1968

Auditor's Approach to Statistical Sampling, Volume 3. Stratified Random Sampling

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

3

AN AUDITOR'S APPROACH TO STATISTICAL SAMPLING

STRATIFIED RANDOM SAMPLING

Individual Study Program
Professional Development Division
American Institute of Certified Public Accountants

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 prepared by David Monroe Miller, programing consultant; Robert E. Healy, CPA, and Morton J. Rossman, CPA, consultants; and Thomas R. Hanley, CPA, Manager, Special Projects. The members of the Committee on Statistical Sampling assisted in an advisory capacity.

AN AUDITOR'S APPROACH TO STATISTICAL SAMPLING

Volume 3

STRATIFIED RANDOM SAMPLING

Programed for the
American Institute of Certified Public Accountants
by
DAVID MONROE MILLER

TABLE OF CONTENTS

*Book is turned over at this point.

PREFACE

The Committee on Statistical Sampling of the American Institute of Certified Public Accountants has undertaken the development of a series of programed texts on statistical sampling techniques in auditing to broaden education in this area. This, the third volume in the series of programed texts, deals with stratified random sampling.

Earlier volumes in the series, *An Introduction to Statistical Concepts and Estimation of Dollar Values* (Volume 1) and *Sampling for Attributes* (Volume 2) cover certain basic statistical concepts and sampling for particular characteristics measured either quantitatively in terms of dollar values or qualitatively in terms of frequency of occurrence. Because Volume 1 serves as introductory material for Volume 3, the reader should have completed that volume before beginning this one.

No attempt has been made to explain the mathematics and theory underlying the formulas used in this volume, nor have criteria been suggested or established for statistical precision and reliability. Precision and reliability are subjective determinations and should be based upon the judgment of the auditor. The determination as to whether statistical sampling is appropriate in the circumstances also must be made in each instance by the auditor based on his knowledge and judgment.

In this volume, as in previous volumes, examples of statistical sampling applications have been constructed. These examples are for instructional purposes only. No implication should be drawn that the techniques discussed in this volume provide the sole means of making estimates in similar situations.

This volume is the result of a joint effort on the part of the entire (1967-1968) Committee on Statistical Sampling with the assistance of Thomas R. Hanley, CPA, Manager, Special Projects. However, special acknowledgement is made of the contribution of two of its members, Robert E. Healy, CPA, and Morton J. Rossman, CPA, to the content and for coordination with the programing consultant, David Monroe Miller.

This booklet is being made available to the Institute membership as part of the continuing education program of the Professional Development Division.

RICHARD A. NEST, CPA
Director of Technical Services

July, 1968

HOW TO USE THIS BOOK

This volume is similar in format to the previous volumes in the AICPA statistical sampling series. In going through the programed text, the reader turns the page after each frame and checks his answer in the left-hand box. The responses called for may be choices, fill-ins or problems to work out. Some frames are marked No Answer Required. These frames often contain important information and should be read as thoroughly as others. As in previous volumes, maximum educational value will be obtained by writing answers directly in the spaces provided, and then changing them if they prove to be incorrect.

This volume is approximately the same length and of the same order of complexity as Volume 1. There is considerable less theory presented, because of the overlap from that volume, but the formulas used in stratified sampling involve more terms than those used in unrestricted sampling. This in itself will not slow the reader down, since an optional feature allows the formulas to be bypassed almost entirely; however, if done by hand, the computations will be considerably more difficult. The reader is therefore advised to have access to a calculating machine when reading Chapters 4 through 7.

The Supplementary Section contains worksheets and other reference material, similar in format to the earlier volumes, designed to be used in conjunction with the programed text. The text will always direct you to the appropriate page at the proper time in the teaching sequence, and you can also refer to the Supplementary Section at any time to examine work you have already done or to get an overview as to what is to come. However, on a few questions clearly indicated in the text, you will be advised not to refer to the Supplementary Section in order to increase the challenge of an exercise.

All pages in the Supplementary Section have been labeled "Exhibits" and are listed, with their titles and exhibit numbers, in the Table of Contents on page S-iii of the Supplementary Section. In addition, page S-v indicates how the Exhibits have been grouped. If you are not certain as to where to find a particular exhibit, consult these two pages and also check in the frame that you are working on.

One of the most frequently-used pages in the Supplementary Section is the "Summary of Stratified Sampling Procedures" on pages S-vii and S-ix. This Summary has general applicability and can be referred to in field problems, but is also keyed to this teaching volume. You can obtain an overview of this volume by skimming through the Summary and Table of Contents in the Supplementary Section right now. Then begin on Page 1 of the programed text.

CHAPTER 1. PRINCIPLES OF STRATIFICATION

1-1. Stratified random sampling is similar in many respects to the technique of unrestricted random sampling presented in Volume I of this series. The major difference is that the population is divided into two or more groups (strata), each of which is then sampled separately. The results can then be combined to give an estimate of the total population value.

The reasons for using this kind of sampling plan, together with some of the factors to be considered in stratifying the population, will be discussed in this chapter.

(No Answer Required)

1-2. In our discussion of unrestricted random sampling in Volume I, we had a few examples in which there were some extremely high and low dollar values within an otherwise homogeneous population. Assuming that the size of the sample remains the same, the presence of extreme values might tend to make it (MORE/LESS) likely that a random sample would be representative of the population as a whole.

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

LESS	REMINDER 1. As you go through this book you will be <u>turning</u> the page each time rather than going <u>down</u> the page. 2. If your answer proves to be incorrect, cross it out and substitute the correct answer. 3. The frames marked "No Answer Required" contain as much information as those which call for an answer. (No Answer Required)
DO	1-44. In addition to knowing <u>how many</u> elements are in each stratum, we also have to know <u>which</u> elements are in each. For example, suppose we knew for certain that 1,432 lots were below $200 in value, and 568 were above $200. We still could not estimate the value of each of the strata because we could not tell, in the case of any given inventory lot, whether it was in the first stratum or the second. Thus, in order to stratify by dollar value in this example, we would first have to inspect each of the 2,000 lots and determine their individual values. This would defeat the purpose of sampling. Instead of stratifying by dollar value, we will attempt stratification by ____________.
SMALLER (The reader may wish to do an experiment with the random number table. Select a sample of 20 out of a population of 50. You may be surprised at the large number of repeats.)	2-27. For this reason, stratified samples are generally taken <u>without</u> replacement. In your own words, what does this mean with respect to the use of the random number table? ____________________________________ ____________________________________

No Answer Required

- - - - - - - - - - -

$ 5
8
12
7
96
13
9
125
10
6
7

$\bar{X} = 27.1$

1-3. As we noticed in several examples in Volume I, the presence of some "unrepresentative" items (as in the miniature population at the left) would result in a high standard deviation. This, in turn, would require us to select a (LARGER/SMALLER) sample size than would otherwise be the case if these elements were not in the population.

---

kind of item

(or similar answer)

1-45. At least as a preliminary decision, then, we will break out the 2,000 inventory lots into two strata, one consisting of the standard items and the other consisting of the hand tools. We now check ourselves by applying certain criteria to the stratification plan. First, the exact number of elements in each stratum must be known. Is this the case in this example? (YES/NO)

EXPLAIN YOUR ANSWER: ______________________

____

Once a number has been selected, it cannot be selected again.

or

No element is selected more than once.

(or similar answer)

2-28. We will not go into the theory on which this discussion is based. The auditor need simply remember the following:

UNRESTRICTED	with replacement	(repeat)
STRATIFIED	without replacement	(no repeats)

The with- or without-replacement decision has been built into the formulas that you are given in Volume I and in this volume. Therefore, the auditor can simply follow the rules stated above without worrying about the effect of replacement on his sample-size computations.

(No Answer Required)

LARGER	1-4. These conclusions follow from the basic principle that the precision and reliability of a statistical estimate depend on the variability of the population and the size of the sample. We can alter the sample size by selecting as many or few elements as we wish. But, given the data in the previous frame -- or indeed, given any population -- can we change the variability? (YES/NO) EXPLAIN YOUR ANSWER: ______________________ ______________________________________
YES (explanation is in the next frame)	1-46. The "elements" in the population are not the individual inventory items, but the 2,000 lots. It is known that 1,500 lots contain only standard items, and 500 contain only hand tools. We are dealing with lot values, not item values, and do not need to know how many individual items there are. Naturally, however, for each lot selected in the sample, we (WILL/WILL NOT) have to add up the individual item values.
No Answer Required	2-29. When sampling without replacement, the auditor can keep track of used numbers either by some system of tick marks, as in Volume I, or by inspection. An opportunity for an additional check exists after the auditor has selected the elements in the preliminary sample. When listing the real lot numbers (as opposed to the random table numbers), he can check for duplication. (No Answer Required)

NO The variability of the population is determined by the values themselves. or Variability is an intrinsic characteristic of a population. (or similar wording)	1-5. Population variability is numerically defined in terms of the <u>standard deviation</u>. This concept is basic to all statistical sampling and especially so in discussing the difference between unrestricted sampling, which was covered in Volume 1, and stratified sampling, which will be covered in this volume. The following sequence reviews the standard deviation concept. Recent readers of Volume 1, or those who have a good background in statistics, may skip to Frame 1-11. (No Answer Required)
WILL (we have an approximate idea of the lot value <u>ranges</u>, but not of each of the lot values.)	1-47. The next criterion, and actually one which implies all the others, is that every element in the population must clearly belong to one, and only one, of the strata. Is this criterion met in our MNO Tool Company stratification plan? (YES/NO) EXPLAIN YOUR ANSWER: ____________________ ____
No Answer Required	2-30. After the preliminary sample has been selected, each of the chosen inventory lots is physically inspected, and the values of the individual items therein are added to give the lot value. These values are then tabulated, as in Volume 1 (see, for example, page S-12 in that volume). For teaching purposes we do not need to make up hypothetical values in a table. If we did, however, the table for each stratum would contain the random numbers, corresponding (actual) lot numbers, lot values, and the values squared. Why is this last item necessary?

No Answer Required - - - - - - - - - - - - - $$S.D. = \sqrt{\frac{\sum (X_j - \bar{X})^2}{N}}$$	1-6. At the left is the formula for the standard deviation of a population. (This is the "definitional" formula, not the short-cut "computational" formula used in Volume 1.) Consider just the term $(X_j - \bar{X})$. This denotes the numerical difference between any value in the population and the ________________ $(\bar{X})$ value of the population. (NOTE: Until we introduce the notation used in stratified sampling, the letters S.D. will stand for standard deviation.)
YES There are 2,000 lots in all, of which 1,500 are known to contain only standard items, and the remaining 500 only hand tools. The lots have been clearly labeled as containing either one kind of product or the other.	1-48. Finally, we must be sure that the strata are clearly named and defined on the basis of some tangible, specifiable difference between them. How have we rigorously defined these strata? ____ ____
The squared values are used in computing the estimated standard deviation of the stratum. (or similar answer)	2-31. In order to continue with the selection of the full sample (after the required size has been determined), you will also have to note, for each stratum, the ____________________________ in the random number table.

mean - - - - - - - - - - - - 5 5 5 5 5 5 30 $\bar{X} = 5$ $S.D. = \sqrt{\frac{\sum (X_j - \bar{X})^2}{N}}$	1-7. In the illustrative population at the left, the mean is clearly 5. The numerical difference between the mean and each of the six population values is, of course, 0. The sum of the squares of these six differences - in other words, $\sum (X_j - \bar{X})^2$ - is ___________, and if we worked out the whole formula, the standard deviation would be ___________.
standard items and hand tools (or similar answer - <u>not</u> "high value" vs. "low value")	1-49. To summarize, a stratification plan must meet these three criteria (no matter how many strata the population is broken into): 1. Every element must belong to one and only one stratum. 2. There must be a tangible, specifiable difference that defines and distinguishes the strata. 3. The exact number of elements in each stratum must be known in advance. (No Answer Required)
stopping point (or similar answer)	2-32. In this chapter, we began by tabulating some of the important data (Exhibit 6), and ended with the selection of the preliminary sample. In review, what do the following symbols refer to? A: ___________________________ U_R: ___________________________ N_i: ___________________________

0

0

- - - - - - - - - - - - -

1	9	1
5	20	2
21	34	3
56	47	5
75	68	96
100	80	151
S.D.=36.8	25.1	59.1

1-8. From inspection of the population in the preceding frame, it is obvious that there is no variability. To prove that the population has a "standard deviation of 0" is simply a mathematical way of stating this fact.

To give the reader a further "feel" for the mathematical concept of standard deviation, three more illustrative populations are listed at the left. All have the same mean, 43.0, but the distribution of the values results in different standard deviations.

(No Answer Required)

No Answer Required

1-50. In general, stratification can be made on the basis of either dollar values or kind of items. The latter does not apply only to inventory. For example, accounts receivable might be stratified on the basis of type of customer or age of account balance. Geographical differences might occasionally be a basis for stratifying a population.

Since the basic purpose of stratification is to reduce variability, the more common kind of stratification is on the basis of dollar values rather than kind of item. But in the MNO Tool Company example, we cannot stratify rigorously on a dollar-value basis. We do not have the information that would enable us to satisfy criterion (#1/#2/#3) in the preceding frame.

A: precision (of the total population estimate)

U_R: reliability coefficient

N_i: Number of elements in a stratum (or in the "i"th stratum.)

2-33. Suppose you had the task of specifying a route and starting point in the random number table. Would it make any difference whether you were taking an unrestricted or a stratified sample? (YES/NO)

No Answer Required	1-9. In an actual sampling problem, the population standard deviation is not computed. In order to do so, we would need to have definitive information about the population values -- which is exactly what we don't have when we decide to estimate by means of a sample. However, in order to use our statistical formulas, we need to estimate the population standard deviation. We do this by computing the standard deviation of a randomly-selected ____________ from the population.
#3 (If you were correct, the following frame may be skipped.)	1-51. Let us suppose that we stratified the MNO Tool Company as follows: (1) All lots with values less than $200; (2) All lots with values of $200 or more. Any lot would have to belong to one or the other of these two strata. This satisfies the first criterion. The difference between the two strata is tangible and specifiable, requiring no subjective judgment (as opposed to saying "high-value" and "low-value.") This satisfies the second criterion. However, we simply do not know which and how many of the lots have values of $200 or over, and which have smaller values. Criterion #3, therefore, cannot be satisfied. (No Answer Required)
NO	2-34. If we were taking an unrestricted sample from a population of 5,000 elements, we could use the random digits 0001 through 5000 (or any 5,000 consecutive digits) to establish correspondence. If the population were, instead, stratified, we could do the same, but we would also establish cut-off points (such as 1,500 in the MNO Tool Co. example) so that each stratum would have its own one-to-one correspondence. Using this method, we would have to go through the random number table: a. only once b. as many times as there are strata

sample $$S.D. = \sqrt{\frac{\sum (X_j - \bar{X})^2}{N}}$$	1-10. The formula at the left is the one we used in discussing the standard deviation of our small illustrative populations. When estimating the population standard deviation from a sample, there are two important differences: 1. The numerator is in a different form which is easier to work with. 2. The numerator is divided not by N, but by the sample size, n, minus one (n-1). (No Answer Required)
No Answer Required	1-52. We have, then, stratified the MNO Tool Company inventory on the basis of kind of item: Standard-item inventory lots and hand-tool lots. You may have noticed in Exhibit 5 that the two strata seem to have some overlapping values, in that the auditor suspects that there is at least one lot of standard items whose value is close to \$200, and at least one lot of hand tools close to \$150. Do you think this might make our stratum division inappropriate? (YES/NO) EXPLAIN YOUR ANSWER: ______________________ ______________________________________
a. only once	2-35. We have just reviewed some of the similarities between unrestricted and stratified sample selection. With respect to using the random number table, what is one way in which stratified sampling <u>differs</u> from unrestricted sampling? (In addition to recalling the phrase, explain what it means.) ______________________________________ ______________________________________ ______________________________________

No Answer Required	1-11. We have reviewed the concept of variability and, earlier in the chapter, discussed its relationship to sample size. This relationship can be summarized as follows: Given any population of size N, the lower the variability, the smaller the sample size required to achieve any given precision and reliability requirements. In the following frames, we will see how this principle applies to certain auditing situations. (No Answer Required)
NO (Explanation is in the next frame.)	1-53. An overlapping of values does not constitute a problem in this case, because the overlap does not conflict with any of the three criteria indicated in Frame 1-49. Moreover, the primary defining characteristic of the two strata was the <u>kind</u> of items contained in the inventory lots, not the <u>values</u> of the lots. Finally, as discussed earlier, the limits of $150 and $200 were only educated guesses so there might not actually be any overlap. (No Answer Required)
Stratified sampling is done <u>without</u> replacement. Once a random digit has been selected, it cannot be used again. (or similar answer)	2-36. We will now assume that preliminary samples have been selected for each stratum, and that the sample values have been totaled. This enables the auditor to make a preliminary estimate of the total population value. In Chapter 3 we will go over this procedure briefly and then begin the process of determining the required final sample size. END OF CHAPTER 2.

No Answer Required

1-12. In the teaching examples in Volume I, the populations were relatively homogeneous -- that is, composed of elements which were generally similar to one another. The "ABC Department Store" problem involved 9,000 accounts receivable of 90 days or over, in which a random sample ranged from $7 to $203 with an estimated standard deviation of $44.0. In order to select our sample, each account was assigned a number that would correspond to an entry in the random number table. This gave every element in the population an equal chance of being selected.

This method is known as (UNRESTRICTED/STRATIFIED) random sampling.

No Answer Required

1-54. Let us consider one final question concerning stratification. If the differences in the two components of the inventory are so clear-cut, why don't we simply make separate estimates of the standard-items inventory and the hand-tool inventory?

We could do this, but each of the estimates would have its own precision and reliability. Since the auditor desires an estimate of the total inventory value with a given precision and reliability, the two categories must be considered as:

a. separate populations

b. strata within one population

CHAPTER 3. SAMPLE-SIZE ALLOCATION

3-1. The most important part of this chapter has to do with the allocation of the total sample size among the strata -- or, in other words, the percentage of the total sample size which each stratum will contribute. This is a new concept which, unlike most of those covered in this book, has no counterpart in Volume I.

First, however, we will make a preliminary estimate of the total population value. In the MNO Tool Company example, what are we referring to when we say "total population value"? (Refer to Exhibit 5 and/or 6, if necessary.)

____

____

UNRESTRICTED	1-13. Now let us assume that of the 9,000 accounts, 500 of them, perhaps for special customers, had 90-day balances ranging between \$1,000 and \$10,000. An unrestricted sample of such a non-homogeneous population would present difficulties. The standard deviation of this population would probably be a few thousand dollars (as opposed to \$44.0 in the original example.) In the original problem, a sample size of 641 was sufficient to achieve a precision of \$20,000 with 80% reliability. In the new hypothetical example as stated above, the same sample size most probably (WOULD/WOULD NOT) be sufficient to achieve the same criteria.
b. strata within one population	1-55. Since the two strata (standard items and hand tools) are clearly part of one population, we are not dealing with "apples and oranges" but simply with two different kinds and sizes of apples. By the same token, we ordinarily (WOULD/WOULD NOT) use stratified sampling as a means of making a combined estimate of accounts receivable and inventory values.
total dollar value of the 2,000 inventory lots (or similar answer)	3-2. When working with an unstratified population, in order to estimate the total value we first computed the sample mean. How did we then compute the estimated total population value? (Answer either in words or in symbols.) ______________________________ ______________________________

WOULD NOT	1-14. In theory, even with the large standard deviation, we could compute the necessary sample size and select the sample using the unrestricted random sampling method. In practice, however, this might result in a sample size of such magnitude that it might be just as efficient to inspect all the accounts on a 100% basis. One alternative is to inspect all high-value accounts and add them up separately, while taking an unrestricted random sample of the 8,500 low value accounts. On statistical or on auditing grounds, can this method be ruled out? (YES/NO)
WOULD NOT	1-56. In review, which of the following criteria must be met in a stratified sampling plan? ____ Every element in the population must definitely be assigned to one, and only one, of the strata. ____ The total population must have a "normal" distribution. ____ The number of elements in each stratum must be known in advance. ____ There must be a tangible, specifiable, known difference between the strata. ____ The strata cannot contain elements with values overlapping those in other strata.
$\hat{X} = \bar{x}N$ or multiply the sample mean by the number of elements in the population (or similar answer)	3-3. When dealing with two or more strata, we do the same thing. The sample mean from each stratum is computed, and multiplied by the number of elements in the stratum. This results in the estimated stratum value. These values are added together to give us an estimate of the total population value. This procedure is tabulated in Exhibit 7 on page S-13. Locate this worksheet and keep it easily accessible. (No Answer Required)

NO (This was discussed briefly in Volume 1, Chapter 3.)	1-15. If, however, there are a large number of elements involved (such as in this case, where we have 500 high-value accounts), the auditor might prefer not to inspect them on a 100% basis, but rather to make a statistical estimate. He would then have two sub-total estimates, one for each of the groups, or <u>strata</u>, in the population. The two estimates can be added together to obtain the best estimate of the total population value. However, in order to evaluate the precision and reliability of a combined estimate from two (or more) strata, a formula has to be applied which is more complex than the formula for evaluating an estimate based on a single sample. (No Answer Required)
√ Every element in the population must definitely be assigned to one, and only one, of the strata. √ The number of elements in each stratum must be known in advance. √ There must be a tangible, specifiable, known difference between the strata.	1-57. The <u>shape</u> -- that is, the distribution of the population values -- may give some clue as to whether or not stratified sampling would be appropriate. Another possible indicator is the ___________ of population values.
No Answer Required	3-4. The instructions opposite Exhibit 7 spell out in more detail the procedures for which we gave an overview in the preceding two frames. Using them as a guide, together with the headings at the top, you will be able to complete Exhibit 7 without referring back to the text. First, however, let us review some of the notation that is used in Exhibit 7 and elsewhere in this book. The following symbols will be used in this book as they were in Volume 1. Next to each, indicate the meaning. Σ ______________________ n ______________________ $\bar{x}$ ______________________

No Answer Required

1-16. The previous frame summarized the method known as <u>stratified</u> random sampling. A more detailed overview, both of the method and of the volume, is given in the Summary of Stratified Sampling Procedures on page S-vii of the Supplementary Section. Read this Summary now for general familiarization purposes. We will be referring to it several times in this book.

(No Answer Required. Go on to Frame 1-17 after reading the Summary.)

---

range

1-58. What is the basic reason for using stratified sampling, when appropriate, rather than unrestricted sampling?

a. to obtain additional information about the major categories within the populatior

b. to make it feasible to sample a non-homogeneous population without requiring an inordinately large sample size

c. to give every element an equal chance of being selected in the population

---

$\sum$ = summation (of all terms following the symbol)

n = sample size

$\bar{x}$ = sample mean

3-5. Let us assume that the preliminary sample totals for the standard-item and hand-tool strata are, respectively, $3,300 and $21,000. These figures, together with the preliminary sample sizes, have already been entered in Exhibit 7.

Referring now to the instructions opposite the Exhibit (page S-12), fill in the top half of Exhibit 7 -- that is, the <u>preliminary</u> estimate of the total value of the MNO Tool Company inventory. Your goal is to complete the equation, $\hat{X}$ = ?, in Column 5.

No Answer Required	1-17. To summarize, the value of stratified sampling hinges on the fact that, although we cannot actually change the variability of any given population, we can break it up into smaller sub-populations, each of which will be more homogeneous than the original population. Each of the strata will therefore have a smaller standard deviation than that of the original population. As a result, the sample size will be smaller than if an unrestricted sample were taken; or, alternatively, the reliability would be higher or the precision limits (NARROWER/WIDER).
b. to make it feasible to sample a non-homogeneous population without requiring an inordinately large sample size	1-59. In this Chapter we have discussed the rationale and principles involved in breaking out a population into strata. We will come back to this subject in Chapter 6, and will continue to work out the MNO Tool Company problem in Chapters 2 through 5. END OF CHAPTER 1
$\bar{x}$ N_i $\bar{x}N_i$ \$110 1,500 \$165,000 \$700 500 <u>\$350,000</u> $\hat{X}$ = \$515,000	3-6. If we wanted to estimate the population <u>mean</u>, how would we go about it? a. divide the estimated total (\$515,000) by the combined N (2,000) b. average the two means (\$110.0 and \$700.0) After circling your choice, do the computation and enter the result in the appropriate space in Exhibit 7.

NARROWER

1-18. Since we are not using actual population data, but simply presenting hypothetical result we cannot <u>prove</u> that stratified sampling would be better in any given case than unrestricted sampling. There is actually no general formula that can substitute for the auditor's judgment in deciding whether or not a stratified sample is appropriate in any given case.

As an aid to applying this judgment in actual problems, we will examine, in the following frames, the characteristics of a population that might influence this decision.

(No Answer Required)

CHAPTER 2. PRELIMINARY SAMPLE SELECTION

2-1. In stratified sampling, the actual mechanics of selecting a random sample are almost identical to those in unrestricted sampling. In this chapter, therefore, we will review the basic principles of random sample selection, but will not go through the entire procedure.

A random number table is not necessary for this chapter, but may be useful in refreshing your memory. Either the two-page teaching table in Volume I, or any published table, may be referred to whenever the reader thinks it will be helpful.

(No Answer Required)

---

a. divide the estimated total ($515,000) by the combined N (2,000)

The figure $257.5 should be entered next to $\hat{X}$ in Exhibit 7. If you were incorrect, make the correction in the Exhibit.

3-7. The population mean computation is not actually necessary in this example since our estimate will be of the population total, not of the mean. However, it does illustrate the important concept of the "weighted mean" which is useful in many applications. Summarizing this example, the sample mean for Stratum 1 is $110.0; for Stratum 2, $700.0. The weighted mean is $257.5. This is much closer to the Stratum 1 mean because ________

____

____

No Answer Required	1-19. The <u>range</u> of the population -- that is, the interval between its lowest and highest value -- is often an important factor in deciding whether or not to stratify a population. As the preceding discussion may indicate, a stratified sample might be relatively efficient (compared to an unrestricted sample) when the range is relatively (SMALL/LARGE).
No Answer Required	2-2. In Chapter 1, we decided to stratify the MNO Tool Company into standard-item inventory lots and hand-tool inventory lots. Before selecting the samples for these two strata, it is advisable (as in Volume 1) for the auditor to record his sampling plan decisions and all available data. For this purpose, re-read Exhibit 5 (page S-9) and then turn to Exhibit 6 (page S-11) and keep it readily accessible. (No Answer Required)
Stratum 1 has a larger number of elements. (or similar answer)	3-8. We now have an estimate of the total population value, but without any idea of how much this may differ from the true value. Before we can know that, we must know how large a sample is required, and in order to compute that we must first compute the estimated ______________________________ of each stratum.

LARGE	1-20. Although a wide range will often suggest the possibility of stratified sampling, the shape of the population, irrespective of range, also is a factor to be considered. Exhibit 1 (page S-1 of the supplementary section) lists four hypothetical populations. The values in all of them range between approximately \$0 and \$200. Basing your answer on the same kind of inspection and analysis as in similar examples in Volume 1, which two populations in Exhibit 1 do you think have relatively high variability a. A and B b. C and D
No Answer Required	2-3. We will go through most of the Exhibits and Worksheets all at once so that the reader does not have to turn back and forth between the supplementary material and the text. Exhibit 6, however, will be used this time not only to record the data but also to review notation and some of the points made in Chapter 1. We will therefore go through this Exhibit a few steps at a time rather than all at once. For the moment, focus just on the three "Auditor's Decisions" and ignore the notation column. The three items of information called for (ARE/ARE NOT) the same as if we were taking an unrestricted, rather than stratified sample of the population.
standard deviation	3-9. For each stratum, the standard deviation is estimated in the same manner as if the stratum were a single population. We will not review the computation procedure, since it was covered in detail in Volume I. Instead, we will simply assume that the estimated standard deviations of Strata 1 and 2, respectively, turn out to be \$36 and \$205. These figures may have somewhat more meaning to the reader if he re-reads the numerical data in Exhibit 5. (No Answer Required)

b. C and D (This is explained in the following frames.)	1-21. In Exhibit 2, (page S-3), the same data are portrayed graphically. The values in Population A are distributed in the familiar "bell-shaped" or "normal" curve. Without going into technicalities, you are probably aware that in general, random sample means taken from a normal distribution generally will tend to approximate the true population mean with smaller sample sizes than are necessary from non-normal populations (other factors being equal). (No Answer Required)
ARE	2-4. Based on the information in Exhibit 5, fill in the first three rows of Exhibit 6. (In specifying the quantity to be estimated, it is important to include the number of elements in the population, and what these elements are.)
No Answer Required	3-10. Having computed the estimated standard deviation of each stratum, we do not need to estimate the standard deviation of the population as a whole. As a review of your understanding of Chapter 1, if we did happen to know the population standard deviation, it would undoubtedly be: a. smaller than \$36 b. between \$36 and \$205 c. greater than \$205 EXPLAIN YOUR ANSWER: ______________________ ______________________________________

No Answer Required	1-22. The reason for this is that in a "normal" distribution, there are approximately equal amounts of very high and very low values. Moreover, most of the values in such a population, as illustrated in Population A, tend to concentrate: a. toward the mean b. at the extremes (Circle the choice you think correct.)
QUANTITY TO BE ESTIMATED: Total inventory value (N = 2,000 inventory lots) DESIRED PRESISION: ± $20,000 DESIRED RELIABILITY: 95%	2-5. In the second column of Exhibit 6, the letters A and R, standing for precision and reliability respectively, are familiar to you from Volume I. The capital letter X, standing for the true population value, is rarely used because at all times we are concerned with the <u>estimate</u> of this true value. What, in general, is the symbol that indicates an estimated value? ____________ What, therefore, would be the symbol for the estimated value of the population? ______________
c. greater than $205 (Explanation is in the first paragraph at the right.)	3-11. In a stratified population, the stratification is made so as to yield a lower variability in each of the strata than in the population as a whole. Readers who did not get the correct answer together with at least an approximation of the explanation, should read the review of the standard deviation concept (Frames 1-5 through 1-8) if they did not do so earlier. It would also be helpful to skim through Frames 1-12 through 1-15. (No Answer Required)

a. toward the mean (For example, in Population A, over 50% of the population values are within $\pm$ \$30 of the mean. You may wish to compute the equivalent percentages for the other three populations.)	1-23. These characteristics are significant in determining what kind of sampling technique to employ. In Volume 1, we studied only unrestricted random sampling. What is the basic principle of sample selection involved in this technique? ____ ____ ____
$\wedge$ (caret) $\hat{X}$	2-6. The next three items in Exhibit 6 are not actually needed at this time. However, They will be used frequently in later computations and it is helpful to record them now so they can be referred to easily. The reliability coefficient (U_R) and its square are found in Exhibit 32 on page S-63. The square of the desired precision can be readily computed. Fill in these items in Exhibit 6.
No Answer Required	3-12. The symbol for the estimated standard deviation of a stratum is S_i. This is an exception to the general rule that estimated values are indicated with a caret. (No Answer Required)

Every element in the population has an equal chance of being selected in the sample. or Every sample of a given size that could possibly be selected has an equal chance of being selected. (or similar answers)	1-24. Since an unrestricted random sampling plan allows each element an equal chance of selection, we assume that in the long run the laws of chance will operate so that if high-value or low-value items appear with equal frequency within the population, they (WILL/WILL NOT) tend to "average out" in the sample as well as in the population.
U_R: 1.96 U_R^2: 3.84 A^2: $400,000,000	2-7. In the remaining part of this worksheet, room for four strata has been provided in case the auditor wants to use this sheet, or a facsimile thereof, for an actual problem. However, there are only two strata in the MNO Tool Company problem. Based on the discussion in Chapter 1, enter the precise definition of each stratum in Exhibit 6, together with the total number of elements (N_i) in each stratum. (NOTE: In this example, and in the second example in this volume, we will arbitrarily denote the low-value, high-N stratum as "Stratum 1.")
No Answer Required	3-13. Given that S_1 = $36 and S_2 = $205, S_1^2 = $1,296 and S_2^2 = $42,025. (In actual practice, as you may recall, S_i^2 is computed first and the square root of that is taken to yield S_i). After S_i^2 and S_i are computed, the figures should be entered in the Data Sheet (Exhibit 6, Page S-11). Do this now. (While doing so, look once again at the N_i figures, since we will be using them shortly.)

WILL	1-25. Population B, like Population A, has approximately an equal number of high and low-value items. We can, therefore, assume that in an unrestricted random sample, they would tend to appear in equal numbers, thereby not distorting the estimate of the population mean. Referring to Exhibit 2, which of the two populations do you think would require a <u>larger</u> sample to achieve the same degree of accuracy? (A/B) EXPLAIN YOUR ANSWER: ______________________ ____
Stratum 1: All standard-item inventory lots N_i = 1,500 Stratum 2: All hand-tool inventory lots N_i = 500	2-8. Why were the strata defined in this fashion, rather than in terms of dollar value ranges? ____ ____ ____
<table><tr><td></td><td>S_i</td><td>S_i^2</td></tr><tr><td>Stratum 1</td><td>36</td><td>1,296</td></tr><tr><td>Stratum 2</td><td>205</td><td>42,025</td></tr></table> (All figures are dollar values.)	3-14. We have one more symbol to introduce: P_i. This is a figure which we will be working with in the rest of this chapter. P_i stands for the <u>percentage</u> of the total sample size that will be allocated to each stratum. It is expressed as a two-place decimal. For example, let us suppose, arbitrarily, that the total required sample size in the MNO Tool Company problem is 200. If 80 of these sample elements are selected from the standard-item stratum (Stratum 1), then P_i would equal 40% or .40. The second stratum would contribute 120 out of the 200 elements, so that P_2 would equal ______________.

B Any of the following is correct: B has more extreme values -- fewer of its elements have values tending to the mean -- higher variability.	1-26. Population C is in the pattern known as a "skewed" distribution. Fifteen per cent of the values are $120 and over, yet these comprise about 37% of the total population value. In this kind of distribution there is some chance that a sample mean would give a distorted estimate of the population mean. If we were to take a random sample (with replacement) of 10 elements from this population, there is a fair chance that <u>none</u> of the high-value items would be selected. (No Answer Required)
In this example, we do not know in advance precisely how many elements are in any given dollar-value range. (or similar answer)	2-9. The subscript "i" is the one new notational feature discussed so far. This is used to refer to individual strata. Thus, N_1 = the size of the first stratum, N_2 the size of the second stratum, and in general, N_i would refer to the size of the ______________ stratum.
.60 (120 ÷ 200)	3-15. The foregoing example illustrates the meaning of the term P_i, but does not reflect the actual procedure. We do <u>not</u> arbitrarily decide on our total sample size or on the P_i values. For an overview of this portion of the sample-size allocation and determination procedure, read steps 8 through 10 in the Summary at the beginning of the Supplementary Section (page S-ix).

No Answer Required	1-27. Naturally, with a population of only 40 items, we would not use statistical estimation to begin with. Assuming, however, that a larger population had a similarly "skewed" distribution, and that we were estimating its total value by means of an <u>unrestricted</u> random sample, how would we attempt to eliminate the potential distortion? a. Select a large enough sample size so that we could be statistically confident that our estimate meets the desired precision and reliability. b. Add up all the high-value items separately and take a random sample of the remaining items.
"i"th	2-10. The remaining three columns will be discussed in Chapter 3. The data which you have entered will be referred to frequently in Chapters 3 through 7. Any reference to the "data sheet" refers to Exhibit 6 for the MNO Tool Company, or Exhibit 20 for the JKL Corporation. (No Answer Required)
No Answer Required	3-16. As implied in what you have just read, and as we will show in the next chapter, once we have determined the percentage of the total sample that will be allocated to each stratum, we can mathematically compute the <u>total sample size</u> that is required in order to make an estimate of the population value at the desired precision and reliability. Then, given the total sample size, and the <u>percentages for each stratum</u>, we can easily compute the actual <u>number of elements</u> required from each stratum. (No Answer Required)

a. Select a large enough sample size so that we could be statistically confident that our estimate meets the desired precision and reliability. (If you were correct, skip the next frame.)	1-28. YOUR ANSWER: b. Add the high-value items separately. This is a good technique to use with skewed distributions, so your answer should not be considered altogether wrong. However, the question specified that unrestricted random sampling was to be used. Therefore, you cannot treat some of the items in a different manner from the others. All must be sampled, with equal chances of being selected. In an unrestricted random sample, the only way to guard against potential distortion is to have a large enough sample size. (No Answer Required)
No Answer Required	2-11. Once the stratification plan has been adopted, and the basic data recorded, each stratum is sampled separately as if it were a population in itself. In order to determine how many elements we will ultimately have to select, we must first estimate the (STANDARD DEVIATION/TOTAL VALUE) of each stratum.
No Answer Required	3-17. One method of allocating the total n among the strata is known as proportional allocation. In this method, the percentage of the sample allocated to each stratum is the same as the percentage of the total population N accounted for by that stratum. For example, with the MNO Tool Company inventory, assume that our total sample size were 200. Using proportional allocation, we would select ____________ standard-item lots and ____________ hand-tool lots.

No Answer Required	1-29. Population D, illustrating the "U-shaped" distribution, is the least applicable of all to unrestricted random sampling. If we were to take a random sample of 10 elements from the population of 40, there is some chance that we would select, say, eight of the extremely high or extremely low values. As with any population, we could diminish the chances of a "freak" result by having a large enough sample size, but it might turn out to be so large as to make it an inefficient technique. (No Answer Required)
STANDARD DEVIATION (The total value will later be estimated, but not for the purpose of determining sample size.)	2-12. In some auditing situations, the standard deviation of a population or stratum can be inferred from knowledge of similar populations and/or past experience. Usually, however, the standard deviation is estimated by computing the sample standard deviation, as already discussed. As discussed in Volume 1, we can assume that the standard deviation computed from the sample data is the best estimate of the population (or stratum) standard deviation, provided that the sample has been randomly selected and contains at least _________ elements.
150 50	3-18. In this book, however, we will use an allocation method known as <u>optimal</u> <u>allocation</u>. This method takes into account the variability of each stratum, as well as the number of elements. Exhibit 8 in the supplementary section (page S-15) is designed to aid the auditor in computing optimal allocation. Locate this exhibit and keep it easily accessible. (No Answer Required)

No Answer Required	1-30. In an actual case, of course, the dollar values of all the elements are not known, or not readily ascertainable in advance. Often, however, there are certain characteristics of a population that give us a clue as to how the population is distributed in terms of dollar values. And, if the population is itself clearly composed of certain major sub-populations, with differing characteristics, the auditor may wish to sample them separately no matter how the dollar values happen to be distributed. (No Answer Required)
30	2-13. The same procedure is used in stratified sampling as in unrestricted sampling. Since we are estimating the standard deviation not of the entire population but of each individual stratum, our preliminary sample in the MNO Tool Company problem should consist of at least: a. 30 elements from each stratum (Turn to Frame 2-14) b. 60 elements in all, but not necessarily 30 from each stratum (Turn to Frame 2-15)
No Answer Required	3-19. Exhibit 8 enables the auditor to compute the proportion (P_i) of the total sample that will be allocated to each stratum. The only data that we need, N_i and S_i, can be found in the Data Sheet (Exhibit 6, page S-11). Referring to the instructions opposite the Exhibit and the headings at the top of the columns, fill out Exhibit 8.

No Answer Required

1-31. Now let us see how this kind of analysis can be applied to an auditing example. Read Exhibit 5 in the Supplementary Section (page S-9).

As far as you can tell from the data presented, which of our illustrative populations does the MNO Tool Company inventory most resemble?

Population A. (If this is your answer, turn to Frame 1-32.)

Population B. (Turn to Frame 1-33.)

Population C. (Turn to Frame 1-34.)

Population D. (Turn to Frame 1-35.)

2-14. YOUR ANSWER: 30 elements from each stratum

Correct. In the preliminary sample stage, each stratum is treated as if it were a population in itself. In order to estimate the standard deviation of a population, or in this case of a stratum considered as a population, at least 30 elements are required.

SKIP NOW TO FRAME 2-16.

N_i	S_i	N_iS_i	P_i
1,500	36	54,000	.35
500	205	102,500	.65
		156,500	1.00

3-20. We have computed the P_i figure for the Standard Item stratum to be .35. What does this mean?

a. 35% of the total sample will come from the Standard Item stratum.

b. 35% of the Standard Item stratum will be sampled.

1-32. YOUR ANSWER: Population A

No. In Population A, as in any "normal" distribution, there are roughly equal numbers of high and low-value items, and most of the values tend to concentrate towards the middle.

In the MNO Tool Company, it seems likely judging from the data, that 25% of the inventory lots contain a good deal more than 25% of the total value.

With this as a clue, return to 1-31 and select another answer.

2-15. YOUR ANSWER: 60 elements in all, but not necessarily 30 from each stratum

No, although in Chapter 3, we will see that the final sample need not contain equal representation from each stratum. At present, however, -- that is, in the preliminary sample stage -- each stratum is treated as if it were a population in itself. In order to estimate the standard deviation of a population, or in this case of a stratum considered as a population, at least 30 elements are required.

(Go on to Frame 2-16.)

---

a. 35% of the total sample will come from the Standard Item stratum.

3-21. Computations are made in exactly the same manner when there are more than two strata. Using optimal allocation, compute the percentage of the total sample size that will be allocated to each stratum when the population contains three strata whose N_i and S_i are, respectively, 1000 and \$50, 800 and \$150, and 400 and \$200.

1-33. YOUR ANSWER: Population B

No. In Population B there are roughly equal numbers of high, low, and intermediate values. This does not seem to be the case with the 2,000 inventory lots of the MNO Tool Company.

Picture how these 2,000 lot values might appear on a graph, and then return to Frame 1-31 and choose another answer.

2-16. We will use the random number table to make our selection, just as in unrestricted sampling. Indeed, at this point we <u>are</u> using unrestricted random sampling, but of each stratum, not of the entire combined population.

Reviewing from Volume 1, what are three steps that must be taken (in order) before using the random number table?

1. ____________________

2. ____________________

3. ____________________

---

1. $N_i = 1{,}000$ $S_i = \$50$
   $N_iS_i = \$50{,}000$ $P_i = .20$
2. $N_i = 800$ $S_i = \$150$
   $N_iS_i = \$120{,}000$ $P_i = .48$
3. $N_i = 400$ $S_i = \$200$
   $N_iS_i = \$80{,}000$ $P_i = .32$

$\sum N_iS_i = \$250{,}000$

3-22. These computations, resulting in P_i figures for each stratum, enable us to apportion the total sample size among the strata -- provided the total n is known. However, this is not the case at this point. For illustrative purposes, then, assume that the population at the left requires a total sample size of 500 elements. Compute the number of elements that will be selected from each stratum.

1: ____________

2: ____________

3: ____________

1-34. YOUR ANSWER: Population C

Correct. In Population C, 15% of the elements contained more than twice that percentage of the total dollar value.

In the case of the MNO Tool Company, it would appear that the hand tools, although representing only 25% of the inventory lots, might comprise a much higher percentage of the total dollar value of the inventory. Another way of looking at it is to picture the distribution of lot values on a graph. This would probably turn out to be "skewed" (to the left) as in Population C.

SKIP NOW TO FRAME 1-36.

---

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

2-17. Since these procedures are basically the same as in Volume I, we will not go through the entire sequence of selecting random numbers. Instead, simply refresh your memory of these procedures by doing the exercise in Exhibit 3 in the Supplementary Section (page S-5).

(Answer is in the <u>second</u> column below. Ignore the last two columns for the moment.)

P_i	n_i	N_i	n_i/N_i
.20	100	1,000	
.48	240	800	.30
.32	160	400	
1.00	500	2,200	

3-23. We are now in a better position to see why stratified sampling is usually done without replacement, especially when optimal allocation is used. Often, a stratum with a relatively small N_i will have a high enough variability to make its n_i relatively large. This will lead to a result similar to that in the previous example. If the total sample size is 500, the third stratum contributes 32% or 160 elements. This represents _____% of its <u>own</u> stratum N, given as 400. Even the first stratum, which has the lowest variability of all, will contribute a sample of 100 elements, representing ______% of its own size.

1-35. YOUR ANSWER: Population D

No. Population D has a large and approximately equal number of high and low values, with virtually nothing in between. The MNO Tool Company inventory might look similar to that if it had 1,000 lots of precision instruments, 1,000 lots of standard items, and only a few lots of hand tools.

Refer again to Exhibit 5, and visualize the distribution of inventory lot values as they might appear on a graph. With that in mind, return to Frame 1-31 and choose another answer.

(Answer to Exhibit 3 exercise is in Exhibit 4, page S-7.)

2-18. The exercise in Exhibit 3 referred only to the first stratum. Assuming that you have numbered those lots 0001 through 1500, how would you number the lots in the second stratum?

a. 1501 through 2000, and select the two samples simultaneously (Frame 2-19).

b. 001 through 500, and select the two samples simultaneously (Frame 2-20).

c. Use any 500 consecutive numbers, and take the second sample after the first has been selected (Frame 2-21).

40% $(160 \div 400 = .4)$

10% $(100 \div 1,000 = .1)$

3-24. When the number of elements in a random sample amounts to at least 10% of the population or stratum from which the sample is taken, sampling is often done without replacement. This is usually the case in a stratified samplin problem, as illustrated by the example just completed and the teaching examples in this volume.

In order to compute our P_i values, we first had to compute the estimated standard deviation of each stratum. In stratified sampling, do we need to estimate the overall population standard deviation? (YES/NO) If YES, when do we do this

____

____

1-36. Let us briefly review why a skewed distribution, as in Population C or the MNO Tool Company inventory, is relatively unsuitable for unrestricted random sampling. Check as many reasons as you think correct.

_____ Unrestricted random samples may not include the proportionate amount of high-value items.

_____ Unrestricted random samples might contain too many high-value items.

_____ In order to assure a representative sample, a relatively large number of elements must be selected.

2-19. YOUR ANSWER: 1501 through 2000, and select the two samples simultaneously.

All three choices would lead to valid results, but this is probably the best. This way you only have to go through the random number table once, yet there is no possibility of confusion. By the same token, if there were a third stratum of, say, 750 elements, you would use the numbers 2001 through 2750 to establish correspondence with that stratum.

SKIP NOW TO FRAME 2-22.

NO

3-25. Also in review, number the following steps 1 through 4 in order. (One has been done to give you a start.)

_____ compute the required overall sample size

__1__ compute the estimated standard deviation of each stratum

_____ compute P_i values for each stratum

_____ compute individual stratum sample sizes

All are correct, but the last is the most general.

1-37. Let us return to our inventory example. (Refer, if necessary, to Exhibit 5.)

We need certain information -- total dollar value -- about the population as a whole, but do not wish to take an unrestricted random sample because of the skewness in the population values. In this example, would it be practical to "weed out" a few extreme values, add them separately, and take an unrestricted random sample of the remainder?

a. YES (turn to Frame 1-38)

b. NO (turn to Frame 1-39)

2-20. YOUR ANSWER: 001 through 500, and select the two samples simultaneously.

You are correct in assuming that there is nothing to prevent you from taking the two samples simultaneously, and with <u>extreme</u> care, this method would work. However, the danger of confusion is great.

Return to Frame 2-18 and select another answer.

The list should read, in order:

3
1
2
4

(corresponding to steps 9, 7, 8 and 10 in the Summary on page S-ix.)

3-26. In this chapter, we have discussed the preliminary estimate of the population value, the estimate of the stratum standard deviations, and optimal allocation of the sample size among the strata.

In review, what does the term "P_i" stand for? (A brief answer such as "percentage" would not be considered sufficient.)

1-38. YOUR ANSWER: Yes

No. If anything, the data given would lead to the opposite conclusion. For one thing, we don't yet know any individual lot values. In order to "weed out" extreme values, we would have to inspect all 500 hand-tool lots. This would defeat the purpose of sampling.

Secondly, even if we knew that there were, say, only a dozen lots of more than $1,000, we would not solve the problem by adding them separately because the rest of the population would remain skewed.

SKIP NOW TO FRAME 1-40.

2-21. YOUR ANSWER: Use any 500 consecutive numbers, and take the second sample after the first has been selected.

You are correct in saying that any 500 consecutive numbers may be used. However, there is no need to go through the random number table two times, as this answer implies. True, with a preliminary sample of only 30 elements in each stratum, the extra time is negligible; however, with a large sample (and samples of 1,000 or higher are by no means uncommon) this method might not be the most efficient. Return to Frame 2-18 and select another answer, bearing in mind, however, that this choice is not to be considered wrong.

P_i for any stratum is the percentage of the total sample size that will come from that stratum.

(or similar answer)

3-27. Without referring to Exhibit 8, place a check next to those figures that are needed to compute P_i, using the optimal allocation method that we have employed.

_____ stratum size

_____ total sample size

_____ estimated population standard deviation

_____ estimated stratum standard deviations

1-39. YOUR ANSWER: No

Correct. We have no way of knowing what the individual lot values are until we check them all. Moreover, even if we knew beforehand that only a few extreme values existed, and added them separately, we would still be left with basically a skewed distribution.

2-22. In general, then, the elements of the population are numbered 1 through N, or, since there is no need to start with 1 all the time, any N consecutive numbers may be used. This is no different from unrestricted sampling.

As a final observation with respect to correspondence, this system does not require the auditor to take his samples simultaneously, but simply makes it possible for him to do so if he desires.

(No Answer Required)

__√__ stratum size

__√__ estimated stratum standard deviations

3-28. Assume that a population is divided into three strata, with sample data as follows:

N	$\sum x_i$	n
800	\$1,500	30
500	\$12,000	30
200	\$150,000	30

The best estimate of the total population value is \$__________.

	1-40. Although it is not feasible in this example to add up the extreme values separately, some kind of stratification is indicated. In general, a population can be stratified on the basis of difference in dollar values in two or more groups, or on the basis of a difference in kind of items. (Other bases of stratification, such as geographical location, will not be discussed in this volume but the procedures would not differ.) (No Answer Required)
No Answer Required	2-23. Using unrestricted random sampling as taught in Volume I, when we come across a number for a second (or more) time we (IGNORE IT/USE IT AGAIN). This procedure is known as sampling (WITH/WITHOUT) replacement.
\$1,240,000 (If you were incorrect, try again using Exhibit 7 as a guide.)	3-29. In the same problem, assume that the standard deviations of the three strata are \$25, \$314, and \$2,400. What could we assume about the size of the standard deviation of the population as a whole? ____ ____

No Answer Required	1-41. If we were to stratify on the basis of kind of items, we would have two strata: all standard-item lots and all hand-tool lots. If we were to attempt to stratify on the basis of dollar values we would run into some problems, as shown in the next few frames. Re-read the next-to-last paragraph of Exhibit 5 and then go on to 1-42. (No Answer Required)
USE IT AGAIN WITH replacement (i.e. the number is replaced back into the pool of usable numbers.)	2-24. Counting the same element more than once requires a larger sample size than would otherwise be necessary. However, this does not present much of a problem in unrestricted sampling. Usually, the required sample size -- n -- is only a small fraction (less than 10%) of the population size. Thus, as you may recall from the sampling examples in Volume I, in an unrestricted sample with replacement, although there are some repeats, there are relatively (FEW/MANY).
It would probably be greater than $2,400. (or similar answer)	3-30. In this chapter, we estimated the total inventory value of the MNO Tool Company to be $515,000. Stratum standard deviations were given as $36 (N = 1,500) and $205 (N = 500.) Although we did not do the computations, these figures were hypothetically estimated from the sample data, in which n for each stratum = 30. We computed P_1 and P_2 to be .35 and .65. In Chapter 4, we will use these results to compute the total required sample size. END OF CHAPTER 3

No Answer Required	1-42. Although in this example, the auditor could stratify by dollar value, such a plan would involve so many difficulties as to be impractical. Let us suppose, for example, that we had one stratum consisting of all inventory lots with values of $200 or less; and another consisting of all inventory lots with values of over $200. Based on the facts in Exhibit 5, do we know how many elements would be in each of these two strata? (YES/NO)
FEW	2-25. In stratified sampling, however, there are often cases in which the sample size of that stratum -- n_i -- is a large proportion of the stratum size. For example, it would not be unusual to require a sample size of, say, 225 elements in a stratum of only 400. In such a case, there would be a very large number of repeats. (No Answer Required)
	CHAPTER 4 BEGINS ON PAGE 45. TURN THE BOOK OVER AND CONTINUE AS BEFORE.

NO	1-43. To be sure, it is known (or strongly suspected, that most of the 1,500 standard-item lots would be in the less-than-$200 stratum, and most of the 500 hand-tool lots in the over-$200 stratum. But in order to make a statistical estimate of the total dollar value, the <u>exact</u> number of elements in the population must be known. With stratified sampling, we actually make separate estimates for each stratum. Therefore, we (DO/DO NOT) need to know the number of elements in each stratum, as well as in the overall population. TURN BACK TO PAGE 3, ROW 2.
No Answer Required	2-26. In the same hypothetical example -- or indeed, in any example -- if we sampled <u>without</u> replacement, there would be, by definition, <u>no</u> repeats. We would therefore use a (SMALLER/LARGER) sample size than if we were sampling <u>with</u> replacement. TURN BACK TO PAGE 3, ROW 3.

Chapter 4. DETERMINATION OF SAMPLE SIZE

4-1. In Chapter 3, on the basis of a preliminary sample consisting of 30 elements from each stratum, we made an estimate of $515,000 for the total value of the MNO Tool Company inventory.

At this point, can we state whether or not this estimate is based on a sample size large enough to satisfy our precision and reliability criteria? (YES/NO)

No Answer Required

5-21. By use of the "divide and average" method, we have discovered that the square root of $365,789,549 is $19,126 (to the nearest dollar). What does this tell us about our estimate of the MNO Tool Company inventory value?

____

____

____

6-28. YOUR ANSWER: NO, he would have to go ahead as originally planned.

This incorrect answer may have been based on the correct principle that one should not tamper with a randomly-selected sample. This would be the case, for example, if the auditor discarded some extreme-value elements from his sample and replaced them with new selections. You can, however, always <u>increase</u> your random sample size to make it more likely that you will meet your precision and reliability criteria -- we do this when we add 10% -- and you can always decide to sample on a 100% basis.

(Follow up the correct answer in Frame 6-27, page 82.)

NO	4-2. In Volume I, for teaching purposes, we evaluated the preliminary estimate -- in other words, we determined what the precision of the estimate would be at our desired reliability level. Only very rarely, however, will the preliminary sample be of sufficient size to enable us to satisfy our precision and reliability requirements. In review, what is the primary purpose of a preliminary sample? ____________________ ____________________
The estimated value does not differ from the true value by more than $19,126 in either direction. (or similar answer)	5-22. The statement at the left is missing an important qualification. What is it? ____________________ ____________________ ____________________
	6-29. As a separate issue from the previous question, suppose that we have selected 30 accounts from the first stratum (under $1,000). After substantiating them and recording their values, we discover that one of them has an actual value of $1,020 instead of its trial-balance value of $920. What should we do with this account? a. Consider it now to be part of the second stratum (Frame 6-30) b. Keep it in the first stratum and proceed normally (Frame 6-31)

To estimate the standard deviation (of a population, in unrestricted sampling; of the strata, in stratified sampling)	4-3. Based on the estimated stratum standard deviations, and the number of elements in each stratum, we then computed ________________ ____.
The reliability of this estimate is 95% (there is a 5% chance that the statement is untrue, and that the estimated value differs from the true value by more than the reported precision). (or similar answer)	5-23. Earlier in this chapter (Frame 5-14 and Exhibit 7), we computed the best estimate of the total population value to be $520,000. We can now state that the true value of the 2,000 lots of inventory of the MNO Tool Company is between $_________ and $_________. We can be _____% confident that this statement is correct; or, phrasing it the other way around, the probability that the true value is <u>not</u> in this range is only _____%.
	6-30. YOUR ANSWER: Consider it now to be part of the second stratum. No. The stratification plan cannot be altered once you have taken a sample. You would have to start all over again, and take a completely new preliminary sample from both strata. Moreover, what would you do if this happened again with another element, as it well might? This problem can be solved as shown in Frame 6-32. Skip to that frame now. (No Answer Required)

the percentage of the total sample size that will be allocated to each stratum (Pi) (or similar answer)	4-4. Having done so, we can now compute the required total sample size. We will begin with a brief review of the relationship between precision, reliability and sample size. Begin by defining these two terms: <u>PRECISION</u> <u>RELIABILITY</u>
\$500,874 and \$539,126 95% 5%	5-24. This concludes the problem. In the following frames, we will review some of the highlights, including material from other chapters. For maximum teaching and reviewing value, answer as fully as possible. The usual lines have been left out to give you more room. Refer freely to the Supplementary Section material. Why did the auditor take a stratified, rather than unrestricted, sample of the MNO Tool Company inventory? (The problem is set forth in Exhibit 5, page S-9).
	6-31. YOUR ANSWER: Keep it in the first stratum and proceed normally. Correct. To do otherwise would be impractical, as shown in the preceding frame. However, if <u>many</u> accounts proved to be in a different range than expected, it could well be that there was some error in the original assumptions underlying the stratification plan. The auditor might want to reconsider the entire problem, on auditing as well as statistical grounds.

PRECISION: range of values, expressed as a + or - difference from the estimated value, within which the true value is expected to lie.

RELIABILITY: degree of confidence, expressed as a percentage, that the true value is contained within the precision limits.

(or similar wording)

4-5. (If you were correct, and feel quite sure of your understanding of the basic concepts in Volume 1, you may skip to Frame 4-9.)

Explain what will happen to the reliability of an estimate if we narrow down the precision limits without changing our sample size:

____

____

____

Your answer should contain most of the following points:

overall distribution skewed...non-homogeneous population...large population variability... each stratum would have much lower variability... sample size required might be impractically large if unrestricted sampling were used

5-25. Other things being equal, stratified sampling normally is most efficient when the population can be stratified by dollar value. Why was the MNO Tool Company stratified by kind of item, rather than by dollar value?

6-32. We can get around the potential problem of a few items turning out to fall within different ranges than expected. Instead of defining our strata in terms of substantiated amounts, we define them in terms of trial-balance amounts.

Thus, we will define our strata as (1) all acounts with <u>trial balance amounts</u> between \$1 and \$999; (2) all accounts with <u>trial balance amounts</u> between \$1,000 and \$9,999; and (3) all accounts with <u>trial balance amounts</u> over \$10,000.

(No Answer Required)

The reliability will decrease (e.g. from 90% to 85%) because we are reducing, or "zeroing in", on the range of values expected to contain the true value. (or similar answer)	4-6. By the same token, how does the precision of the estimate change if we increase the reliability without changing the sample size? ______________________________ ______________________________ ______________________________
dollar-value range of each lot was not definitely known...each element could not be definitively assigned in advance to a stratum...in this example, stratifying by kind of item would be almost equivalent to stratifying by dollar value since there is such a strong correlation between type of product (Standard Items or Hand Tools) and dollar value of the inventory lots	5-26. We recorded our stratification plan (Exhibit 6) and took a preliminary sample. How many elements were in the preliminary sample, and why?
No Answer Required	6-33. To be safe, then, we should define the strata in terms of the ranges of the trial balance values, rather than the substantiated values. However, in the exercise to follow shortly and in the worksheet headings, this qualification can be taken as implicitly understood, so that we can refer to stata 1, 2 and 3 respectively as "$1-999," "$1,000 - 9,999," and "$10,000 and over." To sum up what has been done so far in the JKL Corporation problem, turn to the Summary (page S-vii) and read steps 1 through 3. Keep the Summary easily accessible in the next few frames. (No Answer Required)

The precision limits become wider. We are claiming greater confidence that the estimated range contains the true value; therefore we have to widen that range. (or similar answer)	4-7. When sampling for auditing purposes, of course, we do not simply take a sample and then compute the precision and reliability of the estimate. Instead, we establish our desired precision and reliability in advance. This is done on the basis of: a. statistical formulas b. auditing factors such as materiality and reasonableness c. both a. and b.
30 from each stratum We need to estimate the standard deviation of each stratum. In order to do so, by means of a random sample, at least 30 elements are required.	5-27. We used the random number table to select our samples. The route and starting point were established just as in unrestricted sampling (Exhibit 3). We established correspondence by using the numbers 0001 through 2000, with the number 1501 standing for the first element in the second stratum, and so on. The sample was taken <u>without</u> replacement. Other than the points mentioned above, the mechanics of selecting a random sample do not differ in any way from unrestricted random sampling. (TRUE/FALSE) If you said FALSE, why?
No Answer Required	6-34. The first step is to establish desired precision and reliability. For the JKL Corporation, these were given in Exhibit 18 as ± $200,000 and 98%. The second step, defining the strata unambiguously, has been discussed at length both here and in Chapter 1. We have seen some of the potential problems, all of which, however, can be solved by using appropriate judgment from an auditing as well as statistical point of view. Which type of stratification is more likely to reduce overall variability? a. stratification by dollar value b. stratification by kind of item

b. auditing factors such as materiality and reasonableness (See Appendix 1 of Volume 1 for a full discussion of this point.)	4-8. We therefore have to select a sample size large enough to enable us to report, with a stated __________________ percentage, that the true value does not differ from the estimated value by more than the stated _______________________.
TRUE	5-28. After estimating the stratum standard deviations, we used the optimal allocation method to allocate the total sample size among the strata. This is done (BEFORE/AFTER) the total sample size is computed.
a. stratification by dollar value	6-35. The third step given in the Summary is to determine which, if any, of the strata is to be sampled 100%. Other things being equal, the auditor might lean towards 100% sampling if the values within the stratum are relatively (HIGH/LOW) in magnitude, (HIGH/LOW) in variability, and if the stratum has a (LARGE/SMALL) N.

reliability precision	4-9. We are now ready to compute the required sample size. This will be done entirely with the use of one worksheet. Although in number of text pages this is a very short section, in actual time required it may take about 30 minutes to use the worksheet to solve the equation. However, with repeated use, especially when a calculating machine is employed, the auditor will be able to solve this equation in only a few minutes. In order to get a better grasp of the purpose of this set of computations, and their place in the overall procedure of stratified sampling, the reader is advised to re-read steps 1 through 9 in the Summary on page S-vii. (No Answer Required)
BEFORE	5-29. Exhibit 8 is used to compute P_i values, taking into account the variability as well as the size of each stratum. If P_i turns out to be, say, .42, this means that: a. 42% of the total sample will come from the first stratum. b. 42% of the first stratum will be sampled.
HIGH HIGH SMALL (These terms are, of course, relative, and judgment is required in each individual case.)	6-36. We can now summarize the stratification plan in the Data Sheet. Fill in Exhibit 20, page S-39, referring, if necessary, to Exhibit 18. It may also help to refer to the Data Sheet for the MNO Tool Company (S-9). You can check your answer in the next frame. (NOTE: Leave the S_i and S_i^2 columns blank.)

No Answer Required	4-10. Now locate Exhibit 9 (page S-17) and keep it easily accessible. For the time being, focus only on the equation at the top of the worksheet. All the right hand terms are known, and are easily obtainable from the Data Sheet (Exhibit 6). This leaves us with one unknown, __________, which stands for __________ ______________________________.
a. 42% of the total sample will come from the first stratum.	5-30. After computing our P_i values, we determined the total sample size (Exhibit 9), and then computed the individual stratum sample sizes and added 10% (Exhibit 11). This brought us up to this chapter, in which we selected the additional elements and made a new estimate of the population total and the stratum S_i^2 values. The next, and final, step was to evaluate the new estimate (Exhibits 14 and 16). State fully and precisely the meaning of the word "evaluate" in this context.
Answer is in the frame at the right.	6-37. Your Data Sheet (Exhibit 20) should have the following information: X: Total value of Government accounts receivable (N = 1,800) A: $\pm$ \$200,000 R: 98% U_R: 2.33 U_R^2: 5.43 A^2: \$40,000,000,000 STRATUM 1: \$1-999 N_1 = 1,200 STRATUM 2: \$1,000-9,999 N_2 = 500 STRATUM 3: \$10,000 and over N_3 = 100 (Figures refer to trial balance amounts.)

n number of elements in the required sample size (or similar wording)	4-11. In the first Column of Exhibit 9, the statistical notation shows you what point in the equation you are working on. It is not expected or necessary for you to be able to define precisely what this notation refers to. Auditors who have little or no interest in the mathematical basis of stratified sampling may ignore this column and simply use the second column to guide them. You may also refer to the Instructions opposite the Exhibit. Before beginning Exhibit 9, read the next frame. (No Answer Required)
"Evaluate" means to determine the precision of the estimate at the desired reliability level.	5-31. In the next chapter, we will study an accounts-receivable sampling problem which involves stratification by dollar value (rather than by kind of item), and also involves 100% sampling of one of the strata. These two new aspects, although important from the auditing point of view, actually make no difference in the statistical procedures. Therefore, after we discuss and establish the sampling plan in Chapter 6, you will have the opportunity in Chapter 7 to go through the entire problem virtually unaided, enabling you to review and test the knowledge gained so far. END OF CHAPTER 5.
(Make corrections, if necessary in Exhibit 20.)	6-38. In this chapter, we have explored, at somewhat more length than in Chapter 1, the judgmental factors involved in making a stratification plan. There were several facets of the JKL Corporation that did not apply to the MNO Tool Company -- among them, 100% inspection of a stratum, stratification by dollar value, and a breakout into three strata rather than two. We will see in the next chapter, however, that once the stratification plan has been determined, the statistical procedures will prove to be virtually identical to those you have already employed. END OF CHAPTER 6.

No Answer Required	4-12. Since a wrong computation at the beginning will lead to errors throughout the worksheet, you are advised to check your computations one or two rows at a time. For this purpose, use Exhibit 10 (page S-19), which is identical to Exhibit 9 with the answers filled in. If you are using a machine, your final digit in some of the answers may be different from the printed answer. The latter is correct, having been rounded beyond the capacity of your machine. When that happens, correct the last digit and use that answer as the basis for subsequent computations. Now do Exhibit 9 in its entirety. Then go on to Frame 4-13.
	CHAPTER 6. DOLLAR-VALUE STRATIFICATION and 100% INSPECTION OF A STRATUM 6-1. In Chapter 5, we concluded our analysis of the MNO Tool Company problem by arriving at an estimate of the total inventory value within the desired limits of precision and reliability. In this chapter, we will review the concepts and reasoning behind our approach, and apply the same stratified sampling procedure, with one important variation, to an accounts receivable example. (No Answer Required)
	CHAPTER 7. CONCLUSION AND REVIEW 7-1. In this chapter we will work out the JKL Corporation example and have some review questions covering the entire stratified sampling procedure. To get an overview once again, the reader is advised to scan the Summary of Stratified Sampling Procedures (page S-vii) now, and to refer to it frequently, especially at the conclusion of each major step. (No Answer Required)

Check your answer in Exhibit 10. If the final result is correct, the intermediate computations need not be individually checked.	4-13. We have just computed the sample size that is required to provide an estimate that will vary no more than plus or minus $20,000 at 95% reliability. Before we select our additional elements, two more steps are necessary. First, without actually doing it, indicate how you would ascertain the number of elements required from each individual stratum: ______________________________ ______________________________
No Answer Required	6-2. We have seen that in a non-homogeneous population, a stratified sample might reduce the sample size that would be required in an unrestricted sample. It might mean (depending on the population in question) reducing the size by a relatively small amount, as from 200 to 180. In such a case, it might be just as well to take an unrestricted sample. In the above hypothetical example, suppose we took a stratified sample of 200 elements. In that case, we could have a (WIDER/NARROWER) precision or a (HIGHER/LOWER) reliability than with an unrestricted sample from the same population. (Assume that stratification lowers the overall variability.)
No Answer Required	7-2. In Chapter 6, we decided to stratify the Government accounts receivable of the JKL Corporation on a dollar-value basis, and to sample the high-value stratum on a 100% basis. We will not review the mechanics of sample selection, but will assume that 30 accounts from each of the first two strata have been randomly selected. Hypothetical results are given in Exhibit 19, page S-37. Using Exhibit 21, make a preliminary estimate of the total population value, based on the data in the top half of Exhibit 19.

Multiply the total sample size of 193 by the P_i values of each stratum (.35 and .65 respectively).

4-14. After doing that, we will add 10% to the required sample in each stratum. Reviewing from Volume I, why is this advisable?

____

____

____

NARROWER

HIGHER

6-3. Still focusing on the phrase "reduce the required sample size," an alternative (and probably more common) meaning is "to make feasible and practical a statistical sample of a non-homogeneous population, when such would not be the case using an unrestricted sampling plan."

For example, suppose that in a highly variable population of 5,000 elements, an unrestricted sample would require 4,400 elements to achieve given criteria of precision and reliability. Such a sampling plan would be feasible, in the sense that it could be done and all formulas would apply, but it might not be practical because ________________________.

(Answer is in the frame at the right)

7-3. ANSWER TO EXHIBIT 20 (JKL Corp.)

x_i	n_i	$\bar{x}_i$	N_i	
$15,000	30	$500	1,200	$ 600,000
156,000	30	5,200	500	2,600,000
3,000,000	(Computations not required)			3,000,000
				$\hat{X}$=$6,200,000

to guard against an increase in the estimated standard deviation of the strata (to make sure that the sample size will be sufficient)

(or similar answer)

4-15. Both these operations - computing the individual stratum sample sizes, and adding 10% - are performed in Exhibit 11 on page S-21. Do this worksheet now. Your answer can be checked in Exhibit 12. (Unlike the previous worksheet, you will probably not have to check until you have finished.)

ANY OF THE FOLLOWING:

sample is about 88% of population...might be more practical to examine on a 100% basis... stratifying the population might considerably reduce the required size.

6-4. Before applying the reasoning in Frame 6-3 to any given set of population data, what must the auditor specify first?

a. The strata into which the population might fall, clearly defined so that each population element would belong to one and only one stratum. (Turn to Frame 6-5.)

b. The quantity to be estimated, together with desired precision and reliability. (Turn to Frame 6-6.)

c. Both a. and b. (Turn to Frame 6-7.)

$6,200,000

7-4. We have just estimated the total value of the Government accounts receivable to be $6,200,000. This estimate differs from the trial balance total (Exhibit 18) by $315,000. Barring the possibility of mistake in computing the sample totals, this difference could possibly reflect large discrepancies in the trial balance figures, or could also be the result of normal sampling error. The <u>true</u> value might be anywhere between:

a. $6,200,000 and $6,515,000

b. $6,000,000 and $7,000,000

c. we cannot tell at this time

Answer is in Exhibit 12, page S-23.	4-16. We have determined that the final sample will consist of 74 elements from the standard-item stratum, and 138 elements from the hand-tool stratum. The next step, which for teaching purposes we need not actually do, is to select _______ additional elements from the former and _______ from the latter.
	6-5. YOUR ANSWER: The strata into which the population might fall . . . Probably not. Initially, the auditor decides that he wants to estimate the total value of a certain body of auditing data -- the "population" -- at precision and reliability levels based on his judgment of materiality and reasonableness. These goals are specified first. There is no point in mentally dividing the population into strata until it has been determined that stratification would be helpful. Choice b. was therefore preferable. SKIP TO FRAME 6-8.
c. we cannot tell at this time	7-5. The previous question asked you, in effect, to give the precision limits of an estimate that had not yet been evaluated. Statistically, this is a contradiction in terms. The purpose of the preliminary estimate, then, is not to tell us anything about the true value, but rather as a rough check to see if we are "in the ballpark." Had the estimated total turned out to be, say, half a million dollars, we might conclude that there was an error of procedure or arithmetic. (No Answer Required)

44 (74 - 30) 108 (138 - 30)	4-17. In making additional selections, the auditor would be careful not to make any change in his correspondence plan or route specification. What else would he guard against (which was not necessary in unrestricted sampling)? ________________________________ ________________________________
	6-6. YOUR ANSWER: The quantity to be estimated together with the defined precision and reliability. Correct. This is the primary item of interest to the auditor. Stratification is a <u>method</u> for estimating the total value with a smaller sample than might otherwise be necessary. Whether to stratify, and how to stratify, both depend on the distribution of values within the population. The auditing <u>goals</u> are usually specified first. SKIP TO FRAME 6-8.
No Answer Required	7-6. Exhibit 19 gives hypothetical standard deviations. Transfer them to your JKL Data Sheet (Exhibit 20) and also compute and enter the S_i^2 values.

using a number more than once

(or similar answer)

4-18. In the next chapter, we will assume that the additional selection has been made, and will evaluate the sample results in terms of precision and reliability. That will conclude the MNO Tool Company example.

END OF CHAPTER 4.

6-7. YOUR ANSWER: Both

Not exactly. It is true that in many cases, it is obvious beforehand that certain strata are indicated by the nature and distribution of the population. Logically, however, it is first necessary to decide what the auditor wishes to estimate, then to determine if stratified sampling will be helpful, and finally to decide what kind of stratification would result in the lowest sample size without altering precision and reliability requirements.

Return to Frame 6-4 and follow up the correct answer.

	S_i	S_i^2
Stratum 1	$240	$57,600
Stratum 2	1,200	1,440,000

7-7. The estimated standard deviations, based on our hypothetical preliminary sample data, turned out to be $240 (Stratum 1) and $1,200 (Stratum 2). Given these figures, you can use Exhibit 22 on page S-43 to compute the P_i value for each stratum. Do this now.

CHAPTER 5. EVALUATION OF RESULTS

5-1. In Chapter 4, we computed the sample size for the MNO Tool Company inventory. Once this has been done, the procedure for making a revised estimate and determining its precision and reliability is quite similar to that covered in Volume I for an unrestricted random sample.

For an overview of this process, read Steps 11 through 14 in the Summary on page S-ix.

(No Answer Required)

6-8. Given desired precision and reliability, and with the population size known, the only remaining determinant of required sample size is population variability. There are, however, certain indicators of variability. Assume two populations of equal size. If the ranges are equivalent, then the one with (NORMAL/SKEWED) distribution will probably have greater variability. If the shapes are similar, then the one with (NARROW/WIDE) range will more likely have greater variability.

(NOTE: You may refresh your memory by looking again at Exhibit 2 on page S-3.)

Answer is in the frame at the right.

7-8. The results should have been obtained from Exhibit 22 as follows:

	N_i	S_i	N_iS_i	P_i
Stratum 1.	1,200	$240	$288,000	.32
Stratum 2.	500	1,200	600,000	.68
			$888,000	

What, exactly, do the above computations tell us about the allocation of our sample size?

____

____

No Answer Required	5-2. For teaching purposes, we need not go through the process of selecting the additional elements, but we will discuss some of the important points relating to sample selection. The first point to notice is that the stratum with fewer elements (500 vs. 1,500) actually contributes more elements to the sample (138 vs. 74). This will frequently happen when the optimal allocation method is used, because it takes into account the ___________ of the strata as well as their sizes.
SKEWED WIDE	6-9. With the foregoing discussion in mind, consider the accounts receivable of a hypothetical firm. They could be classified (in the ordinary sense of the word, not necessarily in the sense of "stratified") by age, amount, type of customer, etc. Suppose they range from \$100 to \$100,000 with a population standard deviation of \$20,000. If the accounts were stratified by type of customer, each stratum might still have approximately the same range, shape, and standard deviation as the total population. If this were the case, stratified sampling (WOULD/WOULD NOT) necessarily reduce the required sample size.
32% of the total sample size will come from the first stratum (\$1-999); 68% from the second stratum. (It would <u>not</u> be correct to say, "32% of the first stratum will be sampled.")	7-9. The phrase "total sample size" refers to the first two strata only. Suppose, for example, that we sample 55 elements from the first stratum and 71 from the second. We are also taking all 100 accounts from the third stratum, so that in terms of our total population, we would be substantiating 226 accounts. Our "total sample size," however, in the sense just indicated, would be ___________ elements.

variability (standard deviation)	5-3. This results in a major difference between unrestricted and stratified sampling. Let us suppose that we were taking an <u>unrestricted</u> random sample from this population. With 212 elements to be selected, and 2,000 elements in all, each element in the population would have a probability of selection equal to a little more than ______%.
WOULD NOT	6-10. Let us now assume that the population is stratified by dollar value of the accounts, rather than by type of customer, with the following data known: Number (N_i) — Range — S_i 1,000 — \$100 - \$999 — \$ 200 400 — \$1,000 - \$9,999 — \$ 1,700 25 — \$10,000 - \$100,000 — \$14,000 It makes no difference what the actual numbers and values are. In general, any stratification by dollar value (MUST/NEED NOT) necessarily result in a lower <u>range</u> for each stratum than for the population as a whole.
126 (55 + 71)	7-10. In working with a population containing a stratum that has been sampled 100%, the procedure is as follows: 1. Treat the non-100% strata as if they comprised a population in themselves. In other words, we will make and evaluate an estimate for the \$1-999 and \$1,000-9,999 strata just as we did for the two strata in the MNO Tool Company 2. Given a total-value estimate for those two strata, with precision computed at the desired reliability level, we will add the 100% stratum total to that estimate. The precision and reliability of the overall estimate will remain the same. (No Answer Required)

10% (10.6% to be exact)	5-4.

	n_i	N_i
Standard Items	74	1,500
Hand Tools	138	500

Clearly, this is not the case in this example of stratified sampling. Any given element (inventory lot) in the second stratum has a 27.6% chance of being selected (138 ÷ 500). Any given element in the first stratum has approximately only a _______% chance of being selected.

MUST

(If in doubt, simply compute the stratum ranges and compare them with the population range of $99,900. This kind of analysis would apply to any set of real or hypothetical data.)

6-11. The lower range would most probably indicate lower variability. Moreover, through prior knowledge of the auditing population (i.e. the accounts receivable), the auditor would probably have a good idea as to the distribution of values within each of the categories.

Without going into either the judgmental or theoretical aspects in greater detail, it is safe to say that stratification by dollar value is, in general, mathematically the most efficient kind of stratification. Why, then, might some other basis of stratification sometimes be used? ______________________________

____

No Answer Required

7-11. We have therefore computed P_i values only for the first two strata. We can now determine the total sample size required from these two strata, using Exhibit 23 on page S-45. Fill in this Exhibit now, checking your computations in Exhibit 24.

5% (74 ÷ 1500 = .049)	5-5. In stratified sampling, then, the principle that "every element in the population should have an equal chance of selection in the sample" does not necessarily apply. However, within each stratum every element (SHOULD/SHOULD NOT) have an equal chance of being selected in the sample for that stratum.
dollar value ranges are not always readily ascertainable (or similar answer)	6-12. Look once again at the data in Frame 6-10. This hypothetical (but rather common) distribution indicates another advantage of stratifying by dollar amount. Most of the range and variability is contributed by the high-value stratum which contains only 25 elements. If these 25 accounts were inspected 100%, rather than statistically sampled, most of the potential sampling error in the total population could be eliminated. This will be discussed and illustrated later in this chapter. (No Answer Required)
n = 97 (Exhibit 24)	7-12. The next step, as you may recall or as you can verify in the Summary, is to compute the individual stratum sample sizes and add 10%. Do this now, using Exhibit 25 and checking in Exhibit 26.

SHOULD	5-6. Therefore, considering each stratum as if it were an individual population, we continue to select the additional elements by unrestricted sampling in each stratum. The suggested procedure is to continue using the same correspondence and route as in selecting the preliminary sample. In the random number table, the route resumes where it left off in selecting the preliminary sample. (No Answer Required)
No Answer Required	6-13. Although we have been emphasizing <u>range</u> in the past sequence, range is at best only a possible indicator of, not actually a statistical measure of, variability. The reason for discussing range at some length is that stratifying by dollar value automatically implies stratifying by range -- that is, \$0 - \$1,000, \$1,001 - \$10,000, etc. A stratum could also be established as an "open-ended" range -- for example, "accounts over \$10,000." In the same population, we (COULD/COULD NOT) also have another open-ended stratum such as "accounts over \$5,000."
n_1 = 34 n_2 = 73 (Exhibit 26)	7-13. We have to select 4 additional elements from the first stratum, and 43 from the second stratum. Assume that this has been done, the accounts in question have been substantiated, and the resulting data from the combined samples (preliminary plus additional) turn out as given in the bottom half of Exhibit 19 on page S-37. Read this Exhibit and then look at Exhibit 27 to see how these data are combined with the preliminary data (page S-53). (No Answer Required)

No Answer Required	5-7. Assume that we have selected enough additional random numbers, without duplication, to bring the total up to 212. The next step is physically to inspect each of the 74 lots containing standard items, and the 138 lots containing hand tools. Within each lot, the value of each individual inventory item is recorded. This is the only time (except for the identical procedure after the preliminary sample) that we concern ourselves with individual items, since our "sampling elements" -- that is, the 212 units that make up the total "n" -- are (LOT/ITEM) values.
COULD NOT (since accounts over $10,000 would belong to two strata.)	6-14. As we have seen, when stratifying by dollar value, the number of elements in each range has to be known, or readily ascertainable, in advance. We must also know which items are in each range. This follows from the basic principle: Every element in the population must definitively be assigned, in advance, to one and only one of the strata. This has to be the case in order to take an unrestricted random sample within each stratum. If the exact composition of each stratum were not known, we could not establish ___________ between the elements in the stratum and the random digits.
No Answer Required	7-14. The already-completed Exhibit 27 comprises Step 12 in the Summary. You were not required to do the calculations. All that remains is to enter the new S_i^2 values in your Data Sheet (Exhibit 20 on page S-39). Do this now.

LOT	5-8. For teaching purposes, we will now assume that the auditor has added up the individual elements in each of the 212 lots selected in the final (preliminary plus additional) sample for each stratum. The next step is to record the data and compute the new sample means and estimated S_i^2 figures. To see how this is done, turn to Exhibit 13 on page S-25. (No computations will be necessary.) (No Answer Required)
correspondence	6-15. In the case of the MNO Tool Company we decided not to stratify by dollar value, for the reasons indicated in the preceding frame and in Chapter 1. Now let us suppose that the MNO Tool Company had just taken a complete physical inventory of all 2,000 lots, and that the statistical estimate was to be made as an additional verification. The already-computed figures for each lot (COULD/COULD NOT) be used to assign each lot to a stratum based on dollar value.
You should have entered \$62,500 and \$1,345,600 under S_i^2 in Exhibit 20.	7-15. In the remaining exercises, for additional self-testing purposes, you will not be told which worksheet to use, or where to find the source data. You may, however, refer to the Summary if necessary. Compute the final estimate of the total value of the 1,800 Government accounts receivable of the JKL Component Corporation.

No Answer Required	5-9. The figures in Rows 1 and 2 represent the sums of the sample lot values, these sums having been arrived at as described in Frames 5-7 and 5-8. The figures in Rows 7 and 8 represent the sum of the squares of these lot values. The other computations are explained in the "Source" column. They are identical to the combined-sample computations in Volume 1, except that we are now working with data from two strata rather than one population. At this point in an unrestricted sampling problem we would re-estimate the population standard deviation. Do we also need to do this in a stratified sampling problem? (YES/NO)
COULD	6-16. A similar situation might exist in an accounts receivable problem. This is illustrated in Exhibit 18 on page S-35, with which we will be working extensively. Read this Exhibit and then go on to Frame 6-17. (No Answer Required)
You should have used Exhibit 21, with results as shown at the right.	7-16. FINAL ESTIMATE, JKL Corp. (Exhibit 21) 1. \$18,020 34 \$ 530 1,200 \$636,000 2. 401,500 73 5,500 500 2,750,000 3. (not needed) 3,000,000 $\hat{X}$ = \$6,386,000

NO (As discussed in Frame 3-24, this figure is not needed.)	5-10. We did need to estimate the <u>stratum</u> standard deviations, but do not need to do so again. These figures are required for only one purpose, namely: a. Estimating the total of the stratum values b. allocating the total sample size among the strata
No Answer Required	6-17. Before answering the two questions posed at the end of Exhibit 18, let us note first that these data apply only to the 1,800 Government accounts of the JKL Corporation. What about the manufacturing accounts? Although nothing is said directly, we can deduce, from information in the Exhibit and from principles discussed elsewhere in this book, that they are to be considered as a separate (STRATUM/POPULATION).
No Answer Required	7-17. Now use Exhibit 28 (page S-55) to evaluate this estimate at the desired reliability level, up to but not including the point at which you have to take the square root of A^2. Check your answer in Exhibit 29.

b. allocating the total sample size among the strata	5-11. In statistical language, S_i is used only for the purpose of computing P_i by the optimal allocation method. In determining the total sample size, as you can verify by referring to Exhibit 9, we use S_i^2 rather than S_i. Similarly, in evaluating the estimate, which we will do shortly, we use only S_i^2 and do not have to compute S_i. (No Answer Required)
POPULATION (If you were correct, you may skip to Frame 6-19.)	6-18. From an auditing point of view, it would not seem incorrect to consider the Government and manufacturing accounts to be part of the same "population." However, statistically, this word refers to "the body of data about which the auditor wants to obtain information by means of a statistical estimate." The second sentence of Exhibit 18 indicates that an estimate is required for the Government accounts irrespective of the others. The auditor may wish to treat the manufacturing accounts in an entirely different manner. (No Answer Required)
A^2 = \$35,335,605,551 Exhibit 28; answers in Exhibit 29. (Result for A given in next frame.)	7-18. In the first teaching example (MNO Tool Company) we evaluated the results up to this point, then discussed the procedure for finding the square root, and finally computed the precision (A). In practice, however, you can go right on to the square-root extraction, thus concluding the problem. If you have not already done so, then, solve for A using Exhibit 30, and enter the result "for the record" in the bottom row of Exhibit 28.

No Answer Required	5-12. At this point, the preliminary estimates of S_i and S_i^2 are no longer of any use. In your Data Sheet (Exhibit 6), cross them out, so that you will not use the old figures by mistake in later computations. Then, in the "new S_i^2" column, insert the figures from Row 12 of Exhibit 13.
No Answer Required	6-19. The decision to make a complete substantiation of the 100 accounts of more than $10,000 actually involves three decisions. 1. Making a separate stratum of these accounts, rather than merging them with the 500 accounts of $1,000 to $9,999. 2. Establishing the cut-off point at $10,000 rather than some other figure. 3. Substantiating <u>all</u> 100 accounts, rather than sampling them. These points are discussed in the following frames. (No Answer Required)
A = $187,978	7-19. Given the total estimate of $6,386,000, we can conclude that at 98% reliability, the true value of the JKL Corporation's Government accounts receivable is somewhere between $__________ and $__________.

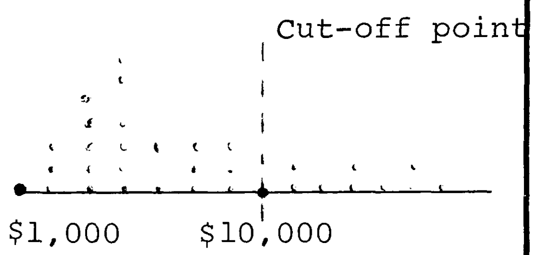
STRATUM	S_i^2 (new)
STRATUM 1	$1,444
STRATUM 2	$40,804

5-13. Summarizing, Exhibit 13 displays the work that the auditor would do after he has taken his additional sample from each stratum. The new sample means are computed in Row 5, and the S_i^2 figures are computed in Row 12.

The above is summarized even more succinctly in Step 12 of the Summary to which you can refer now (page S-ix).

(No Answer Required)

No Answer Required



N=500 N=100

(Assume that all 600 accounts would fall into generally the same patterns.)

6-20. By making a separate stratum of the $10,000+ accounts, the auditor transforms a skewed distribution into two which are much less skewed. This is shown at the left. As a result, each stratum on either side of the cut-off point has a narrower range and (GREATER/LESS) variability than the combined population of all 600 accounts.

$6,198,022 and
$6,573,978

7-20. The precision at 98% reliability was computed using the data from the first two strata only. The estimated sub-total for these two strata, as you could compute from the data in Exhibit 21, is $3,386,000 ± $187,978

When we add in the third stratum, our total estimate is $6,386,000 with the <u>same</u> precision. This can be explained in two ways. You may recall from Volume 1 that precision is a measure of sampling error -- the inevitable difference between the true value and the value estimated by means of a statistical sample. Is there any sampling error associated with the third stratum? (YES/NO)

No Answer Required	5-14. We have gone through Step 12 in the Summary, and are now ready to make our final estimate of the total population value. Do so, using the lower half of Exhibit 7, page S-13. Before beginning that exhibit, read Instruction 6 on page S-12. The answer to Exhibit 7 is in the next frame.
LESS	6-21. Establishing the cut-off point at $10,000 may result from the auditor's approach on previous examinations. The policy might be that all items of over a certain amount -- $10,000 in this case -- must be checked 100% rather than sampled. This is one possible reason for decision #3 in Frame 6-19. (No Answer Required)
NO (It was not statistically sampled, but inspected 100%.)	7-21. We can also see what would happen if we included the third stratum in our other computations. Turn to Exhibit 28. If we were to fill in data for the third stratum, what would be the result in Rows 5, 6, and 7? __________

$\bar{x}$	N	$\bar{x}N$
$112	1,500	$168,000
$704	500	$352,000
		$\hat{X}$ = $520,000

(The first two columns are not needed.)

5-15. We have estimated the total inventory value of the MNO Tool Company to be $520,000, but do not know what precision we could claim for this estimate at our desired reliability level. The final step, then, is to evaluate this estimate. For this purpose, turn to Exhibit 14 on page S-27, but for the moment, focus only on the equation at the top.

This equation is actually the same as the sample-size equation in Exhibit 9. However, the P_i terms do not appear in the evaluation equation. We have already used them to compute n_i. (This computation may be reviewed in Exhibit 11.)

No Answer Required

(Note: "100% check," "100% sample," and "100% inspection" are synonymous.)

6-22. The auditor's approach to the problem may also be based partly on the magnitude of the accounts in this stratum. From a statistical point of view, the magnitudes of population or stratum values have nothing to do with the reliability of the estimate. However, from an auditing point of view, the auditor may well reason that a trial-balance error in even one of these 100 accounts may have so large an effect on the population estimate that it would be better to inspect this stratum 100%.

(No Answer Required)

Answer is in the frame at the right.

7-22. Since N-n is 100-100 or 0, Rows 6 and 7 would also be zero. The figure in Row 8 would remain the same as it is, and the final result would obviously be (DIFFERENT/THE SAME).

No Answer Required	5-16. Another point to be made in connection with the evaluation equation concerns the relationship between reliability and precision. (This discussion is optional; readers interested only in the solution may skip right now to Frame 5-18.) Our desired reliability is 95%, so that U_R = 1.96 (Exhibit 32, page S-63). Suppose now that in the equation, everything to the right of U_R is computed to equal \$10,000. The precision (A) would therefore equal 1.96 x \$10,000, or \$19,600. Our precision requirement of $\pm$\$20,000 (WOULD/WOULD NOT) be satisfied.
No Answer Required	6-23. Another factor which may indicate 100% inspection is the size of the stratum. There are only 100 items in the stratum. If we were to sample them, we would first have to take a preliminary sample of at least _______ items in order to estimate the stratum standard deviation, leaving us with only _______ additional accounts to substantiate.
THE SAME	7-23. The next exercise is one which you have not yet done, but you have seen several examples of it. With respect to the total value of the Government accounts receivable, write a one-paragraph report to the JKL Corporation without using the terms "precision" or "reliability."

WOULD	5-17. We can now illustrate with an arithmetical example the familiar point that as reliability increases, the precision limits become wider (other factors being equal.) In the previous example, assume that our desired reliability was 99%. U_R would therefore be 2.58, as you can verify in Exhibit 32. If the quantity to the right of U_R in the evaluation equation is still assumed to turn out to be $10,000, our computed precision would equal ±$___________, and our requirement (WOULD/ WOULD NOT) be satisfied.
30 70	6-24. We can apply the same reasoning to the 500 accounts ranging between $1,000 and $10,000. Once again, we start by assuming that we will have to sample at least 30 accounts in order to estimate the stratum standard deviation. This will leave us with 470 (as opposed to only 70 in the other stratum). Secondly, the magnitude of the account values is much lower than in the other stratum. Both these points indicate that compared to the high-value, low-N stratum, it is (MORE/ LESS) reasonable to take a sample rather than to substantiate every account.
Answer is in the frame at the right.	7-24. The suggested answer below is based on the wording in previous examples, and is not meant to substitute for the auditing language which you would consider appropriate. "The estimated total value of the 1,800 Government accounts receivable is $6,386,000. There is a 98% chance that the value is between $6,198,022 and $6,573,978. There is a 2% chance that the true value is outside the limits just stated." (No Answer Required)

$25,800 (2.58 x $10,000) WOULD NOT	5-18. We are now ready to determine the precision of our estimate, using Exhibit 14. This worksheet is done in the same manner as the others, using the instructions at the left and the directions in the second column. The first column need not be considered unless you are interested in relating each individual computation to its place in the equation. Do Exhibit 14 (page S-27), checking your computations as often as you desire in Exhibit 15.
MORE	6-25. At this point, then, we can conclude (in response to the first question posed in Exhibit 18): 1. The factors that indicate a 100% substantiation for the third stratum -- namely, the high magnitude (over $10,000) and the low number (100) apply much less, or not at all, to the second stratum. 2. No matter what, we would begin by taking an unrestricted random sample of 30 of the 500 accounts. 3. The same reasoning applies even more to the remaining stratum, in which there are 1,200 accounts with values of under $1,000.
No Answer Required	7-25. This concludes the JKL Corporation example and the volume on stratified sampling. In doing an actual field problem, you probably will not have to refer back to the programed text, but you will need your Supplementary Section. You can use either of the two sets of case-study Exhibits as a basis for making up your own worksheets. You will also need Exhibit 32 (the U_R table) and the Summary of Stratified Sampling Procedures. You will then be equipped to make statistical estimates using the techniqu of stratified random sampling. END OF VOLUME 3

(For answer, refer to Exhibit 15, page S-29.)	5-19. We now have to find the square root of $365,789,549. Although any square-root extraction method may be used, the suggested method, especially when large numbers are involved, is to use the "divide and average" method. Exhibit 16 (page S-31) is provided for this purpose. You need not be already familiar with this method, since the worksheet has detailed instructions like the others. Do Exhibit 16, checking your answer in Exhibit 17. Then enter the result in the last row back in Exhibit 14.
No Answer Required	6-26. Let us now examine a few contingencies that may arise after the preliminary sample has been analyzed. Suppose that the second stratum turns out to have an extremely high estimated variability -- so high that when we do our sample size computations, it develops that, say, 385 elements are required from this stratum to achieve desired precision and reliability. Could we decide, at this point, to dispense with the statistical sample of this stratum, and then substantiate all 500 accounts? YES (Frame 6-27 on the next page) NO (Frame 6-28 in Row 3, page 45)

The figure $19,126 should be entered in the last row of Exhibit 14. Individual computations leading up to this result are in Exhibit 17.

5-20. There would be no point in averaging $19,126 and $19,125 to get $19,125.50, since all our other computations were rounded to the nearest dollar. This is why one of the instructions to the worksheet states that when two numbers are only one dollar apart, simply take the higher rather than averaging the two.

(No Answer Required)

TURN BACK TO PAGE 45, ROW 2.

6-27. Your Answer: YES, we could change our mind and substantiate all 500 accounts.

Correct. This is not to say that it necessarily should be done. However, the auditor might reason that since he is required to sample 77% of the stratum anyway (385/500), he might just as well examine this stratum on a 100% basis. If he did so, he would cut down on sampling error, thereby enabling him to sample fewer accounts from the first stratum and still achieve his original precision and reliability requirements.

(Skip to Frame 6-29 on the third row of page 46.)