USING BENFORD ANALYSIS TO DETECT FRAUD

Benford's Law analysis is a mathematical technique for identifying irregular patterns in data, which might represent red flags that indicate fraudulent activity or material errors.

CLYDE T. STAMBAUGH, MANUEL TIPGOS, FLOYD W. CARPENTER, AND L. MURPHY SMITH

A key role of accountants, including internal auditors, is helping their organizations detect and prevent fraud. Numerous methods have been developed to aid in fraud detection and prevention. A fraud-detection strategy based on Benford's Law is a mathematical technique for identifying irregular patterns in data. These irregular patterns might represent red flags that indicate fraudulent activity or material errors. By using the Benford Analysis Workbook that is described in this paper (and available online), accountants can readily perform an effective test to help identify possible fraud- or error-infested accounts.

A renewed interest in fraud detection

Today, all accountants should be fraud-conscious. In response to the large, widely publicized fraud scandals that occurred in recent years, new professional standards require that CPAs assess the risk of fraud on every audit of financial statements. Both internal and external auditors who believe that fraud may exist have a responsibility to investigate further. A growing segment within the profession is that of forensic accountants who spend much of their time investigating likely incidences of fraud. Moreover, IRS personnel frequently test suspicious returns and supporting data for possible tax evasion. Benford's Law can help identify certain instances of fraudulent activity.

About Benford's Law

Benford's Law was presented in 1938 and describes a phenomenon that was first recognized by the American astronomer Simon Newcomb in 1881.[1] Benford's Law postulates that the positional incidence of a specific digit in a large random file of numbers occurs with predictable frequency. For example, the digit 1 should be the leftmost digit approximately 30 percent of the time and the digit 2 should be the leftmost position approximately 17.6 percent of the time. The leftmost position can never be the digit 0, and the pattern followed by the leftmost position of the values in the dataset with regard to the digits 1 through 9 is logarithmic by nature.

DR. CLYDE T. STAMBAUGH, DR. FLOYD W. CARPENTER, *and* DR. L. MURPHY SMITH *are professors in the Department of Accounting at Murray State University in Murray, Kentucky;* DR. MANUEL TIPGOS *is a retired professor of Indiana University Southeast.*

Lists of random numbers that describe the relative size of similar phenomena should generally correspond to Benford's Law. Examples include a wide range of financial data such as sales transactions, inventory records, disbursements, income tax data, and stock exchange data, to mention a few. Conversely, Benford's Law would not apply to listed numbers with arbitrary maximum or minimum values or to assigned numbers such as Social Security numbers, car license plate numbers, or telephone numbers.[2]

Benford's Law has applications to a broad range of different types of business data,[3] and it has been incorporated as an analytical tool into popular audit and fraud software packages such as ACL (Audit Command Language) and IDEA (Interactive Data Extraction and Analysis). Such programs can be technically challenging and expensive; consequently, many accountants are hesitant to employ them. A Benford Analysis Workbook has been designed by the authors that is simple to use and available online.

Using the Benford Analysis Workbook

The Benford Analysis Workbook, developed by the authors, can be accessed on the Internet (http://goo.gl/EnOIV). After the analysis has been run, the workbook will consist of the following three worksheets:

1. The "Test Data" worksheet is used to hold the test dataset and also performs some calculations involving the isolation of digits by their position in the dataset values.
2. The "Benford Analysis" sheet contains:
   - pivot tables that break the data down into subcategories by digit for analysis purposes,
   - each digit's expected occurrence probabilities according to Benford's Law,
   - actual occurrence rates for each digit by group, and
   - Z-statistics for each digit based on differences between the expected probabilities and actual occurrence rates by each group. The Z-statistics are presented in a bold, italic, red font when they are statistically significant at the risk level determined by the auditor's confidence setting. The user auditor sets the confidence level by clicking one of the radio buttons at the very top of the worksheet.
3. The "Benford Charts" worksheet contains graphs. The graphs depict each digit's expected occurrence rate for a specific numeric position in the dataset values under Benford's Law, along with the actual occurrence rates. Initially, the data is displayed in a bar chart; you can switch back and forth between bar charts and line graphs by clicking the appropriate button at the top of the worksheet. The graphs are useful for discerning unusual trends and anomalies that are potentially indicative of fraud.

Analyzing potentially fraudulent data

To analyze potentially fraudulent data using the workbook apply the following steps:

1. Open the workbook.
2. Enable macros. (You must enable macros for the spreadsheet to work.)
3. Click the "Test Data" tab if it is not already selected.
4. Click the "Setup" button to clear the worksheets and to prepare it to receive data.
5. After the Amount column is cleared
6. Open the file that contains the data that you want to analyze.
7. Select the entire column of the data that you want to test. (The data that is to be analyzed must be in a columnar format to use the worksheet.)
8. Use the copy command (Press Ctrl-C or use the "Copy" command from the "Edit" menu) to put a copy of the data into the Windows Clipboard.
9. Minimize the data file.
10. Click the cell with a cyan background (cell A10) in the "Test Data" worksheet of the *BenfordLaw.xlsm* file.

BENFORD'S LAW HAS APPLICATIONS TO A BROAD RANGE OF DIFFERENT TYPES OF BUSINESS DATA.

EXHIBIT 1 Benford's Law Probabilities Digit

	Digit Positions			
Value	First	Second	Third	Fourth
0	-	0.11968	0.10178	0.10018
1	0.30103	0.11389	0.10138	0.10014
2	0.17609	0.10882	0.10097	0.10010
3	0.12494	0.10433	0.10057	0.10006
4	0.09691	0.10031	0.10018	0.10002
5	0.07918	0.09668	0.09979	0.09998
6	0.06695	0.09337	0.09940	0.09994
7	0.05799	0.09035	0.09902	0.09990
8	0.05115	0.08757	0.09864	0.09986
9	0.04576	0.08500	0.09827	0.09982
Totals	1.00000	1.00000	1.00000	1.00000

11. Press "Ctrl-V" or click "Paste" from the "Edit" menu choice.
12. After you have your data in the "Test Data" worksheet, click the "Update" button. It may take a few minutes for the worksheet to process the data.
13. After the data is processed, a message box will appear that tells you to select either the "Benford Analysis" or the "Benford Chart" worksheet to view the results.

Benford Law probabilities

Clicking the "Benford Analysis" tab will reveal several tables that are useful for analyzing your data. As shown in Exhibit 1, the first table shows the Benford Law probabilities. These probabilities include the expected frequency of each digit in the first, second, third, and fourth position starting from the left. The values indicate the proportions that one would expect from a population that meets the Benford Law criteria, assuming that no fraud or systematic errors have occurred.

The expected results are somewhat counterintuitive, which often foils fraudsters' attempts to cover up their unsavory deeds. For example, those unfamiliar with Benford's Law typically believe that the leftmost digit of a set of random numbers has an equal probability of occurring (i.e., a 1 in 9 probability for any digit 1 through 9). Yet the expected probability of the leftmost digit being a "1" according to Benford's Law is 30.103 percent (for "2" is 17.609 percent, for "3" is only 12.494 percent, etc.). The expected occurrence rates for digits 0 through 9 in the second position from the left are closer but still are not equal (i.e., 11.968 percent for a "0" and only 8.5 percent for a "9").

The dataset used in the exhibits

The remaining tables shown as exhibits in this paper are based on an analysis of the same dataset that was provided by Lehman, Watson, and Jones in their article previously published in the *Journal of Accountancy*.[4] That dataset includes 56,000 hypothetical sales returns over a six-month period. It contains data for fraudulent activity that was perpetrated by an employee who processed sales returns.

Analysis of the example dataset

Exhibit 2 shows the expected and actual digit occurrence rate for a sample population, which contains 56,000 values. The expected occurrence rates assume that all value entries have at least four digits. In this sample, some of the data values are small and have only a few digits. That is why the total of the actual column for the fourth digit is less than 56,000.

The third and fourth tables that are generated by the program provide addi-

EXHIBIT 2 Digit Counts

	Digit Positions							
	1		2		3		4	
Value	Expected	Actual	Expected	Actual	Expected	Actual	Expected	Actual
0	-	0	6,702	6,660	5,700	5,679	5,610	18,651
1	16,858	16,786	6,378	6,395	5,677	5,704	5,608	3,601
2	9,861	9,882	6,094	6,095	5,654	5,612	5,606	3,753
3	6,997	6,992	5,842	5,854	5,632	5,658	5,603	3,781
4	5,427	5,487	5,617	5,625	5,610	5,615	5,601	3,674
5	4,434	4,430	5,414	5,410	5,588	5,594	5,599	3,844
6	3,749	3,748	5,229	5,218	5,566	5,572	5,597	3,815
7	3,248	3,241	5,060	5,046	5,545	5,538	5,594	3,644
8	2,865	2,867	4,904	4,938	5,524	5,540	5,592	3,765
9	2,562	2,567	4,760	4,759	5,503	5,488	5,590	3,707
Totals	**56,000**	**56,000**	**56,000**	**56,000**	**56,000**	**56,000**	**56,000**	**52,235**

EXHIBIT 3 Z-Statistics

	Digit Positions			
Value	First	Second	Third	Fourth
0		0.54079	0.28204	**195.53551**
1	0.65573	0.22154	0.36709	**23.74829**
2	0.22621	0.00673	0.58655	**21.50616**
3	0.05201	0.15268	0.35941	**21.07125**
4	0.85047	0.10185	0.06221	**22.60495**
5	0.05712	0.04898	0.07416	**20.09887**
6	0.00880	0.15228	0.07203	**20.49504**
7	0.10937	0.19474	0.09366	**22.96329**
8	0.03757	0.50196	0.22193	**21.17104**
9	0.08252	0.00533	0.20754	**21.99090**

Note: Bold values indicate significant differences at the .05 level.

tional information about the sample population. The third table shows the proportion of the time that each digit occurs in the sample population with regard to the first four digits positions from the left. The fourth table indicates whether the actual digital occurrence rates are greater or less than the expected rates, assuming Benford's Law. (To conserve space these tables are not shown in this article.)

The fifth table produced by the software is shown as Exhibit 3. It shows the Z-statistics associated with the difference between the actual and the expected occurrence rates. When they are detected, significant differences are shown with a gray font and a light gray background. Note that all of the fourth digit entries are significant at the 5 percent level.

The graphs

The "Benford Chart" worksheet shows four bar charts that are created automatically by the program. The user can switch back and forth between the default bar charts and line graphs by clicking the appropriate button near the top row of the worksheet. Each graph displays the expected probabilities for each digit 0 through 9 with the actual data occurrence rates for one of the four leftmost numeric positions. These graphical representations of occurrence rates with Bedford probabilities may visually alert the

EXHIBIT 4 Benford Probabilities and Actual Occurrence Rates for Leftmost Digit

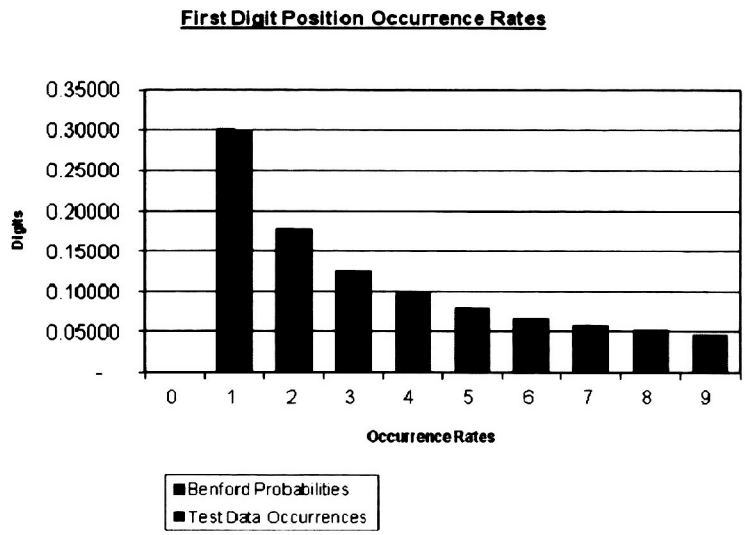
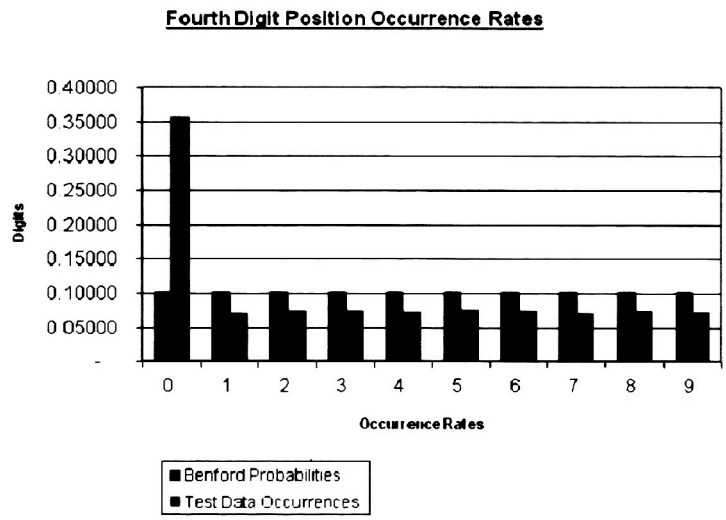


EXHIBIT 5 Benford Probabilities and Actual Occurrence Rates for Fourth Digit from Left



auditor of suspicious and potentially fraudulent numbers in the dataset.

Using the data provided by Lehman, Weidenmier, and Jones,[5] Exhibit 4 shows the expected distribution for the leftmost digit of the dataset's numbers along with the actual occurrence rates. For this digit, the dataset appears to be in compliance with Benford's Law and is logarithmically distributed.

Graphs comparing the actual and expected occurrence rates for the second and third digits from the left are not shown in this paper because no significant differences were detected using Z-tests. However, some differences between sample occurrence rates and the corresponding expected rates under Benford's Law are discernible in the bar chart for the third digit from the left.

Results regarding the fourth digit position were significant in the Z-tests and are shown in Exhibit 5. The expected occurrence rates do not appear to be too different for values 1 through 9. However, in our dataset, the value 0 appears much more frequently than one would expect. Significant deviations in the actual test data occurrence rates from the Benford pattern could signal potential fraudulent activity or could be due to pricing policy, rounding error, or other explainable factors. In any event, the accountant should seek to obtain an understanding as to why this unexpected trend occurred.

Summary and conclusions

Accountants, including internal auditors, external auditors, and forensic accountants, can use the Benford Analysis Workbook that is presented with this paper to help identify possible incidences of fraud or material errors. By comparing the actual occurrence rates of digits with their expected values under Benford's Law, the accountant has a mathematical basis for investigating the risk of material misstatement and for identifying potentially fraudulent activites. ■

NOTES

1 Intuitor.com, Amazing Applications of Probability and Statistics: Benford's Law Part 1: How to Spot Tax Fraud, *http://www.intuitor.com/statistics/Benford's%20Law.html*. Accessed November 11, 2011.

2 Nigrini, M.J. 1996. A Taxpayer Compliance Application of Benford's Law. *Journal of the American Taxation Association* 18(1): 72–91.

3 For example, see Nigrini, M.J. 1999. I've Got Your Number. *Journal of Accountancy*. Online Publications: *http://www.journalofaccountancy.com/Issues/1999/May/nigrini*. Accessed November 11, 2011.

4 Lehman, M., Weidenmier Watson, M., and Jones, T. 2007. Flexing Your Financial Sleuth Power. *Journal of Accountancy. http://www.journalofaccountancy.com/Issues/2007/Jun/FlexingYourSuperFinancialSleuthPower*. Accessed November 11, 2011.

5 *Ibid.*