

USO DA LEI DE BENFORD PARA IDENTIFICAR INSERÇÃO DE DADOS FICTÍCIOS EM REQUERIMENTOS DE SEGURO-DESEMPREGO

Leandro Spier¹

1 Introdução. 2 Fraudes ao seguro-desemprego. 2.1 Vínculo fictício. 3 Lei de Benford. 4 Banco de dados digitais. 5 Busca de anormalidades. 6 Análise dos dados de seguro-desemprego. 7 Considerações finais. Referências.

RESUMO

Evidencia-se que a lei de Benford pode ser utilizada na identificação de casos de inserção de dados fictícios no sistema do seguro-desemprego. Dados fictícios ou vínculos fictícios são informações inventadas e que não têm uma correlação com a realidade, informações inseridas para conseguir pagamentos indevidos. A lei de Benford é uma curiosa constatação empírica de que os primeiros dígitos significativos de alguns conjuntos de números seguem uma distribuição não equitativa e sim uma curva descendente na qual 30,1% dos números começam por 1 e somente 4,58% começam com o dígito 9. A partir do número de endereço residencial do requerente de seguro-desemprego, o trabalho descreve uma metodologia de tratamento dos dados, da separação de subconjuntos e identificação de situações inventadas. Em uma conclusão inicial, com esta técnica foi possível identificar uma região em que há empresas que incluem informações falsas em sistema e identificar um agente cadastrador que incluiu centenas de registros falsos.

Palavras-Chave: Lei de Benford, fraude, auditoria, seguro-desemprego.

1. INTRODUÇÃO

No Brasil, convivemos com notícias de fraudes ao seguro-desemprego. As formas mais elaboradas deste tipo de golpe envolvem a criação de vínculos trabalhistas fictícios.

Vínculo fictício é a inclusão de dados simulados no sistema que é utilizado para processamento e pagamento de benefício. Os dados inseridos não combinam com dados de pessoas reais. Trata-se de invenção de dados para mentir ao sistema e sacar dinheiro público.

¹ Auditor Fiscal do Trabalho.

Com o uso da lei estatística denominada Lei de Benford, é possível separar casos normais, com correspondência na realidade, dos casos anormais, inventados. Para realizar esta separação e localizar dados falsos, foi utilizada a informação do número de endereço residencial do requerente de seguro-desemprego, um campo aparentemente de fácil invenção e qualquer número inventado não levantaria suspeita. De acordo com a regra de Benford, o primeiro dígito significativo dos números de residência deve seguir a uma distribuição teorizada.

Neste trabalho foram realizados testes informatizados para verificar a validade da lei de Benford dados de número de endereço residencial constantes nos requerimentos de seguro-desemprego. Também foram feitos recortes de universos menores, separando os mesmos dados de número de endereço por agentes digitadores de requerimentos, e verificando se estes recortes também seguem a lei de Benford. Desta forma, buscou-se identificar funcionários públicos que digitam requerimentos com informações falsas.

2. FRAUDES AO SEGURO-DESEMPREGO

A história brasileira possui uma vasta lista de fraudes e desvios de dinheiro público. O recebimento indevido de benefícios governamentais por meio da inclusão de dados falsos é uma das formas comuns de desvio. Em 1990, o jornalista Aor Ribeiro² fez um dossiê de fraudes deste tipo, situações de desvio de dinheiro da previdência social. Ao longo de seu livro, o jornalista faz descrição detalhada de milhares de casos em que há declaração falsa de óbito, declaração de situação de invalidez para pessoas saudáveis, uso de certidões de casamento falsas e outros.

Em 1986, o benefício do seguro-desemprego foi introduzido no Brasil por meio dos Decretos-Leis 2283 e 2284. Na Constituição de 1988 houve a previsão de que o seguro-desemprego é um direito dos trabalhadores. E em janeiro de 1990, foi promulgada a Lei 7998, que instituiu o Fundo de Amparo do Trabalhador – FAT e estipulou regras para o pagamento deste benefício.

² RIBEIRO, Aor. *Previdência: terror e mortes no reino das fraudes: subsídios para a história da pilhagem aos cofres públicos no Brasil*. São Paulo : Melhoramentos, 1990.

O objetivo do seguro-desemprego é o pagamento de um auxílio financeiro temporário para o trabalhador que for demitido e não tiver recursos para sustentar a sua família.³

Também o seguro-desemprego passa a ser objeto de fraudes. Em 1993, ao fazer uma análise ampla da recente legislação do seguro-desemprego, Wagner Balera demonstrou preocupação com as fraudes ao benefício e sugeriu o estabelecimento de mecanismos eficientes de punição.⁴

Eduardo Almeida Jeronimo em seu livro *Corrupção velada, Pequenas (grandes) fraudes ao programa do seguro-desemprego*, faz um catálogo de casos de fraudes corriqueiros. Este catálogo de fraudes ao sistema do seguro-desemprego inclui simulação de dispensa por justa causa com alteração do motivo de afastamento, simulação de dispensa nas quais o empregado continua trabalhando para a empresa, manutenção do empregado sem formalização do vínculo trabalhista até o recebimento de parcelas conseguidas em razão de vínculo anterior, omissão de existência de renda suficiente para manutenção da família ou omissão de recolocação no mercado de trabalho.⁵

2.1. VÍNCULO FICTÍCIO

Existem também quadrilhas especializadas. No trabalho de Jeronimo são citadas fraudes que envolvem a inserção de vínculos fictícios, definidas como fraudes maiores e mais profissionais.⁶ Os nomes que a Polícia Federal costuma criar dão um indício da variedade existente: Xequê Duplo, Duas Caras, MAC 70, Fake Work, Chakhal II. Nestas fraudes constatadas, existe a criação de empresas e pessoas fictícias, fantasmas ou laranjas, como são conhecidos.

³ BRASIL. Lei 7998, de 11 de janeiro de 1990. Regula o Programa do Seguro-Desemprego, o Abono Salarial, institui o Fundo de Amparo ao Trabalhador (FAT), e dá outras providências. Disponível em: <http://www.planalto.gov.br/ccivil_03/LEIS/L7998.htm>. Acesso em: 21 jun. 2019.

⁴ BALERA, Wagner. O seguro-desemprego no direito brasileiro. São Paulo : LTr, 1993, p. 171.

⁵ JERONIMO, Eduardo Almeida. *Corrupção velada: pequenas (grandes) fraudes ao programa do seguro-desemprego*. Timburi, SP : Editora Cia do eBook, 2015. Livro eletrônico, não foi possível identificar a página na qual a informação está.

⁶ JERONIMO, Eduardo Almeida. *Corrupção velada: pequenas (grandes) fraudes ao programa do seguro-desemprego*. Timburi, SP : Editora Cia do eBook, 2015. Capítulo 3.6. AS GRANDES (GIGANTES) FRAUDES (OS PROFISSIONAIS).

Houve inclusive criação de vínculos fictícios em lote, aos milhares. Um contabilista com conhecimentos em informática construiu um sistema informatizado, um robô, para incluir informações de requerimentos de seguro-desemprego e desviar cerca de R\$ 14 milhões.⁷ No algoritmo utilizado pelo fraudador, obviamente, houve a necessidade de configurar e comandar o robô para preencher campos necessários com dados inventados.

As operações já realizadas, aparentemente, não inibiram a ação de estelionatários. Podemos citar exemplos recentes da criação de vínculos fictícios para realizar fraudes. Em maio de 2019, a Polícia Federal divulgou operação realizada no Amapá. Conforme constatado:

Com o avanço da investigação, verificou-se que a fraude consistiu na criação de vínculo laboral falso de empregado com empresa prestadora de serviços de publicidade por parte do investigado, e simulação de rescisão contratual dos envolvidos, visando o recebimento do maior valor pago ao benefício de seguro-desemprego à época.⁸

A constituição de um vínculo fictício é feita por meio da inserção de dados falsos em sistema. Em um momento anterior, algumas vezes, são forjados documentos falsos para que o requerente encaminhe o benefício se passando por outrem.⁹ Em todos os casos, os dados do vínculo devem ser inventados. No meio deste processo de invenção, é necessário declarar os dados pessoais, dentre eles o endereço. Afinal, não faz muito sentido para um estelionatário deixar dados verdadeiros e facilitar a localização do fraudador ou de alguma pessoa que não recebeu benefício e poderia contestar a informação inserida.

⁷ MINISTÉRIO PÚBLICO FEDERAL, IPL 022/2014 – DPF/RDO/PA. Disponível em: <http://www.prpa.mpf.mp.br/news/2015/arquivos/Denuncia_MPF_organizacao_crimes_previdenciarios_outros_Redencao_Sao_Felix_Xingu.pdf>. Acesso em 28 jul. 2019.

⁸ POLÍCIA FEDERAL. PF combate fraudes na concessão de seguro-desemprego no Amapá. Disponível em: <<http://www.pf.gov.br/imprensa/noticias/2019/05/pf-combate-fraudes-na-concessao-de-seguro-desemprego-no-amapa>>. Acesso em 21 maio 2019.

⁹ ISTOÉ. “PF caça 21 por fraude de R\$ 20 milhões no seguro-desemprego”. Disponível em <<https://istoe.com.br/pf-caca-21-por-fraude-de-r-20-milhoes-no-seguro-desemprego/>>. Acesso em 21 maio 2019.

G1. “Polícia Civil faz operação contra falsificação de identidades no Paraná”. Disponível em: <<https://g1.globo.com/pr/oeste-sudoeste/noticia/2019/05/10/policia-civil-faz-operacao-contrafalsificacao-de-identidades-no-parana.ghtml>>. Acesso em 21 maio 2019.

3. LEI DE BENFORD

A lei de Benford, como é mais conhecida, foi registrada pela primeira vez em um artigo de 1881. Uma característica de grupos numéricos foi percebida de forma quase que intuitiva por Simon Newcomb ao manusear tábuas de logaritmo para fazer multiplicações com números grandes.¹⁰ Newcomb observou que havia uma consulta muito maior de números iniciados pelo dígito 1, as primeiras páginas do livro estavam mais gastas, quando comparadas com consultas aos números iniciados por 9, últimas páginas do livro de tábuas logarítmicas.

O primeiro dígito significativo de um número é o dígito mais à esquerda e que é diferente de zero.¹¹ Para o número “1234,50”, por exemplo, o primeiro dígito significativo é 1; para o número “94”, o primeiro dígito significativo é 9; para o número “001”, é 1, os zeros não são significativos. Conforme a descoberta de Newcomb, o primeiro dígito significativo de números que aparecem na natureza, que representam proporções de quantidades¹², devem aparecer com as frequências descritas no gráfico 1.

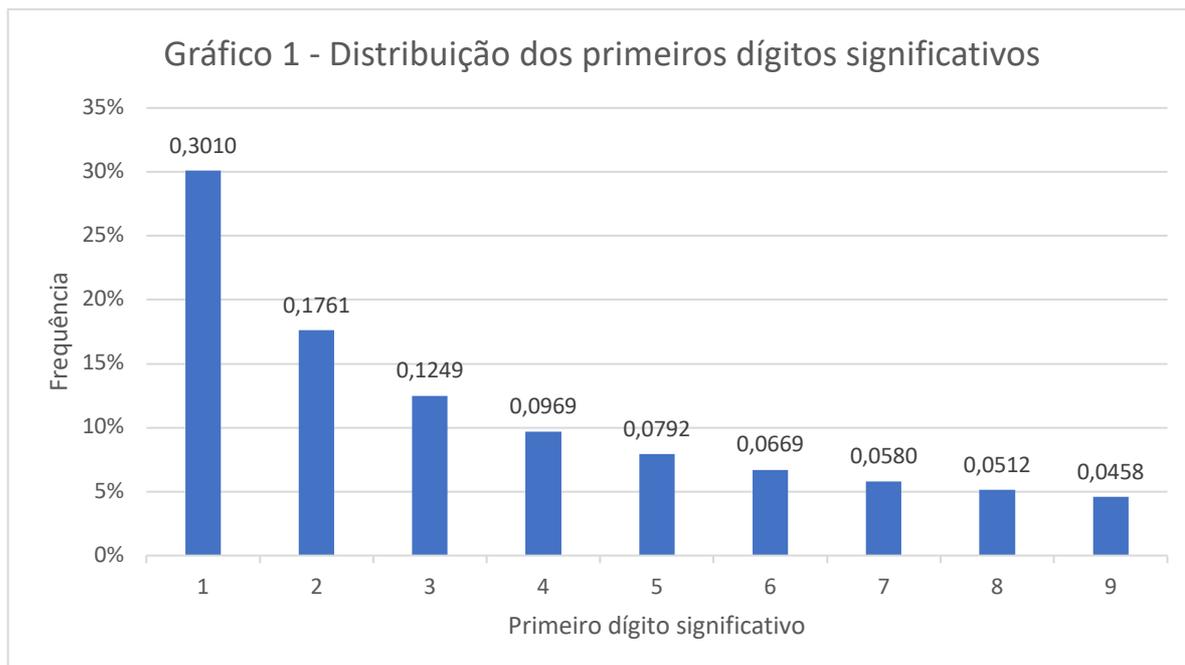
Existem conjuntos de números que apesar de serem proporções de quantidades não seguem a lei de Benford. Datas, altura de pessoas, altura de árvores, número de telefone, não seguem a regra de Benford porque têm limites superiores ou inferiores relativamente bem definidos.¹³

¹⁰ NEWCOMB, Simon. Note on the Frequency of Use of the Different Digits in Natural Numbers. American Journal of Mathematics, Vol 4, n.1/4, pp.39-40, 1881. Disponível em: <<https://www.uvm.edu/pdodds/research/papers/others/1881/newcomb1881a.pdf>>. Acesso em 10 maio 2019.

¹¹ AMARAL, Fernando. Introdução a ciência de dados: mineração de dados e Big Data. Rio de Janeiro : Alta Books, 2016, p. 122.

¹² NEWCOMB, Simon. Note on the Frequency of Use of the Different Digits in Natural Numbers. American Journal of Mathematics, Vol 4, n.1/4, pp.39-40, 1881. Disponível em: <<https://www.uvm.edu/pdodds/research/papers/others/1881/newcomb1881a.pdf>>. Acesso em 10 maio 2019, p. 39.

¹³ MILANI FILHO, Marco Antonio Figueiredo. 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. 47-63, mai/ago 2013, p. 50.



Fonte: elaboração própria com os valores citados por Newcomb.

Conforme Newcomb, em qualquer conjunto de números que representam quantidades, o primeiro dígito significativo será 1 em 30,1% dos casos e o primeiro dígito significativo 9 ocorrerá somente em aproximadamente 4,6% das observações.

Esta conclusão parece desafiar a nossa compreensão porque tendemos a acreditar que há igual chance de aparecer um dígito qualquer no início dos números quando medimos algo ou inventamos um número.

Em 1938, o engenheiro elétrico Frank Benford, também partindo da observação do uso de tábuas logarítmicas, faz um estudo mais aprofundado do fenômeno. No artigo intitulado *The law of anomalous numbers*, demonstrou geometricamente como a distribuição dos primeiros dígitos significativos pode ser deduzida, como se comporta e concluiu que é uma regra estatística invariante de base.¹⁴

Benford chega aos mesmos percentuais de Newcomb e, para a ocorrência do primeiro dígito significativo, apresenta fórmula na qual a frequência (F) de um número de 1 até 9 (a) aparecer como o primeiro dígito significativo segue a seguinte fórmula:¹⁵

¹⁴ BENFORD, Frank. The law of anomalous numbers. Proceedings of the American Philosophical Society, n. 78(4), p. 551–572, mar. 1938. Disponível em: <<http://www.jstor.org/discover/10.2307/984802>>. Acesso em 13 jun. 2019 (requer subscrição), p. 572.

¹⁵ BENFORD, Frank. The law of anomalous numbers. Proceedings of the American Philosophical Society, n. 78(4), p. 551–572, mar. 1938. Disponível em:

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

No artigo de 1937, a lei de Benford, como passou a ser chamada, foi aplicada a um conjunto de mais de 20.000 medições. Benford demonstrou que o tamanho dos rios de um atlas, a população, a área de países, a sequência de números fatoriais, o peso atômico de elementos, números de endereço quando têm as medições tabuladas seguem a distribuição teorizada. Dados de várias origens tendem a seguir esta curiosa regra estatística.

Nos anos 1990, estudos sobre a lei de Benford ganharam impulso e passaram a ser utilizados para a detecção de fraudes financeiras.¹⁶

Os estudos avançam também propondo novos testes estatísticos como o teste dos dois primeiros dígitos significativos¹⁷ e o teste dos últimos dígitos significativos.¹⁸ Há também, e esta possibilidade já estava no trabalho de Newcomb, como testar a proporção de ocorrência do segundo e terceiros dígitos significativos. Neste trabalho, porém, utilizaremos somente o primeiro dígito.

4. BANCOS DE DADOS DIGITAIS

No final do século passado, há uma transformação nos instrumentos de processamento de dados que facilitou a utilização da lei de Benford.

A guarda de informações passou por uma grande transformação com o surgimento e popularização dos computadores com bancos de dados digitais. Ao invés dos registros em papéis e a busca de dados em arquivos físicos, a humanidade

<<http://www.jstor.org/discover/10.2307/984802>>. Acesso em 13 jun. 2019 (requer subscrição), p. 554.

¹⁶ MILANI FILHO, Marco Antonio Figueiredo. 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. 47-63, mai/ago 2013, p. 49.

¹⁷ BENFORD, Frank. The law of anomalous numbers. *Proceedings of the American Philosophical Society*, n. 78(4), p. 551–572, mar. 1938. Disponível em:

<<http://www.jstor.org/discover/10.2307/984802>>. Acesso em 13 jun. 2019 (requer subscrição), p. 78.

¹⁸ NIGRINI, Mark John. *Benford's Law. Applications for Forensic Accounting Auditing, and Fraud Detection*. John Wiley & Sons, Inc. : Hoboken, New Jersey, 2012, p. 129.

passou a ter como consultar e atualizar de forma muito rápida dados arquivados de forma digital.

Estas alterações mudaram a forma de analisar informações. No século 17, por exemplo, os registros de batismos e casamentos mantidos em papéis pelas igrejas eram dados que para serem consultados envolviam muito trabalho. Somente para contar as quantidades e agrupar por características como a causa da morte, como fez o inglês John Graunt, foi necessário um trabalho enorme.¹⁹ Tinha que contar um a um, anotar e somar. Para fazer uma segunda soma, recomeçar tudo.

O trabalho de Benford em 1937, ao agrupar dados pelo primeiro dígito significativo, somar e verificar a conformidade com a distribuição teorizada, também foi enorme.

Mas, após a popularização do armazenamento digital e desenvolvimento de linguagens para consulta aos dados, este tipo de trabalho pode ser feito a partir de comandos relativamente simples que são respondidos quase que instantaneamente mesmo para uma quantidade de milhões de registros.

Esta transformação é recente. O primeiro disco rígido surge em 1956, os bancos de dados relacionais surgem em meados dos anos 1970. Em 1989, por exemplo, Mark Nigrini, que passou a ser referência no estudo da lei de Benford posteriormente, teve que abandonar processos de análise de dados porque o seu computador 286 não suportava o processamento dos dados e o acesso aos computadores maiores, mainframes, era limitado.²⁰

Para quem possui acesso ao banco de dados, a realização destes testes é atualmente bastante facilitada. Com um volume não muito grande de dados, mas bem mais do que os 20.000 números analisados por Benford, é possível fazer com planilhas eletrônicas em computadores pessoais. Nigrini demonstra como utilizar fórmulas automáticas do Excel para analisar dados em busca de fraudes.²¹

¹⁹ BERNSTEIN, Peter. Desafio aos deuses: a fascinante história do risco. Rio de Janeiro : Alta Books, 2018, p. 76.

²⁰ NIGRINI, Mark John. Benford's Law. Applications for Forensic Accounting Auditing, and Fraud Detection. John Wiley & Sons, Inc. : Hoboken, New Jersey, 2012, p. xii-xvi.

²¹ NIGRINI, Mark John. Benford's Law. Applications for Forensic Accounting Auditing, and Fraud Detection. John Wiley & Sons, Inc. : Hoboken, New Jersey, 2012, p. 80.

5. BUSCA DE ANORMALIDADES

Um dos processos para identificar situações fraudulentas é dividir os dados em subgrupos e verificar se todos estes subgrupos seguem a proporção teorizada pela lei de Benford. O universo dos dados pode ser dividido em períodos de tempo e ser executada uma auditoria contínua.²² Ou os dados também podem ser divididos em subgrupos levando em conta outras características.²³ Para verificar a conformidade de grupos de dados à lei de Benford, Nigrini apresenta testes estatísticos como o Z-teste, o qui-quadrado (χ^2), o teste Kolmogorov–Smirnov e o desvio médio absoluto.²⁴ Com estes testes é possível verificar a conformidade do subgrupo com a lei de Benford. Dividir entre dados normais e anormais.

Neste trabalho foi utilizado o teste do qui-quadrado.

Considerando a frequência esperada (E) e a frequência observada (O) para os primeiros dígitos significativos ($k \in 1,2,3,4,5,6,7,8,9$) em uma amostra, o qui-quadrado (χ^2) é calculado a partir da fórmula a seguir:

$$\chi^2 = \sum_{k=1}^n \frac{(O_k - E_k)^2}{E_k}$$

O resultado desta soma, quando aproximar-se de zero, indicará que a distribuição observada tem frequência idêntica a frequência esperada.

A partir do qui-quadrado, com auxílio de tabelas de distribuição, é possível também testar a hipótese de que os valores observados seguem ou não a mesma regra de formação dos valores esperados. Também é possível gerar gráficos e verificar visualmente a similaridade entre as distribuições.

²² COSTA, José Isidrio de Freitas. Análise de Conformidade da Lei de Newcomb-Benford no Ambiente de Auditoria Contínua: Uma Proposta de Identificação de Desvios no Tempo. XXVII encontro da ANPAD. Rio de Janeiro, set. 2013. Disponível em: <http://www.anpad.org.br/admin/pdf/2013_EnANPAD_CON1565.pdf>. Acesso em 28 jun. 2019.

²³ Há um exemplo em: NIGRINI, Mark John. Benford's Law. Applications for Forensic Accounting Auditing, and Fraud Detection. John Wiley & Sons, Inc. : Hoboken, New Jersey, 2012, p. 172.

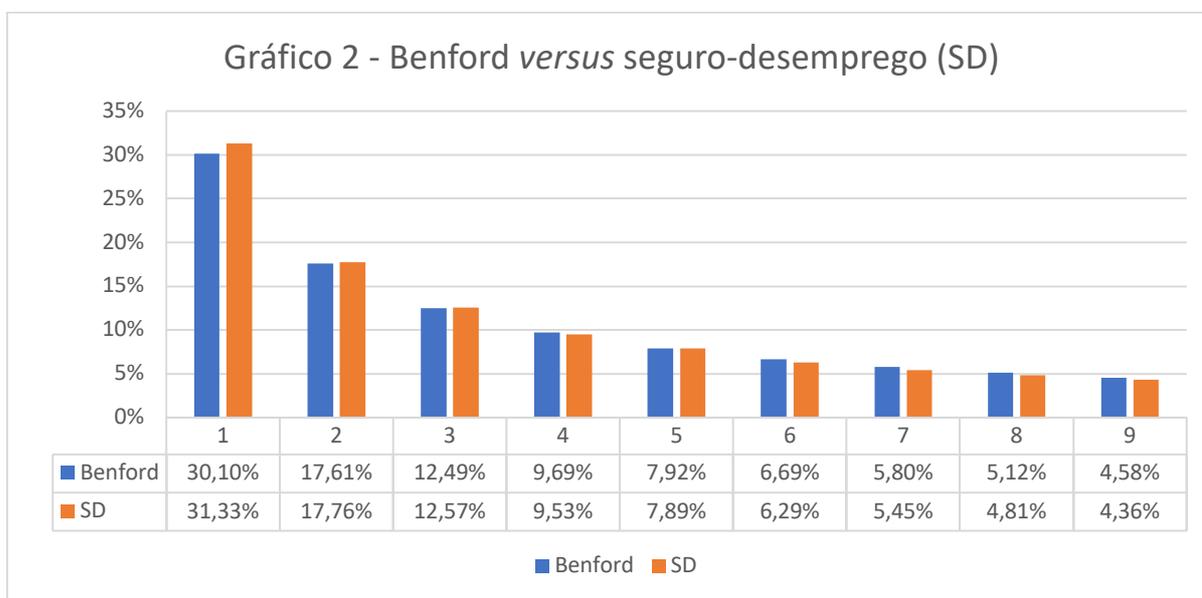
²⁴ NIGRINI, Mark John. Benford's Law. Applications for Forensic Accounting Auditing, and Fraud Detection. John Wiley & Sons, Inc. : Hoboken, New Jersey, 2012, p. 160.

6. ANÁLISE DOS DADOS DE SEGURO DESEMPREGO

Em um ambiente de banco de dados relacional, foram contados os dados de requerimentos de seguro-desemprego dos anos de 2018 em diante (de 01/01/2018 até 31/05/2019). Apesar desta escolha, outros recortes de tempo podem ser utilizados. Ressalta-se, porém, que um recorte de tempo muito longo pode diluir períodos de inserção de dados falsos dentro de períodos que dados válidos eram incluídos.

Foram separados e agrupados do banco de dados: o número que identifica o agente cadastrador do benefício, funcionário público que digitou o requerimento; a quantidade de requerimentos cujo campo de endereço do requerente tinham o primeiro dígito numérico significativo 1, 2, 3, 4, 5, 6, 7, 8 e 9, agrupados por agente cadastrador; e a estatística qui-quadrado para cada um dos dados agrupados por agente cadastrador.

No período analisado existiram 8.818.166 de requerimentos. Após agrupar por primeiro dígito significativo a frequência é quase idêntica a esperada pela Lei de Benford.²⁵



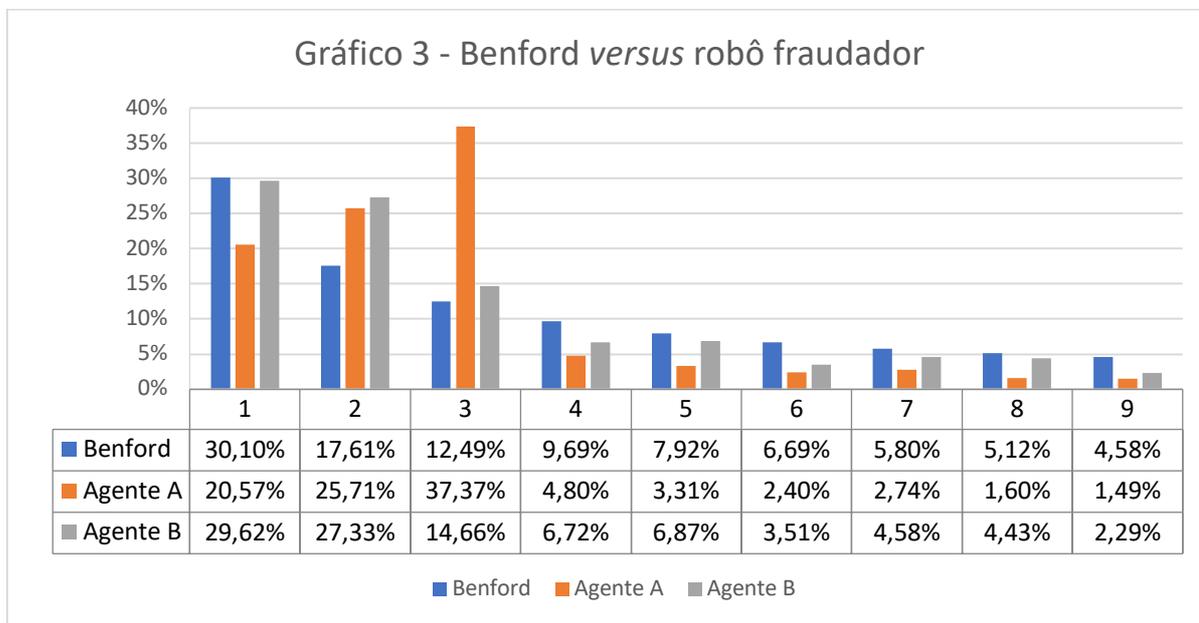
Fonte: elaboração própria.

²⁵ O mesmo teste foi realizado para o período de 01/01/2015 em diante, 29.438.193 requerimentos, e a conclusão foi idêntica. Mantido no texto o período de 2018 e 2019 para, na sequência, fazer a busca de fraudadores que agem atualmente.

Neste ponto, temos uma curiosidade estatística que empiricamente comprova que o número de endereço no banco de dados, como um todo, segue a proporção teorizada. É a mesma conclusão de Benford, que já havia incluído o número de endereço entre os exemplos estudados em 1937.

A partir dos dados como um todo, é possível especular, por exemplo, porque o único desvio mais significativo é uma quantidade um pouco maior de números começados por 1 (31,29% ante o esperado de 30,10%). Talvez ao preencher um requerimento não se sabia o número e incluiu-se o número “1” ou “001”.

Para demonstrar que a inserção de dados falsos segue uma proporção diferente é possível separar os dados de agentes cadastradores já identificados como fraudadores. Os dados gerados pelo contador com conhecimentos de informática que construiu um robô para cadastrar requerimentos simulados²⁶, citado acima, não seguiam a regra de Benford. Estes fraudadores, nos anos de 2013 e 2014 inseriam dados fictícios com uso do perfil de dois agentes cadastradores identificados no gráfico 3.



Fonte: elaboração própria.

²⁶ MINISTÉRIO PÚBLICO FEDERAL, IPL 022/2014 – DPF/RDO/PA. Disponível em: <http://www.prpa.mpf.mp.br/news/2015/arquivos/Denuncia_MPF_organizacao_crimes_previdenciarios_outros_Redencao_Sao_Felix_Xingu.pdf>. Acesso em 28 jul. 2019.

O agrupamento de todos os requerimentos incluídos em sistema em nome dos dois agentes (A e B) se afasta da Lei de Benford.

Mas, esta é uma situação já identificada e de período anterior. Interessa saber se é possível fazer a mesma análise nos dados atuais e tentar identificar situações fraudulentas.

Voltando aos requerimentos cadastrados em 2018 e 2019, os dados foram agrupados por agente cadastrador de benefício. O agente cadastrador é o funcionário público que inclui os dados em sistema para que o empregado tenha direito ao benefício.

Outros agrupamentos, por período, por região ou por empresa também seriam possíveis e talvez até mais adequados. A divisão por grupo, como um todo, serve para direcionar o olhar da auditoria em busca de alguma irregularidade. Um dos grupos estará mais em conformidade com a normalidade e outro não.

Após agrupar os dados por agente e ordenando os dados para evidenciar aqueles que tiveram um número maior no teste do qui-quadrado, foi possível separar situações que são distantes da lei de Benford. Classificou-se para identificar quais agentes tinham distribuições que mais se distanciavam da Lei de Benford.

Na tabela 1, a seguir, o agente é identificado com um número ordinal para não evidenciar o seu cadastro, a quantidade de requerimentos com primeiro dígito significativo 1 até 9 são identificados na coluna Q1 até Q9, a quantidade total de requerimentos do agente está na coluna QT e o resultado do teste qui-quadrado está na coluna intitulada QUI.

Tabela 1-Classificação pela maior distância da distribuição de Benford

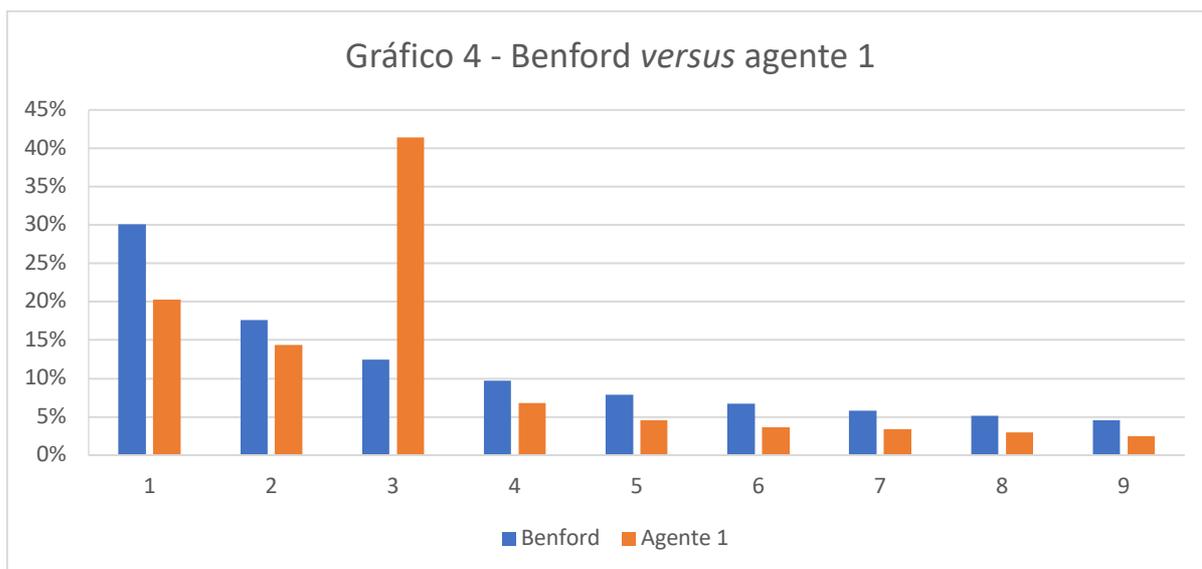
Agente	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	QT	QUI
1	1721	1218	3509	578	383	308	288	255	211	8471	6540,634
2	3612	215	141	94	109	61	54	55	53	4394	5672,765
3	559	398	474	1622	417	817	608	436	541	5872	4258,323
4	3267	273	316	132	104	110	61	47	46	4356	4214,327
5	259	346	736	47	71	52	27	18	39	1595	1800,336
6	654	358	275	229	147	92	112	93	542	2502	1681,339
7	141	235	44	475	17	57	107	88	106	1270	1381,18
8	165	279	60	500	19	51	102	102	116	1394	1373,674
9	963	99	92	58	58	27	47	36	19	1399	1005,467
10	128	217	51	361	16	47	94	78	84	1076	919,8619
11	623	827	595	253	191	58	35	27	36	2645	849,5537
12	449	335	422	515	293	177	21	43	22	2277	830,1746
13	352	346	296	119	100	129	212	300	242	2096	826,9254
14	1117	1228	900	494	332	165	120	70	79	4505	820,3461

Fonte: elaboração própria.

Como resultado, não se chegou a uma lista de fraudadores. A conclusão é de que os números de endereços cadastrados foram inventados. Talvez, quando não sabia o número houve o preenchimento de “9999”. Talvez houve desleixo do agente cadastrador que costuma preencher os requerimentos com qualquer número para acelerar o seu trabalho. Apesar de serem situações talvez reprováveis, não se tratam de fraudes.

O agente cadastrador número 1, pior classificado na comparação com a lei de Benford, por alguma conveniência, cadastrava todos os endereços nos requerimentos com o número de casa “30”. Complementando-se a análise com a verificação de alguns requerimentos e vínculos associados, foi possível constatar que não se trata de situação fraudulenta.

Comparando-se a atuação do agente cadastrador 1 com o esperado pela Lei de Benford temos o gráfico 4.



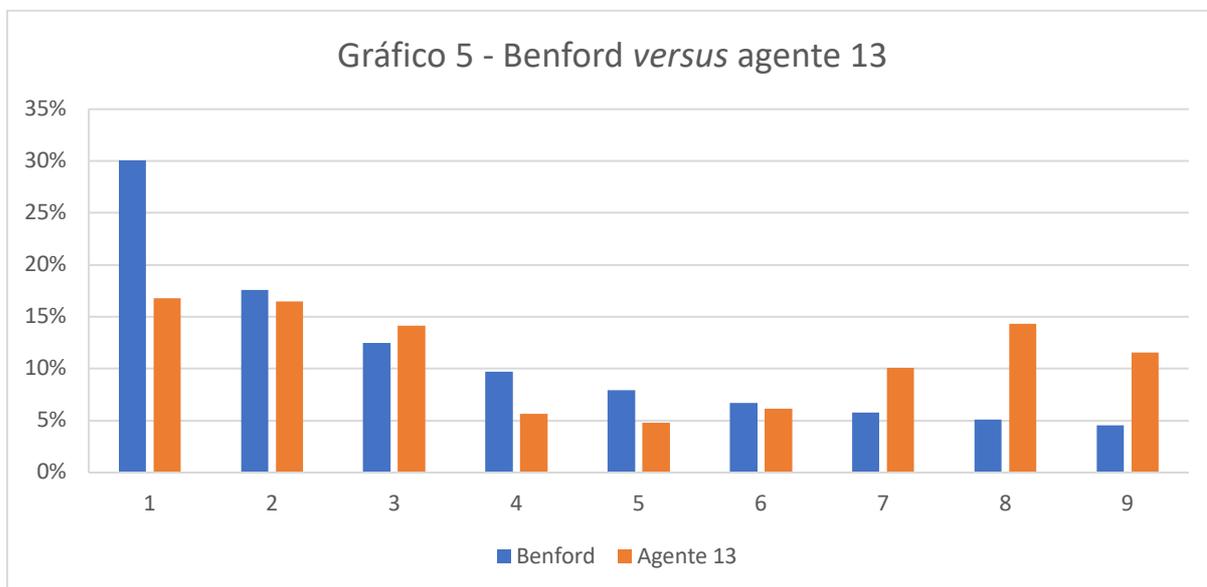
Fonte: elaboração própria.

Apesar da excessiva existência de números iniciados por 3, em razão dos números “30” digitados, a curva dos demais dígitos iniciais é paralela à curva de Benford. De forma geral, pelo menos para os demais números, há adequação à lei de Benford.

A mesma situação de uso de um número como uma espécie de coringa comprometeu o resultado de outros agrupamentos. O agente cadastrador 2, teve um excesso de dígitos 1, porque cadastrou vários endereços com o número “1”. Também não se trata de situação fraudulenta. O agente cadastrador 3, utilizava-se de números de residência “45”, “98”, em requerimentos seguidos, como se houvesse deixado automaticamente preenchido. Quando não tinham o número ou queria evitar o trabalho de busca-lo, os agentes de número 4 e 9, utilizavam “01”. Nos agrupamentos que foram identificados com uso de um número de endereço coringa, a quantidade destes números está com a célula destacada na tabela 1.

Na análise dos dados, o agente cadastrador número 13, linha em destaque na tabela, foi identificado com vários vínculos fictícios inseridos. Existiram requerimentos que o sistema bloqueou em análise de risco, houve tentativas sucessivas de inclusão do mesmo empregado com dados de endereços diferentes e as pesquisas dos vínculos nos dados fornecidos pela empresa não sustentaram a sua existência real.

Ao analisar o gráfico do agente cadastrador 13 em comparação com a lei de Benford, notamos que houve uma distribuição que foge da curva esperada pela regra de Benford.



Fonte: elaboração própria.

Neste gráfico, os demais números não seguem uma curva paralela à de Benford. Na análise do gráfico é possível notar que o desvio da regra é mais significativo.

Na comparação com a distribuição de Benford, de uma forma resumida, a classificação da tabela 1 foi bastante sensibilizada pela inclusão de um número específico pelo agente cadastrador (agentes 1, 2, 4, 5, 6, 9, 11 e 12). Os primeiros colocados da classificação estão lá porque um número qualquer era preenchido quando não se sabia o número somente para o cadastro passar.

A partir da tabela 1, é possível analisar caso a caso, buscar chegar a conclusões parecidas e localizar outros fraudadores.

Outro caminho também foi testado. Na análise dos dados, verificamos se a distribuição dos valores de cada agente foge da distribuição de Benford em direção de uma distribuição proporcional dos primeiros dígitos. Supondo que haja uma preocupação em disfarçar números inventados. Para um fraudador imaginário qualquer, uma estratégia bastante razoável seria distribuir os primeiros números significativos na mesma proporção, incluir a mesma quantidade de cada um dos

números, 11,11% (1/9) dos números começariam com 1, 11,11% começariam com 2, 11,11% com 3 e assim por diante até 9. Para identificar se os números foram incluídos com o primeiro dígito significativo na mesma proporção é possível ajustar o cálculo do qui-quadrado para identificar quais agentes têm requerimentos que seguem esta distribuição proporcional dos primeiros dígitos significativos. A tabela a seguir mostra a classificação dos agentes cadastradores para identificar quais distribuíram os primeiros dígitos mais proporcionalmente. O teste do qui-quadrado (CHI-P) mais próximo a zero indica que os primeiros dígitos significativos estão mais proporcionalmente distribuídos.

Tabela 2 - Classificação pela distribuição proporcional dos dígitos

Agente	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	QT	CHI-P
13	352	346	296	119	100	129	212	300	242	2096	0,1587447
15	129	133	119	85	64	78	60	40	35	743	0,1792223
16	573	456	442	315	296	320	200	154	117	2873	0,1977176
17	525	461	432	288	277	308	185	146	112	2734	0,2011146
18	580	467	448	308	321	346	219	134	94	2917	0,2140166
19	250	189	182	108	107	111	88	64	63	1162	0,2172516
20	406	345	288	230	173	198	133	88	99	1960	0,2280112
21	136	107	85	86	64	51	47	44	19	639	0,2294619
22	243	189	199	99	115	138	82	65	48	1178	0,2305159
23	154	76	71	72	54	55	61	40	34	617	0,2327753
24	310	133	88	92	142	60	121	193	199	1338	0,2331767
25	278	215	202	156	129	115	88	68	53	1304	0,2371284
26	143	110	96	82	64	56	52	32	24	659	0,2441978
27	166	129	140	107	81	86	48	33	33	823	0,2450843
28	374	280	284	200	177	156	125	86	63	1745	0,2463070
29	194	118	46	55	98	76	45	92	67	791	0,2525386

Fonte: elaboração própria.

No resultado, conforme recomendação para consistência das conclusões²⁷, foram incluídos somente agentes que cadastraram mais de 500 benefícios.

O agente cadastrador número 13, que havia aparecido também na tabela 1, possui gráfico mais adequadamente proporcional. Em outras palavras, tentou distribuir os dados uniformemente.

²⁷ NIGRINI, Mark John. Benford's Law. Applications for Forensic Accounting Auditing, and Fraud Detection. John Wiley & Sons, Inc. : Hoboken, New Jersey, 2012, p. 154.

Neste ponto, foi identificado que os agentes 13, 15, 16, 17, 18, 19, 20, 22, 24, 25, 26, 27 e 28 tinham requerimentos fictícios incluídos e eram de uma mesma região geográfica. A partir desta constatação, identificando um dos requerimentos fictícios e a partir dele outros, foi possível descobrir qual era o padrão de inserção de dados falsos e quais empresas estão sendo utilizadas para fraudar o recebimento de seguro-desemprego.

Na análise dos outros agentes cadastradores, o de número 29 havia inserido centenas de vínculos fictícios. Este usuário do sistema inseria quase que exclusivamente vínculos fictícios.

Em uma análise rápida dos outros três agrupamentos (21, 23 e 24) não foram encontradas irregularidades.

7. CONSIDERAÇÕES FINAIS

A utilização da lei de Benford promoveu achados de auditoria úteis para a busca de fraudes em lote nos requerimentos de seguro-desemprego. A constatação de desvios significativos em determinada região, possibilitou a identificação das empresas utilizadas para inserir dados fictícios. Também foi possível a identificação de agente público que insere vínculos fictícios.

A utilização da lei de Benford, obviamente, não encontra todos os casos de fraude. Em possíveis situações que o fraudador não tente esconder o número de endereço do requerente, a regra de Benford é inútil para buscar anormalidades.

Algumas situações de anormalidades em relação à lei de Benford eram inserções de dados para preencher o campo número com algum número qualquer ou até mesmo o desleixo do agente responsável pela inclusão dos dados. Estas situações puderam ser depuradas em análise gráfica da distribuição dos primeiros dígitos e com a verificação de qual o número era mais recorrente no subconjunto de dados.

Como é um trabalho em desenvolvimento, as classificações citadas em parte nas tabelas 1 e 2 poderão revelar outras situações fraudulentas. E também poderão ser testados novos recortes temporais dos dados.

A mesma técnica de utilização da lei de Benford, com o uso do número de endereço ou outro campo que se adequa à regra, pode também subsidiar outras auditorias. Especificamente em relação a bancos de dados públicos, é possível fazer análises parecidas em licenças de pesca, seguro-defeso, bolsa-família, doações eleitorais e outros.

Não é necessária uma especulação teórica para verificar se um conjunto de dados segue a lei de Benford. Como demonstrado neste trabalho, a verificação se um conjunto de números segue esta proporção é de fácil verificação para quem tem acesso ao banco de dados informatizado.

O trabalho demonstra também que a utilização de um campo específico do requerimento, o número de endereço do requerente, já é suficiente para chegar à fraudadores. Mas, há outros campos no requerimento que também seguem alguma distribuição que pode ser facilmente identificada em bancos de dados relacionais. Mesmo que não seja a distribuição de Benford, cada dado segue a uma lógica que pode ser identificada em busca de padrões. Neste ponto, o proprietário do banco de dados está em vantagem em relação aos fraudadores porque pode consultar o comportamento dos dados como um todo. Há também a relação entre os dados que pode ser identificada e testada.

Tentar fraudar o número de endereço em um grupo de requerimentos, por exemplo, levaria a necessidade de estudar como funciona a lei de Benford. Não bastaria um gerador de números aleatórios. Não bastaria o conhecimento do processo descrito neste trabalho, há também outras regras para segundos dígitos significativos, primeiros dois dígitos significativos e últimos dígitos significativos. E, no final, o fraudador teria somente incluído um número de endereço com mais chances de não ser identificado como falso. Mas, existem outros dados em requerimentos, existe a relação entre estes dados, existe o cruzamento com outros bancos de dados. Imitar um dado autêntico não é tarefa tão simples quanto parece.

REFERÊNCIAS

AMARAL, Fernando. Introdução a ciência de dados: mineração de dados e Big Data. Rio de Janeiro: Alta Books, 2016.

BALERA, Wagner. O seguro-desemprego no direito brasileiro. São Paulo: LTr,

Revista da Escola Nacional da Inspeção do Trabalho – Ano 3

1993.

BENFORD, Frank. The law of anomalous numbers. Proceedings of the American Philosophical Society, n. 78(4), p. 551–572, mar. 1938. Disponível em: <<http://www.jstor.org/discover/10.2307/984802>>. Acesso em 13 jun. 2019 (requer subscrição).

BERNSTEIN, Peter. Desafio aos deuses: a fascinante história do risco. Rio de Janeiro: Alta Books, 2018.

BRASIL. Lei 7998, de 11 de janeiro de 1990. Regula o Programa do Seguro-Desemprego, o Abono Salarial, institui o Fundo de Amparo ao Trabalhador (FAT), e dá outras providências. Disponível em: <http://www.planalto.gov.br/ccivil_03/LEIS/L7998.htm>. Acesso em: 21 jun. 2019.

COSTA, José Isidio de Freitas. Análise de Conformidade da Lei de Newcomb-Benford no Ambiente de Auditoria Contínua: Uma Proposta de Identificação de Desvios no Tempo. XXVII encontro da ANPAD. Rio de Janeiro, set. 2013. Disponível em: <http://www.anpad.org.br/admin/pdf/2013_EnANPAD_CON1565.pdf>. Acesso em 28 jun. 2019.

G1. “Polícia Civil faz operação contra falsificação de identidades no Paraná”. Disponível em: <<https://g1.globo.com/pr/oeste-sudoeste/noticia/2019/05/10/policia-civil-faz-operacao-contrafalsificacao-de-identidades-no-parana.ghtml>>. Acesso em 21 maio 2019.

ISTOÉ. “PF caça 21 por fraude de R\$ 20 milhões no seguro-desemprego”. Disponível em <<https://istoe.com.br/pf-caca-21-por-fraude-de-r-20-milhoes-no-seguro-desemprego/>>. Acesso em 21 maio 2019.

JERONIMO, Eduardo Almeida. Corrupção velada: pequenas (grandes) fraudes ao programa do seguro-desemprego. Timburi, SP: Editora Cia do eBook, 2015.

MILANI FILHO, Marco Antonio Figueiredo. 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. 47-63, mai/ago 2013.

MINISTÉRIO PÚBLICO FEDERAL, IPL 022/2014 – DPF/RDO/PA. Disponível em: <http://www.prpa.mpf.mp.br/news/2015/arquivos/Denuncia_MPF_organizacao_crimes_previdenciarios_outros_Redencao_Sao_Felix_Xingu.pdf>. Acesso em 28 jul. 2019.

NEWCOMB, Simon. Note on the Frequency of Use of the Different Digits in Natural Numbers. American Journal of Mathematics, Vol 4, n.1/4, pp.39-40, 1881. Disponível em: <<https://www.uvm.edu/pdodds/research/papers/others/1881/newcomb1881a.pdf>>. Acesso em 10 mai. 2019.

NIGRINI, Mark John. Benford's Law. Applications for Forensic Accounting Auditing, and Fraud Detection. John Wiley & Sons, Inc.: Hoboken, New Jersey, 2012.

POLÍCIA FEDERAL. PF combate fraudes na concessão de seguro-desemprego no Amapá. Disponível em: <<http://www.pf.gov.br/imprensa/noticias/2019/05/pf-combate-fraudes-na-concessao-de-seguro-desemprego-no-amapa>>. Acesso em 21 maio 2019.

RIBEIRO, Aor. Previdência: terror e mortes no reino das fraudes: subsídios para a história da pilhagem aos cofres públicos no Brasil. São Paulo: Melhoramentos, 1990.

THE USE OF BENFORD'S LAW TO IDENTIFY FRAUDULENT DATA INSERTION IN INSURANCE-UNEMPLOYMENT REQUIREMENTS

ABSTRACT

It is evident that Benford's law can be used to identify fraudulent data insertion cases in the unemployment insurance system. Fictitious data or fictitious links are information that is fraudulent and does not have a correlation with reality, information entered to get undue payments. Benford's law is a curious empirical confirmation that the first significant digits of some sets of numbers follow an unequal distribution but a downward curve in which 30.1% of the numbers begin with 1 and only 4.58% begin with digit 9. From the home address of the applicant for unemployment insurance, the proposal describes a methodology for data treatment, separation of subsets and identification of non-real situations. In first conclusion, with this technique it was possible to identify a region with companies that include false information in the system and identify also a registrar agent that included hundreds of false records.

Palavras-Chave: Benford Law's, fraud, audit, unemployment insurance.