CONTINUOUS AUDITING

by

Mark J. Nigrini
606 Rockcrossing Lane
Allen, TX 75002

Ernst & Young Center for Auditing Research and Advanced Technology
University of Kansas

(972) 359-0020
mark_nigrini@msn.com

August 30, 2000

I thank the participants at a University of Kansas, Ernst & Young Center for Auditing Research and Advanced Technology, accounting workshop for their comments on an earlier version of this paper. The paper was improved as a result of discussions with Ted Hill, Ted Mock, Wayne Shaw, and Raj Srivastava. The paper would not have been possible without empirical data and I thank the Internal Audit Director of the hotel company who provided the data sets for the study.

CONTINUOUS AUDITING

ABSTRACT

The Panel on Audit Effectiveness concluded, amongst other things, that the auditing profession needs to develop new approaches to auditing, including some form of continuous auditing, with a greater emphasis on technology-driven analytical and diagnostic procedures (The Panel on Audit Effectiveness, 2000). A CICA report concluded that continuous audits are viable under certain conditions and that automated "alarm triggers" would be needed to signal anomalies and errors. They called for research to show how auditors could effectively use sophisticated automated audit tools (CICA, 1999).

This paper suggests two digital analysis based tests that might be used by auditors as a technology driven analytical and diagnostic procedure. The first-two digits test checks whether the first-two digit proportions of the current data matches that of past audited data. The summation test analyzes the totals of the numbers with specified first-two digits to test for the inclusion of high dollar values in the current set that do not match those in the past audited data. A suggestion for future research is for goodness-of-fit tests to assist auditors in concluding whether material errors might have occurred.

Keywords	Continuous Audit, Analytical procedures, Digital Analysis, Benford's Law.

Data Availability:	The data used in this study is available from the author upon request.

CONTINUOUS AUDITING

The Panel on Audit Effectiveness reviewed and evaluated how financial statement audits of public companies were performed. Their mandate was to assess whether recent trends in audit practices were in the public interest. The Panel made a number of recommendations to the auditing community. The final chapter of the report, *Looking Ahead*, addressed a number of issues that posed challenges for the future. One such issue was the technology explosion. Here the Panel concluded:

> The challenge for the auditing profession will be to develop new approaches to auditing, including some form of continuous auditing, with greater emphasis on the use of technology-driven analytical and diagnostic procedures (The Panel on Audit Effectiveness, 2000, p.160).

The Canadian Institute of Chartered Accountants (CICA) defines a continuous audit as a methodology that enables independent auditors to provide written assurance on a subject matter using a series of auditor's reports issued simultaneously with, or a short period of time after, the occurrence of events underlying the subject matter (CICA, 1999).

The CICA's conclusions noted that automated "alarm triggers" would be used to provide timely reports on anomalies and errors detected by controls and on possible control failures. They did not define an anomaly. The CICA also concluded that highly automated audit procedures would need to be implemented to provide most of the required audit evidence. They also noted that auditors would need to be quickly informed of the results of automated procedures, particularly when the process has identified anomalies or errors requiring follow-up procedures to be performed by audit personnel.

The CICA noted that research by academics, experimentation by practitioners, and guidance from standard setters were all necessary to help continuous audit services evolve. They called for research to show how auditors could most effectively use sophisticated automated audit tools and techniques that are currently not much used in traditional financial statement audits. They also called for researchers to investigate whether auditors can obtain reliable audit evidence when it may not be practicable to readily access external sources or wait for subsequent events to occur (CICA, 1999).

Continuous Audit is still in its infancy given that the Panel on Audit Effectiveness couched its continuous auditing statement in the *Looking Ahead* chapter with a note about developing new approaches. The CICA report concluded that continuous audits are viable under certain conditions. This conclusion would be unnecessary if such audits already existed. The CICA report listed a number of research avenues to help continuous audit services evolve. Such research would not be needed if the evolution of these services were well underway. Ernst & Young LLP filed for a registered trademark for the words Continuous Audit in February,

1998. The mark was registered with the services simply listed as “business audit services.” At the time of writing the firm has been granted its third extension to file a statement of use statement (see www.uspto.gov and search for Continuous Audit or 75436245). This suggests that while the firm plans to market such services it cannot, as yet, show documentary evidence of using these words to market or provide services to clients.

The CICA report noted that research was needed to show how auditors could most effectively use sophisticated automated audit tools in continuous audit. They considered Continuous Auditing viable if, amongst others, auditors could be quickly informed of the results of automated procedures, when the process has identified anomalies or errors requiring follow-up procedures. SAS No. 56, *Analytical Procedures*, requires auditors to use analytical procedures in planning the nature, timing, and extent of other auditing procedures. It would therefore seem that an appropriate initial step in continuous audit research would be to suggest possible Analytical Procedures that might be applied in a Continuous Audit setting. These Analytical Procedures would be aimed at the detection of errors or anomalies requiring follow-up by auditors.

The objective of this paper is to describe two possible Analytical Procedures that could be used in a Continuous Audit setting. These procedures build on prior research advocating the use of Benford’s Law and Digital Analysis.

The first section of this paper reviews Digital Analysis. The second section presents an empirical review of the suggested analytical procedures using actual corporate payroll data. The concluding section summarizes the paper and reviews avenues for additional research.

DIGITAL ANALYSIS

Digital Analysis is introduced by Nigrini and Mittermaier (1997) and is defined to be the analysis of digit and number patterns to detect abnormal recurrences of digits, digit combinations, and number patterns. When using Digital Analysis, the digit patterns of a data set would usually be compared to Benford’s Law but the authors allude to comparing the digit patterns of a data set to historical norms. The Nigrini and Mittermaier (1997) study reviews a set of six digit-based tests performed on the accounts payable data set of an Oil Company and other corporate data sets.

Benford’s Law

Benford (1938) sets out the expected frequencies of the various digits and digit combinations in tabulated data. These frequencies have since become known as Benford’s Law. Under Benford’s Law the formulas for the digit frequencies are as follows with D_1 representing the first digit, D_2 the second digit, and D_1D_2 the first-two digits of a number:

$$P(D_1=d_1)= \log(1 + (1/d_1)); \qquad d_1 \in \{1, 2, ... ,9\} \qquad (1)$$

$$P(D_2=d_2)= \sum_{d_1=1}^{9} \log(1 + (1/d_1d_2)); \quad d_2 \in \{0, 1, ... , 9\} \qquad (2)$$

$$P(D_1D_2=d_1d_2)= \log(1 + (1/d_1d_2)); \qquad d_1d_2 \in \{10, 11, ... , 99\} \qquad (3)$$

where P indicates the probability of observing the event in parentheses and log refers to the log to the base 10.

The first digit is the leftmost digit in a number and, for example, the first digit of 110,364 is a 1. The first-two digits of 110,364 are 11. For first digits the expected frequencies are highly skewed towards the low digits and range from 30.103 percent for the digit 1 down to 4.576 percent for the digit 9. For the first-two digits the expected frequencies are also highly skewed and range from 4.139 percent for the 10 combination down to 0.436 percent for the 99 combination.

Nigrini and Mittermaier (1997, 55) describe the conditions under which Benford's Law forms a valid set of expected digit frequencies. These conditions are summarized as follows:

1. The list of numbers should describe the sizes of similar phenomena, and
2. The numbers should have no built-in maximums or minimums, and
3. The numbers should not be assigned numbers which are numbers used to describe elements in a data set (such as social security or checking account numbers).

It would appear that condition (3) is superfluous since assigned numbers would be ruled out by the size requirement in (1). Nigrini and Mittermaier (1997) demonstrate six Digital Analysis tests that could be used as analytical procedures by auditors. Their thesis is that accounting data should conform to Benford's Law and nonconformity (deviations from Benford's Law) could signal irregularities in the data.

Benford's Law has also been used in other accounting studies. Carslaw (1988) and Thomas (1989) use Benford's Law in cross sectional empirical studies to detect upward income biasing by corporations. Nigrini (1996) uses Benford's Law to detect income tax evasion in US individual tax returns.

A recent review of the theory underlying Benford's Law is found in Hill (1998) and a comprehensive chronological review of the literature is found in Raimi (1976). The premise of the applied Benford's Law literature is that authentic data should follow Benford's Law and deviations from Benford's Law could signal irregularities. In each case Benford's Law functions as an expected distribution and the deviations calculated are relative to this expected distribution. The statistical literature contains no "alternative" expected digit distribution generally applicable in conditions when Benford's Law is not applicable.

An Application-Specific Law

One reason for a significant deviation from Benford's Law is that the numbers produced by an accounting system are too clustered about the mean and are being affected by minimum or maximum values. This would violate condition (2). For example, the total volume of shares traded on the New York Stock Exchange is about 900 million shares per day. The first digits of the total daily volumes are mostly 8s, 9s, and 1s. While the numbers relate to the size of the activity for the day, the current volume numbers do not follow Benford's Law. It is likely that an analysis of the NYSE daily volumes since its formation in 1792 would follow Benford's Law.

One objective of continuous audit analytical procedures would be to test whether current output has the same data characteristics (look and feel) as prior output. The presumption is that the auditor was satisfied with the control risk at some past point in time and that this satisfaction was based on control and substantive tests. A simple look at descriptive statistics would not by themselves be enough to provide evidence that the numbers making up the totals were accurate.

Nigrini and Mittermaier (1997) review six Digital Analysis tests. These tests are:

1. First digits,
2. Second digits,
3. First-two digits,
4. Number duplication,
5. Rounding, and
6. Last-two digits.

This paper proposes that for continuous audit testing that the current digit and number patterns be compared to prior audited digit and number patterns. Such an approach would require the following steps:

1. Select appropriate Digital Analysis tests,
2. Analyze the digit/number patterns of output that has been audited and been found to contain no material errors.
3. Analyze current output, and
4. Formulate statistical tests to test for differences between (2) and (3).

Test selection is reviewed in the next section. Thereafter the selected tests are applied to hypothetical data sets representing the audited past and current system output.

Select Appropriate Digital Analysis Tests

The first and second digit tests are preliminary tests of reasonableness. A close conformity to Benford's Law signals that from a high-level perspective, the data has passed the reasonableness test. For continuous audit these two tests are probably not appropriate since all the information in the first and second digit tests is contained in the first-two digits test.

The first-two digits test is proposed as a relevant test for continuous audit. The test is a balance between not being too high-level (as with the first and second digit tests) and not too focused (as with a first-three digits test). A match between the results of past audited output and current output could be a useful test for continuous audit.

The number duplication test ranks the individual numbers according to the frequency of occurrence. There is a close relationship between this test and the first-two digits test. For example, a spike (excess) on the first-two digits graph of (say) 50 would be matched with a high count of numbers beginning with 50 (perhaps $50, $500, $5,000, or $505). It is suggested that this test only be used in continuous audit when a spike has been detected, and then simply to match a spike with actual transactions.

The rounded numbers test and last-two digits test are designed to detect invented numbers and number estimation. These tests are not expected to find anomalies or errors in continuous auditing.

Nigrini (1996) introduces a Distortion Factor model that tests whether a data set appears to be understated or overstated. An excess of the lower digits (as compared to Benford's Law) signals a possible understatement and an excess of the higher digits signals the converse. This paper introduces a test, which is a variant of the Distortion Factor model, to test for a small count of abnormally large numbers in a data set. This test (named the summation test) guards against a small count of high-dollar numbers which would go undetected by the first-two digits test.

EMPIRICAL STUDY

Continuous auditing is a practice that needs to evolve. There is therefore no empirical data (continuously audited data) on which to test the suggested analytical procedures.

To demonstrate the proposed tests the author obtained two payroll data sets from a company operating nationally in the hotel industry (Hotel Company). Hotel Company has two main payroll runs. Every second week the employees of the Eastern half of the US are paid and every second week the employees of the Western half are paid. The halves alternate so that in one calendar week the one half of the US is paid and in the next week the other half of the US is paid.

To illustrate Digital Analysis tests in a continuous audit situation the payroll runs for two consecutive weeks (one for each half of the country) are compared. The illustration will assume that the first week represents the audited base line (free of material errors) data and that the second week represents a payroll run subject to continuous audit. In continuous audit it would be more appropriate to compare two consecutive eastern half payroll runs. For the purposes of this paper not much would be accomplished by comparing two nearly-identical data sets.

Data Preparation

Digital Analysis requires the auditor to compare like with like. Consequently positive and negative numbers are not analyzed together because from an error perspective the incentive for errors usually operates in opposite directions for positive and negative numbers. For example, management have an incentive to overstate revenues and to understate credit memos allowed. Where positive and negative numbers appear in the same data set (e.g., invoices and credit memos) it is best to analyze each separately. Also, low-value items (usually under $10) are not usually analyzed using Digital Analysis. These items are immaterial and to avoid their driving any audit conclusion they are usually deleted for Digital Analysis and (perhaps) analyzed separately.

The following payroll amounts were deleted from the data sets prior to the analysis:

1. Negative dollar amounts (corrections for prior overpayments), and
2. Positive dollar amounts less than $10 (including zeros).

The first payroll group with a paydate of February 18 was called Group A and the second payroll group with a paydate of February 25 was called Group B. After the deletions there were 5,823 observations in Group A and 9,829 observations in Group B. The details of the data sets are shown in Table 1.

[Insert Table 1 about here]

For purposes of illustrating the Digital Analysis tests we will assume that Group A is the prior audited data, and Group B is the data that is the subject of a current continuous audit.

First-Two Digits Test

The First-Two Digits test is advocated as a test to determine which first-two digit combinations that had actual proportions significantly above the expected proportions of Benford's Law. These excesses (spikes) could be the results of fraud, errors, misstatements, or processing inefficiencies. Audit samples would be drawn from the number duplication test. Numbers corresponding to first-two digit "spikes" would be among the items targeted for audit. For example, if the first-two digits 54 formed a "spike" and the number $546.66 occurred with a high frequency, then $546.66 would be a potential audit target.

The first-two digits test is well-suited as a data integrity test for continuous audit. If the Group A data has been audited and found to be free of material errors then a close fit between the Group B data and the Group A data could, as an analytical procedure, indicate that the

Group B data is free of material errors. The auditor would be substituting an application-specific digit distribution as the expected distribution in place of Benford's Law.

[Insert Figure 1 about here]

The first-two digit graph of Group A is shown in Figure 1. The bars show the actual first-two digit proportions. The Benford's Law line (set of expected proportions) is included as a reference point. If the actual proportions closely followed Benford's Law, then Benford's Law could be used as the expected distribution. Figure 1 shows that the actual proportions deviate from those of Benford's Law. The actual proportions show a clustering that is quite normal for a company where wage rates are relatively uniform. A clustering of dollar amounts might also be normal for an airline company where pilots are paid similar salaries. The actual proportions of Group A now become the *expected* proportions for the continuous audit.

[Insert Figure 2 about here]

The first-two digit graph of Group B is shown in Figure 2. The bars show the actual proportions and the line shows the Benford's Law proportions. Differences are visually evident. None of the Group B proportions exceed 0.030 whereas two of the Group A proportions exceed 0.030. Group B has a larger proportion at 65 and a smaller proportion at 69. Group B also has a relatively large proportion for 84.

The determination of whether the Group B proportions closely approximate those of Group A is a goodness-of-fit problem. The frequently-used tests for goodness-of-fit are reviewed in the next section.

Measuring Goodness-of-Fit

One goodness-of-fit test is the Chi-square test. Because the Chi-square test uses the expected number of observations in the calculation of the Chi-square statistic it suffers from the *excess power problem* which refers to the situation where, for large data sets, the null hypothesis will be rejected when there are only small differences between the observed and expected proportions for practical purposes. Industries such as banking, airlines, and retailing have large transactional data sets and here even where the eye would not detect any noticeable differences between the two graphs, the Chi-square test for goodness-of-fit would reject the null hypothesis of equal proportions. It therefore seems that any goodness-of-fit test would need to be independent of the data set size to be useful in this setting.

A second alternative is to use Z-statistics to test for a significant difference between the Group A and B proportions. Z-statistics are used in Nigrini (1996, 80) to test the significance of the difference between the actual first and second digit proportions and those of Benford's Law for taxpayer reported interest data. The Z-statistic suffers from two problems in this setting. The calculated statistic uses the sample size (n), and for larger samples, all other things being equal, the calculated Z-statistics will be higher. For large samples the Z-statistic will indicate a significant difference, where in fact the difference is immaterial from a practical perspective. A Z-statistic is calculated for each proportion. For the first-two digit graphs there would be 90 Z-statistics and there is no formula for combining the 90 Zs to draw a conclusion about the data taken as a whole.

A third alternative is to use the Mean Absolute Deviation (MAD). The MAD is calculated by summing the absolute differences and then dividing by 90 (there are 90 first-two digit combinations). The MADs for the first, second, and first-two digit tests are reported in Nigrini and Mittermaier (1997, 59). The MAD suffers from two problems. First, there is no objective MAD cut-off level. An auditor cannot conclude that for the first-two digits, a MAD of (say) 0.0006 is acceptable, whereas a MAD of 0.0186 is unacceptable. All that can be concluded is that the first situation is a closer fit than the second situation. Second, the MAD can hide significant problems. Consider, a situation where 88 of the first-two digits deviations are negligible (near-zero) and two of the deviations are plus 4.5 percent and minus 4.5 percent respectively. The MAD would be calculated to be about 0.1 percent. Now consider a situation where all 90 of the absolute deviations are about 0.1 percent. This would also give a MAD of about 0.1 percent. Both situations would give equal MADs but from an auditing perspective the first situation (specifically the one large positive spike) merits more audit attention than the latter situation.

Given the problems with the test alternatives reviewed above, the conclusion is drawn that a valid goodness-of-fit test should have the following attributes:

1. The test should measure the goodness-of-fit of the entire distribution, and not only a single first-two digit combination,
2. The test should *not* use the sample size (n) since the excess power problem will always lead to the rejection of the null hypothesis for large data sets,
3. The test should be understandable by practitioners and programmable into audit software, and
4. The accept/reject decision should be objectively determinable.

Busta and Weinberg (1998) use a neural network to distinguish between numbers conforming to Benford's Law and "contaminated" (i.e., non-Benford) data. Under a mixture of scenarios (different levels of contamination) and a mixture of data attributes (mean, median, standard deviation, kurtosis, skewness) and a mixture of digit tests (first, second, and first-two) digits, the neural network correctly identified the contaminated data in a high proportion of the cases. They conclude that the neural network showed a greater level of discriminant capability than did traditional statistical methods. The technology sounds appealing but auditors have yet to use neural networks with a high enough frequency to make using networks a real contender as a test.

[Insert Figure 3 about here]

The Kolmogorov-Smirnoff test for goodness-of-fit is based on the cumulative distribution function (c.d.f.). If the Group B proportions closely approximate the Group A proportions then the differences between the two c.d.f.s will be minimal. The c.d.f. of Group A and the c.d.f. of Group B is plotted in Figure 3. In both cases the line ends at 99 with a cumulative proportion of 1.00 for each Group. The test statistic is based on the maximum difference between the two c.d.f.s (D_n) and the square root of the sample size ((5,823 + 9,859) / 2). The cut-off score for 5 percent significance level is 1.36. Rearranging the terms this implies that the maximum difference (D_n) can be no more than,

$$D_n = 1.36 / \sqrt{7,841} \qquad (1)$$

$$= 0.0154$$

If the maximum difference does not exceed 0.0154, then the null hypothesis of an equality of the two distributions cannot be rejected. The maximum difference in Figure 3 is 0.0072 and consequently there is not enough evidence to reject the null hypothesis.

The Kolmogorov-Smirnoff test is also rejected as being appropriate. The test allows too much latitude to be of practical use. A 1.54 percent difference at any point in the c.d.f. would seem be too large for the auditor to ignore from a materiality and processing error likelihood. Also the Kolmogorov-Smirnoff test pertains to the situation where the auditor has drawn a sample from a continuous distribution which in the payroll situation would mean drawing a sample of the actual numbers. The first-two digit test represents the results of a discrete distribution.

The final approach to testing for a similarity between the two graphs is to check the correlation between the proportions of Group A and Group B. If low values in Group A correspond to low values in Group B, and high values in Group A match high values in Group B, then the digit distributions are similar. High correlations will imply similarly shaped first-two digit graphs.

[Insert Figure 4 about here]

A scatterplot of the Group A and B observations is shown in Figure 4. The X-axis shows the Group A proportions and the Y-axis shows the Group B proportions. A regression line of the form,

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_I \quad (2)$$

where,
Y_i is the value of the response variable in the *i*th trial
β_0 and β_1 are parameters
X_i is a known constant, namely, the value of the independent variable in the *i*th trial
ε_i is a random error term with mean $E\{\varepsilon_i\} = 0$ and variance $\sigma^2 = \{\varepsilon_i\} = \sigma^2$, ε_i and ε_j are uncorrelated so that the covariance $\sigma\{e_i, e_j\} = 0$ for all i, j; $i \neq j$
$i = 1, 2, \ldots, n$

could be used. A perfect correlation between Groups A and B would give:

$$\beta_0 = 0 \quad (3)$$

$$\beta_1 = 1 \quad (4)$$

The actual calculated parameter values were:

$$\beta_0 = 0.000983$$

$$\beta_1 = 0.911543$$

The standard errors for β_0 and β_1 are 0.000381 and 0.026608 respectively, producing *t*-statistics of 2.58 and 34.26 respectively.

A test for $\beta_0 = 0$ leads to the conclusion that $\beta_0 \neq 0$ at the 0.05 significance level. Also using the *t*-test, the conclusion is reached that $\beta_1 \neq 1$. Since neither parameter value as tested, equals the expected values for the parameters, the conclusion is reached that the distribution of Group B differs materially from that of Group A.

Summation Test

The first-two digits test will identify abnormal duplications of numbers with first-two digits in the Group B data as compared to the Group A data. In a continuous audit environment auditors would then investigate which actual numbers had a (say) first-two digits of 65 and then test the accuracy of these transactions. However the possibility also exists that the current system generates a few abnormally large observations. The *Wall Street Journal* reported the following:

Kenneth L. Steen of Chattanooga, Tenn., was expecting a $513 tax refund. Instead, he recently received a letter from the IRS informing him that he owes the government $300,000,007.57. “It’s mind boggling,” Mr. Steen says. “I thought that they had become the new, friendlier, more efficient IRS, and then this happens.
Mr. Steen has plenty of company. An IRS official says about 3,000 other people around the nation got similar erroneous notices, each showing a balance due of “three hundred million dollars and change.” A “human programming error” caused the problem, he explains (The Wall Street Journal, 1998).

About 3,000 additional numbers with first-two digits of 30 would not cause a spike on the IRS’ Group B graph. However, the sum of these numbers is approximately equal to one-tenth of the US Gross Domestic Product. The payroll example might (say) include six gross pay amounts of $41,200, this amount being the most frequent number ($412.00) with the decimal point moved two places to the right. The small recurrence would have no effect on the first-two digits graph. To formally test for abnormally large dollar recurrences use of the summation theorem is suggested:

The sum of all the elements in a Benford Set with first digit *x*, for all *x* ($1 \leq x \leq 9$: *x*, integer) are equal. Consequently, the sum of all the elements in a Benford Set with a first digit *x*, is 1/9 of the sum of all the elements (Nigrini, 1992, 71).

A Benford Set is defined to be a set of numbers conforming to Benford’s Law. Nigrini also shows that the result is generalizable to the first-two digits. Allaart (1997) reformulates Nigrini’s statement and proves that the sums are constant in *expectation*. Allaart also shows that Benford’s Law is the unique probability distribution such that the expected sums are equal.

In continuous auditing Benford’s Law is no longer the expected distribution of the digits and consequently the Summation theorem is not valid. However, a comparison of the sums appears to be a relevant test.

[Insert Figure 5 about here]

The Group A proportions are shown in Figure 5. For each two-digit combination the actual dollar amounts were summed and then divided by the total sum ($2,041,955). For example, the graph shows that the sum of all the numbers with a first-two digit combination of 10 was 1.3 percent of the total sum.

[Insert Figure 6 about here]

The Group B proportions are shown in Figure 6. If the Group B data included a small number of large dollar amounts (not found in the Group A data) then this would be evidenced by a larger actual proportion in the Group B graph. A visual inspection shows that the Group B proportions for both 65 and 84 are larger than those for Group A.

A test for a similar distribution between the Group A and B graphs could also use a regression test. The results of the analysis were:

$$\beta_0 = 0.001399$$

$$\beta_1 = 0.874047$$

The standard errors for β_0 and β_1 are 0.000476 and 0.033924 respectively, producing *t*-statistics of 2.94 and 25.76 respectively.

A test for $\beta_0 = 0$ leads to the conclusion that $\beta_0 \neq 0$ at the 0.05 significance level. Also using the *t*-test, the conclusion is reached that $\beta_1 \neq 1$. Since neither parameter value as tested, equals the expected values for the parameters, the conclusion is reached that the proportions of Group B differs materially from those of Group A. The results indicate that there are sums that have increased and in a continuous audit the numbers making up those sums could be the subject of follow-up work.

When the current system under continuous audit generates an additional number of large dollar amounts then both tests would flag the observations. For example, the first-two digit 65 proportion showed a significant increase and so too did the summation test show a significant change for the 65 sum. These were both caused by an excessive duplication of $653.85 in the Group B data.

SUMMARY AND DISCUSSION

The Panel on Audit Effectiveness concluded, amongst other things, that the auditing profession needs to develop new approaches to auditing, including some form of continuous auditing, with a greater emphasis on technology-driven analytical and diagnostic procedures (The Panel on Audit Effectiveness, 2000, p. 160). A CICA report concluded that continuous audits are viable under certain conditions. The CICA concluded that automated "alarm triggers" would be needed to signal anomalies and errors. They noted that the research issues included research to show how auditors could effectively use sophisticated automated audit tools (CICA, 1999). Ernst & Young filed for a registered trademark for Continuous Audit in 1998 but have yet to show that they have used this mark in commerce.

This paper suggests two digital analysis based tests that might be used by auditors as a technology driven analytical and diagnostic procedure. The first-two digits test checks whether the first-two digit proportions of the current data matches that of past audited data. The summation test analyzes the totals of the numbers with specified first-two digits to test for the inclusion of high dollar values in the current set that do not match those in the past audited data.

A suggestion for future research is in the area of goodness-of-fit tests to assist auditors in concluding whether material errors might have occurred. Hypothesis testing for statistically based tests are driven by the concept of *statistical significance*. Audit opinions and audit risk are driven by *material* misstatements. Material information is synonymous with important information. Using statistical significance to drive an opinion based on materiality may not be appropriate. Further research into which goodness-of-fit tests perform well or poorly (based on efficiency, auditing the sample costs, and reliability of the results) on different data characteristics (skewness of distribution, frequency and direction of errors, and size of errors) could be of benefit for continuous audit.

The paper represents an extension to the literature on digital analysis whereby Benford's Law is replaced with an application-specific law. In order to be relevant, no external shocks (such as wage rate changes) should influence the current data. The tests are most relevant to systems processing data where the output consists of a set of calculated numbers (e.g., wage amounts, interest payable amounts, and airline ticket prices). It might be argued that the same results could be obtained by stratifying both the prior data and the current data. Stratification would require the auditor to select an arbitrary number of strata and also a different set of strata boundaries for each data set being continuously audited. The proposed tests are linked to Benford's Law, an accepted theoretical construct. Stratification would also require tests for goodness-of-fit similar to those evaluated in this paper.

REFERENCES

Allaart, P. 1997. An invariant-sum characterization of Benford's Law. *The Journal of Applied Probability* 34: 288-291.

Benford, F. 1938. The law of anomalous numbers. *Proceedings of the American Philosophical Society* 78: 551-572.

Busta, B. and R. Weinberg. 1998. Using Benford's Law and neural networks as a review procedure. *Managerial Auditing Journal* 13 (6): 356-366.

Canadian Institute of Chartered Accountants (CICA). 1999. *Continuous Auditing*. Toronto, ON: CICA.

Carslaw, C. 1988. Anomalies in income numbers: Evidence of goal oriented behavior. *The Accounting Review* (April): 321-327.

Hill, T. 1998. The first digit phenomenon. *American Scientist* 86 (July-August): 358-363.

Nigrini, M. J. 1992. The detection of income tax evasion through an analysis of digital distributions. Ph.D. Dissertation, University of Cincinnati, Cincinnati, Ohio.

Nigrini, M. 1996. A taxpayer compliance application of Benford's Law. *The Journal of the American Taxation Association* (18) 1: 72-91.

Nigrini, M. J., and L. J. Mittermaier. 1997. The use of Benford's Law as an aid in analytical procedures. *Auditing: A Journal of Practice and Theory* 16 (Fall): 52-67.

Raimi, R. 1976. The first digit problem. *American Mathematical Monthly* 83 (Aug-Sept): 521-538.

The Panel on Audit Effectiveness. 2000. *Report and Recommendations. Exposure Draft - May 31, 2000*. Stamford, CT: Public Oversight Board.

The Wall Street Journal. 1998. Tax Report: An IRS blooper startles thousands of taxpayers.

Thomas, J. 1989. Unusual patterns in reported earnings. *The Accounting Review* (October): 773-787.

TABLE 1

PAYROLL DATA: DESCRIPTIVE STATISTICS

Details	Paydate 18th	Paydate 25th
Negative pay amounts		
Count	39	36
Total Dollars	- $ 10,561	- $ 8,993
Zero pay amounts		
Count	24	33
Under $10 pay amounts		
Count	6	19
Total dollars	$ 40	$ 115
Over $10 pay amounts		
Count	5,823	9,859
Total dollars	$ 2,041,955	$ 3,489,694
Maximum Gross pay	$ 4,059	$ 1,808
Gross pay $1,000 or higher	46	72

Note: Table 1 provides descriptive statistics of the two data sets used to illustrate the Digital Analysis tests. The Paydate 18th group is designated Group A and Paydate 25th is designated Group B.

FIGURE 1

GROUP A (PAYDATE 18TH)

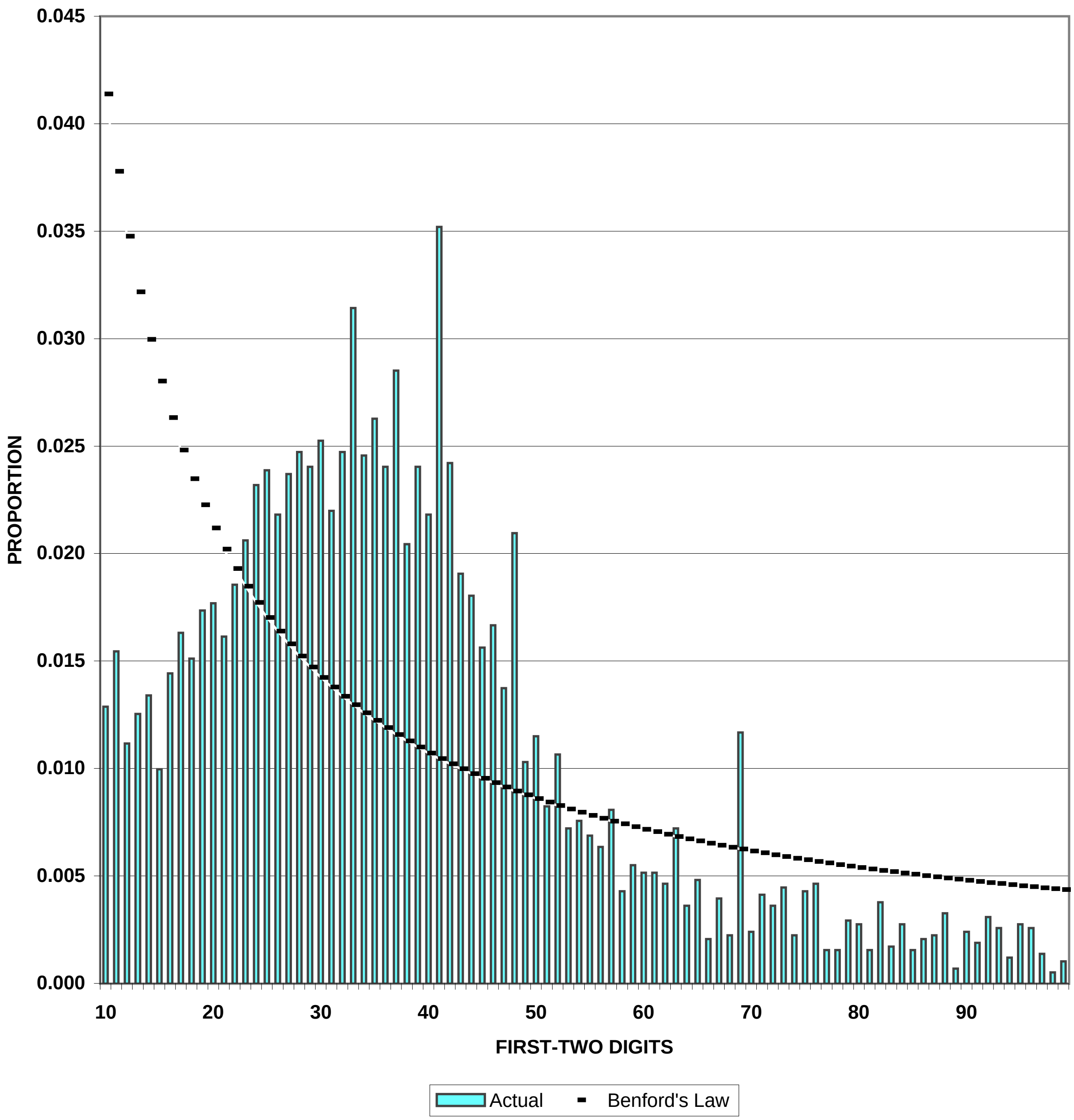

FIGURE 2

GROUP B (PAYDATE 25TH)

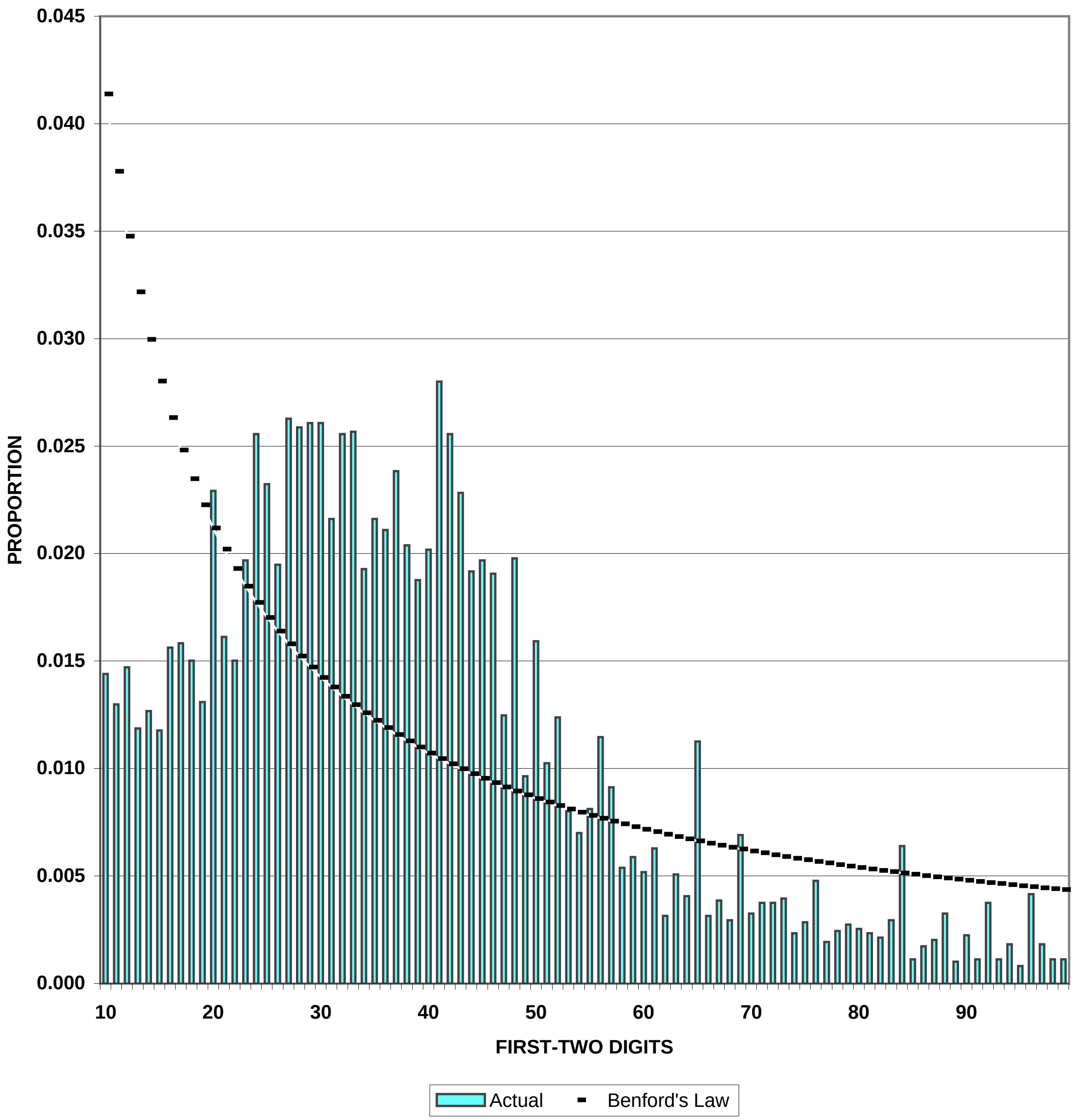

FIGURE 3

CUMULATIVE DENSITY FUNCTIONS OF GROUPS A AND B

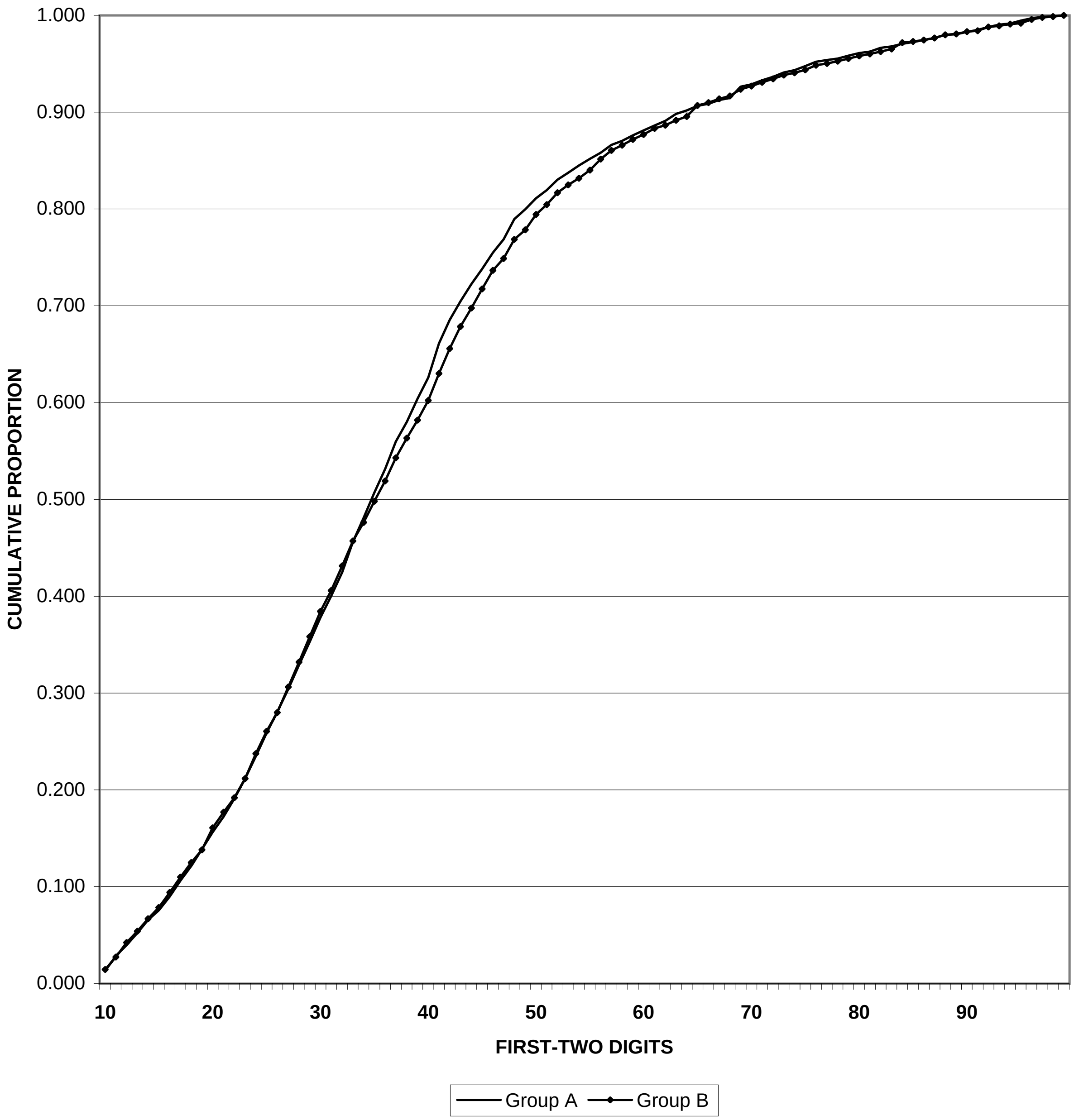

FIGURE 4

PLOT OF GROUP A AND B OBSERVATIONS

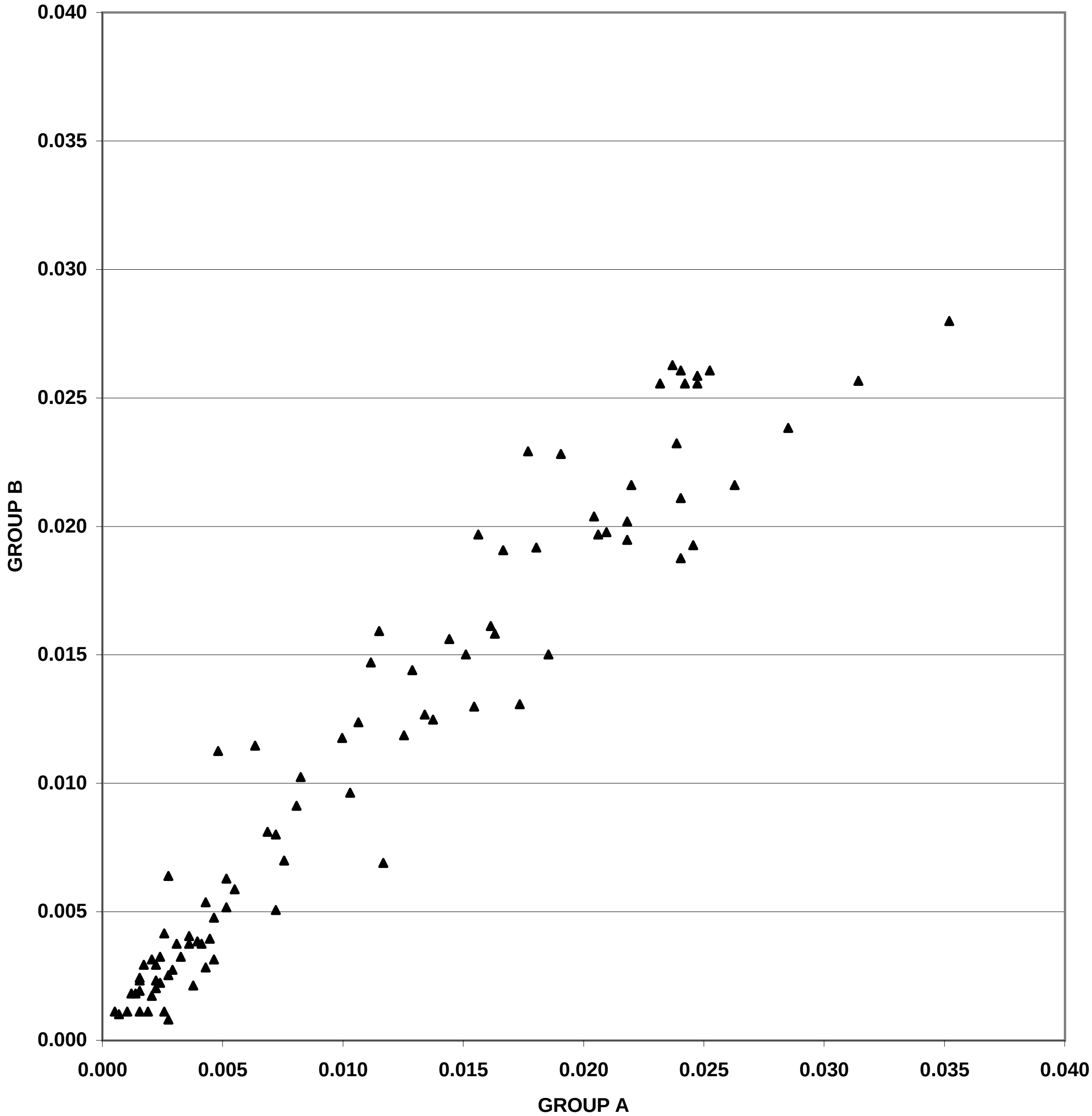

FIGURE 5

SUMMATION TEST: GROUP A

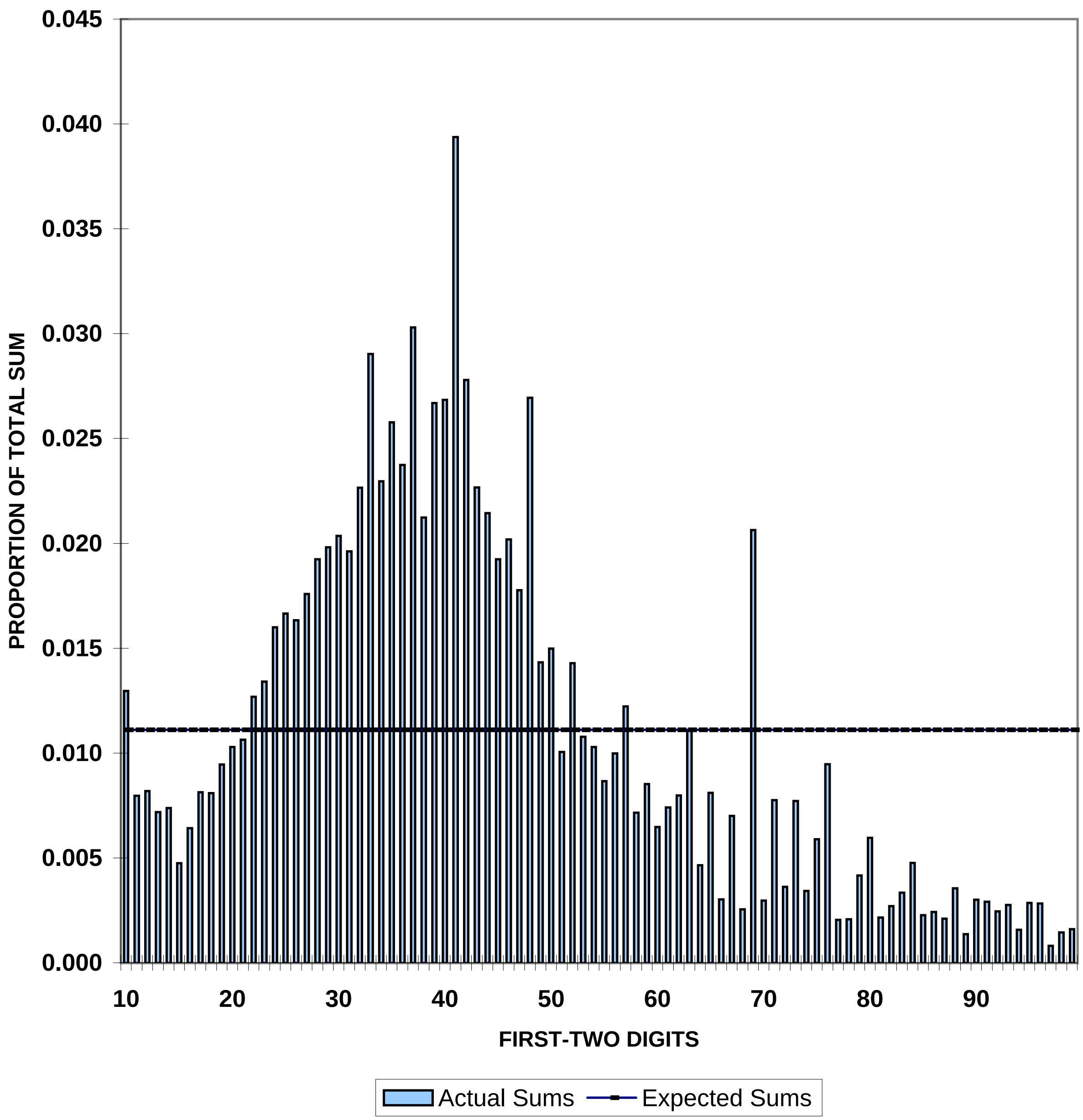

FIGURE 6

SUMMATION TEST: GROUP B