8

Sampling Methods and the Central Limit Theorem

GOALS

When you have completed this chapter, you will be able to:

1 Explain why a sample is often the only feasible way to learn something about a population.

2 Describe methods to select a sample.

3 Define and construct a sampling distribution of the sample mean.

4 Understand and explain the *central limit theorem.*

5 Use the central limit theorem to find probabilities of selecting possible sample means from a specified population.

The Nike annual report says that the average American buys 6.5 pairs of sports shoes per year. Suppose the population standard deviation is 2.1 and that a sample of 81 customers will be examined next year. What is the standard error of the mean in this experiment? (See Goal 5 and Exercise 45.)

Introduction

Chapters 2 through 4 emphasize techniques to describe data. To illustrate these techniques, we organize the prices for the 80 vehicles sold last month at Whitner Autoplex into a frequency distribution and compute various measures of location and dispersion. Such measures as the mean and the standard deviation describe the typical selling price and the spread in the selling prices. In these chapters the emphasis is on describing the condition of the data. That is, we describe something that has already happened.

Chapter 5 starts to lay the foundation for statistical inference with the study of probability. Recall that in statistical inference our goal is to determine something about a *population* based only on the *sample*. The population is the entire group of individuals or objects under consideration, and the sample is a part or subset of that population. Chapter 6 extends the probability concepts by describing three discrete probability distributions: the binomial, the hypergeometric, and the Poisson. Chapter 7 describes the uniform probability distribution and the normal probability distribution. Both of these are continuous distributions. Probability distributions encompass all possible outcomes of an experiment and the probability associated with each outcome. We use probability distributions to evaluate the likelihood something occurs in the future.

This chapter begins our study of sampling. Sampling is a tool to infer something about a population. We begin this chapter by discussing methods of selecting a sample from a population. Next, we construct a distribution of the sample mean to understand how the sample means tend to cluster around the population mean. Finally, we show that for any population the shape of this sampling distribution tends to follow the normal probability distribution.

Sampling Methods

In Chapter 1, we said the purpose of inferential statistics is to find something about a population based on a sample. A sample is a portion or part of the population of interest. In many cases, sampling is more feasible than studying the entire population. In this section, we show major reasons for sampling, and then several methods for selecting a sample.

Reasons to Sample

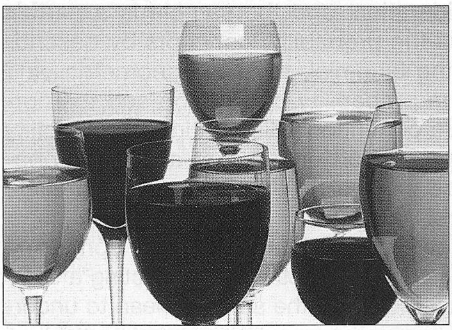
When studying characteristics of a population, there are many practical reasons why we prefer to select portions or samples of a population to observe and measure. Here are some of the reasons for sampling:

1. **To contact the whole population would be time consuming.** A candidate for a national office may wish to determine her chances for election. A sample poll using the regular staff and field interviews of a professional polling firm would take only 1 or 2 days. Using the same staff and interviewers and working 7 days a week, it would take nearly 200 years to contact all the voting population! Even if a large staff of interviewers could be assembled, the benefit of contacting all of the voters would probably not be worth the time.
2. **The cost of studying all the items in a population may be prohibitive.** Public opinion polls and consumer testing organizations, such as Gallup Polls and Roper ASW, usually contact fewer than 2,000 of the nearly 60 million families in the United States. One consumer panel–type organization charges about $40,000 to mail samples and tabulate responses in order to test a product (such as breakfast cereal, cat food, or perfume). The same product test using all 60 million families would cost about $1 billion.

Statistics in Action

With the significant role played by inferential statistics in all branches of science, the availability of large sources of random numbers has become a necessity. The first book of random numbers, containing 41,600 random digits generated by L. Tippett, was published in 1927. In 1938, R. A. Fisher and F. Yates published 15,000 random digits generated using two decks of cards. In 1955, RAND Corporation published a million random digits, generated by the random frequency pulses of an electronic roulette wheel. By 1970, applications of sampling required billions of random numbers. Methods have since been developed for generating, using a computer, digits that are "almost" random and hence are called *pseudo-random*. The question of whether a computer progam can be used to generate numbers that are truly random remains a debatable issue.

3. **The physical impossibility of checking all items in the population.** Some populations are infinite. It would be impossible to check all the water in Lake Erie for bacterial levels, so we select samples at various locations. The populations of fish, birds, snakes, mosquitoes, and the like are large and are constantly moving, being born, and dying. Instead of even attempting to count all the ducks in Canada or all the fish in Lake Pontchartrain, we make estimates using various techniques—such as counting all the ducks on a pond picked at random, making creel checks, or setting nets at predetermined places in the lake.
4. **The destructive nature of some tests.** If the wine tasters at the Sutter Home Winery in California drank all the wine to evaluate the vintage, they would consume the entire crop, and none would be available for sale. In the area of industrial production, steel plates, wires, and similar products must have a certain minimum tensile strength. To ensure that the product meets the minimum standard, the Quality Assurance Department selects a sample from the current production. Each piece is stretched until it breaks and the breaking point (usually measured in pounds per square inch) recorded. Obviously, if all the wire or all the plates were tested for tensile strength, none would be available for sale or use. For the same reason, only a sample of photographic film is selected and tested by Kodak to determine the quality of all the film produced, and only a few seeds are tested for germination by Burpee prior to the planting season.



5. **The sample results are adequate.** Even if funds were available, it is doubtful the additional accuracy of a 100 percent sample—that is, studying the entire population—is essential in most problems. For example, the federal government uses a sample of grocery stores scattered throughout the United States to determine the monthly index of food prices. The prices of bread, beans, milk, and other major food items are included in the index. It is unlikely that the inclusion of all grocery stores in the United States would significantly affect the index, since the prices of milk, bread, and other major foods usually do not vary by more than a few cents from one chain store to another.

Simple Random Sampling

The most widely used type of sampling is a **simple random sample.**

SIMPLE RANDOM SAMPLE A sample selected so that each item or person in the population has the same chance of being included.

A table of random numbers is an efficient way to select members of the sample.

To illustrate simple random sampling and selection, suppose a population consists of 845 employees of Nitra Industries. A sample of 52 employees is to be selected from that population. One way of ensuring that every employee in the population has the same chance of being chosen is to first write the name of each employee on a small slip of paper and deposit all of the slips in a box. After they have been thoroughly mixed, the first selection is made by drawing a slip out of the box without looking at it. This process is repeated until the sample of 52 employees is chosen.

A more convenient method of selecting a random sample is to use the identification number of each employee and a **table of random numbers** such as the one in Appendix B.6. As the name implies, these numbers have been generated by a random process (in this case, by a computer). For each digit of a number,

Statistics in Action

Is discrimination taking a bite out of your paycheck? Before you answer, consider a recent article in *Personnel Journal*. These findings indicate that attractive men and women earn about 5 percent more than average lookers, who in turn earn about 5 percent more than their plain counterparts. This is true for both men and women. It is also true for a wide range of occupations, from construction work to auto repair to telemarketing, occupations for which it would seem that looks would not matter.

the probability of 0, 1, 2, . . . , 9 is the same. Thus, the probability that employee number 011 will be selected is the same as for employee 722 or employee 382. By using random numbers to select employees, bias is eliminated from the selection process.

A portion of a table of random numbers is shown in the following illustration. To select a sample of employees, you first choose a starting point in the table. Any starting point will do. Suppose the time is 3:04. You might look at the third column and then move down to the fourth set of numbers. The number is 03759. Since there are only 845 employees, we will use the first three digits of a five-digit random number. Thus, 037 is the number of the first employee to be a member of the sample. Another way of selecting the starting point is to close your eyes and point at a number in the table. To continue selecting employees, you could move in any direction. Suppose you move right. The first three digits of the number to the right of 03759 are 447—the number of the employee selected to be the second member of the sample. The next three-digit number to the right is 961. You skip 961 because there are only 845 employees. You continue to the right and select employee 784, then 189, and so on.

50525	57454	28455	68226	34656	38884	39018
72507	53380	53827	42486	54465	71819	91199
34986	74297	00144	38676	89967	98869	39744
68851	27305	03759	44723	96108	78489	18910
06738	62879	03910	17350	49169	03850	18910
11448	10734	05837	24397	10420	16712	94496

Starting point (03759) · Second employee (44723) · Third employee (78489) · Fourth employee (18910)

Most statistical software packages have available a routine that will select a simple random sample. The following example uses the Excel System to select a random sample.

Example

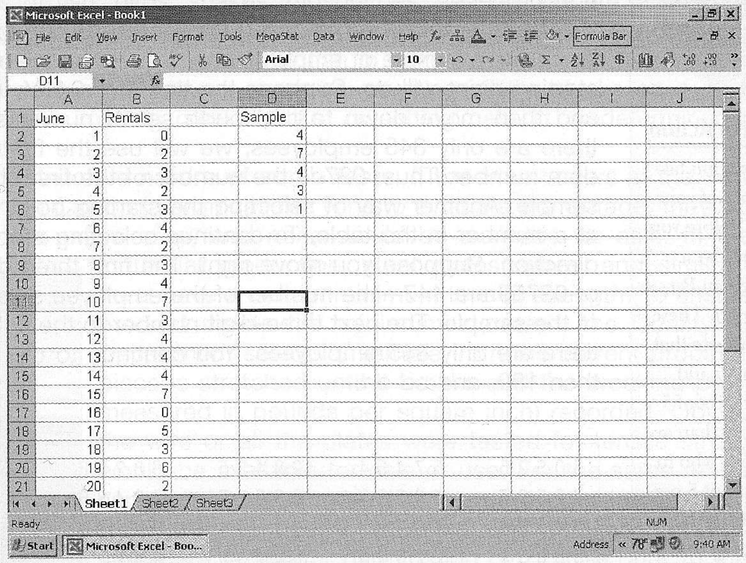
Jane and Joe Miley operate the Foxtrot Inn, a bed and breakfast in Tryon, North Carolina. There are eight rooms available for rent at this B&B. Listed below is the number of these eight rooms rented each day during June 2006. Use Excel to select a sample of five nights during the month of June.

June	Rentals	June	Rentals	June	Rentals
1	0	11	3	21	3
2	2	12	4	22	2
3	3	13	4	23	3
4	2	14	4	24	6
5	3	15	7	25	0
6	4	16	0	26	4
7	2	17	5	27	1
8	3	18	3	28	1
9	4	19	6	29	3
10	7	20	2	30	3

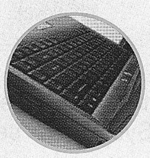
Solution

Excel will select the random sample and report the results. On the first sampled date there were four of the eight rooms rented. On the second sampled date in June, seven of the eight rooms were rented. The information is reported in column D of the Excel spreadsheet. The Excel steps are listed in the **Software Commands** section at the end of the chapter. The Excel system performs the sampling *with* replacement. This means it is possible for the same day to appear more than once in a sample.

Excel



Self-Review 8–1



The following class roster lists the students enrolling in an introductory course in business statistics. Three students are to be randomly selected and asked various questions regarding course content and method of instruction.

(a) The numbers 00 through 45 are handwritten on slips of paper and placed in a bowl. The three numbers selected are 31, 7, and 25. Which students would be included in the sample?

(b) Now use the table of random digits, Appendix B.6, to select your own sample.

(c) What would you do if you encountered the number 59 in the table of random digits?

CSPM 264 01 BUSINESS & ECONOMIC STAT
8:00 AM 9:40 AM MW ST 118 LIND D

RANDOM NUMBER	NAME	CLASS RANK	RANDOM NUMBER	NAME	CLASS RANK
00	ANDERSON, RAYMOND	SO	23	MEDLEY, CHERYL ANN	SO
01	ANGER, CHERYL RENEE	SO	24	MITCHELL, GREG R	FR
02	BALL, CLAIRE JEANETTE	FR	25	MOLTER, KRISTI MARIE	SO
03	BERRY, CHRISTOPHER G	FR	26	MULCAHY, STEPHEN ROBERT	SO
04	BOBAK, JAMES PATRICK	SO	27	NICHOLAS, ROBERT CHARLES	JR
05	BRIGHT, M. STARR	JR	28	NICKENS, VIRGINIA	SO
06	CHONTOS, PAUL JOSEPH	SO	29	PENNYWITT, SEAN PATRICK	SO
07	DETLEY, BRIAN HANS	JR	30	POTEAU, KRIS E	JR
08	DUDAS, VIOLA	SO	31	PRICE, MARY LYNETTE	SO
09	DULBS, RICHARD ZALFA	JR	32	RISTAS, JAMES	SR
10	EDINGER, SUSAN KEE	SR	33	SAGER, ANNE MARIE	SO
11	FINK, FRANK JAMES	SR	34	SMILLIE, HEATHER MICHELLE	SO
12	FRANCIS, JAMES P	JR	35	SNYDER, LEISHA KAY	SR
13	GAGHEN, PAMELA LYNN	JR	36	STAHL, MARIA TASHERY	SO
14	GOULD, ROBYN KAY	SO	37	ST. JOHN, AMY J	SO
15	GROSENBACHER, SCOTT ALAN	SO	38	STURDEVANT, RICHARD K	SO
16	HEETFIELD, DIANE MARIE	SO	39	SWETYE, LYNN MICHELE	SO
17	KABAT, JAMES DAVID	JR	40	WALASINSKI, MICHAEL	SO
18	KEMP, LISA ADRIANE	FR	41	WALKER, DIANE ELAINE	SO
19	KILLION, MICHELLE A	SO	42	WARNOCK, JENNIFER MARY	SO
20	KOPERSKI, MARY ELLEN	SO	43	WILLIAMS, WENDY A	SO
21	KOPP, BRIDGETTE ANN	SO	44	YAP, HOCK BAN	SO
22	LEHMANN, KRISTINA MARIE	JR	45	YODER, ARLAN JAY	JR

Systematic Random Sampling

The simple random sampling procedure may be awkward in some research situations. For example, suppose the sales division of Computer Graphic, Inc., needs to quickly estimate the mean dollar revenue per sale during the past month. It finds that 2,000 sales invoices were recorded and stored in file drawers, and decides to select 100 invoices to estimate the mean dollar revenue. Simple random sampling requires the numbering of each invoice before using the random number table to select the 100 invoices. The numbering process would be a very time-consuming task. Instead, we use **systematic random sampling.**

SYSTEMATIC RANDOM SAMPLE A random starting point is selected, and then every *k*th member of the population is selected.

First, *k* is calculated as the population size divided by the sample size. For Computer Graphic, Inc., we would select every 20th (2,000/100) invoice from the file drawers; in so doing the numbering process is avoided. If *k* is not a whole number, then round down.

Random sampling is used in the selection of the first invoice. For example, a number from a random number table between 1 and *k*, or 20, would be selected. Say the random number was 18. Then, starting with the 18th invoice, every 20th invoice (18, 38, 58, etc.) would be selected as the sample.

Before using systematic random sampling, we should carefully observe the physical order of the population. When the physical order is related to the population characteristic, then systematic random sampling should not be used. For example, if the invoices in the example were filed in order of increasing sales, systematic random sampling would not guarantee a random sample. Other sampling methods should be used.

Statistics in Action

Random and unbiased sampling methods are extremely important to make valid statistical inferences. In 1936, a straw vote to predict the outcome of the presidential race between Franklin Roosevelt and Alfred Landon was done. Ten million ballots in the form of returnable postcards were sent to addresses taken from telephone directories and automobile registrations. A high proportion of the ballots were returned, with 59 percent in favor of Landon and 41 percent favoring Roosevelt. On Election Day, Roosevelt won with 61 percent of the vote. Landon had 39 percent. In the mid-1930s people who had telephones and drove automobiles clearly did not represent American voters!

Stratified Random Sampling

When a population can be clearly divided into groups based on some characteristic, we may use **stratified random sampling.** It guarantees each group is represented in the sample. The groups are also called **strata.** For example, college students can be grouped as full time or part time, male or female, or traditional or nontraditional. Once the strata are defined, we can apply simple random sampling within each group or stratum to collect the sample.

STRATIFIED RANDOM SAMPLE A population is divided into subgroups, called strata, and a sample is randomly selected from each stratum.

For instance, we might study the advertising expenditures for the 352 largest companies in the United States. Suppose the objective of the study is to determine whether firms with high returns on equity (a measure of profitability) spent more of each sales dollar on advertising than firms with a low return or deficit. To make sure that the sample is a fair representation of the 352 companies, the companies are grouped on percent return on equity. Table 8–1 shows the strata and the relative frequencies. If simple random sampling was used, observe that firms in the 3rd and 4th strata have a high chance of selection (probability of 0.87) while firms in the other strata have a low chance of selection (probability of 0.13). We might not select any firms in stratum 1 or 5 *simply by chance*. However, stratified random sampling will guarantee that at least one firm in strata 1 and 5 are represented in the sample. Let's say that 50 firms are selected for intensive study. Then 1 (0.02 × 50) firm from stratum 1 would be randomly selected, 5 (0.10 × 50) firms from stratum 2 would be randomly selected, and so on. In this case, the number of firms sampled from each stratum is proportional to the stratum's relative frequency in the population. Stratified sampling has the advantage, in some cases, of more accurately reflecting the characteristics of the population than does simple random or systematic random sampling.

TABLE 8–1 Number Selected for a Proportional Stratified Random Sample

Stratum	Profitability (return on equity)	Number of Firms	Relative Frequency	Number Sampled
1	30 percent and over	8	0.02	1*
2	20 up to 30 percent	35	0.10	5*
3	10 up to 20 percent	189	0.54	27
4	0 up to 10 percent	115	0.33	16
5	Deficit	5	0.01	1
Total		352	1.00	50

*0.02 of 50 = 1, 0.10 of 50 = 5, etc.

Cluster Sampling

Another common type of sampling is **cluster sampling.** It is often employed to reduce the cost of sampling a population scattered over a large geographic area.

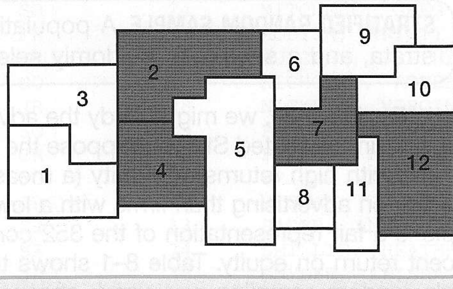
CLUSTER SAMPLE A population is divided into clusters using naturally occurring geographic or other boundaries. Then, clusters are randomly selected and a sample is collected by randomly selecting from each cluster.

Suppose you want to determine the views of residents in Oregon about state and federal environmental protection policies. Selecting a random sample of residents in Oregon and personally contacting each one would be time consuming and very expensive. Instead, you could employ cluster sampling by subdividing the state into small units—either counties or regions. These are often called *primary units*.

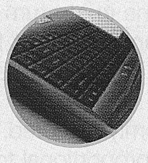
Suppose you divided the state into 12 primary units, then selected at random four regions—2, 7, 4, and 12—and concentrated your efforts in these primary units. You could take a random sample of the residents in each of these regions and interview them. (Note that this is a combination of cluster sampling and simple random sampling.)

Many other sampling methods

The discussion of sampling methods in the preceding sections did not include all the sampling methods available to a researcher. Should you become involved in a major research project in marketing, finance, accounting, or other areas, you would need to consult books devoted solely to sample theory and sample design.



Self-Review 8–2



Refer to Self-Review 8–1 and the class roster on page 264. Suppose a systematic random sample will select every ninth student enrolled in the class. Initially, the fourth student on the list was selected at random. That student is numbered 03. Remembering that the random numbers start with 00, which students will be chosen to be members of the sample?

Exercises

1. The following is a list of Marco's Pizza stores in Lucas County. Also noted is whether the store is corporate-owned (C) or manager-owned (M). A sample of four locations is to be selected and inspected for customer convenience, safety, cleanliness, and other features.

ID No.	Address	Type	ID No.	Address	Type
00	2607 Starr Av	C	12	2040 Ottawa River Rd	C
01	309 W Alexis Rd	C	13	2116 N Reynolds Rd	C
02	2652 W Central Av	C	14	3678 Rugby Dr	C
03	630 Dixie Hwy	M	15	1419 South Av	C
04	3510 Dorr St	C	16	1234 W Sylvania Av	C
05	5055 Glendale Av	C	17	4624 Woodville Rd	M
06	3382 Lagrange St	M	18	5155 S Main	M
07	2525 W Laskey Rd	C	19	106 E Airport Hwy	C
08	303 Louisiana Av	C	20	6725 W Central	M
09	149 Main St	C	21	4252 Monroe	C
10	835 S McCord Rd	M	22	2036 Woodville Rd	C
11	3501 Monroe St	M	23	1316 Michigan Av	M

a. The random numbers selected are 08, 18, 11, 54, 02, 41, and 54. Which stores are selected?

b. Use the table of random numbers to select your own sample of locations.

c. A sample is to consist of every seventh location. The number 03 is the starting point. Which locations will be included in the sample?

d. Suppose a sample is to consist of three locations, of which two are corporate-owned and one is manager-owned. Select a sample accordingly.

2. The following is a list of hospitals in the Cincinnati (Ohio) and Northern Kentucky Region. Also included is whether the hospital is a general medical/surgical hospital (M/S) or a specialty hospital (S). We are interested in estimating the average number of full- and part-time nurses employed in the area hospitals.

a. A sample of five hospitals is to be randomly selected. The random numbers are 09, 16, 00, 49, 54, 12, and 04. Which hospitals are included in the sample?

b. Use a table of random numbers to develop your own sample of five hospitals.

ID Number	Name	Address	Type	ID Number	Name	Address	Type
00	Bethesda North	10500 Montgomery Cincinnati, Ohio 45242	M/S	10	Christ Hospital	2139 Auburn Avenue Cincinnati, Ohio 45219	M/S
01	Ft. Hamilton-Hughes	630 Eaton Avenue Hamilton, Ohio 45013	M/S	11	Deaconess Hospital	311 Straight Street Cincinnati, Ohio 45219	M/S
02	Jewish Hospital-Kenwood	4700 East Galbraith Rd. Cincinnati, Ohio 45236	M/S	12	Good Samaritan Hospital	375 Dixmyth Avenue Cincinnati, Ohio 45220	M/S
03	Mercy Hospital-Fairfield	3000 Mack Road Fairfield, Ohio 45014	M/S	13	Jewish Hospital	3200 Burnet Avenue Cincinnati, Ohio 45229	M/S
04	Mercy Hospital-Hamilton	100 Riverfront Plaza Hamilton, Ohio 45011	M/S	14	University Hospital	234 Goodman Street Cincinnati, Ohio 45267	M/S
05	Middletown Regional	105 McKnight Drive Middletown, Ohio 45044	M/S	15	Providence Hospital	2446 Kipling Avenue Cincinnati, Ohio 45239	M/S
06	Clermont Mercy Hospital	3000 Hospital Drive Batavia, Ohio 45103	M/S	16	St. Francis-St. George Hospital	3131 Queen City Avenue Cincinnati, Ohio 45238	M/S
07	Mercy Hospital-Anderson	7500 State Road Cincinnati, Ohio 45255	M/S	17	St. Elizabeth Medical Center, North Unit	401 E. 20th Street Covington, Kentucky 41014	M/S
08	Bethesda Oak Hospital	619 Oak Street Cincinnati, Ohio 45206	M/S	18	St. Elizabeth Medical Center, South Unit	One Medical Village Edgewood, Kentucky 41017	M/S
09	Children's Hospital Medical Center	3333 Burnet Avenue Cincinnati, Ohio 45229	M/S	19	St. Luke's Hospital West	7380 Turfway Drive Florence, Kentucky 41075	M/S

ID Number	Name	Address	Type	ID Number	Name	Address	Type
20	St. Luke's Hospital East	85 North Grand Avenue Ft. Thomas, Kentucky 41042	M/S	25	Drake Center Rehab—Long Term	151 W. Galbraith Road Cincinnati, Ohio 45216	S
21	Care Unit Hospital	3156 Glenmore Avenue Cincinnati, Ohio 45211	S	26	No. Kentucky Rehab Hospital—Short Term	201 Medical Village Edgewood, Kentucky	S
22	Emerson Behavioral Science	2446 Kipling Avenue Cincinnati, Ohio 45239	S	27	Shriners Burns Institute	3229 Burnet Avenue Cincinnati, Ohio 45229	S
23	Pauline Warfield Lewis Center for Psychiatric Treat.	1101 Summit Road Cincinnati, Ohio 45237	S	28	VA Medical Center	3200 Vine Cincinnati, Ohio 45220	S
24	Children's Psychiatric No. Kentucky	502 Farrell Drive Covington, Kentucky 41011	S				

c. A sample is to consist of every fifth location. We select 02 as the starting point. Which hospitals will be included in the sample?

d. A sample is to consist of four medical and surgical hospitals and one specialty hospital. Select an appropriate sample.

3. Listed below are the 35 members of the Metro Toledo Automobile Dealers Association. We would like to estimate the mean revenue from dealer service departments.

ID Number	Dealer	ID Number	Dealer	ID Number	Dealer
00	Dave White Acura	11	Thayer Chevrolet/Toyota	23	Kistler Ford, Inc.
01	Autofair Nissan	12	Spurgeon Chevrolet Motor Sales, Inc.	24	Lexus of Toledo
02	Autofair Toyota-Suzuki	13	Dunn Chevrolet	25	Mathews Ford Oregon, Inc.
03	George Ball's Buick GMC Truck	14	Don Scott Chevrolet-Pontiac	26	Northtowne Chevrolet
04	Yark Automotive Group	15	Dave White Chevrolet Co.	27	Quality Ford Sales, Inc.
05	Bob Schmidt Chevrolet	16	Dick Wilson Pontiac	28	Rouen Chrysler Jeep Eagle
06	Bowling Green Lincoln Mercury Jeep Eagle	17	Doyle Pontiac Buick	29	Saturn of Toledo
		18	Franklin Park Lincoln Mercury	30	Ed Schmidt Pontiac Jeep Eagle
07	Brondes Ford	19	Genoa Motors	31	Southside Lincoln Mercury
08	Brown Honda	20	Great Lakes Ford Nissan	32	Valiton Chrysler
09	Brown Mazda	21	Grogan Towne Chrysler	33	Vin Divers
10	Charlie's Dodge	22	Hatfield Motor Sales	34	Whitman Ford

a. We want to select a random sample of five dealers. The random numbers are: 05, 20, 59, 21, 31, 28, 49, 38, 66, 08, 29, and 02. Which dealers would be included in the sample?

b. Use the table of random numbers to select your own sample of five dealers.

c. A sample is to consist of every seventh dealer. The number 04 is selected as the starting point. Which dealers are included in the sample?

4. Listed below are the 27 Nationwide Insurance agents in the Toledo, Ohio, metropolitan area. We would like to estimate the mean number of years employed with Nationwide.

ID Number	Agent	ID Number	Agent	ID Number	Agent
00	**Bly Scott** 3332 W Laskey Rd	10	**Heini Bernie** 7110 W Centra	19	**Riker Craig** 2621 N Reynolds Rd
01	**Coyle Mike** 5432 W Central Av	11	**Hinckley Dave** 14 N Holland Sylvania Rd	20	**Schwab Dave** 572 W Dussel Dr
02	**Denker Brett** 7445 Airport Hwy			21	**Seibert John H** 201 S Main
03	**Denker Rollie** 7445 Airport Hwy	12	**Joehlin Bob** 3358 Navarre Av	22	**Smithers Bob** 229 Superior St
04	**Farley Ron** 1837 W Alexis Rd	13	**Keisser David** 3030 W Sylvania Av	23	**Smithers Jerry** 229 Superior St
05	**George Mark** 7247 W Central Av	14	**Keisser Keith** 5902 Sylvania Av	24	**Wright Steve** 105 S Third St
06	**Gibellato Carlo** 6616 Monroe St	15	**Lawrence Grant** 342 W Dussel Dr	25	**Wood Tom** 112 Louisiana Av
07	**Glemser Cathy** 5602 Woodville Rd	16	**Miller Ken** 2427 Woodville Rd	26	**Yoder Scott** 6 Willoughby Av
08	**Green Mike** 4149 Holland Sylvania Rd	17	**O'Donnell Jim** 7247 W Central Av		
09	**Harris Ev** 2026 Albon Rd	18	**Priest Harvey** 5113 N Summit St		

a. We want to select a random sample of four agents. The random numbers are: 02, 59, 51, 25, 14, 29, 77, 69, and 18. Which dealers would be included in the sample?
b. Use the table of random numbers to select your own sample of four agents.
c. A sample is to consist of every seventh dealer. The number 04 is selected as the starting point. Which agents will be included in the sample?

Sampling "Error"

In the previous section we discussed sampling methods that are used to select a sample that is a fair and unbiased representation of the population. In each method, it is important to note that the selection of every possible sample of a specified size from a population has a known chance or probability. This is another way to describe an unbiased sampling method.

Samples are used to estimate population characteristics. For example, the mean of a sample is used to estimate the population mean. However, since the sample is a part or portion of the population, it is unlikely that the sample mean would be *exactly equal* to the population mean. Similarly, it is unlikely that the sample standard deviation would be *exactly equal* to the population standard deviation. We can therefore expect a difference between a *sample statistic* and its corresponding *population parameter*. This difference is called **sampling error.**

SAMPLING ERROR The difference between a sample statistic and its corresponding population parameter.

The following example clarifies the idea of sampling error.

Example

Refer to the previous example on page 263 where we studied the number of rooms rented at the Foxtrot Inn bed and breakfast in Tryon, North Carolina. The population is the number of rooms rented each of the 30 days in June 2006. Find the mean of the population. Use Excel or other statistical software to select three random samples of five days. Calculate the mean of each sample and compare it to the population mean. What is the sampling error in each case?

Solution

During the month there were a total of 94 rentals. So the mean number of units rented per night is 3.13. This is the population mean. Hence we designate this value with the Greek letter μ.

$$\mu = \frac{\Sigma X}{N} = \frac{0 + 2 + 3 + \cdots + 3}{30} = \frac{94}{30} = 3.13$$

The first random sample of five nights resulted in the following number of rooms rented: 4, 7, 4, 3, and 1. The mean of this sample of five nights is 3.8 rooms, which we designate as $\overline{X}_1$. The bar over the X reminds us that it is a sample mean and the subscript 1 indicates it is the mean of the first sample.

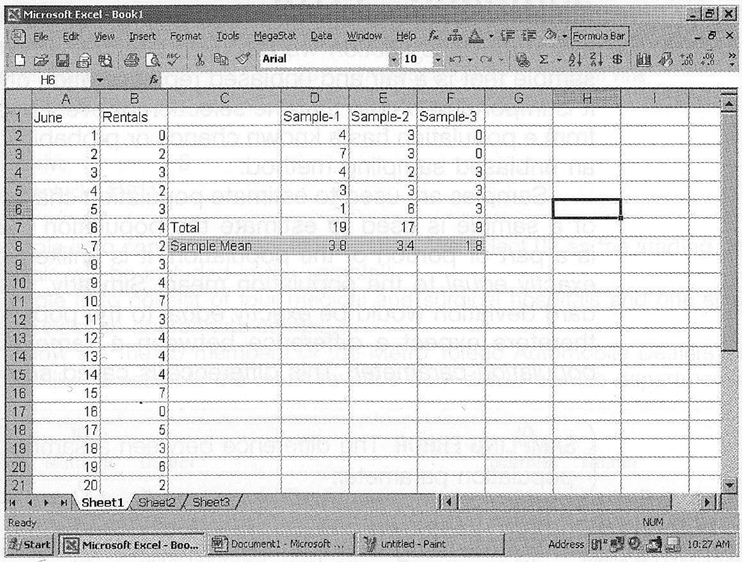
$$\overline{X}_1 = \frac{\Sigma X}{n} = \frac{4 + 7 + 4 + 3 + 1}{5} = \frac{19}{5} = 3.80$$

The sampling error for the first sample is the difference between the population mean (3.13) and the first sample mean (3.80). Hence, the sampling error is ($\overline{X}_1 \quad \mu = 3.80 - 3.13 = 0.67$). The second random sample of five days from the population of all 30 days in June revealed the following number of rooms rented: 3, 3, 2, 3, and 6. The mean of these five values is 3.4, found by

$$\overline{X}_2 = \frac{\Sigma X}{n} = \frac{3 + 3 + 2 + 3 + 6}{5} = 3.4$$

The sampling error is $(\bar{X}_2 - \mu = 3.4 - 3.13 = 0.27)$.

In the third random sample the mean was 1.8 and the sampling error was −1.33. Each of these differences, 0.67, 0.27, and −1.33, is the sampling error made in estimating the population mean. Sometimes these errors are positive values, indicating that the sample mean overestimated the population mean; other times they are negative values, indicating the sample mean was less than the population mean.



In this case where we have a population of 30 values and samples of 5 values there is a very large number of possible samples—142,506 to be exact! To find this value use the combination formula 5–10 on page 168. Each of the 142,506 different samples has the same chance of being selected. Each sample may have a different sample mean and therefore a different sampling error. The value of the sampling error is based on the particular one of the 142,506 different possible samples selected. Therefore, the sampling errors are random and occur by chance. If one were to determine the sum of these sampling errors over a large number of samples the result would be very close to zero. This is true because the sample mean is an unbiased estimator of the population mean.

Sampling Distribution of the Sample Mean

Now that we have discovered the possibility of a sampling error when sample results are used to estimate a population parameter, how can we make an accurate prediction about the possible success of a newly developed toothpaste or other product, based only on sample results? How can the quality-assurance department in a mass-production firm release a shipment of microchips based only on a sample of 10 chips? How can the CNN/*USA Today* or ABC News/*Washington Post* polling organizations make an accurate prediction about a presidential race based on a sample of 1,200 registered voters out of a voting population of nearly 90 million? To answer these questions, we first develop a *sampling distribution of the sample mean*.

Sample means vary from sample to sample.

The sample means in the previous example varied from one sample to the next. The mean of the first sample of 5 days was 3.80 rooms, and the second sample was 3.40 rooms. The population mean was 3.13 rooms. If we organized the means of all possible samples of 5 days into a probability distribution, the result is called the **sampling distribution of the sample mean.**

SAMPLING DISTRIBUTION OF THE SAMPLE MEAN A probability distribution of all possible sample means of a given sample size.

The following example illustrates the construction of a sampling distribution of the sample mean.

Example

Tartus Industries has seven production employees (considered the population). The hourly earnings of each employee are given in Table 8–2.

TABLE 8–2 Hourly Earnings of the Production Employees of Tartus Industries

Employee	Hourly Earnings	Employee	Hourly Earnings
Joe	$7	Jan	$7
Sam	7	Art	8
Sue	8	Ted	9
Bob	8		

1. What is the population mean?
2. What is the sampling distribution of the sample mean for samples of size 2?
3. What is the mean of the sampling distribution?
4. What observations can be made about the population and the sampling distribution?

Solution

Here are the solutions to the questions.

1. The population mean is $7.71, found by:

$$\mu = \frac{\Sigma X}{N} = \frac{\$7 + \$7 + \$8 + \$8 + \$7 + \$8 + \$9}{7} = \$7.71$$

We identify the population mean with the Greek letter μ. Our policy, stated in Chapters 1, 3, and 4, is to identify population parameters with Greek letters.

2. To arrive at the sampling distribution of the sample mean, we need to select all possible samples of 2 without replacement from the population, then compute the mean of each sample. There are 21 possible samples, found by using formula (5–10) on page 168.

$$_NC_n = \frac{N!}{n!(N-n)!} = \frac{7!}{2!(7-2)!} = 21$$

where $N = 7$ is the number of items in the population and $n = 2$ is the number of items in the sample.

The 21 sample means from all possible samples of 2 that can be drawn from the population are shown in Table 8–3. These 21 sample means are used to construct a probability distribution. This is the sampling distribution of the sample mean, and it is summarized in Table 8–4.

TABLE 8–3 Sample Means for All Possible Samples of 2 Employees

Sample	Employees	Hourly Earnings	Sum	Mean	Sample	Employees	Hourly Earnings	Sum	Mean
1	Joe, Sam	$7, $7	$14	$7.00	12	Sue, Bob	$8, $8	$16	$8.00
2	Joe, Sue	7, 8	15	7.50	13	Sue, Jan	8, 7	15	7.50
3	Joe, Bob	7, 8	15	7.50	14	Sue, Art	8, 8	16	8.00
4	Joe, Jan	7, 7	14	7.00	15	Sue, Ted	8, 9	17	8.50
5	Joe, Art	7, 8	15	7.50	16	Bob, Jan	8, 7	15	7.50
6	Joe, Ted	7, 9	16	8.00	17	Bob, Art	8, 8	16	8.00
7	Sam, Sue	7, 8	15	7.50	18	Bob, Ted	8, 9	17	8.50
8	Sam, Bob	7, 8	15	7.50	19	Jan, Art	7, 8	15	7.50
9	Sam, Jan	7, 7	14	7.00	20	Jan, Ted	7, 9	16	8.00
10	Sam, Art	7, 8	15	7.50	21	Art, Ted	8, 9	17	8.50
11	Sam, Ted	7, 9	16	8.00					

TABLE 8–4 Sampling Distribution of the Sample Mean for $n = 2$

Sample Mean	Number of Means	Probability
$7.00	3	.1429
7.50	9	.4285
8.00	6	.2857
8.50	3	.1429
	21	1.0000

3. The mean of the sampling distribution of the sample mean is obtained by summing the various sample means and dividing the sum by the number of samples. The mean of all the sample means is usually written $\mu_{\bar{X}}$. The μ reminds us that it is a population value because we have considered all possible samples. The subscript $\bar{X}$ indicates that it is the sampling distribution of the sample mean.

Population mean is equal to the mean of the sample means

$$\mu_{\bar{X}} = \frac{\text{Sum of all sample means}}{\text{Total number of samples}} = \frac{\$7.00 + \$7.50 + \cdots + \$8.50}{21}$$

$$= \frac{\$162}{21} = \$7.71$$

4. Refer to Chart 8–1, which shows both the population distribution and the distribution of the sample mean. These observations can be made:
   a. The mean of the distribution of the sample mean ($7.71) is equal to the mean of the population: $\mu = \mu_{\bar{X}}$.
   b. The spread in the distribution of the sample mean is less than the spread in the population values. The sample mean ranges from $7.00 to $8.50, while the population values vary from $7.00 up to $9.00. Notice, as we increase the size of the sample, the spread of the distribution of the sample mean becomes smaller.
   c. The shape of the sampling distribution of the sample mean and the shape of the frequency distribution of the population values are different. The distribution of the sample mean tends to be more bell-shaped and to approximate the normal probability distribution.

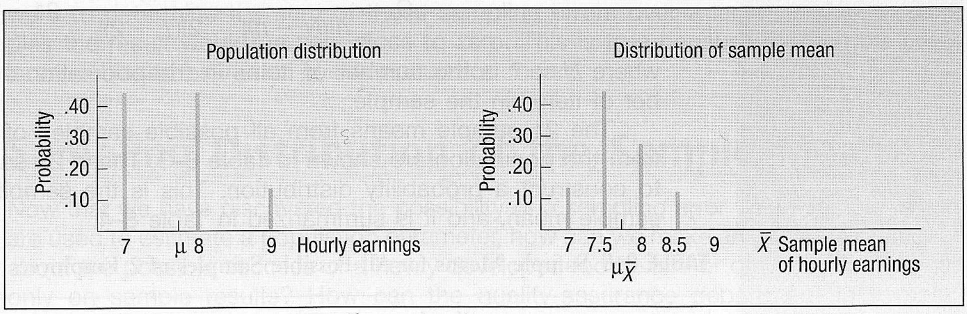


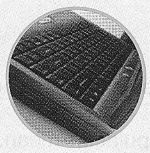
CHART 8–1 Distributions of Population Values and Sample Mean

In summary, we took all possible random samples from a population and for each sample calculated a sample statistic (the mean amount earned). This example illustrates important relationships between the population distribution and the sampling distribution of the sample mean:

1. The mean of the sample means is exactly equal to the population mean.
2. The dispersion of the sampling distribution of sample means is narrower than the population distribution.
3. The sampling distribution of sample means tends to become bell-shaped and to approximate the normal probability distribution.

Given a bell-shaped or normal probability distribution, we will be able to apply concepts from Chapter 7 to determine the probability of selecting a sample with a specified sample mean. In the next section, we will show the importance of sample size as it relates to the sampling distribution of sample means.

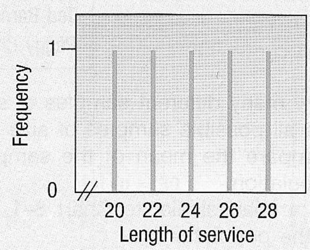
Self-Review 8–3



The lengths of service of all the executives employed by Standard Chemicals are:

Name	Years
Mr. Snow	20
Ms. Tolson	22
Mr. Kraft	26
Ms. Irwin	24
Mr. Jones	28

(a) Using the combination formula, how many samples of size 2 are possible?
(b) List all possible samples of 2 executives from the population and compute their means.
(c) Organize the means into a sampling distribution.
(d) Compare the population mean and the mean of the sample means.
(e) Compare the dispersion in the population with that in the distribution of the sample mean.
(f) A chart portraying the population values follows. Is the distribution of population values normally distributed (bell-shaped)?



(g) Is the distribution of the sample mean computed in part (c) starting to show some tendency toward being bell-shaped?

Exercises

5. A population consists of the following four values: 12, 12, 14, and 16.
   a. List all samples of size 2, and compute the mean of each sample.
   b. Compute the mean of the distribution of the sample mean and the population mean. Compare the two values.
   c. Compare the dispersion in the population with that of the sample mean.
6. A population consists of the following five values: 2, 2, 4, 4, and 8.
   a. List all samples of size 2, and compute the mean of each sample.
   b. Compute the mean of the distribution of sample means and the population mean. Compare the two values.
   c. Compare the dispersion in the population with that of the sample means.
7. A population consists of the following five values: 12, 12, 14, 15, and 20.
   a. List all samples of size 3, and compute the mean of each sample.
   b. Compute the mean of the distribution of sample means and the population mean. Compare the two values.
   c. Compare the dispersion in the population with that of the sample means.
8. A population consists of the following five values: 0, 0, 1, 3, 6.
   a. List all samples of size 3, and compute the mean of each sample.
   b. Compute the mean of the distribution of sample means and the population mean. Compare the two values.
   c. Compare the dispersion in the population with that of the sample means.
9. In the law firm Tybo and Associates, there are six partners. Listed below is the number of cases each associate actually tried in court last month.

Associate	Number of Cases
Ruud	3
Wu	6
Sass	3
Flores	3
Wilhelms	0
Schueller	1

a. How many different samples of 3 are possible?
b. List all possible samples of size 3, and compute the mean number of cases in each sample.
c. Compare the mean of the distribution of sample means to the population mean.
d. On a chart similar to Chart 8–1, compare the dispersion in the population with that of the sample means.

10. There are five sales associates at Mid-Motors Ford. The five representatives and the number of cars they sold last week are:

Sales Representative	Cars Sold
Peter Hankish	8
Connie Stallter	6
Juan Lopez	4
Ted Barnes	10
Peggy Chu	6

a. How many different samples of size 2 are possible?
b. List all possible samples of size 2, and compute the mean of each sample.
c. Compare the mean of the sampling distribution of sample means with that of the population.
d. On a chart similar to Chart 8–1, compare the dispersion in sample means with that of the population.

The Central Limit Theorem

In this section, we examine the **central limit theorem.** Its application to the sampling distribution of the sample mean, introduced in the previous section, allows us to use the normal probability distribution to create confidence intervals for the population mean (described in Chapter 9) and perform tests of hypothesis (described in Chapter 10). The central limit theorem states that, for large random samples, the shape of the sampling distribution of the sample mean is close to the normal probability distribution. The approximation is more accurate for large samples than for small samples. This is one of the most useful conclusions in statistics. We can reason about the distribution of the sample mean with absolutely no information about the shape of the population distribution from which the sample is taken. In other words, the central limit theorem is true for all distributions.

A formal statement of the central limit theorem follows.

CENTRAL LIMIT THEOREM If all samples of a particular size are selected from any population, the sampling distribution of the sample mean is approximately a normal distribution. This approximation improves with larger samples.

If the population follows a normal probability distribution, then for any sample size the sampling distribution of the sample mean will also be normal. If the population distribution is symmetrical (but not normal), you will see the normal shape of the distribution of the sample mean emerge with samples as small as 10. On the other hand, if you start with a distribution that is skewed or has thick tails, it may require samples of 30 or more to observe the normality feature. This concept is summarized in Chart 8–2

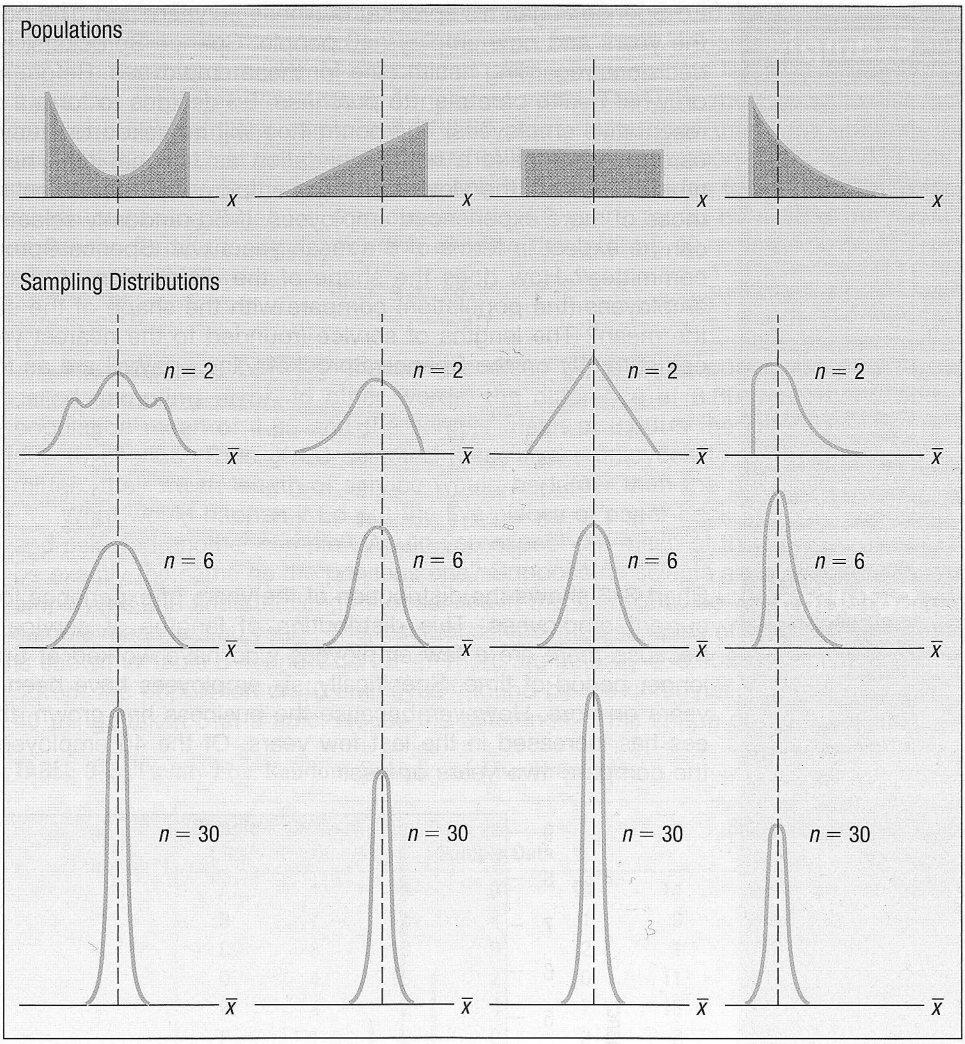


CHART 8–2 Results of the Central Limit Theorem for Several Populations

for various population shapes. Observe the convergence to a normal distribution regardless of the shape of the population distribution. Most statisticians consider a sample of 30 or more to be large enough for the central limit theorem to be employed.

The idea that the distribution of the sample means from a population that is not normal will converge to normality is illustrated in Charts 8–3, 8–4, and 8–5. We will discuss this example in more detail shortly, but Chart 8–3 is a graph of a discrete probability distribution that is positively skewed. There are many possible samples of 5 that might be selected from this population. Suppose we randomly select 25 samples of size 5 each and compute the mean of each sample. These results are shown in Chart 8–4. Notice that the shape of the distribution of sample means has changed from the shape of the original population even though we selected only 25 of the many possible samples. To put it another way, we selected 25 random samples of 5 each from a population that is positively skewed and found the distribution of sample means has changed from the shape of the population. As we take larger samples, that is, $n = 20$ instead of $n = 5$, we will find the distribution of the sample mean will approach the normal distribution. Chart 8–5 shows the results of 25 random samples of 20 observations each from the same population. Observe the clear trend toward the normal probability distribution. This is the point of the central limit theorem. The following example will underscore this condition.

Example

Ed Spence began his sprocket business 20 years ago. The business has grown over the years and now employs 40 people. Spence Sprockets, Inc., faces some major decisions regarding health care for these employees. Before making a final decision on what health care plan to purchase, Ed decides to form a committee of five representative employees. The committee will be asked to study the health care issue carefully and make a recommendation as to what plan best fits the employees' needs. Ed feels the views of newer employees toward health care may differ from those of more experienced employees. If Ed randomly selects this committee, what can he expect in terms of the mean years with Spence Sprockets for those on the committee? How does the shape of the distribution of years of experience of all employees (the population) compare with the shape of the sampling distribution of the mean? The lengths of service (rounded to the nearest year) of the 40 employees currently on the Spence Sprockets, Inc., payroll are as follows.

11	4	18	2	1	2	0	2	2	4
3	4	1	2	2	3	3	19	8	3
7	1	0	2	7	0	4	5	1	14
16	8	9	1	1	2	5	10	2	3

Solution

Chart 8–3 shows the distribution of the years of experience for the population of 40 current employees. This distribution of lengths of service is positively skewed because there are a few employees who have worked at Spence Sprockets for a longer period of time. Specifically, six employees have been with the company 10 years or more. However, because the business has grown, the number of employees has increased in the last few years. Of the 40 employees, 18 have been with the company two years or less.

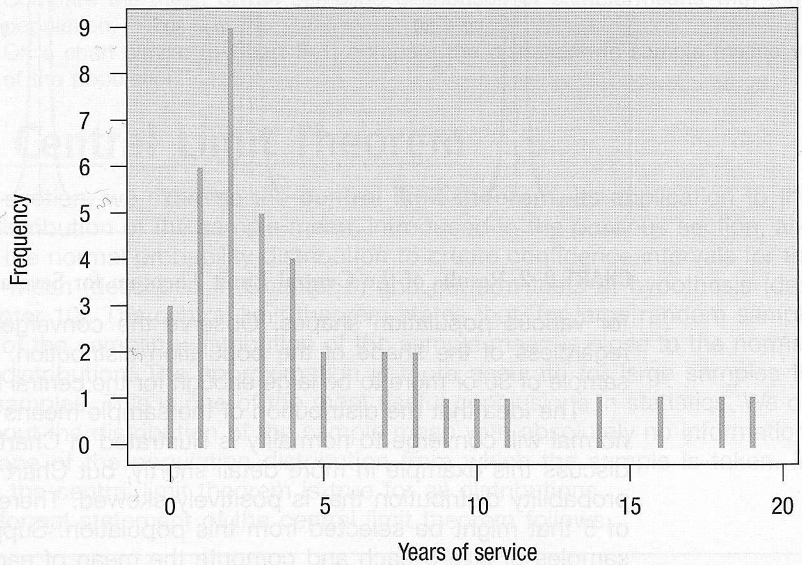


CHART 8–3 Length of Service for Spence Sprockets, Inc., Employees

Let's consider the first of Ed Spence's problems. He would like to form a committee of five employees to look into the health care question and suggest what type of health care coverage would be most appropriate for the majority of workers. How should he select the committee? If he selects the committee randomly, what might he expect in terms of mean length of service for those on the committee?

To begin, Ed writes the length of service for each of the 40 employees on pieces of paper and puts them into an old baseball hat. Next, he shuffles the pieces of paper around and randomly selects five slips of paper. The lengths of service for these five employees are 1, 9, 0, 19, and 14 years. Thus, the mean length of service for these five sampled employees is 8.60 years. How does that compare with the population mean? At this point Ed does not know the population mean, but the number of employees in the population is only 40, so he decides to calculate the mean length of service for *all* his employees. It is 4.8 years, found by adding the lengths of service for *all* the employees and dividing the total by 40.

$$\mu = \frac{11 + 4 + 18 + \cdots + 2 + 3}{40} = 4.80$$

The difference between the sample mean ($\overline{X}$) and the population mean (μ) is called **sampling error.** In other words, the difference of 3.80 years between the population mean of 4.80 and the sample mean of 8.60 is the sampling error. It is due to chance. Thus, if Ed selected these five employees to constitute the committee, their mean length of service would be larger than the population mean.

What would happen if Ed put the five pieces of paper back into the baseball hat and selected another sample? Would you expect the mean of this second sample to be exactly the same as the previous one? Suppose he selects another sample of five employees and finds the lengths of service in this sample to be 7, 4, 4, 1, and 3. This sample mean is 3.80 years. The result of selecting 25 samples of five employees each is shown in Table 8–5 and Chart 8–4. There are actually 658,008 possible samples of 5 from the population of 40 employees, found by the combination formula (5–10) for 40 things taken 5 at a time. Notice the difference in the shape of the population and

TABLE 8–5 Twenty-Five Random Samples of Five Employees

Sample I.D.	Sample Data					Sample Mean
A	1	9	0	19	14	8.6
B	7	4	4	1	3	3.8
C	8	19	8	2	1	7.6
D	4	18	2	0	11	7.0
E	4	2	4	7	18	7.0
F	1	2	0	3	2	1.6
G	2	3	2	0	2	1.8
H	11	2	9	2	4	5.6
I	9	0	4	2	7	4.4
J	1	1	1	11	1	3.0
K	2	0	0	10	2	2.8
L	0	2	3	2	16	4.6
M	2	3	1	1	1	1.6
N	3	7	3	4	3	4.0
O	1	2	3	1	4	2.2
P	19	0	1	3	8	6.2
Q	5	1	7	14	9	7.2
R	5	4	2	3	4	3.6
S	14	5	2	2	5	5.6
T	2	1	1	4	7	3.0
U	3	7	1	2	1	2.8
V	0	1	5	1	2	1.8
W	0	3	19	4	2	5.6
X	4	2	3	4	0	2.6
Y	1	1	2	3	2	1.8

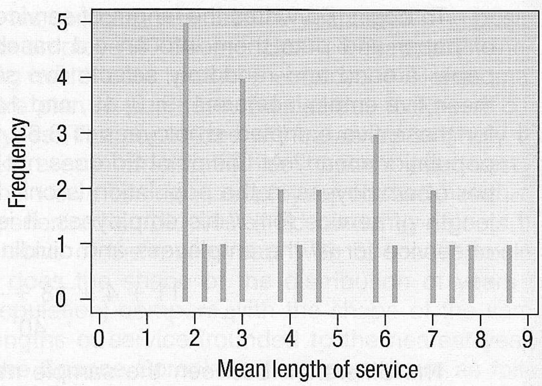


CHART 8–4 Histogram of Mean Lengths of Service for 25 Samples of Five Employees

the distribution of these sample means. The population of the lengths of service for employees (Chart 8–3) is positively skewed, but the distribution of these 25 sample means does not reflect the same positive skew. There is also a difference in the range of the sample means versus the range of the population. The population ranged from 0 to 19 years, whereas the sample means range from 1.6 to 8.6 years.

Table 8–6 reports the result of selecting 25 samples of 20 employees each and computing their sample means. These sample means are shown graphically in Chart 8–5. Compare the shape of this distribution to the population (Chart 8–3) and to the distribution of sample means where the sample is $n = 5$ (Chart 8–4). You should observe two important features:

TABLE 8–6 Random Samples and Sample Means of 25 Samples of 20 Spence Sprocket, Inc., Employees

Sample Number	Sample Data (Length of Service)																				Sample Mean
A	3	8	3	0	2	1	2	3	11	5	1	3	4	2	7	1	1	2	4	16	3.95
B	2	3	8	2	1	5	2	0	3	1	0	7	1	4	3	11	4	4	3	1	3.25
C	14	5	0	3	2	14	11	9	2	2	1	2	19	1	0	1	4	2	19	8	5.95
D	9	2	1	1	4	10	0	8	4	3	2	1	0	8	1	14	5	10	1	3	4.35
E	18	1	2	2	4	3	2	8	2	1	0	19	4	19	0	1	4	0	3	14	5.35
F	10	4	4	18	3	3	1	0	0	2	2	4	7	10	2	0	3	4	2	1	4.00
G	5	7	11	8	11	18	1	1	16	2	2	16	2	3	2	16	2	2	2	4	6.55
H	3	0	2	0	5	4	5	3	8	3	2	5	1	1	2	9	8	3	16	5	4.25
I	0	0	18	2	1	7	4	1	3	0	3	2	11	7	2	8	5	1	2	3	4.00
J	2	7	2	4	1	3	3	2	5	10	0	1	1	2	9	3	2	19	3	2	4.05
K	7	4	5	3	3	0	18	2	0	4	2	7	2	7	4	2	10	1	1	2	4.20
L	0	3	10	5	9	2	1	4	1	2	1	8	18	1	4	3	3	2	0	4	4.05
M	4	1	2	1	7	3	9	14	8	19	4	4	1	2	0	3	1	2	1	2	4.40
N	3	16	1	2	4	4	4	2	1	5	2	3	5	3	4	7	16	1	11	1	4.75
O	2	19	2	0	2	2	16	2	3	11	9	2	8	0	8	2	7	3	2	2	5.10
P	2	18	16	5	2	2	19	0	1	2	11	4	2	2	1	4	2	0	4	3	5.00
Q	3	2	3	11	10	1	1	5	19	16	7	10	3	1	1	1	2	2	3	1	5.10
R	2	3	1	2	7	4	3	19	9	2	2	1	1	2	2	2	1	8	0	2	3.65
S	2	14	19	1	19	2	8	4	2	2	14	2	8	16	4	7	2	9	0	7	7.10
T	0	1	3	3	2	2	3	1	1	0	3	2	3	5	2	10	14	4	2	0	3.05
U	1	0	1	2	16	1	1	2	5	1	4	1	2	2	2	2	2	8	9	3	3.25
V	1	9	4	4	2	8	7	1	14	18	1	5	10	11	19	0	3	7	2	11	6.85
W	8	1	9	19	3	19	0	5	2	1	5	3	3	4	1	5	3	1	8	7	5.35
X	4	2	0	3	1	16	1	11	3	3	2	18	2	0	1	5	0	7	2	5	4.30
Y	1	2	1	2	0	2	7	2	4	8	19	2	5	3	3	0	19	2	1	18	5.05

1. The shape of the distribution of the sample mean is different from that of the population. In Chart 8–3 the distribution of all employees is positively skewed. However, as we select random samples from this population, the shape of the distribution of the sample mean changes. As we increase the size of the sample, the distribution of the sample mean approaches the normal probability distribution. This illustrates the central limit theorem.

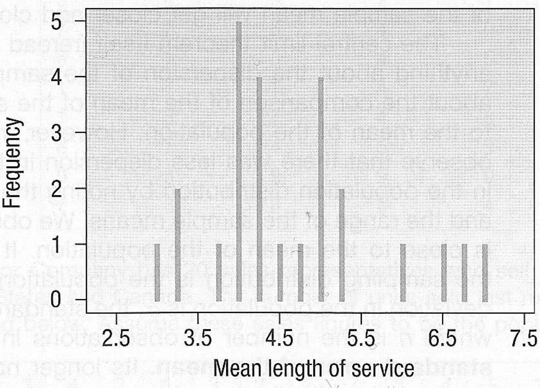


CHART 8–5 Histogram of Mean Lengths of Service for 25 Samples of 20 Employees

2. There is less dispersion in the sampling distribution of sample means than in the population distribution. In the population the lengths of service ranged from 0 to 19 years. When we selected samples of 5, the sample means ranged from 1.6 to 8.6 years, and when we selected samples of 20, the means ranged from 3.05 to 7.10 years.

We can also compare the mean of the sample means to the population mean. The mean of the 25 samples of 20 employees reported in Table 8–6 is 4.676 years.

$$\mu_{\bar{X}} = \frac{3.95 + 3.25 + \cdots + 4.30 + 5.05}{25} = 4.676$$

We use the symbol $\mu_{\bar{X}}$ to identify the mean of the distribution of the sample mean. The subscript reminds us that the distribution is of the sample mean. It is read "mu sub X bar." We observe that the mean of the sample means, 4.676 years, is very close to the population mean of 4.80.

What should we conclude from this example? The central limit theorem indicates that, regardless of the shape of the population distribution, the sampling distribution of the sample mean will move toward the normal probability distribution. The larger the number of observations in each sample, the stronger the convergence. The Spence Sprockets, Inc., example shows how the central limit theorem works. We began with a positively skewed population (Chart 8–3). Next, we selected 25 random samples of 5 observations, computed the mean of each sample, and finally organized these 25 sample means into a graph (Chart 8–4). We observe a change in the shape of the sampling distribution of the sample mean from that of the population. The movement is from a positively skewed distribution to a distribution that has the shape of the normal probability distribution.

To further illustrate the effects of the central limit theorem, we increased the number of observations in each sample from 5 to 20. We selected 25 samples of 20 observations each and calculated the mean of each sample. Finally, we organized these sample means into a graph (Chart 8–5). The shape of the histogram in Chart 8–5 is clearly moving toward the normal probability distribution.

If you go back to Chapter 6 where several binomial distributions with a "success" proportion of .10 are shown in Chart 6–4, you can see yet another demonstration of the central limit theorem. Observe as n increases from 7 through 12 and 20 up to 40 that the profile of the probability distributions moves closer and closer to a normal probability distribution. Chart 8–5 on page 279 also shows the convergence to normality as n increases. This again reinforces the fact that, as more observations are sampled from any population distribution, the shape of the sampling distribution of the sample mean will get closer and closer to a normal distribution.

The central limit theorem itself (reread the definition on page 274) does not say anything about the dispersion of the sampling distribution of the sample mean or about the comparison of the mean of the sampling distribution of the sample mean to the mean of the population. However, in our Spence Sprockets example we did observe that there was less dispersion in the distribution of the sample mean than in the population distribution by noting the difference in the range in the population and the range of the sample means. We observe that the mean of the sample means is close to the mean of the population. It can be demonstrated that the mean of the sampling distribution is the population mean, i.e., $\mu_{\bar{X}} = \mu$, and if the standard deviation in the population is σ, the standard deviation of the sample means is $\sigma/\sqrt{n}$, where n is the number of observations in each sample. We refer to $\sigma/\sqrt{n}$ as the **standard error of the mean.** Its longer name is actually the *standard deviation of the sampling distribution of the sample mean*.

STANDARD ERROR OF THE MEAN $$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad \text{[8–1]}$$

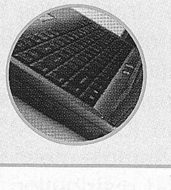
In this section we also came to other important conclusions.

1. The mean of the distribution of sample means will be *exactly* equal to the population mean if we are able to select all possible samples of the same size from a given population. That is:

$$\mu = \mu_{\bar{X}}$$

Even if we do not select all samples, we can expect the mean of the distribution of sample means to be close to the population mean.
2. There will be less dispersion in the sampling distribution of the sample mean than in the population. If the standard deviation of the population is σ, the standard deviation of the distribution of sample means is $\sigma/\sqrt{n}$. Note that when we increase the size of the sample the standard error of the mean decreases.

Self-Review 8–4



Refer to the Spence Sprockets, Inc., data on page 276. Select 10 random samples of 5 employees each. Use the methods described earlier in the chapter and the Table of Random Numbers (Appendix B.6) to find the employees to include in the sample. Compute the mean of each sample and plot the sample means on a chart similar to Chart 8–3. What is the mean of your 10 sample means?

Exercises

11. Appendix B.6 is a table of random numbers. Hence, each digit from 0 to 9 has the same likelihood of occurrence.

a. Draw a graph showing the population distribution. What is the population mean?

b. Following are the first 10 rows of five digits from Appendix B.6. Assume that these are 10 random samples of five values each. Determine the mean of each sample and plot the means on a chart similar to Chart 8–3. Compare the mean of the sampling distribution of the sample means with the population mean.

0	2	7	1	1
9	4	8	7	3
5	4	9	2	1
7	7	6	4	0
6	1	5	4	5
1	7	1	4	7
1	3	7	4	8
8	7	4	5	5
0	8	9	9	9
7	8	8	0	4

12. Scrapper Elevator Company has 20 sales representatives who sell its product throughout the United States and Canada. The number of units sold last month by each representative is listed below. Assume these sales figures to be the population values.

2	3	2	3	3	4	2	4	3	2	2	7	3	4	5	3	3	3	3	5

a. Draw a graph showing the population distribution.
b. Compute the mean of the population.
c. Select five random samples of 5 each. Compute the mean of each sample. Use the methods described in this chapter and Appendix B.6 to determine the items to be included in the sample.
d. Compare the mean of the sampling distribution of the sample means to the population mean. Would you expect the two values to be about the same?
e. Draw a histogram of the sample means. Do you notice a difference in the shape of the distribution of sample means compared to the shape of the population distribution?

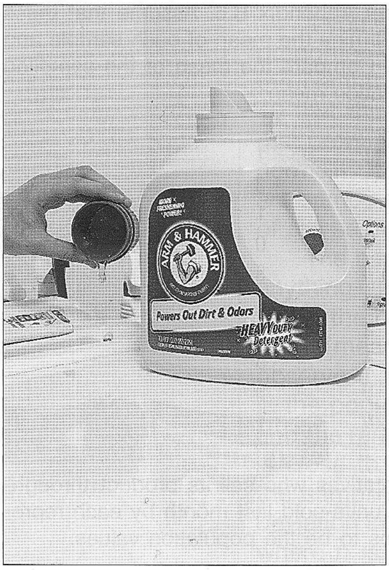
13. Consider all of the coins (pennies, nickels, quarters, etc.) in your pocket or purse as a population. Make a frequency table beginning with the current year and counting backward to record the ages (in years) of the coins. For example, if the current year is 2006, then a coin with 2004 stamped on it is 2 years old.
a. Draw a histogram or other graph showing the population distribution.
b. Randomly select five coins and record the mean age of the sampled coins. Repeat this sampling process 20 times. Now draw a histogram or other graph showing the distribution of the sample means.
c. Compare the shapes of the two histograms.

14. Consider the digits in the phone numbers on a randomly selected page of your local phone book a population. Make a frequency table of the final digit of 30 randomly selected phone numbers. For example, if a phone number is 555-9704, record a 4.
a. Draw a histogram or other graph of this population distribution. Using the uniform distribution, compute the population mean and the population standard deviation.
b. Also record the sample mean of the final four digits (9704 would lead to a mean of 5). Now draw a histogram or other graph showing the distribution of the sample means.
c. Compare the shapes of the two histograms.

Using the Sampling Distribution of the Sample Mean

The previous discussion is important because most business decisions are made on the basis of sampling results. Here are some examples.

1. Arm and Hammer Company wants to ensure that its laundry detergent actually contains 100 fluid ounces, as indicated on the label. Historical summaries from



the filling process indicate the mean amount per container is 100 fluid ounces and the standard deviation is 2 fluid ounces. The quality technician in her 10 A.M. check of 40 containers finds the mean amount per container is 99.8 fluid ounces. Should the technician shut down the filling operation or is the sampling error reasonable?

2. A. C. Nielsen Company provides information to companies advertising on television. Prior research indicates that adult Americans watch an average of 6.0 hours per day of television. The standard deviation is 1.5 hours. For a sample of 50 adults in the greater Boston area, would it be reasonable that we could randomly select a sample and find that they watch an average of 6.5 hours of television per day?
3. Haughton Elevator Company wishes to develop specifications for the number of people who can ride in a new oversized elevator. Suppose the mean weight for an adult is 160 pounds and the standard deviation is 15 pounds. However, the distribution of weights does not follow the normal probability distribution. It is positively skewed. What is the likelihood that for a sample of 30 adults their mean weight is 170 pounds or more?

In each of these situations we have a population about which we have some information. We take a sample from that population and wish to conclude whether the sampling error, that is, the difference between the population parameter and the sample statistic, is due to chance.

Using ideas discussed in the previous section, we can compute the probability that a sample mean will fall within a certain range. We know that the sampling distribution of the sample mean will follow the normal probability distribution under two conditions:

1. When the samples are taken from populations known to follow the normal distribution. In this case the size of the sample is not a factor.
2. When the shape of the population distribution is not known or the shape is known to be nonnormal, but our sample contains at least 30 observations. In this case, the central limit theorem guarantees the sampling distribution of the mean follows a normal distribution.

We can use formula (7–5) from the previous chapter to convert any normal distribution to the standard normal distribution. We also refer to this as a z value. Then we can use the standard normal table, Appendix B.1, to find the probability of selecting an observation that would fall within a specific range. The formula for finding a z value is:

$$z = \frac{X - \mu}{\sigma}$$

In this formula X is the value of the random variable, μ is the population mean, and σ the population standard deviation.

However, most business decisions refer to a sample—not just one observation. So we are interested in the distribution of $\overline{X}$, the sample mean, instead of X, the value of one observation. That is the first change we make in formula (7–5). The second is that we use the standard error of the mean of n observations instead of the population standard deviation. That is, we use $\sigma/\sqrt{n}$ in the denominator rather than σ. Therefore, to find the likelihood of a sample mean with a specified range, we first use the following formula to find the corresponding z value. Then we use Appendix B.1 to locate the probability.

FINDING THE z VALUE OF $\overline{X}$ WHEN THE POPULATION STANDARD DEVIATION IS KNOWN

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \qquad \textbf{[8–2]}$$

The following example will show the application.

Example

The Quality Assurance Department for Cola, Inc., maintains records regarding the amount of cola in its Jumbo bottle. The actual amount of cola in each bottle is critical, but varies a small amount from one bottle to the next. Cola, Inc., does not wish to underfill the bottles, because it will have a problem with truth in labeling. On the other hand, it cannot overfill each bottle, because it would be giving cola away, hence reducing its profits. Its records indicate that the amount of cola follows the normal probability distribution. The mean amount per bottle is 31.2 ounces and the population standard deviation is 0.4 ounces. At 8 A.M. today the quality technician randomly selected 16 bottles from the filling line. The mean amount of cola contained in the bottles is 31.38 ounces. Is this an unlikely result? Is it likely the process is putting too much soda in the bottles? To put it another way, is the sampling error of 0.18 ounces unusual?

Solution

We can use the results of the previous section to find the likelihood that we could select a sample of 16 (n) bottles from a normal population with a mean of 31.2 (μ) ounces and a population standard deviation of 0.4 (σ) ounces and find the sample mean to be 31.38($\overline{X}$). We use formula (8–2) to find the value of z.

$$z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \frac{31.38 - 31.20}{0.4/\sqrt{16}} = 1.80$$

The numerator of this equation, $\overline{X} - \mu = 31.38 - 31.20 = .18$, is the sampling error. The denominator, $\sigma/\sqrt{n} = 0.4/\sqrt{16} = 0.1$, is the standard error of the sampling distribution of the sample mean. So the z values express the sampling error in standard units, in other words, the standard error.

Next, we compute the likelihood of a z value greater than 1.80. In Appendix B.1 locate the probability corresponding to a z value of 1.80. It is .4641. The likelihood of a z value greater than 1.80 is .0359, found by .5000 − .4641.

What do we conclude? It is unlikely, less than a 4 percent chance, we could select a sample of 16 observations from a normal population with a mean of 31.2 ounces and a population standard deviation of 0.4 ounces and find the sample mean equal to or greater than 31.38 ounces. We conclude the process is putting too much cola in the bottles. The quality technician should see the production supervisor about reducing the amount of soda in each bottle. This information is summarized in Chart 8–6.

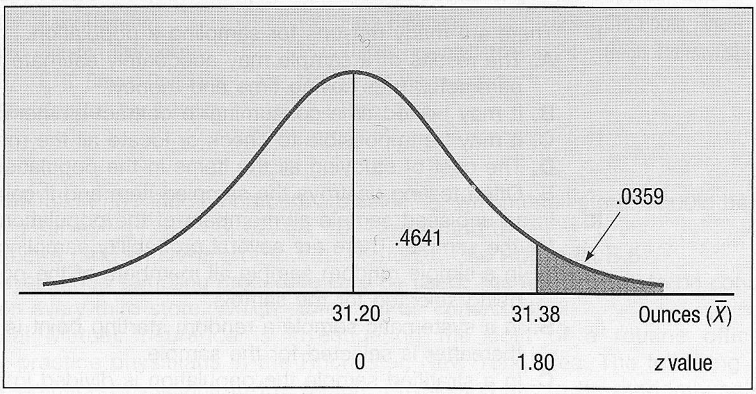
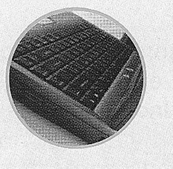


CHART 8–6 Sampling Distribution of the Mean Amount of Cola in a Jumbo Bottle

Self-Review 8–5



Refer to the Cola, Inc., information. Suppose the quality technician selected a sample of 16 Jumbo bottles that averaged 31.08 ounces. What can you conclude about the filling process?

Exercises

15. A normal population has a mean of 60 and a standard deviation of 12. You select a random sample of 9. Compute the probability the sample mean is:
  a. Greater than 63.
  b. Less than 56.
  c. Between 56 and 63.

16. A normal population has a mean of 75 and a standard deviation of 5. You select a sample of 40. Compute the probability the sample mean is:
  a. Less than 74.
  b. Between 74 and 76.
  c. Between 76 and 77.
  d. Greater than 77.

17. The rent for a one-bedroom apartment in Southern California follows the normal distribution with a mean of $2,200 per month and a standard deviation of $250 per month. The distribution of the monthly costs does not follow the normal distribution. In fact, it is positively skewed. What is the probability of selecting a sample of 50 one-bedroom apartments and finding the mean to be at least $1,950 per month?

18. According to an IRS study, it takes a mean of 330 minutes for taxpayers to prepare, copy, and electronically file a 1040 tax form. This distribution of times follows the normal distribution and the standard deviation is 80 minutes. A consumer watchdog agency selects a random sample of 40 taxpayers.
  a. What is the standard error of the mean in this example?
  b. What is the likelihood the sample mean is greater than 320 minutes?
  c. What is the likelihood the sample mean is between 320 and 350 minutes?
  d. What is the likelihood the sample mean is greater than 350 minutes?

Chapter Summary

I. There are many reasons for sampling a population.
  A. The results of a sample may adequately estimate the value of the population parameter, thus saving time and money.
  B. It may be too time consuming to contact all members of the population.
  C. It may be impossible to check or locate all the members of the population.
  D. The cost of studying all the items in the population may be prohibitive.
  E. Often testing destroys the sampled item and it cannot be returned to the population.

II. In an unbiased sample all members of the population have a chance of being selected for the sample. There are several probability sampling methods.
  A. In a simple random sample all members of the population have the same chance of being selected for the sample.
  B. In a systematic sample a random starting point is selected, and then every *k*th item thereafter is selected for the sample.
  C. In a stratified sample the population is divided into several groups, called strata, and then a random sample is selected from each stratum.
  D. In cluster sampling the population is divided into primary units, then samples are drawn from the primary units.

III. The sampling error is the difference between a population parameter and a sample statistic.

IV. The sampling distribution of the sample mean is a probability distribution of all possible sample means of the same sample size.

A. For a given sample size, the mean of all possible sample means selected from a population is equal to the population mean.

B. There is less variation in the distribution of the sample mean than in the population distribution.

C. The standard error of the mean measures the variation in the sampling distribution of the sample mean. The standard error is found by:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \quad [8\text{–}1]$$

D. If the population follows a normal distribution, the sampling distribution of the sample mean will also follow the normal distribution for samples of any size. Assume the population standard deviation is known. To determine the probability that a sample mean falls in a particular region, use the following formula.

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad [8\text{–}2]$$

Pronunciation Key

SYMBOL	MEANING	PRONUNCIATION
$\mu_{\bar{X}}$	Mean of the sampling distribution of the sample mean	*mu sub X bar*
$\sigma_{\bar{X}}$	Population standard error of the sample mean	*sigma sub X bar*

Chapter Exercises

19. The retail stores located in the North Towne Square Mall are:

00	Elder-Beerman	09	Lion Store	18	County Seat
01	Sears	10	Bootleggers	19	Kid Mart
02	Deb Shop	11	Formal Man	20	Lerner
03	Frederick's of Hollywood	12	Leather Ltd.	21	Coach House Gifts
04	Petries	13	B Dalton Bookseller	22	Spencer Gifts
05	Easy Dreams	14	Pat's Hallmark	23	CPI Photo Finish
06	Summit Stationers	15	Things Remembered	24	Regis Hairstylists
07	E. B. Brown Opticians	16	Pearle Vision Express		
08	Kay-Bee Toy & Hobby	17	Dollar Tree		

a. If the following random numbers are selected, which retail stores should be contacted for a survey? 11, 65, 86, 62, 06, 10, 12, 77, and 04

b. Select a random sample of four retail stores. Use Appendix B.6.

c. A systematic sampling procedure is to be used. The first store is to be contacted and then every third store. Which stores will be contacted?

20. Medical Mutual Insurance is investigating the cost of a routine office visit to family-practice physicians in the Rochester, New York, area. The following is a list of family-practice physicians in the region. Physicians are to be randomly selected and contacted regarding their charges. The 39 physicians have been coded from 00 to 38. Also noted is whether they are in practice by themselves (S), have a partner (P), or are in a group practice (G).

Number	Physician	Type of Practice	Number	Physician	Type of Practice
00	R. E. Scherbarth, M.D.	S	20	Gregory Yost, M.D.	P
01	Crystal R. Goveia, M.D.	P	21	J. Christian Zona, M.D.	P
02	Mark D. Hillard, M.D.	P	22	Larry Johnson, M.D.	P
03	Jeanine S. Huttner, M.D.	P	23	Sanford Kimmel, M.D.	P
04	Francis Aona, M.D.	P	24	Harry Mayhew, M.D.	S
05	Janet Arrowsmith, M.D.	P	25	Leroy Rodgers, M.D.	S
06	David DeFrance, M.D.	S	26	Thomas Tafelski, M.D.	S
07	Judith Furlong, M.D.	S	27	Mark Zilkoski, M.D.	G
08	Leslie Jackson, M.D.	G	28	Ken Bertka, M.D.	G
09	Paul Langenkamp, M.D.	S	29	Mark DeMichiei, M.D.	G
10	Philip Lepkowski, M.D.	S	30	John Eggert, M.D.	P
11	Wendy Martin, M.D.	S	31	Jeanne Fiorito, M.D.	P
12	Denny Mauricio, M.D.	P	32	Michael Fitzpatrick, M.D.	P
13	Hasmukh Parmar, M.D.	P	33	Charles Holt, D.O.	P
14	Ricardo Pena, M.D.	P	34	Richard Koby, M.D.	P
15	David Reames, M.D.	P	35	John Meier, M.D.	P
16	Ronald Reynolds, M.D.	G	36	Douglas Smucker, M.D.	S
17	Mark Steinmetz, M.D.	G	37	David Weldy, M.D.	P
18	Geza Torok, M.D.	S	38	Cheryl Zaborowski, M.D.	P
19	Mark Young, M.D.	P			

a. The random numbers obtained from Appendix B.6 are: 31, 94, 43, 36, 03, 24, 17, and 09. Which physicians should be contacted?
b. Select a random sample of four physicians using the random numbers of Appendix B.6.
c. A sample is to consist of every fifth physician. The number 04 is selected as the starting point. Which physicians will be contacted?
d. A sample is to consist of two physicians in solo practice (S), two in partnership (P), and one in group practice (G). Select a sample accordingly. Explain your procedure.

21. What is sampling error? Could the value of the sampling error be zero? If it were zero, what would this mean?
22. List the reasons for sampling. Give an example of each reason for sampling.
23. The manufacturer of eMachines, an economy-priced computer, recently completed the design for a new laptop model. eMachine's top management would like some assistance in pricing the new laptop. Two market research firms were contacted and asked to prepare a pricing strategy. Marketing-Gets-Results tested the new eMachines laptop with 50 randomly selected consumers, who indicated they plan to purchase a laptop within the next year. The second marketing research firm, called Marketing-Reaps-Profits, test-marketed the new eMachines laptop with 200 current laptop owners. Which of the marketing research companies' test results will be more useful? Discuss why.
24. Answer the following questions in one or two well-constructed sentences.
a. What happens to the standard error of the mean if the sample size is increased?
b. What happens to the distribution of the sample means if the sample size in increased?
c. When using the distribution of sample means to estimate the population mean, what is the benefit of using larger sample sizes?
25. There are 25 motels in Goshen, Indiana. The number of rooms in each motel follows:

90	72	75	60	75	72	84	72	88	74	105	115	68	74	80	64	104	82	48	58	60	80	48	58	100

a. Using a table of random numbers (Appendix B.6), select a random sample of five motels from this population.
b. Obtain a systematic sample by selecting a random starting point among the first five motels and then select every fifth motel.
c. Suppose the last five motels are "cut-rate" motels. Describe how you would select a random sample of three regular motels and two cut-rate motels.

26. As a part of their customer-service program, United Airlines randomly selected 10 passengers from today's 9 A.M. Chicago–Tampa flight. Each sampled passenger is to be interviewed in depth regarding airport facilities, service, and so on. To identify the sample, each passenger was given a number on boarding the aircraft. The numbers started with 001 and ended with 250.

a. Select 10 usable numbers at random using Appendix B.6.
b. The sample of 10 could have been chosen using a systematic sample. Choose the first number using Appendix B.6, and then list the numbers to be interviewed.
c. Evaluate the two methods by giving the advantages and possible disadvantages.
d. In what other way could a random sample be selected from the 250 passengers?

27. Suppose your statistics instructor gave six examinations during the semester. You received the following grades (percent correct): 79, 64, 84, 82, 92, and 77. Instead of averaging the six scores, the instructor indicated he would randomly select two grades and compute the final percent correct based on the two percents.

a. How many different samples of two test grades are possible?
b. List all possible samples of size two and compute the mean of each.
c. Compute the mean of the sample means and compare it to the population mean.
d. If you were a student, would you like this arrangement? Would the result be different from dropping the lowest score? Write a brief report.

28. At the downtown office of First National Bank there are five tellers. Last week the tellers made the following number of errors each: 2, 3, 5, 3, and 5.

a. How many different samples of 2 tellers are possible?
b. List all possible samples of size 2 and compute the mean of each.
c. Compute the mean of the sample means and compare it to the population mean.

29. The Quality Control Department employs five technicians during the day shift. Listed below is the number of times each technician instructed the production foreman to shut down the manufacturing process last week.

Technician	Shutdowns
Taylor	4
Hurley	3
Gupta	5
Rousche	3
Huang	2

a. How many different samples of two technicians are possible from this population?
b. List all possible samples of two observations each and compute the mean of each sample.
c. Compare the mean of the sample means with the population mean.
d. Compare the shape of the population distribution with the shape of the distribution of the sample means.

30. The Appliance Center has six sales representatives at its North Jacksonville outlet. Listed below is the number of refrigerators sold by each last month.

Sales Representative	Number Sold
Zina Craft	54
Woon Junge	50
Ernie DeBrul	52
Jan Niles	48
Molly Camp	50
Rachel Myak	52

a. How many samples of size two are possible?
b. Select all possible samples of size two and compute the mean number sold.
c. Organize the sample means into a frequency distribution.
d. What is the mean of the population? What is the mean of the sample means?
e. What is the shape of the population distribution?
f. What is the shape of the distribution of the sample mean?

31. Mattel Corporation produces a remote-controlled car that requires three AA batteries. The mean life of these batteries in this product is 35.0 hours. The distribution of the battery lives closely follows the normal probability distribution with a standard deviation of 5.5 hours. As a part of its testing program Sony tests samples of 25 batteries.
 a. What can you say about the shape of the distribution of the sample mean?
 b. What is the standard error of the distribution of the sample mean?
 c. What proportion of the samples will have a mean useful life of more than 36 hours?
 d. What proportion of the sample will have a mean useful life greater than 34.5 hours?
 e. What proportion of the sample will have a mean useful life between 34.5 and 36.0 hours?

32. CRA CDs, Inc., wants the mean lengths of the "cuts" on a CD to be 135 seconds (2 minutes and 15 seconds). This will allow the disk jockeys to have plenty of time for commercials within each 10-minute segment. Assume the distribution of the length of the cuts follows the normal distribution with a population standard deviation of 8 seconds. Suppose we select a sample of 16 cuts from various CDs sold by CRA CDs, Inc.
 a. What can we say about the shape of the distribution of the sample mean?
 b. What is the standard error of the mean?
 c. What percent of the sample means will be greater than 140 seconds?
 d. What percent of the sample means will be greater than 128 seconds?
 e. What percent of the sample means will be greater than 128 but less than 140 seconds?

33. Recent studies indicate that the typical 50-year-old woman spends $350 per year for personal-care products. The distribution of the amounts spent follows a normal distribution with a standard deviation of $45 per year. We select a random sample of 40 women. The mean amount spent for those sampled is $335. What is the likelihood of finding a sample mean this large or larger from the specified population?

34. Information from the American Institute of Insurance indicates the mean amount of life insurance per household in the United States is $110,000. This distribution follows the normal distribution with a standard deviation of $40,000.
 a. If we select a random sample of 50 households, what is the standard error of the mean?
 b. What is the expected shape of the distribution of the sample mean?
 c. What is the likelihood of selecting a sample with a mean of at least $112,000?
 d. What is the likelihood of selecting a sample with a mean of more than $100,000?
 e. Find the likelihood of selecting a sample with a mean of more than $100,000 but less than $112,000.

35. The mean age at which men in the United States marry for the first time follows the normal distribution with a mean of 24.8 years. The standard deviation of the distribution is 2.5 years. For a random sample of 60 men, what is the likelihood that the age at which they were married for the first time is less than 25.1 years?

36. A recent study by the Greater Los Angeles Taxi Drivers Association showed that the mean fare charged for service from Hermosa Beach to Los Angeles International Airport is $18.00 and the standard deviation is $3.50. We select a sample of 15 fares.
 a. What is the likelihood that the sample mean is between $17.00 and $20.00?
 b. What must you assume to make the above calculation?

37. Crossett Trucking Company claims that the mean weight of its delivery trucks when they are fully loaded is 6,000 pounds and the standard deviation is 150 pounds. Assume that the population follows the normal distribution. Forty trucks are randomly selected and weighed. Within what limits will 95 percent of the sample means occur?

38. The mean amount purchased by a typical customer at Churchill's Grocery Store is $23.50 with a standard deviation of $5.00. Assume the distribution of amounts purchased follows the normal distribution. For a sample of 50 customers, answer the following questions.
 a. What is the likelihood the sample mean is at least $25.00?
 b. What is the likelihood the sample mean is greater than $22.50 but less than $25.00?
 c. Within what limits will 90 percent of the sample means occur?

39. The mean SAT score for Division I student-athletes is 947 with a standard deviation of 205. If you select a random sample of 60 of these students, what is the probability the mean is below 900?

40. Suppose we roll a fair die two times.
 a. How many different samples are there?
 b. List each of the possible samples and compute the mean.
 c. On a chart similar to Chart 8–1, compare the distribution of sample means with the distribution of the population.
 d. Compute the mean and the standard deviation of each distribution and compare them.

41. The following table lists the per capita personal income for each of the 50 states in 2004.

Number	State	2004	Number	State	2004
0	Alabama	$27,795	25	Montana	$26,857
1	Alaska	34,454	26	Nebraska	31,339
2	Arizona	28,442	27	Nevada	33,405
3	Arkansas	25,725	28	New Hampshire	37,040
4	California	35,019	29	New Jersey	41,332
5	Colorado	36,063	30	New Mexico	26,191
6	Connecticut	45,398	31	New York	38,228
7	Delaware	35,861	32	North Carolina	29,246
8	Florida	31,455	33	North Dakota	31,398
9	Georgia	30,051	34	Ohio	31,322
10	Hawaii	32,160	35	Oklahoma	28,089
11	Idaho	27,098	36	Oregon	29,971
12	Illinois	34,351	37	Pennsylvania	33,348
13	Indiana	30,094	38	Rhode Island	33,733
14	Iowa	30,560	39	South Carolina	27,172
15	Kansas	30,811	40	South Dakota	30,856
16	Kentucky	27,709	41	Tennessee	30,005
17	Louisiana	27,581	42	Texas	30,222
18	Maine	30,566	43	Utah	26,606
19	Maryland	39,247	44	Vermont	32,770
20	Massachusetts	41,801	45	Virginia	35,477
21	Michigan	31,954	46	Washington	35,299
22	Minnesota	35,861	47	West Virginia	25,872
23	Mississippi	24,650	48	Wisconsin	32,157
24	Missouri	30,608	49	Wyoming	34,306

a. You wish to select a sample of eight from this list. The selected random numbers are 45, 15, 81, 09, 39, 43, 90, 26, 06, 45, 01, and 42. Which states are included in the sample?

b. You wish to use a systematic sample of every sixth item and the digit 02 is chosen as the starting point. Which states are included?

42. Human Resource Consulting (HRC) is surveying a sample of 60 firms in order to study health care costs for a client. One of the items being tracked is the annual deductible that employees must pay. The state Bureau of Labor reports the mean of this distribution is $502 with a standard deviation of $100.

a. Compute the standard error of the sample mean for HRC.

b. What is the chance HRC finds a sample mean between $477 and $527?

c. Calculate the likelihood that the sample mean is between $492 and $512.

d. What is the probability the sample mean is greater that $550?

43. Over the past decade the mean number of members of the Information Systems Security Association who have experienced a denial-of-service attack each year is 510 with a standard deviation of 14.28 attacks. Suppose nothing in this environment changes.

a. What is the likelihood this group will suffer an average of more than 600 attacks in the next 10 years?

b. Compute the probability the mean number of attacks over the next 10 years is between 500 and 600.

c. What is the possibility they will experience an average of less than 500 attacks over the next 10 years?

44. The Oil Price Information Center reports the mean price per gallon of regular gasoline is $3.26 with a population standard deviation of $0.18. Assume a random sample of 40 gasoline stations is selected and their mean cost for regular gasoline is computed.

a. What is the standard error of the mean in this experiment?

b. What is the probability that the sample mean is between $3.24 and $3.28?

c. What is the probability that the difference between the sample mean and the population mean is less than 0.01?

d. What is the likelihood the sample mean is greater than $3.34?

45. Nike's annual report says that the average American buys 6.5 pairs of sports shoes per year. Suppose the population standard deviation is 2.1 and that a sample of 81 customers will be examined next year.
   a. What is the standard error of the mean in this experiment?
   b. What is the probability that the sample mean is between 6 and 7 pairs of sports shoes?
   c. What is the probability that the difference between the sample mean and the population mean is less than 0.25 pairs?
   d. What is the likelihood the sample mean is greater than 7 pairs?

exercises.com

WWW

46. You need to find the "typical" or mean annual dividend per share for large banks. You decide to sample six banks listed on the New York Stock Exchange. These banks and their trading symbols follow.

Bank	Symbol	Bank	Symbol
AmSouth Bancorporation	ASO	National City Corp.	NCC
Bank of America Corp.	BAC	Northern Trust Corp.	NTRS
Bank of New York	BK	PNC Financial Services Group	PNC
BB&T Corp.	BBT	Regions Financial Corp.	RF
Charter One Financial	CF	SouthTrust Corp.	SOTR
Comerica, Inc.	CMA	SunTrust Banks	STI
Fifth Third Bancorp	FITB	Synovus Financial Corp.	SNV
Golden West Financial	GDW	U.S. Bancorp	USB
Huntington Bancshares	HBAN	Wachovia Corp.	WB
JP Morgan Chase	JPM	Washington Mutual, Inc.	WM
KeyCorp	KEY	Wells Fargo & Co.	WFC
Mellon Financial Corp.	MEL	Zions Bancorp	ZION

   a. After numbering the banks from 01 to 24, which banks would be included in a sample if the random numbers were 14, 08, 24, 25, 05, 44, 02, and 22? Go to the following website: http://bigcharts.marketwatch.com. Enter the trading symbol for each of the sampled banks and record the price earnings ratio (P/E ratio). Determine the mean annual dividend per share for the sample of banks.
   b. Which banks are selected if you use a systematic sample of every fourth bank starting with the random number 03?

47. There are several websites that will report the 30 stocks that make up the Dow Jones Industrial Average (DJIA). One site is http:www.bloomberg.com/markets/stocks/movers_index_dow.ht. Compute the mean of the 30 stocks.
   a. Use a random number table, such as Appendix B.6, to select a random sample of five companies that make up the DJIA. Compute the sample mean. Compare the sample mean to the population mean. What did you find? What did you expect to find?
   b. You should not expect to find that the mean of these 30 stocks is the same as the current DJIA. Go to the website: http://www.investopedia.com/articles/02/082702 and read the reasons.

Data Set Exercises

48. Refer to the Real Estate data, which report information on the homes sold in the Denver, Colorado, area last year.
   a. Compute the mean and the standard deviation of the distribution of the selling prices for the homes. Assume this to be the population. Develop a histogram of the data. Would it seem reasonable from this histogram to conclude that the population of selling prices follows the normal distribution?
   b. Let's assume a normal population. Select a sample of 10 homes. Compute the mean and the standard deviation of the sample. Determine the likelihood of finding a sample mean this large or larger from the population.

49. Refer to the CIA data, which report demographic and economic information on 46 countries. Select a random sample of 10 countries. For this sample calculate the mean GDP/capita. Repeat this sampling and calculation process five more times. Then find the mean and standard deviation of your six sample means.

a. How do this mean and standard deviation compare with the mean and standard deviation of the original "population" of 46 countries?

b. Make a histogram of the six means and discuss whether the distribution is normal.

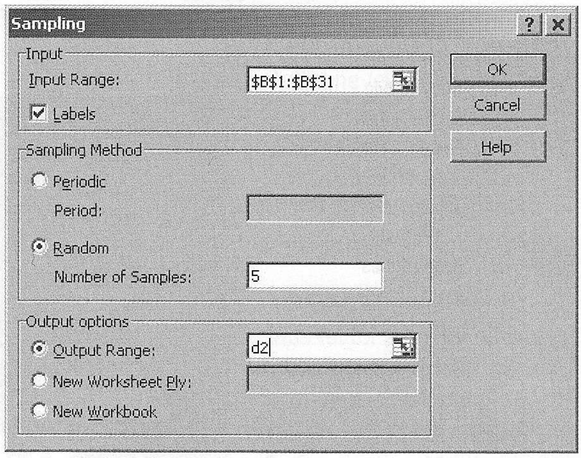
c. Suppose the population distribution is normal. For the first sample mean you computed, determine the likelihood of finding a sample mean this large or larger from the population.

Software Commands

1. The Excel commands to select a simple random sample on page 264 are:

a. Select **Tools, Data Analysis,** and then **Sampling,** and click **OK.**

b. For **Input Range** insert *B1:B31*. Since the column is named, click the **Labels** box. Select **Random,** and enter the sample size for the **Number of Samples,** in this case *5*. Click on **Output Range** and indicate the place in the spreadsheet you want the sample information. Note that your sample results will differ from those in the text. Also recall that Excel samples with replacement, so it is possible for a population value to appear more than once in the sample.



Chapter 8 Answers to Self-Review

8–1 **a.** Students selected are Price, Detley, and Molter.
b. Answers will vary.
c. Skip it and move to the next random number.

8–2 The students selected are Berry, Francis, Kopp, Poteau, and Swetye.

8–3 **a.** 10, found by:

$$_5C_2 = \frac{5!}{2!(5-2)!}$$

b.

	Service	Sample Mean
Snow, Tolson	20, 22	21
Snow, Kraft	20, 26	23
Snow, Irwin	20, 24	22
Snow, Jones	20, 28	24
Tolson, Kraft	22, 26	24
Tolson, Irwin	22, 24	23
Tolson, Jones	22, 28	25
Kraft, Irwin	26, 24	25
Kraft, Jones	26, 28	27
Irwin, Jones	24, 28	26

c.

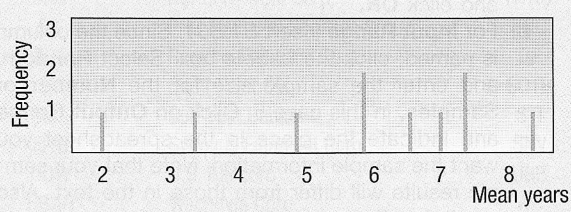
Mean	Number	Probability
21	1	.10
22	1	.10
23	2	.20
24	2	.20
25	2	.20
26	1	.10
27	1	.10
	10	1.00

d. Identical: population mean, μ, is 24, and mean of sample means, $\mu_{\bar{X}}$, is also 24.
e. Sample means range from 21 to 27. Population values go from 20 to 28.
f. Nonnormal.
g. Yes.

8–4 The answers will vary. Here is one solution.

	Sample Number									
	1	2	3	4	5	6	7	8	9	10
	8	2	2	19	3	4	0	4	1	2
	19	1	14	9	2	5	8	2	14	4
	8	3	4	2	4	4	1	14	4	1
	0	3	2	3	1	2	16	1	2	3
	2	1	7	2	19	18	18	16	3	7
Total	37	10	29	35	29	33	43	37	24	17
$\bar{X}$	7.4	2	5.8	7.0	5.8	6.6	8.6	7.4	4.8	3.4

Mean of the 10 sample means is 5.88.



8–5 $$z = \frac{31.08 - 31.20}{0.4/\sqrt{16}} = -1.20$$

The probability that z is greater than -1.20 is $.5000 + .3849 = .8849$. There is more than an 88 percent chance the filling operation will produce bottles with at least 31.08 ounces.