DETECÇÃO DE FRAUDES: O USO DA LEI DE BENFORD PARA AVALIAR DADOS EDUCACIONAIS E FINANCEIROS

Thiago Duarte Nascimento
Universidade Federal Rural do Rio de Janeiro (UFRRJ)
Av. Governador Roberto da Silveira, s/n
thin504@hotmail.com

Erito Marques de Souza Filho
Universidade Federal Rural do Rio de Janeiro (UFRRJ)
Av. Governador Roberto da Silveira, s/n
eritomarques@yahoo.com.br

Luiz Buscacio
Universidade Federal Fluminense (UFF)
Rua Marques de Paraná, 303
luiz.buscacio@gmail.com

RESUMO

A corrupção no Brasil tem raízes históricas e muitas das vezes é referida como um câncer, uma doença ou uma praga social. Um estudo da Federação das Indústrias do Estado de São Paulo, elaborado em 2010, projetava que entre 1,38% e 2,3% do Produto Interno Bruto se perdiam entre ações corruptas no país. Uma tendência crescente de recursos desviados implica em maior custo de corrupção, uma vez que esses recursos não se transformam em crescimento econômico. Nesse contexto, o presente trabalho apresenta um conjunto de ferramentas quantitativas a serem utilizadas no campo da auditoria contábil como instrumento para detecção de fraudes ou irregularidades a partir de amostras de dados numéricos. Essas ferramentas incluem a Lei de Benford e testes estatísticos de conformidade, os quais serão aplicados a dados obtidos do Instituto Nacional de Pesquisas e Estudos Educacionais e da Secretaria da Fazenda do Estado do Rio de Janeiro.

PALAVRAS CHAVE. Fraudes, Lei de Benford, Auditoria. Área principal: AdP, SE, GF

ABSTRACT

Corruption in Brazil has many historical roots and is often referred to as a cancer, a disease or a social plague. A study by the Federation of Industries of the State of São Paulo, developed in 2010, projected that between 1.38% and 2.3% of GDP were lost between corrupt actions in the country. A growing trend of diverted resources implies higher cost of corruption, since these features are not transformed into economic growth. In this context, this paper presents a set of quantitative tools to be used in the field of accounting audit as a tool for detection of fraud or irregularities from samples of numerical data. These tools include Benford's Law and statistical tests of compliance, which will be applied to data obtained from the National Institute for Research and Educational Studies and the Department of Finance of the State of Rio de Janeiro.

KEYWORDS. Fraud. Benford's Law. Accounting audit. Main area: AdP, SE, GF

1. Introdução

A corrupção no Brasil tem raízes históricas e muitas das vezes é referida como um câncer, uma doença ou uma praga social. Na época em que o país era colônia de Portugal, por exemplo, já existiam funcionários públicos que em vez de fiscalizar contrabando e transgressões contra a coroa Portuguesa praticavam o comércio ilegal de produtos brasileiros, como o pau-brasil. No governo de Juscelino Kubitschek, a construção de Brasília teve como um de seus pontos negativos, a acusação de desvio de recursos e superfaturamento das obras públicas. Mesmo no século XXI, suspeitas de desvios e fraudes nos mais diversos níveis são bastante comuns, com destaque para os recentes casos do que foi chamado de "MENSALÃO do PT" e do "MENSALÃO DO DEMOCRATAS". Situações congêneres, praticamente idênticas, ainda se fazem presentes no cenário nacional sem que se sirva de qualquer instrumento eficiente. Vários países no mundo se utilizam de outras ferramentas diversas aos mecanismos legais tradicionais, as quais por si só não garantem autenticidade. O refinamento dos instrumentos financeiros originou contextos tão amplos e complexos que ultrapassam nossa capacidade de geri-los, abrindo margem para comportamento burlista. O exercício da função pública oferece a inafastável ideia de satisfação do interesse coletivo, sob os auspícios da legalidade dos direitos fundamentais. De fato, em um Estado que se diz Democrático e de Direito, como o nosso, que anuncia a semente popular de seu poder, só pode ser o povo o destinatário de toda atividade pública, uma vez que é ao povo que deve a Administração Pública servir, na busca do bem-estar social. Portanto, os agentes públicos, no exercício do mandato, cargo, emprego ou função pública, atuam em nome do povo, sendo inconcebível que se valham da função pública para satisfação de interesses pessoais, em uma total inversão de valores (Miranda, 2012). Nesse ínterim, seria esperada, em função do alto custo de fraudes corporativas, uma considerável atividade de pesquisa com o intuito de investigar a eficácia dos mecanismos destinados a coibir ou minimizar fraudes. Entretanto, uma revisão da literatura atesta uma falta relevante de mecanismos de prevenção, salvo o papel das auditorias. Essa falta de trabalhos tem como possível explicação o fato de as fraudes serem crimes, que muitas das vezes são desconhecidos dentro de uma organização e que envolvem, em muitos casos, indivíduos com muito poder e prestígio – fato que, inclusive, torna difícil a avaliação do sucesso de mecanismos propostos para combatê-la (DAVIS e PESCH, 2013; NEU *et al*., 2013).

Um estudo da Federação das Indústrias do Estado de São Paulo (FIESP), elaborado em 2010, projetava que entre 1,38% e 2,3% do Produto Interno Bruto (PIB) se perdiam entre ações corruptas no país. Eles concluíram que em 2008 houve uma perda nominal entre 41,5 e 69,1 bilhões de reais - montante esse que poderia ser investido em educação, saúde dentre outros benefícios para a sociedade. O estudo também ressalta que uma maior omissão no controle da corrupção implica em uma tendência crescente de recursos desviados, e, portanto, maior é o custo da corrupção, uma vez que esses recursos não se transformam em crescimento econômico. Por outro lado, a existência de um sentimento de impunidade também contribui para a ocorrência de uma conduta ilícita. O Índice de Percepção da Corrupção em 2013, por exemplo, elaborado pelo grupo Transparência Internacional, coloca o país na posição 72, dentre 177 países avaliados no mundo.

Nesse contexto, o presente trabalho apresenta um conjunto de ferramentas quantitativas a serem utilizadas no campo da auditoria contábil como instrumento para detecção de fraudes ou irregularidades a partir de amostras de dados numéricos. Essas ferramentas incluem a Lei de Benford e testes estatísticos de conformidade, os quais serão aplicados a dados obtidos do Instituto Nacional de Pesquisas e Estudos Educacionais (INEP) e da Secretaria da Fazenda do Estado do Rio de Janeiro (SEFAZ-RJ). Na próxima seção serão apresentados detalhes sobre a Lei de Benford seguida de uma breve revisão sobre o tema. Na quarta seção é descrita a metodologia e, na sequência, são apresentados, respectivamente, os resultados computacionais e discussão e, por fim, as conclusões.

2. Lei de Benford

Em 1881, o astrônomo e matemático Simon Newcomb publicou um artigo no *American Journal of Mathematics* chamado: "Nota sobre a frequência do uso de diferentes dígitos nos números naturais." Ele percebeu que os dez dígitos não ocorriam com igual frequência ao observar constantemente tabelas logarítmicas. Newcomb observou também que as primeiras páginas das tabelas de logaritmos desgastavam-se mais rápido que as últimas, o que o levou a concluir que o dígito "1" aparecia com mais frequência do que qualquer outro e que a ocorrência dos dígitos diminuíam progressivamente do 1 para o 9. Cinquenta e seis anos mais tarde, em abril de 1937, o engenheiro da *General Electric Company*, Frank Albert Benford publicava um artigo com o título: "A lei de anomalia dos números". O engenheiro fez a mesma observação que seu predecessor havia feito no que se referia as páginas das tabelas de logaritmos concluindo também que as primeiras páginas desgastavam-se mais rapidamente que as últimas, devido ao fato de ser mais comum o uso de números começados com o dígito "1" do que números começados com o dígito 9. Benford compilou mais de 20.000 primeiros dígitos tomados de fontes diferentes e concluiu que existe uma distribuição logarítmica dos primeiros dígitos. O engenheiro, entretanto ressaltou que quando os números são compostos de quatro ou mais dígitos a conformidade com a lei é maior. Com bases nas observações dos seus estudos ele postulou, empiricamente, que a frequência dos primeiros dígitos de uma massa aleatória de dados atenderia a lei expressa na Equação 1.

$$F(a) = \log\left(\frac{a+1}{a}\right),\ a \in \{1,2,3,4,5,6,7,8,9\} \qquad (1)$$

A variável *a* representa o a-ésimo primeiro dígito e F(a) representa a frequência esperada do dígito *a*. Isto significa que em uma sequência numérica de dados aleatórios, são esperados aproximadamente que 30% dos primeiros dígitos sejam "1", uma vez que $F(1) = \log\left(\frac{1+1}{1}\right) = \log 3 \approx 0{,}301$.

Hill (1995) estimou que a frequência do segundo dígito dependesse do primeiro dígito e avaliou que esse cálculo podia ser feito por meio da Equação 2.

$$F(b) = \sum_{a=1}^{9} \log\left(1 + \frac{1}{\overline{ab}}\right),\ b \in \{0,1,2,3,4,5,6,7,8,9\} \qquad (2)$$

A variável *b* representa o *b*-ésimo segundo dígito, F(*b*) representa a frequência esperada desse dígito *b* e $\overline{ab}$ indica o número cujo primeiro dígito é *a* e o segundo é *b*. . Isto significa que em uma sequência numérica de dados aleatórios, são esperados aproximadamente que 12% dos segundos dígitos sejam "0", uma vez que $F(0) = \log\left(1+\frac{1}{10}\right) + \log\left(1+\frac{1}{20}\right) + \log\left(1+\frac{1}{30}\right) + \ldots + \log\left(1+\frac{1}{90}\right) \approx 0{,}12$. Hill (1995) também generalizou o resultado anterior para a frequência do *n*-ésimo dígito, entretanto no âmbito desse trabalho apenas serão utilizadas as frequências do primeiro e segundo dígitos, as quais são exibidas na Tabela 1.

Tabela 1: Frequências esperadas dos dígitos na primeira e segunda posição

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)

3. Revisão bibliográfica

Pinkham (1961) mostrou que a Lei de Benford possui a propriedade de invariância escalar. Ele demonstrou ainda que os dígitos que ocorrem com probabilidade logarítmica são os únicos que mantém sua probabilidade de ocorrência quando multiplicados por um fator de escala. Hamming (1970) examinou as mantissas de pontos flutuantes e mostrou como as operações de um computador levam várias distribuições em direção ao limite de uma distribuição logarítmica. Além disso, ele propôs aplicações em *hardwares* e *softwares* que elucidavam a otimização do custo computacional. Hill (1988) mostrou evidências experimentais de que quando pessoas inventam números aleatórios, estes números nunca estão em conformidade com a Lei de Benford, ainda que compartilhem algumas propriedades da mesma. Hill (1995) desenvolveu um aparato teórico matemático mais sólido para a Lei de Benford ancorado no conceito de α-álgebra de mantissas, as quais foram usadas para provar a invariância por escala e a unicidade da distribuição.

Nigrini e Mittermaier (1997) iniciaram os estudos da aplicação da Lei de Benford como um modelo de procedimentos analíticos para auditores capaz de norteá-los na busca de possíveis manipulações fraudulentas. A ideia matriz é a análise dos desvios encontrados entre frequências observadas e esperadas. Nigrini (1999) analisou um caso de fraude ocorrido no estado do Arizona em 1993, na qual um gerente de uma empresa desviava fundos para uso próprio. Nigrini (2000) desenvolveu um modelo chamado contabilométrico, que sugere a utilização de cinco principais testes estatísticos para avaliar a frequência relativa do primeiro dígito, do segundo dígito, dos dois primeiros dígitos, dos primeiros três dígitos e dos últimos dois dígitos. Dentre esses modelos, pode-se destacar dois testes de hipóteses: Z-teste e o teste qui-quadrado.

Santos *et. al* (2003) desenvolveram uma adaptação do modelo contabilométrico, a partir do modelo desenvolvido por Nigrini (2000), que se fundamenta apenas na relação entre a Lei de Benford e os testes de hipótese: Z-teste e qui-quadrado. No âmbito do trabalho aqui desenvolvido foi feito uso do modelo de Santos *et al.* (2003) em duas situações dinstintas – sendo a primeira na análise dos dados do INEP e a segunda na análise dos dados da SEFAZ-RJ. Ribeiro *et al.* (2005) aplicaram o modelo contabilométrico criado por Nigrini (2000) e aperfeiçoado por Santos *et a.l* (2003) em um estudo de caso referente a notas de empenho de 20 municípios do estado da Paraíba. Diniz *et al*. (2006) afirmam que a aplicação de métodos quantitativos na forma de modelos contabilométricos tem se acentuado nesses últimos anos. Essa metodologia apresentou-se poderosa no planejamento do trabalho de campo de auditoria nas contas públicas. Dessa forma eles aplicaram o modelo contabilométrico de Santos *et al*. (2003) com o intuito de comprovar a eficácia do modelo na auditoria digital. Os resultados revelaram que há uma significativa correlação entre a quantidade de irregularidades apontadas nas auditorias e os valores apresentados pelo modelo contabilométrico construído a partir da Lei de Benford.

Madureira (2012), em sua dissertação de mestrado, analisou a aplicabilidade de uma metodologia de investigação de problemas existentes no conjunto de dados eleitorais de uma eleição para o cargo majoritário de Presidente da República. Ele constatou em sua pesquisa a possibilidade de se reduzir o risco de fraude eleitoral através da utilização da análise digital e avaliou também que o teste do 2º dígito seria interessante para a detecção de manipulação ou fraude no conjunto de dados eleitorais.

4. Metodologia

Inicialmente foram extraídos do sítio do INEP dados relativos ao número total de alunos matriculados nas escolas das redes municipais e estaduais, das áreas urbanas e rurais, em cada município brasileiro. Esses dados integram o censo escolar 2012. O arquivo utilizado consta de 30000 números divididos nas categorias anos iniciais do ensino fundamental, anos finais do ensino fundamental e ensino médio, que funcionam em regime parcial, isto é, em apenas um único turno. Além disso, foram analisados os dados referentes a pagamentos efetuados pela administração direta e indireta do estado do Rio de Janeiro nos anos de 2008, 2009, 2010, 2011 e 2012, que foram extraídos do sítio da SEFAZ. Esses dados constituem-se dos valores das notas de empenho pagas pela administração direta e indireta do estado nos referidos anos - foram extraídos 20 arquivos contendo ao todo um total de 200.545 registros. Em seguida, fez-se uso do modelo contabilométrico proposto por Nigrini (2000) e modificado por Santos *et al.* (2003). Dessa forma, foram calculadas as frequências de cada primeiro e segundo dígito (para todos os dígitos) e, na sequência, foram utilizados o Z-teste e o teste qui-quadrado para avaliar a conformidade dos dados à Lei de Benford. A ferramenta utilizada para avaliar a frequência de cada dígito foi o Microsoft Excel 2007. Para extrair o primeiro ou o segundo dígito de cada número da sequência de dados foi utilizado a função “EXT.TEXTO” e para fazer a contagem de ocorrências utilizou-se da função estatística “FREQUENCIA”. Dessa forma foi possível contar a quantidade de ocorrências dos dígitos 1,2,3,4,5,6,7,8 e 9 na primeira posição e dos dígitos 0,1,2,3,4,5,6,7,8 e 9 na segunda posição.

De acordo com Santos *et al.* (2003), dois testes estatísticos são interessantes para avaliar a conformidade de um conjunto de dados a Lei de Benford. O teste-z avalia sob um determinado nível de significância as diferenças entre a probabilidade observada (*po*) e a probabilidade esperada (*pe*) no tocante ao ocorrência dos dígitos numéricos, conforme se visualiza na Equação 3.

$$Z = \frac{|po - pe| - \frac{1}{2n}}{\sqrt{\frac{pe(1-pe)}{n}}} \tag{3}$$

Nessa estatística de teste “*n*” é o número de observações, o valor $\frac{1}{2n}$ é o termo de correção de continuidade e só é utilizado quando for menor que $|po - pe|$. Nesse teste, para um nível de significância de 5% (α=0,05), o valor crítico Z_c é igual a 1,96. Sob o ponto de vista prático, se um determinado dígito possui valor superior ao crítico, sugere-se uma discrepância em relação aos valores esperados caso os dados seguissem a distribuição de Benford. De maneira similar, outro teste importante e com função similar é o teste qui-quadrado, que é definido na Equação 4.

$$X^2 = \sum_{d=1}^{9} \frac{(PO - PE)^2}{PE} \tag{4}$$

Os valores de PO e PE são as proporções observadas e esperadas multiplicadas pela população de todos os números avaliados, conforme se verifica nas Equações 5 e 6.

$$PO = po * (população) \tag{5}$$

$$PO = po * (população) \tag{6}$$

Para um nível de significância de 5% (α=0,05) e para um grau de liberdade igual a 8, o valor de corte do teste qui-quadrado é igual a "15,507". Novamente, sob o ponto de vista prático, se um determinado dígito possui valor superior ao crítico, sugere-se uma discrepância em relação aos valores esperados caso os dados seguissem a distribuição de Benford. Cabe ressaltar que no caso da aplicação desse teste para o segundo dígito o grau de liberdade muda de 8 para 9, já que no estudo dos segundos dígitos, o dígito "0" (zero) passa a ser incluído. O grau de liberdade igual a "9"(nove) implica em um novo valor de corte igual a "16,92".

5. Resultados obtidos e Discussão

A Tabela 1 exibe os resultados obtidos após a análise dos dados extraídos do INEP. A partir dela, pode-se perceber que, em relação ao total de alunos matriculados no ensino fundamental nos anos iniciais e ensino médio, houve grande proximidade entre as frequências obtidas, tanto para primeiro quanto para o segundo dígito, e as frequências esperadas pela Lei de Benford. O teste-z, conforme ressaltado anteriormente, tem como valor crítico 1,96. Os resultados obtidos para esses dois conjuntos de dados indicam para cada dígito valores inferiores ao valor crítico. Similarmente, o teste qui-quadrado tem como valores críticos 15,95 e 16,92 dependendo se a análise se refere ao primeiro ou ao segundo dígito. Nesse caso, todos os valores obtidos também foram inferiores a esses valores críticos. Dessa forma, os resultados obtidos indicam a existência de conformidade dos dados com a distribuição de Benford. Retomando o trabalho de Hill(1988), destaca-se empiricamente que quando se inventam números aleatórios a conformidade com a Lei de Benford não é alcançada. Isso significa, sob o ponto de vista contábil, que esses valores tem menor probabilidade de terem origem fraudulenta. A análise do número de alunos matriculados no ensino fundamental na categoria anos finais apresentou conformidade com a Lei de Benford, exceto para o dígito "2", que violou o teste-z, mas apresentou conformidade segundo o teste qui-quadrado.

A Tabela 2 exibe os resultados relativos à análise dos valores das notas de empenho disponibilizadas no sítio da Secretaria de Fazenda. A partir dela, percebe-se que valor crítico do teste qui-quadrado foi violado em todos os anos analisados (de 2008 a 2012). O valor crítico para o teste-z foi ultrapassado diversas vezes tanto para o primeiro quanto para o segundo dígito – as violações são destacadas em amarelo. Quando um conjunto de dados relativamente grande de dados, não apresenta conformidade com a lei, é importante que se procure uma melhor compreensão sobre o fato – note-se que, os dados avaliados são de natureza aleatória e apresentam quatro ou mais dígitos. Esse maior entendimento pode ser obtido quando se conjuga os resultados obtidos com o que é expresso na Lei federal n.º 8.666/93. Da redação dos artigos 23 e 24 se depreende que quando os gastos com compras e serviços que forem maiores do que R$ 8000, faz-se necessário de realização de licitação. Dessa forma, retomando novamente a Tabela 2, percebe-se que o dígito "8" não apresentou conformidade com a distribuição de Benford em todos os anos analisados. Nesse contexto, pode-se sugerir que essa discrepância possa ser fruto de atividades ilícitas: para não se sujeitar ao processo licitatório, em vez de se emitir uma nota como valor de 8000 reais, faz-se a opção por emitir duas notas com valores menores, o que impacta também em não conformidade em outros dígitos.

É importante salientar que, o trabalho aqui desenvolvido não permite afirmar categoricamente que um indivíduo ou empresa possuem atividades fraudulentas. Na verdade, o principal objetivo é apresentar subsídios para que a equipe de auditoria possa maximizar sua chance de detecção desse tipo de corrupção. Em um ambiente de escassez de recursos, a equipe de auditoria geralmente tem uma quantidade grande informações a serem analisadas e, muitas das vezes, não possui tempo e/ou recursos para fazê-lo no curto prazo. Dessa forma, o modelo contabilométrico se mostra de grande valia no tocante a determinação de quais conjuntos de dados deveriam ser auditados em função da sua maior ou menor conformidade com a Lei de Benford. Um ideia bem simples é aplicar o modelo contabilométrico e, na sequência, fazer uso do clássico Problema da Mochila (PM), definido nesse contexto, pelos seguintes parâmetros e variáveis de decisão binárias:

x_i : variável binária que indica se ocorrerá uma auditoria no local *i*

r_i : parâmetro que indica quantidade de recurso utilizado para auditar o local *i*

s_i : parâmetro que indica a probabilidade de fraude associada ao local *i*

n : número total de locais possíveis de serem auditados

R : quantidade total de recursos

O valor de s_i a ser utilizado pode ser obtido após a aplicação do modelo contabilométrico utilizando-se a Equação 7. Dado um conjunto Ψ de dígitos encontrados no local *i* e que após aplicação do teste-z, não estavam em conformidade com a lei, deve-se fazer o somatório da diferenças (em módulo) entre as probabilidades observadas e esperadas na análise dos dados numéricos.

$$s_i = \sum_{d \in \Psi} | po - pe | \qquad (7)$$

Após a obtenção dos parâmetros de auditoria s_i por meio da aplicação do modelo contabilométrico e supondo conhecidos a priori os parâmetros r_i, o Problema da Mochila a ser utilizado é descrito nas Equações 8, 9 e 10.

$$Max \sum_{i=1}^{n} s_i x_i \qquad (8)$$

Tal que: $\sum_{i=1}^{n} r_i x_i \leq R$ (9)

$$x_i \in \{0,1\} \qquad (10)$$

A título de exemplo, pode-se utilizar os dados da SEFAZ para avaliar os documentos de quais anos deveriam auditados, supondo-se que se gaste uma quantidade de recurso em cada ano de 2008 a 2012, dada por, respectivamente, 1, 3, 4, 6, 7 e 9 (unidades arbitrárias), e que apenas 13 unidades de recurso estejam disponíveis. Os parâmetros de probabilidade podem ser obtidos a partir da da Tabela 2 (destacados em verde). Para obtenção desses coeficientes, ao longo de cada ano, foram somadas, para os dígitos que não estavam em conformidade com a Lei de Benford, as diferenças em módulo entre as probabilidades observadas e as probabilidades esperadas. Nesse exemplo, os valores dos parâmetros relativos a probabilidade de fraude para cada ano, são respectivamente, 0,222805, 0,142338, 0,126744, 0,098887 e 0,073393. Esse exemplo foi implementado em linguagem Mosel e resolvido no XPRESS-MP. Os resultados obtidos indicaram que a equipe de auditoria, considerando seus recursos disponíveis, deveria auditar os dados da SEFAZ referentes aos anos de 2008, 2009 e 2010, de modo que seriam usadas 8 unidades de recurso dentre as 13 disponíveis.

5. Conclusões

"*A corrupção de homens que eventualmente chegaram ao poder traz consigo uma doença, que a torna contagiosa para a multidão*" – A frase de Alexis de Tocqueville em seu livro "A democracia na América" parece se adequar perfeitamente aos diversos casos de irregularidades no setor público e privado brasileiro. Entretanto, o crescimento das fraudes não se refletiram em métodos e mecanismos eficientes para coibir suas práticas. Nesse contexto, o presente trabalho apresentou uma aplicação da Lei de Benford e sua utilização como uma possível ferramenta para orientar a busca por disparidades em conjuntos de dados numéricos. Essa lei juntamente com teste-z e o teste qui-quadrado compõem o chamado modelo contabilométrico aqui utilizado. Os dados referentes ao censo escolar, fornecidos pelo INEP mostraram consistência em relação aos métodos aplicados. Por outro lado, a análise dos dados da SEFAZ mostrou várias violações. Em particular, foi registrado uma possível explicação para o fato de diversas violações relativas a frequência esperada de dígito "8": a existência de irregularidades em relação ao cumprimento do preconizado nos artigos 22 e 23 da Lei Federal n.º 8.666/93. Esses resultados podem servir de subsídio para detecção de fraudes, apesar de não representar uma garantia de suas existências. Além disso, foi evidenciada a possibilidade de utilização dos resultados do modelo contabilométrico como parâmetro relacionado à probabilidade de fraude em um Problema da Mochila.

Ademais, com o constante crescimento dos gastos públicos, a tarefa de auditoria é cada vez mais árdua e o descaminho mais oportuno. Nesse contexto, sugere-se com trabalhos futuros a utilização das ferramentas aqui desenvolvidas em outros conjuntos de dados visando avaliar o desempenho de outros órgãos públicos. Além disso, o uso dessas ferramentas para avaliar a lisura de eleições ainda pode ser considerado incipiente. Sob a ótica do setor privado, além da auditoria de empresas propriamente dita, aplicações interessantes da Lei de Benford incluem seu uso para compor *portfólios* de ativos (ações e opções, por exemplo), as quais podem ser baseadas no Problema da Mochila aqui desenvolvido. Faz-se necessário acrescentar ainda, que empiricamente o bom funcionamento da lei está atrelado a dados aleatórios e com mais de quatro dígitos. Isso implica na existência de uma lacuna sob o ponto de vista teórico, relacionado o uso da Lei de Benford e de suas possíveis derivações sobre conjuntos de dados de outra natureza.

Tabela 1: Resultados após análise dos dados do INEP para as categorias anos iniciais(*), anos finais (**) e ensino médio (***)

1º Dig.*	Freq	po	pe	|po-pe|	Z	χ^2	2º Dig.*	Freq	po	pe	|po-pe|	Z	χ^2
0	--	--	--	--	--	--	**0**	1521	0,11522	0,12000	0,004780	1,677	2,514
1	4019	0,30141	0,30100	0,000410	0,094	0,007	**1**	1550	0,11742	0,11400	0,003420	1,223	1,354
2	2362	0,17714	0,17600	0,001140	0,334	0,098	**2**	1423	0,10779	0,10900	0,001210	0,432	0,177
3	1639	0,12292	0,12500	0,002080	0,713	0,462	**3**	1401	0,10613	0,10400	0,002130	0,787	0,576
4	1251	0,09382	0,09700	0,003180	1,226	1,390	**4**	1327	0,10052	0,10000	0,000520	0,185	0,036
5	1086	0,08145	0,07900	0,002450	1,033	1,013	**5**	1304	0,09878	0,09700	0,001780	0,676	0,431
6	921	0,06907	0,06700	0,002070	0,939	0,853	**6**	1281	0,09704	0,09300	0,004040	1,583	2,317
7	776	0,05820	0,05800	0,000200	0,081	0,009	**7**	1200	0,09090	0,09000	0,000900	0,346	0,119
8	663	0,04972	0,05100	0,001280	0,652	0,428	**8**	1098	0,08318	0,08800	0,004820	1,939	3,485
9	617	0,04627	0,04600	0,000270	0,128	0,021	**9**	1096	0,08302	0,08500	0,001980	0,800	0,609
$\sum$	13334	1,000	1,000	---	---	4,281	$\sum$	13201	1,000	1,000	---	---	9,104
1º Dig.**	**Freq**	**po**	**pe**	**|po-pe|**	**Z**	χ^2	**2º Dig.****	**Freq**	**po**	**pe**	**|po-pe|**	**Z**	χ^2
0	--	--	--	--	--	--	**0**	1537	0,11549	0,12000	0,004510	1,588	2,256
1	4046	0,30097	0,30100	0,000030	0,008	0,000	**1**	1566	0,11766	0,11400	0,003660	1,315	1,564
2	2388	0,17764	0,17600	0,001640	0,488	0,205	**2**	1437	0,10797	0,10900	0,001030	0,367	0,130
3	1654	0,12304	0,12500	0,001960	0,674	0,413	**3**	1411	0,10602	0,10400	0,002020	0,749	0,522
4	1259	0,09365	0,09700	0,003350	1,298	1,555	**4**	1337	0,10046	0,10000	0,000460	0,162	0,028
5	1097	0,08160	0,07900	0,002600	1,102	1,150	**5**	1312	0,09858	0,09700	0,001580	0,601	0,343
6	932	0,06933	0,06700	0,002330	1,063	1,089	**6**	1287	0,09670	0,09300	0,003700	1,455	1,959
7	779	0,05795	0,05800	0,000050	0,006	0,001	**7**	1210	0,09092	0,09000	0,000920	0,356	0,125
8	669	0,04977	0,05100	0,001230	0,629	0,399	**8**	1108	0,08325	0,08800	0,004750	1,919	3,412
9	619	0,04605	0,04600	0,000050	0,007	0,001	**9**	1104	0,08295	0,08500	0,002050	0,832	0,658
$\sum$	13443	1,000	1,000	---	---	4,813	$\sum$	13309	1,000	1,000	---	---	8,741
1º Dig.***	**Freq**	**po**	**pe**	**|po-pe|**	**Z**	χ^2	**2º Dig.*****	**Freq**	**po**	**pe**	**|po-pe|**	**Z**	χ^2
0	--	--	--	--	--	--	**0**	798	0,11431	0,12000	0,005690	1,444	1,883
1	2075	0,29664	0,30100	0,004360	0,782	0,442	**1**	769	0,11016	0,11400	0,003840	0,991	0,903
2	1316	0,18813	0,17600	0,012130	2,648	5,848	**2**	750	0,10743	0,10900	0,001570	0,402	0,158
3	893	0,12766	0,12500	0,002660	0,655	0,396	**3**	765	0,10958	0,10400	0,005580	1,508	2,090
4	638	0,09121	0,09700	0,005790	1,616	2,418	**4**	699	0,10013	0,10000	0,000130	0,016	0,001
5	532	0,07605	0,07900	0,002950	0,893	0,771	**5**	642	0,09196	0,09700	0,005040	1,403	1,828
6	432	0,06176	0,06700	0,005240	1,729	2,867	**6**	651	0,09325	0,09300	0,000250	0,051	0,005
7	419	0,05990	0,05800	0,001900	0,654	0,435	**7**	668	0,09569	0,09000	0,005690	1,640	2,511
8	354	0,05061	0,05100	0,000390	0,121	0,021	**8**	637	0,09125	0,08800	0,003250	0,937	0,838
9	336	0,04803	0,04600	0,002030	0,782	0,627	**9**	602	0,08623	0,08500	0,001230	0,347	0,124
$\sum$	6995	1,000	1,000	---	---	13,825	$\sum$	6981	1,000	1,000	---	---	8,458

Tabela 2: Resultados obtidos para os dados da SEFAZ em 2008(*), 2009(**) e 2010(***)

1º Dig.*	Freq	po	pe	|po-pe|	Z	χ^2	2º Dig.*	Freq	po	pe	|po-pe|	Z	χ^2
0	--	--	--	--	--	--	**0**	7539	0,18487	0,12000	0,064870	40,305	1430,060
1	11058	0,27116	0,30100	0,029838	13,131	120,618	**1**	3967	0,09728	0,11400	0,016722	10,617	100,026
2	7177	0,17599	0,17600	0,000007	0,004	0,000	**2**	4442	0,10893	0,10900	0,000074	0,040	0,002
3	4682	0,11481	0,12500	0,010189	6,214	33,868	**3**	3825	0,09380	0,10400	0,010204	6,742	40,828
4	3689	0,09046	0,09700	0,006539	4,453	17,976	**4**	3732	0,09152	0,10000	0,008485	5,703	29,357
5	3641	0,08928	0,07900	0,010284	7,690	54,594	**5**	3799	0,09316	0,09700	0,003842	2,613	6,204
6	3064	0,07513	0,06700	0,008135	6,561	40,279	**6**	3508	0,08602	0,09300	0,006977	4,843	21,348
7	2938	0,07205	0,05800	0,014045	12,124	138,698	**7**	3292	0,08073	0,09000	0,009274	6,536	38,972
8	2573	0,06309	0,05100	0,012095	11,091	116,967	**8**	3344	0,08200	0,08800	0,005999	4,268	16,677
9	1958	0,04801	0,04600	0,002014	1,929	3,595	**9**	3332	0,08171	0,08500	0,003293	2,376	5,203
$\sum$	40780	1,000	1,000		---	526,593	$\sum$	40780	1,000	1,000		---	258,618

1º Dig.**	Freq	po	pe	|po-pe|	Z	χ^2	2º Dig.**	Freq	po	pe	|po-pe|	Z	χ^2
0	--	--	--	--	--	--	**0**	4727	0,16863	0,12000	0,048635	25,048	552,523
1	8487	0,30277	0,30100	0,001772	0,640	0,292	**1**	2954	0,10538	0,11400	0,008617	4,530	18,256
2	4860	0,17338	0,17600	0,002621	1,144	1,094	**2**	2748	0,09803	0,10900	0,010966	5,882	30,923
3	3412	0,12172	0,12500	0,003278	1,650	2,409	**3**	2459	0,08772	0,10400	0,016276	8,917	71,398
4	2427	0,08658	0,09700	0,010417	5,883	31,360	**4**	2869	0,10235	0,10000	0,002351	1,302	1,549
5	2286	0,08155	0,07900	0,002553	1,573	2,312	**5**	2855	0,10185	0,09700	0,004852	2,734	6,802
6	1955	0,06974	0,06700	0,002744	1,826	3,151	**6**	2314	0,08255	0,09300	0,010449	6,013	32,905
7	1758	0,06272	0,05800	0,004716	3,365	10,750	**7**	2164	0,07720	0,09000	0,012800	7,478	51,027
8	1522	0,05430	0,05100	0,003297	2,496	5,975	**8**	2354	0,08398	0,08800	0,004022	2,366	5,152
9	1324	0,04723	0,04600	0,001233	0,972	0,927	**9**	2587	0,09229	0,08500	0,007291	4,366	17,529
$\sum$	28031	1,000	1,000	---	---	58,269	$\sum$	28031	1,000	1,000	---	---	235,541

1º Dig.***	Freq	po	pe	|po-pe|	Z	χ^2	2º Dig.***	Freq	po	pe	|po-pe|	Z	χ^2
0	--	--	--	--	--	--	**0**	7259	0,15769	0,12000	0,037695	24,880	545,052
1	13737	0,29842	0,30100	0,002577	1,200	1,016	**1**	4741	0,10299	0,11400	0,011006	7,423	48,916
2	8091	0,17577	0,17600	0,000231	0,124	0,014	**2**	4709	0,10230	0,10900	0,006702	4,606	18,967
3	5568	0,12096	0,12500	0,004041	2,614	6,013	**3**	4330	0,09406	0,10400	0,009935	6,975	43,688
4	4054	0,08807	0,09700	0,008931	6,466	37,851	**4**	4462	0,09693	0,10000	0,003067	2,186	4,331
5	3715	0,08070	0,07900	0,001705	1,347	1,693	**5**	4421	0,09604	0,09700	0,000958	0,687	0,436
6	3238	0,07034	0,06700	0,003342	2,859	7,675	**6**	4039	0,08774	0,09300	0,005257	3,875	13,677
7	2990	0,06495	0,05800	0,006955	6,374	38,389	**7**	3643	0,07914	0,09000	0,010859	8,133	60,316
8	2486	0,05401	0,05100	0,003006	2,921	8,155	**8**	3916	0,08507	0,08800	0,002929	2,210	4,487
9	2153	0,04677	0,04600	0,000772	0,779	0,596	**9**	4512	0,09802	0,08500	0,013019	10,007	91,787
$\sum$	46032	1,000	1,000	---	---	101,401	$\sum$	46032	1,000	1,000	---	---	286,604

Tabela 3: Resultados obtidos para os dados da SEFAZ em 2011(*) e 2012(**)

1º Dig.*	Freq	po	pe	|po-pe|	Z	χ^2	2º Dig.*	Freq	po	pe	|po-pe|	Z	χ^2
0	--	--	--	--	--	--	0	6694	0,15271	0,12000	0,032713	21,069	390,893
1	12805	0,29212	0,30100	0,008875	4,046	11,471	1	4422	0,10088	0,11400	0,013119	8,635	66,181
2	7760	0,17703	0,17600	0,001032	0,561	0,265	2	4619	0,10537	0,10900	0,003625	2,428	5,285
3	5675	0,12947	0,12500	0,004466	2,820	6,993	3	4501	0,10268	0,10400	0,001317	0,896	0,731
4	4141	0,09447	0,09700	0,002530	1,782	2,892	4	4039	0,09214	0,10000	0,007857	5,475	27,059
5	3672	0,08377	0,07900	0,004771	3,694	12,628	5	4509	0,10287	0,09700	0,005865	4,141	15,546
6	2972	0,06780	0,06700	0,000801	0,661	0,420	6	3965	0,09045	0,09300	0,002545	1,826	3,053
7	2567	0,05856	0,05800	0,000562	0,493	0,239	7	3683	0,08402	0,09000	0,005978	4,365	17,408
8	2380	0,05430	0,05100	0,003296	3,126	9,336	8	3647	0,08320	0,08800	0,004800	3,539	11,475
9	1862	0,04248	0,04600	0,003522	3,508	11,817	9	3755	0,08566	0,08500	0,000664	0,490	0,227
$\sum$	43834	1,000	1,000	---	---	56,061	$\sum$	43834	1,000	1,000	---	---	146,967
1º Dig.**	**Freq**	**po**	**pe**	**|po-pe|**	**Z**	χ^2	**2º Dig.****	**Freq**	**po**	**pe**	**|po-pe|**	**Z**	χ^2
0	--	--	--	--	--	--	0	6014	0,14364	0,12000	0,023642	14,879	195,014
1	12763	0,30484	0,30100	0,003839	1,707	2,050	1	4330	0,10342	0,11400	0,010580	6,804	41,108
2	7282	0,17393	0,17600	0,002072	1,107	1,022	2	4639	0,11080	0,10900	0,001801	1,174	1,245
3	5518	0,13180	0,12500	0,006795	4,197	15,466	3	4316	0,10309	0,10400	0,000914	0,605	0,336
4	4096	0,09783	0,09700	0,000831	0,566	0,298	4	4167	0,09953	0,10000	0,000473	0,314	0,094
5	3258	0,07782	0,07900	0,001184	0,889	0,743	5	4283	0,10230	0,09700	0,005298	3,654	12,114
6	2700	0,06449	0,06700	0,002512	2,046	3,942	6	3820	0,09124	0,09300	0,001761	1,232	1,396
7	2506	0,05985	0,05800	0,001855	1,613	2,483	7	3464	0,08274	0,09000	0,007264	5,185	24,545
8	1947	0,04650	0,05100	0,004497	4,171	16,600	8	3413	0,08152	0,08800	0,006482	4,673	19,990
9	1798	0,04294	0,04600	0,003056	2,973	8,498	9	3422	0,08173	0,08500	0,003267	2,388	5,257
$\sum$	41868	1,000	1,000	---	---	51,101	$\sum$	41868	1,000	1,000	---	---	106,085

Referências

Benford, F. (1938), The law of anomalous numbers. Proceedings of the American Philosophical Society, vol. 78, p. 551-572.

Brasil, (1993) Lei Federal n.º 8.666/93.

Davis, J.S., Pesch, H. L. (2013), Fraud dynamics and controls in organizations. Accounting, Organizations and Society, nº 38, pp. 469-483.

Diniz, J. A., Santos, J., Dieng, M., Diniz, M. A. A., (2006), 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. Fundação Instituto de Pesquisas Contábeis (FIPECAF), Universidade de São Paulo, disponível em: http://www.congressousp.fipecafi.org/web/artigos62006/261.pdf

Fiesp. (2010), Corrupção: custos econômicos e propostas de combate. Relatório disponível em: http://www.fiesp.com.br/indices-pesquisas-e-publicacoes/relatorio-corrupcao-custos-economicos-e-propostas-de-combate/

Hamming, R.W. (1970), One the distribution of numbers. The Bell System - Technical Journal, vol. 49, n.º 8, pp. 1609-1625.

Hill, T. (1988), Random-number guessing and the first digit phenomenon. Psychological Reports, Vol. 62, pp. 967-971.

Hill, T. (1995a), The Significant-Digit Phenomenon. American Mathematical Monthly, vol. 102, nº 4, pp. 322-327.

Hill, T. (1995b), Base-Invariance Implies Benford's Law. Proceedings of the American Mathematical Society, Vol. 123, n.º 3, pp. 887-895.

Hill, T. (1995), A statistical derivation of the significant-digit law." Statistical Science, vol. 10, nº. 4, pp. 354-363.

Madureira, N. L. (2012), Aplicabilidade da Lei de Benford na Análise de um conjunto de dados eleitorais. Dissertação de Mestrado, Universidade do Estado do Rio de Janeiro.

Neu, D., Everett, J., Rahaman, A. S., Martinez, D. (2013), Fraud dynamics and controls in organizations. Accounting, Organizations and Society, nº 38, pp. 505-524.

Newcomb, S. (1881), Note on the frequency of use of the different digits in natural numbers. American Journal of Mathematics, vol. 4, p.39-40, nº 1.

Nigrini, M. J. (2000). Digital Analysis Using Benford's Law – Tests statistics for auditors, Canadá, Global Audit Publication.

Nigrini, M., Mittermaier, L. J. (1997), The use of Benford's Law as an aid in analitycal procedures. Auditing: a journal of practice and theory, vol. 16, n.º 2.

Nigrini, M. J. (1999), I've got your number. Journal of Accountancy, 187, p. 79-83.

Miranda, G. S. (2012) *Corrupção Pública*: Uma Pandemia Nacional. In: Associação Nacional dos Membros do Ministério Público. Disponível em: http://www.conamp.org.br

Pikham, R.S. (1961), On the distribution of first significant digits. The annals of mathematical statistics, vol.32, nº 4, pp. 1223-1230.

Portal da Transparencia Internacional, (2013). Índice de percepção da corrupção, disponível em: http://issuu.com/transparencyinternational

Ribeiro, J.C., Monteiro, G.B., Santos, J., Galvão, K.S. (2005), Aplicação da lei de Newcomb-Benford na auditoria - Caso notas de empenho dos municípios do estado da Paraíba. In. Congresso USP contabilidade e controladoria, 5, São Paulo.

Santos, J., Tenório, J. N. B., Silva, L. G. C. (2003), "Uma aplicação da teoria das probabilidades na contabilometria – A lei de Newcomb-Benford como medida para análise de dados no campo da auditoria contábil. UNB Contábil, vol. 6, nº 1.

Tocqueville, A. (2005), A democracia na América. 2ª Edição. Tradução. São Paulo, Ed. Martins Fontes.