chapter

7 Confidence Intervals and Sample Size

Outline

Objectives

After completing this chapter, you should be able to

1. Find the confidence interval for the mean when σ is known or $n \geq 30$.
2. Determine the minimum sample size for finding a confidence interval for the mean.
3. Find the confidence interval for the mean when σ is unknown and $n < 30$.
4. Find the confidence interval for a proportion.
5. Determine the minimum sample size for finding a confidence interval for a proportion.
6. Find a confidence interval for a variance and a standard deviation.

Statistics Today

Would You Change the Channel?

A survey by the Roper Organization found that 45% of the people who were offended by a television program would change the channel, while 15% would turn off their television sets. The survey further stated that the margin of error is 3 percentage points and 4000 adults were interviewed.

Several questions arise:

1. How do these estimates compare with the true population percentages?
2. What is meant by a margin of error of 3 percentage points?
3. Is the sample of 4000 large enough to represent the population of all adults who watch television in the United States?

After reading this chapter, you will be able to answer these questions, since this chapter explains how statisticians can use statistics to make estimates of parameters.

Source: The Associated Press.

7–1 Introduction

One aspect of inferential statistics is **estimation,** which is the process of estimating the value of a parameter from information obtained from a sample. For example, *The Book of Odds,* by Michael D. Shook and Robert L. Shook (New York: Penguin Putnam, Inc.), contains the following statements:

"One out of 4 Americans is currently dieting." (Calorie Control Council.)

"Seventy-two percent of Americans have flown on commercial airlines." ("The Bristol Meyers Report: Medicine in the Next Century.")

"The average kindergarten student has seen more than 5000 hours of television." (U.S. Department of Education.)

"The average school nurse makes \$32,786 a year." (National Association of School Nurses.)

"The average amount of life insurance is \$108,000 per household with life insurance." (American Council of Life Insurance.)

Since the populations from which these values were obtained are large, these values are only *estimates* of the true parameters and are derived from data collected from samples.

The statistical procedures for estimating the population mean, proportion, variance, and standard deviation will be explained in this chapter.

An important question in estimation is that of sample size. How large should the sample be in order to make an accurate estimate? This question is not easy to answer since the size of the sample depends on several factors, such as the accuracy desired and the probability of making a correct estimate. The question of sample size will be explained in this chapter also.

7–2 Confidence Intervals for the Mean (σ Known or $n \geq 30$) and Sample Size

Objective 1. Find the confidence interval for the mean when σ is known or $n \geq 30$.

Suppose a college president wishes to estimate the average age of students attending classes this semester. The president could select a random sample of 100 students and find the average age of these students, say, 22.3 years. From the sample mean, the president could infer that the average age of all the students is 22.3 years. This type of estimate is called a *point estimate.*

A **point estimate** is a specific numerical value estimate of a parameter. The best point estimate of the population mean μ is the sample mean $\overline{X}$.

One might ask why other measures of central tendency, such as the median and mode, are not used to estimate the population mean. The reason is that the means of samples vary less than other statistics (such as medians and modes) when many samples are selected from the same population. Therefore, the sample mean is the best estimate of the population mean.

Sample measures (i.e., statistics) are used to estimate population measures (i.e., parameters). These statistics are called **estimators.** As previously stated, the sample mean is a better estimator of the population mean than the sample median or sample mode.

A good estimator should satisfy the three properties described now.

Three Properties of a Good Estimator

1. The estimator should be an **unbiased estimator.** That is, the expected value or the mean of the estimates obtained from samples of a given size is equal to the parameter being estimated.
2. The estimator should be consistent. For a **consistent estimator,** as sample size increases, the value of the estimator approaches the value of the parameter estimated.
3. The estimator should be a **relatively efficient estimator.** That is, of all the statistics that can be used to estimate a parameter, the relatively efficient estimator has the smallest variance.

Confidence Intervals

As stated in Chapter 6, the sample mean will be, for the most part, somewhat different from the population mean due to sampling error. Therefore, one might ask a second question: How good is a point estimate? The answer is that there is no way of knowing how close the point estimate is to the population mean.

This answer places some doubt on the accuracy of point estimates. For this reason, statisticians prefer another type of estimate, called an *interval estimate.*

An **interval estimate** of a parameter is an interval or a range of values used to estimate the parameter. This estimate may or may not contain the value of the parameter being estimated.

In an interval estimate, the parameter is specified as being between two values. For example, an interval estimate for the average age of all students might be $26.9 < \mu < 27.7$, or 27.3 ± 0.4 years.

Either the interval contains the parameter or it does not. A degree of confidence (usually a percent) can be assigned before an interval estimate is made. For instance, one may wish to be 95% confident that the interval contains the true population mean. Another question then arises. Why 95%? Why not 99% or 99.5%?

Historical Notes

Point and interval estimates were known as long ago as the late 1700s. However, it wasn't until 1937 that a mathematician, J. Neyman, formulated practical applications for them.

If one desires to be more confident, such as 99% or 99.5% confident, then the interval must be larger. For example, a 99% confidence interval for the mean age of college students might be $26.7 < \mu < 27.9$, or 27.3 ± 0.6. Hence, a tradeoff occurs. To be more confident that the interval contains the true population mean, one must make the interval wider.

The **confidence level** of an interval estimate of a parameter is the probability that the interval estimate will contain the parameter.

A **confidence interval** is a specific interval estimate of a parameter determined by using data obtained from a sample and by using the specific confidence level of the estimate.

Intervals constructed in this way are called *confidence intervals.* Three common confidence intervals are used: the 90%, the 95%, and the 99% confidence intervals.

The algebraic derivation of the formula for determining a confidence interval for a mean will be shown later. A brief intuitive explanation will be given first.

The central limit theorem states that when the sample size is large, approximately 95% of the sample means will fall within ± 1.96 standard errors of the population mean, that is,

$$\mu \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

Now, if a specific sample mean is selected, say, $\bar{X}$, there is a 95% probability that it falls within the range of $\mu \pm 1.96(\sigma/\sqrt{n})$. Likewise, there is a 95% probability that the interval specified by

$$\bar{X} \pm 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

will contain μ, as will be shown later. Stated another way,

$$\bar{X} - 1.96\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + 1.96\left(\frac{\sigma}{\sqrt{n}}\right)$$

Hence, one can be 95% confident that the population mean is contained within that interval when the values of the variable are normally distributed in the population.

The value used for the 95% confidence interval, 1.96, is obtained from Table E in Appendix C. For a 99% confidence interval, the value 2.58 is used instead of 1.96 in the formula. This value is also obtained from Table E and is based on the standard normal distribution. Since other confidence intervals are used in statistics, the symbol $z_{\alpha/2}$ (read "zee sub alpha over two") is used in the general formula for confidence intervals. The Greek letter α (alpha) represents the total area in both tails of the standard normal distribution curve, and $\alpha/2$ represents the area in each one of the tails. More will be said after Examples 7–1 and 7–2 about finding other values for $z_{\alpha/2}$.

The relationship between α and the confidence level is that the stated confidence level is the percentage equivalent to the decimal value of $1 - \alpha$, and vice versa. When the 95% confidence interval is to be found, $\alpha = 0.05$, since $1 - 0.05 = 0.95$, or 95%. When $\alpha = 0.01$, then $1 - \alpha = 1 - 0.01 = 0.99$, and the 99% confidence interval is being calculated.

Formula for the Confidence Interval of the Mean for a Specific α

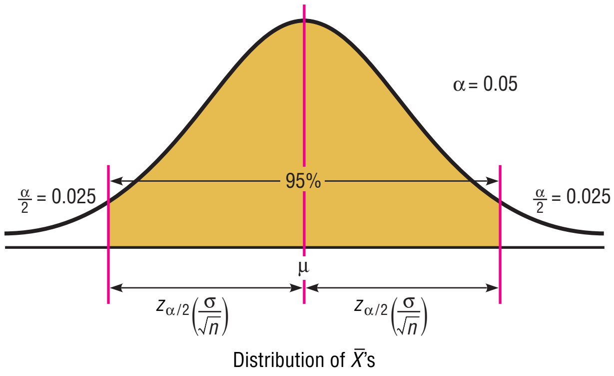
$$\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

For a 90% confidence interval, $z_{\alpha/2} = 1.65$; for a 95% confidence interval, $z_{\alpha/2} = 1.96$; and for a 99% confidence interval, $z_{\alpha/2} = 2.58$.

The term $z_{\alpha/2}(\sigma/\sqrt{n})$ is called the *maximum error of estimate.* For a specific value, say, $\alpha = 0.05$, 95% of the sample means will fall within this error value on either side of the population mean, as previously explained. See Figure 7–1.

Figure 7–1

95% Confidence Interval



The **maximum error of estimate** is the maximum likely difference between the point estimate of a parameter and the actual value of the parameter.

A more detailed explanation of the maximum error of estimate follows Examples 7–1 and 7–2, which illustrate the computation of confidence intervals.

Rounding Rule for a Confidence Interval for a Mean When you are computing a confidence interval for a population mean by using *raw data,* round off to one more decimal place than the number of decimal places in the original data. When you are computing a confidence interval for a population mean by using a sample mean and a standard deviation, round off to the same number of decimal places as given for the mean.

Example 7–1

The president of a large university wishes to estimate the average age of the students presently enrolled. From past studies, the standard deviation is known to be 2 years. A sample of 50 students is selected, and the mean is found to be 23.2 years. Find the 95% confidence interval of the population mean.

Solution

Since the 95% confidence interval is desired, $z_{\alpha/2} = 1.96$. Hence, substituting in the formula

$$\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

one gets

$$23.2 - 1.96\left(\frac{2}{\sqrt{50}}\right) < \mu < 23.2 + 1.96\left(\frac{2}{\sqrt{50}}\right)$$

$$23.2 - 0.6 < \mu < 23.2 + 0.6$$

$$22.6 < \mu < 23.8$$

or 23.2 ± 0.6 years. Hence, the president can say, with 95% confidence, that the average age of the students is between 22.6 and 23.8 years, based on 50 students.

Example 7–2

A survey of 30 adults found that the mean age of a person's primary vehicle is 5.6 years. Assuming the standard deviation of the population is 0.8 year, find the 99% confidence interval of the population mean.

Source: Based on information in *USA TODAY.*

Solution

$$5.6 - 2.58\left(\frac{0.8}{\sqrt{30}}\right) < \mu < 5.6 + 2.58\left(\frac{0.8}{\sqrt{30}}\right)$$

$$5.6 - 0.38 < \mu < 5.6 + 0.38$$

$$5.22 < \mu < 5.98$$

or

$$5.2 < \mu < 6.0 \text{ (rounded)}$$

Hence, one can be 99% confident that the mean age of all primary vehicles is between 5.2 years and 6.0 years, based on 30 vehicles.

Another way of looking at a confidence interval is shown in Figure 7–2. According to the central limit theorem, approximately 95% of the sample means fall within 1.96 standard deviations of the population mean if the sample size is 30 or more or if σ is known when n is less than 30 and the population is normally distributed. If it were possible to build a confidence interval about each sample mean, as was done in Examples 7–1 and 7–2 for μ, 95% of these intervals would contain the population mean, as shown in Figure 7–3. Hence, one can be 95% confident that an interval built around a specific sample mean would contain the population mean.

Figure 7–2

95% Confidence Interval for Sample Means

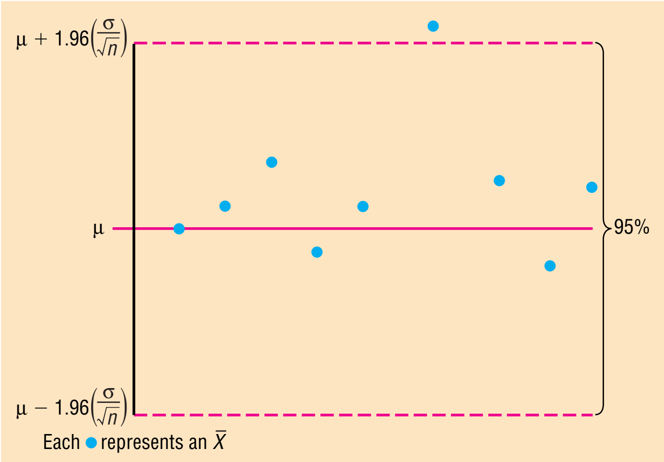
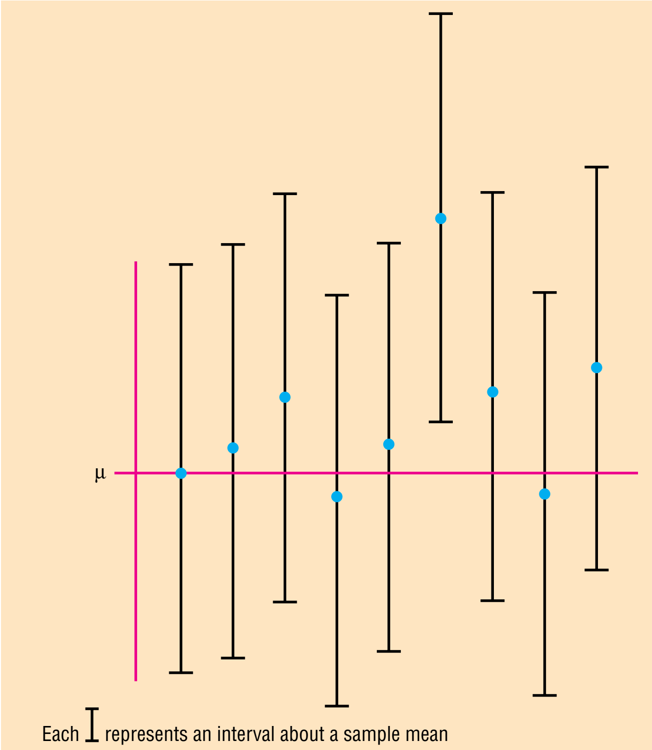


Figure 7–3

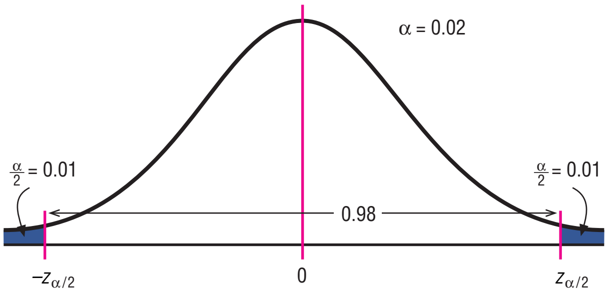
95% Confidence Intervals for Each Sample Mean



If one desires to be 99% confident, the confidence intervals must be enlarged so that 99 out of every 100 intervals contain the population mean.

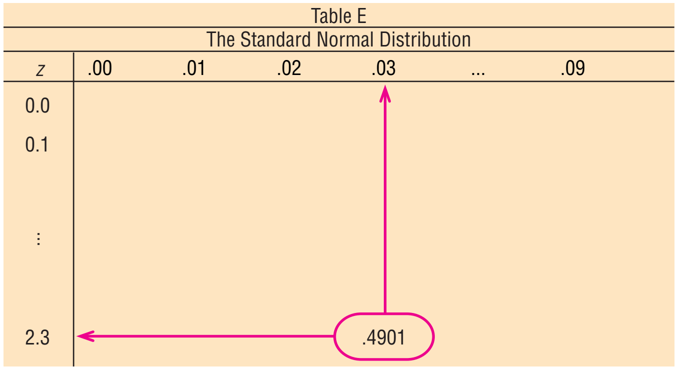
Since other confidence intervals (besides 90%, 95%, and 99%) are sometimes used in statistics, an explanation of how to find the values for $z_{\alpha/2}$ is necessary. As stated previously, the Greek letter α represents the total of the areas in both tails of the normal distribution. The value for α is found by subtracting the decimal equivalent for the desired confidence level from 1. For example, if one wanted to find the 98% confidence interval, one would change 98% to 0.98 and find $\alpha = 1 - 0.98$, or 0.02. Then $\alpha/2$ is obtained by dividing α by 2. So $\alpha/2$ is 0.02/2, or 0.01. Finally, $z_{0.01}$ is the z value that will give an area of 0.01 in the right tail of the standard normal distribution curve. See Figure 7–4.

Figure 7–4
Finding α/2 for a 98% Confidence Interval



Once $\alpha/2$ is determined, the corresponding $z_{\alpha/2}$ value can be found by using the procedure shown in Chapter 6 (see Example 6–17), which is reviewed here. To get the $z_{\alpha/2}$ value for a 98% confidence interval, subtract 0.01 from 0.5000 to get 0.4900. Next, locate the area that is closest to 0.4900 (in this case, 0.4901) in Table E, and then find the corresponding z value. In this example, it is 2.33. See Figure 7–5.

Figure 7–5
Finding $z_{\alpha/2}$ for a 98% Confidence Interval



For confidence intervals, only the positive z value is used in the formula.

When the original variable is normally distributed and α is known, the standard normal distribution can be used to find confidence intervals regardless of the size of the sample. When $n \geq 30$, the distribution of means will be approximately normal even if the original distribution of the variable departs from normality. Also, if $n \geq 30$ (some authors use $n > 30$), s can be substituted for σ in the formula for confidence intervals; and the standard normal distribution can be used to find confidence intervals for means, as shown in Example 7–3.

Example 7–3

The following data represent a sample of the assets (in millions of dollars) of 30 credit unions in southwestern Pennsylvania. Find the 90% confidence interval of the mean.

12.23	16.56	4.39
2.89	1.24	2.17
13.19	9.16	1.42
73.25	1.91	14.64
11.59	6.69	1.06
8.74	3.17	18.13
7.92	4.78	16.85
40.22	2.42	21.58
5.01	1.47	12.24
2.27	12.77	2.76

Source: *Pittsburgh Post Gazette.*

Solution

STEP 1 Find the mean and standard deviation for the data. Use the formulas shown in Chapter 3 or your calculator. The mean $\overline{X} = 11.091$. The standard deviation $s = 14.405$.

STEP 2 Find $\alpha/2$. Since the 90% confidence interval is to be used, $\alpha = 1 - 0.90 = 0.10$, and

$$\frac{\alpha}{2} = \frac{0.10}{2} = 0.05$$

STEP 3 Find $z_{\alpha/2}$. Subtract 0.05 from 0.5000 to get 0.4500. The corresponding z value obtained from Table E is 1.65. (*Note:* This value is found by using the z value for an area between 0.4495 and 0.4505. A more precise z value obtained mathematically is 1.645 and is sometimes used; however, 1.65 will be used in this textbook.)

STEP 4 Substitute in the formula

$$\overline{X} - z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \overline{X} + z_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

(Since $n \geq 30$, s is used in place of σ when σ is unknown.)

$$11.091 - 1.65\left(\frac{14.405}{\sqrt{30}}\right) < \mu < 11.091 + 1.65\left(\frac{14.405}{\sqrt{30}}\right)$$

$$11.091 - 4.339 < \mu < 11.091 + 4.339$$

$$6.752 < \mu < 15.430$$

Hence, one can be 90% confident that the population mean of the assets of all credit unions is between \$6.752 million and \$15.430 million, based on a sample of 30 credit unions.

Comment to Computer and Statistical Calculator Users

This chapter and subsequent chapters include examples using raw data. If you are using computer or calculator programs to find the solutions, the answers you get may vary somewhat from the ones given in the textbook. This is due to the fact that computers and calculators do not round the answers in the intermediate steps and can use 12 or more decimal places for computation. Also, they use more exact values than those given in the tables in the back of this book. These discrepancies are part and parcel of statistics.

Sample Size

Objective 2. Determine the minimum sample size for finding a confidence interval for the mean.

Sample size determination is closely related to statistical estimation. Quite often, one asks, "How large a sample is necessary to make an accurate estimate?" The answer is not simple, since it depends on three things: the maximum error of estimate, the population standard deviation, and the degree of confidence. For example, how close to the true mean does one want to be (2 units, 5 units, etc.), and how confident does one wish to be (90%, 95%, 99%, etc.)? For the purpose of this chapter, it will be assumed that the population standard deviation of the variable is known or has been estimated from a previous study.

The formula for sample size is derived from the maximum error of estimate formula

$$E = z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

and this formula is solved for n as follows:

$$E\sqrt{n} = z_{\alpha/2}(\sigma)$$

$$\sqrt{n} = \frac{z_{\alpha/2} \cdot \sigma}{E}$$

Hence, $$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

Formula for the Minimum Sample Size Needed for an Interval Estimate of the Population Mean

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

where E is the maximum error of estimate. If necessary, round the answer up to obtain a whole number. That is, if there is any fraction or decimal portion in the answer, use the next whole number for sample size n.

Example 7–4

The college president asks the statistics teacher to estimate the average age of the students at their college. How large a sample is necessary? The statistics teacher would like to be 99% confident that the estimate should be accurate within 1 year. From a previous study, the standard deviation of the ages is known to be 3 years.

Solution

Since $\alpha = 0.01$ (or $1 - 0.99$), $z_{\alpha/2} = 2.58$, and $E = 1$, substituting in the formula, one gets

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2 = \left[\frac{(2.58)(3)}{1}\right]^2 = 59.9$$

which is rounded up to 60. Therefore, to be 99% confident that the estimate is within 1 year of the true mean age, the teacher needs a sample size of at least 60 students.

Notice that when one is finding the sample size, the size of the population is irrelevant when the population is large or infinite or when sampling is done with replacement. In other cases, an adjustment is made in the formula for computing sample size. This adjustment is beyond the scope of this book.

The formula for determining sample size requires the use of the population standard deviation. What, then, happens when σ is unknown? In this case, an attempt is made to estimate σ. One such way is to use the standard deviation s obtained from a sample taken previously as an estimate for σ. The standard deviation can also be estimated by dividing the range by 4.

Sometimes, interval estimates rather than point estimates are reported. For instance, one may read a statement such as "On the basis of a sample of 200 families, the survey estimates that an American family of two spends an average of \$84 per week for groceries. One can be 95% confident that this estimate is accurate within \$3 of the true mean." This statement means that the 95% confidence interval of the true mean is

$$\$84 - \$3 < \mu < \$84 + \$3$$
$$\$81 < \mu < \$87$$

The algebraic derivation of the formula for a confidence interval is shown next. As explained in Chapter 6, the sampling distribution of the mean is approximately normal when large samples ($n \geq 30$) are taken from a population. Also,

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Furthermore, there is a probability of $1 - \alpha$ that a z will have a value between $-z_{\alpha/2}$ and $+z_{\alpha/2}$. Hence,

$$-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}$$

Using algebra, one finds

$$-z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Subtracting $\bar{X}$ from both sides and from the middle, one gets

$$-\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < -\mu < -\bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Multiplying by -1, one gets

$$\bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} > \mu > \bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Reversing the inequality, one gets the formula for the confidence interval:

$$\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Exercises 7–2

1. What is the difference between a point estimate and an interval estimate of a parameter? Which is better? Why?

2. What information is necessary to calculate a confidence interval?

3. What is the maximum error of estimate?

4. What is meant by the 95% confidence interval of the mean?

5. What are three properties of a good estimator?

6. What statistic best estimates μ?

7. What is necessary to determine the sample size?

8. When one is determining the sample size for a confidence interval, is the size of the population relevant?

9. Find each.

a. $z_{\alpha/2}$ for the 99% confidence interval
b. $z_{\alpha/2}$ for the 98% confidence interval
c. $z_{\alpha/2}$ for the 95% confidence interval
d. $z_{\alpha/2}$ for the 90% confidence interval
e. $z_{\alpha/2}$ for the 94% confidence interval

10. Find the 95% confidence interval for the mean paid attendance at the Major League All-Star games. A random sample of the paid attendances is shown.

47,596	68,751	5,838
69,831	28,843	53,107
31,391	48,829	50,706
62,892	55,105	63,974
56,674	38,362	51,549
31,938	31,851	56,088
34,906	38,359	72,086
34,009	50,850	43,801
46,127	49,926	54,960
32,785	48,321	49,671

Source: *Time Almanac.*

11. A sample of the reading scores of 35 fifth-graders has a mean of 82. The standard deviation of the sample is 15.

a. Find the 95% confidence interval of the mean reading scores of all fifth-graders.
b. Find the 99% confidence interval of the mean reading scores of all fifth-graders.
c. Which interval is larger? Explain why.

12. Find the 90% confidence interval of the population mean for the incomes of western Pennsylvania credit unions. A random sample of 50 credit unions is shown. The data are in thousands of dollars.

84	14	31	72	26
49	252	104	31	8
3	18	72	23	55
133	16	29	225	138
85	24	391	72	158
4340	346	19	5	846
461	254	125	61	123
60	29	10	366	47
28	254	6	77	21
97	6	17	8	82

Source: *Pittsburgh Post Gazette.*

13. A study of 40 English composition professors showed that they spent, on average, 12.6 minutes correcting a student's term paper.

a. Find the 90% confidence interval of the mean time for all composition papers when $\sigma = 2.5$ minutes.
b. If a professor stated that he spent, on average, 30 minutes correcting a term paper, what would be your reaction?

14. A study of 40 bowlers showed that their average score was 186. The standard deviation of the population is 6.

a. Find the 95% confidence interval of the mean score for all bowlers.
b. Find the 95% confidence interval of the mean score if a sample of 100 bowlers is used instead of a sample of 40.
c. Which interval is smaller? Explain why.

15. A survey of individuals who passed the seven exams and obtained the rank of Fellow in the actuarial field finds the average salary to be $150,000. If the standard deviation for the sample of 35 Fellows was $15,000, construct a 95% confidence interval for all Fellows.

Source: BeAnActuary.org.

16. A random sample of the number of farms (in thousands) in various states is found below. Estimate the mean number of farms per state with 90% confidence.

47	95	54	33	64	4	8	57	9	80
8	90	3	49	4	44	79	80	48	16
68	7	15	21	52	6	78	109	40	50
29									

Source: *N.Y. Times Almanac.*

17. Find the 90% confidence interval for the mean number of local jobs for top corporations in southwestern Pennsylvania. A sample of 40 selected corporations is shown.

7,685	3,100	725	850
11,778	7,300	3,472	540
11,370	5,400	1,570	160
9,953	3,114	2,600	2,821
6,200	3,483	8,954	8
1,000	1,650	1,200	390
1,999	400	3,473	600
1,270	873	400	713
11,960	1,195	2,290	175
887	1,703	4,236	1,400

Source: *Pittsburgh Tribune Review.*

18. A random sample of 48 days taken at a large hospital shows that an average of 38 patients were treated in the emergency room (ER) per day. The standard deviation of the population is 4.

a. Find the 99% confidence interval of the mean number of ER patients treated each day at the hospital.

b. Find the 99% confidence interval of the mean number of ER patients treated each day if the standard deviation were 8 instead of 4.

c. Why is the confidence interval for part *b* wider than the one for part *a*?

19. Noise levels at various area urban hospitals were measured in decibels. The mean of the noise levels in 84 corridors was 61.2 decibels, and the standard deviation was 7.9. Find the 95% confidence interval of the true mean.

Source: M. Bayo, A. Garcia, and A. Garcia, "Noise Levels in an Urban Hospital and Workers' Subjective Responses," *Archives of Environmental Health* 50, no. 3, p. 249 (May–June 1995). Reprinted with permission of the Helen Dwight Reid Educational Foundation. Published by Heldref Publications, 1319 Eighteenth St. N.W., Washington, D.C. 20036-1802. Copyright © 1995.

20. The growing seasons for a random sample of 35 U.S. cities were recorded, yielding a sample mean of 190.7 days and a sample standard deviation of 54.2 days. Estimate the true mean population of the growing season with 95% confidence.

Source: *The Old Farmer's Almanac.*

21. How many cities' growing seasons would have to be sampled in order to estimate the true mean growing season with 95% confidence within 2 days? (Use the standard deviation from Exercise 20.)

22. In the hospital study cited in Exercise 19, the mean noise level in the 171 ward areas was 58.0 decibels, and the standard deviation was 4.8. Find the 90% confidence interval of the true mean.

Source: M. Bayo, A. Garcia, and A. Garcia, "Noise Levels in an Urban Hospital and Workers' Subjective Responses," *Archives of Environmental Health* 50, no. 3, p. 249 (May–June 1995). Reprinted with permission of the Helen Dwight Reid Educational Foundation. Published by Heldref Publications, 1319 Eighteenth St. N.W., Washington, D.C. 20036-1802. Copyright © 1995.

23. An insurance company is trying to estimate the average number of sick days that full-time food service workers use per year. A pilot study found the standard deviation to be 2.5 days. How large a sample must be selected if the company wants to be 95% confident of getting an interval that contains the true mean with a maximum error of 1 day?

24. A restaurant owner wishes to find the 99% confidence interval of the true mean cost of a dry martini. How large should the sample be if she wishes to be accurate within \$0.10? A previous study showed that the standard deviation of the price was \$0.12.

25. A health care professional wishes to estimate the birth weights of infants. How large a sample must she select if she desires to be 90% confident that the true mean is within 6 ounces of the sample mean? The standard deviation of the birth weights is known to be 8 ounces.

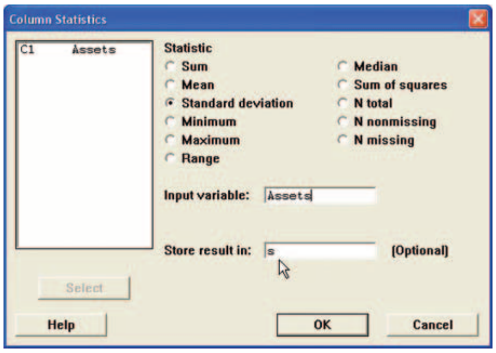
Technology Step by Step

MINITAB Step by Step

Finding a *z* Confidence Interval for the Mean

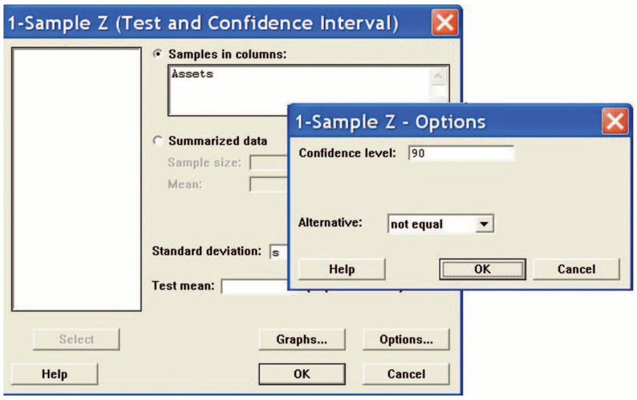
For Example 7–3, find the 90% confidence interval estimate for the mean amount of assets for credit unions in southwestern Pennsylvania.

1. Maximize the worksheet then enter the data into C1 of a MINITAB worksheet. If sigma is known, skip to step 3.
2. Calculate the standard deviation for the sample. It will be used as an estimate for sigma.



a) Select **Calc>Column statistics.**
b) Click the option for Standard deviation.
c) Enter **C1 Assets** for the Input variable and **s** for Store in:.

3. Select **Stat>Basic Statistics>1-Sample Z.**



4. Select C1 Assets for the Samples in Columns.
5. Click in the box for Standard Deviation and enter **s.** Leave the box for Test mean empty.
6. Click the [Options] button. In the dialog box make sure the Confidence Level is 90 and the Alternative is not equal.
7. Optional: Click [Graphs], then select Boxplot of data. The boxplot of these data would clearly show the outliers!
8. Click [OK] twice. The results will be displayed in the session window.

One-Sample Z: Assets

```
The assumed sigma = 14.4054
Variable     N      Mean     StDev   SE Mean          90% CI
Assets      30   11.0907   14.4054    2.6301   (6.7646, 15.4167)
```

TI-83 Plus Step by Step

Finding a *z* Confidence Interval for the Mean (Data)

1. Enter the data into L_1.
2. Press **STAT** and move the cursor to TESTS.
3. Press **7** for ZInterval.
4. Move the cursor to Data and press **ENTER.**
5. Type in the appropriate values.
6. Move the cursor to Calculate and press **ENTER.**

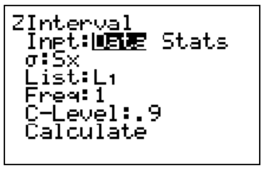
Example TI7–1

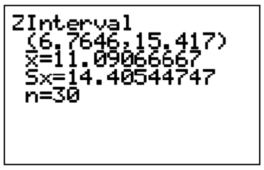
This is Example 7–3 from the text. Find the 90% confidence interval for the population mean, given the data values

12.23	2.89	13.19	73.25	11.59	8.74	7.92	40.22	5.01	2.27
16.56	1.24	9.16	1.91	6.69	3.17	4.78	2.42	1.47	12.77
4.39	2.17	1.42	14.64	1.06	18.13	16.85	21.58	12.24	2.76

The population standard deviation σ is unknown. Since the sample size is $n = 30$, one can use the sample standard deviation s as an approximation for σ. After the data values are entered in L_1 (step 1 above), press **STAT,** move the cursor to CALC, press **1** for 1-Var Stats, then press

ENTER. The sample standard deviation of 14.40544747 will be one of the statistics listed. Then continue with step 2. At step 5 on the line for σ, press **VARS** for variables, press **5** for Statistics, press **3** for S_x.





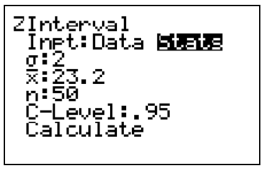
The 90% confidence interval is $6.765 < \mu < 15.417$. The difference between these limits and the ones in Example 7–3 is due to rounding.

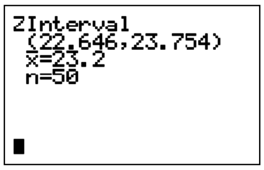
Finding a *z* Confidence Interval for the Mean (Statistics)

1. Press **STAT** and move the cursor to TESTS.
2. Press **7** for ZInterval.
3. Move the cursor to Stats and press **ENTER.**
4. Type in the appropriate values.
5. Move the cursor to Calculate and press **ENTER.**

Example TI7–2

This is Example 7–1 from the text. Find the 95% confidence interval for the population mean, given $\sigma = 2$, $\overline{X} = 23.2$, and $n = 50$.





The 95% confidence interval is $22.6 < \mu < 23.8$.

Excel
Step by Step

Finding a *z* Confidence Interval for the Mean

Excel has a procedure to produce the maximum error of the estimate. But it does not produce confidence intervals. However, you may determine confidence intervals for the mean by using the Mega-Stat Add-in available on your CD and Online Learning Center. If you have not installed this add-in, do so by following the instructions on page 24.

Find the 95% confidence interval when $\sigma = 11$, using this sample:

43	52	18	20	25	45	43	21	42	32	24	32	19	25	26
44	42	41	41	53	22	25	23	21	27	33	36	47	19	20

1. From the toolbar, select **Mega-Stat>Confidence Intervals/Sample Size.**
2. In the dialog box, select the Confidence interval mean tab.
3. Enter **32.03** for the mean of the data, and then select *z.* Enter **11** for the standard deviation and **30** for *n,* the sample size.
4. Either type in or scroll to 95% for the confidence level, and then click [OK].

The result of the procedure is shown next.

Confidence interval - mean

```
  95% confidence level
32.03 mean
   11 std. dev.
   30 n
1.960 z
3.936 half-width
35.966 upper confidence limit
28.094 lower confidence limit
```

7–3

Confidence Intervals for the Mean (σ Unknown and $n < 30$)

Objective 3. Find the confidence interval for the mean when σ is unknown and $n < 30$.

When σ is known and the variable is normally distributed or when σ is unknown and $n \geq 30$, the standard normal distribution is used to find confidence intervals for the mean. However, in many situations, the population standard deviation is not known and the sample size is less than 30. In such situations, the standard deviation from the sample can be used in place of the population standard deviation for confidence intervals. But a somewhat different distribution, called the ***t* distribution,** must be used when the sample size is less than 30 and the variable is normally or approximately normally distributed.

Some important characteristics of the *t* distribution are described now.

Historical Notes

The *t* distribution was formulated in 1908 by an Irish brewing employee named W. S. Gosset. Gosset was involved in researching new methods of manufacturing ale. Because brewing employees were not allowed to publish results, Gosset published his finding using the pseudonym *Student;* hence, the *t* distribution is sometimes called *Student's t distribution.*

Characteristics of the *t* Distribution

The *t* distribution shares some characteristics of the normal distribution and differs from it in others. The *t* distribution is similar to the standard normal distribution in these ways.

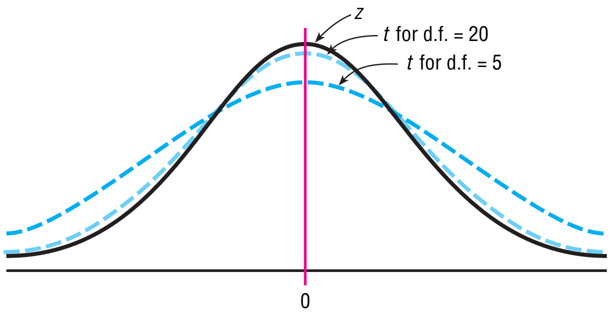
1. It is bell-shaped.
2. It is symmetric about the mean.
3. The mean, median, and mode are equal to 0 and are located at the center of the distribution.
4. The curve never touches the *x* axis.

The *t* distribution differs from the standard normal distribution in the following ways.

1. The variance is greater than 1.
2. The *t* distribution is actually a family of curves based on the concept of *degrees of freedom,* which is related to sample size.
3. As the sample size increases, the *t* distribution approaches the standard normal distribution. See Figure 7–6.

Figure 7–6

The *t* Family of Curves



Many statistical distributions use the concept of degrees of freedom, and the formulas for finding the degrees of freedom vary for different statistical tests. The **degrees**

of freedom are the number of values that are free to vary after a sample statistic has been computed, and they tell the researcher which specific curve to use when a distribution consists of a family of curves.

For example, if the mean of 5 values is 10, then 4 of the 5 values are free to vary. But once 4 values are selected, the fifth value must be a specific number to get a sum of 50, since $50 \div 5 = 10$. Hence, the degrees of freedom are $5 - 1 = 4$, and this value tells the researcher which t curve to use.

The symbol d.f. will be used for degrees of freedom. The degrees of freedom for a confidence interval for the mean are found by subtracting 1 from the sample size. That is, d.f. $= n - 1$. *Note:* For some statistical tests used later in this book, the degrees of freedom are not equal to $n - 1$.

The formula for finding a confidence interval about the mean by using the t distribution is given now.

Formula for a Specific Confidence Interval for the Mean When σ Is Unknown and $n < 30$

$$\bar{X} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{X} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

The degrees of freedom are $n - 1$.

The values for $t_{\alpha/2}$ are found in Table F in Appendix C. The top row of Table F, labeled "Confidence Intervals," is used to get these values. The other two rows, labeled "One Tail" and "Two Tails," will be explained in Chapter 8 and should not be used here.

Example 7–5 shows how to find the value in Table F for $t_{\alpha/2}$.

Example 7–5

Find the $t_{\alpha/2}$ value for a 95% confidence interval when the sample size is 22.

Solution

The d.f. $= 22 - 1$, or 21. Find 21 in the left column and 95% in the row labeled "Confidence intervals." The intersection where the two meet gives the value for $t_{\alpha/2}$, which is 2.080. See Figure 7–7.

Figure 7–7

Finding $t_{\alpha/2}$ for Example 7–5

Table F
The t Distribution

	Confidence Intervals	50%	80%	90%	95%	98%	99%
d.f.	One Tail α	0.25	0.10	0.05	0.025	0.01	0.005
	Two Tails α	0.50	0.20	0.10	0.05	0.02	0.01
1							
2							
3							
⋮							
21					2.080	2.518	2.831
⋮							
z ∞		.674	1.282[a]	1.645[b]	1.960	2.326[c]	2.576[d]

Note: At the bottom of Table F where d.f. = ∞, the $z_{\alpha/2}$ values can be found for specific confidence intervals. The reason is that as the degrees of freedom increase, the *t* distribution approaches the standard normal distribution.

Examples 7–6 and 7–7 show how to find the confidence interval when one is using the *t* distribution.

Example 7–6

Ten randomly selected automobiles were stopped, and the tread depth of the right front tire was measured. The mean was 0.32 inch, and the standard deviation was 0.08 inch. Find the 95% confidence interval of the mean depth. Assume that the variable is approximately normally distributed.

Historical Note

Gosset derived the *t* distribution by selecting small random samples of measurements taken from a population of incarcerated criminals. For the measures he used the lengths of one of their fingers.

Solution

Since σ is unknown and s must replace it, the t distribution (Table F) must be used for 95% confidence interval. Hence, with 9 degrees of freedom, $t_{\alpha/2} = 2.262$.

The 95% confidence interval of the population mean is found by substituting in the formula

$$\bar{X} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{X} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

Hence,

$$0.32 - (2.262)\left(\frac{0.08}{\sqrt{10}}\right) < \mu < 0.32 + (2.262)\left(\frac{0.08}{\sqrt{10}}\right)$$

$$0.32 - 0.057 < \mu < 0.32 + 0.057$$

$$0.26 < \mu < 0.38$$

Therefore, one can be 95% confident that the population mean tread depth of all right front tires is between 0.26 and 0.38 inch based on a sample of 10 tires.

Example 7–7

The data represent a sample of the number of home fires started by candles for the past several years. (Data are from the National Fire Protection Association.) Find the 99% confidence interval for the mean number of home fires started by candles each year.

5460 5900 6090 6310 7160 8440 9930

Solution

STEP 1 Find the mean and standard deviation for the data.
Use the formulas in Chapter 3 or your calculator.
The mean $\bar{X} = 7041.4$.
The standard deviation $s = 1610.3$.

STEP 2 Find $t_{\alpha/2}$ in Table F. Use the 99% confidence interval with d.f. = 6. It is 3.707.

STEP 3 Substitute in the formula and solve.

$$\bar{X} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{X} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

$$7041.4 - 3.707\left(\frac{1610.3}{\sqrt{7}}\right) < \mu < 7041.4 + 3.707\left(\frac{1610.3}{\sqrt{7}}\right)$$

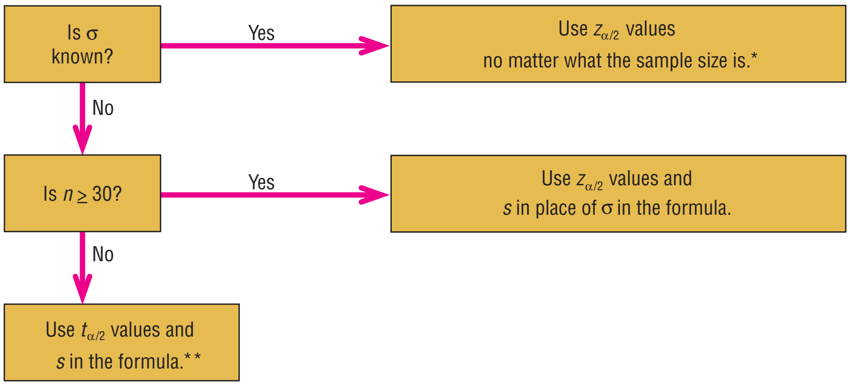
$$7041.4 - 2256.2 < \mu < 7041.4 + 2256.2$$

$$4785.2 < \mu < 9297.6$$

One can be 99% confident that the population mean of home fires started by candles each year is between 4785.2 and 9297.6, based on a sample of home fires occurring over a period of 7 years.

Students sometimes have difficulty deciding whether to use $z_{\alpha/2}$ or $t_{\alpha/2}$ values when finding confidence intervals for the mean. As stated previously, when σ is known, $z_{\alpha/2}$ values can be used *no matter what the sample size is,* as long as the variable is normally distributed or $n \geq 30$. When σ is unknown and $n \geq 30$, s can be used in the formula and $z_{\alpha/2}$ values can be used. Finally, when σ is unknown and $n < 30$, s is used in the formula and $t_{\alpha/2}$ values are used, as long as the variable is approximately normally distributed. These rules are summarized in Figure 7–8.

Figure 7–8
When to Use the z or t Distribution



*Variable must be normally distributed when $n < 30$.
**Variable must be approximately normally distributed.

It should be pointed out that some statisticians have a different point of view. They use $z_{\alpha/2}$ values when σ is known and $t_{\alpha/2}$ values when σ is unknown. In these circumstances, a t table that contains t values for sample sizes greater than or equal to 30 would be needed. The procedure shown in Figure 7–8 is the one used throughout this textbook.

Exercises 7–3

1. What are the properties of the t distribution?

2. What is meant by degrees of freedom?

3. When should the t distribution be used to find a confidence interval for the mean?

4. (**ans**) Find the values for each.

a. $t_{\alpha/2}$ and $n = 18$ for the 99% confidence interval for the mean

b. $t_{\alpha/2}$ and $n = 23$ for the 95% confidence interval for the mean

c. $t_{\alpha/2}$ and $n = 15$ for the 98% confidence interval for the mean

d. $t_{\alpha/2}$ and $n = 10$ for the 90% confidence interval for the mean

e. $t_{\alpha/2}$ and $n = 20$ for the 95% confidence interval for the mean

For Exercises 5 through 20, assume that all variables are approximately normally distributed.

5. The average hemoglobin reading for a sample of 20 teachers was 16 grams per 100 milliliters, with a sample standard deviation of 2 grams. Find the 99% confidence interval of the true mean.

6. A sample of 17 states had these cigarette taxes (in cents):

112	120	98	55	71	35	99	124	64
150	150	55	100	132	20	70	93	

Find a 98% confidence interval for the cigarette tax in all 50 states.

Source: Federation of Tax Administrators.

7. A state representative wishes to estimate the mean number of women representatives per state legislature. A random sample of 17 states is selected, and the number of women representatives is shown. Based on the sample, what is the point estimate of the mean? Find the 90% confidence interval of the mean population. (*Note:* The population mean is actually 31.72, or about 32.) Compare this value to the point estimate and the confidence interval. There is something unusual about the data. Describe it and state how it would affect the confidence interval.

5	33	35	37	24
31	16	45	19	13
18	29	15	39	18
58	132			

8. A random sample of the number of barrels (in millions) of oil produced per day by world oil-producing countries is listed here. Estimate the mean oil production with 95% confidence.

3.56	1.90	7.83	2.83	1.91	5.88	2.91	6.08

Source: *N.Y. Times Almanac.*

9. A sample of six adult elephants had an average weight of 12,200 pounds, with a sample standard deviation of 200 pounds. Find the 95% confidence interval of the true mean.

10. The daily salaries of substitute teachers for eight local school districts is shown. What is the point estimate for the mean? Find the 90% confidence interval of the mean for the salaries of substitute teachers in the region.

60	56	60	55	70	55	60	55

Source: *Pittsburgh Tribune Review.*

11. A recent study of 28 city residents showed that the mean of the time they had lived at their present address was 9.3 years. The standard deviation of the sample was 2 years. Find the 90% confidence interval of the true mean.

12. An automobile shop manager timed six employees and found that the average time it took them to change a water pump was 18 minutes. The standard deviation of the sample was 3 minutes. Find the 99% confidence interval of the true mean.

13. A recent study of 25 students showed that they spent an average of \$18.53 for gasoline per week. The standard deviation of the sample was \$3.00. Find the 95% confidence interval of the true mean.

14. For a group of 10 men subjected to a stress situation, the mean number of heartbeats per minute was 126, and the standard deviation was 4. Find the 95% confidence interval of the true mean.

15. For the stress test described in Exercise 14, six women had an average heart rate of 115 beats per minute. The standard deviation of the sample was 6 beats. Find the 95% confidence interval of the true mean for the women.

16. For a sample of 24 operating rooms taken in the hospital study mentioned in Exercise 19 in Section 7–2, the mean noise level was 41.6 decibels, and the standard deviation was 7.5. Find the 95% confidence interval of the true mean of the noise levels in the operating rooms.

Source: M. Bayo, A. Garcia, and A. Garcia, "Noise Levels in an Urban Hospital and Workers' Subjective Responses," *Archives of Environmental Health* 50, no. 3, p. 249 (May–June 1995). Reprinted with permission of the Helen Dwight Reid Educational Foundation. Published by Heldref Publications, 1319 Eighteenth St. N.W., Washington, D.C. 20036-1802. Copyright © 1995.

17. The number of grams of carbohydrates in a 12-ounce serving of a regular soft drink is listed here for a random sample of sodas. Estimate the mean number of carbohydrates in all brands of soda with 95% confidence.

48	37	52	40	43	46	41	38
41	45	45	33	35	52	45	41
30	34	46	40				

Source: *The Doctor's Pocket Calorie, Fat, and Carbohydrate Counter.*

18. For a group of 20 students taking a final exam, the mean heart rate was 96 beats per minute, and the standard deviation was 5. Find the 95% confidence interval of the true mean.

19. The average yearly income for 28 married couples living in city C is \$58,219. The standard deviation of the sample is \$56. Find the 95% confidence interval of the true mean.

20. The number of unhealthy days based on the AQI (Air Quality Index) for a random sample of metropolitan areas is shown. Construct a 98% confidence interval based on the data.

61	12	6	40	27	38	93	5	13	40

Source: *N.Y. Times Almanac.*

Extending the Concepts

21. A *one-sided confidence* interval can be found for a mean by using

$$\mu > \bar{X} - t_\alpha \frac{s}{\sqrt{n}} \qquad \text{or} \qquad \mu < \bar{X} + t_\alpha \frac{s}{\sqrt{n}}$$

where t_α is the value found under the row labeled "One Tail." Find two one-sided 95% confidence intervals of the population mean for the data shown and interpret the answers. The data represent the daily revenues in dollars from 20 parking meters in a small municipality.

2.60	1.05	2.45	2.90
1.30	3.10	2.35	2.00
2.40	2.35	2.40	1.95
2.80	2.50	2.10	1.75
1.00	2.75	1.80	1.95

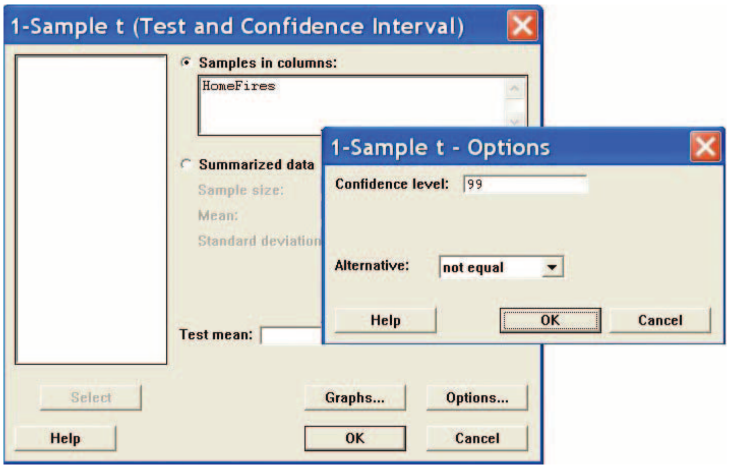
Technology Step by Step

MINITAB Step by Step

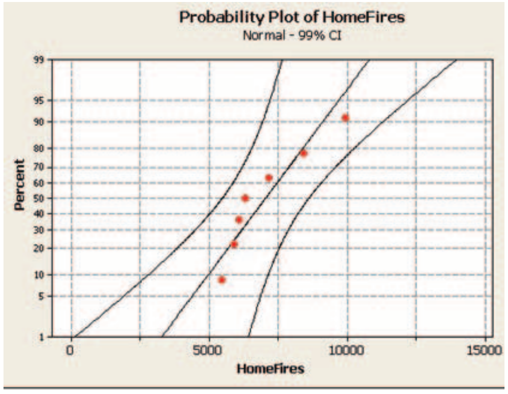
Find a *t* Interval for the Mean

For Example 7–7, find the 99% confidence interval for the mean number of home fires started by candles each year.

1. Type the data into C1 of a MINITAB worksheet. Name the column **HomeFires.**
2. Select **Stat>Basic Statistics>1-Sample t.**



3. Double-click C1 HomeFires for the Samples in Columns.
4. Click on [Options] and be sure the Confidence Level is 99 and the Alternative is not equal.
5. Click [OK] twice.
6. Check for normality:
   a) Select **Graph>Probability Plot,** then Single.
   b) Select C1 HomeFires for the variable. The normal plot is concave, a skewed distribution.



In the session window you will see the results. The 99% confidence interval estimate for μ is between 500.4 and 626.0. The sample size, mean, standard deviation, and standard error of the mean are also shown.

However, this small sample appears to have a nonnormal population. The interval is less likely to contain the true mean.

One-Sample T: HomeFires

```
Variable    N      Mean      StDev    SE Mean          99% CI
HomeFires   7   7041.43    1610.27     608.63    (4784.99, 9297.87)
```

TI-83 Plus Step by Step

Finding a t Confidence Interval for the Mean (Data)

1. Enter the data into L_1.
2. Press **STAT** and move the cursor to TESTS.
3. Press **8** for TInterval.
4. Move the cursor to Data and press **ENTER.**
5. Type in the appropriate values.
6. Move the cursor to Calculate and press **ENTER.**

Finding a t Confidence Interval for the Mean (Statistics)

1. Press **STAT** and move the cursor to TESTS.
2. Press **8** for TInterval.
3. Move the cursor to Stats and press **ENTER.**
4. Type in the appropriate values.
5. Move the cursor to Calculate and press **ENTER.**

Excel Step by Step

Finding a t Confidence Interval for the Mean

Excel has a procedure to produce the maximum error of the estimate. But it does not produce confidence intervals. However, you may determine confidence intervals for the mean by using the Mega-Stat Add-in available on your CD and Online Learning Center. If you have not installed this add-in, do so by following the instructions on page 24.

Find the 95% confidence interval, using this sample:

625	675	535	406	512	680	483	522	619	575

1. From the toolbar, select **Mega-Stat>Confidence Intervals/Sample Size.**
2. In the dialog box, select the Confidence interval mean tab.
3. Enter **563.2** for the mean of the data, and then select *t.* Enter **87.9** for the standard deviation and 10 for *n,* the sample size.
4. Either type in or scroll to 95% for the confidence level, and then click [OK].

The result of the procedure will show the output in a new chart as indicated below.

Confidence interval - mean

```
     95% confidence level
   563.2 mean
    87.9 std. dev.
      10 n
   2.262 t (df = 9)
  62.880 half-width
 626.080 upper confidence limit
 500.320 lower confidence limit
```

7–4 Confidence Intervals and Sample Size for Proportions

A *USA TODAY* Snapshots feature stated that 12% of the pleasure boats in the United States were named *Serenity.* The parameter 12% is called a **proportion.** It means that of all the pleasure boats in the United States, 12 out of every 100 are named *Serenity.* A proportion represents a part of a whole. It can be expressed as a fraction, decimal, or percentage. In this case, $12\% = 0.12 = \frac{12}{100}$ or $\frac{3}{25}$. Proportions can also represent probabilities.

Objective 4. Find the confidence interval for a proportion.

In this case, if a pleasure boat is selected at random, the probability that it is called *Serenity* is 0.12.

Proportions can be obtained from samples or populations. The following symbols will be used.

Symbols Used in Proportion Notation

p = symbol for population proportion

$\hat{p}$ (read "p hat") = symbol for sample proportion

For a sample proportion,

$$\hat{p} = \frac{X}{n} \quad \text{and} \quad \hat{q} = \frac{n - X}{n} \quad \text{or} \quad 1 - \hat{p}$$

where X = number of sample units that possess the characteristics of interest and n = sample size.

For example, in a study, 200 people were asked if they were satisfied with their job or profession; 162 said that they were. In this case, $n = 200$, $X = 162$, and $\hat{p} = X/n = 162/200 = 0.81$. It can be said that for this sample, 0.81, or 81%, of those surveyed were satisfied with their job or profession. The sample proportion is $\hat{p} = 0.81$.

The proportion of people who did not respond favorably when asked if they were satisfied with their job or profession constituted $\hat{q}$, where $\hat{q} = (n - X)/n$. For this survey, $\hat{q} = (200 - 162)/200 = 38/200$, or 0.19, or 19%.

When $\hat{p}$ and $\hat{q}$ are given in decimals or fractions, $\hat{p} + \hat{q} = 1$. When $\hat{p}$ and $\hat{q}$ are given in percentages, $\hat{p} + \hat{q} = 100\%$. It follows, then, that $\hat{q} = 1 - \hat{p}$, or $\hat{p} = 1 - \hat{q}$, when $\hat{p}$ and $\hat{q}$ are in decimal or fraction form. For the sample survey on job satisfaction, $\hat{q}$ can also be found by using $\hat{q} = 1 - \hat{p}$, or $1 - 0.81 = 0.19$.

Similar reasoning applies to population proportions; that is, $p = 1 - q$, $q = 1 - p$, and $p + q = 1$, when p and q are expressed in decimal or fraction form. When p and q are expressed as percentages, $p + q = 100\%$, $p = 100\% - q$, and $q = 100\% - p$.

Example 7–8

In a recent survey of 150 households, 54 had central air conditioning. Find $\hat{p}$ and $\hat{q}$, where $\hat{p}$ is the proportion of households that have central air conditioning.

Solution

Since $X = 54$ and $n = 150$,

$$\hat{p} = \frac{X}{n} = \frac{54}{150} = 0.36 = 36\%$$

$$\hat{q} = \frac{n - X}{n} = \frac{150 - 54}{150} = \frac{96}{150} = 0.64 = 64\%$$

One can also find $\hat{q}$ by using the formula $\hat{q} = 1 - \hat{p}$. In this case, $\hat{q} = 1 - 0.36 = 0.64$.

As with means, the statistician, given the sample proportion, tries to estimate the population proportion. Point and interval estimates for a population proportion can be made by using the sample proportion. For a point estimate of p (the population proportion), $\hat{p}$ (the sample proportion) is used. On the basis of the three properties of a good estimator, $\hat{p}$ is unbiased, consistent, and relatively efficient. But as with means, one is not able to decide how good the point estimate of p is. Therefore, statisticians also use an interval estimate for a proportion, and they can assign a probability that the interval will contain the population proportion.

Confidence Intervals

To construct a confidence interval about a proportion, one must use the maximum error of estimate, which is

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Confidence intervals about proportions must meet the criteria that $np \geq 5$ and $nq \geq 5$.

Formula for a Specific Confidence Interval for a Proportion

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

when np and nq are each greater than or equal to 5.

Rounding Rule for a Confidence Interval for a Proportion Round off to three decimal places.

Example 7–9

A sample of 500 nursing applications included 60 from men. Find the 90% confidence interval of the true proportion of men who applied to the nursing program.

Solution

Since $\alpha = 1 - 0.90 = 0.10$ and $z_{\alpha/2} = 1.65$, substituting in the formula

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

when $\hat{p} = 60/500 = 0.12$ and $\hat{q} = 1 - 0.12 = 0.88$, one gets

$$0.12 - 1.65\sqrt{\frac{(0.12)(0.88)}{500}} < p < 0.12 + 1.65\sqrt{\frac{(0.12)(0.88)}{500}}$$

$$0.12 - 0.024 < p < 0.12 + 0.024$$

$$0.096 < p < 0.144$$

or $$9.6\% < p < 14.4\%$$

Hence, one can be 90% confident that the percentage of applicants who are men is between 9.6% and 14.4%.

When a specific percentage is given, the percentage becomes $\hat{p}$ when it is changed to a decimal. For example, if the problem states that 12% of the applicants were men, then $\hat{p} = 0.12$.

Example 7–10

A survey of 200,000 boat owners found that 12% of the pleasure boats were named *Serenity*. Find the 95% confidence interval of the true proportion of boats named *Serenity*.

Source: *USA TODAY* Snapshot.

Solution

From the Snapshot, $\hat{p} = 0.12$ (i.e., 12%), and $n = 200{,}000$. Since $z_{\alpha/2} = 1.96$, substituting in the formula

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$\text{yields } 0.12 - 1.96\sqrt{\frac{(0.12)(0.88)}{200{,}000}} < p < 0.12 + 1.96\sqrt{\frac{(0.12)(0.88)}{200{,}000}}$$

$$0.119 < p < 0.121$$

Hence, one can say with 95% confidence that the true percentage of boats named *Serenity* is between 11.9% and 12.1%.

Sample Size for Proportions

Objective 5. Determine the minimum sample size for finding a confidence interval for a proportion.

To find the sample size needed to determine a confidence interval about a proportion, use this formula:

Formula for Minimum Sample Size Needed for Interval Estimate of a Population Proportion

$$n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2$$

If necessary, round up to obtain a whole number.

This formula can be found by solving the maximum error of estimate value for *n:*

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

There are two situations to consider. First, if some approximation of $\hat{p}$ is known (e.g., from a previous study), that value can be used in the formula.

Second, if no approximation of $\hat{p}$ is known, one should use $\hat{p} = 0.5$. This value will give a sample size sufficiently large to guarantee an accurate prediction, given the confidence interval and the error of estimate. The reason is that when $\hat{p}$ and $\hat{q}$ are each 0.5, the product $\hat{p}\hat{q}$ is at maximum, as shown here.

$\hat{p}$	$\hat{q}$	$\hat{p}\hat{q}$
0.1	0.9	0.09
0.2	0.8	0.16
0.3	0.7	0.21
0.4	0.6	0.24
0.5	**0.5**	**0.25**
0.6	0.4	0.24
0.7	0.3	0.21
0.8	0.2	0.16
0.9	0.1	0.09

Example 7–11

A researcher wishes to estimate, with 95% confidence, the proportion of people who own a home computer. A previous study shows that 40% of those interviewed had a computer at home. The researcher wishes to be accurate within 2% of the true proportion. Find the minimum sample size necessary.

Solution

Since $z_{\alpha/2} = 1.96$, $E = 0.02$, $\hat{p} = 0.40$, and $\hat{q} = 0.60$, then

$$n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2 = (0.40)(0.60)\left(\frac{1.96}{0.02}\right)^2 = 2304.96$$

which, when rounded up, is 2305 people to interview.

Example 7–12

The same researcher wishes to estimate the proportion of executives who own a car phone. She wants to be 90% confident and be accurate within 5% of the true proportion. Find the minimum sample size necessary.

Solution

Since there is no prior knowledge of $\hat{p}$, statisticians assign the values $\hat{p} = 0.5$ and $\hat{q} = 0.5$. The sample size obtained by using these values will be large enough to ensure the specified degree of confidence. Hence,

$$n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2 = (0.5)(0.5)\left(\frac{1.65}{0.05}\right)^2 = 272.25$$

which, when rounded up, is 273 executives to ask.

In determining the sample size, the size of the population is irrelevant. Only the degree of confidence and the maximum error are necessary to make the determination.

Exercises 7–4

1. In each case, find $\hat{p}$ and $\hat{q}$.
a. $n = 80$ and $X = 40$
b. $n = 200$ and $X = 90$
c. $n = 130$ and $X = 60$
d. $n = 60$ and $X = 35$
e. $n = 95$ and $X = 43$

2. (**ans**) Find $\hat{p}$ and $\hat{q}$ for each percentage. (Use each percentage for $\hat{p}$.)
a. 12%
b. 29%
c. 65%
d. 53%
e. 67%

3. A U.S. Travel Data Center survey conducted for *Better Homes and Gardens* of 1500 adults found that 39% said that they would take more vacations this year than last year. Find the 95% confidence interval for the true proportion of adults who said that they will travel more this year.

Source: *USA TODAY.*

4. A recent study of 100 people in Miami found 27 were obese. Find the 90% confidence interval of the population proportion of individuals living in Miami who are obese.

Source: Based on information from the Center for Disease Control and Prevention, *USA TODAY.*

5. The proportion of students in private schools is around 11 percent. A random sample of 450 students from a wide geographic area indicated that 55 attended private schools. Estimate the true proportion of students attending private schools with 95% confidence. How does your estimate compare to 11%?

Source: National Center for Education Statistics (nces.ed.gov).

6. The Gallup Poll found that 27% of adults surveyed nationwide said they had personally been in a tornado. How many adults should be surveyed to estimate the true proportion of adults who have been in a tornado with a 95% confidence interval 5% wide?

Source: www.pollingreport.com.

7. A survey found that out of 200 workers, 168 said they were interrupted three or more times an hour by phone messages, faxes, etc. Find the 90% confidence interval of the population proportion of workers who are interrupted three or more times an hour.

Source: Based on information from *USA TODAY* Snapshot.

8. A CBS News/*N.Y. Times* poll found that 329 out of 763 adults said they would travel to outer space in their lifetime, given the chance. Estimate the true proportion of adults who would like to travel to outer space with 92% confidence.

Source: www.pollingreport.com.

9. A study by the University of Michigan found that one in five 13- and 14-year-olds is a sometime smoker. To see how the smoking rate of the students at a large school district compared to the national rate, the superintendent surveyed two hundred 13- and 14-year-old students and found that 23% said they were sometime smokers. Find the 99% confidence interval of the true proportion and compare this with the University of Michigan's study.

Source: *USA TODAY.*

10. A survey of 80 recent fatal traffic accidents showed that 46 were alcohol-related. Find the 95% confidence interval of the true proportion of fatal alcohol-related accidents.

11. A survey of 90 families showed that 40 owned at least one gun. Find the 95% confidence interval of the true proportion of families who own at least one gun.

12. In a certain state, a survey of 500 workers showed that 45% belonged to a union. Find the 90% confidence interval of the true proportion of workers who belong to a union.

13. In a Gallup Poll of 1005 individuals, 452 thought they were worse off financially than a year ago. Find the 95% confidence interval for the true proportion of individuals that feel they are worse off financially.

Source: Gallup Poll.

14. In a poll of 1000 likely voters, 560 say that the United States spends too little on fighting hunger at home. Find a 95% confidence interval for the true proportion of voters who feel this way.

Source: Alliance to End Hunger.

15. A medical researcher wishes to determine the percentage of females who take vitamins. He wishes to be 99% confident that the estimate is within 2 percentage points of the true proportion. A recent study of 180 females showed that 25% took vitamins.

a. How large should the sample size be?

b. If no estimate of the sample proportion is available, how large should the sample be?

16. A recent study indicated that 29% of the 100 women over age 55 in the study were widows.

a. How large a sample must one take to be 90% confident that the estimate is within 0.05 of the true proportion of women over 55 who are widows?

b. If no estimate of the sample proportion is available, how large should the sample be?

17. A researcher wishes to estimate the proportion of adult males who are under 5 feet 5 inches tall. She wants to be 90% confident that her estimate is within 5% of the true proportion.

a. How large a sample should be taken if in a sample of 300 males, 30 were under 5 feet 5 inches tall?

b. If no estimate of the sample proportion is available, how large should the sample be?

18. Obesity is defined as a *body mass index* (BMI) of 3 kg/m^2 or more. A 95% confidence interval for the percentage of U.S. adults aged 20 years and over who were obese was found to be 22.4% to 23.5%. What was the sample size?

Source: National Center for Health Statistics (www.cdc.gov/nchs).

19. How large a sample should be surveyed to estimate the true proportion of college students who do laundry once a week within 3% with 95% confidence?

20. A federal report indicated that 27% of children ages 2 to 5 years had a good diet—an increase over previous years. How large a sample is needed to estimate the true proportion of children with good diets within 2% with 95% confidence?

Source: Federal Interagency Forum on Child and Family Statistics, *Washington Observer-Reporter.*

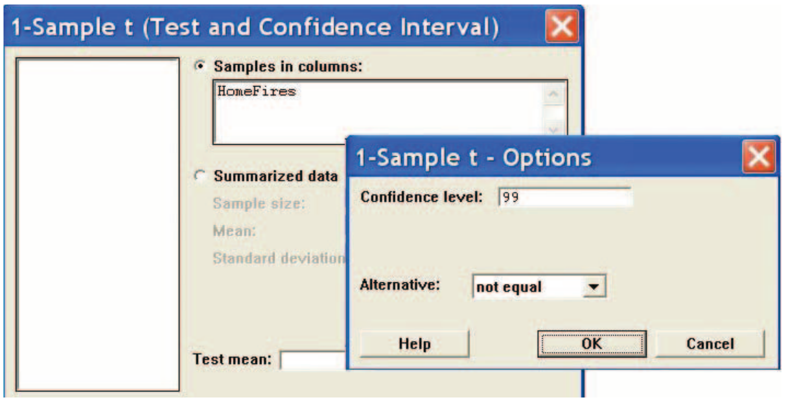
Technology Step by Step

MINITAB
Step by Step

Find a Confidence Interval for a Proportion

MINITAB will calculate a confidence interval, given the statistics from a sample *or* given the raw data. In Example 7–9, in a sample of 500 nursing applications 60 were from men. Find the 90% confidence interval estimate for the true proportion of male applicants.

1. Select **Stat>Basic Statistics>1 Proportion.**
2. Click on the button for Summarized data. No data will be entered in the worksheet.
3. Click in the box for Number of trials and enter **500.**
4. In the Number of events box, enter **60.**
5. Click on [Options].
6. Type **90** for the confidence level.



7. Check the box for Use test and interval based on normal distribution.
8. Click [OK] twice.

The results for the confidence interval will be displayed in the session window.

Test and CI for One Proportion

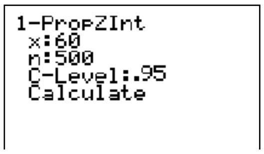
```
Test of p = 0.5 vs p not = 0.
Sample    X     N   Sample p          90% CI           Z-Value   P-Value
1        60   500   0.120000   (0.096096, 0.143904)     -16.99     0.000
```

TI-83 Plus
Step by Step

Finding a Confidence Interval for a Proportion

1. Press **STAT** and move the cursor to TESTS.
2. Press **A (ALPHA, MATH)** for 1-PropZInt.
3. Type in the appropriate values.
4. Move the cursor to Calculate and press **ENTER.**

Input

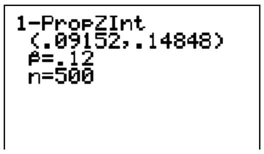


Example TI7–5

Find the 95% confidence interval of p when $X = 60$ and $n = 500$, as in Example 7–9.

The 95% confidence level for p is $0.09152 < p < 0.14848$. Also $\hat{p}$ is given.

Output



Excel
Step by Step

Finding a Confidence Interval for a Proportion

Excel does not produce confidence intervals for a proportion. However, you may determine confidence intervals for a proportion by using the Mega-Stat Add-in available on your CD and Online

Speaking of STATISTICS

Here is a survey about college students' credit card usage. Suggest several ways that the study could have been more meaningful if confidence intervals had been used.

OTHER PEOPLE'S MONEY

Undergrads love their plastic. That means—you guessed it—students are learning to become debtors. According to the Public Interest Research Groups, only half of all students pay off card balances in full each month, 36% sometimes do and 14% never do. Meanwhile, 48% have paid a late fee. Here's how undergrads stack up, according to Nellie Mae, a provider of college loans:

Undergrads with a credit card	**78%**
Average number of cards owned . .	**3**
Average student card debt	**$1236**
Students with 4 or more cards. . . .	**32%**
Balances of $3000 to $7000	**13%**
Balances over $7000	**9%**

Learning Center. If you have not installed this add-in, do so by following the instructions on page 24.

There were 500 nursing applications in a sample, including 60 from men. Find the 90% confidence interval for the true proportion of male applicants.

1. From the toolbar, select **Mega-Stat>Confidence Intervals/Sample Size.**
2. In the dialog box, select the Confidence interval — *p*.
3. Enter **60** in the first box; *p* will automatically switch to *x*.
4. Enter **500** in the second box for *n*.
5. Either type in or scroll to 90% for the confidence level, and then click [OK].

The result of the procedure will show the output in a new chart, as indicated here.

Confidence interval - proportion

```
  90% confidence level
 0.12 proportion
  500 n
1.645 z
0.024 half-width
0.144 upper confidence limit
0.096 lower confidence limit
```

7–5

Confidence Intervals for Variances and Standard Deviations

Objective 6. Find a confidence interval for a variance and a standard deviation.

In Sections 7–2 through 7–4, confidence intervals were calculated for means and proportions. This section will explain how to find confidence intervals for variances and standard deviations. In statistics, the variance and standard deviation of a variable are as important as the mean. For example, when products that fit together (such as pipes) are manufactured, it is important to keep the variations of the diameters of the products as small as possible; otherwise, they will not fit together properly and will have to be scrapped. In the manufacture of medicines, the variance and standard deviation of the medication in the pills play an important role in making sure patients receive the proper dosage. For these reasons, confidence intervals for variances and standard deviations are necessary.

To calculate these confidence intervals, a new statistical distribution is needed. It is called the **chi-square distribution.**

The chi-square variable is similar to the *t* variable in that its distribution is a family of curves based on the number of degrees of freedom. The symbol for chi-square is χ^2 (Greek letter chi, pronounced "ki"). Several of the distributions are shown in Figure 7–9, along with the corresponding degrees of freedom. The chi-square distribution is obtained from the values of $(n - 1)s^2/\sigma^2$ when random samples are selected from a normally distributed population whose variance is σ^2.

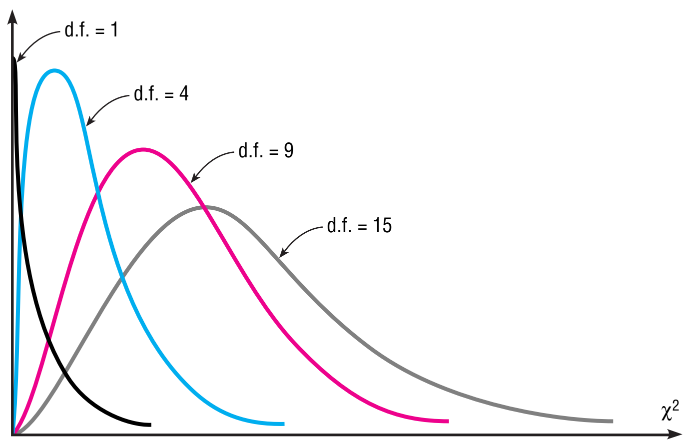
A chi-square variable cannot be negative, and the distributions are positively skewed. At about 100 degrees of freedom, the chi-square distribution becomes somewhat symmetric. The area under each chi-square distribution is equal to 1.00, or 100%.

Table G in Appendix C gives the values for the chi-square distribution. These values are used in the denominators of the formulas for confidence intervals. Two different values are used in the formula. One value is found on the left side of the table, and the other is on the right. For example, to find the table values corresponding to the 95% confidence interval, one must first change 95% to a decimal and subtract it from

Historical Note

The χ^2 distribution with 2 degrees of freedom was formulated by a mathematician named Hershel in 1869 while he was studying the accuracy of shooting arrows at a target. Many other mathematicians have since contributed to its development.

Figure 7–9
The Chi-Square Family of Curves



1 (1 − 0.95 = 0.05). Then divide the answer by 2 ($\alpha/2 = 0.05/2 = 0.025$). This is the column on the right side of the table, used to get the values for χ^2_{right}. To get the value for χ^2_{left}, subtract the value of $\alpha/2$ from 1 (1 − 0.05/2 = 0.975). Finally, find the appropriate row corresponding to the degrees of freedom $n - 1$. A similar procedure is used to find the values for a 90% or 99% confidence interval.

Example 7–13

Find the values for χ^2_{right} and χ^2_{left} for a 90% confidence interval when $n = 25$.

Solution

To find χ^2_{right}, subtract 1 − 0.90 = 0.10 and divide by 2 to get 0.05.

To find χ^2_{left}, subtract 1 − 0.05 to get 0.95. Hence, use the 0.95 and 0.05 columns and the row corresponding to 24 d.f. See Figure 7–10.

Figure 7–10
χ^2 Table for Example 7–13

Table G
The Chi-square Distribution

Degrees of freedom	α 0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1										
2										
⋮										
24				13.848 (χ^2 left)			36.415 (χ^2 right)			

The answers are

$$\chi^2_{\text{right}} = 36.415$$
$$\chi^2_{\text{left}} = 13.848$$

Useful estimates for σ^2 and σ are s^2 and s, respectively.

To find confidence intervals for variances and standard deviations, one must assume that the variable is normally distributed.

The formulas for the confidence intervals are shown here.

Formula for the Confidence Interval for a Variance

$$\frac{(n-1)s^2}{\chi^2_{\text{right}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\text{left}}}$$

d.f. $= n - 1$

Formula for the Confidence Interval for a Standard Deviation

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\text{right}}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{\text{left}}}}$$

d.f. $= n - 1$

Recall that s^2 is the symbol for the sample variance and s is the symbol for the sample standard deviation. If the problem gives the sample standard deviation s, be sure to *square* it when you are using the formula. But if the problem gives the sample variance s^2, *do not square it* when using the formula, since the variance is already in square units.

Rounding Rule for a Confidence Interval for a Variance or Standard Deviation When you are computing a confidence interval for a population variance or standard deviation by using raw data, round off to one more decimal place than the number of decimal places in the original data.

When you are computing a confidence interval for a population variance or standard deviation by using a sample variance or standard deviation, round off to the same number of decimal places as given for the sample variance or standard deviation.

Example 7–14 shows how to find a confidence interval for a variance and standard deviation.

Example 7–14

Find the 95% confidence interval for the variance and standard deviation of the nicotine content of cigarettes manufactured if a sample of 20 cigarettes has a standard deviation of 1.6 milligrams.

Solution

Since $\alpha = 0.05$, the two critical values, respectively, for the 0.025 and 0.975 levels for 19 degrees of freedom are 32.852 and 8.907. The 95% confidence interval for the variance is found by substituting in the formula:

$$\frac{(n-1)s^2}{\chi^2_{\text{right}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\text{left}}}$$

$$\frac{(20-1)(1.6)^2}{32.852} < \sigma^2 < \frac{(20-1)(1.6)^2}{8.907}$$

$$1.5 < \sigma^2 < 5.5$$

Hence, one can be 95% confident that the true variance for the nicotine content is between 1.5 and 5.5.

For the standard deviation, the confidence interval is

$$\sqrt{1.5} < \sigma < \sqrt{5.5}$$
$$1.2 < \sigma < 2.3$$

Hence, one can be 95% confident that the true standard deviation for the nicotine content of all cigarettes manufactured is between 1.2 and 2.3 milligrams based on a sample of 20 cigarettes.

Example 7–15

Find the 90% confidence interval for the variance and standard deviation for the price in dollars of an adult single-day ski lift ticket. The data represent a selected sample of nationwide ski resorts. Assume the variable is normally distributed.

59	54	53	52	51
39	49	46	49	48

Source: *USA TODAY*.

Solution

STEP 1 Find the variance for the data. Use the formulas in Chapter 3 or your calculator. The variance $s^2 = 28.2$.

STEP 2 Find χ^2_{right} and χ^2_{left} from Table G in Appendix C. Since $\alpha = 0.10$, the two critical values are 3.325 and 16.919, using d.f. $= 9$ and 0.95 and 0.05.

STEP 3 Substitute in the formula and solve.

$$\frac{(n-1)s^2}{\chi^2_{\text{right}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\text{left}}}$$
$$\frac{(10-1)(28.2)}{16.919} < \sigma^2 < \frac{(10-1)(28.2)}{3.325}$$
$$15.0 < \sigma^2 < 76.3$$

For the standard deviation

$$\sqrt{15} < \sigma < \sqrt{76.3}$$
$$3.87 < \sigma < 8.73$$

Hence one can be 90% confident that the standard deviation for the price of all single-day ski lift tickets of the population is between \$3.87 and \$8.73 based on a sample of 10 nationwide ski resorts. (Two decimal places are used since the data are in dollars and cents.)

Note: If you are using the standard deviation instead (as in Example 7–14) of the variance, be sure to square the standard deviation when substituting in the formula.

Exercises 7–5

1. What distribution must be used when computing confidence intervals for variances and standard deviations?

2. What assumption must be made when computing confidence intervals for variances and standard deviations?

3. Using Table G, find the values for χ^2_{left} and χ^2_{right}.

a. $\alpha = 0.05, n = 16$
b. $\alpha = 0.10, n = 5$
c. $\alpha = 0.01, n = 23$
d. $\alpha = 0.05, n = 29$
e. $\alpha = 0.10, n = 14$

4. Find the 95% confidence interval for the variance and standard deviation for the lifetime of batteries if a sample of 20 batteries has a standard deviation of 1.7 months. Assume the variable is normally distributed.

5. Find the 90% confidence interval for the variance and standard deviation for the time it takes an inspector to check a bus for safety if a sample of 27 buses has a standard deviation of 6.8 minutes. Assume the variable is normally distributed.

6. Find the 99% confidence interval for the variance and standard deviation of the weights of 25 one-gallon containers of motor oil if a sample of 14 containers has a variance of 3.2. The weights are given in ounces. Assume the variable is normally distributed.

7. The sugar content (in grams) for a random sample of 4-ounce containers of applesauce is shown. Find the 99% confidence interval for the population variance and standard deviation. Assume the variable is normally distributed.

18.6	19.5	20.2	20.4	19.3
21.0	20.3	19.6	20.7	18.9
22.1	19.7	20.8	18.9	20.7
21.6	19.5	20.1	20.3	19.9

8. Find the 90% confidence interval for the variance and standard deviation of the ages of seniors at Oak Park College if a sample of 24 students has a standard deviation of 2.3 years. Assume the variable is normally distributed.

9. The number of calories in a 1-ounce serving of various kinds of regular cheese is shown. Estimate the population variance and standard deviation with 90% confidence.

110	45	100	95	110
110	100	110	95	120
130	100	80	105	105
90	110	70	125	108

Source: *The Doctor's Pocket Calorie, Fat, and Carbohydrate Counter.*

10. A random sample of stock prices per share (in dollars) is shown. Find the 90% confidence interval for the variance and standard deviation for the prices. Assume the variable is normally distributed.

26.69	13.88	28.37	12.00
75.37	7.50	47.50	43.00
3.81	53.81	13.62	45.12
6.94	28.25	28.00	60.50
40.25	10.87	46.12	14.75

Source: *Pittsburgh Tribune Review.*

11. A service station advertises that customers will have to wait no more than 30 minutes for an oil change. A sample of 28 oil changes has a standard deviation of 5.2 minutes. Find the 95% confidence interval of the population standard deviation of the time spent waiting for an oil change.

12. Find the 95% confidence interval for the variance and standard deviation of the ounces of coffee that a machine dispenses in 12-ounce cups. Assume the variable is normally distributed. The data are given.

12.03 12.10 12.02 11.98 12.00 12.05 11.97 11.99

Extending the Concepts

13. A confidence interval for a standard deviation for large samples taken from a normally distributed population can be approximated by

$$s - z_{\alpha/2}\frac{s}{\sqrt{2n}} < \sigma < s + z_{\alpha/2}\frac{s}{\sqrt{2n}}$$

Find the 95% confidence interval for the population standard deviation of calculator batteries. A sample of 200 calculator batteries has a standard deviation of 18 months.

Technology Step by Step

TI-83 Plus Step by Step

The TI-83 Plus does not have a built-in confidence interval for the variance or standard deviation. However, the downloadable program named SDINT is available on your CD and Online Learning Center. Follow the instructions with your CD for downloading the program.

Finding a Confidence Interval for the Variance and Standard Deviation (Data)

1. Enter the data values into L_1.
2. Press **PRGM,** move the cursor to the program named SDINT, and press **ENTER** twice.
3. Press **1** for Data.
4. Type $\mathbf{L_1}$ for the list and press **ENTER.**
5. Type the confidence level and press **ENTER.**
6. Press **ENTER** to clear the screen.

Example TI7–3

This refers to Example 7–15 in the text. Find the 90% confidence interval for the variance and standard deviation for the data:

59 54 53 52 51 39 49 46 49 48

```
LIST ?L1
```

```
ENTER CONF LEVEL
(0 < CL <1)

CONF LEVEL=.9
```

```
CONF LEVEL .9
S= 5.31   n=10

CONF INT FOR σx²
{15.01,76.39}
CONF INT FOR σx
{3.87,8.74}
ENTER TO CLEAR
```

Finding a Confidence Interval for the Variance and Standard Deviation (Statistics)

1. Press **PRGM,** move the cursor to the program named SDINT, and press **ENTER** twice.
2. Press **2** for Stats.
3. Type the sample standard deviation and press **ENTER.**
4. Type the sample size and press **ENTER.**
5. Type the confidence level and press **ENTER.**
6. Press **ENTER** to clear the screen.

Example TI7–4

This refers to Example 7–14 in the text. Find the 95% confidence interval for the variance and standard deviation, given $n = 20$ and $s = 1.6$.

```
S= 1.6
N= 20
```

```
ENTER CONF LEVEL
(0 < CL <1)

CONF LEVEL=.95■
```

```
CONF LEVEL .95
S= 1.6   n=20

CONF INT FOR σx²
{1.48,5.46}
CONF INT FOR σx
{1.22,2.34}
ENTER TO CLEAR
```

7–6 Summary

An important aspect of inferential statistics is estimation. Estimations of parameters of populations are accomplished by selecting a random sample from that population and choosing and computing a statistic that is the best estimator of the parameter. A good estimator must be unbiased, consistent, and relatively efficient. The best estimators of μ and p are $\overline{X}$ and $\hat{p}$, respectively. The best estimators of σ^2 and σ are s^2 and s, respectively.

There are two types of estimates of a parameter: point estimates and interval estimates. A point estimate is a specific value. For example, if a researcher wishes to estimate the average length of a certain adult fish, a sample of the fish is selected and measured. The mean of this sample is computed, for example, 3.2 centimeters. From this sample mean, the researcher estimates the population mean to be 3.2 centimeters.

The problem with point estimates is that the accuracy of the estimate cannot be determined. For this reason, statisticians prefer to use the interval estimate. By computing an interval about the sample value, statisticians can be 95% or 99% (or some other percentage) confident that their estimate contains the true parameter. The confidence level is determined by the researcher. The higher the confidence level, the wider the interval of the estimate must be. For example, a 95% confidence interval of the true mean length of a certain species of fish might be

$$3.17 < \mu < 3.23$$

whereas the 99% confidence interval might be

$$3.15 < \mu < 3.25$$

When the confidence interval of the mean is computed, the z or t values are used, depending on whether the population standard deviation is known and depending on the size of the sample. If σ is known or $n \geq 30$, the z values can be used. If σ is not known, the t values must be used when the sample size is less than 30 and the population is normally distributed.

Closely related to computing confidence intervals is the determination of the sample size to make an estimate of the mean. This information is needed to determine the minimum sample size necessary.

1. The degree of confidence must be stated.
2. The population standard deviation must be known or be able to be estimated.
3. The maximum error of estimate must be stated.

Confidence intervals and sample sizes can also be computed for proportions, using the normal distribution; and confidence intervals for variances and standard deviations can be computed, using the chi-square distribution.

Important Terms

chi-square distribution 354
confidence interval 328
confidence level 328
consistent estimator 327
degrees of freedom 340
estimation 326
estimator 327
interval estimate 328
maximum error of estimate 329
point estimate 327
proportion 346
relatively efficient estimator 327
t distribution 340
unbiased estimator 327

Important Formulas

Formula for the confidence interval of the mean when σ is known (when $n \geq 30$, s can be used if σ is unknown):

$$\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Formula for the sample size for means:

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

where E is the maximum error.

Formula for the confidence interval of the mean when σ is unknown and $n < 30$:

$$\bar{X} - t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right) < \mu < \bar{X} + t_{\alpha/2}\left(\frac{s}{\sqrt{n}}\right)$$

Formula for the confidence interval for a proportion:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

where $\hat{p} = X/n$ and $\hat{q} = 1 - \hat{p}$.

Formula for the sample size for proportions:

$$n = \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2$$

Formula for the confidence interval for a variance:

$$\frac{(n-1)s^2}{\chi^2_{\text{right}}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\text{left}}}$$

Formula for confidence interval for a standard deviation:

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\text{right}}}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi^2_{\text{left}}}}$$

Review Exercises

1. A study of 36 members of the Central Park Walkers showed that they could walk at an average rate of 2.6 miles per hour. The sample standard deviation is 0.4. Find the 95% confidence interval for the mean for all walkers.

2. In a survey of 1004 individuals, 442 felt that President George W. Bush spent too much time away from Washington. Find a 95% confidence interval for the true population proportion.

Source: *USA TODAY*/CNN/Gallup Poll.

3. A U.S. Travel Data Center survey reported that Americans stayed an average of 7.5 nights when they went on vacation. The sample size was 1500. Find the 95% confidence interval of the true mean. Assume the standard deviation was 0.8 night.

Source: *USA TODAY.*

4. A sample of 12 funeral homes in Toledo, Ohio, found these costs for cremation (in dollars).

1320	1052	1090	1285	1570	995	1150
1585	820	1195	1160	590		

Find the 95% confidence interval for the population mean.

Source: *Toledo Blade.*

5. For a certain urban area, in a sample of 5 months, an average of 28 mail carriers were bitten by dogs each month. The standard deviation of the sample was 3. Find the 90% confidence interval of the true mean number of mail carriers who are bitten by dogs each month. Assume the variable is normally distributed.

6. A researcher is interested in estimating the average salary of teachers in a large urban school district. She wants to be 95% confident that her estimate is correct. If the standard deviation is $1050, how large a sample is needed to be accurate within $200?

7. A researcher wishes to estimate, within $25, the true average amount of postage a community college spends each year. If she wishes to be 90% confident, how large a sample is necessary? The standard deviation is known to be $80.

8. A U.S. Travel Data Center's survey of 1500 adults found that 42% of respondents stated that they favor historical sites as vacations. Find the 95% confidence interval of the true proportion of all adults who favor visiting historical sites as vacations.

Source: *USA TODAY.*

9. In a study of 200 accidents that required treatment in an emergency room, 40% occurred at home. Find the 90% confidence interval of the true proportion of accidents that occur at home.

10. In a recent study of 100 people, 85 said that they were dissatisfied with their local elected officials. Find the 90% confidence interval of the true proportion of individuals who are dissatisfied with their local elected officials.

11. The federal report in the previous exercise also stated that 88% of children under age 18 were covered by health insurance in 2000. How large a sample is needed to estimate the true proportion of covered children with 90% confidence with a confidence interval 0.05 wide?

Source: *Washington Observer-Reporter.*

12. A survey by *Brides* magazine found that 8 out of 10 brides are planning to take the surname of their new husband. How large a sample is needed to estimate the true proportion within 3% with 98% confidence?

Source: *Time* magazine.

13. The standard deviation of the diameter of 28 oranges was 0.34 inch. Find the 99% confidence interval of the true standard deviation of the diameters of the oranges.

14. A random sample of 22 lawn mowers was selected, and the motors were tested to see how many miles per gallon of gasoline each one obtained. The variance of the measurements was 2.6. Find the 95% confidence interval of the true variance.

15. A random sample of 15 snowmobiles was selected, and the lifetime (in months) of the batteries was measured. The variance of the sample was 8.6. Find the 90% confidence interval of the true variance.

16. The heights of 28 police officers from a large-city police force were measured. The standard deviation of the sample was 1.83 inches. Find the 95% confidence interval of the standard deviation of the heights of the officers.

Statistics Today

Would You Change the Channel?—Revisited

The estimates given in the survey are point estimates. However, since the margin of error is stated to be 3 percentage points, an interval estimate can easily be obtained. For example, if 45% of the people changed the channel, then the confidence interval of the true percentages of people who changed channels would be $42\% < p < 48\%$. The article fails to state whether a 90%, 95%, or some other percentage was used for the confidence interval.

Using the formula given in Section 7–4, a minimum sample size of 1068 would be needed to obtain a 95% confidence interval for p, as shown. Use $\hat{p}$ and $\hat{q}$ as 0.5, since no value is known for $\hat{p}$.

$$\begin{aligned} n &= \hat{p}\hat{q}\left(\frac{z_{\alpha/2}}{E}\right)^2 \\ &= (0.5)(0.5)\left(\frac{1.96}{0.03}\right)^2 = 1067.1 \\ &= 1068 \end{aligned}$$

Data Analysis

The Data Bank is found in Appendix D, or on the World Wide Web by following links from www.mhhe.com/math/stat/bluman/.

1. From the Data Bank choose a variable, find the mean, and construct the 95% and 99% confidence intervals of the population mean. Use a sample of at least 30 subjects. Find the mean of the population, and determine whether it falls within the confidence interval.

2. Repeat Exercise 1, using a different variable and a sample of 15.

3. Repeat Exercise 1, using a proportion. For example, construct a confidence interval for the proportion of individuals who did not complete high school.

4. From Data Set III in Appendix D, select a sample of 30 values and construct the 95% and 99% confidence intervals of the mean length in miles of major North American rivers. Find the mean of all the values, and determine if the confidence intervals contain the mean.

5. From Data Set VI in Appendix D, select a sample of 20 values and find the 90% confidence interval of the mean of the number of acres. Find the mean of all the values, and determine if the confidence interval contains the mean.

6. From Data Set XIV in Appendix D, select a sample of 20 weights for the Pittsburgh Steelers, and find the proportion of players who weigh over 250 pounds. Construct a 95% confidence interval for this proportion. Find the proportion of all Steelers who weigh more than 250 pounds. Does the confidence interval contain this value?

Chapter Quiz

Determine whether each statement is true or false. If the statement is false, explain why.

1. Interval estimates are preferred over point estimates since a confidence level can be specified.

2. For a specific confidence interval, the larger the sample size, the smaller the maximum error of estimate will be.

3. An estimator is consistent if, as the sample size decreases, the value of the estimator approaches the value of the parameter estimated.

4. To determine the sample size needed to estimate a parameter, one must know the maximum error of estimate.

Select the best answer.

5. When a 99% confidence interval is calculated instead of a 95% confidence interval with n being the same, the maximum error of estimate will be

a. Smaller
b. Larger
c. The same
d. It cannot be determined

6. The best point estimate of the population mean is

a. The sample mean
b. The sample median
c. The sample mode
d. The sample midrange

7. When the population standard deviation is unknown and sample size is less than 30, what table value should be used in computing a confidence interval for a mean?

a. z
b. t
c. Chi-square
d. None of the above

Complete the following statements with the best answer.

8. A good estimator should be _______, _______, and _______.

9. The maximum difference between the point estimate of a parameter and the actual value of the parameter is called _______.

10. The statement "The average height of an adult male is 5 feet 10 inches" is an example of a(n) _______ estimate.

11. The three confidence intervals used most often are the _______%, _______%, and _______%.

12. A random sample of 49 shoppers showed that they spend an average of $23.45 per visit at the Saturday Mornings Bookstore. The standard deviation of the sample was $2.80. Find the 90% confidence interval of the true mean.

13. An irate patient complained that the cost of a doctor's visit was too high. She randomly surveyed 20 other patients and found that the mean amount of money they spent on each doctor's visit was $44.80. The standard deviation of the sample was $3.53. Find the 95% confidence interval of the population mean. Assume the variable is normally distributed.

14. The average weight of 40 randomly selected school buses was 4150 pounds. The standard deviation was 480 pounds. Find the 99% confidence interval of the true mean weight of the buses.

15. In a study of 10 insurance sales representatives from a certain large city, the average age of the group was 48.6 years and the standard deviation was 4.1 years. Find the 95% confidence interval of the population mean age of all insurance sales representatives in that city.

16. In a hospital, a sample of 8 weeks was selected, and it was found that an average of 438 patients were treated in the emergency room each week. The standard deviation was 16. Find the 99% confidence interval of the true mean. Assume the variable is normally distributed.

17. For a certain urban area, it was found that in a sample of 4 months, an average of 31 burglaries occurred each month. The standard deviation was 4. Find the 90% confidence interval of the true mean number of burglaries each month.

18. A university dean wishes to estimate the average number of hours that freshmen study each week. The standard deviation from a previous study is 2.6 hours. How large a sample must be selected if he wants to be 99% confident of finding whether the true mean differs from the sample mean by 0.5 hour?

19. A researcher wishes to estimate within $300 the true average amount of money a county spends on road repairs each year. If she wants to be 90% confident, how large a sample is necessary? The standard deviation is known to be $900.

20. A recent study of 75 workers found that 53 people rode the bus to work each day. Find the 95% confidence interval of the proportion of all workers who rode the bus to work.

21. In a study of 150 accidents that required treatment in an emergency room, 36% involved children under 6 years of age. Find the 90% confidence interval of the true proportion of accidents that involve children under the age of 6.

22. A survey of 90 families showed that 40 owned at least one television set. Find the 95% confidence interval of the true proportion of families who own at least one television set.

23. A nutritionist wishes to determine, within 3%, the true proportion of adults who do not eat any lunch. If he wishes to be 95% confident that his estimate contains the population proportion, how large a sample will be necessary? A previous study found that 15% of the 125 people surveyed said they did not eat lunch.

24. A sample of 25 novels has a standard deviation of 9 pages. Find the 95% confidence interval of the population standard deviation.

25. Find the 90% confidence interval for the variance and standard deviation for the time it takes a state police inspector to check a truck for safety if a sample of 27 trucks has a standard deviation of 6.8 minutes. Assume the variable is normally distributed.

26. A sample of 20 automobiles has a pollution by-product release standard deviation of 2.3 ounces when 1 gallon of gasoline is used. Find the 90% confidence interval of the population standard deviation.

Critical Thinking Challenges

A confidence interval for a median can be found by using the formulas

$$U = \frac{n+1}{2} + \frac{z_{\alpha/2}\sqrt{n}}{2} \qquad \text{(round up)}$$

$$L = n - U + 1$$

to define positions in the set of ordered data values.

Suppose a data set has 30 values, and one wants to find the 95% confidence interval for the median. Substituting in the formulas, one gets

$$U = \frac{30+1}{2} + \frac{1.96\sqrt{30}}{2} = 21 \qquad \text{(rounded up)}$$

$$L = 30 - 21 + 1 = 10$$

when $n = 30$ and $z_{\alpha/2} = 1.96$.

Arrange the data in order from smallest to largest, and then select the 10th and 21st values of the data array; hence, $X_{10} < \text{med} < X_{21}$.

Find the 90% confidence interval for the median for the data in Exercise 12 in Section 7–2.

Data Projects

Use MINITAB, the TI-83 Plus, or a computer program of your choice to complete these exercises.

1. Select several variables, such as the number of points a football team scored in each game of a specific season, the number of passes completed, or the number of yards gained. Using confidence intervals for the mean, determine the 90%, 95%, and 99% confidence intervals. (Use z or t, whichever is relevant.) Decide which you think is more appropriate. When this is completed, write a summary of your findings by answering the following questions.

a. What was the purpose of the study?
b. What was the population?
c. How was the sample selected?
d. What were the results obtained by using confidence intervals?
e. Did you use z or t? Why?

2. Using the same data or different data, construct a confidence interval for a proportion. For example, you might want to find the proportion of passes completed by the quarterback or the proportion of passes that were intercepted. Write a short paragraph summarizing the results.

You may use the following websites to obtain raw data:

http://www.mhhe.com/math/stat/bluman/
http://lib.stat.cmu.edu/DASL
http://www.statcan.ca/english/