



Universidade de Brasília

Faculdade de Administração, Contabilidade e Economia – FACE

Graduação em Ciências Econômicas

O uso da Lei de Benford na auditoria de obras públicas: o caso do VLP

Renata Motta Café

Orientador: Maurício Soares Bugarin

Brasília, 2015

Resumo

Grandes obras públicas são necessárias ao desenvolvimento de um país. Dada a falta de coordenação dos agentes privados, o investimento governamental torna-se altamente necessário. A fim de prezar pela melhor alocação dos recursos escassos, os Tribunais de Contas realizam auditoria nas obras públicas. Visto que a auditoria representa em si um custo para o governo, é imprescindível que ela seja realizada de forma eficiente. O presente trabalho testa a utilização da Lei de Benford na etapa de planejamento da auditoria de uma obra pública do Distrito Federal, o VLP ou Expresso Sul. Acredita-se que dados que não sofreram manipulação humana seguem a distribuição de Benford. São apresentados e aplicados os testes: (i) primeiro dígito, (ii) segundo dígito, (iii) dois primeiros dígitos e (iv) teste da soma, de acordo com Nigrini (2012). Ainda, é apresentada e aplicada uma versão do algoritmo de Bugarin e Cunha (2015) para detecção das rubricas com maiores evidências de sobrepreço na planilha contratual do VLP. O algoritmo apresenta resultados bastante favoráveis ao uso da Lei de Benford na auditoria. São revelados 73,40% do superfaturamento apontado pelo TCDF em apenas 38,17% do valor total da obra pública. Acredita-se que o uso dessa ferramenta simples para a seleção da amostra confira maior rendimento à auditoria.

Palavras-chave: Lei de Benford; auditoria; detecção de fraude; VLP DF; BRT.

Abstract

Major public works are crucial for the development of a country. Given the lack of coordination of private agents, government investment becomes highly necessary. In order to achieve the better allocation of scarce resources, the Courts of Accounts perform auditing public works. Since the audit is itself a cost to the government, it is essential to perform efficiently. This study tests the use of Benford's Law in the planning stage of the audit of a public work of Distrito Federal, the BRT. It is believed that data that have not undergone human manipulation will follow the Benford distribution. The following tests described in Nigrini (2012) are presented and applied to data: (i) first digit, (ii) the second digit, (iii) the first two digits and (iv) sum test. Still, a version of Bugarin and Cunha algorithm (2015) for the detection of items with more evidence of overpricing is presented and tested. The results of the algorithm are very favorable to the use of Benford's Law in auditing. About 73.40% of the overcost appointed by TCDF are revealed in just 38.17% of the total amount by the algorithm. The use of this simple tool for the selection of the sample seems to provide better results to auditing.

Key words: Benford's Law; auditing; fraud detection; BRT.

Sumário

1. INTRODUÇÃO	6
2. A LEI DE BENFORD	7
3. O MÉTODO	11
3.1 Perfil dos Dados	11
3.2 Teste do Primeiro Dígito.....	12
3.3 Teste do Segundo Dígito.....	15
3.4 Teste dos Dois Primeiros Dígitos.....	17
3.5 Teste da Soma	20
3.6 Algoritmo	21
4. APLICAÇÃO DA LEI DE BENFORD À PLANILHA DE OBRA PÚBLICA E COMPARAÇÃO COM A AUDITORIA DO TCDF	24
4.1 Perfil dos Dados	25
4.2 Teste do Primeiro Dígito.....	26
4.3 Teste do Segundo Dígito	29
4.4 Teste dos Dois Primeiros Dígitos.....	31
4.5 Teste da Soma	36
4.6 Algoritmo de seleção de itens candidatos à auditoria.....	40
4.6.1 Primeira rodada.....	41
4.6.2 Segunda rodada.....	41
4.6.3 Terceira rodada	42
4.6.4 Quarta rodada	42
4.6.5 Quinta rodada.....	43
4.6.6 Sexta rodada	43
5. CONCLUSÃO.....	45
6. REFERÊNCIAS BIBLIOGRÁFICAS	46
6.1 Artigos e livros	46
6.2 Dados	48

6.3 Sítios eletrônicos 48

**7. APÊNDICE - ITENS DETECTADOS EM COMUM PELO ALGORITMO DE
BENFORD E PELO TCDF 48**

1. Introdução

Considera-se que o investimento em grandes obras públicas por parte do governo é necessário ao desenvolvimento e à competitividade de um país. Abundantes evidências existem que, se deixada para o setor privado, a provisão de bens públicos será inferior ao ótimo social. A necessidade de coordenação pelo agente público se coloca porque as pessoas têm incentivo para ser “caronas” no uso dos bens públicos, que são por natureza não-rivais e não-excludentes (Mueller, 2003). Isso leva à provisão desse tipo de bem pelo governo, que pode usar o poder coercitivo para cobrar impostos e obter uma melhora de Pareto.

Entretanto, é evidente que o Estado deve prezar pela melhor alocação dos recursos, dado que estes são escassos. Assim, para evitar o desperdício e o superfaturamento, são realizadas auditorias nas obras públicas.

Os Tribunais de Contas zelam pela legalidade, legitimidade, efetividade, eficácia, eficiência e economicidade na gestão dos recursos públicos. Dentre outras funções, realizam a fiscalização de obras e serviços de engenharia como forma de coibir a corrupção e reaver perdas.

As planilhas orçamentárias podem ser extensas e de difícil análise, ocupando semanas de trabalho do auditor na realização de uma análise de preço. Visto que o trabalho de auditoria representa em si um custo para o Estado, é imprescindível que ele seja realizado da forma mais eficiente possível.

O presente trabalho testa a utilização da Lei de Benford na etapa de planejamento da auditoria de uma obra pública. Nigrini (2012) apresenta testes construídos a partir da Lei de Benford e defende seu uso para nortear o trabalho do auditor, apontando onde parecem haver indícios de fraude. Seriam ferramentas simples para prover maior assertividade à auditoria.

A obra analisada consta no documento denominado “*Relatório de obras e serviços de engenharia com indícios de irregularidades graves*” divulgado pelo Tribunal de Contas do Distrito Federal (TCDF). Trata-se da implantação do Expresso Sul, que passa pelas

regiões administrativas Gama, Santa Maria e Plano Piloto, e ficou mais conhecido pelo acrônimo VLP - Veículo Leve sobre Pneus. Utilizou-se a planilha de custos da obra referente à 26ª medição, que data de janeiro de 2014, a mais recente disponibilizada e a mesma utilizada pela auditoria do Tribunal. A soma de custos da planilha totaliza a quantia de R\$ 648.774.183,21. De acordo com a Informação nº 25/2014 – NFO do TCDF, foi encontrado um superfaturamento de R\$ 139.404.455,95, resultante da soma do montante superfaturado de R\$ 105.441.823,55 e do prejuízo em potencial restante de R\$ 33.962.632,40.

Neste trabalho, é aplicado um algoritmo ligeiramente modificado daquele sugerido por Bugarin e Cunha (2015) para a seleção dos dígitos e, posteriormente, feita uma comparação com a auditoria do TCDF. Para a realização dos testes é utilizada a versão 2013 do programa *Excel*®.

2. A Lei de Benford

A Lei de Newcomb-Benford, ou simplesmente Lei de Benford, trata da frequência dos primeiros dígitos de dados aleatórios, que não sofreram intervenção humana. A aplicação é diversa: vale para fatos encontrados na literatura de várias ciências e também para questões ordinárias do cotidiano. Suas origens remontam ao século XIX, mas apenas recentemente ela foi utilizada como ferramenta de detecção de manipulação de dados, a partir dos trabalhos de Nigrini (1999; 2009; 2012). Diz-se que dados em conformidade com o previsto pela lei seguem a distribuição de Benford.

Newcomb (1881) nota que os 10 dígitos não aparecem com a mesma frequência nos números naturais, a partir da observação de que as últimas páginas da tabela logarítmica são progressivamente menos utilizadas que as primeiras. No que concerne aos primeiros dígitos, o 1 aparece com maior frequência, de aproximadamente 30%, decaindo gradativamente até o 9, que aparece em cerca de 4,58% dos casos.

A Tabela 1 abaixo apresenta a probabilidade de ocorrência dos dois primeiros dígitos significantes de um número natural, conforme consta no artigo seminal.

Tabela 1.

Frequência dos primeiros dígitos conforme apresentada por Newcomb (1881)

Dígito	Frequência do Primeiro Dígito	Frequência do Segundo Dígito
0	-	0,1197
1	0,3010	0,1139
2	0,1761	0,1088
3	0,1249	0,1043
4	0,0969	0,1003
5	0,0792	0,0967
6	0,0669	0,0934
7	0,0580	0,0904
8	0,0512	0,0876
9	0,0458	0,0850

Newcomb (1881) ressalta que no caso do terceiro dígito, a probabilidade seria aproximadamente a mesma para cada dígito e, a partir do quarto dígito a diferença de frequência seria inapreciável. O autor apresenta o resultado sem, no entanto, apresentar evidências matemáticas ou empíricas. O trabalho não contou com muita notoriedade até meio século após sua publicação, com as descobertas de Benford.

A partir da mesma observação de que as primeiras páginas da tabela logarítmica são mais utilizadas que as últimas, mas aparentemente de maneira independente, Benford (1938) realiza um teste empírico. A compilação de 20.000 primeiros dígitos de fontes diversas indica distribuição logarítmica para os primeiros dígitos dos números compostos por 4 ou mais dígitos. Os números sem nenhuma relação aparente apresentam maior conformidade com a distribuição que dados formais ou provindos de tabulações matemáticas, o que confere ao artigo o título de “Lei dos Números Anômalos”.

Benford (1938) apresenta uma análise mais avançada que revela que os dados numéricos diversos apresentam forte tendência de seguir séries geométricas. Se a série contém 3 ou mais dígitos, os primeiros formam distribuição logarítmica. Benford (1938) ainda constrói uma equação para a frequência dos primeiros dígitos em diferentes “ordens” de números (dezenas, centenas, etc). A equação é estendida para a frequência dos outros dígitos em números com múltiplos dígitos.

Apesar da Lei de Benford ter sido descoberta de maneira empírica através do exame de tabelas com dados numéricos, há uma série de modelos matemáticos que explicam o fenômeno com base teórica. Raimi (1976) apresenta uma revisão bibliográfica de provas da Lei de Benford que, no entanto, não explicam consistentemente o fenômeno, e conclui que a resposta permanece obscura.

A prova mais robusta e aceita atualmente é a apresentada por Hill (1995), que deriva a Lei de Benford a partir da hipótese de invariância de base. A prova apresentada por Hill baseia-se no fato de que os números cujos primeiros dígitos seguem a distribuição de Benford são gerados a partir da combinação de outras distribuições, e ainda que uma distribuição isolada não siga perfeitamente a distribuição de Benford, a combinação das distribuições o fará.

Anos antes de Hill, Pinkham (1961) afirma que se uma lei rege a distribuição dos dígitos, ela deve ser invariante à escala. Tal fato permite que uma amostra de dados não manipulada seja testada independente da unidade de medida utilizada. Boyle (1994) revela que números de diferentes fontes que sofram operação matemática com potências inteiras seguem a distribuição de Benford.

Durtschi, Hillison e Pacini (2004) resumem os casos em que a Lei de Benford é relevante e aplicável: (i) quando os dados provêm de duas ou mais distribuições, como em análises de custo total, isto é, preços multiplicados por quantidades, (ii) quando os dados referem-se à resultados micro, como as vendas de uma empresa, (iii) quando o número de dados é grande e (iv) quando a base de dados é assimétrica, de modo que a média é maior que a mediana. Por outro lado, a lei não será relevante caso os números (i) sejam pré-concebidos ou sofram influência do comportamento humano, (ii) possuam números específicos que se repetem, ou (iii) possuam máximos ou mínimos pré-

determinados. Assim, não se espera que números de telefone, endereços e folhas de pagamento sigam a distribuição de Benford.

Varian (1972) foi um dos primeiros a sugerir o uso da Lei de Benford em auditoria. Neste ponto, vale ressaltar que a conformidade do conjunto de dados com a Lei de Benford não necessariamente implica autenticidade, mas a não-conformidade deve levantar algum nível de suspeita. Logo após, Carslaw (1988) e Thomas (1989) realizaram trabalhos visando inicialmente detectar a manipulação de dados via arredondamentos excessivos no balanço de empresas. Seriam arredondamentos para cima, no caso de entradas de dinheiro, e para baixo no caso de saídas de caixa.

Mittermaier e Nigrini (1997) defendem o uso de uma técnica de análise de dígitos baseada na Lei de Benford no estágio de planejamento da auditoria. A técnica compara a frequência do dígito (para números com mais de um dígito) com a frequência prevista pela Lei de Benford e levanta suspeitas de fraude ou irregularidades para aqueles que não estão em conformidade.

Mittermaier e Nigrini (1997) e Nigrini (2012) apresentam uma série de testes que possibilitam um uso extensivo da Lei de Benford em auditorias. No presente trabalho utilizaremos os testes: (i) primeiros dígitos, (ii) segundos dígitos, (iii) dois primeiros dígitos e (iv) teste da soma.

Cunha e Bugarin (2015) aplicam os testes propostos por Nigrini (2012) à auditoria de uma obra pública brasileira, o estádio de futebol Maracanã e encontram resultados admiráveis: os dígitos selecionados pelos testes correspondem a mais de 70% do volume de fraudes encontrado pelo Tribunal de Contas da União (TCU). Bugarin e Cunha (2015) aplicam os testes a mais uma obra pública, o estádio Arena Amazônia e também obtém bons resultados. Ainda, Bugarin e Cunha (2015) propõem um algoritmo para a seleção dos dígitos a serem auditados. O método proposto é utilizado no presente trabalho com os dados de uma obra pública do Distrito Federal, o chamado VLP.

3. O método

Esta seção apresenta a estrutura básica dos testes detalhados em Nigrini (2012) e uma versão do algoritmo proposto por Bugarin e Cunha (2015) para seleção dos dígitos para auditoria. A título de ilustração, os testes são aplicados a uma base de dados contendo as áreas dos 5.561 municípios brasileiros, em quilômetros quadrados, conforme conta na Resolução nº 5, de 10 de outubro de 2002, do Instituto Brasileiro de Geografia e Estatística (IBGE).

3.1 Perfil dos Dados

Nigrini (2012) ressalta a importância de verificar a viabilidade da aplicação dos testes baseados na Lei de Benford por meio do exame do perfil dos dados. Consiste na contagem e soma dos números em cinco categorias (igual ou acima de 10,00; entre 0 e 10; igual a 0; entre 0 e -10; abaixo de menos 10), além do levantamento de valores muito altos (acima de 100.000,00) ou muito baixos (até 50,00). As categorias podem ser adaptadas de acordo com o tipo de dados.

A verificação do perfil dos dados permite que o usuário entenda melhor a composição da base de dados, evita erros e detecta problemas sérios que apontem para o abandono da abordagem de planejamento da auditoria por Benford. A seguir, procedemos com a verificação do perfil da base de dados do IBGE referente às áreas dos municípios brasileiros.

Tabela 2.

Perfil dos dados para a base de áreas dos municípios brasileiros

PERFIL DOS DADOS				
Detalhamento	Contagem	% do Total	Soma	% do Total
10,00 ou acima	5.559	99,96	8521857,68	100,00
De 0,01 a 9,99	2	0,04	0,00	0,00
igual à zero	0	0,00	0,00	0,00
-0,01 a -9,99	0	0,00	0,00	0,00
-1000 ou abaixo	0	0,00	0,00	0,00

Soma:	5.561	100,00	8521857,68	100,00
Valores baixos				
0,01 to 50,00	0	0,00	0,00	0,00
Valores muito altos				
100.000 ou acima	4	0,07	498959,55	5,86

O exame do perfil dos dados não revela problemas com a base. Apenas 2 das 5561 observações estão abaixo de 10, o que não provoca viés na realização do teste dos dois primeiros dígitos se esses forem ignorados. Conforme esperado, não há valores muito baixos, negativos ou iguais a zero. Além disso, os quatro valores muito altos encontrados correspondem a 5,86%, o que não gera viés significativo para o teste da soma.

3.2 Teste do Primeiro Dígito

O teste do primeiro dígito encontra-se dentre os testes primários descritos por Nigrini (2012) e é um primeiro teste de conformidade geral dos dados com a distribuição de Benford.

Conforme consta na Tabela 1 deste trabalho, a frequência esperada dos primeiros dígitos pela Lei de Benford segue a seguinte expressão matemática:

$$\text{Prob}(\text{primeiro dígito} = D_1) = \log_{10}\left(1 + \frac{1}{D_1}\right), D_1 = 1, \dots, 9.$$

A Figura 1 a seguir mostra a frequência dos primeiros dígitos para a base de dados do IBGE relativa à área, em quilômetros quadrados, dos 5.561 municípios brasileiros.

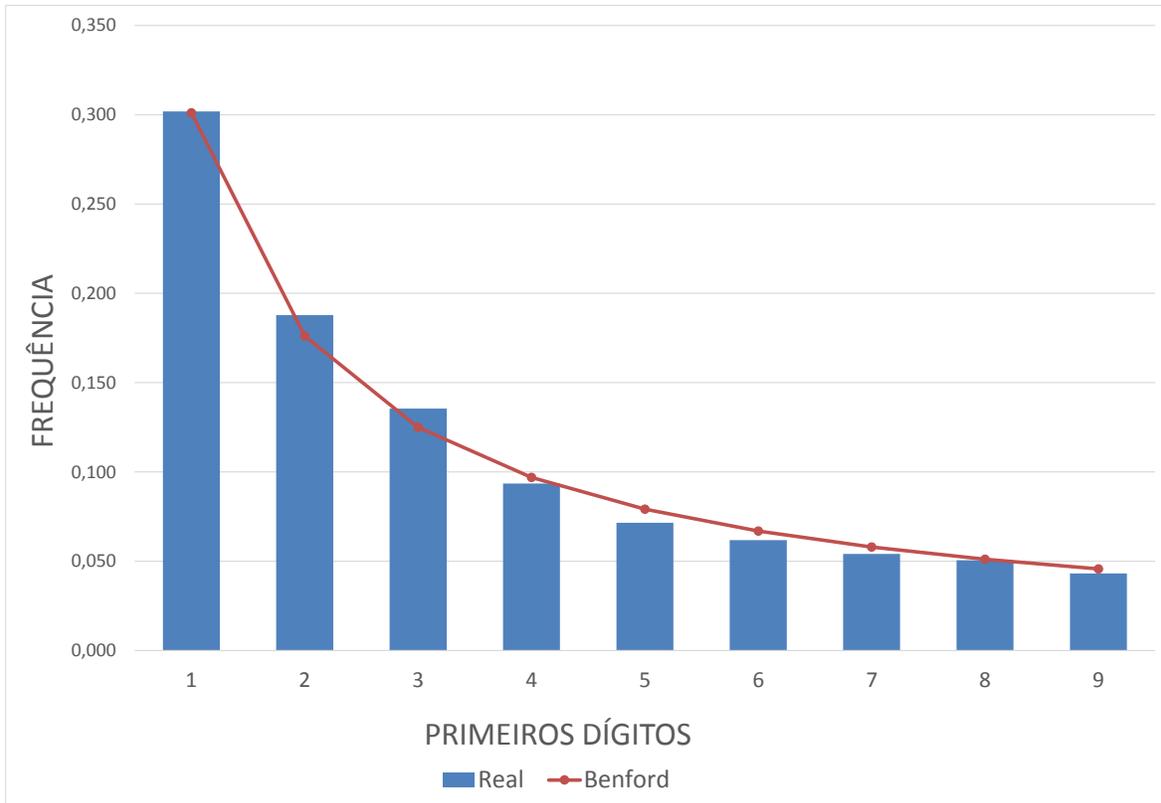


Figura 1. Frequência dos primeiros dígitos na base de dados de áreas dos municípios brasileiros e a comparação com o previsto por Benford.

Em seguimento a uma primeira análise de conformidade visual através do gráfico, comparam-se as frequências observadas com as esperadas para cada dígito através do cálculo das estatísticas Z. A estatística Z leva em conta a magnitude absoluta da diferença (entre as frequências reais e esperadas por Benford), o tamanho da base de dados e a frequência esperada, conforme a fórmula a seguir:

$$(i) \quad Z = \frac{|F_{Ri} - F_{REi}| - \frac{1}{2N}}{\sqrt{\frac{F_{REi}(1 - F_{REi})}{N}}}$$

na qual N denota o número de observações válidas, FR_i a frequência relativa encontrada para o primeiro dígito i e FRE_i a frequência relativa esperada para i de acordo com a Lei de Benford. O último termo do numerador $\frac{1}{2N}$ é usado apenas se for menor que o primeiro ($|FR_i - FRE_i|$), para correção de continuidade. Compara-se a estatística Z encontrada com o valor tabulado para um nível de significância. Se a estatística Z for menor que o limite tabulado, não é possível rejeitar a hipótese de conformidade com a distribuição de Benford. Caso contrário, rejeita-se a hipótese de conformidade. Para os dados de áreas, todas as estatísticas Z ficaram abaixo do valor crítico de 1,96 para 5% de significância, o mais utilizado.

Há duas medidas sugeridas em Nigrini (2012) para verificar a conformidade geral dos primeiros dígitos com a distribuição de Benford. A primeira é o desvio médio absoluto (DMA), estatística calculada a partir da média das frequências relativas e esperadas para cada dígito descrita pela fórmula abaixo.

$$(ii) \quad DMA = \frac{\sum_{i=1}^9 |FR_i - FRE_i|}{9},$$

onde FR_i refere-se à frequência relativa encontrada para o primeiro dígito i e FRE_i refere-se à frequência relativa esperada para i de acordo com a Lei de Benford.

Nigrini (2012) propõe o seguinte critério para avaliar a conformidade da base de dados de acordo com o teste DMA para os primeiros dígitos: se a estatística estiver entre 0 e 0,006 tem-se conformidade estrita; entre 0,006 e 0,012, conformidade aceitável; 0,012 a 0,015 conformidade marginalmente aceitável e, por fim, se for maior que 0,015 tem-se não-conformidade. Para a base de dados ilustrada, o valor da estatística DMA é 0,00512, sinalizando conformidade estrita.

A outra medida sugerida é calcular a estatística qui-quadrado, que compara um conjunto de resultados obtidos com esperados. O resultado esperado em questão e, logo, a hipótese nula, é a conformidade com a distribuição de Benford. A estatística qui-quadrado é calculada pela fórmula a seguir.

$$(iii) \quad \chi^2 = \sum_{i=1}^k \frac{(C-CE)^2}{CE},$$

onde k é cada categoria (neste caso, cada um dos 9 dígitos), C é a contagem do números pertencentes à categoria e CE a contagem esperada por Benford para a categoria. O valor crítico da estatística qui-quadrado para 5% de significância e 8 graus de liberdade é 15,51. Se a estatística resultante for menor que o limite tabulado, não é possível rejeitar a hipótese de conformidade com a distribuição de Benford. Caso contrário, rejeita-se a hipótese de conformidade. Neste caso, diferente do resultado encontrado com o uso da estatística DMA, para a base de dados das áreas dos municípios encontra-se o valor 18,35 e a hipótese de conformidade é rejeitada, por uma pequena margem, demonstrando que o qui-quadrado é um teste bem mais excludente.

De qualquer maneira, de acordo com Nigrini (2012), o teste do primeiro dígito não costuma ser de muita utilidade para identificar os dados com indícios de fraude a serem auditados por dividir a base em poucas categorias. Costuma ser melhor para um conjunto de dados pequeno (até 300 observações) e para comparar duas bases de dados quanto a indícios de manipulação.

3.3 Teste do Segundo Dígito

A frequência esperada pela Lei de Benford para os segundos dígitos aparece na Tabela 1 e segue a seguinte expressão matemática:

$$\text{Prob}(\text{segundo dígito} = D_2) = \sum_{D_1=1}^9 \log\left(1 + \frac{1}{D_1 D_2}\right), \quad D_2 = 0,1,2,3,4,5,6,7,8,9.$$

As frequências da base de dados das áreas dos municípios brasileiros para o segundo dígito estão plotadas na Figura 2, a seguir.

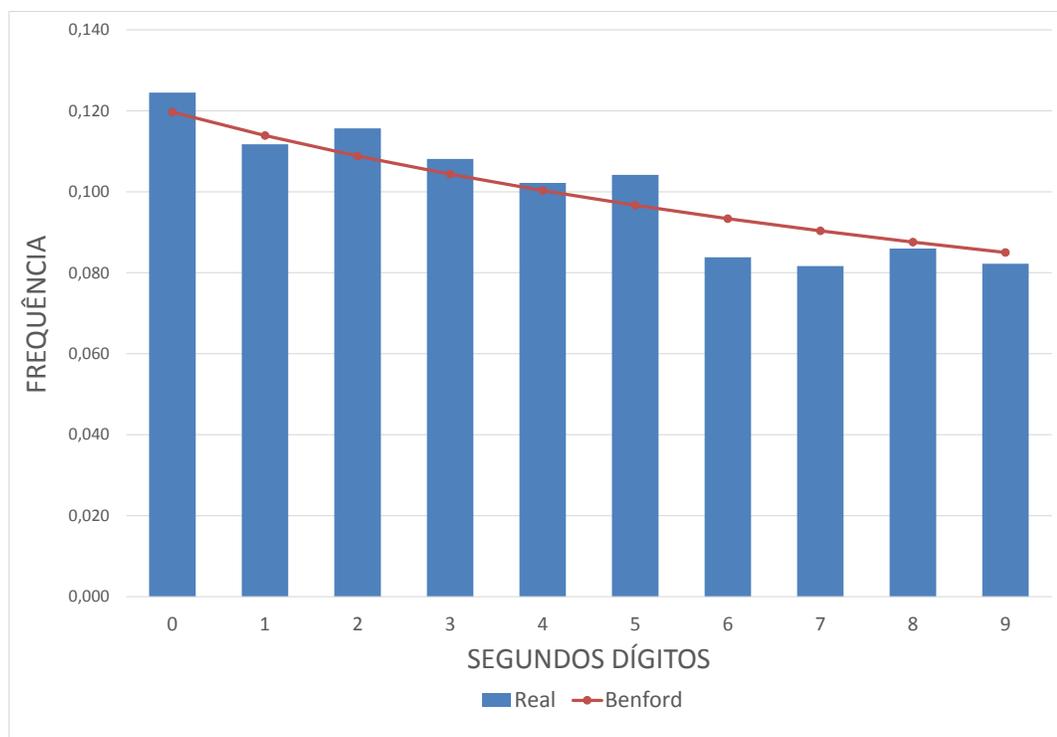


Figura 2. Frequência dos segundos dígitos na base de dados de áreas dos municípios brasileiros e a comparação com o previsto por Benford.

Novamente, pode-se calcular as estatísticas Z para cada dígito e as medidas de conformidade geral DMA e qui-quadrado. Para a base de dados das áreas dos municípios brasileiros, os dígitos 6 e 7 possuem estatística Z maior que 1,96, respectivamente, 2,422 e 2,234, medida de DMA igual a 0,0049 e estatística qui-quadrado igual a 18,59.

Os limites para conformidade propostos por Nigrini (2012) para avaliar o DMA dos segundos dígitos são ligeiramente diferentes: se a medida estiver entre 0 e 0,008 tem-se conformidade estrita; entre 0,008 e 0,010, conformidade aceitável; 0,010 a 0,012 conformidade marginalmente aceitável e, por fim, se for maior que 0,012 o diagnóstico é de não-conformidade. O valor crítico da estatística qui-quadrado a um nível de significância de 5% para 9 graus de liberdade é 16,92. Novamente, para a base de dados ilustrada, o valor da estatística qui-quadrado sugere não-conformidade enquanto o DMA sugere conformidade estrita com a distribuição de Benford.

O teste do segundo dígito é bom para detectar arredondamentos excessivos na base de dados, se for constatada a presença excessiva de dígitos 0 e 5 (Carslaw 1988; Thomas 1989).

3.4 Teste dos Dois Primeiros Dígitos

O teste dos dois primeiros dígitos é um teste mais focado e, portanto, mais apropriado para detectar as categorias de dígitos com evidências de fraude que os anteriores. A base de dados é dividida em 90 categorias (10, 11, ..., 98, 99) e é feita uma comparação da frequência de cada uma delas com a frequência relativa esperada por Benford. A frequência relativa para os dois primeiros dígitos segue a seguinte expressão matemática:

$$\text{Prob}(\text{dois primeiros dígitos} = D_1D_2) = \log\left(1 + \frac{1}{D_1D_2}\right), D_1D_2 = 10, 11, 12, \dots, 98, 99.$$

A Tabela 3 apresenta essas frequências relativas, também traçadas em vermelho na Figura 3.

Tabela 3.**Frequência esperada para os dois primeiros dígitos**

Primeiros dois dígitos	Frequência esperada por Benford	Primeiros dois dígitos	Frequência esperada por Benford	Primeiros dois dígitos	Frequência esperada por Benford
10	0,041	40	0,011	70	0,006
11	0,038	41	0,010	71	0,006
12	0,035	42	0,010	72	0,006
13	0,032	43	0,010	73	0,006
14	0,030	44	0,010	74	0,006
15	0,028	45	0,010	75	0,006
16	0,026	46	0,009	76	0,006
17	0,025	47	0,009	77	0,006
18	0,023	48	0,009	78	0,006
19	0,022	49	0,009	79	0,005
20	0,021	50	0,009	80	0,005
21	0,020	51	0,008	81	0,005
22	0,019	52	0,008	82	0,005
23	0,018	53	0,008	83	0,005
24	0,018	54	0,008	84	0,005
25	0,017	55	0,008	85	0,005
26	0,016	56	0,008	86	0,005
27	0,016	57	0,008	87	0,005
28	0,015	58	0,007	88	0,005
29	0,015	59	0,007	89	0,005
30	0,014	60	0,007	90	0,005
31	0,014	61	0,007	91	0,005
32	0,013	62	0,007	92	0,005
33	0,013	63	0,007	93	0,005
34	0,013	64	0,007	94	0,005
35	0,012	65	0,007	95	0,005
36	0,012	66	0,007	96	0,005
37	0,012	67	0,006	97	0,004
38	0,011	68	0,006	98	0,004
39	0,011	69	0,006	99	0,004

Para fins de ilustração, a Figura 3 apresenta a frequência dos dois primeiros dígitos para a base de dados das áreas dos municípios brasileiros.

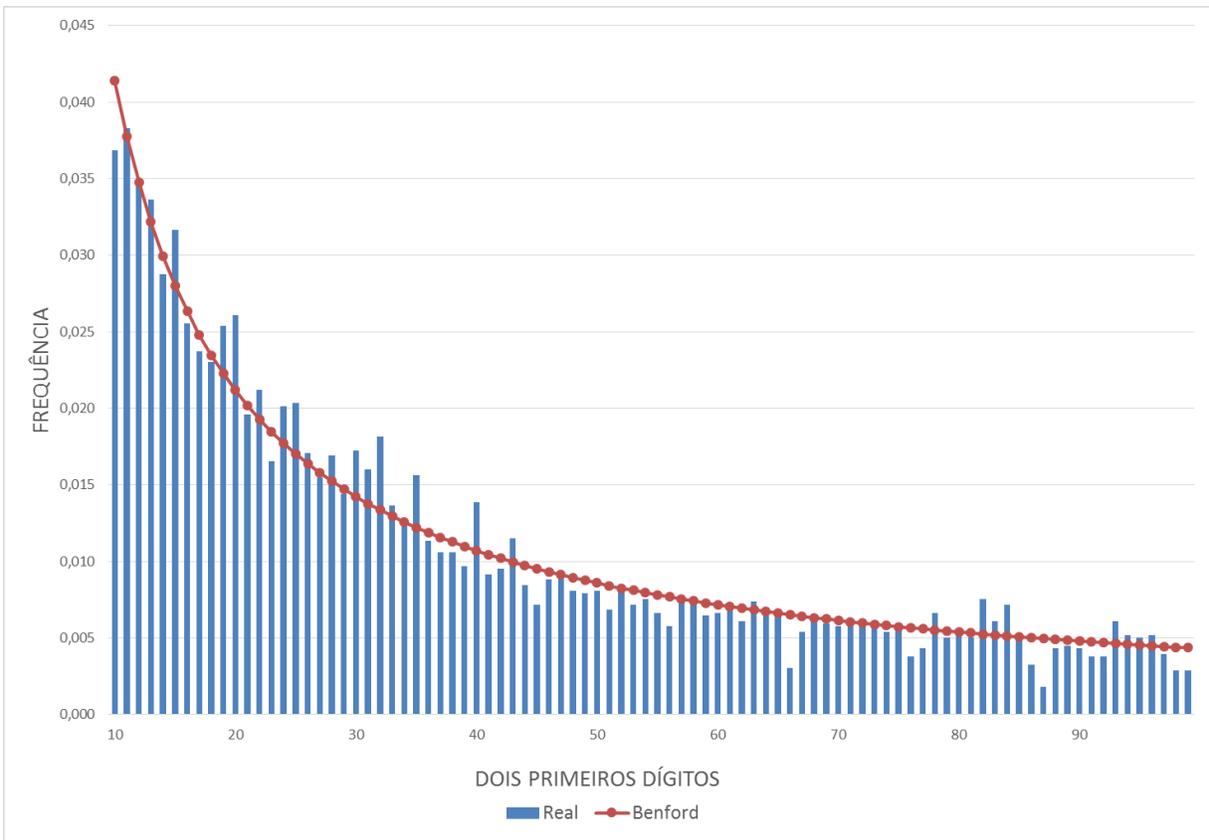


Figura 3. Frequência dos dois primeiros dígitos na base de dados de áreas dos municípios brasileiros e a comparação com o previsto por Benford.

Nigrini (2012) sugere que se até 5 categorias de dígitos dentre as 90 não obedecerem a Lei de Benford, não há evidência forte de manipulação. Na base de dados de áreas analisada, ao calcular as estatísticas Z e adotando o valor crítico de 1,96 (5% de significância), encontramos oito categorias de dígitos que não estão em conformidade, são elas: 20, 32, 35, 40, 66, 82, 84, 87. Por esse primeiro critério, a base de dados não apresenta conformidade com Benford.

Há ainda dois critérios sugeridos por Nigrini (2012) para avaliar a conformidade geral com a Lei de Benford a partir do teste dos dois primeiros dígitos: as estatísticas qui-quadrado e DMA. O valor crítico da estatística qui-quadrado a 5% de significância e 89 graus de liberdade é 112,02. Os limites para avaliar o DMA dos dois primeiros dígitos para conformidade propostos por Nigrini (2012) são: entre 0 e 0,0012 conformidade

estrita; entre 0,0012 e 0,0018, conformidade aceitável; 0,0018 a 0,0022 conformidade marginalmente aceitável e maior que 0,0022 não-conformidade.

A estatística qui-quadrado para a base de dado das áreas dos municípios brasileiras é 124,35 e o DMA 0,0012. A primeira estatística sugere não-conformidade enquanto a segunda se situa no limite entre conformidade estrita e aceitável.

3.5 Teste da Soma

O teste da soma é um teste avançado proposto por Nigrini (2012) para detectar números excessivamente grandes na base de dados. Trata-se de uma modificação do teste usual para os dois primeiros dígitos, no qual os dados são divididos em 90 categorias (10, 11, ..., 98, 99) e as observações são somadas para cada grupo. O resultado esperado é que cada classe de dígitos some aproximadamente o mesmo montante, isto é, $\frac{1}{90} = 0,011$ ou 1,1% do total da soma de todos os números da base.

Apesar de matematicamente provada, empiricamente os dados não costumam conformar com a expectativa teórica. Ainda assim, quando a soma da categoria é anormalmente grande, os dados merecem maior escrutínio. A Figura 4 traça o teste da soma na base de dados das áreas dos municípios brasileiros.

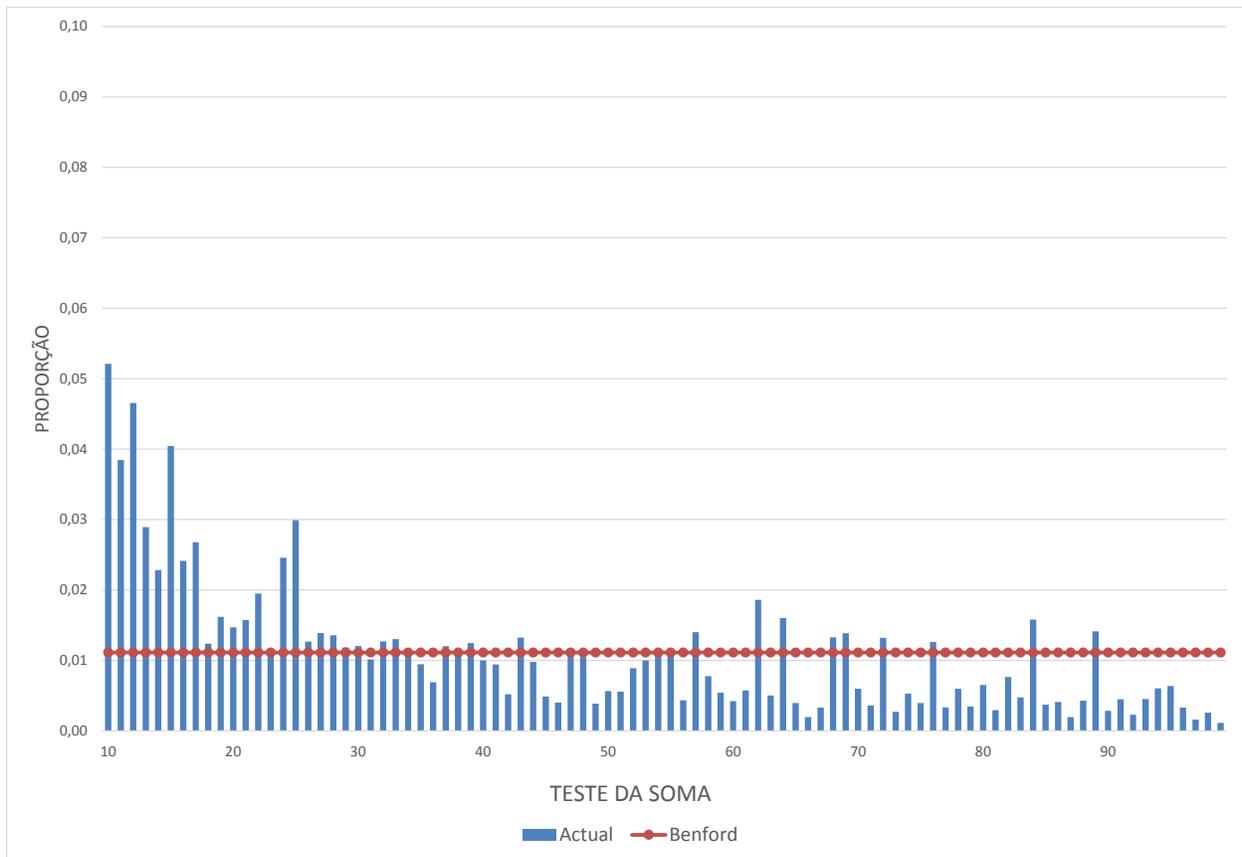


Figura 4. Teste da soma aplicado à base de dados de áreas dos municípios brasileiros e a comparação com o previsto por Benford.

O teste da soma é utilizado em conjunto com o teste dos dois primeiros dígitos para detectar categorias de dígitos relevantes para a auditoria no algoritmo proposto por Bugarin e Cunha (2015).

3.6 Algoritmo

A auditoria de grandes obras é essencial para assegurar o bom uso dos recursos públicos, porém é igualmente um trabalho extenso e custoso para os Tribunais de Contas. Neste contexto, a etapa de planejamento da auditoria e seleção da amostra dos dados que passará por maior escrutínio é crucial. Quanto mais precisa for essa amostra,

melhores serão os resultados obtidos pela auditoria. Testes baseados na Lei de Benford podem ser úteis para determinar a amostra a ser auditada.

O algoritmo apresentado nesta seção baseia-se naquele proposto por Bugarin e Cunha (2015). Obtém-se uma seleção de rubricas com maiores evidências de manipulação humana a partir da combinação dos testes da soma e dos dois primeiros dígitos, dentro de uma porcentagem do total preestabelecida para auditoria.

O primeiro passo é estabelecer os valores iniciais dos parâmetros-chave. São eles:

- i. O parâmetro σ estabelece a porcentagem da soma total (T) que será auditada;
- ii. O parâmetro ε estabelece a precisão da regra de parada, isto é, quão próxima a soma da amostra selecionada se encontra do parâmetro σ para a porcentagem da planilha a ser auditada;
- iii. O parâmetro λ estabelece o nível de significância usado no teste dos dois primeiros dígitos. Seguindo Nigrini (2012) e o habitual de trabalhos empíricos, λ se inicia em 5%;
- iv. O parâmetro μ estabelece o limite a ser usado como critério de seleção de categorias de dígitos no teste da soma. Bugarin e Cunha (2015) sugerem μ inicial de 100%;
- v. O parâmetro δ estabelece o ajuste a ser feito no parâmetro μ de seleção de categorias de dígitos através do teste da soma. Bugarin e Cunha (2015) sugerem δ de 25%.

Roda-se o teste dos dois primeiros dígitos aplicando-se o critério de significância λ e selecionam-se as categorias de dígitos correspondentes.

Roda-se também o teste da soma utilizando o limite estabelecido μ e selecionam-se as categorias de dígitos cuja frequência relativa se encontram acima de $0,011(1 + \mu)$, recordando que 0,011 é a frequência prevista por Benford.

A fim de garantir a relevância da amostra selecionada, é feito o confronto entre as categorias de dígitos selecionadas pelos testes dos dois primeiros dígitos e da soma. Se utilizado apenas o teste dos dois primeiros dígitos, uma categoria de dígitos pode ser apontada simplesmente por conter poucas observações na amostra, o que não a torna uma boa candidata para a auditoria.

O confronto sugerido por Bugarin e Cunha (2015) consiste na seleção de valores cujos dígitos revelem alta frequência em ao menos um dos testes. Isto é, caso a categoria de dígitos seja selecionada pelo teste da soma e/ou apresente uma frequência maior que a esperada por Benford no teste dos dois primeiros dígitos, ela é selecionada na etapa do confronto.

O presente trabalho realiza de maneira distinta a etapa do confronto. Aqui, optou-se por uma seleção mais estrita, isto é, a categoria de dois dígitos somente será selecionada caso seja indicada não conformidade nos dois testes.

Após o confronto, verifica-se o valor total selecionado para auditoria (S) e procede-se à análise quanto a realizar mais uma rodada dos testes. O critério de decisão neste ponto é dado pela comparação entre a soma selecionada para auditoria até o momento (S) e a soma pretendida para ser auditada (σT), resumido pela fórmula:

$$(iv) \quad p = \frac{S - \sigma T}{\sigma T}.$$

Compara-se o módulo do p obtido com o critério ε . Se $|p| \leq \varepsilon$, verifica-se se o valor S está abaixo de σT . Se estiver, o método foi eficaz em obter a amostra a ser auditada dentro dos critérios pretendidos. Caso contrário, o método não foi capaz de sinalizar um número alto o suficiente de categorias de dígitos para corresponder a um custo próximo ao pretendido, σT .

Se $|p| > \varepsilon$ e $S < T$, sugere-se uma próxima rodada de testes. Nesta segunda rodada, o parâmetro λ , relativo ao teste dos dois primeiros dígitos é mantido em 5%, e o parâmetro μ relativo ao teste da soma deve ser diminuído em $\delta=25\%$, caindo para $\mu=75\%$. É feito o confronto e os dígitos são selecionados. Após, o valor p , calculado a partir de (iv) é comparado novamente com ε para decidir quanto a novas rodadas.

Se $|p| > \varepsilon$ e $S > T$, a nova rodada deve aumentar em μ em $\delta=25\%$. É o caso em que a rodada de dígitos resulta em uma amostra selecionada para a auditoria maior do que a pretendida.

Procedem-se com demais rodadas enquanto $|p| > \varepsilon$ e $S < T$. Quando o valor do parâmetro μ for igual a zero, pode-se ajustar λ para 10%. Este é a última flexibilização permitida.

A Tabela 4 resume a utilização dos parâmetros a cada rodada para o caso usual, em que S é sempre menor do que T.

Tabela 4.

Tabela-resumo dos parâmetros utilizados nos testes a cada rodada

PRIMEIRA RODADA	
$\lambda = 5\%$	$\mu = 100\%$
SEGUNDA RODADA	
$\lambda = 5\%$	$\mu = 75\%$
TERCEIRA RODADA	
$\lambda = 5\%$	$\mu = 50\%$
QUARTA RODADA	
$\lambda = 5\%$	$\mu = 25\%$
QUINTA RODADA	
$\lambda = 5\%$	$\mu = 0\%$
SEXTA RODADA	
$\lambda = 10\%$	$\mu = 0\%$
FIM DO ALGORITMO	

4. Aplicação da Lei de Benford à planilha de obra pública e comparação com a auditoria do TCDF

A presente seção procede uma análise da planilha de custos uma obra pública com base na Lei de Benford. Serão aplicados os testes e o logaritmo descritos na seção anterior. Novamente, o intuito dos procedimentos adotados é apontar onde há maiores evidências de manipulação humana, indicando que os dados devem passar por maior escrutínio.

A obra em questão, referente à implantação do Expresso Sul (VLP), já foi analisada pelo TCDF, que encontrou superfaturamento vultoso. O valor da licitação assinado em contrato em 2009 era de R\$ 587.400.719,83 (data-base maio de 2009). O valor foi repactuado em 2011 e acrescentado um segundo termo aditivo que reduziu o valor total para R\$ 533.619.830,71 (data-base julho de 2010). Ao longo da execução, o contrato sofreu outros termos aditivos até 2012 que culminaram no valor de R\$ 648.774.183,21

(data-base julho de 2010), referente aos dados trabalhados¹. A planilha foi escolhida para a aplicação dos testes por ser a mesma auditada pelo TCDF na Informação nº 25/2014 – NFO. É denominada na sequência de planilha contratual do VLP. Ao final, é realizada uma comparação com a análise do Tribunal.

O TCDF utiliza o princípio de Pareto, também conceito como curva ABC, para a análise da planilha e seleção das rubricas a serem auditadas. Tal método consiste em ordenar os itens de acordo com seu custo total em ordem decrescente e auditar até 20% destes itens, a começar pelo de maior valor, de forma que o valor auditado some até 80% do total. A utilização de um método completamente independente pela auditoria do TCDF torna comparável a análise feita aqui com base na Lei de Benford.

Os testes sugeridos por Nigrini (2012) e a adaptação do algoritmo proposto por Bugarin e Cunha (2015) são aplicados na base dos preços totais de cada rubrica, isto é, preços multiplicados por quantidades. Como explicitado na seção 2 deste trabalho, a Lei de Benford é relevante e aplicável no caso em que os dados vêm de duas ou mais distribuições, como em análises de custo total (Durtschi, Hillison e Pacini; 2004).

4.1 Perfil dos Dados

Antes de realizar os testes, procede-se com o exame do perfil dos dados, detalhado na Tabela 5.

¹ O valor final contratado é maior, já que houve novos aditamentos após 2012.

Tabela 5.

Perfil dos Dados para a planilha contratual do VLP

PERFIL DOS DADOS				
Detalhamento	Contagem	% do Total	Soma	% do Total
10,00 ou acima	2.467	100,00	R\$ 648.774.183,21	100,00
De 0,01 a 9,99	0	0,00	0,00	0,00
igual à zero	0	0,00	0,00	0,00
-0,01 a -9,99	0	0,00	0,00	0,00
-10,00 ou abaixo	0	0,00	0,00	0,00

Soma:	2.467	100,00	R\$ 648.774.183,21	100,00
Valores baixos				
0,01 to 50,00	12	0,49	R\$ 467,73	0,00
Valores muito altos				
100.000 ou acima	563	22,82	R\$ 615.786.589,11	94,92

O exame do perfil dos dados para a planilha contratual do VLP não revela problemas com a base. Todos os valores possuem ao menos dois dígitos significantes, o que confere cobertura total dos dados nos testes de Benford. Conforme esperado, há pouquíssimos valores baixos, visto que se trata do custo total de cada rubrica. Não há valores negativos ou iguais a zero, que, no caso desta planilha representaria erro. É notável, porém, que 22,82% das rubricas correspondam à aproximadamente 95% da soma total².

4.2 Teste do Primeiro Dígito

O primeiro teste a ser feito consiste na comparação das frequências relativas de cada primeiro dígito na base de dados, denominada a seguir de “Real”, com a frequência prevista por Benford. A Tabela 6 apresenta as frequências relativas dos primeiros dígitos e a Figura 5 as ilustra visualmente.

² Esse perfil poderia indicar que a planilha estimativa poderia ser revista já na licitação.

Tabela 6.

Frequências relativas dos primeiros dígitos da planilha contratual do VLP e a comparação com Benford.

FREQUÊNCIA DOS PRIMEIROS DÍGITOS		
Dígito	Real	Benford
1	0,285	0,301
2	0,195	0,176
3	0,130	0,125
4	0,094	0,097
5	0,072	0,079
6	0,056	0,067
7	0,060	0,058
8	0,060	0,051
9	0,048	0,046

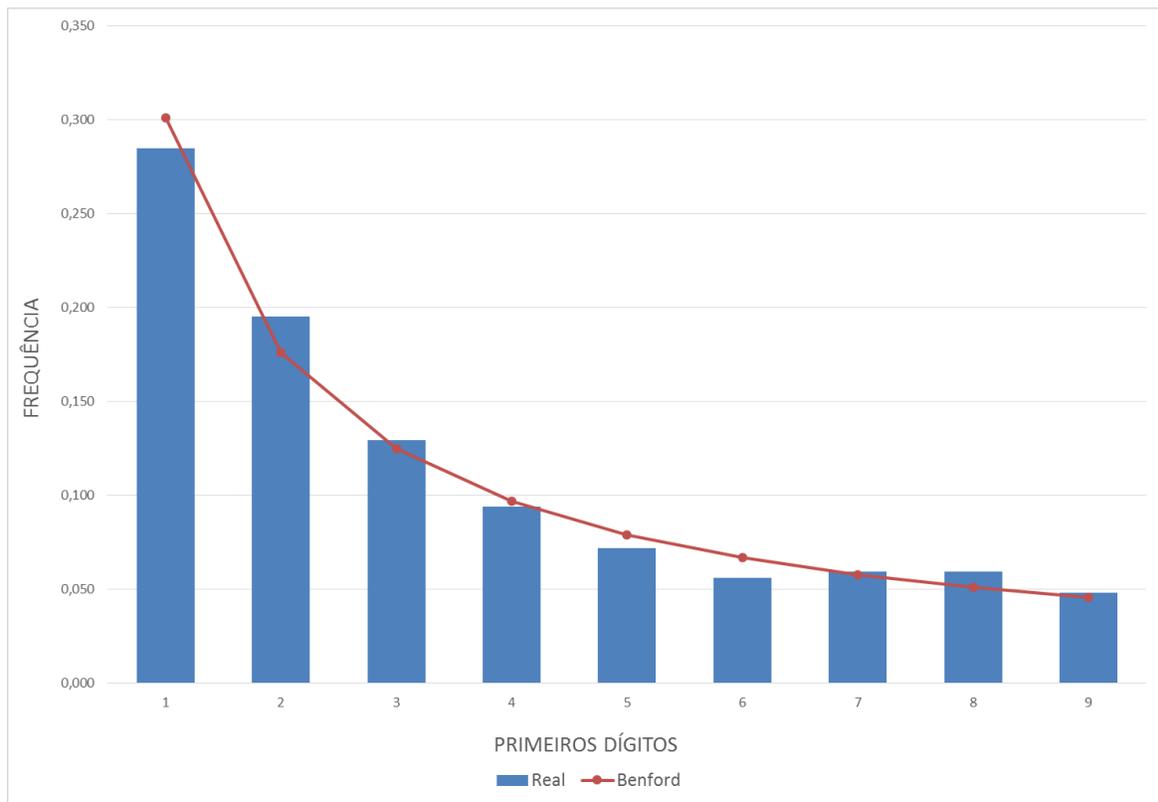


Figura 5. Frequência dos primeiros dígitos na planilha contratual do VLP e a comparação com o previsto por Benford.

Ao primeiro exame e ao aspecto visual, os dados seguem a distribuição de Benford. Em seqüência, procede-se o cálculo das estatísticas Z, DMA e qui-quadrado.

Os resultados do teste do primeiro dígito estão apresentados na Tabela 7. O cabeçalho “Dígito” refere-se às classes de primeiro dígito 1 a 9, “Contagem” refere-se à quantidade de rubricas relacionadas os dígito, “Real” refere-se à frequência relativa de cada classe de dígitos (contagem da classe de dígitos dividida pelo total de rubricas), “Benford” refere-se à frequência prevista pela Lei de Benford para o primeiro dígito, “Estatística Z” registra o valor calculado para o teste Z de diferença entre os valores previsto e real, “DMA_i” é um registro intermediário valor do desvio médio padrão para cada dígito e “ χ^2 ” o valor intermediário para cada dígito da estatística qui-quadrado.

Tabela 7.

Aplicação do teste do primeiro dígito à planilha contratual do VLP

Dígito	Contagem	Real	Benford	Estatística Z	DMA _i	χ^2
1	703	0,285	0,301	1,718	0,016	2,116
2	482	0,195	0,176	2,489	0,019	5,213
3	320	0,130	0,125	0,686	0,005	0,450
4	232	0,094	0,097	0,448	0,003	0,209
5	178	0,072	0,079	1,255	0,007	1,539
6	139	0,056	0,067	2,067	0,011	4,145
7	147	0,060	0,058	0,296	0,002	0,108
8	147	0,060	0,051	1,856	0,008	3,433
9	119	0,048	0,046	0,541	0,002	0,331

A estatística Z é maior do que 1,96 para os dígitos 2 e 6, respectivamente, 2,489 e 2,067, rejeitando a 5% de significância a hipótese de conformidade para tais dígitos. Para encontrar o DMA, somam-se os valores intermediários “DMA_i”. O valor resultante é 0,008 e o diagnóstico é conformidade aceitável com Benford de acordo com o critério apresentado em Nigrini (2012). A estatística qui-quadrado é obtida somando os valores intermediários χ^2 . O resultado obtido é 17,54, acima do valor crítico 15,50 para 5% de

significância e 8 graus de liberdade, provendo um diagnóstico de não-conformidade. Este teste é aparentemente mais estrito que os demais.

Neste ponto, deve-se enfatizar que o teste do primeiro dígito não possui sensibilidade suficiente para indicar quais rubricas que devem passar por maior escrutínio em bases com um número grande de dados como esta. Trata-se apenas de uma primeira visão geral de conformidade dos dados com Benford, sem especificar claramente onde se encontram as maiores não-conformidades.

4.3 Teste do Segundo Dígito

A Tabela 8 apresenta as frequências relativas dos primeiros dígitos para a planilha contratual da obra, abaixo do cabeçalho “Real”, e a comparação com o valor previsto por Benford, e a Figura 6 as ilustra visualmente.

Tabela 8.

Frequências relativas dos segundos dígitos da planilha contratual do VLP e a comparação com Benford.

FREQUÊNCIA DOS SEGUNDOS DÍGITOS		
Dígito	Real	Benford
0	0,137	0,120
1	0,142	0,114
2	0,116	0,109
3	0,093	0,104
4	0,091	0,100
5	0,095	0,097
6	0,090	0,093
7	0,073	0,090
8	0,088	0,088
9	0,074	0,085

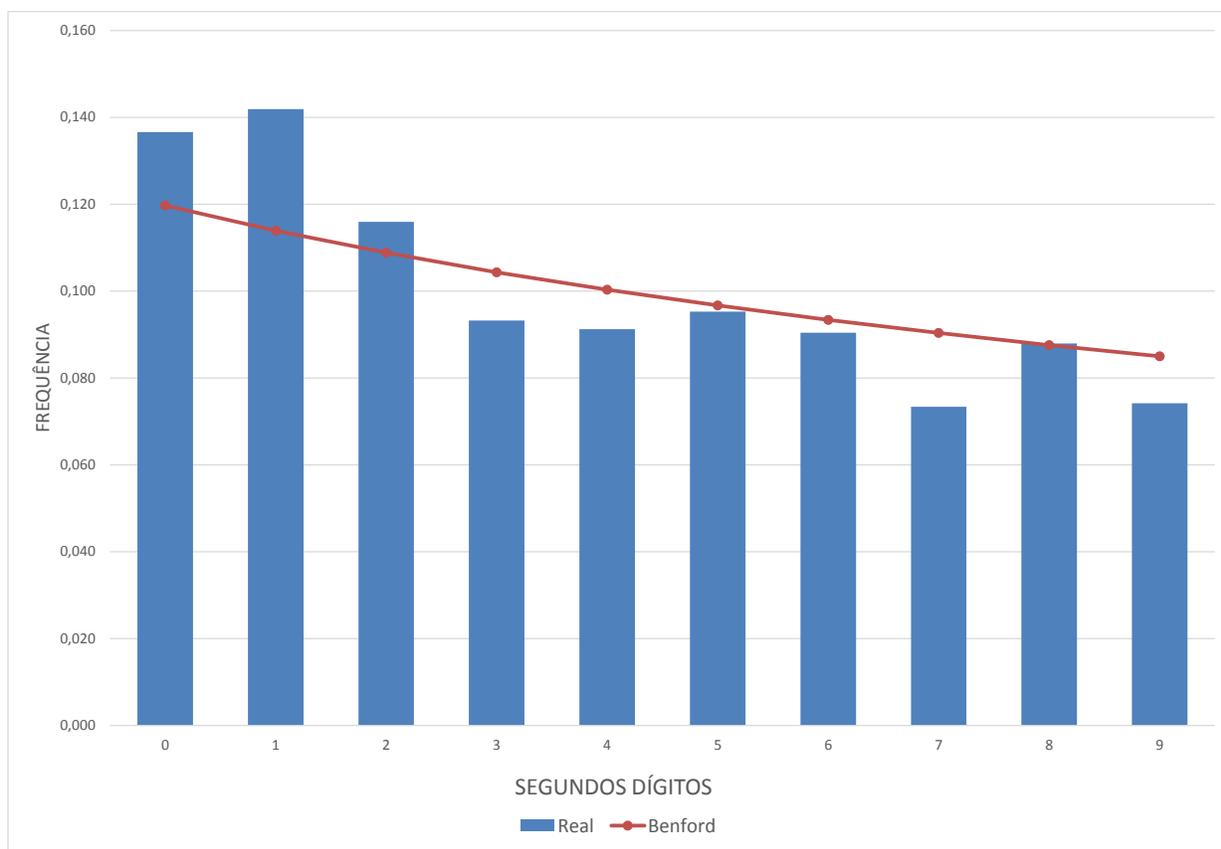


Figura 6. Frequência dos segundos dígitos na planilha contratual do VLP e a comparação com o previsto por Benford.

O exame visual da Figura 6 revela picos nos dígitos 0, 1 e 2. Procede-se abaixo o cálculo das estatísticas.

Os resultados do teste do segundo dígito estão apresentados na Tabela 9. Seguindo a legenda da Tabela 7, de resultados do teste do primeiro dígito, temos: “Dígito” referindo-se às classes de segundo dígito 0 a 9, “Contagem” referindo-se à quantidade de rubricas relacionadas ao dígito, “Real” referindo-se à frequência relativa de cada classe de dígitos (contagem da classe de dígitos dividida pelo total de rubricas), “Benford” referindo-se à frequência prevista pela Lei de Benford para o segundo dígito, “Estatística Z” registrando o valor calculado para o teste Z de diferença entre os valores previsto e real, “DMA_i” como um registro intermediário valor do desvio médio padrão para cada dígito e “ χ_i^2 ” como o valor intermediário para cada dígito da estatística qui-quadrado.

Tabela 9.

Aplicação do teste do segundo dígito à planilha contratual do VLP

Dígito	Contagem	Real	Benford	Estatística Z	DMA _i	χ^2
0	337	0,137	0,120	2,559	0,017	5,904
1	350	0,142	0,114	4,343	0,028	16,961
2	286	0,116	0,109	1,102	0,007	1,146
3	230	0,093	0,104	1,771	0,011	2,913
4	225	0,091	0,100	1,472	0,009	2,039
5	235	0,095	0,097	0,205	0,001	0,052
6	223	0,090	0,093	0,474	0,003	0,234
7	181	0,073	0,090	2,907	0,017	7,874
8	217	0,088	0,088	0,033	0,000	0,004
9	183	0,074	0,085	1,891	0,011	3,398

No que concerne a estatística Z, esta ultrapassa o valor crítico para 5% de significância, 1,96, para os dígitos 0, 1 e 7. O DMA, obtido pela soma dos DMA_i's é 0,0104 e a estatística qui-quadrado, obtida a partir da soma dos χ^2 's é 40,53. O diagnóstico do DMA por Nigrini (2012) é de conformidade marginalmente aceitável, enquanto a comparação do qui-quadrado com o valor crítico a 5% de significância para 9 graus de liberdade - 16,92, confere amplamente o diagnóstico de não-conformidade. A título de comparação, recorda-se que o valor da estatística qui-quadrado dos segundos dígitos calculada para a base de dados das áreas dos municípios brasileiros foi 18,59. Ainda que a conformidade seja rejeitada para as duas bases de dados a 5% de significância, a estatística calculada aqui é significativamente maior.

4.4 Teste dos Dois Primeiros Dígitos

A Tabela 10 mostra os resultados da aplicação do teste dos dois primeiros dígitos. Seguindo as tabelas anteriores, "Dígitos" refere-se aos dois primeiros dígitos (90 categorias), "Contagem" refere-se à quantidade de rubricas relacionadas ao dígito, "Real" refere-se à frequência relativa de cada classe de dígitos (contagem da classe de dígitos

dividida pelo total de rubricas), “Benford” refere-se à frequência prevista pela Lei de Benford para os dois primeiros dígitos, “Estatística Z” registra o valor calculado para o teste Z de diferença entre os valores previsto e real, “DMA_i” é um registro intermediário valor do desvio médio padrão para cada dígito e “ χ^2 ” o valor intermediário para cada dígito da estatística qui-quadrado. A Figura 7 ilustra as frequências dos primeiros dígitos da planilha contratual do VLP e sua comparação com Benford.

Tabela 10.

Aplicação do teste dos dois primeiros dígitos à planilha contratual do VLP

Dígitos	Contagem	Real	Benford	Estatística Z	DMA _i	χ^2
10	90	0,036	0,041	1,174	0,005	1,438
11	95	0,039	0,038	0,135	0,001	0,034
12	105	0,043	0,035	2,060	0,008	4,317
13	59	0,024	0,032	2,270	0,008	5,241
14	65	0,026	0,030	0,994	0,004	1,076
15	89	0,036	0,028	2,361	0,008	5,700
16	50	0,020	0,026	1,817	0,006	3,443
17	51	0,021	0,025	1,260	0,004	1,712
18	62	0,025	0,023	0,475	0,002	0,286
19	37	0,015	0,022	2,381	0,007	5,867
20	84	0,034	0,021	4,365	0,013	19,255
21	65	0,026	0,020	2,098	0,006	4,610
22	47	0,019	0,019	0,018	0,000	0,008
23	27	0,011	0,018	2,705	0,008	7,586
24	53	0,021	0,018	1,337	0,004	1,962
25	52	0,021	0,017	1,475	0,004	2,370

Dígitos	Contagem	Real	Benford	Estatística Z	DMA _i	χ ²
26	41	0,017	0,016	0,010	0,000	0,008
27	23	0,009	0,016	2,497	0,006	6,541
28	38	0,015	0,015	0,066	0,000	0,004
29	52	0,021	0,015	2,537	0,006	6,767
30	58	0,024	0,014	3,801	0,009	14,887
31	44	0,018	0,014	1,637	0,004	2,931
32	29	0,012	0,013	0,608	0,002	0,478
33	19	0,008	0,013	2,222	0,005	5,271
34	24	0,010	0,013	1,184	0,003	1,604
35	28	0,011	0,012	0,308	0,001	0,158
36	47	0,019	0,012	3,183	0,007	10,606
37	19	0,008	0,012	1,707	0,004	3,207
38	37	0,015	0,011	1,653	0,004	3,021
39	15	0,006	0,011	2,245	0,005	5,420
40	9	0,004	0,011	3,314	0,007	11,517
41	31	0,013	0,010	0,926	0,002	1,040
42	23	0,009	0,010	0,342	0,001	0,194
43	48	0,019	0,010	4,631	0,009	22,171
44	26	0,011	0,010	0,291	0,001	0,154
45	20	0,008	0,010	0,631	0,001	0,535
46	26	0,011	0,009	0,515	0,001	0,380
47	11	0,004	0,009	2,339	0,005	5,921
48	15	0,006	0,009	1,409	0,003	2,276
49	23	0,009	0,009	0,185	0,001	0,085
50	28	0,011	0,009	1,370	0,003	2,169
51	34	0,014	0,008	2,795	0,005	8,369
52	25	0,010	0,008	0,909	0,002	1,033
53	12	0,005	0,008	1,689	0,003	3,217
54	11	0,004	0,008	1,848	0,004	3,814
55	10	0,004	0,008	2,012	0,004	4,485
56	12	0,005	0,008	1,490	0,003	2,557
57	7	0,003	0,008	2,589	0,005	7,263
58	18	0,007	0,007	0,074	0,000	0,005
59	21	0,009	0,007	0,590	0,001	0,497
60	12	0,005	0,007	1,242	0,002	1,841
61	52	0,021	0,007	8,194	0,014	68,631
62	12	0,005	0,007	1,125	0,002	1,543
63	9	0,004	0,007	1,801	0,003	3,673
64	9	0,004	0,007	1,751	0,003	3,487
65	7	0,003	0,007	2,197	0,004	5,353
66	15	0,006	0,007	0,153	0,000	0,077
67	6	0,002	0,006	2,360	0,004	6,141
68	7	0,003	0,006	2,065	0,004	4,774
69	10	0,004	0,006	1,256	0,002	1,903
70	23	0,009	0,006	1,879	0,003	4,006

Dígitos	Contagem	Real	Benford	Estatística Z	DMA _i	χ ²
71	8	0,003	0,006	1,680	0,003	3,256
72	21	0,009	0,006	1,493	0,003	2,619
73	20	0,008	0,006	1,293	0,002	2,017
74	15	0,006	0,006	0,031	0,000	0,027
75	9	0,004	0,006	1,249	0,002	1,899
76	10	0,004	0,006	0,939	0,002	1,146
77	20	0,008	0,006	1,531	0,003	2,758
78	8	0,003	0,006	1,398	0,002	2,338
79	13	0,005	0,005	0,130	0,000	0,017
80	17	0,007	0,005	0,877	0,001	1,023
81	17	0,007	0,005	0,927	0,002	1,130
82	12	0,005	0,005	0,135	0,000	0,075
83	25	0,010	0,005	3,266	0,005	11,540
84	10	0,004	0,005	0,614	0,001	0,566
85	8	0,003	0,005	1,142	0,002	1,638
86	4	0,002	0,005	2,246	0,003	5,678
87	36	0,015	0,005	6,662	0,010	46,086
88	12	0,005	0,005	0,031	0,000	0,001
89	6	0,002	0,005	1,585	0,002	2,978
90	16	0,006	0,005	1,067	0,002	1,463
91	4	0,002	0,005	2,112	0,003	5,076
92	12	0,005	0,005	0,123	0,000	0,015
93	11	0,004	0,005	0,136	0,000	0,018
94	12	0,005	0,005	0,048	0,000	0,039
95	12	0,005	0,005	0,084	0,000	0,054
96	18	0,007	0,005	1,924	0,003	4,285
97	8	0,003	0,004	0,752	0,001	0,813
98	20	0,008	0,004	2,620	0,004	7,651
99	6	0,002	0,004	1,303	0,002	2,111

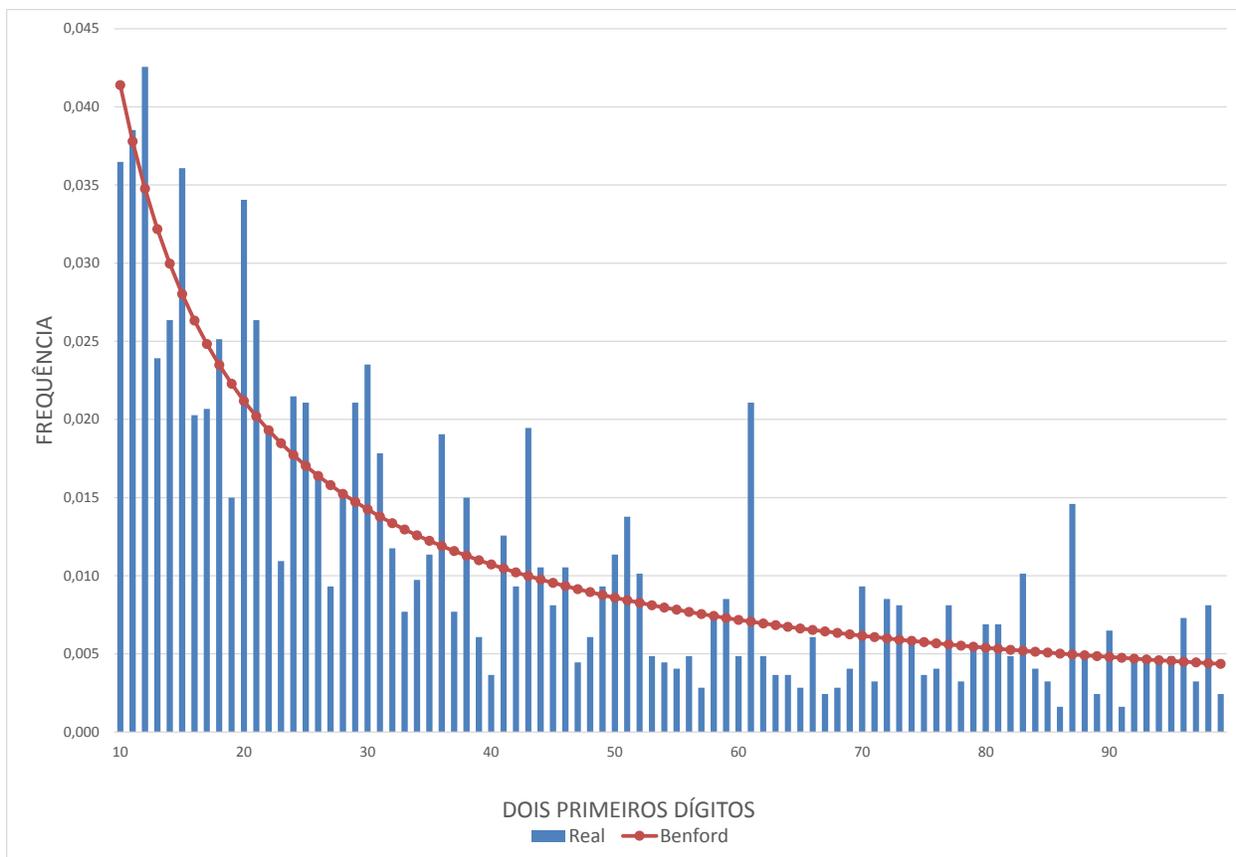


Figura 7. Frequência dos dois primeiros dígitos na planilha contratual do VLP e a comparação com o previsto por Benford.

A análise visual da Figura 7 permite notar vários picos em relação às frequências previstas por Benford. De fato, o teste Z revela 28 categorias cujas estatísticas são maiores que 1,96. São elas: 12, 13, 15, 19, 20, 21, 23, 27, 29, 30, 33, 36, 39, 40, 43, 47, 51, 55, 57, 61, 65, 67, 68, 83, 86, 87, 91, 98.

Os testes de conformidade geral da base de dados, DMA e qui-quadrado, resultam nos valores 0,0034 e 413,24, respectivamente. O diagnóstico pelo critério para o DMA sugerido Nigrini (2012) resulta em não-conformidade. O valor crítico da estatística qui-quadrado para 5% de significância e 89 graus de liberdade é 112,02. A hipótese de conformidade é rejeitada com ampla margem por este critério. Recorda-se que o valor da estatística qui-quadrado para a base de dados das áreas dos municípios brasileiros era 124,35. Apesar de também ter sido rejeitada pelo teste qui-quadrado, a estatística é

notavelmente maior para a planilha contratual do VLP, sugerindo maior evidência de manipulação dos dados.

4.5 Teste da Soma

O teste da soma certifica a importância dos dígitos em termos de montante total. Ele garante que o tempo útil do auditor seja dispendido com categorias de valor monetário significativo e é usado em combinação com o teste dos dois primeiros dígitos no método apresentado para detecção da amostra a ser auditada.

A Tabela 11 mostra os resultados da aplicação do teste da soma à planilha contratual do VLP, onde: “Dígitos” refere-se aos dois primeiros dígitos (90 categorias), “Soma” refere-se à soma dos valores correspondentes à cada categoria, “Real” refere-se à frequência relativa de cada classe de dígitos e “Benford” refere-se à proporção prevista para a soma de cada categoria.

A Figura 8 ilustra as proporções encontradas para a soma de cada categoria em comparação com a esperada pelo teste.

Tabela 11.**Aplicação do teste da soma à planilha contratual do VLP**

	Soma	Real	Benford
10	R\$ 22.395.376,35	0,035	0,011
11	R\$ 24.205.167,85	0,037	0,011
12	R\$ 8.111.343,03	0,013	0,011
13	R\$ 40.033.454,60	0,062	0,011
14	R\$ 25.294.984,72	0,039	0,011
15	R\$ 26.648.050,94	0,041	0,011
16	R\$ 11.049.794,63	0,017	0,011
17	R\$ 26.926.718,60	0,042	0,011
18	R\$ 27.801.371,79	0,043	0,011
19	R\$ 25.894.326,03	0,040	0,011
20	R\$ 11.564.513,13	0,018	0,011
21	R\$ 9.957.442,94	0,015	0,011
22	R\$ 11.035.569,64	0,017	0,011
23	R\$ 6.156.963,78	0,009	0,011
24	R\$ 5.197.410,38	0,008	0,011
25	R\$ 6.991.224,50	0,011	0,011
26	R\$ 6.366.999,19	0,010	0,011
27	R\$ 2.984.659,17	0,005	0,011
28	R\$ 8.960.469,33	0,014	0,011
29	R\$ 5.585.165,54	0,009	0,011
30	R\$ 3.994.262,34	0,006	0,011
31	R\$ 4.780.456,98	0,007	0,011
32	R\$ 12.629.533,13	0,019	0,011
33	R\$ 9.241.300,46	0,014	0,011
34	R\$ 3.182.848,65	0,005	0,011
35	R\$ 1.558.168,61	0,002	0,011
36	R\$ 12.681.847,66	0,020	0,011
37	R\$ 6.003.811,18	0,009	0,011
38	R\$ 11.964.348,70	0,018	0,011
39	R\$ 9.621.925,58	0,015	0,011
40	R\$ 546.548,15	0,001	0,011
41	R\$ 6.890.075,57	0,011	0,011
42	R\$ 1.202.018,43	0,002	0,011
43	R\$ 1.621.646,29	0,002	0,011
44	R\$ 7.083.276,52	0,011	0,011
45	R\$ 7.124.926,31	0,011	0,011

Dígitos	Soma	Real	Benford
46	R\$ 1.682.865,11	0,003	0,011
47	R\$ 1.208.220,26	0,002	0,011
48	R\$ 351.886,79	0,001	0,011
49	R\$ 8.580.795,74	0,013	0,011
50	R\$ 15.575.685,59	0,024	0,011
51	R\$ 2.535.546,98	0,004	0,011
52	R\$ 1.900.555,86	0,003	0,011
53	R\$ 1.121.657,56	0,002	0,011
54	R\$ 61.441.906,83	0,095	0,011
55	R\$ 2.496.946,25	0,004	0,011
56	R\$ 1.484.123,70	0,002	0,011
57	R\$ 1.851.072,66	0,003	0,011
58	R\$ 11.339.183,00	0,017	0,011
59	R\$ 1.201.360,63	0,002	0,011
60	R\$ 1.897.360,60	0,003	0,011
61	R\$ 2.801.387,85	0,004	0,011
62	R\$ 1.427.653,79	0,002	0,011
63	R\$ 114.694,69	0,000	0,011
64	R\$ 786.008,27	0,001	0,011
65	R\$ 920.396,65	0,001	0,011
66	R\$ 2.061.106,11	0,003	0,011
67	R\$ 880.355,44	0,001	0,011
68	R\$ 6.968.154,29	0,011	0,011
69	R\$ 1.749.849,45	0,003	0,011
70	R\$ 2.058.344,77	0,003	0,011
71	R\$ 572.010,68	0,001	0,011
72	R\$ 184.634,24	0,000	0,011
73	R\$ 8.946.864,68	0,014	0,011
74	R\$ 2.040.272,38	0,003	0,011
75	R\$ 3.106.256,48	0,005	0,011
76	R\$ 15.627.352,51	0,024	0,011
77	R\$ 1.748.514,25	0,003	0,011
78	R\$ 238.855,09	0,000	0,011
79	R\$ 4.325.224,71	0,007	0,011
80	R\$ 1.232.870,64	0,002	0,011
81	R\$ 2.817.890,00	0,004	0,011
82	R\$ 990.034,46	0,002	0,011
83	R\$ 2.152.593,35	0,003	0,011
84	R\$ 1.970.120,70	0,003	0,011
85	R\$ 1.141.097,92	0,002	0,011

Dígitos	Soma	Real	Benford
86	R\$ 268.519,46	0,000	0,011
87	R\$ 9.403.684,13	0,014	0,011
88	R\$ 3.101.090,57	0,005	0,011
89	R\$ 11.042.568,07	0,017	0,011
90	R\$ 9.308.667,10	0,014	0,011
91	R\$ 102.414,40	0,000	0,011
93	R\$ 2.459.049,43	0,004	0,011
94	R\$ 9.743.415,98	0,015	0,011
95	R\$ 506.184,81	0,001	0,011
96	R\$ 1.263.339,64	0,002	0,011
97	R\$ 515.868,64	0,001	0,011
98	R\$ 3.107.987,73	0,005	0,011
99	R\$ 2.099.886,25	0,003	0,011

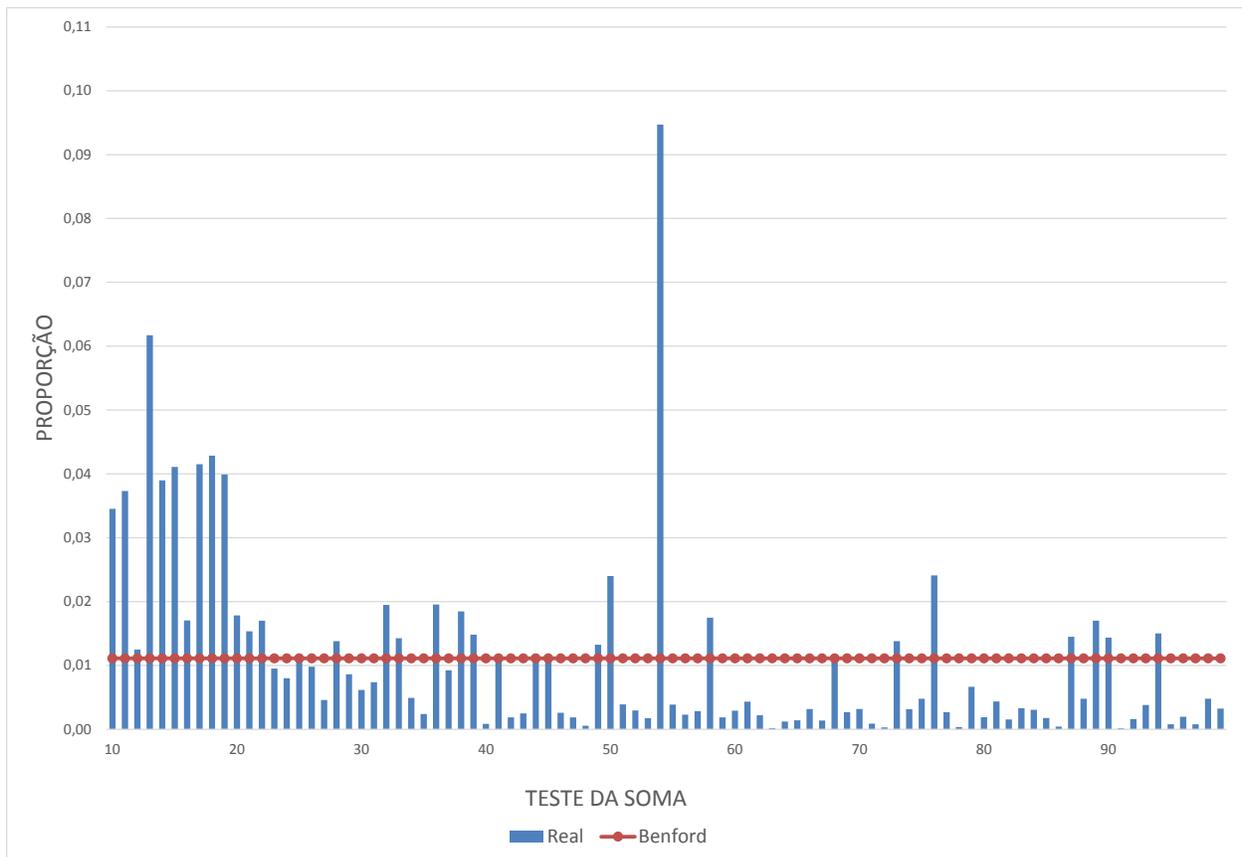


Figura 8. Proporção da soma de cada categoria e a comparação com o valor esperado.

O teste da soma revela 11 categorias de dígitos que ultrapassam o limite μ inicial de 100% estabelecido pelo algoritmo, isto é, a proporção ultrapassa 2,2%. São elas: 10, 11, 13, 14, 15, 17, 18, 19, 50, 54, 76. Juntas, tais categorias correspondem a 44,61% do valor total da planilha. O teste da soma revela não-conformidade com a Lei de Benford.

4.6 Algoritmo de seleção de itens candidatos à auditoria

O algoritmo de Bugarin e Cunha (2015) modificado, conforme apresentado na seção 3 deste trabalho, é aplicado à planilha contratual do VLP para seleção das rubricas candidatas à auditoria. A cada rodada de seleção de dígitos é realizada uma comparação com a auditoria realizada pelo TCDF, para aferir se o método de seleção dos dígitos aponta onde estão as fraudes.

O método utilizado pelo Tribunal, como explicado anteriormente, baseia-se na curva ABC. As rubricas de preços unitários são agregadas e classificadas do maior para o menor custo. Após, são selecionados até 20% dos itens, a começar do mais caro, de forma que somem até 80% do valor total. As rubricas de preços selecionadas são comparadas com valores de mercado para a detecção de sobrepreço.

A comparação com a auditoria do TCDF requer alguns comentários metodológicos. Primeiramente, como é natural a uma planilha de obra de grandes proporções, há itens que se repetem em vários momentos da planilha. Por exemplo, a rubrica “Escavação, carga e transporte de material de 1ª categoria DMT até 5km” tem 10 ocorrências. O preço unitário da rubrica é sempre o mesmo, mas a quantidade e, logo, o preço total podem variar. A análise do Tribunal é feita sobre o preço unitário e para tal, os registros de um mesmo item são analisados de maneira agregada. A análise realizada neste trabalho por Benford é feita a partir dos preços totais (preços multiplicados pelas quantidades). Assim sendo, a rubrica com sobrepreço detectada pelo Tribunal é considerada indicada pela análise de Benford caso ao menos um registro de preço total referente ao item possua os primeiros dígitos selecionados.

A planilha contratual do VLP soma R\$ 648.774.183,21, mas a coluna de quantidades está incompleta³. Assim, vários preços finais resultaram em zero, o que os inviabiliza para o teste de Benford. Desta forma, eles foram retirados da análise. Uma planilha completa provavelmente geraria dados mais robustos.

Por fim, como no máximo 80% do valor total é auditado pelo TCDF, há margem para que o algoritmo utilizado aponte itens que não passaram pela análise do Tribunal. Possivelmente, seria descoberto um superfaturamento adicional caso esses também passassem por auditoria.

4.6.1 Primeira rodada

Para ser coerente com a auditoria do TCDF, estabelece-se $\sigma = 80\%$ e $\varepsilon = 5\%$. Ou seja, audita-se até 80% do valor total da planilha, utilizando a precisão de 5% para o critério de parada.

A primeira rodada do algoritmo seleciona as categorias de dígitos cuja estatística Z do teste dos dois primeiros dígitos é maior que 1,96 ($\lambda=5\%$ de significância) e a diferença no teste da soma é maior que $\mu=100\%$ (proporção maior que 2,2%).

Os dígitos selecionados são: 13, 15 e 19. O custo da amostra selecionada é R\$ 92.575.831,58, o que corresponde a 14,27% do total da planilha. São detectados R\$ 55.460.158,98 em comum com o TCDF, que equivale a 39,78% do superfaturamento detectado. Logo, a auditoria do TCDF aponta sobrepreço em 59,91% da amostra selecionada pelo algoritmo em primeira rodada.

Como a amostra selecionada se encontra abaixo do limite estabelecido de 80% do total, procede-se à segunda rodada.

4.6.2 Segunda rodada

A segunda rodada mantém o critério de 1,96 para a estatística Z do teste dos dois primeiros dígitos ($\lambda=5\%$) e diminui a diferença no teste da soma pelo critério $\delta=25\%$ para $\mu=75\%$ (proporção maior que 1,925%), conforme previsto pelo algoritmo.

³ Deve-se à supressão de itens inicialmente contratados.

Em adição aos selecionados anteriormente, a rodada indica o dígito 36. O custo da amostra selecionada passa para R\$ 105.257.679,24, isto é, 16,22% do total da planilha. O montante de R\$ 57.683.798,98 foi detectado em comum com o TCDF, ou seja 41,38% do superfaturamento. É apontado pelo TCDF sobrepreço em 54,80% da amostra selecionada em segunda rodada.

Como a amostra selecionada ainda está distante do limite estabelecido de 80% do total, passa-se para a terceira rodada.

4.6.3 Terceira rodada

O algoritmo recomenda a manutenção do critério de 1,96 para a estatística Z do teste dos dois primeiros dígitos ($\lambda=5\%$) e a diminuição da diferença no teste da soma em $\delta=25\%$ para $\mu=50\%$ (proporção maior que 1,65%) em terceira rodada.

Em adição às categorias indicadas anteriormente, foi selecionado o dígito 20. O custo total da amostra passa para R\$ 116.822.192,37 e o valor detectado em comum pelo TCDF é R\$ 73.686.554,15. Chega-se ao apontamento de 52,44% do superfaturamento encontrado em auditoria com 18,01% do total da planilha. Em 63% da amostra sinalizada por Benford está indicado sobrepreço.

O algoritmo determina o prosseguimento das rodadas, dado que o valor selecionado ainda está longe do limite estabelecido.

4.6.4 Quarta rodada

A quarta rodada mantém o critério de 1,96 para a estatística Z do teste dos dois primeiros dígitos ($\lambda=5\%$) e a diminui a diferença no teste da soma em $\delta=25\%$ para $\mu=25\%$ (proporção maior que 1,375%).

A amostra selecionada pelo algoritmo passa a incluir os dígitos 21, 33, 39 e 87. O custo da amostra sobe para R\$ 155.046.545,48, 23,90% do total da planilha. Foram detectados R\$ 75.429.212,39 em comum com o TCDF, que equivale a 54,11% do superfaturamento detectado. A auditoria do TCDF aponta sobrepreço em 48,65% da amostra apontada pelo algoritmo em quarta rodada.

Como a amostra selecionada ainda está longe do limite estabelecido, procede-se à quinta rodada.

4.6.5 Quinta rodada

A quinta rodada ainda mantém o critério de 1,96 para a estatística Z do teste dos dois primeiros dígitos ($\lambda=5\%$) e a diminui a diferença no teste da soma em $\delta=25\%$ para $\mu=0\%$ (proporção maior que 1,1%).

Em adição aos demais, seleciona-se o dígito 12 e o custo da amostra passa para R\$ 163.157.888,52, 25,15% do total. A categoria, no entanto, não está relacionada com nenhuma nova rubrica de sobrepreço apontada pela auditoria do TCDF. A proporção entre o valor detectado em comum pelo TCDF e a amostra selecionada pelo algoritmo é de 46,23%.

A amostra selecionada ainda está longe do limite estabelecido e procede-se à sexta rodada.

4.6.6 Sexta rodada

Na sexta rodada, não é possível diminuir o critério de diferença do teste da soma, pois $\mu=0$. Assim, o algoritmo recomenda a adoção de $\lambda=10\%$ para o teste dos dois primeiros dígitos, passando para 1,64 o valor limite para a estatística Z.

As categorias de dígitos selecionadas em adição às demais são: 16, 38 e 54. O custo da amostra selecionada pelo algoritmo é de R\$ 247.613.938,67, ou seja, 38,17%. Destes, são detectados R\$ 102.320.832,18 pela auditoria do TCDF. Em 41,32% da amostra selecionada é constatado sobrepreço.

O algoritmo revela admiráveis 73,40% do superfaturamento apontado pelo Tribunal em menos de 40% do custo total da planilha. Não são aceitas demais flexibilizações e as rodadas chegam ao fim. São indicados a auditoria de preços totais iniciados pelos dígitos: 12, 13, 15, 16, 19, 20, 21, 33, 36, 38, 39, 54, 87 (13 categorias). Outro método, como a curva ABC, pode complementar a escolha da amostra a ser auditada para chegar a 80% do total.

A Tabela 12 a seguir resume o procedimento de rodadas do algoritmo para detecção da amostra com maiores evidências de fraude de acordo com Benford.

Tabela 12

Aplicação do teste da soma à planilha contratual do VLP

PRIMEIRA RODADA Teste dos dois dígitos. $Z > 1.96$ ($\lambda = 5\%$) Teste da soma. Diferença $> 100\%$. ($> 2.2\%$)	
Categorias de dígitos selecionadas:	13, 15, 19
Custo da amostra:	\$92.575.831,58
(em porcentagem do total):	14,27%
Valor detectado em comum pelo TCDF	\$55.460.158,98
(em porcentagem do superfaturamento):	39,78%
É comprovado (pela auditoria do TCDF) sobrepreço em 59.91% da amostra apontada por Benford em 1a rodada	
SEGUNDA RODADA Teste dos dois dígitos. $Z > 1.96$ ($\lambda = 5\%$) Teste da soma. Diferença $> 75\%$. ($> 1.925\%$)	
Categorias de dígitos selecionadas:	13, 15, 19, 36
Custo da amostra:	\$105.257.679,24
(em porcentagem do total):	16,22%
Valor detectado em comum pelo TCDF	\$57.683.798,98
(em porcentagem do superfaturamento):	41,38%
É comprovado (pela auditoria do TCDF) sobrepreço em 54.80% da amostra apontada por Benford em 2a rodada	
TERCEIRA RODADA Teste dos dois dígitos. $Z > 1.96$ ($\lambda = 5\%$) Teste da soma. Diferença $> 50\%$. ($> 1.65\%$)	
Categorias de dígitos selecionadas:	13, 15, 19, 20, 36
Custo da amostra:	R\$ 116.822.192,37
(em porcentagem do total):	18,01%
Valor detectado em comum pelo TCDF	\$73.686.554,15
(em porcentagem do superfaturamento):	52,86%
É comprovado (pela auditoria do TCDF) sobrepreço em 63.08% da amostra apontada por Benford em 3a rodada	
QUARTA RODADA Teste dos dois dígitos. $Z > 1.96$ ($\lambda = 5\%$) Teste da soma. Diferença $> 25\%$. ($> 1.375\%$)	
Categorias de dígitos selecionadas:	13, 15, 19, 20, 21, 33, 36, 39, 87
Custo da amostra:	\$155.046.545,48
(em porcentagem do total):	23,90%
Valor detectado em comum pelo TCDF	\$75.429.212,39
(em porcentagem do superfaturamento):	54,11%
É comprovado (pela auditoria do TCDF) sobrepreço em 48.65% da amostra apontada por Benford em 4a rodada	

QUINTA RODADA	
Teste dos dois dígitos. $Z > 1.96$ ($\lambda = 5\%$)	
Teste da soma. Diferença $> 0\%$ ($> 1.1\%$)	
Categorias de dígitos selecionadas:	12, 13, 15, 19, 20, 21, 33, 36, 39, 87
Custo da amostra:	\$163.157.888,52
(em porcentagem do total):	25,15%
Valor detectado em comum pelo TCDF	\$75.429.212,39
(em porcentagem do superfaturamento):	54,11%
É comprovado (pela auditoria do TCDF) sobrepreço em 46.23% da amostra apontada por Benford em 5ª rodada	
SEXTA RODADA	
Teste dos dois dígitos. $Z > 1.64$ ($\lambda = 10\%$)	
Teste da soma. Diferença $> 0\%$ ($> 1.1\%$)	
Categorias de dígitos selecionadas:	12, 13, 15, 16, 19, 20, 21, 33, 36, 38, 39, 54, 87
Custo da amostra:	\$247.613.938,67
(em porcentagem do total):	38,17%
Valor detectado em comum pelo TCDF	\$102.320.832,18
(em porcentagem do superfaturamento):	73,40%
É comprovado (pela auditoria do TCDF) sobrepreço em 41.32% da amostra apontada por Benford em 6ª rodada	
FIM DO ALGORITMO	

5. Conclusão

O método de seleção de rubricas candidatas à auditoria pela Lei de Benford apresentou bons resultados. Ao fim da sexta rodada, foram revelados 73,40% do superfaturamento apontado pelo TCDF auditando-se apenas 38,17% do valor total da planilha. Para cada real apontado como suspeito de fraude por Benford, 41 centavos tiveram sobrepreço indicado pela auditoria do Tribunal. Acredita-se, portanto, que o uso dos testes de Benford na etapa de planejamento da auditoria de uma obra pública tornem o trabalho do auditor mais assertivo. O algoritmo que combina os testes dos dois primeiros dígitos e da soma forneceu um diagnóstico bastante preciso das rubricas que apresentavam irregularidades na planilha contratual do VLP.

Dentre os critérios que atestam conformidade geral com a distribuição de Benford para o teste dos dois primeiros dígitos, o diagnóstico por Nigrini (2012) a partir do desvio médio absoluto (DMA) se mostrou mais adequado que a estatística qui-quadrado a 5% de significância. O critério DMA revela conformidade para a base de dados das áreas dos

municípios brasileiros e não-conformidade para a planilha contratual do VLP. Já a estatística qui-quadrado se revelou muito sensível a diferenças de frequência esperada, diagnosticando não-conformidade para as duas bases de dados analisadas. É notável, porém, a diferença de escala entre os valores calculados para as duas bases. Para a base de dados das áreas dos municípios, temos $\chi^2=124,35$, e para a planilha contratual do VLP, $\chi^2=413,24$. Lembra-se que quanto maior o valor da estatística, maiores são as evidências de manipulação. Ainda assim, como o valor crítico a 5% é 112,02, a conformidade foi rejeitada para ambos.

Os testes do primeiro e do segundo dígito não se mostraram muito úteis para a seleção de itens para a auditoria, por repartirem a base de dados em um número baixo de categorias. São indicados para uma primeira visão geral dos números. O teste da soma de maneira isolada também não foi capaz de sinalizar precisamente as categorias candidatas a maior escrutínio. No entanto, a combinação do teste dos dois primeiros dígitos, que divide a base de dados em 90 categorias e indica as que não estão em conformidade com Benford, e o teste da soma, que seleciona as categorias de dígitos com maior importância pecuniária, levou a um apontamento acurado de rubricas com evidências de manipulação de dados.

Portanto, julga-se benéfica a utilização do algoritmo apresentado para a escolha dos itens de planilhas de obras públicas para auditoria. Além disso, por usar somente a plataforma *Excel*®, a técnica é bastante acessível. Caso necessário, pode ser completado por outro procedimento para completar a seleção da porcentagem total pretendida para análise.

6. Referências Bibliográficas

6.1 Artigos e livros

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

Boyle, J. (1994), "An application of Fourier series to the most significant digit problem". *American Mathematical Monthly*. **101**, p. 879-886.

Bugarin, M. e Cunha, F. (2016), "A didactic note on the use of Benford's Law in public works auditing, with an application to the construction of Brazilian Amazon Arena 2014 World Cup soccer stadium". *Economia* (Yokohama), no prelo.

Cunha, F. e Bugarin, M. (2015), "Benford Law for audit of public works: An analysis of overpricing in Maracanã soccer arena's renovation". *Economics Bulletin*. **35**, p.120-129.

Cunha, F.C.R. (2013), "Aplicações da Lei Newcomb-Benford à Auditoria de Obras Públicas". Dissertação (Mestrado em Regulação e Gestão de Negócios) – Universidade de Brasília, Brasília, 486p.

Carslaw, C. (1988), "Anomalies in income numbers: evidence from goal oriented behaviour". *The Accounting Review*. **63**, p. 321-327.

Durtschi, C., W. Hillison e C. Pacini (2004), "The Effective Use of Benford's Law to Assist in Detecting Fraud in Accounting Data". *Journal of Forensic Accounting*. **5**, p. 17–34.

Hill, T. (1995) "Base-Invariance Implies Benford's Law". *Proceedings of the American Mathematical Society*. **123**, p. 887-895.

Jamain, A. (2001) "Benford's Law". Dissertação - Departamento de Matemática, Imperial College of London, Reino Unido. 62 p.

Leemis, L., B. Schmeiser e D. Evans (2000), "Survival Distributions Satisfying Benford's Law", *American Statistician*, **54**, p. 236–241.

Newcomb, S. (1881) "Note on the frequency of the different digits in natural numbers". *The American Journal of Mathematics*, **4**, p. 39-40.

Mittermaier, L. e Nigrini, M. (1997) "The use of Benford's Law as an aid in analytical procedures." *Auditing: A Journal of Practice & Theory*. **Fall**. p. 52+.

Mueller, D. (2003). "Public choice III". Cambridge University Press; 3ª edição.

Nigrini, M. (2012) "Benford's Law. Applications for Forensic Accounting Auditing, and Fraud Detection". John Wiley & Sons, Inc.: Hoboken, New Jersey.

Pinkham, R. (1961), "On the Distribution of First Significant Digits". *The Annals of Mathematical Statistics*. **32**, p. 1223-1230.

Raimi, R. (1976), "The first digit problem". *American Mathematical Monthly*. **83**, p. 521-538.

Thomas, J. (1979), "Unusual patterns in reported earnings". *The Accounting Review*. **64**, p. 773-787.

Varian, H. (1972), "Benford's Law". *The American Statistician*. **26**, p. 55-56.

6.2 Dados

Área territorial dos municípios brasileiros, em km^2 . Arquivo em formato xls. Disponível em: http://www.ibge.gov.br/home/geociencias/areaterritorial/resolucao_04_2014.shtm

Processo nº 25778/12. Anexo XVIII – CD de medição de janeiro/2014.

Informação nº 25/2014 – Núcleo de Fiscalização de Obras e Serviços de Engenharia (NFO), referente ao processo 889/2009. Tribunal de Contas do Distrito Federal. Disponível online em: <http://www.tc.df.gov.br>.

6.3 Sítios eletrônicos

<http://www.benfordonline.net/>

<http://www.ibge.gov.br/>

<http://www.nigrini.com/>

<http://www.tc.df.gov.br>

7. Apêndice A - Itens apontados em comum pelo algoritmo de Benford e pelo TCDF

Esta última seção visa aclarar o processo de correspondência dos dígitos selecionados pelo algoritmo com os itens com sobrepreço detectado pelo TCDF. Recorda-se que, na planilha contratual do VLP há itens que se repetem em vários momentos. O preço unitário da rubrica é sempre o mesmo, mas a quantidade e, logo, o preço total podem variar. A

análise do tribunal é feita sobre o preço unitário e para tal, os registros de um mesmo item são analisados de maneira agregada. A análise realizada por Benford é feita a partir dos preços totais (preços multiplicados pelas quantidades). Desta forma, a rubrica com sobrepreço detectada pelo tribunal é considerada indicada pela análise de Benford caso ao menos um registro de preço total referente ao item possua os primeiros dígitos selecionados.

A Tabela 13 resume os itens com sobrepreço detectado pelo TCDF de forma agregada, bem como o sobrepreço detectado em casa uma. A coluna da direita, denominada “Dígitos correspondentes” esclarece os dígitos detectados por Benford para preços totais que revelam a rubrica em questão.

Tabela 13.

Itens superfaturados detectados pelo TCDF e a correspondência com Benford

Item	Preço unitário contratado	Quantidade total	Preço unitário ajustado	Sobrepreço total	% do superfaturamento total	Dígitos correspondentes
Pavimentação em concreto simples com forma deslizante inclusive transporte até 5 km	\$643,56	\$85.148,05	\$354,56	\$24.607.786,45	17,65%	54
Fornecimento e aplicação de aço ca 50/60	\$8,13	\$4.403.694,36	\$4,93	\$14.091.821,94	10,11%	15
Estacas raiz com dn 0,40m	\$727,95	\$55.582,93	\$623,41	\$5.810.639,50	4,17%	13, 33
Momento extraordinário de transp. De solos moles (DMT além de 5km)	\$1,53	\$2.417.000,00	\$0,61	\$2.223.640,00	1,60%	36
Momento extraordinário de transporte de BGTC (além de 5km)	\$1,94	\$2.021.091,45	\$0,81	\$2.283.833,34	1,64%	16
Momento extraordinário de transporte de BGS	\$1,85	\$9.992.992,46	\$0,83	\$10.192.852,31	7,31%	Não detectado por Benford
Momento extraordinário de transporte de solo-cal (DMT além de 5km)	\$1,37	\$54.423,50	\$0,47	\$48.981,15	0,04%	Não detectado por Benford
Momento extraordinário de transporte material 1a cat. ou solo de jazida (DMT além de 5km)	\$1,38	\$18.822.819,07	\$0,62	\$14.305.342,50	10,26%	Não detectado por Benford
Concreto Fck 30 Mpa	\$612,13	\$30.484,20	\$477,27	\$4.111.099,21	2,95%	13, 15
Concreto Fck 22,5 Mpa	\$496,44	\$9.496,04	\$378,15	\$1.123.286,57	0,81%	13, 16
Concreto magro	\$353,61	\$17.114,83	\$279,28	\$1.272.145,31	0,91%	19, 36

Item	Preço unitário contratado	Quantidade total	Preço unitário ajustado	Sobrepço total	% do superfaturamento total	Dígitos correspondentes
Brita graduada tratada com cimento - BGTC (pav. Flexível)	\$255,95	\$55.017,45	\$153,17	\$5.654.693,51	4,06%	19
Brita graduada tratada com cimento - BGTC (pav. Rígida)	\$255,95	\$77.972,10	\$164,31	\$7.145.363,24	5,13%	20
Escavação, carga e transporte de material 1a cat., até 5 km	\$16,08	\$1.655.587,28	\$10,73	\$8.857.391,93	6,35%	18
Lançamento e montagem de vigas metálicas	\$10,22	\$743.828,10	\$9,34	\$654.568,73	0,47%	Não detectado por Benford
Perfil metálico I 120 x 45 - SAC 350 ou similar	\$10,22	\$1.079.366,60	\$7,14	\$3.324.449,13	2,38%	Não detectado por Benford
Perfil metálico I 220 x 60 - SAC 350 ou similar	\$9,54	\$726.107,60	\$7,14	\$1.742.658,24	1,25%	16, 21
Brita graduada simples (BGS)	\$178,37	\$128.578,56	\$114,49	\$8.213.598,42	5,89%	13
Banco de dutos de solo-cimento com 3 tubos PVC 4"	\$252,35	\$34.800,00	\$63,65	\$6.566.760,00	4,71%	Não detectado por Benford
Cimbramento metálico	\$47,69	\$142.558,28	\$35,01	\$1.807.639,02	1,30%	19, 21
Forma plana	\$80,45	\$114.147,39	\$51,04	\$3.357.074,74	2,41%	15, 19
Asfalto diluído CM-30	\$2.805,97	\$2.177,80	VR	\$1.471.009,25	1,06%	Não detectado por Benford
CAP 50/70 - Fornecimento	\$1.987,37	\$7.818,39	VR	\$3.365.969,87	2,41%	15
Emulsão asfáltica RR-2C	\$1.736,66	\$447,89	VR	\$212.759,32	0,15%	Não detectado por Benford
Emulsão asfáltica RR-2C modificada por polímero-fornecimento	\$2.257,66	\$520,70	VR	\$306.901,39	0,22%	Não detectado por Benford
Estrutura em aço com perfis metálicos laminados (tipo I 150 x 18)	\$17,42	\$559.186,95	\$8,52	\$4.976.763,86	3,57%	15
Estrutura em aço com perfis metálicos laminados (tipo U 6" x 19,4)	\$17,42	\$29.109,02	\$8,52	\$259.070,28	0,19%	19
Concreto projetado consumo 440 kg de cimento/m3 para aplicação em trincheira	\$1.366,33	\$1.899,29	\$620,60	\$1.416.356,75	1,02%	13
Total:				\$139.404.455,96	100,00%	

Dados extraídos de Informação nº 25/2014 – NFO do TCDF.

8. Apêndice B – Listagem de itens apontados pelo algoritmo de Benford

- Abrigo provisório de madeira
- Acabamentos
- Aceitação Na Fábrica
- Aluguel de Impressora HP
- Aparelho Ar Condicionado 12000BTUs
- Aparelho de apoio de neoprene fretado
- Aplicação de primer sintético em estrutura de aço carbono, duas demãos, à revólver
- Aprovação do Projeto
- Aproveitamento De Material
- Argamassa de cimento e areia traço 1:4
- Armação
- Armação em tela de aço CA-60, tipo Q196
- Arrasamento de estaca
- Auxiliar de Pessoal
- Banco de dutos de solo-cimento com 3 tubos PVC 4"
- Bota-fora de material demolido até a distância de 5 km

- Brita Graduada Simples (BGS)
- Brita Graduada Tratada com Cimento - BGTC
- CAIXA DISSIPADORA DEB03
- Calha de chapa galvanizada nº 26 L.D. = 90cm
- Camada de desgaste - CBUQ, exclusive CAP 50/70
- Caminhão F-4000
- Caminhão Munck
- CAP 50/70 - FORNECIMENTO
- Carga e transporte de mat. 1ª cat. de emp. até 5 Km
- Centro de controle operacional
- Cerca de arame galvanizado com mourão de concreto
- Chapa de aço perfurada com furo redondo alternado - $\varnothing 2.8\text{mm}$, #7/64", permeabilidade visual 44%, com pintura esmalte fosca
- Chapisco em paredes e tetos traço 1:3
- Cimbramento metálico
- Cobertura em telha metálica em chapa trapezoidal de aço zincado, pré-pintada, tipo "sanduiche" com enchimento em espuma de poliuretano expandido - Perfilor - linha Termilor TP - cor cinza metálico.
- Concreto Fck 22,5 Mpa
- Concreto Fck 30 Mpa
- Concreto magro
- Concreto projetado consumo 440 kg de cimento/m³ para aplicação em trincheira
- Confecção de tubos concreto armado D=0,40 m PA1
- Confecção de tubos concreto armado D=0,40 m PS1
- Consolidação Do Projeto Básico
- Consumo de Energia
- Copeiro
- Cópia e Encadernações
- Corpo BSTC D=0,40 m, sem fornecimento de tubo
- Corpo BSTC, com lastro, D=1,00 m, sem fornecimento de tubo
- Corrimão em perfil tubular de aço inox - $\varnothing 2"$, inclusive grade palito (ver projeto)
- Cortes
- Demolição de concreto armado
- Demolição de concreto simples, inclusive passeio e meio-fio
- Descida d'água aterros em degraus - DAD 01
- Descida d'água tipo rápida, calha de concreto - DAR 01
- Descidas de águas pluviais de PVC dn 100 mm (4")
- Desenhista
- Desmobilização
- Despesas com Moradia Administração
- Detecção E Combate À Incêndio
- Diversos
- Drenagem
- Elétricas
- Emassamento de paredes com massa corrida PVA, duas demãos
- Emassamento de portas de madeira
- Emboço
- Encaregado de Terraplanagem
- Engenharia / Administração / Produção
- Entrega ao Cliente
- Escavação manual de vala
- Escavação mecanizada de vala

- Escavação, carga e transporte de mat. 1ª cat. de emp. até 5 km - Aterro
- Escavação, carga e transporte de mat. 1ª cat. de emp. até 5 km - Empréstimo ou Jazida
- Escavação, carga e transporte de material 1ª cat., até 5 km
- Escavação, carga e transporte de material de 1a categoria invertida - cut and cover, até 5km sem transporte vertical
- Esquadria alumínio nº25 com anodização natural acetinada, vidro fantasia incolor, basculante com contramarco e ferragens.
- Esquadria Zenital em vidro laminado 8mm - (EVL02)
- Estacas raiz com dn 0,40 m
- Estrutura de saída da bacia
- Estrutura em aço com perfis metálicos laminados (tipo I 150x18)
- Estrutura em aço com perfis metálicos laminados (tipo U 6"x19.4)
- Execução da gabião colchão reno H = 0,30m
- Fachadas
- Fixo de vidro temperado para guichê - e=10mm - caixilho de alumínio anodizado - cor natural fosca - gaxeta de neoprene - 2.225m x 1.00m
- Forma plana
- Formas
- Fornecimento e aplicação de aço ca 50 / 60
- Fornecimento e aplicação de geogrelhas
- Fundações
- Geometria
- Gerência Fase 2
- Grade do vertedor (barra chata de ferro, 10x25mm, espaçamento=7,5cm , 4,00x1,00m)
- GRANITINA para revestimento de piso moldado "in loco", inclusive contrapiso
- Hidráulicas
- Instalação E Montagem
- Instalações de combate a incêndio
- Instalações elétricas
- Internet
- Janela de alumínio nº 25 com anodização natural acetinada, vidro fantasia 5mm incolor, basculante, 0.50m X 0.40m com contramarco e ferragens
- Kombi
- Limpeza da obra
- Locação da obra
- Locação de estacas
- Meio-fio conforme desenho 01/67 - DU - NOVACAP, fornecimento e assentamento
- Momento de transporte de concreto para além de 5 km
- Momento de transporte de mat.1ª cat. além de 5 Km
- Momento Extraordinário de transp. de solos moles (DMT além de 5km)
- Momento extraordinário de transporte de BGTC (além de 5 Km)
- Momento Extraordinário de transporte de material (Bota Fora)
- OAE's
- Passarelas
- Passeio em concreto esp.=5cm
- Pastilha de porcelana 25x25mm, Linha Colors, Jatobá
- Pavimentação em Concreto Simples (fctM,k=4,5 Mpa) com forma deslizante inclusive transporte até 5 km
- Pavimento Flexível
- Pavimento Rígido
- Perfil Metálico I 220 x 60 - SAC 350 ou similar
- Perfuração e montagem de tirantes em solo, inclusive fornecimento de acessórios

- Pilar em concreto pr moldado $\varnothing 40\text{cm}$ e $L=2,70\text{m}$
- Pintura de liga o, exclusive RR-2C
- Pintura em esmalte sint tico sobre ferro, duas dem os
- PISO CER MICO em placas, anti- derrapante PEI IV, assentado sobre argamassa de cimento colante, inclusive contrapiso
- PISO CER MICO t til de alerta, anti-derrapante, assentado sobre argamassa de cimento colante, inclusive contrapiso
- Piso em chapa corrugada
- Placa de sinaliza o vi ria permanente
- Placas indicativas da obra
- Plantas E Situa o
- Plantio de arbustos rasteiros
- Plantio de  rvores ornamentais de pequeno porte
- Plantio de  rvores ornamental de grande porte
- Po o de visita - PVI 02
- Po o de visita - PVI 09
- Porta de alum nio n  25 com anodiza o natural acetinada, veneziana, de abrir, 01 folha 0.90m X 2.10m, com contramarco e ferragens
- Porta de alum nio n  25 com anodiza o natural acetinada, veneziana, de abrir, 02 folhas 2.10m X 2.10m, com contramarco e ferragens
- PORTA interna de madeira de uma folha com batente, guarni o e ferragem, 0,70 x 2,10 m
- PORTA interna de madeira de uma folha com batente, guarni o e ferragem, 0,80 x 2,10 m
- Port o em perfil tubular de a o inox - $\varnothing 2"$, inclusive grade palito (Bloqueio PNE)
- Raspagem e limpeza manual do terreno - OAE e OC
- Reaterro compactado de vala com mat. reaproveitado
- Rebaixamento de len ol fre tico com ponteiras filtrantes
- Recepcionista
- Rede "wireless" para comunica o com equipamentos embarcados em  nibus
- Regulariza o de fundo de vala
- Regulariza o e compacta o do subleito
- Rel De Planejamento
- Rel Planejamento
- Remo o mecanizada de camada granular de pavimento (base e sub-base)
- Retirada de tubos de concreto com di metro at  0,80m
- Rodap  em GRANITINA para revestimento de piso moldado "in loco"
- Sa da das caixas (bueiro simples tubular $\varnothing 60$)
- Sarjetas canteiro central concreto - SCC 03
- Sarjetas de Canteiro (STC-01 Triangular)
- Sinaliza o
- Sistema De Informa es E Dados
- Sistema de Monitoramento de Imagens
- Sistema Embarcado em  nibus
- Spda
- Terraplenagem
- Tirantes Fred diam. 22,2mm
- Transporte BGTC at  5 km inclusive carga e descarga
- Transporte em caminh o de mat. 1a. Cat. At  a dist ncia de 5 km - Bota fora
- Trecho Catetinho - Parkway
- Trecho Esta o Metr  - Tas
- Trecho Gama - Catetinho
- Trecho Parkway - Asa Sul
- Trecho Santa Maria - Catetinho

- Urbanização
- Verga de concreto armado
- Viga em aço perfil I (W 610 x 174,0)