

## Get M.A.D. with the Numbers!

### Moving Benford's Law from Art to Science

BY DAVID G. BANKS, CFE, CIA

September/October 2000

Until recently, using Benford's Law was as much of an art as a science. Fraud examiners and auditors performed digital frequency analyses (DFA) and subjectively viewed the resulting data.

In my article, "Benford's Law Made Easy," in the Sept./Oct. 1999 issue of *The White Paper*, I described how to use Microsoft Excel's macro functions to extract the initial digits from a data table for analysis. In this article I go a step further and show how to use commonly available spreadsheet software to quantify data from a digital frequency analysis and distill it down to a single meaningful number. A fraud examiner or auditor can use that number to quickly perform time period or unit comparisons of DFA results and compile evidence against suspects.

#### **Digital Frequency Analysis and Benford's Law**

When people are asked the chances that the first digit of any number in a table will be the digit 9, most people readily assume that the odds are one in nine (or 11.1 percent). However, Dr. Frank Benford, a physicist, demonstrated in the 1930s that the odds actually were less than one in 20.

Without the aid of a computer, Benford examined first-digit frequencies of 20 lists covering 20,299 observations of natural numbers in a diverse

group of tables.

He worked only with tables of numbers, which weren't manipulated by a particular numbering scheme and weren't generated by a random number generator. The data in the tables included, among others, street numbers of scientists listed in an edition of American Men of Science, the numbers contained in the articles of one issue of Reader's Digest, and such natural phenomena as the surface areas of lakes and molecular weights.

Benford noted that the frequency of the first digits in any table of unmanipulated data followed a predictable pattern, which now bears his name. He calculated the expected rate of occurrence for the first digit with this logarithmic distribution formula:

$$\text{Probability (X is the first digit)} = \frac{\log_{10}(x+1) - \log_{10}(x)}{\log_{10}(10)}$$

When data is manipulated, as it is in a fraud, the frequency of appearance of the initial digits usually differs from Benford's predicted frequency, which makes his law a potentially powerful tool for fraud detection. Using Benford's formula, these are the probabilities of a number appearing as the initial digit:

1. 30.1 percent
2. 17.61 percent
3. 12.49 percent
4. 9.69 percent
5. 7.92 percent
6. 6.69 percent
7. 5.80 percent
8. 5.12 percent
9. 4.58 percent

Using this information and commonly available spreadsheet software such as Microsoft Excel or Lotus 123, we can calculate and graph the anticipated frequency of occurrence for any population. For example, suppose we have the following small population of invoice amounts from a fictitious firm called Hankypanky Co.

**Invoice Amount (\$)**

916.00

818.00

778.00

615.00

505.00

500.00

440.00

415.00

334.00

331.00

330.00

282.00

258.00

249.00

238.00

234.00

222.00

212.00

212.00

202.00

144.00

112.00

15.00

12.00

Extracting the initial digits from this table and summarizing brings these results:

### Digit Actual Frequency of Occurrence

1.	4
2.	9
3.	3
4.	2
5.	2
6.	2
7.	1
8.	1
9.	1

There are 25 items in the population. Using that figure “25” and Benford’s law, we calculate what the frequency of occurrence *should* be for the above table. Using the theoretical percentages derived from Benford’s formula and allowing for rounding, we arrive at the following:

### Digit Theoretical Frequency of Occurrence

1.	$25 * .3010 = 7.5250$
2.	$25 * .1761 = 4.4025$
3.	$25 * .1249 = 3.1225$
4.	$25 * .0969 = 2.4225$
5.	$25 * .0792 = 1.9800$
6.	$25 * .0669 = 1.6725$
7.	$25 * .0580 = 1.4500$
8.	$25 * .0512 = 1.2800$
9.	$25 * .0458 = 1.1450$

To visualize the two tables, let’s combine them (*Exhibit 1*) and graph the results, using, in this instance, Excel’s Chart Wizard. Looking at the line graph, it’s obvious that there’s a difference between the curve that Benford’s Law predicted and the curve defined by the actual data. The differences could signify irregularities but how does a fraud examiner attach a number to those differences in curves?

There are numerous ways to calculate and express measurement between the “two curves” or “closeness of fit” as statisticians refer to it. I’ve tried several with varying degrees of success but one measure of closeness of fit recently has come to the forefront. Mark J. Nigrini, Ph.D., in his new book, “Digital Analysis Using Benford’s Law” (Global Audit Publications, 2000) suggests the use of Mean Absolute Deviation (MAD) as the best

measure of closeness of fit for DFA. Fortunately, calculating MAD isn't difficult, particularly when using spreadsheet software. The result gives us a number that helps to tell a story about our data. Here's how to do it.

Begin by copying the previous test data into a spreadsheet as in Exhibit 2. To use the spreadsheet to perform your MAD calculations, simply fill in the column that shows the "Actual Number of Initial Digits" with your actual initial digit frequencies. If you correctly filled in the test data, you'll arrive at a MAD of .04395556 as in Exhibit 1.

To see what that figure signifies, let's review two more MADs. In the first MAD, we'll examine a population that conforms exactly to Benford's prediction. In the second MAD, we'll examine a population that strays far from Benford's standard. A look at the data, sample graphs, and resulting MADs will give a strong sense of what the DFA figures represent.

In the conformist sample (Exhibit 3), the actual and predicted frequencies of appearance are almost exactly the same — the graph appears almost as one line not two overlapped lines. You'll also notice that the MAD is quite small — close to zero.

In the wildly nonconformist sample (Exhibit 4), the graph and the MAD calculation show the frequencies to be quite different. Because we use the *absolute* deviation (that is we give positive values to negative deviations) the fact that the area of actual experience above Benford's curve is approximately equal to the area below Benford's curve doesn't impair the summary of the deviation.

If we simply averaged the deviations, the negative deviations would offset the positive deviations. Thus, if the deviations above the curve equaled the deviations below the curve, the mean (but not absolute) average deviation would be zero, which would tell us nothing. By giving positive (absolute) values to all the deviations, we arrive at a positive figure that accurately reflects the difference between the actual experience and Benford's predicted experience.

(Beyond discovering MAD, the Chart Wizard graphs themselves are helpful in finding explanations for variances. In Exhibit 5, for instance, since Hankypanky Co. has many more invoiced amounts beginning with the digit 2 than those beginning with the digit 1, we could examine those invoices with amounts beginning with either of those two digits.)

## **Wrestling with the Output**

When is a MAD abnormal and therefore could indicate fraud? In his book, "Digital Analysis Using Benford's Law," Nigrini provides the following guidelines for measuring conformity using the MAD:

**Close conformity** — 0.000 to 0.004

**Acceptable conformity** — 0.004 to 0.008

**Marginally acceptable conformity** — 0.008 to 0.012

**Nonconformity** — greater than 0.012

(For reasons that can't be explained in detail here, the thresholds of acceptability or conformity may vary with sample size and with the nature of the sample population.)

I've found that any deviation from Benford's Law beyond a MAD of .020 is a red flag and I should scrutinize the population. Also, when I've used DFA in accounts payables audits each vendor's digital frequency profile will remain remarkably consistent from period to period. Thus, the fraud examiner or auditor should analyze not just with simple benchmarking but period-to-period comparisons of activity.

A thorough knowledge of the subject's population is important for interpretation. For example, many services such as property or equipment rental companies submit repetitive amounts on invoices that deviate widely from Benford's predictions but don't indicate fraud. Similarly, I've discovered that invoices from freight companies for shipping company products can be skewed by bills for repetitive trips to the same locations, which could just indicate loyal clusters of customers and not fraud.

DFA can be applied in numerous situations. Companies can examine MADs for the initial digits of transactions for each retail clerk or bank teller. And firms can develop internal MAD standards. Suppose that a general merchandise retailer has several dozen cashiers; a baseline analysis of the cashiers' transactions for a one-month period might produce a MAD of .013. By analyzing each cashier's transactions during the same period, the fraud examiner or auditor may find that the .013 MAD is indicative of a combination of fraudulent and non-fraudulent activity. One cashier may have a .369 MAD and all others may have MADs around .008. Also, DFA is a perfect analysis tool for retail stores, restaurants, and other businesses with high transaction volumes.

Though Benford's Law has been known for decades, today's personal computers and software packages have given fraud examiners and auditors a practical tool to find irregularities in data tables. And now with Mean Absolute Deviation, they can distill that data down to a single number, which can point a finger at strong suspects. Benford's Law research will continue to refine the principle for practical investigations.

**David G. Banks, CFE, CIA**, is the director of internal audit for Weirton Steel Corporation in Weirton, W.V.

*The Association of Certified Fraud Examiners assumes sole copyright of any article published on ACFE.com. ACFE follows a policy of exclusive publication. Permission of the publisher is required before an article can be copied or reproduced. Requests for reprinting an article in any form must be emailed to: [FraudMagazine@ACFE.com](mailto:FraudMagazine@ACFE.com).*