9

Estimation and Confidence Intervals

The American Restaurant Association collected information on the number of meals eaten outside the home per week by young married couples. A survey of 60 couples showed the sample mean number of meals eaten outside the home was 2.76 meals per week, with a standard deviation of 0.75 meals per week. Construct a 97 percent confidence interval for the population mean. (See Goal 3 and Exercise 36.)

GOALS

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

1 Define a *point estimate*.

2 Define *level of confidence*.

3 Construct a confidence interval for the population mean when the population standard deviation is known.

4 Construct a confidence interval for a population mean when the population standard deviation is unknown.

5 Construct a confidence interval for a population proportion.

6 Determine the sample size for attribute and variable sampling.

Statistics in Action

On all new cars, a fuel economy estimate is prominently displayed on the window sticker as required by the Environmental Protection Agency (EPA). Often, fuel economy is a factor in a consumer's choice of a new car because of fuel costs or environmental concerns. For example, a 2006 Toyota Camry's (4 cylinder, automatic) fuel estimates are 34 miles per gallon on the highway and 24 mpg in the city. The EPA recognizes that actual fuel economy may differ from the estimates by noting, "No test can simulate all possible combinations of conditions and climate, driver behavior, and car care habits. Actual mileage depends on how, when, and where the vehicle is driven. EPA has found that the mpg obtained by most drivers will be within a few mpg of the estimates . . . " In fact, the window sticker also includes an interval estimate for fuel economy: 19 to 27 mpg in the city and 27 to 37 mpg on the highway. http://www.fueleconomy.gov/

Introduction

The previous chapter began our discussion of statistical inference. It introduced the reasons and methods of sampling. The reasons for sampling were:

- To contact the entire population is too time consuming.
- The cost of studying all the items in the population is often too expensive.
- The sample results are usually adequate.
- Certain tests are destructive.
- Checking all the items is physically impossible.

There are several methods of sampling. Simple random sampling is the most widely used method. With this type of sampling, each member of the population has the same chance of being selected to be a part of the sample. Other methods of sampling include systematic sampling, stratified sampling, and cluster sampling.

Chapter 8 assumes information about the population, such as the mean, the standard deviation, or the shape of the population. In most business situations, such information is not available. In fact, the purpose of sampling may be to estimate some of these values. For example, you select a sample from a population and use the mean of the sample to estimate the mean of the population.

This chapter considers several important aspects of sampling. We begin by studying **point estimates.** A point estimate is a single value (point) derived from a sample and used to estimate a population value. For example, suppose we select a sample of 50 junior executives and ask each the number of hours they worked last week. Compute the mean of this sample of 50 and use the value of the sample mean as a point estimate of the unknown population mean. However, a point estimate is a single value. A more informative approach is to present a range of values in which we expect the population parameter to occur. Such a range of values is called a **confidence interval.**

Frequently in business we need to determine the size of a sample. How many voters should a polling organization contact to forecast the election outcome? How many products do we need to examine to ensure our quality level? This chapter also develops a strategy for determining the appropriate size of the sample.

Point Estimates and Confidence Intervals for a Mean

We begin the study of point estimates and confidence intervals by studying estimates of the population mean. We will consider two cases, where

- The population standard deviation (σ) is known,
- The population standard deviation is unknown. In this case we substitute the sample standard deviation (s) for the population standard deviation (σ).

There are important distinctions in the assumptions between these two situations. We consider the case where σ is known first.

Population Standard Deviation, Known (σ)

In the previous chapter, the data on the length of service of Spence Sprockets employees, presented in the example on page 276, is a population because we present the length of service for all 40 employees. In that case we can easily compute the population mean. We have all the data and the population is not too large. In most situations, however, the population is large or it is difficult to identify all members of the population, so we need to rely on sample information. In other words, we do not know the population parameter and we therefore want to estimate the value from a sample statistic. Consider the following business situations.

1. Tourism is a major source of income for many Caribbean countries, such as Barbados. Suppose the Bureau of Tourism for Barbados wants an estimate of the mean amount spent by tourists visiting the country. It would not be feasible to contact each tourist. Therefore, 500 tourists are randomly selected as they depart the country and asked in detail about their spending while visiting the island. The mean amount spent by the sample of 500 tourists is an estimate of the unknown population parameter. That is, we let $\overline{X}$, the sample mean, serve as an estimate of μ, the population mean.
2. Centex Home Builders, Inc., builds quality homes in the southeastern region of the United States. One of the major concerns of new buyers is the date on which the home will be completed. In recent times Centex has been telling customers, "Your home will be completed 45 working days from the date we begin installing drywall." The customer relations department at Centex wishes to compare this pledge with recent experience. A sample of 50 homes completed this year revealed the mean number of working days from the start of drywall to the completion of the home was 46.7 days. Is it reasonable to conclude that the population mean is still 45 days and that the difference between the sample mean (46.7 days) and the proposed population mean is sampling error?
3. Recent medical studies indicate that exercise is an important part of a person's overall health. The director of human resources at OCF, a large glass manufacturer, wants an estimate of the number of hours per week employees spend exercising. A sample of 70 employees reveals the mean number of hours of exercise last week is 3.3. The sample mean of 3.3 hours estimates the unknown population mean, the mean hours of exercise for all employees.

A point estimate is a single statistic used to estimate a population parameter. Suppose Best Buy, Inc., wants to estimate the mean age of buyers of HD plasma televisions. It selects a random sample of 50 recent purchasers, determines the age of each purchaser, and computes the mean age of the buyers in the sample. The mean of this sample is a point estimate of the mean of the population.

POINT ESTIMATE The statistic, computed from sample information, which is used to estimate the population parameter.

The sample mean, $\overline{X}$, is a point estimate of the population mean, μ; p, a sample proportion, is a point estimate of π, the population proportion; and s, the sample standard deviation, is a point estimate of σ, the population standard deviation.

A point estimate, however, tells only part of the story. While we expect the point estimate to be close to the population parameter, we would like to measure how close it really is. A confidence interval serves this purpose.

CONFIDENCE INTERVAL A range of values constructed from sample data so that the population parameter is likely to occur within that range at a specified probability. The specified probability is called the *level of confidence*.

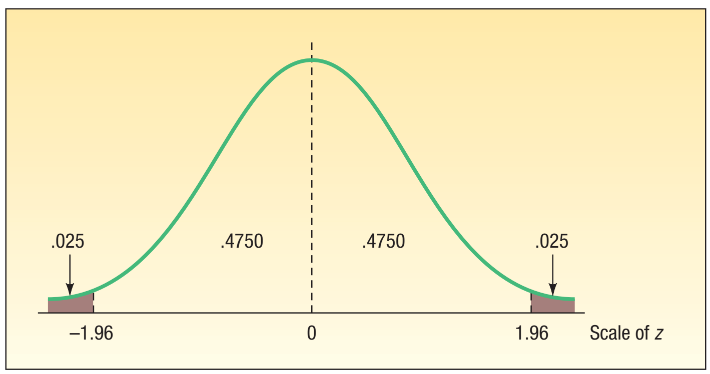
For example, we estimate the mean yearly income for construction workers in the New York–New Jersey area is \$65,000. The range of this estimate might be from \$61,000 to \$69,000. We can describe how confident we are that the population parameter is in the interval by making a probability statement. We might say, for instance, that we are 90 percent sure that the mean yearly income of construction workers in the New York–New Jersey area is between \$61,000 and \$69,000.

The information developed about the shape of a sampling distribution of the sample mean, that is, the sampling distribution of $\overline{X}$, allows us to locate an interval that has a specified probability of containing the population mean, μ. For reasonably large samples, the results of the central limit theorem allow us to state the following:

1. Ninety-five percent of the sample means selected from a population will be within 1.96 standard deviations of the population mean μ.
2. Ninety-nine percent of the sample means will lie within 2.58 standard deviations of the population mean.

The standard deviation discussed here is the standard deviation of the sampling distribution of the sample mean. It is usually called the *standard error*. Intervals computed in this fashion are called the **95 percent confidence interval** and the **99 percent confidence interval**. How are the values of 1.96 and 2.58 obtained? The *95 percent* and *99 percent* refer to the percent of similarly constructed intervals that would include the parameter being estimated. The *95 percent,* for example, refers to the middle 95 percent of the observations. Therefore, the remaining 5 percent are equally divided between the two tails.

See the following diagram.



We use Appendix B.1 to find the appropriate z values. Locate .4750 in the body of the table. Read the corresponding row and column values. The value is 1.96. Thus, the probability of finding a z value between 0 and 1.96 is .4750. Likewise, the probability of being in the interval between −1.96 and 0 is also .4750. When we combine these two, the probability of being in the interval −1.96 to 1.96 is .9500. Following is a portion of Appendix B.1. The z value for the 90 percent level of confidence is determined in a similar manner. It is 1.65. For a 99 percent level of confidence the z value is 2.58.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
⋮	⋮	⋮	⋮	⋮	⋮	⋮	·	⋮	⋮	⋮
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936

How do we determine a 95 percent confidence interval? The width of the interval is determined by the level of confidence and the size of the standard error of the mean. We've described above how to find the *z* value for a particular level of confidence. Recall from the previous chapter (see formula 8–1 on page 280) the standard error of the mean reports the variation in the distribution of sample means. It is really the standard deviation of the distribution of sample means. The formula is repeated below:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where

- $\sigma_{\bar{x}}$ is the symbol for the standard error of the mean. We use a Greek letter because it is a population value, and the subscript $\bar{x}$ reminds us that it refers to a sampling distribution of the sample means.
- σ is the population standard deviation.
- n is the number of observations in the sample.

The size of the standard error is affected by two values. The first is the standard deviation of the population. The larger the population standard deviation, σ, the larger $\sigma/\sqrt{n}$. If the population is homogenous, resulting in a small population standard deviation, the standard error will also be small. However, the standard error is also affected by the number of observations in the sample. A large sample will result in a small standard error of estimate, indicating that there is less variability in the sample means.

To explain these ideas, consider the following example. Del Monte Foods, Inc., distributes diced peaches in 4-ounce cans. To be sure each can contains at least the required amount Del Monte sets the filling operation to dispense 4.01 ounces of peaches and syrup in each can. So 4.01 is the population mean. Of course not every can will contain exactly 4.01 ounces of peaches and syrup. Some cans will have more and others less. Suppose the standard deviation of the process is .02 ounces. Let's also assume that the process follows the normal probability distribution. Now, we select a random sample of 16 cans and determine the sample mean. It is 4.015 ounces of peaches and syrup. The 95 percent confidence interval for the population mean of this particular sample is:

$$4.015 \pm 1.96(.02/\sqrt{16}) = 4.015 \pm .0098$$

The 95 percent confidence interval is between 4.0052 and 4.0248. Of course in this case we observe that the population mean of 4.01 ounces is in this interval. But that will not always be the case. Theoretically, if we selected 100 samples of 16 cans from the population, calculated the sample mean, and developed a confidence interval based on each *sample* mean, we would expect to find the *population* mean in about 95 of the 100 intervals.

We can summarize the following calculations for a 95 percent confidence interval in the following formula:

$$\bar{X} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

Similarly a 99 percent confidence interval is computed as follows.

$$\bar{X} \pm 2.58 \frac{\sigma}{\sqrt{n}}$$

As we discussed earlier, the values 1.96 and 2.58 are *z* values corresponding to the middle 95 percent and the middle 99 percent of the observations, respectively.

We are not restricted to the 95 and 99 percent levels of confidence. We can select any confidence level between 0 and 100 percent and find the corresponding value for *z*. In general, a confidence interval for the population mean when the population standard deviation is known is computed by:

CONFIDENCE INTERVAL FOR POPULATION MEAN WITH σ KNOWN

$$\bar{X} \pm z \frac{\sigma}{\sqrt{n}} \qquad [9\text{–}1]$$

where *z* depends on the level of confidence. Thus for a 92 percent level of confidence, the value of *z* in formula (9–1) is 1.75. The value of *z* is from Appendix B.1. This table is based on half the normal distribution, so .9200/2 = .4600. The closest value in the body of the table is .4599 and the corresponding *z* value is 1.75.

Frequently, we also use the 90 percent level of confidence. In this case, we want the area between 0 and *z* to be .4500, found by .9000/2. To find the *z* value for this level of confidence, move down the left column of Appendix B.1 to 1.6 and then over to the columns headed 0.04 and 0.05. The area corresponding to a *z* value of 1.64 is .4495, and for 1.65 it is .4505. To be conservative, we use 1.65. Try looking up the following levels of confidence and check your answers with the corresponding *z* values given on the right.

Confidence Level	Nearest Half Probability	*z* Value
80 percent	.3997	1.28
94 percent	.4699	1.88
96 percent	.4798	2.05

The following example shows the details for calculating a confidence interval and interpreting the result.

Example

The American Management Association wishes to have information on the mean income of middle managers in the retail industry. A random sample of 256 managers reveals a sample mean of $45,420. The standard deviation of this population is $2,050. The association would like answers to the following questions:

1. What is the population mean?
2. What is a reasonable range of values for the population mean?
3. What do these results mean?

Solution

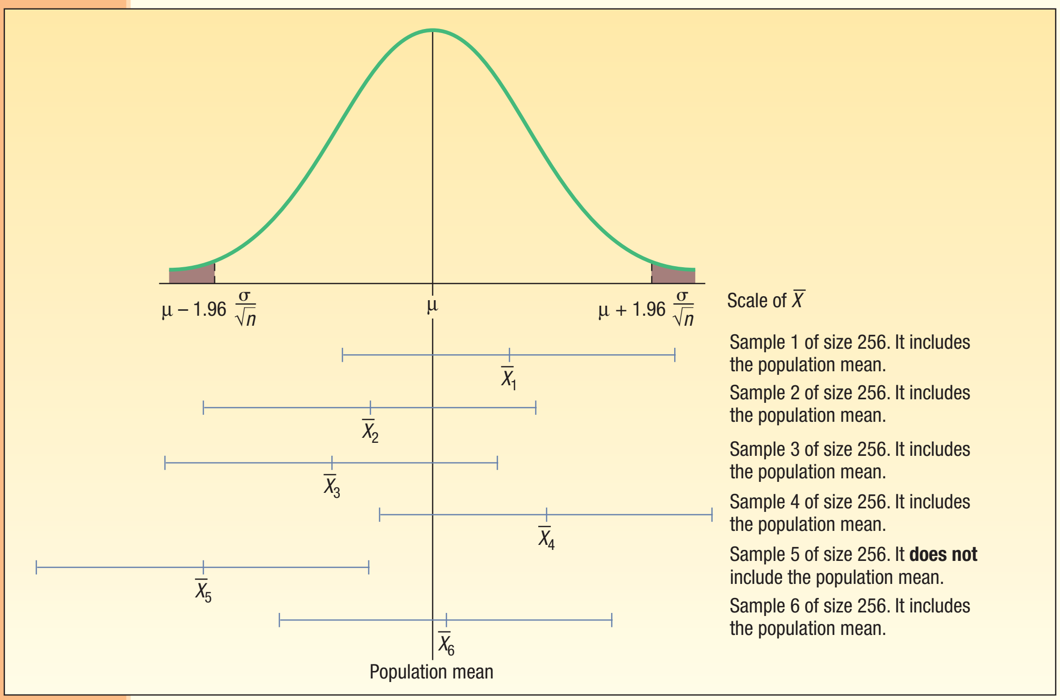
Generally, distributions of salary and income are positively skewed, because a few individuals earn considerably more than others, thus skewing the distribution in the positive direction. Fortunately, the central limit theorem stipulates that if we select a large sample, the distribution of the sample means will follow the normal distribution. In this instance, a sample of 256 middle managers is large enough that we can assume that the sampling distribution will follow the normal distribution. Now to answer the questions posed in the example.

1. **What is the population mean?** In this case, we do not know. We do know the sample mean is $45,420. Hence, our best estimate of the unknown population value is the corresponding sample statistic. Thus the sample mean of $45,420 is a *point estimate* of the unknown population mean.

2. **What is a reasonable range of values for the population mean?** The association decides to use the 95 percent level of confidence. To determine the corresponding confidence interval we use formula (9–1).

$$\bar{X} \pm z\frac{\sigma}{\sqrt{n}} = \$45{,}420 \pm 1.96\frac{\$2{,}050}{\sqrt{256}} = \$45{,}420 \pm \$251$$

   The usual practice is to round these endpoints to $45,169 and $45,671. These endpoints are called the *confidence limits*. The degree of confidence or the *level of confidence* is 95 percent and the confidence interval is from $45,169 to $45,671. The ±$251 is often referred to as the *margin of error*.
3. **What do these results mean?** Suppose we select many samples of 256 managers, perhaps several hundred. For each sample, we compute the mean and then construct a 95 percent confidence interval, such as we did in the previous section. We could expect about 95 percent of these confidence intervals to contain the *population* mean. About 5 percent of the intervals would not contain the population mean annual income, which is μ. However, a particular confidence interval either contains the population parameter or it does not. The following diagram shows the results of selecting samples from the population of middle managers in the retail industry, computing the mean of each and then, using formula (9–1), determining a 95 percent confidence interval for the population mean. Note that not all intervals include the population mean. Both the endpoints of the fifth sample are less than the population mean. We attribute this to sampling error, and it is the risk we assume when we select the level of confidence.



A Computer Simulation

With the aid of a computer, we can randomly select samples from a population, quickly compute the confidence interval, and show how confidence intervals usually, but not always, include the population parameter. The following example will help to explain.

Example

From many years in the automobile leasing business, Town Bank knows the mean distance driven on a four-year lease is 50,000 miles and the standard deviation is 5,000. Suppose, using the MINITAB statistical software system, we want to find what proportion of the 95 percent confidence intervals will include the population mean of 50. To make the calculations easier to understand, we'll conduct the study in thousands of miles, instead of miles. We select 60 random samples of size 30 from a population with a mean of 50 and a standard deviation of 5.

Solution

The results of 60 random samples of 30 automobiles each are summarized in the computer output below. Of the 60 confidence intervals with a 95 percent confidence level, 2, or 3.33 percent, did not include the population mean of 50. The intervals (C3 and C59) that do *not* include the population mean are highlighted. 3.33 percent is close to the estimate that 5 percent of the intervals will not include the population mean, and the 58 of 60, or 96.67 percent, is close to 95 percent.

To explain the first calculation in more detail: MINITAB began by selecting a random sample of 30 observations from a population with a mean of 50 and a standard deviation of 5. The mean of these 30 observations is 50.053. The sampling error is 0.053, found by $\bar{X} - \mu = 50.053 - 50.000$. The endpoints of the confidence interval are 48.264 and 51.842. These endpoints are determined by using formula (9–1):

$$\bar{X} \pm 1.96\frac{\sigma}{\sqrt{n}} = 50.053 \pm 1.96\frac{5}{\sqrt{30}} = 50.053 \pm 1.789$$

One-Sample Z:

The assumed sigma = 5

Variable	N	Mean	StDev	SE Mean	95.0% CI
C1	30	50.053	5.002	0.913	(48.264, 51.842)
C2	30	49.025	4.450	0.913	(47.236, 50.815)
C3	30	52.023	5.918	0.913	(50.234, 53.812)
C4	30	50.056	3.364	0.913	(48.267, 51.845)
C5	30	49.737	4.784	0.913	(47.948, 51.526)
C6	30	51.074	5.495	0.913	(49.285, 52.863)
C7	30	50.040	5.930	0.913	(48.251, 51.829)
C8	30	48.910	3.645	0.913	(47.121, 50.699)
C9	30	51.033	4.918	0.913	(49.244, 52.822)
C10	30	50.692	4.571	0.913	(48.903, 52.482)
C11	30	49.853	4.525	0.913	(48.064, 51.642)
C12	30	50.286	3.422	0.913	(48.497, 52.076)
C13	30	50.257	4.317	0.913	(48.468, 52.046)
C14	30	49.605	4.994	0.913	(47.816, 51.394)
C15	30	51.474	5.497	0.913	(49.685, 53.264)
C16	30	48.930	5.317	0.913	(47.141, 50.719)
C17	30	49.870	4.847	0.913	(48.081, 51.659)
C18	30	50.739	6.224	0.913	(48.950, 52.528)
C19	30	50.979	5.520	0.913	(49.190, 52.768)
C20	30	48.848	4.130	0.913	(47.059, 50.638)
C21	30	49.481	4.056	0.913	(47.692, 51.270)
C22	30	49.183	5.409	0.913	(47.394, 50.973)
C23	30	50.084	4.522	0.913	(48.294, 51.873)
C24	30	50.866	5.142	0.913	(49.077, 52.655)
C25	30	48.768	5.582	0.913	(46.979, 50.557)
C26	30	50.904	6.052	0.913	(49.115, 52.694)
C27	30	49.481	5.535	0.913	(47.691, 51.270)
C28	30	50.949	5.916	0.913	(49.160, 52.739)
C29	30	49.106	4.641	0.913	(47.317, 50.895)
C30	30	49.994	5.853	0.913	(48.205, 51.784)
C31	30	49.601	5.064	0.913	(47.811, 51.390)
C32	30	51.494	5.597	0.913	(49.705, 53.284)
C33	30	50.460	4.393	0.913	(48.671, 52.249)
C34	30	50.378	4.075	0.913	(48.589, 52.167)
C35	30	49.808	4.155	0.913	(48.019, 51.597)
C36	30	49.934	5.012	0.913	(48.145, 51.723)
C37	30	50.017	4.082	0.913	(48.228, 51.806)

Variable	N	Mean	StDev	SE Mean	95.0% CI
C38	30	50.074	3.631	0.913	(48.285, 51.863)
C39	30	48.656	4.833	0.913	(46.867, 50.445)
C40	30	50.568	3.855	0.913	(48.779, 52.357)
C41	30	50.916	3.775	0.913	(49.127, 52.705)
C42	30	49.104	4.321	0.913	(47.315, 50.893)
C43	30	50.308	5.467	0.913	(48.519, 52.097)
C44	30	49.034	4.405	0.913	(47.245, 50.823)
C45	30	50.399	4.729	0.913	(48.610, 52.188)
C46	30	49.634	3.996	0.913	(47.845, 51.424)
C47	30	50.479	4.881	0.913	(48.689, 52.268)
C48	30	50.529	5.173	0.913	(48.740, 52.318)
C49	30	51.577	5.822	0.913	(49.787, 53.366)
C50	30	50.403	4.893	0.913	(48.614, 52.192)
C51	30	49.717	5.218	0.913	(47.927, 51.506)
C52	30	49.796	5.327	0.913	(48.007, 51.585)
C53	30	50.549	4.680	0.913	(48.760, 52.338)
C54	30	50.200	5.840	0.913	(48.410, 51.989)
C55	30	49.138	5.074	0.913	(47.349, 50.928)
C56	30	49.667	3.843	0.913	(47.878, 51.456)
C57	30	49.603	5.614	0.913	(47.814, 51.392)
C58	30	49.441	5.702	0.913	(47.652, 51.230)
C59	30	47.873	4.685	0.913	(46.084, 49.662)
C60	30	51.087	5.162	0.913	(49.297, 52.876)

Self-Review 9–1

The Bun-and-Run is a franchise fast-food restaurant located in the Northeast specializing in half-pound hamburgers, fish sandwiches, and chicken sandwiches. Soft drinks and french fries are also available. The Planning Department of Bun-and-Run, Inc., reports that the distribution of daily sales for restaurants follows the normal distribution. The standard deviation of the distribution of daily sales is $3,000. A sample of 40 showed the mean daily sales to be $20,000.

(a) What is the population mean?
(b) What is the best estimate of the population mean? What is this value called?
(c) Develop a 99 percent confidence interval for the population mean.
(d) Interpret the confidence interval.

Exercises

1. A sample of 49 observations is taken from a normal population with a standard deviation of 10. The sample mean is 55. Determine the 99 percent confidence interval for the population mean.
2. A sample of 81 observations is taken from a normal population with a standard deviation of 5. The sample mean is 40. Determine the 95 percent confidence interval for the population mean.
3. A sample of 10 observations is selected from a normal population for which the population standard deviation is known to be 5. The sample mean is 20.
   a. Determine the standard error of the mean.
   b. Explain why we can use formula (9–1) to determine the 95 percent confidence interval even though the sample is less than 30.
   c. Determine the 95 percent confidence interval for the population mean.
4. Suppose you want an 85 percent confidence level. What value would you use to multiply the standard error of the mean by?
5. A research firm conducted a survey to determine the mean amount steady smokers spend on cigarettes during a week. They found the distribution of amounts spent per week followed the normal distribution with a standard deviation of $5. A sample of 49 steady smokers revealed that $\overline{X} = \$20$.
   a. What is the point estimate of the population mean? Explain what it indicates.

b. Using the 95 percent level of confidence, determine the confidence interval for μ. Explain what it indicates.

6. Refer to the previous exercise. Suppose that 64 smokers (instead of 49) were sampled. Assume the sample mean remained the same.
a. What is the 95 percent confidence interval estimate of μ?
b. Explain why this confidence interval is narrower than the one determined in the previous exercise.

7. Bob Nale is the owner of Nale's Texaco GasTown. Bob would like to estimate the mean number of gallons of gasoline sold to his customers. Assume the number of gallons sold follows the normal distribution with a standard deviation of 2.30 gallons. From his records, he selects a random sample of 60 sales and finds the mean number of gallons sold is 8.60.
a. What is the point estimate of the population mean?
b. Develop a 99 percent confidence interval for the population mean.
c. Interpret the meaning of part (b).

8. Dr. Patton is a professor of English. Recently she counted the number of misspelled words in a group of student essays. She noted the distribution of misspelled words per essay followed the normal distribution with a standard deviation of 2.44 words per essay. For her 10 A.M. section of 40 students, the mean number of misspelled words was 6.05. Construct a 95 percent confidence interval for the mean number of misspelled words in the population of student essays.

Population Standard Deviation σ Unknown

In the previous section we assumed the population standard deviation was known. In the case involving Del Monte 4-ounce cans of peaches, there would likely be a long history of measurements in the filling process. Therefore, it is reasonable to assume the standard deviation of the population is available. However, in most sampling situations the population standard deviation (σ) is not known. Here are some examples where we wish to estimate the population means and it is unlikely we would know the population standard deviations. Suppose each of these studies involves students at West Virginia University.

- The Dean of the Business College wants to estimate the mean number of hours full-time students work at paying jobs each week. He selects a sample of 30 students, contacts each student and asks them how many hours they worked last week. From the sample information he can calculate the sample mean, but it is not likely he would know or be able to find the *population* (σ) standard deviation required in formula (9–1). He could calculate the standard deviation of the sample and use that as an estimate, but he would not likely know the population standard deviation.
- The Dean of Students wants to estimate the distance the typical commuter student travels to class. She selects a sample of 40 commuter students, contacts each, and determines the one-way distance from each student's home to the center of campus. From the sample data she calculates the mean travel distance, that is $\bar{X}$. It is unlikely the standard deviation of the population would be known or available, again making formula (9–1) unusable.
- The Director of Student Loans wants to know the mean amount owed on student loans at the time of his/her graduation. The director selects a sample of 20 graduating students and contacts each to find the information. From the sample information he can estimate the mean amount. However, to develop a confidence interval using formula (9–1), the population standard deviation is necessary. It is not likely this information is available.

Fortunately we can use the sample standard deviation to estimate the population standard deviation. That is, we use s, the sample standard deviation, to estimate σ, the population standard deviation. But in doing so, we cannot use formula

(9–1). Because we do not know σ we cannot use the z distribution. However, there is a remedy. We use the sample standard deviation and replace the z distribution with the t distribution.

The t distribution is a continuous probability distribution, with many similar characteristics to the z distribution. William Gosset, an English brewmaster, was the first to study the t distribution.

He was particularly concerned with the exact behavior of the distribution of the following statistic:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

where s is an estimate of σ. He was especially worried about the discrepancy between s and σ when s was calculated from a very small sample. The t distribution and the standard normal distribution are shown graphically in Chart 9–1. Note particularly that the t distribution is flatter, more spread out, than the standard normal distribution. This is because the standard deviation of the t distribution is larger than the standard normal distribution.

Statistics in Action

William Gosset was born in England in 1876 and died there in 1937. He worked for many years at Arthur Guinness, Sons and Company. In fact, in his later years he was in charge of the Guinness Brewery in London. Guinness preferred its employees to use pen names when publishing papers, so in 1908, when Gosset wrote "The Probable Error of a Mean," he used the name "Student." In this paper he first described the properties of the t distribution.

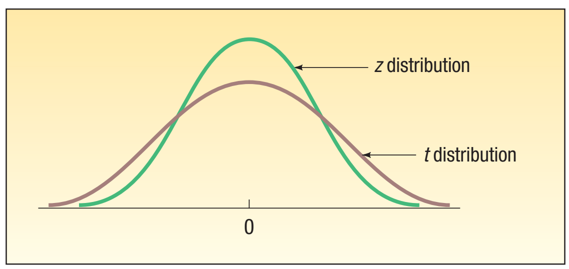


CHART 9–1 The Standard Normal Distribution and Student's t Distribution

The following characteristics of the t distribution are based on the assumption that the population of interest is normal, or nearly normal.

1. It is, like the z distribution, a continuous distribution.
2. It is, like the z distribution, bell-shaped and symmetrical.
3. There is not one t distribution, but rather a family of t distributions. All t distributions have a mean of 0, but their standard deviations differ according to the sample size, n. There is a t distribution for a sample size of 20, another for a sample size of 22, and so on. The standard deviation for a t distribution with 5 observations is larger than for a t distribution with 20 observations.
4. The t distribution is more spread out and flatter at the center than the standard normal distribution (see Chart 9–1). As the sample size increases, however, the t distribution approaches the standard normal distribution, because the errors in using s to estimate σ decrease with larger samples.

Because Student's t distribution has a greater spread than the z distribution, the value of t for a given level of confidence is larger in magnitude than the corresponding z value. Chart 9–2 shows the values of z for a 95 percent level of confidence and of t for the same level of confidence when the sample size is $n = 5$. How we obtained the actual value of t will be explained shortly. For now, observe that for the same level of confidence the t distribution is flatter or more spread out than the standard normal distribution.

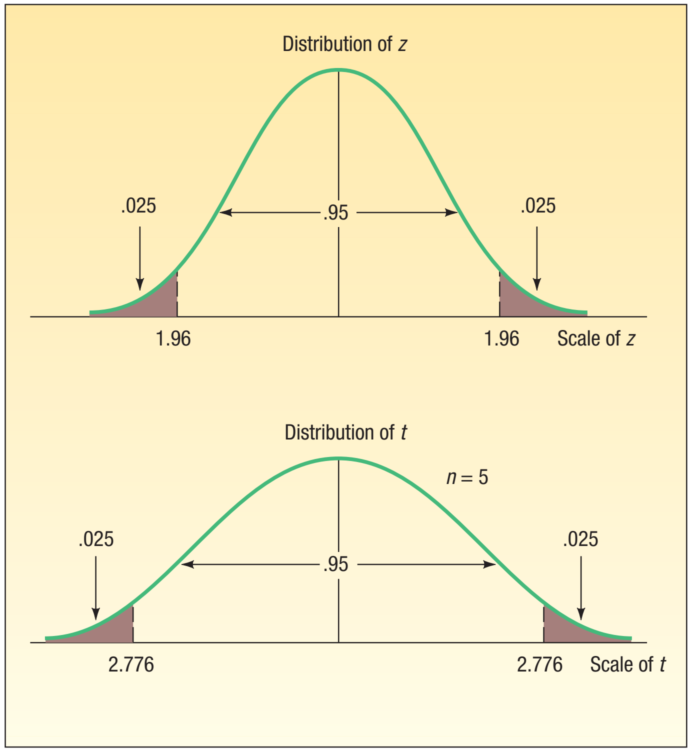


CHART 9–2 Values of z and t for the 95 Percent Level of Confidence

To develop a confidence interval for the population mean using the t distribution, we adjust formula (9–1) as follows.

CONFIDENCE INTERVAL FOR THE POPULATION MEAN, σ UNKNOWN

$$\bar{X} \pm t\frac{s}{\sqrt{n}} \quad [9\text{–}2]$$

To develop a confidence interval for the population mean with an unknown standard deviation we:

1. Assume the sampled population is either normal or approximately normal.
2. Estimate the population standard deviation (σ) with the sample standard deviation (s).
3. Use the t distribution rather than the z distribution.

We should be clear at this point. We base the decision on whether to use the t or the z on whether or not we know σ, the population standard deviation. If we know the population standard deviation, then we use z. If we do not know the population standard deviation, then we must use t. Chart 9–3 summarizes the decision-making process.

The following example will illustrate a confidence interval for a population mean when the population standard deviation is unknown and how to find the appropriate value of t in a table.

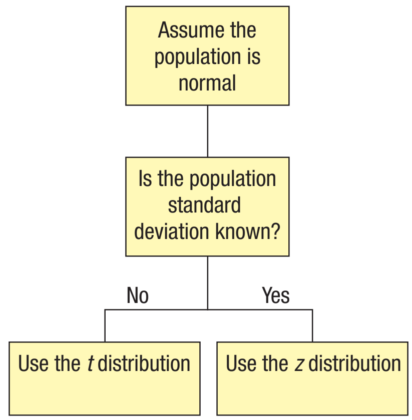


CHART 9–3 Determining When to Use the z Distribution or the t Distribution

Example

A tire manufacturer wishes to investigate the tread life of its tires. A sample of 10 tires driven 50,000 miles revealed a sample mean of 0.32 inch of tread remaining with a standard deviation of 0.09 inch. Construct a 95 percent confidence interval for the population mean. Would it be reasonable for the manufacturer to conclude that after 50,000 miles the population mean amount of tread remaining is 0.30 inches?

Solution

To begin, we assume the population distribution is normal. In this case, we don't have a lot of evidence, but the assumption is probably reasonable. We do not know the population standard deviation, but we know the sample standard deviation, which is .09 inches. We use formula (9–2):

$$\bar{X} \pm t \frac{s}{\sqrt{n}}$$

From the information given, $\bar{X} = 0.32$, $s = 0.09$, and $n = 10$. To find the value of t we use Appendix B.2, a portion of which is reproduced in Table 9–1. Appendix B.2 is also reproduced on the inside back cover of the text. The first step for locating t is to move across the columns identified for "Confidence Intervals" to the level of confidence requested. In this case we want the 95 percent level of confidence, so we move to the column headed "95%." The column on the left margin is identified as "*df.*" This refers to the number of degrees of freedom. The number of degrees of freedom is the number of observations in the sample minus the number of samples, written $n - 1$. In this case it is $10 - 1 = 9$. Why did we decide there were 9 degrees of freedom? When sample statistics are being used, it is necessary to determine the number of values that are *free to vary.* To illustrate: assume that the mean of four numbers is known to be 5. The four numbers are 7, 4, 1, and 8. The deviations of these numbers from the mean must total 0. The deviations of $+2$, -1, -4, and $+3$ do total 0. If the deviations of $+2$, -1, and -4 are known, then the value of $+3$ is fixed (restricted) in order to satisfy the condition that the sum of the deviations must equal 0. Thus, 1 degree of freedom is lost in a sampling problem involving the standard deviation of the sample because one number (the arithmetic mean) is known. For a 95 percent level of confidence and 9 degrees of freedom, we select the row with 9 degrees of freedom. The value of t is 2.262.

TABLE 9–1 A Portion of the *t* Distribution

	Confidence Intervals				
	80%	90%	95%	98%	99%
	Level of Significance for One-Tailed Test				
df	0.100	0.050	0.025	0.010	0.005
	Level of Significance for Two-Tailed Test				
	0.20	0.10	0.05	0.02	0.01
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169

To determine the confidence interval we substitute the values in formula (9–2).

$$\bar{X} \pm t\frac{s}{\sqrt{n}} = 0.32 \pm 2.262\frac{0.09}{\sqrt{10}} = 0.32 \pm .064$$

The endpoints of the confidence interval are 0.256 and 0.384. How do we interpret this result? It is reasonable to conclude that the population mean is in this interval. The manufacturer can be reasonably sure (95 percent confident) that the mean remaining tread depth is between 0.256 and 0.384 inches. Because the value of 0.30 is in this interval, it is possible that the mean of the population is 0.30.

Here is another example to clarify the use of confidence intervals. Suppose an article in your local newspaper reported that the mean time to sell a residential property in the area is 60 days. You select a random sample of 20 homes sold in the last year and find the mean selling time is 65 days. Based on the sample data, you develop a 95 percent confidence interval for the population mean. You find that the endpoints of the confidence interval are 62 days and 68 days. How do you interpret this result? You can be reasonably confident the population mean is within this range. The value proposed for the population mean, that is, 60 days, is not included in the interval. It is not likely that the population mean is 60 days. The evidence indicates the statement by the local newspaper may not be correct. To put it another way, it seems unreasonable to obtain the sample you did from a population that had a mean selling time of 60 days.

The following example will show additional details for determining and interpreting a confidence interval. We used MINITAB to perform the calculations.

Example

The manager of the Inlet Square Mall, near Ft. Myers, Florida, wants to estimate the mean amount spent per shopping visit by customers. A sample of 20 customers reveals the following amounts spent.

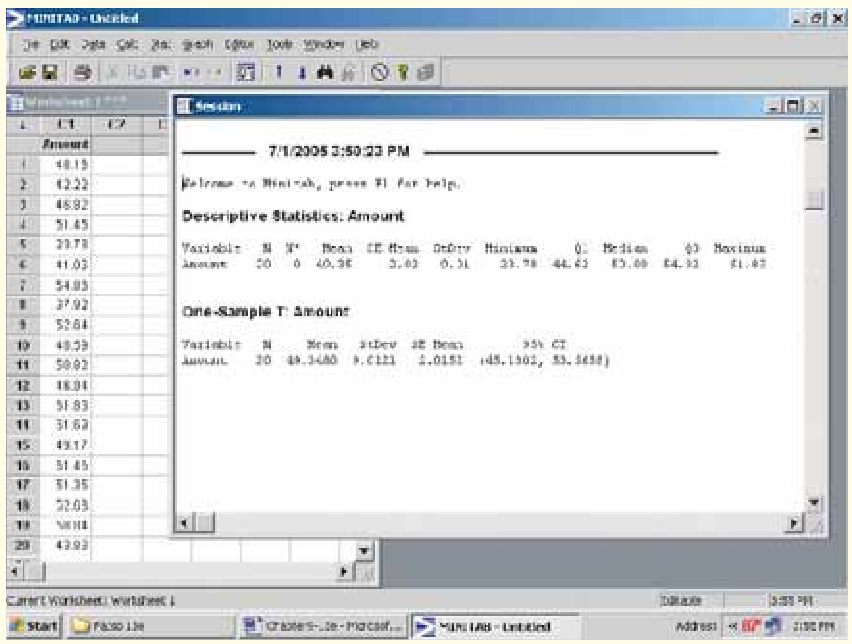
$48.16	$42.22	$46.82	$51.45	$23.78	$41.86	$54.86
37.92	52.64	48.59	50.82	46.94	61.83	61.69
49.17	61.46	51.35	52.68	58.84	43.88	

What is the best estimate of the population mean? Determine a 95 percent confidence interval. Interpret the result. Would it be reasonable to conclude that the population mean is $50? What about $60?

Solution

The mall manager assumes that the population of the amounts spent follows the normal distribution. This is a reasonable assumption in this case. Additionally, the confidence interval technique is quite powerful and tends to commit any errors on the conservative side if the population is not normal. We should not make the normality assumption when the population is severely skewed or when the distribution has "thick tails." In Chapter 18 we present methods for handling this problem if we cannot make the normality assumption. In this case, the normality assumption is reasonable.

The population standard deviation is not known. Hence, it is appropriate to use the *t* distribution and formula (9–2) to find the confidence interval. We use the MINITAB system to find the mean and standard deviation of this sample. The results are shown below.



Minitab

The mall manager does not know the population mean. The sample mean is the best estimate of that value. From the pictured MINITAB output, the mean is \$49.35, which is the best estimate, the *point estimate,* of the unknown population mean.

We use formula (9–2) to find the confidence interval. The value of *t* is available from Appendix B.2. There are $n - 1 = 20 - 1 = 19$ degrees of freedom. We move across the row with 19 degrees of freedom to the column for the 95 percent confidence level. The value at this intersection is 2.093. We substitute these values into formula (9–2) to find the confidence interval.

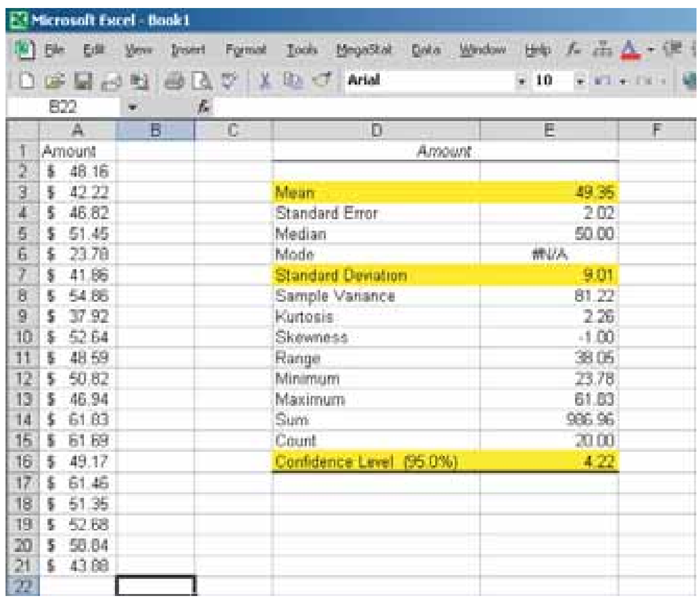
$$\bar{X} \pm t\frac{s}{\sqrt{n}} = \$49.35 \pm 2.093\frac{\$9.01}{\sqrt{20}} = \$49.35 \pm \$4.22$$

The endpoints of the confidence interval are \$45.13 and \$53.57. It is reasonable to conclude that the population mean is in that interval.

The manager of Inlet Square wondered whether the population mean could have been \$50 or \$60. The value of \$50 is within the confidence interval. It is reasonable that the population mean could be \$50. The value of \$60 is not in the confidence interval. Hence, we conclude that the population mean is unlikely to be \$60.

The calculations to construct a confidence interval are also available in Excel. The output is below. Note that the sample mean ($49.35) and the sample standard deviation ($9.01) are the same as those in the Minitab calculations. In the Excel information the last line of the output also includes the margin of error, which is the amount that is added and subtracted from the sample mean to form the endpoints of the confidence interval. This value is found from

$$t\frac{s}{\sqrt{n}} = 2.093\frac{\$9.01}{\sqrt{20}} = \$4.22.$$



	A	B	C	D	E	F
1	Amount			Amount		
2	$ 48.16					
3	$ 42.22			Mean	49.35	
4	$ 46.82			Standard Error	2.02	
5	$ 51.45			Median	50.00	
6	$ 23.78			Mode	#N/A	
7	$ 41.86			Standard Deviation	9.01	
8	$ 54.86			Sample Variance	81.22	
9	$ 37.92			Kurtosis	2.26	
10	$ 52.64			Skewness	-1.00	
11	$ 48.59			Range	38.05	
12	$ 50.82			Minimum	23.78	
13	$ 46.94			Maximum	61.83	
14	$ 61.83			Sum	986.96	
15	$ 61.69			Count	20.00	
16	$ 49.17			Confidence Level (95.0%)	4.22	
17	$ 61.46					
18	$ 51.35					
19	$ 52.68					
20	$ 58.04					
21	$ 43.88					
22						

Excel

Self-Review 9–2

Dottie Kleman is the "Cookie Lady." She bakes and sells cookies at 50 different locations in the Philadelphia area. Ms. Kleman is concerned about absenteeism among her workers. The information below reports the number of days absent for a sample of 10 workers during the last two-week pay period.

4	1	2	2	1	2	2	1	0	3

(a) Determine the mean and the standard deviation of the sample.
(b) What is the population mean? What is the best estimate of that value?
(c) Develop a 95 percent confidence interval for the population mean.
(d) Explain why the *t* distribution is used as a part of the confidence interval.
(e) Is it reasonable to conclude that the typical worker does not miss any days during a pay period?

Exercises

9. Use Appendix B.2 to locate the value of *t* under the following conditions.
  a. The sample size is 12 and the level of confidence is 95 percent.
  b. The sample size is 20 and the level of confidence is 90 percent.
  c. The sample size is 8 and the level of confidence is 99 percent.

10. Use Appendix B.2 to locate the value of *t* under the following conditions.
  a. The sample size is 15 and the level of confidence is 95 percent.
  b. The sample size is 24 and the level of confidence is 98 percent.
  c. The sample size is 12 and the level of confidence is 90 percent.

11. The owner of Britten's Egg Farm wants to estimate the mean number of eggs laid per chicken. A sample of 20 chickens shows they laid an average of 20 eggs per month with a standard deviation of 2 eggs per month.
a. What is the value of the population mean? What is the best estimate of this value?
b. Explain why we need to use the *t* distribution. What assumption do you need to make?
c. For a 95 percent confidence interval, what is the value of *t*?
d. Develop the 95 percent confidence interval for the population mean.
e. Would it be reasonable to conclude that the population mean is 21 eggs? What about 25 eggs?

12. The American Sugar Producers Association wants to estimate the mean yearly sugar consumption. A sample of 16 people reveals the mean yearly consumption to be 60 pounds with a standard deviation of 20 pounds.
a. What is the value of the population mean? What is the best estimate of this value?
b. Explain why we need to use the *t* distribution. What assumption do you need to make?
c. For a 90 percent confidence interval, what is the value of *t*?
d. Develop the 90 percent confidence interval for the population mean.
e. Would it be reasonable to conclude that the population mean is 63 pounds?

13. Merrill Lynch Securities and Health Care Retirement, Inc., are two large employers in downtown Toledo, Ohio. They are considering jointly offering child care for their employees. As a part of the feasibility study, they wish to estimate the mean weekly child-care cost of their employees. A sample of 10 employees who use child care reveals the following amounts spent last week.

$107	$92	$97	$95	$105	$101	$91	$99	$95	$104

Develop a 90 percent confidence interval for the population mean. Interpret the result.

14. The Greater Pittsburgh Area Chamber of Commerce wants to estimate the mean time workers who are employed in the downtown area spend getting to work. A sample of 15 workers reveals the following number of minutes spent traveling.

29	38	38	33	38	21	45	34
40	37	37	42	30	29	35	

Develop a 98 percent confidence interval for the population mean. Interpret the result.

A Confidence Interval for a Proportion

The material presented so far in this chapter uses the ratio scale of measurement. That is, we use such variables as incomes, weights, distances, and ages. We now want to consider situations such as the following:

- The career services director at Southern Technical Institute reports that 80 percent of its graduates enter the job market in a position related to their field of study.
- A company representative claims that 45 percent of Burger King sales are made at the drive-through window.
- A survey of homes in the Chicago area indicated that 85 percent of the new construction had central air conditioning.
- A recent survey of married men between the ages of 35 and 50 found that 63 percent felt that both partners should earn a living.

These examples illustrate the nominal scale of measurement. When we measure with a nominal scale an observation is classified into one of two or more mutually exclusive groups. For example, a graduate of Southern Tech either entered the job market in a position related to his or her field of study or not. A particular Burger King customer either made a purchase at the drive-through window or did not make a purchase at the drive-through window. There are only two possibilities, and the outcome must be classified into one of the two groups.

PROPORTION The fraction, ratio, or percent indicating the part of the sample or the population having a particular trait of interest.

As an example of a proportion, a recent survey indicated that 92 out of 100 surveyed favored the continued use of daylight savings time in the summer. The sample proportion is 92/100, or .92, or 92 percent. If we let p represent the sample proportion, X the number of "successes," and n the number of items sampled, we can determine a sample proportion as follows.

SAMPLE PROPORTION
$$p = \frac{X}{n} \qquad [9\text{–}3]$$

The population proportion is identified by π. Therefore, π refers to the percent of successes in the population. Recall from Chapter 6 that π is the proportion of "successes" in a binomial distribution. This continues our practice of using Greek letters to identify population parameters and Roman letters to identify sample statistics.

To develop a confidence interval for a proportion, we need to meet the following assumptions.

1. The binomial conditions, discussed in Chapter 6, have been met. Briefly, these conditions are:
   a. The sample data is the result of counts.
   b. There are only two possible outcomes. (We usually label one of the outcomes a "success" and the other a "failure.")
   c. The probability of a success remains the same from one trial to the next.
   d. The trials are independent. This means the outcome on one trial does not affect the outcome on another.
2. The values $n\pi$ and $n(1 - \pi)$ should both be greater than or equal to 5. This condition allows us to invoke the central limit theorem and employ the standard normal distribution, that is, z, to complete a confidence interval.

Developing a point estimate for a population proportion and a confidence interval for a population proportion is similar to doing so for a mean. To illustrate, John Gail is running for Congress from the third district of Nebraska. From a random sample of 100 voters in the district, 60 indicate they plan to vote for him in the upcoming election. The sample proportion is .60, but the population proportion is unknown. That is, we do not know what proportion of voters in the *population* will vote for Mr. Gail. The sample value, .60, is the best estimate we have of the unknown population parameter. So we let p, which is .60, be an estimate of π, which is not known.

To develop a confidence interval for a population proportion, we use:

CONFIDENCE INTERVAL FOR A POPULATION PROPORTION
$$p \pm z\sqrt{\frac{p(1-p)}{n}} \qquad [9\text{–}4]$$

An example will help to explain the details of determining a confidence interval and the result.

Statistics in Action

Many survey results reported in newspapers, in news magazines, and on TV use confidence intervals. For example, a recent survey of 800 TV viewers in Toledo, Ohio, found 44 percent watched the evening news on the local CBS affiliate. The article went on to indicate the margin of error was 3.4 percent. The margin of error is actually the amount that is added and subtracted from the point estimate to find the endpoints of a confidence interval. From formula (9–4) and the 95 percent level of confidence:

$$z\sqrt{\frac{p(1-p)}{n}}$$
$$= 1.96\sqrt{\frac{.44(1-.44)}{800}}$$
$$= 0.034$$

Example

The union representing the Bottle Blowers of America (BBA) is considering a proposal to merge with the Teamsters Union. According to BBA union bylaws, at least three-fourths of the union membership must approve any merger. A random sample of 2,000 current BBA members reveals 1,600 plan to vote for the merger proposal. What is the estimate of the population proportion? Develop a 95 percent confidence interval for the population proportion. Basing your decision on this sample information, can you conclude that the necessary proportion of BBA members favor the merger? Why?

Solution

First, calculate the sample proportion from formula (9–3). It is .80, found by

$$p = \frac{X}{n} = \frac{1{,}600}{2{,}000} = .80$$

Thus, we estimate that 80 percent of the population favor the merger proposal. We determine the 95 percent confidence interval using formula (9–4). The z value corresponding to the 95 percent level of confidence is 1.96.

$$p \pm z\sqrt{\frac{p(1-p)}{n}} = .80 \pm 1.96\sqrt{\frac{.80(1-.80)}{2{,}000}} = .80 \pm .018$$

The endpoints of the confidence interval are .782 and .818. The lower endpoint is greater than .75. Hence, we conclude that the merger proposal will likely pass because the interval estimate includes values greater than 75 percent of the union membership.

The interpretation of a confidence interval can be very useful in decision making and play a very important role especially on election night. For example, Cliff Obermeyer is running for Congress from the 6th District of New Jersey. Suppose 500 voters are contacted upon leaving the polls and 275 indicate they voted for Mr. Obermeyer. We will assume that the exit poll of 500 voters is a random sample of those voting in the 6th District. That means that 55 percent of those in the sample voted for Mr. Obermeyer. Based on formula (9–3):

$$p = \frac{X}{n} = \frac{275}{500} = .55$$

Now, to be assured of election, he must earn *more than* 50 percent of the votes in the population of those voting. At this point we know a point estimate, which is .55, of the population of voters that will vote for him. But we do not know the percent in the population that will vote for the candidate. So the question is could we take a sample of 500 voters from a population where 50 percent or less of the voters support Mr. Obermeyer and find that 55 percent of the sample support him? To put it another way, could the sampling error, which is $p - \pi = .55 - .50 = .05$ be due to chance, or is the population of voters who support Mr. Obermeyer greater than .50. If we develop a confidence interval for the sample proportion and find that .50 is *not* in the interval, then we conclude that the proportion of voters supporting Mr. Obermeyer is greater than .50. What does that mean? Well, it means he should be elected! What if .50 is in the interval? Then we conclude that it is possible that 50 percent or less of the voters support his candidacy and we cannot conclude that he will be elected based on the sample information. In this case using the 95 percent significance level and formula (9–4):

$$p \pm z\sqrt{\frac{p(1-p)}{n}} = .55 \pm 1.96\sqrt{\frac{.55(1-.55)}{500}} = .55 \pm .044$$

So the endpoints of the confidence interval are $.55 - .044 = .506$ and $.55 + .044 = .594$. The value of .50 is not in this interval. So we conclude that probably *more than 50 percent* of the voters support Mr. Obermeyer and that is enough to get him elected.

Is this procedure ever used? Yes! It is exactly the procedure used by television networks, news magazines, and polling organizations on election night.

Self-Review 9–3

A market survey was conducted to estimate the proportion of homemakers who would recognize the brand name of a cleanser based on the shape and the color of the container. Of the 1,400 homemakers sampled, 420 were able to identify the brand by name.
(a) Estimate the value of the population proportion.
(b) Develop a 99 percent confidence interval for the population proportion.
(c) Interpret your findings.

Exercises

15. The owner of the West End Kwick Fill Gas Station wishes to determine the proportion of customers who use a credit card or debit card to pay at the pump. He surveys 100 customers and finds that 80 paid at the pump.
- **a.** Estimate the value of the population proportion.
- **b.** Develop a 95 percent confidence interval for the population proportion.
- **c.** Interpret your findings.

16. Ms. Maria Wilson is considering running for mayor of the town of Bear Gulch, Montana. Before completing the petitions, she decides to conduct a survey of voters in Bear Gulch. A sample of 400 voters reveals that 300 would support her in the November election.
- **a.** Estimate the value of the population proportion.
- **b.** Develop a 99 percent confidence interval for the population proportion.
- **c.** Interpret your findings.

17. The Fox TV network is considering replacing one of its prime-time crime investigation shows with a new family-oriented comedy show. Before a final decision is made, network executives commission a sample of 400 viewers. After viewing the comedy, 250 indicated they would watch the new show and suggested it replace the crime investigation show.
- **a.** Estimate the value of the population proportion.
- **b.** Develop a 99 percent confidence interval for the population proportion.
- **c.** Interpret your findings.

18. Schadek Silkscreen Printing, Inc., purchases plastic cups on which to print logos for sporting events, proms, birthdays, and other special occasions. Zack Schadek, the owner, received a large shipment this morning. To ensure the quality of the shipment, he selected a random sample of 300 cups. He found 15 to be defective.
- **a.** What is the estimated proportion defective in the population?
- **b.** Develop a 95 percent confidence interval for the proportion defective.
- **c.** Zack has an agreement with his supplier that he is to return lots that are 10 percent or more defective. Should he return this lot? Explain your decision.

Finite-Population Correction Factor

The populations we have sampled so far have been very large or infinite. What if the sampled population is not very large? We need to make some adjustments in the way we compute the standard error of the sample means and the standard error of the sample proportions.

A population that has a fixed upper bound is *finite.* For example, there are 21,376 students enrolled at Eastern Illinois University, there are 40 employees at Spence Sprockets, DaimlerChrysler assembled 917 Jeep Wranglers at the Alexis Avenue plant yesterday, or there were 65 surgical patients at St. Rose Memorial Hospital in Sarasota yesterday. A finite population can be rather small; it could be all the students registered for this class. It can also be very large, such as all senior citizens living in Florida.

For a finite population, where the total number of objects or individuals is N and the number of objects or individuals in the sample is n, we need to adjust the standard errors in the confidence interval formulas. To put it another way, to find the confidence interval for the mean we adjust the standard error of the mean in formulas (9–1) and (9–2). If we are determining the confidence interval for a proportion, then we need to adjust the standard error of the proportion in formula (9–3).

This adjustment is called the **finite-population correction factor.** It is often shortened to *FPC* and is:

$$FPC = \sqrt{\frac{N - n}{N - 1}}$$

Why is it necessary to apply a factor, and what is its effect? Logically, if the sample is a substantial percentage of the population, the estimate is more precise. Note the effect of the term $(N - n)/(N - 1)$. Suppose the population is 1,000 and the sample is 100. Then this ratio is (1,000 − 100)/(1,000 − 1), or 900/999. Taking the square root gives the correction factor, .9492. Multiplying this correction factor by the standard error *reduces* the standard error by about 5 percent (1 − .9492 = .0508). This reduction in the size of the standard error yields a smaller range of values in estimating the population mean or the population proportion. If the sample is 200, the correction factor is .8949, meaning that the standard error has been reduced by more than 10 percent. Table 9–2 shows the effects of various sample sizes. Note that when the sample is less than about 5 percent of the population, the impact of the correction factor is quite small. The usual rule is if the ratio of n/N is less than .05, the correction factor is ignored.

TABLE 9–2 Finite-Population Correction Factor for Selected Samples When the Population Is 1,000

Sample Size	Fraction of Population	Correction Factor
10	.010	.9955
25	.025	.9879
50	.050	.9752
100	.100	.9492
200	.200	.8949
500	.500	.7075

So if we wished to develop a confidence interval for the mean from a finite population and the population standard deviation was unknown, we would adjust formula (9–2) as follows:

$$\bar{X} \pm t\frac{s}{\sqrt{n}}\left(\sqrt{\frac{N - n}{N - 1}}\right)$$

We would make a similar adjustment to formula (9–3) in the case of a proportion.

The following example summarizes the steps to find a confidence interval for the mean.

Example

There are 250 families in Scandia, Pennsylvania. A random sample of 40 of these families revealed the mean annual church contribution was $450 and the standard deviation of this was $75. Could the population mean be $445 or $425?

1. What is the population mean? What is the best estimate of the population mean?
2. Discuss why the finite-population correction factor should be used.

3. Develop a 90 percent confidence interval for the population mean. What are the endpoints of the confidence interval?
4. Interpret the confidence interval.

Solution

First, note the population is finite. That is, there is a limit to the number of people in Scandia, in this case 250.

1. We do not know the population mean. This is the value we wish to estimate. The best estimate we have of the population mean is the sample mean, which is $450.
2. The sample is 16 percent of the population, found by $n/N = 40/250 = .16$. Because the sample constitutes more than .05 of the population, we should use the *FPC* to adjust the standard error in determining the confidence interval.
3. The formula to find the confidence interval for a population mean follows.

$$\bar{X} \pm t\frac{s}{\sqrt{n}}\left(\sqrt{\frac{N-n}{N-1}}\right)$$

In this case we know $\bar{X} = 450$, $s = 75$, $N = 250$, and that $n = 40$. We do not know the population standard deviation so we use the t distribution. To find the appropriate value of t we use Appendix B.2, and move across the top row to the column headed 90 percent. The degrees of freedom is $df = n - 1 = 40 - 1 = 39$, so we move to the cell where the df row of 39 intersects with the column headed 90 percent. The value is 1.685. Inserting these values in the formula:

$$\bar{X} \pm t\frac{s}{\sqrt{n}}\left(\sqrt{\frac{N-n}{N-1}}\right)$$

$$= \$450 \pm 1.685\frac{\$75}{\sqrt{40}}\left(\sqrt{\frac{250-40}{250-1}}\right) = \$450 \pm \$19.98\sqrt{.8434} = \$450 \pm \$18.35$$

The endpoints of the confidence interval are $431.65 and $468.35.
4. It is likely that the population mean is more than $431.65 but less than $468.35. To put it another way, could the population mean be $445? Yes, but it is not likely that it is $425. Why is this so? Because the value $445 is within the confidence interval and $425 is not within the confidence interval.

Self-Review 9–4

The same study of church contributions in Scandia revealed that 15 of the 40 families sampled attend church regularly. Construct the 95 percent confidence interval for the proportion of families attending church regularly. Should the finite-population correction factor be used? Why or why not?

Exercises

19. Thirty-six items are randomly selected from a population of 300 items. The sample mean is 35 and the sample standard deviation 5. Develop a 95 percent confidence interval for the population mean.

20. Forty-nine items are randomly selected from a population of 500 items. The sample mean is 40 and the sample standard deviation 9. Develop a 99 percent confidence interval for the population mean.

21. The attendance at the Savannah Colts minor league baseball game last night was 400. A random sample of 50 of those in attendance revealed that the mean number of soft drinks consumed per person was 1.86 with a standard deviation of 0.50. Develop a 99 percent confidence interval for the mean number of soft drinks consumed per person.

22. There are 300 welders employed at Maine Shipyards Corporation. A sample of 30 welders revealed that 18 graduated from a registered welding course. Construct the 95 percent confidence interval for the proportion of all welders who graduated from a registered welding course.

Choosing an Appropriate Sample Size

A concern that usually arises when designing a statistical study is how many items should be in the sample. If a sample is too large, money is wasted collecting the data. Similarly, if the sample is too small, the resulting conclusions will be uncertain. The necessary sample size depends on three factors:

1. The level of confidence desired.
2. The margin of error the researcher will tolerate.
3. The variability in the population being studied.

The first factor is the *level of confidence*. Those conducting the study select the level of confidence. The 95 percent and the 99 percent levels of confidence are the most common, but any value between 0 and 100 percent is possible. The 95 percent level of confidence corresponds to a z value of 1.96, and a 99 percent level of confidence corresponds to a z value of 2.58. The higher the level of confidence selected, the larger the size of the corresponding sample.

The second factor is the *allowable error*. The maximum allowable error, designated as E, is the amount that is added and subtracted to the sample mean (or sample proportion) to determine the endpoints of the confidence interval. It is the amount of error those conducting the study are willing to tolerate. It is also one-half the width of the corresponding confidence interval. A small allowable error will require a larger sample. A large allowable error will permit a smaller sample.

The third factor in determining the size of a sample is the *population standard deviation*. If the population is widely dispersed, a large sample is required. On the other hand, if the population is concentrated (homogeneous), the required sample size will be smaller. However, it may be necessary to use an estimate for the population standard deviation. Here are three suggestions for finding that estimate.

1. **Use a comparable study.** Use this approach when there is an estimate of the dispersion available from another study. Suppose we want to estimate the number of hours worked per week by refuse workers. Information from certain state or federal agencies who regularly sample the workforce might be useful to provide an estimate of the standard deviation. If a standard deviation observed in a previous study is thought to be reliable, it can be used in the current study to help provide an approximate sample size.
2. **Use a range-based approach.** To use this approach we need to know or have an estimate of the largest and smallest values in the population. Recall from Chapter 3, where we described the Empirical Rule, that virtually all the observations could be expected to be within plus or minus 3 standard deviations of the mean, assuming that the distribution follows the normal distribution. Thus, the distance between the largest and the smallest values is 6 standard deviations. We could estimate the standard deviation as one-sixth of the range. For example, the director of operations at University Bank wants an estimate of the number of checks written per month by college students. She believes that the distribution of the number of checks written follows the normal distribution. The minimum number written per month is 2 and 50 is the most. The range of the number of checks written per month is 48, found by $50 - 2$. The estimate of the standard deviation then would be 8 checks per month, 48/6.
3. **Conduct a pilot study.** This is the most common method. Suppose we want an estimate of the number of hours per week worked by students enrolled in

the College of Business at the University of Texas. To test the validity of our questionnaire, we use it on a small sample of students. From this small sample we compute the standard deviation of the number of hours worked and use this value to determine the appropriate sample size.

We can express the interaction among these three factors and the sample size in the following formula.

$$E = z\frac{\sigma}{\sqrt{n}}$$

Solving this equation for n yields the following result.

SAMPLE SIZE FOR ESTIMATING THE POPULATION MEAN

$$n = \left(\frac{z\sigma}{E}\right)^2 \quad [9\text{–}5]$$

where:

n is the size of the sample.
z is the standard normal value corresponding to the desired level of confidence.
σ is the population standard deviation.
E is the maximum allowable error.

The result of this calculation is not always a whole number. When the outcome is not a whole number, the usual practice is to round up *any* fractional result. For example, 201.22 would be rounded up to 202.

Example

A student in public administration wants to determine the mean amount members of city councils in large cities earn per month as remuneration for being a council member. The error in estimating the mean is to be less than $100 with a 95 percent level of confidence. The student found a report by the Department of Labor that estimated the standard deviation to be $1,000. What is the required sample size?

Solution

The maximum allowable error, E, is $100. The value of z for a 95 percent level of confidence is 1.96, and the estimate of the standard deviation is $1,000. Substituting these values into formula (9–5) gives the required sample size as:

$$n = \left(\frac{z\sigma}{E}\right)^2 = \left(\frac{(1.96)(\$1{,}000)}{\$100}\right)^2 = (19.6)^2 = 384.16$$

The computed value of 384.16 is rounded up to 385. A sample of 385 is required to meet the specifications. If the student wants to increase the level of confidence, for example to 99 percent, this will require a larger sample. The z value corresponding to the 99 percent level of confidence is 2.58.

$$n = \left(\frac{z\sigma}{E}\right)^2 = \left(\frac{(2.58)(\$1{,}000)}{\$100}\right)^2 = (25.8)^2 = 665.64$$

We recommend a sample of 666. Observe how much the change in the confidence level changed the size of the sample. An increase from the 95 percent to the 99 percent level of confidence resulted in an increase of 281 observations. This could greatly increase the cost of the study, both in terms of time and money. Hence, the level of confidence should be considered carefully.

The procedure just described can be adapted to determine the sample size for a proportion. Again, three items need to be specified:

1. The desired level of confidence.
2. The margin of error in the population proportion.
3. An estimate of the population proportion.

The formula to determine the sample size of a proportion is:

SAMPLE SIZE FOR THE POPULATION PROPORTION

$$n = p(1 - p)\left(\frac{z}{E}\right)^2 \quad [9\text{–}6]$$

If an estimate of p is available from a pilot study or some other source, it can be used. Otherwise, .50 is used because the term $p(1 - p)$ can never be larger than when $p = .50$. For example, if $p = .30$, then $p(1 - p) = .3(1 - .3) = .21$, but when $p = .50$, $p(1 - p) = .5(1 - .5) = .25$

Example

The study in the previous example also estimates the proportion of cities that have private refuse collectors. The student wants the margin of error to be within .10 of the population proportion, the desired level of confidence is 90 percent, and no estimate is available for the population proportion. What is the required sample size?

Solution

The estimate of the population proportion is to be within .10, so $E = .10$. The desired level of confidence is .90, which corresponds to a z value of 1.65. Because no estimate of the population proportion is available, we use .50. The suggested number of observations is

$$n = (.5)(1 - .5)\left(\frac{1.65}{.10}\right)^2 = 68.0625$$

The student needs a random sample of 69 cities.

Self-Review 9–5

Will you assist the college registrar in determining how many transcripts to study? The registrar wants to estimate the arithmetic mean grade point average (GPA) of all graduating seniors during the past 10 years. GPAs range between 2.0 and 4.0. The mean GPA is to be estimated within plus or minus .05 of the population mean. The standard deviation is estimated to be 0.279. Use the 99 percent level of confidence.

Exercises

23. A population is estimated to have a standard deviation of 10. We want to estimate the population mean within 2, with a 95 percent level of confidence. How large a sample is required?

24. We want to estimate the population mean within 5, with a 99 percent level of confidence. The population standard deviation is estimated to be 15. How large a sample is required?

25. The estimate of the population proportion is to be within plus or minus .05, with a 95 percent level of confidence. The best estimate of the population proportion is .15. How large a sample is required?

26. The estimate of the population proportion is to be within plus or minus .10, with a 99 percent level of confidence. The best estimate of the population proportion is .45. How large a sample is required?

27. A survey is being planned to determine the mean amount of time corporation executives watch television. A pilot survey indicated that the mean time per week is 12 hours, with a standard deviation of 3 hours. It is desired to estimate the mean viewing time within one-quarter hour. The 95 percent level of confidence is to be used. How many executives should be surveyed?

28. A processor of carrots cuts the green top off each carrot, washes the carrots, and inserts six to a package. Twenty packages are inserted in a box for shipment. To test the weight of the boxes, a few were checked. The mean weight was 20.4 pounds, the standard deviation 0.5 pounds. How many boxes must the processor sample to be 95 percent confident that the sample mean does not differ from the population mean by more than 0.2 pound?

29. Suppose the U.S. president wants an estimate of the proportion of the population who support his current policy toward revisions in the Social Security system. The president wants the estimate to be within .04 of the true proportion. Assume a 95 percent level of confidence. The president's political advisors estimated the proportion supporting the current policy to be .60.

a. How large of a sample is required?

b. How large of a sample would be necessary if no estimate were available for the proportion supporting current policy?

30. Past surveys reveal that 30 percent of tourists going to Las Vegas to gamble during a weekend spend more than $1,000. Management wants to update this percentage.

a. The new study is to use the 90 percent confidence level. The estimate is to be within 1 percent of the population proportion. What is the necessary sample size?

b. Management said that the sample size determined above is too large. What can be done to reduce the sample? Based on your suggestion recalculate the sample size.

Chapter Summary

I. A point estimate is a single value (statistic) used to estimate a population value (parameter).

II. A confidence interval is a range of values within which the population parameter is expected to occur.

A. The factors that determine the width of a confidence interval for a mean are:

1. The number of observations in the sample, n.

2. The variability in the population, usually estimated by the sample standard deviation, s.

3. The level of confidence.

a. To determine the confidence limits when the population standard deviation is known we use the z distribution. The formula is

$$\bar{X} \pm z\frac{\sigma}{\sqrt{n}} \qquad \textbf{[9–1]}$$

b. To determine the confidence limits when the population standard deviation is unknown we use the t distribution. The formula is

$$\bar{X} \pm t\frac{s}{\sqrt{n}} \qquad \textbf{[9–2]}$$

III. The major characteristics of the t distribution are:

A. It is a continuous distribution.

B. It is mound-shaped and symmetrical.

C. It is flatter, or more spread out, than the standard normal distribution.

D. There is a family of t distributions, depending on the number of degrees of freedom.

IV. A proportion is a ratio, fraction, or percent that indicates the part of the sample or population that has a particular characteristic.

A. A sample proportion is found by X, the number of successes, divided by n, the number of observations.

B. We construct a confidence interval for a sample proportion from the following formula.

$$p \pm z\sqrt{\frac{p(1-p)}{n}} \qquad \textbf{[9–4]}$$

V. For a finite population, the standard error is adjusted by the factor $\sqrt{\frac{N-n}{N-1}}$.

VI. We can determine an appropriate sample size for estimating both means and proportions.

A. There are three factors that determine the sample size when we wish to estimate the mean.

1. The desired level of confidence, which is usually expressed by z.

2. The maximum allowable error, E.

3. The variation in the population, expressed by s.

4. The formula to determine the sample size for the mean is

$$n = \left(\frac{z\sigma}{E}\right)^2 \qquad \textbf{[9–5]}$$

B. There are three factors that determine the sample size when we wish to estimate a proportion.
  1. The desired level of confidence, which is usually expressed by z.
  2. The maximum allowable error, E.
  3. An estimate of the population proportion. If no estimate is available, use .50.
  4. The formula to determine the sample size for a proportion is

$$n = p(1 - p)\left(\frac{z}{E}\right)^2 \quad \textbf{[9–6]}$$

Chapter Exercises

31. A random sample of 85 group leaders, supervisors, and similar personnel at General Motors revealed that, on the average, they spent 6.5 years on the job before being promoted. The standard deviation of the sample was 1.7 years. Construct a 95 percent confidence interval.

32. A state meat inspector in Iowa has been given the assignment of estimating the mean net weight of packages of ground chuck labeled "3 pounds." Of course, he realizes that the weights cannot be precisely 3 pounds. A sample of 36 packages reveals the mean weight to be 3.01 pounds, with a standard deviation of 0.03 pounds.
- **a.** What is the estimated population mean?
- **b.** Determine a 95 percent confidence interval for the population mean.

33. A recent study of 50 self-service gasoline stations in the Greater Cincinnati–Northern Kentucky metropolitan area revealed that the mean price of unleaded gas was $2.029 per gallon. The sample standard deviation was $0.03 per gallon.
- **a.** Determine a 99 percent confidence interval for the population mean price.
- **b.** Would it be reasonable to conclude that the population mean was $1.50? Why or why not?

34. A recent survey of 50 executives who were laid off from their previous position revealed it took a mean of 26 weeks for them to find another position. The standard deviation of the sample was 6.2 weeks. Construct a 95 percent confidence interval for the population mean. Is it reasonable that the population mean is 28 weeks? Justify your answer.

35. Marty Rowatti recently assumed the position of director of the YMCA of South Jersey. He would like some current data on how long current members of the YMCA have been members. To investigate, suppose he selects a random sample of 40 current members. The mean length of membership of those included in the sample is 8.32 years and the standard deviation is 3.07 years.
- **a.** What is the mean of the population?
- **b.** Develop a 90 percent confidence interval for the population mean.
- **c.** The previous director, in the summary report she prepared as she retired, indicated the mean length of membership was now "almost 10 years." Does the sample information substantiate this claim? Cite evidence.

36. The American Restaurant Association collected information on the number of meals eaten outside the home per week by young married couples. A survey of 60 couples showed the sample mean number of meals eaten outside the home was 2.76 meals per week, with a standard deviation of 0.75 meals per week. Construct a 97 percent confidence interval for the population mean.

37. The National Collegiate Athletic Association (NCAA) reported that the mean number of hours spent per week on coaching and recruiting by college football assistant coaches during the season was 70. A random sample of 50 assistant coaches showed the sample mean to be 68.6 hours, with a standard deviation of 8.2 hours.
- **a.** Using the sample data, construct a 99 percent confidence interval for the population mean.
- **b.** Does the 99 percent confidence interval include the value suggested by the NCAA? Interpret this result.
- **c.** Suppose you decided to switch from a 99 to a 95 percent confidence interval. Without performing any calculations, will the interval increase, decrease, or stay the same? Which of the values in the formula will change?

38. The Human Relations Department of Electronics, Inc., would like to include a dental plan as part of the benefits package. The question is: How much does a typical employee and his or her family spend per year on dental expenses? A sample of 45 employees reveals the mean amount spent last year was $1,820, with a standard deviation of $660.

a. Construct a 95 percent confidence interval for the population mean.
b. The information from part (a) was given to the president of Electronics, Inc. He indicated he could afford $1,700 of dental expenses per employee. Is it possible that the population mean could be $1,700? Justify your answer.

39. A student conducted a study and reported that the 95 percent confidence interval for the mean ranged from 46 to 54. He was sure that the mean of the sample was 50, that the standard deviation of the sample was 16, and that the sample was at least 30, but could not remember the exact number. Can you help him out?

40. A recent study by the American Automobile Dealers Association revealed the mean amount of profit per car sold for a sample of 20 dealers was $290, with a standard deviation of $125. Develop a 95 percent confidence interval for the population mean.

41. A study of 25 graduates of four-year colleges by the American Banker's Association revealed the mean amount owed by a student in student loans was $14,381. The standard deviation of the sample was $1,892. Construct a 90 percent confidence interval for the population mean. Is it reasonable to conclude that the mean of the population is actually $15,000? Tell why or why not.

42. An important factor in selling a residential property is the number of people who look through the home. A sample of 15 homes recently sold in the Buffalo, New York, area revealed the mean number looking through each home was 24 and the standard deviation of the sample was 5 people. Develop a 98 percent confidence interval for the population mean.

43. Warren County Telephone Company claims in its annual report that "the typical customer spends $60 per month on local and long-distance service." A sample of 12 subscribers revealed the following amounts spent last month.

$64	$66	$64	$66	$59	$62	$67	$61	$64	$58	$54	$66

a. What is the point estimate of the population mean?
b. Develop a 90 percent confidence interval for the population mean.
c. Is the company's claim that the "typical customer" spends $60 per month reasonable? Justify your answer.

44. The manufacturer of a new line of ink-jet printers would like to include as part of its advertising the number of pages a user can expect from a print cartridge. A sample of 10 cartridges revealed the following number of pages printed.

2,698	2,028	2,474	2,395	2,372	2,475	1,927	3,006	2,334	2,379

a. What is the point estimate of the population mean?
b. Develop a 95 percent confidence interval for the population mean.

45. Dr. Susan Benner is an industrial psychologist. She is currently studying stress among executives of Internet companies. She has developed a questionnaire that she believes measures stress. A score above 80 indicates stress at a dangerous level. A random sample of 15 executives revealed the following stress level scores.

94	78	83	90	78	99	97	90	97	90	93	94	100	75	84

a. Find the mean stress level for this sample. What is the point estimate of the population mean?
b. Construct a 95 percent confidence level for the population mean.
c. Is it reasonable to conclude that Internet executives have a mean stress level in the dangerous level, according to Dr. Benner's test?

46. As a condition of employment, Fashion Industries applicants must pass a drug test. Of the last 220 applicants 14 failed the test. Develop a 99 percent confidence interval for the proportion of applicants that fail the test. Would it be reasonable to conclude that more than 10 percent of the applicants are now failing the test? In addition to the testing of applicants, Fashion Industries randomly tests its employees throughout the year. Last year in the 400 random tests conducted, 14 employees failed the test. Would it be reasonable to conclude that less than 5 percent of the employees are not able to pass the random drug test?

47. There are 20,000 eligible voters in York County, South Carolina. A random sample of 500 York County voters revealed 350 plan to vote to return Louella Miller to the state senate. Construct a 99 percent confidence interval for the proportion of voters in the county who plan to vote for Ms. Miller. From this sample information, can you confirm she will be re-elected?

48. In a poll to estimate presidential popularity, each person in a random sample of 1,000 voters was asked to agree with one of the following statements:
1. The president is doing a good job.
2. The president is doing a poor job.
3. I have no opinion.
A total of 560 respondents selected the first statement, indicating they thought the president was doing a good job.
a. Construct a 95 percent confidence interval for the proportion of respondents who feel the president is doing a good job.
b. Based on your interval in part (a), is it reasonable to conclude that a majority (more than half) of the population believes the president is doing a good job?

49. Police Chief Edward Wilkin of River City reports 500 traffic citations were issued last month. A sample of 35 of these citations showed the mean amount of the fine was $54, with a standard deviation of $4.50. Construct a 95 percent confidence interval for the mean amount of a citation in River City.

50. The First National Bank of Wilson has 650 checking account customers. A recent sample of 50 of these customers showed 26 to have a Visa card with the bank. Construct the 99 percent confidence interval for the proportion of checking account customers who have a Visa card with the bank.

51. It is estimated that 60 percent of U.S. households subscribe to cable TV. You would like to verify this statement for your class in mass communications. If you want your estimate to be within 5 percentage points, with a 95 percent level of confidence, how large of a sample is required?

52. You need to estimate the mean number of travel days per year for outside salespeople. The mean of a small pilot study was 150 days, with a standard deviation of 14 days. If you must estimate the population mean within 2 days, how many outside salespeople should you sample? Use the 90 percent confidence level.

53. You are to conduct a sample survey to determine the mean family income in a rural area of central Florida. The question is, how many families should be sampled? In a pilot sample of 10 families, the standard deviation of the sample was $500. The sponsor of the survey wants you to use the 95 percent confidence level. The estimate is to be within $100. How many families should be interviewed?

54. *Families USA,* a monthly magazine that discusses issues related to health and health costs, surveyed 20 of its subscribers. It found that the annual health insurance premiums for a family with coverage through an employer averaged $10,979. The standard deviation of the sample was $1,000.
a. Based on this sample information, develop a 90 percent confidence interval for the population mean yearly premium.
b. How large a sample is needed to find the population mean within $250 at 99 percent confidence?

55. Passenger comfort is influenced by the amount of pressurization in an airline cabin. Higher pressurization permits a closer-to-normal environment and a more relaxed flight. A study by an airline user group recorded the corresponding air pressure on 30 randomly chosen flights. The study revealed a mean equivalent pressure of 8,000 feet with a standard deviation of 300 feet.
a. Develop a 99 percent confidence interval for the population mean equivalent pressure.
b. How large a sample is needed to find the population mean within 25 feet at 95 percent confidence?

56. A random sample of 25 people employed by the Florida state authority established they earned an average wage (including benefits) of $65.00 per hour. The sample standard deviation was $6.25 per hour.
a. What is the population mean? What is the best estimate of the population mean?
b. Develop a 99 percent confidence interval for the population mean wage (including benefits) for these employees.
c. How large a sample is needed to assess the population mean with an allowable error of $1.00 at 95 percent confidence?

57. A film alliance used a random sample of 50 U.S. citizens to estimate that the typical American spent 78 hours watching videos and DVDs last year. The standard deviation of this sample was 9 hours.
a. Develop a 95 percent confidence interval for the population mean number of hours spent watching videos and DVDs last year.

b. How large a sample should be used to be 90 percent confident the sample mean is within 1.0 hour of the population mean?

58. You plan to conduct a survey to find what proportion of the workforce has two or more jobs. You decide on the 95 percent confidence level and state that the estimated proportion must be within 2 percent of the population proportion. A pilot survey reveals that 5 of the 50 sampled hold two or more jobs. How many in the workforce should be interviewed to meet your requirements?

59. The proportion of public accountants who have changed companies within the last three years is to be estimated within 3 percent. The 95 percent level of confidence is to be used. A study conducted several years ago revealed that the percent of public accountants changing companies within three years was 21.

a. To update this study, the files of how many public accountants should be studied?

b. How many public accountants should be contacted if no previous estimates of the population proportion are available?

60. Huntington National Bank, like most other large banks, found that using automatic teller machines (ATMs) reduces the cost of routine bank transactions. Huntington installed an ATM in the corporate offices of Fun Toy Company. The ATM is for the exclusive use of Fun's 605 employees. After several months of operation, a sample of 100 employees revealed the following use of the ATM machine by Fun employees in a month.

Number of Times ATM Used	Frequency
0	25
1	30
2	20
3	10
4	10
5	5

a. What is the estimate of the proportion of employees who do not use the ATM in a month?

b. Develop a 95 percent confidence interval for this estimate. Can Huntington be sure that at least 40 percent of the employees of Fun Toy Company will use the ATM?

c. How many transactions does the average Fun employee make per month?

d. Develop a 95 percent confidence interval for the mean number of transactions per month.

e. Is it possible that the population mean is .7? Explain.

61. In a recent Zogby poll of 1,000 adults nationwide, 613 said they believe other forms of life exist elsewhere in the universe. Construct the 99 percent confidence interval for the population proportion of those believing life exists elsewhere in the universe. Does your result imply that a majority of Americans believe life exists outside of Earth?

62. As part of an annual review of its accounts, a discount brokerage selects a random sample of 36 customers. Their accounts are reviewed for total account valuation, which showed a mean of $32,000, with a sample standard deviation of $8,200. What is a 90 percent confidence interval for the mean account valuation of the population of customers?

63. A sample of 352 subscribers to *Wired* magazine shows the mean time spent using the Internet is 13.4 hours per week, with a sample standard deviation of 6.8 hours. Find the 95 percent confidence interval for the mean time *Wired* subscribers spend on the Internet.

64. The Tennessee Tourism Institute (TTI) plans to sample information center visitors entering the state to learn the fraction of visitors who plan to camp in the state. Current estimates are that 35 percent of visitors are campers. How large a sample would you take to estimate at a 95 percent confidence level the population proportion with an allowable error of 2 percent?

exercises.com

65. Yahoo is an excellent source of business information. It includes daily summaries as well as information about various industries and specific companies. Go to the site http://finance.yahoo.com. About halfway down the page on the left side, click on **Industries** and then select **Chemicals—Major Diversified** and then click on **Industry Browser.** This should give you a list of companies. Use a table of random numbers, such as Appendix B.6, to randomly select at least five companies. Find the mean price/earnings ratio for the selected companies and develop a confidence interval for the mean.

66. The online edition of the *Information Please Almanac* is a valuable source of business information. Go to the website at www.infoplease.com. On the left, click on **Business.** Then in the **Almanac** section, click on **Taxes,** then on **State Taxes on Individuals.** The result is a listing of the 50 states and the District of Columbia. Use a table of random numbers to randomly select 5 to 10 states. Compute the mean state income tax rate on individuals. Develop a confidence interval for the mean amount. Because the sample is a large part of the population, you will want to include the finite-population correction factor. Interpret your result. You might, as an additional exercise, download all the information and use Excel or MINITAB to compute the population mean. Compare that value with the results of your confidence interval.

Data Set Exercises

67. Refer to the Real Estate data, which report information on the homes sold in Denver, Colorado, last year.
- **a.** Develop a 95 percent confidence interval for the mean selling price of the homes.
- **b.** Develop a 95 percent confidence interval for the mean distance the home is from the center of the city.
- **c.** Develop a 95 percent confidence interval for the proportion of homes with an attached garage.

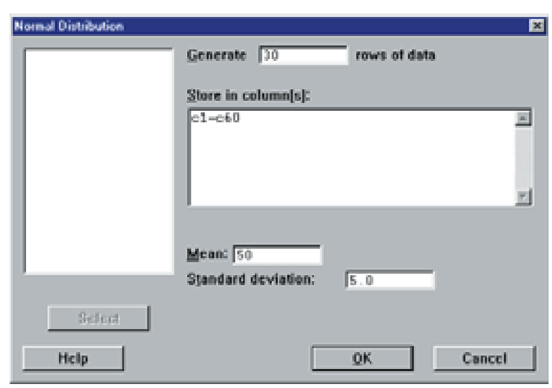
68. Refer to the Baseball 2005 data, which report information on the 30 Major League Baseball teams for the 2005 season.
- **a.** Develop a 95 percent confidence interval for the mean number of home runs per team.
- **b.** Develop a 95 percent confidence interval for the mean number of errors committed by each team.
- **c.** Develop a 95 percent confidence interval for the mean number of stolen bases for each team.

69. Refer to the Wage data, which report information on annual wages for a sample of 100 workers. Also included are variables relating to industry, years of education, and gender for each worker.
- **a.** Develop a 95 percent confidence interval for the mean wage of the workers. Is it reasonable to conclude that the population mean is $35,000?
- **b.** Develop a 95 percent confidence interval for the mean years of education. Is it reasonable to conclude that the population mean is 13 years?
- **c.** Develop a 95 percent confidence interval for the mean age of the workers. Could the mean age be 40 years?

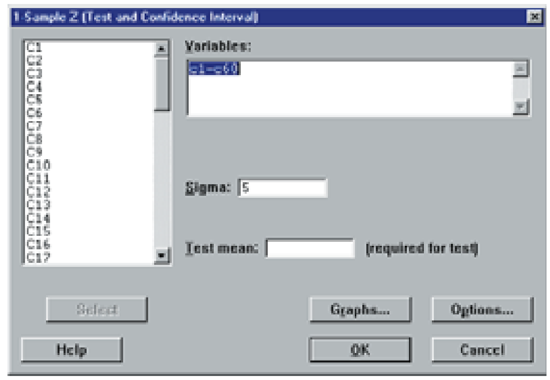
70. Refer to the CIA data, which report demographic and economic information on 46 countries.
- **a.** Develop a 90 percent confidence interval for the mean percent of the population age 65 years or older.
- **b.** Develop a 90 percent confidence interval for the mean Gross Domestic Product (GDP) per capita.
- **c.** Develop a 90 percent confidence interval for the mean imports.

Software Commands

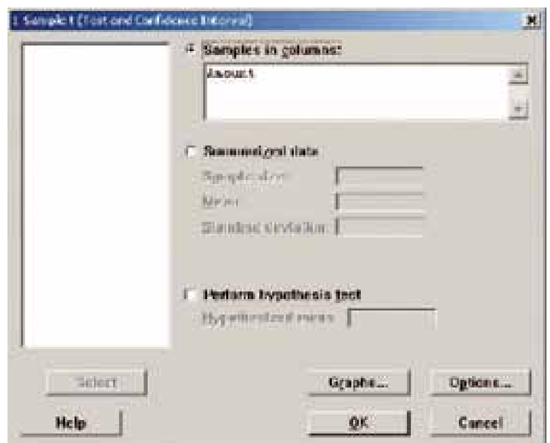
1. The MINITAB commands for the 60 columns of 30 random numbers used in the example/solution on page 300 are:
   - **a.** Select **Calc, Random Data,** and then click on **Normal.**
   - **b.** From the dialog box click in the **Generate** box and type *30* for the number of rows of data, **Store in column(s)** is *C1-C60,* **Mean** is *50,* **Standard deviation** is *5.0,* and finally click **OK.**



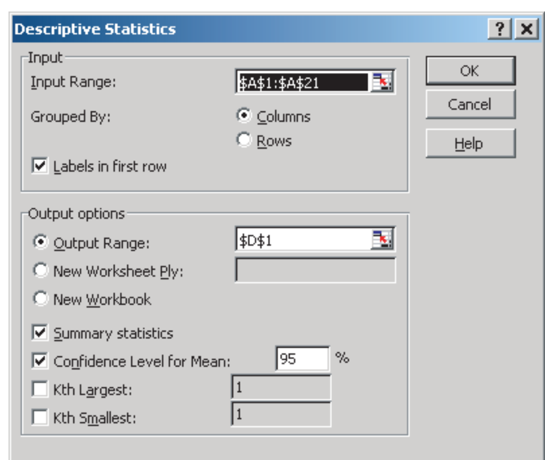
2. The MINITAB commands for the 60 confidence intervals on page 300 follow.
   a. Select **Stat, Basic Statistics,** and then click on **1-Sample Z.**
   b. In the dialog box indicate that the **Variables** are *C1-C60* and that **Sigma** is *5.* Next, click on **Options** in the lower right corner, and in the next dialog box indicate that the **Confidence level** is *95,* and then click **OK.** Click **OK** in the main dialog box.



3. The MINITAB commands for the descriptive statistics on page 307 follow. Enter the data in the first column and label this column *Amount*. On the Toolbar select **Stat, Basic Statistics,** and **Display Descriptive Statistics.** In the dialog box select *Amount* as the **Variable** and click **OK.**
4. The MINITAB commands for the confidence interval for the amount spent at the Inlet Square Mall on page 307 are:
   a. Enter the 20 amounts spent in column C1 and name the variable *Amount,* or locate the data on the student data disk. It is named **Shopping** and is found in the folder for Chapter 8.
   b. On the Toolbar select **Stat, Basic Statistics,** and click on **1-Sample t.**
   c. Select **Samples in columns:** and select **Amount** and click **OK.**



5. The Excel commands for the confidence interval for the amounts spent at the Inlet Square Mall on page 308 are:
   a. From the menu bar select **Tools, Data Analysis,** and **Descriptive Statistics,** and then click **OK.**
   b. For the **Input Range** type *A1:A21,* click on **Labels in first row,** type *D1* as the **Output Range,** click on **Summary statistics** and **Confidence Level for Mean,** and then on **OK.**



Chapter 9 Answers to Self-Review

9–1 **a.** Unknown. This is the value we wish to estimate.
b. \$20,000, point estimate.
c. $\$20{,}000 \pm 2.58 \frac{\$3{,}000}{\sqrt{40}} = \$20{,}000 \pm \$1{,}224$
d. The endpoints of the confidence interval are \$18,776 and \$21,224. About 99 percent of the intervals similarly constructed would include the population mean.

9–2 **a.** $\bar{X} = \frac{18}{10} = 1.8 \qquad s = \sqrt{\frac{11.6}{10 - 1}} = 1.1353$
b. The population mean is not known. The best estimate is the sample mean, 1.8 days.
c. $1.80 \pm 2.262 \frac{1.1353}{\sqrt{10}} = 1.80 \pm 0.81$
The endpoints are 0.99 and 2.61
d. t is used because the population standard deviation is unknown.
e. The value of 0 is not in the interval. It is unreasonable to conclude that the mean number of days of work missed is 0 per employee.

9–3 **a.** $p = \frac{420}{1{,}400} = .30$
b. $.30 \pm 2.58(.0122) = .30 \pm .03$
c. The interval is between .27 and .33. About 99 percent of the similarly constructed intervals would include the population mean.

9–4 $.375 \pm 1.96 \sqrt{\frac{.375(1 - .375)}{40}} \sqrt{\frac{250 - 40}{250 - 1}} =$

$.375 \pm 1.96(.0765)(.9184) = .375 \pm .138$

The correction factor should be applied because $40/250 > .05$.

9–5 $n = \left(\frac{2.58(.279)}{.05}\right)^2 = 207.26$. The sample should be rounded to 208.

A Review of Chapters 8 and 9

We began Chapter 8 by describing the reasons sampling is necessary. We sample because it is often impossible to study every item, or individual, in some populations. It would be too expensive and time consuming, for example, to contact and record the annual incomes of all U.S. bank officers. Also, sampling often destroys the product. A drug manufacturer cannot test the properties of each tablet manufactured, because there would be none left to sell. To estimate a population parameter, therefore, we select a sample from the population. A sample is a part of the population. Care must be taken to ensure that every member of our population has a chance of being selected; otherwise, the conclusions might be biased. A number of probability-type sampling methods can be used, including *simple random, systematic, stratified,* and *cluster sampling.*

Regardless of the sampling method selected, a sample statistic is seldom equal to the corresponding population parameter. For example, the mean of a sample is seldom exactly the same as the mean of the population. The difference between this sample statistic and the population parameter is the *sampling error.*

In Chapter 8 we demonstrated that, if we select all possible samples of a specified size from a population and calculate the mean of these samples, the result will be exactly equal to the population mean. We also showed that the dispersion in the distribution of the sample means is equal to the population standard deviation divided by the square root of the sample size. This result is called the standard error of the mean. There is less dispersion in the distribution of the sample means than in the population. In addition, as we increase the number of observations in each sample, we decrease the variation in the sampling distribution.

The central limit theorem is the foundation of statistical inference. It states that, if the population from which we select the samples follows the normal probability distribution, the distribution of the sample means will also follow the normal distribution. If the population is not normal, it will approach the normal probability distribution as we increase the size of the sample.

Our focus in Chapter 9 was point estimates and interval estimates. A point estimate is a single value used to estimate a population parameter. An interval estimate is a range of values within which we expect the population parameter to occur. For example, based on a sample, we estimate that the mean annual income of all professional house painters in Atlanta, Georgia (the population), is $45,300. That estimate is called a *point estimate.* If we state that the population mean is probably in the interval between $45,200 and $45,400, that estimate is called an *interval estimate.* The two endpoints ($45,200 and $45,400) are the *confidence limits* for the population mean. We described the procedure for establishing a confidence interval for both large and small sample means as well as for sample proportions. In this chapter we also provided a method to determine the necessary sample size based on the dispersion in the population, the level of confidence desired, and the desired precision of the estimate.

Glossary

Bias A possible consequence if certain members of the population are denied the chance to be selected for the sample. As a result, the sample may not be representative of the population.

Central limit theorem If the size of the sample is sufficiently large, the sampling distribution of the sample mean will approach a normal distribution regardless of the shape of the population.

Cluster sampling A method often used to lower the cost of sampling if the population is dispersed over a wide geographic area. The area is divided into smaller units (counties, precincts, blocks, etc.) called primary units. Then a few primary units are chosen, and a random sample is selected from each primary unit.

Finite-population correction factor (FPC) When sampling without replacement from a finite population, a correction term is used to reduce the standard error of the mean according to the relative size of the sample to the size of the population. The correction factor is used when the sample is more than 5 percent of a finite population.

Interval estimate The interval within which a population parameter probably lies, based on sample information. Example: According to sample data, the population mean is in the interval between 1.9 and 2.0 pounds.

Point estimate A single value computed from a sample and used to estimate a population parameter. Example: If the sample mean is 1,020 psi, it is the best estimate of the mean tensile strength of the population.

Probability sample A sample of items or individuals chosen so that each member of the population has a chance of being included in the sample.

Sampling distribution of the sample mean A probability distribution consisting of all possible means of samples of a given size selected from the population.

Sampling error The difference between a sample statistic and the corresponding population parameter. Example: The sample mean income is $22,100; the population mean is $22,000. The sampling error is $22,100 − $22,000 = $100. This error can be attributed to sampling, that is, chance.

Simple random sampling A sampling scheme such that each member of the population has the *same* chance of being selected as part of the sample.

Stratified random sampling A population is first divided into subgroups called strata. A sample is then chosen from each stratum. If, for example, the population of interest consisted of all undergraduate students, the sample design might call for sampling 62 freshmen, 51 sophomores, 40 juniors, and 39 seniors.

Systematic random sampling Assuming the population is arranged in some way, such as alphabetically, by height, or in a file drawer, a random starting point is selected, then every *k*th item becomes a member of the sample. If a sample design called for interviewing every ninth household on Main Street starting with 932 Main, the sample would consist of households at 932 Main, 941 Main, 950 Main, and so on.

Exercises

Part I—Multiple Choice

1. Each new employee is given an identification number. The personnel files are arranged sequentially starting with employee number 0001. To sample the employees, the number 0153 was first selected. Then numbers 0253, 0353, 0453, and so on become members of the sample. This type of sampling is called:
a. Simple random sampling.
b. Systematic sampling.
c. Stratified random sampling.
d. Cluster sampling.

2. You divide a precinct into blocks. Then you select 12 blocks at random and concentrate your sampling efforts in those 12 blocks. This type of sampling is called:
a. Simple random sampling.
b. Systematic sampling.
c. Stratified random sampling.
d. Cluster sampling.

3. The sampling error is:
a. Equal to the population mean.
b. A population parameter.
c. Always positive.
d. The difference between the sample statistic and the population parameter.

4. Which of the following are correct statements about confidence intervals?
a. They cannot contain negative numbers.
b. They are always based on the *z* distribution.
c. They must always include the population parameter.
d. None of the above are always correct.

5. The endpoints of a confidence interval are called:
a. Confidence levels.
b. The test statistics.
c. The degrees of confidence.
d. The confidence limits.

6. We compute the mean and the standard deviation from a sample of 16 observations. Assume the population follows a normal probability distribution. We wish to develop a confidence interval for the mean. Which of the following statements is correct?
a. We cannot develop a confidence interval because we do not know the population standard deviation.
b. We use the *z* distribution because we know the population standard deviation.
c. We can use the *t* distribution to develop the confidence interval.
d. None of the above statements are correct.

7. Which of the following is *not* a correct statement about the *t* distribution?
a. It is positively skewed.
b. It is a continuous distribution.
c. It has a mean of 0.
d. There is a family of *t* distributions.

8. As the number of degrees of freedom increases in the *t* distribution:
a. It approaches the standard normal distribution.
b. The level of confidence increases.
c. It becomes a continuous distribution.
d. It becomes flatter.

9. The degrees of freedom are:
a. The total number of observations.
b. The number of observations minus the number of samples.
c. The number of samples.
d. The number of samples minus one.

10. We select a sample of 15 observations from a normal population and wish to develop a 98 percent confidence interval for the mean. The appropriate value of t is:
a. 2.947
b. 2.977
c. 2.624
d. None of the above.

Part II—Problems

11. A recent study indicated that women took an average of 8.6 weeks of unpaid leave from their jobs after the birth of a child. Assume that this distribution follows the normal probability distribution with a standard deviation of 2.0 weeks. We select a sample of 35 women who recently returned to work after the birth of a child. What is the likelihood that the mean of this sample is at least 8.8 weeks?

12. The manager of Tee Shirt Emporium reports that the mean number of shirts sold per week is 1,210, with a standard deviation of 325. The distribution of sales follows the normal distribution. What is the likelihood of selecting a sample of 25 weeks and finding the sample mean to be 1,100 or less?

13. The owner of the Gulf Stream Café wished to estimate the mean number of lunch customers per day. A sample of 40 days revealed a mean of 160 per day, with a standard deviation of 20 per day. Develop a 98 percent confidence interval for the mean number of customers per day.

14. The manager of the local Hamburger Express wishes to estimate the mean time customers spend at the drive-through window. A sample of 80 customers experienced a mean waiting time of 2.65 minutes, with a standard deviation of 0.45 minutes. Develop a 90 percent confidence interval for the mean waiting time.

15. The office manager for a large company is studying the usage of its copy machines. A random sample of six copy machines revealed the following number of copies (reported in 000s) made yesterday.

826	931	1,126	918	1,011	1,101

Develop a 95 percent confidence interval for the mean number of copies per machine.

16. John Kleman is the host of KXYZ Radio 55 AM drive-time news in Denver. During his morning program, John asks listeners to call in and discuss current local and national news. This morning, John was concerned with the number of hours children under 12 years of age watch TV per day. The last 5 callers reported that their children watched the following number of hours of TV last night.

3.0	3.5	4.0	4.5	3.0

Would it be reasonable to develop a confidence interval from these data to show the mean number of hours of TV watched? If yes, construct an appropriate confidence interval and interpret the result. If no, why would a confidence interval not be appropriate?

17. Historically, Widgets Manufacturing, Inc., produces 250 widgets per day. Recently the new owner bought a new machine to produce more widgets per day. A sample of 16 days' production revealed a mean of 240 units with a standard deviation of 35. Construct a confidence interval for the mean number of widgets produced per day. Does it seem reasonable to conclude that the mean daily widget production has increased? Justify your conclusion.

18. The manufacturer of a power chip used in expensive stereo equipment wishes to estimate the useful life of the chip (in thousands of hours). The estimate is to be within 0.10 (100) hours. Assume a 95 percent level of confidence and that the standard deviation of the useful life of the chip is 0.90 (900 hours). Determine the required sample size.

19. The manager of a home improvement store wishes to estimate the mean amount of money spent in the store. The estimate is to be within $4.00 with a 95 percent level of confidence. The manager does not know the standard deviation of the amounts spent. However, he does estimate that the range is from $5.00 up to $155.00. How large of a sample is needed?

20. In a sample of 200 residents of Georgetown County, 120 reported they believed the county real estate taxes were too high. Develop a 95 percent confidence interval for the proportion of residents who believe the tax rate is too high. Would it be reasonable to conclude that the majority of the taxpayers feel that the taxes are too high?

21. In recent times, the percent of buyers purchasing a new vehicle via the Internet has been large enough that local automobile dealers are concerned about its impact on their business. The information needed is an estimate of the proportion of purchases via the Internet. How large of a sample of purchasers is necessary for the estimate to be within 2 percentage points with a 98 percent level of confidence? Current thinking is that about 8 percent of the vehicles are purchased via the Internet.

22. Historically, the proportion of adults over the age of 24 who smoke has been .30. In recent years, much information has been published and aired on radio and TV that smoking is not good for one's health. A sample of 500 adults revealed only 25 percent of those sampled smoked. Develop a 98 percent confidence interval for the proportion of adults who currently smoke. Would you agree that the proportion is less than 30 percent?

23. The auditor of the State of Ohio needs an estimate of the proportion of residents who regularly play the state lottery. Historically, about 40 percent regularly play, but the auditor would like some current information. How large a sample is necessary for the estimate to be within 3 percentage points, with a 98 percent level of confidence?

Case

Century National Bank

Refer to the description of Century National Bank at the end of the Review of Chapters 1–4 on page 136. When Mr. Selig took over as president of Century several years ago, the use of debit cards was just beginning. He would like an update on the use of these cards. Develop a 95 percent confidence interval for the proportion of customers using these cards. On the basis of the confidence interval, is it reasonable to conclude that more than half of the customers use a debit card? Interpret the results.