O Uso da Lei Newcomb-Benford na contabilidade e auditoria

Ângelo Henrique Lopes da Silva

A Lei Newcomb-Benford encontra aplicação relevante na atividade de auditoria, podendo aumentar as possibilidades de detecção de fraudes em balanços contábeis. A referida lei proporciona uma maneira adicional de averiguação de registros numéricos, através da comparação das freqüências observadas de aparecimento dos dígitos nos balanços àquelas preconizadas pela lei. Entretanto, como todo modelo quantitativo, principalmente quando aplicado às ciências sociais, devemos identificar a qual universo dos fenômenos humanos e sociais aplicam-se os seus resultados. É importante que o auditor conheça a quais dados se aplicam, os tipos de fraudes detectáveis, bem como realizar os testes estatísticos que permitam a inferência de conclusões advindas da Lei Newcomb-Benford.

A análise de Benford é uma arma bastante difundida entre empresas de investigação e órgãos de fiscalização internacionais, sendo empregada na detecção de fraudes e desvios em grandes bancos de dados, sobretudo os contábeis. A análise digital de Benford é contemplada em alguns softwares de auditoria, a exemplo do *Audit Command Language* (ACL). Este texto pretende apresentar a Lei Newcomb-Benford, através de sua formulação; indicar o conjunto de dados aos quais se aplica; introduzir os testes estatísticos usados na investigação; encerrando com uma explicação inicial acerca de seu uso pelo ACL.

1. A LEI NEWCOMB-BENFORD E SUA APLICABILIDADE

Suponha que um auditor se depare com um grande conjunto de dados, podendo ser um balancete contábil, de retornos financeiros de mercado de ações ou censos populacionais. Escolhe-se aleatoriamente um registro contábil e pergunta-se qual a probabilidade de que esse número se inicie com o algarismo 1? Intuitivamente, tendo-se em mente uma possível distribuição uniforme dos dígitos, a resposta natural seria uma probabilidade de 1/9 (11,1%). Surpreendentemente, a resposta é equivocada. A Lei Newcomb-Benford afirma que essa probabilidade é de 30,1%. Para o algarismo 2, a mesma probabilidade é de 17,6%; para o 3, 12,5%; decaindo finalmente para 4,6%, no caso do 9.

Ângelo Henrique Lopes da Silva é servidor do TCU, graduado em Engenharia Mecânica no Instituto Tecnológico de Aeronáutica (ITA), especialista em Contabilidade e Auditoria pela Universidade Federal do Piauí (UFPI) e mestre em Economia pela Universidade de Brasília (UnB).

"O primeiro a perceber tal lei natural foi o astrônomo e matemático Simon Newcomb, que, em 1881, observou que, em uma biblioteca pública, as primeiras páginas dos livros de logaritmo eram mais usadas, encontrando-se mais sujas e desgastadas." (DURTSCHI et.al., 2004, p.18; SANTOS et.al.,2003, p.36). Ou seja, os leitores procuravam mais freqüentemente por números que começavam por 1 do que por números que começavam por 2, 3 e assim por diante.

"Cerca de 50 anos mais tarde, o físico Frank Benford identificou o fenômeno observado por Newcomb e buscou testá-lo com mais de 20.000 observações de diversas origens, como: áreas de rios, pesos atômicos de elementos e números apresentados em artigos da Reader's Digest." (DURTSCHI et.al., 2004, p.20; SANTOS et.al., 2003, p.36). Confirmando o fenômeno, Benford concluiu que os números apresentavam um padrão de comportamento em que algarismos baixos aparecem com mais freqüência na primeira posição que nas últimas. Benford não só aprofundou o estudo sobre o tema, como também divulgou amplamente seus resultados, o que lhe conferiu o título do fenômeno, como Lei de Benford. Mas em respeito ao precursor, Simon Newcomb, preferimos chamá-la de Lei Newcomb-Benford, no entanto, adotando a denominação de distribuição de Benford, por este ter se aprofundado matematicamente no tema.

Desde então, vários matemáticos e estatísticos estudam o fenômeno e têm descoberto uma maior gama de aplicações. O matemático Hill (1995 apud DURTSCHI et.al., 2004, p.20) não só demonstrou a Lei Newcomb-Benford, assim como indicou aplicações para dados no mercado de ações, em censos estatísticos e em dados contábeis. Afirmou ainda que a distribuição de Benford, assim como outras distribuições, tal como a normal, é um fenômeno empírico. Boyle (1994 apud DURTSCHI et.al., 2004, p.20) demonstrou que dados que são obtidos de fontes diversas e que são multiplicados ou divididos seguem a lei Newcomb-Benford. Isso explica por que a maioria dos dados contábeis apresenta comportamento bem próximo à distribuição de Benford, visto que frequentemente submetem-se a operações matemáticas simples. "No entanto, é Nigrini que se destaca pela ampla pesquisa da aplicação da lei de Newcomb-Benford na detecção de fraudes contábeis." (DURTSCHI et.al., 2004, p.22).

2. A FÓRMULA DA LEI NEWCOMB-BENFORD

Antes de passarmos à fórmula que nos leva aos resultados da Lei Newcomb-Benford, superemos o golpe que o seu resultado traz à nossa intuição comum. Imagine o valor de um determinado registro contábil em 1 milhão e seu crescimento a uma taxa de 5% ao ano:

Intuição correta para Lei Newcomb-Benford					
Ano	Valor ($)	Ano	Valor ($)	Ano	Valor ($)
1	1.000.000	11	1.628.895	21	2.653.298
2	1.050.000	12	1.710.339	22	2.785.963
3	1.102.500	13	1.795.856	23	2.925.261
4	1.157.625	14	1.885.649	24	3.071.524
5	1.215.506	15	1.979.932	25	3.225.100
6	1.276.282	16	2.078.928	26	3.386.355
7	1.340.096	17	2.182.875	27	3.555.673
8	1.407.100	18	2.292.018	28	3.733.456
9	1.477.455	19	2.406.619	29	3.920.129
10	1.551.328	20	2.526.950	30	4.116.136

Crescimento estável é uma hipótese razoável para dados contábeis. Podemos notar que o primeiro algarismo permanecerá 1 até que o valor alcance 2 milhões, o que acontecerá no 16º ano. O primeiro algarismo permanecerá 2 por 8 anos; permanecerá 3 po 6 anos e assim por diante sempre numa frequência decrescente. Portanto, os valores menores para primeiro dígito são mais prováveis que os valores maiores.

Algebricamente, a probabilidade de que um determinado algarismo *d* inicie um número qualquer é dada pela seguinte fórmula:

$$\text{Prob}\ (d) = \log\left(1 + \frac{1}{d}\right)$$

onde d = 1, 2,3,4,5,6,7,8 ou 9

Isto é, a probabilidade de que um número comece com o algarismo d=1 é de:

$$\text{Prob}\ (1) = \log\left(1 + \frac{1}{1}\right) = \log 2 = 0{,}30103$$

Para d=2:

$$\text{Prob}\ (2) = \log\left(1 + \frac{1}{2}\right) = \log 1{,}5 = 0{,}17609$$

A probabilidade também pode ser calculada para as posições seguintes de um número pela seguinte fórmula geral:

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

onde D_1 é o primeiro dígito, D_2 é o segundo dígito, sendo D_k é k-ésimo dígito. Por exemplo, a probabilidade de que um número inicie por 175 é de:

$$\text{Prob}\,(D_1 = 1, D_2 = 7, D_3 = 5) = \log\left[1 + \frac{1}{175}\right] = 0{,}00247$$

Para calcularmos a probabilidade de que o 2º dígito de um número seja 3, por exemplo, teríamos que calcular pela fórmula logo acima a soma das probabilidades do número começar por 13, 23, 33, 43, 53, 63, 73, 83 e 93. Ou seja:

$$\text{Prob}\ (D_2 = 3) = \sum_{j=1}^{9} \log\left[1 + \frac{1}{10 \cdot j + 3}\right] = 0{,}10433$$

Por meio da formulação apresentada, podemos construir a tabela abaixo com a probabilidade prevista pela Lei Newcomb-Benford até o quarto dígito. A tabela demonstra que a distribuição uniforme apenas passa a ser plausível a partir da quarta posição. A média tenderá a 4,5, valor esperado para uma distribuição uniforme, e a variância aproximar-se-á assintoticamente de 8,25. Entretanto, do primeiro ao terceiro dígito, predomina uma distribuição menos aderente a uniforme, de forma que testes de distribuição de Benford possuem mais chances de sucesso.

Freqüências esperadas baseadas na Lei de Newcomb-Benford				
Algarismo	1ª posição	2ª posição	3ª posição	4ª posição
0	-	0,11968	0,10178	0,10018
1	0,30103	0,11389	0,10138	0,10014
2	0,17609	0,10882	0,10097	0,10010
3	0,12494	0,10433	0,10057	0,10006
4	0,09691	0,10031	0,10018	0,10002
5	0,07918	0,09668	0,09979	0,09998
6	0,06695	0,09337	0,09940	0,09994
7	0,05799	0,09035	0,09902	0,09990
8	0,05115	0,08757	0,09864	0,09986
9	0,04576	0,08500	0,09827	0,09982

3. QUANDO APLICAR A LEI NEWCOMB-BENFORD

Como já mencionado, a Lei Newcomb-Benford tem aplicações interessantes à auditoria e à contabilidade. Por um teste baseado nessa lei, pode-se detectar, por exemplo, manipulação de valores em balanços contábeis, seja por inserção, remoção ou alteração de registros. No entanto, as pesquisas recomendam cautela nas conclusões respaldadas pela lei, no sentido de identificar corretamente quais tipos de dados se prestam à utilização da distribuição de Benford.

Durtschi et. al. (2004, p.23-24) comentam sobre quais características identificam oportunidade para o uso da referida lei e quais aquelas que a afastam. Afirmam que, em geral, dados relacionados à contabilidade obedecem à distribuição de Benford. Contas de receita e de pagamento, que são resultado de multiplicação de quantidades e preços, se amoldam bem à Lei Newcomb-Benford. Mas os mesmos autores também alertam para exceções da Lei Newcomb-Benford em dados contábeis. Tais exceções correspondem a números que podem sofrer deliberadamente influência humana (como saques em caixa eletrônico e preços de serviços e produtos) ou aqueles números sequenciais, como números de cheque e de ordem de compra. Números sequenciais tendem a ser números distribuídos uniformemente, bem diferente do perfil de Benford. Em relação aos preços, é prática comercial considerar a barreira psicológica de se fixar preços na casa dos noves sem arredondá-los para o inteiro superior. Números dessa natureza apresentam um perfil de distribuição parabólico em torno dos números mais prováveis. Registros contábeis que possuem restrição intervalar, isto é, que variam forçadamente entre limites máximo e mínimo pré-estabelecidos, também não se distribuem segundo Benford. Por exemplo, carteiras de investimento com *guidelines* ou conjunto de ativos que em razão da baixa de materialidade se exige limite mínimo para que sejam registrados contabilmente não teriam distribuição de Benford.

Seria interessante, preliminarmente aos testes, identificar se um particular conjunto de dados apresenta a distribuição de Benford, com vistas à sua aplicação, haja vista que, ao aplicarmos equivocadamente o teste a dados incompatíveis com a citada distribuição, ele poderá acusar uma fraude inexistente. Wallace (2002 apud DURTSCHI et. al., 2004, p.24) propõe que, se a média de um determinado conjunto de dados é maior que a sua mediana e se a sua dispersão for positiva, será aplicável o teste de Benford. Intuitivamente, percebemos que é esperado que a média dos dados deva ser maior que a sua mediana, já que os valores menores são mais prováveis na distribuição de Benford. "No entanto, o método de identificação sugerido esbarra em uma natural inconsistência, pois um banco de dados com uma quantidade suficiente de números manipulados, que deveria ser investigado, poderia apresentar uma resposta negativa a tal teste de identificação." (DURTSCHI et. al. 2004, p.24). Portanto, a identificação deve ter embasamento também no exame de outras características, algumas delas já citadas. Entendemos que a análise de Benford deve servir como mais uma ferramenta importante e como apoio às demais análises convencionais de controle interno.

As características a serem observadas são assim resumidas na tabela abaixo:

Quando a análise de Benford é ou não é provavelmente útil	
Quando a análise de Benford é provavelmente útil	**Exemplos**
Conjunto de números que resultam de combinações de números – resultados advindos de duas distribuições	Receitas (preço x quantidade vendida) Despesas (preço x quantidade comprada)
Dados referentes ao nível de transação	Pagamentos, vendas, despesas
Em grande conjunto de dados – quanto mais observações, melhor	Transações do ano inteiro
Registros que parecem se amoldarem – quando a média do conjunto de dados é maior do que a mediana e a dispersão é positiva	Maioria dos conjuntos de dados contábeis
Quando a análise de Benford não é provavelmente útil	Exemplos
Conjunto de dados envolve números sequenciais	Números de cheques, números de notas fiscais, CEPs
Números que são influenciados pelo pensamento humano	Preços sem arredondamento com noves, saques em caixas eletrônicos
Registros com uma grande quantidade de números específicos	Uma conta que recebe apenas um pequeno número de valores
Registros com mínimo ou máximo embutidos	Números que apenas são registrados pela sua alta materialidade, guidelines em carteiras de investimento
Onde a transação não é registrada	Propina, suborno, roubos, caixa dois

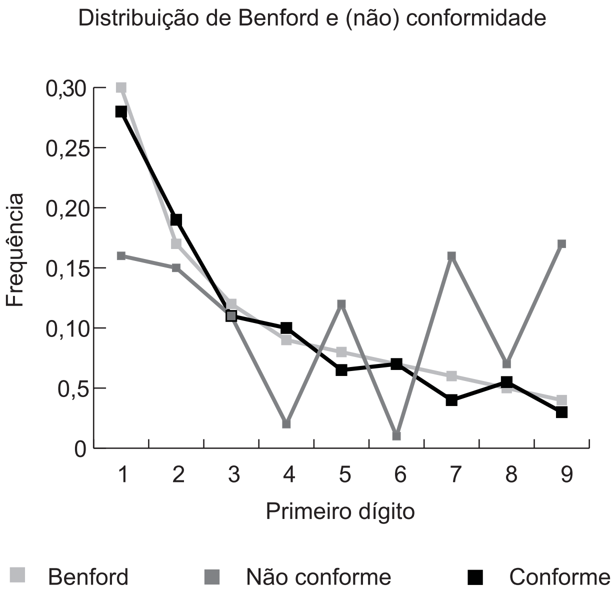
Fonte: Durtschi et. al. (2004, p.24, tradução nossa)

Finalmente, há de ser ter em mente que dados gerados aleatoriamente, como loterias, assim como datas, não se prestam à análise de Benford, pois não são processos de contagem. Aliás, números escolhidos pelo homem apresentam, em geral, uma aleatoriedade que foge significativamente da distribuição de Benford, fato que proporciona a sua utilidade em investigações e auditorias. Mas se um fraudador conseguir incluir e/ou remover registros mantendo a conformidade com a Lei Newcomb-Benford, naturalmente a fraude não poderá ser identificada pela lei.

4. COMO TESTAR A CONFORMIDADE COM A LEI NEWCOMB-BENFORD

Naturalmente, o teste da Lei Newcomb-Benford baseia-se na confrontação entre as freqüências observadas dos dados sob investigação e as freqüências esperadas de acordo com a distribuição de Benford estabelecida pela lei. Abaixo, podemos observar a distribuição teórica de Benford, assim como, dados hipotéticos que se enquadrariam (conforme) e outros que não se enquadrariam (não-conforme). Os dados não-conformes seriam candidatos a uma investigação mais apurada.

Figura 1- Gráfico da distribuição de Benford e de distribuições hipotéticas



Uma maneira mais rigorosa de aferir quão conforme estão os dados coletados é com o uso de testes estatísticos. Essa confrontação dos dados investigados com a distribuição predita por Newcomb e Benford, obviamente, sempre apresentará um desvio. Devemos nos perguntar se os desvios se devem à variação aleatória ou à manipulação fraudulenta. Em que magnitude dos desvios deve-se passar a desconfiar de uma existência de uma fraude?

O primeiro teste é o χ^2 de aderência, em que se compara as probabilidades observadas e a serem testadas com as probabilidades esperadas para todos os dígitos *d* em conjunto. Segundo Santos et. al. (2003, p.46) o teste para a primeira posição é realizado pela fórmula a seguir:

$$\chi^2_{8gl} = \sum_{d=1}^{9} \frac{\left[\log\left(1+\frac{1}{d}\right) - \text{Prob}\left(D_1 = d\right)\right]^2}{\log\left(1+\frac{1}{d}\right)} \times S$$

onde S é o tamanho da amostra.

Entretanto, o teste mais empregado é o z. Para um dígito particular, podemos usar a estatística z para examinar se ele foge do padrão predito pelo teste. Ou seja, podemos identificar se um dígito está ou não condizente com a freqüência esperada para esse dígito pela distribuição de Benford. Segundo Santos et. al. (2003, p.46) o teste para a primeira posição obedece à seguinte formulação:

$$z = \frac{\left|\text{Prob}\,(D_1 = d) - \log\left(1+\frac{1}{d}\right)\right| - \frac{1}{2S}}{\sqrt{\log\left(1+\frac{1}{d}\right)\left(1-\log\left(1+\frac{1}{d}\right)\right) \Big/ S}}$$

Este último teste apresenta a vantagem de identificar qual é dígito mais suspeito para averiguação, no entanto, pode apresentar maior quantidade de falsos resultados positivos que o teste qui-quadrado, isto é, apontando desconformidade com a distribuição de Benford onde não existe fraude. Ao contrário, devemos também sempre ter em mente que o teste pode não revelar fraudes, embora existam, como por exemplo, nos casos em que as transações não são em grande quantidade.

5. ANÁLISE DIGITAL NO ACL

O aplicativo de auditoria ACL conta com análise digital fundamentada na distribuição de Benford. De acordo com o seu ***help***, "As ferramentas de análise digital como o comando Benford permitem que auditores e outros analistas de dados se concentrem em possíveis anomalias, em conjuntos de dados volumosos." (DURTSCHI et. al., 2009). O comando da análise de Benford pode ser acessado a partir do botão **Analisar** na barra de menus e selecionando-se em seguida o botão **Efetuar análise de Benford**.

Das estatísticas mencionadas anteriormente, o ACL produz a estatística z para cada dígito. É possível a análise até o dígito da sexta posição, no entanto, a partir do quarto dígito o comando de Benford envia o resultado para um arquivo. Já a inclusão do quinto e sexto dígitos acarreta muito tempo de processamento. O programa sugere que, quando mais de uma coluna de resultados foge aos limites estabelecidos pela distribuição de Benford, os dados podem ser anômalos, merecendo apuração mais detalhada.

A análise de Benford do ACL possui funcionalidades importantes. Como qualquer incursão estatística necessita amostras minimamente grandes, o programa alerta quando o tamanho da amostra é pequeno demais para a análise. Também, o programa alerta para melhor análise quando os valores positivos e negativos são analisados separadamente, fornecendo um filtro para a segregação entre eles. A análise de Benford desconsidera automaticamente os registros iguais a zero na estatística final, porém, informando o número de registros nessa situação. Da mesma forma, sempre considera apenas o primeiro dígito significativo à esquerda de cada registro, ou seja, ignorando os zeros à esquerda do registro. O programa utiliza como saída padrão dos resultados uma tabela, no entanto, pode ser enviado para um arquivo ou gráfico.

O ACL alerta que as ferramentas de análise digital "não provam a existência de um erro ou fraude, mas identificam itens que merecem estudo posterior, com base em fundamentos estatísticos. A análise digital complementa as ferramentas e técnicas analíticas existentes, e não deve ser usada isoladamente delas". (DURTSCHI et. al., 2009).

6. COMENTÁRIOS FINAIS

A análise de Benford é uma ferramenta auxiliar na pesquisa de anomalias numéricas que possam levantar suspeitas.Todavia, deve ser usado com parcimônia, principalmente quando não há certeza de que banco de dados sob exame segue as propriedades da distribuição de Benford. No entanto, afastado esse risco, é uma ferramenta preliminar bastante prática, sobretudo quando a quantidade de registros for volumosa, possibilitando concentrarmos os exames em uma amostra menor.

Contrastando com a sua grande aplicação internacional na detecção de fraudes, a Lei Newcomb-Benford ainda é bastante desconhecida no nosso país. Poucos trabalhos nacionais existem sobre o tema. Essa é uma oportunidade para que futuros trabalhos realizados em auditoria avancem na comprovação e na definição das aplicabilidades da Lei Newcomb-Benford no âmbito da administração pública de nosso país.

REFERÊNCIAS

ACL SERVICES LTD. Ajuda do ACL Versão 8.

DURTSCHI, Cindy; HILLISON, William; PACINI, Carl. Journal of Forensic Accounting. *The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data*, v. 5, p. 17-34. 2004. Disponível em: <http://www.auditnet.org/articles/JFA-V-1-17-34.pdf>. Acesso em: 12 ago.2009.

SANTOS, Josenildo dos; TENÓRIO, José Nelson Barbosa; SILVA, Luis Gustavo Cordeiro da. *Uma Aplicação da Teoria das Probabilidades na Contabilometria:* a Lei Newcomb-Benford como medida para análise de dados no campo da auditoria contábil. Unb Contábil, v. 6, n. 1, p. 35-54. 2003. Disponível em: < http://www.unbcontabil.unb.br/Volumes/v6n1/9a32.pdf> Acesso em: 12 ago.2009.