

Statistics helping to fight fraud

Author: Edson R. Montoro

Increasingly, good citizens demand fairness in electoral processes and the use of public money, our money, by the way. Statistics has helped a lot in this and more and more it is used, thus contributing to the appreciation of ethics and morality in all media.

Simon Newcomb (1835 - 1909), an American astronomer and mathematician, in his natural curiosity as a scientist, discovered in the Harvard library that the students' logarithm board was dirtier and worn out on the front pages, while the last pages gave signs little use, showing that smaller numbers were more common than larger ones. This showed that probably the number probability distribution model was not uniform as previously believed. But he didn't gather any more data to evidence his discovery, perhaps for lack of time. This fact was forgotten for many years and it was not until the 1930s that Frank Benford (1883–1948), a long-time electrical engineer at GE, came back with this idea and worked more at GE, generating a law known as the First Digit Act or Benford law, but the most correct would be Newcomb-Benford law.

Frank Benford has shown that this result applies to a wide variety of data sets, including electricity bills, addresses, stock prices, house prices, population numbers, mortality rates, river lengths, physical and mathematical constants, among others.

His work analyzed the first digits of various collections of data collected and all showed that about 30% of the numbers had 1 as the first digit, and in contrast less than 5% had 9 as the first digit.

Obviously, this law does not apply to any and all datasets, as it is an empirical law; for example, if we take a phone book with prefix (98), we won't find any digits first 1. But it's a very

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro

edsonmontoro@ermontoro.com

+55 (35) 99161-8141

Flavio A. Montoro

flaviomontoro@ermontoro.com

+55 (16) 98102-6651

www.ermontoro.com

interesting law when it comes to random and large amounts of data. Benford’s formula for calculating the probability of each first digit is the logarithm of $(1 + 1 / d)$ in base 10, or

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

where P (d) is the probability of the digit d occurring.

Doing this for digits 1 through 9, we have the following table (Table 1):

Table 1 - Probability of first digits according to Newcomb-Benford law.

First Digit	Probability (%)
1	0.301
2	0.176
3	0.125
4	0.097
5	0.079
6	0.067
7	0.058
8	0.051
9	0.046

Represented by the curve shown on Figure 1.

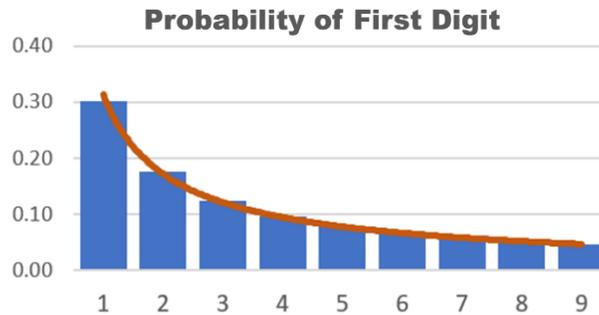


Figure 1 – Newcomb-Benford’s Distribution of probability graph.

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro
edsonmontoro@ermontoro.com
+55 (35) 99161-8141

Flavio A. Montoro
flaviomontoro@ermontoro.com
+55 (16) 98102-6651

www.ermontoro.com

To get an idea of the force of this law, after the 1980s it began to be used to verify the accuracy of random data. That is, it has begun to be used to check crime at gaming tables, in electronic gambling, as it is cheap and only a small collection of data is enough to attest to the true randomness of data or not. So, the Internal Revenue Service (IRS) in 1998 implemented this law in its fraud tracking system! Yes, because if, for example, expense amounts are appearing with too many leading 3 digits above 12% this is a sign that something is wrong with the tax return. And most striking is that historical fraud in the US comprising data from large companies was uncovered and correct, applying only the Newcomb-Benford law.

Several studies were carried out assuming that fabricated data are identified by the deviation of digits in relation to the Benford distribution. Nigrini (1992; 1997; 1999; 2011; 2012), assuming that reliable accounting data closely matched Benford's distribution, demonstrated that substantial deviations from this Law would suggest possible fraud or manipulated data. The author developed several tests to measure compliance with Benford Law, and frauds were detected in seven New York companies by the Brooklyn Attorney's office using these tests. As evidence, it was found in this case that fraudulent and random data had few values starting at 1 and many numbers starting at 6. Based on these previous successes, Nigrini was called on to advise tax collection agencies in various countries and install Benford Law tests on most computer fraud detection programs.

Rauch, et al. (2011) published an article in the German Economic Review, in which they suggested that Benford's Law could be used to test macroeconomic data, revealing which ones needed closer inspection. They analyzed the compliance with the Benford Law of the first digits of macroeconomic data reported to the European Union Statistics Office (Eurostat) by European Union (EU) member countries. They built a ranking of the 27 member countries according to the extent of the deviation found. The country that had the largest deviation was Greece, whose data

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro

edsonmontoro@ermontoro.com

+55 (35) 99161-8141

Flavio A. Montoro

flaviomontoro@ermontoro.com

+55 (16) 98102-6651

www.ermontoro.com

manipulation had been officially confirmed by the European Commission (2010) at an earlier time.

Walter Mebane, a University of Michigan statistician, has studied election data from several countries, including the United States, Russia, and Mexico. In 2006, he found that vote counting tended to follow Benford's Law in the second digit. The researcher analyzed Iran's 2009 election data and found abnormalities that strongly indicated the occurrence of fraud in the victory of politician Ahmadinejad. Mebane found that in cities with few invalid votes, Ahmadinejad's numbers were far from Benford's distribution, and that the candidate in these situations had a large vote advantage, raising suspicion of electoral fraud.

In Brazil, we are still starting to apply this law, but we already have some great initiatives. At TCU (Federal Court of Accounts), a master's dissertation was presented by server Flávia Ceccato, guided by Professor Maurício Bugarin, addressing the application of this methodology to public works audits. Relevant works were tested in the context of the 2014 World Cup, such as the renovation of Maracanã Stadium, the construction of the Amazon Arena and the renovation of the Minas Gerais International Airport. These three works were previously audited by the Court and analyzes based on the Newcomb-Benford Act were compared with the over-price detected by TCU. The services indicated by this Law as having suffered possible manipulation in their prices corresponded, on average, to 80% of the over-price identified by the Court.

There are several publications in Brazil showing the application of this tool to identify frauds and contributing to the corrupt to be sent to the deserved place. The duo Ceccato & Bugarin developed articles for: the TCU Magazine (Benford Law and Public Works Audit: an analysis of price over Maracanã reform); the Economics Bulletin Magazine (Benford's law for audit of public works: an analysis of overpricing in Maracanã soccer arena's renovation); the NDJ

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro

edsonmontoro@ermontoro.com

+55 (35) 99161-8141

Flavio A. Montoro

flaviomontoro@ermontoro.com

+55 (16) 98102-6651

www.ermontoro.com

Magazine (Benford Law for Public Works Auditing: price analysis in the construction of the Amazon arena); among others.

The Newcomb-Benford Act is already in the interest of the CADE (Administrative Council for Economic Defense) as a cartel filter and also of the Federal Police in criminal engineering expertise. Excuse me, but I get excited when I see the application of scientific techniques in practice, especially in our country, where corruption has been the tonic for some time. There are hopes!

But back to the Newcomb-Benford law, we enforce some rules for its application, namely:

1. Numbers must be generated naturally. The law does not apply, for example, to analyze CPFs (Brazilian citizen id), but is perfectly suitable for application in accounts payable, invoices, rate cards, etc.
2. Values cannot have a maximum limit value set.
3. Must be at least 4 digits.
4. Must have at least 1.000 records.

An example (1) with partial data (600 values) from the Fibonacci series will be presented using an Excel spreadsheet (if anyone wants the spreadsheet, please contact us).

The table (Table 2) will be presented partially by space saving.

Table 2 - Data from Example 1.

Random	First Digit	one	two	three	four	five	six	seven	eight	nine
1	1	1	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	0	0
2	2	0	1	0	0	0	0	0	0	0
3	3	0	0	1	0	0	0	0	0	0
5	5	0	0	0	0	1	0	0	0	0
8	8	0	0	0	0	0	0	0	1	0

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro

edsonmontoro@ermontoro.com

+55 (35) 99161-8141

Flavio A. Montoro

flaviomontoro@ermontoro.com

+55 (16) 98102-6651

www.ermontoro.com

13	1	1	0	0	0	0	0	0	0	0
21	2	0	1	0	0	0	0	0	0	0
34	3	0	0	1	0	0	0	0	0	0
55	5	0	0	0	0	1	0	0	0	0
89	8	0	0	0	0	0	0	0	1	0
144	1	1	0	0	0	0	0	0	0	0

The calculations follow a simple Chi-square adherence test to see if the experimental data fit the theoretical distribution defined by Newcomb-Benford (NB) law, presented in Table 2.

With a significance level of 5%, the assumptions of the adherence test are as follows:

H₀: Data follow NB law;

H₁: Data do not follow NB law;

The rejection criterion of H₀ is if the observed value is greater than the critical.

The final result is presented in Table 3.

Table 3 - Result of the Chi-square test of example 1.

DIGITS	Absolute Frequency # OCCURRED	Relative Frequency (%) STATISTICS (%Of)	Absolute Frequency BENFORD LAW (#)	Relative Frequency (%) BENFORD LAW (Ef)	((Of-Ef) ²)/Ef
1	180	30.00	180.6	30.10	0,000
2	106	17.67	105.7	17.61	0,000
3	75	12.50	75.0	12.49	0,000
4	57	9.50	58.1	9.69	0,004
5	48	8.00	47.5	7.92	0,001
6	40	6.67	40.2	6.69	0,000
7	34	5.67	34.8	5.80	0,003
8	33	5.50	30.7	5.12	0,029
9	27	4.50	27.5	4.58	0,001
				χ^2 Obs =	0,038
				χ^2 Crit(0.05;8) =	15.51
				P- value =	1.000
Of =	Observed Frequency				
Ef =	Expected Frequency (Theory)				

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro
edsonmontoro@ermontoro.com
+55 (35) 99161-8141

Flavio A. Montoro
flaviomontoro@ermontoro.com
+55 (16) 98102-6651

www.ermontoro.com

Conclusion α 5%= Data appear to be coherent

It is possible to visualize the curve of the series data with the Newcomb-Benford law bars in Figure 2.

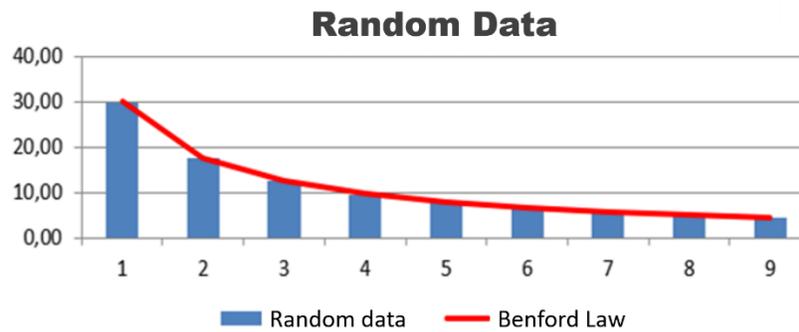


Figure 2 – Data from the Fibonacci series follows Newcomb-Benford law.

As it can be seen from the graph in Figure 2, the adjustment of Fibonacci series to the law is very good, as confirmed by the Chi-square test, in which the observed value (0.038) was lower than the critical one (15.51).

We can now see a second example (2) with a series of 900 Excel-generated random number data using the function “= RANDOM (0; 9000)” (Table 4). In real life, when data is expected to follow the law and it does not, we have a high chance that the data being manipulated, certainly in a real case, would be chosen for an audit. Again, the 5% significance test has the following hypotheses:

- H₀: Data follow NB law;
- H₁: Data do not follow NB law;

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro
edsonmontoro@ermontoro.com
+55 (35) 99161-8141

Flavio A. Montoro
flaviomontoro@ermontoro.com
+55 (16) 98102-6651

www.ermontoro.com

Table 4 - "Random" data from example 2.

Random	First digit	one	two	three	four	five	six	seven	eight	nine	ten
8283	8	0	0	0	0	0	0	0	1	0	0
646	6	0	0	0	0	0	1	0	0	0	0
2260	2	0	1	0	0	0	0	0	0	0	0
2219	2	0	1	0	0	0	0	0	0	0	0
2878	2	0	1	0	0	0	0	0	0	0	0
2156	2	0	1	0	0	0	0	0	0	0	0
5094	5	0	0	0	0	1	0	0	0	0	0
1785	1	1	0	0	0	0	0	0	0	0	0
2218	2	0	1	0	0	0	0	0	0	0	0
1620	1	1	0	0	0	0	0	0	0	0	0
3440	3	0	0	1	0	0	0	0	0	0	0

The calculations are presented in Table 5 and the graph in Figure 3.

Table 5 - Example 2 Chi-Square Test.

DIGITS	# OCCURRED	STATISTICS (oF)	BENFORD LAW (#)	BENFORD LAW (Fe)	$((oF-Ef)^2)/Ef$	
1	125	12.55	299.8	30.10	10,235	
2	116	11.65	175.4	17.61	2,019	
3	124	12.45	124.4	12.49	0,000	
4	123	12.35	96.5	9.69	0,729	
5	120	12.05	78.9	7.92	2,154	
6	142	14.26	66.7	6.69	8,542	
7	114	11.45	57.8	5.80	5,498	
8	119	11.95	50.9	5.12	9,126	
9	13	1.31	45.6	4.58	2,338	
				χ^2 Obs =	40,642	
Of =	Observed Frequency				χ^2 Crit(0.05;8) =	15.51
Ef =	Expected Frequency (Theory)				P- value =	0.000
Conclusion α 5% =				There is some data distortion		

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro
edsonmontoro@ermontoro.com
+55 (35) 99161-8141

Flavio A. Montoro
flaviomontoro@ermontoro.com
+55 (16) 98102-6651

www.ermontoro.com

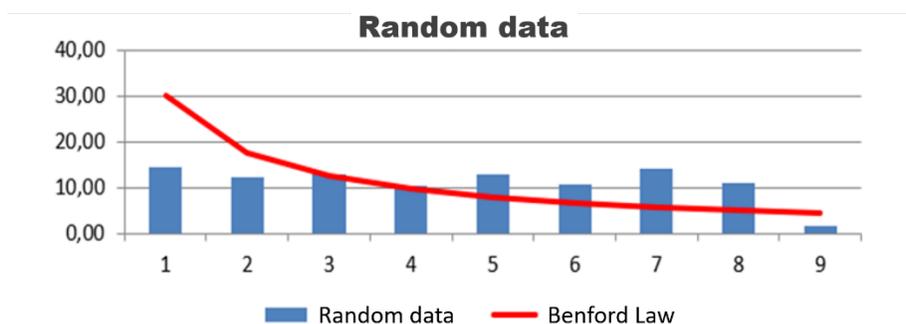


Figure 3 – Newcom-Benford Law applied to a “random” data set.

As it can be observed, the data do not fit the Law and Newcomb-Benford, because the Chi-square tests had the null hypothesis rejected, being the observed value (40.64) higher than the critical (15.51) and also from the graph in Figure 3 it is possible to observe that the data do not fit the expected curve of the law.

As seen, the random numbers generated by Excel, like other software, are pseudorandom.

This technique, although simple, can be very useful in several areas, as mentioned above, in accounting, election results, economics and can even be used in process control reports and laboratory data, and may indicate the existence of forgery.

References

- (1) The Law of Anomalous Numbers; Frank Benford; Proceedings of the American Philosophical Society; Vol. 78, No. 4 (Mar. 31, 1938), pp. 551-572.

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro
edsonmontoro@ermontoro.com
 +55 (35) 99161-8141

Flavio A. Montoro
flaviomontoro@ermontoro.com
 +55 (16) 98102-6651

www.ermontoro.com

- (2) Nigrini, M. (1992). The Detection of Income Tax Evasion Through an Analysis of Digital Frequencies. Cincinnati, OH: Ph.D. thesis. University of Cincinnati.
- (3) Nigrini, M., & Mittermaier, L. (1997). The Use of Benford's Law as an Aid in Analytical Procedures. Auditing: A Journal of Practice & Theory, 16, 52-67.
- (4) Nigrini, M. (1999). Adding Value with Digital Analysis. The Internal Auditor, 56, 21-23.
- (5) Nigrini, M. J. 2011. Forensic Analytics – Methods and Techniques for Forensic Accounting Investigations. Hoboken, NJ: John Wiley & Sons.
- (6) Nigrini, M. J. 2012. Benford's Law – Applications for Forensic Accounting, Auditing, and Fraud Detection. Hoboken, NJ: John Wiley & Sons.
- (7) Fact and Fiction in EU-Governmental Economic Data; Bernhard Rauch, Max Götsche, Gernot Brähler and Stefan Engel; German Economic Review, 2011, vol. 12, issue 3, 243-255.
- (8) Comment on “Benford's Law and the Detection of Election Fraud”; Walter R. Mebane Jr.; Political Analysis (2011) 19:269–272.
- (9) Aplicações da lei Newcomb-Benford à auditoria de obras Públicas; CUNHA, Flávia Ceccato Rodrigues da; <http://repositorio.unb.br/handle/10482/16379>.
- (10) Lei de Benford aplicada à auditoria da reforma do Aeroporto Internacional de Minas Gerais; Maurício Soares Bugarin Universidade de Brasília (UnB), Flávia Ceccato Rodrigues da Cunha, Tribunal de Contas da União (TCU) Rev. Serv. Público Brasília 68 (4) 915-940 out/dez 2017.

About the author:

Edson R. Montoro is Technical Director of ERMontoro Consulting and Training Ltda, a company focused on people development and process improvement consulting using Applied Statistics and Lean Manufacturing.

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro

edsonmontoro@ermontoro.com

+55 (35) 99161-8141

Flavio A. Montoro

flaviomontoro@ermontoro.com

+55 (16) 98102-6651

www.ermontoro.com

The author is a Chemist from UNESP (Paulista State University “Júlio de Mesquita Filho) - Araraquara, MBA in Business Management from FGV (Getulio Vargas Foundation), Master Black Belt from Air Academy Associates, Quality Engineer from ASQ (America Society for Quality) and Postgraduate in Production Management from UFSC (Federal University of Santa Catarina).

ERMontoro Consultoria e Treinamento Ltda.

Edson R. Montoro

edsonmontoro@ermontoro.com

+55 (35) 99161-8141

Flavio A. Montoro

flaviomontoro@ermontoro.com

+55 (16) 98102-6651

www.ermontoro.com
