Benford’s Law in Appraisal
by Mark Pomykacz, MAI, AI-GRS, Chris Olmsted, and Katherine Tantinan

Abstract
Benford’s Law has emerged as a tool for forensic financial work. It has been proven in mathematics and science for the detection of database anomalies, corruption, and accounting fraud. This article shows how this tool also can be applied in appraisal review. Appraisal managers, reviewers, and users can utilize Benford’s Law as an appraisal review technique to locate data error or manipulation. Benford’s Law states that the digits in many types of important databases have predictable frequencies, while erroneous, corrupted, or fraudulent databases do not. Many types of real estate and valuation databases are consistent with Benford’s Law, thereby affording its use in the detection of valuation data anomalies.

In recent decades, Benford’s Law analysis has emerged as a tool for the detection of financial fraud and database corruption. This tool succeeds at detecting error and fraud by using a combination of psychology and mathematical principles. Benford’s Law states that the digits in many types of important databases have predictable frequencies. When used as a forensic tool, Benford’s Law identifies databases that do not have the predictable frequencies. Frequency anomalies suggest the data may contain errors, or may be fraudulent or otherwise corrupted, and call for further investigation. Many types of real estate and valuation databases are consistent with the parameters of Benford’s Law, thereby affording the opportunity for its use to detect appraisal data error, corruption, and fraud. This article explains Benford’s Law and examines how it can be applied to appraisal and appraisal review.

The Psychology

Before getting into the specifics of Benford’s Law, let’s look at how numbers are generally perceived. Most people assume the probability for the numerals 1–9 occurring in a database is the same. Also, when asked to report the number in the middle between one and nine, most traditionally educated individuals will report the number five \( \frac{(1 + 9)}{2} \). These individuals have learned that one plus four equals five and nine minus four equals five. Five is the number you get when you add and subtract the same number—four—from the two endpoint numbers in the database. Most individuals are trained to think in terms of pluses and minuses; in other words, they think arithmetically.

There is another way to think about the relationship between numbers however—namely, logarithmically. Examples of logarithmical perceptions of numbers can be found among some peoples in the world who have not been formally trained to think arithmetically. When these individuals are asked to report the number in the middle between one and nine, they frequently report three not five. These individuals perceive that three is the number you get when you multiply and divide one and nine, respectively, by the same number \( 1 \times 3 = 3 \) and \( 9/3 = 3 \).

One result of this difference in intuitive mathematical behavior is that when traditionally educated people are asked to randomly make up numbers, they do so in different ways than people who think logarithmically about numbers. In fact, when motivated for nefarious purposes or otherwise, traditionally educated people often randomly create numbers within the confines of their arithmetical bias. This bias comes from the belief that there must be an even chance that any given number will occur, i.e., there must be an even chance of producing a one, or a two, or a three, etc. While this is correct for many databases, in many important databases of numbers—such as financial, economic, social, and scientific databases—the numerals are not evenly, or arithmetically, distributed. In databases where the numbers are logarithmically distributed, there is a much greater prevalence of some digits than
others. Consequently, when people are psychologically predisposed to create, predict, and interpret numbers arithmetically, they incorrectly create and predict and interpret numbers that naturally form logarithmic data sets.

Benford’s Law

In 1938, physicist Frank Benford1 observed that for a great number of interesting databases, the numbers are not arithmetically distributed but instead are logarithmically distributed. Benford’s Law tells us that in logarithmic databases, numbers beginning with the numeral 1 (i.e., numbers where the first digit is 1, such as 100) are much more likely than numbers with a first digit of 2. Further, numbers beginning with the digit 2 are much more likely than numbers that begin with the digit 3, and so on, until we find that numbers beginning with the numeral 9 are the least common. Benford’s Law can be expressed in the following formula for the frequency of any first digit in a logarithmic database:

\[
\log_{10} \left(1 + \frac{1}{\text{digit}_1}\right),
\]

where the digit, is a number 1–9. Application of Benford’s Law indicates that the prevalence of first digits in logarithmic databases is as shown in Exhibit 1.

<table>
<thead>
<tr>
<th>Number</th>
<th>Prevalence of Number as First Digit (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30.1</td>
</tr>
<tr>
<td>2</td>
<td>17.6</td>
</tr>
<tr>
<td>3</td>
<td>12.5</td>
</tr>
<tr>
<td>4</td>
<td>9.7</td>
</tr>
<tr>
<td>5</td>
<td>7.9</td>
</tr>
<tr>
<td>6</td>
<td>6.7</td>
</tr>
<tr>
<td>7</td>
<td>5.8</td>
</tr>
<tr>
<td>8</td>
<td>5.1</td>
</tr>
<tr>
<td>9</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Benford’s Law has been proven correct both with mathematical formulae and with empirical evidence.2 To demonstrate the validity of Benford’s Law, consider the two annual expenses at a retail building: common area maintenance (CAM) at $100,000, and property taxes at $90,000. If inflation is 3% per year, then CAM will be in the $100,000s (between $100,000 and $199,999) for 23.5 years before it inflates to over $199,999 and its first digit is no longer a 1. But property taxes will be in the $90,000s (between $90,000 and $99,999) for about 3.5 years before the taxes inflate above $99,999 and its first digit is no longer 9. The first digit of the CAM expense will spend twenty years longer with 1 as a first digit than the expense for property taxes. This observation is consistent with Benford’s Law. Therefore, if at any given moment an appraiser examines a property’s expense statements, the appraiser will find that many more expenses have 1 as a first digit than 9. Under Benford’s Law, the appraiser can expect 30.1% of the expenses will begin with the digit 1, while only 4.6% will begin with the digit 9.

Because Benford’s Law is most often used to examine first digits, it is also known as the First Digit Law. However, the principles behind Benford’s Law have been proven for every type of digit (first digit, second digit, last digit, etc.), and for combinations of digits (first two digits, last two digits, etc.), but the expected frequencies change for each type of digit and for each digit combination.

A look at sale price data also can help illustrate Benford’s Law. For example, most people might erroneously expect that the digits 1–9 are equally likely to be the first digit of the prices in a database of sale prices. An examination of sale prices of apartment buildings in New Jersey, however, shows such first digits do not occur evenly. In fact, the numbers in the database are naturally consistent with Benford’s Law, with many more sale prices having a first digit of 1, and with 9 as the least-common first digit, as can be seen in Exhibit 2.

---

Application of Benford’s Law to Types of Databases

The characteristics and types of real estate databases that comport with Benford’s Law include large databases; databases of similar items; demographic data; assessment rolls; large databases of sale prices, loan amounts, or appraised values; appraisal assumptions; income and expense data; cash flows; databases of values, incomes, and expenses that grow exponentially over time (i.e., compounding and discounting); and databases that range over multiple orders of magnitude.

Not all databases are naturally consistent with Benford’s Law however. Benford’s Law is not applicable to databases that are not logarithmic; databases that represent normal distributions around averages or some other central tendency; databases that do not change in value over time; databases that change linearly over time; databases that do not range over multiple orders of magnitude; databases with minimums and maximums; and databases with assigned numbers or identification numbers (phone numbers, social security numbers, or zip codes). Additionally, while the type of data in a database may be naturally consistent with Benford’s Law, a database may not be adequately large to test for compliance; this is a statistical sample size issue.

A data set consisting of historical weekly average mortgage rates is one example of a database that is naturally inconsistent with Benford’s law. Even over long time frames, this data has central tendencies, as can be seen in Exhibit 3.

3. For example, a database of $10,000, $100,000, $1,000,000 is three orders of magnitude.
Benford’s Law in Real Estate Databases

The following types of real estate data are known to be naturally consistent with Benford’s Law:
- Demographics
- Market area data
- Prices, values, and rents
- Income and expense histories
- Some property performance histories

The following real estate databases are not consistent with Benford’s Law:
- Consumer price indices
- Market value indices and averages
- Real estate market averages
- Mortgage rates, capitalization and discount rates, equity rates
- Dates, parcel identification numbers, addresses
- Many databases of percentages
- Dates, parcel identification numbers, addresses
- Many databases of percentages
- Some real estate databases are consistent with Benford’s Law but are often too small for statistical verification. The individual items listed below will be consistent with Benford’s Law and but can be verified only if included in large databases.
- Individual narrative appraisals
- Individual form appraisals, such as residential/URAR appraisals or HUD rent comparability studies
- An annual income and expense statement
- Individual cost, sales, or income approaches

### Exhibit 3 First Digit Frequency, 1971–2014 Federal Reserve Conventional Mortgage Rates Weekly Averages

<table>
<thead>
<tr>
<th>First Digit</th>
<th>Actual Frequency</th>
<th>Benford’s Law Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.350</td>
<td>0.400</td>
</tr>
<tr>
<td>2</td>
<td>0.300</td>
<td>0.300</td>
</tr>
<tr>
<td>3</td>
<td>0.250</td>
<td>0.200</td>
</tr>
<tr>
<td>4</td>
<td>0.200</td>
<td>0.150</td>
</tr>
<tr>
<td>5</td>
<td>0.150</td>
<td>0.100</td>
</tr>
<tr>
<td>6</td>
<td>0.100</td>
<td>0.050</td>
</tr>
<tr>
<td>7</td>
<td>0.050</td>
<td>0.000</td>
</tr>
<tr>
<td>8</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

\( n=2,239 \quad p=0.274 \)

- Actual Frequency
- Benford’s Law Frequency
Spotting Database Errors, Corruption, or Fraud

Sometimes databases get corrupted accidentally. Computers and humans may do unintended things to databases. Humans sometimes accidentally mishandle the data. For example, numbers may be mistyped (wrong keystroke). Benford’s Law can be used to identify such corrupted data and to find errors in databases. The principles and practice for error identification are simple. First, identify whether the type of database should be consistent with Benford’s Law. Second, test the database. If the database shows inconsistencies when it should be consistent with Benford’s Law, then the data may be corrupted. Follow-up analyses then can be used to identify where the problems lie within the data. For example, since Benford’s Law indicates that the digit 4 should occur as a first digit 9.7% of the time, if the actual frequency is 12%, then this is an indication of possible issues within the database. Once the problem is located, it can be fixed. The follow-up examination could include various types of Benford’s Law analyses as well as other kinds of research, analyses, and algorithms. For example, one common algorithm can search for accidental double keystrokes by looking for two identical digits side by side within a number.

Another common research step would be to simply interview management about the numbers in the noncompliant digit category in search of legitimate reasons for the noncompliance. For example, a Benford’s Law examiner might discover that appraisers legitimately use the digit 5 more frequently as second and third digits in their forecasts and conclusions because appraisers often round forecasts and conclusions to either five or zero in second and third digits.

In addition to identifying errors and legitimate anomalies, Benford’s Law can be used to identify intentional manipulation of data. In the 1970s, Hal Varian4 argued that Benford’s Law could be used to identify fraud in economic and social data. The concept is that if the data has been purposefully altered for wrongful purposes, the data manipulator may have unwittingly created the data according to arithmetical intuitions, when only logarithmic intuitions would match the natural frequencies in the data type.

The fraud detection concept is simple. The data in question is first examined to determine whether that type of data should be consistent with Benford’s Law. If so, then the data is analyzed for consistency with Benford’s Law. If the frequency of digits in a database is consistent, the database is called natural, and fraud or corruption cannot be inferred. If the digit frequency is not consistent with Benford’s Law, it is called unnatural, and the data may be corrupt or fraudulent. When the data is not consistent, additional research and analysis are completed to reveal the cause. If the cause is not accidental, then it is fraud.

In his 1999 article, accountant Mark Nigrini uses the case of State of Arizona v. Wayne James Nelson (1993, CV92-18841) to illustrate how Benford’s Law can draw attention to fraudulent numbers.5 In that case, the accused was found guilty of trying to defraud the state based, in part, on the testimony of an expert employing Benford’s Law analysis. The defendant embezzler, Nelson, had created bogus transaction entries that used digit patterns and digits combinations that fell outside the actual expected frequencies. Nigrini reports that Nelson used the numbers 7, 8, or 9 as a first digit in over 90% of the fraudulent transactions.6

---

6. No other court cases upholding or rejecting the use of Benford’s Law have been identified in the authors’ research, and there appear to be no court rulings concerning the use of Benford’s Law in appraisal.
Benford’s Law in Appraisal

An individual appraised value is not subject to Benford’s Law, but databases of appraisals, parts of appraisals, and sets of appraised values may be. Generally, a database needs to have more than several hundred data points before the database can be reliably statistically analyzed for consistency. Therefore, a database of appraised values from a large portfolio of loans could be analyzed. Also, tax assessment roles could be analyzed. The first digits of a portfolio of appraised values from an appraisal company could be analyzed, unless, for example, that appraisal company only appraises homes in neighborhoods where the homes are homogeneous, and therefore the database does not cover a range of multiple magnitudes of value. Even if the first digit cannot be analyzed, however, this same portfolio of values from the appraisal company could be analyzed for compliance with Benford’s Law as to the second digit or some combination of digits.

In applying Benford’s Law in appraisal, as in other circumstances, there are two considerations: (1) whether the type of database should be consistent with Benford’s Law or not, and (2) whether the database is large enough to be reliably statistically tested or not. In real estate–related databases, some groups of numbers may be a type consistent with Benford’s Law, such as a database of historical income and expenses, or projected income and expense data in a discounted cash flow analysis. Nonetheless, database size is still a consideration to complete a reliable Benford’s Law consistency test. Direct capitalization forecasts are rarely large enough to be tested using Benford’s Law, but some discounted cash flow forecasts will be large enough. Each appraisal’s historical data and forecasts must be examined to determine its applicability.

Interestingly, Benford’s Law analyses will never apply to some digits in appraised values. Given that appraised values are usually rounded off, the last digit(s) will not be naturally consistent with Benford’s Law, as appraisers artificially (but not fraudulently) make the final digits 0 (just as many retail prices customarily end in 99 cents). In a sense, zeroes are the naturally found last digits for legitimate appraised values, but they are inconsistent with Benford’s Law.

Benford’s Law Math and Analysis

The mathematical analysis needed to test for consistency with Benford’s Law is not difficult when using a sophisticated data analysis package such as Stata, RStudio, Microsoft Excel, or similar software. Further, basic interpretation of the results requires only intermediate math competency. However, a high degree of knowledge and experience with real estate and statistics may be required to interpret some advanced analyses. Since real-world data never perfectly matches the Benford’s Law expectation, the issue for the analyst lies in discerning how much variance is material or significant before the analyst can determine whether a data set is or is not consistent with Benford’s Law, and if the data is corrupted or possibly fraudulent. The advanced analysis may require proficiency with several statistical concepts, such as scale invariance, sum invariance, and transformation, Chi-squared and t-testing, and with the statistical works of Kolmogorov-Smirnov, Kuiper, Leemis (max (m) statistic), Cho and Gaines (distance (d) statistic), and Nigrini (z-statistics, mean average deviation).

Procedures of the Basic Analysis

The actual procedures for and interpretation of a basic Benford’s Law analysis are remarkably simple. As previously mentioned, the threshold step is to examine the database and confirm that this type of database is consistent with Benford’s Law. For example, imagine a database of building sale prices with various details about the buildings. For the Benford’s Law analysis, the analyst may focus only on the sale prices; no other data from the database is needed in the
Benford's Law test. Let’s assume there are 2,000 records in the database. This type of database is known to be consistent with Benford's Law and 2,000 records can be statistically sufficient.

The second step is to extract the digit to be analyzed, for example, the first digit. If the database ranges from millions of dollars to tens of millions of dollars (from $1,000,000 to $99,999,999), then only the first digit (the digit reporting the millionth or ten millionth place) is required. Note that in those sales over $9,999,999, the second digit, denoting the millions, is not the target of a first-digit analysis.

Next, software can be used to count how many of the digits in the millionth or ten millionth place are 1s, 2s, 3s, and so on through 9. Then, the analyst computes the percentage each digit represents of the total count for the first digit. This is the actual frequency of each number in the first digit in the database. If the database has 600 records where the sale price is a number between $1,000,000 and $1,999,999 or between $10,000,000 and $19,999,999, then the frequency of 1 as the first digit is 30% (600/2,000). The analyst can then compare the results for the database’s actual digit frequencies to Benford's Law predicted frequencies and draw conclusions about compliance. (Benford's Law frequencies for first digits are shown in Exhibit 1.) In this example, since the actual frequency in the database for the digit 1 is 30% and Benford's Law says it should be 30.1%, there is a very close correlation between the expected and actual frequencies, and the database is consistent with Benford's Law for the first digit 1. If the other first digits correlate similarly well, then the analyst would conclude that the database is consistent with Benford's Law and there is no indication of errors, corruption, or fraud.

If the data is inconsistent with Benford's Law, then the analyst must conduct further investigations to determine the reasons. As previously discussed, the possible reasons can be (1) this database is not the type or size needed for Benford's Law analysis, (2) there are unintentional data errors or corruptions, or (3) there is fraud. Additional next steps include various interpretive research and analyses.

The final step for the analyst is to report the findings. Given that the analysis and findings are essentially the comparison of the frequencies between the Benford's Law predictions and the actual percentages, the reporting of the findings may be a table or graph comparing the two sets of percentages and/or frequency counts.

Like many analytical tools, a test using Benford's Law can yield false positive results and false negative results. For example, false positives may occur when a data set under analysis is too small or when a data set is thought to naturally comply with Benford's Law when in fact it does not. False negatives are possible when there is relatively little erroneous or fraudulent data in the database. By itself, Benford's Law is not a definitive test. Like appraisal practice generally, additional analyses, analyst judgment, and reconciliation of all indications are required to support a final opinion.

Benford's Law in Real Estate—Case Study Example

The following case study illustrates how Benford's Law can be applied to real estate data and how data manipulation will be revealed in an analysis. In the case study example, the database is comprised of 817 sale prices of Class A and B office buildings in New York City from 2000 to 2011. This data set meets the conditions that indicate it should comply with Benford's Law (i.e., a large database of similar items, sale prices that range over multiple orders of magnitude). An analysis of the data indicates that the digit frequency does in fact conform with the expected distributions under Benford's Law. Exhibit 4 shows the relative frequency of each numeral from 1 to 9 as the first digit of the sale prices in the data set, alongside the expected numeral frequency under Benford's Law.

7. A false positive is a test result indicating the data does not conform with the expectations of Benford's Law when in fact the data has not been corrupted or manipulated in some way. A false negative is a test result indicating the data does conform with the expected digit frequencies under Benford's Law when in fact the data has been corrupted or manipulated.
The graph in Exhibit 4 indicates the data appears to be consistent with Benford’s Law, and this consistency is further proven with the calculated correlation \( p \) of 0.995 (no correlation at all would be 0.0, and a perfect correlation would be 1.0). The necessary level of correlation to satisfy any particular analysis depends on the amount and quality of available data, the expected conformity of the data, and the requirements of the researcher performing the analysis, but typical minimum correlation levels are 0.95 and above, 0.98 and above, or 0.99 and above. In the case of the office building sale prices analyzed in Exhibit 4, the correlation level clearly exceeds any and all of these typical minimum levels.

To demonstrate how anomalous data can be identified, the exact same data set is reexamined (817 sale prices of Class A and B office buildings sold in New York City between 2000 and 2011), only this time the data has been artificially manipulated by simply rounding the sale prices to the nearest $100,000. Exhibit 5 compares the relative frequency of 1 to 9 as the first digit of the rounded sale prices to the expected frequency of those numerals under Benford’s Law.

Examination of the graph in Exhibit 5 indicates that the data set now is not consistent with Benford’s Law, and this indication is further supported by the comparatively low calculated correlation of 0.746. Given the expectation that this data set should in fact comply with Benford’s Law, an analyst or appraiser would immediately know that this data set had been manipulated or corrupted in some way, either intentionally or unintentionally, and would therefore be aware of the need to further investigate the reliability and validity of the data.
The Future of Benford’s Law

The awareness of the potential applications for Benford’s Law in financial review, like appraisal review, is accelerating. Currently, some audit software used extensively around the world, such as ACL and IDEA, incorporates Benford’s Law algorithms to find corruption and fraud. Independent and internal auditors, including those of the Internal Revenue Service, are beginning to establish regular programs of financial review that utilize Benford’s Law. It is anticipated that major corporate and government accounting software will adopt Benford’s Law algorithms in the future. Similar usage of Benford’s Law can be expected for major appraisal database creators and users interested in ensuring appraisal and data integrity. In the future, Benford’s Law algorithms, analyses, and audits may be employed by lenders and banks, property tax assessors, major appraisal companies and appraisal management companies, data distributors and researchers (e.g., NAREIT, NAIOP, NAR, NCREIF, IREM, BOMA, CoStar, CoreLogic, Case Shiller), and governmental agencies.

Benford’s Law offers a newly emerging tool for forensic appraisal work and appraisal review. Benford’s Law has been proven in mathematics and science, and it has been accepted in financial audits and used in court for the detection of accounting fraud. Appraisal managers, reviewers, and users now have a whole new appraisal review technique to identify appraisal errors and even fraud.
About the Authors

Mark Pomykacz, MAI, AI-GRS, is the managing partner at Federal Appraisal, LLC. Pomykacz specializes in the appraisal of power plants, refineries, utilities, infrastructure and other specialty properties, and he frequently provides expert testimony on such matters. He leads the litigation and forensic services group at Federal Appraisal and has over 30 years’ experience in business and real estate appraisal and advisory services. Contact: mark@federalappraisal.com

Chris Olmsted is a senior manager at Federal Appraisal, LLC. Olmsted is licensed as a state-certified general real estate appraiser and is pursuing the Appraisal Institute’s MAI designation. Olmsted has been with Federal Appraisal since 2007. He has developed expertise in the valuation of a wide range of commercial property types, including railroads, power plants, oil refineries, and other complex properties, and he provides litigation support and forensic valuation services. Contact: chris@federalappraisal.com

Katherine Tantinan is an appraiser at Federal Appraisal, LLC. Tantinan is licensed as a state-certified general real estate appraiser and is pursuing the Appraisal Institute’s MAI designation. Tantinan joined Federal Appraisal in 2013 after earning a bachelor’s degree in mathematics. Contact: katherine@federalappraisal.com

Additional Reading

Suggested by the Authors


SEE NEXT PAGE FOR ADDITIONAL RESOURCES >
Additional Resources
Suggested by the Y. T. and Louise Lee Lum Library

Institute of Internal Auditors, “Instructions for Using Excel to Apply Benford’s Law”

ISACA Journal
- “Understanding and Applying Benford’s Law”
- “Using Spreadsheets and Benford’s Law to Test Accounting Data”

Nigrini.com—Data Analysis Technology for the Audit Community
- Benford’s Law Software
  http://www.nigrini.com/datas_software.htm
- Benford’s Law—Texts, Presentations, and Downloads
  http://www.nigrini.com/benfordslaw.htm