

Aplicação da Lei de Benford em Auditoria de Bases Governamentais

Marcelo H. Oliveira¹, Leonardo V. Ribeiro¹

¹Instituto de Educação Continuada (IEC)
Pontifícia Universidade Católica de Minas Gerais (PUC MINAS)
Belo Horizonte – MG – Brasil

{marcelohonoliveira, leonardo.leovilela}@gmail.com

Abstract. *This article describes the use of Benford's Law for the analysis of data in order to find possible deviations that could indicate fraud by manipulation the displayed numbers. To demonstrate its application in auditing public data were analyzed. The data published by the Federal Government of Brazil used in this article came from the databases of the programs “Minha Casa, Minha Vida” and “Programa de Aceleração do Crescimento”.*

Resumo. *Este artigo descreve a utilização da Lei de Benford para a análise de dados com o objetivo de encontrar possíveis desvios que pudessem indicar fraude por manipulação dos números apresentados. Para demonstrar sua aplicação em auditoria, foram analisados dados públicos divulgados pelo Governo Federal do Brasil: base de dados do programa “Minha Casa, Minha Vida” e “Programa de Aceleração do Crescimento”.*

1. A Lei de Benford

A Lei de Benford, intitulada em homenagem ao físico Frank Benford, que trabalhou na teoria em 1938, é a teoria matemática de dígitos iniciais. Especificamente, a ocorrência do (s) dígito (s) inicial (ais) é dada conforme uma distribuição específica e não uniforme.

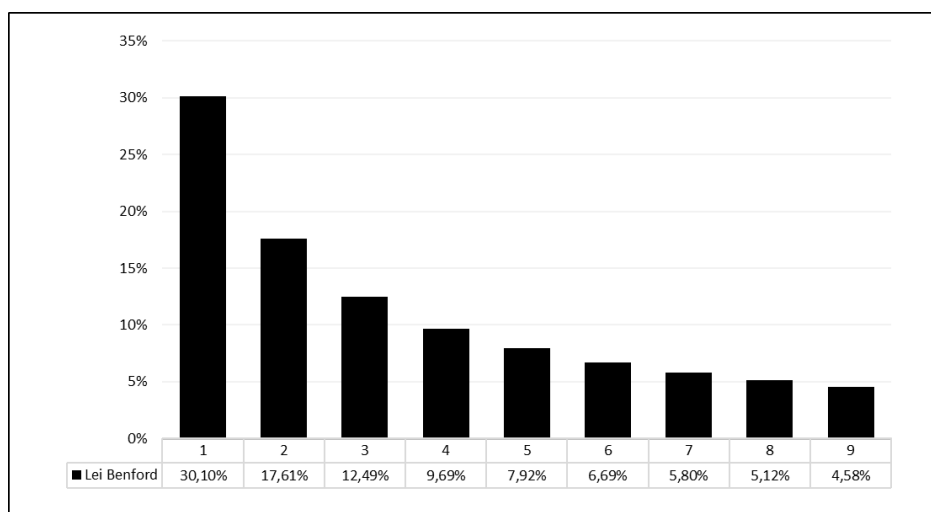


Figure 1. Lei de Benford e a distribuição do primeiro dígito

Embora se possa pensar que o número 1 seria exibido como o primeiro dígito 11% das vezes (isto é, um dos nove números possíveis), ele parece cerca de 30% das vezes (ver figura 1). O número 9, por outro lado, é o primeiro dígito em menos de 5% das vezes. A teoria abrange o primeiro dígito, o segundo dígito, os primeiros dois dígitos, último dígito e outras combinações de dígitos já que a teoria se baseia no logaritmo da probabilidade de ocorrência de dígitos sob a fórmula: $P(n) = \log_{10} \left(1 + \frac{1}{n}\right)$.

A Lei de Benford é válida para um conjunto de dados que cresce exponencialmente, mas também parece ser verdade para muitos casos em que um padrão de crescimento exponencial não é óbvio. Ele é melhor aplicado a conjuntos de dados que variam em diversas ordens de magnitude (por exemplo, as populações de vilas ou cidades, as distribuições de renda).

A teoria não é verdadeira para conjuntos de dados em que os dígitos estão predispostos a começar com um conjunto limitado de dígitos. Por exemplo, a Lei de Benford não será uma realidade para conjuntos de alturas de humanos, pesos de humanos e pontuação do Quociente Intelectual (QI). Outro exemplo seria pequenos créditos de seguros (por exemplo, entre US \$ 50 e US \$ 100). A teoria também não é verdadeira quando um conjunto de dados abrange apenas uma ou duas ordens de magnitude.

Para facilitar o entendimento, seguem algumas regras básicas:

- Os números devem ser gerados de forma natural;
- Os números não devem ter valores máximo e mínimo definidos;
- Devem possuir, no mínimo, 4 dígitos;
- Devem possuir mil registros pelo menos.

2. Cenários de Aplicação

Em quais circunstâncias seria interessante o uso da Lei de Benford? Daria para identificar possíveis fraudes ou manipulação má intencionada de dados numéricos? Sim! Quase desde o início, os defensores da Lei de Benford sugeriram que seria uma ferramenta útil para detecção de fraude e seu principal cenário é em auditorias.

Verificar o comportamento dos dígitos de números em relatórios fiscais e/ou financeiros é uma técnica prevista em muitos softwares de auditoria para evitar fraudes. Esses softwares fazem uma inspeção na massa de dados e, caso seja encontrado um comportamento inesperado, os auditores são notificados e convocados a uma inspeção mais detalhada.

Pontos de atenção após a aplicação da Lei de Benford:

- Se há mais dígitos baixos do que o previsto na lei, há indícios que os números foram modificados para baixo;
- Se os dados foram inventados, espera-se encontrar mais dígitos 5, 6 e 7;
- Se os números originais foram alterados, espera-se encontrar mais ocorrências de números que iniciam com 1 a 5.

3. Análise de bases do governo brasileiro

Diante de um contexto onde se tem colocado em dúvida como o governo brasileiro em suas diversas esferas tem gerido o dinheiro público, esse trabalho fez uso de dados abertos divulgados pelo próprio governo federal no portal **dados.gov.br**. O portal tem o objetivo promover a interlocução entre atores da sociedade e o governo para pensar uma melhor forma de uso dos dados e o acesso a essas informações é previsto pela Constituição Federal e na Declaração Universal dos Direitos Humanos.

Para realizar a aplicação da Lei de Benford, foram utilizados dados do Ministério do Planejamento, Desenvolvimento e Gestão (MP) - órgão integrante do poder Executivo Federal, que atua basicamente na formulação do planejamento estratégico nacional e na avaliação dos impactos socioeconômicos das políticas e programas do governo federal e na elaboração de estudos especiais para a reformulação de políticas.

Foram coletados e analisados os dados de dois programas do governo federal a saber:

1. Minha Casa Minha Vida (MCMV) – Empreendimentos do Programa Minha Casa, Minha Vida. Empreendimentos pertencentes a Carteira do PAC referente a março de 2012 – *Valor do Imóvel*.
2. Programa de Aceleração do Crescimento (PAC) – Os dados do 3º balanço de 2015-2018, referente a junho de 2016, e fornecem informações acerca de todos os empreendimentos presentes na carteira de projetos do programa – *Valor da Obra*.

A escolha dos dados destes programas foi intencional dada a visibilidade em âmbito nacional. Cada base possui um número considerável de registros e seus valores atendem às premissas da Lei de Benford listadas no tópico 1.

De cada base, foram extraídas as colunas de interesse e aplicadas nelas a análise da quantidade de ocorrências dos números de 1 a 9 como primeiro dígito do valor. A distribuição encontrada foi comparada com a distribuição de Benford.

3.1. Ferramenta de Análise e Resultados

A análise das bases a partir da aplicação da Lei de Benford foi realizada com o auxílio de uma linguagem de programação denominada “R” e um ambiente de desenvolvimento integrado para cálculos estatísticos e gráficos denominado “R Studio”.

Além das funções básicas e nativas do software R, é possível incluir conjuntos de funções específicas denominados Pacotes. Para essa análise, foi adicionado o pacote “*benford.analysis*” cujo objetivo é prover algumas funções que facilitem a validação de dados utilizando a Lei de Benford.

A seguir, serão apresentados os gráficos que demonstram a distribuição da ocorrência do primeiro dígito para os campos de interesse. Para cada massa de dados, foram plotados gráficos que comparam a distribuição da amostra com a distribuição prevista por Benford.

Os gráficos resultantes se encontrarão nos tópicos abaixo. Os dados de interesse estão representados pelas barras e ‘x’ e os valores esperados pela Lei de Benford pela linha tracejada.

3.2. Programa Minha Casa, Minha Vida

Criado em 2009, o Programa Minha Casa, Minha Vida veio com o objetivo de tornar a moradia acessível às famílias organizadas por meio de cooperativas habitacionais, associações e demais entidades privadas. A base de dados é constituída por uma listagem dos imóveis repassados a partir do programa até março de 2012 e foram analisados o valor final do imóvel.

Seguem os gráficos plotados pela ferramenta:

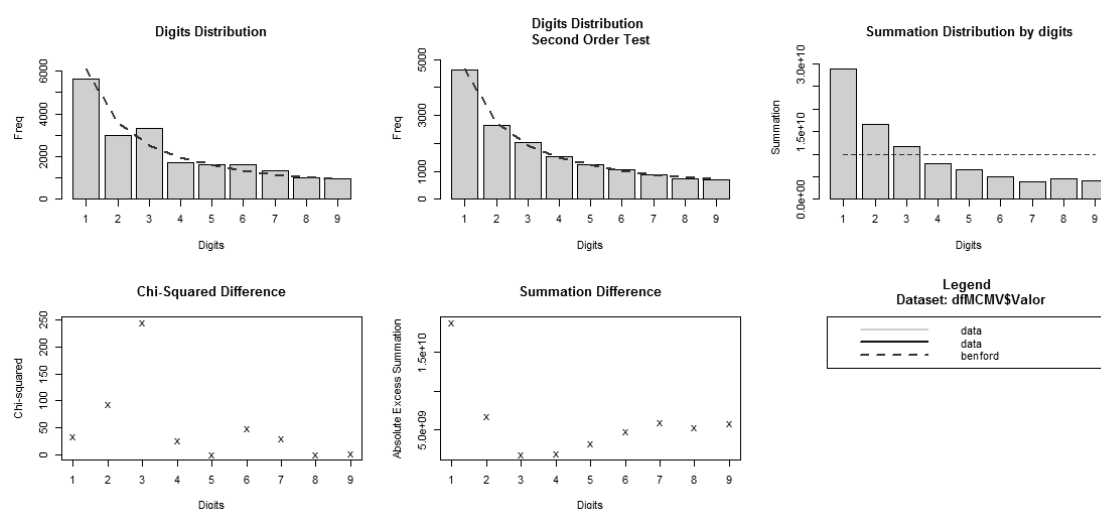


Figure 2. Análise de Benford para o campo “Valor” – Base MCMV

O primeiro gráfico se refere à contagem de observações com relação ao seu primeiro dígito, comparando-a com o valor esperado pela Lei de Benford. Percebe-se que os dados do governo se ajustam à Lei, mas, também, que há um desvio discreto no dígito 3! O segundo gráfico é análogo ao primeiro, mas faz esta contagem para a diferença dos dados ordenados e não demonstra nada estranho. Por fim, o terceiro gráfico tem um objetivo diferente e, em geral, não se deve esperar encontrar um bom ajuste dos dados à reta tracejada.

3.3. Programa de Aceleração do Crescimento

Criado em 2007, o Programa de Aceleração do Crescimento – PAC planeja e executa grandes obras de infraestrutura social, urbana, logística e energética do país. A base de dados é a lista geral de todas as obras da carteira de projetos do programa, cada registro indica uma obra e foi analisado o valor total do investimento.

A análise é análoga a do tópico anterior a partir do mesmo modelo de gráficos fornecidos pela ferramenta:

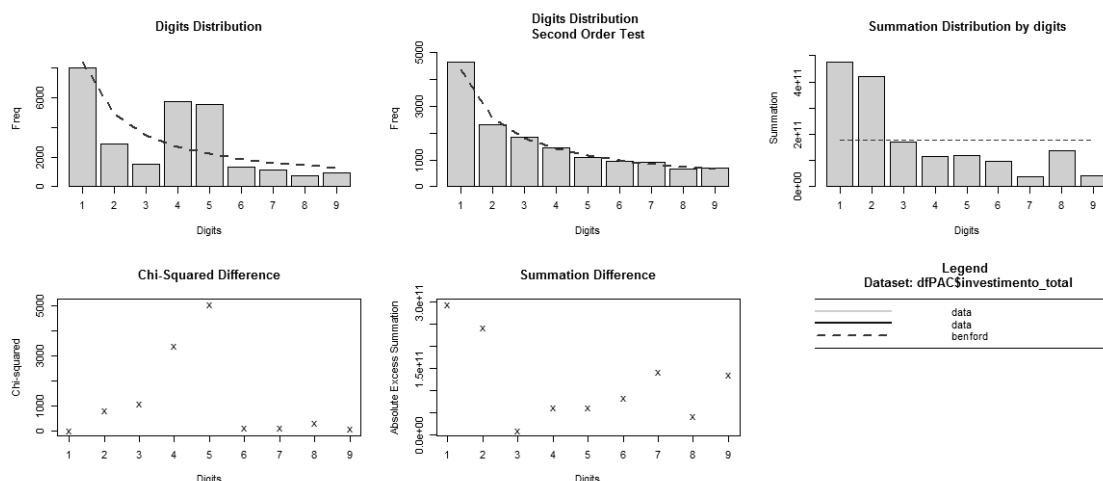


Figure 3. Análise de Benford para o campo “investimento_total” – Base PAC

Nesta segunda amostra, foram encontrados desvios claros para os dígitos 2, 3, 4 e 5. Tal comportamento denota uma forte possibilidade de incorreção dos dados ou por coleta incorreta ou por manipulação intencional. Neste caso, seria indicada uma inspeção detalhada nos valores dos investimentos com o foco nas obras cujo valor inicia-se com esses dígitos.

Outra saída interessante da ferramenta são as Estatísticas da Mantissa conforme Tabela 1 abaixo. A mantissa é a parte decimal do logaritmo e tais estatísticas tem um valor esperado quando a base analisada possui dados realmente aleatórios.

Tabela 1. Comparação das Estatísticas da Mantissa

Item	Valor Esperado	Valor Encontrado
média	0,5	0,516
variância	0,083	0,073
curtose	1,2	-1,095
assimetria	0	-0,279

Nota-se que dois itens, curtose e assimetria, estão bem diferentes e colabora para demonstrar se uma base de dados de interesse tem características suspeitas.

4. Conclusão

Frente ao que o trabalho expôs, constata-se que a Lei de Benford é um método aderente à detecção de erros em bases de dados numéricos, seja por erros de digitação ou fraudes contábeis ou financeiras no contexto de uma auditoria. Dessa forma, comprova-se a eficácia do uso da teoria da Lei de Benford na detecção de interferências nos dados.

A aplicação desta metodologia, especialmente em dados financeiros das organizações (governamentais ou não), possibilita uma maior profundidade de detalhes para o acompanhamento e controle das suas divulgações. Isso permite que se possam detectar problemas e distorções de resultados, não deixando que qualquer desvio crie

consequências negativas e insustentáveis no futuro destas organizações. Nesse cenário, recomenda-se com trabalhos futuros a utilização das ferramentas aqui citadas em outras bases de dados visando avaliar o desempenho de outros órgãos públicos.

Referências Bibliográficas

Singleton, T. (2011) “Understanding and Applying Benford’s Law”, In: Isaca Journal - Volume 3 - IT Audit Basics, United States of America. Publishing Press.

Portal Brasileiro de Dados Abertos (2016) – Ferramenta Web de acesso livre, <http://dados.gov.br>, Outubro – Acesso em 27 de out. de 2016.

Tribunal de Contas da União – Portal Inova TCU (1995) “Aplicações da Lei de Benford à auditoria de obras públicas”, <http://portal.tcu.gov.br/innovatcu/noticias/aplicacoes-da-lei-de-benford-a-auditoria-de-obras-publicas.htm>, Maio – Acesso em 27 de out. de 2016.