la.

Os resultados dos testes mostram que os valores de despesas da base do Fornecedor 1, não apresentam sinais de manipulação humana. Essa constatação está em linha com os resultados identificados pela auditoria de conformidade efetuada sobre as notas fiscais desse fornecedor.

Em complemento às análises apresentadas, o resultado do teste dos Dois Primeiros Dígitos para a base de notas fiscais emitidas pelo Fornecedor 2, pode ser observado no gráfico 09 em anexo.

Os valores de despesas da base de notas fiscais do Fornecedor 2, apresentam uma tendência diferente da distribuição conforme a Lei de *Newcomb-Benford*, ou seja, podem ser observados picos de ocorrências dos dígitos maiores em relação aos menores. De forma a comprovar a conformidade estatística com a distribuição de Benford, foram aplicados aos resultados os testes estatísticos Teste-Z, Qui-Quadrado e Média dos Desvios Absoluto. A **tabela 07** em anexo apresenta os resultados obtidos:

4.2.3. Teste da Soma

Neste trabalho, foram levantados os resultados do Teste dos Dois Primeiros Dígitos e aplicados os testes estatísticos Teste Z, Qui-Quadrado e Média dos Desvios Absolutos para a base total de notas fiscais de fornecedores analisada. Em seguida, o Teste da Soma foi aplicado por meio do cálculo da proporção dos montantes, conforme **tabela 08** em anexo.

Realizados os filtros dos primeiros dígitos com valor z acima do limiar 1,96 (nível de significância de 5%) e com relevância financeira acima da proporção de 0,011 (1/90), tem-se o seguinte resultado para a base total de fornecedores, conforme quadro 01 em anexo.

Conforme pode ser observado, os primeiros dígitos com maiores picos no Teste dos Dois Primeiros Dígitos e que apresentaram maior relevância financeira no Teste da Soma foram 11, 12, 13, 14, 16, 17, 20, 23, 26, 34, 39, 45 e 49. Chama a atenção a proporção identificada para a categoria z, que, em média, representa 4,89 e sugere grande desconformidade com a Lei de *Newcomb-Benford*. Destaca-se, ainda, a representatividade financeira dos montantes que possuem os primeiros dígitos com maiores discrepâncias em relação à Lei de *Newcomb-Benford*, que somam 30,8% do montante total da base de notas fiscais de fornecedores analisada.

5. CONCLUSÃO

A pesquisa consistiu na avaliação, documental e estudo de caso, do método de *Data Mining* Lei de *Newcomb-Benford* para seleção amostral de dados com indícios de fraudes, na condução de avaliações de conformidade nas organizações, sendo plenamente satisfeito.

Quanto aos objetivos específicos, situação problema e sua pergunta, estes foram plenamente satisfeitos, uma vez que o método de *Data Mining* Lei de *Newcomb-Benford* para seleção amostral de dados com indícios de fraudes se apresentou como uma ferramenta efetiva de controle e apoio à realização de auditoria para gestão financeira do negócio, sobretudo no direcionamento de avaliações de conformidade nas organizações.

Os valores das notas fiscais da referida base de dados tiveram suas frequências dos primeiros dígitos comparadas às frequências segundo a Lei de *Newcom-Benford* e o resultado mostrou que a base não apresentava conformidade em relação à distribuição da lei.

Dessa forma, para testar a efetividade do método da Lei de *Newcomb-Benford* no levantamento de dados com indícios de fraude, os testes do Primeiro Dígito, dos Dois Primeiros

Dígitos e da Soma, foram aplicados aos registros de um fornecedor que cometeu fraude e aos registros de um fornecedor que não cometeu fraude, segundo os resultados da auditoria mencionada.

Comparando os resultados, que passaram pelos testes estatísticos de hipótese Teste Z, Qui-Quadrado e Média dos Desvios Absolutos, constatou-se que os indícios de manipulação dos números obtidos por meio da aplicação da Lei de *Benford* reforçam tal posicionamento encontrando no relatório da referida auditoria de conformidade.

O **fornecedor 1**, que não cometeu fraude teve seus respectivos valores de notas fiscais em conformidade com a Lei de *Newcomb-Benford*. E o **fornecedor 2** que cometeu, em desconformidade com a Lei de *Newcomb-Benford*, visto que a ocorrência de desvios entre as frequências esperadas e as frequências dos dados analisados indica que estes podem ter sofrido algum tipo de manipulação.

Este estudo da aplicação da Lei de *Newcomb-Benford* em auditorias de conformidade reforça a importância da aplicação da referida ferramenta de *Data Mining* em auditorias e, como sugestão para pesquisas futuras, pode ser comparada com o método Curva ABC, comumente utilizado em auditorias.

REFERÊNCIAS BIBLIOGRÁFICAS

ALBRECHT, Steve. *Fraud Examination*. Editora Thomson, 2003.

AMARAL, F. **Introdução a ciência de dados: mineração de dados e Big Data**. 1ª ed. Rio de Janeiro: Alta Books, 2016.

ASHFORTH, Blake E.; GIOIA, Dennis A.; ROBINSON, Sandra L.; TREVIÑO, Linda K. **Re-viewing organizational corruption**. *Academy of Management Review*, vol. 33, nº 3, p. 670-684, 2008.

AUBRIL, Ony; LILIAN, Page; REMAUD, Valentine. **Loi de Benford**. *Université d'Angers, Département de Mathématiques*, 2022.

BENFORD, F. **The law of anomalous numbers**. *Proceedings of the American Philosophical Society*, v. 78, nº 4, p. 551-572, 1938.

BENFORD, Frank. **Base dependence of Benford random variables**. *Stats*, v. 4, p. 578–617, 2021. Disponível em: <https://doi.org/10.3390/stats4030034>.

BENFORD, Frank. **Base dependence of Benford random variables**. 2020. Disponível em: <https://arxiv.org/abs/2010.07944>. Acesso em: 07 jun. 2025.

BENFORD, Frank. **Construction of Benford random variables: generators and seed functions**. 2016. Disponível em: <https://arxiv.org/abs/1609.04852>.

BENFORD, Frank. **Fourier analysis and Benford random variables**. *Benford Applied Mathematics*, 2020. Disponível em: <https://www.researchgate.net/publication/342666725>.

BENFORD, Frank. **Fourier analysis and Benford random variables (Preprint revisado)**. *Benford Applied Mathematics*, 2020. Disponível em: <https://arxiv.org/abs/2006.07136>.

BOSSOMAIR, Thomas; SCHNEIDER, Lenny. **Gauss-Benford rules and their harmonic peers**. 2021. Disponível em: <https://arxiv.org/abs/2107.12493>. Acesso em: 07 jun. 2025.

BUGARIN, M.; CUNHA, F.C.R. **Benford's Law for audit of public works: an analysis of overpricing in Maracanã soccer arena's renovation**. *Economics Bulletin*, v. 35, 2015.

CAFE, R. M. **O uso da lei de Benford na auditoria de obras públicas: o caso do VLP**. Trabalho de Conclusão de Curso para bacharelado em Economia, 2015.

CARSLAW, C. A. P. N. **Anomalies in income numbers: evidence of goal oriented behavior**. *The Accounting Review*, v. 2, nº 63, p. 321-327, 1988.

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**. (Dissertação de Mestrado). Universidade Federal de Pernambuco, Recife, 2012.

CRESSEY, D. R. **Other People's Money: A study in the social psychology of embezzlement**. Glencoe, IL: The Free Press, 1953.

CUNHA, P. R.; BEUREN, I. M.; HEIN, N. **Técnicas de amostragem utilizadas nas empresas de Auditoria Independente Estabelecidas em Santa Catarina vs no Rio de Janeiro**. *Revista UnB Contábil*, Brasília, UnB, v. 9, n. 1, 2006.

DAVIC, Robert D. **Newcomb-Benford number law and ecological processes**. *PLOS ONE*, v. 20, n. 3, p. 1–25, 2025. Disponível em: <https://doi.org/10.1371/journal.pone.0310205>.

GIL, Antônio Carlos de Loureiro. **Como evitar fraudes, pirataria e convivência**. 2ª ed. São Paulo: Atlas, 1999.

GOMES, Marcelo Alcides Carvalho. **Uma contribuição à Prevenção de Fraudes Contra as Empresas**. Tese de doutorado – FEA – USP, São Paulo, 2000.

HILL, T. P. **Base-invariance Implies Benford's law**. *Proceedings of the American Mathematical Society*, v. 123, nº 3, p. 887-895, 1995.

HILL, Theodore P.; SCHNEIDER, Lenny. **Fourier analysis and Benford random variables**. 2022. Disponível em: <https://arxiv.org/abs/2207.01722>. Acesso em: 07 jun. 2025.

- KARSIOTIS, Panagiotis; ADAMOPOULOS, Antonios. *Εφαρμογή του νόμου Benford στην ανάλυση δεδομένων απογραφής πληθυσμού*. In: Congresso Internacional de Estatística. Grécia: University of Western Macedonia, 2024.
- KNORR, Edwin M.; NG, Raymond T. *Algorithms for mining distance-based outliers in large datasets*. In: Proceedings of the 2000 International Conference on Very Large Data Bases (VLDB). 2000. p. 392–403.
- KRAUS, Peter; VALVERDE, Carmen. *Accounting fraud and Benford's Law: theoretical reflections and practical implications*. International Journal of Business and Social Science, v. 5, n. 4, p. 21–32, 2014.
- LAGIOIA, U. C. T. et al. *Aplicabilidade da Lei de Newcomb-Benford nas fiscalizações do imposto sobre serviços – ISS*. Revista de Contabilidade & Finanças, v. 22, n. 56, 2011.
- MILLS, A. C. *A auditoria da qualidade: uma ferramenta para avaliação constante e sistemática da manutenção da qualidade*. 5ª ed. São Paulo: Makron Books, 1994.
- MORENO, David; VALLEJOS, Rocío; DI RIENZO, Julio. *Newcomb–Benford number law and ecological processes*. Ecological Modelling, v. 443, p. 109460, 2021. DOI: <10.1016/j.ecolmodel.2021.109460>.
- MURCIA, Fernando Dal-Ri; CARVALHO, Luis Nelson. *Conjecturas acerca do gerenciamento de lucros, republicação das demonstrações contábeis e fraude contábil*. Contabilidade Vista & Revista, v. 18, n. 4, p. 61–82, 2007.
- NEWCOMB, S. *Note on the frequency of the different digits in natural numbers*. The American Journal of Mathematics, v. 4, nº 1, p. 39–40, 1881.
- NIGRINI, M. J. *A taxpayer compliance application of Benford's law*. The American Taxation Association, v. 1, p. 72–91, 1996.
- NIGRINI, M. J. *Digital Analysis Using Benford's Law: Tests? Statistics for Auditors*. Global Audit Publication. Canada, 2000.
- NIGRINI, M. J.; MITTERMAIER, L. J. *The use of Benford's Law as an aid in analytical procedures*. Auditing, v. 16, n. 2, p. 52–68, 1997.
- NIGRINI, M. J.; WELLS, Joseph T. *Benford's law: applications for forensic accounting, auditing, and fraud detection*. Wiley Corporate F&A, 2012.
- PEDROSA, Isabel; COSTA, Carlos J.; LAUREANO, Raul M. S. *Use of information technology on statutory auditors' work: new profiles beyond spreadsheets' users*. Procedia Computer Science, v. 64, p. 1313–1321, 2015. Disponível em: <https://doi.org/10.1016/j.procs.2015.08.587>. Acesso em: 07 jun. 2025.
- PINKHAM, R. S. *On the distribution of first significant digits*. Annals of Mathematical Statistics, v. 32, n. 4, p. 1223–1230, 1961.

RIBEIRO, J. C.; MONTENEGRO, G. B.; SANTOS, J.; GALVÃO, K. S. **Aplicação da Lei de Newcomb-Benford na Auditoria: Caso notas de empenho dos Municípios do Estado da Paraíba.** Anais do Congresso USP de Iniciação Científica em Contabilidade. São Paulo, 2005.

RIVES, Julio. *Gauss-Benford rules and their harmonic peers.* Mathematics Letters, v. 10, n. 3, p. 24–35, 2024. Disponível em: <https://doi.org/10.11648/j.ml.20241003.11>.

SANTOS, J.; RIBEIRO FILHO, J. F.; LAGIOIA, U.; ALVES FILHO, B. F. A.; ARAÚJO, I. J. C. **Aplicações da Lei de Newcomb-Benford na auditoria tributária do Imposto sobre Serviços de Qualquer Natureza (ISS).** Revista de Contabilidade e Finanças, USP, São Paulo, v. 20, n. 49, p. 79–94, 2009.

SANTOS, J.; TENÓRIO, J. N. B.; SILVA, L. G. C. **Uma aplicação da teoria das probabilidades na contabilometria: a lei Newcomb-Benford como medida para análise de dados no campo da auditoria contábil.** Revista Contabilidade, Gestão e Governança, Brasília, v. 6, n. 1, 2003.

TARTER, Timothy. *A generalized Benford framework for threat identification in counter-intelligence.* James Madison University, 2025.

TASSINARI, Francisco. *Loi de Benford et comportement numérique des phénomènes naturels.* Lyon: INSA Lyon, 2010.

TORRES-ZÚÑIGA, Vicente. **Excel, ley de Benford-Newcomb, ChatGPT, censo y COVID-19 en México.** 2024. Documento interno de pesquisa.

TORRES-ZÚÑIGA, Vicente. *Excel ley de Benford Newcomb ChatGTP comparación. In: SOMECE (org.). Argumentos y usos tecnopedagógicos de la inteligencia artificial.* México: UNAM, 2024. p. 370–385.

TRAN, Quoc T.; HUYNH, Tuan V. *Fourier analysis and Benford random variables.* Journal of Mathematical Analysis and Applications, 2021. Disponível em: <https://doi.org/10.1016/j.jmaa.2021.125471>. Acesso em: 07 jun. 2025.

WELLS, J. T. *Principles of Fraud Examination.* 2ª ed. New Jersey: Wiley, 2012.