1974

Auditor's Approach to Statistical Sampling, Volume 6. Field Manual for Statistical Sampling

American Institute of Certified Public Accountants. Continuing Professional Education Division. Individual Study Program

6

AN AUDITOR'S APPROACH TO STATISTICAL SAMPLING

FIELD MANUAL FOR STATISTICAL SAMPLING

Individual Study Program
Continuing Professional Education Division
American Institute of Certified Public Accountants

AICPA Continuing Professional Education

NOTICE TO READERS

This manual is a publication of the staff of the American Institute of Certified Public Accountants and is not to be regarded as an official pronouncement of the Institute. It was prepared by the members of the Committee on Statistical Sampling.

AN AUDITOR'S APPROACH TO STATISTICAL SAMPLING

<u>Volume 6</u>

FIELD MANUAL FOR STATISTICAL SAMPLING

Individual Study Program

Continuing Professional Education Division

American Institute of Certified Public Accountants

FIELD MANUAL FOR STATISTICAL SAMPLING

CONTENTS PAGE

PREFACE

Previous volumes of An Auditor's Approach to Statistical Sampling present the fundamentals of several statistical sampling techniques that are useful to auditors. These volumes are training documents, prepared in programed-instruction format. They deal with the technical aspects of statistical sampling in auditing, but do not discuss the audit related aspects fully. The reader is referred to Section 320 of Statement on Auditing Standards No. 1 for further discussion.

Volume 6 is a field reference manual, intended to serve as a natural extension of its predecessors. Volume 6 alone will not suffice to train the auditor and should not be used for that purpose. For each subject introduced in the training volumes, Volume 6 contains a chapter that generally presents a technical summary of the subject, followed by one or more illustrative case studies. The technical summaries are intentionally brief. The primary usefulness of this volume lies in the case studies. These case studies utilize the worksheets and tables, as applicable, contained in Appendices A and B respectively. The worksheets in Appendix A are blank, so they may be photocopied by the auditor for use on audit engagements. If legal-sized paper is used, there will be room for headings.

The cases used in this volume are illustrative of good practices in the circumstances described and are intended only to provide guidance. Thus, it is the auditor's clear responsibility to judge the practices which are appropriate in each situation that confronts him.

Appendix C contains listings for computer time-sharing programs that perform certain of the statistical functions described in this volume. They may be given to a time-sharing vendor of the user's choice, to be entered into his library for continuous application. These programs are all conversational in nature. Their use is illustrated in Appendix C. There are other time-sharing programs that pertain to statistical sampling which are not included in this volume but can be equally useful to the auditor.

This book is offered to members as part of the Institute's continuing professional education program.

Robert E. Schlosser, PhD, CPA, Director
Continuing Professional Education Division

Chapter 1
SELECTING A RANDOM SAMPLE

The body of data the auditor is concerned with is called the population. Statistical sampling permits the auditor to estimate some characteristics of the population by selecting a sample rather than by examining every item in the population.

In order to use statistical techniques to evaluate results, it is necessary to select the sample in accordance with sound statistical principles. In this chapter, attention is confined to a simple method - unrestricted random sampling without replacement; namely, an item once selected is not replaced in the population for possible reselection.

In this method, each element of the population is afforded the same chance of selection, and in addition each possible sample of a particular size has an equal chance of being selected. These two characteristics would also hold if the sample were selected with replacement. The "with replacement" method is used for teaching purposes only. In practice the more practical "without replacement" method is used.

A random number table or a computer time-sharing program may be used to help select an unrestricted random sample. To use either, it is necessary to establish correspondence between (1) the elements in the population and (2) either the digits in the random number table or the computer program limits. If a table is used, it is also necessary to establish a starting point and to specify a route within the table.

It is important to understand that once a sample element is selected according to the sampling plan, it cannot be discarded or replaced but must be followed to a conclusion in order to make the statistical inferences valid. Prior to selecting his sample, therefore, the auditor should exclude any elements from the population which are irrelevant to the purpose of the sample or are of such large dollar value that 100% examination is desired. It may also be desirable to divide the population into two or more groups which are sampled separately. This procedure, called stratification, is explained in Chapter 4. Stratification may be used to reduce the sample size required to achieve the sampling objective or to develop conclusions about various segments of the population.

The application of the principles of unrestricted random sampling is illustrated in Case Study 1.

Case Study 1

Selecting a Random Sample

Background

The BIG COMPANY operates on a voucher system for disbursements from its main bank account. All checks must be supported by a voucher. The details of the voucher vary, depending on the nature of the item being paid; e.g., purchased materials, payroll account reimbursements, and monthly recurring items such as rent and utilities. Since this is a highly seasonal business, the amounts of many regular expenditures (payroll, for example) vary significantly from month to month even though their number may not. Certain kinds of checks, however, (purchased materials, for example), do diminish in number during the slack season. Checks are prenumbered in a continuous series of numbers, but vouchers are numbered by typewriter as they are written, with a new series beginning each month. All January vouchers begin with the prefix "1," February, "2," etc., through December, which begins with "12." Each year the numbering system starts over, but a new color of stationery is used.

Vouchers originate in different sections of the accounting department, depending on the type of item being paid. Each section is assigned a batch of numbers; if all are used, additional numbers are assigned. The vouchers follow different routings because different kinds of documents and approvals are needed for different types of expenditures. All vouchers come to the check-writing station. There the check number is typed on the voucher, and the voucher is filed by voucher number.

In evaluating the routine supporting the writing of checks, you, as BIG COMPANY's auditor on the engagement, decided to examine a sample of vouchers, rather than all vouchers written. A sample size of 300 was specified, representing about 2% of the total number of vouchers written during the year. (The method used to specify a sample size of 300 is not relevant to this case.)

You tentatively decided to sample from the population of all vouchers written during the year. You initially considered block and random selection techniques as alternatives. Block sampling, however, was discarded in order to obtain a valid inference about population for the entire audit period. (Consistent with Section 320 of SAS No. 1.)

Random Selection

To select a random sample, you considered the possibility of just reaching into the filing cabinet drawers and boxes where the vouchers are filed, and selecting 300 items haphazardly. You decided against this because you realized that you could not select truly at random. Some vouchers are physically thicker than others because of the number of documents attached. These, you felt, would be more likely to "find your fingers." You also feared that you would have a tendency to reach near the middle of the drawer, and as a consequence any voucher at the front or back of the drawer would not be as likely to be selected. The fact that fifteen thousand vouchers were filed in a number of drawers and boxes at various locations would also influence your selection. And you realized that if any vouchers were not filed in these boxes and drawers, you would have no chance at all of including them in your sample by using this technique. Therefore, you decided to obtain 300 random numbers and select vouchers bearing these numbers.

Random numbers are available from several different random number tables, including those in the Supplementary Section of Volume 1. Also, a number of computer programs have random number generators which will select the desired quantity of random numbers from a specified range (see Appendix C on computer time-sharing); many computer routines also have a feature that will arrange these numbers in ascending order for ease of pulling the vouchers.

Random number tables are typically set up so that the numbers are in a single series. For practical reasons in this situation, you need numbers from a different series for each month. At first you considered using the random number tables, with the following plan:

1. Select a two-digit random number; if it lies in the range from 1 to 12, use it as the month designation. Otherwise, discard the number.

2. Select a four-digit random number to represent the voucher number within the month every time a "hit" is made in step 1. If this random number is one of the numbers used during the month, the combination of month number and voucher number represents a voucher to be examined. Otherwise select a different voucher number to go with the selected month number, repeating until a "hit" is made.

However, you discarded this plan because of its inefficiency. First, only twelve percent of the numbers you chose from the random number table made a hit in step 1, and second - even when you did find a hit in step 1 - most of the numbers you looked up in step 2 had to be discarded as being numbers not used. There certainly were less than 9999 numbers used, even in the heaviest month. To achieve efficiency, you adopted a new plan for step 1. If the random number was any one of 1 through 8, the month was taken to be January; if any one of 9 through 16, February; and so on, with 89 to 96 representing December, and only 97 to 00 being discarded. This increased efficiency, and you were considering a similar plan for selecting vouchers within a month, when you realized that even if you did achieve a high degree of efficiency, <u>this plan would not be random</u>. Each month did have an equal opportunity of being chosen, but because some months had fewer vouchers than others, an individual voucher in a light month would have a greater chance of being selected than would an individual voucher in a heavy month. Because once a month had been selected at random, you would be bound by the plan to select a voucher from that month. You realized that you could overcome this weakness by discarding the number selected in step 1 if you didn't achieve a hit on the first selection of a number in step 2. After making adjustment for this, your plan would be satisfactory from a probabilistic viewpoint; however, you feared it would be extremely inefficient.

Several plans came to mind for finding an efficient set of random numbers that would be truly random, with each voucher having an equal chance of being selected. One was a combination of the two steps of the above plan. In this revised plan, you looked up a single digit random number of six digits. The first two were to represent the month, as you had decided earlier, and the next four digits were to represent the voucher number within the month. Since there were not more than 3,000 vouchers in any month, the third digit was to stand for the "thousand" series. When the number was 1, 2 or 3, the "thousand" series was the first; when the number was 4, 5 or 6, the "thousand" series was the second, and so on. The conversion would not be overly complicated, once it was started.

Another plan was to renumber each voucher issued throughout the year into a single series, and to use the random number tables in the ordinary manner. Long before you found a consecutive numbering machine to renumber each of the 15,000 vouchers, you had discarded the plan of physically placing the number on the voucher. You did, however, devise a plan for renumbering each voucher by a worksheet method. Since there were 2,116 numbers allocated for January vouchers and 1,007 for February vouchers, you set up a table as follows:

Month	Orignal Voucher Number	Subtract From Random Number
January	1-1 to 1-2116	zero
February	2-1 to 2-1007	2116
March	3-1 to 3-1887	3123

With this simple table, you would subtract from each random number selected the highest number possible from the right-hand column of the table. By reference to the table, you could tell in which month you were to find the voucher, and the remainder from the subtraction would tell you the number of the voucher to use in that month. For example, when you drew the number 3000, you subtracted 2116 and used voucher 2-884. You were still a little fearful, because you knew that with the block system of assigning vouchers to the different divisions of the accounting department, there might be some unused vouchers toward the end of the month. You decided to use this method anyway, planning that if you did hit one of these unused numbers, you would discard the random number selected and go on to the next one. To cover this contingency, you planned to look up 400 random numbers and to use the vouchers from the first 300 "hits" obtained.

This feature of looking up a few extra random numbers, "just in case," requires a word of warning. Typically, the random numbers are rearranged in sequence for ease in pulling the vouchers. If the numbers are generated by a computer program, the initial print-out may show them in ascending sequence. Then, if the first 300 numbers are all that are needed, instead of all 400, there is almost certainty that the first 300 in ascending sequence are not the first 300 drawn. Perhaps this sounds like overrefinement. But if the highest one hundred numbers are discarded, the high end of the file is being exempted from examination; this is of course improper. Therefore, the order in which the numbers are acquired must be retained so that the proper ones may be dropped if not needed. Many computer programs have the capacity to print the selected random numbers in both orders. Often, auditors simply will examine all sample items selected because it is easier than reducing the sample size.

Knowing that your method of selecting the vouchers was not only sound but also efficient, you were ready to proceed when you were troubled by another point which you realized should have been considered at the outset. Did you have only a single population from which you were selecting items, or did you have several populations? You were not worried about the boxes versus the filing cabinets in which the vouchers were filed. You were worried about the several sections of the accounting department, each of which prepares a different kind of voucher. The routines in each section are different because different types of data are being handled. In addition, the people are different. Is it possible to miss entirely one type of voucher in the random selection process? Since it is possible, even though unlikely, should additional steps be taken to assure that one of each type is selected? There are only 52 weekly payroll vouchers prepared during the year. Referring to attribute sampling tables, you can see that from among 15,000 vouchers, the probability of selecting one of these 52 with a random sample of 300 is about 60 percent. (This percentage is determined from the discovery sampling tables given in Volume 2, Revised.) This illustrates the importance of careful planning to assure that all of the auditor's objectives are satisfied by the sampling plan.

In summary, this case illustrates the type of problems and considerations which should be kept in mind when in selecting a random sample:

1. A random sample must be obtained in order to use any of the evaluation methods described in Volumes 1 through 5.

2. Once a random sample is obtained, however, it may be evaluated by using any or all of these methods.

3. Regardless of the details of the plan used to obtain a random sample, it must have the characteristic that, at the point of selection, each item in the population has an equal opportunity for selection.

4. Finally, proper documentation of the methods followed for selection, and contraction or expansion of random samples should be maintained.

Chapter 2
ATTRIBUTE SAMPLING

Estimation Sampling

One type of sampling for attributes involves estimating the frequency of occurrence of a specified characteristic in the population. This population occurrence rate is estimated from the sample occurrence rate.

The sample estimate is evaluated in terms of its precision and reliability. The auditor may choose to calculate either the precision interval or only the upper precision limit. In either case, the calculation is made at a specified reliability level.

When the auditor's objective is to estimate the population occurrence rate, he will calculate a precision interval and utilize both precision limits. When he wishes to decide whether the population occurrence rate exceeds some tolerable limit, he will make an estimate and concern himself only with the upper precision limit.

Tables permit the auditor to determine the sample size required to achieve a desired precision interval at a specified reliability level.

To use the tables, the auditor is required to specify an anticipated sample occurrence rate. This rate will often be based on his previous experience, and should be conservative (high).

By specifying the anticipated sample occurrence rate, the desired upper precision limit, and the reliability level, the appropriate sample size can be determined from Tables 1 and 2 in Appendix B. The closer the anticipated rate is to the desired upper precision limit, the greater the sample size will be.

The tables in this volume provide a slightly conservative approximation. When the sampling fraction (n/N) exceeds 10%, the tables will yield a more conservative evaluation; that is, the actual upper precision limit is lower than the table indicates. The use of exact tables would gain little in precision but would substantially proliferate the number of tables needed.

The tables are based on unrestricted random sampling, so the auditor must use that or an equivalent method with these tables. Furthermore, it is important to identify carefully the characteristics being measured, so the auditor can readily determine whether the sampling unit possesses them. If, for example, the sampling unit is an invoice, and the audit objective is to estimate the proportion of invoices with pricing errors, then multiple occurrences of the particular error on a single invoice constitute only a single occurrence of the designated attribute for the sampling unit.

To use the evaluation table (Table 3 in Appendix B), the auditor specifies the reliability and the observed number of occurrences for the sample size involved. He then reads the achieved upper precision limit from the table.

In such sample evaluation, the reliability indicates the confidence that the population occurrence rate is no larger than the achieved upper precision limit. The auditor's risk of being incorrect in this decision is equal to the complement of the reliability.

Discovery Sampling

Discovery Sampling is a type of attribute sampling, although the objective is not to estimate an occurrence rate. Rather, the basic objective of discovery sampling is to provide a sample size large enough for the auditor to have a prescribed chance of seeing at least one occurrence of some designated attribute, provided the population occurrence rate exceeds some specified level.

Discovery sampling is used when the auditor believes that the population occurrence rate is quite small, and he wants the sample to be large enough so that he will observe at least one occurrence in the sample with some desired probability when the occurrence rate exceeds some tolerable level.

Regardless of the objective in determining the sample size, the auditor may evaluate the results of a discovery sample in exactly the same manner as in any other attribute sample. If he audits each item in the sample, he can establish the upper precision limit at a specified reliability level based on the number of occurrences observed in the sample. If the sole objective is discovery, however, the auditor may stop auditing the sample items as soon as he finds an occurrence.

The appropriate sample size required to achieve the auditor's objective in discovery sampling is determined from Table 4 in Appendix B.

To use Table 4, the auditor must specify the critical occurrence rate and the desired probability of discovery. The critical occurrence rate is defined as the rate in the population at which he desires the specified probability of seeing at least one occurrence in the sample. If the population occurrence rate is below the specified critical occurrence rate, the chances of an occurrence in the sample will be smaller than the specified reliability level.

If the auditor does not observe any occurrences in the sample, the achieved upper precision limit will coincide with the critical rate of occurrence when the evaluation is made at the same reliability as the desired probability of discovery.

Case Study 2

Use of Attribute Estimation in Compliance Tests

Background and Objectives

You have made a preliminary review and evaluation of the BETTER PRODUCTS COMPANY's system of internal control, and based on this, you have designed a tentative audit program. Your next step is to perform compliance tests in order to confirm or challenge your evaluation. If these tests support your evaluation, you will proceed with the audit as planned; if they do not, you will alter the nature, extent or timing of your planned substantive tests, as appropriate.

A major area of compliance testing is cash disbursements. Monthly, the Company processes an average of 360 vouchers, paying 600 invoices. Each voucher contains a copy of the check along with supporting documentation, such as vendor's invoices, purchase orders, and receiving reports.

You intend to examine a sample of vouchers to evaluate compliance with several processing controls in which you are interested. These tests are set forth in your audit program as follows:

Each voucher selected (there may be several invoices supporting one voucher) should be examined for:

1. The existence of purchase orders, vendor's invoices, expense reports, receiving reports, and other supporting data.
2. Evidence of proper authorization for the expenditure.
3. Evidence of approval of all supporting documents; for example, purchase order, invoice, and receiving report.
4. Evidence that prices on the invoices were compared with a purchase order or other

authorization for the acquisition of goods or services.

5. Evidence of clerical check.

6. Approval of account distribution.

7. Cancellation of supporting documents to prevent their reuse.

8. Proper disposition of voucher; for example, by examination of the cancelled check.

In planning substantive tests, you presumed that compliance with these controls was about the same as in the preceding three years. In each of those years you sampled 50 vouchers, with the following results:

	Number of Occurrences		
	19X0	19X1	19X2
Lack of receiving report	2	-	1
Lack of approval	1	1	-
No evidence of clerical check	1	-	2
Unapproved account distribution	-	2	-

From this, you expect a noncompliance rate of about 2% for steps 1, 2, 5 and 6, and complete compliance with all others.

You realize, however, that the occurrence rate of a sample is only an indication of the true population occurrence rate, and not a perfect representation, whether that sample is judgmental or statistical. As a consequence, when you planned your substantive tests, you allowed for noncompliance at a higher rate. Specifically, after consideration of a number of related audit steps, you decide you can tolerate an upper precision limit for the noncompliance rate of 6% for controls 1 through 6 and 3% for controls 7 and 8. If actual results are within these desired levels, no alteration of your year-end audit plans will be made.

Approach and Determination of Sample Size

In this year's audit, you decide to use the attribute estimation statistical sampling method. Each of the controls outlined above will be considered an attribute. The voucher is the sampling unit, and the population consists of the 4320 vouchers issued during the year.

These data are entered on Work Sheet 16 and used in conjunction with Table 1 or Table 2 to determine sample size. A copy of Work Sheet 16 prepared to this point follows. Note that a reliability of 95% is used. A high level of reliability is elected because compliance tests, as part of your review and evaluation of internal control, serve as the foundation for the greater part of the audit yet to come.

Tables 1 and 2 are not precise, and indeed, they need not be. For example, Table 2 (95% reliability) indicates that a sample of 150 is required for an occurrence rate of 2% and an upper precision limit of 5.1%, rather than the 6.0% as originally specified. Also, the table is based on an infinite population size whereas our population numbers only 4320 vouchers. However, the purpose of sample size determination is to give us a sample size that is adequate but not necessarily exact. This is further indicated by the fact that estimated occurrence rates cannot be exact. Accordingly, you decide on a sample of 150 vouchers as a reasonable and conservative approximation of the sample size relative to a 6% upper precision limit.

ATTRIBUTE SAMPLING

Summary Worksheet

Description of population:

Vouchers issued during year.

Size of population:

4320

Description of an occurrence:
(Frequently in evaluating sample items, the auditor is making several different tests. On a separate work sheet, describe "an occurrence" for each of these tests.)

	Occurrence Type 1	2	3	4	5	6	7	8	9	10	11	12
Auditor's decisions												
a. Anticipated occurrence rate	2	2	0	0	2	2	0	0				
b. Specified upper precision limit	6	6	6	6	6	6	3	3				
c. Specified reliability (confidence level)	95%	95%	95%	95%	95%	95%	95%	95%				
Table Reference (Table 1, 2 or 4)												
d. Required sample size	150	150	50	50	150	150	100	100				
Actual sample size												
e. (Consider using largest value on line d)												
Audit results												
f. Number of occurrences in sample												
g. Actual upper precision limit (Table 2)												

Selection of Sample and Evaluation of Results

Since you can examine the same set of vouchers for each attribute, you decide it would be easiest to use a sample size of 150 for every attribute, as suggested on Work Sheet 16. Accordingly, you choose 150 random numbers to identify voucher numbers, pull the vouchers, and perform your tests.

The only errors you find are in the test of approval of account distribution (attribute 6). Of the 150 vouchers examined, account distribution was not approved on 8. These results are also entered on Work Sheet 16 to provide a sample evaluation. Completed Work Sheet 16 follows.

ATTRIBUTE SAMPLING

Summary Worksheet

Description of Population:

Vouchers issued during year.

Size of population:

4320

Description of an occurrence:
(Frequently in evaluating sample items, the auditor is making several different tests. On a separate work sheet, describe "an occurrence" for each of these tests.)

	Occurrence Type 1	2	3	4	5	6	7	8	9	10	11	12
Auditor's decisions												
a. Anticipated occurrence rate	2	2	0	0	2	2	0	0				
b. Specified upper precision limit	6	6	6	6	6	6	3	3				
c. Specified reliability (confidence level)	95%	95%	95%	95%	95%	95%	95%	95%				
Table Reference (Table 1, 2 or 4)												
d. Required sample size	150	150	50	50	150	150	100	100				
Actual sample size												
e. (Consider using largest value on line d)	150	150	150	150	150	150	150	150				
Audit results												
f. Number of occurrences in sample	0	0	0	0	0	8	0	0				
g. Actual upper precision limit (Table 3)	2	2	2	2	2	9 1/2	2	2				

The tests performed indicate all controls except approval of account distribution are being complied with adequately. The sample evidence enables you to be 95% confident that approval of account distribution errors may be occurring at a rate as high as but not exceeding 9.42%. Your audit plan, however, can tolerate no more than a 6% error rate. Therefore, you must alter those substantive procedures affected by this type of error in the system of internal control.

A possible reaction to this kind of result is to consider expanding the sample as it relates to attribute 6, in hopes that the upper precision limit will be reduced to an acceptable level. For example, Table 3 indicates that a sample of about 300 (double our present sample of 150), with 8 errors, will have an upper precision limit of 4.76%. This approach is not advisable because if we found 8 errors in the first 150 items examined, we would also expect to find approximately 8 errors in the <u>next</u> 150 items examined, bringing our upper precision limit down to only about 8.0%. Chances of obtaining no additional errors are extremely small. Since internal control in this area cannot be relied on to the extent expected, other work must be expanded. Furthermore, this procedure involves sequential sampling evaluation methods not covered in this volume.

Conclusion

Compliance tests are a very important phase of most audit examinations. Attribute estimation is particularly suited to these tests. Further, since attribute estimation involves the use of tables rather than formula, it is not difficult to apply.

Case Study 3

Use of Attribute Sampling with Positive Confirmations (For Purposes of Discovery)

Background and Objectives

Your client is a mutual fund. You have completed your review of internal control and performed tests of compliance and other tests. All results have been satisfactory. The fund has a computerized system that processes a high volume of transactions. One of the areas for which it is used is shareholders' accounts - a record of the number of shares each holder owns. Your judgment at this point in the audit is that the shareholders' accounts are accurately stated; however, you will corroborate this preliminary judgment by sending confirmation requests to certain shareholders selected by statistical sampling techniques. Since you believe the system is excellent, and you expect to find no errors, you elect to use attribute sampling for purposes of discovery.

Your objective is to obtain additional corroboration of the accuracy of shareholders' account balances. Since the discovery of a single error could have great significance in appraising the effectiveness of such a highly structured system, you decide on a critical occurrence rate of 0.5%. Because the system is evaluated as excellent, you decide to use an 80% probability of discovery.

The fund has 20,000 shareholders and 14,000,000 shares, for an average of 700 shares per shareholder. No one shareholder has more than 100,000 shares.

You believe shareholders are vitally interested in the accuracy of their accounts, and will treat the requests accordingly. Your experiences in prior years support this assumption.

The company's experience in sending proxy requests to stockholders, however, indicates the initial response rate is not high enough to justify the use of negatives, as stipulated in Section 331.05 of SAS No. 1. So you decide to use positive confirmation requests.

Determination of Sample Size and Evaluation of Results

Given the above factors, you refer to Table 4 and find that a sample of 320 items will provide an 80% probability of including at least one error if errors occur in the population at a rate of 0.5%. You obtain a random sample of 320 items and send positive requests. You send follow-up requests at a later date and perform alternative procedures to accounts of nonrespondents. Ultimately, you become satisfied that each of the 320 account balances is accurately stated. Consequently, your audit objective is met.

Chapter 3
MEAN-PER-UNIT ESTIMATION

In this chapter, we will cover estimating the total dollar value population characteristic, and the statistical technique we will use is mean-per-unit estimation. That is, we compute the sample mean, and multiply this by the number of items in the population. This technique is appropriate in auditing either when a book value for each population item is not available or when the footed total of the book value cannot be relied on. Mean-per-unit estimation is seldom used without stratification, and generally extreme values are audited 100%.

To assess the accuracy of a statistical estimate, we compute the precision and the reliability of the estimate. The precision is a measure of the maximum difference between the estimate and the true population value. Reliability expresses the frequency with which this difference will not exceed the precision. In other words, with a reliability of 95%, we expect 95% of all possible samples to result in the maximum difference between the sample estimate and the true population value that is equal to or less than precision.

Frequently, the auditor will express these results by using a precision interval. This interval has a lower endpoint given by the estimate minus the precision and an upper endpoint given by the estimate plus the precision. The precision interval will contain the true population value with a frequency equal to the reliability. That is, among all the possible samples, the fraction represented by the reliability has the property that the precision interval contains the true population value.

A key determinant of the appropriate sample size where mean per-unit estimation is to be used is an estimate of the population variability. The population variability is measured by the standard deviation. One method of obtaining this estimate is to select and evaluate a preliminary sample. Although some auditors use as few as 30 items for this purpose, we generally should use substantially more. This preliminary sample will ultimately become part of the final sample.

The population standard deviation can be estimated from the sample data by means of the formula

$$s_{X_j} = \sqrt{\frac{\sum (x_j - \bar{x})^2}{n-1}}$$

or, as expressed in a form more suitable for computation,

$$s_{X_j} = \sqrt{\frac{\sum x_j^2 - n\bar{x}^2}{n-1}}$$

The sample mean, $\bar{x}$, is an estimate of the population mean $\bar{X}$. Multiplying the sample mean by N, the number of items in the population, the quantity $N \cdot \bar{x}$ is a sample estimate, $\hat{X}$, of the total population value $N \cdot \bar{X}$.

If we were to choose a large number of different samples from the population, and compute a sample mean for each of them, we would observe that although these sample means would vary, there would be a tendency for them to cluster and to approximate the normal distribution. The variability of the sample means is measured by the standard error of the mean, and the center of all the possible sample means is located at the population mean.

From the sample, we can estimate the standard error of the mean as

$$\hat{\sigma}_{\bar{x}} = \frac{s_{X_j}}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

that is, the estimated population standard deviation divided by the square root of the sample size, multiplied by the finite population correction factor. The factor $\sqrt{1 - \frac{n}{N}}$ occurs because the sampling is done without replacement. In cases where the population is large as compared to the sample size, this factor has little effect. Using statistical tables, we can calculate the chance that a sample mean will differ from the population mean by less than some multiple of the standard error of the mean. For example, the chance that $\left| \bar{x}-\bar{X} \right|$ is less than one multiple of $\hat{\sigma}_{\bar{x}}$ is about 68%, where $\left| \bar{x}-\bar{X} \right|$ is the difference between the sample mean and the population mean, disregarding the algebraic sign.

Using these basic statistical principles, we define the precision of our estimated total value as

$$A = \frac{S_{X_j} \cdot U_R \cdot N}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

where $\frac{S_{X_j}}{\sqrt{n}} \cdot \sqrt{1 - \frac{n}{N}}$ is our estimated standard error of the mean, U_R represents the factor (i.e., the Normal Deviate from Table 5) associated with a reliability of R%, and N is the number of items in the population. Manipulating this formula, we also obtain

$$n = \frac{S_{X_j}^{\,2} \cdot U_R^{\,2} \cdot N^2}{A^2 + S_{X_j}^{\,2} \cdot U_R^{\,2} \cdot N}$$

This formula gives the sample size required to achieve, at a reliability corresponding to U_R, a desired precision around an estimate of a total population value. From this formula, we observe the necessity of obtaining an estimate of the population standard deviation.

After obtaining the estimate S_{X_j}, we use the formula to determine the required sample size, n. We suggest adding some items to the determined sample size to compensate for the fact that we have estimated only the population standard deviation. We suggest 10% or more.

After selecting at random the items for the sample, and after establishing an audited value for each of them, the auditor can evaluate the results. The estimate of total value of the population is $\hat{X} = N \cdot \bar{x}$ and the achieved precision of this estimate is calculated by means of the previously stated formula. The statistical conclusion is that the true population total is contained in the interval $\hat{X} - A$ to $\hat{X} + A$ with a reliability of R%.

Case Study 4

Mean-Per-Unit Estimation

Background and Approach

Your client, ABC Leasing Corporation, leases trucks and automobiles both on a fleet basis (with each vehicle on a separate lease) and on an individual basis. For internal purposes, the Company follows the practice of computing interest earned on leases on a straight-line basis. For statement purposes, it makes an adjustment to accrue additional interest income on a "rule of 78's" basis. This is a very material adjustment, and one of your major audit objectives is to determine that the amount is computed accurately.

The crux of the rule of 78's adjustment is average vehicle lease life, which the client contends is 35 months. Of course, there is no way of directly establishing the lease life of a vehicle on the road, but this can be estimated by observing the time period of terminated leases. You decide that an appropriate approach is to determine the average life of all leases terminated during the past two years, and to compare that average to the 35-month period as a test of its reasonableness. The appropriateness of this method is based on the fact that the number of vehicles leased and the rate of turnover have remained reasonably stable over the past several years.

First, you determine the following characteristics about the population.

1. There are 15,000 leases currently active.
2. There were 14,000 leases terminated during the two-year period.
3. Data concerning the leases terminated are available on an EDP listing consisting of 280 pages with 50 items (lines) per page.
4. There are no unusual lives shown on the EDP listing.

Second, you select a preliminary sample of fifty items from the EDP listing, using a computer time-sharing-based random number generator and specifying the following parameters:

1. Numbers are to be five digits, broken into two sets of three digits and two digits, respectively.

2. The parameters for set one are 1 through 280, representing the number of pages.

3. The parameters for set two are 1 through 50, representing the number of items on a page.

Third, you locate the 50 items on the client's listing and record the vehicle number, lease date, termination date, and lease life. These data are verified by tracing to appropriate lease files.

Evaluation of Results of Preliminary Samples

The details of the 50 preliminary sample items are summarized, and Work Sheet 2 is used to compute the mean value of the sample and the estimated standard deviation of the population. You will use this data to determine final sample size.

MEAN-PER-UNIT ESTIMATION

Estimated Standard Deviation of Items in Population

$$S_{X_j} = \sqrt{\frac{\sum x_j^2 - n\bar{x}^2}{n-1}}$$

☒ Preliminary sample
☐ Total sample

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	$\sum x_j$	Sum of all sample items		1,131
2	$\sum x_j^2$	Sum of squares of all sample items (square before summing)		35,341
3	n	Sample size		50
4	$n-1$		③ - 1.0	49
5	$\bar{x}$	Mean	① ÷ ③	22.62
6	$\bar{x}^2$		⑤^2	511.66
7	$n\bar{x}^2$		③ x ⑥	25,583
8	$\sum x_j^2 - n\bar{x}^2$		② - ⑦	9,758
9	$\frac{\sum x_j^2 - n\bar{x}^2}{n-1}$	Variance	⑧ ÷ ④	199.14
10	S_{X_j}	Standard deviation of sample items and estimated standard deviation of population	$\sqrt{⑨}$	14.11

A close inspection of the results on Work Sheet 2 produces a surprise - the average life of these 50 leases is only 22.62 months. This is so far below the client's estimate of 35, it seems a certainty that his figure is too high. (In fact, if you were to complete Work Sheet 4 to obtain a one-sided estimate from this preliminary sample, at a reliability level of 95%, you would conclude that the true average lease life does not exceed 25.8 months.) You discuss the possible reasons for this with the client; i.e., reasons why current vehicles will be leased longer than they have been in the past. Your judgment is that no such valid reasons exist, considering the stable nature of the client's business. (Had there been substantial growth or change in vehicle mix, your approach would have been different.)

As a result of this finding, you revise your audit objective from that of supporting the client's estimate to that of making a precise enough estimate of your own to determine the amount of the adjustment needed. You elect to do this by making a two-sided estimate at a reliability level of 95% and with a precision of 1.5 months. (Precision expressed in total months will be 1.5 months x 14,000 leases or 21,000 lease months.)

Determination of Final Sample Size

Using these criteria, you determine final sample size with Work Sheet 3. A safety factor is not considered necessary in this situation because, if desired precision is not achieved, it will not be difficult to expand the sample.

MEAN-PER-UNIT ESTIMATION
NON-STRATIFIED

Determination of Sample Size from Estimated Standard Deviation

$$n = \frac{S_{X_j}^{\ 2}\ U_R^{\ 2}\ N^2}{A^2 + S_{X_j}^{\ 2}\ U_R^{\ 2}\ N}$$

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	A	Precision desired	Auditor's decision	±21,000
2	R	Reliability desired ☐ one-sided estimate ☒ two-sided estimate	Auditor's decision	95%
3	$U_R^{\ 2}$	Square of reliability factor	Table 5	3.84
4	N	Number of items in population		14,000
5	S_{X_j}	Estimated standard deviation	Work Sheet 2, line (10)	14.11
6	$S_{X_j}^{\ 2}$		Work Sheet 2, line (9) = (5)^2	199.14
7	$S_{X_j}^{\ 2}\ U_R^{\ 2}\ N$		(6) x (3) x (4)	10,705,758
8	$S_{X_j}^{\ 2}\ U_R^{\ 2}\ N^2$		(4) x (7)	149,880,612,000
9	A^2		(1)^2	441,000,000
10	$A^2 + S_{X_j}^{\ 2}\ U_R^{\ 2}\ N$		(9) + (7)	451,725,758
11	n	Sample Size	(8) ÷ (10)	332
12	final n	Sample size with safety factor	(11) plus safety factor (auditor's decision)	N/A

Evaluation of Final Results

Next, using the time-sharing program to generate random numbers, you select 282 additional items for examination. You summarize the entire sample, enter the data on a new Work Sheet 2, and obtain a mean of 23.8 and standard deviation of 14.7. Then you determine the final result from the following Work Sheet 4.

MEAN PER-UNIT ESTIMATION
NON-STRATIFIED

Evaluation of Sample Results

$$A = \frac{U_R \; N \; S_{x_j}}{\sqrt{n}} \sqrt{1 - \frac{n}{N}} = \text{precision}$$

$$\hat{X} = N\bar{x} = \text{point estimate of population total}$$

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	N	Population size		14,000
2	n	Sample size		332
3	S_{x_j}	Standard deviation	Work Sheet 2, line ⑩	14.7
4	$\bar{x}$	Sample mean	Work Sheet 2, line ⑤	23.8
5	n/N	Proportion of population sampled	② ÷ ①	.024
6	$1 - \frac{n}{N}$		1.0 - ⑤	.976
7	$\sqrt{1 - \frac{n}{N}}$	Finite correction factor	$\sqrt{⑥}$	.988
8	$\sqrt{n}$		$\sqrt{②}$	18.22
9	R	Reliability ☐ one-sided estimate ☒ two-sided estimate	Auditor's decision	95%
10	U_R	Reliability factor	Table 5	1.96
11	A	Precision	⑩ x ① x ③ x ⑦ ÷ ⑧	21,873
12	$\hat{X} = N\bar{x}$	Point estimate of population total	① x ④	333,200
13*	$\hat{X} + A$	Upper precision limit	⑫ + ⑪	355,073
14*	$\hat{X} - A$	Lower precision limit	⑫ - ⑪	311,327

*If one-sided estimate, enter "not applicable" in line ⑬ or line ⑭ as appropriate.

The precision on Work Sheet 4 of 21,873 is expressed in total months. This is converted into average months per vehicle by dividing by 14,000 to arrive at 1.6 months. Thus, you have 95% reliability that the true average lease life is between 22.2 (23.8 - 1.6) months and 25.4 (23.8 + 1.6) months. This precision of 1.6 average months is a little worse than you had intended. This is because the standard deviation of the final sample is somewhat larger than that upon which sample size determination was based. (14.70 vs. 14.11)

At this point, you can either determine the amount of the audit adjustment or expand your sample in an attempt to obtain tighter precision and determine the adjustment from those results. This is a matter of judgment. Achieved precision will rarely be exactly equal to desired precision, so planning will often be conservative (as by adding a safety factor to the sample size) in order to allow for this result.

Chapter 4
STRATIFICATION

The population variability, as measured by the standard deviation, is a key determinant of the sample size required to achieve a desired precision and reliability. The larger the standard deviation, the larger the required sample size.

In stratified sampling, we divide the population into relatively homogeneous groups or strata. Constructing these internally homogeneous groups has the effect of reducing overall sample size. The most common basis for stratification in auditing is dollar book value, although any criterion that can be defined adequately can be used.

Stratification is always desirable when the population has a wide range of values, or is skewed.

The auditor must establish a basis for stratification before conducting the audit test; that is, he must establish criteria before sampling, so that he knows which stratum each item in the population belongs to.

To calculate an appropriate total sample size, it is necessary to obtain an estimate of the standard deviation within each stratum. This may be done on the basis of previous experience or from data in a preliminary sample.

Once the total or overall sample size is determined, it must be allocated among the strata. One method of allocation is known as proportional allocation. In this method, the percentage of the sample allocated to each stratum is the same as the percentage of the total population accounted for by that stratum. That is,

$$n_i = n \cdot \frac{N_i}{N}$$

where n_i represents the sample size in the i^{th} stratum, and N_i the number of population items in the i^{th} stratum.

A generally more effective method, however, is optimal allocation. In this method, the relative variability of each stratum is taken into account. Specifically,

$$n_i = n \cdot \frac{N_i S_i}{\sum N_i S_i}$$

or $n_i = n \cdot P_i$ where $P_i = \dfrac{N_i S_i}{\sum N_i S_i}$ and $S_i = \sqrt{\dfrac{\sum x_{ij}^2 - n_i \bar{x}_i^2}{n_i - 1}}$

Work Sheet 6 is used to determine the allocation P_i and the allocation of the total sample size required to achieve a desired precision and reliability. Specifically, the work sheet guides the user through the formula:

$$n = \frac{\left(\sum N_i S_i\right)^2}{\left(A/U_R\right)^2 + \sum N_i S_i^2}$$

This formula takes into account the effect of sampling without replacement.

In selecting the sample, the principles of unrestricted random sampling outlined in Chapter 1 are applied to each stratum.

Having established an audited value for each item in both the preliminary sample and the subsequent sample of each stratum, we can now form the estimate of the total audited value of the population. This is done by the formula:

$$\hat{X} = \sum N_i \bar{x}_i$$

where $\bar{x}_i$ represents the sample mean of audited values in stratum i.

To evaluate the achieved precision of this estimate at the desired reliability level, we use the following formula:

$$A = U_R \sqrt{\sum \frac{N_i (N_i - n_i)}{n_i} S_i^2}$$

To use this formula it is necessary first to compute the standard deviation of audited values (S_i) for each stratum.

Finally, we can assert that with confidence of R%, the interval $\hat{X}$ - A to $\hat{X}$ + A contains the true population value.

In some applications, such as accounts receivable, the auditor can use the recorded or book amount of each item in the population as the basis for stratification. Together with this possibility, he may also decide to examine all the items in some stratum. This can occur for a number of reasons - the computed sample size is nearly all or more than all the items in the stratum, the stratum is small, or the recorded values are so large that any error could be very important.

When every item in one of the strata is examined, the sampling error of that stratum is zero. In computing the estimate, we add the value obtained from this stratum to our estimate from the remaining strata. However, the precision of this estimate is equal to the precision of the estimated portion only.

In reporting the results of a stratified sampling plan, we can use the following language:

The estimated total value of the N accounts is $\hat{X}$ based on our sample of n accounts. The margin of error in this estimate is A. Thus, if each account were examined, the true total value obtained would lie between $\hat{X}$ - A and $\hat{X}$ + A with R% reliability. There is an R% probability that the precision interval does contain the true value.

Case Study 5

Stratified Mean-Per-Unit Estimation

This case is an extension of Case Study 4 in the previous chapter. It illustrates the use of stratification with mean-per-unit estimation.

Suppose in planning the audit of lease lives of the vehicles of ABC LEASING CORPORATION, you notice that automobiles and trucks appear to constitute two definable groups with internally homogeneous values (lease lives). This leads you to believe you might benefit from stratifying the population into two strata - automobiles and trucks.

Instead of selecting one preliminary sample as you did in Case Study 4, you select two, of 50 each. You prepare Work Sheet 2 for each preliminary sample with the following results:

	Total	Automobiles	Trucks
Stratum size	14,000	12,000	2,000
Stratum standard deviation	N/A	10.1	16.2

These data are entered on Work Sheets 5 and 6 which follow:

MEAN-PER-UNIT ESTIMATION
STRATIFIED

Data Sheet

A.

AUDITOR'S DECISIONS		
Quantity to be estimated	X	Lease Months
Desired precision	A	21,000
Desired reliability	R	95%

B.

REQUIRED FACTORS		
Reliability factor (Table 5)	U_R	1.96
Square of reliability factor (Table 5)	U^2_R	3.84
Square of desired precision	A^2	441,000,000

C.

STRATA	PRECISE DESCRIPTION	N_i	S_i PRELIMINARY	S_i^2* PRELIMINARY	S_i^2 FINAL
Stratum 1	Autos	12,000	10.1	102.01	
Stratum 2	Trucks	2,000	16.2	262.44	
Stratum 3					
Stratum 4					

*Enter 0 for stratum examined 100%. For each other stratum, calculate by the method set forth on Work Sheet 2 for S_{X_j}.

MEAN-PER-UNIT ESTIMATION
STRATIFIED

Determination of Sample Size

$$n = \frac{\left(\sum N_i S_i\right)^2}{\left(\frac{A}{U_R}\right)^2 + \sum N_i S_i^2}$$

LINE	NOTATION	SOURCE	STRATUM 1*	STRATUM 2*	STRATUM 3*	STRATUM 4*
1	N_i	Work Sheet 5	12,000	2,000		
2	S_i	Work Sheet 5	10.1	16.2		
3	$N_i S_i$	① x ②	121,200	32,400		
4	S_i^2	Work Sheet 5	102.01	262.44		
5	$N_i S_i^2$	① x ④	1,224,120	524,880		
6	$\sum N_i S_i$	Total of line 3 figures	153,600			
7	$(\sum N_i S_i)^2$	Line 6 squared	23,592,960,000			
8	$\sum N_i S_i^2$	Total of line 5 figures	1,749,000			
9	A^2	Work Sheet 5	441,000,000			
10	Reliability ☐ one-sided estimate ☒ two-sided estimate	Auditor's decision	95%			
11	U^2_R	Work Sheet 5	3.84			
12	$\left(\frac{A}{U_R}\right)^2$	⑨ ÷ ⑪	114,843,750			
13	$\left(\frac{A}{U_R}\right)^2 + \sum N_i S_i^2$	⑧ + ⑫	116,592,750			
14	n	⑦ ÷ ⑬	202			
15	P_i	③ ÷ ⑥	.789	.211		
16	n_i - sample size for each stratum	⑭ x ⑮	159	43		
17	Final n_i - sample size within safety factor					

*Leave blank for strata examined 100%.

Worksheet 6 shows a required sample size of 202 (159 + 43). The total sample size required without stratification was 332. Thus, a significant degree of efficiency was accomplished by a stratified design. This would be expected because of the low variability found in the stratum for automobiles.

The optimal allocation of the sample of 202 is 159 automobiles and 43 trucks. This will require 109 additional automobile leases and no more for trucks. In effect, the preliminary sample of leases for trucks was more than adequate, leading to a slightly greater than optimal sample size. Of course, as in nonstratified sampling, the adequacy of the final results will depend on the accuracy of the preliminary standard deviation from each stratum.

Chapter 5
RATIO AND DIFFERENCE ESTIMATION

Ratio and difference estimation are techniques used to estimate dollar amounts, as is mean-per-unit estimation. And they frequently are more efficient. Ratio estimation is also useful in estimating a ratio - for example, sales margin. As will be discussed, both can be used with or without stratification. Also, as with mean-per-unit estimation, extreme values should be audited 100%.

Ratio Estimation

In addition to the special circumstance of estimating a ratio, ratio estimation is appropriate when:

1. There is a book value for each population item.
2. The total book value is known and corresponds to the addition of all the individual book values.
3. There are some observed differences between audited and book values.
4. The audited values are nearly proportional to the book values.

The estimated ratio is computed by dividing the total sample audited value by the total sample book value when using unstratified random sampling. In symbols,

$$\hat{R} = \frac{\sum a_j}{\sum b_j}$$

where $\hat{R}$ represents the estimated population ratio, a_j represents sample item audited value, and b_j represents sample item book value. To obtain the estimated total audited value, the estimated ratio is multiplied by the total book value. This is expressed symbolically by

$$\hat{X} = \hat{R}B$$

The ratio estimate of the total audited value generally will have better precision than the similar mean-per-unit estimate. This occurs because the standard deviation of the ratio is smaller than the standard deviation of the gross-dollar values.

To determine an appropriate sample size for a specified precision and reliability with ratio estimation, a preliminary sample may be selected, using a random number table or a computer time-sharing program exactly as in previous chapters. The primary purpose of the preliminary sample is to estimate the standard deviation of the ratio population.

The preliminary sample should consist of at least 50 items. This number is generally adequate if the incidence of difference between book and audited values exceeds 20%. When the incidence of difference lies between 5% and 20%, 80 to 100 items should be selected in the preliminary sample. When the incidence of difference is smaller than 5%, the size of a preliminary sample may be even larger than 100 in order to observe some differences.

The calculation of the estimated standard deviation of the ratio population is simplified by using the difference for each item in the sample. This difference is defined as $d_j = a_j - b_j$; the difference in the j^{th} sample item is its audited value minus its book value. The estimated standard deviation is then calculated by means of the formula

$$S_{R_j} = \sqrt{\frac{\sum d_j^2 + (\hat{R}-1)^2 \sum b_j^2 - 2\,(\hat{R}-1)\sum b_j d_j}{n-1}}$$

where $\hat{R} = \frac{\sum a_j}{\sum b_j}$ as indicated above.

Once the estimated standard deviation of the ratio population is available, and given the desired precision and reliability, the required sample size is calculated from the formula

$$n = \frac{n'}{1 + \frac{n'}{N}} \quad \text{where } n' = \left(\frac{S_{R_j} \cdot U_R \cdot N}{A} \right)^2$$

In this formula, n' represents the sample size required when sampling with replacement. This formula is equivalent to the formula used in Chapter 3 with S_{R_j} taking the place of S_{X_j}.

The extra calculation involved in computing n provides for the fact that sampling is without replacement. If the calculated value of n' is smaller than 5% of the population size N, these extra calculations can be omitted, because n is so close to n'.

To compute the precision of the ratio estimate, it is necessary to obtain an estimate of the standard deviation of the ratio population based on the entire sample. This involves using the same formula as used for the preliminary sample.

The calculation of the achieved precision is accomplished by following the formula:

$$A = \frac{S_{R_j} \cdot U_R \cdot N}{\sqrt{n}} \quad \sqrt{1 - \frac{n}{N}}$$

Again, this is the same formula used in Chapter 3 with S_{R_j} taking the place of S_{X_j}, the estimated standard deviation of the population dollar values.

The precision interval with lower limit $\hat{X}$ - A and upper limit $\hat{X}$ + A will contain the true total audited value at the reliability level specified.

To use stratification with ratio estimation it is necessary to:

1. Choose a basis for stratification. The book values will often serve this purpose effectively.

2. Select the number of strata. Usually three to five, possibly with the top stratum sampled 100%, will be reasonable.

3. Select the stratum boundaries. One reasonable procedure is to select boundaries so as to equalize the dollars within each stratum (except the 100% stratum).

If feasible, a preliminary sample from each stratum may be used to obtain an estimate of the stratum standard deviation of ratios. These estimates may be used in place of the S_i of Chapter 4 in computing an appropriate sample size and in allocating the sample to the strata.

If it is not feasible to audit a preliminary sample, it is possible to examine the book values of a sample from each stratum, compute the estimated standard deviation of book values, and use these estimates instead of the S_i values in Chapter 4. This generally provides a more than adequate sample size.

Another alternative would be to calculate a sample size without considering stratification; that is, use the procedures outlined in Chapter 3 and then allocate the resulting sample to the strata in proportion to the dollar amount in the stratum.

After audited values have been determined for each sample item, two ratio estimates of the total audited value are possible - a separate estimate and a combined estimate.

The separate ratio estimate is advantageous when the ratios vary greatly from stratum to stratum. As the name implies, it is formed by computing an estimated ratio for each stratum separately, extending each of the ratios by the total book value in the stratum, and adding the results over all the strata. To compute the precision of the separate ratio estimate, you need an estimate of the standard deviation of the ratio population in each stratum. As a working rule, this requires at least 30 observations in each stratum as well as several observed differences in each stratum.

The point estimate and precision of the separate ratio estimate are computed by the following formulas:

$$\hat{X} = \sum \hat{R}_i \; B_i$$

$$A = U_R \sqrt{\sum \frac{N_i \; (N_i - n_i)}{n_i} \cdot S_i^{\,2}}$$

where $S_i^{\,2}$ equals the amount computed for $S_{R_i}^{\;2}$ in stratum i.

In auditing, the combined ratio estimate is probably used more often than the separate ratio estimate - primarily because it does not require as many observed errors per stratum. As a working rule, it is suggested that at least 50 observations be used, with a significant number of observed differences (at least 10 and preferably more) when using the combined ratio estimate.

As the name implies, the combined ratio estimate is formed by combining the data from each stratum to form a single ratio, which is then extended by the total book value to obtain the combined ratio estimate of the total audited value.

The combined ratio is equal to the estimated total audited value divided by the estimated total book value. These estimates are formed in exactly the same way as the estimated total value was formed in Chapter 4, using mean-per-unit estimation. In symbols, if there are multiple strata:

$$\hat{R}_C = \frac{N_1\bar{a}_1 + N_2\bar{a}_2 + N_3\bar{a}_3}{N_1\bar{b}_1 + N_2\bar{b}_2 + N_3\bar{b}_3} = \frac{\sum N_i \bar{a}_i}{\sum N_i \bar{b}_i}$$

where N_i equals the number of items in the i^{th} stratum (and in this case, i = 1, 2 and 3), $\bar{a}_i$ = sample average audited value within the i^{th} stratum, and $\bar{b}_i$ = sample average book value within the i^{th} stratum.

The formula for computing the precision of this estimate is as follows:

$$A = U_R \sqrt{\sum \frac{N_i\ (N_i - n_i)}{n_i} \qquad S_i^{\,2}}$$

where $S_i^{\,2}$ equals the amount computed for $S_{R_j}^{\,2}$ in stratum i using $\hat{R}_c$ instead of $\hat{R}_i$.

Since the formula for the standard deviation of either of these estimates is complicated, we present an alternative procedure known as replicated sampling. This procedure consists of dividing the total sample into several random subsamples (usually 8 to 10), so that each subsample is a replication of the sampling procedure. In other words, if a stratified sampling procedure calls for 120 items in stratum 1 and 80 items in stratum 2, then ten subsamples from each stratum would result in 12 items in each subsample from stratum 1 and in 8 items in each subsample from stratum 2.

For each subsample, it is necessary to calculate an estimated total value, using either the separate or combined ratio method, then compute the range of these estimates (value of largest estimate minus value of smallest estimate), and divide this range by the number of subsamples. This number when extended by the reliability factor serves as the precision of the estimate. The estimate itself is calculated on the basis of the whole sample.

<u>Difference Estimation</u>

Difference estimation is appropriate and advantageous when:

1. There is a book value for each population item.

2. The total book value is known and corresponds to the addition of all the individual book values.

3. There are some observed differences between audited and book values.

4. The audited values are not proportional to the book values, or similarly, the differences are not proportional to the book values.

In using unrestricted random sampling without replacement, the estimate of the total population difference is formed by extending the sample average difference by the number of items in the population. In symbols,

$$\hat{D} = N\bar{d}, \text{ where } \bar{d} = \frac{1}{n} \sum_{1}^{n} d_j \text{ and } d_j = a_j - b_j$$

To determine an appropriate sample size, a preliminary sample may be used if it is large enough to observe several differences. A preliminary sample of 50 will suffice if there are at least ten differences in the sample.

The formula for estimating the standard deviation of the difference population is exactly the same as that used in Chapter 3, with d_j taking the place of x_j and $\bar{d}$ taking the place of $\bar{x}$. In symbols,

$$S_{D_j} = \sqrt{\frac{\sum_{j=1}^{n} (d_j - \bar{d})^2}{n - 1}} = \sqrt{\frac{\sum_{j=1}^{n} d_j^2 - n\bar{d}^2}{n - 1}}$$

The necessary sample size to achieve a precision A at reliability of R% is given by the formula

$$n = \frac{n'}{1 + \frac{n'}{N}} \quad \text{where } n' = \left(\frac{S_{D_j} \cdot U_R \cdot N}{A}\right)^2$$

After the sample has been selected and audited values established for each of the sample items, the estimated total difference is calculated by means of the formula

$$\hat{D} = N\bar{d}$$

where N represents the number of items in the population and $\bar{d}$ represents the average sample difference. For each item in the sample, the difference equals the audited value minus the book value.

The estimated total audited value can be obtained by adding the estimated total difference to the known book value. In symbols,

$$\hat{X} = B + \hat{D}$$

The precision of this estimate is calculated by means of the formula

$$A = \frac{S_{D_j} \cdot U_R \cdot N}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

Again, this formula is similar to that used in Chapter 3 with S_{D_j} replacing S_{X_j}. This formula is also similar to the one used in ratio estimation, with S_{D_j} replacing S_{R_j}.

Difference estimation may be used with stratified random sampling. The basis for stratification should be some known quantity or quality that is related to the magnitude of the differences in the population. Frequently, the size of the book amount is used as the basis.

In manual application, three to five strata are commonly employed, with stratum boundaries selected so that each stratum contains nearly an equal dollar amount - except the top stratum if it is sampled 100%. Preliminary samples should be selected from each stratum to provide estimates of the standard deviation of differences. These estimates take the place of the S_i in Chapter 4, and are used to calculate an appropriate sample size and to allocate the sample to the strata.

When it is not feasible to audit a preliminary sample, a sample of book values may be selected from each stratum and used to estimate the standard deviation of the book value population with each stratum. These estimates take the place of the S_i in Chapter 4 and can be used to calculate a sample size and to allocate it. The resulting sample size will generally be more than adequate.

After the sample has been selected and audited values established, the estimated total difference is calculated from the following formula, assuming three strata:

$$\hat{D} = N_1\bar{d}_1 + N_2\bar{d}_2 + N_3\bar{d}_3 = \sum N_i\bar{d}_i$$

where N_i represents the number of items in the i^{th} stratum, $\bar{d}_i$ represents the average sample difference in the i^{th} stratum, and $i = 1, 2$ and 3.

To compute the precision of this estimate, the procedures of Chapter 4 can be followed substituting the differences for the dollar values, or the approximation involving subsamples outlined above can be used.

Case Study 6

Estimating a Difference Using Ratio and Difference Estimation

Background and Objectives

Arlo Supply Company is an audit client of your firm. You plan to use statistical sampling in connection with the inventory work in the current year's examination. Specifically, you plan to check quantities, prices and extensions on a selected random sample of inventory tags, and to evaluate your results by using difference or ratio estimation. You will have to foot the client's inventory listing because both these methods depend on having a known book value. You can use either or both ratio and difference estimation in this type of audit test because both use the same basic data; whichever provides the more precise estimate should be planned and used.

A brief summary of the procedure Arlo uses to count its physical inventory follows:

1. The general manager and the warehouse foreman supervise the inventory taking.
2. Two-man teams perform the counts, using two-part inventory tags.
3. The general manager and the foreman place the tags, pre-numbered, on the inventory items.
4. One member of the two-man team counts each item and the other verifies it. They mark each portion of the tag with the agreed-upon quantity.
5. After counting is completed, the general manager and the foreman remove the upper part of each tag and assure themselves that all items were counted and that counts appear reasonable. If appropriate, recounts are made. The lower portion is sent to the accounting department for pricing.
6. On receiving the inventory tags, the

client's accounting department prepares a list showing the tag number, identification data, quantity, and unit price. These sheets are extended, footed and totaled to determine the total inventory value.

In prior years you have observed the inventory and tested its compilation in several separate steps. You always noted various minor errors, and had them listed and given to the client's staff so that totals could be corrected. Although numerous errors have always been found, they have been considered immaterial, in part because they tended to be offsetting. About 5 percent of the inventory items were usually tested.

This year you decide to use statistical sampling to evaluate the reasonableness of the inventory. In terms of your overall audit objective, you believe that if the client's book amount ($86,857) of inventory is within $4,000 of the estimated true value indicated by this test, the book amount can be considered reasonable. This decision results from your determination that about $8,000 would represent a material difference. Your assessment of the difficulty of obtaining additional sample items after the listing has been prepared leads you to select a reliability level of 95%.[1]

To determine an appropriate sample size, you analyze last year's results. Last year's sample of 50 tags, although not randomly chosen, showed 8 differences in quantities, 3 differences in pricing, and none in extensions. Realizing the twin dangers that last year's sample was not random and that the small number of observed differences might lead to a poor estimate of the standard deviation, you decide to take 90 observations, which should allow you to observe about 20 differences.

---

[1] This plan is based on a beta risk of 2½% and reflects an expectation that there will be a high rate of error. See Elliott and Rogers, "Relating Statistical Sampling to Audit Objectives," <u>Journal of Accountancy</u>, July 1972, pp. 46 - 55.

Selection of Sample and Evaluation of Results

You determine in advance that the client has 868 inventory items this year. Using a random number table, you select 90 numbers between 1 and 868 which will correspond with pre-numbered inventory tags. You use a simple matrix to place these in numerical sequence as they are obtained from the table. Duplicate numbers are discarded. Thus:

10 / 100	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99
0- 99	6				42 46	56 55 51		74 70 77	87 83	
100-199	101	118 111	126 123 128	136 134				177 179	187	
200-299	207	218		238 237	241		267	274 271	288 284	290
300-399		316	326	338	343	357	366			
400-499	405		428	438		451 458		473	489	496 491 492
500-599	502	512 519	526	530		552 557	562 561		588 583	591
600-699	604	612	622 621	632 630	646 644				684	697 695
700-799	707 709	716		738 735	742 748	752	764	774		792
800-899			820	833 838		857 851 854	862			

You then make up from this schedule an audit work paper on which to enter the results of your counts of the selected items, extension of prices, and comparison of audited value to book value. You enter some of these values on Work Sheet 8, which is used to compile the required information for the statistical evaluation.

You then enter these data on standard schedules to evaluate results.

These schedules are attached. A key to them is as follows:

Work Sheet 1 -	Sampling objectives
Work Sheet 8 -	Data collection
Work Sheet 9 -	Summary of data for input into evaluation schedules
Work Sheet 10 -	Computation of standard deviation for ratio estimate
Work Sheet 12 -	Evaluation of ratio estimation sample results
Work Sheet 13 -	Computation of standard deviation for difference estimate
Work Sheet 15 -	Evaluation of difference estimation sample results

The aggregate net difference is a credit amount on Work Sheet 8, indicating that the books are overstated. You carry it forward to the standard schedules as a credit, as you do the amount of the cross-products. (It is essential that the integrity of the signs of various input values be maintained from start to finish.)

Since the differences produce a lower standard deviation than do the ratios (5.6 vs. 6.4) (Work Sheets 13 and 10), difference estimation produces the more precise result. The precision interval of the total inventory value based on the differences, at 95% reliability, is $83,317 to $85,223 (Work Sheet 15). The book value of the inventory of $86,857 is not included in the achieved precision interval. However, the planned precision of $4,000 would produce an interval that does include the client's book value ($80,270 to $88,270). This latter interval has a very high reliability (greater than 99.9%) of containing the true value, and consequently the statistical evidence supports the reasonableness of the inventory value; that is, on the basis of $8,000 materiality.

If you use ratio estimation, the point estimate of the inventory balance is $84,190. You can be 95% certain that the actual balance will fall somewhere between $83,101 and $85,279 ($84,190 ± $1,089). Earlier you stated that you would accept any value if you could be 95% certain the recorded inventory value was within ± $4,000 of the estimated true value. Therefore, you would accept any value between $80,190 and $88,190 ($84,190 ± $4,000), and since the client's inventory value of $86,857 is within those limits, you accept his value as being fairly stated.

Conclusion

Ratio and difference estimation are useful methods in audit situations where two values are available for each sample item - for example, recorded and audited values. These methods are generally efficient ones where the book balance and the audited balance for the majority of the sample items agree. Furthermore, these methods are useful as alternatives, giving you flexibility in your estimation approach. The only caveat to the use of difference and ratio estimation is that the sample must contain sufficient differences to assure a good approximation of the distribution of differences and thereby assure a reliable estimate.

Sampling Objectives

Population (define): Physical inventory listing

Population size (N): 868

Information to be obtained: Physical count, unit price, extension

Desired precision (A): $4,000
Desired reliability (R): 95%

Sampling Plan

Sampling method and sample design

Random? [X] yes
[ ] no (Describe method used)

Stratified? [X] no Use Work Sheets 3-4*
[ ] yes Use Work Sheets 5-7

Replacement? [ ] with
[X] without

*Apply to mean-per-unit estimation only.

Evaluation method

Mean per-unit estimation?	[ ] yes	Use Work Sheets 2-7
Ratio estimation?	[X] yes	Use Work Sheets 8-12
Difference estimation?	[X] yes	Use Work Sheets 8 & 9 and 13-15
Attribute estimation?	[ ] yes	Use Work Sheet 16

Correspondence

Describe method of correspondence between the items in the population and the numbers in the random tables.
Last three digits of random number will correspond to the inventory ticket number.

Random number table

What table used?
Random number table U. S. Dept. of Commerce
Route through table:
Left to right, then down to next line
Starting point in table: Row: 6 Column: 3
How selected?
Blind stab

Random numbers from computer program
What program?

Starting specifications:

Other method used for selecting random items - describe

RATIO AND/OR DIFFERENCE ESTIMATION

Data Collection Sheet

Identification	a_j Audited Value	b_j Book Value	Difference $a_j - b_j$	b_j^2	d_j^2	$b_j d_j$
6	\$ 100	\$ 100	—	10,000	—	—
42	85	85	—	7,225	—	—
46	120	120	—	14,400	—	—
51	420	450	\$ <30>	202,500	900	<13,500>
55	18	18	—	324	—	—
56	10	10	—	100	—	—
851	25	25	—	625	—	—
854	152	150	2	22,500	4	300
857	85	85	—	7,225	—	—
862	76	86	<10>	7,396	100	<860>
Totals	\$ 8,456	\$ 8,723	\$ <268>	2,865,425	3,616	<41,250>

Note: There was a total noted of eighteen differences, which comprise the net difference of \$<268>. The footed total of the client's book balance of the inventory is \$86,857.

RATIO AND/OR DIFFERENCE ESTIMATION

Data Summary Sheet

Data from preliminary sample ☐
Data from total sample ☒

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	$\sum b_j$	Sum of sample book values	Work Sheet 8	8,723
2	$\sum b_j^2$	Sum of squares of sample book values	Work Sheet 8	2,865,425
3	$\sum d_j$	Sum of sample differences	Work Sheet 8	<268>
4	$\sum d_j^2$	Sum of squares of sample differences	Work Sheet 8	3,616
5	$\sum b_j d_j$	Sum of cross products	Work Sheet 8	<41,250>
6	$(\hat{R}-1)$	Estimated ratio - 1.0	③ ÷ ①	<.0307>
7	$(\hat{R}-1)^2$		⑥ squared	.0009
8	$2(\hat{R}-1)$		⑥ doubled	<.0614>
9	n	Sample size	Work Sheet 11	90

RATIO ESTIMATION

Estimated Standard Deviation of Population Ratios

$$s_{R_j} = \sqrt{\frac{\sum d_j^2 + (\hat{R}-1)^2 \sum b_j^2 - 2(\hat{R}-1)\sum b_j d_j}{n - 1}}$$

From preliminary sample ☐
From total sample ☒

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	$\sum b_j^2$	Sum of squares of sample book values	Work Sheet 9, line (2)	2,865,425
2	$\sum d_j^2$	Sum of squares of sample differences	Work Sheet 9, line (4)	3,616
3	$(\hat{R}-1)^2$		Work Sheet 9, line (7)	.0009
4	$(\hat{R}-1)^2 \sum b_j^2$		(1) x (3)	2,579
5	$\sum b_j d_j$	Sum of cross products	Work Sheet 9, line (5)	<41, 250>
6	$2(\hat{R}-1)$		Work Sheet 9, line (8)	<.0614>
7	$2(\hat{R}-1)\sum b_j d_j$		(5) x (6)	2,533
8	$\sum d_j^2 + (\hat{R}-1)^2 \sum b_j^2 - 2(\hat{R}-1)\sum b_j d_j$		(2) + (4) - (7)	3,662
9	n	Sample size		90
10	n-1		(9) - 1.0	89
11	$s_{R_j}^2$		(8) ÷ (10)	41.15
12	s_{R_j}	Estimated standard deviation of population of ratio	$\sqrt{(11)}$	6.4

RATIO ESTIMATION

Evaluation of Sample Results

$$A = \frac{U_R \, S_{R_j} \, N}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	$(\hat{R} - 1)$	Estimated ratio - 1.0	Work Sheet 9, line (6)	<.0307>
2	B	Total book value	Client records	86,857
3	$\hat{D} = (\hat{R} - 1)B$	Estimated total difference	(1) x (2)	< 2,667 >
4	$\hat{X} = B + \hat{D}$	Estimated total audited value	(2) + (3)	84,190
5	N	Population size	Work Sheet 11, line (1)	868
6	R	Desired reliability ☐ one-sided estimate ☒ two-sided estimate	Auditor's decision	95%
7	U_R	Reliability factor	Table 5	1.96
8	S_{R_j}	Estimated standard deviation of population of ratios	Work Sheet 10, line (12)	6.4
9	n	Sample size		90
10	$\sqrt{n}$	Square root of sample size	$\sqrt{(9)}$	9.5
11	n/N	Sampling fraction	(9) ÷ (5)	.10
12	1-n/N		1.0 - (11)	.90
13	$\sqrt{1-n/N}$	Finite correction factor	$\sqrt{(12)}$	.95
14	A	Precision	(5)x(7)x(8)x (13) ÷ (10)	1,089
*15	$\hat{X} - A$	Lower precision limit	(4) - (14)	83,101
*16	$\hat{X} + A$	Upper precision limit	(4) + (14)	85,279

*If one-sided estimate, enter "not applicable" in line (15) or (16), as appropriate.

DIFFERENCE ESTIMATION

Estimated Standard Deviation of Population of Differences

$$s_{D_j} = \sqrt{\frac{\sum (d_j - \bar{d})^2}{n - 1}} = \sqrt{\frac{\sum d_j^2 - n\bar{d}^2}{n - 1}}$$

From preliminary sample ☐
From total sample ☒

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	$\sum d_j$	Sum of differences	Work Sheet 9, line ③	<268>
2	$\sum d_j^2$	Sum of squares of differences	Work Sheet 9, line ④	3,616
3	n	Sample size		90
4	n-1		③ - 1.0	89
5	$\bar{d}$	Average difference	① ÷ ③	<2.98>
6	$(\bar{d})^2$	Square of average difference	⑤ x ⑤	8.88
7	$n\bar{d}^2$		③ x ⑥	799
8	$\sum d_j^2 - n\bar{d}^2$		② - ⑦	2,817
9	$s_{D_j}^2$	$\frac{\sum d_j^2 - n\bar{d}^2}{n - 1}$	⑧ ÷ ④	31.7
10	s_{D_j}	Estimated standard deviation of difference population	$\sqrt{⑨}$	5.6

DIFFERENCE ESTIMATION

Evaluation of Sample Results

$$A = \frac{S_{D_j} U_R N}{\sqrt{n}} \sqrt{1 - \frac{n}{N}} = \text{precision}; \quad \hat{X} = N \frac{\sum d_j}{n} + B = \text{point estimate of population total}$$

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	N	Population size		868
2	R	Reliability ☐ one-sided estimate ☒ two-sided estimate	Auditor's decision	95%
3	U_R	Reliability factor	Table 5	1.96
4	n	Sample size		90
5	$\sqrt{n}$		$\sqrt{④}$	9.5
6	S_{D_j}	Estimated standard deviation of population of difference	Work Sheet 13, line ⑩	5.6
7	$\frac{S_{D_j} U_R N}{\sqrt{n}}$	Precision (with replacement)	① x ③ x ⑥ ÷ ⑤	1,003
8	$\frac{n}{N}$	Sampling fraction	④ ÷ ①	.10
9	$1 - \frac{n}{N}$		1.0 - ⑧	.90
10	$\sqrt{1 - \frac{n}{N}}$	Finite correction factor	$\sqrt{⑨}$	.95
11	A	Precision	⑦ x ⑩	953
12	$\sum d_j$	Sum of differences	Worksheet 9, line ③	⟨268⟩
13	$\frac{\sum d_j}{n}$	Average difference	⑫ ÷ ④	⟨2.98⟩
14	$N \frac{\sum d_j}{n} = \hat{D}$	Estimated total difference	① x ⑬	⟨2,587⟩
15	B	Book value	Client's record	86,857
16	$\hat{X}$	Estimated total audited value	⑭ + ⑮	84,270
*17	$\hat{X} + A$	Upper precision limit	⑯ + ⑪	85,223
*18	$\hat{X} - A$	Lower precision limit	⑯ - ⑪	83,317

*If one-sided estimate, enter "not applicable" in line ⑰ or ⑱, as appropriate.

Case Study 7

Estimating a Ratio

Background and Approach

You have been engaged as the first examination auditor for the ABC COMPANY, a seller of building materials that plans to go public shortly after the end of the current year, 19X3. You had been given notice of this two years ago, and consequently, you made an observation of physical inventories at the end of 19X1 and 19X2. However, inventories at the beginning of 19X1 were not observed. Furthermore, the company does not maintain perpetual inventory records. Your dilemma is how to satisfy yourself that cost of goods sold is reasonable for 19X1, as the 19X1 earnings statement will be included in the summary of earnings in the registration statement on an audited basis.

In addition to other tests, you decide to use statistical sampling to test the client's recorded cost of goods sold. Your plan is to take a sample of sales invoices from 19X1, test and cost them out from purchase data, and - using ratio estimation with 95% reliability - estimate the cost of goods sold for the year. If this estimate indicates the recorded cost of goods sold is reasonable before consideration of obsolescence, your audit objective will be fulfilled.

Overall data for year 19X1 is as follows:

Number of sales invoices	14,000
Total amount of sales	$6,000,000
Total amount of cost of sales	4,560,000

A random sample - without replacement - of 100 invoices is taken. This is an arbitrary amount, since you do not have previous knowledge upon which to base your estimate of variance. If this sample size is not adequate, additional items will be selected.

Evaluation of Results

The sample of 100 items is selected and examined, and yields the following:

Sum of sales prices $(\sum b_j)$		\$ 40,117.28
Sum of cost amounts $(\sum a_j)$		30,689.72
Sum of the squares of the sales prices $(\sum b_j^2)$		16,249,783.00
Sum of the squares of the cost amounts $(\sum a_j^2)$		9,489,612.00
Sum of the cross-products of sales prices and cost amounts $(\sum b_j a_j)$		12,351,268.00

Notice that the sale prices assume the role of the book values, and the cost amounts assume the role of the audited values in this application of ratio estimation. Furthermore, Work Sheets 10 and 12 are designed to facilitate the computation of the precision when most of the differences (d_j) between audited values (a_j) and book values (b_j) are zero. The formula shown on Work Sheet 10 may be adapted to your problem by some algebraic manipulation, using the fact that $d_j = a_j - b_j$. Thus,

1. The sample ratio

$$\hat{R} = \frac{\sum a_j}{\sum b_j} = \frac{30{,}689.72}{40{,}117.28} = .765$$

2. The standard deviation

$$S_{R_j} = \sqrt{\frac{\sum(a_j - \hat{R}b_j)^2}{n-1}}$$

or in computational form with a_j and b_j

$$S_{R_j} = \sqrt{\frac{\sum a_j^2 - 2\hat{R}\,(\sum b_j a_j) + \hat{R}^2 \sum b_j^2}{n-1}}$$

$$= \sqrt{\frac{9{,}489{,}612 - (2)\ (.765)\ (12{,}351{,}268) + (.765)^2\ (16{,}249{,}783)}{99}}$$

$$= \quad 32.09$$

3. Precision

$$A = \frac{S_{R_j} \cdot U \cdot N}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

$$= \frac{32.09 \cdot 1.96 \cdot 14{,}000}{\sqrt{100}} \sqrt{1 - \frac{100}{14{,}000}} = 87{,}703$$

The estimated cost of sales is then $4,590,000 ($6,000,000 x .765) and the precision of this estimate at 95% reliability is $87,700. Since the client's recorded amount of $4,560,000 is contained in the precision interval which ranges from $4,502,300 to $4,677,700, the statistical estimate indicates the recorded cost of goods sold is reasonable. Whether this evidence is adequate depends on the magnitude of the achieved precision of $87,700 in relation to an amount you consider material.

Conclusion

Ratio estimation is a useful tool for estimating ratios as well as differences. This case study presents one situation where the estimation of a ratio is required. Other situations would include determining an accounts receivable aging, determining the reasonableness of departmental mark-ons in a retail inventory, and determining the relationship between FIFO and LIFO valuation of inventories.

Case Study 8

Difference Estimation Resulting in an Audit Adjustment

Morris-Walker, Inc., is one of your firm's audit clients that has 900 individual accounts receivable, with a total recorded value of $895,420. You have taken a statistical approach to confirming receivables and a judgmental sampling approach on all other accounts. Based on your appraisal of results of the judgmental tests and the overall financial position and results of operations of Morris-Walker, you believe a misstatement in the receivables of more than $38,000 would be material. You are equally concerned with both under- and overstatements.

In planning your confirmation procedures, you elect to take a primarily substantive approach; that is, not rely heavily on internal control in that area. Accordingly, this leads you to set the desired precision at $20,000, based on the informal guidelines provided by Section 320 of SAS No. 1. You set the reliability at 95%. Also, you decide to use the previous year's results as a means of estimating the population standard deviation in planning sample size. In order to achieve a precision of $20,000, your computations (not included here) indicate a sample of 87 items is required. To be conservative, you round this to 100 and proceed.[1]

---

[1] The reader should note that the $20,000 desired precision was based on a planned alpha risk of 5%. The beta risk is 4%. A detailed discussion of this process can be found in Elliott and Rogers, "Relating Statistical Sampling to Audit Objectives," Journal of Accountancy, July 1972, pp. 46-55. Also see Roberts, "A Statistical Appraisal of SAP #54," Journal of Accountancy, March 1974, pp. 47-53.

First and second requests are sent, and satisfactory alternative procedures are performed on all non-replies. You obtain an audited value for each of the 100 sample items, compare it with each item's book balance, and evaluate the results, using difference estimation. A summary of your evaluation is as follows:

Population size	900
Population total book value	$895,420
Sample size	100
Number of differences found	30
Aggregate amount of differences (Note that a credit difference represents an overstatement in the books)	$ <6,500>
Average difference in sample	$ <65>
Point estimate of population total difference - $<65> x 900	$<58,500>
Point estimate of population total audited value - $895,420 - $58,500	$836,920
Precision of estimate at 95% reliability - computation not included	$ 18,714
95% confidence interval of true total population audited value - $836,920 ± $18,714	$818,206 to $855,634

Based on this result, you conclude, with 95% certainty, that Morris-Walker's accounts receivable are overstated by an amount between $39,786 and $77,214. Both of these limits exceed the materiality limit of $38,000, and therefore an adjustment must be made.

Determining the amount of the adjustment can be approached in several ways:

1. It could be based on these statistical results,

such that the adjusted book amount of the accounts receivable would be no farther than $38,000 from the farthest limit. In this case, any adjustment to an amount within the precision interval would achieve this. You must note, however, that unless achieved precision is equal to or less than materiality, no adjustment will bring the book figure to within a material amount of both precision limits.

In this situation, many auditors prefer to adjust to the point estimate ($836,920), because this shows an equal concern for both over- and understatements.

2. You could use your determination of the cause of the differences found as a basis for additional tests to quantify more exactly errors in certain segments of the population, thus removing those differences and that segment of the population from your statistical test.

   For example, if most or many differences in the accounts receivable were due to the late recording of credit memos, you might attack the problem by a more thorough review of subsequent credit memos.

3. The client, on being confronted with the sample evidence, may wish to make a thorough analysis of his records and correct all specific items involved.

It should also be noted that some precision intervals may be obtained that do contain the client's book value, but where the book value is a material amount away from one of the confidence limits. In such a case, there is a measurable risk that the books may be materially misstated. This risk is called the beta risk and may be controlled by selecting an appropriate precision in relation to materiality. When the auditor uses a low beta risk together with a high reliability, his statistical sample will provide a good basis for making an adjustment. When this is not the case, he should consider either expanding the sample or using the procedure described as alternative 2 above. (See references previously given.)

APPENDIX A

NOTATION AND WORK SHEETS

Note - Work Sheets 2, 9, 10 and 13 are used for both preliminary and final samples. Where there is a preliminary sample, a second copy of the work sheet will be used to combine the preliminary and additional sample items to produce total sample results.

Notation

X_j	Value of the j^{th} item in the population
x_j	Value of the j^{th} item in the sample
$^{S}X_j$	Estimate of standard deviation of items in the population
$\hat{\sigma}_{\bar{x}}$	Estimate of standard error of the mean
n	Number of items in the sample
$\bar{x}$	Mean of items in the sample (pronounced "x bar")
$\bar{X}$	Mean of items in the population ("capital X bar")
R	Reliability - confidence
U_R	Reliability factor - a table reference from R (See Table 5)
N	Number of items in population
A	Precision
X	Population total value
$\hat{X}$	Estimate of population total value (caret indicates estimate)

Additional Notation for Stratified Sampling

x_{i_j}	Value of the j^{th} sample item in stratum i
S_i	Estimate of standard deviation of items in stratum i
N_i	Number of population items in stratum i
P_i	Proportion of total sample items in stratum i
$\bar{x}_i$	Mean of items in sample of stratum i
$\hat{X}_i$	Estimated total of items in stratum i
n_i	Number of items in sample of stratum i

Additional Notation for Ratio and Difference Estimation

a_j	Audited value of the j^{th} sample item
b_j	Book value of the j^{th} sample item

d_j	Excess of audited amount over book value of the j^{th} sample item $(a_j - b_j)$
$\hat{R}$	Estimate of ratio of audited values to book values
n'	Sample size with replacemnt
S_{R_j}	Estimated standard deviation of population of ratios
$\bar{d}$	Sample mean difference
S_{D_j}	Estimated standard deviation of population difference
$\hat{D}$	Estimated total difference
B	Total population book value
B_j	Book value of j^{th} item in population
$\bar{d}_i$	Mean of sample item differences in i^{th} stratum
$\hat{R}_i$	Ratio of audited values to book values for sample items in i^{th} stratum
$\hat{R}_c$	Ratio of audited values to book values for sample items from all strata combined

Sampling Objectives

Population (define):

Population size (N):

Information to be obtained:

Desired precision (A):
Desired reliability (R):

Sampling Plan

Sampling method and sample design

Random?	☐ yes	
	☐ no	(Describe method used)
Stratified?	☐ no	Use Work Sheets 3-4*
	☐ yes	Use Work Sheets 5-7

*Apply to mean-per-unit estimation only.

Evaluation method

Mean-per-unit estimation?	☐ yes	Use Work Sheets 2-7
Ratio estimation?	☐ yes	Use Work Sheets 8-12
Difference estimation?	☐ yes	Use Work Sheets 8 & 9 and 13-15
Attribute estimation?	☐ yes	Use Work Sheet 16

Correspondence

Describe method of correspondence between the items in the population and the numbers in the random tables.

Random number table

What table used?

Route through table:

Starting point in table: Row: Column:
How selected?

Random numbers from computer program
What program?

Starting specifications:

Other method used for selecting random items - describe

MEAN-PER-UNIT ESTIMATION

Estimated Standard Deviation of Items in Population

$$S_{X_j} = \sqrt{\frac{\sum x_j^2 - n\bar{x}^2}{n-1}}$$

☐ Preliminary sample
☐ Total Sample

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	$\sum x_j$	Sum of all sample items		
2	$\sum x_j^2$	Sum of squares of all sample items (square before summing)		
3	n	Sample size		
4	$n-1$		③ - 1.0	
5	$\bar{x}$	Mean	① ÷ ③	
6	$\bar{x}^2$		⑤²	
7	$n\bar{x}^2$		③ x ⑥	
8	$\sum x_j^2 - n\bar{x}^2$		② - ⑦	
9	$\frac{\sum x_j^2 - n\bar{x}^2}{n-1}$	Variance	⑧ ÷ ④	
10	S_{X_j}	Standard deviation of sample items and estimated standard deviation of population	$\sqrt{⑨}$	

MEAN-PER-UNIT ESTIMATION
NON-STRATIFIED

Determination of Sample Size from Estimated Standard Deviation

$$n = \frac{S_{X_j}^2 \; U_R^2 \; N^2}{A^2 + S_{X_j}^2 \; U_R^2 \; N}$$

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	A	Precision desired	Auditor's decision	
2	R	Reliability desired ☐ one-sided estimate ☐ two-sided estimate	Auditor's decision	
3	U_R^2	Square of reliability factor	Table 5	
4	N	Number of items in population		
5	S_{X_j}	Estimated standard deviation	Work Sheet 2, line ⑩	
6	$S_{X_j}^2$		Work Sheet 2, line ⑨ = $⑤^2$	
7	$S_{X_j}^2 \; U_R^2 \; N$		⑥ x ③ x ④	
8	$S_{X_j}^2 \; U_R^2 \; N^2$		④ x ⑦	
9	A^2		$①^2$	
10	$A^2 + S_{X_j}^2 \; U_R^2 \; N$		⑨ + ⑦	
11	n	Sample Size	⑧ ÷ ⑩	
12	final n	Sample size with safety factor	⑪ plus safety factor (auditor's decision)	

MEAN-PER-UNIT ESTIMATION
NON-STRATIFIED

Evaluation of Sample Results

$$A = \frac{U_R \, N \, S_{X_j}}{\sqrt{n}} \sqrt{1 - \frac{n}{N}} = \text{precision}$$

$$\hat{X} = N\bar{x} = \text{point estimate of population total}$$

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	N	Population size		
2	n	Sample size		
3	S_{X_j}	Standard deviation	Work Sheet 2, line (10)	
4	$\bar{x}$	Sample mean	Work Sheet 2, line (5)	
5	n/N	Proportion of population sampled	(2) ÷ (1)	
6	$1 - \frac{n}{N}$		1.0 - (5)	
7	$\sqrt{1 - \frac{n}{N}}$	Finite correction factor	$\sqrt{(6)}$	
8	$\sqrt{n}$		$\sqrt{(2)}$	
9	R	Reliability ☐ one-sided estimate ☐ two-sided estimate	Auditor's decision	
10	U_R	Reliability factor	Table 5	
11	A	Precision	(10) x (1) x (3) x (7) ÷ (8)	
12	$\hat{X} = N\bar{x}$	Point estimate of population total	(1) x (4)	
13*	$\hat{X} + A$	Upper precision limit	(12) + (11)	
14*	$\hat{X} - A$	Lower precision limit	(12) - (11)	

*If one-sided estimate, enter "not applicable" in line (13) or line (14) as appropriate.

MEAN—PER-UNIT ESTIMATION STRATIFIED

Data Sheet

A. AUDITOR'S DECISIONS		
Quantity to be estimated	X	
Desired precision	A	
Desired reliability	R	

B. REQUIRED FACTORS		
Reliability factor (Table 5)	U_R	
Square of reliability factor (Table 5)	U^2_R	
Square of desired precision	A^2	

C.

STRATA	PRECISE DESCRIPTION	N_i	S_i PRELIMINARY	S_i^2* PRELIMINARY	S_i^2 FINAL
Stratum 1					
Stratum 2					
Stratum 3					
Stratum 4					

*Enter 0 for stratum examined 100%. For each other stratum, calculate by the method set forth on Work Sheet 2 for S_{x_j}.

MEAN-PER-UNIT ESTIMATION
STRATIFIED

Determination of Sample Size

$$n = \frac{\left(\sum N_i S_i\right)^2}{\left(\frac{A}{U_R}\right)^2 + \sum N_i S_i^2}$$

LINE	NOTATION	SOURCE	STRATUM 1*	STRATUM 2*	STRATUM 3*	STRATUM 4*
1	N_i	Work Sheet 5				
2	S_i	Work Sheet 5				
3	$N_i S_i$	① x ②				
4	S_i^2	Work Sheet 5				
5	$N_i S_i^2$	① x ④				
6	$\sum N_i S_i$	Total of line 3 figures				
7	$\left(\sum N_i S_i\right)^2$	Line 6 squared				
8	$\sum N_i S_i^2$	Total of line 5 figures				
9	A^2	Work Sheet 5				
10	Reliability ☐ one-sided estimate ☐ two-sided estimate	Auditor's decision				
11	U^2_R	Work Sheet 5				
12	$\left(\frac{A}{U_R}\right)^2$	⑨ ÷ ⑪				
13	$\left(\frac{A}{U_R}\right)^2 + \sum N_i S_i^2$	⑧ + ⑫				
14	n	⑦ ÷ ⑬				
15	P_i	③ ÷ ⑥				
16	n_i - sample size for each stratum	⑭ x ⑮				
17	Final n_i - sample size within safety factor					

*Leave blank for strata examined 100%.

MEAN-PER-UNIT ESTIMATION
STRATIFIED

Evaluation of Results

$$A = U_R \sqrt{\sum \frac{N_i\ (N_i - n_i)}{n_i}\ S_i^{\,2}} = \text{precision}$$

$$\hat{X} = \sum N_i \bar{x}_i \quad = \text{point estimate of population total}$$

LINE	NOTATION	SOURCE	STRATUM 1	STRATUM 2	STRATUM 3	STRATUM 4
1	N_i	Work Sheet 5				
2	$\sum x_i$	Sum of items in sample				
3	n_i	Size of sample				
4	$\bar{x}_i$	② ÷ ③				
5	$\hat{X}_i$	① x ④				
6	$S_i^{\,2}$ (final)	Work Sheet 5				
7	$N_i - n_i$	① - ③				
8	$N_i(N_i - n_i)$	① x ⑦				
9	$N_i(N_i - n_i)S_i^{\,2}$	⑥ x ⑧				
10	$\frac{N_i(N_i - n_i)}{n_i} S_i^{\,2}$	⑨ ÷ ③				

LINE	NOTATION	SOURCE	
11	$\sum \frac{N_i(N_i - n_i)}{n_i} S_i^{\,2}$	Total of line ⑩ figures	
12	$\hat{X}$ point estimate of total	Total of line ⑤ figures	
13	R reliability ☐ one-sided estimate ☐ two-sided estimate	Auditor's decision	
14	U_R	Table 5	
15	$\sqrt{\sum \frac{N_i(N_i - n_i)}{n_i} S_i^{\,2}}$	$\sqrt{⑪}$	
16	A	⑮ x ⑭	
17*	$\hat{X}$-A lower limit at given reliability level	⑫ - ⑯	
18*	$\hat{X}$+A upper limit at given reliability level	⑫ + ⑯	

*If one-sided estimate, enter "not applicable" in line ⑰ or ⑱, as appropriate.

RATIO AND/OR DIFFERENCE ESTIMATION

Data Collection Sheet

Identification	a_j Audited Value	b_j Book Value	Difference $a_j - b_j$	b_j^2	d_j^2	$b_j d_j$

RATIO AND/OR DIFFERENCE ESTIMATION

Data Summary Sheet

Data from preliminary sample ☐
Data from total sample ☐

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	$\sum b_j$	Sum of sample book values	Work Sheet 8	
2	$\sum b_j^2$	Sum of squares of sample book values	Work Sheet 8	
3	$\sum d_j$	Sum of sample differences	Work Sheet 8	
4	$\sum d_j^2$	Sum of squares of sample differences	Work Sheet 8	
5	$\sum b_j d_j$	Sum of cross products	Work Sheet 8	
6	$(\hat{R}-1)$	Estimated ratio - 1.0	③ ÷ ①	
7	$(\hat{R}-1)^2$		⑥ squared	
8	$2(\hat{R}-1)$		⑥ doubled	
9	n	Sample size	Work Sheet 11	

RATIO ESTIMATION

Estimated Standard Deviation of Population Ratios

$$s_{R_j} = \sqrt{\frac{\sum d_j^2 + (\hat{R}-1)^2 \sum b_j^2 - 2(\hat{R}-1)\sum b_j d_j}{n-1}}$$

From preliminary sample ☐
From total sample ☐

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	$\sum b_j^2$	Sum of squares of sample book values	Work Sheet 9, line (2)	
2	$\sum d_j^2$	Sum of squares of sample differences	Work Sheet 9, line (4)	
3	$(\hat{R}-1)^2$		Work Sheet 9, line (7)	
4	$(\hat{R}-1)^2 \sum b_j^2$		(1) x (3)	
5	$\sum b_j d_j$	Sum of cross Products	Work Sheet 9, line (5)	
6	$2(\hat{R}-1)$		Work Sheet 9, line (8)	
7	$2(\hat{R}-1)\sum b_j d_j$		(5) x (6)	
8	$\sum d_j^2 + (\hat{R}-1)^2 \sum b_j^2 - 2(\hat{R}-1)\sum b_j d_j$		(2) + (4) - (7)	
9	n	Sample size		
10	$n-1$		(9) - 1.0	
11	$s_{R_j}^2$		(8) ÷ (10)	
12	s_{R_j}	Estimated standard deviation of population of ratio	$\sqrt{(11)}$	

RATIO ESTIMATION

Determination of Sample Size

Sample size with replacement = n' = $\left(\frac{S_{R_j} \, U_R \, N}{A}\right)^2$

Sample size without replacement = n = $\frac{n'}{1 + \frac{n'}{N}}$

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	N	Population size		
2	S_{R_j}	Estimated standard deviation of population of ratios	Work Sheet 10, line (9)	
3	R	Desired reliability ☐ one-sided estimate ☐ two-sided estimate	Auditor's decision	
4	U_R	Reliability factor	Table 5	
5	A	Desired precision	Auditor's decision	
6	$\frac{S_{R_j} \, U_R \, N}{A}$		(1) x (2) x (4) ÷ (5)	
7	n'	Sample size (with replacement)	(6) squared	
8	n'/N	Adjustment factor	(7) ÷ (1)	
9	1 + n'/N		1.0 + (8)	
10	n	Sample size (without replacement)	(7) ÷ (9)	
11	final n	Sample size with safety factor		

RATIO ESTIMATION

Evaluation of Sample Results

$$A = \frac{U_R \; S_{R_j} \; N}{\sqrt{n}} \sqrt{1 - \frac{n}{N}}$$

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	$(\hat{R} - 1)$	Estimated ratio - 1.0	Work Sheet 9, line ⑥	
2	B	Total book value	client records	
3	$\hat{D} = (\hat{R} - 1)B$	Estimated total difference	① x ②	
4	$\hat{X} = B + \hat{D}$	Estimated total audited value	② + ③	
5	N	Population size	Work Sheet 11, line ①	
6	R	Desired reliability ☐ one-sided estimate ☐ two-sided estimate	Auditor's decision	
7	U_R	Reliability factor	Table 5	
8	S_{R_j}	Estimated standard deviation of population of ratios	Work Sheet 10, line ⑫	
9	n	Sample size		
10	$\sqrt{n}$	Square root of sample size	$\sqrt{⑨}$	
11	n/N	Sampling fraction	⑨ ÷ ⑤	
12	1 - n/N		1.0 - ⑪	
13	$\sqrt{1 - n/N}$	Finite correction factor	$\sqrt{⑫}$	
14	A	Precision	⑤ x ⑦ x ⑧ x ⑬ ÷ ⑩	
*15	$\hat{X} - A$	Lower precision limit	④ - ⑭	
*16	$\hat{X} + A$	Upper precision limit	④ + ⑭	

*If one-sided estimate, enter "not applicable" in line ⑮ or ⑯, as appropriate.

DIFFERENCE ESTIMATION

Estimated Standard Deviation of Population of Differences

$$S_{D_j} = \sqrt{\frac{\sum (d_j - \bar{d})^2}{n - 1}} = \sqrt{\frac{\sum d_j^2 - n\bar{d}^2}{n - 1}}$$

From preliminary sample ☐
From total sample ☐

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	$\sum d_j$	Sum of differences	Work Sheet 9, line ③	
2	$\sum d_j^2$	Sum of squares of differences	Work Sheet 9, line ④	
3	n	Sample size		
4	$n-1$		③ - 1.0	
5	$\bar{d}$	Average difference	① ÷ ③	
6	$(\bar{d})^2$	Square of average difference	⑤ x ⑤	
7	$n\bar{d}^2$		③ x ⑥	
8	$\sum d_j^2 - n\bar{d}^2$		② - ⑦	
9	$S_{D_j}^2$	$\frac{\sum d_j^2 - n\bar{d}^2}{n - 1}$	⑧ ÷ ④	
10	S_{D_j}	Estimated standard deviation of difference population	$\sqrt{⑨}$	

DIFFERENCE ESTIMATION

Determination of Sample Size

$$\text{Sample size with replacement} = n' = \left(\frac{S_{R_j} \; U_R \; N}{A} \right)^2$$

$$\text{Sample size without replacement} = n = \frac{n}{1 + \frac{n'}{N}}$$

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	N	Population Size		
2	S_{R_j}	Estimated standard deviation of population of ratios	Work Sheet 13, line ⑩	
3	R	Desired reliability ☐ one-sided estimate ☐ two-sided estimate	Auditor's decision	
4	U_R	Reliability factor	Table 5	
5	A	Desired precision	Auditor's decision	
6	$\frac{S_{R_j} \; U_R \; N}{A}$		① x ② x ④ ÷ ⑤	
7	n'	Sample size (with replacement)	⑥ squared	
8	n'/N	Adjustment factor	⑦ ÷ ①	
9	1 + n'/N		1.0 + ⑧	
10	n	Sample size (without replacement)	⑦ ÷ ⑨	
11	final n	Sample size with safety factor		

DIFFERENCE ESTIMATION Work Sheet 15

Evaluation of Sample Results

$$A = \frac{S_{D_j} U_R N}{\sqrt{n}} \sqrt{1 - \frac{n}{N}} = \text{precision}; \quad \hat{X} = N \frac{\sum d_j}{n} + B = \text{point estimate of population total}$$

LINE	NOTATION	DESCRIPTION	SOURCE	RESULT
1	N	Population size		
2	R	Reliability ☐ one-sided estimate ☐ two-sided estimate	Auditor's decision	
3	U_R	Reliability factor	Table 5	
4	n	Sample size		
5	$\sqrt{n}$		√④	
6	S_{D_j}	Estimated standard deviation of population of differences	Work Sheet 13, line ⑩	
7	$\frac{S_{D_j} U_R N}{\sqrt{n}}$	Precision (with replacement)	① x ③ x ⑥ ÷ ⑤	
8	$\frac{n}{N}$	Sampling fraction	④ ÷ ①	
9	$1 - \frac{n}{N}$		1.0 - ⑧	
10	$\sqrt{1 - \frac{n}{N}}$	Finite correction factor	√⑨	
11	A	Precision	⑦ x ⑩	
12	$\sum d_j$	Sum of differences	Work Sheet 9, line ③	
13	$\frac{\sum d_j}{n}$	Average difference	⑫ ÷ ④	
14	$N \frac{\sum d_j}{n} = \hat{D}$	Estimated total difference	① x ⑬	
15	B	Book value	Client's record	
16	$\hat{X}$	Estimated total audited value	⑭ + ⑮	
*17	$\hat{X} + A$	Upper precision limit	⑯ + ⑪	
*18	$\hat{X} - A$	Lower precision limit	⑯ - ⑪	

*If one-sided estimate, enter "not applicable" in line ⑰ or ⑱, as appropriate

ATTRIBUTE SAMPLING

Summary Worksheet

Description of Population:

Size of population:

Description of an occurrence:

	Occurrence Type											
	1	2	3	4	5	6	7	8	9	10	11	12
Auditor's decisions												
a. Anticipated occurrence rate												
b. Specified upper precision limit												
c. Specified reliability (confidence level)												
Table reference (Table 1, 2 or 4)												
d. Required sample size												
Actual sample size												
e. (Consider using largest value on line d)												
Audit results												
f. Number of occurrences in sample												
g. Actual upper precision limit (Table 2 or 3)												

APPENDIX B

TABLES

Contents

Use of Tables 1, 2 and 3

Tables 1, 2 and 3 are based on the binomial distribution, which does not take into account the size of the population from which the sample is selected. Whenever the auditor wishes to adjust the tabled results to reflect the population size, the following procedure may be used:

Step 1 Subtract the sample occurence rate (SOR) from the tabled value of the upper precision limit (TUPL) [SOR = number of sample occurences (NSO) ÷ sample size]

Step 2 Multiply this number by the square root of: one minus the sample size, n, divided by the population size, N $\left(\sqrt{1 - \frac{n}{N}}\right)$

Step 3 Compute the adjusted upper precision limit (AUPL) by adding the result of Step 2 to the sample occurence rate (SOR)

The procedure is summarized by the following expression:

$$AUPL = SOR + (TUPL - SOR) \sqrt{1 - \frac{n}{N}}$$

In general, this procedure will be worthwhile doing only when the sample size exceeds 10% of the population size.

To facilitate choosing an appropriate sample size, the necessary information is presented in both graphical and tabular form. In using either form, the auditor specifies a desired reliability level, an anticipated occurrence rate, and a desired upper precision limit.

If he is using the graphical form (Table 1), the auditor selects the graph covering the desired reliability level and follows the curve corresponding to his anticipated occurrence rate down until it intersects with the vertical line corresponding to his desired upper precision limit. The horizontal line at which the intersection occurs corresponds to the required sample size.

If he is using the tabular form (Table 2), the auditor selects the tabulation covering the desired reliability level, finds his anticipated occurrence rate along the top row, then proceeds down the corresponding column until he finds the desired upper precision limit. Reading across the row to the left from this value, he will find the indicated sample size.

This table can also be used to evaluate the results of a sample. When used for this purpose, the top row represents the observed sample occurrence rate. The achieved upper precision limit is found at the point where the column headed by the sample occurrence rate intersects with the row corresponding to the sample size.

Table 3 is also available for evaluating the results of the sample. To use this table, the auditor locates the sample size, the row corresponding to the number of observed occurrences, and the column corresponding to the desired reliability. The indicated number represents the upper precision limit at the specified reliability level.

If the auditor uses the graphs to choose an appropriate sample size, there are no problems with interpolation. To avoid interpolating sample results, sample sizes should be set at even increments of 25 through 400. When this is not possible, the following interpolation rules can be used:

1. Consult the table corresponding to next lower tabled sample size n_L. The upper precision limit (LUPL) corresponding to the observed number of occurrences is larger than the achieved upper precision limit.

2. Consult the table corresponding to next higher tabled sample size, n_H. The upper precision limit (SUPL) corresponding to the observed number of occurrences is smaller than the achieved upper precision limit.

3. Use linear interpolation:

$$UPL = LUPL - \left(\frac{n - n_L}{n_H - n_L}\right) \cdot (LUPL - SUPL)$$

Tables 1 and 3 have been adapted from materials furnished by The American Group of Certified Public Accountant Firms.

Table 1

ATTRIBUTE SAMPLING - DETERMINATION OF SAMPLE SIZE-
GRAPHICAL FORM

Reliability Level - 90 Percent

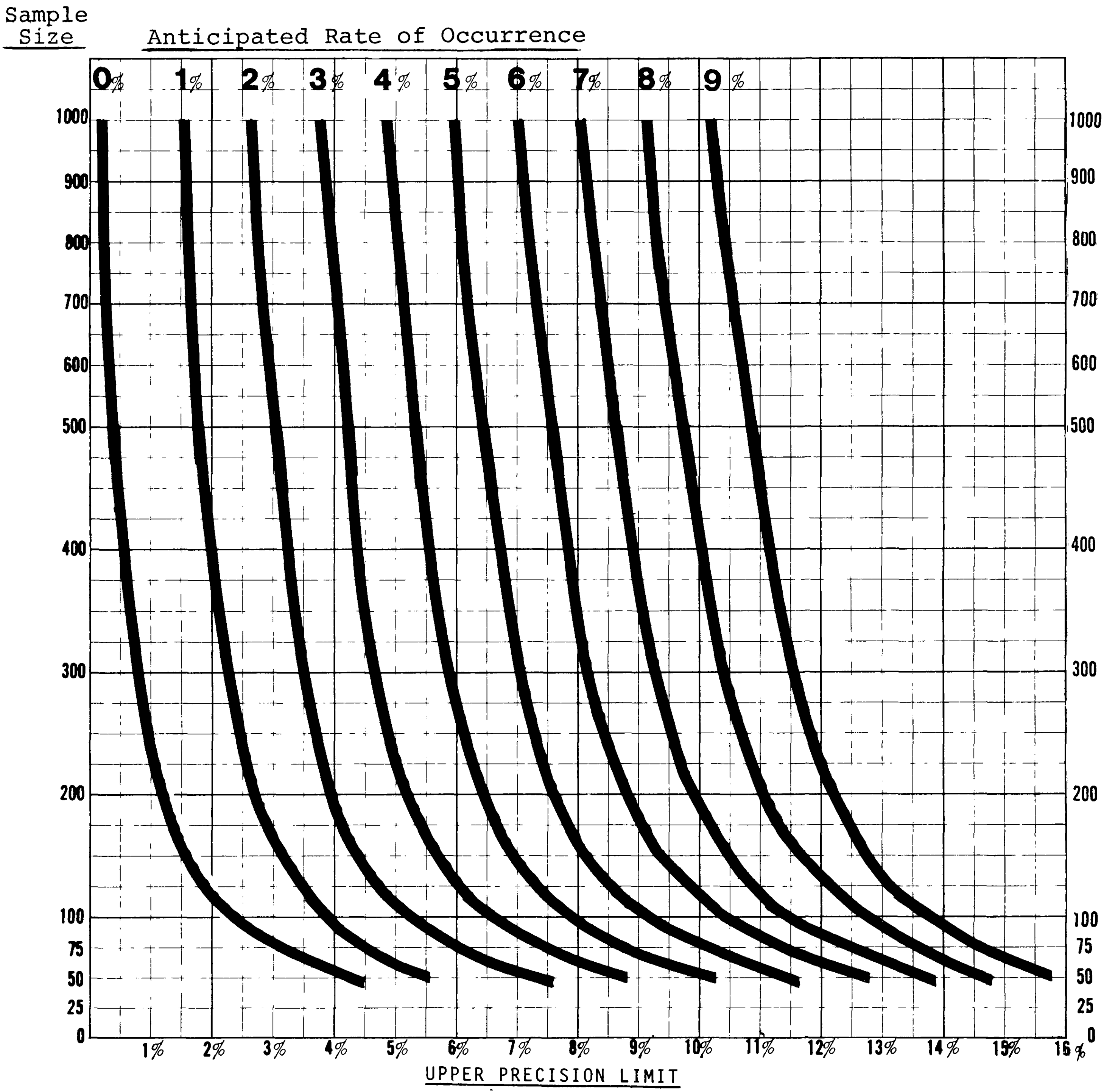

Table 1

ATTRIBUTE SAMPLING - DETERMINATION OF SAMPLE SIZE-
GRAPHICAL FORM

Reliability Level - 95 Percent

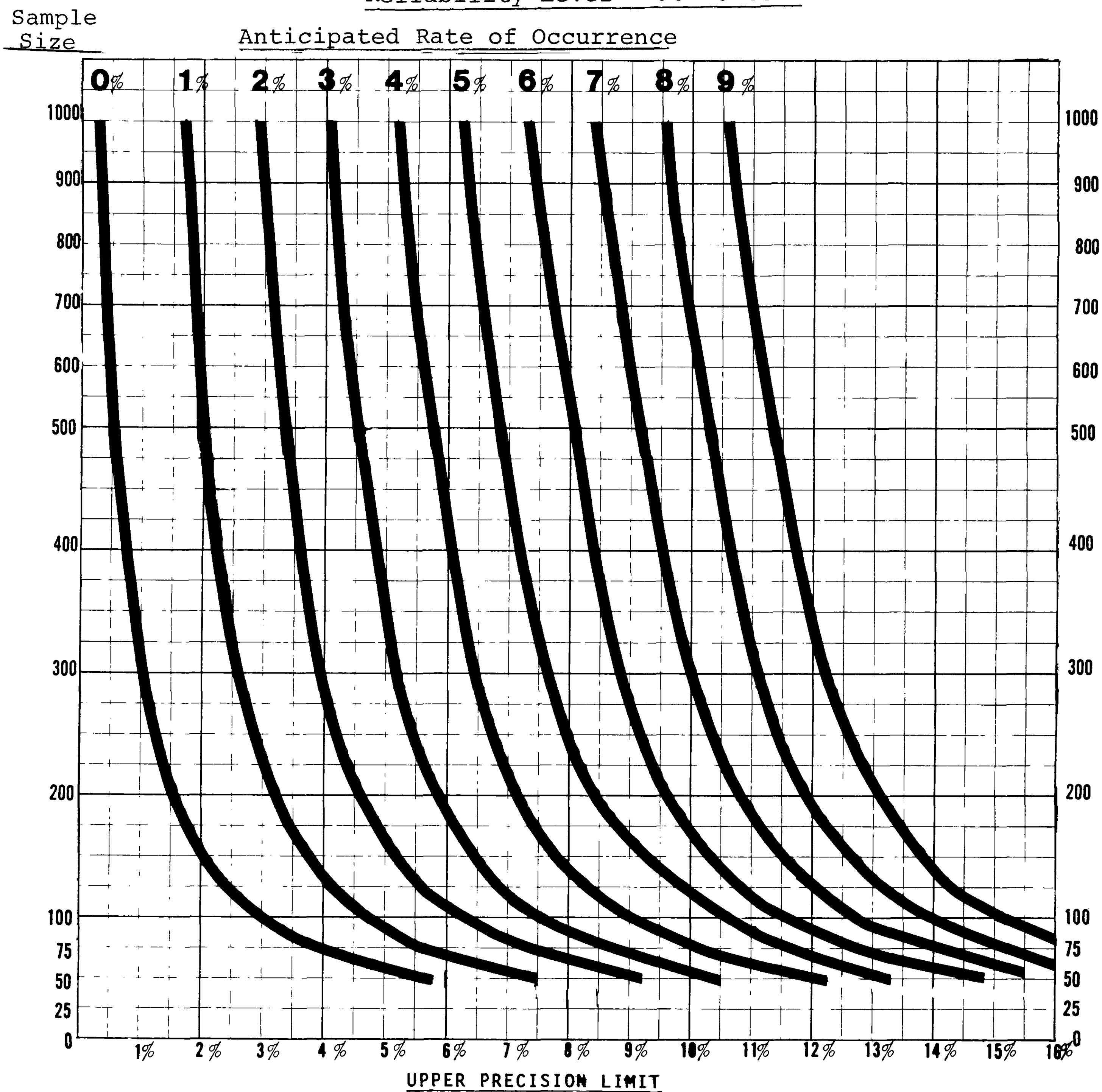

Table 1

ATTRIBUTE SAMPLING - DETERMINATION OF SAMPLE SIZE—
GRAPHICAL FORM

Reliability Level - 99 Percent

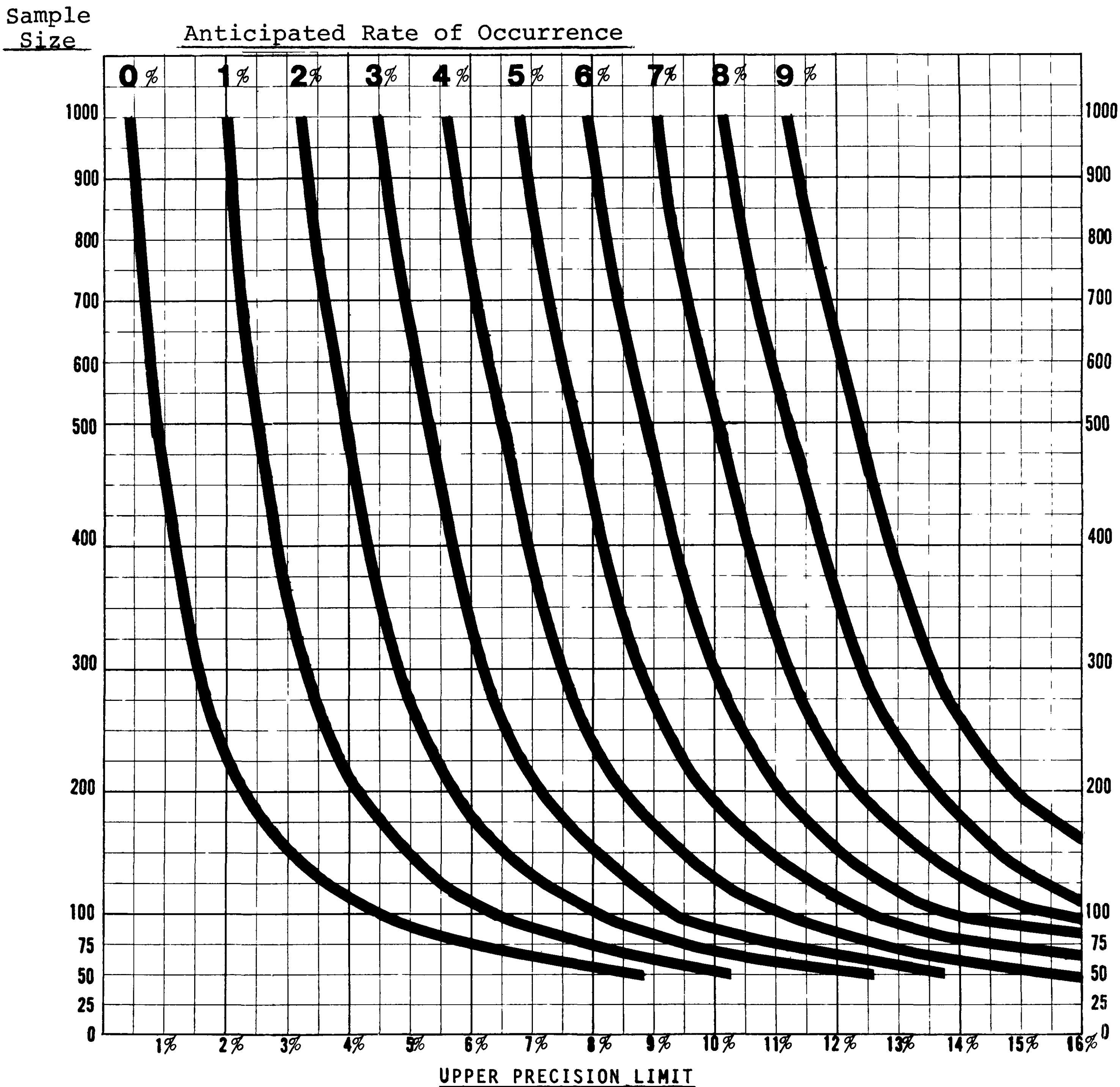

Table 2

DETERMINATION OF SAMPLE SIZE - TABULAR FORM

ONE-SIDED UPPER PRECISION LIMITS

Reliability Level - 90.0 Percent

* *

SAMPLE	OCCURRENCE RATE																			
SIZE	0.0	.5	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	12.0	14.0	16.0	18.0	20.0	25.0	30.0	40.0	50.0
50	4.5			7.6		10.3		12.9		15.4		17.8	20.1	22.7	24.7	27.2	29.1		39.8	50.0	59.9
100	2.3		3.3	5.2	6.6	7.8	9.1	10.3	11.7	12.7	14.0	15.0	17.3	19.6	21.7	24.0	26.1	31.4	36.6	46.9	56.8
150	1.5			4.4		6.9		9.3		11.6		13.9	16.1	18.4	20.5	22.7	24.8		35.2	45.5	55.4
200	1.1	1.9	2.6	4.0	5.2	6.4	7.6	8.8	10.0	11.0	12.2	13.3	15.5	17.7	19.8	22.0	24.0	29.3	34.5	44.4	54.4
250	.9			3.7		6.1		8.4		10.7		12.9	15.1	17.2	19.3	21.5	23.6		33.7	43.7	53.7
300	.8		2.2	3.5	4.7	5.9	7.0	8.2	9.3	10.4	11.5	12.6	14.7	16.9	19.0	21.1	23.2	28.2	33.2	43.2	53.2
350	.7			3.3		5.7		8.0		10.2.		12.3	14.5	16.7	18.8	20.9	22.8		32.8	42.8	52.8
400	.6	1.3	2.0	3.2	4.4	5.6	6.7	7.8	8.9	10.0	11.1	12.2	14.3	16.5	18.5	20.5	22.5	27.5	32.5	42.5	52.5
450	.5			3.1		5.5		7.7		9.9		12.0	14.2	16.3	18.3	20.3	22.3		32.3	42.3	52.2
500	.5		1.8	3.1	4.2	5.4	6.5	7.6	8.7	9.8	10.9	11.9	14.1	16.1	18.1	20.1	22.1	27.1	32.1	42.1	52.0
550	.4			3.0		5.3		7.5		9.7		11.8	13.9	15.9	17.9	19.9	21.9		31.9	41.9	51.9
600	.4	1.1	1.7	2.9	4.1	5.2	6.3	7.4	8.5	9.6	10.7	11.7	13.7	15.7	17.7	19.7	21.7	26.7	31.7	41.7	51.7
650	.4			2.9		5.2		7.4		9.5		11.6	13.6	15.6	17.6	19.6	21.6		31.6	41.6	51.6
700	.3		1.7	2.9	4.0	5.1	6.2	7.3	8.4	9.5	10.5	11.5	13.5	15.5	17.5	19.5	21.5	26.5	31.5	41.5	51.5
750	.3			2.8		5.1		7.3		9.4		11.4	13.4	15.4	17.4	19.4	21.4		31.4	41.4	51.4
800	.3	1.0	1.6	2.8	3.9	5.0	6.1	7.2	8.3	9.3	10.3	11.3	13.3	15.3	17.3	19.3	21.3	26.3	31.3	41.3	51.3
850	.3			2.8		5.0		7.2		9.2		11.2	13.2	15.3	17.3	19.3	21.3		31.3	41.3	51.3
900	.3		1.6	2.7	3.9	5.0	6.0	7.1	8.2	9.2	10.2	11.2	13.2	15.2	17.2	19.2	21.2	26.2	31.2	41.2	51.2
950	.2			2.7		4.9		7.1		9.1		11.1	13.1	15.1	17.1	19.1	21.1		31.1	41.1	51.1
1000	.2	.9	1.5	2.7	3.8	4.9	6.0	7.1	8.1	9.1	10.1	11.1	13.1	15.1	17.1	19.1	21.1	26.1	31.1	41.1	51.1
1500	.2		1.4	2.5	3.6	4.7	5.7	6.7	7.7	8.7	9.7	10.7	12.7	14.7	16.7	18.7	20.7	25.7	30.7	40.7	50.7
2000	.1	.8	1.3	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5	12.5	14.5	16.5	18.5	20.5	25.5	30.5	40.6	50.6
2500	.1		1.3	2.4	3.4	4.4	5.4	6.4	7.4	8.4	9.4	10.4	12.4	14.4	16.4	18.4	20.4	25.4	30.4	40.4	50.4
3000	.1	.7	1.3	2.4	3.4	4.4	5.4	6.4	7.4	8.4	9.4	10.4	12.4	14.4	16.4	18.4	20.4	25.4	30.4	40.4	50.4
4000	.1	.7	1.2	2.3	3.3	4.3	5.3	6.3	7.3	8.3	9.3	10.3	12.3	14.3	16.3	18.3	20.3	25.3	30.3	40.3	50.3
5000	.0	.7	1.2	2.3	3.2	4.2	5.2	6.2	7.2	8.2	9.2	10.2	12.2	14.2	16.2	18.2	20.2	25.2	30.2	40.2	50.2

* *

Table 2

DETERMINATION OF SAMPLE SIZE - TABULAR FORM

ONE-SIDED UPPER PRECISION LIMITS

Reliability Level - 95.0 Percent

* *

SAMPLE SIZE	OCCURRENCE RATE 0.0	.5	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0	12.0	14.0	16.0	18.0	20.0	25.0	30.0	40.0	50.0
50	5.8			9.1		12.1		14.8		17.4		19.9	22.3	25.1	27.0	29.6	31.6		42.4	52.6	62.4
100	3.0		4.7	6.2	7.6	8.9	10.2	11.5	13.0	14.0	15.4	16.4	18.7	21.2	23.3	25.6	27.7	33.1	38.4	48.7	56.6
150	2.0			5.1		7.7		10.2		12.6		15.0	17.3	19.6	21.7	24.0	26.1		36.7	47.0	56.8
200	1.5	2.4	3.1	4.5	5.8	7.1	8.3	9.5	10.8	11.9	13.1	14.2	16.4	18.7	20.9	23.1	25.2	30.5	35.7	45.7	55.6
250	1.2			4.2		6.7		9.1		11.4		13.7	15.9	18.1	20.3	22.4	24.6		34.8	44.8	54.7
300	1.0		2.6	3.9	5.2	6.4	7.6	8.8	10.0	11.1	12.2	13.3	15.5	17.7	19.8	22.0	24.1	29.1	34.1	44.1	54.1
350	.9			3.7		6.2		8.5		10.8		13.0	15.2	17.4	19.5	21.7	23.6		33.6	43.6	53.6
400	.7	1.6	2.3	3.6	4.8	6.0	7.2	8.3	9.5	10.6	11.7	12.8	15.0	17.2	19.2	21.2	23.2	28.2	33.2	43.2	53.2
450	.7			3.5		5.9		8.2		10.4		12.6	14.8	16.8	18.9	20.9	22.9		32.9	42.9	52.9
500	.5		2.1	3.4	4.6	5.8	6.9	8.0	9.2	10.3	11.4	12.5	14.6	16.7	18.6	20.7	22.6	27.6	32.6	42.6	52.6
550	.5			3.3		5.7		7.9		10.1		12.3	14.4	16.4	18.4	20.4	22.4		32.4	42.4	52.4
600	.5	1.3	2.0	3.2	4.4	5.6	6.7	7.8	9.0	10.0	11.2	12.2	14.2	16.2	18.2	20.2	22.2	27.2	32.2	42.2	52.2
650	5			3.2		5.5		7.7		10.0		12.1	14.1	16.1	18.1	20.1	22.1		32.1	42.1	52.1
700	.4		1.9	3.1	4.3	5.4	6.6	7.7	8.8	9.9	10.8	11.9	13.9	15.9	17.9	19.9	21.9	26.9	31.9	41.9	51.9
750	.4			3.1		5.4		7.6		9.8		11.8	13.8	15.8	17.8	19.8	21.8		31.8	41.8	51.8
800	.4	1.1	1.8	3.0	4.2	5.3	6.4	7.5	8.7	9.7	10.7	11.7	13.7	15.7	17.7	19.7	21.7	26.7	31.7	41.7	51.7
850	.4			3.0		5.3		7.5		9.6		11.6	13.6	15.6	17.6	19.6	21.6		31.6	41.6	51.6
900	.3		1.7	3.0	4.1	5.2	6.3	7.5	8.5	9.5	10.5	11.5	13.5	15.5	17.5	19.5	21.5	26.5	31.5	41.5	51.5
950	.3			2.9		5.2		7.4		9.4		11.4	13.4	15.5	17.4	19.5	21.4		31.5	41.5	51.5
1000	.3	1.0	1.7	2.9	4.0	5.2	6.3	7.4	8.4	9.4	10.4	11.4	13.4	15.4	17.4	19.4	21.4	26.4	31.4	41.4	51.4
1500	.2		1.5	2.7	3.8	4.9	5.9	6.9	7.9	8.9	9.9	10.9	12.9	14.9	16.9	18.9	20.9	25.9	30.9	40.9	50.9
2000	.1	.8	1.4	2.6	3.7	4.7	5.7	6.7	7.7	8.7	9.7	10.7	12.7	14.7	16.7	18.7	20.7	25.7	30.7	40.7	50.7
2500	.1		1.4	2.6	3.6	4.6	5.6	6.6	7.6	8.6	9.6	10.6	12.6	14.6	16.6	18.6	20.6	25.6	30.6	40.6	50.6
3000	.1	.8	1.4	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5	10.5	12.5	14.5	16.5	18.5	20.5	25.5	30.5	40.5	50.5
4000	.1	.7	1.3	2.4	3.4	4.4	5.4	6.4	7.4	8.4	9.4	10.4	12.4	14.4	16.4	18.4	20.4	25.4	30.4	40.4	50.4
5000	.1	.7	1.3	2.3	3.3	4.3	5.3	6.3	7.3	8.3	9.3	10.3	12.3	14.3	16.3	18.3	20.3	25.3	30.3	40.3	50.3

* *

Table 3

EVALUATION OF SAMPLE RESULTS

	Number of Occurrences	UPL* RELIABILITY LEVEL 90%	95%	99%
SAMPLE SIZE 25	0	8.80	11.29	16.82
	1	14.69	17.61	23.75
	2	19.91	23.10	29.59
	3	24.80	28.17	34.88
	4	29.47	32.96	39.79

	Number of Occurrences	UPL* RELIABILITY LEVEL 90%	95%	99%
SAMPLE SIZE 50	0	4.50	5.82	8.80
	1	7.56	9.14	12.55
	2	10.30	12.06	15.77
	3	12.88	14.78	18.72
	4	15.35	17.38	21.50
	5	17.76	19.88	24.15
	6	20.11	22.32	26.71
	8	24.69	27.02	31.61

	Number of Occurrences	UPL* RELIABILITY LEVEL 90%	95%	99%
SAMPLE SIZE 75	0	3.02	3.92	5.96
	1	5.09	6.17	8.53
	2	6.94	8.16	10.74
	3	8.69	10.01	12.78
	4	10.38	11.79	14.70
	5	12.02	13.51	16.55
	6	13.62	15.18	18.34
	7	15.20	16.82	20.08
	8	16.75	18.42	21.77
	12	22.78	24.63	28.25

	Number of Occurrences	UPL* RELIABILITY LEVEL 90%	95%	99%
SAMPLE SIZE 100	0	2.28	2.95	4.50
	1	3.83	4.66	6.45
	2	5.23	6.16	8.14
	3	6.56	7.57	9.70
	4	7.83	8.92	11.17
	5	9.08	10.23	12.58
	6	10.29	11.50	13.95
	7	11.49	12.75	15.29
	8	12.67	13.97	16.59
	9	13.83	15.18	17.87
	10	14.99	16.37	19.13
	11	16.13	17.55	20.37
	15	20.61	22.15	25.18

	Number of Occurrences	UPL* RELIABILITY LEVEL 90%	95%	99%
SAMPLE SIZE 125	0	1.83	2.37	3.62
	1	3.08	3.74	5.19
	2	4.20	4.95	6.55
	3	5.27	6.09	7.81
	4	6.29	7.17	9.00
	5	7.29	8.23	10.15
	6	8.27	9.25	11.26
	7	9.24	10.26	12.34
	8	10.19	11.25	13.40
	9	11.13	12.23	14.44
	10	12.06	13.19	15.47
	11	12.98	14.15	16.48
	12	13.89	15.09	17.47
	13	14.80	16.03	18.45
	19	20.14	21.50	24.16

	Number of Occurrences	UPL* RELIABILITY LEVEL 90%	95%	99%
SAMPLE SIZE 150	0	1.52	1.98	3.02
	1	2.57	3.12	4.34
	2	3.51	4.14	5.49
	3	4.40	5.09	6.54
	4	5.26	6.00	7.54
	5	6.10	6.88	8.50
	6	6.92	7.74	9.44
	7	7.72	8.59	10.35
	8	8.52	9.42	11.24
	9	9.31	10.24	12.12
	10	10.09	11.05	12.98
	11	10.86	11.85	13.83
	12	11.62	12.64	14.67
	13	12.39	13.43	15.50
	14	13.14	14.21	16.32
	15	13.89	14.98	17.13
	16	14.64	15.75	17.94
	23	19.79	21.02	23.42

*Upper Precision Limit

Table 3

EVALUATION OF SAMPLE RESULTS

SAMPLE SIZE 175

Number of Occurrences	UPL RELIABILITY LEVEL 90%	95%	99%
0	1.31	1.70	2.60
1	2.20	2.68	3.73
2	3.01	3.55	4.72
3	3.78	4.37	5.63
4	4.52	5.15	6.49
5	5.24	5.91	7.32
6	5.94	6.65	8.12
7	6.63	7.38	8.91
8	7.32	8.10	9.68
9	8.00	8.80	10.43
10	8.67	9.50	11.18
11	9.33	10.19	11.91
12	9.99	10.87	12.64
13	10.65	11.55	13.36
14	11.30	12.22	14.07
15	11.95	12.89	14.77
16	12.59	13.55	15.47
17	13.23	14.21	16.16
18	13.87	14.87	16.85
27	19.51	20.64	22.84

SAMPLE SIZE 200

Number of Occurrences	UPL RELIABILITY LEVEL 90%	95%	99%
0	1.14	1.49	2.28
1	1.93	2.35	3.27
2	2.64	3.11	4.14
3	3.31	3.83	4.93
4	3.96	4.52	5.69
5	4.59	5.18	6.42
6	5.21	5.83	7.13
7	5.82	6.47	7.82
8	6.42	7.10	8.50
9	7.01	7.72	9.16
10	7.60	8.33	9.82
11	8.18	8.94	10.46
12	8.76	9.54	11.10
13	9.34	10.14	11.73
14	9.91	10.73	12.36
15	10.48	11.31	12.98
16	11.04	11.90	13.59
17	11.61	12.48	14.20
18	12.17	13.05	14.81
19	12.73	13.63	15.41
20	13.28	14.20	16.01
21	13.84	14.77	16.60
30	18.75	19.79	21.82

Table 3

EVALUATION OF SAMPLE RESULTS

SAMPLE SIZE 225

Number of Occurrences	UPL RELIABILITY LEVEL 90%	95%	99%
0	1.02	1.32	2.03
1	1.72	2.09	2.91
2	2.35	2.77	3.68
3	2.94	3.41	4.39
4	3.52	4.02	5.07
5	4.08	4.62	5.72
6	4.63	5.19	6.35
7	5.18	5.76	6.97
8	5.71	6.32	7.57
9	6.24	6.88	8.17
10	6.76	7.42	8.75
11	7.28	7.96	9.33
12	7.80	8.50	9.90
13	8.31	9.03	10.46
14	8.82	9.56	11.02
15	9.33	10.08	11.58
16	9.83	10.60	12.12
17	10.34	11.12	12.67
18	10.84	11.63	13.21
19	11.33	12.14	13.75
20	11.83	12.65	14.28
21	12.32	13.16	14.81
22	12.82	13.67	15.34
23	13.31	14.17	15.87
34	18.63	19.60	21.50

SAMPLE SIZE 250

Number of Occurrences	UPL RELIABILITY LEVEL 90%	95%	99%
0	.92	1.19	1.83
1	1.55	1.88	2.63
2	2.11	2.50	3.32
3	2.65	3.07	3.96
4	3.17	3.62	4.57
5	3.68	4.16	5.16
6	4.17	4.68	5.73
7	4.66	5.19	6.29
8	5.15	5.70	6.83
9	5.62	6.20	7.37
10	6.10	6.69	7.89
11	6.56	7.18	8.42
12	7.03	7.66	8.93
13	7.49	8.14	9.44
14	7.95	8.62	9.95
15	8.41	9.09	10.45
16	8.86	9.56	10.94
17	9.32	10.03	11.43
18	9.77	10.49	11.92
19	10.22	10.95	12.41
20	10.66	11.41	12.89
21	11.11	11.87	13.37
22	11.55	12.33	13.85
23	12.00	12.78	14.33
24	12.44	13.23	14.80
25	12.88	13.69	15.27
26	13.32	14.14	15.74
38	18.52	19.44	21.23

Table 3

EVALUATION OF SAMPLE RESULTS

SAMPLE SIZE 275

Number of Occurrences	UPL Reliability Level 90%	95%	99%
0	.83	1.08	1.66
1	1.41	1.71	2.39
2	1.92	2.27	3.02
3	2.41	2.80	3.61
4	2.89	3.30	4.16
5	3.35	3.78	4.70
6	3.80	4.26	5.22
7	4.24	4.73	5.72
8	4.68	5.19	6.22
9	5.12	5.64	6.71
10	5.55	6.09	7.19
11	5.97	6.53	7.67
12	6.40	6.97	8.14
13	6.82	7.41	8.60
14	7.24	7.84	9.06
15	7.65	8.27	9.52
16	8.07	8.70	9.97
17	8.48	9.13	10.42
18	8.89	9.55	10.87
19	9.30	9.97	11.31
20	9.71	10.39	11.75
21	10.11	10.81	12.19
22	10.52	11.23	12.62
23	10.92	11.64	13.06
24	11.32	12.05	13.49
25	11.73	12.47	13.92
26	12.13	12.88	14.35
27	12.53	13.28	14.77
28	12.92	13.69	15.20
42	18.42	19.30	21.00

SAMPLE SIZE 300

Number of Occurrences	UPL Reliability Level 90%	95%	99%
0	.76	.99	1.52
1	1.29	1.57	2.19
2	1.76	2.08	2.77
3	2.21	2.56	3.31
4	2.65	3.02	3.82
5	3.07	3.47	4.31
6	3.48	3.91	4.79
7	3.89	4.34	5.25
8	4.30	4.76	5.71
9	4.69	5.18	6.16
10	5.09	5.59	6.60
11	5.48	6.00	7.04
12	5.87	6.40	7.47
13	6.26	6.80	7.90
14	6.64	7.20	8.32
15	7.02	7.59	8.74
16	7.40	7.99	9.16
17	7.78	8.38	9.57
18	8.16	8.77	9.98
19	8.53	9.15	10.39
20	8.91	9.54	10.79
21	9.28	9.92	11.20
22	9.65	10.31	11.60
23	10.02	10.69	12.00
24	10.39	11.07	12.39
25	10.76	11.44	12.79
26	11.13	11.82	13.18
27	11.50	12.20	13.57
28	11.86	12.57	13.96
29	12.23	12.95	14.35
30	12.59	13.32	14.74
31	12.96	13.69	15.13
45	17.98	18.81	20.42

Table 3

EVALUATION OF SAMPLE RESULTS

SAMPLE SIZE 325

Number of Occurrences	UPL RELIABILITY LEVEL 90%	95%	99%
0	.71	.92	1.41
1	1.19	1.45	2.02
2	1.63	1.92	2.56
3	2.04	2.37	3.06
4	2.44	2.79	3.53
5	2.83	3.21	3.98
6	3.22	3.61	4.42
7	3.60	4.01	4.85
8	3.97	4.40	5.28
9	4.34	4.78	5.69
10	4.70	5.16	6.10
11	5.06	5.54	6.51
12	5.42	5.91	6.90
13	5.78	6.28	7.30
14	6.13	6.65	7.69
15	6.49	7.02	8.08
16	6.84	7.38	8.46
17	7.19	7.74	8.85
18	7.54	8.10	9.23
19	7.88	8.46	9.60
20	8.23	8.82	9.98
21	8.58	9.17	10.35
22	8.92	9.52	10.72
23	9.26	9.88	11.09
24	9.60	10.23	11.46
25	9.94	10.58	11.83
26	10.28	10.93	12.19
27	10.62	11.27	12.55
28	10.96	11.62	12.92
29	11.30	11.97	13.28
30	11.64	12.31	13.64
31	11.97	12.66	13.99
32	12.31	13.00	14.35
33	12.64	13.34	14.71
49	17.93	18.73	20.27

SAMPLE SIZE 350

Number of Occurrences	UPL RELIABILITY LEVEL 90%	95%	99%
0	.66	.85	1.31
1	1.11	1.35	1.88
2	1.51	1.79	2.38
3	1.90	2.20	2.84
4	2.27	2.60	3.28
5	2.63	2.98	3.70
6	2.99	3.36	4.11
7	3.34	3.72	4.51
8	3.69	4.09	4.91
9	4.03	4.44	5.29
10	4.37	4.80	5.67
11	4.70	5.15	6.05
12	5.04	5.50	6.42
13	5.37	5.84	6.79
14	5.70	6.18	7.15
15	6.03	6.52	7.51
16	6.36	6.86	7.87
17	6.68	7.20	8.23
18	7.00	7.53	8.58
19	7.33	7.86	8.93
20	7.65	8.20	9.28
21	7.97	8.53	9.63
22	8.29	8.85	9.97
23	8.61	9.18	10.32
24	8.92	9.51	10.66
25	9.24	9.83	11.00
26	9.56	10.16	11.34
27	9.87	10.48	11.68
28	10.19	10.80	12.01
29	10.50	11.13	12.35
30	10.82	11.45	12.68
31	11.13	11.77	13.02
32	11.44	12.09	13.35
33	11.75	12.40	13.68
34	12.06	12.72	14.01
35	12.37	13.04	14.34
36	12.68	13.36	14.67
53	17.89	18.66	20.14

Table 3

EVALUATION OF SAMPLE RESULTS

SAMPLE SIZE 375

Number of Occur-rences	UPL RELIABILITY LEVEL 90%	95%	99%
0	.61	.80	1.22
1	1.03	1.26	1.76
2	1.41	1.67	2.22
3	1.77	2.05	2.65
4	2.12	2.42	3.06
5	2.46	2.78	3.46
6	2.79	3.13	3.84
7	3.12	3.48	4.22
8	3.44	3.82	4.58
9	3.76	4.15	4.94
10	4.08	4.48	5.30
11	4.39	4.81	5.65
12	4.70	5.13	6.00
13	5.01	5.45	6.34
14	5.32	5.77	6.68
15	5.63	6.09	7.02
16	5.94	6.41	7.35
17	6.24	6.72	7.69
18	6.54	7.03	8.02
19	6.84	7.35	8.35
20	7.14	7.66	8.67
21	7.44	7.96	9.00
22	7.74	8.27	9.32
23	8.04	8.58	9.64
24	8.34	8.88	9.96
25	8.63	9.19	10.28
26	8.93	9.49	10.60
27	9.22	9.79	10.91
28	9.52	10.09	11.23
29	9.81	10.39	11.54
30	10.10	10.69	11.86
31	10.39	10.99	12.17
32	10.69	11.29	12.48
33	10.98	11.59	12.79
34	11.27	11.89	13.10
35	11.56	12.18	13.41
36	11.85	12.48	13.71
37	12.14	12.77	14.02
38	12.43	13.07	14.33
57	17.85	18.59	20.01

SAMPLE SIZE 400

Number of Occur-rences	UPL RELIABILITY LEVEL 90%	95%	99%
0	.57	.75	1.14
1	.97	1.18	1.65
2	1.32	1.57	2.08
3	1.66	1.93	2.49
4	1.99	2.27	2.87
5	2.31	2.61	3.24
6	2.62	2.94	3.60
7	2.92	3.26	3.95
8	3.23	3.58	4.30
9	3.53	3.89	4.64
10	3.83	4.20	4.97
11	4.12	4.51	5.30
12	4.41	4.82	5.63
13	4.70	5.12	5.95
14	4.99	5.42	6.27
15	5.28	5.72	6.59
16	5.57	6.01	6.90
17	5.85	6.31	7.21
18	6.14	6.60	7.52
19	6.42	6.89	7.83
20	6.70	7.18	8.14
21	6.98	7.47	8.44
22	7.26	7.76	8.75
23	7.54	8.05	9.05
24	7.82	8.33	9.35
25	8.10	8.62	9.65
26	8.38	8.90	9.95
27	8.65	9.19	10.25
28	8.93	9.47	10.54
29	9.20	9.75	10.84
30	9.48	10.03	11.13
31	9.75	10.32	11.42
32	10.03	10.60	11.71
33	10.30	10.88	12.01
34	10.57	11.15	12.30
35	10.84	11.43	12.59
36	11.12	11.71	12.87
37	11.39	11.99	13.16
38	11.66	12.27	13.45
39	11.93	12.54	13.74
40	12.20	12.82	14.02
41	12.47	13.09	14.31
60	17.54	18.25	19.62

Table 4

SAMPLE SIZES FOR A ZERO EXPECTED OCCURRENCE RATE

Where no errors are found, the upper precision limit, reliability and sample size can be determined by establishing two of the three parameters in the following formula, and solving for the unknown (desired) variable.

$$n = \frac{RF}{UPL}$$

where

n = sample size
RF = reliability factor
UPL = upper precision limit

A table of reliability factors for this purpose is:

Reliability Level	RF	Reliability Level	RF
50%	.7	86%	2.0
55%	.8	88%	2.1
59%	.9	89%	2.2
63%	1.0	90%	2.3
67%	1.1	91%	2.4
70%	1.2	92%	2.5
73%	1.3	93%	2.6
75%	1.4	94%	2.8
78%	1.5	95%	3.0
80%	1.6	96%	3.2
82%	1.7	97%	3.4
83%	1.8	98%	3.7
85%	1.9	99%	4.3
		99%+	5.4

To illustrate the use of this table, suppose the auditor desires an upper precision limit of 5% at a reliability of 95%, and expects to observe no occurrences. The sample size would be 60, as follows:

$$n = \frac{3.0}{.05}$$
$$= 60$$

If an auditor selected 100 sample items and found zero occurrences but could tolerate a 4% occurrence rate, he could evaluate his achieved reliability at a 4% upper precison limit as follows:

$$100 = \frac{RF}{.04}$$

$$CLF = 4.0$$

Looking at the table, he finds the closest entry smaller than 4.0 is 3.7, and the reliability is 98%. Notice that in using the table in this way, you must always use the tabled factor equal to or smaller than the computed factor.

This table may also be used as a discovery sampling table. When used for this purpose, the interpretation of the symbols changes. The "reliability level" becomes the "probability of discovery" and the corresponding RF may be called the "discovery factor." The upper precision limit, UPL, becomes the population occurrence rate for which a desired probability of discovering an occurrence is assessed.

To illustrate, suppose the auditor desires a 95% probability of discovering a sample occurrence when the population occurrence rate is 3%. In this case, he chooses a UPL equal to .03, RF = 3.0 corresponding to the 95% probability of discovery, and consequently the required sample size is 100, since

$$n = \frac{3.0}{.03}$$

$$= 100$$

Table 5

Values of U_R and U_R^2

Percent Desired Reliability R	One-Sided Estimates: Normal Deviate U_R	One-Sided Estimates: Square Of The Normal Deviate U_R^2	Two-Sided Estimates: Normal Deviate U_R	Two-Sided Estimates: Square Of The Normal Deviate U_R^2
70	.52	.27	1.04	1.08
75	.67	.45	1.15	1.32
80	.84	.71	1.28	1.64
85	1.04	1.08	1.44	2.07
90	1.28	1.64	1.64	2.69
95	1.64	2.69	1.96	3.84
99	2.33	5.43	2.58	6.66

APPENDIX C

COMPUTER TIME-SHARING

Contents

INTRODUCTION

Computer techniques have proven invaluable to users of statistical sampling. These techniques are of two general types: (1) programs which perform statistical routines on files of machine readable data - for example, as part of a generalized audit software program - and (2) computer time-sharing based programs which generally deal with sample items only.

This section presents five time-sharing programs designed to select and evaluate sample items using the mean-per-unit, ratio, and difference methods described earlier in the volume. These programs have been tested with two examples involving real data. Programs for attribute estimation and stratified sampling are not included. Attribute estimation generally does not require computer assistance because of the availability of practical tables. Where stratification is used, each stratum can be evaluated separately with these programs and then combined manually, using the appropriate forms included in the volume.

Readers should follow the procedures listed below to implement these time-sharing programs in their practices:

1. Select a time-sharing vendor. Factors to be considered are service, efficiency, storage and operating cost. The system selected must have sufficient capability to handle large numerical amounts. Often vendors will implement time-sharing programs at no initial set-up charge. Many also offer free training. Several articles have been published in the *Journal of Accountancy* and similar publications regarding the selection of time-sharing vendors.

2. Review the program listings with the vendor and agree on changes. Mark your copy with all changes made.

3. After the vendor loads the program, produce a listing and compare it to your copy.

4. Run the program, using appropriate data from "Example Data Table" examples included herein. Compare results- they should agree with examples.

5. Run programs with more extensive data of your choosing. Make intentional errors. Become familiar with the vendor's operating system under various conditions. Become comfortable with the program.

At this point you will be ready to use these programs in live situations. As you do, there are several user caveats you should always keep in mind:

1. It is possible for distorted time-sharing transmission lines to result in erroneous output. You should always re-dial whenever you receive an unclear reception tone or "garbage" printout.

2. The program may protect you against format errors, but it cannot recognize erroneous or invalid input data values. You should always check input for accuracy. This is easy to do because it is printed out for your inspection.

3. You should also guard against accepting results on blind faith. Always examine them for overall reasonableness even though input details may check out.

4. The Program for Random Number Generation uses the vendor's random number generator. You may want assurance from the vendor that his random number generator has been tested for randomness of the output.

EXAMPLE DATA TABLE

ITEM NUMBER	BOOK VALUE	AUDITED VALUE	DIFFERENCE	BOOK VALUE SQUARED	AUDITED VALUE SQUARED	BOOK TIMES AUDITED VALUE	DIFFERENCE SQUARED
1	$ 110	$ 100	$ 10	$ 12,100	$ 10,000	$ 11,000	$ 100
2	150	150	0	22,500	22,500	22,500	0
3	198	200	(2)	39,204	40,000	39,600	4
4	170	170	0	28,900	28,900	28,900	0
5	180	180	0	32,400	32,400	32,400	0
6	200	172	28	40,000	29,584	34,400	784
7	300	302	(2)	90,000	91,204	90,600	4
8	800	715	85	640,000	511,225	572,000	7,225
9	312	312	0	97,344	97,344	97,344	0
10	100	101	(1)	10,000	10,201	10,100	1
11	89	89	0	7,921	7,921	7,921	0
12	125	126	(1)	15,625	15,876	15,750	1
13	300	201	99	90,000	40,401	60,300	9,801
14	198	198	0	39,204	39,204	39,204	0
15	160	166	(6)	25,600	27,556	26,560	36
16	400	400	0	160,000	160,000	160,000	0
17	112	112	0	12,544	12,544	12,544	0
18	8	8	0	64	64	64	0
19	88	88	0	7,744	7,744	7,744	0
20	310	310	0	96,100	96,100	96,100	0
21	260	260	0	67,600	67,600	67,600	0
22	100	112	(12)	10,000	12,544	11,200	144
23	90	70	20	8,100	4,900	6,300	400
24	113	113	0	12,769	12,769	12,769	0
25	156	156	0	24,336	24,336	24,336	0
26	240	244	(4)	57,600	59,536	58,560	16
27	190	190	0	36,100	36,100	36,100	0
28	200	198	2	40,000	39,204	39,600	4
29	116	116	0	13,456	13,456	13,456	0
30	200	200	0	40,000	40,000	40,000	0
Totals	$5,975	$5,759	$216	$1,777,211	$1,591,213	$1,674,952	$18,520

NOTE: Appropriate data items from this table are used as sample input to the computer programs BBB, CCC, DDD and EEE.

PROGRAM FOR RANDOM NUMBER GENERATION

I - Program Description and Operating Instructions

General Description

This program - designated "AAA" for purposes of this volume - generates random numbers and allows the user to select one of the following three output options:

Option 1 - Will produce RANDOM DAYS for a given period.

Option 2 - Will produce RANDOM PAGE AND LINE NUMBERS.

Option 3 - Will produce RANDOM DOCUMENT NUMBERS FROM ONE POPULATION OF BROKEN SEQUENCES.

The program rejects duplications, which results in sampling without replacement. Because of rejections, the number of samples printed may vary slightly from the number requested. Excessive samples may be reduced by eliminating the last items selected, as indicated in the order of selection column. It is unlikely that too few samples will be obtained. If this happens, the auditor may either reduce the sample accordingly, or generate a few additional numbers using the same parameters.

The maximum number of random numbers obtainable is approximately twelve hundred (1200).

All output from "AAA" is in a format for use as an audit working paper. Transcribing the numbers obtained from this program should be unnecessary. Note in Example 3 the use of a lined acetate overlay and xerox machine to provide the basic checklist for the audit procedures and to summarize the results for evaluation.

Note that this program is designed so that a different set of random numbers is generated even though the same input parameters are given in successive applications.

By eliminating instruction 100 RANDOM, the program may be modified so that an identical set is always obtained.

OPTION 1 - Random days of the Year

Using Option 1, the auditor may generate random days from a fiscal or calendar period of one to twelve months for one or more locations comprising a single population. He may exclude Sundays or weekends from the sample. The program provides for leap years when applicable.

Data Requirements

The user must furnish, via the keyboard, the following:

1. Number of random samples desired.
2. Number of months in the sample (1 to 12).
3. Number of locations in the same population.
4. Days to be excluded (D)

Saturdays, Sundays	(2)
Sundays only	(1)
No excluded days	(0)

5. For D = 2 or D = 1:

   Month, day, and year of first Sunday in the period.

   For D = 0:

   Month, day and year the period begins.

Operating Instructions

The sample run (Example 1) illustrates the use of the program. User input is underlined. Multiple input (samples, months, locations - month, day, year) must be separated by commas.

Comments, Limitations, Uses

All periods must be calendar months. If a fiscal period extends into 13 or 14 months (for example, a 53-week year from March 28, 1971 to April 1, 1972), the auditor may wish to regard the extra days (March 28, 29, 30 and 31, 1971, and April 1, 1972) as a separate population, or may devise some method of establishing correspondence between random numbers and the population, by using some other program.

OPTION 2 - Random Page and Line Numbers

Option 2 is a fast method of obtaining a sample if the population is from a uniform listing (journal pages, inventory sheets, etc.).

Data Requirements

Data is supplied, via the keyboard, as follows:

1. Number of random samples desired.
2. Number of pages in the population.
3. Number of lines per page.

The sample run (Example 2) with user supplied information underlined, illustrates the use of the program.

Extra numbers may have to be generated if the population includes unused lines or pages.

II - Examples of Operation

Example 1

```
LOA AAA
READY
RUN

* AAA  20:06   09/13/73


THIS PROGRAM GENERATES YOUR CHOICE OF:

    1...RANDOM DAYS OF THE YEAR
    2...RANDOM PAGE AND LINE NUMBERS
    3...RANDOM DOCUMENT NUMBERS FROM ONE POPULATION
        OF BROKEN SEQUENCES

WHICH DO YOU WANT OPTION 1, 2, OR 3 TO STOP PRESS ESCAPE KEY?  1

HOW MANY SAMPLES, MONTHS, LOCATIONS?    10, 2, 2

EXCLUDE WHICH DAYS:

       2 = SATURDAYS AND SUNDAYS
       1 = SUNDAYS ONLY
       0 = NO EXCLUDED DAYS?  2

ENTER THE MONTH, DAY, YEAR OF THE FIRST SUNDAY IN THE PERIOD? 1, 2, 72

THANK YOU




RANDOM
 ORDER  MONTH  DAY


LOCATION   #1

   11          1     4

    1          2     8
    4          2    21
    7          2    25
```

Example 1, contd.

```
LOCATION  #2

   5      1     6
   9      1    10
   3      1    12

   8      2     2
  10      2    15
   2      2    22


RANDOM NUMBERS PRINTED - - - 10

TOTAL POPULATION - - - 84

OPTION 1, 2, OR 3 TO STOP PRESS ESCAPE KEY?

NOW AT 260

READY
```

Example 2

```
RUN

AAA    20:11    09/13/73

THIS PROGRAM GENERATES YOUR CHOICE OF:

    1...RANDOM DAYS OF THE YEAR
    2...RANDOM PAGE AND LINE NUMBERS
    3...RANDOM DOCUMENT NUMBERS FROM ONE POPULATION
        OF BROKEN SEQUENCES

WHICH DO YOU WANT OPTION 1, 2, OR 3 TO STOP PRESS ESCAPE KEY? 2

HOW MANY SAMPLES, PAGES, LINES PER PAGE? 10, 10, 10


THANK YOU



ORDER        PAGE & LINE

   1           1      7
   9           1      8
   4           1     10

   5           2     10

   6           3      1

   3           4      3

   2           5      3

   8          10      3

RANDOM NUMBERS PRINTED - - - 8

TOTAL POPULATION - - - 100

OPTION 1, 2, OR 3 TO STOP PRESS ESCAPE KEY?

NOW AT 260

READY
```

Example 3

Example of using an acetate overlay and xerox machine to provide an audit working paper from the output to the program.

ORDER OF SELECTION	PAGE NUMBER	LINE NUMBER	Audit Test #1	#2	#3	#4	#5	Comments
6	3	6	✓	✓	✓	✓	✓	
10	3	9	✓	✓		✓	✓	Missing Voucher
1	4	2	✓	✓	✓	✓	✓	
9	4	8	✓	✓	✓	✓	✓	
11	7	6	✓	✓	✓		✓	#4 NOT APPLICABLE
5	7	9	✓	✓	✓	✓	✓	
4	9	5	✓	✓	✓	✓	X	SEE W/P G-2
7	9	9	✓	✓	✓	✓	✓	
2	10	2	✓	✓	✓	✓	✓	
3	10	5	✓	✓	✓	✓	✓	
8	10	8	✓	✓	✓	✓	✓	
Total Exceptions			0	0	1	0	1	
Sample Size			11	11	11	10	11	

RANDOM NUMBERS PRINTED - - - - 11

TOTAL POPULATION - - - 100

OPTION 1, 2, 3, 4, OR STOP? STOP

PROGRAM HALTED

OPTION 3 - Random Document Numbers From One Population of Broken Sequences

This program option is a most useful random sample selector. Random numbers are generated from the number of items in the population and made to correspond to the serial number in the population. Input is by keyboard or by data file. If a large number of sequences is included, preparation and use of a data file greatly reduces terminal time.

Data Requirements

The user must furnish:

1. The number of random numbers desired.
2. The number of groups of numbers in the population.
3. For each group, the lowest and highest number.

For Large Numbers of Sequences

Prepare a tape "offline" in data file format. The data required is listed above. Example 4A illustrates the steps. In this example, the user requested a selection of 80 items from a population of 367 sequences. Accordingly, the first data item is 80 (number of samples), the second is 367 (number of groups of sequences) and the remaining items are the lowest and highest numbers in each of the 367 groups.

When the tape has been completed, call the time-sharing service, name the data file, load the tape and "save" the file. Call AAA and run the program. The computer will ask for the data file name, which must be entered via the keyboard. The program will then be executed with no further user-supplied information.

For Small Numbers of Sequences

Example 5E illustrates the input via keyboard. User-supplied information is underlined.

Program Error Messages

Excessive or insufficient data in the data file will cause the program to disregard the file, after printing an appropriate message. (Examples 5A and 5B demonstrate these errors.) The data file should be corrected before execution is attempted. If the second number of a set of numbers is smaller than the first, the program will request a correction (see Example 5C). Example 5D shows old file before correction (see line 30). If the error is due to an omitted number, enter "STOP" and correct the data file; if not, enter the corrected values. Example 5D shows old file before correction (see line 30). The transposition of numbers caused the error message shown on Example 5C. The keyboard input replaced the incorrect values.

Keyboard input will likewise be tested by the program. If the second number of the pair is less than the first, the user will be asked for corrected values (see Example 5E).

Example 4A

OFFLINE TAPE PREPARATION...

ACCESS THE SYSTEM...

LOADING THE DATA FILE...

```
TAPE NN   DATA 1
READ PAPER TAPE
TP ON
100     80,   367
110    101,    99999
120    101,    99999
130    101,    99999
140    101,    99999
150    101,    99999
160    101,    99999

    THIS FILE ACTUALLY CONTAINS 367 DIFFERENT SEQUENCES

3730   101,    99999
3740   101,    99999
3750   101,    99999
3760   101,    99999
3770   101,    99999

FILE DATA 1 SAVED
END OF PAPER TAPE INPUT
READY
```

Example 4A, contd.

SAVING THE FILE...

```
SAVE DATA 1
FILE DATA 1 SAVED
READY
```

CALLING THE PROGRAM...

```
LOA AAA
READY
RUN
AAA  20:41   09/13/73

THIS PROGRAM GENERATES YOUR CHOICE OF:

    1...RANDOM DAYS OF THE YEAR
    2...RANDOM PAGE AND LINE NUMBERS
    3...RANDOM DOCUMENT NUMBERS FROM ONE POPULATION
        OF BROKEN SEQUENCES

WHICH DO YOU WANT OPTION 1, 2, OR 3 TO STOP PRESS ESCAPE KEY? 3

IS THE INPUT STORED IN A DATA FILE? YES

ENTER THE DATA FILE NAME? DATA 1

THANK YOU
```

Example 4B - Output

```
ORDER DOCUMENT

GROUP   # 2        101 - 99999

 56      34729

GROUP   # 20       101 - 99999

 17      47860

GROUP   # 27       101 - 99999

 35      45726

GROUP   # 36       101 - 99999

 40      20331
 45      92525

GROUP   # 365      101 - 99999

 27      34616
 47      94515

GROUP   # 367      101 - 99999

 57      73641

RANDOM NUMBERS PRINTED - - -  80

TOTAL POPULATION - - - 36662933

PURGE FILE TEST 1

OPTION 1, 2, OR 3 TO STOP PRESS ESCAPE KEY? @

NOW AT 260

READY
```

Example 5 - Examples of System Messages

```
5A  IS THE INPUT STORED IN A DATA FILE? YES

    ENTER THE DATA FILE NAME? DATA 2

    INSUFFICIENT DATA IN FILE DATA 2.  CHECK YOUR PARAMETERS.

    OPTION 1, 2, OR 3 TO STOP PRESS ESCAPE KEY? @

    NOW AT 260

    RUN
```

```
5B  IS THE INPUT STORED IN A DATA FILE? YES

    ENTER THE DATA FILE NAME? DATA 1

    DATA IN FILE DATA 1 NOT EXHAUSTED.  CHECK YOUR PARAMETERS.

    OPTION 1, 2, OR 3 TO STOP PRESS ESCAPE KEY? @
```

```
5C  IS THE INPUT STORED IN A DATA FILE? YES

    ENTER THE DATA FILE NAME? DATA 3

    GROUP 17 BEGINNING - ENDING 420 240 CORRECTION PLEASE
     240,420
```

```
5D  OLD DATA 3
    READY
    EDI LIS

    10 500,50,345,876,234,987,5768,8654,345,876,123,321,987,1098,345,987
    20 23,54,65,78,45,65,45,98,123,543,234,765,345,876,143,987,456,876
    30 420,240
    40 1,10,9876,10987,234,654,1,7,4,10,236,873,109987,109999,23,65,1,10
    50 2,12,13,15,876,987,456,789,101,202,303,404,123,345,876,921,2,8
    60 1,4,3,8,10,20,43,65,20,30,76,85,1001,1901,5000,6000,87,98
    70 240,420,1,10,2,23,4,45,6,61,3,79
```

```
5E  IS THE INPUT STORED IN A DATA FILE?  NO

    HOW MANY SAMPLES, HOW MANY GROUPS? 10,2

    FOR GROUP   ENTER LOWEST AND HIGHEST NUMBER
      # 1            10,9
    PLEASE CORRECT   10,90
      # 2            100,180
```

```
97.        REM*****PROGRAM AAA 9/20/74
98.         REM*****PRELIMINARY PROCESSING
99.         REM*****PROGRAM CONSTANTS
100.        RANDOM
101.        LET X9=1500
102.        LET X0=X9+1
150.        PRINT
160.        PRINT"THIS PROGRAM GENERATES YOUR CHOICE OF:"
170.        PRINT
180.        PRINT"    1...RANDOM DAYS OF THE YEAR"
190.        PRINT"    2...RANDOM PAGE AND LINE NUMBERS"
200.        PRINT"    3...RANDOM DOCUMENT NUMBERS FROM ONE POPULATION"
210.        PRINT"        OF BROKEN SEQUENCES"
230.        PRINT
240.        PRINT"WHICH DO YOU WANT ";
250.        PRINT "OPTION 1, 2, OR 3";
251.        PRINT " TO STOP PRESS ESCAPE KEY";
260.        INPUT N7
290.        IF N7>3 THEN 240
295.        IF N7<1 THEN 240
310.        PRINT
350.        ON N7 GO TO 940,1510,650
360.        REM*****OPTION THREE-PHASE 1 (DATA FILE INPUT SECTION)
640.        DIM X(2,1500),S(1500),B(1500),J(366),A(12)
650.        PRINT
660.        PRINT"IS THE INPUT STORED IN A DATA FILE";
670.        INPUT G$
680.        IF G$ ="NO" THEN 1630
690.        PRINT
700.        PRINT"ENTER THE DATA FILE NAME";
710.        INPUT F$
720.        PRINT
730.        FILE #1 = F$
740.        INPUT #1,N1,S
750.       S1=S
760.        FOR L= 1 TO S
770.        INPUT #1,B(L),S(L)
780.        IF B(L)<>0 THEN 810
790.        PRINT"INSUFFICIENT DATA IN FILE ";F$;" . CHECK YOUR PARAMETERS."
800.        GO TO 2860
810.        IF S(L)-B(L)>=0 THEN 860
820.        PRINT"GROUP ";L;" BEGINNING - ENDING ";B(L);S(L);
830.        PRINT" CORRECTION PLEASE "
840.        INPUT B(L),S(L)
850.        PRINT
860.       S(L)=S(L)-B(L)+1
870.       O=O+S(L)
880.        NEXT L
890.        IF END #1 THEN 1790
900.        PRINT"DATA IN FILE ";F$;" NOT EXHAUSTED. CHECK YOUR PARAMETERS."
910.        GO TO 2860
920.       :###  #######
930.       :#####      #### ####
935.        REM*****OPTION ONE-PHASE 1
```

```
940.        PRINT
950.        PRINT"HOW MANY SAMPLES,MONTHS,LOCATIONS";
960.        INPUT N1,M8,L8
970.        PRINT
980.        PRINT"EXCLUDE WHICH DAYS:"
995.        PRINT
1000.       PRINT "      2 = SATURDAYS AND SUNDAYS"
1010.       PRINT "      1 = SUNDAYS ONLY"
1020.       PRINT "      0 = NO EXCLUDED DAYS";
1030.       INPUT X5
1040.       PRINT
1050.      D5=29
1060.       IF X5<>0 THEN 1110
1070.       PRINT"ENTER THE MONTH,DAY,YEAR THE PERIOD BEGINS";
1080.       INPUT M5,D6,Y8
1090.       IF D6<>1 THEN 1070
1100.       GO TO 1140
1110.       PRINT"ENTER THE MONTH,DAY,YEAR OF THE FIRST SUNDAY IN THE PERIOD";
1120.        INPUT M5,X6,Y8
1130.       IF X6>7 THEN 1110
1140.       IF M5<=2 THEN 1160
1150.      Y8=Y8+1
1160.       IF Y8/4=INT(Y8/4)THEN 1180
1170.      D5=28
1180.       PRINT
1190.       IF X5 = 1 THEN 1250
1200.       IF X5= 0 THEN 1250
1210.       IF X6=1 THEN 1240
1220.      Y5=X6-1
1230.       GO TO 1250
1240.      Y5=7
1250.       FOR L=1 TO (M5-1)
1260.       READ A(L),D6
1270.       NEXT L
1280.       FOR L = 1 TO M8
1290.      M= L-1
1300.       READ A(L),S(L)
1310.       IF S(L)>0 THEN 1330
1320.      S(L)=D5
1330.      S(L)=S(M)+S(L)
1340.      Q9=S(L)
1350.       NEXT L
1360.      N=0
1370.       FOR L = 1 TO Q9
1380.       IF L<>Y5 THEN 1410
1390.      Y5=Y5+7
1400.       GO TO 1460
1410.       IF L<>X6 THEN 1440
1420.      X6=X6+7
1430.       GO TO 1460
1440.      N=N+1
1450.      J(N)=L
1460.       NEXT L
1470.      Q9=N
```

```
1480.      O=N*L8
1490.      S=S1=12
1500.       GO TO 1790
1505.       REM*****OPTION TWO-PHASE 1
1510.       PRINT
1520.       PRINT"HOW MANY SAMPLES,PAGES,LINES PER PAGE";
1530.       INPUT N1,S,O9
1532.       IF O9<2  THEN 1545
1533.       IF O9<10000 THEN1541
1541.       IF S<1 THEN 1545
1542.       IF S<X0 THEN 1550
1545.       GOSUB 3480
1546.       GO TO 1510
1550.       S1=S
1560.      O=S*O9
1570.       FOR L= 1 TO S
1580.      S(L)=O9
1590.      B(L)=1
1600.       NEXT L
1610.       PRINT
1620.       GO TO 1790
1625.       REM*****OPTION THREE-PHASE 2 (INTERACTIVE INPUT SECTION)
1630.       PRINT
1640.       PRINT"HOW MANY SAMPLES,HOW MANY GROUPS";
1650.       INPUT N1,S
1660.       S1=S
1670.       PRINT
1680.       PRINT"FOR GROUP      ENTER LOWEST AND HIGHEST NUMBER"
1690.       FOR L=1 TO S
1700.       PRINT"   #";L,
1710.       INPUT B(L),S(L)
1720.       IF S(L)-B(L)>=0 THEN 1750
1730.       PRINT "PLEASE CORRECT";
1740.       GO TO 1710
1750.      S(L)=S(L)-B(L)+1
1760.      O=O+S(L)
1770.       NEXT L
1780.       PRINT
1790.       PRINT "THANK YOU"
1795.       REM*****OPTION THREE-PHASE 3 (SAMPLE GENERATION SECTION)
1800.       PRINT
1810.       PRINT
1820.       IF N1<O THEN 1845
1830.       PRINT"SAMPLE SIZE EXCEEDS POPULATION"
1840.       GO TO 2860
1845.       LET M4=N1
1850.      N1=N1*(1+(N1/O))
1852.       IF N1<1 THEN 1855
1853.       IF N1<X0 THEN 1860
1855.       PRINT "FOR THIS PROBLEM, THE ORIGINAL SAMPLE SIZE IS";M4;".  THE"
1856.       PRINT "POPULATION SIZE IS ";O;".  THIS RESULTS IN A DERIVED FUNCTION"
1857.       PRINT "OF ";N1;" WHICH IS GREATER THAN ";X9;"-- THE PRESENT LIMIT."
1858.       GOSUB 3480
1859.       GO TO 2860
```

```
1860.      FOR I=1 TO N1
1870.     X(1,I)=INT((O*RND)+1)
1880.     X(2,I)=I
1890.      NEXT I
1895.      REM*****OPTION THREE-PHASE 4 (SHELL SORT SECTION)
1900.     D=8191
1910.      FOR K=1 TO 12
1920.     D=(D-1)/2
1930.      FOR I = D+1 TO N1
1940.     T=X(1,I)
1950.     T2 =X(2,I)
1960.      J = I
1970.      J = J-D
1980.      IF J>=1 THEN 2010
1990.      J = J+D
2000.      GO TO 2050
2010.      IF T>=X(1,J)THEN 2080
2020.     X(1,J+D)=X(1,J)
2030.     X(2,J+D)=X(2,J)
2040.      GO TO 1970
2050.     X(1,J)=T
2060.     X(2,J)=T2
2070.      GOTO 2100
2080.     X(1,J+D)=T
2090.     X(2,J+D)=T2
2100.      NEXT I
2110.      NEXT K
2120.      D = K = T = T2=J =0
2130.     J(X(1,N1+1))=J(X(2,N1+1))=10E7
2140.     I = 1
2150.      H = 1
2160.      GOSUB 3010
2170.      ON N7 GO TO 2180,2470,2470
2175.      REM*****OPTION ONE--PHASE 2
2180.      FOR H = 1 TO L8
2190.      IF B8 <55 THEN 2220
2200.      GOSUB 3010
2210.      GO TO 2240
2220.      IF H9 = 0 THEN 2240
2230.      GOSUB 3200
2240.      FOR L = 1 TO M8
2245.      H9 = 1
2250.      M9 =A(L)
2260.      M = L-1
2270.      K=X(1,I)-((INT((X(1,I)-1)/09))*09)
2280.      IF J(K)>S(L) THEN 2440
2290.      K = X(1,I)-((INT((X(1,I)-1)/09))*09)
2300.      IF J(K)>S(L) THEN 2420
2310.      GOSUB 3340
2320.      PRINT USING 930,X(2,I),M9,J(K)-S(M)
2340.      B8 = B8+1
2350.      IF B8<70 THEN 2370
2360.      GOSUB 3010
2370.     N9 =N9+1
```

```
2380.      I = I+1
2390.      IF I>N1 THEN 2730
2400.      IF X(1,I)>H*09 THEN 2450
2410.      GO TO 2290
2420.      PRINT
2430.      B8 = B8+1
2440.      NEXT L
2450.      NEXT H
2460.      GO TO 2730
2465.      REM*****OPTION TWO AND THREE
2470.     X(1,N1+1)=X(2,N1+1)=10E7
2480.      PRINT
2490.      FOR L= 1 TO S1
2500.     T1=T1+S(L)
2510.      IF X(1,I)>T1 THEN 2720
2520.      IF B8 <60 THEN 2540
2530.      GOSUB 3010
2540.      IF N7 = 2 THEN 2590
2550.      PRINT"GROUP #";L,B(L);"-";S(L)+B(L)-1
2560.      PRINT
2570.      B8 = B8+2
2580.      IF X(1,I)>T1 THEN 2700
2590.      GOSUB 3340
2600.      IF N7 = 3 THEN 2630
2610.      PRINT USING 930,X(2,I),L,X(1,I)+B(L)-(T1-S(L))-1
2620.      GO TO 2640
2630.      PRINT USING 920,X(2,I),X(1,I)+B(L)-(T1-S(L))-1
2640.     B8=B8+1
2650.     N9 = N9 + 1
2660.     I=I+1
2670.      IF B8<70 THEN 2690
2680.      GOSUB 3010
2690.      GO TO 2580
2700.      PRINT
2710.      B8 = B8+1
2720.      NEXT L
2730.      PRINT
2740.      PRINT
2750.      PRINT
2760.      PRINT "RANDOM NUMBERS PRINTED - - - -";N9
2770.      PRINT
2780.      PRINT"TOTAL POPULATION - - -";O
2790.      IF G$ = "YES" THEN 2810
2800.      GO TO 2860
2810.      PRINT
2820.      PRINT
2830.      PRINT"PURGE FILE "F$
2840.      PRINT
2850.      PRINT
2860.      MAT X= ZER
2870.      FOR L = 1 TO S1
2880.     B(L)=S(L)=0
2890.      NEXT L
2900.      FOR L= 1 TO 366
```

```
2910.     J(L)=0
2920.      NEXT L
2930.     N7=0=N1=M8=S=S1=T1=I=I1=N9=0
2940.      B8 = Q9 = L8 = H9 = 0
2950.      RESTORE
2960.      PRINT
2965.      FILE
2970.      GO TO 250
2980.      DATA 1,31,2,0,3,31,4,30,5,31,6,30,7,31,8,31
2990.      DATA 9,30,10,31,11,30,12,31,1,31,2,0,3,31,4,30
3000.      DATA 5,31,6,30,7,31,8,31,9,30,10,31,11,30
3005.      REM*****SUBROUTINE ONE-ENTRY POINT A
3010.      FOR M6 = 1 TO 6
3020.      PRINT
3030.      NEXT M6
3040.      B8 = 0
3160.      PRINT
3170.      B8 = 6
3180.      ON N7 GO TO 3190,3270,3300
3190.      PRINT "RANDOM"
3191.      PRINT" ORDER  MONTH DAY"
3195.      REM*****SUBROUTINE ONE-ENTRY POINT B
3200.      PRINT
3210.      IF L8 = 1 THEN 3320
3220.      B8 = B8+2
3230.      PRINT
3240.      PRINT"LOCATION #"H
3250.      PRINT
3260.      GO TO 3320
3270.      PRINT "ORDER     PAGE & LINE"
3280.      PRINT
3290.      GO TO 3320
3300.      PRINT "ORDER  DOCUMENT"
3310.      PRINT
3320.      B8 = B8+3
3330.      RETURN
3335.      REM*****SUBROUTINE TWO
3340.      IF X(1,I+1)<>X(1,I) THEN 3400
3350.      I1=I1+1
3360.      IF X(2,I)<X(2,I+1) THEN 3390
3370.      A1 = X(2,I+1)
3380.      GO TO 3400
3390.      A1 = X(2,I)
3400.      IF I1 = 0 THEN 3460
3410.      IF X(1,I+1)<>X(1,I) THEN 3440
3420.      I = I+1
3430.      GO TO 3340
3440.      X(2,I)=A1
3450.      I1=0
3460.      RETURN
3470.      REM*****SUBROUTINE THREE
3480.      PRINT
3490.      PRINT "***YOU HAVE JUST ENTERED SPECIFICATIONS WHICH"
3500.      PRINT "***IF EXECUTED WOULD REQUIRE WORK BEYOND THE"
3510.      PRINT "***CAPACITY OF THIS PROGRAM TO PERFORM.  PLEASE"
3520.      PRINT "***RECHECK YOUR ANSWER AND TRY AGAIN."
3530.      PRINT
3540.      RETURN
9999.      END
```

MEAN-PER-UNIT ESTIMATION - PROGRAM FOR
DETERMINATION OF SAMPLE SIZE

I - Program Description and Operating Instructions

General Description

This program - designated "BBB" for purposes of this volume - is a computational program designed to determine sample size where a statistical evaluation is to be made using mean-per-unit estimation without stratification. The program makes requests for two options:

Option 1 - The PRELIMINARY SAMPLE option - whether the details of a preliminary sample will be entered - is answered yes or no.

Option 2 - The ESTIMATE TYPE option - whether the estimate to be made from the sample is one-sided or two-sided - is answered 1 or 2. If some other number is entered, an error message and request for re-entry will result.

Inputs Required

The following inputs are required for all options:

1. Desired reliability percentage - 70, 75, 80, 85, 90, 95, and 99 (entered as whole numbers) are acceptable. Any other number will produce an error message and request for re-entry.

2. Desired precision - entered as an amount. Any number of decimal places is acceptable, but precision is usually in whole dollars.

3. Total population size - (not to be confused with preliminary sample size).

Where a preliminary sample is specified, additional inputs are:

1. Size of preliminary sample.
2. Sample details - these are entered as DATA statements, using lines 1 through 999. Follow the instructions contained in the program listing. If an error is made in the data statement, merely correct the offending item and rerun the program.

Where no preliminary sample is specified, an estimate of the population variance $\left(S_{X_j}^{2}\right)$ must be entered.

Outputs Produced

1. Computed sample size - given as a whole number.
2. When preliminary sample details are entered, two additional outputs are produced:
   a. Preliminary sample statistics: mean, variance, and standard deviation.
   b. Supplementary information: sum of the sample items and sum of the squares of the sample items.

II - Examples of Operation

Example 1 - Sample details (audited values from Example Data Table) are entered with the following data:

Reliability - 90%

Precision - ± 50,000

Population size - 10,000

A two-sided estimate is desired.

Example 2 - Sample totals are entered with the same data as used in Example 1.

Example 1

```
1 DATA 100 , 150 , 200 , 170 , 180 , 172 , 302 , 715 , 312 , 101
2 DATA 89 , 126 , 201 , 198 , 166 , 400 , 112 , 8 , 88 , 310
3 DATA 260 , 112 , 70 , 113 , 156 , 244 , 190 , 198 , 116 , 200
READY

BBB   19:26   09/19/74

THIS PROGRAM IS USED TO DETERMINE SAMPLE SIZE FOR MEAN PER UNIT
ESTIMATES WHERE STRATIFICATION IS NOT USED.


IS THERE A PRELIMINARY SAMPLE (ENTER YES OR NO)
?YES

WHAT IS THE SIZE OF THE PRELIMINARY SAMPLE
?30

IS THIS A ONE-SIDED ESTIMATE OR A TWO-SIDED ESTIMATE
ENTER 1 OR 2
?2

WHAT RELIABILITY PERCENTAGE IS DESIRED (EXAMPLE 90)
?90

WHAT PRECISION (AMOUNT) IS DESIRED
?50000

ENTER TOTAL POPULATION SIZE
?10000

HAVE PRELIMINARY SAMPLE ITEMS BEEN STORED IN DATA
STATEMENTS (YES OR NO)?YES


THE PRELIMINARY SAMPLE STATISTICS ARE:

          MEAN                        191.97
          VARIANCE                  16747.
          STANDARD DEVIATION          129.41


THE SAMPLE SIZE FOR A TWO-SIDED MEAN PER UNIT ESTIMATE WITH
A RELIABILITY OF  90% AND A PRECISION OF PLUS AND MINUS $  50000
 IS    1527.

SUM OF X        5759
SUM OF X SQR    1591213


PROGRAM COMPLETE

NOW AT 2880
SRU'S:0.5
READY
```

Example 2

```
RUN

BBB   19:29   09/19/74

THIS PROGRAM IS USED TO DETERMINE SAMPLE SIZE FOR MEAN PER UNIT
ESTIMATES WHERE STRATIFICATION IS NOT USED.


IS THERE A PRELIMINARY SAMPLE (ENTER YES OR NO)
?NO

IS THIS A ONE-SIDED ESTIMATE OR A TWO-SIDED ESTIMATE
ENTER 1 OR 2
?2

WHAT RELIABILITY PERCENTAGE IS DESIRED (EXAMPLE 90)
?90

WHAT PRECISION (AMOUNT) IS DESIRED
?50000

ENTER TOTAL POPULATION SIZE
?10000

ENTER THE ESTIMATED STATISTICAL VARIANCE OF THE POPULATION?16747


THE SAMPLE SIZE FOR A TWO-SIDED MEAN PER UNIT ESTIMATE WITH
A RELIABILITY OF  90% AND A PRECISION OF PLUS AND MINUS $  50000
 IS    1527.



PROGRAM COMPLETE

NOW AT 2880
SRU'S:0.2
READY
```

III - Program Listing

```
1010.      REM***PROGRAM BBB***MARCH-12-74
1020.        PRINT "THIS PROGRAM IS USED TO DETERMINE SAMPLE SIZE FOR MEAN PER UNIT"
1030.        PRINT "ESTIMATES WHERE STRATIFICATION IS NOT USED."
1040.        PRINT
1050.        PRINT
1060.        PRINT "IS THERE A PRELIMINARY SAMPLE (ENTER YES OR NO)"
1070.        INPUT Q$
1080.        PRINT
1090.        IF Q$="NO" GOTO 1130
1100.        PRINT "WHAT IS THE SIZE OF THE PRELIMINARY SAMPLE"
1110.        INPUT N9
1120.        PRINT
1130.        PRINT "IS THIS A ONE-SIDED ESTIMATE OR A TWO-SIDED ESTIMATE"
1140.        PRINT "ENTER 1 OR 2"
1150.       INPUT T
1160.       PRINT
1170.       IF T=1 GOTO 1210
1180.       IF T=2 GOTO 1210
1190.        PRINT "ERROR. ENTER AGAIN"
1200.       GOTO 1150
1210.        PRINT "WHAT RELIABILITY PERCENTAGE IS DESIRED (EXAMPLE 90)"
1220.        INPUT R
1230.        PRINT
1240.        IF T=1 GOTO 1370
1250.        IF T=2 GOTO 1620
1260.      PRINT "WHAT PRECISION (AMOUNT) IS DESIRED"
1270.        INPUT A
1280.        PRINT
1290.        PRINT"ENTER TOTAL POPULATION SIZE"
1300.        INPUT N2
1310.        PRINT
1320.        IF Q$="YES" GOTO 2080
1330.        PRINT "ENTER THE ESTIMATED STATISTICAL VARIANCE OF THE POPULATION";
1340.        INPUT S2
1350.        GOTO 1870
1360.        REM TABLE FOR ONE-SIDED U VALUES
1370.        IF R=70 GOTO 1470
1380.        IF R=75 GOTO 1490
1390.        IF R=80 GOTO 1510
1400.        IF R=85 GOTO 1530
1410.        IF R=90 GOTO 1550
1420.        IF R=95 GOTO 1570
1430.        IF R =99 GOTO 1590
1440.        PRINT "VALUE OF RELIABILITY PERCENTAGE NOT ACCEPTABLE"
1450.        PRINT "ENTER NEW VALUE"
1460.        GOTO 1220
1470.        U=.52
1480.        GOTO 1260
1490.        U=.67
1500.        GOTO 1260
1510.        U=.84
1520.        GOTO 1260
1530.        U=1.04
1540.        GOTO 1260
```

```
1550.         U=1.28
1560.         GOTO 1260
1570.         U=1.64
1580.         GOTO 1260
1590.         U=2.33
1600.         GOTO 1260
1610.         REM TABLE FOR TWO-SIDED U VALUES
1620.         IF R=70 GOTO 1720
1630.         IF R=75 GOTO 1740
1640.         IF R=80 GOTO 1760
1650.         IF R=85 GOTO 1780
1660.         IF R=90 GOTO 1800
1670.         IF R=95 GOTO 1820
1680.         IF R=99 GOTO 1840
1690.         PRINT "VALUE OF RELIABILITY PERCENTAGE NOT ACCEPTABLE"
1700.         PRINT "ENTER NEW VALUE"
1710.         GOTO 1220
1720.         U=1.04
1730.         GOTO 1260
1740.         U=1.15
1750.         GOTO 1260
1760.         U=1.28
1770.         GOTO 1260
1780.         U=1.44
1790.         GOTO 1260
1800.         U=1.64
1810.         GOTO 1260
1820.         U=1.96
1830.         GOTO 1260
1840.         U=2.58
1850.         GOTO 1260
1860.         REM COMPUTATION OF SAMPLE SIZE
1870.        C1=N2↑2*U↑2*S2
1880.         C2=A↑2+(N2*U↑2*S2)
1890.         N1=C1/C2
1900.         PRINT
1910.         PRINT
1920.         IF T=1 GOTO 1940
1930.         IF T=2 GOTO 2000
1940.         PRINT "THE SAMPLE SIZE FOR A ONE-SIDED MEAN PER UNIT ESTIMATE WITH"
1950.         PRINT USING 1960,R,A,N1
1960.       :A RELIABILITY OF ###% AND A PRECISION OF PLUS $####### IS #######.
1970.         PRINT
1980.         PRINT
1990.       GOTO 2780
2000.         PRINT "THE SAMPLE SIZE FOR A TWO-SIDED MEAN PER UNIT ESTIMATE WITH"
2010.         PRINT USING 2020,R,A
2020.        :A RELIABILITY OF ###% AND A PRECISION OF PLUS AND MINUS $#######
2030.         PRINT USING 2040,N1
2040.       : IS #######.
2050.         PRINT
2060.         PRINT
2070.       GOTO 2780
2080.         REM***** COMPUTATION OF VARIANCE FROM PRELIMINARY SAMPLE
```

```
2090.       PRINT "HAVE PRELIMINARY SAMPLE ITEMS BEEN STORED IN DATA"
2100.       PRINT "STATEMENTS (YES OR NO)";
2110.       INPUT K$
2120.       IF K$="YES "THEN 2350
2130.       IF K$="NO" THEN 2150
2140.       GO TO 2090
2150.       PRINT
2160.       PRINT "ENTER VALUES OF PRELIMINARY SAMPLE ITEMS IN DATA STATE-"
2170.       PRINT "MENTS AS FOLLOWS.  LINE NUMBER RANGE USED MUST BE FROM"
2180.       PRINT "LINE 1 TO LINE 699.  SEPARATE ALL VALUES BY COMMAS; MAKE"
2190.       PRINT "SURE THAT THE LAST VALUE ON A LINE IS NOT FOLLOWED BY A "
2200.       PRINT "COMMA."
2210.       PRINT
2220.       PRINT "     LLL DATA V(1) , V(2) , V(3), . . . . V(N)"
2230.       PRINT
2240.       PRINT "          WHERE LLL  REPRESENTS A LINE NUMBER IN THE RANGE"
2250.       PRINT "          1 TO 999."
2260.       PRINT "          V(N) REPRESENTS A SAMPLE ITEM ENTERED IN THE"
2270.       PRINT"          DATA STATEMENT."
2280.       PRINT
2290.       PRINT "DURING RUNNING OF THIS PROGRAM YOU ARE ASKED THE SIZE"
2300.       PRINT "OF THE PRELIMINARY SAMPLE.  MAKE SURE THAT THE NUMBER"
2310.       PRINT "YOU ENTER AT THAT  TIME IS THE SAME AS THE NUMBER OF "
2320.       PRINT "SAMPLE ITEMS ENTERED IN THE DATA STATEMENT."
2330.       PRINT
2340.       GO TO 2780
2350.       LET J=X1=X2=0
2360.       FOR K8=1 TO N9
2370.       READ Z9
2380.       LET J=J+1
2390.       LET X1=X1+Z9
2400.       LET X2=X2+Z912
2410.       NEXT K8
2420.       REM*****END OF DATA STATEMENT READ
2430.       IF J=N9 GOTO 2470
2440.       PRINT
2450.       PRINT USING 2460,N9,J
2460.      :WARNING--SAMPLE SIZE GIVEN AS ####;#### SAMPLE ITEMS WERE ENTERED
2470.       C3=X1/N9
2480.       C4 = N9*C312
2490.        S2=(X2-C4)/(N9-1)
2500.        S1=SQR(S2)
2510.        PRINT
2520.        PRINT
2530.        PRINT "THE PRELIMINARY SAMPLE STATISTICS ARE:"
2540.        PRINT
2550.        PRINT USING 2560,C3
2560.       :          MEAN                     ######.##
2570.        PRINT USING 2580,S2
2580.      :          VARIANCE            ########.
2590.        PRINT USING 2600,S1
2600.       :          STANDARD DEVIATION    ######.##
2610.        GOTO 1870
2620.        REM N1=COMPUTED SAMPLE SIZE
```

```
2630.        REM N2=POPULATION SIZE
2640.        REM N9=PRELIMINARY SAMPLE SIZE
2650.        REM X1=SAMPLE ITEM
2660.        REM X2=SAMPLE ITEM SQUARED
2670.        REM X3=POPULATION VALUE
2680.        REM R=RELIABILITY %
2690.        REM U=NORMAL DEVIATE OF R
2700.        REM S1=STANDARD DEVIATION
2710.        REM S2=VARIANCE
2720.        REM S3=STANDARD ERROR OF THE MEAN
2730.        REM S4=STANDARD ERROR OF THE TOTAL
2740.        REM T=ONE-SIDED/TWO-SIDED OPTION
2750.        REM Q$=PRELIMINARY SAMPLE OPTION
2760.        REM C1-C9=COMPUTATIONAL WORK FIELDS
2770.        REM Z1-Z5=PRELIMINARY SAMPLE WORK FIELDS
2780.      IF Q$="NO" GOTO 2830
2790.        PRINT
2800.        PRINT "SUM OF X",X1
2810.        PRINT "SUM OF X SQR",X2
2820.        PRINT
2830.        PRINT
2840.        PRINT
2850.        PRINT "PROGRAM COMPLETE"
2860.        PRINT
2870.      END
```

MEAN-PER-UNIT ESTIMATION - PROGRAM FOR EVALUATION OF SAMPLE RESULTS

I - Program Description and Operating Insturctions

General Description

This program - designated "CCC" for purposes of this volume - is a computational program designed to statistically evaluate a sample with mean-per-unit estimation without stratification. The program makes requests for two options:

Option 1 - The ESTIMATE TYPE option - whether the estimate to be made from the sample is one-sided or two-sided - is answered 1 or 2. If some other number is entered an error message and request for re-entry will result. When a one-sided estimate is specified, a request is made to indicate upper or lower limit with a U or an L.

Option 2 - The SAMPLE DETAIL option - whether the details of the sample will be entered - is answered yes or no.

Inputs Required

The following inputs are required for all options:

1. Desired reliability percentage - 70, 75, 80, 85, 90, 95, and 99 (entered as whole numbers) are acceptable. Any other number will produce an error message and request for re-entry.
2. Total population size.
3. Size of the sample.

Where sample details are specified, such details must be entered as DATA statements, using lines 1 through 999. Follow instructions contained in the program listing. If an error is made in the data statement, merely correct the offending item and rerun the program.

Where no sample details are specified, the following sample data must be entered:

1. Sum of the sample item values.
2. Sum of the squares of the sample item values.

Outputs Produced

1. Sample statistics (when details are entered): mean, variance, and standard deviation.
2. Estimated population statistics: point estimate of the total, standard error of the mean, standard error of the total, precision, and confidence limits.
3. Supplementary information (where details are entered): sum of the sample items and sum of the squares of the sample items.

II - Examples of Operation

Example 1 - Sample details (audited values from Example Data Table) are entered with the following data:

Reliability - 90%

Population size - 10,000

Sample size - 30

A two-sided estimate is requested.

Example 2 - Sample totals are entered with the same data as used in Example 1.

Example 1

```
1 DATA 100 , 150 , 200 , 170 , 180 , 172 , 302 , 715 , 312 , 101
2 DATA 89 , 126 , 201 , 198 , 166 , 400 , 112 , 8 , 88 , 310
3 DATA 260 , 112 , 70 , 113 , 156 , 244 , 190 , 198 , 116 , 200
READY
RUN

CCC   19:52   09/19/74

THIS PROGRAM IS USED TO EVALUATE A SAMPLE USING MEAN PER UNIT
ESTIMATION WHERE STRATIFICATION IS NOT USED.


IS THIS A ONE-SIDED ESTIMATE OR A TWO-SIDED ESTIMATE
ENTER 1 OR 2
?2

WHAT RELIABILITY PERCENTAGE IS DESIRED (EXAMPLE 90)
?90


ENTER TOTAL POPULATION SIZE
?10000

ENTER THE SIZE OF YOUR SAMPLE
?30

HAVE SAMPLE ITEMS BEEN STORED IN DATA
STATEMENTS (YES OR NO)?YES


THE SAMPLE STATISTICS ARE:

              MEAN                                                191.97
              VARIANCE                                           16747.
              STANDARD DEVIATION                                  129.41


THE ESTIMATED POPULATION STATISTICS FOR A TWO-SIDED ESTIMATE
WITH A RELIABILITY OF  90% ARE AS FOLLOWS:

               POINT ESTIMATE OF TOTAL                         1919667.
               STANDARD ERROR OF THE MEAN                          23.59
               STANDARD ERROR OF THE TOTAL                      235918.
               PRECISION                                        386906.
               CONFIDENCE INTERVAL            1532761. TO  2306573.


SUM OF X         5759
SUM OF X SQR     1591213


PROGRAM COMPLETE

NOW AT 3310
SRU'S:0.6
READY
```

Example 2

```
CCC   20:00   09/19/74

THIS PROGRAM IS USED TO EVALUATE A SAMPLE USING MEAN PER UNIT
ESTIMATION WHERE STRATIFICATION IS NOT USED.


IS THIS A ONE-SIDED ESTIMATE OR A TWO-SIDED ESTIMATE
ENTER 1 OR 2
?2

WHAT RELIABILITY PERCENTAGE IS DESIRED (EXAMPLE 90)
?90


ENTER TOTAL POPULATION SIZE
?10000

ENTER THE SIZE OF YOUR SAMPLE
?30

HAVE SAMPLE ITEMS BEEN STORED IN DATA
STATEMENTS (YES OR NO)?NO

ARE SAMPLE DETAILS TO BE ENTERED (YES OR NO)
?NO

WHAT IS THE TOTAL OF THE SAMPLE VALUES
?5759

WHAT IS THE TOTAL OF THE SQUARES OF THE SAMPLE VALUES
?1591213


THE SAMPLE STATISTICS ARE:

              MEAN                                          191.97
              VARIANCE                                     16747.
              STANDARD DEVIATION                            129.41


THE ESTIMATED POPULATION STATISTICS FOR A TWO-SIDED ESTIMATE
WITH A RELIABILITY OF  90% ARE AS FOLLOWS:

               POINT ESTIMATE OF TOTAL                   1919667.
               STANDARD ERROR OF THE MEAN                    23.59
               STANDARD ERROR OF THE TOTAL                 235918.
               PRECISION                                   386906.
               CONFIDENCE INTERVAL          1532761. TO  2306573.


PROGRAM COMPLETE

NOW AT 3310
SRU'S:0.8
READY
```

III - Program Listing

```
1010.      REM***PROGRAM CCC***12-MARCH-74
1020.        PRINT "THIS PROGRAM IS USED TO EVALUATE A SAMPLE USING MEAN PER UNIT"
1030.        PRINT "ESTIMATION WHERE STRATIFICATION IS NOT USED."
1040.        PRINT
1050.        PRINT
1060.        PRINT "IS THIS A ONE-SIDED ESTIMATE OR A TWO-SIDED ESTIMATE"
1070.        PRINT "ENTER 1 OR 2"
1080.       INPUT T
1090.       PRINT
1100.       IF T=1 GOTO 1140
1110.       IF T=2 GOTO 1170
1120.       PRINT "ERROR.  PLEASE RE-ENTER"
1130.       GO TO 1080
1140.        PRINT "UPPER OR LOWER LIMIT (U OR L)"
1150.        INPUT L$
1160.        PRINT
1170.        PRINT "WHAT RELIABILITY PERCENTAGE IS DESIRED (EXAMPLE 90)"
1180.        INPUT R
1190.        PRINT
1200.        IF T=1 GOTO 1340
1210.        IF T=2 GOTO 1610
1220.        INPUT R
1230.        GOTO 1340
1240.        GOTO 1610
1250.       PRINT
1260.        PRINT"ENTER TOTAL POPULATION SIZE"
1270.        INPUT N2
1280.        PRINT
1290.        PRINT "ENTER THE SIZE OF YOUR SAMPLE"
1300.        INPUT N1
1310.        PRINT
1320.      GOTO 1900
1330.        REM TABLE FOR ONE-SIDED U VALUES
1340.        IF R=70 GOTO 1460
1350.        IF R=75 GOTO 1480
1360.        IF R=80 GOTO 1500
1370.        IF R=85 GOTO 1520
1380.        IF R=90 GOTO 1540
1390.        IF R=95 GOTO 1560
1400.        IF R =99 GOTO 1580
1410.        PRINT "VALUE OF RELIABILITY PERCENTAGE NOT ACCEPTABLE"
1420.        PRINT "ENTER NEW VALUE"
1430.        GOTO 1180
1440.        PRINT
1450.        GOTO 1220
1460.        U=.52
1470.        GOTO 1250
1480.        U=.67
1490.        GOTO 1250
1500.        U=.84
1510.        GOTO 1250
1520.        U=1.04
1530.        GOTO 1250
1540.        U=1.28
```

```
1550.          GOTO 1250
1560.          U=1.64
1570.          GOTO 1250
1580.          U=2.33
1590.          GOTO 1250
1600.          REM TABLE FOR TWO-SIDED U VALUES
1610.          IF R=70 GOTO 1720
1620.          IF R=75 GOTO 1740
1630.          IF R=80 GOTO 1760
1640.          IF R=85 GOTO 1780
1650.          IF R=90 GOTO 1800
1660.          IF R=95 GOTO 1820
1670.          IF R=99 GOTO 1840
1680.          PRINT "VALUE OF RELIABILITY PERCENTAGE NOT ACCEPTABLE"
1690.          PRINT "ENTER NEW VALUE"
1700.          GOTO 1180
1710.          PRINT
1720.          U=1.04
1730.          GOTO 1250
1740.          U=1.15
1750.          GOTO 1250
1760.          U=1.28
1770.          GOTO 1250
1780.          U=1.44
1790.          GOTO 1250
1800.          U=1.64
1810.          GOTO 1250
1820.          U=1.96
1830.          GOTO 1250
1840.          U=2.58
1850.          GOTO 1250
1860.        PRINT "ARE SAMPLE DETAILS TO BE USED (YES OR NO)"
1870.        INPUT W$
1880.        PRINT
1890.        IF W$="NO" GOTO 2990
1900.        REM COMPUTATION OF VARIANCE FROM SAMPLE ITEMS
1910.        PRINT "HAVE SAMPLE ITEMS BEEN STORED IN DATA"
1920.         PRINT "STATEMENTS (YES OR NO)";
1930.         INPUT K$
1940.         IF K$="YES "THEN 2220
1950.         IF K$="NO" THEN 1970
1960.         GO TO 1910
1970.         PRINT
1980.          REM ACCUMULATION OF SAMPLE VALUES/COMPUTATIONS
1990.          PRINT "ARE SAMPLE DETAILS TO BE ENTERED (YES OR NO)"
2000.          INPUT W$
2010.          PRINT
2020.          IF W$="NO" GOTO 2990
2030.        PRINT "ENTER VALUES OF SAMPLE ITEMS IN DATA STATE-"
2040.         PRINT "MENTS AS FOLLOWS.  LINE NUMBER RANGE USED MUST BE FROM"
2050.        PRINT "LINE 1 TO LINE 999.  SEPARATE ALL VALUES BY COMMAS;MAKE"
2060.         PRINT "SURE THAT THE LAST VALUE ON A LINE IS NOT FOLLOWED BY A "
2070.         PRINT "COMMA."
2080.         PRINT
```

```
2090.     PRINT "     LLL DATA V(1) , V(2) , V(3), . . . . V(N)"
2100.     PRINT
2110.     PRINT "          WHERE LLL  REPRESENTS A LINE NUMBER IN THE RANGE"
2120.    PRINT "          1 TO 999."
2130.    PRINT "          V(N) REPRESENTS A SAMPLE ITEM ENTERED IN THE"
2140.    PRINT "          DATA STATEMENT."
2150.    PRINT
2160.     PRINT "DURING RUNNING OF THIS PROGRAM YOU ARE ASKED THE SIZE"
2170.     PRINT "OF THE PRELIMINARY SAMPLE.  MAKE SURE THAT THE NUMBER"
2180.     PRINT "YOU ENTER AT THAT  TIME IS THE SAME AS THE NUMBER OF "
2190.     PRINT "SAMPLE ITEMS ENTERED IN THE DATA STATEMENT."
2200.     PRINT
2210.    GOTO 3310
2220.     LET J=X1=X2=0
2230.    FOR K8=1 TO N1
2240.     READ Z9
2250.     LET J=J+1
2260.     LET X1=X1+Z9
2270.     LET X2=X2+Z9↑2
2280.     NEXT K8
2290.     REM*****END OF DATA STATEMENT READ
2300.     IF J=N1 GOTO 2340
2310.     PRINT
2320.     PRINT USING 2330,N1,J
2330.    :WARNING--SAMPLE SIZE GIVEN AS ####;#### SAMPLE ITEMS WERE ENTERED
2340.     C3=X1/N1
2350.     C4=N1*C3↑2
2360.      S2=(X2-C4)/(N1-1)
2370.      S1=SQR(S2)
2380.      PRINT
2390.      PRINT
2400.      PRINT "THE SAMPLE STATISTICS ARE:"
2410.      PRINT
2420.      PRINT USING 2430,C3
2430.     :              MEAN                                         ######.##
2440.      PRINT USING 2450,S2
2450.    :               VARIANCE                                    ########.
2460.      PRINT USING 2470,S1
2470.     :              STANDARD DEVIATION                           ######.##
2480.      PRINT
2490.      PRINT
2500.      REM COMPUTATION OF POPULATION ESTIMATES
2510.      C1=SQR(N1)
2520.      C2=S1/C1
2530.      C6=(N2-N1)/N2
2540.      C5=SQR(C6)
2550.      S3=C2*C5
2560.      S4=S3*N2
2570.      A=U*S4
2580.      X3=C3*N2
2590.      X4=X3-A
2600.      X5=X3+A
2610.      IF T=1 GOTO 2630
2620.      IF T=2 GOTO 2820
```

```
2630.       PRINT "THE ESTIMATED POPULATION STATISTICS FOR A ONE-SIDED ESTIMATE"
2640.       PRINT USING 2650,R
2650.     :WITH A RELIABILITY OF ###% ARE AS FOLLOWS:
2660.       PRINT
2670.       PRINT USING 2680,X3
2680.     :                 POINT ESTIMATE OF TOTAL                        ########.
2690.       PRINT USING 2700,S3
2700.     :                 STANDARD ERROR OF THE MEAN                       ######.##
2710.       PRINT USING 2720,S4
2720.     :                 STANDARD ERROR OF THE TOTAL                    ########.
2730.       PRINT USING 2740,A
2740.     :                 PRECISION                                      ########.
2750.       IF L$="L" GOTO 2790
2760.       PRINT USING 2770,X5
2770.     :                 UPPER CONFIDENCE LIMIT                         ########.
2780.       GOTO 2960
2790.       PRINT USING 2800,X4
2800.     :                 LOWER CONFIDENCE LIMIT                         ########.
2810.       GOTO 2960
2820.       PRINT "THE ESTIMATED POPULATION STATISTICS FOR A TWO-SIDED ESTIMATE"
2830.       PRINT USING 2840,R
2840.     :WITH A RELIABILITY OF ###% ARE AS FOLLOWS:
2850.       PRINT
2860.       PRINT USING 2870,X3
2870.     :                 POINT ESTIMATE OF TOTAL                        ########.
2880.       PRINT USING 2890,S3
2890.     :                 STANDARD ERROR OF THE MEAN                       ######.##
2900.       PRINT USING 2910,S4
2910.     :                 STANDARD ERROR OF THE TOTAL                    ########.
2920.       PRINT USING 2930,A
2930.     :                 PRECISION                                      ########.
2940.       PRINT USING 2950,X4,X5
2950.     :                 CONFIDENCE INTERVAL          ########. TO ########.
2960.       PRINT
2970.       PRINT
2980.     GOTO 3230
2990.       REM INPUT OF SAMPLE TOTALS WHERE DETAILS NOT SPECIFIED
3000.       PRINT "WHAT IS THE TOTAL OF THE SAMPLE VALUES"
3010.       INPUT X1
3020.       PRINT
3030.       PRINT "WHAT IS THE TOTAL OF THE SQUARES OF THE SAMPLE VALUES"
3040.       INPUT X2
3050.      GOTO 2340
3060.       REM N1=COMPUTED SAMPLE SIZE
3070.       REM N2=POPULATION SIZE
3080.       REM X1=SAMPLE ITEM
3090.       REM X2=SAMPLE ITEM SQUARED
3100.       REM X3=POPULATION VALUE
3110.       REM R=RELIABILITY %
3120.       REM U=NORMAL DEVIATE OF R
3130.       REM S1=STANDARD DEVIATION
3140.       REM S2=VARIANCE
3150.       REM S3=STANDARD ERROR OF THE MEAN
3160.       REM S4=STANDARD ERROR OF THE TOTAL
3170.       REM T=ONE-SIDED/TWO-SIDED OPTION
3180.       REM W$=SAMPLE DETAILS/TOTALS OPTION
3190.       REM C1-C9=COMPUTATIONAL WORK FIELDS
3200.       REM Z1-Z9=SAMPLE WORK FIELDS
3210.       REM X4=LOWER CONFIDENCE LIMIT
3220.       REM X5=UPPER CONFIDENCE LIMIT
3230.      IF W$="NO" GOTO 3270
3240.       PRINT "SUM OF X",X1
3250.     PRINT "SUM OF X SQR",X2
3260.       PRINT
3270.       PRINT
3280.       PRINT "PROGRAM COMPLETE"
3290.       PRINT
3300.       '
3310.     END
```

RATIO AND/OR DIFFERENCE ESTIMATION - PROGRAM FOR DETERMINATION OF SAMPLE SIZE

I - Program Description and Operating Instructions

General Description

This program - designated "DDD" for purposes of this volume - is a computational program designed to determine sample size where a statistical evaluation is to be made using ratio estimation and/or difference estimation without stratification. The program makes requests for two options:

Option 1 - The ESTIMATE TYPE option - whether the estimate to be made from the sample is one-sided or two-sided - is answered 1 or 2. If some other number is entered, an error message and request for re-entry will result.

Option 2 - The PRELIMINARY SAMPLE option - whether the details of a preliminary sample will be entered - is answered yes or no.

Inputs Required

The following inputs are required for all options:

1. Desired reliability percentage - 70, 75, 80, 85, 90, 95, and 99 (entered as whole numbers) are acceptable. Any other number will produce an error message and request for re-entry.
2. Desired precision - entered as an amount. Any number of decimal places is acceptable, but precision is usually in whole dollars.
3. Population size.

Where a preliminary sample is specified, additional inputs are:

1. Size of preliminary sample.
2. Sample details - these are entered in pairs of items as DATA statements in lines 1 through 999. For each pair of items, the sample audited amount is followed by the sample book amount. Refer to the program listing for further details. If an error is made in the data statement, merely correct the offending item and rerun the program.

Where no preliminary sample is specified, an estimate of the population variance $\left(S_{R_j}^{\ 2} \text{ or } S_{D_j}^{\ 2}\right)$ must be entered.

Outputs Produced

1. Computed sample size - given as a whole number. The program will indicate, where a preliminary sample is used, whether the computed sample size was determined on the basis of ratios or differences. (The smallest of the two will be printed.) For the ratio information to be valid, all sample items must be entered, whereas difference information will be valid if only sample items where differences exist are entered.

2. When preliminary sample details are entered, two additional outputs are produced.

   a. Preliminary sample statistics: variance (on both bases) and standard deviation.

   b. Supplementary information: sum of the sample item audited values, sum of the sample item book values, sum of the sample item differences, sum of the squares of each of the above, and sum of the cross-products of the sample item audited and book values.

II - Examples of Operation

Example 1 - Sample details (audited values and book values from Example Data Table) are entered with the following data:

Reliability - 90%

Precision - ± 50,000

Population size - 10,000

A two-sided estimate is desired.

Example 2 - Sample totals are entered with the same data as used in Example 1.

Example 1

```
1 DATA 100 , 110 , 150 , 150 , 200 , 198 , 170 , 170 , 180 , 180
2 DATA 172 , 200 , 302 , 300 , 715 , 800 , 312 , 312 , 101 , 100
3 DATA 89 , 89 , 126 , 125 , 201 , 300 , 198 , 198 , 166 , 160
4 DATA 400 , 400 , 112 , 112 , 8 , 8 , 88 , 88 , 310 , 310
5 DATA 260 , 260 , 112 , 100 , 70 , 90 , 113 , 113 , 156 , 156
6 DATA 244 , 240 , 190 , 190 , 198 , 200 , 116 , 116 , 200 , 200
READY
RUN

DDD   20:08   09/19/74

THIS PROGRAM IS USED TO DETERMINE SAMPLE SIZE FOR THE
ESTIMATION OF POPULATION DIFFERENCES WHERE STRATIFICATION
IS NOT USED. IT USES EITHER THE DIFFERENCE OR THE RATIO
OF THE SAMPLE ITEM VALUES AND ELECTS THE ALTERNATIVE
WHICH PRODUCES THE SMALLEST SAMPLE SIZE.


IS THERE A PRELIMINARY SAMPLE (ENTER YES OR NO)
?YES

WHAT IS THE SIZE OF THE PRELIMINARY SAMPLE
?30

IS THIS A ONE-SIDED ESTIMATE OR A TWO-SIDED ESTIMATE
ENTER 1 OR 2
?2

WHAT RELIABILITY PERCENTAGE IS DESIRED (EXAMPLE 90)
?90

WHAT PRECISION (AMOUNT) IS DESIRED
?50000

ENTER TOTAL POPULATION SIZE
?10000

HAVE PRELIMINARY SAMPLE ITEMS BEEN STORED IN DATA
STATEMENTS (YES OR NO)?YES
```

Example 1, contd.

```
THE PRELIMINARY SAMPLE STATISTICS ARE:

          VARIANCE                   464.
          STANDARD DEVIATION          21.54


NOTE THAT THE ABOVE WERE COMPUTED USING SAMPLE ITEM RATIOS
THE VARIANCE ON THE BASIS OF SAMPLE ITEM DIFFERENCES IS      585.


THE SAMPLE SIZE FOR A TWO-SIDED ESTIMATE WITH
A RELIABILITY OF  90% AND A PRECISION OF PLUS AND MINUS $  50000
IS     50.



SUM OF X          5759
SUM OF Y          5975
SUM OF D         -216
SUM OF X SQR      1591213
SUM OF Y SQR      1777211
SUM OF D SQR      18520
SUM OF XY         1674952


PROGRAM COMPLETE

NOW AT 3230
SRU'S:0.6
READY
```

Example 2

```
RUN

DDD   20:11   09/19/74

THIS PROGRAM IS USED TO DETERMINE SAMPLE SIZE FOR THE
ESTIMATION OF POPULATION DIFFERENCES WHERE STRATIFICATION
IS NOT USED. IT USES EITHER THE DIFFERENCE OR THE RATIO
OF THE SAMPLE ITEM VALUES AND ELECTS THE ALTERNATIVE
WHICH PRODUCES THE SMALLEST SAMPLE SIZE.


IS THERE A PRELIMINARY SAMPLE (ENTER YES OR NO)
?NO

IS THIS A ONE-SIDED ESTIMATE OR A TWO-SIDED ESTIMATE
ENTER 1 OR 2
?2

WHAT RELIABILITY PERCENTAGE IS DESIRED (EXAMPLE 90)
?90

WHAT PRECISION (AMOUNT) IS DESIRED
?50000

ENTER TOTAL POPULATION SIZE
?10000

ENTER THE ESTIMATED STATISTICAL VARIANCE OF THE POPULATION
?464


THE SAMPLE SIZE FOR A TWO-SIDED ESTIMATE WITH
A RELIABILITY OF  90% AND A PRECISION OF PLUS AND MINUS $  50000
IS      50.



PROGRAM COMPLETE

NOW AT 3230
SRU'S:0.2
READY
```

```
1010.      REM***PROGRAM DDD***12-MARCH-74
1020.        PRINT "THIS PROGRAM IS USED TO DETERMINE SAMPLE SIZE FOR THE"
1030.        PRINT "ESTIMATION OF POPULATION DIFFERENCES WHERE STRATIFICATION"
1040.        PRINT "IS NOT USED. IT USES EITHER THE DIFFERENCE OR THE RATIO"
1050.        PRINT "OF THE SAMPLE ITEM VALUES AND ELECTS THE ALTERNATIVE"
1060.        PRINT "WHICH PRODUCES THE SMALLEST SAMPLE SIZE."
1070.        PRINT
1080.        PRINT
1090.        PRINT "IS THERE A PRELIMINARY SAMPLE (ENTER YES OR NO)"
1100.        INPUT Q$
1110.        PRINT
1120.        IF Q$="NO" GOTO 1160
1130.        PRINT "WHAT IS THE SIZE OF THE PRELIMINARY SAMPLE"
1140.        INPUT N9
1150.        PRINT
1160.        PRINT "IS THIS A ONE-SIDED ESTIMATE OR A TWO-SIDED ESTIMATE"
1170.        PRINT "ENTER 1 OR 2"
1180.       INPUT T
1190.       PRINT
1200.       IF T=1 GOTO 1240
1210.       IF T=2 GOTO 1240
1220.        PRINT "ERROR. ENTER AGAIN"
1230.       GOTO 1180
1240.        PRINT "WHAT RELIABILITY PERCENTAGE IS DESIRED (EXAMPLE 90)"
1250.        INPUT R
1260.        PRINT
1270.        IF T=1 GOTO 1410
1280.        IF T=2 GOTO 1660
1290.        PRINT "WHAT PRECISION (AMOUNT) IS DESIRED"
1310.        INPUT A
1320.        PRINT
1330.        PRINT"ENTER TOTAL POPULATION SIZE"
1340.        INPUT N2
1350.        PRINT
1360.      IF Q$="YES" GOTO 2120
1370.        PRINT "ENTER THE ESTIMATED STATISTICAL VARIANCE OF THE POPULATION"
1380.        INPUT S2
1390.        GOTO 1900
1400.        REM TABLE FOR ONE-SIDED U VALUES
1410.        IF R=70 GOTO 1510
1420.        IF R=75 GOTO 1530
1430.        IF R=80 GOTO 1550
1440.        IF R=85 GOTO 1570
1450.        IF R=90 GOTO 1590
1460.        IF R=95 GOTO 1610
1470.        IF R =99 GOTO 1630
1480.        PRINT "VALUE OF RELIABILITY PERCENTAGE NOT ACCEPTABLE"
1490.        PRINT "ENTER NEW VALUE"
1500.        GOTO 1250
1510.        U=.52
1520.        GOTO 1290
1530.        U=.67
1540.        GOTO 1290
1550.        U=.84
```

```
1560.        GOTO 1290
1570.        U=1.04
1580.        GOTO 1290
1590.        U=1.28
1600.        GOTO 1290
1610.        U=1.64
1620.        GOTO 1290
1630.        U=2.33
1640.        GOTO 1290
1650.        REM TABLE FOR TWO-SIDED U VALUES
1660.        IF R=70 GOTO 1760
1670.        IF R=75 GOTO 1780
1680.        IF R=80 GOTO 1800
1690.        IF R=85 GOTO 1820
1700.        IF R=90 GOTO 1840
1710.        IF R=95 GOTO 1860
1720.        IF R=99 GOTO 1880
1730.        PRINT "VALUE OF RELIABILITY PERCENTAGE NOT ACCEPTABLE"
1740.        PRINT "ENTER NEW VALUE"
1750.        GOTO 1250
1760.        U=1.04
1770.        GOTO 1290
1780.        U=1.15
1790.        GOTO 1290
1800.        U=1.28
1810.        GOTO 1290
1820.        U=1.44
1830.        GOTO 1290
1840.        U=1.64
1850.        GOTO 1290
1860.        U=1.96
1870.        GOTO 1290
1880.        U=2.58
1890.        GOTO 1290
1900.        REM COMPUTATION OF SAMPLE SIZE
1910.        C1=N2|2*U|2*S2
1920.        C2=A|2+(N2*U|2*S2)
1930.        N1=C1/C2
1940.        PRINT
1950.        PRINT
1970.        IF T=2 GOTO 2040
1980.        PRINT "THE SAMPLE SIZE FOR A ONE-SIDED ESTIMATE WITH"
1990.        PRINT USING 2000,R,A,N1
2000.      :A RELIABILITY OF ###% AND A PRECISION OF PLUS $####### IS #######.
2010.        PRINT
2020.        PRINT
2030.       GOTO 3080
2040.        PRINT "THE SAMPLE SIZE FOR A TWO-SIDED ESTIMATE WITH"
2050.        PRINT USING 2060,R,A
2060.       :A RELIABILITY OF ###% AND A PRECISION OF PLUS AND MINUS $#######
2070.        PRINT USING 2080,N1
2080.      :IS #######.
2090.        PRINT
2100.        PRINT
```

```
2110.     GOTO 3080
2120.     REM***** COMPUTATION OF VARIANCE FROM PRELIMINARY SAMPLE
2130.    PRINT "HAVE PRELIMINARY SAMPLE ITEMS BEEN STORED IN DATA"
2140.    PRINT "STATEMENTS (YES OR NO)";
2150.    INPUT K$
2160.    IF K$="YES "THEN 2390
2170.    IF K$="NO" THEN 2190
2180.    GO TO 2130
2190.    PRINT
2200.    PRINT "ENTER VALUES OF PRELIMINARY SAMPLE ITEMS IN DATA STATE-"
2210.   PRINT "MENTS - AUDITED VALUE,BOOK VALUE - AS FOLLOWS. LINE NUMBER RANGE MUST"
2220.   PRINT "BE FROM 1 TO 999. SEPARATE ALL VALUES BY COMMAS; MAKE"
2230.    PRINT "SURE THAT THE LAST VALUE ON A LINE IS NOT FOLLOWED BY A "
2240.    PRINT "COMMA."
2250.    PRINT
2260.    PRINT "     LLL DATA V(1) , V(2) , V(3), . . . . V(N)"
2270.    PRINT
2280.    PRINT "           WHERE LLL  REPRESENTS A LINE NUMBER IN THE RANGE"
2290.   PRINT "          1 TO 999."
2300.    PRINT "           V(N) REPRESENTS A SAMPLE ITEM ENTERED IN THE"
2310.    PRINT"           DATA STATEMENT."
2320.    PRINT
2330.    PRINT "DURING RUNNING OF THIS PROGRAM YOU ARE ASKED THE SIZE"
2340.    PRINT "OF THE PRELIMINARY SAMPLE.  MAKE SURE THAT THE NUMBER"
2350.    PRINT "YOU ENTER AT THAT  TIME IS THE SAME AS THE NUMBER OF "
2360.    PRINT "SAMPLE ITEMS ENTERED IN THE DATA STATEMENT."
2370.    PRINT
2380.   GOTO 3230
2390.    LET J=X1=X2=Y1=Y2=D1=D2=Z7=0
2400.    FOR K8=1 TO N9
2410.   READ Z9,Z8
2420.    LET J=J+1
2430.    LET X1=X1+Z9
2440.    LET X2=X2+Z9|2
2450.   LET Y1=Y1+Z8
2460.   LET Y2=Y2+Z8|2
2470.   LET D1=D1+(Z9-Z8)
2480.   LET D2=D2+(Z9-Z8)|2
2490.   LET Z7=Z7+(Z9*Z8)
2500.    NEXT K8
2510.    REM*****END OF DATA STATEMENT READ
2520.    IF J=N9 GOTO 2570
2530.    PRINT
2540.    PRINT USING 2550,N9,J
2550.   :WARNING--SAMPLE SIZE GIVEN AS ####; ####SAMPLE ITEMS ENTERED
2560.    PRINT "DIFFERENCE BASIS DATA MAY BE VALID; RATIO BASIS DATA WILL NOT."
2570.     R1=X1/Y1
2580.     C3=X2-(2*R1*Z7)+(R1|2*Y2)
2590.     C4-C3/(N9-1)
2600.     C5=D1/N9
2610.     C6=N9*C5|2
2620.     C7=(D2-C6)/(N9-1)
2630.     IF C4<=C7 GOTO 2660
2640.     S2=C7
2650.     GOTO 2670
2660.     S2=C4
```

```
2670.        S1=SQR(S2)
2680.        PRINT
2690.        PRINT
2700.        PRINT "THE PRELIMINARY SAMPLE STATISTICS ARE:"
2710.        PRINT
2720.        PRINT USING 2730,S2
2730.     :             VARIANCE                  ########.
2740.        PRINT USING 2750,S1
2750.      :             STANDARD DEVIATION     ######.##
2760.        PRINT
2770.        PRINT
2780.        IF C4<C7 GOTO 2830
2790.        PRINT "NOTE THAT THE ABOVE WERE COMPUTED USING SAMPLE ITEM DIFFERENCES"
2800.        PRINT USING 2810,C4
2810.     :THE VARIANCE ON THE BASIS OF SAMPLE ITEM RATIOS IS ########.
2820.        GOTO 1910
2830.        PRINT "NOTE THAT THE ABOVE WERE COMPUTED USING SAMPLE ITEM RATIOS"
2840.        PRINT USING 2850,C7
2850.     :THE VARIANCE ON THE BASIS OF SAMPLE ITEM DIFFERENCES IS ########.
2860.        GOTO 1910
2870.        REM N1=COMPUTED SAMPLE SIZE
2880.        REM N2=POPULATION SIZE
2890.        REM N9=PRELIMINARY SAMPLE SIZE
2900.        REM X1=SAMPLE ITEM AUDITED VALUE
2910.        REM Y1=SAMPLE ITEM BOOK VALUE
2920.        REM X2=SAMPLE ITEM AUDITED VALUE SQUARED
2930.        REM Y2=SAMPLE ITEM BOOK VALUE SQUARED
2940.        REM R1=RATIO OF SUMS OF SAMPLE ITEM AUDITED AND BOOK VALUES
2950.        REM X3=POPULATION VALUE
2960.        REM R=RELIABILITY %
2970.        REM U=NORMAL DEVIATE OF R
2980.        REM S1=STANDARD DEVIATION
2990.        REM S2=VARIANCE
3000.        REM S3=STANDARD ERROR OF THE MEAN
3010.        REM S4=STANDARD ERROR OF THE TOTAL
3020.        REM T=ONE-SIDED/TWO-SIDED OPTION
3030.        REM Q$=PRELIMINARY SAMPLE OPTION
3040.        REM C1-C9=COMPUTATIONAL WORK FIELDS
3050.        REM Z1-Z5=PRELIMINARY SAMPLE WORK FIELDS
3060.        REM D1=SAMPLE ITEM DIFFERENCE
3070.        REM D2=SAMPLE ITEM DIFFERENCE SQUARED
3080.       PRINT
3090.       IF Q$="NO" GOTO 3180
3100.        PRINT "SUM OF X",X1
3110.        PRINT "SUM OF Y",Y1
3120.        PRINT "SUM OF D",D1
3130.        PRINT "SUM OF X SQR",X2
3140.        PRINT "SUM OF Y SQR",Y2
3150.        PRINT "SUM OF D SQR",D2
3160.        PRINT "SUM OF XY",Z7
3170.        PRINT
3180.        PRINT
3190.        PRINT
3200.        PRINT "PROGRAM COMPLETE"
3210.        PRINT
3220.        '
3230.     END
```

RATIO AND/OR DIFFERENCE ESTIMATION - PROGRAM FOR EVALUATION OF SAMPLE RESULTS

I - Program Description and Operating Instructions

General Description

This program - designated "EEE" for purposes of this volume - is a computational program designed to statistically evaluate a sample using ratio estimation and/or difference estimation without stratification. The program makes requests for two options:

Option 1 - The ESTIMATE TYPE option - whether the estimate to be made from the sample is one-sided or two-sided - is answered 1 or 2. If some other number is entered, an error message and request for re-entry will result. When a one-sided estimate is specified, a request is made to indicate upper or lower limit with a U or an L.

Option 2 - The SAMPLE DETAIL option - whether the details of the sample will be entered - is answered yes or no.

Inputs Required

The following inputs are required for all options:

1. Desired reliability percentage - 70, 75, 80, 85, 90, 95, and 99 (entered as whole numbers) are acceptable. Any other number will produce an error message and request for re-entry.
2. Population size.
3. Population total book value.
4. Size of the sample.

Where sample details are specified, they are entered in pairs as DATA statements. Within each pair, the audited value is followed by the book value, using lines 1 through 999. Refer to the program listing for further details. If an error is made in the data statement, merely correct the offending item and rerun the program.

Where no sample details are specified, the following sample data must be entered:

1. Sum of the sample item audited values.
2. Sum of the squares of the audited values.
3. Sum of the sample item book values.
4. Sum of the squares of the book values.
5. Sum of the sample item differences.
6. Sum of the squares of the differences.
7. Sum of the cross-products of the sample item audited and book values.

If the user has this data as it pertains to only ratio estimation or only difference estimation, he should enter a 1 in response to the request for other values. In so doing, the results for the method not used will be clearly artificial.

Outputs Produced

Outputs are produced on the basis of sample item ratios and differences, as follows:

1. Sample statistics (where details are entered): mean, variance, and standard deviation.
2. Estimated population statistics: point estimate of the population difference, point estimate of the population total, standard error of the mean, standard error of the total, precision, confidence limits for the population difference, and confidence limits for the population total.
3. Supplementary information (when details are entered): sum of the sample item audited values, sum of the sample item book values, sum of the differences, sum of the squares of each of the above, and sum of the cross-products of the sample item audited and book values.

II - Examples of Operation

Example 1 - Sample details (audited values and book values from Example Data Table) are entered with the following data:

Reliability - 90%

Population size - 10,000

Population total book value - 2,000,000

Sample size - 30

A two-sided estimate is requested.

Example 2 - Sample totals are entered with the same data as used in Example 1.

Example 1

```
1 DATA 100 , 110 , 150 , 150 , 200 , 198 , 170 , 170 , 180 , 180
2 DATA 172 , 200 , 302 , 300 , 715 , 800 , 312 , 312 , 101 , 100
3 DATA 89 , 89 , 126 , 125 , 201 , 300 , 198 , 198 , 166 , 160
4 DATA 400 , 400 , 112 , 112 , 8 , 8 , 88 , 88 , 310 , 310
5 DATA 260 , 260 , 112 , 100 , 70 , 90 , 113 , 113 , 156 , 156
6 DATA 244 , 240 , 190 , 190 , 198 , 200 , 116 , 116 , 200 , 200
READY
RUN

EEE   20:29   09/19/74

THIS PROGRAM IS USED TO EVALUATE A SAMPLE DESIGNED TO ESTIMATE
THE POPULATION DIFFERENCE WHERE STRATIFICATION IS NOT USED.
RESULTS ARE GIVEN ON THE BASIS OF BOTH THE DIFFERENCES AND
THE RATIOS OF THE SAMPLE ITEMS. HOWEVER, IF ONLY THOSE SAMPLE
ITEMS WHERE DIFFERENCES OCCURRED ARE ENTERED, ONLY THE
RESULTS ON THE BASIS OF DIFFERENCES CAN BE USED.


ARE SAMPLE DETAILS TO BE ENTERED (YES OR NO)
?YES

IS THIS A ONE-SIDED ESTIMATE OR A TWO-SIDED ESTIMATE
ENTER 1 OR 2
?2

WHAT RELIABILITY PERCENTAGE IS DESIRED (EXAMPLE 90)
?90

ENTER TOTAL POPULATION SIZE
?10000

ENTER THE POPULATION TOTAL BOOK VALUE
?2000000

ENTER THE SIZE OF YOUR SAMPLE
?30

HAVE PRELIMINARY SAMPLE ITEMS BEEN STORED IN DATA
STATEMENTS (YES OR NO)?YES


THE SAMPLE STATISTICS ARE:

                                    ON THE BASIS OF SAMPLE
                                   DIFFERENCES       RATIOS
         MEAN                              -7.2          1.0
         VARIANCE                         585.0        463.8
         STANDARD DEVIATION               24.19        21.54
```

Example 1, contd.

```
THE ESTIMATED POPULATION STATISTICS FOR A TWO-SIDED ESTIMATE
WITH A RELIABILITY OF   90% ARE AS FOLLOWS:

                                          ON THE BASIS OF SAMPLE
                                         DIFFERENCES       RATIOS
          POINT ESTIMATE OF THE
            POPULATION DIFFERENCE          -72000.        -72301.
          POINT ESTIMATE OF THE
            POPULATION TOTAL              1928000.       1927699.
          STANDARD ERROR OF THE MEAN           4.4            3.9
          STANDARD ERROR OF THE TOTAL       44092.         39259.
          PRECISION                         72311.         64384.
          CONFIDENCE INTERVAL FOR
            POPULATION DIFFERENCE        -144311. TO    -136685. TO
                                              311.          -7917.
          CONFIDENCE INTERVAL FOR
            POPULATION TOTAL              1855689.TO     1863315.TO
                                           2000311.       1992083.

NOTE: NEGATIVE DIFFERENCES REPRESENT AN OVERSTATEMENT IN
      THE BOOK AMOUNT


SUM OF X         5759
SUM OF Y         5975
SUM OF D        -216
SUM OF X SQR     1591213
SUM OF Y SQR     1777211
SUM OF D SQR     18520
SUM OF XY        1674952


PROGRAM COMPLETE

NOW AT 4160
SRU'S:1.0
READY
```

Example 2

```
RUN

EEE   20:34   09/19/74

THIS PROGRAM IS USED TO EVALUATE A SAMPLE DESIGNED TO ESTIMATE
THE POPULATION DIFFERENCE WHERE STRATIFICATION IS NOT USED.
RESULTS ARE GIVEN ON THE BASIS OF BOTH THE DIFFERENCES AND
THE RATIOS OF THE SAMPLE ITEMS. HOWEVER, IF ONLY THOSE SAMPLE
ITEMS WHERE DIFFERENCES OCCURRED ARE ENTERED, ONLY THE
RESULTS ON THE BASIS OF DIFFERENCES CAN BE USED.


ARE SAMPLE DETAILS TO BE ENTERED (YES OR NO)
?NO

IS THIS A ONE-SIDED ESTIMATE OR A TWO-SIDED ESTIMATE
ENTER 1 OR 2
?2

WHAT RELIABILITY PERCENTAGE IS DESIRED (EXAMPLE 90)
?90

ENTER TOTAL POPULATION SIZE
?10000

ENTER THE POPULATION TOTAL BOOK VALUE
?2000000

ENTER THE SIZE OF YOUR SAMPLE
?30

WHAT IS THE TOTAL OF THE SAMPLE AUDITED VALUES
?5759

WHAT IS THE TOTAL OF THE SQUARES OF THE SAMPLE AUDITED VALUES
?1591213

WHAT IS THE TOTAL OF THE SAMPLE BOOK VALUES
?5975

WHAT IS THE TOTAL OF THE SQUARES OF THE SAMPLE BOOK VALUES
?1777211

WHAT IS THE TOTAL OF THE SAMPLE DIFFERENCES
?-216

WHAT IS THE TOTAL OF THE SQUARES OF THE SAMPLE DIFFERENCES
?18520

WHAT IS THE TOTAL OF THE CROSS-PRODUCTS OF THE SAMPLE
AUDITED AND BOOK VALUES
?1674952
```

Example 2, contd.

THE SAMPLE STATISTICS ARE:

	ON THE BASIS OF SAMPLE DIFFERENCES	RATIOS
MEAN	-7.2	1.0
VARIANCE	585.0	463.8
STANDARD DEVIATION	24.19	21.54

THE ESTIMATED POPULATION STATISTICS FOR A TWO-SIDED ESTIMATE WITH A RELIABILITY OF 90% ARE AS FOLLOWS:

	ON THE BASIS OF SAMPLE DIFFERENCES	RATIOS
POINT ESTIMATE OF THE POPULATION DIFFERENCE	-72000.	-72301.
POINT ESTIMATE OF THE POPULATION TOTAL	1928000.	1927699.
STANDARD ERROR OF THE MEAN	4.4	3.9
STANDARD ERROR OF THE TOTAL	44092.	39259.
PRECISION	72311.	64384.
CONFIDENCE INTERVAL FOR POPULATION DIFFERENCE	-144311. TO 311.	-136685. TO -7917.
CONFIDENCE INTERVAL FOR POPULATION TOTAL	1855689.TO 2000311.	1863315.TO 1992083.

NOTE: NEGATIVE DIFFERENCES REPRESENT AN OVERSTATEMENT IN THE BOOK AMOUNT

PROGRAM COMPLETE

NOW AT 4160
SRU'S:0.5
READY

III - Program Listing

```
1000.     REM***PROGRAM FEE***12-MARCH-74
1010.     PRINT "THIS PROGRAM IS USED TO EVALUATE A SAMPLE DESIGNED TO ESTIMATE"
1020.     PRINT "THE POPULATION DIFFERENCE WHERE STRATIFICATION IS NOT USED."
1030.     PRINT "RESULTS ARE GIVEN ON THE BASIS OF BOTH THE DIFFERENCES AND"
1040.     PRINT "THE RATIOS OF THE SAMPLE ITEMS. HOWEVER, IF ONLY THOSE SAMPLE"
1050.     PRINT "ITEMS WHERE DIFFERENCES OCCURRED ARE ENTERED, ONLY THE"
1060.     PRINT "RESULTS ON THE BASIS OF DIFFERENCES CAN BE USED."
1070.     PRINT
1080.     PRINT
1090.     PRINT "ARE SAMPLE DETAILS TO BE ENTERED (YES OR NO)"
1100.     INPUT Q$
1110.     PRINT
1120.     PRINT "IS THIS A ONE-SIDED ESTIMATE OR A TWO-SIDED ESTIMATE"
1130.     PRINT "ENTER 1 OR 2"
1140.     INPUT T
1150.     PRINT
1160.     IF T=1 GOTO 1200
1170.     IF T=2 GOTO 1230
1180.     PRINT "ERROR. ENTER AGAIN"
1190.     GOTO 1140
1200.     PRINT "UPPER OR LOWER LIMIT (U OR L)"
1210.     INPUT L$
1220.     PRINT
1230.     PRINT "WHAT RELIABILITY PERCENTAGE IS DESIRED (EXAMPLE 90)"
1240.     INPUT R
1250.     PRINT
1260.     IF T=1 GOTO 1430
1270.     IF T=2 GOTO 1700
1280.     INPUT R
1290.     GOTO 1430
1300.     INPUT R
1310.     GOTO 1700
1320.     PRINT"ENTER TOTAL POPULATION SIZE"
1330.     INPUT N2
1340.     PRINT
1350.     PRINT "ENTER THE POPULATION TOTAL BOOK VALUE"
1360.     INPUT A2
1370.     PRINT
1380.     PRINT "ENTER THE SIZE OF YOUR SAMPLE"
1390.     INPUT N1
1400.     PRINT
1410.     GOTO 1960
1420.     REM TABLE FOR ONE-SIDED U VALUES
1430.     IF R=70 GOTO 1550
1440.     IF R=75 GOTO 1570
1450.     IF R=80 GOTO 1590
1460.     IF R=85 GOTO 1610
1470.     IF R=90 GOTO 1630
1480.     IF R=95 GOTO 1650
1490.     IF R =99 GOTO 1670
1500.     PRINT "VALUE OF RELIABILITY PERCENTAGE NOT ACCEPTABLE"
1510.     PRINT "ENTER NEW VALUE"
1520.     GOTO 1240
1530.     PRINT
```

```
1540.     GOTO 1280
1550.     LET U=.52
1560.     GOTO 1320
1570.     LET U=.67
1580.     GOTO 1320
1590.     LET U=.84
1600.     GOTO 1320
1610.     LET U=1.04
1620.     GOTO 1320
1630.     LET U=1.28
1640.     GOTO 1320
1650.     LET U=1.64
1660.     GOTO 1320
1670.     LET U=2.33
1680.     GOTO 1320
1690.     REM TABLE FOR TWO-SIDED U VALUES
1700.     IF R=70 GOTO 1820
1710.     IF R=75 GOTO 1840
1720.     IF R=80 GOTO 1860
1730.     IF R=85 GOTO 1880
1740.     IF R=90 GOTO 1900
1750.     IF R=95 GOTO 1920
1760.     IF R=99 GOTO 1940
1770.     PRINT "VALUE OF RELIABILITY PERCENTAGE NOT ACCEPTABLE"
1780.     PRINT "ENTER NEW VALUE"
1790.     GOTO 1240
1800.     PRINT
1810.     GOTO 1300
1820.     LET U=1.04
1830.     GOTO 1320
1840.     LET U=1.15
1850.     GOTO 1320
1860.     LET U=1.28
1870.     GOTO 1320
1880.     LET U=1.44
1890.     GOTO 1320
1900.     LET U=1.64
1910.     GOTO 1320
1920.     LET U=1.96
1930.     GOTO 1320
1940.     LET U=2.58
1950.     GOTO 1320
1960.     REM***** COMPUTATION OF VARIANCE FROM PRELIMINARY SAMPLE
1970.     IF Q$="NO" GOTO 3620
1980.     PRINT "HAVE PRELIMINARY SAMPLE ITEMS BEEN STORED IN DATA"
1990.     PRINT "STATEMENTS (YES OR NO)";
2000.     INPUT K$
2010.     IF K$="YES "THEN 2240
2020.     IF K$="NO" THEN 2040
2030.     GO TO 1980
2040.     PRINT
2050.     PRINT "ENTER VALUES OF PRELIMINARY SAMPLE ITEMS IN DATA STATE-"
2060.     PRINT "MENTS - AUDITED VALUE,BOOK VALUE - AS FOLLOWS. LINE NUMBER
          RANGE MUST"
2070.     PRINT "BE FROM 1 TO 999. SEPARATE ALL VALUES BY COMMAS; MAKE"
```

```
2080.     PRINT "SURE THAT THE LAST VALUE ON A LINE IS NOT FOLLOWED BY A "
2090.     PRINT "COMMA."
2100.     PRINT
2110.     PRINT "     LLL DATA V(1) , V(2) , V(3), . . . . V(N)"
2120.     PRINT
2130.     PRINT "            WHERE LLL  REPRESENTS A LINE NUMBER IN THE RANGE"
2140.     PRINT "            1 TO 999."
2150.     PRINT "            V(N) REPRESENTS A SAMPLE ITEM ENTERED IN THE"
2160.     PRINT"            DATA STATEMENT."
2170.     PRINT
2180.     PRINT "DURING RUNNING OF THIS PROGRAM YOU ARE ASKED THE SIZE"
2190.     PRINT "OF THE PRELIMINARY SAMPLE.  MAKE SURE THAT THE NUMBER"
2200.     PRINT "YOU ENTER AT THAT  TIME IS THE SAME AS THE NUMBER OF "
2210.     PRINT "SAMPLE ITEMS ENTERED IN THE DATA STATEMENT."
2220.     PRINT
2230.     GOTO 4160
2240.     LET J=0
2250.     LET X1=0
2260.     LET X2=0
2270.     LET Y1=0
2280.     LET Y2=0
2290.     LET D1=0
2300.     LET D2=0
2310.     LET Z7=0
2320.     FOR K8=1 TO N1
2330.     READ Z9,Z8
2340.     LET J=J+1
2350.     LET X1=X1+Z9
2360.     LET X2=X2+Z9↑2
2370.     LET Y1=Y1+Z8
2380.     LET Y2=Y2+Z8↑2
2390.     LET D1=D1+(Z9-Z8)
2400.     LET D2=D2+(Z9-Z8)↑2
2410.     LET Z7=Z7+(Z9*Z8)
2420.     NEXT K8
2430.     REM*****END OF DATA STATEMENT READ
2440.     IF J=N1 GOTO 2480
2450.     PRINT
2460.     PRINT USING 2470, N1, J
2470.     :WARNING--SAMPLE SIZE GIVEN AS ####; #### SAMPLE ITEMS ENTERED
2480.     LET R1=X1/Y1
2490.     LET C3=X2-(2*R1*Z7)+(R1↑2*Y2)
2500.     LET C4=C3/(N1-1)
2510.     LET C5=D1/N1
2520.     LET C6=N1*C5↑2
2530.     LET C7=(D2-C6)/(N1-1)
2540.     LET S1=SQR(C7)
2550.     LET V1=SQR(C4)
2560.     PRINT
2570.     PRINT
2580.     PRINT "THE SAMPLE STATISTICS ARE:"
2590.     PRINT
2600.     PRINT "                    ON THE BASIS OF SAMPLE"
2610.     PRINT "                   DIFFERENCES        RATIOS"
```

```
2620.   PRINT USING 2630,C5,R1
2630.   :                MEAN                                      ######.#      #####.#
2640.   PRINT USING 2650,C7,C4
2650.   :                VARIANCE                                 #######.#    #######.#
2660.   PRINT USING 2670,S1,V1
2670.   :                STANDARD DEVIATION                       ######.##    ######.##
2680.   PRINT
2690.   PRINT
2700.   REM COMPUTATION OF POPULATION ESTIMATES
2710.   LET D3=C5*N2
2720.   LET V3=(R1*A2)-A2
2730.   LET D4=A2+D3
2740.   LET W1=(N2-N1)/N2
2750.   LET V4=V3+A2
2760.   LET W2=SQR(W1)
2770.   LET W3=SQR(N1)
2780.   LET D5=(S1/W3)*W2
2790.   LET V5=(V1/W3)*W2
2800.   LET D6=N2*D5
2810.   LET V6=N2*V5
2820.   LET D7=U*D6
2830.   LET V7=U*V6
2840.   LET D8=D3+D7
2850.   LET V8=V3+V7
2860.   LET D9=D4+D7
2870.   LET V9=V4+V7
2880.   LET D0=D3-D7
2890.   LET V0=V3-V7
2900.   LET E1=D4-D7
2910.   LET F1=V4-V7
2920.   IF T=1 GOTO 2940
2930.   IF T=2 GOTO 3270
2940.   PRINT "THE ESTIMATED POPULATION STATISTICS FOR A ONE-SIDED ESTIMATE"
2950.   PRINT USING 2960,R
2960.   :WITH A RELIABILITY OF ###% ARE AS FOLLOWS:
2970.   PRINT
2980.   PRINT "                                                ON THE BASIS OF SAMPLE"
2990.   PRINT "                                               DIFFERENCES         RATIOS"
3000.   PRINT "               POINT ESTIMATE OF THE"
3010.   PRINT USING 3020,D3,V3
3020.   :                  POPULATION DIFFERENCE          ########.     ########.
3030.   PRINT "               POINT ESTIMATE OF THE"
3040.   PRINT USING 3050,D4,V4
3050.   :                  POPULATION TOTAL               ########.     ########.
3060.   PRINT USING 3070,D5,V5
3070.   :                STANDARD ERROR OF THE MEAN       #######.#     #######.#
3080.   PRINT USING 3090,D6,V6
3090.   :                STANDARD ERROR OF THE TOTAL      ########.     ########.
3100.   PRINT USING 3110,D7,V7
3110.   :                PRECISION                        ########.     ########.
3120.   IF L$="L" GOTO 3200
3130.   PRINT "              UPPER CONFIDENCE LIMIT FOR"
3140.   PRINT USING 3150,D8,V8
3150.   :                  POPULATION DIFFERENCE          ########.     ########.
```

```
3160.    PRINT "              UPPER CONFIDENCE LIMIT FOR"
3170.    :                POPULATION TOTAL            ########.    ########.
3180.    PRINT USING 3170,E1,F1
3190.    GOTO 3560
3200.    PRINT "              LOWER CONFIDENCE LIMIT FOR"
3210.    PRINT USING 3220,D0,V0
3220.    :                POPULATION DIFFERENCE       ########.    ########.
3230.    PRINT "              LOWER CONFIDENCE LIMIT FOR"
3240.    PRINT USING 3250,E1,F1
3250.    :                POPULATION TOTAL            ########.    ########.
3260.    GOTO 3560
3270.    PRINT "THE ESTIMATED POPULATION STATISTICS FOR A TWO-SIDED ESTIMATE"
3280.    PRINT USING 3290,R
3290.    :WITH A RELIABILITY OF ####% ARE AS FOLLOWS:
3300.    PRINT
3310.    PRINT "                                       ON THE BASIS OF SAMPLE"
3320.    PRINT "                                      DIFFERENCES       RATIOS"
3330.    PRINT "              POINT ESTIMATE OF THE"
3340.    PRINT USING 3350,D3,V3
3350.    :                POPULATION DIFFERENCE      ########.     ########.
3360.    PRINT "              POINT ESTIMATE OF THE"
3370.    PRINT USING 3380,D4,V4
3380.    :                POPULATION TOTAL           ########.     ########.
3390.    PRINT USING 3400,D5,V5
3400.    :              STANDARD ERROR OF THE MEAN     #######.#      #######.#
3410.    PRINT USING 3420,D6,V6
3420.    :              STANDARD ERROR OF THE TOTAL  ########.      ########.
3430.    PRINT USING 3440,D7,V7
3440.    :              PRECISION                    ########.      ########.
3450.    PRINT "              CONFIDENCE INTERVAL FOR"
3460.    PRINT USING 3470,D0,V0
3470.    :                POPULATION DIFFERENCE     ########. TO   ########. TO
3480.    PRINT USING 3490,D8,V8
3490.    :                                             ########.      ########.
3500.    PRINT "              CONFIDENCE INTERVAL FOR"
3510.    PRINT USING 3520,E1,F1
3520.    :                POPULATION TOTAL           ########.TO    ########.TO
3530.    PRINT USING 3540,D9,V9
3540.    :                                             ########.      ########.
3550.    IF F$="YES" GOTO 3600
3560.    PRINT
3570.    PRINT "NOTE: NEGATIVE DIFFERENCES REPRESENT AN OVERSTATEMENT IN"
3580.    PRINT "      THE BOOK AMOUNT"
3590.    PRINT
3600.    PRINT
3610.    GOTO 4030
3620.    REM INPUT OF SAMPLE TOTALS WHERE DETAILS NOT SPECIFIED
3630.    PRINT "WHAT IS THE TOTAL OF THE SAMPLE AUDITED VALUES"
3640.    INPUT X1
3650.    PRINT
3660.    PRINT "WHAT IS THE TOTAL OF THE SQUARES OF THE SAMPLE AUDITED VALUES"
3670.    INPUT X2
3680.    PRINT
3690.    PRINT "WHAT IS THE TOTAL OF THE SAMPLE BOOK VALUES"
```

```
3700.     INPUT Y1
3710.     PRINT
3720.     PRINT "WHAT IS THE TOTAL OF THE SQUARES OF THE SAMPLE BOOK VALUES'
3730.     INPUT Y2
3740.     PRINT
3750.     PRINT "WHAT IS THE TOTAL OF THE SAMPLE DIFFERENCES"
3760.     INPUT D1
3770.     PRINT
3780.     PRINT "WHAT IS THE TOTAL OF THE SQUARES OF THE SAMPLE DIFFERENCES"
3790.     INPUT D2
3800.     PRINT
3810.     PRINT "WHAT IS THE TOTAL OF THE CROSS-PRODUCTS OF THE SAMPLE"
3820.     PRINT "AUDITED AND BOOK VALUES"
3830.     INPUT Z7
3840.     GOTO 2480
3850.     REM N1=COMPUTED SAMPLE SIZE
3860.     REM N2=POPULATION SIZE
3870.     REM X1=SAMPLE ITEM AUDITED VALUE
3880.     REM Y1=SAMPLE ITEM BOOK VALUE
3890.     REM X2=SAMPLE ITEM AUDITED VALUE SQUARED
3900.     REM Y2=SAMPLE ITEM BOOK VALUE SQUARED
3910.     REM R1=RATIO OF SUMS OF SAMPLE ITEM AUDITED AND BOOK VALUES
3920.     REM X3=POPULATION VALUE
3930.     REM R=RELIABILITY %
3940.     REM U=NORMAL DEVIATE OF R
3950.     REM S1=STANDARD DEVIATION
3960.     REM T=ONE-SIDED/TWO-SIDED OPTION
3970.     REM C1-C9=COMPUTATIONAL WORK FIELDS
3980.     REM Z1-Z6=SAMPLE WORK FIELDS
3990.     REM D1=SAMPLE ITEM DIFFERENCE
4000.     REM D2=SAMPLE ITEM DIFFERENCE SQUARED
4010.     REM D3-0;V3-0;W1-3:E1;F1=COMPUTATIONAL FIELDS
4020.     REM Q$=SAMPLE DETAILS/TOTALS OPTION
4030.     IF Q$="NO" GOTO 4120
4040.     PRINT "SUM OF X",X1
4050.     PRINT "SUM OF Y".Y1
4060.     PRINT "SUM OF D",D1
4070.     PRINT "SUM OF X SQR",X2
4080.     PRINT "SUM OF Y SQR",Y2
4090.     PRINT "SUM OF D SQR",D2
4100.     PRINT "SUM OF XY",Z7
4110.     PRINT
4120.     PRINT
4130.     PRINT "PROGRAM COMPLETE"
4140.     PRINT
4160.     END
4170.     REM K$=DATA STORAGE TEST
4180.     REM L$=UPPER/LOWER LIMIT OPTION
```