

LUẬT BENFORD VÀ ỨNG DỤNG TRONG VIỆC PHÁT HIỆN SAI SÓT, GIAN LẬN TRONG THỐNG KÊ, KẾ TOÁN

TS. Phạm Hạ Thủy



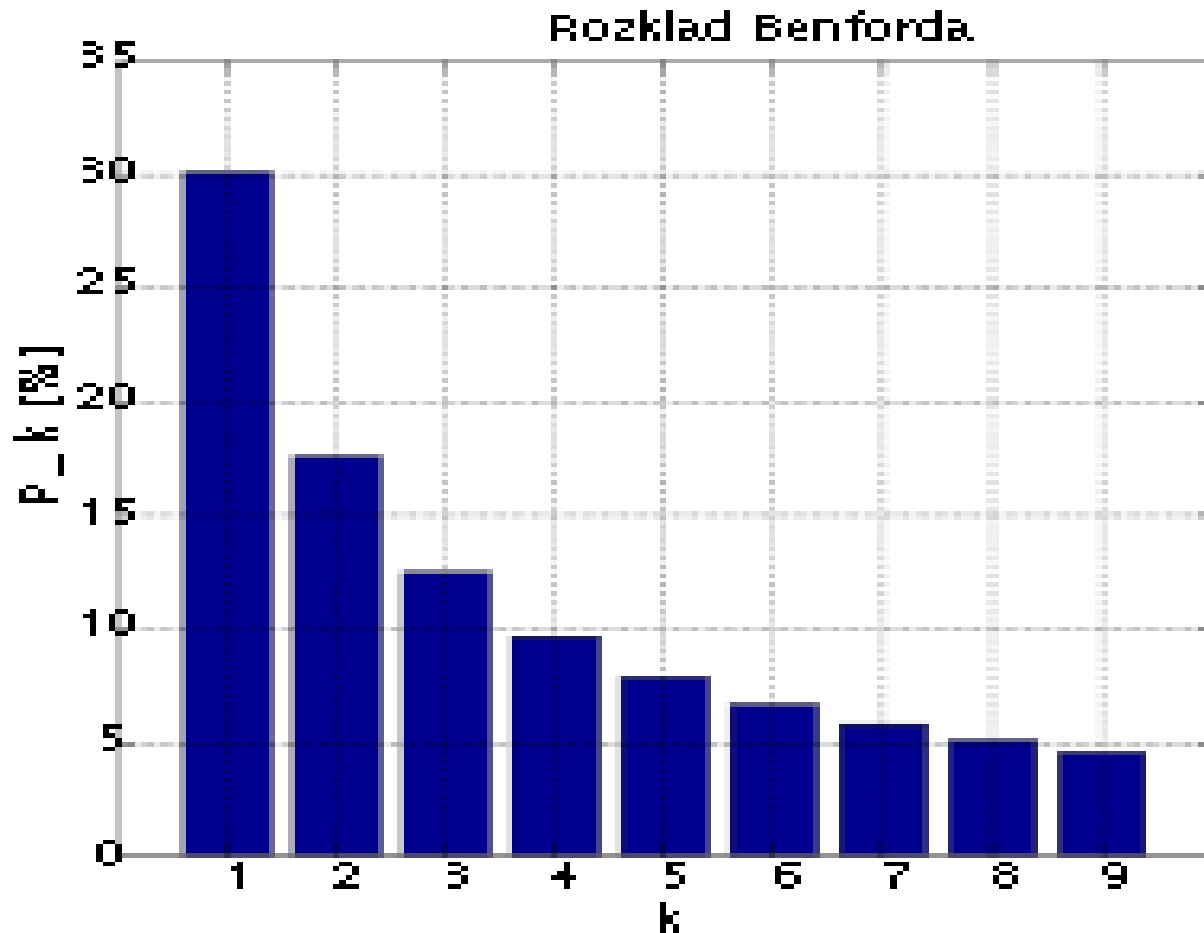
NỘI DUNG



- ✓ Giới thiệu
- ✓ Luật Benford
- ✓ Ứng dụng trong việc phát hiện sai sót, gian lận trong thống kê, kế toán
- ✓ Tài liệu tham khảo

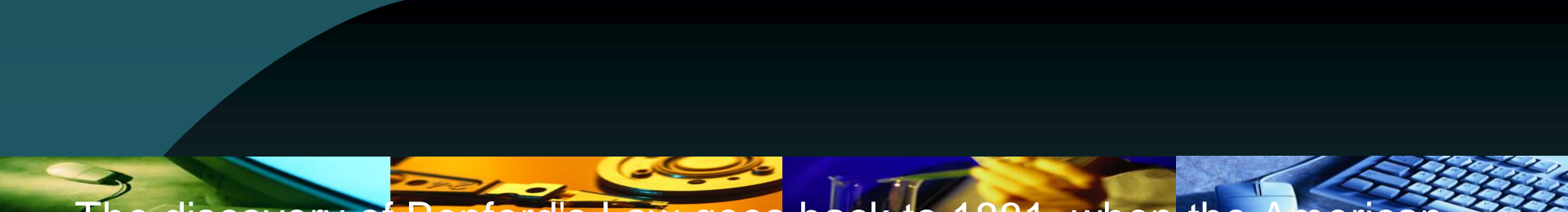
Giới thiệu

- **Benford's Law**, also called the **First-Digit Law**, refers to the frequency distribution of digits in many (but not all) real-life sources of data. In this distribution, 1 occurs as the leading digit about 30% of the time, while larger digits occur in that position less frequently: 9 as the first digit less than 5% of the time. It has been shown that this result applies to a wide variety of data sets, including electricity bills, street addresses, stock prices, population numbers, death rates, lengths of rivers, physical and mathematical constants, and processes described by power laws (which are very common in nature). It tends to be most accurate when values are distributed across multiple orders of magnitude.
- It is named after physicist Frank Benford, who stated it in 1938,^[1] although it had been previously stated by Simon Newcomb in 1881.^[2]



The distribution of first digits, according to Benford's law.

Each bar represents a digit, and the height of the bar is the percentage of numbers that start with that digit.

- 
- The discovery of Benford's Law goes back to 1881, when the American astronomer [Simon Newcomb](#) noticed that in [logarithm](#) tables (used at that time to perform calculations) the earlier pages (which contained numbers that started with 1) were much more worn than the other pages.^[2] Newcomb's published result is the first known instance of this observation and includes a distribution on the second digit, as well. Newcomb proposed a law that the probability of a single number N being the first digit of a number was equal to $\log(N + 1) - \log(N)$.
 - The phenomenon was again noted in 1938 by the physicist [Frank Benford](#),^[1] who tested it on data from 20 different domains and was credited for it. His data set included the surface areas of 335 rivers, the sizes of 3259 US populations, 104 [physical constants](#), 1800 [molecular weights](#), 5000 entries from a mathematical handbook, 308 numbers contained in an issue of [Reader's Digest](#), the street addresses of the first 342 persons listed in *American Men of Science* and 418 death rates. The total number of observations used in the paper was 20,229. This discovery was later named after Benford (making it an example of [Stigler's Law](#)).

Một số nghiên cứu ở nước ngoài

Nigrini, M. (1996). "A taxpayer compliance application of Benford's Law". *J Amer Tax Assoc* 18: 72–91.

Buyse, M, George, SL, Evans, S, Geller, NL, Edler, L and Hutton, J (1999). The Role of Biostatistics in the Prevention, Detection and Treatment of Fraud in Clinical Trials. *Statistics in Medicine* 18 (24), 3435-3451. ISSN:0277-6715

- Smith, CA (2002). Detecting Anomalies in Your Data Using Benford's Law. Paper 249 (Statistics and Data Analysis) in: Proceedings of SUGI 27, Orlando, USA, April 14-17.

Kumar, K and Bhattacharya, S (2003). Benford's law and its application in financial fraud detection. *Advances in Financial Planning and Forecasting* 11, 57-70

Durtschi, C; Hillison, W; Pacini, C (2004). "The effective use of Benford's Law to assist in detecting fraud in accounting data". *J Forensic Accounting* 5: 17–34.

Diekmann A (2007) Not the First Digit! Using Benford's Law to detect fraudulent scientific data. *J Appl Stat* 34 (3) 321–329, [doi:10.1080/02664760601004940](https://doi.org/10.1080/02664760601004940)

Một số nghiên cứu ở nước ngoài

Ni, D and Ren, Z (2008). Benford's law and half-lives of unstable nuclei. *Eur. Phys. J. A* 38, 251–255

Bhattacharya, S, Xu, D and Kumar, K (2010). An ANN-based auditor decision support system using Benford's Law. *Decision support systems*, 50 (3), 576-584.

Joseph Deckert, Mikhail Myagkov and Peter C. Ordeshook, (2010) [The Irrelevance of Benford's Law for Detecting Fraud in Elections](#), Caltech/MIT Voting Technology Project Working Paper No. 9

Friar, JL; Goldman, T; Pérez-Mercader, J (2012). "Genome sizes and the benford distribution". *PLOS ONE* 7 (5): e36624.

[doi:10.1371/journal.pone.0036624](https://doi.org/10.1371/journal.pone.0036624)

Alex Ely Kossovsky. [Benford's Law: Theory, the General Law of Relative Quantities, and Forensic Fraud Detection Applications](#), 2014, World Scientific Publishing. [ISBN 978-981-4583-68](#)

Đến nay có hơn 800 công trình nghiên cứu được xuất bản có liên quan hoặc sử dụng luật Benford (xem trong <http://www.benfordonline.net/list/>

• LUẬT BENFORD

Với nhiều tập hợp số được tạo bởi do thống kê, tính toán tuân theo qui luật sau:

$$P(D_1=d_1) = \log(1+1/d_1);$$

$$d_1 \in \{1,2,\dots,9\}$$

$$P(D_2=d_2) = \sum_{d_1=1}^9 \log(1+1/(d_1 d_2));$$

$$d_2 \in \{1,2,\dots,9\}$$

$$P(D_1 D_2=d_1 d_2) = \log(1+1/(d_1 d_2));$$

$$d_1 d_2 \in \{10,11,\dots,99\}$$

D1: first significant digit,

D2 :second digit of numbers.

Number	D1	D2
--------	----	----

613 •	6	1
-------	---	---

0.0002867 •	2	8
-------------	---	---

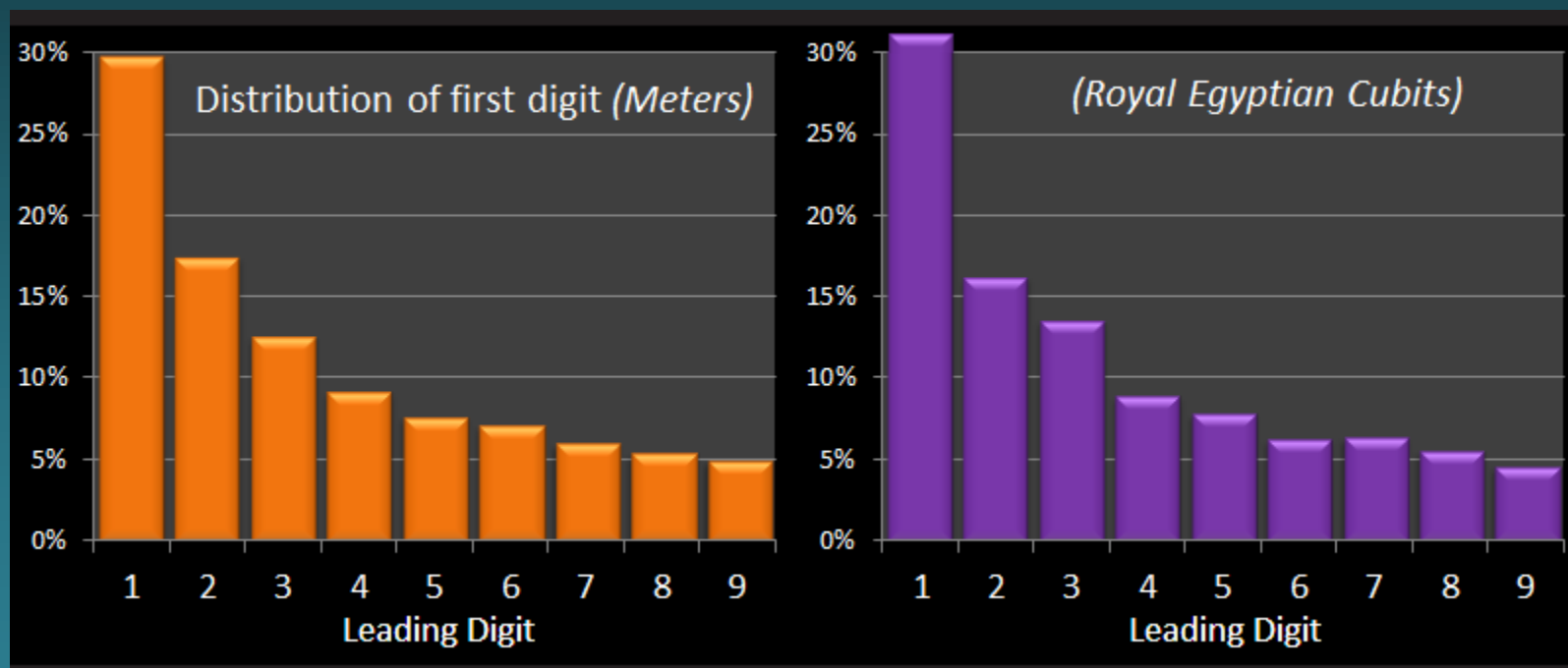
-71 •	7	1
-------	---	---

1,053,832 •	1	0
-------------	---	---

d	$\Pr(D_1 = d)$	$\Pr(D_2 = d)$
0	—	0.1197
1	0.3010	0.1139
2	0.1761	0.1088
3	0.1249	0.1043
4	0.0969	0.1003
5	0.0792	0.0967
6	0.0669	0.0934
7	0.0580	0.0904
8	0.0512	0.0876
9	0.0458	0.0850

Benford, F. (1938), The law of anomalous numbers, Proc. Amer.Philosophical Soc. 78, 551–572.

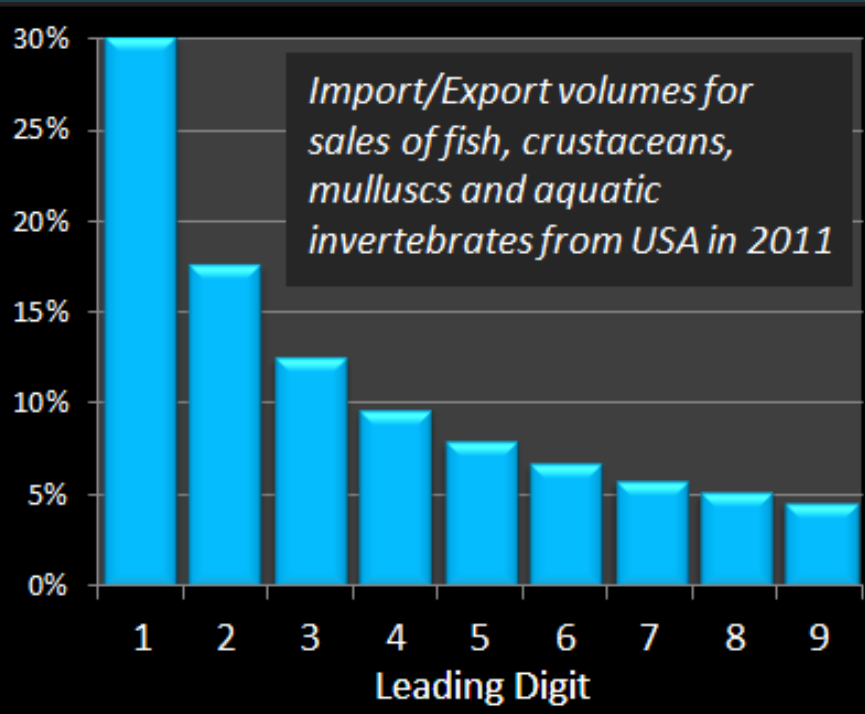
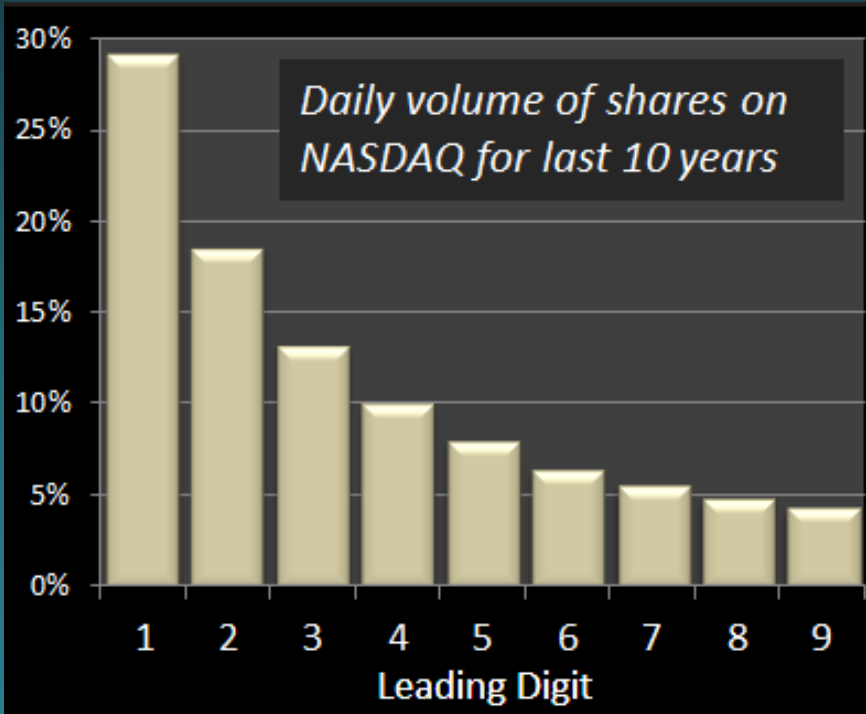
- Đã có nhiều thử nghiệm đối với Luật Benford, ví dụ: hình ảnh sau là biểu đồ về tần số các chữ số đầu tiên của độ cao của 122000 thành phố trên thế giới (với 2 đơn vị đo khác nhau, meters và Cubits (One Royal Egyptian Cubit = 523 mm))




- (<http://datagenetics.com/blog/march52012/index.html>)



Khối lượng giao dịch cổ phiếu trên sàn NASDAQ trong 10 năm(2002-2012) và khối lượng xuất/nhập khẩu các loại cá, loài giáp xác, động vật thân mềm của Mỹ trong năm 2011 – tuân theo luật Benford.



(nguồn từ : <http://datagenetics.com/blog/march52012/index.html>)

- 
- ❖ Có nhiều nghiên cứu mở rộng luật Benford dạng tổng quát hơn với cơ số 10 hoặc cơ số bất kỳ b :

In a form more complete than (1.1), BL is a statement about the joint distribution of *all* decimal digits: For every positive integer m ,

$$\text{Prob}((D_1, D_2, \dots, D_m) = (d_1, d_2, \dots, d_m)) = \log_{10} \left(1 + \left(\sum_{j=1}^m 10^{m-j} d_j \right)^{-1} \right) \quad (1.2)$$

holds for all m -tuples (d_1, d_2, \dots, d_m) , where d_1 is an integer in $\{1, 2, \dots, 9\}$ and for $j \geq 2$, d_j is an integer in $\{0, 1, \dots, 9\}$; here D_2, D_3, D_4 etc. represent the second, third, fourth etc. significant decimal digit, e.g.

- ❖ Các khái niệm về dãy số Benford (Benford sequences), tập hợp Benford (Benford set) được đưa ra và nghiên cứu.

(Arno Berger, Theodore P. Hill (2011). *A basic theory of Benford's Law. Probability Surveys, Vol. 8 (2011) 1–126;ISSN: 1549-5787;DOI: 10.1214/11-PS175*)

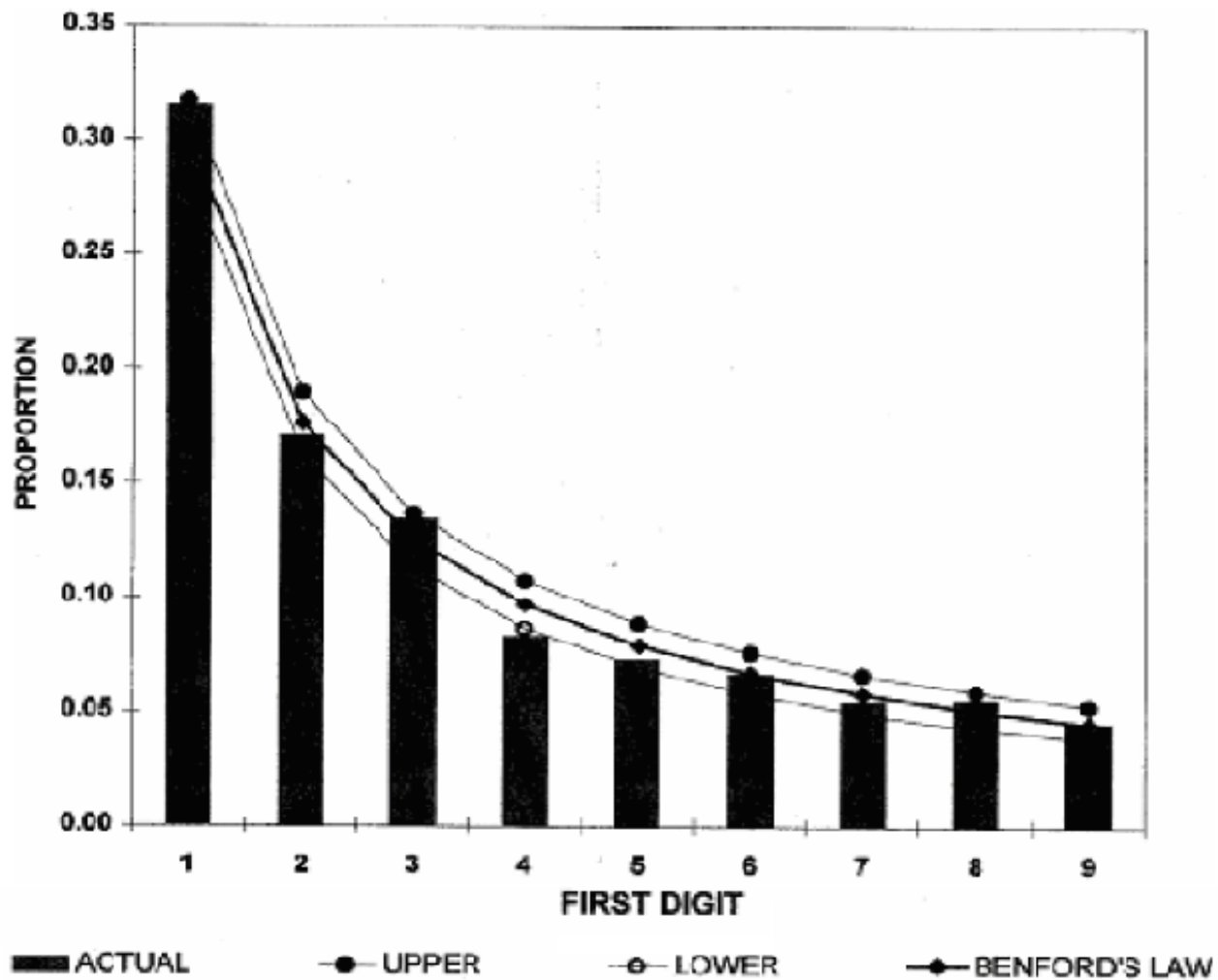
ỨNG DỤNG LUẬT BENFORD TRONG PHÁT HIỆN GIAN LẬN VÀ SAI SÓT TRONG THỐNG KÊ VÀ KẾ TOÁN

✓ Nguyên tắc chung:

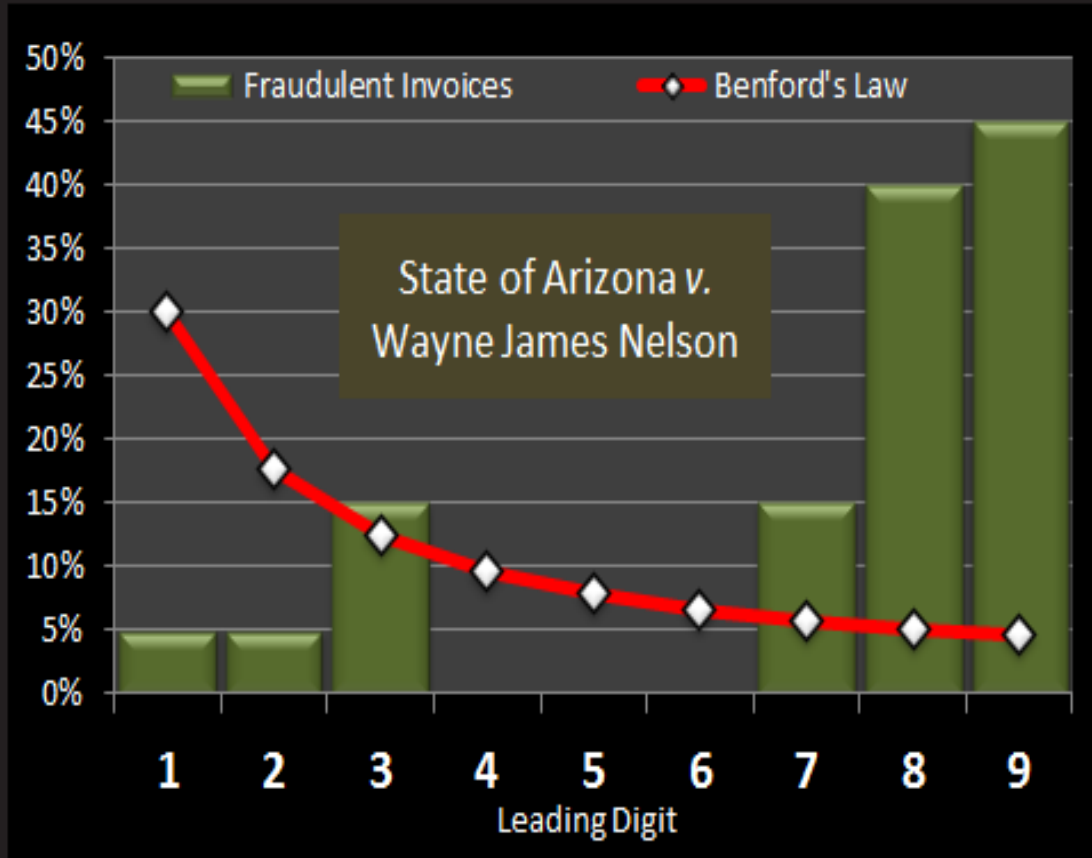
- ✓ *Một tập hợp số thực tuân theo luật Benford(*) nếu:*
- ✓ + *Cùng một hệ đo*
- ✓ + *Không bị hạn chế về giá trị lớn nhất và nhỏ nhất*
- ✓ + *Không thuộc vào một hệ thống qui định trước*
- ✓ *(ví dụ số hiệu tài khoản, số CMND, ...)*
- ✓ + *Có số lượng đủ lớn*

✓ Phương pháp

- ✓ - Tập hợp các khoản mục chi tiêu trong sổ sách kế toán tuân theo luật Benford
- ✓ - Tập các số liệu điều tra, thống kê, thực nghiệm về một lĩnh vực (thỏa mãn một số điều kiện (*)) tuân theo luật Benford
- ✓ Tổng hợp số các chữ số đầu và chữ số thứ hai của tập dữ liệu, tính tỉ lệ % so với tổng số - so sánh với tỉ lệ % lý thuyết của Luật Benford (lập biểu đồ để dễ nhìn)
- ✓ + Nếu chúng không phù hợp và vượt quá một ngưỡng nào đó thì cần nghi ngờ và đi sâu kiểm tra tại những giá trị vượt quá.



Ví dụ về việc phát hiện gian lận của Wayne James Nelson

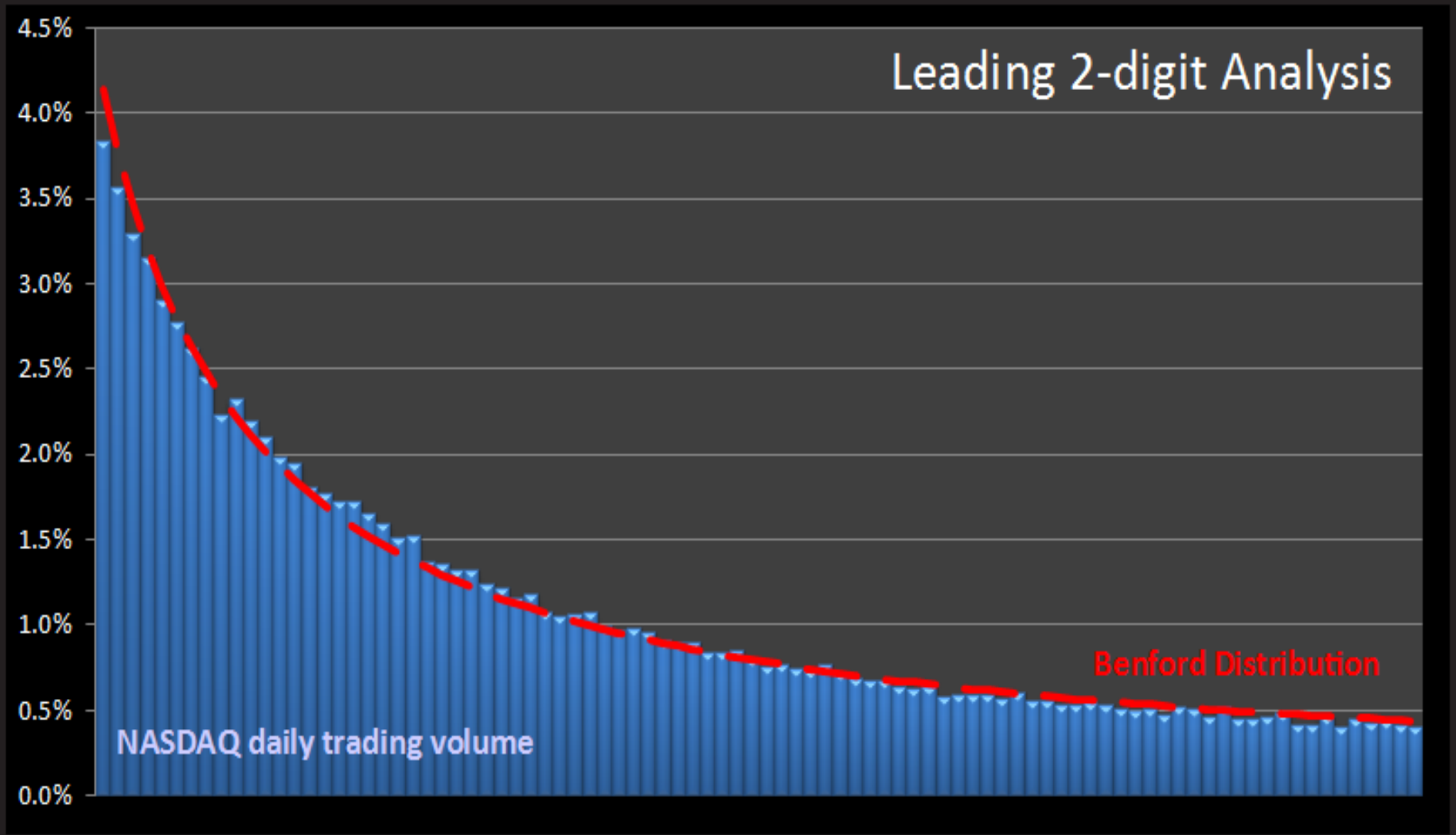


In 1993, in *State of Arizona v. Wayne James Nelson (CV92-18841)*, the accused was found guilty of trying to defraud the state of nearly \$2 million, by diverting funds to a bogus vendor.

The perpetrator selected payments with the intention of making them appear random: None of the check amounts was duplicated, there were no round-numbers, and all the values included dollars and cents amounts. However, as Benford's Law is esoterically counterintuitive, he did not realize that his seemingly random looking selections were far from random.

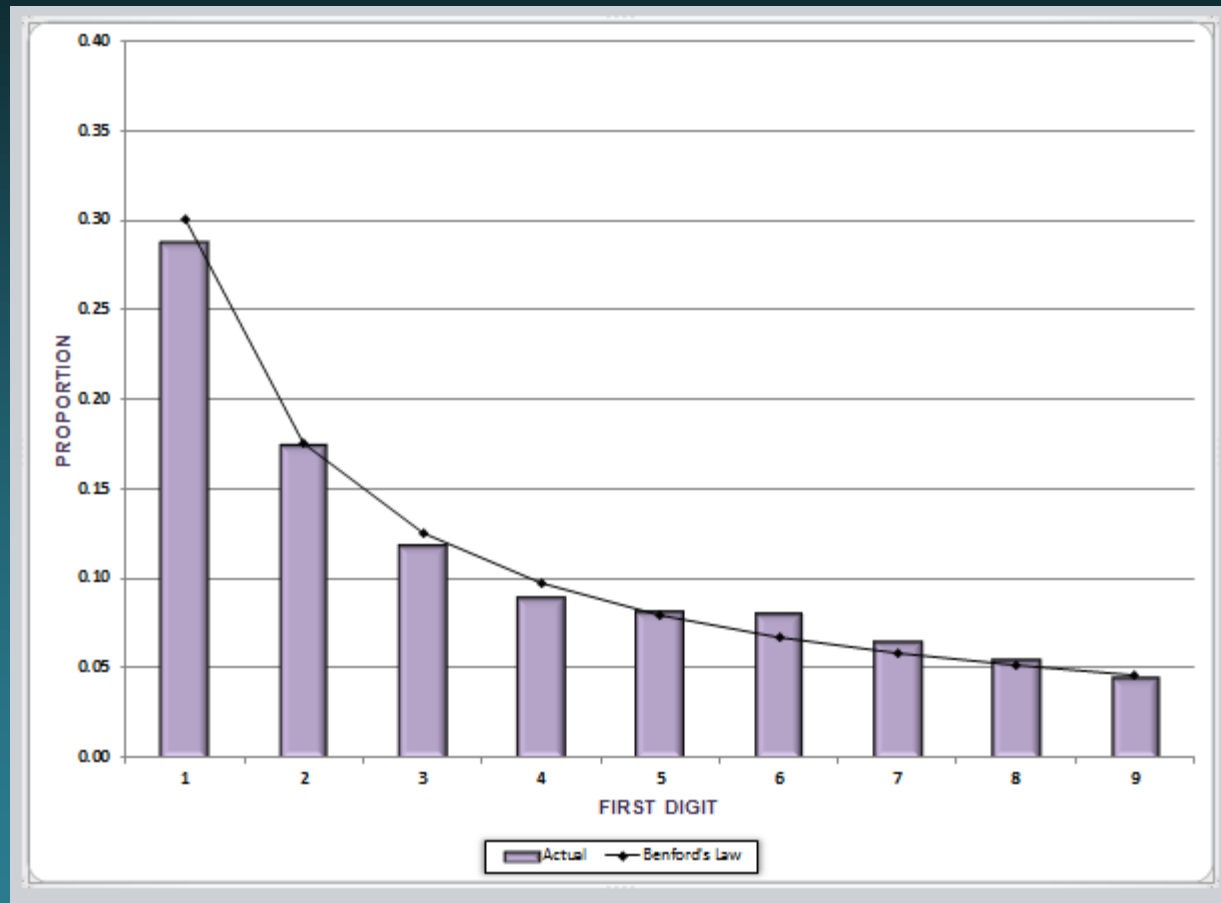
(<http://datagenetics.com/blog/march52012/index.html>)

Benford's Law does not just apply to the first digit. The system can be applied to look at the expected distributions for the first n -digits of numbers. Here, for instance is the same NASDAQ data plotted based on the leading 2-digits alongside the theoretical Benford distribution.



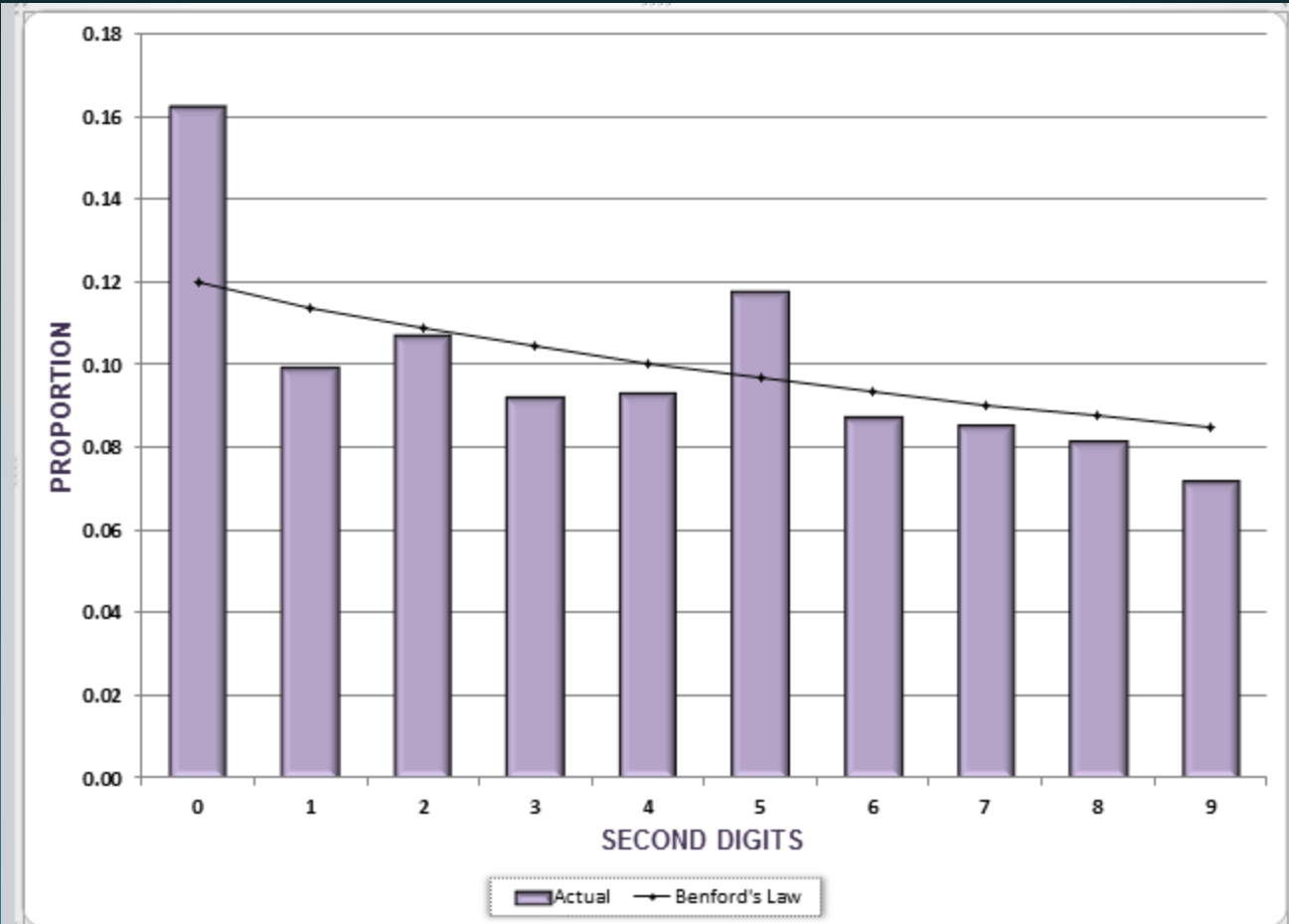
(<http://datagenetics.com/blog/march52012/index.html>)

Nghiên cứu về các khoản phải trả của công ty GM

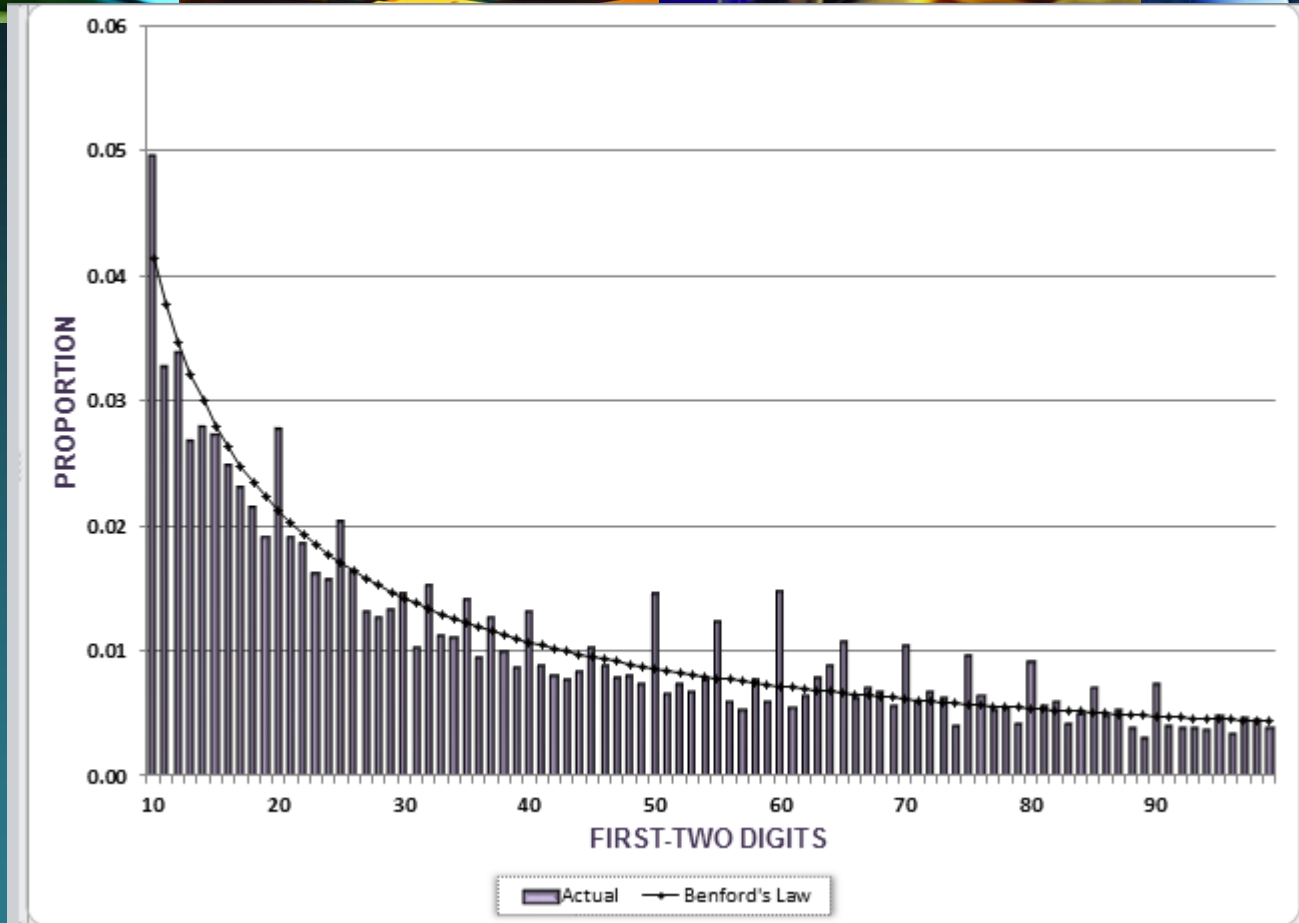


(<http://www.nigrini.com/benfordslaw.htm>)

Nghiên cứu về các khoản phải trả của công ty GM



Nghiên cứu về các khoản phải trả của công ty GM





- Hiện nay chức năng kiểm tra sự phù hợp với luật Benford đã được tích hợp trong nhiều phần mềm kiểm toán, thống kê:
 - IDEA
 - ACL
 - Benford's Law Excel 2007/2010
 -

Hướng dẫn sử dụng Excel kiểm tra Luật Benford

- Các bước tiến hành
- 1- Tải file dữ liệu vào
- 2- Ở cột số cần kiểm tra cần chuyển các số về dạng có chữ số đầu khác 0
- 3- Dùng hàm Left(X,1) để lấy ra chữ số đầu tiên của các số
- 4- Dùng Subtotal (count) để tính số các chữ số khác nhau.
- 5- Copy kết quả ra vị trí khác và tính toán tỉ lệ % của mẫu chọn và đồng thời tính giá trị% theo luật Benford
- 6- Dùng chức năng vẽ biểu đồ (Chart) để vẽ biểu đồ cho mẫu kiểm tra và biểu đồ luật Benford.
- Chi tiết, xem http://www.theiia.org/intAuditor/media/files/Step-by-step_Instructions_for_Using_Benford%27s_Law.pdf
- (Phần sau chúng tôi trình bày việc thực hành sử dụng Luật Benford trên Excel)

Tài liệu tham khảo

- Berger, A and Eshun, G (2014). Benford solutions of linear difference equations. Theory and Applications of Difference Equations and Discrete Dynamical Systems, Springer Proceedings in Mathematics & Statistics Volume 102, pp. 23-60. ISSN:978-3-662-44139-8. DOI:10.1007/978-3-662-44140-4_2.
- Benford, F. (1938), The law of anomalous numbers, Proc. Amer.Philosophical Soc. 78, 551–572.
- Bhattacharya, S, Xu, D and Kumar, K (2010). An ANN-based auditor decision support system using Benford's Law. Decision support systems, 50 (3), 576-584.
- Block, HW and Savits, TH (2010). A General Example for Benford Data. The American Statistician 64(4), 335-339.
- Brooks, D (2002). Naked-Eye Quantum Mechanics: Practical Applications of Benford's Law for Integer Quantities. Frequencies - The Journal of Size Law Applications, Special Paper #1, 1-8.
-




- Brown, P and Mitchell, J (2008). Culture and stock price clustering: evidence from the peoples' republic of China. *Pacific-Basin Finance Journal* 16(1), 95-120 . ISSN:0927-538X
- Busta, B and Weinberg, R (1998). Using Benford's law and neural networks as a review procedure. *Managerial Auditing Journal* 13(6), 356-366.
- Buyse, M, George, SL, Evans, S, Geller, NL, Edler, L and Hutton, J (1999). The Role of Biostatistics in the Prevention, Detection and Treatment of Fraud in Clinical Trials. *Statistics in Medicine* 18 (24), 3435-3451. ISSN:0277-6715.
- Cho, WKT and Gaines, BJ (2007). Breaking the (Benford) law: Statistical fraud detection in campaign finance. *American Statistician* 61(3), 218-223. ISSN:0003-1305. DOI:10.1198/000313007X223496.
- Cleary, R and Thibodeau, JC (2005). Applying Digital Analysis Using Benford's Law to Detect Fraud: The Dangers of Type I Errors. *Auditing - A Journal of Practice & Theory* 24(1), 77-81. ISSN:0278-0380.



- Conway, D (2010). Benford's Law Tests for Wikileaks Data . Zero Intelligence Agents website posting August 1, 2010.
- de Vocht, F and Kromhout, H (2013). The use of Benford's law for evaluation of quality of occupational hygiene data. *Ann Occup Hyg.* 57(3), pp. 296-304. DOI:10.1093/annhyg/mes067.
- Geyer, A and Martì, J (2012). Applying Benford's law to volcanology. *Geology* 40(4), pp. 327-330.
- Geyer, CL and Williamson, PP (2004). Detecting Fraud in Data Sets Using Benford's Law. *Communications in Statistics: Simulation and Computation* 33(1), 229-246. ISSN:0361-0918. DOI:10.1081/SAC-120028442.
- <http://www.benfordonline.net/list/>



- <http://www.nigrini.com/benfordslaw.htm>
- Geyer, CL and Williamson, PP (2004). Detecting Fraud in Data Sets Using Benford's Law. *Communications in Statistics: Simulation and Computation* 33(1), 229-246. ISSN:0361-0918. DOI:10.1081/SAC-120028442.
- Gorodnichenko, Y, Goriunov, D and Mylovanov, T (2014). Election fraud in 2014 Parliamentary Elections. VoxUkraine.org posting November 5, 2014; last accessed November 8, 2014.
- Kumar, K and Bhattacharya, S (2003). Benford's law and its application in financial fraud detection. *Advances in Financial Planning and Forecasting* 11, 57-70.
- Long, J (2014). Testing Benford's Law. Website: <http://testingbenfordslaw.com/>. Last accessed Sep 10, 2014.
- Mebane, WR Jr (2010). Fraud in the 2009 presidential election in Iran?. *Chance* 23(1), 6-15.

- 
- Moore, GB and Benjamin, CO (2004). Using Benford's Law for fraud detection. *Internal Auditing* 19(1), 4-9.
 - Ni, D and Ren, Z (2008). Benford's law and half-lives of unstable nuclei. *Eur. Phys. J. A* 38, 251–255.
 - Nigrini, MJ and Miller, SJ (2009). Data Diagnostics Using Second-Order Tests of Benford's Law. *Auditing: A Journal of Practice & Theory* 28(2), 305-324.
 - Pimbley, JM (2014). Benford's Law and the Risk of Financial Fraud. Global Association of Risk Professionals (garp.org)
 - http://www.garp.org/media/1301213/benfords_law_5.2.14.pdf. Last retrieved 10 September 2014.
 - Smith, CA (2002). Detecting Anomalies in Your Data Using Benford's Law. Paper 249 (Statistics and Data Analysis) in: *Proceedings of SUGI 27, Orlando, USA, April 14-17*.
 - Zorzi, A (2011). Benford's law and pi. *The Mathematical Gazette*, vol. 95, no. 533, July 2011, pp. 264-266.



END



