Using Digital Analysis to Enhance Data Integrity

Michael L. Ettredge *
Rajendra P. Srivastava *

*University of Kansas

June, 1998

The authors would like to acknowledge the assistance of Mikhail Leshchiner in developing the software that is an integral part of this case, and of Mark Nigrini for suggesting the statistical test. The authors also wish to thank Ernst & Young for providing funding through the Ernst & Young Center for Auditing Research and Advanced Technology at the University of Kansas.

Using Digital Analysis to Enhance Data Integrity

ABSTRACT

Digital analysis is a method of analyzing the patterns of digits in a sample of numbers to determine if the sample is similar to a population of numbers. This article introduces digital analysis into the educational literature and provides instructors with materials that can be used to teach the concepts and power of digital analysis as a tool for improving data integrity. The article discusses how digital analysis is being used today and how it might be used in future. It then discusses two digital frequency distributions that can be used as benchmarks to test data validity using digital analysis. Suggestions for class use are provided in a teaching note. Free Excel-based software that performs digital analysis can be downloaded via the internet and distributed to students for use in one of the assignments.

Using Digital Analysis to Enhance Data Integrity

INTRODUCTION

What is digital analysis? Consider the following batch of three sales transaction amounts: $235, $68, and $22. Two-thirds of these numbers begin with the first digit '2' and one-third begins with the first digit '6.' If you were to imagine one thousand different numbers, you might expect that each possible first digit (1, 2, 3, … 7, 8, 9) would occur with equal frequency. Such is not the case. The frequency distribution of first digits differs dramatically from the uniform distribution for many populations of numbers. The frequency distribution of digits exhibited by a batch of numbers can be compared with the frequency distribution for a population of numbers to test whether the batch is likely valid (i.e., similar to the population). In recent years many large businesses and some governmental agencies have begun to use digital analysis to detect unusual or fraudulent transactions.

This article introduces digital analysis into the educational literature and provides instructors with materials that can be used to teach the concepts and power of digital analysis as a tool for improving data integrity. In the next section we discuss how digital analysis is being used today and might be used in future. Then we discuss two frequency distributions that can be used as benchmarks to test data validity using digital analysis. Additional information and suggestions for class use are provided in the teaching note.

USES OF DIGITAL ANALYSIS

The use of digital analysis as an analytical technique is based on the premise that human tampering leaves traces in the digits of the resulting numbers. There is some evidence that supports this premise. Hill (1988) found that the distribution of first digits of numbers invented by subjects differed from the distribution, discussed below, that is exhibited by many naturally occurring sets of numbers. Carslaw (1988) and Thomas (1989) used digital analysis to examine whether earnings numbers and earnings per share numbers, reported by companies, appear to have been manipulated by managers. These studies found evidence suggesting that managers tend to "round earnings up" when earnings are just below a barrier (i.e., earnings just under one million dollars would be manipulated upward to achieve one million). Although these results were interesting, they did not lead to any immediate real-world application.

Recent years have seen a variety of digital analysis applications that were proposed by academics and adopted by business and government entities. Christian and Gupta (1993) used digital analysis to determine that taxpayers appear to bias their taxable incomes downward in certain circumstances. Nigrini (1996) also examined the potential for using digital analysis to detect taxpayers' manipulation of tax return data. A variety of governmental agencies currently use digital analysis for this purpose. Nigrini (1997) proposed that external auditors employ digital analysis to detect employee fraud. Nigrini and Mittermaier (1997) provided examples of the use of digital analysis as an analytical procedure in auditing. Numerous internal and external auditors now use digital analysis to search for fraud, as do some governmental entities. For example, the borough of Brooklyn in New York City uses digital analysis to detect fraudulent checks (Berton, 1995).

Fraud rarely occurs, but digital analysis can also be used to detect errors and other anomalies in data. As described by Searing (1997) digital analysis is being used to detect operational inefficiency and circumvention of internal control. For example, if a manager has authority to sign for disbursements up to $5,000 but desires to spend several times that amount on a project, (s)he might sign numerous checks in amounts just under the authorized limit until the desired total is reached. This would create an unusual pattern of first digits.

Some experts predict that information systems of the future increasingly will include software that identifies unusual events or relationships. Vasarhelyi (1997, p. 48), for example, projects that, "data flowing through the system [will be] monitored and analyzed continuously (i.e., daily) using a set of auditor defined rules. System alarms and reports [will] call the auditor's attention to any deterioration or anomalies in the system." Digital analysis constitutes one of the methods that likely will be used in future to analyze numerical data and to test its validity on a day-to-day basis.

EXPECTED FREQUENCIES OF DIGITS

Digital analysis requires a benchmark frequency distribution of digits against which distributions of samples of numbers can be compared. This section discusses two sources from which a benchmark can be derived. The first is the Benford distribution.

In 1938, Frank Benford, a physicist, discovered that the digits of naturally occurring numbers such as death rates, areas drained by rivers, populations of cities, and many other phenomena are distributed in a predictable non-uniform manner. If one were to examine the leading or first digit of a large set of such data, the number '1' would appear in about 30.1 percent of the cases, '2' would appear in about 17.6 percent, '3' would appear in about 12.5 percent and so on in decreasing fashion. The number '9' would occur in only about 4.6 percent of the cases. This distribution of first digits is known as the Benford distribution, and data exhibiting this distribution are said to conform to Benford's Law.

Benford's Law specifies that the probability that the first digit is 'd_1' is given by:

$$P(\text{first digit is } 'd_1') = \text{Log}_{10}(1 + 1/d_1). \quad (1)$$

Although our discussion has focused on the distribution of first digits for simplicity, it is possible to derive expected distributions of combinations of digits. For example, the probability that the first and second digits will be 'd_1d_2' is given by

$$P(\text{first two digits are } 'd_1d_2') = \text{Log}_{10}(1 + 1/(d_1d_2)). \quad (2)$$

Benford's Law also can be used to derive the distributions for just the second digit, just the third digit, and so on, as well as conditional distributions, such as the probability of the second digit being '2' given that the first digit is '1.' Additional discussion of the mathematics underlying Benford's Law can be found in Hill (1995a, 1995b, 1996) among other sources.

Interestingly, many kinds of accounting data exhibit distributions of digits that are in close agreement with Benford's Law. Consider the future values at different points in time of $1 invested to earn 5% interest as shown in Table 1. Notice that it takes about fourteen periods for the investment to grow from $1 to $2. It only takes about eight additional periods for the investment to grow from $2 to $3. It takes about six more periods to reach $4, and even shorter periods to reach each successive new digit. Clearly, the distribution of first digits is not uniform for this set of data.

- Insert Table 1 Here -

Any process that involves exponential growth, such as that shown in table 1, will follow Benford's Law exactly. Most accounting data sets, however, do not consist of a time-series of numbers exhibiting exponential growth. Why, then, does Benford's Law often describe other accounting data? Boyle (1994) provides a possible explanation. He found that random variables that have been repeatedly multiplied, divided, or raised to integer powers display the Benford distribution. In fact, the Benford distribution is the limiting distribution for such numbers. Perhaps that is why large sets of accounting numbers, derived through various arithmetic operations, often approximate the Benford distribution. For example, we obtained a sample of 3,000 accounts payable invoice amounts from a large U.S. corporation. The distribution of the digits conforms well to Benford's Law as shown in Figure 1. About 33% of the 3,000 numbers have a first digit of '1' versus an expected 30.1% under the Benford distribution. In summary, the

Benford distribution frequently can be used as a benchmark for comparison with batches of accounting data.

- Insert Figure 1 here-

Although many data sets of interest approximate the Benford distribution, there is no theoretical reason why all should do so. Fortunately there is an alternative. If there is a large set of data believed to be valid, frequency distributions of digits can be derived from the valid data set, even if it does not conform to the Benford distribution. These observed distributions can be used as expected distributions for digital analysis. The expected distributions can then be compared with the distributions of a batch of data to determine if the batch resembles the population of valid transactions. Consider, for example, the set of 3,000 transactions described above. If the corporation that provided the data is confident that these transactions are valid, benchmark distributions of digits can be derived from this data set. Rather than using the Benford percentage of 30.1% for a first digit of '1' as a benchmark, the corporation could use the population percentage of 33%. The software included with the case uses this approach.

CONCLUSIONS

In this manuscript, we have provided a background on digital analysis and how it is being used today and might be used in future to enhance data integrity. In the accompanying teaching note, we discuss how digital analysis could be taught in auditing and systems courses. We provide two assignments that can be used to help students understand basic digital analysis. The first assignment uses our software program (available through the Internet and based on Excel) that performs the digital analysis for a small sample of numbers. Although the provided software's tests are based on a specific set of data, alternative versions could be designed to work with a wide variety of data sets. One Big Six firm has developed and currently is using digital analysis software. The second assignment requires students to analyze the first digits of a large sample of numbers that they obtain themselves from the Internet. A third assignment that we suggest, but do not discuss in detail, involves having students write their own software that conducts digital analysis. In the sections to follow, we provide the details on the three assignments.

TEACHING NOTE

Overview

These teaching notes provide several assignments (see the Appendix) that can be used, together with information in the article, to teach students about digital analysis. In the first assignment, the student is asked to enter a series of invented accounts payable invoice amounts into a simulated business database. The Excel-based software then tests the data for validity using digital analysis, and accepts or rejects the student's transactions. The student prints out the results and brings them to class, where the instructor explains how digital analysis works.

When students have a basic understanding of digital analysis, a second activity can be assigned. Students are asked to obtain a large set of real-world numerical data in an electronic file format that can be imported into a spreadsheet. Such data can be downloaded from a number of Internet sites, and we suggest two such sites. Students are asked to analyze their data using spreadsheets and to determine whether the distributions of first digits are approximately Benford. This assignment should reinforce the concepts by convincing students that many real-world data sets exhibit non-uniform distributions of first digits.

A final assignment asks students to create their own software that implements digital analysis. Students who complete all three assignments should have not only an intuitive understanding of how digital analysis works, and a belief that it can be used in many settings, but also will have first-hand experience in translating an abstract concept into a usable piece of software.

Learning Objectives and Course Usage

Learning objectives for the first assignment include:

- Students should understand that the frequency distributions of first digits are non-uniform for many types of data.

- Students should understand how digital analysis can be used to assess data integrity, for internal control purposes and as an external auditing tool.

The first assignment (use of the software) is suitable for almost any auditing course. In that context, it provides students with hands-on experience, and subsequent understanding, of a novel analytical review technique that is coming into use both by internal and external auditors (Searing, 1997). The first assignment does not require any prior knowledge of digital analysis or auditing, and could be used in other courses. For example, the software provides a relatively interesting way to demonstrate internal controls to MBA students. In many MBA programs, students are not exposed to internal control concepts. The first assignment provides an interesting high-tech context in which to discuss internal control issues related to the integrity and reliability of accounting numbers, along with a hands-on experience using the software. The assignment might also be used in accounting information systems courses. The first assignment does not require much time either inside or outside of class (about 30 minutes in each instance).

Learning objectives for the second assignment include:

- Students should be able to download a set of numerical data from an internet Web site.
- Students should be able to import the numerical text file into a spreadsheet.
- Students should be able to use spreadsheet functions to determine the relative frequencies of each of the nine possible first digits for their data sets.
- Students should understand that non-uniform distributions of first digit frequencies are pervasive in real-world data sets.

The second assignment re-enforces the students' beliefs that digital analysis is a valuable technique because it can be used in a variety of settings. A secondary objective is to provide students with a modest software challenge. The student must download some numerical data from the Internet, import it into a spreadsheet, and figure out how to determine the relative frequency of each of the nine possible first digits. This second assignment is appropriate for auditing courses and accounting information systems courses in which the instructor desires students to have a deeper appreciation for the potential usefulness of digital analysis.

Learning objectives for the third assignment include:

- Students should be able to translate an abstract concept, distributions of first digits, into a usable piece of software.

The third assignment (write your own software) is appropriate for some information systems courses. Instructors wishing to use this assignment probably will want to have the students complete the first two assignments as background. Completion of all three assignments will provide students with an understanding of the potential uses of digital analysis in detecting and controlling errors and irregularities, on a continuous basis, as an internal control device in an electronic commerce environment.

The First Assignment

The software provided for use with the first assignment is intended for educational use only. The tests employed are inappropriate for samples other than the data included with the software. The software is an Excel workbook (Office 97 version) that can be obtained in several ways. (1) Using a Web browser such as Netscape, open the following URL location: www.bschool.ukans.edu. Click on the Faculty/Staff button, then on the Faculty FTP Sites button. Find and open the folder labeled "mettredg." Open the folder labeled "Digital Analysis" and select the Digital Analysis file. Save the file to your hard drive. (2) Send an e-mail message requesting the file to mettredge@bschool.wpo.ukans.edu. You will receive the file as an e-mail attachment. (3) Send a blank, formatted, virus-free disk, with stamped return envelope, to either author.

- Insert Figure 2 Here -

The workbook consists of several spreadsheets. Students can see two of the sheets, but others are hidden. The instructor has full access to all sheets by typing a special password obtained from the authors. The current password is "Instructor" (case sensitive). The first sheet (labeled 'Instructions') contains instructions to be read before commencing the exercise. The 'Printout' sheet summarizes the student's results. The student is instructed to print out this sheet and bring it to class. See Figure 2 for an example. The remaining sheets are hidden from students but visible to the instructor. The third sheet is labeled ‘Input Window’ and contains the template for the data input window. The ‘Engine’ sheet contains a test of observed versus expected distributions of first digits. As mentioned earlier, the ‘Payables Data’ sheet contains 3,000 accounts payable invoice amounts obtained from a major U.S. corporation. The instructor can provide these data to students for use in the second assignment if desired.

Insert Figure 3 Here -

On the ‘Engine’ sheet, observed digit frequencies are compared with expected digit frequencies. For example, as Figure 2 indicates, 4 numbers out of the 30 invented by student Jane Doe have the first digit ‘one.’ We would expect 9.85 out of 30 numbers to have the first digit ‘one’ if they were drawn from the valid population of 3,000. The graph in Figure 2 is typical for student-generated data: the frequencies observed are less than expected for low digits (digits one through four) and more than expected for high digits (digits five through nine).

The software contains two tests of student-generated data. The first (Test 1) rejects the student's data if any amount is less than the minimum, or greater than the maximum, in the population of 3,000 valid transactions. The digital analysis test (Test 2) requires some explanation. A chi-square test might seem appropriate to compare observed versus expected frequencies such as those depicted in figures 1 and 2. However, the chi-square test requires a minimum of five observations per cell. This requirement frequently would be violated given the students only are asked to enter 30 observations. Instead, the test is based on the sum of the absolute deviations between observed and expected digital frequencies (see Figure 3). In Jane Doe’s case, the absolute deviation between observed and expected frequency for the digit one is 9.85 - 4= 5.85. The sum of absolute deviations, across all nine first digits, is 18.44. Greater summed absolute deviations imply a worse 'fit' between the student data and valid data.

A bootstrapping technique was used to map the magnitudes of summed absolute deviations into probability measures. One hundred samples of 30 observations each were randomly drawn from the population of 3,000 valid transactions. A sample size of 30 was chosen because this is the number of observations that students are asked to enter. For each sample, the sum of absolute deviations was computed. The resulting distribution of summed absolute deviations for the valid transactions was used to determine thresholds at which known proportions of valid samples of size 30 would be rejected. This information is provided in the top right corner of figure 3. For example, about 10% of random samples drawn from the valid population of 3,000 will have absolute deviations that sum to 16 or more (i.e., probability of type one error is approximately 0.10 at a cutoff level of 16).

We have chosen 16 as the default cutoff level for pedagogical purposes because it provides a good balance between type one and type two errors. It is difficult to predict how the test will perform in rejecting your students' data since these data are idiosyncratic. However, as stated above, student data generally tend to have first digits that are more uniformly distributed than are the first digits of valid data (i.e., too few 'ones' and too many 'nines'). The figure 3 column headed "Approx. Probability of Type 2 Error" is based on 100 random samples of size 30 drawn from a uniform distribution over the interval $1 to $1,000. At a cutoff level of 16, the test will fail to reject 30 numbers drawn from a uniform distribution about 9% of the time. The instructor can modify the cutoff level if desired by entering a desired cutoff level into the indicated cell.

Class Coverage of the First Assignment

The instructor should distribute the first assignment (see Appendix) to the students in advance. Students should *not* be required to read the accompanying article in advance unless the instructor desires to see if students can circumvent the digital analysis test.

On the date the assignment is due, the instructor can begin by asking students how many got a 'reject' verdict as their Test 1 result (see the top row in Figure 2). Most students, however, should have an 'accept.' The instructor then explains that Test 1 only gives a 'reject' verdict if any of the student's amounts were larger or smaller than the valid maximum or minimum for the account (less than $1 or more than $425,784). Test 1 is an example of a test of reasonableness that is based on distributional properties of the numbers themselves, rather than of the numbers' digits.

The instructor next asks how many students received a 'reject' verdict as their Test 2 result (see the top row in Figure 2). The students are likely to be surprised by how many got a 'reject.' The instructor then explains how digital analysis works, using the material in the article.

First, use Table 1 to provide some intuition why frequencies of first digits might not be uniformly distributed for some kinds of business data. Figure 1 is used to introduce the Benford distribution of first digits, and to show how the 3,000 valid transactions conform well to that distribution. The instructor can use the graph in Figure 2 to show how the first digits of a sample of invented numbers ('Observed') can differ from those of valid transactions (‘Expected’). The students are asked to examine the graphs in their own printouts. Those who received an ‘accept’ verdict will have graphs more similar to the 'Expected' distribution in Figure 2 than those who got a ‘reject’.

The instructor proceeds to explain how such software could be used in the ‘real world.’ A ‘reject’ result causes a system alarm to alert either the internal or external auditor. Such tests would be conducted daily, or even more frequently, as data are entered into the system. This provides opportunities to catch errors as they occur on a day-to-day basis. The instructor then wraps up the case by reiterating some of the points mentioned in the article's introduction.

The Second Assignment

The second assignment should be distributed after the first has been completed. In this assignment students are asked to obtain a large electronic file of numerical data from the internet (at least 100 observations), to import the data into a spreadsheet, and to determine the relative frequencies of the nine possible first digits for their samples. The assignment suggests two internet sites where such data can be obtained. However, the instructor should attempt to visit the site in advance because site addresses (URLs) sometimes change. The assignment assumes that students are able to import a file of numerical data into a spreadsheet. If not, the instructor should provide information on how to do this. Finally, the assignment asks students to determine the relative frequencies of the nine possible first digits, and to compare these with the Benford and uniform distributions. The key to accomplishing this step is to use appropriate spreadsheet functions. In Excel, the LEFT() function can be used to return each number's first digit. When a column of first digits has been obtained, the COUNTIF() function can be used to determine the frequency of each of the nine possible first digits. If students have access to Excel's data analysis tools, the Histogram tool can be used to obtain frequencies. Finally, the student divides the frequency for each digit by the total number of observations to obtain a relative frequency.

Class Coverage of the Second Assignment

The second assignment can be covered quickly. Assuming that students obtained data sets from a variety of sources, the instructor could put up an overhead looking something like figure 4. For each possible first digit (1 - 9) inquire who has the lowest proportion, and who has the highest proportion, of numbers beginning with that digit. Write these proportions into the max and min columns, making sure that each is based on a sample of at least 100 total observations. Then determine the approximate median for the class and write that in. The median proportion for each possible first digit can be compared with the uniform distribution (0.11 per digit) and with the Benford distribution (provided in figure 4). The median proportions obtained by your students should approximate the Benford distribution rather than the uniform.

- Insert Figure 4 -

The Third Assignment

We do not provide guidelines for the third assignment (write your own software) because instructors who choose this option are likely to vary widely in their expectations of what the students should accomplish, and over what time period.

USING DIGITAL ANALYSIS TO ENHANCE DATA INTEGRITY

APPENDIX

Assignment 1

In this assignment, you will enter a series of fictitious accounts payable invoice amounts into a simulated business database. The software will test the data that you enter for 'reasonableness' using several criteria, and will accept or reject your batch of transactions. Then you will be asked to print out your results and bring them to class. Your instructor will explain to you at that time how the software examines data for 'reasonableness.'

Ideally, the software should reject your transactions since they are fictitious. However, it is possible that your amounts will be accepted since you are free to enter data that are similar to valid transactions. The valid amounts with which your entries will be compared are a sample of 3,000 accounts payable invoice amounts obtained from a large U.S. corporation. The minimum amount in the sample is $1 and the maximum is $425,784. The mean and median values are $5,161 and $396 respectively. Standard deviation of the amounts is $27,480. Clearly the distribution of the valid amounts is highly skewed. Most transactions are less than $396 but there are a few very large amounts.

The software is an Excel workbook file (Office 97 version). Your instructor will tell you how to obtain this file. The instructions for using the software are given below. Be sure to have these instructions with you when you are ready to use the software. As the workbook opens, you might see a notice that it contains macros. If so, click on the button that says "Enable Macros." Once you have opened the workbook in Excel, insert the cursor in the open ‘Input’ window and type your name carefully, then click on the ‘OK’ button. The following message will appear: “You may start reading the instructions.” Click on the OK button again.

After reading the instructions, click on the button that says, “Click if you’re ready.” An Input window will appear. It contains a date field and a transaction number field that are pre-filled. It also contains an empty amount field in which you will enter your thirty fictitious transaction amounts. A counter field at the upper right corner of the input window helps you keep track of how many amounts you have entered.

To enter an amount, insert the cursor in the amount field and carefully type in the amount. For convenience, use only whole dollars (no cents) and do not use a dollar sign or commas (i.e., type 1000 instead of $1,000). If you make a mistake while typing you can backspace to correct it. When the amount has been typed correctly, click on the "(1) Enter Amount" button. Do not hit the "Enter" key on your keyboard. It is necessary to hit the button so that the software will process the data correctly. The input window disappears briefly as your transaction amount is posted to another location in the workbook where the batch of transactions will be collected. As each transaction is posted, the transaction number and the transaction counter advance. Inventing 30 numbers is tedious, but please continue until the task is complete. The results of the experiment are interesting.

If you are interrupted and must stop before the task is completed, click on the ‘End Session’ button and begin the task again when you are able. When you finish entering your numbers (i.e., the transaction counter reads ‘31’), click on the "(2) Entries Completed" button. This triggers a software macro routine that analyses your transaction amounts. Allow two or three seconds for the analysis, then click on the "(3) End Session" button. The Input window will disappear and you will be looking at the Printout page of the workbook.

This new page should contain your name, the date, your transaction amounts, your two test results (accept or reject), and a graph. Click on the ‘Print’ button located under the test results and above the graph. Bring your printout to class where your instructor will explain the graph and tell you how the software tested your data.

Assignment 2

This assignment requires you to obtain a sample of numbers and to analyze the distributions of first digits. Numerous downloadable files of numbers can be found at the Bureau of Economic Analysis' internet web site. That site is located at www.bea.doc.gov. A second useful site is the Stat-USA database, found at www.stat-usa.gov/BEN/databases.html. You might know alternative sites. Find a file of at least 100 numbers, preferably more, and download it to your computer as a text file. Open a spreadsheet and import the text file into one or more columns.

Next, determine what proportion of your numbers begin with each of the nine possible first digits. In Excel, the LEFT() function can be used to return each number's first digit. For example, if a transaction valued at 345 is in the cell address 'C15', one can type the function '=LEFT(C15, 1)' in Excel at the location where the first digit is to be displayed. This will yield the first digit '3'. This formula can be copied and pasted, creating a column of first digit numbers. Then, the COUNTIF() function can be used to determine the frequency of each of the nine possible first digits. Or, if you have access to Excel's data analysis tools, the Histogram tool can be used to obtain frequencies. Finally, divide the frequency of each digit by the total number of your observations to determine relative frequencies (proportions). How do your proportions compare with a uniform distribution and with Benford's distribution? Fill out table 2 and bring your results to class where they will be compared with those of your classmates.

- Insert Table 2 about Here -

REFERENCES

Berton, L. 1995. He's Got Their Number: Scholar Uses Math to Foil Financial Fraud. *Wall Street Journal* (July 10).

Boyle, J. 1994. An application of Fourrier series to the most significant digit problem. *American Mathematical Monthly* (November): 879-886.

Carslaw, C. 1988. Anomalies in Income Numbers: Evidence of Goal Oriented Behavior. *The Accounting Review,* Vol. 63, No. 2 (April): 321-327.

Christian, C. and S. Gupta. 1993. New evidence on "secondary evasion." *The Journal of the American Taxation Association* (Spring):72-92.

Hill, T. 1988. Random-number guessing and the first digit phenomenon. *Psychological Reports* 62: 967-971.

Hill, T. 1995a. The Significant-Digit Phenomenon. *The Mathematical Monthly* (April): 322-327.

Hill, T. 1995b. Base-invariance Implies Benford's Law. *Proceedings of American Mathematical Society*, Vol. 123: 887-895.

Hill, T. 1996. A statistical derivation of the significant digit law. *Statistical Science* (4): 354-363.

Nigrini, M. 1996. A taxpayer compliance application of Benford's Law. *The Journal of the American Taxation Association* 18 (Spring): 72-92.

Nigrini, M. 1997. Digital Analysis and the Reduction of Auditor Litigation Risk. *Proceedings of the 1996 Deloitte & Touche/University of Kansas Symposium on Auditing Problems*: 69-81.

Nigrini, M. and L. Mittermaier. 1997. The Use of Benford's Law as an Aid in Analytical Procedures. *Auditing: A Journal of Practice & Theory* (Fall): 52-67.

Searing, J. 1997. Discussion of 'Digital Analysis and the Reduction of Auditor Litigation Risk.' *Proceedings of the 1996 Deloitte & Touche/University of Kansas Symposium on Auditing Problems*: 83-86.

Thomas, J. 1989. Unusual Patterns in Reported Earnings. *The Accounting Review.* Vol. 64, No. 4 (October): 773-787.

Vasarhelyi, Miklos. 1997. The CPAS/CCM experiences: Prospects for AI/ES research in Accounting. *Proceedings of the 1996 Deloitte & Touche/University of Kansas Symposium on Auditing Problems*: 45-65.

Table 1
Future Value of $1 Invested at 5%

Period	Future Value	First Digit	Period	Future Value	First Digit
1	$1.05	1	25	$3.39	3
2	$1.10	1	26	$3.56	3
3	$1.16	1	27	$3.73	3
4	$1.22	1	28	$3.92	3
5	$1.28	1	29	$4.12	4
6	$1.34	1	30	$4.32	4
7	$1.41	1	31	$4.54	4
8	$1.48	1	32	$4.76	4
9	$1.55	1	33	$5.00	5
10	$1.63	1	34	$5.25	5
11	$1.71	1	35	$5.52	5
12	$1.80	1	36	$5.79	5
13	$1.89	1	37	$6.08	6
14	$1.98	1	38	$6.39	6
15	$2.08	2	39	$6.70	6
16	$2.18	2	40	$7.04	7
17	$2.29	2	41	$7.39	7
18	$2.41	2	42	$7.76	7
19	$2.53	2	43	$8.15	8
20	$2.65	2	44	$8.56	8
21	$2.79	2	45	$8.99	8
22	$2.93	2	46	$9.43	9
23	$3.07	3	47	$9.91	9
24	$3.23	3	48	$10.40	1

Table 2
Results of Assignment 2

First digit	Uniform Distribution	Benford's Distribution	My Distribution
1	0.11	.301	
2	0.11	.176	
3	0.11	.125	
4	0.11	.097	
5	0.11	.079	
6	0.11	.067	
7	0.11	.058	
8	0.11	.051	
9	0.11	.046	

Figure 1
Distribution of First Digits of 3,000 Accounts Payable Invoice Amounts (Actual) Compared with Benford's Distribution (Expected)

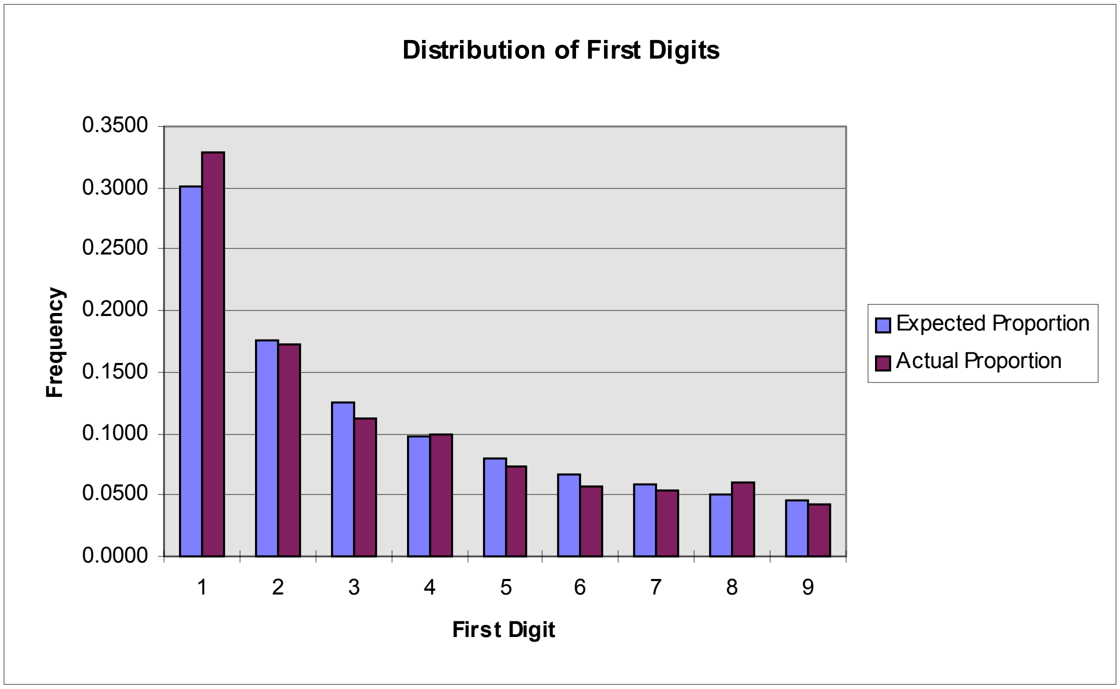

Figure 2
Example Printout Sheet

Student:	Jane Doe		Test 1 Accept	Test 2 Reject

Date	Control N	Amount
9-Jun	85932	704
9-Jun	85933	715
9-Jun	85934	5456
9-Jun	85935	112
9-Jun	85936	59
9-Jun	85937	549
9-Jun	85938	190
9-Jun	85939	43
9-Jun	85940	33
9-Jun	85941	45
9-Jun	85942	504
9-Jun	85943	37
9-Jun	85944	53
9-Jun	85945	15
9-Jun	85946	19
9-Jun	85947	48
9-Jun	85948	28
9-Jun	85949	91
9-Jun	85950	82
9-Jun	85951	24
9-Jun	85952	59
9-Jun	85953	70
9-Jun	85954	26
9-Jun	85955	22
9-Jun	85956	69
9-Jun	85957	59
9-Jun	85958	62
9-Jun	85959	94
9-Jun	85960	51
9-Jun	85961	66

Print

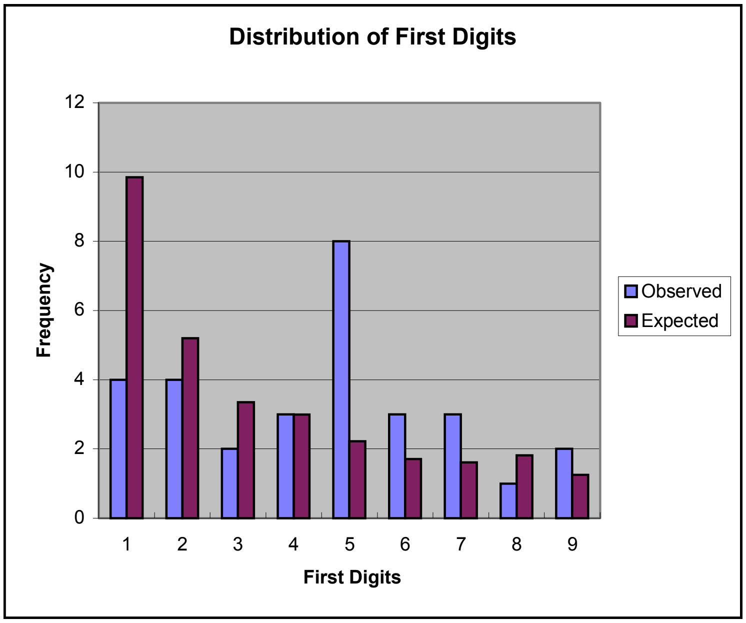

Figure 3
Example Engine Sheet

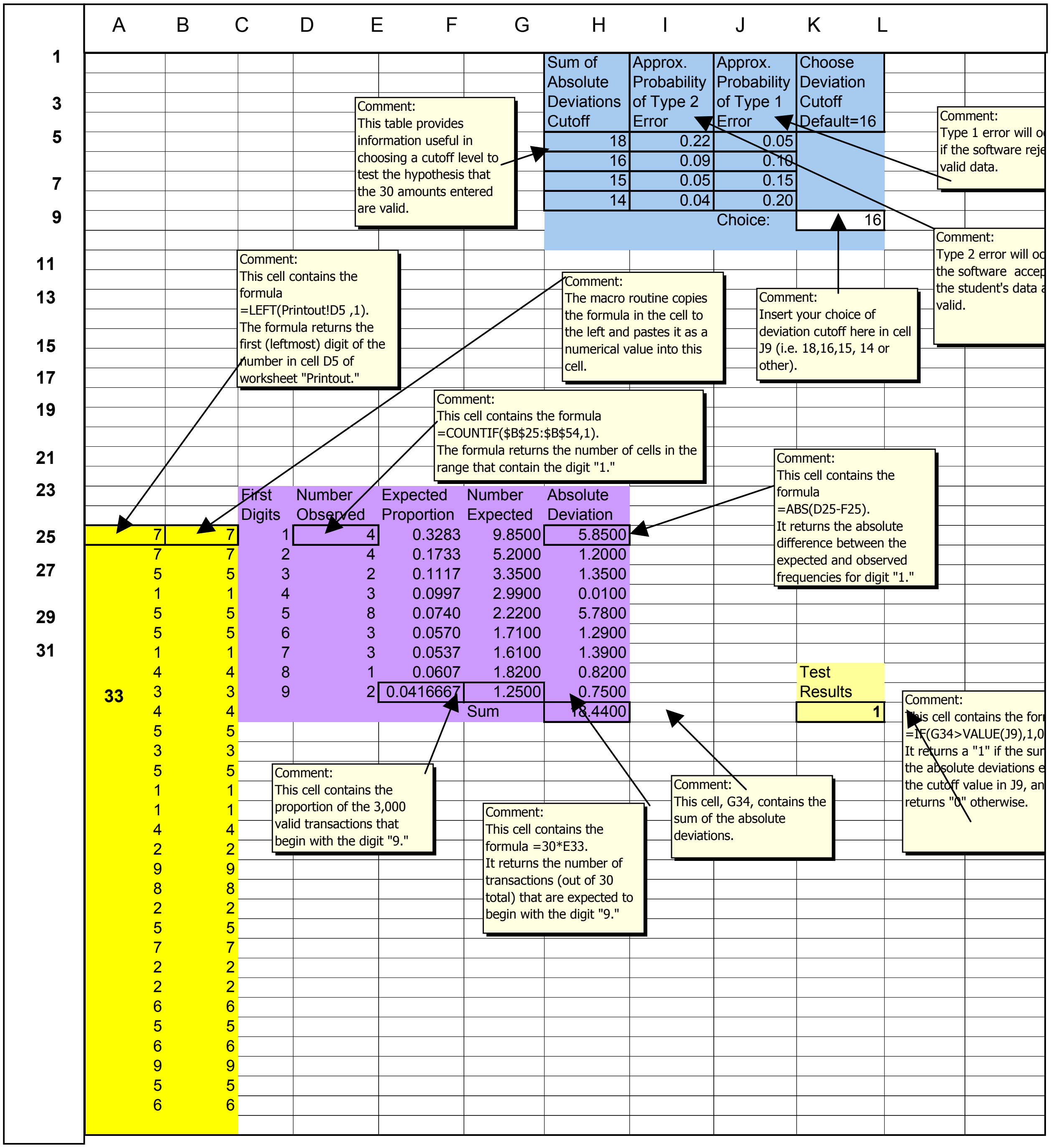

Figure 4
Assignment 2 Overhead

First Digit	Benford's Proportion	Maximum For the Class	Minimum For the Class	Median For the Class
1	.301			
2	.176			
3	.125			
4	.097			
5	.079			
6	.067			
7	.058			
8	.051			
9	.046			