Physica A 567 (2021) 125626

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

A Benford's Law based methodology for fraud detection in social welfare programs: Bolsa Familia analysis

Caio da Silva Azevedo, Msc [a,*], Rodrigo Franco Gonçalves, Phd [b], Vagner Luiz Gava, Phd [c], Mauro de Mesquita Spinola, Phd [a]

[a] *University of São Paulo, Brazil*
[b] *Universidade Paulista, Brazil*
[c] *Instituto de Pesquisas Tecnologicas, Brazil*

A R T I C L E I N F O

Article history:
Received 19 June 2020
Received in revised form 26 September 2020
Available online 31 December 2020

Keywords:
Big data analytics
Newcomb–Benford's Law
Statistical antifraud analysis
Anomaly detection
Bolsa Familia
Social welfare programs

A B S T R A C T

This paper aims to introduce a data science approach for guiding auditors to accurately select regions suspected of frauds in welfare programs benefits distribution. The technique relies on Newcomb–Benford's Law (NBL) for significant digits. It has been analysed Bolsa Familia data from Federal Government Transparency Portal, a tool that aims to increase fiscal transparency of the Brazilian Government through open budget data. The methodology consists in submit four data samples to null hypothesis statistical methods and thereby evaluate the conformity with the law as well as the summation test which looks for excessively large numbers in the dataset. Research results in this paper are that beneficiaries' cash transfer per se is not a good test variable. Besides, once payment data are grouped by municipalities, they fit NBL, and finally, when submitted to the summation test, the distribution of the Bolsa Familia payments in several municipalities shows some fraud evidence. In this sense, we conclude the NBL can be an appropriate method to fraud investigation of welfare programs' benefits distribution having beneficiaries' payment geographically grouped.

1. Introduction

Newcomb–Benford's Law (NBL) is that kind of counter-intuitive law. Some authors compared it with Newton's law of gravitation, saying it is a more simple observation of reality rather than a provable mathematical result [1].

This law states that lower first significant digits, also known as, leading digits occur more frequently than higher ones in natural phenomena, contradicting the common sense that all digits occur with the same frequency in a uniform distribution. For example, in a set of numerical data or anomalous numbers as postulated by Benford [2], the number 1 would appear as the most significant digit about 30% of the time, while digit 9 would appear less than 5% of the time, such as demonstrated in the first 3500 Fibonacci numbers on Fig. 1.

Nigrini's works [3–5] propagated among scholars and accounting practitioners that NBL can be used as a forensic accounting and auditing tool for financial data. Since then, it has been used as a sophisticated statistical technique in the fraud detection process.

* Correspondence to: University of São Paulo, Polytechnic School, Production Engineering Department, Av. Prof. Luciano Gualberto, 1380, Butantã, São Paulo, 05508-010, Brazil

E-mail addresses: caio.aze@usp.br (C. da S. Azevedo), rodrigo.goncalves@docente.unip.br (R.F. Gonçalves), vlgava@ipt.br (V.L. Gava), mauro.spinola@usp.br (M. de M. Spinola).

https://doi.org/10.1016/j.physa.2020.125626
0378-4371/

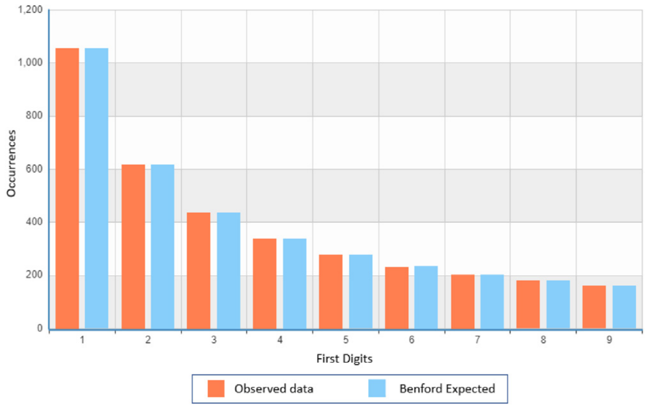

Fig. 1. First digit occurrences in the Fibonacci sequence observed data and NBL distribution.
Source: Created by the author, 2019.

In 2004, the Brazilian Government instituted the Bolsa Familia as a social assistance program. Bolsa Familia is a welfare program which is the flagship of social protection to the poor by providing financial aid to low-come families. By the time of our research, it reached approximately 13.4 million households monthly, corresponding to the poorest population of Brazil. Its primary goals are to fight hunger and poverty; strengthen access to the public service network, especially to education, health, and social assistance; promote intersectionality integration and public policy synergy, and encourage sustained empowerment of beneficiary families (Brazil 2004).[1]

The program consists of conditional cash transfer being given preferentially to females in exchange for children attending school and receiving regular medical check-ups in public clinics.

Corruption is a significant and traditional problem in Brazilian politics [6] and not even social programs are free from this unpleasant situation. There is suspicion of corruption and fraud on preventing beneficiaries from receiving 100% of the money and local officials could be reporting inaccurate information on eligibility to receive kickbacks. Hence, there is the need to study data anomalies, if any, on reported information provided by Federal Government Transparency Portal, a tool that aims to increase fiscal transparency of the Brazilian Government through open budget data, including monthly benefits payments all over the country.

Cerioli et al. [7] have been shown that NBL is a valuable starting point for forensic accountants and it is also applicable in account contexts, such as external, internal, and governmental auditing.

In view of the above, this study is based on the following questions: Is the accounting model, based on the Newcomb–Benford Law, applicable to the Bolsa Familia program benefits payment? How to be timely and accurate on select regions suspected of frauds on benefits distribution?

The objective of the study was to evaluate the application of NBL to fraud detection in the context of Bolsa Familia regular payments in order to guide auditors to accurately select regions under suspicion of frauds in welfare benefits distribution.

The paper is organized as follow: first, we present the literature review about NBL and its fundamental. Then we present the research method (Section 3). Section 4 contains a description of the data. The findings are collected in Section 5 and the discussion in Section 6. The last section also allows us to offer suggestions for further research lines.

2. Literature review and hypothesis development

Although widely used over, especially after Nigrini's and other authors publications, as described by Cunha [8], NBL was first announced more than a century ago, when the American astronomer and mathematician Newcomb [9] published the first known article in the American Journal of Mathematics, about what is today known as Newcomb–Benford's Law. As said by Benford [2] ''It has been observed that the first pages of a table of common logarithms show more wear than do the last pages, indicating that more used numbers begin with the digit 1 than with the digit 9''.

In his paper, Newcomb [9] questioned about the probability that in a set of natural numbers one of them be randomly taken and its first significant digit is n, its second n′, etc.

Newcomb [9] concluded an intriguing fact that in many natural and human phenomena the leading — that is, the first significant digits are not uniformly scattered, as one could naively expect, but follow a logarithmic-type distribution.

Despite its relevance, Newcomb [9] provided no theoretical explanation, and 57 years later, in 1938, Frank Benford came with the article ''The Law of Anomalous Numbers'' published in American Philosophical Society.

[1] Source: Decree No. 5.209, of 17 September 2004, regulates Law No. 10.836, which creates the Bolsa Família program and sets out other provisions, http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2004/Decreto/D5209.htm (accessed June 22, 2019).

Frank Benford was a General Electric Company physicist which assembled numbers from sources as diverse as Readers' Digest articles, the first 342 street addresses of American Men of Science, atomic weights, population sizes, death rates, back body radiation and physical constants [2]. These sets of data showed that leading digits from a wide range of sources produced an uncanny adherence to the logarithmic rule that Newcomb had penned, apparently unnoticed, decades earlier. More than Newcomb's observation tests, and based on substantial empirical evidence, Benford stated a base-10 logarithmic law of frequency of leading digits. This law gives

$$P\left(d\right)=\log_{10}\left(1+\frac{1}{d}\right),d=1,2,\ldots,9 \tag{1}$$

where P is the probability of leading digit (d) occurrence.

That means that, given a randomly selected natural number, it should begin with the digit 1 in about 30% of the time: more precisely, the proportion should be 0.301. The frequency of numbers with leading digit 2 should be about 18% (obtained from 3/2), those with leading digit 3 should be about 12% (from 4/3), and so on until the frequency of 8's should be 5.1% and that of 9's should be 4.6%.

Almost six decades later, the first rigorous proof of Benford's Law was developed by Theodore Hill in 1995. In his work Hill proves, based on probability theory, that scale invariance implies base invariance and base invariance, in turn, implies the Benford's Law. In another paper, Hill [10] expanded the law not only for the first digit but also to the others.

$$P\left(D_1=d_{1,\ldots,}D_k=d_k\right)=log_{10}\left(1+\frac{1}{\sum_{i=1}^{k}10^{k-i}d_i}\right) \tag{2}$$

where d1 ∈ {1, …, 9}, and all other dj ∈ {0,…,9}, j = 2,…,k.

Thus, the probability of any combination of digits can be easily found using the expression (2) as exemplified:

$$P\left(D_1=3,D_2=1,D_3=4\right)=log_{10}\left(1+\frac{1}{(314)}\right)=log_{10}\left(\frac{315}{314}\right)\cong 0.0014$$

A corollary derived by (2) is that the significant digits are dependent, Hill [10].

Based on researches in the field of probability theory, Hill [10–12], Pinkham [13] and Raimi [14] demonstrated that NBL data sets contain the following properties: (a) scale and base invariance; (b) come from a choice provided by a variety of different data sources. This result is reached from a rigorous analysis of central limit theory in form of theorems to mantissas of random variables over multiplication effect. Hence, as the number of variables grow, the density function tends to a logarithm distribution. Hill [12] demonstrated "random samples from random distributions", i.e. a data set collected from arbitrary samples provided from a variety of different distributions is the distribution of Newcomb–Benford [15]. Pietronero et al. [16] and Whyman et al. [17] also demonstrate NBL scale invariance and that it is the most usual among other natural data sets whose distributions are scale-invariant as well. Moreover, they propose NBL generalization in terms of power laws to a series of numbers generated from more complex systems. And finally, they highlight the relation between NBL and Zipf's law.

Aiming to meet the proposed problem and to achieve the general objective of this research, it is applied hypothesis testing based on Chi-square ($\chi 2$), Z-statistic and Mean Absolute Deviation (MAD) tests to validate observed frequency of leading digits on sample datasets (P) against predicted frequency according to NBL (P0). Nevertheless, there are other methods, such as the work of Biau [18] that measures data set discrepancies with NBL using Shannon's entropy. Moreover, the summation test is used to identify numbers that are large compared to the norm for that data.

Nigrini [4,5,19], assuming that trustworthy accounting data reporting individual incomes follows NBL, argue that meaningful deviation in regard of the law suggests possible fraud or manipulated data. Thus, the presentation and demonstration of the law, as a powerful methodology in the audit field, were further emphasized by Hill [20], Pinkham [13], Raimi [21], Durtschi et al. [22], among others, and also in Nigrini and Miller [23], Pimbley [24], Amiram et al. [25], Ausloos et al. [26]. In fact, NBL is also applied outside the financial audit realm; e.g. see Fu et al. [27] for image forensics, Ausloos et al. [28] for birth rate anomalies, Pollach et al. [29] for maternal mortality rates, or elsewhere in the natural sciences Sambridge, et al. [30], and on religious activities Mir [31]. In Brazil, as proposed by Gamermann and Antunes [32], NBL was used to analyse money donations for the electoral campaigns and the election results. That work shows either possible manipulations in the declarations or a significant number of donations that might not have been spontaneous from the donors. Like our study, governmental indexes were analysed by Zanetti et al. [33]. They used NBL to validate Argentine inflation statistics from 2006 to 2015 as they suppose these numbers were manipulated by the government.

It is noticed that there is a wide literature: see Alali and Romero [34], for approximately the last decade, Costa et al. [35], or Beebe [36], which contains a rather exhaustive list of references.

As suggested by Nigrini [5], the requirements for conformity are that the data should represent the sizes of facts or events, such as the populations of towns and cities, the flow rates of rivers, or the sizes of heavenly bodies. Financial examples include market values, companies' revenues, or daily trading volumes.

There should be no built-in minimum or maximum values in the data, except that a minimum of zero is acceptable for data that can only be positive numbers, for instance, election results, population counts, or inventory counts.

Data should not be numbers used as identification numbers or labels, such as social security numbers, bank account numbers, and flight numbers. The data should have more small numbers than larger numbers, which implies that data

Table 1
Nonmathematical guideline for determining whether a data set should follow NBL.
Source: Wallace [35] and Nigrini [5].

No minimum or maximum values	No numbers used as identification, such as social security, bank account or flight numbers
At least 1000 records	Four or more digits
Mean should be greater than the median and positive skewness	

should not be too clustered around its mean value. Salary data, for instance, does not conform to Benford's Law because most people in the same organization are paid approximately the same amount Nigrini [5].

In addition, the data set size, as proposed by Nigrini [37], in a general rule should be at least 1000 records. Tests can be run with less than this, but it is likely to have larger deviation from the NBL. The author also shows by his research that each numeric amount should have four or more digits for a good fit. When numbers have fewer than four digits there is a slightly bias in favour of the lower digits.

Moreover, Wallace [38] suggested that if the mean of test variable in a data set is larger than the median and skewness value is positive, the data set is likely to be suitable for Benford's Law distribution (see Table 1).

The present study uses financial statements from benefits payments of Bolsa Familia Program on their published accounts and balance files for examining the reliability of reported numbers using Benford's Law distribution. For this purpose, the null following hypotheses were considered for which it is desired to prove the falsehood.
H_{0A}: The distribution of payment values significant digits conforms to Benford's law
H_{0B}: Once dataset conforms to NBL (H_{0A}) its proportions conform to summation test

3. Methodology

As described by Benford [2], the geometric foundation of the anomalous numbers means that its data series will have Benford-like properties if ordered records approximate a geometric sequence.

This study uses the financial statement items like benefit payments among Brazilian municipalities population for examining reliability using Benford's Law distribution. It submits the samples to a set of statistical tests proposed by Nigrini [5,35], i.e., Chi-square ($\chi 2$), Z-statistic and Mean Absolute Deviation (MDA) to validate their distribution according to Benford-Law. Finally, the summation theorem, developed by Nigrini as a Ph.D. student, is used to look for excessively large numbers.

3.1. Statistical significance level

Every statistical analysis lay down whether results have statistical significance according to pre-established limits. Significance levels show how likely a pattern in your data is due to chance. The significance level, also denoted as alpha or α, is the probability of rejecting the null hypothesis when it is true. The most common level used to mean something is good enough to be believed is 95%. Thus, a significance level of 95% indicates a 5% risk of concluding that a difference exists when there is no actual difference. The researcher determines the significance level before experimenting.

3.2. Statistical tests

This study uses the financial benefit payments on four different periods, January 2018, March 2018, January 2019, and March 2019 for examining the reliability using Benford Law distribution. According to Nigrini [5], numbers beginning with low first digits (1, 2 and 3) are more frequent than numbers starting with higher first digits (such as 7, 8 and 9). Important to mention that the first digit of a number is the leftmost digit, and 0 is not acceptable for it. For instance, numbers 64 774, 0.00717, the first digits are 6 and 7, respectively. In the case of first-two digits, they are 64 and 71.

Our analysis uses three statistics tests – Chi-square ($\chi 2$), Z-statistic and Mean Absolute Deviation (MAD) for examining the statistical significance.

Those tests are quite different, first, the Z-statistic check whether the individual distribution significantly differs from an expected distribution, i.e. Benford's Law distribution. Individual distribution refers to a digit, first-two digits, or any other first-n digit combination. Mathematically, the Z-statistic considers the absolute magnitude of the difference (the numeric distance from the actual to the expected), the size of the data set, and the expected proportion. The formula adapted from Fleiss et al. [39] is shown in Eq. (3).

$$Z = \frac{(|p - p_0|) - \left(\frac{1}{2n}\right)}{\sqrt{\frac{p_{0(1-p_0)}}{n}}} \tag{3}$$

where P is the observed frequency of leading digits.

Table 2
Critical values to chi-square test applied to first and first-two digits.
Source: Adapted from Engineering Statistics Handbook[2].

Significance level	First digit test	First-two digits test
0.01	20.090	122.942
0.05	15.507	112.022
0.10	13.362	106.469

p_0 is the expected frequency under Benfordś law.
n is the number of records.

The term $\left(\frac{1}{2n}\right)$ is a continuity correction term and should be used only when it is smaller than ($|p\text{-}p_0|$).

If the values of Z-statistic exceed the critical value 1.96, the null hypothesis (H_{0A}) is rejected at 5% of significance level, the most usual level as proposed by Nigrini [5]. It is also possible to adopt critical values of 2.57 and 1.64 to 1% and 10% significance levels, respectively.

Let us consider a significance level of 5 per cent. This means that we should not be too concerned if we have four or five (90 possible first-two digits 5 per cent) significant spikes for any data set, Nigrini [5].

Second, chi-square is used to test the statistical significance to the whole distribution in the observed frequency of first digit and first two digits against their expected frequency under Benford's Law. The null hypothesis (H_{0A}) is that the digits conform to Benford's Law. The chi-square statistic is calculated as is shown in Eq. (4).

$$Chi - square = \sum_{i=1}^{K} \frac{(Pi - P0i)^2}{Pi} \tag{4}$$

where P and P0 represent the actual count and expected count respectively, and K represents the number of bins (90, because of there are 90 first-two digits). The number of degrees of freedom is K–1, which means that for the first-two digits, the test is evaluated using 89 degrees of freedom. A calculated chi-square is compared to a critical value and the higher it is, the more data deviates from Benford's Law.

As well as Z-statistic, Chi-square also uses 5% of significance level as Nigrini [5] recommendation for Benford's Law validation, in this case, 15.507 and 112.022 as critical values for first and first-two digits tests, respectively. Once they exceed, the null hypothesis (H_{0A}) is rejected.

In this work, just like Z-statistic, it is assumed 5%, although chi-square limits may also consider 1%, 5% and 10% as detailed in Table 2, extracted from Engineering Statistics Handbook.[2]

Lastly, there is the Mean Absolute Deviation (MAD) test. This test, as opposite as previous ones ignores the dataset size and thus it is indicated by Nigrini [5] to large databases. It is mathematically expressed as:

$$\text{MAD} = \frac{\sum_{i=1}^{K} |P_i - P0_i|}{K} \tag{5}$$

where K means the number of bins (which equals 90 for the first two digits), P represents the actual proportion and P_0 the expected proportion under NBL.

The absolute symbol means that the deviation is given a positive sign irrespective of whether it is positive or negative. The absolute deviations need to be added together and divided by the number of bins, i.e. the average (or mean) absolute deviation.

However, there are no objective critical scores for the MAD test, Nigrini [5] proposed critical scores for nonconformity, for conformity, and for some in-between categories based on personal experience with everyday data tables that were tested against Benford's Law as detailed in Table 3. This table gives the critical values for the first and first-two digits tests.

3.3. Summation test

Classified by Nigrini [5] as one of the advanced Benford's Law tests, as well as the second-order test, the summation test looks for excessively large numbers in a dataset. It identifies numbers that are large compared to the norm for that data. The test was proposed by Nigrini [5] and it is based on the fact that the sums of all numbers in a Benford distribution with first-two digits (10, 11, 12, …99) should be the same. Therefore, each of the 90 first-two digits groups sum proportions should be equal, i.e. 1/90 or 0.011, and spikes indicate that there are some large single numbers or set of numbers. To the scope of this work, let us assume 25% of upper tolerance and thus the proportional critical value is 0.01375.

Nevertheless, Nigrini [5] verified that summation theorem has shown that real-world data sets seldom show the neat, once they have abnormal duplications of large numbers. In general, it is not possible to say whether the summation spikes are caused by a handful (one, two, or three) of noticeably big numbers or an abnormal duplication of a few hundred moderately big numbers without a closer look at the data.

[2] Source: NIST/SEMATECH e-Handbook of Statistical Methods, https://itl.nist.gov/div898/handbook/eda/section3/eda3674.htm (accessed May 24, 2019).

Table 3
Cut-off scores and conclusions for calculated MAD values.
Source: Adapted from Nigrini [5].

Digits	Range	Conclusion
First digits	0.000 to 0.006	Close conformity
	0.006 to 0.012	Acceptable conformity
	0.012 to 0.015	Marginally acceptable conformity
	Above 0.015	Nonconformity
First-two digits	0.000 to 0.012	Close conformity
	0.012 to 0.018	Acceptable conformity
	0.018 to 0.022	Marginally acceptable conformity
	Above 0.022	Nonconformity

Table 4
Withdrawal data file dictionary.
Source: Adapted from Brazilian transparency portal, 2019. Available at http://www.portaltransparencia.gov.br/pagina-interna/603401-dicionario-de-dados-bolsa-familia-saques.

Column	Description
Reference year/month	Payroll year/month
Accounting period year/month	Year/month of the payment
UF	Program beneficiary federation unit initials
Municipality SIAFI code	Program beneficiary municipality SIAFI code. SIAFI (Financial Administrative Integrated System)
Municipality name	Program beneficiary municipality name
Beneficiary NIS	Program beneficiary NIS. Created by Caixa Econômica Federal, NIS means a social identification number provided by Brazilian citizen join any social program, i.e., Bolsa Familia, FGTS, working papers, became INSS taxpayer or started labouring life for private or public companies. Source: Caixa Econômica Federal
Beneficiary name	Bolsa família program beneficiary name
Withdrawal date	Date in which the withdrawal was realized
Withdrawal value	Beneficiary payment value

4. Data collection

In this section, analysed data is explained. All data used in this study is publicly available on Federal Government Transparency Portal, a tool that aims to increase fiscal transparency of the Brazilian Government through open budget data.

4.1. Data

Real-life data are used to determine conformity to Benford's Law. Our results are meant to provide an assessment of the likelihood of misstatement or fraud.

The Transparency Portal provides both information — monthly provisioned payment as well as effective cash withdrawal made per month. This work uses withdrawal data, as it consists of real benefits amount monthly paid to the citizens. All data files were downloaded on May 10th, 2020, from Transparency Portal and they are available at http://www.portaltransparencia.gov.br/download-de-dados/bolsa-familia-saques. Brazilian government Transparency Portal also provides many other budget data files at http://www.portaltransparencia.gov.br/download-de-dados. However, we notice that data changes over time due to accounting adjustments and thus we provide datasets exactly as used in this work at: https://www.kaggle.com/caioazevedo/brazilian-welfareprogram-bolsafamilia-dataanalysis.

It was analysed four samples provided by Brazilian Transparency Portal — January and March of 2018 as well as January and March of 2019. All those files have the same data dictionary which is:

This work sample withdrawal values, in Brazilian national currency Real (R$). As a first analytical step, their frequencies of leading digits are calculated and goodness-of-fit tests (Chi-square, Z-statistic, and Mean Absolute Deviation) are used to study the degree of fit to Benford's Law. Besides, the summation test is used to identify numbers that are large compared to the norm for the dataset.

4.2. Analytic tool

Real-life data are used to determine conformity to Benford's Law. Our results are meant to provide an assessment of the likelihood of misstatement or fraud.

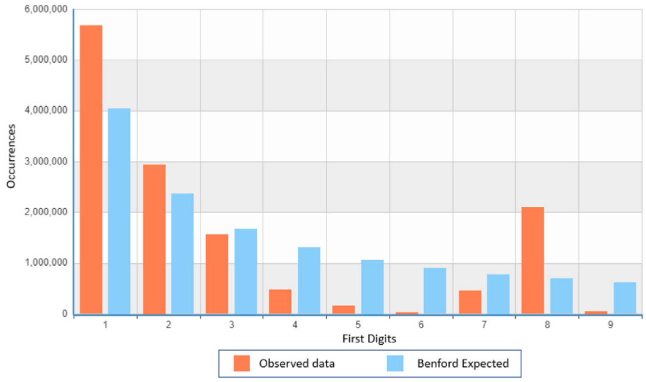

Fig. 2. Comparison through bar charts of January 2018 withdrawal values and NBL first digit occurrences. The orange bars show the actual proportions and the blue bars show the proportions of NBL.

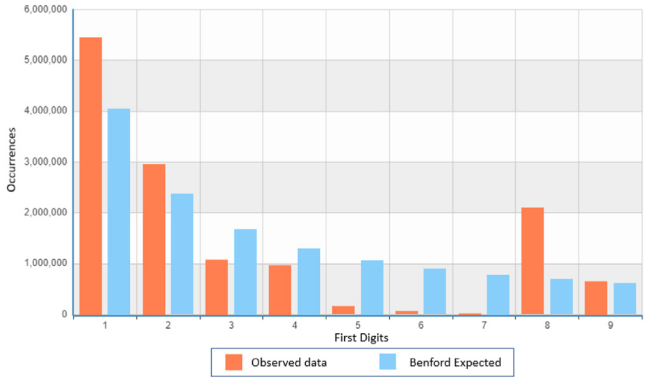

Fig. 3. Comparison through bar charts of January 2019 withdrawal values and NBL first digit occurrences. The orange bars show the actual proportions and the blue bars show the proportions of NBL.

5. Analysis of result

This section discusses the results. We have grouped the analyses into two perspectives. At first, we submitted withdrawal values to NBL tests. Once these tests failed, we changed our viewpoint and we geographically grouped data into the Brazilian municipalities and thus tested under NBL rules.

5.1. Exploratory data analysis

In this section, we report the result of Benford's Law analysis over test variable, which is the beneficiaries' withdrawal values. Initially "first digit analysis" is reported using Chi-square. Having this test succeed a second round is performed with "first-two digits analysis" with more accurate tests such as Z-statistic and Mean Absolute Deviation as well as Chi-square for first-two digits.

Figs. 2 and 3 illustrate original data first digit test results, having as test variable the beneficiaries' withdrawal values. These data are provided by Transparency Portal and they are regarding January 2018 and January 2019. Orange bars represent the observed distributions, the blue bars represent the Benford's law theoretical distribution. The bar charts show that all first significative digits (FSD) distributions are different from the expected NBL distribution.

In case of January 2018, the sample contains 13,442,529 records. It is verified that observed and expected first significative digits distributions are different as reinforced by chi-square test which value is 6,480,128.958 and overwhelmingly above the critical value of 15.507, having us used 5%, of significance level. Not differently, January 2019 sample, composed of 13,497,267 records do also have distributions far from goodness-of-fitting Benford's Law. Where chi-square test scored 6,180,301.189 and again far above the critical value of 15.507. Once the first tests failed, first-two digits tests were aborted.

At first glance it is easy to suppose evidence of fraud based on these observations, however, we realized that cash transfer value per se is not a good test variable, not only because the results point to a drastic breakdown of Benford's law but also and mainly because it failed on Benford's data set validation.

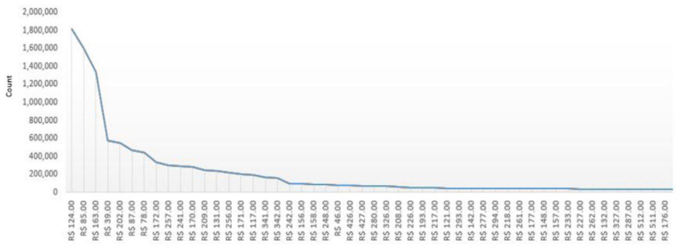

Fig. 4. Histogram of the number of cash transfer values per family in January 2018 .
Source: Created by the author, 2019.

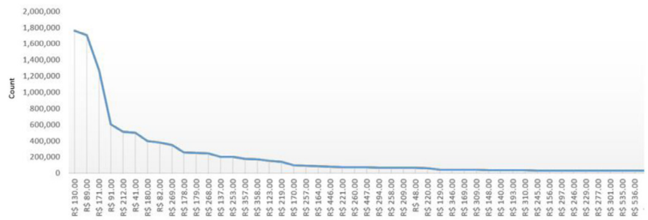

Fig. 5. Histogram of the number of cash transfer values per family in January 2019.
Source: Created by authors, 2019.

After first significative digit tests, we realized that cash transfer value does not conform to a Benford's data set. The reasons are that first, it contains a minimum value. In addition, most of the paid values concentrate on no more than five different values, as Figs. 4 and 5 demonstrate, and thus overall data clustered around its mean value. Finally, withdrawal value may be compared as salary data in the sense of most people receive the same amount and according to Nigrini [5], this data behaviour does not conform to Benford's Law.

The reasons above are corroborated by Brazil (2015),[1] which describe that each month the PBF transfers money to families living in extreme poverty (per capita family monthly income of up to R$ 89.00) or poverty (per capita family income between R$ 89.01 and R$ 178.00) through a bank card. The PBF grant structure varies according to the degree of the family's poverty and its age composition. In short, the program transfers a monthly amount to families living in extreme poverty that allows each family member to rise above the extreme poverty line (R$ 89.00). Poor families are eligible to take part in the program if they have children or adolescents up to the age of 17, in which case they receive the so-called variable grant – R$ 41.00 per child or adolescent aged between 0 and 15 years or pregnant or nursing woman, limited to five grants per family – and a variable grant of R$ 46.00 per adolescent aged between 16 and 17 who attends school, limited to three per family.

5.2. Another perspective

A significant breakthrough has been achieved when data are geographically aggregated into the 5570 Brazilian municipalities by summing up beneficiaries' payments for each municipality. Thus, instead of individual withdrawals values, our test variable henceforth is the amount each municipality have spent. Doing so, our analyses run over samples of the population, i.e. the beneficiaries, which represents a natural data set as pointed out by Nigrini [5].

In order to group data, original records were combined using R-Script aggregation statements. The script is available at GitHub repository at https://github.com/cazevedo1977/brazilian-welfareprogram-bolsafamilia-dataanalysis as well as CSV files generated from aggregated data. The CSV files (BolsaFamilia_Municipalities_Jan_2018.csv, BolsaFamilia_Municipalities_Mar_2018.csv, BolsaFamilia_Municipalities_Jan_2019.csv and BolsaFamilia_Municipalities_Mar_2019.csv) and their data dictionary is described on Table 5.

Original data files from January and March of 2018 and 2019 periods were aggregated as data dictionary described on Table 4 and afterwards municipalities withdrawal values leading digits frequencies are calculated and submitted to goodness-of-fit tests, as well as to summation test.

Figs. 6 and 7 illustrate aggregated data file from January 2018 withdrawal values first and first-two significative digits (FSD) frequencies respectively and the expected frequency according to Benford's Law. Following Wallace [38] and Durtschi et al. [20], we find that mean of withdrawal values are larger than the median and skewness value is positive. So, our data set is suitable for Benford's Law analysis.

The "first digit analysis" and "first-two digits analyses" are reported using Chi-square, Z-statistic and Mean Absolute Deviation, respectively. For individual digit distribution, Z-statistic test is used and for whole distribution, Chi-square and MDA tests are used.

Table 5
Geographically aggregated data dictionary.
Source: Created by the authors, 2019.

Column	Description
Accounting.period	Year/month of the payment
Municipality.Name	Program beneficiary municipality name
State	Brazilian state, i.e. a geographical region
Number.of.Beneficiaries	Number of beneficiaries that were paid in that municipality
Withdrawal.value	Municipality accumulated payment value

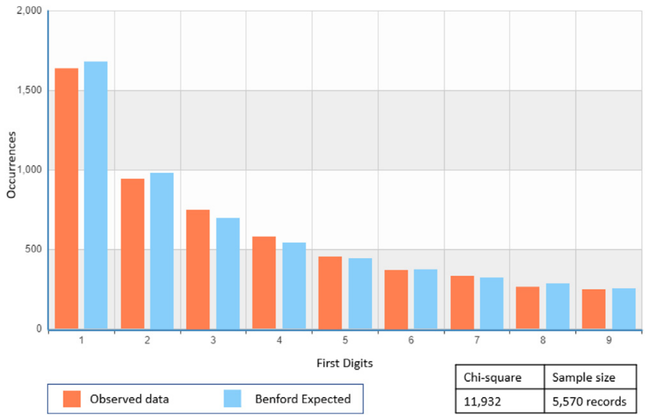

Fig. 6. Comparison through bar charts of January 2018 withdrawal values and NBL first digit occurrences. The orange bars show the actual proportions and the blue bars show the proportions of NBL.

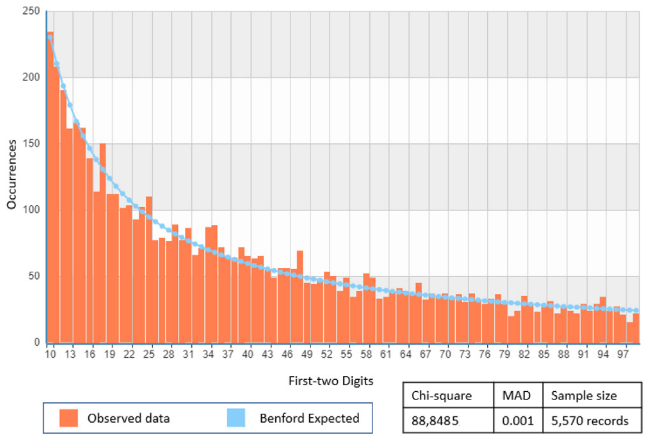

Fig. 7. Comparison through bar charts of January 2018 withdrawal values and NBL first-two digits occurrences. The orange bars show the actual proportions and the blue bars show the proportions of NBL.

The orange bars represent the observed distributions, while the blue bars and line represent the Benford's law theoretical distribution. The bar charts show that all FSD distribution fits the expected Benford's Law FSD distribution.

Additionally, statistical tests corroborate bar charts visual analysis as in case of first-digit analysis chi-square of 11.932 is under the critical value of 15.507 with 5% of significance level. Not differently, first-two digits analysis statistical tests are also under critical values, i.e., chi-square of 80.8485 less than the critical value of 112.022 when significance level equals 5%. Moreover, there are only four Z-statistics values above the cut-off of 1.969 which are 48 (2.6486), 35 (2.3590), 17 (2.0468) and 34 (1.9683) what do not clash NBL as suggested by Nigrini [5] where at 5% of significance level it is expected four or five significance spikes. Last but not least, MAD test scored 0.0010 far below 0.0022 limit value.

Thus, according to the methodology adopted in the scope of this work, aggregated data from January 2018 conforms NBL and Table 6 summarizes their NBL conformity tests (H_{0A}), not rejecting the null hypothesis.

To reinforce the methodology and likewise validate the results three other geographically aggregated data sets are submitted to the tests. Like January 2018 data analysis, those three datasets also conform NBL and the results are detailed on Tables 7–9.

Table 6
NBL tests result summary for January 2018 Brazilian municipalities withdrawals.
Source: Created by authors, 2019.

January 2018 – NBL goodness-of-fit tests result summary			
Sample size: 5570	Mean: 432,198.482	Median: 165,355.5	Skewness: 27.711
First digit test			Chi-square 11.932
Significance level 5%: Chi-square (critical) = 15.507			
First-two digits tests	Chi-square 80.8485	Z-statistics[a] 17 (2.0468) 34 (1.9683) 35 (2.3590) 48 (2.6486)	MAD 0.001
Significance level 5%: Chi-square (critical) = 112.022, Z (critical) = 1.969 and MAD (critical) = 0.0022			

[a]The methodology of this works assumes five spikes as NBL limit.

Table 7
NBL tests result summary for March 2018 Brazilian municipalities withdrawals.
Source: Created by authors, 2019.

March 2018 – NBL goodness-of-fit tests result summary			
Sample size: 5570	Mean: 424,614.614	Median: 162,741	Skewness: 27.150
First digit test			Chi-square 6.402
Significance level 5%: Chi-square (critical) = 15.507			
First-two digits tests	Chi-square 73.7051	Z-statistics[a] 17 (2.0468) 34 (1.9683)	MAD 0.001
Significance level 5%: Chi-square (critical) = 112.022, Z (critical) = 1.969 and MAD (critical) = 0.0022			

[a]The methodology of this works assumes five spikes as NBL limit.

Table 8
NBL tests result summary for January 2019 Brazilian municipalities withdrawals.
Source: Created by authors, 2019.

January 2019 – NBL goodness-of-fit tests result summary			
Sample size: 5570	Mean: 455,091.997	Median: 171,758	Skewness: 26.43
First digit test			Chi-square 10.681
Significance level 5%: Chi-square (critical) = 15.507			
First-two digits tests	Chi-square 79.7722	Z-statistics[a] 47 (2.3328) 72 (2.0605)	MAD 0.001
Significance level 5%: Chi-square (critical) = 112.022, Z (critical) = 1.969 and MAD (critical) = 0.0022			

[a]The methodology of this works assumes five spikes as NBL limit.

Table 9
NBL tests result summary for March 2019 Brazilian municipalities withdrawals.
Source: Created by authors, 2019.

March 2019 – NBL goodness-of-fit tests result summary			
Sample size: 5570	Mean: 457,253.508	Median: 172,404.5	Skewness: 25,715
First digit test			Chi-square 12.672
Significance level 5%: Chi-square (critical) = 15.507			
First-two digits tests	Chi-square 78.8933	Z-statistics[a] 40 (2.1813) 60 (1.9864)	MAD 0.001
Significance level 5%: Chi-square (critical) = 112.022, Z (critical) = 1.969 and MAD (critical) = 0.0022			

[a]The methodology of this works assumes five spikes as NBL limit.

Table 10
Top 10 worst summation test result, number of municipalities per digits group and the total amount spent in each of the analysed samples.
Source: Created by the authors, 2019.

Jan 2018	Digits	Municipalities	Total amount	Summation test
	10	234	R$ 109,877,366	0.0456
	12	190	R$ 104,132,939	0.0433
	11	208	R$ 87,720,460	0.0364
	17	114	R$ 85,948,151	0.0357
	68	37	R$ 80,584,943	0.0335
	15	162	R$ 70,752,611	0.0294
	16	139	R$ 67,394,183	0.0280
	14	166	R$ 66,918,817	0.0278
	25	110	R$ 66,881,381	0.0278
	38	64	R$ 66,550,185	0.0276
Mar 2018	Digits	Municipalities	Total amount	Summation test
	11	213	R$ 127,922,364	0.0541
	10	239	R$ 100,210,843	0.0424
	66	35	R$ 83,096,964	0.0351
	16	147	R$ 80,890,286	0.0342
	15	140	R$ 73,323,370	0.0310
	13	177	R$ 71,198,374	0.0301
	24	105	R$ 68,382,890	0.0289
	12	174	R$ 65,410,569	0.0277
	14	174	R$ 65,578,539	0.0277
	36	71	R$ 59,148,489	0.0250
Jan 2019	Digits	Municipalities	Total amount	Summation test
	10	238	R$ 122,100,486	0.0482
	11	215	R$ 116,716,804	0.0460
	12	206	R$ 112,070,372	0.0442
	69	45	R$ 84,470,797	0.0333
	17	135	R$ 83,878,058	0.0331
	25	109	R$ 72,838,575	0.0287
	19	118	R$ 70,530,707	0.0278
	35	83	R$ 69,867,279	0.0276
	14	155	R$ 68,976,357	0.0272
	40	65	R$ 68,556,247	0.0270
Mar 2019	Digits	Municipalities	Total amount	Summation test
	12	192	R$ 122,539,107	0.0481
	10	222	R$ 114,091,863	0.0448
	11	217	R$ 108,891,623	0.0428
	18	148	R$ 89,895,376	0.0353
	68	32	R$ 79,156,572	0.0311
	25	103	R$ 79,004,806	0.0310
	14	160	R$ 70,581,452	0.0277
	40	77	R$ 70,151,431	0.0275
	16	143	R$ 67,003,252	0.0263
	19	111	R$ 66,750,354	0.0262

Once goodness-of-fit results demonstrate the four data set conforms to NBL. The test variable, i.e. withdrawal values, are submitted to summation test where according to Nigrini [5], the sums of all the numbers in a Benford distribution with first-two digits 10, 11, 12... 99 should be equal, as well as their proportions, i.e. 1/90 or 0.011. The methodology of this work assumes 25% of upper tolerance and thus the proportion critical value of 0.01375.

When submitted to the summation test, we select the subset from the 90 first-two digits in which the sum proportion value overcome the critical value of 0.01375. Thus, in case of Bolsa Familia benefits payment, we verify the materiality and relevance of each group of digits, in order to select those that deserve a more detailed and critical look.

The method proposed in this paper, adapted from Nigrini [5], select the top 10 highest summation test results regarding digits groups selection as shown in Table 10.

Finally, given the digits highlighted on Table 10, we select municipalities in which test variable leading digits begin with.

In practical terms, in case of auditing or fraud detection, this work suggests starting by the chosen municipalities.

6. Conclusion

We conclude that Benfordś analysis, when used correctly, is a powerful tool for identifying fraud suspect accounts for further analysis and investigation in nationwide money distribution welfare programs such as the Brazilian Bolsa Familia program. To this end, we analysed payments during four different periods, January and March of 2018 and 2019 and suspicion of accounting irregularities were found.

When submitted to goodness-of-fit with NBL, beneficiary's withdrawal values in January 2018 and 2019 data, provided by Transparency Portal, do not conform NBL. However, we concluded that cash transfer per se is not a good test variable because it is not a valid Benford-set. On the other hand, when these data are geographically aggregated, they fit Benford data set requirements. Therefore, the payment total amount per municipality in four different periods became our test variable.

As detailed on Tables 6–9, all four datasets we use, perfectly succeed on NBL goodness-of-fit tests the methodology adopt — first digit chi-square, first-two digits chi-square, Z-test and MAD, having 5% of the significance level. For the reasons mentioned above the data sets accept the null hypothesis (H_{0A}) of conformity to Benford's Law.

However, when submitted to the summation test, several leading digits groups failed, suggesting, according to Nigrini [5], that these groups exceed expected value and hence null hypothesis (H_{0B}) is rejected. Consequently, considering the result of conformity tests combined with summation test, we conclude that municipalities, in which withdrawal values significant digits begin with, paid the benefits to those who are not allowed to receive it, a hint of fraud.

It is important to highlight that the methodology we proposed, which is based on NBL tests, may not be considered the final word in terms of fraud detection. Other techniques may be used, such as machine learning as proposed by Badal-Valero et al. [40]. Moreover, the group of municipalities under suspicion of fraud, pointed by this work, may be considered the starting point in case of Bolsa Familia program auditing and that they should be analysed in more depth.

Although we have selected four data sets from different periods of time, in case of continuous auditing process, we recommend submitting datasets from all months to the methodology.

This research has proposed an NBL-based tool for goodness-of-fit and summation tests. All tests presented on this work were performed on a web-based platform developed on Microsoft. net framework in C# language, however other tools, such as Microsoft Excel, R or Python programming languages provide the same results. That tool may be use not only for anti-fraud purposes but also to any kind of NBL conformity validation. R-Script files, as well as CSV files used on this research, are available at https://github.com/cazevedo1977/brazilian-welfareprogram-bolsafamilia-dataanalysis and https://www.kaggle.com/caioazevedo/brazilian-welfareprogram-bolsafamilia-dataanalysis, respectively.

R-scripts were written having "benford.analysis" package as a supporter. This library was provided by Cinelli [41].

We hope that the presented method opens the door to a wide range of social welfare programs auditing, in which assistance is provided to individuals and families through programs such as health care, food stamps, unemployment compensation, pandemic crisis support, such as caused by COVID-19, housing assistance, and childcare assistance. We also expect that this paper can support auditors applying NBL based analysis to increase their ability to detect fraud on such important welfare programs.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors gratefully acknowledge the financial support from grants #2019/14011-0 and #2017/50343-2, São Paulo Research Foundation (FAPESP), Brazil. We would like to thank Glauber Rocha Romão and Carlos Prado for suggestions and support on designed software used for the Bolsa Família Program analysis.

References

[1] A. Jamain, Benford's Law, Imperial College, 2001, http://wwwf.imperial.ac.uk/~nadams/classificationgroup/Benfords-Law.pdf.

[2] F. Benford, The law of anomalous numbers Author(s): Frank Benford the law of anomalous numbers has been observed that the pages of a much used table of common logarithms show evidences of a selective use of the natural numbers. The pages containing the logarithms, Proc. Am. Phil. Soc. 78 (1938) 551–572.

[3] M. Nigrini, Using digital frequencies to detect fraud, 1994, pp. 72–91.

[4] M. Nigrini, A taxpayer compliance application of Benford's law, J. Am. Tax. Assoc. 18 (1996) 72.

[5] M. Nigrini, Benford's Law: Applications for Forensic Accounting, Auditing and Fraud Detection, first ed., John Wiley and Sons, Inc., 2012.

[6] I. Jucá, M.A. Melo, L. Rennó, The political cost of corruption: Scandals, campaign finance, and reelection in the Brazilian chamber of deputies, J. Politics Lat. Am. 8 (2016) 3–36, http://dx.doi.org/10.1177/1866802x1600800201.

[7] A. Cerioli, L. Barabesi, A. Cerasa, M. Menegatti, D. Perrotta, Newcomb–Benford law and the detection of frauds in international trade, Proc. Natl. Acad. Sci. USA 116 (2019) 106–115, http://dx.doi.org/10.1073/pnas.1806617115.

[8] F.C.R. da Cunha, Aplicações da Lei Newcamb-Benford à Auditoria de Obras Públicas, Universidade de Brasilia, 2013.

[9] S. Newcomb, Note on the frequency of use of the different digits in natural numbers Author(s): Simon Newcomb source : American journal of mathematics, Amer. J. Math. 4 (1) (1881) 39–40, Published by : The Johns Hopkins University Press Stable URL : http://www., http://www.jstor.org/stable/2369148.

[10] T.P. Hill, The significant digit phenomenon, 102 (1995) 322–327.

[11] B.S. Law, P. Hill, S. Newcomb, Base-invariance implies Benford's law, 123 (1995) 887–895.

[12] T.P. Hill, A statistical derivation of the significant-digit law, Stat. Sci. 10 (1995) 354–363, http://dx.doi.org/10.2307/2246134.

[13] R.S. Pinkham, On the distribution of first significant digits, Ann. Math. Stat. 32 (1961) 1223–1230, http://dx.doi.org/10.1214/aoms/1177704862.

[14] Halph A. Raimi, The peculiar distribution of first digit', Sci. Am. 221 (1969) 109–121.

[15] J.I. de F. Costa, J. dos Santos, S.K. de M. Travassos, Análise de conformidade nos gastos públicos dos entes federativos: aplicação da lei de Newcomb-Benford para o primeiro e segundo dígitos dos gastos em dois estados brasileiros, Rev. Contab. Finanç. 23 (2012) 187–198, http://dx.doi.org/10.1590/s1519-70772012000300004.
[16] L. Pietronero, E. Tosatti, V. Tosatti, A. Vespignani, Explaining the uneven distribution of numbers in nature: The laws of Benford and Zipf, Phys. A 293 (2001) 297–304, http://dx.doi.org/10.1016/S0378-4371(00)00633-6.
[17] G. Whyman, N. Ohtori, E. Shulzinger, E. Bormashenko, Revisiting the Benford law: When the Benford-like distribution of leading digits in sets of numerical data is expectable? Phys. A 461 (2016) 595–601, http://dx.doi.org/10.1016/j.physa.2016.06.054.
[18] D. Biau, The first-digit frequencies in data of turbulent flows, Phys. A 440 (2015) 147–154, http://dx.doi.org/10.1016/j.physa.2015.08.016.
[19] M. Nigrini, Digital Analysis Using Benford's Law, Global Audit Publications, 2000.
[20] T.P. Hill, The first digit phenomenon: A century-old observation about an unexpected pattern in many numerical tables applies to the stock market, census statistics and accounting data, Am. Sci. 86 (1998) 358–363.
[21] R.A. Raimi, The first digit phenomenon again, Am. Phil. Soc. 2 (1985) 211–219.
[22] C. Durtschi, W. Hillison, C. Pacini, The effective use of Benford's law to assist in detecting fraud in accounting data, 99 (2004) 17–34.
[23] S.J. Nigrini, Miller Mark, Data diagnostics using second-order tests of Benford's law, Audit.: J. Pract. Theory 28 (2009) 305–324, http://dx.doi.org/10.2308/aud.2009.28.2.305.
[24] J.M. Pimbley, Benford's law and the risk of financial fraud, Risk Prof. (2014) 1–7, http://www.maxwell-consulting.com/Benfords_Law_GARP_May_2014.pdf.
[25] D. Amiram, Z. Bozanic, E. Rouen, Financial statement irregularities: Evidence from the distributional properties of financial statement numbers, Rev. Acc. Stud. 20 (2015) 1540–1593, http://dx.doi.org/10.2139/ssrn.2374093.
[26] M. Ausloos, R. Castellano, R. Cerqueti, Regularities and discrepancies of credit default swaps: a data science approach through Benford's law, Chaos Solitons Fractals 90 (2016) 8–17, http://dx.doi.org/10.1016/j.chaos.2016.03.002.
[27] D. Fu, Y.Q. Shi, W. Su, A generalized Benford's law for JPEG coefficients and its applications in image forensics, in: Proc. SPIE - Int. Soc. Opt. Eng., 2007, http://dx.doi.org/10.1117/12.704723.
[28] M. Ausloos, C. Herteliu, B. Ileanu, Breakdown of Benford's law for birth data, Phys. A 419 (2015) 736–745, http://dx.doi.org/10.1016/j.physa.2014.10.041.
[29] G. Pollach, K. Jung, F. Namboya, C. Pietruck, Maternal mortality rate—A reliable indicator? Int. J. Clin. Med. 06 (2015) 342–346, http://dx.doi.org/10.4236/ijcm.2015.65044.
[30] M. Sambridge, H. Tkalčić, A. Jackson, Benford's law in the natural sciences, Geophys. Res. Lett. 37 (2010) http://dx.doi.org/10.1029/2010GL044830.
[31] T.A. Mir, The law of the leading digits and the world religions, Phys. A 391 (2012) 792–798, http://dx.doi.org/10.1016/j.physa.2011.09.001.
[32] D. Gamermann, F.L. Antunes, Statistical analysis of Brazilian electoral campaigns via Benford's law, Phys. A 496 (2018) 171–188, http://dx.doi.org/10.1016/j.physa.2017.12.120.
[33] M. Miranda-Zanetti, F. Delbianco, F. Tohmé, Tampering with inflation data: A Benford law-based analysis of national statistics in Argentina, Phys. A 525 (2019) 761–770, http://dx.doi.org/10.1016/j.physa.2019.04.042.
[34] F.A. Alali, S. Romero, Benford's law: Analyzing a decade of financial data, J. Emerg. Technol. Account. 10 (2013) 1–39, http://dx.doi.org/10.2308/jeta-50749.
[35] J.I. de F. Costa, S.K. de M. Travassos, J. dos Santos, Application of Newcomb-Benford law in accounting audit: a bibliometric analysis in the period from 1988 to 2011, in: 10th International Conference on Information Systems and Technology Management, June 12–14, Sao Paulo, Brazil, 2013, http://dx.doi.org/10.5748/9788599693094-10contecsi/ps-80.
[36] N.H.F. Beebe, A Bibliography of Publications About Virtual Machines Title Word Cross-Reference, 2007.
[37] M. Nigrini, Forensic Analytics Methods and Techniques for Forensic Accounting Investigations, 2011, http://dx.doi.org/10.1017/CBO9781107415324.004.
[38] W. Wallace, Assessing the quality of data used for benchmarking and decision-making, J. Gov. Financ. Manag. 31 (2002) 16–22.
[39] T.F. de C. Marshall, J.L. Fleiss, Statistical Methods for Rates and Proportions, 2003, http://dx.doi.org/10.2307/2988144.
[40] E. Badal-Valero, J.A. Alvarez-Jareño, J.M. Pavía, Combining Benford's law and machine learning to detect money laundering. An actual spanish court case, Forensic Sci. Int. 282 (2018) 24–34, http://dx.doi.org/10.1016/j.forsciint.2017.11.008.
[41] C. Cinelli, benford.analysis: Benford Analysis for data validation and forensic analytics, (n.d.). https://cran.r-project.org/package=benford.analysis.