10-2020

Applying visual analytics to fraud detection using Benford's law

Clarence GOH
Singapore Management University, clarencegoh@smu.edu.sg

Applying visual analytics to fraud detection using Benford's law

Clarence Goh

Published in JOURNAL OF CORPORATE ACCOUNTING AND FINANCE, Advance online

DOI: 10.1002/jcaf.22440

Abstract

Benford's law has been examined as a useful tool for detecting potential accounting fraud. In this article, I provide an introduction to Benford's law and examine how the first digit, second digit, and first-two digits tests in Benford's law can be employed to detect potential accounting fraud. In addition, I also highlight, through a worked example, how Tableau, a visual analytics tool, can be used to perform Benford's law's first-digit test to detect potential fraud.

Keywords

Benford's law, fraud, tableau, visual analytics

1 INTRODUCTION

In 1938, in examining over 20,000 observations from a diverse range of datasets, including datasets on the areas of rivers and the atomic weights of element, Frank Benford, an American Physicist, observed a consistent pattern where small digits occurred more frequently in the first position of numbers than larger digits (Benford, 1938). This observation laid the foundation for the mathematical tenet that has become known as Benford's law, which defines the expected frequency that digits appear in data.

Benford's law has been examined extensively in a wide range of areas including in mathematics (Hill, 1995; Newcomb, 1881), the physical sciences (Sambridge, Tkalčić, & Jackson, 2010), and business (Giles, 2007; Judge & Schechter, 2009). Research has also examined Benford's law in the accounting setting. For example, in the area of tax accounting, Nigrini (1996) examined how Benford's law can be used to investigate tax compliance among tax payers. In the area of audit, Nigrini and Mittermaier (1997) examined how Benford's law could be used as an effective aid in analytical procedures in the planning stage of an audit while Nigrini and Miller (2009) examined how second-order tests of Benford's law can be used to detect unusual issues related to data integrity that might not have been easily detectable using traditional audit analytical procedures.

Benford's law has also been examined as a useful tool to detect potential accounting fraud. In a survey conducted on 86 accountants to gain insights into the perceptions of fraud detection and prevention methods, Bierstaker, Brody, and Pacini (2006) found that the accountants rated "digital analysis," which is based on Benford's law, as the tenth (out of 34) most effective fraud detection procedure. Consistent with this, various accounting studies (e.g., Durtschi, Hillison, & Pacini, 2004; Kumar & Bhattacharya, 2007) highlight the applications of Benford's law as an easy to implement data mining technique that can effectively determine the authenticity or otherwise of a set of accounting data.

Beyond the research setting, Benford's law has also been demonstrated to have been an effective tool for fraud detection in practice. In June 2009, Bernie Madoff was sentenced to 150 years in prison for operating the largest Ponzi scheme in United States history (Nasaw, 2011). While the scale of the US$65 billion scam was eye-catching, what was perhaps more surprising was that Madoff managed to run the Ponzi scheme for decades without getting caught. In the aftermath of the high-profile case, forensic investigators have been left asking themselves whether there were any clues that could have alerted them to the scam earlier. As it turns out, the relatively obscure Benford's law could have pointed investigators to the fraud even before it began to unravel.

A Ponzi scheme often falsely leads investor to believe that profits derive from an underlying business when it in fact generates returns for early investors simply by acquiring new investors and using investments from these new investors to pay profits to older investors. Accordingly, any profits or returns published by a Ponzi scheme are likely to be fabricated because no real underlying business actually exists. To the extent that data on profits or returns are fabricated, they are not likely to obey Benford's law. Consistent with this, a first-digit test performed on the monthly returns from 1990 to 2008 of Fairfield Sentry Fund, one of Bernie Madoff's largest feeder funds, reveals a mean absolute deviation (MAD) of 0.030 which would suggests nonconformity with Benford's law. This nonconformity would, in turn, be indicative of potential fraud (Nigrini, 2012).[1]

In the rest of this article, I provide an introduction to Benford's law (Nigrini, 2005) and discuss how it can be employed to design tests to detect potential accounting fraud. In addition, I also highlight, through a worked example, how Tableau (Janvrin, Raschke, & Dilla, 2014), a visual analytics tool, can be used to perform Benford's law's first-digit test to detect potential fraud.

2 FRAUD DETECTION USING BENFORD'S LAW

In 1881, Simon Newcomb, an astronomer and mathematician, published the first article on Benford's law (Newcomb, 1881). In the article, Newcomb (1881) established that, according to Benford's law, the probability that a randomly generated number has any particular nonzero first digit is given by the following formula:

$$P(d) = \log_{10}(1 + 1/d)$$

where d represents a number between 1 and 9, and P represents the given probability. In applying the formula, Benford's law provides that in a randomly generated set of data, numbers within the dataset should have 1 as the most frequently appearing first digit (about 30.1% of the time) and have 9 as the least frequently appearing first digit (about 4.6% of the time). Exhibit 1 summarizes some expected digit frequencies based on Benford's law.

EXHIBIT 1. Expected digit frequencies based on Benford's law

Digit	First position	Second position	Third position
0	n/a	12.0%	10.2%
1	30.1%	11.4%	10.1%
2	17.6%	10.9%	10.1%
3	12.5%	10.4%	10.1%
4	9.7%	10.0%	10.0%
5	7.9%	9.7%	10.0%
6	6.7%	9.3%	9.9%
7	5.8%	9.0%	9.9%
8	5.1%	8.8%	9.9%
9	4.6%	8.5%	9.8%

Knowledge of Benford's law has allowed forensic investigators to design related tests that can effectively detect fraud (Diekmann & Jann, 2010). In particular, Benford's law based tests represent tests of abnormal duplication, where the actual frequencies with which digits appear in numbers in a dataset are tabulated and compared with the expected distribution of these digits as predicted by Benford's law. Where significant deviations between these actual and expected frequencies are detected, Benford's law tests would highlight these deviations as anomalies that should be investigated for potential fraud (Nigrini, 2012).

Three commonly employed Benford's law based tests are the first-digit test, the second-digit test, and the first-two digits test. In the first-digit and second-digit tests, the actual frequencies with which a number's first and second digits (respectively) appear in a dataset are compared with their respective expected frequencies predicted by Benford's law. In contrast, the first-two digits tests compares the actual frequencies with which a number's first-two digits appear in a data set with their respective frequencies predicted by Benford's law. While the first-digit and second-digit tests examine a relatively small range of digits—the first digit tests examines digits from 1 to 9 and the second digit test examines digits from 0 to 9—and represent high level tests which are design to only provide a general indication of abnormal duplications in data, the first-two digits test examines a wider range of digits—it examines digits from 10 to 99—and is a more focused test that can detect abnormal duplication and possible bias in the data (Diekmann, 2007; Nigrini, 2012).

To evaluate the results from Benford's law based tests, accounting researchers often rely on MAD as a test to assess the extent of a dataset's conformity to Benford's law (Drake & Nigrini, 2000). The formula to compute MAD is given by:

$$\frac{1}{N}\sum_{1}^{N} f_i \lfloor X_i - \bar{X} \rfloor$$

where N represents the number of digits being examined, X_i represents the actual digit from the dataset, $\bar{X}$ represents the expected digit, and f_i represents the frequency (of X_i and $\bar{X}$). The larger the MAD, the larger the deviations between actual and expected frequencies in a given dataset, and, correspondingly, the larger the deviations from Benford's law. Exhibit 2 summarizes the critical MAD range values for the first digit, second digit, and first-two digits tests developed by Drake and Nigrini (2000).

EXHIBIT 2. Critical range values for mean absolute deviation

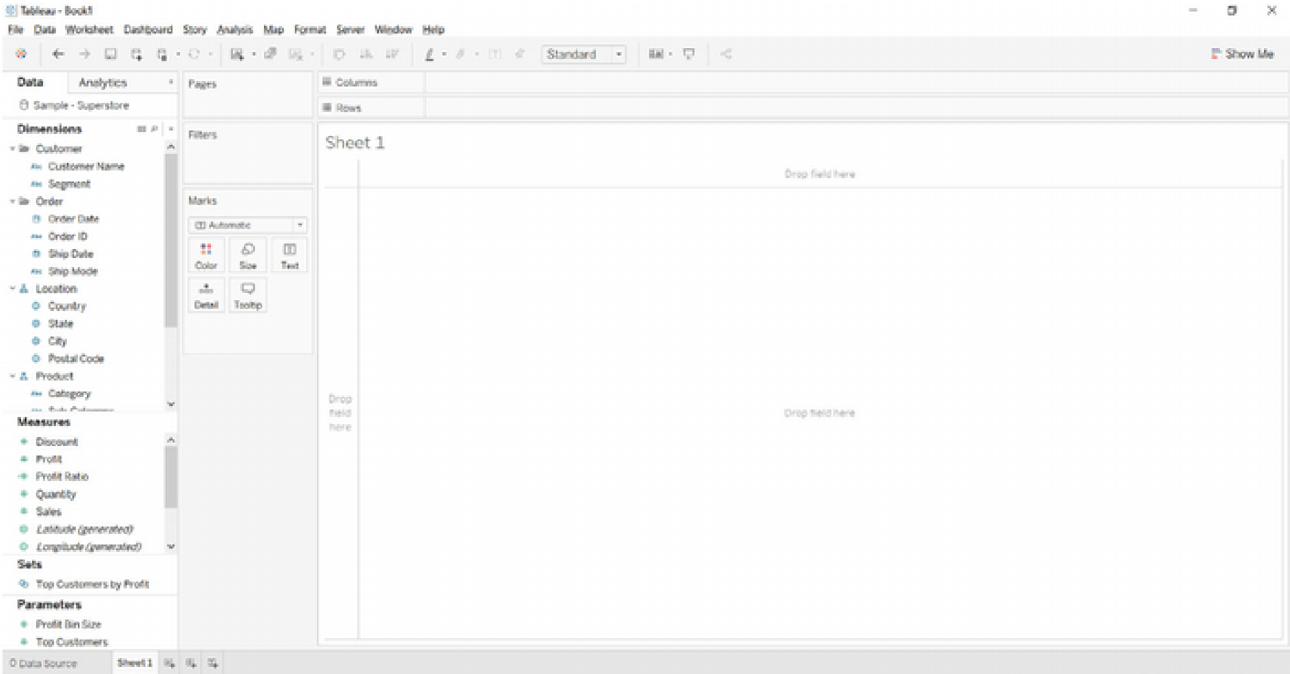
Conformity range	First digits	Second digits	First-two digits
Close conformity	0.000–0.006	0.000–0.008	0.0000–0.0012
Acceptable conformity	0.006–0.012	0.008–0.010	0.0012–0.0018
Marginally acceptable conformity	0.012–0.015	0.010–0.012	0.0018–0.0022
Nonconformity	Above 0.015	Above 0.012	Above 0.0022

3 APPLYING BENFORD'S LAW USING TABLEAU

Tableau is a visual analytics tool that can be used to perform visual analytics and create data visualizations (Murray, 2014). It has been widely used in the area of accounting, including for detecting fraud (Dilla & Raschke, 2015; Hoelscher & Mortimer, 2018; Janvrin et al., 2014). In this section, I illustrate, using a worked example, how Tableau can be used as a visual analytics tool to detect potential fraud using Benford's law. In particular, I conduct Benford's law's first-digit test on a set of sales data from the "Sample-Superstore" dataset that comes preinstalled with Tableau. To connect to this dataset, I launch the Tableau software and click on "Data" ➔ "New Data Source" ➔ "Sample-Superstore."

Tableau visualizations are created on worksheets. To create a new worksheet, I clock on "Worksheet" ➔ "New Worksheet." All the data fields available in the "Sample-Superstore" dataset for use in building a visualization are displayed on the left panel of the screen, and are categorized as "Dimensions" and "Measures." Dimensions contain qualitative values while Measures contain numeric, quantitative values that can be measured. Exhibit 3 presents the Tableau worksheet interface along with corresponding Dimensions and Measures from the "Sample-Superstore" dataset.

EXHIBIT 3. A Tableau worksheet interface connected to the "Sample-Superstore" dataset



To begin my analysis, I create two calculated fields, *Leftmost Integer* and *Benford's Law*. A calculated field in Tableau allows a user to create a new field in the dataset, the values or members of which are determined by a calculation formulated by the user. Calculated fields are created by clicking on "Analysis" ➔ "Create Calculated Field" and entering the relevant formula. *Leftmost Integer* is created using the formula *LEFT*(*STR*([*Sales*]),1), which extracts the leftmost digit from the sales value of each recorded sales transaction in the dataset. *Benford's Law* is created using the formula

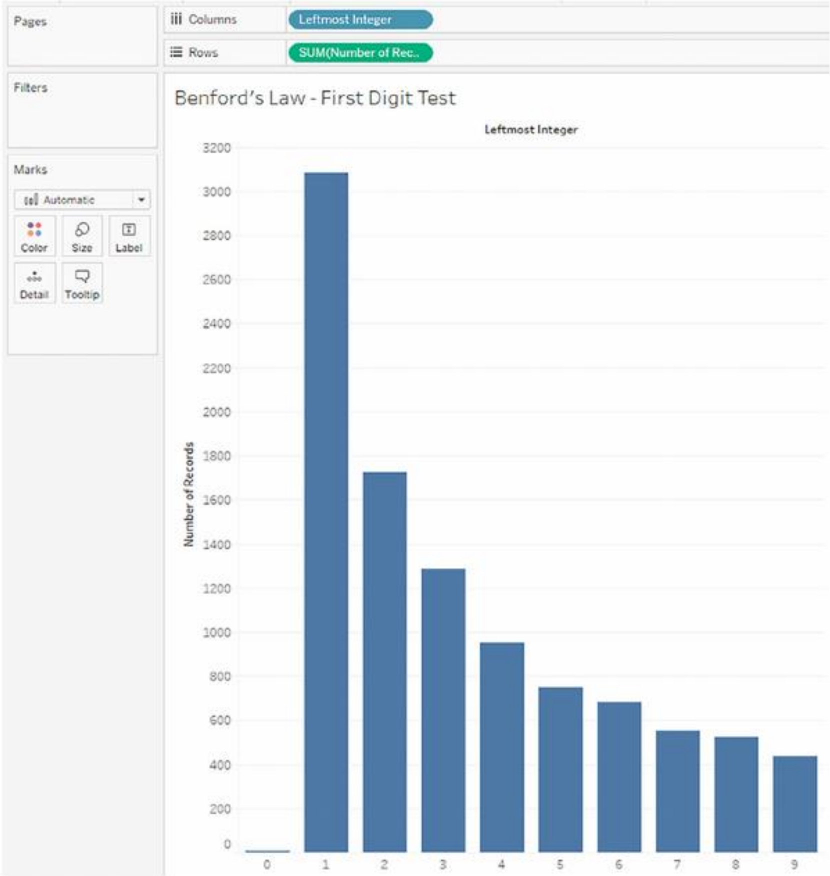
(*LOG*(*INT*([*Leftmost Integer*]) + 1) − *LOG*(*INT*([*Leftmost Integer*]))) * 100, which computes the expected proportion of each first-digit integer according to Benford's law. Exhibit 4 presents the formulas used in creating these two calculated fields.

EXHIBIT 4. Formulas for the calculated fields "*Leftmost Integer*" and "*Benford's Law*"

Calculated field	**Formula**
Leftmost Integer	*LEFT*(*STR*([*Sales*]),1)
Benford's Law	(*LOG*(*INT*([*Leftmost Integer*]) + 1) − *LOG*(*INT* ([*Leftmost Integer*]))) * 100

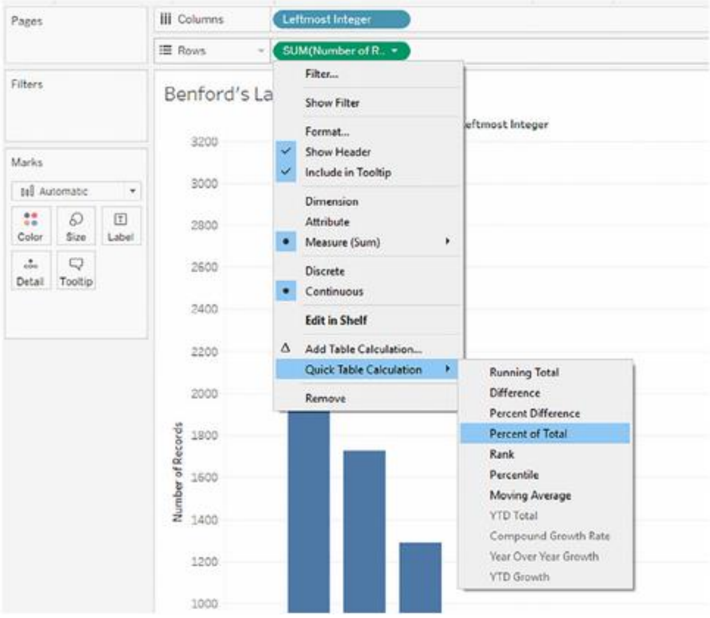
Following the creation of the calculated fields, I create a bar chart by dragging and dropping the *Leftmost Integer* field into the Columns shelf and the *Number of Records* field into the Rows shelf. This bar chart is displayed in Exhibit 5. This visualization presents the integers being examine in a first-digit test on the horizontal axis and the frequency of sales transactions with first digits corresponding to these integers on the vertical axis.

EXHIBIT 5. Bar chart with "*Leftmost Integer*" in the Columns shelf and "*Benford's Law*" in the Rows shelf



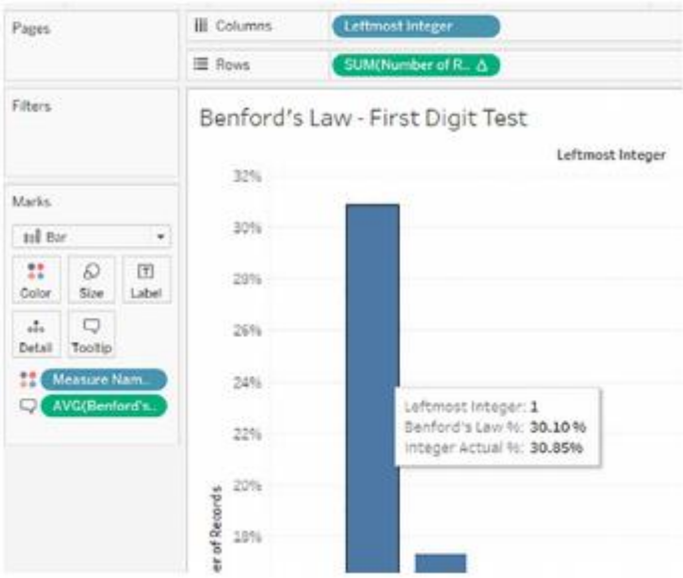
In Exhibit 5, the vertical axis displays the frequency of sales transactions corresponding to each integer being examined (on the horizontal axis). I modify the bar chart to instead display these frequencies as a percentage of the total number of sales transactions in the dataset. I do so by clicking on the dropdown arrow on the *Sum*(*Number of Records*) pill (in the Rows shelf) ➔ "Quick Calculation Table" ➔ "Percent of Total." Exhibit 6 presents the sequence of steps in performing this quick table calculation.

EXHIBIT 6. Performing a quick table calculation on *Sum*(*Number of Records*)



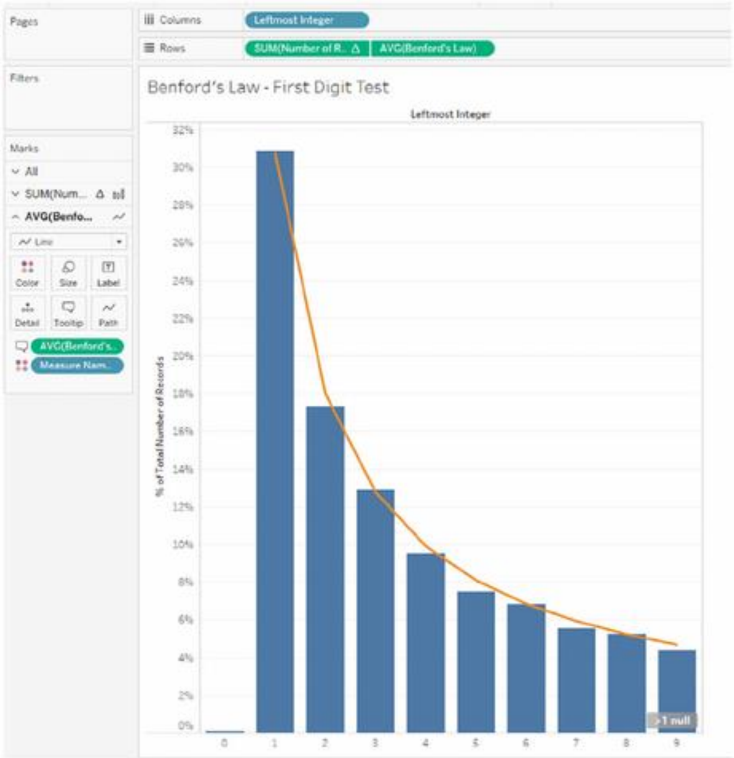
Next, I drag and drop the *Benford's Law* field into "Tooltip" in the Marks area. I then click on "Tooltip" in the Marks area and edit the tooltip to display information about the actual proportion of sales transactions corresponding to each integer under examination along with the expected proportions predicted by Benford's law. Exhibit 7 presents the tooltip information displayed when I hover the curser over the bar corresponding to the integer 1.

EXHIBIT 7. Tooltip information displayed when curser hovers over the bar representing the integer 1



Finally, I add a line graph to the visualization to display the expected proportions of each integer as predicted by Benford's law. I do so by dragging and dropping *Benford's Law* to the Rows shelf, clicking on the *Benford's Law* pill and selecting to utilize a dual axis, and then selecting a "Line" in the Marks area for the Benford's law visualization. Exhibit 8 presents my modified bar chart. Overall, the visualization presented in Exhibit 8 displays, for each integer I examine, the extent to which the proportion of actual sales transactions conforms to the proportions predicted by Benford's law. Inspecting this visualization will allow a forensic investigator to conduct a first-digit test and more easily detect deviations from Benford's law in the data.

EXHIBIT 8. Visualization for Benford's law's first-digit test



4 CONCLUSION

In this article, I examine Benford's law and discuss how it can be effectively employed to detect fraud. In addition, I illustrate, using a worked example, how Tableau can be used as a visual analytics tool to perform Benford's law's first-digit test on a sample of data.

Knowledge of Benford's law is important because it allows forensic investigators to design tests that can effectively detect fraud. In particular, I highlight Benford's law's first digit, second digit, and first-two digits tests as examples of commonly employed tests of abnormal duplication, where the actual frequencies with which digits appear in numbers in a dataset are tabulated and compared with the expected distribution of these digits as predicted by Benford's law. Where significant deviations between these actual and expected frequencies are detected, Benford's law tests would flag these deviations as anomalies that should be investigated for potential fraud.

REFERENCES

Benford, F. (1938). The law of anomalous numbers. Proceedings of the American Philosophical Society, 78, 551–572.

Bierstaker, J. L., Brody, R. G., & Pacini, C. (2006). Accountants' perceptions regarding fraud detection and prevention methods. Managerial Auditing Journal, 21(5), 520–535.

Diekmann, A. (2007). Not the first digit! Using Benford's law to detect fraudulent scientific data. Journal of Applied Statistics, 34 (3), 321–329.

Diekmann, A., & Jann, B. (2010). Benford's law and fraud detection: Facts and legends. German Economic Review, 11(3), 397–401.

Dilla, W. N., & Raschke, R. L. (2015). Data visualization for fraud detection: Practice implications and a call for future research. International Journal of Accounting Information Systems, 16,1–22.

Drake, P. D., & Nigrini, M. J. (2000). Computer assisted analytical procedures using Benford’s law. Journal of Accounting Education, 18(2), 127–146.

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.

Giles, D. E. (2007). Benford’s law and naturally occurring prices in certain ebay auctions. Applied Economics Letters, 14(3), 157–161.

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

Hoelscher, J., & Mortimer, A. (2018). Using Tableau to visualize data and drive decision-making. Journal of Accounting Education, 44, 49–59.

Janvrin, D. J., Raschke, R. L., & Dilla, W. N. (2014). Making sense of complex data using interactive data visualization. Journal of Accounting Education, 32(4), 31–48.

Judge, G., & Schechter, L. (2009). Detecting problems in survey data using Benford’s law. Journal of Human Resources, 44(1), 1–24.

Kumar, K., & Bhattacharya, S. (2007). Detecting the dubious digits: Benford’s law in forensic accounting. Significance, 4(2), 81–83.

Murray, D. (2014). What Tableau is and is not. Interworks. Retrieved from https://interworks.com/blog/dmurray/2014/12/29/what-tableau-and-not/

Nasaw, D. (2011). Timeline: Key dates in the Bernard Madoff case. The Guardian. Retrieved from https://www.theguardian.com/business/2009/mar/12/bernard-madoff-timeline-fraud

Newcomb, S. (1881). Note on the frequency of the use of the different digits in natural numbers. American Journal of Mathematics, 4,39–40.

Nigrini, M. (2012). Benford’s law: Applications for forensic accounting, auditing, and fraud detection. Hoboken, NJ: John Wiley & Sons.

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

Nigrini, M. J., & Miller, S. J. (2009). Data diagnostics using second-order tests of Benford’s law. Auditing: A Journal of Practice & Theory, 28(2), 305–324.

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), 52–67.

Nigrini, M. J. (2005). An assessment of the change in the incidence of earnings management around the Enron-Andersen episode. Review of Accounting and Finance, 4,92–110.

Sambridge, M., Tkalcic, H., & Jackson, A. (2010). Benford’s law in the natural sciences. Geophysical Research Letters, 37(22), L22301.

Biography

Clarence Goh is Assistant Professor of Accounting (Practice) at the School of Accountancy, Singapore Management University.