Statistical sampling in auditing

Richard Terry Stegner

The University of Montana

Let us know how access to this document benefits you.
Follow this and additional works at: https://scholarworks.umt.edu/etd

Recommended Citation

This Thesis is brought to you for free and open access by the Graduate School at ScholarWorks at University of Montana. It has been accepted for inclusion in Graduate Student Theses, Dissertations, & Professional Papers by an authorized administrator of ScholarWorks at University of Montana. For more information, please contact scholarworks@mso.umt.edu.
STATISTICAL SAMPLING

IN

AUDITING

by

Richard T. Stegner
B.A., Montana State University, 1947

Presented in partial fulfillment of the requirement for the degree of Master of Arts.

Montana State University

1949

Approved:

Chairman of Board of Examiners

Dean, Graduate School
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. STATISTICAL STANDARDS</td>
<td>1</td>
</tr>
<tr>
<td>Need for standards</td>
<td>1</td>
</tr>
<tr>
<td>Movement for standards</td>
<td>3</td>
</tr>
<tr>
<td>What are standards?</td>
<td>3</td>
</tr>
<tr>
<td>Basis for statistical theory</td>
<td>5</td>
</tr>
<tr>
<td>Statistics in sampling</td>
<td>6</td>
</tr>
<tr>
<td>Statistical goals and benefits</td>
<td>7</td>
</tr>
<tr>
<td>II. STATISTICAL SAMPLING THEORY</td>
<td>10</td>
</tr>
<tr>
<td>Extent of tests</td>
<td>10</td>
</tr>
<tr>
<td>Choice of items tested</td>
<td>25</td>
</tr>
<tr>
<td>Test for accounting control</td>
<td></td>
</tr>
<tr>
<td>Test for standard populations</td>
<td>33</td>
</tr>
<tr>
<td>Procedure of setting up the standard</td>
<td>36</td>
</tr>
<tr>
<td>Likelihood ratio</td>
<td>44</td>
</tr>
<tr>
<td>Sequential sampling</td>
<td>47</td>
</tr>
<tr>
<td>III. APPLICATION OF STATISTICAL SAMPLING THEORY</td>
<td>53</td>
</tr>
<tr>
<td>Present procedure</td>
<td>53</td>
</tr>
<tr>
<td>Critical examination of present audit procedures</td>
<td>55</td>
</tr>
<tr>
<td>Application of the statistical approach</td>
<td>57</td>
</tr>
<tr>
<td>Conclusions</td>
<td>70</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>72</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Probability ratios for ten items</td>
<td>16</td>
</tr>
<tr>
<td>II. Probability ratios with approximations</td>
<td>18</td>
</tr>
<tr>
<td>III. Percentage probability ratios</td>
<td>20</td>
</tr>
<tr>
<td>IV. Economy chart for two false items</td>
<td>23</td>
</tr>
<tr>
<td>V. Economy chart with percentages tested</td>
<td>24</td>
</tr>
<tr>
<td>VI. Unit-month probability ratios</td>
<td>30</td>
</tr>
<tr>
<td>VII. Unit-item probability ratio</td>
<td>30</td>
</tr>
<tr>
<td>VIII. Areas under the normal curve</td>
<td>41</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Normal Curve of Distribution</td>
<td>40</td>
</tr>
<tr>
<td>2. Sequential Sampling Graph</td>
<td>48</td>
</tr>
</tbody>
</table>
CHAPTER I

STATISTICAL STANDARDS

Need for standards. Statistical standards are one phase of auditing technique, yet they are among the most important and least emphasized problems the auditor has to face. The auditor must delve into the books of account of his client, and on the basis of his research into the transactions of the business, he must submit an opinion as to the correctness of his client's financial statements. To certify to their absolute accuracy would necessitate a complete audit of all the transactions of the client as well as a complete search for suppressed transactions. However, the auditor will have many limitations placed upon him such as those of time, money and personal conditions that would render a complete audit of all transactions impossible. He must, then, check a portion of the transactions that he considers the most important. If he fails to note large errors or systematic fraud, he may find himself the object of a law-suit which will injure his reputation irreparably regardless of the legal outcome. The auditor would naturally like to have an unreserved and honest belief that the accounts are correct; that there are no "sins of omission."

Every auditor has at times had the despondent feeling that perhaps fraud may have been committed and errors of a
serious nature may be undiscovered. How then, can the auditor feel secure in issuing his certificate of correctness of the accounts when he can test only a portion of the total items?

It has been stated that, "The character and extent of the tests an auditor will perform will be governed by the circumstances of each case and should be so designed as to satisfy the auditor of the general correctness of the recorded transactions for the period, although such tests will not necessarily disclose every irregularity."

It would seem from the above quotation that the determination of the size of the sample, the selection of the items to be tested, and the drawing of statistical inferences are left solely to the judgement of the individual auditor. That is precisely the case as it stands today; there are few generally accepted basic assumptions as to the size and extent of test checks. The general consensus is that every situation requires its own solution so that no specific rule can be laid down which could apply to all situations.

To a great extent all business transactions are composed of estimates and probabilities. For example, a business man must estimate his stock size and variety according to the

---

1"Tenative Classification of Accountancy Services" from a Special Bulletin of New York State Society of C.P.A.'s. (January 12, 1933)
probabilities of the customers' purchasing his stock at the prices set. The successful practitioners in auditing are those whose weighing of the probabilities of finding fraud or errors in the books of account are most correct and whose judgments are most sound. These features depend largely upon experience, but the amount of experience that can be carried in one's head is limited.

If there were some basic principles or standards that could be applied to evaluate the findings of both the veteran and the beginning auditor, it would prove of immense value. The client would no longer be at the mercy of the auditor's personal ability. The auditor would have a definite program to make his decisions sound, and would be able to withstand an attack if subsequent findings of fraud should invalidate his findings. This would allow the courts to have a basis for judgment of the adequacy of the audit.

Movement for standards. There has been very little information available to the auditor as to the extent of testing items in an audit by statistical methods. There seems to be a recent trend to acquaint the practitioner of a need for auditing standards as can be seen by the fact that the great bulk of information that pertains to statistical standards has been written in the past two years.

What are standards? The definition of an auditing
standard as expressed by the Committee on Auditing Procedure for the American Institute of Accountants is as follows:

"Auditing standards may be regarded as the underlying principles of auditing which control the nature and extent of the evidence to be obtained by auditing procedures."

Actual requirements have been few but the American Institute of Accountants has gone along with the Securities and Exchange Commission in requiring an expression in the accountant's certificate with respect to the use of standard procedures by providing, in its suggested form of certificate, the phrase, "Our examination was made in accordance with generally accepted auditing standards applicable in the circumstances." However, the "generally accepted auditing standards" are a miscellany consisting primarily of a booklet outlining typical auditing procedures and pronouncements of committees of the Institute, none of which permit themselves to be specific as to the size of the sample required. Therefore, our problem is to set up a standard degree of statistical testing which could apply to audit situations. It is with this aim that this study has been undertaken.

---

2 In "The Revised Securities and Exchange Commission's Rule on 'Accountants Certificates'", Statement on Auditing Procedure No 6. (March, 1941)

Basis for Statistical Theory. The basis of all sampling theory is that a small number of items taken from the total group of items assumes the characteristics of the larger group.\(^4\) This can be shown very simply. Thus, a load of coal is accepted or rejected by a handful of coal tested; the behavior of all rats is deduced from the experimentation with a very few; a doctor may make a decision from as little as one drop of blood. The analogy can be drawn in auditing. Thus, if we could test a very small number of items that have the same characteristics of the total group, it might then be necessary only to pick seven or eight items and draw conclusions as to the audit outcome. The difficulty is getting representative items. If we could depend on one or two false entries regularly every week, it would be a simple matter to check all entries for a period of a week. Unfortunately, in auditing, the distribution of errors is anything but regular. The reason for this situation may be due to the variety of types of errors which may be listed as follows:

1. Human fatigue
2. Misunderstanding
3. Mechanical slips
4. Failures of equipment
5. Employee dishonesty

6. Carelessness

These are the types of errors that must be discovered in the sample tested. It can be seen that the biggest difficulty in sampling will be to make the sample representative, of the total items. Two other difficulties encountered are to have the sample adequate and to show stability. The sample is adequate when each item has had the same chance of inclusion as all other items and enough items are included to show the same result in successive samples. The sample is stable when the results stay the same regardless of the increase in size. Other problems that should be mentioned and will be discussed in Chapter II are:

1. Relevancy of the sample to the problem
2. Use of homogeneous data
3. Accuracy of work in gathering and compiling the sample

These are some of the problems that will have to be solved by the auditor to justify the use of sample items to represent the total items.

Statistics in sampling. In order to obtain a representative sample that will include material errors, our problem is to find the percentage of items that are needed to be tested. For any given percentage of errors in a group of items, mathematical formulas can be applied to calculate the probabilities of discovering those errors for various
amounts tested. The answer given is expressed in probabilities, not in certainties. That is, the mathematical answer can be given for the probabilities of obtaining either a head or a tail in a single flip of a coin. The answer, $\frac{1}{2}$, does not mean that heads and tails will turn up exactly the same for any given number of tosses, but it refers to the theoretical frequency if the coin was tossed an unlimited number of times. Thus the mathematical formula can be applied exactly only when enough items have been tested to conform with the theoretical probabilities.

This field is within the realm of the statistician when the mathematical formulas are applied to business data. The field of statistics has been quite highly developed but the theory has not been applied to auditing problems. Perhaps it is because there are very few statisticians with a working knowledge of accounting and very few accountants with enough knowledge of statistics.

If the standards are to be established, it will be by two means:

1. Cooperation of accountants and statisticians in relation to the various areas of audit procedure, i.e., the statistical method of calculating probabilities.
2. Collating the experiences of many accountants over a great number of engagements.

Statistical goals and benefits. If statistical
standards are accepted, what can we expect them to accomplish? The following three applications of sampling theory may be set up as our goals:

1. **Determination of sample size.** A mathematical answer will determine the most economical percentage of items to be tested, will alleviate much overtesting and establish a safe minimum amount of testing for all auditors to use.

2. **Test for accounting control.** Statistical standards will give a means for evaluating the actual false entries found with a preconceived standard acceptable both to the client and auditor.

3. **Test for bias in the direction of error.** If there is a consistent value gained or lost in the errors found, it may be found to be a significant misstatement. Then if a more complete audit was not desired, it would be possible to make an adjustment proportionate to the direction and amount of past errors.

If the size of the sample can be determined, the following benefits may occur:

1. Assurance that the sample is adequate.

2. Obtaining the economical cost for the sample.

---

3. An individual practitioner, new or old, would know definitely how large a sample to draw.

4. The individual practitioner would be considerably strengthened in his efforts to resist improper pressures from clients to cut work improperly or to issue a certificate of correctness where the sample shows sub-grade accounting techniques and control.

5. State boards of accounting, governmental agencies, and courts would have a basis for judgment on at least one basic aspect of the adequacy of auditing work.

The following benefits might be obtained for the second and third statistical goals:

1. Provide a means for evaluating the efficiency and adequacy of an accounting system.

2. Provide a means for possible correction of the books of account on the basis of past errors without the necessity of a detailed audit.

This chapter has attempted to point out the need for a standard procedure in testing in audits, the basis for such standards, and the resulting possible benefits if those standards were adopted. In Chapter II we will attempt to show the statistical theory necessary to obtain the goals we have set up.
CHAPTER II

STATISTICAL SAMPLING THEORY

Extent of tests. Our first goal in statistical sampling is to find the sample size. By that we refer to the most economical size of sample, yet with a reasonable assurance of discovering false entries. The mathematical theories behind this problem have been understood by the mathematicians for many years but few have attempted to apply them to accounting problems. The basis of the theory in its application to auditing was set out by Carman¹ in 1933. The subsequent articles merely rely on Carman's approach for substance, with the possible exception of Prytherch² who attempted to take up where Carman left off and analyze the results of the findings in order to apply the theory to practical auditing.

The basic theory rests on the simple rules of probability. For example, suppose we have ten balls in a bowl; nine white balls and one black ball. If we were to draw blindly from the bowl, we would have one chance in ten of drawing the black ball. Stating it a little differently we would have a 10% or .1 probability of success in drawing the black ball.


Since in one draw we must either succeed or fail in drawing the black ball, the probability of failure must be 90% or .9. This can be stated in the formula:

\[ p + q = 1 \]

where \( p \) = probability of success
and \( q \) = probability of failure

If we let \( a \) = number of ways in which a favorable outcome can appear,
and \( b \) = number of ways in which an unfavorable outcome can appear,
Then \( a + b = n \)
Where \( n \) = total number of possible events.
Then:

\[ p = \frac{a}{a+b} \text{ or } \frac{a}{n} \]
and

\[ q = \frac{b}{a+b} \text{ or } \frac{b}{n} \]

Probability of success can, therefore, be stated as the ratio of number of ways an event may succeed to the total number of ways the event may happen. An alternative would be to state it in terms of combinations if we desired the probability of success for any number of trials desired.
Thus, \( p = \frac{vC_t}{nC_t} \)

Where \( v \) = number of successful events  
\( C \) = symbol for combination  
\( t \) = number tested  
\( n \) = total number of events

This means that it is stated by the number of ways a group of tested (\( t \)) events may be selected from the successful (\( v \)) events divided by the number of ways a group of tested events may be selected from the total (\( n \)) events.

Example:

There are 8 white balls; two black balls in a bowl. If we pick 3 balls blindly from the bowl, what is the probability of success in not picking a black ball?

\[ p = \frac{vC_t}{nC_t} \]

There are 8 successful events in combinations of 3 that would pick only white balls or

\[ 8 \ C 3 = \frac{8 \times 7 \times 6}{3 \times 2 \times 1} = 56 \] combinations of just white balls.\(^3\)

There are 10 total possible events in combinations of 3 or 10 \( C 3 = \frac{10 \times 9 \times 8}{3 \times 2 \times 1} = 120 \) total combinations

\[ p = \frac{56}{120} = \frac{7}{15} \]

In other words we would have 7 out of 15 chances of success; that is, in not picking a black ball.

---

Since \( p \neq q \neq 1 \)

\[
P = \frac{7}{15}
\]

Then \( q = 1 - \frac{7}{15} = \frac{8}{15} \) or \( .5333 \) chance of picking at least one black ball.

We can expand our formula in general terms:

\[
P = \frac{\binom{v}{Lt}}{\binom{n}{Lt}} \cdot \frac{(v-1)(v-2)\ldots(v-Lt+1)}{(n-1)(n-2)\ldots(n-Lt+1)}
\]

Formula (1)

With this formula we could calculate the probability of success in drawing a false item out of the total false items for any given number tested in an audit. However, the number tested in auditing is usually so large that the formula becomes too cumbersome so we can recast it:

\[
P = \frac{pCf}{nCf}
\]

Where

- \( p \) = number of items not tested
- \( C \) = symbol for combination
- \( f \) = estimated number of false entries
- \( n \) = total items in audit

Since this formula will find the probability of success of finding the number of false items (\( f \)) that can be selected from the number not tested (\( p \)) divided by the number of ways a group of false items can be selected from the

\[\text{Carman, op. cit., p. 10.}\]
total events, we will come to the same result. If we expand this formula in general terms:

\[ P_g = pC_f = \frac{p(p-1)(p-2)\ldots(p-f+1)}{n(n-1)(n-2)\ldots(n-t+1)} \]

To check the results of this formula we will use the same example as before; 8 white balls, 2 black balls, and drawing 3 balls from the bowl as before.

Then

- \( t = 3 \) = number tested
- \( p = 7 \) = number not tested
- \( n = 10 \) = total number
- \( f = 2 \) = black balls
- \( v = 8 \) = white balls

\[ pC_f = 7C2 \text{ or } \frac{7 \times 6}{2 \times 1} = 21 \text{ possible combinations of black balls out of the number not tested.} \]

\[ nC_f = 10C2 = \frac{10 \times 9}{2 \times 1} = 45 \text{ total possible combinations} \]

a group of false items can be selected from the total events.

\[ P_g = \frac{21}{45} = 0.4444 \text{ or } 7 \text{ out of } 15 \text{ chances of there being a black ball in those items not picked or in not picking a black ball which is the same result as before.} \]

To show more clearly the amount of work saved in using formula 2, we will take another example in terms of an audit:

\[ \text{Carman, op. cit., p. 10.} \]
n = 100 = total items in accounts payable

\( t = 10 \) = number tested

\( p = 90 \) = number not tested

\( f = 2 \) = false entries

\( v = 98 \) = valid entries

By number 1 formula

\[
\frac{v^t \cdot (n-t)^{n-t}}{n^{n-t}} = \frac{98 \times 97 \times 96 \times 95 \times 94 \times 93 \times 92 \times 91 \times 90 \times 89}{100 \times 99 \times 98 \times 97 \times 96 \times 95 \times 94 \times 93 \times 92 \times 91}
\]

\[= \frac{90 \times 89}{100 \times 99}\]

By number 2 formula

\[
\frac{p^t \cdot (n-t)^{n-t}}{n^{n-t}} = \frac{90 \times 89}{100 \times 99}
\]

This illustrates the difference in work for only ten entries tested while in the usual audit there might be hundreds tested but only a few false entries to be found. Then for any given size of total items checked, if we know the number of false items we can compute the probability of success of finding at least one false item for any value of \( t \) (number tested). Table 1 shows the probabilities of success of finding at least one false item out of a group of 10 items, given various values for \( f \) and \( t \).

---

6Supra, p. 13.

7Supra, p. 14.

8Procedure for estimating false entries is shown. Infra, p. 56.
To show the derivation of the probability percentages we can take a hypothetical case:

**Given**  
\[ n = 10 \]  total items  
\[ f = 2 \]  estimated false entries  
\[ t = 4 \]  number tested

What is the probability of finding 1 false item?  
\[ P_g = \frac{pC_f}{nC_f} = \frac{6 \times 5}{10 \times 9} = .33 \]  probability of not finding a false item.

\[ 1 - P_g = .67 \]  as shown on line 4, column 2. This is the probability that if four items are checked we will discover at least 1 false item, or 6 out of 9 times if we check 4 items, at least one of the 4 will be false.

---

Taking a larger example:

\[ n = 10,000 \quad t = 4,000 \quad f = 5 \]

\[ p_n = \frac{6,000 \times 5.999 \times 5.998 \times 5.997 \times 5.996}{10,000 \times 9,999 \times 9,998 \times 9,997 \times 9,996} = .077708 \]

\[ p_s = 1 - p_n = .922292 \]

This is obviously a very awkward calculation, so we can approximate and still get a very close result:

**Approximation A.**

\[ \frac{6,000 \times 6,000 \times 6,000 \times 6,000 \times 6,000}{10,000 \times 10,000 \times 10,000 \times 10,000 \times 10,000} = .077760 \]

\[ s = .922240 - \text{correct to 4 places on an average.} \]

**Approximation B.**

\[ \frac{5.998 \times 5.998 \times 5.998 \times 5.998 \times 5.998}{9.998 \times 9.998 \times 9.998 \times 9.998 \times 9.998} = .077708 \]

\[ s = .92292 - \text{correct to 6 places using probability of success as } S_a \text{ and probability of failure as } N_a \text{ - Approximation A can be summarized as follows:} \]

\[ S_a = 1 - N_a = 1 - \left(\frac{p}{n}\right)^f = 1 - \left(\frac{n-t}{n}\right)^f \quad \text{Formula (3)} \]

**Approximation B can be written:**

\[ S_b = 1 - N_b = 1 - \left(\frac{p-f-1}{n-f-1}\right)^{\frac{f}{2}} \quad \text{or} \]

\[ 1 - \left(\frac{2p - f + 1}{2n - f + 1}\right)^{\frac{f}{2}} \quad \text{Formula (4)} \]

\[ ^{10}\text{Carman, op. cit., p. 10.} \]

\[ ^{11}\text{Op. cit., p. 10.} \]
Table II  Probability Ratios With Approximations

<table>
<thead>
<tr>
<th>N</th>
<th>T</th>
<th>F</th>
<th>Approximation</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>A</td>
<td>B</td>
<td>Exact</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>30</td>
<td>2</td>
<td>.510000</td>
<td>.512109</td>
<td>.512121</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>40</td>
<td>3</td>
<td>.784000</td>
<td>.788334</td>
<td>.788374</td>
<td></td>
</tr>
<tr>
<td>1,000</td>
<td>200</td>
<td>2</td>
<td>.360000</td>
<td>.360160</td>
<td>.360160</td>
<td></td>
</tr>
<tr>
<td>1,000</td>
<td>250</td>
<td>3</td>
<td>.578125</td>
<td>.578547</td>
<td>.578547</td>
<td></td>
</tr>
<tr>
<td>1,000</td>
<td>400</td>
<td>5</td>
<td>.922240</td>
<td>.922758</td>
<td>.922759</td>
<td></td>
</tr>
<tr>
<td>10,000</td>
<td>2,000</td>
<td>3</td>
<td>.485000</td>
<td>.488038</td>
<td>.488038</td>
<td></td>
</tr>
<tr>
<td>10,000</td>
<td>3,000</td>
<td>5</td>
<td>.831930</td>
<td>.832002</td>
<td>.832002</td>
<td></td>
</tr>
<tr>
<td>10,000</td>
<td>3,000</td>
<td>10</td>
<td>.971752</td>
<td>.971807</td>
<td>.971807</td>
<td></td>
</tr>
<tr>
<td>10,000</td>
<td>4,000</td>
<td>5</td>
<td>.922240</td>
<td>.922292</td>
<td>.922292</td>
<td></td>
</tr>
<tr>
<td>50,000</td>
<td>12,500</td>
<td>4</td>
<td>.683594</td>
<td>.683606</td>
<td>.683606</td>
<td></td>
</tr>
<tr>
<td>50,000</td>
<td>10,000</td>
<td>10</td>
<td>.892625</td>
<td>.892650</td>
<td>.892650</td>
<td></td>
</tr>
<tr>
<td>50,000</td>
<td>20,000</td>
<td>5</td>
<td>.922240</td>
<td>.922250</td>
<td>.922250</td>
<td></td>
</tr>
<tr>
<td>50,000</td>
<td>15,000</td>
<td>20</td>
<td>.999202</td>
<td>.999203</td>
<td>.999203</td>
<td></td>
</tr>
<tr>
<td>100,000</td>
<td>10,000</td>
<td>2</td>
<td>.190000</td>
<td>.190001</td>
<td>.190001</td>
<td></td>
</tr>
<tr>
<td>100,000</td>
<td>40,000</td>
<td>5</td>
<td>.922240</td>
<td>.922245</td>
<td>.922245</td>
<td></td>
</tr>
<tr>
<td>100,000</td>
<td>30,000</td>
<td>10</td>
<td>.971752</td>
<td>.971758</td>
<td>.971758</td>
<td></td>
</tr>
<tr>
<td>100,000</td>
<td>30,000</td>
<td>30</td>
<td>.999977</td>
<td>.999977</td>
<td>.999977</td>
<td></td>
</tr>
<tr>
<td>100,000</td>
<td>40,000</td>
<td>50</td>
<td>.999999*</td>
<td>.999999*</td>
<td>.999999*</td>
<td></td>
</tr>
</tbody>
</table>

*Same percentages checked.

The approximations are calculated from formulas 3 and 4\(^{13}\) while the exact probabilities are calculated from formula 2 \(^{14}\).

In comparison, using Table II, approximation B is extremely accurate to 6 decimal places above 1,000 and may be depended upon for 3 or 4 places when N lies between 100 and 1,000.

Approximation A is not so accurate but in all cases above 1,000 is accurate from 2 to 4 decimals or more. It is

\(^{12}\)Carman, op. cit., p. 10.

\(^{13}\)Supra, p. 17.

\(^{14}\)Supra, p. 14.
accurate to 2 decimals from 100 to 1,000 which is the class the average audit of any size will be in. Since probability expressed in more than two decimals is of no value to the auditor, Approximation A can be used in that class range with confidence. It should be noticed in Table II, that the approximations approach the exact as N increases but they are always conservative, as the probability of success is always less than the exact probability. It should also be noticed that whenever the percentage of items tested to the total items is the same, (as shown in Table II the items marked by an asterisk), the probability of success is exactly the same. It seems odd that the probability would stay the same regardless of whether there are 1,000, 10,000 or 100,000 items checked.

For this to be true, we must make one basic assumption, that the false items are distributed evenly throughout the total items. If we were making a 25% test, we would assume that the total items (commonly called the universe) is split into four equal parts. If there were four false items, then there would be one false item in each part. We would then be assured of discovering a false item by checking one of the parts.

In order to be assured that the false items will be uniformly distributed in the universe, a method of sampling is employed called random sampling. It is a method of picking
the items at random. This would divide the universe into four equal parts so the probability of success remains the same for the same percentage tested. We can therefore draw up a chart based on Approximation A, to show the probabilities of success for any given percentage of items tested.

<table>
<thead>
<tr>
<th>Percentage Tested</th>
<th>.1</th>
<th>.2</th>
<th>.3</th>
<th>.4</th>
<th>.5</th>
<th>.6</th>
<th>.7</th>
<th>.8</th>
<th>.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
<td>.1</td>
</tr>
<tr>
<td>3</td>
<td>.37</td>
<td>.37</td>
<td>.37</td>
<td>.37</td>
<td>.37</td>
<td>.37</td>
<td>.37</td>
<td>.37</td>
<td>.37</td>
</tr>
<tr>
<td>4</td>
<td>.56</td>
<td>.56</td>
<td>.56</td>
<td>.56</td>
<td>.56</td>
<td>.56</td>
<td>.56</td>
<td>.56</td>
<td>.56</td>
</tr>
<tr>
<td>5</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
</tr>
<tr>
<td>6</td>
<td>.94</td>
<td>.94</td>
<td>.94</td>
<td>.94</td>
<td>.94</td>
<td>.94</td>
<td>.94</td>
<td>.94</td>
<td>.94</td>
</tr>
<tr>
<td>7</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
</tr>
</tbody>
</table>

Table III shows the probabilities of success for percentages tested for various values of \( f \) (number of false items). This chart can be extended to any number of false entries. To obtain the percentage that should be tested for a given number of false entries, the probability levels are examined in the false item row until the desired level is shown. The column heading this probability figure is the stated percentage items that should be tested. If the desired probability figure is not stated exactly in the chart, the

\[15\text{As adapted from Carman, op. cit., p. 10.}\]
difference can be interpolated. For example: If there are 4 estimated false entries and the desired probability level is 90%, the percentage to be tested would be between 40 and 50%.

$$\begin{align*}
5 \left( \frac{87}{10} = 40\% \right) & \quad 3 \left( \frac{92}{10} = 50\% \right) \\
\frac{3}{5} \times 10 & = 6 \\
90 & = 40 + 6 = 46\%
\end{align*}$$

Table III is entirely satisfactory for groups of 1,000 or more and acceptable in most cases down to groups of 100. Groups lower than 100 would probably be inaccurate but always conservatively so. However, in cases of small groups, it is best to use the exact formula to determine the probabilities of success.

An objection to these tables should be discussed at this time; that raised by Abrams. It is his contention that practically none of the items tested by an auditor form a normal distribution. Because there is such a small percentage of defectives among the items sampled, the type of distribution is very much skewed, and the Poisson distribution

---

16 *Supra*, p. 18

17 Jerome Abrams "Sampling Theory applied to the Test Audit." *New York Certified Public Accountant*. (October, 1947) P.

18 The Poisson distribution is an approximation of the binomial distribution which is the basis of a normal distribution. The mathematical differences in the two types of distributions is brought out by Arne Fisher, *The Mathematical Theories of Probabilities*. New York: Macmillan Company, 1922. 110-115.
would fit the data much closer. This would make an entirely different chart of probabilities.

The Poisson distribution is applicable in cases where the happening of an event is very rare. Examples of rare events are: the drawing of the Ace, King, Queen and Jack of Diamonds with the first four cards from a pack of cards; throwing a double six with a pair of dice twice in succession; or contracting a case of Scurvy in Boston. The circumstances in which the Poisson distribution best fits the data in auditing is when \( N \) (number of entries) approaches infinity and \( F \) (false entries) approaches 0.

Our problem is to decide whether false items can be termed rare. That brings up another question. What are false items? If we considered false items fraud only, then the possibility of fraud occurring in a business is rare and would undoubtedly be a good example of a Poisson distribution. But in auditing a client's books, the client desires all material errors brought to light regardless of the reason. Types of errors other than fraud\(^{19}\) are present in every accounting system so the possibility of a false item being present if all material errors were included could not be termed rare. Also, if a business did have fraud present, it is generally continuous and would follow more or less a normal distribution. The Poisson distribution is designed for very large groups while the auditing universe would be com-

\(^{19}\text{Supra, p. 6.}\)
paratively small, with chief concern that fraud is present and is continuous. For these reasons this writer feels that the objection raised by Abrams will not apply in the majority of audits as the majority will contain enough false items in the universe to form a normal distribution.

To be able to apply Table III, we must know the desired probability of success. A high probability is often desired but the extent of the tests necessary might result in excessive costs of the audit. The most economical sample can be calculated but often the probability of detection might be unsatisfactory to the client. The difference between the percent tested and the resulting probability of success is the economy of effort. Thus, if for a 10% test for two false items, the resulting probability of success would be 19%, then the economy of effort is 9%. The economies are illustrated below in Table IV for various percentages tested in order to find at least one false item if there is a total of two false items in the universe.

<table>
<thead>
<tr>
<th>Test Column 1</th>
<th>Resulting Probability Column 2</th>
<th>Difference or Economy Column 3 (1-2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>.19</td>
<td>.09</td>
</tr>
<tr>
<td>20</td>
<td>.36</td>
<td>.16</td>
</tr>
<tr>
<td>30</td>
<td>.51</td>
<td>.21</td>
</tr>
<tr>
<td>40</td>
<td>.64</td>
<td>.24</td>
</tr>
<tr>
<td>50</td>
<td>.75</td>
<td>.25</td>
</tr>
<tr>
<td>60</td>
<td>.84</td>
<td>.24</td>
</tr>
<tr>
<td>70</td>
<td>.91</td>
<td>.21</td>
</tr>
<tr>
<td>80</td>
<td>.96</td>
<td>.16</td>
</tr>
<tr>
<td>90</td>
<td>.99</td>
<td>.09</td>
</tr>
<tr>
<td>100</td>
<td>1.00</td>
<td>--</td>
</tr>
</tbody>
</table>

20Carman, op. Cit., p. 10
Thus it can be seen that the most economical sample would be 50% with a resulting probability of success of 75%. If 10% more effort was expended there will be only a 9% increase in probability of detection. A table can be drawn up to show the economy for any level of tests or number of false items.

<table>
<thead>
<tr>
<th>Percent Tested</th>
<th>Number of False Items</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$P_s$</td>
<td>$E$</td>
<td>$P_s$</td>
<td>$E$</td>
<td>$P_s$</td>
<td>$E$</td>
</tr>
<tr>
<td>10%</td>
<td></td>
<td>19</td>
<td>09</td>
<td>27</td>
<td>17</td>
<td>34</td>
<td>24</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>36</td>
<td>16</td>
<td>49</td>
<td>29</td>
<td>59</td>
<td>39</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>51</td>
<td>21</td>
<td>66</td>
<td>36</td>
<td>76</td>
<td>46</td>
</tr>
<tr>
<td>40</td>
<td></td>
<td>64</td>
<td>24</td>
<td>78</td>
<td>38*</td>
<td>87</td>
<td>47*</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>75</td>
<td>25*</td>
<td>87</td>
<td>37</td>
<td>94</td>
<td>44</td>
</tr>
<tr>
<td>60</td>
<td></td>
<td>84</td>
<td>24</td>
<td>94</td>
<td>34</td>
<td>97</td>
<td>37</td>
</tr>
<tr>
<td>70</td>
<td></td>
<td>91</td>
<td>21</td>
<td>97</td>
<td>27</td>
<td>99</td>
<td>29</td>
</tr>
<tr>
<td>80</td>
<td></td>
<td>96</td>
<td>16</td>
<td>99*</td>
<td>19*</td>
<td>99*</td>
<td>19*</td>
</tr>
<tr>
<td>90</td>
<td></td>
<td>99</td>
<td>09</td>
<td>99*</td>
<td>9*</td>
<td>99*</td>
<td>9*</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>100</td>
<td>--</td>
<td>100</td>
<td>--</td>
<td>100</td>
<td>--</td>
</tr>
</tbody>
</table>

*Most economical sample

Table V shows the probability of finding a false item for various tests with the resulting economies. The economies marked with an asterisk show the most economical size of the sample for universes with false items of from 1 to 7 items. It should be cautioned that the most economical sample may have a success probability that would be considered too small.

21This chart is derived by expanding Table IV for universes with false items from two to seven items.
Choice of Items Tested. There are two main methods for choosing the items to be tested.

1. Unit-months---this is the commonest method in practice at the present time. All the entries in all the books of account are checked for a period, usually a month's entries at a time for as many months as are desired. It is generally thought that the greatest danger of fraud is in the first and last months of the accounting period, so if it were a 25% check, the first, last and one other month would be checked completely. The objection to this procedure is obvious. If it became standard procedure to check the first and last months, the defaulter could confine his activities to the other months or concentrate his false entries in one month, thus minimizing the chance of detection. Let us examine the effect of block testing in the probability ratio.

Example:

12,000 items in accounts payable or an average of 1,000 per month.

\[ f = 6 \text{--estimated number of false entries} \]
\[ t = 3,000 \text{--number of items tested} \]

Using approximation A, the probability of success in discovering at least one false item is:

\[ P_s = 1 - \frac{(p)^f}{(n)} = 1 - \frac{(9,000)^6}{(12,000)} = .82 \]

\[ ^{22} \text{Supra, p. 17.} \]
If we check 3 months solidly, however, we would have 6 probability ratios, each ratio depending on the number of false items in each month. If the false items are in 6 separate months, the probability is .91.

Proof:

\[ n = 12 \] - total items in terms of months
\[ t = 3 \] - number of months tested
\[ p = 9 \] - number of months not tested
\[ f = 6 \] - number of false items

\[ P_s = \frac{pC_f}{nC_f} = \frac{9C_6}{12C_6} = .09 \]

\[ 1 - P_s = .91 \] - probability of success if distribution is in 6 separate months.

If distribution is in:

- 5 months the probability is .84
- 4 months the probability is .75
- 3 months the probability is .62
- 2 months the probability is .45
- 1 month the probability is .25

It can be seen that if the items are evenly distributed the block testing offers a higher probability ratio but the more concentrated the items get, the smaller the ratio becomes. A defaulter with this knowledge could evade detection the easiest by concentrating his items in one month, which he would undoubtedly do.
2. The other means of choosing items is called the Unit-item method. This method merely chooses items throughout the universe. It is broken down further into 3 types of sampling by unit items.

1. random sampling
2. stratified random sampling
3. purposive sampling

By random sampling is meant the selection of a sample in such a way that each item in the universe has an equal chance of being part of the sample.

Stratified random sampling is the breakdown of the universe into sub-universes, usually according to dollar value and testing various percentages according to the dollar value.

Purposive sampling is the deliberate attempt on the auditor's part to choose a sample that is representative of the universe.

Random sampling is the system which is best suited to draw a representative sample. In both the other methods the bias of the auditor may give a very unrepresentative sample. Professor Vance\textsuperscript{23} suggests there is a strong case for "subjective randomizing," (purposive sampling). He admits that the statistician claims the personality of an individual almost always introduces a bias which partially nullifies the

attempt to have a representative sample, but Vance claims that that objection applies only when the drawer can see the attribute sampled, i.e., the size of watermelons, or the shade of paint. In auditing, however, the auditor cannot see the attribute, that is, a fraudulent entry; every entry is potentially false. Thus by a conscious effort for randomness, he should be able to draw a good sample. This is undoubtedly very true but there is a very real objection to purposive sampling. It gives the auditor an opportunity to use his own judgment and apply some little ability in picking representative items to be tested. Thus if an auditor had a bad day, he might be careless and choose a poor sample. There is a great temptation in an auditor's path in present day auditing methods to be careless. The feeling that "there probably isn't any fraud anyway so I'll save a little time" is a very easy pitfall. The alternative for the auditor is to draw from his universe items by choice, giving every item an equal opportunity to be drawn. This would be quite laborious and expensive, but random series have been drawn up in various publications.  

24A random series is a group of numbers that have selected entirely by chance. Three different series have been incorporated into one publication by L. H. C. Tippett. *Tracts for Computers, No. 15, Random Sampling Numbers*. London: Cambridge University Press, 1921. The three series of numbers are: (1) L. H. C. Tippett numbers which comprise 41,600 digits taken from census reports combined into fours to make 10,400 four figure numbers.  

(continued)
The procedure of selecting the numbers from the random numbers in the tables is simple. For example, let us assume that we are going to test 100 items out of a total universe of accounts receivable of 500, a 20% check. We first number the universe from 1-500, numbering every 10th item. We then open the table of random numbers to any page and reading in any direction list the first 100 numbers of 3 digits each, ignoring any digit numbers over 500 since our numbering of items is only to 500. The first 200 numbers of the Kendall Babington Smith tables of random numbers are presented below:

Random Sampling Numbers

<table>
<thead>
<tr>
<th>2315</th>
<th>7548</th>
<th>5901</th>
<th>8372</th>
<th>5993</th>
<th>7624</th>
<th>9708</th>
<th>8695</th>
<th>2303</th>
<th>6744</th>
</tr>
</thead>
<tbody>
<tr>
<td>0554</td>
<td>5550</td>
<td>4310</td>
<td>5374</td>
<td>3508</td>
<td>9061</td>
<td>1837</td>
<td>4410</td>
<td>9622</td>
<td>1343</td>
</tr>
<tr>
<td>1487</td>
<td>1603</td>
<td>5032</td>
<td>4043</td>
<td>6223</td>
<td>5005</td>
<td>1003</td>
<td>2211</td>
<td>5438</td>
<td>0834</td>
</tr>
<tr>
<td>3897</td>
<td>6749</td>
<td>5194</td>
<td>0517</td>
<td>5853</td>
<td>7880</td>
<td>5901</td>
<td>9432</td>
<td>4287</td>
<td>1695</td>
</tr>
<tr>
<td>9731</td>
<td>2617</td>
<td>1899</td>
<td>7553</td>
<td>0870</td>
<td>9425</td>
<td>1258</td>
<td>4154</td>
<td>8821</td>
<td>0513</td>
</tr>
</tbody>
</table>

Starting from the first digit, the first 25 items to be selected from the universe would be:

321, 57, 54, 85, 90, 183, 72, 59, 93, 76, 249, 70, 88, 69, 52, 303, 67, 440, 55, 455, 50, 431, 053, 74, 350.

Random numbers put the level of testing on an objective basis and leave no room for the whims of an individual.

---

24 (continued)
(2) Kendall Babington Smith numbers comprising 100,000 digits grouped in twos and fours in 100 separate thousands.
(3) R. A. Fisher and F. Yates numbers comprising 15,000 digits arranged in twos.

25 op. cit., p. 28.
practitioner. As stated previously, it is the best system for obtaining a normal distribution upon which the tables of probability are based. Unit-item methods of stratified and purposive sampling are effective only if the sample is not too biased.

In order to show clearly the difference in the probability ratios between the unit-month method and the unit item method, tables VI and VII have been drawn up.

Table VI  Unit-Month Probability Ratios

<table>
<thead>
<tr>
<th>No. of Months Tested</th>
<th>No. of Months containing False Items</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>.08</td>
</tr>
<tr>
<td>2</td>
<td>.17</td>
</tr>
<tr>
<td>3</td>
<td>.25</td>
</tr>
<tr>
<td>4</td>
<td>.33</td>
</tr>
<tr>
<td>5</td>
<td>.42</td>
</tr>
<tr>
<td>6</td>
<td>.50</td>
</tr>
<tr>
<td>7</td>
<td>.58</td>
</tr>
<tr>
<td>8</td>
<td>.67</td>
</tr>
<tr>
<td>9</td>
<td>.75</td>
</tr>
</tbody>
</table>

Table VII  Unit-Item Probability Ratios - 12,000 items

<table>
<thead>
<tr>
<th>False items</th>
<th>NO. of items tested</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000</td>
<td>.08</td>
<td>16</td>
<td>23</td>
<td>29</td>
<td>35</td>
<td>41</td>
<td>46</td>
<td>50</td>
<td>54</td>
<td>58</td>
<td>62</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>2,000</td>
<td>.17</td>
<td>31</td>
<td>42</td>
<td>52</td>
<td>60</td>
<td>67</td>
<td>72</td>
<td>77</td>
<td>81</td>
<td>84</td>
<td>87</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>3,000</td>
<td>.25</td>
<td>44</td>
<td>58</td>
<td>68</td>
<td>76</td>
<td>82</td>
<td>87</td>
<td>96</td>
<td>92</td>
<td>94</td>
<td>96</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>4,000</td>
<td>.33</td>
<td>56</td>
<td>70</td>
<td>80</td>
<td>87</td>
<td>91</td>
<td>94</td>
<td>96</td>
<td>97</td>
<td>98</td>
<td>99</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>5,000</td>
<td>.42</td>
<td>66</td>
<td>80</td>
<td>88</td>
<td>93</td>
<td>96</td>
<td>98</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>6,000</td>
<td>.50</td>
<td>75</td>
<td>88</td>
<td>94</td>
<td>97</td>
<td>98</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>7,000</td>
<td>.58</td>
<td>83</td>
<td>93</td>
<td>97</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>8,000</td>
<td>.67</td>
<td>89</td>
<td>96</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>9,000</td>
<td>.75</td>
<td>94</td>
<td>98</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td></td>
</tr>
</tbody>
</table>

26Carman, op. cit., p. 10.
27Ibid.
To illustrate the difference:
In order to find 5 estimated false entries, the auditor decides to check 33 1/3% of the items. If he chooses the Unit-month method, he will check 4 months completely. The probability ratios are .33, .58, .75, .82, .93, depending on the number of months the false items are contained in, whether 1, 2, 3, 4, or 5 months.

If he chooses the unit-item method, the probability ratio is 87%. Therefore if the items are dispersed in 5 different months, the unit-month method is more accurate. Otherwise the unit-item is the most accurate. It should be pointed out that the probability ratio for the unit-item method remains at 87% regardless of the concentration of items but the ratios go down as low as 33% if the items become concentrated.

To summarize, the extent of tests can be found in two cases from Table III. The percentage of items to be tested to give a desired probability level for a universe with a given number of false items can be found and the level that previous testing has already reached can be determined. To illustrate Case 1, if there are 5 estimated false entries out of a universe of 1,000 entries and we desire a 95% probability of discovering at least one false entry in our audit, by Table V on page 24, we must make a 45% test. To illustrate Case 2, if there were 4 estimated false entries out of a universe of 1,000 but none were found in a sample of 350,
our probability for having found at least one false item in a 35% check is 82%. Therefore there is 18% chance of missing all false entries.

The most economical sized sample can be found in Table V or for any desired probability level the economy is shown. If there are 5 estimated false entries, the most economical sample by Table V would be a 30% check, which given a probability level of 83%. If a 95% probability is desired for finding 5 false entries, approximately a 46% test is necessary. The economy is 49%, or in other words, we would have a 95% probability of finding at least one false entry with only a 46% test.

Depending on the method of choosing the items to be tested, either Table VI or Table VII will be used to find the probability level. If the Unit-month method is used, an assumption of the amount of months that the items are contained in is necessary to obtain the probability level. The more concentrated the items become, the lower probability of detection while the ratio remains constant in the unit-item method if there is an even distribution. We assure ourselves of an even distribution by making the selection of items at random, using tables of random numbers.

The problems still unsolved by the Carman theory are:
1. False entries have to be estimated.\(^\text{28}\)

\(^{28}\) *Infra*, p. 56.
2. Drawing a distinction between types of errors, whether honest errors or fraud.  

3. Making a distinction between accounting systems.  

A firm with a high degree of control over the books would not need the same extent of sampling as a firm with little or no internal control. This brings us into the second statistical problem--accounting control.

**Test for accounting control.**

**Test for Standard Population.** That the quality of accounting should be a factor in the extent of tests is evident, but the determination of the quality is very difficult for the auditor unfamiliar with the accounting system. The relationship between internal control and the extent of the auditors tests has not been set forth in sufficient definite terms either to assist the auditor in the exercise of judgment or to furnish any standard by which his judgment may be appraised. This lack would not be so serious if the standards to estimate the size of the sample that would be adequate under any set of conditions and without regard to internal control were accepted. Recently the American Institute of Accountants recognized this problem and published the following report of which this is a part.

---

29*Infra*, p. 50.

30*Infra*, p. 37.

"There is to be a proper study and evaluation of the existing internal control as a basis for reliance thereon and for the determination of the result and extent of tests to which auditing procedures are to be restricted."

A theory to test for accounting control has been explored by Vance and to date is the only article of value that this writer has discovered. The fundamental conception for testing internal control is to assume that for each strength of internal control there is a normal or a standard which should be attained. Having set a standard for our sample to attain if it is satisfactory, we can then compare the actual sample drawn with the standard. We may, therefore, consider the sample drawn in the light that it may have been drawn by chance out of all other possibilities from the population, and upon the relative rarity or frequency of getting the actual sample, decide whether or not, we think it came from the standard population.

Thus, if a person claimed that he had a special skill in tossing heads in a flipping coin, and proceeded to flip two heads in a row, we could not be sure, for he would attain the same result by chance one out of 4 times. If, however, he tossed 6 heads in succession, we would be inclined to believe he did have a special skill as the result by chance.

---

32 Vance, op. cit., p. 8.
would occur only $1 \frac{1}{2}$ times in 100 approximately. This is the fundamental reasoning in the test for accounting control. It is to determine whether the variations of the measures of the sample from the measures of the standard might have occurred as chance events or whether some non-chance factor has produced a significant variation. A difference is said to be significant if the probability of its occurrence as a chance variation is so small that a hypothesis of the existence of non-chance factors is more tenable. The common practice is to suspect significant causes when $P^{33}$ is less than 5%, to search actively for such causes when $P$ is less than 1% and to call value of $P$ that are less than 1% almost certain evidence of a significant difference. In other words, if the probability is 5% or less that the differences are due just to chance in picking the sample, then the odds are too small to believe that it was by chance. Some non-chance factors may be:

1. presence of some bias in taking the sample
2. failure to provide in the sample for all of the characteristics of the universe
3. an inadequate number of cases in the sample
4. sample more defective than the standard sample

$P^{33}$ refers to the probability that the sample drawn might have been drawn out of all other possibilities from the standard population that we have set up. Supra, p. 34.
Our objective is number 4, so if we can eliminate
the first three non-chance factors, then the significant
non-chance factor must be due to a more defective sample
than the standard. The first non-chance factor may be elim­
inated if we draw the sample at random. If we accept a high
probability of success ratio in a postulated sample, that is,
making the percentage tested high enough to be reasonably
sure of finding an estimated defective population, the sec­
ond and third non-chance factors will be eliminated. Then
assuming our sample is representative, significant differences
in our drawn sample with the standard sample must be due to
the sample being more defective than the standard population.

Procedure of setting up the standard. In finding the
normal or standard sample, we must make an assumption as to
the allowable defective amount or percentage within a uni­
verse. This applies to two types of errors:

1. procedural-failures to follow prescribed methods of
   internal check of the system of accounting adopted.
   Example: Failure to countersign a check

2. substantive-errors that result in a misstatement of
   values in accounts or statements.
   Example: Extension of 2.00 x 300 as 60.00.

The standard sample that is set up will be governed
by the purpose of the audit and the degree to which each pro­
posed step justifies its own cost by its contribution to the
objective. Thus the standard for a large publicly owned
company with an efficient organization and strong internal
control will vary greatly with a small closely held corporation without the degree of control, organization and system.\textsuperscript{34}

The objective in the first case is:

1. that financial statements have been fairly stated.
2. that acceptable accounting procedures have been followed.
3. that assets are fairly stated and do exist.
4. that all liabilities are included.

The auditor would not concern himself with most minor errors or even unimportant instances of fraud since that is presumed to be the internal control business. Therefore, a very liberal standard would be set up, possibly as much as \(10\%\) defective errors allowed.

The objective in the smaller, less efficient organization would be enough check to disclose systematic fraud or gross carelessness. The sample would only be as large as it is economical and useful. The extent in these cases will vary under different conditions but will gradually standardize itself with the auditor's knowledge of the accounts and the client's objective.

The standard sample extent will also be governed by the particular body of items tested. To be effective, random

\textsuperscript{34}William Cranston, "A New Look at Basic Auditing Techniques." \textit{Journal of Accountancy}, (October, 1948), 274.
sampling must deal with homogeneous items. A suggested breakdown is the following:

1. accounts receivable balances
2. accounts payable balances
3. inventory types
4. inventory footings
5. purchase vouchers
6. cash disbursements
7. payrolls
8. cash book postings
9. voucher register footings.

Each of the preceding classifications is considered a universe and the test results from each universe should be able to stand alone. It is a general rule that the greater liquidity an account is concerned with, the greater the extent of the checking. Thus, cash and items easily converted into cash will have to have a higher standard of testing than classifications in which fraud is less likely to occur.

If a standard was set up for 5% defective items or a 5% defective dollar amount of the total, we could assume that the actual sample was normal if the defective items were about 5% of the total. The probability that the defective items will fall exactly at the amount allowed is very small,

---

35 By Leo Herbert "Practical Sampling for Auditors." N. Y. Certified Public Accountant. (January, 1947) 57-61
but if we have a series of samples that are all normal, they will average close to the standard amount. Thus, if each sample were representative of the universe, each sample's defective items would approximate 5% and vary only slightly due to not having exactly the right proportion of right and wrong items due to incomplete testing. The amount of variation from the 5% of the universe can be measured in terms of per item variation by the following formula:

\[
SE^{36} = \sqrt{\frac{pq}{n}}
\]

This is called the standard error of the universe.

By assuming a standard of a universe, the variations will form a normal curve of distribution, that is, there will be variations over 5% and variations under 5% but their average will be 5%. It is common knowledge in statistics\(^{37}\) that plus or minus one standard deviation (error as we call it) from the exact normal value will contain 68% of the items if the universe is normal. Plus or minus 2 standard deviations contains 95.5% of the items and plus or minus 3 standard deviations would contain 99.73% of the items. Stating it a little differently, if the tested sample's standard deviation is three times that of the normal standard's deviation, then the result could have occurred by chance out of 100 only 100-


99.73 times, or .27 since by definition plus or minus three standard deviations from the normal will contain 99.73% of the items. It should be noted that that result or worse could have occurred in that .27 percentage. We can show this relationship more clearly with the aid of the normal curve.

Figure 1

Normal curve of distribution

The deviation from the normal is measured from point D toward points A and G in the above illustration. Deviations greater than 5% will be shown to the right of D and deviations less than 5% will be shown to the left. Since our normal universe will have exactly 5% defective items, the area on either side of the standard (represented in the figure by DD') will be equal. From C to E represents one standard deviation on
either side of the standard and will contain 68% of all the items if the universe is normal. The actual sample drawn will probably show a deviation, either greater or less than the normal 5% so the percentage of items (or area) that will contained with the deviation found can be stated as a percentage of half the total items. A table can be drawn up to show the exact percentage of area in a normal distribution for any given variation from the standard deviation.

<table>
<thead>
<tr>
<th>Table VIII Normal Curve Area Table</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>0.0</td>
</tr>
<tr>
<td>0.1</td>
</tr>
<tr>
<td>0.2</td>
</tr>
<tr>
<td>0.3</td>
</tr>
<tr>
<td>0.4</td>
</tr>
<tr>
<td>0.5</td>
</tr>
<tr>
<td>0.6</td>
</tr>
<tr>
<td>0.7</td>
</tr>
<tr>
<td>0.8</td>
</tr>
<tr>
<td>0.9</td>
</tr>
<tr>
<td>1.0</td>
</tr>
<tr>
<td>1.1</td>
</tr>
<tr>
<td>1.2</td>
</tr>
<tr>
<td>1.3</td>
</tr>
<tr>
<td>1.4</td>
</tr>
<tr>
<td>1.5</td>
</tr>
<tr>
<td>1.6</td>
</tr>
<tr>
<td>1.7</td>
</tr>
<tr>
<td>1.8</td>
</tr>
<tr>
<td>1.9</td>
</tr>
<tr>
<td>2.0</td>
</tr>
<tr>
<td>2.1</td>
</tr>
<tr>
<td>2.2</td>
</tr>
<tr>
<td>2.3</td>
</tr>
<tr>
<td>2.4</td>
</tr>
<tr>
<td>2.5</td>
</tr>
<tr>
<td>2.6</td>
</tr>
<tr>
<td>2.7</td>
</tr>
<tr>
<td>2.8</td>
</tr>
<tr>
<td>2.9</td>
</tr>
<tr>
<td>3.0</td>
</tr>
<tr>
<td>4.0</td>
</tr>
</tbody>
</table>

The column on the left represents the number of standard deviations the drawn sample varies from the normal population. The percentage of area shown is expressed as area on one side of the curve. The percentage found will be subtracted from .500 to find the area not included with a deviation as great or greater than that of the sample drawn. To show clearly how the normal curve area table is used, we can illustrate in terms of an audit:

The standard is 5% allowable defective items. Out of a total of 1,000 accounts payable tested, we find 6.38% defective items.

Standard Error: (deviation)

\[ SE = \sqrt{\frac{pq}{n}} = \sqrt{.05 \times .95} = \sqrt{0.000475} = .0069 \text{ or } .69\% \]

As we are assuming the sample is representative of the universe, what is the probability that the difference between the actual percentage and the standard percentage is due to chance factors? Looking in the chart the percentage of the area under the normal curve when the actual value is 2 times that of the standard error (6.38 - 5 = 1.38; \(\frac{1.38}{.69} = 2\)) is .477250. Our result shows us that 47.725% of 50% total items (or area) is contained up to our actual error. Then our sample, if it is within the standard sample must have as great or greater deviation only 2.275% of the time. Then by chance we would pick our sample from a normal sample about
2 times out of 100. Since the odds for picking our results are so small, we would be inclined to believe that the difference isn't due just to chance factors.

It is now possible to indicate more clearly what we can do in auditing. We can measure the quality of accounting entries or calculations in terms of the percentage of a correct as against incorrect items in the population, and upon the basis of a standard limit of defective entries acceptable in the books of account and the consequent probability of drawing a given sample from it, decide whether having the sample, we are willing to accept the population as no worse than the standard. The probability limits most generally used by statisticians run from 90 - 99 percent of the distribution, the limits of 95 - 99% being commonly used. In other words, if a sample is so rare as to be drawn only 1% to 5% of the time by chance alone, it is considered too rare to give credence to the assumption that it came from the standard population.

It is at once apparent that a sample may come from a number of different standards and still may appear to come from our standard sample set up with acceptable frequencies. In other words the results from a sample may vary up to the results of a 90 or 95% probability of significant differences.

39 This method of measuring the drawn sample with a normal universe has been termed the "normal variate" by Professor Vance, op. cit., p. 8.
and still be acceptable to our standard. For auditing purposes it would be better to have a somewhat more discriminat­ing device which we find in the likelihood ratio.

Likelihood Ratio. By testing the significance of the difference found between the standard population and the actual population we can decide whether the variation is due to chance or is significant. The method of the likelihood ratio is to make such computations for a particular sample with respect to two hypotheses, for example:

- $H_1$: that the population was .005 defective
- $H_2$: that the population was .03 defective

If $P_1$ is the probability of drawing the sample from $H_1$ and $P_2$ the probability of drawing it from $H_2$, then the likelihood ratio is $\frac{P_2}{P_1}$. Then assuming we wish to make de­cisions with a risk of error of something like 10 percent, we will choose $H_1$ as our estimate of the actual population from which the sample came when the likelihood ratio is $\frac{1}{9}$ or smaller and $H_2$ when it is nine or larger. In other words, we will make a decision when the probability of getting the sample from one of the postulated populations is nine times that of getting it from the other.

40 These hypotheses are chosen arbitrarily for purposes of illustration and do not necessarily represent significant limits.
Example:

\[ H_1 \] - population .005 defective
\[ H_2 \] - population .03 defective

Total items tested in accounts payable = 1000

12 false items

\[ S_e = \sqrt{\frac{.005 \times .995}{1000}} = \sqrt{.000004975} = .00223 \]

\[ \frac{\text{observed difference}}{S_e H_1} = \frac{.005 - .012}{.00223} = -.007 = -3.14 \]

\[ P_1 = .50 - .499155 = .000845 \]

\[ P_1 = \text{probability that the sample drawn is from the standard} \]
\[ -H, \text{ Will be drawn about 8 times every 10,000 times} \]

\[ S_e H_2 = \sqrt{\frac{.03 \times .97}{1000}} = \sqrt{.0000291} = .00539 \]

\[ \frac{\text{observed difference}}{S_e H_2} = \frac{.03 - .012}{.00539} = .018 = 3.34 \]

(Table of areas) 3.34 = .499581

\[ P_2 = .50 - .499581 = .000419 \]

\[ P_2 = \text{probability that the sample will be drawn about 4 times} \]
\[ \text{every 10,000 times.} \]

The likelihood ratio is \[ \frac{P_1}{P_2} = \frac{.000845}{.000419} = 2 \]

In other words there is 2 times the likelihood of the sample came from the population, \( H_1 \) than from the population \( H_2 \).

It should be noted that the probabilities are that the sample tested came from neither population as \( P_1 \) was 8 times in 10,000 by chance and 4 times in 10,000 in \( P_2 \). The proba-
bilities show that the population sampled was worse than .005 but better than .03. This is the information that is desired as we have set an upper and a lower limit on which we can find the likelihood of which direction the actual sample is taking.

If we desired to be 90% correct in choosing either $H_1$ or $H_2$ we would have to continue our sampling until the probability of one was 9 times that of the other. It can be seen that for small test groups, the results will not be conclusive if the results neither fall into $H_1$ or $H_2$ but lie in between. Consequently the likelihood ratio is most effective for large samples. It should be used in conjunction with the most economical sample, that is the amount of testing necessary to find a certain standard of errors.

It is apparent that it is quite laborious to figure the likelihood ratio for the different valued throughout the testing but we are fortunate in having the method of sequential sampling which, in effect, enables us to calculate the likelihood ratio at each step of the drawing of the sample by merely looking into a table or by plotting points on a chart.

---

41 The calculations of the likelihood ratio on page 45 are based on the normal distribution. In sequential sampling the likelihood ratio is calculated from the binomial distribution. Since the binomial distribution approximates the normal distribution very closely as the number of items in the population increases above 100 items, the results will be approximately the same but the difference in methods of calculations should be pointed out. This is brought out by Irving Gavett, op. cit., p. 12.
Sequential Sampling. Sequential sampling developed during World War II and permits the computation of the likelihood ratio at every step of drawing the sample, which means that the sample required to make a decision is held to the minimum under the circumstances of each drawing. This treatment of the sample is possible because of the mathematical work of A. Wald,\(^{42}\) who worked out formulas which permit a simple graphical (or tabular) solution of the problem. Where a graphical procedure is used, two parallel lines are drawn on common coordinates, the X axis representing the number of items drawn and the Y axis representing number of errors found. The position and slope of the parallels are determined by the upper and lower limits desired. If we desired the upper limit to be a maximum of .03 defective items that could be tolerated, and a lower limit of .005, which would represent an efficient accounting system, then the upper parameter would be constructed for any desired value of \(\frac{P_1}{P_2}\). That is, if we wished to have a 90% probability that we are correct in deducing that the drawn sample came from the .03 population and not the .005 population, the values of \(\frac{P_1}{P_2}\) would be equal to 9 for every point along the parameter; the converse is true for the lower parameter; we would desire 9 times the probability that the drawn sample came from the .005 population than the .03 population so \(\frac{P_1}{P_2}\) would have to

be 1.9 for every point on the parameter.

Figure 2 shows it in graphic form:

In Figure 2 the upper limit of defective items that can be tolerated is represented by parameter A while the lower limit, representing an efficient set of books, is parameter B. At frequent intervals the number of false items found, to the total number tested is plotted on the graph. This procedure is represented by line M. If the results of the sampling are shown to the right of the lower limit, then there is more than 9 times the probability that the sample is efficient and so should be accepted. Conversely, if the

\[43\] L. Vance, *op. cit.*, p. 8
plotting placed the results above the upper limit, the sample should be rejected as an intolerable percentage of defective items.

An equivalent procedure has been worked for a tabular solution. The formulas which derive the tabular solution are available at the present time but as they are highly mathematical and of little value to the auditor, they will not be discussed in this paper.

In summary, the test for accounting control is the auditing analogy to tests for statistical control. It is applied to two samples to determine whether they might have come from the same population, i.e., whether the conditions producing the two populations appear constant. A more discriminating device is used in the likelihood ratio computed by sequential analysis. This test compares the relative probabilities of assuming two hypotheses on the same sample.

This brings us to our third goal in statistical sampling, that of the testing for bias in the direction of error.

**Test for bias in the direction of error.** It would be ideal to be able to infer from a sample the over and understatement of assets or income but there seems to be no

---

consistent pattern of accounting errors against which a particular sample could be interpreted. Errors in accounting arise from a variety of sources - from human fatigue, misunderstanding, mechanical slips and failures of equipment. The amount of the resulting error depends upon the digits originally involved, the amount of displacement of decimal points, transposition, and many other factors.

This would seem to deny the construction of a sampling distribution of accounting errors in terms of value. However, it can be seen that for any type of error, if the error is consistent, the error might be in one direction and have an average value. If this were so, if for any reason, the cost of the audit to find the actual errors was exorbitant, then the auditor could make a correction of the account for the total estimated error. Let us examine the types of errors:

1. human fatigue
2. misunderstanding
3. mechanical slips
4. failures of equipment
5. fraud
6. carelessness

---

45 Supra, p. 6.
46 Supra, p. 6.
Numbers 1 and 6 will probably not be biased, that is, there will be about as many positive errors as negative errors. Any of the other types of errors could have a consistent error of either positive or negative amounts in value. If the error is accidental, then in a normal distribution the negative errors will approximate the positive errors. If there is a significant difference between the positive and negative values, it may be that the errors are not due to chance factors; that there is a definite bias in one direction.

To test the significance of the difference, we will employ the normal variate.\textsuperscript{47}

Example: Percentage of the number of errors due to failures of equipment. Standard: Equal percentage of positive and negative errors.

In examining 1000 accounts payable, 10 errors due to the failure of equipment were found; 8 positive and 2 negative. Is the difference significant?

\[ S_e = \sqrt{\frac{.5 \times .5}{10}} = \sqrt{.025} = .158 \]

\[ \frac{\text{observed difference}}{S_e} = \frac{.5 - .8}{.158} = \frac{-3}{.158} = -1.79 \]

\[ P = .50 - .4633 = .0367 \text{ or 3.67 chances out of 100 that the difference isn't significant.} \]

\textsuperscript{47}See the footnote on page 43.
An objection can be seen immediately. There are not enough false items to form a normal distribution so the slightest variation between positive and negative errors will be noted as significant. This is a very real objection and invalidates our findings unless there are a sufficient number tested to form a normal distribution of errors. Therefore, this test for bias is not used unless there is a large number of items tested and a reasonable assurance that the distribution has had a chance to be normal. Correction of the books of record upon an estimation is objectionable even if a bias is definite if it can be avoided, but there is merit in the possible correction of the books if the results can be reached in no other way.

This completes the theory in the solving of our three-statistical problems. Let us compare present policies with our new approach and apply the statistical theory to a practical problem.

---

48 Supra, p. 8.
Present Procedure. The present auditing procedure, outlined below, is as presented by William Bell and Ralph Johns, CPA's, who have written a text\(^1\) that attempts to conform to the principles of present auditing theory as closely as possible. They make the following statements, "In most cases, if good judgment is exercised in the selection of the number and variety of entries to be examined, the auditor will be fairly certain to discover some evidence of any fraud or of material errors that may exist. The justification for testing rests largely upon the probability that irregularities are recurrent; that, once committed, they will be repeated. As has been stated, the process of testing and sampling in auditing practice has been recognized by the courts as reasonably adequate, with the proviso that the tests made must be sufficiently comprehensive to cover all accounts."

They go on to say that if the tests have led to the slightest suspicion of fraud or have indicated the existence of many errors, it may be necessary to make a complete audit, depending on the wishes of the client. In deciding the proportion of items to be tested, consideration should be given

to the probability of there being any irregularity or serious error in the particular records. The extent of tests would vary as to the extent of internal control, i.e., if there is such good internal control over the handling and recording of cash that it would be virtually impossible for anyone to misappropriate cash, the tests can be restricted to a low percentage of the total items. Degree of control over sales and inventory records would be another deciding factor. The known ability and intelligence of the employees would govern the tests also. That is, if clerk A, is intelligent and careful while clerk B is inclined to be negligent, more attention will be paid to B's work than to A's.

The auditor should, of course, be careful not to allow these apparent differences in probabilities to influence him exclusively in planning his work. No part of the usual procedure of the audit may be completely dispensed without an understanding with the client and a qualification of the certificate. Apparent internal control is not always effective; the internal control must be actively effective to lower the extent of tests.

In an average situation, where the conditions are neither especially conducive to nor protection against fraud and serious error, it is thought that from about one-sixth to one-half of the entries in the general-ledger accounts should be examined. The tests should not always be confined
to the entries for certain months. It is usually satisfactory to select from one to four months and examine them on block, that is, every item down to the smallest item. This procedure is sometimes supplemented by a random selection of other items, usually the larger items. The first and last months of the accounting period are those most conducive to fraud, so they are as a general rule examined but supplemented by at least one other month. The extent of the tests should be kept from the client's office force so the testing cannot be nullified by a previous knowledge of what items would be tested.

Critical Examination of Present Audit Procedures. The present auditing procedure may be summed up in the following statement.²

"In most cases, if good judgment is exercised in the selection of the number and variety of entries to be examined, the auditor will be fairly certain to discover some evidence of any fraud or of material errors that may exist."

The fact that if good judgment is used, the errors will be found, cannot be disputed but not every auditor has the "good judgment" to pick out the false entries. It gives no means for the inexperienced auditor to judge the extent of his tests, and it gives the veteran auditor entirely too much

²Procedure as set out by Bell and Johns, op. cit., p. 53.
latitude in his decision. In the "average set of books" from one-sixth to one-half should be tested.\(^3\) Does that mean if there are 600 entries, the auditor can test from 100 items to 300 items and with the exercise of good judgment, find the errors whether he tested 101 or 299? Obviously, he wouldn't have the same percentage of chance of finding the fraud.

From the chart on page 20, the probability figure for 3 false items would be 49% for a 20\(^\text{rd}\) test but 87% for a 50\(^\text{th}\) test. According to the theory the auditor has the perogative of varying his tests that amount and more. It can be seen that the client would be concerned with the probablilities of success as well as the auditor.

The theory\(^4\) also states that the extent of the tests will be governed by the degree of internal control and the efficiency of the employees. That is true but how will the auditor judge the internal control and an individual's efficiency? There has to be a basis for judgment or there would be a splendid opportunity for bias on the auditor and client's part to influence the testing results.

In summary of present-day auditing procedures, it can be seen that the auditor has all but turned his back on the decision as to the extent and variety of entries to be tested.

\(^3\)Supra, p. 54.

\(^4\)As summarized, by Bell and Johns, op. cit., p. 53.
Good judgment is the criteria of deciding whether the test will be one-sixth of the items or one-half of the items. There is no means for -

1. new auditors to judge the extent of their tests or the results of others.
2. the courts to decide whether an auditor is liable for negligence.
3. the client to have confidence in the auditor's findings.
4. outside interested parties to know if the certificate of auditor is justified.

Application of the Statistical Approach. To apply the new approach to auditing cases, two problems will have to be solved.

1. to find the percentage of items that should be tested.
2. to judge the efficiency of the internal control.

In order to arrive at the percentage figure, the following information must be known:

N - the total number of cases
F - the estimated false entries

and the desired level of probability that the auditor will succeed. To obtain N we must know what the universe will consist of. Since in random sampling the items tested should be homogeneous, we must divide our books of account into groups with similar items. As was previously suggested,^5

^5 _Supra_, p. 38.
the breakdown can be into nine homogeneous groups. Each of these groups is a universe. The auditor should number every tenth item in each universe and obtain the total items in each group.

There are three methods of estimating the false items in a universe. First, it is possible to estimate by past experience the average number of errors found in past or similar audits. This method is objectionable since it involves the use of judgment on the part of the auditor. Although it will never be possible to eliminate entirely the use of judgment in the field of auditing, the less reliance that has to be placed on judgment as a criteria of tests, the more standard the procedure can become. The second method would be to decide the amount in dollar value that would be a material error. This amount divided by the average dollar amount of all the items in the universe would give the number of items, which, in the aggregate, would be considered material.

To illustrate, assume that we are testing purchases which in dollar value total $80,000.00. There are 400 entries in the purchase book so the average entry is $200.00. If we considered that a $4,000.00 error would be material, then $4,000.00 divided by $200.00 or 20 entries would be the estimated false entries. Turning to Table V on page 246,

---

6 The table as illustrated shows universes with a maximum of 7 defective items. This table can be extended to include any desired number of defective items.
the most economical sample would be a 15% test or 60 entries. This would give a 95% probability of encountering at least one false entry.

The third method is very closely related to the second method. In universes that the dollar value cannot be applied, a percentage of allowable errors is set up. In the above example if less than 5% of the universe could be acceptably false, then 5% of 400 items or 20 items would be the maximum, so a 15% test would also be desirable.

In summary an auditor should divide his books of account into homogeneous groups, number the items, find the total in each universe and estimate the false entries, preferably by the second or third methods. By referring to Table III on page 20 we can find the percentage of items that should be tested for various probabilities of success. The level of probability that should be reached will be determined by the purpose of the audit, the type of business, the auditor's judgment and the extent of internal control.

The majority of audits will need a probability level from 90 to 99% with 95 - 99% used the most often. If the most economical size of the sample is used, the probability level will always be above 90% if there are 7 false items or more. The economical level may be considered insufficient if internal control is lax, fraud is suspected, or to completely satisfy both the client and the auditor that material
errors will be found.

Using the average dollar value of each item to estimate the number of false entries has a draw-back. There is the possibility that only one item was false but that it was large enough to be a material error. To guard against that possibility the auditor should test all items over a certain amount in addition to the items picked at random. Thus any item that would be a material error in itself should be tested. This raises our probability level but the auditor should not lower the percentage tested at random in order to be doubly confident of success.

The next problem is to judge the efficiency of the internal control.

To judge the efficiency of the internal control, a standard must first be set up to make a comparison. In order to determine the degree of internal control that a business is using effectively, a questionnaire can be drawn up to determine what methods are being used. This will be a general questionnaire that could apply to any business.7 A means to evaluate the answers will have to be worked out. One possible solution would be to place a business in a class if a certain

7Two questionnaires that have been published that could be used for such a purpose are:

percentage of the questions were answered affirmatively. Recognizing that most every business has their own system and that they would be difficult to compare, a further discriminating system can be devised by dividing all businesses into three general classes; large, medium and small. A questionnaire can be drawn up for each class.

The work of evaluating the percentage of questions answered affirmatively will mean the collating of the experiences of auditors over a great number of engagements. The class that the answers put a business into, will determine the level of probability that an auditor will want to reach and the percentage of errors that would be allowable for the internal control the business has or wants. The results of the questionnaires could put the businesses into five classes. The first class would be a very strong internal control with a highly efficient organization. The fifth class a business with little or no control or organization, and the other classes in between. The first class would set a standard with a fairly low probability of success, probably not below 75%, however, and a low percentage of errors that would be tolerated.

The errors will be in two classes, procedural and

---

8 Supra, p. 8.

9 The stronger the internal control is, the less need for a high probability of success.

10 Supra, p. 36.
substantive. The present emphasis is on the substantive error, the amount of error in accounts or statements but the procedural error in determining internal control efficiency should be emphasized. Procedural errors and omissions are the causes for inefficiency in an accounting system, thus allowing the opportunity for substantive errors. The procedural errors and omissions should be noted and later analyzed.

Each class of internal control should have an upper and lower limit. The upper limit is the maximum error that could be tolerated with the alternative of a 100% check if the percentage is higher. The lower limit would be the percentage of error that would be considered efficient for internal control. The exact percentages will have to be worked out over the experiences of many engagements. The likelihood ratio is employed in the method of sequential sampling. A graph is drawn up for the limits of each class for a 90% probability of success\(^1\) (when there is 9 times the probability of the drawn sample being from one of the limits than from the other) so there would be five standard graphs, one for each class\(^2\) of internal control. Let us examine more closely what we expect to learn from the

---

\(^1\)This is an arbitrary probability level for the purposes of illustration.

\(^2\)Supra, p. 51.
sequential graph.

If we have set up a lower limit that would be considered efficient and an upper limit that would be considered intolerable, then we may expect to find one of these two extremes if the results indicate that a decision to accept one of the two limits should be made. The decision would be that the system is either efficient or not acceptable. This immediately changes our method of testing. If the system is efficient, then we have found the results that were desired, **so the testing is complete** (with respect to the internal control). If the system is shown to be intolerable, a **decision to make a 100% check** is indicated. Further conclusions will be discussed but first we must show the effects of testing if a decision to accept one of the limits is **not** reached.

The possibilities that a system will either be efficient or inefficient to a point of rejection is rather small. The great majority of systems will have results of testing that lie in between the two extremes. Consequently, in order to force a decision with the use of the sequential graph, the amount of items tested will have to be very large. This would cause excessive costs in testing and would nullify the benefits of calculating the most economical size of a sample. This writer believes that an alternative procedure can be set up.

Using the most economical size of a sample as a basis
for testing, we can plot the results from time to time on the sequential graph. If at any time during the testing we reach a decision, we will change the method of testing as suggested on page 63. If the results of testing stay in the area on the sequential graph that forces further testing, and we have already tested the number set up for the most economical sample, then we must throw the results out. Instead of using sequential analysis as a test for accounting control, we can assume that our sample is representative of the universe. That is to say, if the results show that the sample is 6% defective, we can assume that the universe is also 6% defective.

To make sure that this assumption is approximately correct, we will determine the percentage of items that should be tested in order to find a 6% defective population. We will then continue our testing up to the new percentage of items. If the percentage of defective items in the sample remains at 6%, we can be reasonably sure that the sample is representative of the universe and that the actual percentage of defective items is approximately 6%. Thus, if our lower limit (the efficient system) was 3%, then we could surmise that the accounting system was approximately 3% inefficient. This should be explained by examining procedural
errors,\textsuperscript{13} that is, analyzing the reasons for each substantive error. Thus, inefficiency in the operation of the system should be reported to the client with an explanation of the cause. The inefficiency should be compared from year to year to check whether the system is improving.

If the client wishes his books corrected on the basis of past error, we must ascertain if there is a definite bias as to the understating or overstating of the accounts. The limitations and the undesirability of correcting the books according to past error has been pointed out previously,\textsuperscript{14} but if the individual conditions justify it, the following procedure is used.

Each type of error\textsuperscript{15} should be considered individually. There must be enough of the same type of error to have formed a normal distribution in order to justify the correction on the basis of past error. The negative errors should be approximately the same as positive errors so we can compare the errors by means of the normal variate to see if the difference is significant. If the normal variate indicates a significant difference, then the accounts can be adjusted for the average error that it is being misstated. The easiest

\begin{itemize}
\item \textsuperscript{13}Supra, p. 36.
\item \textsuperscript{14}Supra, p. 52.
\item \textsuperscript{15}Supra, p. 50.
\end{itemize}
method would be to find the net difference between positive and negative errors and divide this by the total number of errors found. This will give us the average net error. The percentage of defective items to good items found in the testing should be applied to the universe to find the estimated false items for the universe. Multiplying this figure by the average net error will give us the estimated amount of error for the universe. This can be illustrated by the simple example below:

Percentage of number of errors due to failures of equipment. Standard: Equal percentage of positive and negative errors. In examining 1000 accounts payable, 10 errors due to the failure of equipment were found: 8 positive and 2 negative. The total positive errors was $200.00 while the negative error was $20.00. Is the difference between positive and negative errors significant and if so, what is the correction?

\[ S_e = \sqrt{\frac{.5 \times 5}{10}} = .158 \]

\[ \frac{\text{observed difference}}{S_e} = \frac{.5 - .8}{.158} = -1.79 \]

\[ P = .50 - .4633 = .367 \text{ or } 3.67 \text{ chances out of 100 that difference isn't significant.} \]

The average net error is \( \frac{200-20}{10} = 18 \). The percent of defective items found is 1% so if the total universe has 5000 items, the estimated number of false items is 50. 50 times $18 would be the estimated amount of
error or $300 correction.

We have determined the percentage of items to be tested, showed the procedure for choosing those items, with the resulting probability of succeeding in finding material errors in a given defective population, and determined the extent and efficiency of the accounting system. We have further shown the procedure for the possible correction of the books with a minimum of cost through the estimation of the direction of a bias in the errors found. To summarize the statistical approach to sampling as developed, we can take a simple illustration of accounting data which will apply in one example the applications learned.

Example:

We are auditing a medium sized business. We are testing the accounts receivable in order to certify them for the purpose of a proposed sale. In the answering of our questionnaire the answers place the business in class NO. 5, a business with little or no internal control over the accounts receivable. The prospective buyer wishes a 90% probability of success which is agreed upon. The total of the accounts receivable is $40,000. It is agreed that $2,000 would be a material error.

First, we must determine the size of the sample. Numbering every tenth item we find that there are 500 items in the accounts receivable. The average entry would be $80 so the number of estimated false entries would be $2000 ÷ 80 or
25. This would necessitate a 20% check or 100 items tested. We choose the items to be tested by picking our items from a random table.\textsuperscript{16}

Upon testing the 100 items, 12 false entries were found, all were from carelessness. In using a sequential graph of limits from 5% to 8%, the results crossed the upper limit to be rejected. There were 10 negative errors and 2 positive errors. The value of the positive errors was $4,000 while the negative errors were $100. As the percentage allowable was in terms of total value, that is, 8% of $40,000, or $3200, it can be seen that the total value of the errors ($4100) would be greater than the maximum.

We now inform the client that the percentage of errors is greater than can be allowed so there should be a 100% check. The client informs us that the prospective buyer has agreed to adjust the books on the basis of our present findings to determine the actual value of the accounts.

We must find if the difference in positive and negative errors is significant.

\[
S_e = \sqrt{\frac{.5 \times .5}{12}} \approx .145
\]

\[
\frac{.5 - .833}{.145} = 2.3 \quad 2.3 \text{ (area table)} = .489276
\]

\[
P = .50 - .489276 = .010724 \text{ or}
\]

one chance in 100 that the difference isn't significant. The

\textsuperscript{16} Supra, p. 29.
percentage of errors was 12 + 100 or 12%. The average negative error (understatement) was $4000 - $100 + 12 = $325.

If the average error for 100 items is $325, then the total estimated error for 500 items is 5 x 325 = $1615 total negative error. The corrected account would be $40,000 + $1615 or $41,615. In our certification we would have to state on what basis we estimated the false entries with the resulting probabilities of success in finding a material error. We should state the class of internal control (number 5) and the result efficiency or inefficiency (inefficiency of approximately 5%). We should state how much we corrected the accounts receivable with the basis for so doing. The reasons for the inefficiency should be pointed out in the certificate, in this case carelessness. The errors show that Employee X has a tendency to be careless in his extensions, with large errors occurring. There wasn’t sufficient internal control to discover the errors or to check the total of accounts receivable.

By using this method of standard procedure, it is possible for the client, the interested party and the auditor, to understand the procedure, the extent and the results of the audit. The auditor will have a basis for defending his procedure should it come to a test. The client can see clearly the degree of internal control that his accounting system has and the efficiency with which it is operating.
We have adjusted the books to the satisfaction of both parties with a minimum of cost. We have solved the three problems that we set up for statistical sampling in the introduction.17

**Conclusions.** We have compared the present auditing procedure of testing with the new method of statistical sampling. We have found that the present procedure sets little basis for estimating the amount of testing necessary to make the audit adequate or to have confidence in the results other than reliance upon the auditor's personal judgment. It is a sadly neglected part of auditing procedure but is showing a recent trend in recognizance of the problem. It is admitted that a standard procedure can never replace judgment entirely. It provides, however, a tool in the accountant's hands that is backed by mathematical reasoning and has not deprived the accountant of anything he has had before. It should be cautioned that statistical formulas are based on probabilities, not certainties. We can only be certain if there is a 100% test and then only if the auditor finds the false entry upon its examination. That our entire theory rests on the assumption that the false entry will be discovered if it is examined should be emphasized strongly.

---

17Supra, p. 8
That is the situation in which judgment will still play a big part in the audit. It is true that where one auditor can find no evidence for fraud or false entries, another auditor can locate them at a glance. The theories are based on a theoretically perfect procedure of sampling and a normal distribution. The perfect results will seldom be attained but the results can be useful, practical and based on a sound foundation.

We can determine a size to be sampled that all auditors should use for a business, determine the degree and efficiency of the internal control and make a correction of the accounts according to past error. The field is new but potentially powerful. The change to a statistical basis can only be done with a close cooperation of accountants and statisticians in relating their fields and a cooperation of accountants in collating their experiences in the audits. Those are two big steps to be taken but they should and will be taken in the near future.
BIBLIOGRAPHY


