

FRAUD

A PUBLICATION OF THE ASSOCIATION OF CERTIFIED FRAUD EXAMINERS MAGAZINE®

Benford's Law Made Easy

Common Software Eases use of Fraud Search Theorem

BY DAVID G. BANKS, CFE, CIA

September/October 1999

Sam was the ideal employee at the manufacturing plant. He worked overtime, needed minimal supervision, was congenial, and often did favors for his colleagues. But behind the façade, he and two others pocketed more than \$100,000 from a plant fund in a 10-year period. Sam's crime could have been caught much earlier if a fraud examiner had used both a mathematical theorem called Benford's Law and a software program found on most personal computers today.

In the April/May 1994 issue of *The White Paper*, Mark J. Nigrini, Ph.D., presented a novel analytical method of detecting numerical anomalies that could indicate fraudulent activity. Using Benford's Law, the method involves finding a pattern in the frequency of digits in a list of figures, beginning from the far-left digit in a figure. As he wrote in the article, digital frequencies refer to the proportion of numbers that have a 1, 2, | 9 as a first digit, and the proportion of numbers that have a 0, 1, | 9 as a second, third, and so forth digit.

The Failure of Common Sense

What are the chances that the first digit of a multi-digit number in a list or table will be nine? Many may assume that the odds are one in nine. Intuitively, we may believe that numbers in large lists are composed of

random digits, each having an equal chance of appearance. Unfortunately, sometimes common sense fails us.

During the 1930s, a physicist named Dr. Frank Benford examined the frequency of certain numbers appearing as initial digits in lists of "natural" numbers. Benford divided all the information into two categories "natural" and "non-natural" numbers. Natural numbers are those numbers that are not ordered in a particular numbering scheme and are *not* generated from a random number system. For example, most accounts payable files will be populated by dollar values that are natural numbers. On the other hand, Social Security numbers and telephone numbers "non-natural numbers" are designed systematically to convey information that restricts the natural nature of the number.

Without the aid of a computer, Benford examined first-digit frequencies of 20 lists covering 20,299 observations of natural numbers. His lists covered data such as 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 discovered that the distribution of the initial digits in natural numbers is not random but rather follows a predictable pattern, which is now known by his name. Benford derived a formula to predict the appearance of the initial digit in any table of natural numbers. The expected occurrence for the first digit is:

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

Using this formula, we see that there's approximately a 30.1 percent chance that the first digit will be a 1, and a 4.6 percent chance that the first digit will be a 9. This discovery obviously defies the intuitive belief that initial digits are random. For decades, Benford's discovery laid dormant. Without computers, it was uneconomical and cumbersome to apply it to practical problems.

The Key to Detection

Recent research clearly has demonstrated that it's possible to detect fraudulent numbers (or errors in bookkeeping or accounting) by comparing the frequency of the appearance of initial digits in a list of numbers to those anticipated by Benford.

With the aid of computers, pioneers such as Nigrini and Dr. Bruce Busta showed in the 1990s that Benford's Law could be used to detect anomalies in tax and economic data. However, while the computer was capable of

rapidly analyzing data, the cost of software and the time spent in learning to apply it prohibited the majority of fraud examiners and auditors from using Nigrini's and Busta's discoveries.

Enter Excel

Intrigued with Nigrini's 1994 article in *The White Paper*, I began a search for a simple, low-cost, effective software solution that would give most fraud examiners and auditors the ability to perform Benford analyses. While a graduate student at West Virginia University under the sponsorship of Dr. Paul Speaker, I began to experiment with various software packages. My associate, Stephen C. Tedrick, and I discovered that Microsoft Excel was the only available software solution.

By the end of 1996, Tedrick and I had developed an Excel macro that extracted the left-most four digits from numbers in a list and placed them each in a separate column. Using the simplest of Excel's functions, it now is possible to analyze quickly and accurately tables containing more than 50,000 records. (The Office 2000 version of Excel will run up to 65,536 records.)

Years later, I used Benford's Law and Excel to reexamine the fraud involving Sam, the ideal worker at the manufacturing plant. In that case, any employee cash request to pay an invoice greater than \$500 required the review and approval of a manager. Using Excel, I found among employee requests an extraordinarily high incidence of the digit four coupled with an equally low incidence of the digits one, two, and three. I discovered that Sam and two other employees were processing invoices for dummy companies "owned" by him. (One of the companies was named A.F.A.B., which Sam later told me actually was Anything For A Buck.) To avoid management scrutiny and maximize return, the fraudsters kept all the false claims below the \$500 threshold and in the upper \$400 range.

How it's Done

Performing a Benford's Law analysis with Excel is actually a five-step process. First, select a population for analysis. Second, assemble the raw data in a format acceptable to Excel. Third, clean non-numeric leading characters (such as letters or dollar signs) and decimal points. Fourth, extract leading digits and store for analysis. Fifth, execute the final analysis. While these steps can be executed manually in Excel, the task is shortened through the use of Excel macros.

A macro is a program designed to execute a series of commands within a

software package such as Excel. Macros are especially useful for the execution of complex repetitive tasks.

1. Select a Population

Let's examine some of the potential pitfalls. To perform a useful analysis, the data should be natural, and therefore, useful. Remember, Benford's Law only applies to natural numbers. Most experts agree that dates are not natural numbers. Attempts to analyze dates normally are frustrating and often futile. Invoice numbers may or may not be natural but it's unlikely that they will be meaningful for our purposes. (I've never heard of a fraud that involved the manipulation of invoice numbers.) This leaves the fraud examiner or auditor with the dollar amount on invoices or other instruments such as checks, financial reports, and receipts as possible natural numbers suitable for analysis.

(Note: There are constraints on dollar amounts. For instance, in Sam's case, there was a \$500 per transaction limit in the manufacturing plant fund, which restricted the appearance of six through nine as initial digits. When manipulating such funds, there is a tendency to maximize the amounts, thus leading to an abundance of \$49X transactions, which could skew the data. In any case, the auditor or fraud examiner must select a population of data that *should* be completely natural.

2. Assemble Raw Data

Fortunately, beginning with Excel's "FILE OPEN" command, the program leads the user through a series of steps that nearly guarantee the successful use of most common data files. Generally, there should be no problems with Microsoft's Excel and Access file types, and most Lotus, Comma Delimited Files (files that are used to transmit spreadsheet information between brands of spreadsheets), and text files are also readable by Excel with minor coaxing.

3. Cleaning Raw Data

Many data files contain non-numeric characters (such as letters or dollar signs) and decimal points, and spaces that can impede the analysis. Excel contains functions, "=CLEAN" and "=TRIM," that can be used to remove these characters. "=CLEAN" removes extraneous characters from the designated cell and "=TRIM" removes leading spaces. Decimal points are eliminated by multiplying each cell by 100 and then using Excel's "FORMAT" command to format the column with "0" decimal places.

While it's possible to perform each of the above functions manually, using a macro will shorten the process. Enter the macros into one sheet of the spreadsheet. On another sheet of the spreadsheet, enter the data to be tested in the "B" column of the spreadsheet. Place the cursor on the topmost cell of the data and execute the command through the "TOOLS," "MACRO," "MACROS," "RUN" command sequence. The result is a cleaned and sorted data file ready for analysis.

4. Digital Extraction

This step lies at the heart of the analysis. The object is to separate each left-most digit, placing it in a separate column where it then can be analyzed. This is accomplished using Excel's "=LEFT" function. Because Excel will only apply the "=LEFT" function to non-numeric strings, it's necessary to use functions under the "EDIT" command to convert the numeric string to a non-numeric string; after using the "=LEFT" function, the macro converts the non-numeric string back to a numeric string. Again, this macro should be entered in one sheet of the Excel spreadsheet. The cleaned data should be in the "B" column of another spreadsheet. Place the cursor on the topmost cell of the data and execute the "TOOLS," "MACRO," "MACROS," "RUN" command sequence. The result will be four columns, each containing leading digits from the target data column.

5. Execute Final Analysis

The final analysis consists of three steps: data summarization, graphing, and interpretation. While it is possible to use a macro to summarize the data, it's just as easy to sort each of the columns of data produced in the digital extraction using the row numbers to calculate frequencies of appearance. Then produce a line graph using Excel's Chart Wizard. Finally, examine the digital frequencies against those predicted by Benford to reveal any deviations.

One for the Tool Kit

Benford's Law cannot be used to extract specific fraudulent transactions, nor can it differentiate between fraudulently manipulated data and legitimate changes. However, fraud examiners and auditors, using software programs like Microsoft Excel, can put Benford's Law in their detection tool kits to discover major fraud schemes that couldn't be found any other way.

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 e-mailed to: FraudMagazine@ACFE.com.