1973

Auditor's Approach to Statistical Sampling, Volume 1. Introduction of Statistical Concepts and Estimation of Dollar Values

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

1

AN AUDITOR'S APPROACH TO STATISTICAL SAMPLING

AN INTRODUCTION TO STATISTICAL CONCEPTS AND ESTIMATION OF DOLLAR VALUES

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

AICPA Continuing Professional Education

NOTICE TO READERS

This programed learning text 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 programed by Teaching Systems Corporation. The members of the Committee on Statistical Sampling assisted in an advisory capacity.

AN AUDITOR'S APPROACH TO STATISTICAL SAMPLING

Volume 1

AN INTRODUCTION TO STATISTICAL CONCEPTS AND ESTIMATION OF DOLLAR VALUES

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

TABLE OF CONTENTS

PREFACE

In recent years interest in applying statistical techniques to auditing problems has increased. Many CPAs, however, have not had courses in statistics. Others find they have forgotten the concepts already learned.

To help fill the educational void, we have prepared a series of programed instruction texts. This volume is a part of that series. It is designed for auditors who want to use statistics in audit engagements; it is *not* a statistics text.

The purpose of this volume is two-fold. First, it discusses some basic statistical concepts.

Second, it illustrates how mean value estimation techniques can be used in an auditing environment. However, other techniques may be more useful in some situations, so all volumes in this series should be completed before attempting to apply any of them in practice.

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

ROBERT E. SCHLOSSER, PH.D., CPA, *Director*
Continuing Professional Education Division

Revised November 1973
(original printing November 1966)

INTRODUCTION

An Introduction to Statistical Concepts and Estimation of Dollar Values is a "programed instruction" text. The essential feature of this relatively new method is a step-by-step approach that requires the reader to absorb each point thoroughly before going on to the next. You will notice that the pages are divided into three "frames." Most frames require you to answer a question, or work out a problem, in addition to reading the material. Complete instructions are given below.

Although this method may seem slow at first, it virtually guarantees that the reader who conscientiously works through each frame will, upon completing the book, be able to solve problems, answer questions, and perform tasks that involve all the knowledge required for applying the concepts taught to actual auditing situations. *If you turn right now to the Questions and Problems beginning on page S-27 of the Supplementary Section, you will see from the "Giant Franchise Company" problem and the questions following it that the objectives of this book are almost entirely practical rather than theoretical.* The reader who has invested the time required (approximately eight to ten hours) in the prescribed manner will be able to work out field problems of this nature without further instruction.

The text has been designed to accommodate those with no statistical background as well as those who have had an elementary course in statistics. Since each reader will have a different background, you will probably find yourself wishing at some point that the instruction would proceed at a faster or slower pace. Some exercises have therefore been made optional, and certain frames have been designated to be skipped by the reader who has evidenced mastery of the topic under discussion. However, except for these clearly specified cases, maximum value will be obtained from this book only by going through each frame according to the sequence "programed" by the authors. If the early chapters seem relatively easy, it is only in order that the reader will be fully prepared for the more difficult later chapters.

INSTRUCTIONS FOR USE

A. Programed Text

Each frame contains either a *blank* to fill in, a *choice* to circle, a *direction* to do an exercise or read an Exhibit in the Supplementary Section, or the phrase "No answer required." In the latter case, simply read the material and go on to the next frame. The "no answer" frames are, however, as important in the teaching sequence as the others, and usually are followed up with a question in a later frame.

If an answer is required, write it directly in the space provided. When a choice is given, it is recommended that you circle the correct one rather than crossing out the wrong one or answering mentally. All questions are designed to be answered easily, based on the material immediately preceding or on your own judgment. The correct answer appears on the following page in the space adjacent to the next frame. If your response is incorrect, you are advised to cross it out and substitute the correct one.

A special format used occasionally in later chapters is the "branching" frame. This consists of a multiple-choice question together with an instruction to turn to a specified frame depending on the answer you give. Except for these clearly indicated cases, however, all answers appear on the page immediately following the question and should be checked before you proceed to the next frame.

The frames are not to be read *down* the page, but rather go in numerical order *across* the book. Thus, every frame in the top row will be read before going on to the middle row. After all three rows are completed, the book is turned over and the same procedure begins again on the pages that previously were upside down. Until that time, these pages are ignored.

B. Supplementary Section

The smaller booklet serves as both a workbook and review book, and is also designed to be a reference manual for on-the-job use. The Worksheets in this booklet have additional space for use in your own unrestricted sampling applications. The programed text directs you to the relevant pages in this book at the proper time, and it may also be referred to whenever necessary. As a helpful overview before beginning the programed text, the reader may wish to look at the "Reference Guide for Unrestricted Random Sampling with Replacement" found on pages S-15 and S-16 in the Supplementary Section, and to skim through the chapter summaries beginning on page S-22. In addition, the individual summaries should be read immediately before and immediately after each chapter. The recommended procedure is to do one chapter at one uninterrupted sitting.

A NOTE ON CONTENT

This book is not designed to be a statistics course. Rather, it examines in detail one particular statistical technique — unrestricted random sampling with replacement — solely for the purpose of aiding in an auditing situation. Mathematical proofs and derivations have been omitted or made optional. Even the "basic statistical concepts" referred to in the title have been screened so that only those with practical relevance are covered. Certain basic statistical concepts such as Median, Mode, Normal Curve, and everything having to do with probability theory and small-sample applications are not discussed (although the reader with some statistical background will be able to see how the normal curve relates to the concepts presented in Chapter 5).

Due to this approach, questions may arise due to curiosity or philosophical considerations while you are going through this book. In addition, questions will almost certainly occur to you having to do with types of situations not included under the topic of unrestricted sampling. The fact that these questions cannot be answered in this book has nothing to do with the method of programed instruction, but rather results from the decision to make this instruction as effective as possible for practical applications with a minimum investment of the reader's time. Questions of this nature, therefore, will either be answered in future volumes in this series, or will remain unanswered without detracting from the book's practical effectiveness.

CHAPTER 1

CHAPTER 1. SPECIFYING THE SAMPLING OBJECTIVES

1-1. Statistical sampling consists of a variety of techniques for obtaining information about a certain body of data, based on inferences from a relatively small portion of the data.

If you look again at the above paragraph, you may notice that to some extent this same description (WOULD NOT/WOULD ALSO) apply to auditing.

CIRCLE THE CHOICE YOU THINK CORRECT. THEN TURN THE PAGE.

WOULD ALSO

(The correct answer always appears in this box.)

Now go on to the Frame at the right.

1-2. An auditor is accustomed to making inferences and drawing conclusions about a company's account balances without examining every single voucher, check or journal entry. Similarly, by means of statistical technique, the auditor can make reasonable estimates on the basis of relatively small samples.

This series will not attempt to make you into a professional statistician. What it will do is teach the basic techniques that can give you a sound basis for making statistically valid estimates.

NO ANSWER REQUIRED. TURN THE PAGE.

No answer required	REMINDER As you go through this book, remember that you will ignore the material in the rows above and below the one you are reading. If you make a mistake, the recommended procedure is to go back and cross out your answer and substitute the correct answer. TURN THE PAGE AND GO ON TO #1-3.
	2-10. If you were to draw several samples from the same population, the estimate and precision limits would probably be slightly different each time. The proportion of cases in which the precision interval will contain the actual value is known as the (PRECISION/RELIABILITY) of the estimate.
	3-21. The next two numbers, 669 and 963, are also unusable. We then come to four numbers that correspond to elements in the population. Place an asterisk next to each of the corresponding Lot designations in Exhibit 3-A. (You will probably find it easier to list the numbers first and then put in the asterisks rather than switching back and forth each time.)

Go on to #1-3.	1-3. There are many kinds of estimates that an auditor may find useful, but basically every estimate is either of a quantity, or of a rate of occurrence. The statistical terms that roughly correspond to "quantities" and "rates of occurrence" are, respectively, <u>variables</u> and <u>attributes</u>. In this book we discuss estimates of variables. An example of a variable, as distinct from an attribute, would be (circle one): a. proportion of receivables that are overdue b. total dollar value of an inventory
RELIABILITY	2-11. If you wish to increase the reliability of an estimate (assuming no change in the sample size), what must you do with your desired precision? a. Make it greater (as from $5,000 to $10,000) b. Make it smaller (as from $10,000 to $5,000) EXPLAIN YOUR CHOICE: ______________________________ ______________________________
*HT-1E6E (190) *VM-3Q2W (280) *SN-9A8N (290) *LP-5Y6W (157)	3-22. Having come to the end of Column 1, we start again at the top of Column 2 and continue in the same manner. Why? a. This happens to be the route we specified. b. Columns in a random number table are always read left to right.

b. total dollar value of an inventory	1-4. In estimating a numerical quantity or "variable," there is always a range of values within which the estimate may vary. Thus, if you were to make several estimates of the same quantity, using the same procedures but basing your estimate on a different sample each time, these estimates would tend to ___________ between some lower and some upper limit.
a. Make it greater Explanation: By widening the precision limits, you increase the chances of the actual value falling within the precision limits.	2-12. Once the objectives of the sampling task have been specified, the next step is to choose a sampling plan that will result in a statistically valid sample. In this book we will consider only the most basic type of sampling procedure, unrestricted random sampling with replacement. The word "random," of course, suggests that the sample selection should be governed by: a. chance b. intuition c. analysis
a. This happens to be the route we specified.	3-23. Continue in the same manner, placing an asterisk next to each element selected, but stop after you have come to 260 in Row 484 and have entered the appropriate asterisk.

vary (range, fall, etc.)	1-5. There are certain other distinctions between "variables" and "attributes" that we do not need to go into here. In this book we will only be concerned with estimating the former. For our purposes, then, a "variable" is a (QUANTITY/RATE OF OCCURRENCE) that can fall: a. only into certain distinct categories b. anywhere within a given range
a. chance	2-13. The selection of the sample on a random basis guarantees that the person selecting the sample will not influence or bias the sample selection either consciously or unconsciously. Moreover, a <u>random</u> sample enables us to make estimates that have a valid mathematical basis, since the mathematical theorems that are used in statistics are based on the laws of chance. (No answer required)
You should have selected 064, 148, 250, 276, 158 and 260.	3-24. The next usable digit, 276, is one that has already appeared. This in itself is not unusual. What may be surprising is that we use it <u>again</u>. In our computations we will count the value of this twice, just as if it were two items, and will count this item twice in making up our required preliminary sample size of 30. In other words, we may end up with only 28 or 29 different Lots, but we are not counting Lots -- we are actually counting the number of "sampling elements" as indicated by the number of asterisks on our worksheet. Therefore, place a second asterisk next to the Lot that corresponds to number 276 in Exhibit 3-A.

QUANTITY b. anywhere within a given range	1-6. Now let's see how statistical estimation might be applied in practice. Turn to Exhibit 1 in the Supplementary Section. Read the entire Exhibit to get an overall view of the auditor's task. There is no need to memorize the details since we will be referring to this Exhibit frequently. (No answer required. After reading Exhibit 1, go on to Frame 1-7).
No answer required	2-14. The two requirements of unrestricted random sampling are first that every element in the population should have an equal chance of being included in the sample and second, that each combination of elements has an equal chance of being chosen. Thus, although it may turn out that inventory lot FM-1I6N is not selected in the sample while lot JH-5E6E is, a properly designed random sampling plan will insure that their <u>chances</u> of selection are ____________. We can be sure that this is the case only if the selection method is (ARBITRARY/RANDOM).
**NC-1E6S	3-25. This procedure is known as "replacement." After drawing the number 276 the first time, we did not "take it out of circulation." Rather, we <u>replaced</u> it back into our pool of usable numbers (all numbers from 001 through 300). Thus, in sampling <u>with</u> <u>replacement</u>, a random number and its corresponding element are selected and counted: a. only once b. as often as they happen to appear

No answer required	1-7. In this volume, we will use the "ABC Store" problem as an example of <u>how</u> to apply the technique of unrestricted random sampling. (NOTE: This example, together with all the others in this book, has been constructed for teaching purposes only.) (No answer required)
equal RANDOM	2-15. However, it is difficult to achieve true "randomness" without some kind of aid. One device which is frequently used to help insure random selection is a <u>random number table</u>. Locate your two-page table in the Supplementary Section (page S-38) and keep it readily accessible throughout this chapter. (No answer required) <u>NOTE</u> This two-page table has been produced for teaching purposes only, and has not been certified for randomness. For actual field work use the table of random numbers beginning on page S-41 or any other published table.
b. as often as they happen to appear	3-26. In practice, you will use sampling without replacement. However, this introductory volume uses sampling with replacement in order to simplify the formulas. These simple formulas will provide conservative answers when sampling without replacement is used. However, the agreement between these simple formulas and the more complex ones which will be introduced in later volumes is satisfactory whenever the sample size does not exceed 5% of the population. (No answer required)

No answer required	1-8. It has been assumed, then, that we will not go to the expense of examining all 9,000 accounts in order to make a complete aging. Instead, we can solve the problem by means of statistical estimation. This would involve examining a sample of the 9,000 accounts and on the basis of the sample making an ______________ of the total dollar value of the three-month-or-more overdue amounts.
No answer required	2-16. The digits in the table have been randomly generated by a computer. In a properly randomized table, there is approximately an equal number of each of the digits 0 through 9, and the order in which the digits appear is completely random. The numbers at the extreme left (401, 402, etc.) and top (0, 1, etc.) of each page are used only for purposes of identification. For example, line 415-1 of this table consists of the digits ____________.
No answer required	3-27. If we happen to draw 276 a third time, what will we do? ______________________________________ ______________________________________

estimate	1-9. Making an estimate on the basis of a sample is a common occurrence in everyday life. For example, suppose that in half an hour you have finished reading ten pages of a 65-page report. Including the half hour already spent, how long do you think it would take to read the entire report? YOUR ESTIMATE: __________
965749	2-17. We can illustrate the use of the random number table with a simple example. Suppose you wish to select a random sample of 10 invoices out of a population of 100, in order to estimate the total dollar amount of the 100 invoices. Assume that the invoice numbers run consecutively from 1 through 100. Let us suppose (and here we can be arbitrary for the moment) that we start in row 411, column 2, and decide to go down the column and take the last two digits in each line. The first two-digit number drawn in this fashion is ______.
Place a third asterisk next to Lot NC-1E6S and count it once again as "another" element in our preliminary sample of 30.	3-28. As an optional exercise, which is especially recommended if you had trouble with any of the preceding questions, continue from where you left off (Row 485, Column 2) and complete the selection of the preliminary sample of 30 items. The last number to be used is 132 in 499-3.

195 minutes (3¼ hours)	1-10. Obviously there are several assumptions that have to be made before the question can be answered with any degree of confidence. But assuming for the time being that you would continue reading uninterruptedly at the same rate, the problem is one of simple arithmetic. The average reading time per page was ____ minutes. Multiplying this by the number of pages, 65, you would arrive at an estimate of 195 minutes or three hours and 15 minutes.
73 (last two digits of Row 411, Column 2)	2-18. This means that Invoice #73 will be the first element selected in our sample. Going down to the next line (412-2), the last two digits are 43. And, just to be sure you are in the right place, the last two digits in the line below that are _____.
Instructions for checking your work are given in the next frame.	3-29. It is usually advisable to enter the sampling information in a worksheet. Turn to Exhibit 5, in which this has been done for you. Focus only on the second and third columns. Then, if you wish, check off each element listed in the worksheet against your list in Exhibit 3-A to make sure that you have selected the sample accurately. (No answer required)

three (3)	1-11. In this example, we first computed an ___________ reading time per page, amounting to three minutes, and then used this as a basis for estimating the ___________ reading time required.
38	2-19. By letting these random numbers correspond to the invoice numbers, we have thus far selected three elements of the sample as shown below. Now, continuing where we left off (Row 414, Column 2), complete the sample selection by entering the next seven sets of two-digit numbers in the box below.
No answer required	3-30. For further practice, we will now go through essentially the same sampling procedure for our ABC Store accounts receivable example. Refresh your memory of this problem by rereading Exhibit 1, with special emphasis on the third paragraph. Can we establish correspondence by having the actual account numbers correspond to the random numbers, as we did in the invoice example? (YES/NO) EXPLAIN: ______________________________ ______________________________

INVOICES SELECTED IN THE SAMPLE (box in frame 2-19):

Invoice #73	Invoice #____
" 43	" ____
" 38	" ____
" ____	" ____
" ____	" ____

average total	1-12. We can apply the same reasoning to the accounts receivable example in Exhibit 1. Suppose that on the basis of a sample, you compute the average value of the three-month overdue amounts in each account to be $50.00. If this is the average for one account, then for the 9,000 accounts you could estimate the total value of the amounts that are overdue three months or more to be $________. Of course you would like to be reasonably confident that your estimate is close to the actual overdue amount.
73 47 43 14 38 93 4 (or 04) 13 61 84	2-20. These were the invoices that happened to be selected, but the <u>chances</u> of selection were equal for each item. For instance, if we had started in Row 416, Column 6, we would have obtained ten different sets of digits. Or, if we had started in Column 1 of Row 411 instead of Column 2, we would once again have selected invoices #43 and #84, but the other eight would have been different. You can verify this yourself as an <u>optional</u> exercise.
NO The account numbers might well be in an irregular or broken sequence which would make it difficult to establish correspondence.	3-31. Instead, should we follow the procedure used in the shoe inventory problem, listing the accounts and artificially numbering them 1 through 9,000 to correspond to the digits in the table? (YES/NO) EXPLAIN: ____________________ ____________________

$450,000 ($50 x 9,000)	1-13. This leaves several important questions unanswered. Among them are: 1. What do we mean by "reasonably confident?" 2. What do we mean by "close to the actual overdue amount?" These questions have to do with evaluating the accuracy of an estimate, and are a major concern of the remainder of the book. What you have seen so far is that the step of making an estimate is actually quite simple, and involves finding the ____________ value for each element, and then multiplying by the number of elements to arrive at an estimate of the ____________ value.
Beginning in 411-1, the ten numbers are: 99 17 84 16 41 19 68 24 49 43	2-21. In this example, we randomized the selection process by letting the random number 01 represent invoice #1, random number 02 represent invoice #2, and so on. Could we let the random number 100 represent invoice #100? a. YES (If this is your answer, turn to Frame 2-23.) b. NO (If this is your answer, turn to Frame 2-22.)
NO (explanation is in the frame at the right)	3-32. This is not necessary in this case. The accounts are stored on punch card equipment, so there is no need for a master list, or at least no evidence of one given in the problem. Fortunately, these accounts are stored in a separate file, so we can dispense with the extra labor of physically numbering them 1 through 9,000, and do the same thing mentally. That is, if we draw random numbers 2254 and 7843, for example, we will simply have our equipment locate the 2254th and 7843rd accounts in the order in which they appear in the file. Notice how this is worded in Exhibit 6. (No answer required)

average total	1-14. There are still some more important questions to be answered. In the "Reading" example, two such questions undoubtedly occurred to you: 1. Were the ten pages a fair sample of all 65 pages, or was the sample "biased" in some way? 2. Even if they were chosen fairly, is ten a sufficient number of pages to take as a sample? If we were to compute the average value of a sample of the ABC Store accounts receivable, the same kinds of questions (WOULD/ WOULD NOT) apply.
	2-22. YOUR ANSWER: NO, in this example we could not let the random number 100 correspond to invoice #100. Correct. Now explain why not: ________________________________ ________________________________ (The correct explanation appears in the next frame.)
No answer required	3-33. After establishing correspondence, the next step is to select a route through the table. We will use the route which you may have noticed in Exhibit 6, but for additional practice, specify another possibility: ________________________________ ________________________________ ________________________________

WOULD	1-15. These questions are among those that will be answered in this book. In the remainder of this chapter, we will define some of the basic concepts in statistical sampling and then apply them to our accounts receivable example. In Chapters 2 and 3 we will show how to guard against selection bias by making a random sample selection. In the remaining chapters we will discuss how to choose a sample size and evaluate the accuracy of an estimate. (No answer required)
	2-23. YOUR ANSWER: YES, we could let the random number 100 correspond to invoice #100. We could -- but not in this example, in which we are only working with two-digit numbers. (No answer required)
There is, of course, no one correct answer. Evaluate your response based on the material in the frame at the right.	3-34. Any route is satisfactory as long as it is unambiguous to other persons, and in particular specifies the following: 1. Which places in each line will be used? (first four, last four, middle four, last two and first two in next column, etc.) 2. What direction does the route take? (down, up, sideways, etc.) 3. What happens at the end of a row or column? (doubling back is not permissible) (No answer required)

No answer required	1-16. In solving our ABC Store problem, or any statistical estimation problem, the first major step after defining the auditing objectives is to specify the sampling objectives. We have already stated that the purpose of statistical sampling is "to obtain information about a certain body of data." We therefore begin by specifying: 1. the body of data that is under consideration 2. the ___________ we wish to obtain about this body of data.
No answer required	2-24. We cannot simply ignore invoice #100, since it is part of the population whose total we wish to estimate. If we wish to restrict ourselves to two-digit numbers, therefore, we can let invoice #100 be represented by the digits 00. These are just as likely to appear in the table as 14, 33, or any other two digits, so that each of the invoices will have an equal chance of being selected in the sample. Moreover, there are a very large number of (1 followed by 20 zeros or 10^{20}) different possible samples of ten invoices chosen from the 100, and this procedure will give an equal chance to each of these possibilities. (No answer required)
No answer required	3-35. We will use the random start method discussed in the last section. Describe briefly how you would go about selecting a starting point randomly. ______________________________ ______________________________ ______________________________

information	1-17. The "body of data about which we need certain information" is known as the population. In the problem described in Exhibit 1, how many elements are in the population? a. 9,000 b. 51,000 c. 240,000 EXPLAIN YOUR ANSWER: ____ ____
No answer required	2-25. The example of the 100 invoices illustrated the three steps involved in using a random number table to select an unrestricted random sample. These are as follows: 1. Establish correspondence between the elements in the population and the digits in the table. We did this by having each invoice number correspond to the identical two-digit random number. (Invoice #100 was represented by 00.) 2. Select a route through the table. We took the last two digits of each line and proceeded down consecutive lines. 3. Select a starting point. We chose 411-2. (In practice you would make the selection randomly, as shown later in this book.) (No answer required)
Make a "blind stab" with the pencil; let the first three digits on the nearest usable line represent the row, and the next digit represent the column. (You could also let the last three, or any three consecutive digits, represent the row, and any other digit represent the column provided this is specified in advance.)	3-36. Using the first method indicated in the space at the left, suppose that the pencil lands near Row 465, Column 5, as indicated by the dot in your table. Where would we start selecting our random digits for the ABC Store sample? a. 465-5 (Frame 3-37) b. try again (Frame 3-38) c. 431-5 (Frame 3-39)

a. 9,000

The sample would be drawn only from the 9,000 accounts that contain 3-month overdue balances. The remaining accounts are of no interest in this problem.

(or similar answer)

1-18. In this example, as in most estimates of a variable for auditing purposes, we are interested in obtaining a total and not an average value. In order to make such an estimate, the exact number of elements in the population must be known. We therefore define our population in the ABC Store problem to be:

a. all accounts that contain three-month-or-more overdue amounts

b. the overdue amounts themselves

No answer required

2-26. In the XYZ Shoe Company problem described in Exhibits 3 and 3-A, the first step is not as easy as it was in the invoice example. There is no automatic correspondence between the random digits and the Lot designations such as "LT-5A2W," so we have to establish this correspondence ourselves.

This has been done in Exhibit 3-A. Instead of thinking of each Lot in terms of the company's code designation, the auditor has simply ________________________________.

3-37. YOUR ANSWER: 465-5

No. The "blind stab" result is not the starting point; it only determines the point according to a pre-specified correspondence.

Of course, there is nothing to prevent the auditor from deciding to start directly where the pencil lands, as we did in the XYZ Shoe Company sample. However, the suggested method takes no extra time other than turning to the determined page, and has the advantage of adding an extra layer of randomization to the entire process.

Return to 3-36 and select another answer.

a. all accounts that contain three-month-or-more overdue amounts	1-19. We don't know how many amounts there are that are overdue three months or more. All we know is that each of the 9,000 accounts contains at least one such amount. By defining our population to be these 9,000 accounts, we make it feasible to take a sample of these accounts, add up the amounts that are overdue three months or more on each account, average this figure for the accounts in our sample, and multiply by ______ to obtain an estimate of the total.
numbered each lot from 1 to 300 (or similar answer)	2-27. We therefore have established a correspondence similar to that in the invoice example, but instead of working with two-digit numbers, we will work with ________-digit numbers. Thus, the first Lot listed, PJ-3A6N, will correspond to the random number ________.
	3-38. YOUR ANSWER: try again No, this is not necessary. The pencil will not often land directly and unambiguously on a line, and we were fortunate in this case that the first attempt landed near a usable line. Since the rows in our two-page table all begin with 4, it is clear that 465-5 (431542) is the nearest usable line. Return to 3-36 and choose again.

9,000 (as in Frame 1-12)

1-20. In frame 1-1 we said, "Statistical estimation consists of a variety of techniques for obtaining information about a certain body of data, based on inferences from a relatively small portion of the data."

We can now replace the phrase "certain body of data" with the more rigorous term "__________."

Also, instead of the phrase "relatively small portion of the data," we can substitute the single word "__________."

3-digit

001 (This is not the only answer possible, but is clearly the most convenient.)

2-28. The last Lot on the list, CJ-5A6E, could correspond to the random number 000, but that device isn't necessary this time since all elements in the population are designated by the same number of digits (three). Instead, then, we will ignore 000 if we come across it in the table, and have Lot CJ-5A6E correspond to the random number ________.

3-39. YOUR ANSWER: 431-5

Right. These are the first four digits in the nearest usable line, and we don't have to measure it to the nearest millimeter since in this example the other nearby lines do not begin with the digit 4. Thus, according to our pre-specified plan, the digits 431 correspond to Row 431 on the first page, and the digit 5 gives us the starting column. Enter this in Exhibit 6.

With the starting point, and route given in Exhibit 6, what is the first random number to be used? __________

population sample	1-21. Having defined the population and the information he wishes to obtain, the auditor usually begins by specifying these items in writing. This has been done in Exhibit 2. Before turning to this Exhibit, think out or jot down the words you would use.
300	2-29. We now know what to do with any random number between 001 and 300 inclusive. For example, if we draw the random number 272 we will select Lot __________ as part of our sample. (Look up the original code designation in Exhibit 3-A.)
0733 (last two digits of 431-5 and first two digits of 431-6)	3-40. The exercise suggested below is designed to increase your facility with the random number table. If you feel that you do not need further practice, you may skip the exercise and go right on to Frame 3-45. Using the sampling plan specified in Exhibit 6, draw the preliminary sample of 30 items. List the numbers on a separate piece of paper. Although you will have to inspect the table carefully, this exercise should not take more than a minute or two. When you are finished, check your results against the list in the first column of Exhibit 7. You should have the same numbers in the same order. (No answer required)

Look at Exhibit 2 if you have not already done so; then continue with Frame 1-22.	1-22. After specifying the population and the information to be obtained, we next specify the maximum amount of sampling error that will still leave us with a useful estimate. In estimating the value of the three-month overdue accounts, for instance, a certain amount of sampling error will be introduced because we (DO/DO NOT) intend to add up the three-month overdue amounts on all 9,000 accounts. Instead, we have decided to make an estimate on the basis of a sample. Is it reasonable to expect that we will hit the exact figure "on the nose?" (YES/NO) For most auditing purposes, do we need this degree of accuracy? (YES/NO)
HC-3I6S	2-30. What are we to do with random numbers 301 through 999? If we wished, we could have had 301 and 601 serve the same function as 001; 302 and 602 the same as 002;...and 600 and 900 the same as 300. In this system, Lot TK-1A2S would be represented in the random number table by the digits 118, 418, and ________.
No answer required	3-41. In Chapters 1 and 2 you have learned how to take an unrestricted random sample with replacement. What are the basic principles of an unrestricted random sample? ________________________________ ________________________________

DO NOT NO NO (although the auditor's judgment is the final determining factor in any specific instance)	1-23. There is, of course, one and only one actual dollar value of the amounts that are overdue three months or more. Let us assume that you have statistically estimated the value to be $453,300 while the actual value is $453,306. This would be an amazingly close estimate, but technically we would say that the sampling "error" in this case amounted to $6.00. Thus, a "sampling ____________ " is not actually a mistake but is simply the difference between the estimated value based on a sample and the ____________ value that can be determined exactly only be examining the entire population.
118, 418, and 718	2-31. This procedure would save a little time, since we could make use of all numbers between 001 and 900 inclusive. However, we shall instead keep things simple in this example and only use numbers 001 through 300, ignoring 000 and all other numbers above 300. In review, then, the first step in preparing to use the random number table is to establish ____________ between the elements of the population and the digits in the table.
every element in the population must have an equal chance of being selected in the sample and every combination of elements must have an equal chance of being selected. (or similar wording)	3-42. What is meant by "sampling with replacement"? ________________________________ ________________________________

error actual (true, real, etc.) (NOTE: In this book we are assuming that there are no non-sampling errors such as mistakes in computation.)	1-24. Since we ordinarily do not know the <u>actual</u> value, it is not possible to compute the sampling error in this manner. Nevertheless, while we cannot compute the sampling error for a particular estimate, we can compute the sampling error for a particular <u>procedure</u>. We use the term <u>precision</u> to measure the sampling error associated with a particular procedure. The desired precision is usually specified before any information is obtained. Thus, in the ABC Store example (Exhibit 1), the auditor says that he would like to obtain enough information so that his estimate will not deviate from the actual value by more than ______ in either direction.
correspondence	2-32. The correspondence scheme should be specified in writing before the sample is selected. Exhibit 4 shows how this is done. As indicated in this Exhibit, the next procedure to be specified in writing after "correspondence" is the ___________ to be taken through the table.
a random number is used every time it appears, and its corresponding element in the population is counted each time (or similar wording)	3-43. What are the three steps that are necessary before beginning to use the random number table? (Name them in order.) 1. ________________________________ 2. ________________________________ 3. ________________________________

$20,000 (The $\pm$ sign could be omitted from this answer since the words "in either direction" mean exactly that.)

1-25. Notice that the desired precision is not dependent on the value of the estimate - we want a precision of $20,000 regardless of the value of the estimate. This is the reason we say that precision depends upon the procedure rather than on the estimate.

This is true even though we often express the magnitude of the precision as a percentage of the estimate. Thus, if you calculate an estimate as $400,000 and the precision as $12,000, you could also report the precision as ______.

---

route

2-33. In the invoice example we took the last two digits in each line and moved down one line at a time. We could also take the first two digits, or the middle two digits, or start at the bottom and proceed upwards. More complex routes are possible, of course, but it is best to have as simple a route as possible, and to work with consecutive digits in the table. Would this include the possibility of using the last digit in one column and the first digit in the next?

a. YES (If this is your answer, turn to Frame 2-34)
b. NO (If this is your answer, turn to Frame 2-35)

---

1. establish correspondence
2. decide on a route
3. select a starting point randomly

3-44. What is the major purpose of the preliminary sample of at least 30 items?

____

____

3%	1-26. In the example we just worked with, the figures \$388,000 (\$400,000 - \$12,000) and \$412,000 (400,000 + 12,000) are called the lower and upper <u>precision limits</u>. Thus, for example, with an estimate of \$100,000 and precision <u>limits</u> of \$97,000 and \$103,000, the precision is (\$6,000/\$3,000) or ______%.
	2-34. YOUR ANSWER: YES, the last digit in one column and the first digit in the next could be considered "consecutive." Correct. The columnar divisions have no significance except as a visual aid. You can go across them as if the divisions did not exist. By the same token, there is no need to limit ourselves to consecutive numbers. This limitation is suggested only as a convenience. (Now skip to 2-36.)
estimate the variability of the population	3-45. In order to achieve as much randomization as possible, and to increase the number of usable digits in the table, one further procedure is suggested in establishing the sampling plan. Rather than specifying our digit places arbitrarily -- "first four, last two and first two of next column," etc., we can let the random number table do it for us. Reviewing for a moment, we let the first three digits determine the starting ____________, and the next digit corresponds to the starting ____________.

$3,000

3%

1-27. PRACTICE EXERCISE: PRECISION LIMITS

(The following exercise is optional for the reader who desires the practice. If you do not do the exercise, go to Frame 1-28.)

	ESTIMATE	PRECISION	PRECISION LIMITS
1.	$ 73,600	$2,500	$_____ and $_____
2.	$1,350,000	3%	$_____ and $_____
3.	$ 25,000	$____ or ____%	$24,000 and $26,000

2-35. YOUR ANSWER: NO, the last digit in one column and the first digit in the next are not "consecutive."

Wrong. The columnar division in this or any random number table has no significance except as a visual aid. If we were to limit ourselves to working only within columns, we would substantially reduce the number of usable digits in the table. By the same token, of course, we do not have to limit ourselves to consecutive numbers. This limitation is suggested only as a convenience. (Go on to the next frame.)

---

row

column

3-46. In selecting our ABC Store starting point, our "blind stab" landed near 465-5. The digits of that line are 431542.

We can make use of the fifth digit number by having it represent the <u>starting digit</u> in the row and column selected. The fifth number is <u>4</u>, so in our starting line (431-5), we start with the <u>4</u>th, 5th and 6th digit plus the 1st digit of the next column.

If we had used this method, the first account selected would have corresponded to the four-digit random number ______.

1. $71,100 $76,100

2. $1,309,500
   $1,390,500

3. $1,000 or 4%

1-28. By specifying the precision, we are saying that the procedure used will provide that the difference between the (unknown) actual value and our estimate will not exceed the value of the precision more than a stipulated number of times out of 100.

This is equivalent to stating that the precision interval - ranging from the lower ____________ limit to the upper precision limit - will include the actual value with the prescribed frequency.

2-36. Bearing in mind, then, that many other routes would be equally feasible, in the XYZ Shoe Inventory problem we shall simply take the first three digits in a line to be selected randomly, and then proceed downward one line at a time as we did before. The last three digits in each line will be (INSPECTED/IGNORED).

4073 (4th through 6th digits of 431-5, plus the first digit of the adjacent column)

3-47. Let us suppose that in another example we need four-digit numbers (0001 through 9999). We take a "blind stab" and land on line 441-2. We would therefore want to start selecting our sample in Row ______, Column ______. However, the next number is 9 and there is no "9th" digit in this or any column.

We therefore stab again.

precision	1-29. It is, of course, impossible to be certain that the precision interval includes the actual value. The best that we can do is state that this will occur with a given frequency. This means that were we to repeat the same sampling procedure many times, the fraction of precision intervals including the actual value would approximate the stipulated frequency. Thus, in our estimate of the total dollar value of the amounts overdue by three months or more, we want the difference between the actual value and our estimate to be no more than $20,000 with a frequency of ________%.
IGNORED	2-37. If we were to begin in Row 473, Column 7, we would start right off by selecting Lot number ______ as part of our sample. You can see for yourself that we would hit only 5 more usable random numbers and then come to the last usable number in the column, number _____ in Row _____.
Row: 467 Column: 2	3-48. This time we land in 408-0. Our starting point is Row _____, Column _____. With which digits do we begin? ____________________________

80%	1-30. You probably have an intuitive feeling as to what it means to be "80% certain" of something. It's the same as saying that the odds are 80 to 20 (4 to 1) in your favor, or in other words, 8 times out of 10 the precision interval (WILL/WILL NOT) contain the actual value.
146 (FM-1I6N) 125 in Row 498 (The intervening usable numbers are 055, 050, 085, 008, and 248.)	2-38. If we needed more digits to complete our sample, we could specify starting again in the next column. The first number to be used in Column 8 would be the number ______ in Row ______.
Row: 445 Column: 0 Digits: The fifth and sixth plus the next two of Column 1. (8564)	3-49. In this chapter we have seen how to select a preliminary sample using unrestricted random sampling. In the following chapters, we will see how this data is used. END OF CHAPTER 3

WILL	1-31. The percentage figure, such as 80%, is known as the reliability of the estimate. The reliability is not a guess or a mere hope, but is a mathematically determined amount that expresses the proportion of cases in which: a. the actual and estimated values will be the same b. the precision interval will include the actual value c. the actual value will not be somewhere within the stated precision limits
177 in Row 464	2-39. Proceeding in this manner, if we came to the end of Column 9 and still needed more digits, we would simply go on to the next page and begin at the top of Column 0. In this two-page illustrative table, we would consider the first page to be "next." In Exhibit 4, write in the route we have just described. Do not refer to any row or column numbers, since we have not yet selected our starting point. You may refer back to the preceding frames if necessary. When you specify the route, the directions should be clear and complete so that another person following the same route would select the same sample that you would.
CHAPTER 4	CHAPTER 4: SAMPLE MEAN AND STANDARD DEVIATION 4-1. While the preliminary sample is being selected, the sample data are entered on a worksheet such as Exhibit 7. Locate this Exhibit, which we will be referring to frequently in this chapter. (No answer required)

b. the precision interval will include the actual value.

1-32. The mathematical basis of the reliability will be discussed later. The following sequence gives a further explanation of the meaning of the term, as you might explain it to a colleague.

1. "I am going to make an estimate, with certain precision limits, and state flatly that the calculated precision interval contains the actual value. For example, if my estimate is $100,000 ± 2%, I state that the actual value is somewhere between $_______ and $_______."

ROUTE: Considering only the first three digits in each line, proceed downward until reaching the end of the starting column; then begin again at the top of the column to the right. (Column 0 on page 1 is considered to be the next column to the right of Column 9 on page 2.)

2-40. The third and final step in preparing to use a random number table is to select a starting point randomly. We can let the digits in the table correspond to row and column numbers. Thus, the digits 4387 could represent (in this two-page table) Row _______, Column _______.

No answer required

4-2. Notice the listing in the first column. The letter x in statistics (unlike its usage in algebra) does not stand for the unknown, but simply refers to the elements in the sample. Thus, x_1 is the first, x_2 is the second, and x_j is the $_j$th element when j stands for a whole number. This notation is illustrated in the first column of Exhibit 7.

The small letter n at the bottom-left always refers to the number of elements in the (SAMPLE/POPULATION).

$98,000 and $102,000	1-33. 2. "But, as we have discussed, I am making this estimate with these precision limits on the basis of computing the average value of: a. all the elements of the population b. a sample drawn from the population."
Row 438, Column 7	2-41. The digits that determine the starting point can be selected by closing your eyes and making a "blind stab" at the table. Since your pencil will rarely land conveniently on a line, you might specify that the starting point will be determined by "the first four digits in the nearest line," or "the last four digits in the line above," or any other determinant that is non-ambiguous and results in a __________ selection.
SAMPLE	4-3. The random numbers are listed in the next column. The order in which they are listed makes no difference, but it is usually most convenient to list them as they are being drawn. This makes it easy to double-check the selection against the numbers in the table. What do you think is the reason for listing the stopping point? ______________________________ ______________________________

b. a sample drawn from the population."	1-34. 3. "Now assume that I were to go through the entire estimating process once again. This time I select another random sample of accounts to examine. I go through the same kind of computation to estimate the total value, but since my sample is different, my estimate most probably (WILL/WILL NOT) be identical to my first estimate."
random	2-42. If the result of your "blind stab" is a set of unusable digits (i.e., digits that do not correspond to a row number in your table), you could go up or down the column until you hit on usable digits. However, the recommended procedure is simply to take another stab. (In Chapter 3 we will discuss a more sophisticated method of insuring a random start.) For our XYZ Shoe Inventory sample, then, let us suppose that after a few stabs we hit the digits 4931. Enter the starting point in the appropriate box in Exhibit 4.
This is only a preliminary sample. More elements will probably be added later, so we know where we have left off and where to start our additional selection. (The starting point had already been entered in Exhibit 6.)	4-4. The actual account numbers are not really needed, but are good to have as part of the record. Thus, as each account is located by the electronic equipment, we note its number and the (AVERAGE/TOTAL) of the amounts on it that are overdue three months or more.

WILL NOT	1-35. 4. "Let us suppose that my second estimate is $95,000 with precision limits of $93,100 and $96,900. Once again I make the same flat statement that the calculated precision interval contains the ______ value."
Row: 493 Column: 1	2-43. In review, preferably without looking back at Exhibit 4, what are the three major steps, in order, that are necessary in preparing to use the random number table? 1. ______________________ 2. ______________________ 3. ______________________
TOTAL (If you were correct, you may skip to 4-6 now.)	4-5. Once each account in the sample has been located, the auditor adds up the individually itemized amounts that are overdue three months or more. This is our x_j value for each account. Consider an account that has three-month-or-more overdue amounts of $70, $55, and $25. For this account, x_j = a. $150 b. $ 50

actual (true)	1-36. 5. "I do this 100 times, each time selecting a different sample. The chances are that each time I will arrive at a slightly different estimate with a slightly different set of precision limits. But each time, I make the same flat statement that the precision interval will ____________ ____________________________."
1. establish correspondence 2. determine a route 3. select a starting point	2-44. These three steps, and the other procedures discussed in connection with unrestricted random sampling, are all designed to guarantee the basic principle that every element in the population should have ________________________________ and that every sample combination should have ________________________________.
a. $150	4-6. Mark each statement True or False. ____ 1. The 4608th account in the file was the 10th element we happened to select in the sample. ____ 2. Account #026197 has $21 worth of three-month-or-more overdue amounts. ____ 3. X_{10} = $21

cover the actual value

(or similar wording)

1-37. At this point your colleague breaks in: "You wouldn't expect to be right 100 times out of 100, would you?"

Your best reply is: "No, not at all. In this example (Exhibit 1, paragraph 4) I would only expect to be right _____ times out of 100. In other words, we have specified _____% reliability."

---

an equal chance of being selected in the sample

an equal chance of being selected

(or similar wording)

2-45. In this chapter we have discussed the principles of random sample selection and the use of the random number table. You can apply this knowledge to auditing even if you do not use statistical estimation. For instance, assume that you are to select 35 out of 824 branch offices of a company for a surprise audit, and wish to give each of the 824 an equal chance of being chosen. Describe briefly how you would use the table to make a random selection.

All are true.

4-7. The fourth column lists the sample values. Even the most casual inspection of these values reveals that they are not distributed symmetrically like the heights of American men. Given the problem as described in Exhibit 1, what can we say about the 25th account in the sample?

a. $203 is an extreme value and should have been "weeded out" in advance

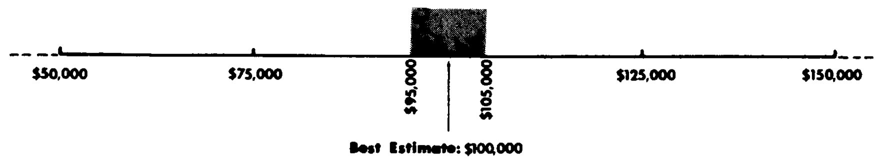
b. this is simply normal variability within the population

80 80%	1-38. SUMMARY: RELIABILITY 1. Reliability expresses the proportion of cases in which the precision interval will contain the actual value if a large number of random samples were drawn from the same population, and the same statistical estimating procedures were employed each time. 2. In common sense language, the reliability figure expresses the degree of confidence that the actual value will be somewhere within the precision limits.
Answer is in the frame at the right.	2-46. Each office could be assigned a number from 1 through 824. A route would then be selected that would take you up or down the columns, using any three consecutive digits specified in advance. The starting point could be chosen by the "blind stab correspondence" method. If these procedures are specified in advance, and written out in a format similar to Exhibit 4, you are left with a record of the selection process. END OF CHAPTER 2
b. this is simply normal variability within the population (If you were correct, you can skip to 4-9.)	4-8. It seems reasonable to expect that most of the amounts in the population will be relatively small, with a few being in the hundreds. Since we have a population of 9,000 accounts, there might be, say, three or four hundred with three-month overdues adding up to a few hundred dollars. We would therefore not be surprised to see one in the sample. (No answer required)

No answer required

1-39. Reliability has no meaning unless associated with a specified precision. In the sketch below, the auditor has made an estimate of $100,000 ± $5,000. If the reliability of this estimate is 90%, and he asserts that the (SHADED/UNSHADED) area contains the actual value, he is taking a _______% risk of being incorrect.

(NOTE: In this diagram and all others in this book, notice only the width of the shaded area. The height is irrelevant.)



CHAPTER 3

CHAPTER 3. SELECTING THE PRELIMINARY SAMPLE

3-1. Having discussed the principles of unrestricted random sampling, and seen in a few examples how the random number table helps achieve randomness, we can now select the samples for the accounts receivable and shoe inventory estimates. First, however, we will get an advance look at the kind of information -- and potential misinformation -- that can be obtained from analysis of a sample.

(No answer required)

No answer required

4-9. In the extreme right-hand column of the worksheet, the values are squared. This is important in a later calculation but can be ignored now.

At the bottom of the last two columns, the values are summed. Why are these sums listed as subtotals rather than totals?

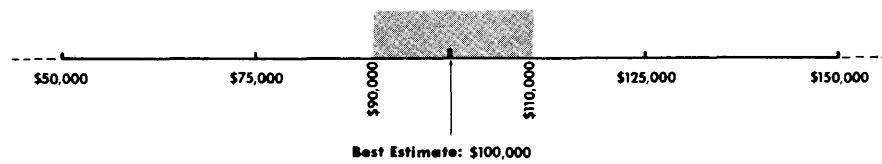
____

____

SHADED

10% since the risk is the complement of the reliability, that is (100-90)%.

1-40. Suppose that holding all other factors constant, the precision limits of this same estimate were widened, as in the diagram below. Now we are saying that the estimated value may differ from the actual value by as much as ± $______. This is the new (PRECISION/RELIABILITY) of the estimate.



No answer required

Invoice	Value
# 73	$ 11
43	10
38	11
04	9
61	18
47	11
14	10
93	14
13	10
84	6
	$110

3-2. Let us return to our earlier example in which we wanted to estimate the total value of 100 invoices, based on a random sample of 10 of them. (This is a simplified problem for illustrative purposes. In practice, populations as small as 100 and samples as small as 10 are rarely dealt with by sampling.)

Suppose that our random sample gives the results as shown in the box at the left. The average value of the sample invoices is $11, so the best estimate of the total value of all 100 invoices is $_____.

This is a preliminary sample. We will probably add more elements to it.

(or similar answer)

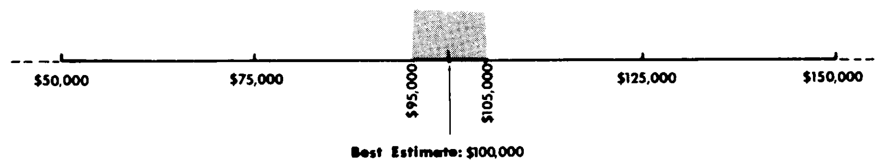
4-10. Given the sample values and the subtotal we can compute the "average" value, or more precisely, the "arithmetic mean." This is simply the sum of all the values divided by the number of elements. Thus, for the sample in Exhibit 7, the mean is ________.

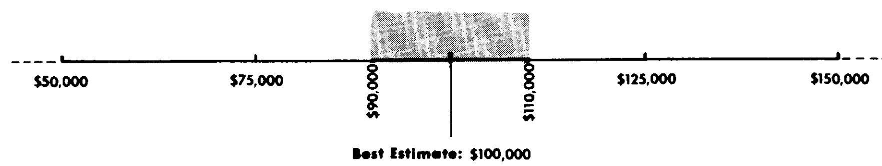
NOTE

In this book, when a mean is computed from whole-number data, it is shown to the nearest tenth.

When a mean is not being computed, but is simply referred to or mentioned in passing, it can be shown to the nearest whole number.

$10,000 PRECISION	1-41. In the bottom diagram, we have widened the precision interval. Holding all other factors constant, you would expect a wider interval has a (BETTER/WORSE) chance of containing the actual value.





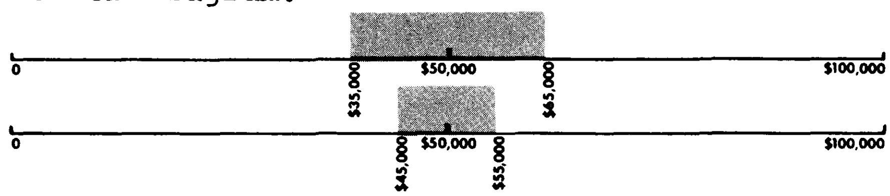
$1,100	3-3. Notice that the invoice values in the sample vary over a narrow range - the smallest value is 6 while the largest value is 18. If the population of 100 invoices showed a similar small variation, you would expect relatively (SMALL/LARGE) variation among the possible sample means. On the other hand, suppose that invoice #93 amounted to $204 instead of $14. The sample average would be $30 instead of $11 and the estimated population total would be raised by $1,900.
$48.0 ($1,440 ÷ 30)	4-11. In statistical notation, the sample mean is indicated by the symbol $\bar{x}$ (pronounced "x-bar"). The formula for its calculation, which we have simplified for the moment, is: $\bar{x} = \frac{\Sigma x_j}{n}$ What does "n" stand for? ______________________________

BETTER	1-42. At the same time, however, we are saying that the estimated value may differ from the actual value by as much as ± $10,000. In the first diagram we specified a precision of $5,000. Thus, in order to increase our reliability, we (GAINED/SACRIFICED) some accuracy in our estimate; or in other words, we (INCREASED/DECREASED) the acceptable precision limits or "margin of error."
SMALL	3-4. Your first reaction to finding such an isolated extreme value in the sample might be either to ignore it or choose another invoice to replace it. Your might seek to justify this by saying that the high value was rare in the population and hence its inclusion in the sample gives a false impression with respect to the actual total value in the population. Is this response consistent with using unrestricted random sampling? (YES/NO). If YES turn to frame 3-5, if NO to frame 3-6.
number of elements in the sample (<u>not</u> simply "number of elements")	4-12. $\bar{x} = \frac{\sum_{j=1}^{j=n} x_j}{n}$ The Greek letter Σ (sigma) means "the sum of the following elements." The notations below and above the Σ indicate that all sample elements are to be summed from the first (j=1) through the last (j=n). In this book you will not be required to write symbolic notation, but it will be necessary to recognize it. A reference glossary is included in this volume as Appendix 2. (No answer required)

SACRIFICED

INCREASED

1-43. Below are two hypothetical sets of precision limits for the same estimate based on the same sample. Without certain other information that we don't yet have, we cannot state the actual percentage figures for the reliability of these estimates. We can say, however, that in the top diagram the reliability -- that is, the probability of the precision limits containing the actual value -- is (GREATER/SMALLER) than in the bottom diagram.



3-5. Your "YES" response is incorrect. To understand why, recall the following two conditions required in unrestricted random sampling:

1. Each element in the population has an equal chance of being selected;
2. Each possible sample combination has an equal chance of being selected.

Either of the actions suggested violates both of these conditions, and hence the response is not consistent with the principles of unrestricted random sampling. Turn to 3-7.

No answer required

4-13. As we saw earlier in this book, the sample mean can be used as a basis for a preliminary estimate of the population total. For example, suppose that a random sample of 30 petty cash vouchers yields a sum of $240. The mean value for each voucher would be $_________.

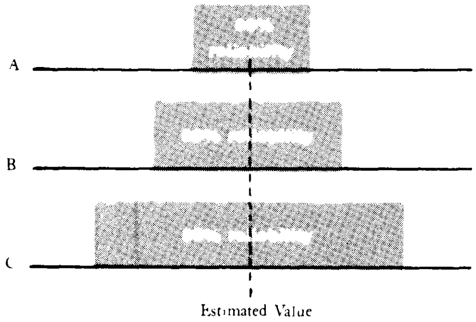
If the population includes 1,000 vouchers, our preliminary estimate of the total amount disbursed would be $_________.

GREATER	1-44. As these examples illustrate, the precision and reliability statements have no meaning unless they are paired with each other. Thus, it is of little value to state the estimate together with its precision unless you simultaneously state the reliability of this estimate. Similarly, the reliability (such as "95%") only applies to the probability of the precision interval covering the ______ value.
	3-6. In answering "NO," you are correct in saying that the response is not consistent with the principles of unrestricted random sampling that require that each element in the population have an equal chance of being selected and all possible samples have an equal chance of selection.
\$8 (\$240 ÷ 30) \$8,000 (\$8 x 1,000)	4-14. This calculation could be symbolically notated as follows: $\hat{X} = \bar{x}N$ Study only the right-hand side of the equation. The symbol $\bar{x}$ refers to the sample ___________. The capital letter N stands for the number of elements in the (SAMPLE/POPULATION).

actual	1-45. In the ABC Store example with which we are working (Exhibit 1), the desired reliability has been set at 80% and the desired precision at $20,000. Suppose we gave ourselves more leeway and specified a precision of $30,000. The reliability of the estimate would then be (GREATER/LESS) than 80%, because the wider precision limits would (INCREASE/DECREASE) the range of values within which the true value might fall.
	3-7. The problem of finding an isolated unusual item in the sample cannot be solved <u>after</u> the sample has been selected, but there are ways of avoiding it <u>beforehand</u>. If you know the population contains a few unusual items, these may be removed from the population prior to selecting the sample. The items removed are no longer part of the population and must be audited separately. (No answer required)
$\bar{x}$ = sample mean N = number of elements in population	4-15. $\hat{X} = \bar{x}N$ A capital X with no subscript always stands for the population total. A caret (^) over any symbol indicates an <u>estimated</u> value. In multiplying our sample mean by the number of elements in the population, we have (COMPUTED/ESTIMATED) X, the population total.

GREATER

INCREASE



1-46. Diagrams A, B and C at the left show that "the more confident we are, the less we have to be confident about." In each diagram, the shaded area represents the range of values that might include the true value with the indicated degree of confidence.

In which diagram are we trying to attain the closest estimate? ___ In which are we <u>least</u> sure that the true value lies within the stated limits? ___

No answer required

3-8. Any qualifications of this nature must be specified in the initial stage. We will assume that our shoe inventory and accounts receivable do not contain extreme-value items, so we will not have to change our original specifications in Exhibit 2.

The extreme values are evaluated separately as an auxiliary population, and we can then use unrestricted random sampling for the main population. However, the two populations cannot be routinely added. The techniques for estimates based on stratified populations are covered in Volume III in this series.

(No answer required)

ESTIMATED

4-16. We can apply the same procedure to our ABC Store problem. Given the sample mean of $48 for each account, we could estimate that the 9,000 accounts would contain a total of $_________ of amounts that are overdue three months or more.

A A	1-47. Of course, we cannot specify our precision and reliability and then automatically assume that our estimate will be accurate within the limits stated. We can make these positive statements only after performing certain computations to be covered later. We can, however, specify the <u>desired</u> precision and reliability beforehand. As we will see, this must be done in order to know how many elements have to be in the sample. For the record, then, enter in Exhibit 2 the desired precision and reliability for the three-month overdue amount estimate. (You may refer back to the statement of the problem in Exhibit 1.)
No answer required	3-9. In the XYZ Shoe Company audit, the auditor has already ascertained that the 300 lots range in value between approximately $400 and $600 dollars. If it turns out that one of these lots actually has a value of say, $679, no great distortion would result even if it were selected in the sample. If the auditor knows that there are also a few lots with values of over $1,000, perhaps because they are composed only of imported shoes, he would most probably (INCLUDE/EXCLUDE) these lots when he originally defines the population.
$432,000	4-17. Naturally, we immediately ask ourselves: "To what extent does this estimate differ from the actual population total?" There is no way to answer this question exactly without adding up the overdue amounts on all 9,000 accounts -- and that, of course, would defeat our purpose in taking the sample to begin with. On the other hand we can answer the question probabilistically. Instead of dealing directly with totals we ask the following equivalent question: "To what extent can we expect the sample <u>mean</u> to differ from the population <u>mean</u> if we were to sample repeatedly?" (No answer required)

DESIRED PRECISION: $20,000 DESIRED RELIABILITY: 80%	1-48. There is nothing special about the figures "$20,000" and "80%." These simply happened to be levels of precision and reliability desired by the auditor in our example. Indeed, in <u>any</u> auditing estimate, whether or not it is made with the aid of statistical sampling technique, the auditor always is faced with the potential question, "What is the probability that the estimate is accurate within such-and-such an amount?" In any given example, establishing the desired answer to this question is a matter for the (AUDITOR'S/STATISTICIAN'S) judgment.
EXCLUDE	3-10. Special cases of extreme values should not be confused with normal variability within the population. Consider, for example, that classic favorite of statisticians -- the heights of American men. The large majority will be within a few inches of the average height. A small percentage will differ from the average by, say, six inches, and the distribution gradually "tapers off" as we find only a few men over seven and under five feet in height. It stands to reason that if a random sample of men contained enough members, this pattern of variability in the population (WOULD/WOULD NOT) be reflected in the sample.
No answer required	4-18. The importance of this question can be made clear with an arithmetical example. Based on a sample mean of $48, we have initially estimated that the 9,000 accounts in the population contain a total of $432,000 in amounts that are overdue three months or more. If the population mean were actually, say, $60, then the actual population total would be $540,000. Thus, a difference between the sample mean ($\bar{x}$) and the true population mean ($\bar{X}$) of only $12 results in an estimating error of $12 x N or $________.

AUDITOR'S

1-49. We are now ready to begin answering our two basic questions pertaining to the selection of the sample: First, how do we select the sample? Second, how do we determine the appropriate sample size?

The first question will be the subject of the next Chapter.

END OF CHAPTER 1

WOULD

Invoice	Value
# 73	$ 11
43	10
38	11
04	9
61	18
47	11
14	10
93	14
13	10
84	6
	$110

3-11. By the same token, then, a random sample enables us to estimate the variability of the population. For example, look again at the invoice sample in the answer space at the left. Although ten elements is not a large enough number for a definitive statement, we would have reason to believe that the population is probably:

a. regularly distributed, with most values fairly close to the average and the rest tapering off to the extremes

b. irregularly distributed, with no particular pattern and no concentration around the average

---

$108,000 (N = 9,000)

4-19. An error of this magnitude would, of course, greatly exceed our desired precision of $20,000.

If the true population mean were $45, the population total would be $__________. Our estimate of $432,000 would be in error by $__________. Would this be acceptable? (YES/NO)

CHAPTER 2

CHAPTER 2. BASIC RANDOM SAMPLING PROCEDURE

2-1. In this Chapter we shall discuss using a random number table for selecting a sample. We shall also review the major concepts covered up to this point.

Read Exhibit 3, "XYZ SHOE COMPANY." Then go on to Frame 2-2.

(No answer required)

---

a. regularly distributed, with most values fairly close to the average and the rest tapering off to the extremes

$ 10	$ 11	
6	10	
15	11	(First
7	9	Sample)
18	18	
16	11	
8	10	
7	14	
17	10	
6	6	
$110	$110	

3-12. Suppose, however, that we had drawn a sample whose values were as shown at the left. The range is the same ($6 to $18), and the total and average are the same ($110 and $11). However, in comparison with the preceding example, this sample would lead us to estimate a (GREATER/SMALLER) variability within the population.

$405,000

$27,000

4-20. We have just seen that in this example, a difference of only $3 between the true population mean and the computed sample mean would be too big a difference if our desired precision is $20,000. How big a difference can we tolerate? This is easy to compute if we postpone the question of reliability, which will be discussed in the next chapter. If we wish to limit the difference between the estimated total and the actual total to no more than ± $20,000, then the difference between the sample mean and the actual population mean must be no more than ± $__________ (N = 9,000).

No answer required	2-2. As we discussed in Chapter 1, the first major task in statistical estimation is to specify the sampling objectives. As an exercise in doing this, turn to Exhibit 2, and, on the basis of the information you have just read, fill in all four rows in the XYZ SHOE COMPANY column. You may refer at any time to Exhibit 3 if you need help in recalling the facts of the problem. When you are finished with Exhibit 2, go on to Frame 2-3.
GREATER (There is clearly a wider "spread," with few of the values near the average.)	3-13. Variability is usually defined and measured by calculating a quantity known as the <u>standard</u> <u>deviation</u>. This will be covered in detail in later chapters. For the moment, we can be content with the observation that other things being equal, the greater the variability, the (EASIER/HARDER) it is to make an estimate close to the actual value.
± \$2.22 (\$20,000 ÷ N)	4-21. What, then, are the factors that influence the difference between the sample mean and the population mean? Pure <u>chance</u> is one. If we had selected a different route or starting point in the random number table, we would probably not have selected these thirty accounts. The chances are that the values in the fourth column of Exhibit 7 would have been different and would have added up to (THE IDENTICAL/A DIFFERENT) subtotal.

The second column in Exhibit 2 should read as follows: POPULATION: The 300 lots that comprise the shoe inventory in the warehouse of the XYZ Shoe Company INFORMATION TO BE OBTAINED: Total value at cost DESIRED PRECISION: $2,000 DESIRED RELIABILITY: 90% (Make all necessary corrections in Exhibit 2 before going on.)	2-3. The following sequence, from 2-3 through 2-11, reviews the concepts of precision and reliability. This review is strongly recommended; however, if your prior background gives you enough confidence in dealing with these concepts, you can skip to Frame 2-12 on page 5. For purposes of review, let us "jump the gun" and suppose that you have already selected a sample of the required size, and estimated the average value of each lot to be $500. What would be your estimate of the total value of the population? __________ With a precision of $2,000, what would be the range of values that would be expected to include the true value? $______ to $______
HARDER	3-14. From the foregoing examples, we can see that a few obviously extreme and unusual values in the sample would affect not only our estimate of the total value, but also of the variability (standard deviation) of the population. We have seen that a way to prevent this from occurring is to eliminate such unusual cases from the (SAMPLE/POPULATION).
A DIFFERENT	4-22. We cannot control the chance factor of what elements are included in the sample. Indeed, our sampling plan using the random number table is designed to prevent just that. We can, however, lessen the effects of chance, and make the probable error smaller, by (INCREASING/DECREASING) the size of our sample.

$150,000 ($500 x 300)

$148,000 to $152,000

2-4. If you were to take another sample from the same population, would your estimate and precision limits probably turn out to be exactly the same as those based on the first sample? (YES/NO)

POPULATION (once the population has been rigorously defined, elements cannot be arbitrarily excluded from a sample)

3-15. We can, however, compensate for the inevitable variability within a population by making our sample size large enough both to reflect this variability and to "average it out." It follows from this discussion that if we do not know the standard deviation of the population, the exact sample size necessary to provide the desired accuracy (CAN/CANNOT) be determined.

INCREASING

A	B
$ 11	$ 10
10	6
11	15
9	7
18	18
11	16
10	8
14	7
10	17
6	6
$110	$110

4-23. Another important factor that influences the difference between the sample mean and the population mean is (as we have mentioned previously) the variability of the population. At the left are our two groups of invoices once again. Which group seems to have <u>less</u> variability? (A/B)

On what do you base your answer? ____________

____

NO

2-5. Suppose you were to take a large number of samples. If the reliability of your estimate were 90%, approximately how many times out of every ten would you expect the actual value to be exactly the same as your estimated value?

a. 0 (If this is your answer, turn to Frame 2-6.)

b. 9 (If this is your answer, turn to Frame 2-7.)

c. cannot be determined (If this is your answer, turn to Frame 2-8.)

CANNOT

3-16. We begin virtually any statistical estimation by selecting a preliminary random sample to estimate the (VARIABILITY/TOTAL VALUE) of the population. In general, the larger the preliminary sample the better - but since very small samples can yield poor estimates, you should have at least 30.

A (explanation is at the right)

A	B
$ 11	$ 10
10	6
11	15
9	7
18	18
11	16
10	8
14	7
10	17
6	6
$110	$110

4-24. In group A, the values tend to be concentrated around the mean, while in group B they tend more to the extremes. This concept is easily grasped if we pretend that the two digits represent points scored in successive games by two basketball players. Their high and low scores are the same, and their average is the same, but B is less predictable, more erratic, or in other words, more ___________ in his scoring than A.

<table>
<tr><td></td><td>2-6. YOUR ANSWER: 0

Correct! This question contained a trap. With a reliability of 90% you would expect that 9 times out of 10 the precision interval would contain the actual value. However, the odds against your hitting the actual value exactly "on the nose" are enormous.

Now skip directly to Frame 2-9.</td></tr>
<tr><td>VARIABILITY (We need to know this <u>before</u> we can say anything about the accuracy of our total value estimate.)</td><td>3-17. In review, an unrestricted random sampling plan gives every element in the population an equal chance of being selected as well as insures that each possible combination of elements is equally likely. If the auditor has reason to believe that there are some extreme elements that may result in a distorted estimate, these are excluded at the time the population is being defined. However, normal and inevitable variability within a population is compensated for simply by making the sample size large enough according to a formula to be presented later in this volume.

(No answer required)</td></tr>
<tr><td><u>variable</u></td><td>4-25. Since B's scoring is less predictable than A's, a scout watching this team might report on A (EARLIER/LATER) than on B. Considering each game as a sample element, the scout might well delay his report on B until he has had an opportunity to watch more games -- or in statistical terms, to increase ____________ ________________________.</td></tr>
</table>

	2-7. YOUR ANSWER: 9 It is true that with 90% reliability, you could expect that 9 times out of 10 the precision interval would contain the actual value. This answer shows that you understand the important concept taught in Chapter 1, but also shows that you did not read the question in 2-5 carefully enough. Return to Frame 2-5 and select a different answer.
No answer required	3-18. We are now ready to begin selecting the sample of the XYZ shoe inventory lots. Since we do not know the standard deviation of the population, we cannot tell at the outset exactly how large a sample size is necessary to give us the required information at the desired levels of precision and reliability. In order to estimate the population variability, we begin by selecting a preliminary random sample containing at least ___________ elements.
EARLIER the sample size (or similar wording)	4-26. We have seen that there is a relationship between the variability of the population (as estimated from a sample) and the <u>size</u> of the sample required to make an estimate at a given level of accuracy. What we now need is a <u>mathematical measure</u> of variability. We shall demonstrate how this is calculated in the concluding portion of this chapter, but first we shall review the concepts and notation presented thus far. (No answer required)

	2-8. YOUR ANSWER: cannot be determined This is not the best answer. To be sure, you cannot determine the exact number of times you will be successful unless you actually try it. However, based on the contents of the previous chapter, you can still have a reasonable expectation of one of the other two answers. Return to Frame 2-5 and try again.
30 (more would be better, but 30 will be used here for teaching purposes)	3-19. According to the starting point established in the preceding chapter for the shoe inventory sample selection (refer to Exhibit 4 if necessary), what is our first random number? ___________ Place an asterisk next to the appropriate Lot designation in Exhibit 3-A to indicate that it has been selected in the sample.
No answer required	4-27. We began by showing a convenient format for tabulating the sample data (Exhibit 7). In statistical notation, x_j refers to ________ ______________________________________.

2-9. As discussed in Chapter 1, hitting the actual value "on the nose" is not a realistic goal of statistical estimation. Whenever you make an estimate based on a sample (in contrast to inspecting every item in the population), there will almost always be some difference between the estimated value and the actual value. As long as this difference is equal to or smaller than the desired precision, the estimate can be considered satisfactory. Thus, if the desired precision is $3,000 and the difference is $2,500, the estimate (IS/IS NOT) satisfactory.

184

*RL-5Q6N

3-20. Going down the column as established in our sampling plan, the next number to be encountered in our route is 546. What do we do with it? ______________________________

(the value of) the jth element in the sample

4-28. We then computed the sample mean for the accounts receivable sample. The sample mean is indicated by the symbol _______.

IS	NOW TURN BACK TO PAGE 3 AND BEGIN THE SECOND ROW.
nothing (ignore it)	NOW TURN BACK TO PAGE 3 AND BEGIN THE THIRD ROW.
$\bar{x}$	NOW TURN THE PAGE AND FLIP THE BOOK OVER SO THAT PAGE 61 IS RIGHT SIDE UP. CONTINUE WITH CHAPTER 4 EXACTLY AS BEFORE, BEGINNING WITH FRAME 4-29.

GO ON TO FRAME 4-29.

4-29. Similarly, the symbol $\overline{X}$ (capital X-bar) stands for the population mean. In general, what is the distinction between small and capital letters?

____

____

5-26. Information of this kind is extremely important because it enables us to assign a degree of confidence to an estimate based on a sample mean. For example, suppose that in a given case the estimated standard error of the mean is computed to be $10. What is the probability that $|\overline{x} - \overline{X}|$ will not be more than $15? In this example, $15 is equal to ________ times the standard error of the mean. According to Table 2, a sample mean will differ from the population mean by no more than 1.50 times the standard error of the mean in _____% of all possible cases.

6-9. If we wish to solve for n directly, rather than $\sqrt{n}$, the formula becomes:

$$\text{(3a)} \quad n = \frac{S_{X_j}{}^2 \cdot U_R{}^2 \cdot N^2}{A^2}$$

This involves squaring four values rather than one, thereby increasing the work involved and the chance of arithmetical error. It is therefore suggested that you solve for $\sqrt{n}$ as shown in the worksheet, and round n up to the next whole number.

(No answer required)

small letters used for sample values; capitals used for population values	4-30. $$\bar{x} = \frac{\sum_{j=1}^{j=n} x_j}{n}$$ Above is the formula for the mean. What is meant by each of the following symbols? 1. $\sum$ ______________________ 2. j=1 and j=n ______________________ ______________________ 3. n ______________________
1.50 86%	5-27. Here's how the need for this kind of information may have arisen: Suppose an auditor wishes to estimate the total value of $7,500. An estimate based on a sample mean will be within these precision limits mean will be within these precision limits provided that the difference between the sample mean and the true population mean is not more than ________.
No answer required	6-10. Most statistical estimates for auditing purposes involve the estimation of a <u>total</u> rather than a <u>mean</u> value. However, the auditor can just as easily state a desired precision for the mean as for the total. For instance, in the Shoe Inventory problem, if we wish to report on the estimated mean value of the inventory lots, we would set our desired precision not at $2,000, but at this amount divided by the number of lots (300). The sample size formula would then be: (3b) $$\sqrt{n} = \frac{S_{X_j} \cdot U_R}{A}$$ Compare this with equation (3) in Frame 6-8. The sample size (WOULD/WOULD NOT) be the same.

1. $\sum$ means "the sum of the following"

2. j=1 and j=n means "include in the summation all elements from the first through the nth"

3. n = number of elements in the sample

4-31. $\hat{X} = \bar{x}N$

Explain the meaning of the above equation in your own words:

____

____

____

$15

5-28. What is the probability that a sample mean will not differ from a population mean by more than $15? We don't know, for $15 could represent an extremely large or a very small relative difference. We can answer the question, however, if we can change the dollar amount into a number times the ______________________________.

WOULD (In equation 3b, the original A has been divided by N, and the latter has been taken out of the numerator.)

6-11. There will be no confusion about the "A" term in the equation if the desired precision and the information to be obtained about the population are specified clearly, preferably in writing, as shown in Chapter 1 of this book (see Exhibit 2 for examples).

(No answer required)

END OF OPTIONAL SECTION

An estimate (^) of the population total (X) is obtained by multiplying the sample mean ($\bar{x}$) by the number of elements in the population (N). (or similar wording)	4-32. Let us illustrate this with one more numerical example. If a preliminary sample of 30 inventory tickets adds up to $27,600, then the mean value is $_________ and the preliminary estimate of the total inventory value of all 500 tickets in the population is _________.
standard error of the mean (This example was worked out back in Frame 5-26.)	5-29. Let us see how this kind of information is applied in another example. A random sample of 36 invoices drawn from a population of 1,000 results in a mean value of $20 per invoice. Our preliminary estimate is therefore $\hat{X}$ = $20,000. If our desired precision is $3,000, then the sample mean must not differ from the population mean by more than $_______.
No answer required	6-12. The Sample Size Worksheet, like the others, is simply a helpful aid to working out the formula. For the ABC Store example, we have begun by entering information that has been given or derived in preceding chapters. When we do the computations for Column 4 and then Column 6, what will we have? a. an arbitrary constant b. the standard error of the mean c. the numerator in the right-hand term of the formula

$\bar{x} = \$920$

$\hat{X} = \$460,000$

4-33. In this example, with N equal to 500, if the actual population mean were \$905, our estimate of the total would be off by \$______. This would not upset us if our desired __________ were equal to or greater than this amount.

\$3.00

5-30. The estimated standard deviation of the population is \$9.00, based on a calculation of the sample standard deviation. With a sample size of 36, how much is the estimated standard error of the mean? _______

---

c. the numerator in the right-hand term of the formula

6-13. Do the computations required in Columns 4 and 6. You could undoubtedly complete the rest of the formula at this point, but it is suggested that you do it step by step since we will discuss each step separately.

$7,500 ($15 x 500) precision	4-34. The amount by which the sample mean ($\bar{x}$) differs from the actual population mean ($\bar{X}$) is to some extent a matter of chance. Explain why. ________________________________ ________________________________
$1.50	5-31. Combining the information in the two previous exercises, our estimate based on the sample mean of $20 will be satisfactory; that is, will have a precision of $3,000, provided that the sample mean does not differ from the population mean by more than ______ times the standard error of the mean.
4: $56.32 6: $506,880	6-14. The amount in Column 6 is divided by the desired precision. This means that the smaller the precision figure, the larger the sample size. Explain why this should be the case: ________________________________ ________________________________

Sample selection is governed purely by the random numbers. A different sample will usually result in a different mean. (or similar answer)	4-35. In addition to chance, what are two factors that influence the difference between $\bar{x}$ and $\bar{X}$? 1. ____________________ 2. ____________________
2.00 (\$3.00 ÷ \$1.50)	5-32. Judging from previous examples, and without looking at the table, what do you think the chances are of this estimate falling within its precision limits? a. poor b. fair c. good EXPLAIN: ____________________ ____________________
a smaller (narrower) precision means that we are trying to come closer to the true value; therefore, we need to sample more elements in order to obtain more information (or similar answer)	6-15. Do the division in Column 8; then square the result to get the desired sample size in Column 9. Note the suggested procedures for rounding off. (Do not do Column 10 at this time.)

/EITHER ORDER/ sample size variability of the population	4-36. Thus, if we can mathematically measure the variability of a population, we are then in a position to predict the probable range of difference between the sample mean and the true population mean, and to compute the required sample size. Such a mathematical measure of variability exists, and is known as the <u>standard</u> <u>deviation</u>. (No answer required)
c. good Judging from previous examples, it seems that a large proportion of random samples fall within ± 2.00 standard errors of the mean.	5-33. Looking up the R value for 2.00 in Table 2, we get 95%. Explain in your own words, either in general terms or in terms of this particular example beginning in 5-29, what this means. ______________________________ ______________________________ ______________________________ ______________________________ ______________________________ ______________________________
7: \$20,000 $\sqrt{n}$: 25.3 n: 640.09 → 641	6-16. This may seem like a large number of elements, but with N = 9,000 the sample is only slightly more than 7% of the population. In another case, we might have to sample, let us say, 15% of the population; and in still another only one or two percent or even less might suffice. What characteristic of a population could influence this (other than size)? ______________________________

	A	B
X_1	1	1
X_2	2	1
X_3	4	3
X_4	5	3
X_5	6	6
X_6	6	8
X_7	7	9
X_8	9	9
ΣX_j =	40	40
$\bar{X}$ =	5.0	5.0

4-37. Consider A and B at the left to be total populations rather than samples. Note that for this reason the elements are indicated by capital X_j.

Based on the same kind of inspection as in the preceding similar examples, which population seems to have greater variability? (A/B)

Answer is at the right. If you have any one of these three points in sufficient detail, your understanding can be considered good enough.

5-34.

1. Approximately 95% of all means of a given size sample (in this case 36) will not differ from the true mean by more than 2.00 times the standard error of the mean.

2. The estimated standard error of the mean was computed to be $1.50; therefore, there is a 95% chance that the sample mean does not differ from the true population mean by more than ± $3.00.

3. There is the same 95% chance that the estimate of the total does not differ by more than $3.00 times N or $3,000.

This third point is a statement of the type discussed in Chapter 1. The figure 95% is the ______________ of this estimate.

---

standard deviation (variability)

o

NOTE: Precision and reliability are characteristics of an estimate, not of the population itself.

6-17. As you can see from the filled-in worksheet, the sample size of 641 is based on the assumption that the standard deviation of the population is $44.0. Now suppose that after selecting the additional sample elements, we discover that the estimated standard deviation of the population is actually greater than $44.0. In that case, our sample size of 641 would be (SUFFICIENT/INSUFFICIENT) at the desired levels of precision and reliability.

EXPLAIN: ______________________________

____

B (We will verify this mathematically in the following frames.)	4-38. As an overview, the standard deviation of a population is computed as follows: 1. The arithmetic mean is computed. 2. For each item, the difference from the arithmetic mean is computed. 3. The differences from the arithmetic mean are individually squared. 4. The squared differences are added and then divided by the number of items. This result is defined as the <u>variance</u>. 5. The square root of the variance is then extracted. This last result is the <u>standard deviation</u>. (No answer required)
reliability	5-35. We will now go through the same procedure to see if our preliminary estimate of the ABC Store inventory has the required reliability with the desired precision of $20,000. The purpose of this exercise is to develop a "feel" for what sample results mean. In practice, a preliminary sample will rarely be large enough to obtain an estimate of the total value at the desired precision and reliability. However, the purpose of the preliminary sample is to estimate ______________________.
INSUFFICIENT greater variability leads to greater chance of sampling error (other things being equal)	6-18. As a precaution against this, the practice sometimes employed is to add some additional items to the n computed by the formula. We will follow this practice in this volume by adding an additional 10%. Fill in Column 10 of Worksheet 4.

No answer required

	(1) X_j	(2) $X_j-\bar{X}$	(3) $(X_j-\bar{X})^2$
X_1	1	-4	16
X_2	2	-3	9
X_3	4	-1	1
X_4	5	0	0
X_5	6	1	1
X_6	6	1	1
X_7	7	2	4
X_8	9	4	16

$\bar{X} = 5.0$ $\quad \Sigma(X_j-\bar{X})^2 = 48$

4-39. In the space at the left, this procedure is followed to compute the variability, or more exactly the standard deviation, of our population A. In column 2, each X_j value is subtracted from the mean of 5.0. The result is squared in column 3. The population standard deviation is computed by dividing the sum of the squared deviations, or _______ in this example, by the number of elements in the population, _______, and then taking the square root of the quotient.

---

the population standard deviation (S_{X_j})

5-36. Now locate Worksheet #2 and keep it accessible. This Worksheet indicates the logic and the computations that result in determining the reliability of an estimate.

With N equal to 9,000 and the desired precision established at $20,000, the first two columns can be filled in. Do this in the "ABC Store (preliminary)" row.

NOTE: The capital letter A is used in this volume to stand for the precision of the total estimate.

706

641 + 64.1 = 705.1 ⟶ 706

6-19. Summarizing thus far,

1. The sample size formula is derived from the concepts discussed in Chapter 5. The derivation is shown in Frames 6-4 through 6-11.
2. Worksheet #4 in the Supplementary Section is an aid to working out the formula.
3. After solving for the square root of n, we square the result and then add 10% as a precaution against the possibility that the population standard deviation is higher than estimated.

(No answer required)

48 8	4-40. 1. $48 \div 8 = 6.0$ 2. $\sqrt{6.0} = 2.45$ The figure 2.45 is the ________ ________ of population A.
A: \$20,000 maximum $\lvert \bar{x} - \bar{X} \rvert$: \$2.22	5-37. From previous worksheets or from memory, you know that the estimated standard deviation of the population is \$44.0 and the sample size is 30. Compute and enter in the worksheet the estimated standard error of the mean. (Use 5.5 as the square root of 30.)
No answer required	6-20. For further practice in working with this important formula, determine the required sample size for population "Q" which consists of 5,000 elements with an estimated standard deviation of \$30. The estimate is to have a precision of \$12,000 at a reliability level of 75%.

standard deviation

	(1) X_j	(2) $X_j-\bar{X}$	(3) $(X_j-\bar{X})^2$
X_1	1	-4	16
X_2	1	-4	16
X_3	3	-2	4
X_4	3	-2	4
X_5	6	1	1
X_6	8	3	9
X_7	9	4	16
X_8	9	4	16

$\bar{X} = 5.0$ $\Sigma(X_j-\bar{X})^2 = 82$

4-41. The same computations have been made for population B. The computed population standard deviation, symbolized by the Greek letter σ (small sigma), is 3.20 as compared with 2.45 for population A. (The final computations are shown in the box below.) Thus, sample means from population B will, in general, result in (SMALLER/WIDER) deviations from the true mean than in population A.

$$\sigma X_j^2 = 82/8 = 10.25$$

$$\sigma X_j = \sqrt{10.25} = 3.20$$

$\hat{\sigma}_{\bar{x}} = \$8.0$

5-38. In order for our estimate to be within the desired precision limits, the sample mean in this example must differ from the true population mean by no more than:

a. a fraction of the standard error of the mean

b. the standard error of the mean

(Study Columns 3 and 4 of the worksheet if the answer is not apparent.)

R: 75%

U_R: 1.15

S_{X_j}: $30

#4: $34.50

N: 5,000

#6: $172,500

A: $12,000

$\sqrt{n}$: 14.4

n: 208 (not 207)

#10: 229

6-21. Once the required sample size has been determined, the remainder of the estimating process consists essentially of a repetition of procedures that we have already discussed. Concluding our ABC Store problem, then, the next step is to draw the 676 additional random numbers necessary to complete our sample of 706 items. Naturally we need not go through this entire procedure for teaching purposes. Instead, simply find your stopping point in Exhibit 7 and your route in Exhibit 6, and continue as before. What are the next three numbers drawn?

____

____

____

WIDER	4-42. In actual practice we never compute the standard deviation of the population, for in order to do so we first have to find the population total and mean. This is exactly what we are trying to avoid doing when we decide to use statistical estimation procedures in the first place. We therefore have to estimate the population standard deviation. In the absence of any other information about the population, the only way to do this is to calculate the estimated standard deviation of the population using the values of our ___________.
a. a fraction of the standard error of the mean	5-39. With the preceding answer as a guide, compute the U value for this example as directed in Column 4.
8044 0800 3069	6-22. The punched card equipment then locates the 8044th account in this file as well as the 800th, 3069th, and all the others in the newly-selected sample. In each account, the amounts that are overdue three months or more are totaled. The total of these amounts for each account is the x_j value, listed and squared as in Exhibit 7. There is obviously no difference in procedure except that this time we are dealing with a sample of 706 instead of 30 elements. (No answer required)

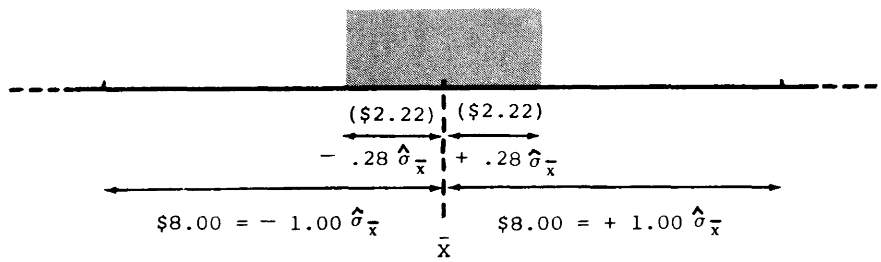
sample

4-43. We could calculate the estimated population standard deviation in a similar manner as before; that is, by subtracting each value from the sample mean, squaring the difference. summing the squares, and so on. Instead, however, we use the "short-cut" computational formula given in Worksheet 1. Locate this Worksheet and keep it accessible.

(No answer required)

.28 ($2.22 ÷ $8.00)

5-40. Explain why the true population mean must be somewhere in the shaded area in order for the estimate to achieve the desired precision.



No answer required

6-23. The preliminary sample of 30 and the additional sample of 676 can be considered as a single sample of 706 items. The next step, then, is to calculate the mean and standard deviation of this combined sample. As before, then, we list the sum of the amounts that are overdue three months or more on each of the accounts. What other numerical data do we list?

____

No answer required	4-44. The formula has been broken down into its individual components in the Standard Deviation Worksheet. For the ABC Store problem, columns A through C have been filled in for you. This information is obtained from the sampling worksheet (Exhibit 7). Compute and fill in columns D, E, F, and G.
Difference between $\bar{x}$ and $\bar{X}$ is magnified by N (9,000) when estimating the total. A difference of more than $\pm$ .28 $\hat{\sigma}_{\bar{x}}$, which in this example would equal \$2.22, would result in an error of more than \$20,000 in the total estimate.	5-41. What is the probability that the shaded area does contain the true population mean? Look up the answer in Table 2 and enter it in Column 5 of the worksheet. (Let .28 = .3) What is the reliability of the estimate based on this sample of 30 items at a precision of \$20,000? __________
the square of each of the values (as in Exhibit 7)	6-24. Once again a worksheet has been provided to facilitate tabulation and computation. Locate Worksheet 5 and keep it easily accessible. In the ABC Store column, we have entered the data that is already known. In addition, we are assuming that Σx_j and Σx_j^2 are \$33,860 and \$3,049,824 respectively and have entered these figures in the appropriate places. Compute the new sample mean and enter it in the worksheet.

D: \$48.0 E: \$2,304 F: \$69,120 G: \$56,056	4-45. Columns H and I give the estimated variance and standard deviation, respectively. We have worked with these concepts previously; but note the following differences due to the fact that we are working with a sample rather than a population. 1. The estimated standard deviation is symbolized not by σ but by the letter S to indicate the estimated nature of this quantity. 2. The estimated variance is computed by dividing the quantity in column G not by n, but by ______, which in this example is ______.
23%	5-42. This is a far lower reliability than we desire, for it means that the odds are more than 3 to 1 (77 to 23) that the true value (IS/IS NOT) within $\pm$ \$20,000 of the estimated value according to the results we have obtained.
$\bar{x}$ = \$50.0 (\$35,300 ÷ 706)	6-25. At this point we could make a new estimate of the total population value. However, instead we will first get all our data, and then analyze the results. Rows 8 through 11 involve the same work as in the Standard Deviation Worksheet with which you are already familiar. We have therefore done all computations except the last two, which are important enough to merit further exercise. Compute and enter in Worksheet #5 the variance and standard deviation of the combined sample.

n-1 29	4-46. Now compute the variance and standard deviation (columns H and I) for the ABC Store. Calculate the variance to two decimal places; the standard deviation to one. A square root table is included for your reference in the Supplementary Section (page S-40).
IS NOT	5-43. What was the estimated total value of the accounts that were overdue three months or more? (If you do not remember, reconstruct it from the appropriate data.)
$s_{x_j}^2 = \$2,000$ $s_{x_j} = \$44.7$	6-26. Most of the figures in this worksheet are obtained from computation from known data. Which two items are obtained from the new sample? ____________________ ____________________ All of the computations are directed at finding which two items in the worksheet? ____________________ ____________________

$1,932.97

$44.0

4-47. If you had any difficulty in the preceding examples, or desire greater facility in computing the standard deviation, you may wish as an optional exercise to fill in the worksheet for Sample "Q" below. Otherwise, go right on.

Sample "Q"

	x_j	x_j^2
x_1	3	9
x_2	5	25
x_3	8	64
x_4	10	100
x_5	14	196
	40	394

$432,000 ($\bar{x}$ = $48.0; N = 9,000)

5-44. It may actually be the case, of course, that the true value is within ± $20,000 of this estimated figure. Further information may reveal that this might be highly probable. But based on the information we now have, we can only be 23% confident that this is the case. That is, following the same procedure repeatedly, only 23% of the calculated precision intervals would contain the true value.

The next question is, what precision statement can we make with 80% confidence? Locate Worksheet 3, which will help us with this question.

(No answer required)

---

1. Σx_j (additional)

   $\Sigma (x_j)^2$ (additional)

   (Rows 2 and 7)

2. $\bar{x}$ (Row 5)

   S_{X_j} (Row 13)

6-27. The first use to which we will put some of this data is to solve the basic equation $\hat{X} = \bar{x}N$. Explain verbally the meaning of this notation and compute $\hat{X}$.

____

____

A	B	C
40	394	5
D	**E**	**F**
8.0	64	320
G	**H**	**I**
74	18.50	4.3

4-48. We have estimated the standard deviation of the ABC Store sample to be $44.0. This quantity, S_{X_j}, is of interest in that it is the estimate of the standard deviation of the population.

Usually, we place a caret (^) over a symbol that indicates an estimated value. The one exception is the symbol S_{X_j}, which is the (ESTIMATED/ACTUAL) standard deviation of the __________.

No answer required

5-45. As an aid to understanding, and also as a possible check against miscomputations, it is helpful to predict the result in advance. With 80% reliability, we can tell from the preceding exercise that the resulting precision will be a (LARGER/SMALLER) figure than $20,000.

EXPLAIN YOUR ANSWER:

____

____

---

the estimate of the total population value is obtained by multiplying the sample mean by the number of elements in the population

(or similar wording)

$\hat{X}$ = $450,000

6-28. This is the figure which we have been trying to obtain from the very beginning.

There is, however, one more step. For purposes of reporting, we need to know the attained precision and reliability of the estimate. These figures should come very close to our <u>desired</u> figures of $20,000 and 80% respectively, but may not be identical. Give one reason why not.

____

ESTIMATED

population

4-49. We now have almost all the information we need in order to evaluate the accuracy of an estimate based on a mean of a random sample. Below is a summary of the main points discussed so far:

1. An estimate of the population total is obtained by multiplying the sample mean ($\bar{x}$) by the number of elements in the population (N).

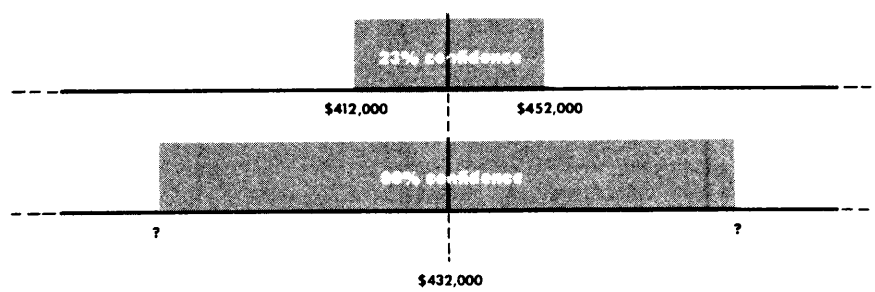
2. The closer the sample mean is to the actual population mean, the better the estimate will be.

3. The extent to which a sample mean will tend to approximate the actual mean depends in part on the ____________ of the population.

LARGER (the precision limits must be wider in order to increase our confidence that the precision interval will contain the true value)

(If you missed this answer, study the diagrams at the right.)

5-46.



We haven't calculated the 80% precision limits, but we know that they must be wider than the 23% precision limits.

(No answer required)

/ANY OR ALL OF THE FOLLOWING/

1. Estimated standard deviation based on the combined sample (n=706) is different from the estimate based on the preliminary sample (n=30).
2. We added 10% to the sample size that was computed by formula to be sufficient for the desired precision and reliability.
3. Very slight differences may result due to rounding.

6-29. Since precision and reliability vary together, we cannot find both simultaneously, for neither value really has meaning unless we know the other one. First we will find the reliability at the desired precision level of $20,000. Locate the appropriate worksheet for this task and fill in the first column in the row labeled "ABC Store (Final)."

variability (standard deviation)	4-50. The following "semi-review" questions have not been covered directly, but their answers can be deduced from what has been discussed. The sample size can obviously be increased or decreased at will. Can the same be said about the population standard deviation? (YES/NO) EXPLAIN: ______________________________ ______________________________________
No answer required	5-47. From Table 1 we determine that 80% of all sample means of a given size will differ from the true population mean by no more than $\pm$ _______ times the standard error of the mean. Fill in the answer in the appropriate space of Worksheet 3. (We are working with the ABC Store preliminary sample.)
The figure $20,000 should be entered in the first column of Worksheet #2.	6-30. The remainder of the worksheet involves exactly the same kind of work that we did earlier. The only difference is that we have a larger sample and a different standard deviation. Before doing any computations, then, you should be aware that with respect to the preliminary entry for the ABC Store problem, which was based on a desired precision of $20,000, the figure to be entered in Column 2 will be (THE SAME/DIFFERENT).

NO The standard deviation is an inherent characteristic of the population. We can <u>compute</u> a standard deviation, but cannot <u>control</u> it.	4-51. We have stated that at least 30 elements, if randomly selected, should be used for an estimate of the population standard deviation. Would, say, 50 elements give us an even better estimate? (YES/NO) If NO, why not? ____________________ ____________________ If YES, why don't we select 50 rather than 30? ____________________ ____________________
1.28	5-48. 1. What is the estimated standard error of the mean in this example? (Fill in the answer in the appropriate space in the worksheet.) 2. Why is it the same as the figure given in the preceding worksheet for the ABC Store problem? ____________________ ____________________
THE SAME (If you were correct and understand why, skip to Frame 6-32.)	6-31. Our first sample mean was \$48.0, and our mean from the larger sample was \$50.0. In either case, the simple statement of fact at the top of Column 2 holds true: If the true mean differs from the sample mean by more than \$2.22, the estimate of the total will be in error by more than \$20,000. The only difference is that with the larger sample of 706, the probability of the sample mean actually being in error by this amount or more has considerably (INCREASED/DECREASED).

YES

Each element in the sample adds more to our time and cost. The additional information may not be worth it.

4-52. The statement at the left refers only to our estimate of the standard deviation. A useful estimate of the population total value usually (DOES/DOES NOT) require more than 30 elements in the sample.

1. $8.00

2. The estimated standard deviation of the population and the sample size are the same; therefore, $\hat{\sigma}_{\bar{x}}$ is the same. (The only thing that has changed is the information we are looking for.)

5-49. There is therefore an 80% chance that the difference between $\bar{x}$ and $\bar{X}$ is no more than _____ times _____. Numerically, this works out to $_______. (Fill in the last answer in the worksheet.)

DECREASED

6-32. The new estimate of the population standard deviation is obtained from the Combined Sample Worksheet (#5). The square root of n (706) must be extracted or looked up.

Using these figures, compute the new estimated standard error of the mean and enter it in Column 3. Also fill in the correct figure for Column 2 if you have not already done so.

DOES	4-53. In the next chapter, we will see how the standard deviation relates to the precision and reliability of an estimate, and in the final chapter, we will see how to calculate the required sample size for an estimate at any given level of precision and reliability. END OF CHAPTER 4
1.28 times \$8.00 \$10.24	5-50. We can be 80% confident that the true population mean is within \$_______ of the mean that was calculated from our preliminary sample. We can therefore be equally confident that the true population total is within \$_______ of the total we estimated from our preliminary sample.
Column 2: \$2.22 $\hat{\sigma}_{\bar{x}} = \frac{S_{x_j}}{\sqrt{n}}$ $= \frac{\$44.7}{26.6}$ $= \$1.68$	6-33. If the standard error of the mean is \$1.68, then a difference between $\bar{x}$ and $\bar{X}$ of \$2.22 is equal to _______ times the standard error of the mean. (Compute as directed in Column 4 and enter the answer there.)

CHAPTER 5

CHAPTER 5. STANDARD ERROR, PRECISION, AND RELIABILITY

5-1. In Chapter 4, we calculated the sample mean and standard deviation of the ABC Store accounts receivable sample, and discussed the significance of these concepts. We observed that when an estimate of a total population value is based on a sample mean, any difference between the sample mean ($\bar{x}$) and the population mean ($\bar{X}$) is magnified ________ times.

$10.24

$92,160 ($10.24 x 9,000)

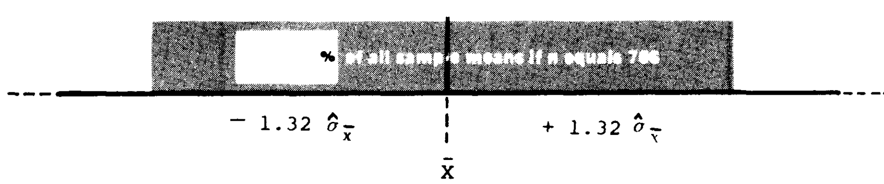
5-51. Is $92,160 the figure that should be entered in the fifth column of Worksheet 3?

a. YES (Frame 5-52)

b. NO (Frame 5-53)

1.32 ($2.22 ÷ $1.68)

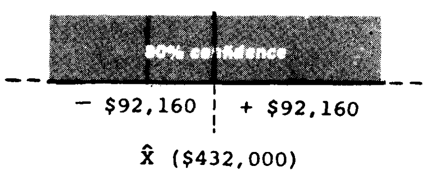
6-34. What percentage of all sample means from this population when n = 706 will differ from the true population mean by no more than 1.32 times the standard error of the mean? This is the question asked in the diagram below. Answer it by referring to Table 2 and recording the resulting reliability figure in Column 5 of the worksheet.



N	5-2. The sample mean, of course, might be either larger or smaller than the actual population mean. In either case, the difference between the two is expressed as follows: $$\left\|\bar{x} - \bar{X}\right\|$$ The vertical lines stand for "absolute value." This means that only the <u>magnitude</u> of the difference is considered. For example, $\|40 - 50\|$ (DOES/DOES NOT) have the same absolute value as $\|50 - 40\|$.
	5-52. YOUR ANSWER: YES Correct. This is the maximum error at the given level of reliability (80%). There is an 80% chance that the true value differs from the estimated value by no more than this amount. Enter it in Column 5 of the worksheet, and go on to Frame 5-54.
R = 81% (This percentage was obtained through the process of interpolating in Table 2. A more conservative procedure would be to round 1.32 down to the next lower U figure given in the table. This would make the R figure 80%.)	6-35. We have just seen that with our desired precision of $20,000, we can claim 81% reliability. If we stay with our original desired reliability of 80%, what precision could we attain? You can work this out in Worksheet #3. Read the column headings carefully and remember that we are using our <u>original</u> desired reliability and our <u>new</u> sample data.

DOES

5-3. In order to know the precision and reliability of our estimate, therefore, we need to have some way of estimating the probable difference between $\bar{x}$ and $\bar{X}$. This can only be accomplished by applying certain laws of probability, for in any given case we (CAN/CANNOT) tell how much our sample mean differs from the true population mean.



5-53. YOUR ANSWER: NO

Yes, it should. We are 80% confident that the sample mean does not differ from the population mean by more than $10.24. We can therefore have the identical degree of confidence that the estimate of the total does not differ from the true total by this amount times N (9,000). This amount, $92,160, is the precision of the estimate and should be entered in Column 5. The diagram at the left can be compared with the diagrams in Chapter 1, Frames 39-43, which are recommended if you are still uncertain as to the meaning of precision.

R: 80%

U: 1.28

$\hat{\sigma}_{\bar{x}}$: $1.68

4: $2.15

A: $19,350

6-36. The auditor can therefore report as follows:

1. The estimate of the total dollar value of all three-month-or-more overdue amounts is $450,000 ± $20,000, with a reliability of 81%.

AND/OR

2. The estimate is $450,000 ± $19,350, with a reliability of 80%.

Since either of these statements meets the requirements given in the statement of the problem, further sampling (IS/IS NOT) necessary.

CANNOT

A	B
0	0
4	6
10	12
11	21
13	22
19	25
38	26
41	31
47	40
50	50
$\Sigma X_j = 233$	$\Sigma X_j = 233$
$\bar{X} = 23.3$	$\bar{X} = 23.3$

5-4. We do have reason to believe, however, that the probable difference between $\bar{x}$ and $\bar{X}$ depends to some extent on the variability of the population. For example, consider the two populations at the left. By calculating the sum of squared deviations from the mean, or simply by inspecting the values as in previous examples of this kind, you could conclude readily enough that even though they have the same range (0 to 50), (A/B) has a higher standard deviation.

No answer required

5-54. The auditor can therefore report as follows: "An unrestricted random sample of 30 of the accounts produced a mean of $48.0 and a standard deviation of $44.0. Our preliminary estimate for the total of amounts that are three months or more overdue in all 9,000 accounts is therefore $432,000. The attained precision interval at 80% reliability is $________ to $________."

IS NOT

6-37. The steps by which we arrived at this conclusion are summarized in Exhibit 8. This summarizes the entire unrestricted random sampling procedure from statement of the problem to calculation and evaluation of the final results. It is suggested that the reader at this point simply glance over the pages of Exhibit 8 and use them later to refer to as needed.

(No answer required)

A

A	B
0	0
4	6
10	12
11	21
13	22
19	25
38	26
41	31
47	40
50	50
$\Sigma X_j = 233$	$\Sigma X_j = 233$
$\bar{X} = 23.3$	$\bar{X} = 23.3$

5-5. From which population would sample means tend in general more to approximate the true mean of 23.3?

a. A

b. B

c. No way of telling

\$339,840 to \$524,160

(\$432,000 ± \$92,160)

5-55. At the same time, we have seen that if our estimate is to be within ± \$20,000 of the true value, our confidence in this estimate would be only 23%. At this point, in practice, the auditor would want to be in a position to make a better estimate. How can he do this?

____

No answer required

6-38. We will now use the same procedures applied to the shoe inventory problem. This was described in Exhibit 3. Notice that the desired reliability is considerably higher than in the accounts receivable problem. This might be because of the consideration of materiality in the two situations. Whatever the reason, establishing the desired precision and reliability levels is obviously a decision of the (STATISTICIAN/AUDITOR).

b. B	5-6. One way to tell is to try it out. Select half a dozen samples of four elements from population A. List the means that result. Then do the same for population B. The means from the less variable population, B, will be less "scattered" compared to those means selected from population A. Moreover, they will tend to deviate (MORE/ LESS) from the true population mean of 23.3.
add more elements to the sample (or similar answer)	5-56. This is what we will do in Chapter 6. First, as additional practice and for teaching purposes only, let us change our precision requirement to $54,000, and our reliability requirement to 70%. We will now determine whether our preliminary sample gives sufficient information to meet these requirements. Turn to Worksheet 2 and fill in the desired precision in the row labeled "ABC Store (Hypothetical.)" Keep the worksheet easily accessible.
AUDITOR (This point is discussed more fully in Appendix 1.)	6-39. Our preliminary sample that we selected back in Chapter 3 yielded the results shown in Exhibit 5. Compute the following: $\bar{x}$ ____________ $\hat{X}$ ____________

LESS	5-7. We conclude, therefore, that the greater the standard deviation of the population, the (GREATER/SMALLER) the difference between most sample means and the true population mean.
You should have filled in \$54,000 in Column 1 of Worksheet 2.	5-57. If you wish to complete the entire exercise now, do so and check your answer on page 98 opposite Frame 5-61. If any uncertainty remains about either the concepts or the computational procedures, it is suggested that you do it step by step in the following frames. If our desired precision is \$54,000 for the total estimate, then the sample mean cannot differ from the true population mean by more than \$________. (Answer here first; then in the worksheet.)
$\bar{x}$ = \$497 $\hat{X}$ = \$149,100 (N = 300)	6-40. The sum of the squares of x_j values has been done for you in Exhibit 5. Use the appropriate worksheet to calculate the estimated standard deviation of the population based on this sample. (NOTE: $497^2 = 247,009$)

GREATER	5-8. However, even if the standard deviation of the population is very large, an increase in the sample size will tend to (INCREASE/ REDUCE) the probable difference between $\bar{x}$ and $\bar{X}$.
\$6.00	5-58. We had computed the estimated standard error of the mean to be \$8.00. This time it will be: a. greater b. smaller c. the same EXPLAIN YOUR ANSWER: ______________________ ______________________________________
S_{X_j}: \$26.9 F: \$7,410,270 G: \$20,910 H: \$721.03 I: \$26.9	6-41. Notice once again that although we are working with the sample values we use the symbol S_{X_j} to refer to the estimated standard deviation. (No answer required)

REDUCE

5-9. In fact, it is known (although we shall not prove it in this book) that dividing the estimated standard deviation of the population by the square root of the sample size gives us an estimate of a quantity known as the "standard error" of the mean. It is also known that for large sample sizes approximately 68% of the time a sample mean will not differ from the true population mean by more than this amount.

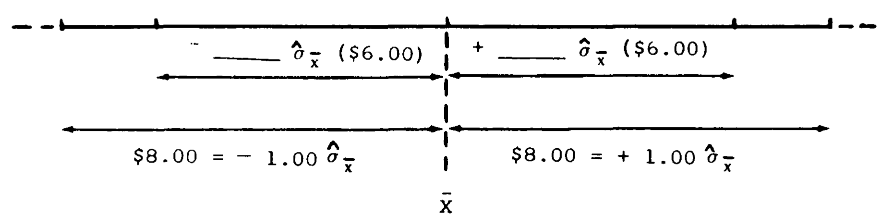
(No answer required)

---

c. the same

(the estimated standard deviation and the sample size have not changed)

5-59. The next step on the worksheet is equivalent to filling in the blank in the diagram. You can do the arithmetic either mentally or by following the hint at the top of Column 4 of the worksheet. Fill in the blanks in the diagram and enter the answer in the appropriate space in the worksheet.



No answer required

6-42. Now that we have an estimate of the population standard deviation, the preliminary sample has served its purpose. Ordinarily we would go right on to the sample size worksheet. However, there may be cases when the auditor might wish to calculate the precision and reliability of the preliminary estimate (which in this case was $149,100).

(No answer required)

No answer required

5-10.

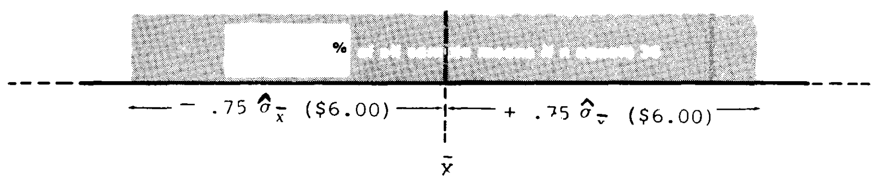
$$\hat{\sigma}_{\bar{x}} = \frac{S_{x_j}}{\sqrt{n}}$$

The term at the left is the estimated standard error of the mean. The standard error of the mean measures the variability among all possible sample means based on samples of size n. The above relationship tells us that we can estimate this variability by dividing the estimated population ________ ________ by the square root of the sample size.

For the present we are interested in the fact that for large random samples of size n, approximately 68% of the calculated sample means (WILL/WILL NOT) differ from the true population mean by more than this amount.

.75 ($6.00 ÷ $8.00)

5-60. Once again, fill in the answer both in the diagram and in the worksheet. You will have to interpolate in Table 2 and round <u>down</u> to the nearest percent.



No answer required

6-43. In the ABC Store preliminary sample, the x_j values ranged from $7.00 to $203.00 with a standard deviation of $44.00. The population contained 9,000 elements. There was little chance, then, that a preliminary sample of 30 items would give us a usable estimate.

In the XYZ Shoe Inventory problem, on the other hand, simple inspection of Exhibit 5 will show that there is much smaller spread around the sample mean (S_{x_j} = $26.9). The population contained only 300 items. In comparison to the ABC Store example, then, there is relatively (MORE/LESS) chance that the preliminary sample will yield results that are useful to the auditor.

Standard deviation WILL NOT	5-11. NOTE In the following examples, as in all examples in this book, we are assuming that the sample under consideration has been selected using unrestricted random sampling with replacement, and that the estimate of the population standard deviation has been based on a sample of at least 30 elements. (No answer required)
53% (If you were correct, skip to Frame 5-62.)	5-61. U: .7, .8 — R: 51%, 56% Above is the relevant section of Table 2. Our U value for .75 must lie between 51% and 56%. (It does not actually lie halfway between, but if you simply take the halfway value and round down to the nearest percent, the error, if any, will be on the conservative side.) The midpoint between 51% and 56% is _____%, which we round down to _____%.
MORE	6-44. Secondly, a statistical estimate of an inventory might be made as a check against another independent estimate or complete count. If, for instance, you were checking a reported figure of $150,000, you might be able to satisfy yourself just with the preliminary sample provided the reliability were reasonable and the precision not too large. Statistical experience will give you a "feel" for the significance of sample figures but, as we have already seen, any decision based on the figures is a matter of the auditor's judgment. (No answer required)

No answer required	5-12. Before we can show their applications to auditing, we must first acquire some arithmetical and verbal facility in working with these quantities. The formula for the estimated standard error of the mean is: $$\hat{\sigma}_{\bar{x}} = \frac{s_{x_j}}{\sqrt{n}}$$ The expression in the numerator stands for ________________________________. The expression in the denominator stands for ________________________________.
53.5% 53%	5-62. We cannot, therefore, achieve a precision of \$54,000 with a reliability of 70%, for we can only be 53% confident that the error will be no more than this amount. We could use the next worksheet to determine what the precision would be for a reliability of 70%. First, what do you think will be the result? a. more than \$54,000 b. less than \$54,000 c. no way of telling
No answer required	6-45. We will continue to use the XYZ Shoe Company inventory sample as a review exercise covering all procedures discussed in Chapters 4 through 6. We have already computed the preliminary estimate of the total to be \$149,100 and have just seen that this is a case in which it might be appropriate to calculate the precision and reliability of the preliminary sample. Begin your review, then, by using the appropriate worksheet to compute the reliability of this estimate at our desired precision of \$2,000. (Use 5.5 as the square root of 30.)

estimated standard deviation of the population square root of the sample size	5-13. $\hat{\sigma}_{\bar{x}} = \frac{s_{x_j}}{\sqrt{n}}$ 1. If a randomly-selected sample containing 36 elements has an estimated standard deviation of $60, how much is the estimated standard error of the mean? ________ 2. What would it be if the sample contained 49 elements? ________ 3. 49 elements with an estimated standard deviation of $70? ________
a. more than $54,000	5-63. Since we know that the precision will be more than $54,000, we have negatively answered our hypothetical question: "Would this preliminary sample enable us to make an estimate with a precision of $54,000 and a reliability of 70%?" However, as an <u>optional</u> exercise you may wish to find the precision at this reliability. If so, fill in the first two columns in the "ABC Store (Hypothetical)" row in Worksheet 3. Otherwise, skip to Frame 5-66.
A: $\pm$ $2,000 $\lvert \bar{x} - \bar{X} \rvert$: $6.67 $\hat{\sigma}_{\bar{x}}$: $4.89 ($26.9 ÷ 5.5) U: 1.36 R: 82% (Worksheet 2)	6-46. Explain in your own words the meaning of the R figure in this example. ________________________________ ________________________________

1. \$10 $\frac{\$60}{\sqrt{36}}$

2. \$8.57 $\frac{\$60}{\sqrt{49}}$

3. \$10 $\frac{\$70}{\sqrt{49}}$

5-14. Notice how the standard error of the mean varies with both standard deviation <u>and</u> sample size by comparing examples 1 and 3 at the left. To see the effects of sample size, compare 1 with 2; to see the effects of standard deviation, 2 with 3.

(No answer required)

70% (given)

1.04 (from Table 1)

5-64. The estimated standard error of the mean is still \$8.00. Show the arithmetic involved in reaching this result.

There is an 82% chance that the estimated value ± \$2,000 will contain the true value.

(or similar value)

6-47. If we used the same preliminary sample for the same estimate (\$149,100) at 90% reliability, what would be the precision? Use the appropriate worksheet.

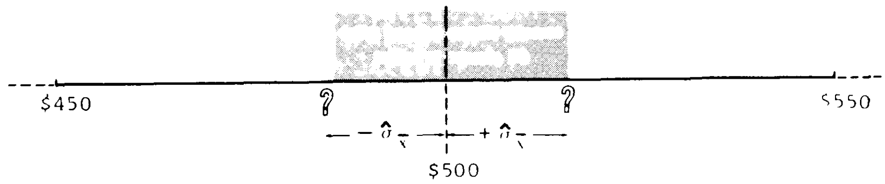
No answer required	5-15. It is known (although we shall not prove it here) that if we took every possible sample of a given size from a population and calculated the mean for each of these samples, approximately 68% of all these sample means would differ from the true population means by no more than $\hat{\sigma}_{\bar{x}}$, the standard error of the mean. In any given case, the standard error of the mean can be estimated by ______________ ______________________. (Answer may be either in symbols or words.)
$\hat{\sigma}_{\bar{x}} = \frac{s_{x_j}}{\sqrt{n}}$ $= \frac{\$44.0}{5.5}$ $= \$8.00$	5-65. Fill in the remaining two columns. The precision at this reliability is $________.
R: 90% U: 1.64 $\hat{\sigma}_{\bar{x}}$: $4.89 #4: $8.02 A: $2,406 (Worksheet 3)	6-48. At this point, the auditor can say: 1. The estimate is $149,100 with a precision of $2,000 at a reliability of 82% and/or 2. With 90% reliability, the estimate is $149,100 with a precision of $2,406 Could he stop right there? YES (Frame 6-49) NO (Frame 6-50)

dividing the estimated standard deviation of the population by the square root of the sample size

or

$$\frac{S_{X_j}}{\sqrt{n}}$$

5-16. If the estimated population standard deviation is $90 and the sample size is 36, the estimated standard error of the mean would be $15.00. If the true population mean is actually $500 as shown in the diagram, then 68% of all possible sample means will be between _______ and _______. (Fill them in in place of the question marks in the diagram.)



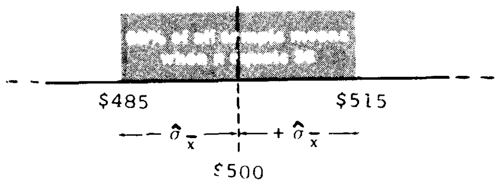
$8.32 ($8.00 x 1.04)

$74,880 (Precision)

5-66. In this chapter we have seen how the computation of the estimated standard error of the mean, $S_{X_j} \div \sqrt{n}$, can enable us to determine the precision of an estimate at a given reliability or the reliability of an estimate at a given precision. In practice, the preliminary sample is usually not worked with in such detail, for it is rare that the initial sample of 30 items will be enough to achieve our desired degree of accuracy. Moreover, there is rarely any valid reason for changing the desired precision and reliability. What this means, then, is that we have to add to our sample in (ALL/MOST/FEW) cases.

6-49. YOUR ANSWER: YES

The statistician would certainly find no fault with either of these statements. However, once again the decision is one requiring auditor's judgment. Having originally established a desired reliability of 90% for a precision of $2,000, is there any good reason for the auditor to wish to lower his confidence to 82%? Or if he wants to be sure that his estimate has a reliability of 90%, can he justify extending his precision limits by more than $400? Only the individual circumstance can indicate the yes or no answer to this question.



5-17. By the same token, if the <u>sample</u> mean is \$500, then there is a 68% chance that this differs from the true population mean by no more than \$15. Why this amount?

____

____

MOST

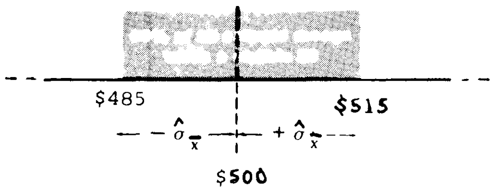
5-67. In the next and concluding chapter, therefore, we will see how to determine the appropriate sample size at the desired precision and reliability for a population whose variability has been estimated on the basis of the standard deviation of a preliminary sample.

END OF CHAPTER 5

6-50. YOUR ANSWER: NO

This depends on the individual circumstance. For instance, in the checking situation described in Frame 6-44, a precision of \$2,406 might be just as good as a precision of \$2,000. On the other hand, holding the precision constant and lowering the reliability is a little harder to justify, since reliability is a less tangible concept than precision and may seem more subjective to the interested party who is not a statistician. However, there can be no yes or no answer to this question other than that supplied by the auditor's judgment.

$15 in this example is the estimated standard error of the mean ($S_{X_j} \div \sqrt{n}$)



5-18. If the estimated population standard deviation were $50, and the sample size were 100, then $\hat{\sigma}_{\bar{x}}$ would equal ______. Cross out the figures 485 and 515 and in the box at the left insert the resulting figures from this example.

CHAPTER 6

CHAPTER 6. DETERMINATION OF SAMPLE SIZE

6-1. At the end of the last chapter, we had an example in which the preliminary sample did not give the auditor sufficient information. This will usually be the case, for the basic purpose of the preliminary sample is to:

a. provide an estimate of the population standard deviation

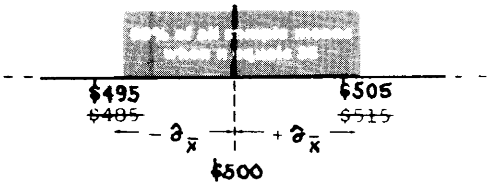
b. provide an estimate of the total population value at desired precision and reliability

---

6-51. Whatever the auditor's decision, we still have before us the purely statistical question: "What sample size is required for an estimate of the total population at 90% reliability with a precision of $2,000?"

Answer this question in your sample size worksheet. All source data is available to you from the various worksheets, tables and exhibits.

$5 ($50 ÷ $\sqrt{100}$)



5-19. There is no reason to confine ourselves only to the figure "68%." That simply happens to be the (approximate) proportion of cases in which a mean calculated from a random sample of size n will not differ from the true population mean by more than:

a. $10.00

b. $ 5.00

c. $S_{X_j} \div \sqrt{n}$, whatever they happen to be

---

a. provide an estimate of the population standard deviation

6-2. In order to meet our desired criteria, therefore, we have to increase our sample size. The following is an overview of the process by which we ultimately arrive at an estimate with the desired precision and reliability: (1) determine appropriate sample size by use of a formula; (2) continue with the random sample selection until the correct number of elements have been selected; (3) combine the preliminary and additional samples and calculate the mean and standard deviation of this combined sample; (4) after estimating the total value, calculate the precision of this new estimate at the desired reliability level; (5) evaluate the new results. (No answer required)

R: 90%

U_R: 1.64

S_{X_j}: $26.9

#4: $44.12

N: 300

6: $13,236

A: $2,000

$\sqrt{n}$: 6.6

n: 43.56 → 44

10% of n = 4.4

sample size = 48.4 → 49

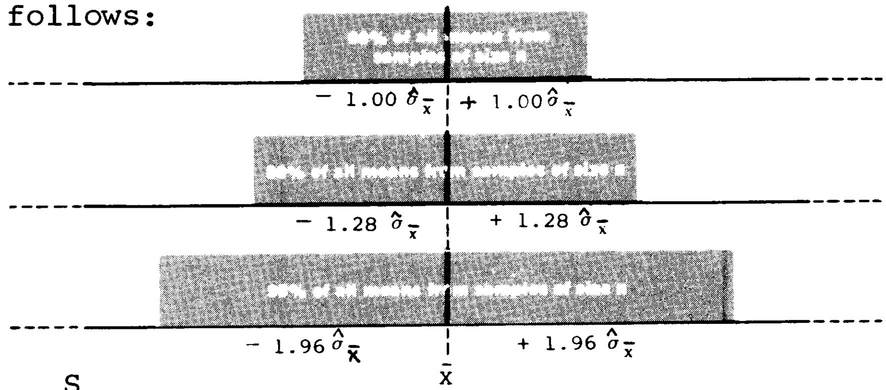
6-52. This calculation tells us that an additional 19 elements must be randomly selected. The following question is easy in itself, but is a test of your recollection of where to look for all the relative sampling information: What is the next lot number selected in the sample? ____________

c. $S_{X_j} \div \sqrt{n}$, whatever they happen to be	5-20. The same statistical tables that tell us that 68% of the time a sample mean will not differ from the true population mean by more than ± 1.00 times $\hat{\sigma}_{\bar{x}}$ also tell us that 80% of the time the difference will be no more than ± 1.28 times $\hat{\sigma}_{\bar{x}}$ and that approximately 95% of the time, the difference will be no more than 1.96 times $\hat{\sigma}_{\bar{x}}$. Thus, if the estimated standard error of the mean in a given population for a given sample size is $10.00, approximately 80% of all samples of that size would result in means that differ from the true population mean by no more than $10 x 1.28. Approximately 95% of all these sample means would be within ± _______ of the true population mean.
No answer required	6-3. Locate Worksheet #4 and keep it readily accessible. The formula at the top is the sample-size formula. After solving for the square root of n, we square the result and round off as indicated in the note. This formula is easily derived from the basic concepts under discussion thus far. If you are interested in this derivation, you may read the following optional sequence from 6-4 through 6-11. Otherwise, skip now to Frame 6-12. (No answer required)
BD-1Q6W (If you were correct, go on to Frame 6-54; if incorrect, the answer appears at the right.)	6-53. The route through the random number table, as indicated in Exhibit 4, specifies using the first three digits in each line and proceeding downward. As indicated in Exhibit 5, we had concluded the preliminary sample in Row 499, Column 3. The next random number to be selected is 257 in Row 500, Column 3. This corresponds to Lot #BD-1Q6W (Exhibit 3-A). (No answer required)

$19.60

($10, the standard error of the mean, times 1.96 which is obtained from a table)

5-21. This is shown diagramatically as follows:



If $\frac{S_{X_j}}{\sqrt{n}}$ is $20, then approximately 80% of all means from samples of size n will differ from the true mean by no more than ± $_______.

No answer required

6-4. DERIVATION OF SAMPLE SIZE FORMULA

In Chapter 4 we saw that the precision of the estimated total could be expressed as the estimated standard error of the mean multiplied by the appropriate reliability factor, magnified N times. Thus,

$$(1) \quad A = N \cdot U_R \cdot \hat{\sigma}_{\bar{x}}$$

where A is the maximum error (±) in the (MEAN/TOTAL) estimate at any given level of reliability and the period is the algebraic sign for multiplication.

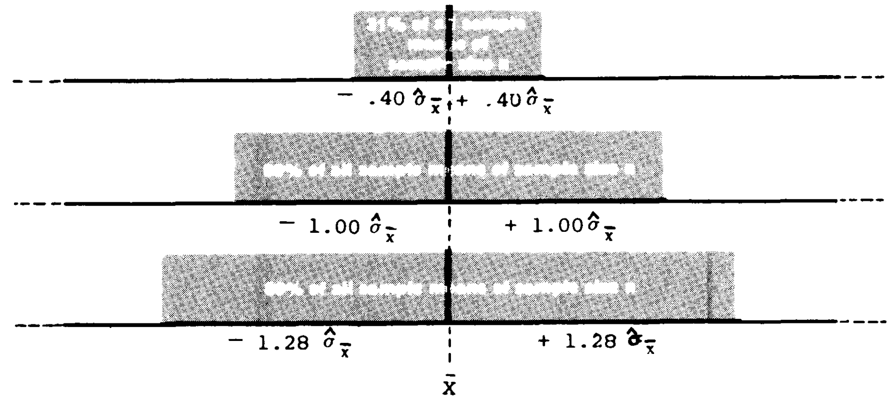
No answer required

6-54. Assume that the additional 19 inventory lots have a total value of $9,394 and the sum of the squared values is $4,658,644. Fill in the Combined Sample Worksheet (#5) to get the estimate of the standard deviation of the population. (Note: $496^2 = 246,016$)

\$25.60 (\$20 x 1.28)	5-22. How frequently will the sample mean differ from the population mean by <u>less</u> than 1 standard error of the mean (1.00 times $\hat{\sigma}_{\bar{x}}$)? From the preceding diagram, or from the discussion thus far, we could predict that this would happen (MORE/LESS) than 68% of the time.
TOTAL	6-5. The precision, A, measures in a probabalistic sense the extent of the difference between the estimated population total and the actual total. This means that were we to select samples repeatedly, in approximately R% of the cases the quantity $\lvert N\bar{x} - N\bar{X} \rvert$ would be less than $N \cdot U_R \cdot \hat{\sigma}_{\bar{x}}$. You see that this is equivalent to the statement that $\lvert \bar{x} - \bar{X} \rvert$ will be less than $U_R \cdot \hat{\sigma}_{\bar{x}}$ approximately R% of the time. The vertical lines indicate ____ ________________________________.
(computed values only) Row 3: \$24,304 Row 8: \$12,089,824 Row 9: \$246,016 Row 10: \$12,054,784 Row 11: \$35,040 $s_{X_j}^2$: \$730.00 s_{X_j}: \$27.0	6-55. Our new estimate of the total inventory value is, of course, ________ x ________ or \$148,800.

LESS

5-23. This is illustrated by the top diagram. A case in which the sample mean differs from the true population mean by only four-tenths of the standard error or less is relatively (FREQUENT/INFREQUENT), occurring in the long run approximately _____% of the time.



absolute value (magnitude of the difference irrespective of whether the difference is positive or negative)

6-6. The standard error of the mean is, by definition, the standard deviation that would be calculated from the population of the means of every possible sample of a given size n that could be drawn from the original population. It obviously (IS/IS NOT) practical to attempt this calculation.

$496 x 300

6-56. Once again we calculate the precision and reliability of the estimate, this time using the combined sample data. Should we expect our desired criteria to be met? (YES/NO)

EXPLAIN: ______________________________

____

INFREQUENT 31%	5-24. How do we know that a sample mean will differ from the population mean by no more than four-tenths of the standard error only 31% of the time? Only through certain mathematical theorems. However, the auditor can find this information in Tables 1 and 2 on page 37 of the Supplementary Section. For instance, Table 1 tells us that 99% of the time, the sample mean will differ from the population mean by no more than $\pm$ _____ times the ______________________________.
IS NOT	6-7. We therefore <u>estimate</u> the standard error of the mean in the following formula: $$\hat{\sigma}_{\bar{x}} = \frac{s_{x_j}}{\sqrt{n}}$$ s_{x_j} is ______________________________
YES (sample size was chosen by a formula designed to meet the desired criteria, and we added 10% for additional insurance)	6-57. What is the reliability of the estimate if we wish to hold our precision at $2,000? Use the appropriate worksheet. (The standard error of the mean must be computed from the new sample data.)

2.58 standard error of the mean	5-25. What is the probability that the difference between a sample mean and the population mean will not be greater than, say, 1.15 times the standard error of the mean? Take a guess based on the diagram in Frame 5-21: ______% Then look it up in Table 2: ______%
the estimated standard deviation of the population	6-8. Substituting for $\hat{\sigma}_{\bar{x}}$ in equation (1), $$(2) \quad A = \frac{S_{X_j} \cdot U_R \cdot N}{\sqrt{n}}$$ and by simple algebraic transposition, $$(3) \quad \sqrt{n} = \frac{S_{X_j} \cdot U_R \cdot N}{A}$$ This is the equation for determining sample size as given in Worksheet 4. (No answer required)
A: $2,000 $\lvert \bar{x} - \bar{X} \rvert$: $6.67 $\hat{\sigma}_{\bar{x}}$: $3.86 ($27.0 ÷ 7) U: 1.73 R: 91%	6-58. If he wishes, the auditor could stop at this point. However, auditors ordinarily find it is preferable to stick to the desired round-number reliability (80%, 90%, etc.) and then determine the precision. Do this now.

Guess: somewhere between 68% and 80% Table 2: 74%	NOW TURN BACK TO PAGE 61 AND BEGIN THE SECOND ROW.
No answer required	NOW TURN BACK TO PAGE 61 AND BEGIN THE THIRD ROW.
R: 90% U: 1.64 $\hat{\sigma}_{\bar{x}}$: \$3.86 #4: \$6.33 A: \$1,899	6-59. Thus, thanks to the extra 10% in the sample, we can attain our desired reliability and narrow our original precision by close to a hundred dollars. Our final report, then, using the data at the left and the combined sample mean of \$496.0, is that there is a _______ chance that the inventory value is between __________ and __________. ($\hat{X}$ = \$148,800)

90%

$146,901 and $150,699

6-60. This concludes our coverage of unrestricted random sampling for variables. Since this is an introductory and non-theoretical volume, it is not expected that all of the underlying mathematical theory involved be perfectly clear to you at this point, but only that the practical work has been thorough enough so that you could work out a similar estimating problem from the beginning. An opportunity to do just that is provided in the Questions and Problems in the Supplementary Section.

END OF VOLUME 1