https://doi.org/10.24867/SYMOPIS-2024-51-087

Faculty of Technical Sciences
University of Novi Sad
51st Symposium on Operational Research
SYM-OP-IS 2024
Tara, 16-19. 9. 2024.

FRAUD DETECTION IN FINANCIAL REPORTS OF THREE PRIVATE HOSPITALS OPERATING IN SERBIA, TESTING WITH THE NEW TESTS

JELENA STANOJEVIĆ[1], DRAGANA RADOJIČIĆ[1]

Stanojević, ORCID: 0000-0001-5668-5297
Radojičić, ORCID: 0000-0001-7850-2623

[1] University of Belgrade – Faculty of Economics and Business, Serbia, jelena.stanojevic@ekof.bg.ac.rs

Abstract: *In this paper, we use Benford's law as the main methodology for fraud detection in the financial reports of three private hospitals operating in Serbia. This law is a statistical method that can be used to quickly and efficiently detect suspicious positions and figures in large datasets. The law states that the leading digits from 1 to 9 appear in a decreasing logarithmic distribution. It states that the digit 1 appears most frequently with 30% of the time and the digit 9 has the lowest frequency with about 5% of the time according to this law. We also investigate the conformity of an empirical distribution of observed real data with the Benford's distribution. The main goal of the paper is to present an application of the permutation tests as well as the bootstrap tests for checking conformity with the law, and the application of resampling methods is proposed.*

Keywords: *Fraud detection, Benford's law, statistical tests, resampling methods*

1. INTRODUCTION

Benford's law is a phenomenon that is frequently used in various fields, especially to detect fraud, manipulation or errors in financial data. Intuitively, one might assume that each digit from 1 to 9 has the same probability of being a leading digit of the number in the data set under consideration, i.e. that the first digits of the numbers are uniformly distributed. Originally, the American scientist Simon Newcomb (see [13]) used statistical tables in his research and found that the first pages were more worn than the others. This observation led him to assume that the digit 1 occurs more frequently than other digits and that lower digits tend to occur more frequently than higher digits, ignoring the zero as the first digit. At the time, this discovery did not attract the attention of people from academia and industry. Later, Frank Benford (see [2]) formally defined and proved this law of the first digit, which is now known as Benford's law. Benford's law, often referred to as the "law of the first digit", determines the frequency with which each digit appears as the leading digit of a number in observed data, as well as the occurrence of other digits within the data. Benford's law can be applied to numerous datasets from various fields, e.g. river lengths, birth rates, tax returns, numbers appearing in newspapers, sports statistics, etc. Given the presence of random errors, manipulations, and fraud in financial statements, including data entry mistakes and rounding discrepancies, there is a need for a methodology that can efficiently identify and locate suspicious entries. Mark Nigrini noted that the distribution of digits in a database containing fraud or errors is not random and does not conform to Benford's law, see [14]. Therefore, Benford's law has important applications in the fields of accounting, auditing and taxation (see [12], [4], [15], [18], etc.). Benford's law is a mathematical technique used to detect irregularities, errors or potential fraud in numerical datasets, including the analysis of financial reports. More recently, Benford's law has been effectively applied to detect fraud and manipulation in business and administrative data, such as balance sheets and tax declarations. Several papers (e.g. [17], [11]) discuss the application of Benford's law and its potential benefits for international and governmental macroeconomic statistics. The law was also applied to determine whether countries have manipulated COVID-19 data during the pandemic by verifying the reported number of COVID-19 cases ([6], [20]). In [3], the authors investigated whether the regression coefficients and other statistics derived from sociological empirical research are in compliance with Benford's law. The paper [8] explores the application of Benford's law in the context of prices in eBay auctions. In the paper [1], the authors analyze the reliability of inspections with the help of Benford's law using a data set consisting of the balance sheets and income statements of Italian companies balance sheets and income statements. In his works [10] and [9], Hill provided a mathematical proof of Benford's law. He also investigated the question of which common distributions conform to Benford's law and established that a combination of two distributions can lead to a Benford's distribution, even if the individual distributions do not adhere to the law.

1.1. Fundamental characteristics and properties of Benford's law

This law can be useful to detect intentional or unintentional errors that may occur during data collection and recording or as a consequence of data manipulation. Therefore, it stands as one of the most important and effective methods for verifying data accuracy. However, it should be noted that non-compliance with Benford's law does not automatically imply that the data contains manipulations, errors or fraud. Even compliance with the law is no guarantee that no manipulation has taken place. In practice, Benford's law is used as a preliminary step, followed by a more in-depth analysis of the data or figures that do not conform to the law. It is important to remark that positive and negative values have the same treatment in testing with Benford's law, as are values with decimals. Note that the concept that the law remains applicable even if the scale of measurement is changed, see [19].

In this section, we summarize the main definitions to get the reader convenient with the basic concept that stands behind Benford's law.

The next two definitions, originally presented in the paper [7], describe the notations for leading significant digits and the significance of a number.

► **Definition 1.** For any positive number $x > 0$ and the base B, x is represented as $x = S_B(x) \cdot B^{k(x)}$, where $S_B(x) \in [1, B)$ is the significand of x and the integer $k(x)$ (necessarily unique) is the exponent. For negative number x, $S_B(x) = S_B(-x)$ and for convenience, $S_B(0) = 0$.

► **Definition 2.** A (real-valued) random variable X follows Benford's law in the base B if and only if for all $t \in [1, B)$,

$$P\{S_B(X) \leq t\} = \log_B(t),$$

in particular,

$$P\{FSD = d\} = \log_B\left(\frac{d+1}{d}\right) = \log_B\left(1 + \frac{1}{d}\right), \quad d \in \{1, 2, ..., B-1\},$$

where FSD stands for the first significant digit of X, i.e. the first (leftmost) digit of $S_B(X)$.
Please note that hereafter we restrict our attention to the general case $B = 10$.

The calculated probabilities of occurrence for the first five digits are presented in Table 1.

Table 1: The probability of an occurrence of a digit in different positions in a number (percentage)

Digit	Percentage				
	1st position	2nd position	3rd position	4th position	5th position
0	-	11.968	10.178	10.018	10.000
1	30.103	11.389	10.138	10.014	10.000
2	17.609	10.882	10.097	10.010	10.000
3	12.494	10.433	10.057	10.006	10.000
4	9.691	10.031	10.018	10.002	10.000
5	7.918	9.668	9.979	9.998	10.000
6	6.695	9.337	9.940	9.994	10.000
7	5.799	9.035	9.902	9.990	10.000
8	5.115	8.757	9.864	9.986	10.000
9	4.576	8.500	9.827	9.982	10.000
Σ	100.000	100.000	100.000	100.000	100.000

The probability of an occurrence of the second digit is obtained by the following formula (see [16]):

$$P\{D_2(X) = d_2\} = \sum_{d_1=1}^{9} \log\left(1 + \frac{1}{10d_1 + d_2}\right), d_2 \in \{0, 1, ..., 9\}, \tag{1}$$

where $D_2(X)$ is a random variable modeling the second order significant digit of the number X and d_2 is a digit on the second position, and for the k-th position:

$$P\{D_k(X)=d_k\}=\sum_{d_1=1}^{9}\sum_{d_2=0}^{9}...\sum_{d_{k-1}=0}^{9}\log\left(1+\frac{1}{\sum_{i=1}^{k}10^{k-i}d_i}\right), \tag{2}$$

where $D_k(X)$ is random variable modeling the k-th order significant digit of the number X, and $d_k \in \{0,1,...,9\}$.

Further, Benford's law can be extended, and in [10] the joint distribution of the first and higher-order significant digits is determined, so the following formula holds:

$$P\{D_1(X)=d_1,D_2(X)=d_2,...,D_k(X)=d_k\}=\log\left(1+\frac{1}{\sum_{i=1}^{k}10^{k-i}d_i}\right), \tag{3}$$

for all $k \in N$, all $d_1 \in \{1,2,...,9\}$ and all $d_i \in \{0,1,...,9\}$, $i \geq 2$.

Table 2 shows the probability of an occurrence of the combination of the first two digits in the number.

Table 2: The probability of an occurrence of the first two digits in a number (percentage)

2nd	1st position									Σ
position	1	2	3	4	5	6	7	8	9	
0	4.139	2.119	1.424	1.072	0.860	0.718	0.616	0.540	0.480	11.968
1	3.779	2.020	1.379	1.047	0.843	0.706	0.607	0.533	0.475	11.389
2	3.476	1.931	1.336	1.022	0.827	0.695	0.599	0.526	0.470	10.882
3	3.218	1.848	1.296	0.998	0.812	0.684	0.591	0.520	0.464	10.433
4	2.996	1.773	1.259	0.976	0.797	0.673	0.583	0.514	0.460	10.031
5	2.803	1.703	1.223	0.955	0.783	0.663	0.575	0.508	0.455	9.668
6	2.633	1.639	1.190	0.934	0.769	0.653	0.568	0.502	0.450	9.337
7	2.482	1.579	1.158	0.914	0.755	0.643	0.560	0.496	0.445	9.035
8	2.348	1.524	1.128	0.895	0.742	0.634	0.553	0.491	0.441	8.757
9	2.228	1.472	1.100	0.877	0.730	0.625	0.546	0.485	0.436	8.500
Σ	30.103	17.609	12.494	9.691	7.918	6.695	5.799	5.115	4.576	100.000

2. RESAMPLING METHODS FOR FRAUD DETECTION

In this paper, we will consider the application of permutation and bootstrap tests as the main resampling methods known in the literature (see [5]), but applied for the first time for the purpose of fraud detection.

2.1. Permutation tests

Permutation tests were introduced by Fisher in the 1930s as an important theoretical argument and support for the Student's t-test. With the development of powerful software, permutation tests could be used as routine statistical tests in the future. Their advantage over the standard statistical methodology is the simple basic idea behind the method and the fact that the tests are free of mathematical assumptions.

The two-sample problem for testing the null hypothesis H_0 of the difference between the two distributions F and G is applied to the two independent random samples $z=(z_1,z_2,...,z_n)$ and $y=(y_1,y_2,...,y_m)$. The equality between F and G means that booth distributions assign equal probabilities to all subsets of the common sample space of z and y. The null hypothesis states that there is no difference between the probabilities of the random variables z and y, $H_0: F=G$, and like any hypothesis test begins with the test statistic $\widehat{\theta}$. In this paper, we will consider the mean difference as the test statistic $\widehat{\theta}=\bar{z}-\bar{y}$. We assume that if the null hypothesis is not true, we expect the value of the test statistic to be larger than the null hypothesis is true. More precisely, we cannot quantify what "large" means, we can only say that the larger the observed value of the test statistic $\widehat{\theta}$, the stronger the evidence against the null hypothesis. In other situations, we can also consider a smaller value instead of a larger value to represent stronger evidence.

We may define the achieved significance level (p-value) of the test with the following formula:

$$ASL = P_{H_0}\{\widehat{\theta}^* \geq \widehat{\theta}\}, \tag{4}$$

where the quantity $\widehat{\theta}$ is fixed at its observed value and $\widehat{\theta}^*$ is a random variable and has the distribution of $\widehat{\theta}$ if H_0 is true. The main difference between $\widehat{\theta}$ and $\widehat{\theta}^*$ is that the first is an observed mean difference and the latter is a hypothetically generated value.

The main problem in practice is to calculate *ASL*. If the null hypothesis H_0 defines a single distribution from which we can calculate the corresponding probability given by equation (4), we can calculate *ASL*. For most practical problems, the null hypothesis provides a family of possible null hypothesis distributions. To calculate *ASL*, for example, we usually have to approximate the null hypothesis variance or use Student's method. This method is a good approach, but only applicable in the normal case.

For the general null hypothesis $H_0 : F = G$, Fisher's permutation test is a solution to the problem of calculating *ASL*. If the null hypothesis is correct, we combine all $n+m$ observations from both groups together, take a sample of size n without replacement to represent the first group, and the last m observations for the second group. We then calculate the difference between the group means and repeat this process a large number of times. This two-sided permutation test rejects the null hypothesis at a 5% level if the original difference in sample means falls outside the middle 95% of the distribution of differences.

Let $N = n+m$, let $v = (v_1, v_2, ..., v_N)$ be the combined and ordered vector of all values, and let $g = (g_1, g_2, ..., g_N)$ be the vector indicating to which group each ordered observation belongs. Both vectors v and g together give us the same informations as the vectors z and y together.

The vector g consists of n z's and m y's and there are $\binom{N}{n}$ possible g vectors. Another result is important for the permutation test.

Permutation Lemma: Under $H_0 : F = G$ the vector g has the probability $1/\binom{N}{n}$ that all possible values are equal.

We can imagine a test statistic $\widehat{\theta} = S(g, v)$ as a function of the vectors g and v. For example: $\widehat{\theta} = \bar{z} - \bar{y} = \frac{1}{n}\sum_{g_i=z} v_i - \frac{1}{m}\sum_{g_i=y} v_i$.

Let g^* denote one of the $\binom{N}{n}$ possible vectors of n z's and m y's, and define the permutation replication of $\widehat{\theta}$ in the next line:

$$\widehat{\theta}^* = \widehat{\theta}(g^*) = S(g^*, v).$$

The appropriate permutation p-value is defined with the probability:

$$ASL = P\{\widehat{\theta}^* \geq \widehat{\theta}\} = \sharp\{\widehat{\theta}^* \geq \widehat{\theta}\} / \binom{N}{n},$$

where $\binom{N}{n}$ is the number of permutation replications $\widehat{\theta}^*$. In practise, Monte Carlo simulation is used for this calculation, based on the following algorithm.

Algorithm for calculating the statistical two-sample permutation test, in three steps:

1. Choose B independent vectors $g^*(1), g^*(2), ..., g^*(B)$, each consisting of n z's and m y's and each being randomly selected from the set of all $\binom{N}{n}$ possible such vectors.
2. Evaluate the permutation replications of $\widehat{\theta}$ corresponding to each permutation vector,

$$\widehat{\theta}^*(b) = S(g^*(b), v), b = 1, 2, ..., B.$$

3. Approximate p-value with

$$\widehat{ASL} = \sharp\{\widehat{\theta}^*(b) \geq \widehat{\theta}\}/B.$$

2.2. Bootstrap tests

Compared to permutation tests, bootstrap tests provide similar results when both are available, and they are more widely applicable, even though they are less accurate. In this two-sample problem, we have samples z and y from possibly different probability distributions F and G, and we want to test the null hypothesis $H_0 : F = G$.

A test statistic does not have to be an estimate of a parameter, we denote it here by $t(x)$, and for the example $t(x) = \bar{z} - \bar{y}$. We seek for a p-value:

$$ASL = P_{H_0}\{t(x^*) \geq t(x)\}.$$

The value $t(x)$ is an observed value and it is fixed, while x^* is a random value and has a distribution under the null hypothesis H_0, F_0. The goal is to estimate F_0. In bootstrap tests, a "plug-in" estimate for F_0 is used, whereas in permutation tests the order statistic v is fixed and the distribution F_0 is defined as the distribution of possible orders of ranks g. Let x be a combined sample and its empirical distribution denoted by $\widehat{F_0}$, where the probability $1/(n+m)$ is placed on each member of x. $\widehat{F_0}$ gives a nonparametric estimate of the common population. In the next lines, we give an algorithm for calculation of p-value.

Algorithm for computation of the bootstrap test statistic, in three steps:

1. Draw B samples of size $n+m$ with replacement from x. Call the first n observations z^* and the remaining m observations y^*.
2. Evaluate $t(\cdot)$ on each sample,
$$t(x^{*b}) = \bar{z}^* - \bar{y}^*, b = 1, 2, ..., B.$$
3. Approximate p-value with
$$\widehat{ASL} = \sharp\{t(x^{*b}) \geq t_{obs}\}/B,$$
where $t_{obs} = t(x)$ the observed value of the statistic.

The only difference between this algorithm and the permutation tests algorithm is that samples are drawn with replacement rather than without replacement.

3. RESULTS

The observed datasets were downloaded from the website of the Serbian Business Registers Agency (SBRA, https://apr.gov.rs). The next table shows the p-values for both tests, permutation and bootstrap, in the case of the three considered hospitals operating in Serbia: Bel Medic hospital, Euromedic hospital and Medigroup hospital. Both tests provide very similar results.

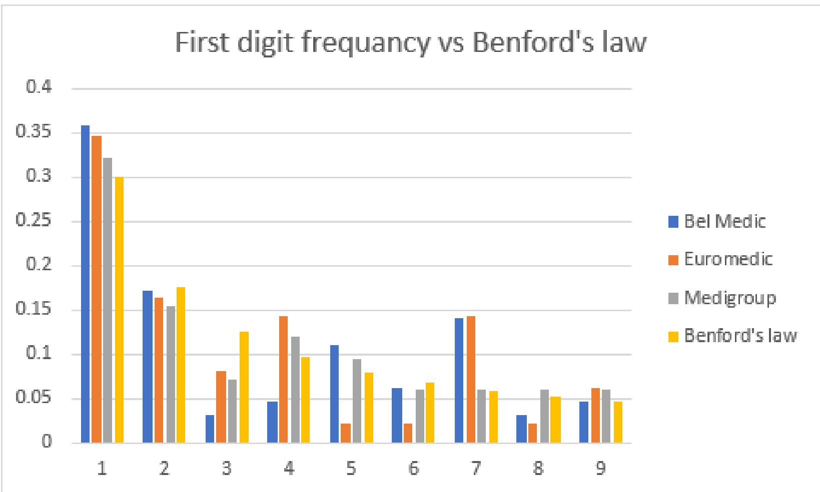
Table 3: p-values for two tests in a case of three hospitals

	Bel Medic	Euromedic	Medigroup
Permutation test	1	1	1
Bootstrap test	0.499	0.483	0.511

We may conclude that the first digit for all considered hospitals are confirmed with the Benford's law.

In the next picture we present first digit frequency of the considered hospitals vs Benford's law frequency.

Picture 1: First digit frequency for all hospitals



4. CONCLUSION

To test the agreement of the data set with Benford's law, the following four tests have been used in the literature so far: Z-test, Chi-square test, Kolmogorov Smirnov test and Mean Absolute Deviation test. All of them are well known and well described in the literature. In this paper we use two known tests from the literature, but for the first time in this purpose, to find out conformity of the considered data set with the Benford's law. For both test we calculate p-values, and for all considered datasets we get that p-values are larger than 0.05, what mean that on the 5% confidence level we accept conformity of the first digit with the Benford's law.

The advantage of these two resampling tests, applied in this paper, is their simple application with a suitable algorithm in three steps. A possible limitation for these tests is the limitation of datasets available during the research. Also, we can recommend these two methods as a first step to detect fraud and possible data manipulation in collected datasets, and that is their additional application.

For future research, we suggest the development of new tests and their comparison with the old tests in terms of their strength.

REFERENCES

[1] Arezzo, M. F., & Cerqueti, R. (2023). A Benford's Law view of inspections' reasonability. Physica a: Statistical Mechanics and Its Applications, 632, 129294.

[2] Benford, F. (1938). The law of anomalous numbers. Proceedings of the American philosophical society, 551-572.

[3] Diekmann, A. (2007). Not the first digit! Using benford's law to detect fraudulent scientif ic data. Journal of Applied Statistics, 34(3), 321-329.

[4] Durtschi, C., Hillison, W., & Pacini, C. (2004). The effective use of Benford's law to assist in detecting fraud in accounting data. Journal of forensic accounting, 5(1), 17-34.

[5] Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. Chapman and Hall/CRC.

[6] Eutsler, J., Harris, M. K., Williams, L. T., & Cornejo, O. E. (2023). Accounting for partisanship and politicization: Employing Benford's Law to examine misreporting of COVID-19 infection cases and deaths in the United States. Accounting, Organizations and Society, 108, 101455.

[7] Fang, G., & Chen, Q. (2020). Several common probability distributions obey Benford's law. Physica A: Statistical Mechanics and its Applications, 540, 123129.

[8] Giles, D. E. (2007). Benford's law and naturally occurring prices in certain ebaY auctions. Applied Economics Letters, 14(3), 157-161.

[9] Hill, T. P. (1995). Base-invariance implies Benford's law. Proceedings of the American Mathematical Society, 123(3), 887-895.

[10] Hill, T. P. (1995). A statistical derivation of the significant-digit law. Statistical science, 354-363.

[11] Hindls, R., & Hronová, S. (2015). Benford's Law and Possibilities for Its Use in Governmental Statistics. Statistika: Statistics & Economy Journal, 95(2).

[12] Law, N. M. B. S. (2012). Applications for forensic accounting, auditing, and fraud detection, vol. 586.

[13] Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. American Journal of mathematics, 4(1), 39-40.

[14] Nigrini, M. J. (1996). A taxpayer compliance application of Benford's law. The Journal of the American Taxation Association, 18(1), 72.

[15] Nigrini, M. J., & Mittermaier, L. J. (1997). The use of Benford's Law as an Aid in Analytical Procedures. Auditing: A journal of practice & theory, 16(2).

[16] Nigrini, M. J. (2012). Benford's Law: Applications for forensic accounting, auditing, and fraud detection (Vol. 586). John Wiley & Sons.

[17] Nye, J., & Moul, C. (2007). The political economy of numbers: on the application of Benford's law to international macroeconomic statistics. The BE Journal of Macroeconomics, 7(1).

[18] Papić, M., Vudrić, N., & Jerin, K. (2017). Benfordov zakon i njegova primjena u forenzičkom računovodstvu. Zbornik sveučilišta Libertas, 1(1-2), 153-172.

[19] Pinkham, R. S. (1961). On the distribution of first significant digits. The Annals of Mathematical Statistics, 32(4), 1223-1230.

[20] Zhang, J. (2020). Testing case number of coronavirus disease 2019 in China with Newcomb-Benford law. arXiv preprint arXiv:2002.05695.