

METHOD ARTICLE

A data transformation process for using Benford's Law with bounded data [version 1; peer review: 2 approved]

Daniel McCarville ^{1,2}¹Educational Methodology, Policy, and Leadership, University of Oregon, Eugene, OR, 97403, United States²Corporate Audit Services, Jack Henry, Inc., Shawnee, KS, 66227, United States

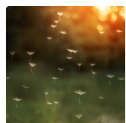
V1 First published: 25 Nov 2021, 3:29
<https://doi.org/10.35241/emeraldopenres.14374.1>
Latest published: 25 Nov 2021, 3:29
<https://doi.org/10.35241/emeraldopenres.14374.1>

Abstract

Benford's Law is an empirical observation about the frequency of digits in a variety of naturally occurring data sets. Auditors and forensic scientists have used Benford's Law to detect erroneous data in accounting and legal usage. One well-known limitation is that Benford's Law fails when data have clear minimum and maximum values. Many kinds of education data, including assessment scores, typically include hard maximums and therefore do not meet the parametric assumptions of Benford's Law. This paper implements a transformation procedure which allows for assessment data to be compared to Benford's Law. As a case study, a data quality assessment of oral language scores from the Early Childhood Longitudinal Study, Kindergarten (ECLS-K) study is used and higher risk data segments detected. The same method could be used to evaluate other concerns, such as test fraud, or other bounded datasets.

Keywords

Benford's Law, educational assessment, data reliability, data quality, secondary education, audit




This article is included in the [Quality Education for All](#) gateway.

Open Peer Review

Approval Status  

	1	2
version 1 25 Nov 2021		

1. **Mark Nigrini**, West Virginia University,
Morgantown, United States

2. **Nooraslinda Abdul Aris** , Universiti
Teknologi MARA, Shah Alam, Malaysia

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Daniel McCarville (daniel.mccarville@gmail.com)

Author roles: McCarville D: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2021 McCarville D. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: McCarville D. **A data transformation process for using Benford's Law with bounded data [version 1; peer review: 2 approved]** Emerald Open Research 2021, 3:29 <https://doi.org/10.35241/emeraldopenres.14374.1>

First published: 25 Nov 2021, 3:29 <https://doi.org/10.35241/emeraldopenres.14374.1>

Introduction

Benford's Law is an empirical observation about the frequency of digits in a variety of naturally occurring data sets. Although previously observed by [Simon Newcomb \(1881\)](#), it was more famously published in scholarly literature by [Frank Benford \(1938\)](#) who collected 20,000 first digits from an impressively diverse set of sources such as the distances between cities in a road atlas, numbers printed in an arbitrary issue of *Reader's Digest*, tables of mathematical constants, addresses of persons listed in the *Annals of Science*, and a dozen more. Across all of these data sources he noted that the first significant digit of numbers followed the same pattern. He formalized this pattern in what we now call Benford's Law, in which the probability of the first significant digit being $d \in \{0, 9\}$ is defined by a logarithmic function $pr(D = d) = \log(1 + 1/d)$. Analyses which attempt to detect deviations from Benford's Law are called Benford's analyses. Benford's analyses have been adopted by auditors and financial examiners as a way of detecting problems in accounting records, but have expanded to numerous other disciplines as well.

This paper implements a novel data transformation technique which allows Benford's analyses for bounded data sets. Academic assessment data is often bounded, meaning that the range of scores includes a set minimum or maximum value. This may be the case in simple assessments in which the score is the number of items answered correctly, number of words read, number of pushups, etc. Bounded data do not conform to Benford's Law, limiting its application to assessment, operational, and other kinds of data. Although the basis for a data transformation which overcomes this limitation exists, it has not yet been implemented. Although Benford's Law has exploded across a variety of disciplines in the last 10 years, it is relatively unknown in educational assessment. Accordingly, this paper is positioned both as a methodological improvement in Benford's Law, as well as an introduction to Benford's Law for academic assessment usage.

Benford's analysis in auditing

Benford's Law was treated as a trivial mathematical observation for most of its history. While it was initially observed in 1881, the first applications weren't developed until the 1980s. Early applications involved detecting irregularities in engineering data ([Becker, 1982](#); [Nelson, 1984](#)). In 1988 the first accounting application was developed. [Carlsaw \(1988\)](#) examined financial statements of corporations in New Zealand under the premise that accounting values represent natural activities and should conform to Benford's Law. He found that too many entries ended in 0 and too few ended with 9, consistent with improperly rounding up earnings to create the appearance of stronger financial performance. [Nigrini \(1996a\)](#) popularized Benford's analysis with auditors by demonstrating its usage in detecting fraudulent income tax filings. Auditing and accounting literature in the late 1990s - early 2000s is replete with Benford's Law applications in a variety of usages.

Benford's Law is not only of interest to the academic community. The auditing community has adopted it as a common

tool for evaluating the risks of data irregularities, including fraud. Training materials such as books, trade publication articles, and courses in Benford's Law are provided regularly by major professional organizations such as the American Institute of Certified Public Accountants, Institute of Internal Auditors, Association of Certified Fraud Examiners, and ISACA (an organization for professionals in auditing information technology and security). This training has become ubiquitous enough that Benford's Law is now a common subject in undergraduate accounting courses in the United States, and perhaps elsewhere. Common auditing software such as Audit Command Language (ACL), Caseware IDEA, and TeamMate all include Benford's Law as a standard feature. Outside of training, Benford's Law provides for real needs of accounting firms. [Nigrini \(1996b\)](#) described Benford's Law as a tool for accountants and auditors to demonstrate that they have met professional obligations to consider fraud. This view was tested during a class action lawsuit related to the Bernie Madoff ponzi scheme fraud ([New York Law School v. J. Ezra Merkin, 2010](#)). Investors sued J. Ezra Merkin, an investment advisor whose firm was responsible for decisions to invest at least \$2.5 billion into the ponzi scheme. They claimed that Merkin had failed to provide basic due diligence when choosing investments, including his failure to utilize Benford's Law.

Beyond auditing

Benford's Law has expanded to a wide variety of disciplines. *Benford Online Bibliography*, a web-accessible reference for Benford's Law-related research, includes approximately 1,600 published works across a variety of disciplines ([Berger et al., 2009](#)). Summarizing this literature would not be economical; key areas of research are highlighted here. One such broad area is physical measurement. [Benford \(1938\)](#) initially noted that measurements of atomic weight, molecular weight, black body radiation, specific heat, the lengths, areas, and drainage of rivers, and other physical values conform to Benford's Law. Numerous other physical constants followed in the 21st century. For example, [Sambridge et al. \(2010\)](#) published a similar summary of 15 other physical measurements which conform to Benford's Law including the mass of exoplanets, greenhouse gas emissions by country, and the rotational frequency of pulsars. The knowledge that these measures follow Benford's Law has been applied to test the quality of data. For example, by knowing that physical measures of river and lake size conform to Benford's Law it was possible to evaluate hydrological databases to uncover omitted data ([Nigrini & Miller, 2007](#)).

Computer security and forensics have been a fertile field for Benford's Law. Activity along both computer and social networks follows Benford's Law under some conditions, making it possible to detect anomalous activity. As a kind of anomalous network traffic, denial of service attacks can be detected with Benford's Law ([Prandl, 2017](#)). Social network activity such as friend or follower counts conforms to Benford's Law, and it can be used to quickly locate automated bots in a pool of normal human users ([Madahali & Hall, 2020](#)). Benford's Law can be used to examine the data underlying images and videos, including biometric data. It can help detect double compression

schemes (Frick *et al.*, 2020) which can accompany image manipulation, as well as steganographic techniques for concealing messages inside images. In addition to locating manipulated images, Benford's Law also has a use in locating synthetic computer-generated images ("deep fakes") (Bonetti *et al.*, 2020). Approaches for applying Benford's Law to images have been extended to biometrics data, which allows for the detection of tampering of fingerprints (Aamo & Caleb, 2017).

Benford's Law in educational research

Educational research is one field where Benford's Law has not been previously explored. Although there is a small, but extant body of literature on Benford's Law in education it focuses on pedagogical issues related to teaching Benford's Law to students in various course levels and disciplines (Drake & Nigrini, 2000; Linville, 2008; Mills *et al.*, 2020). Only a single extant publication listed in *Benford Online Bibliography* uses Benford's Law analytically and none appear in the EdArXiv preprints. Perhaps the only existing example in assessment is Slepko *et al.* (2015), who analyzed the potential answers to multiple-choice questions in physics and chemistry textbooks and found that those answers conformed to Benford's Law. Notable perhaps for not using Benford's Law are Ashcroft & Kelly (1995) who counted the frequencies of numbers appearing in grade 1–6 mathematics problems. No existing applications in the analysis of assessment scores presently exist.

Challenges and advancements

Auditors' focus on fraud and compliance led to growing pains. Critics claimed that Benford's analysis often rejected data even when no fraud was present, leading to a high false-positive rate (Diekmann & Jann, 2010). Benford's analysis literature addressed this in three ways. First, improved statistical procedures were created that were more appropriate for Benford's analysis and could limit the risk of false positives (Cho & Gaines, 2007; Fu *et al.*, 2007; Morrow, 2014; Winter *et al.*, 2012). Second, cognition research was able to articulate a theory describing how recalling and inventing numbers are fundamentally different psychological processes (Burns, 2009; Chi, 2020). This theory linked the act of committing certain kinds of financial fraud to numerical deviations from Benford's Law, which provided a stronger footing for claims that it could actually succeed in detecting fraud. Finally, Benford's analysis was reconceptualized as a general-purpose data integrity tool which could detect a variety of different irregularities such as missing data (Nigrini & Miller, 2007), imprecise data (Judge & Schechter, 2009) as well as distinguish between data produced by different processes (Kreuzer *et al.*, 2014).

Benford's analysis is a parametric test in which empirical data is tested for its conformity to the logarithmic distribution specified in Benford's Law. When parametric assumptions are not met, analysts have a few options: they may alter their parametric assumptions, transform their data to meet the parametric assumptions, or opt for non-parametric tests. Most attempts to adapt Benford's analysis to non-Benford distributed digits have focused on adjusting Benford's Law. One such approach is to re-parameterize Benford's Law to allow for shape and rate

parameters (Fu *et al.*, 2007). Alternatively, if the distribution of the data is known Winter *et al.* (2012) showed that exact expected digit frequencies could be derived analytically.

Data transformation methods are particularly important for educational researchers. Hard minimum or maximum values in a dataset make it unsuitable for Benford's analysis (Nigrini, 2011). Unfortunately, many assessment instruments produce scores with clear minimum or maximum values based either on the number of items on the assessment or the scoring mechanism. Similarly, Nigrini and others have observed that data which range over many orders of magnitude are more amenable to Benford's analysis (Fewster, 2009; Nigrini, 2011); this range is also uncommon in assessment instruments.

Techniques for transforming data that does not satisfy Benford's Law into a more appropriate form are less common. It's been long-known that multiplying Benford-compliant data by a constant results in data that still follows Benford's Law (Pinkham, 1961), but multiplying non-Benford-compliant data by a constant results in data that is still not Benford-compliant (Raimi, 1969). Extending this property to produce a useful transformation method has been elusive. Jamain (2001) demonstrated that if a random value is divided by a random number from a Benford-compliant distribution, then the result is a new value that is also Benford-compliant. However, this stochastic method is unsuitable for analytic needs because the transformed values can't be traced back to a non-transformed value, which is problematic for field usage. An alternative line of research proved that exponentiation to an arbitrary large power would eventually result in a Benford-compliant data (Adhikari & Sarkar, 1968). However, a specific procedure wasn't practicable until Morrow (2014) showed that raising a dataset to the tenth power was sufficient to result in a Benford-compliant distribution with satisfying characteristics. This paper presents the first application of this data transformation technique.

Methods

Early Childhood Longitudinal Study Kindergarten Oral Language Development Scale (ECLS-K OLDS) data and analysis

ECLS-K OLDS (National Center for Education Statistics, 2009) scores were used to test whether Benford's analysis could be used with assessment data. The ECLS is a series of four longitudinal studies which track educational and socioemotional development. Each study included a different age or grade range; ECLS-K followed a cohort of kindergarten students in 1999 through the 8th grade. The cohort contains approximately 21,000 children drawn from across the United States.

The OLDS was administered as a language screener to students whose primary home language was not English. If the student's OLDS score was above a threshold they were administered the full English-language assessment battery. Although in principle any assessment score that is an interval or ratio measurement could be used with Benford's Law, OLDS scores presented several advantages. First, the ECLS-K is a well-known dataset. Using an established dataset aids interpretation and

provides a clear application of new analytical techniques. Second, OLDS scores are significant to the rest of the ECLS-K assessment scores. If the OLDS scores prove to be unreliable for any reason, it could cascade to later assessments and disrupt the subject-specific assessment scores. Focusing on the quality of the OLDS screener data is therefore more practical than focusing on the subject-specific assessments. Third, the OLDS instrument and items have already undergone reliability assessments. For the purpose of the analysis, the initial belief is that this data is free from material errors. That is, it possesses no errors which would alter users' conclusions. If the analysis does not reject conformance to Benford's Law, this would match the existing belief.

CISCTOT is the variable representing fall 2010 total OLDS scores from the English-language form. 2, 849 numerical responses were available, excluding non-responses and procedural flags. Scores of zero were also removed, because the probability distribution in Benford's Law is not defined for zero. After removing those values 2,563 scores remained. No weighting procedure was applied. Data was analyzed using Python 3.9.1 (Van Rossum & Drake, 2009) and the Benfords package version 1.0.2 (McCarville, 2021). Field practitioners often conduct Benford's analyses by visually inspecting the frequency of digits. For analytical usage a measure of fit is more appropriate. An increasingly common test statistic in the literature is the *d* statistic created by Cho & Gaines (2007). Morrow (2014) provided a modified *d* statistic, *d*^{*}, which is adjusted for sample size. Analysis will be conducted both visually and using the *d* statistic, or *d*^{*} when comparisons to samples or subsets of different sizes are needed.

Results

The naive (untransformed) OLDS scores do not adhere to Benford's Law. Figure 1 below compares the expected first-digit frequencies from Benford's Law to the empirical frequency of first digits among the fall 2010 OLDS scores. Benford's Law anticipates a monotonically decreasing frequency of digits; the OLDS scores instead increase, then decrease. The *d* statistic is 0.31. No hypothesis test is conducted, because the goal of this analysis is not to accept or reject the fit of the distribution, but to prioritize quality assurance resources. This matches with the general field usage of Benford's Law in assessing risks, not testing hypotheses.

The transformation process from Adhikari & Sarkar (1968) was used to transform this non-Benford compliant data into a more appropriate form. This is accomplished by raising the data to arbitrarily high power. Morrow (2014) found that *X*¹⁰ was sufficient for simulated data following known distribution. One key concern is whether the transformation procedure can possibly result in an array of scores in which each first significant digit is possible. OLDS scores range from 1–60. Raising each of those to the tenth power shows that it is possible for transformed values to begin with any digit 1–9, although a first-significant digit of 7 is only possible when the OLDS score is 49.

Figure 2 shows the results of the Benford's analysis using the transformed OLDS scores. The distribution now more closely resembles the monotonically-decreasing pattern expected by Benford's Law. The *d* statistic is 0.10, about 1/3 of the value of the untransformed data. Clearly the transformation procedure

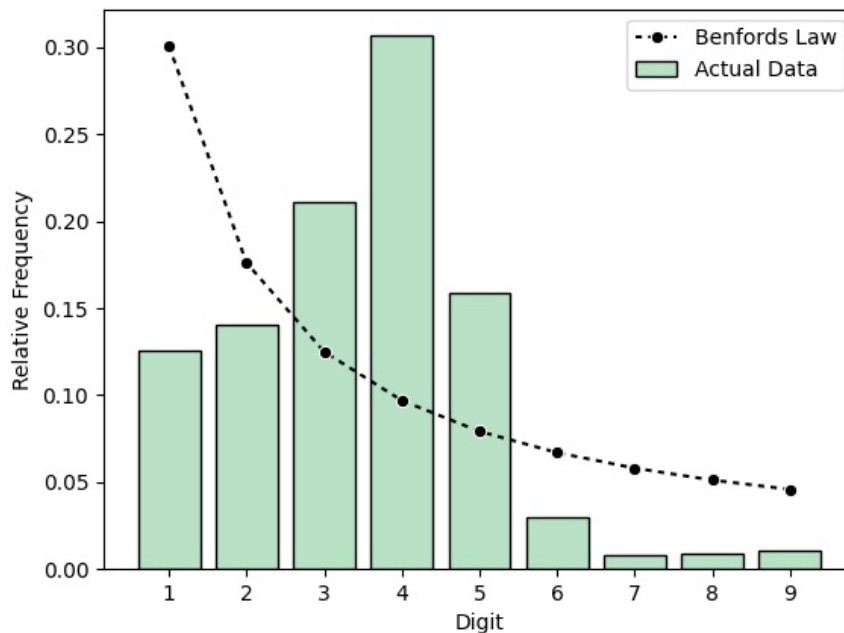


Figure 1. Comparison of Oral Language Development Scale (OLDS) scores to Benford's Law before being transformed.

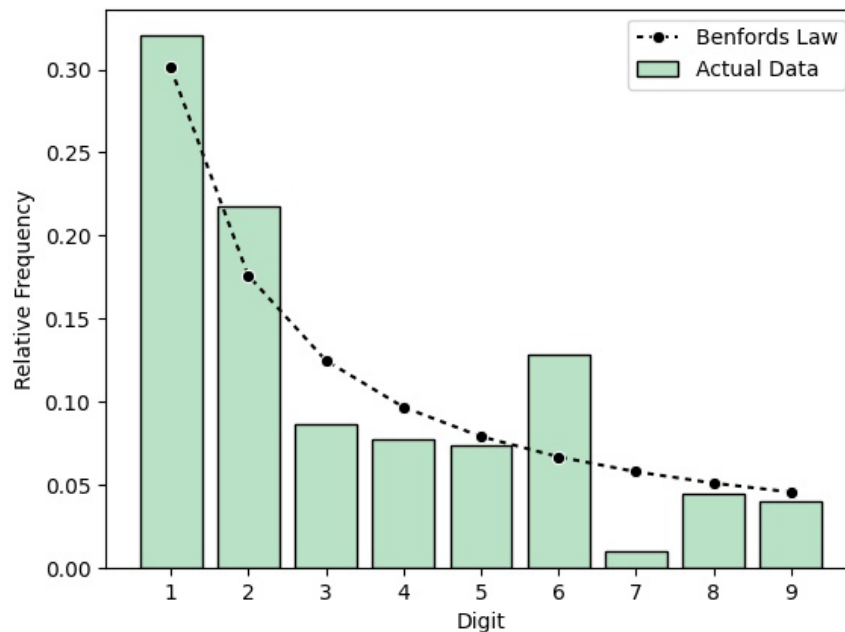


Figure 2. Comparison of Oral Language Development Scale (OLDS) scores to Benford's Law after being transformed.

has resulted in a better fit to Benford's Law. There is a large spike at the digit 6 and a lack of 7s; since the transformation procedure is unlikely to result in a first-significant digit of 7 in this dataset, this may be an artifact of transformation.

If the data transformation process is successful it should discriminate between high and low-risk subsets. High-risk subsets are where practitioners would want to deploy QA resources to ensure that data integrity is maintained. Table 1 shows the results of a Benford's analysis when subsetting the data by race and region of the United States. Using the d^* statistic, variance between regions and ethnicity categories is apparent. The least-conforming subsets are the western region, Asians, and Hispanic (Race not Specified). These categories are also among the largest. The same process could be used to refine smaller subsets for investigation. For example, within the western region an investigator could look for patterns in racial categories, resulting in a finer-grained subset of potentially erroneous data.

Discussion

The two goals of this paper were to implement a novel data transformation procedure for Benford's Law, and to introduce Benford's Law as a tool for evaluating academic assessment data. The data transformation procedure was capable of taking assessment scores which were intrinsically not suited for Benford's analysis and render them amenable to it. The analysis was also capable of distinguishing between higher risk and lower risk subsets (i.e., regions and racial categories). This same analysis could be conducted on any unit of aggregation such as assessment administrator, classroom, school district, or

any other category which may be of interest. Although a variety of reliability and fidelity measures have been published already, Benford's analysis is a well-established tool in other disciplines which can expand assessment analysts' options. It is uniquely well-positioned to detect irregularities in aggregates, such as an administration error which slightly alters students' scores in a single building, or test fraud by an administrator or educator.

This research can be placed in a broader research context which seeks to develop methods to extend Benford's analysis to datasets that intrinsically do not conform to Benford's law. Morrow (2014) as well as Adhikari & Sarkar's (1968) analytical work describe a data transformation method based on exponentiation, but this method had lacked an extant application until now. Additionally, this paper also provides a similar application of Morrow's (ibid) d^* statistic, which improves the more common d statistic by making it responsive to sample size. These methodological advancements both expand the universe of applications for Benford's Law and improve rigor.

Auditors may be interested in these results, because they extend the level of assurance that can be provided in a variety of engagements. Auditors already use Benford's Law to examine financial data such as transactions and financial statements. Data that does not conform to Benford's Law may be considered a higher risk and audited in more depth. However, there is a variety of financial data that is restricted to only a few orders of magnitude or otherwise unsuitable for Benford's analysis, such as payment messages for foreign currencies (Krakar & Zgela, 2009). Financial auditors now have a simple way to

Table 1. Comparison of transformed and untransformed Oral Language Development Scale (OLDS) scores, segmented by student region and race.

		N	D* Statistic			% Change
			X	X ^o	X-X ^o	
Region (n=2,563)	Northeast	434	8.58	2.19	6.39	74.53%
	Midwest	384	7.33	1.74	5.58	76.20%
	South	553	6.84	2.75	4.09	59.75%
	West	1,192	9.39	3.82	5.57	59.30%
Race (n=2,563)	White, Non-Hispanic	191	6.36	2.01	4.35	68.35%
	Black or African American, Non-Hispanic	39	3.60	1.80	1.80	50.08%
	Hispanic, Race Specified	578	6.65	2.58	4.07	61.19%
	Hispanic, Race Not Specified	907	8.30	2.92	5.38	64.82%
	Asian	748	9.32	3.01	6.31	67.74%
	Native Hawaiiin, Other Pacific Islander	50	2.47	1.16	1.31	53.14%
	American Indian or Alaska Native	9	1.05	0.49	0.56	53.14%
	More Than One Race, Non-Hispanic	30	2.66	1.21	1.45	54.47%
	Not ascertained	11	1.69	1.02	0.66	39.30%

apply a well-known tool in their field to a greater variety of data. Performance auditors and others who typically deal in non-financial data are similarly benefitted from the extension of Benford's analysis into their domains. As the auditing profession moves toward more robust analytical procedures and computer-assisted auditing techniques, use of null-hypothesis testing and adequate critical values will be increasingly important. This paper demonstrates these tools.

One limitation of this project is that a more detailed field audit could not be conducted to determine the nature of the irregularity detected. Typically, a substantive review would occur to determine why the data was not compliant with Benford's Law and whether the non-compliance represents a weakness in the data set. In auditing applications, this may also be used to

determine whether fraud or some other kind of human-manipulation of data is likely to have occurred. In this case, no such field review was possible. Nonetheless, the transformation process was successful in turning data that was not expected to match Benford's Law into a form that did. The transformation process also retains a necessary property: that not all subsets were rendered Benford-compliant. Although the nature of the discrepancy is not known, the transformation process was successful.

Data availability

Source data

ECLS-K data used in this study is made publicly available by the National Center for Education Statistics: <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009005>.

References

Aamo I, Caleb SF: **On the Use of Benford's Law to Detect JPEG Biometric Data Tampering.** *J Inform Secur.* 2017; **8**(3): 240–256.

[Publisher Full Text](#)

Adhikari AK, Sarkar BP: **Distribution of most significant digit in certain functions whose arguments are random variables.** *Sankhya Ser. B.* 1968; **30**: 47–58.

[Reference Source](#)

Ashcroft MH, Christy KS: **The Frequency of Arithmetic Facts in Elementary Texts: Addition and Multiplication in Grades 1-6.** *J Res Math Educ.* 1995; **26**(5): 396–421.

[Publisher Full Text](#)

Becker PW: **Patterns in Listing of Failure-Rate and MTTF Values and Listings of Other Data.** *IEEE Trans Reliab.* 1982; **R-31**(2): 132–134.

[Publisher Full Text](#)

Benford F: **The Law of Anomalous Numbers**. *Proc Am Philos Soc*. 1938; **78**(4): 551–572.

[Reference Source](#)

Berger A, Hill TP, Rogers E: **Benford Online Bibliography**. 2009; Last accessed April 25.

[Reference Source](#)

Bonetti N, Bestagini P, Milani S, et al.: **On the use of Benford's law to detect GAN-generated images**. *Pre-print*. Retrieved September 10, 2020 from the arXiv database. 2020.

[Reference Source](#)

Burns B: **Sensitivity to statistical regularities: People (largely) follow Benford's law**. *Proceedings of the Annual Meeting of the Cognitive Science Society*. 2009; **31**(31).

[Reference Source](#)

Carlsaw CAPN: **Anomalies in income numbers: Evidence of goal oriented behavior**. *The Accounting Review*. 1988; **63**(2): 321–327.

[Reference Source](#)

Chi D: **First Digit Phenomenon in Number Generation Under Uncertainty: Through the Lens of Benford's Law**. [Published master's thesis]. University of Sydney. 2020.

[Reference Source](#)

Cho WKT, Gaines BJ: **Breaking the (Benford) Law: Statistical Fraud Detection in Campaign Finance**. *Am Stat*. 2007; **61**(3): 218–223.

[Reference Source](#)

Diekmann A, Jann B: **Benford's Law and Fraud Detection: Facts and Legends**. *Ger Econ Rev*. 2010; **11**(3): 397–401.

[Publisher Full Text](#)

Drake PD, Nigrini MJ: **Computer assisted analytical procedures using Benford's Law**. *J Account Educ*. 2000; **18**(2): 127–146.

[Publisher Full Text](#)

Fewster RM: **A Simple Explanation of Benford's Law**. *Am Stat*. 2009; **63**(1): 26–32.

[Publisher Full Text](#)

Frick RA, Liu H, Steinebach M: **Proceedings of the 15th International Conference on Availability, Reliability and Security**. 2020.

[Reference Source](#)

Fu D, Shi YQ, Su W: **A generalized Benford's Law for JPEG Coefficients and its applications to image forensics**. *Proc SPIE 6505, Security, Steganography, and Watermarking of Multimedia Contents IX 65051L*. 2007.

[Publisher Full Text](#)

Jamain A: **Benford's Law**. [Published master's thesis]. Imperial College of London. 2001.

Judge G, Schechter L: **Detecting Problems in Survey Data Using Benford's Law**. *J Hum Resour*. 2009; **44**(1): 1–24.

[Publisher Full Text](#)

Krakar Z, Zgela M: **Evaluation of Benford's Law Application in Stock Prices and Stock Turnover**. *Informatologia*. 2009; **42**(3): 158–165.

[Reference Source](#)

Kreuzer M, Jordan D, Antkowiak D, et al.: **Brain Electrical Activity Obeys Benford's Law**. *Anesth Analg*. 2014; **118**(1): 183–91.

[PubMed Abstract](#) | [Publisher Full Text](#)

Linville M: **Introducing digit analysis with an interactive class exercise**. *Academy of Educational Leadership Journal*. 2008; **12**(3): 55–69.

Madahali L, Hall M: **Application of the Benford's law to Social bots and Information Operations activities**. *2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*. 2020.

[Publisher Full Text](#)

McCarville DJ: **Benford's Law**. 2021.

[Reference Source](#)

Mills RJ, Beaulieu TY, Feldon DF, et al.: **Implications of Prelecture Material on Cognitive Load and Instructional Effectiveness in Cross-Disciplinary IS Education: The Nexus of Benford's Law and SQL**. *Decis Sci*. 2020; **18**(2): 313–338.

[Publisher Full Text](#)

Morrow J: **Benford's Law Families of Distributions and a Test Basis**. *CEP Discussion Paper No. 1291*. Center for Economic Performance, London School of Economics and Political Science. 2014; Accessed online July 7, 2020.

[Reference Source](#)

National Center for Education Statistics: **Early Childhood Longitudinal Study, Kindergarten Class of 1998-99**. NCEES 2009005. National Center for Education Statistics, Washington D.C. 2019; (accessed December 28, 2020).

[Reference Source](#)

Nelson L: **Technical aids**. *Journal of Quality Technology*. 1984; **16**: 175–176.

Newcomb S: **Note on the frequency of use of the different digits in natural numbers**. *Am J Math*. 1881; **4**: 39–40.

[Publisher Full Text](#)

New York Law School v. J. Ezra Merkin: **Third Consolidated Amended Class Action Complaint**. No. 08 Civ. 10922 (DAB) (U.S. District Court, Southern District of New York). 2010.

Nigrini MJ: **A Taxpayer Compliance Application of Benford's Law**. *Journal of the American Taxation Association*. 1996a; **18**(1): 72–91.

[Reference Source](#)

Nigrini MJ: **Digital Analysis and the Reduction of Auditor Litigation Risk**. *Deloitte & Touche/University of Kansas Symposium on Auditing Problems*. 1996b; 69–81.

[Reference Source](#)

Nigrini MJ: **Forensic Analytics: Methods and Techniques for Forensic Investigations**. John Wiley & Sons, Inc. 2011.

[Reference Source](#)

Nigrini MJ, Miller S: **Benford's Law Applied to Hydrology Data - Results and Relevance to Other Geophysical Data**. *J Int Ass Math Geol*. 2007; **39**: 469–490.

[Publisher Full Text](#)

Pinkham RS: **On the Distribution of First Significant Digits**. *Ann Math Stat*. 1961; **32**(4): 1223–1230.

[Reference Source](#)

Prandl S: **PEIMA: Harnessing Power Laws to Detect Malicious Activities from Denial of Service to Intrusion Detection Traffic Analysis and Beyond**. *Black Hat USA 2017*. 2017.

[Reference Source](#)

Raimi RA: **On the Distribution of First Significant Figures**. *Am Math Mon*. 1969; **76**(4): 342–348.

[Publisher Full Text](#)

Sambridge M, Tkalcic H, Jackson A: **Benford's law in the natural sciences**. *Geophys Res Lett*. 2010; **37**(22).

[Publisher Full Text](#)

Slepko AD, Ironside KB, DeBattista D: **Benford's Law: Textbook Exercises and Multiple-choice Testbanks**. *PLoS One*. 2015; **10**(2): e0117972.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Van Rossum G, Drake FL: **Python 3 Reference Manual**. Scotts Valley, CA: CreateSpace. 2009.

[Reference Source](#)

Winter C, Schneider M, Yanniks Y: **Model-Based Digit Analysis for Fraud Detection overcomes Limitations of Benford's Analysis**. *Proceedings of the Seventh International Conference on Availability, Reliability, and Security*. 2012.

[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:  

Version 1

Reviewer Report 22 June 2022

<https://doi.org/10.21956/emeraldopenres.15466.r28063>

© 2022 Aris N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Nooraslinda Abdul Aris 

Faculty of Accountancy, Universiti Teknologi MARA, Shah Alam, Malaysia

The introduction on Benford's Law was clearly explained in the article. The usage and limitation of Benford highlighted was crucial should future enhancement or even new detection model were to be conducted or introduced. The main issues in the new era lies on data reliability and quality using Big Data. Transforming data and bounded the dataset may result in the need to propose a new model.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Is the argument information presented in such a way that it can be understood by a non-academic audience?

Yes

Does the piece present solutions to actual real world challenges?

Not applicable

Is real-world evidence provided to support any conclusions made?

Yes

Could any solutions being offered be effectively implemented in practice?

Yes

Competing Interests: No competing interests were disclosed.**Reviewer Expertise:** My research interests include sustainability (management, strategy and accounting), financial fraud (using ratios analysis, Benford's law and Beneish model), governance/CSR and small medium enterprises.**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 24 January 2022

<https://doi.org/10.21956/emeraldopenres.15466.r27887>

© 2022 Nigrini M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Mark Nigrini**

Department of Accounting, West Virginia University, Morgantown, WV, United States

The author provides a good description of Benford's Law and the literature that preceded this study. The suggested method was applied to the results of a series of scores related to education. After the transformation the data had a much better conformity to Benford's Law. The author claims that this can distinguish between higher and lower risk subsets. This, of course, assumes that the higher risk is the lowest level of conformity and not that the transformation did not use a high enough power for the transformation.

Is the rationale for developing the new method (or application) clearly explained?

Yes

Is the description of the method technically sound?

Yes

Are sufficient details provided to allow replication of the method development and its use by others?

Yes

If any results are presented, are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions about the method and its performance adequately supported by the findings presented in the article?

Yes

Is the argument information presented in such a way that it can be understood by a non-academic audience?

Yes

Does the piece present solutions to actual real world challenges?

Not applicable

Is real-world evidence provided to support any conclusions made?

Yes

Could any solutions being offered be effectively implemented in practice?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Audit data analytics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
