Chapter 8

Estimation

Chapter Learning Objectives

- Understanding the concept of estimation and the reasons for it
- Estimating confidence intervals for means
- Understanding the concept of risk and how to reduce it
- Estimating confidence intervals for proportions

In this chapter, we discuss the procedures involved in estimating population means and proportions. These procedures are based on the principles of sampling and statistical inference discussed in Chapter 7. Knowledge about the sampling distribution allows us to estimate population means and proportions from sample outcomes and to assess the accuracy of these estimates.

Example 1: Based on a survey of 862 likely Democratic primary voters, Zogby's political pollsters predicted that the then Senator Barack Obama would defeat Senator Hillary Clinton in the New Hampshire primary election by 13 percentage points. However, when official results were declared the day after the election, we saw that Senator Clinton actually ended up defeating Senator Obama in New Hampshire by just under 3 percentage points. The failure of the polls to predict the outcome of the race is considered as one of the worst failures in polling history.

Example 2: Each month, the Bureau of Labor Statistics interviews a sample of about 50,000 adult Americans to determine job-related activities. Based on these interviews, monthly estimates are made of statistics such as the unemployment rate (the proportion who are unemployed), average earnings, the percentage of the workforce working part-time, and the percentage collecting unemployment benefits. These estimates are considered so vital that they cause fluctuations in the stock market and influence economic policies of the federal government.

Example 3: Shortly after the state of Arizona introduced its controversial immigration law in 2010, the Gallup Research Organization conducted a national telephone survey to assess public opinion of the new law. Gallup interviewed 1,013 adults over the telephone and found 51% of Americans in favor of the new law, whereas 39% opposed it (*Source:* "More Americans Favor Than Oppose Arizona Immigration Law," Gallup Poll, April 27–28, 2010).

Example 4: Every other year, the National Opinion Research Center (NORC) conducts the General Social Survey (GSS) on a representative sample of about 1,500 respondents. The GSS, from which many of the examples in this book are selected, is designed to provide social science researchers with a readily accessible database of socially relevant attitudes, behaviors, and attributes of a cross section of the U.S. adult population. For example, in analyzing the responses to the 2008 GSS, researchers found that the average respondent's education was about 13.43 years. This average probably differs from the average of the population from which the GSS sample was drawn. However, we can establish that in most cases the sample mean (in this case, 13.43 years) is fairly close to the actual true average in the population.

As you read these examples, you may have questioned the reliability of some of the numbers. Are preelection pollsters inaccurate seeing that they incorrectly predicted which senator would win from New Hampshire? What is the actual percentage of unemployed Americans? How do all Americans feel about Arizona's immigration law? What is the actual average level of education in the United States?

ESTIMATION DEFINED

The official outcome of the New Hampshire primary, the actual percentage of unemployed Americans, Americans' attitude toward Arizona's new immigration law, and the actual average level of education in the United States are all population parameters. Preelection predictions of which senator would win New Hampshire, the percentage of unemployed Americans as estimated by the Bureau of Labor Statistics, Americans' attitude toward Arizona's immigration law as assessed by the Gallup organization, and the average level of education in the United States as calculated from the GSS are all sample estimates of population parameters. Sample estimates are used to calculate population parameters; the mean number of years of education of 13.43 calculated from the GSS sample can be used to estimate the mean education of all adults in the United States. Similarly, based on a national sample of adult Americans, the Gallup organization estimated the attitudes of Americans toward Arizona's controversial immigration law.

These are all illustrations of estimation. **Estimation** is a process whereby we select a random sample from a population and use a sample statistic to estimate a population parameter. We can use sample proportions as estimates of population proportions, sample means as estimates of population means, or sample variances as estimates of population variances.

Estimation A process whereby we select a random sample from a population and use a sample statistic to estimate a population parameter.

Reasons for Estimation

Why estimate? The goal of most research is to find the population parameter. Yet we hardly ever have enough resources to collect information about the entire population. We rarely know the value of the population parameter. On the other hand, we can learn a lot about a population by randomly selecting a sample from that population and obtaining an estimate of the population parameter. The major objective of sampling theory and statistical inference is to provide estimates of unknown population parameters from sample statistics.

Point and Interval Estimation

Estimates of population characteristics can be divided into two types: point estimates and interval estimates. **Point estimates** are sample statistics used to estimate the exact value of a population parameter. When the Gallup organization reports that 51% of Americans support Arizona's immigration law, they are using a point estimate. Similarly, if we reported the average level of education of the population of adult Americans to be exactly 13.43 years, we would be using a point estimate.

Point estimate A sample statistic used to estimate the exact value of a population parameter.

The problem with point estimates is that sample estimates usually vary, and most result in some sort of sampling error. As a result, when we use a sample statistic to estimate the exact value of a population parameter, we never really know how accurate it is.

One method of increasing accuracy is to use an interval estimate rather than a point estimate. In interval estimation, we identify a range of values within which the population parameter may fall. This range of values is called a **confidence interval (CI)**. Instead of using a single value, 13.43 years, as an estimate of the mean education of adult Americans, we could say that the population mean is somewhere between 12 and 14 years.

Confidence interval (CI) A range of values defined by the confidence level within which the population parameter is estimated to fall. Sometimes confidence intervals are referred to as margin of error.

When we use confidence intervals to estimate population parameters, such as mean educational levels, we can also evaluate the accuracy of this estimate by assessing the likelihood that any given interval will contain the mean. This likelihood, expressed as a percentage or a probability, is called a **confidence level**. Confidence intervals are defined in terms of confidence levels. Thus, by selecting a 95% confidence level, we are saying that there is a .95 probability—or 95 chances out of 100—that a specified interval will contain the population mean. Confidence intervals can be constructed for any level of confidence, but the most common ones are the 90%, 95%, and 99% levels. You should also know that confidence intervals are sometimes referred to in terms of **margin of error**. In short, margin of error is simply the radius of a confidence interval. If we select a 95% confidence level, we would have a margin of error of ±5 percentage points.

Confidence level The likelihood, expressed as a percentage or a probability, that a specified interval will contain the population parameter.

Margin of error The radius of a confidence interval.

Confidence intervals can be constructed for many different parameters based on their corresponding sample statistics. In this chapter, we describe the rationale and the procedure for the construction of confidence intervals for means and proportions.

What is the difference between a point estimate and a confidence interval?

✓ *Learning Check*

PROCEDURES FOR ESTIMATING CONFIDENCE INTERVALS FOR MEANS

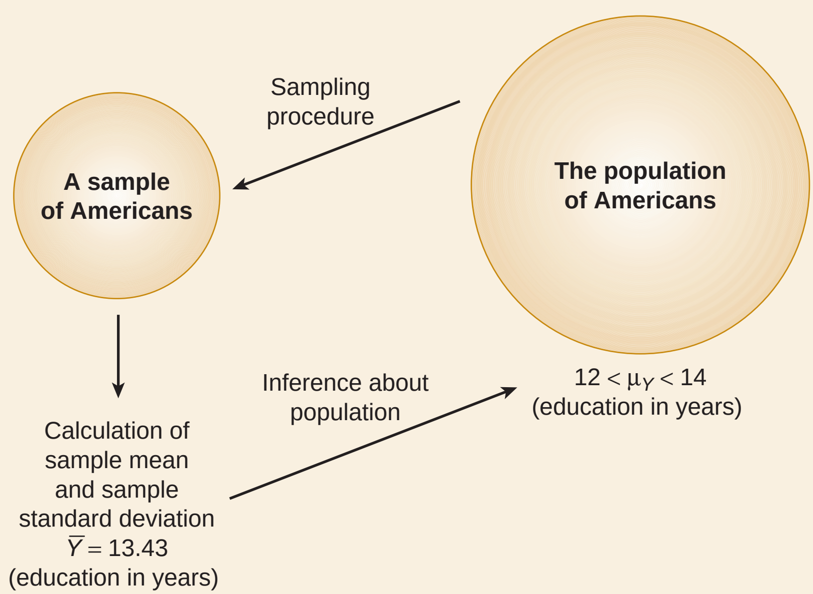
To illustrate the procedure for establishing confidence intervals for means, we'll reintroduce one of the research examples mentioned in Chapter 7—assessing the needs of commuting students on our campus.

Recall that we have been given enough money to survey a random sample of 500 students. One of our tasks is to estimate the average commuting time of all 15,000 commuters on our campus—the population parameter. To obtain this estimate, we calculate the average commuting time for the sample. Suppose the sample average is $\bar{Y} = 7.5$ hr/week, and we want to use it as an estimate of the true average commuting time for the entire population of commuting students.

Because it is based on a sample, this estimate is subject to sampling error. We do not know how close it is to the true population mean. However, based on what the central limit theorem tells us about the properties of the sampling distribution of the mean, we know that with a large enough sample size, most sample means will tend to be close to the true population mean. Therefore, it is unlikely that our sample mean, $\bar{Y} = 7.5$ hr/week, deviates much from the true population mean.

A Closer Look 8.1
Estimation as a Type of Inference

The goal of inferential statistics is to say something meaningful about the population based entirely on information from a sample of that population. A confidence interval attempts to do just that: By knowing a sample mean, sample size, and sample standard deviation, we are able to say something about the population from which that sample was drawn.



We know exactly what our sample mean is. Combining this information with the sample standard deviation and sample size gives us a range within which we can confidently say that the population mean falls.

We know that the sampling distribution of the mean is approximately normal with a mean $\mu_{\bar{Y}}$ equal to the population mean μ_Y and a standard error $\sigma_{\bar{Y}}$ (standard deviation of the sampling distribution) as follows:

$$\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{N}} \tag{8.1}$$

This information allows us to use the normal distribution to determine the probability that a sample mean will fall within a certain distance—measured in standard deviation (standard error) units or *Z* scores—of μ_Y or $\mu_{\bar{Y}}$. We can make the following assumptions:

- A total of 68% of all random sample means will fall within ±1 standard error of the true population mean.
- A total of 95% of all random sample means will fall within ±1.96 standard errors of the true population mean.
- A total of 99% of all random sample means will fall within ±2.58 standard errors of the true population mean.

On the basis of these assumptions and the value of the standard error, we can establish a range of values—a confidence interval—that is likely to contain the actual population mean. We can also evaluate the accuracy of this estimate by assessing the likelihood that this range of values will actually contain the population mean.

The general formula for constructing a confidence interval (CI) for any level is

$$\text{CI} = \bar{Y} \pm Z(\sigma_{\bar{Y}}) \tag{8.2}$$

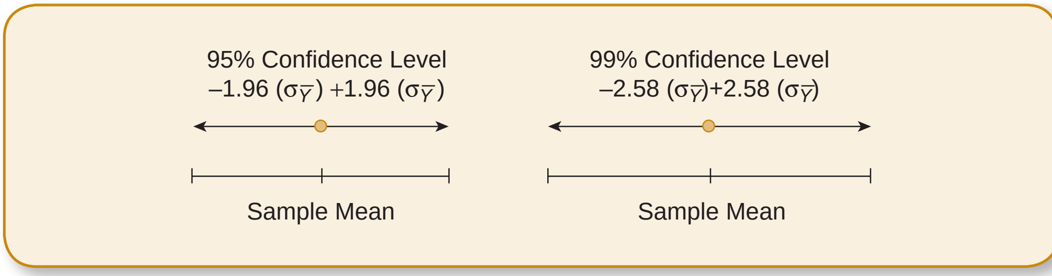
Note that to calculate a confidence interval, we take the sample mean and add to or subtract from it the product of a *Z* value and the standard error.

The *Z* score we choose depends on the desired confidence level. For example, to obtain a 95% confidence interval we would choose a *Z* of 1.96 because we know (from Appendix B) that 95% of the area under the curve lies between ±1.96. Similarly, for a 99% confidence level, we would choose a *Z* of 2.58. The relationship between the confidence level and *Z* is illustrated in Figure 8.1 for the 95% and 99% confidence levels.

To understand the relationship between the confidence level and Z, review the material in Chapter 6. What would be the appropriate Z value for a 98% confidence interval?

✓ **Learning Check**

Figure 8.1 Relationship Between Confidence Level and Z for 95% and 99% Confidence Intervals



Source: From David Freedman, Robert Pisani, Roger Purves, and Ani Akhikari, *Statistics,* 2nd ed. (New York: W. W. Norton, 1991). Copyright ©1991 by W. W. Norton & Company, Inc. Used by permission of W. W. Norton and Company, Inc.

Determining the Confidence Interval

To determine the confidence interval for means, follow these steps:

1. Calculate the standard error of the mean.
2. Decide on the level of confidence, and find the corresponding Z value.
3. Calculate the confidence interval.
4. Interpret the results.

Let's return to the problem of estimating the mean commuting time of the population of students on our campus. How would you find the 95% confidence interval?

Calculating the Standard Error of the Mean

Let's suppose that the standard deviation for our population of commuters is $\sigma_Y = 1.5$. We calculate the standard error for the sampling distribution of the mean:

$$\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{N}} = \frac{1.5}{\sqrt{500}} = 0.07$$

Deciding on the Level of Confidence and Finding the Corresponding Z Value

We decide on a 95% confidence level. The Z value corresponding to a 95% confidence level is 1.96.

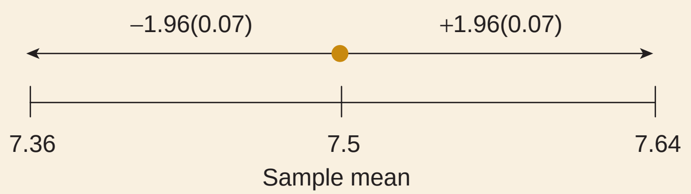
Calculating the Confidence Interval

The confidence interval is calculated by adding and subtracting from the observed sample mean the product of the standard error and Z:

$$
\begin{aligned}
95\%\ \text{CI} &= 7.5 \pm 1.96(0.07) \\
&= 7.5 \pm 0.14 \\
&= 7.36 \text{ to } 7.64
\end{aligned}
$$

The 95% CI for the mean commuting time is illustrated in Figure 8.2.

Figure 8.2 Ninety-Five Percent Confidence Interval for the Mean Commuting Time ($N = 500$)



Interpreting the Results

We can be 95% confident that the actual mean commuting time—the true population mean—is not less than 7.36 hr and not greater than 7.64 hr. In other words, if we collected a large number of samples ($N = 500$) from the population of commuting students, 95 times out of 100, the true population mean would be included within our computed interval. With a 95% confidence level, there is a 5% risk that we are wrong. Five times out of 100, the true population mean will not be included in the specified interval.

Remember that we can never be sure whether the population mean is actually contained within the confidence interval. Once the sample is selected and the confidence interval defined, the confidence interval either does or does not contain the population mean—but we will never be sure.

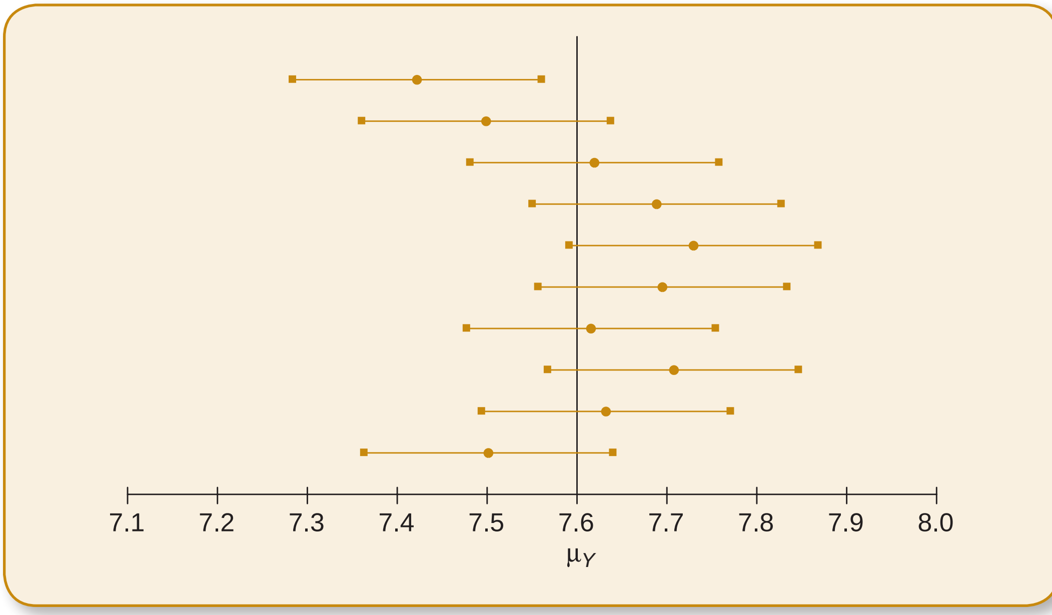
What is the 90% confidence interval for the mean commuting time? (Hint: First, find the Z value associated with a 90% confidence level.)

✓ ***Learning Check***

To further illustrate the concept of confidence intervals, let's suppose that we draw 10 different samples ($N = 500$) from the population of commuting students. For each sample mean, we construct a 95% confidence interval. Figure 8.3 displays these confidence intervals. Each horizontal line represents a 95% confidence interval constructed around a sample mean (marked with a circle).

The vertical line represents the population mean. Note that the horizontal lines that intersect the vertical line are the intervals that contain the true population mean. Only 1 out of the 10 confidence intervals does not intersect the vertical line, meaning it does not contain the population mean. What would happen if we continued to draw samples of the same size from this population and constructed a 95% confidence interval for each sample? For about 95% of all samples the specified interval would contain the true population mean, but for 5% of all samples it would not.

Figure 8.3 Ninety-Five Percent Confidence Intervals for 10 Samples



Reducing Risk

One way to reduce the risk of being incorrect is by increasing the level of confidence. For instance, we can increase our confidence level from 95% to 99%. The 99% confidence interval for our commuting example is

$$\begin{aligned} 99\%\ \text{CI} &= 7.5 \pm 2.58(0.07) \\ &= 7.5 \pm 0.18 \\ &= 7.32 \text{ to } 7.68 \end{aligned}$$

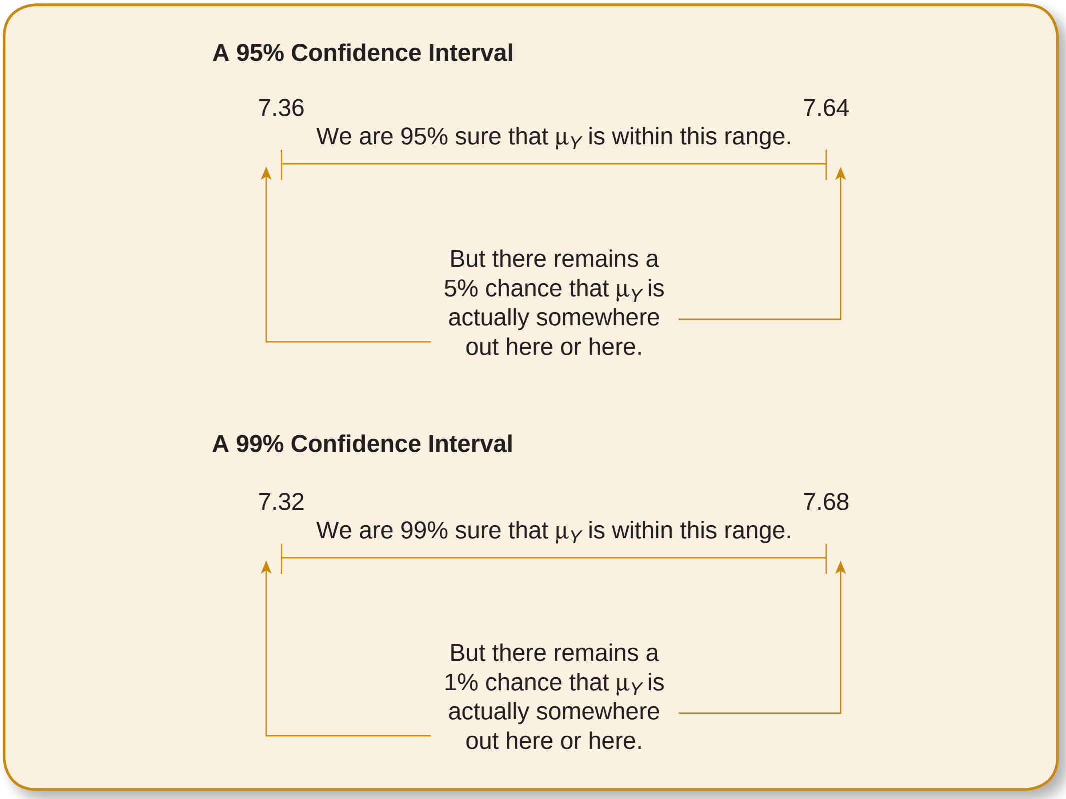
When using the 99% confidence interval, there is only a 1% risk that we are wrong and the specified interval does not contain the true population mean. We can be almost certain that the true population mean is included in the interval ranging from 7.32 to 7.68 hr/week. Note that by increasing the confidence level, we have also increased the width of the confidence interval from 0.28 (7.36–7.64) to 0.36 hr (7.32–7.68), thereby making our estimate less precise.

You can see that there is a trade-off between achieving greater confidence in an estimate and the precision of that estimate. Although using a higher level of confidence (e.g., 99%) increases our confidence that the true population mean is included in our confidence interval, the estimate becomes less precise as the width of the interval increases. Although we are only 95% confident that the interval ranging between 7.36 and 7.64 hr includes the true population mean, it is a more precise estimate than the 99% interval ranging from 7.32 to 7.68 hr. The relationship between the confidence level and the precision of the confidence interval is illustrated in Figure 8.4. Table 8.1 lists three commonly used confidence levels along with their corresponding Z values.

Table 8.1 Confidence Levels and Corresponding Z Values

Confidence Level	*Z Value*
90%	1.65
95%	1.96
99%	2.58

Figure 8.4 Confidence Intervals, 95% Versus 99% (mean commuting time)



Estimating Sigma

To calculate confidence intervals, we need to know the standard error of the sampling distribution, $\sigma_{\bar{Y}}$. The standard error is a function of the population standard deviation and the sample size:

$$\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{N}}$$

In our commuting example, we have been using a hypothetical value, $\sigma_Y = 1.5$, for the population standard deviation. Typically, both the mean (μ_Y) and the standard deviation (σ_Y) of the population are unknown to us. When $N \geq 50$, however, the sample standard deviation, S_Y, is a good estimate of $\sigma_{\bar{Y}}$. The standard error is then calculated as follows:

$$S_{\bar{Y}} = \frac{S_Y}{\sqrt{N}} \tag{8.3}$$

As an example, we'll estimate the mean hours per day that Americans spend watching television based on the 2008 GSS survey. The mean hours per day spent watching television for a sample of $N = 562$ is $\bar{Y} = 2.98\ \text{hr}$, and the standard deviation $S_Y = 2.66$ hr. Let's determine the 95% confidence interval for these data.

Calculating the Estimated Standard Error of the Mean

The estimated standard error for the sampling distribution of the mean is

$$S_{\bar{Y}} = \frac{S_Y}{\sqrt{N}} = \frac{2.66}{\sqrt{562}} = 0.11$$

Deciding on the Level of Confidence and Finding the Corresponding Z Value

We decide on a 95% confidence level. The Z value corresponding to a 95% confidence level is 1.96.

Calculating the Confidence Interval

The confidence interval is calculated by adding to and subtracting from the observed sample mean the product of the standard error and Z:

$$\begin{aligned} 95\%\ \text{CI} &= 2.98 \pm 1.96(0.11) \\ &= 2.98 \pm 0.22 \\ &= 2.76 \text{ to } 3.20 \end{aligned}$$

Interpreting the Results

We can be 95% confident that the actual mean hours spent watching television by Americans from which the GSS sample was taken is not less than 2.76 hr and not greater than 3.20 hr. In other words, if we drew a large number of samples ($N = 562$) from this population, then 95 times out of 100, the true population mean would be included within our computed interval.

Sample Size and Confidence Intervals

Researchers can increase the precision of their estimate by increasing the sample size. In Chapter 7, we learned that larger samples result in smaller standard errors and, therefore, sampling distributions are more clustered around the population mean (Figure 7.6). A more tightly clustered sampling distribution means that our confidence intervals will be narrower and more precise. To illustrate the relationship between sample size and the standard error, and thus the confidence interval, let's calculate the 95% confidence interval for our GSS data with (1) a sample of $N = 195$ and (2) a sample of $N = 1{,}987$.

With a sample size $N = 195$, the estimated standard error for the sampling distribution is

$$S_{\bar{Y}} = \frac{S_Y}{\sqrt{N}} = \frac{2.66}{\sqrt{195}} = 0.19$$

and the 95% confidence interval is

$$\begin{aligned} 95\%\text{ CI} &= 2.98 \pm 1.96(0.19) \\ &= 2.98 \pm 0.37 \\ &= 2.61 \text{ to } 3.35 \end{aligned}$$

With a sample size $N = 1{,}987$, the estimated standard error for the sampling distribution is

$$S_{\bar{Y}} = \frac{S_Y}{\sqrt{N}} = \frac{2.66}{\sqrt{1{,}987}} = 0.06$$

and the 95% confidence interval is

$$\begin{aligned} 95\%\text{ CI} &= 2.98 \pm 1.96(0.06) \\ &= 2.98 \pm 0.12 \\ &= 2.86 \text{ to } 3.10 \end{aligned}$$

In Table 8.2, we summarize the 95% confidence intervals for the mean number of hours watching television for these three sample sizes: $N = 195$, $N = 562$, and $N = 1{,}987$.

Table 8.2 Ninety-Five Percent Confidence Interval and Width for Mean Number of Hours Watching Television for Three Different Sample Sizes

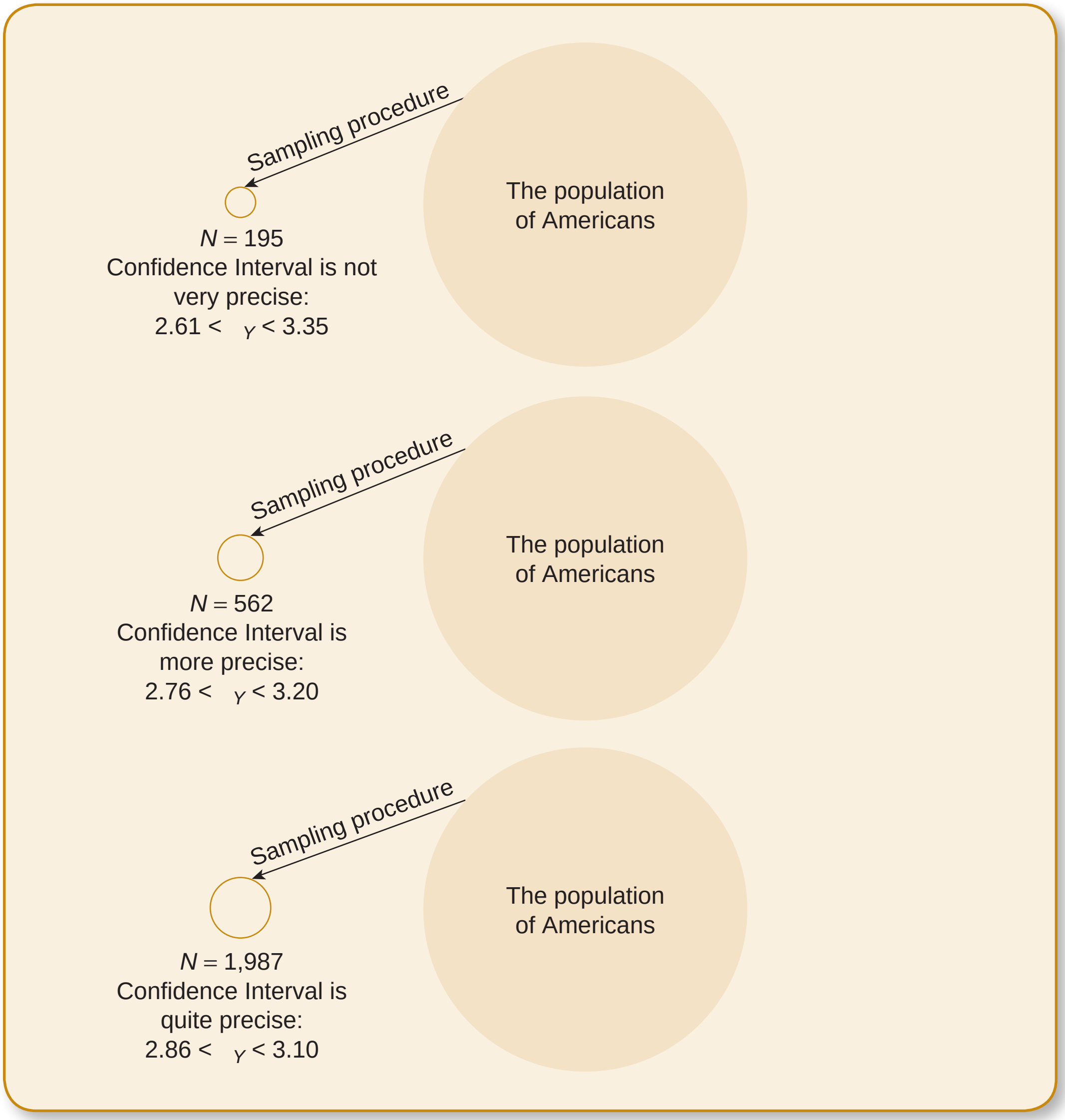
Sample Size (N)	*Confidence Interval*	*Interval Width*	S_Y	$S_{\bar{Y}}$
195	2.61–3.35	0.74	2.66	0.19
562	2.76–3.20	0.44	2.66	0.11
1,987	2.86–3.10	0.24	2.66	0.06

Note that there is an inverse relationship between sample size and the width of the confidence interval. The increase in sample size is linked with increased precision of the confidence interval. The 95% confidence interval for the GSS sample of 195 cases is 0.74 hr. But the interval widths decrease to 0.44 and 0.24 hr, respectively, as the sample sizes increase to $N = 562$ and then to $N = 1{,}987$. We had to nearly quadruple the size of the sample (from 562 to 1,987) to reduce the confidence interval by about one half (from 0.44 to 0.24 hr). In general, although the precision of estimates increases steadily with sample size, the gains would appear to be rather modest after N reaches 1,987. An important factor to keep in mind is the increased cost associated with a larger sample. Researchers have to consider at what point the increase in precision is too small to justify the additional cost associated with a larger sample.

✓ *Learning Check*

Why do smaller sample sizes produce wider confidence intervals? (See Figure 8.5.) (Hint: Compare the standard errors of the mean for the three sample sizes.)

Figure 8.5 The Relationship Between Sample Size and Confidence Interval Width



A Closer Look 8.2
What Affects Confidence Interval Width? Summary

"Holding other factors constant . . ."

If the sample size goes up	$\uparrow$	the confidence interval becomes more precise.	$\rightarrow\ \leftarrow$
If the sample size goes down	$\downarrow$	the confidence interval becomes less precise.	$\leftarrow\ \rightarrow$
If the value of the sample standard deviation goes up	$\uparrow$	the confidence interval becomes less precise.	$\leftarrow\ \rightarrow$
If the value of the sample standard deviation goes down	$\downarrow$	the confidence interval becomes more precise.	$\rightarrow\ \leftarrow$
If the level of confidence goes up (from 95% to 99%)	$\uparrow$	the confidence interval becomes less precise.	$\leftarrow\ \rightarrow$
If the level of confidence goes down (from 99% to 95%)	$\downarrow$	the confidence interval becomes more precise.	$\rightarrow\ \leftarrow$

STATISTICS IN PRACTICE: HISPANIC MIGRATION AND EARNINGS

Tienda and Wilson investigated the relationship between migration and the earnings of Hispanic men.[1] Based on a sample of the 1980 Census, they argued that Mexicans, Puerto Ricans, and Cubans varied markedly in socioeconomic characteristics because of differences in the timing and circumstances of their immigration to the United States. The authors claimed that the period of entry and the circumstances prompting migration affected the geographical distribution and the employment opportunities of each group. For example, Puerto Ricans were disproportionately located in the Northeast, where the labor market was characterized by the highest unemployment rates, whereas the majority of Cuban immigrants resided in the Southeast, where the unemployment rate was the lowest in the United States.

Tienda and Wilson also noted persistent differences in educational levels among Mexicans and Puerto Ricans compared with Cubans. About 60% of Mexicans and Puerto Ricans had not completed high school, compared with 42% of Cuban men. At the other extreme, 17% of Cuban men were college graduates, compared with about 4% of Mexican men and Puerto Rican men.

These differences in migrant status and educational level were likely to be reflected in disparities in earnings among the three groups. Tienda and Wilson anticipated that the earnings of Cubans would be higher than the earnings of Mexicans and Puerto Ricans. We tested these ideas based on a sample from the 2000 Census that included 29,233 Cubans, 34,620 Mexican Americans, and 66,933 Puerto Ricans. As hypothesized, with average earnings of \$24,018 (S_Y = \$36,298), Cubans were at the top of the income hierarchy. Puerto Ricans were intermediate among the groups with earnings averaging \$18,748 (S_Y = \$25,694). Mexican men were at the bottom of the income hierarchy with average annual earnings of \$16,537 (S_Y = \$23,502).

Although Tienda and Wilson did not calculate confidence intervals for their estimates, we will use the updated income data from the 2000 Census[2] to calculate a 95% confidence interval for the mean income of the three groups of Hispanic men.

To find the 95% confidence interval for Cuban income, we first estimate the standard error:

$$S_{\bar{Y}} = \frac{S_Y}{\sqrt{N}} = \frac{36{,}298}{\sqrt{29{,}233}} = 212.29$$

Then, we calculate the confidence interval:

$$\begin{aligned} 95\%\ \text{CI} &= 24{,}018 \pm 1.96(212.29) \\ &= 24{,}018 \pm 416 \\ &= 23{,}602 \text{ to } 24{,}434 \end{aligned}$$

For Puerto Rican income, the estimated standard error is

$$S_{\bar{Y}} = \frac{S_Y}{\sqrt{N}} = \frac{25{,}694}{\sqrt{66{,}933}} = 99.32$$

and the 95% confidence interval is

$$\begin{aligned} 95\%\ \text{CI} &= 18{,}748 \pm 1.96(99.32) \\ &= 18{,}748 \pm 195 \\ &= 18{,}553 \text{ to } 18{,}943 \end{aligned}$$

Finally, for Mexican income, the estimated standard error is

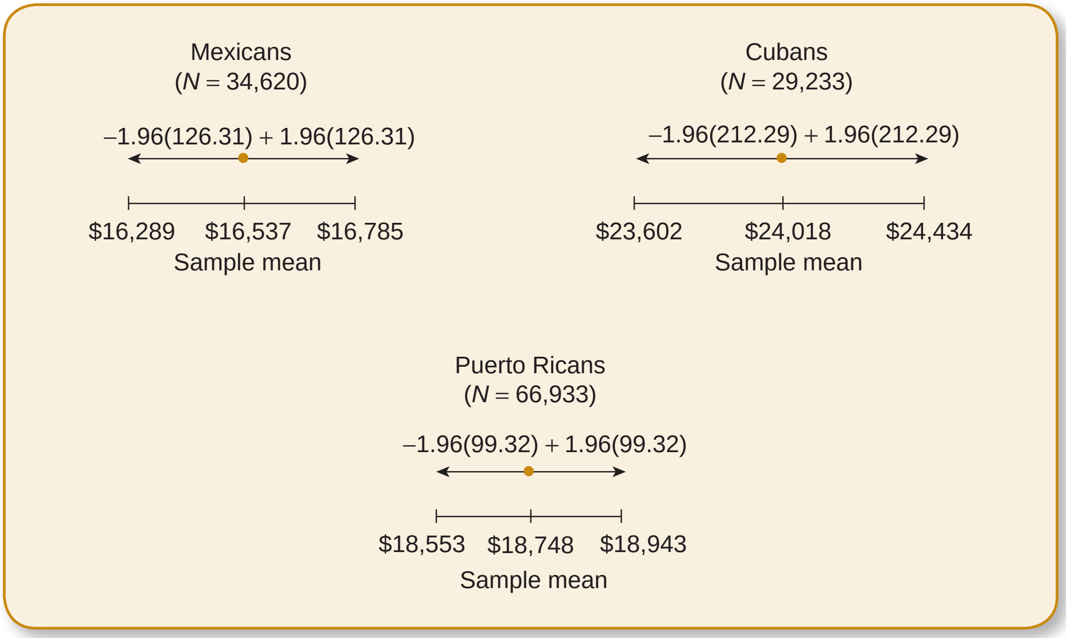
$$S_{\bar{Y}} = \frac{S_Y}{\sqrt{N}} = \frac{23{,}502}{\sqrt{34{,}620}} = 126.31$$

and the 95% confidence interval is

$$\begin{aligned} 95\%\ \text{CI} &= 16{,}537 \pm 1.96(126.31) \\ &= 16{,}537 \pm 248 \\ &= 16{,}289 \text{ to } 16{,}785 \end{aligned}$$

The confidence intervals for mean annual income of Cuban, Puerto Rican, and Mexican immigrants are illustrated in Figure 8.6. We can say with 95% confidence that the true income mean for each Hispanic group lies somewhere within the corresponding confidence interval. Note that the confidence intervals do not overlap, thus revealing great disparities in earnings among the three groups. Highest interval estimates are for Cubans, followed by Mexicans and then Puerto Ricans.

Figure 8.6 Ninety-Five Percent Confidence Intervals for the Mean Income of Puerto Ricans, Mexicans, and Cubans



CONFIDENCE INTERVALS FOR PROPORTIONS

Confidence intervals can also be computed for sample proportions or percentages to estimate population proportions or percentages. The procedures for estimating proportions and percentages are identical. Any of the formulas presented for proportions can be applied to percentages, and vice versa. We can obtain a confidence interval for a percentage by calculating the confidence interval for a proportion and then multiplying the result by 100.

The same conceptual foundations of sampling and statistical inference that are central to the estimation of population means—the selection of random samples and the special properties of the sampling distribution—are also central to the estimation of population proportions.

Earlier, we saw that the sampling distribution of the means underlies the process of estimating population means from sample means. Similarly, the *sampling distribution of proportions* underlies the estimation of population proportions from sample proportions. Based on the central limit theorem, we know that with sufficient sample size the sampling distribution of proportions is approximately normal, with mean μ_p equal to the population proportion π and with a standard error of proportions (the standard deviation of the sampling distribution of proportions) equal to

$$\sigma_p = \sqrt{\frac{(\pi)(1-\pi)}{N}} \tag{8.4}$$

where

σ_p = the standard error of proportions

π = the population proportion

N = the population size

However, since the population proportion, π, is unknown to us (that is what we are trying to estimate), we can use the sample proportion, p, as an estimate of π. The estimated standard error then becomes

$$S_p = \sqrt{\frac{(p)(1-p)}{N}} \tag{8.5}$$

where

S_p = the estimated standard error of proportions

p = the sample proportion

N = the sample size

As an example, let's calculate the estimated standard error for the survey by Gallup. Based on a random sample of 1,013 adults, the percentage who favors the new immigration law introduced by the state of Arizona in 2010 was estimated to be 51%. Based on Formula 8.5, with $p = 0.51$, $1 - p = (1 - 0.51) = 0.49$, and $N = 1{,}013$, the standard error is $S_p = \sqrt{(0.51)(1-0.51)/1{,}013} = 0.016$. We will have to consider two factors to meet the assumption of normality with the sampling distribution of proportions: (1) the sample size N and (2) the sample proportions p and $1 - p$. When p and $1 - p$ are about 0.50, a sample size of at least 50 is sufficient. But when $p > 0.50$ (or $1 - p < 0.50$), a larger sample is required to meet the assumption of normality. Usually, a sample of 100 or more is adequate for any single estimate of a population proportion.

Procedures for Estimating Proportions

Because the sampling distribution of proportions is approximately normal, we can use the normal distribution to establish confidence intervals for proportions in the same manner that we used the normal distribution to establish confidence intervals or means.

The general formula for constructing confidence intervals for proportions for any level of confidence is

$$\text{CI} = p \pm Z(S_p) \tag{8.6}$$

where

CI = the confidence interval
p = the observed sample proportion
Z = the Z corresponding to the confidence level
S_p = the estimated standard error of proportions

Let's examine this formula in more detail. Note that to obtain a confidence interval at a certain level, we take the sample proportion and add to or subtract from it the product of a Z value and the standard error. The Z value we choose depends on the desired confidence level. We want the area between the mean and the selected $\pm Z$ to be equal to the confidence level.

For example, to obtain a 95% confidence interval, we would choose a Z of 1.96 because we know (from Appendix B) that 95% of the area under the curve is included between ±1.96. Similarly, for a 99% confidence level, we would choose a Z of 2.58. (The relationship between confidence level and Z values is illustrated in Figure 8.1.)

To determine the confidence interval for a proportion, we follow the same steps that were used to find confidence intervals for means:

1. Calculate the estimated standard error of the proportion.
2. Decide on the desired level of confidence, and find the corresponding Z value.
3. Calculate the confidence interval.
4. Interpret the results.

To illustrate these steps, we use the results of the Gallup survey on the percentage of Americans who favor the new immigration law introduced by the state of Arizona in 2010.

Calculating the Estimated Standard Error of the Proportion

The standard error of the proportion 0.51 (51%) with a sample $N = 1{,}013$ is 0.016.

Deciding on the Desired Level of Confidence and Finding the Corresponding Z Value

We choose the 95% confidence level. The Z corresponding to a 95% confidence level is 1.96.

Calculating the Confidence Interval

We calculate the confidence interval by adding to and subtracting from the observed sample proportion the product of the standard error and Z:

$$\begin{aligned} 95\%\ \text{CI} &= 0.51 \pm 1.96(0.016) \\ &= 0.51 \pm 0.03 \\ &= 0.48 \text{ to } 0.54 \end{aligned}$$

Interpreting the Results

We are 95% confident that the true population proportion is somewhere between 0.48 and 0.54. In other words, if we drew a large number of samples from the population of adults, then 95 times out of 100, the confidence interval we obtained would contain the true population proportion. We can also express this result in percentages and say that we are 95% confident that the true population percentage of Americans who favor the new immigration law introduced by the state of Arizona in 2010 is included somewhere within our computed interval of 48% to 54%.

✓ *Learning Check*

Calculate the confidence interval for the Gallup survey using percentages rather than proportions. Your results should be identical with ours except that they are expressed in percentages.

Note that with a 95% confidence level, there is a 5% risk that we are wrong. If we continued to draw large samples from this population, in 5 out of 100 samples the true population proportion would not be included in the specified interval.

We can decrease our risk by increasing the confidence level from 95% to 99%.

$$\begin{aligned}99\%\ \text{CI} &= 0.51 \pm 2.58(0.016)\\ &= 0.51 \pm 0.04\\ &= 0.47 \text{ to } 0.55\end{aligned}$$

When using the 99% confidence interval, we can be almost certain (99 times out of 100) that the true population proportion is included in the interval ranging from 0.47 (47%) to 0.55 (55%). However, as we saw earlier, there is a trade-off between achieving greater confidence in making an estimate and the precision of that estimate. Although using a 99% level increased our confidence level from 95% to 99% (thereby reducing our risk of being wrong from 5% to 1%), the estimate became less precise as the width of the interval increased.[3,4]

STATISTICS IN PRACTICE: HEALTH CARE REFORM

Poll or survey results may be limited to a single estimate of a parameter. For instance, political pollsters could estimate the percentage of Americans who approve of the new health care law. Most survey studies, however, are not limited to single estimates for the overall population. Often, separate estimates are reported for subgroups within the overall population of interest. In a report released on April 8–11, 2010, Gallup compared Americans' reaction to the passage of Healthcare Reform Bill by Party Identification. They were interested in exploring whether or not there were differences across groups with different political affiliations.[5]

When estimates are reported for subgroups, the confidence intervals are likely to vary from subgroup to subgroup. Each confidence interval is based on the confidence level, the standard error of the proportion (which can be estimated from p), and the sample size. Even when a confidence interval is reported only for the overall sample, we can easily compute separate confidence intervals for each of the subgroups if the confidence level and the size of each of the subgroups are included.

To illustrate this, let's calculate the 95% confidence intervals for the proportions of Democrats and Republicans who think it is a good thing that Congress passed the new health care legislation. Out of 470 Democrats in the sample, 0.81 (or 81%) indicated support for the legislation. In contrast, of the 459 Republicans surveyed, only 0.10 (or 10%) expressed support for the new health care legislation.

Calculating the Estimated Standard Error of the Proportion

The estimated standard error for the proportion of Democrats is

$$S_p = \sqrt{\frac{(0.81)(1-0.81)}{470}} = 0.02$$

The estimated standard error for the proportion of Republicans is

$$S_p = \sqrt{\frac{(0.10)(1-0.10)}{459}} = 0.01$$

Deciding on the Desired Level of Confidence and Finding the Corresponding *Z* Value

We choose the 95% confidence level, with a corresponding Z value of 1.96.

Calculating the Confidence Interval

For Democrats,

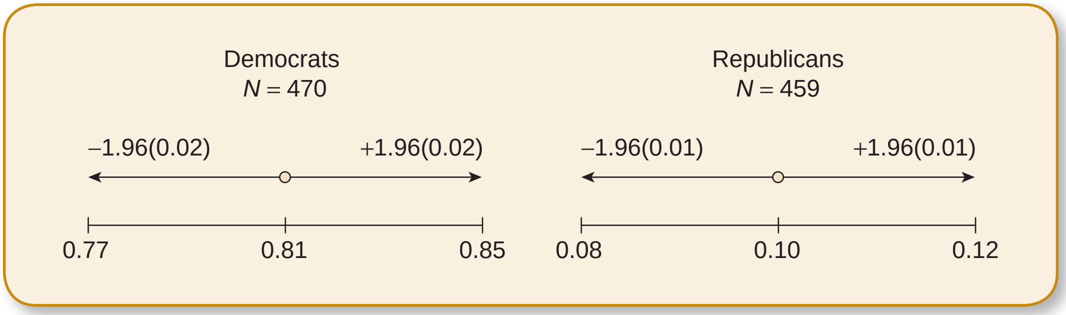
$$\begin{aligned} 95\%\ \text{CI} &= 0.81 \pm 1.96(0.02) \\ &= 0.81 \pm 0.04 \\ &= 0.77 \text{ to } 0.85 \end{aligned}$$

and for Republicans,

$$\begin{aligned} 95\%\ \text{CI} &= 0.10 \pm 1.96(0.01) \\ &= 0.10 \pm 0.02 \\ &= 0.08 \text{ to } 0.12 \end{aligned}$$

The 95% confidence interval for the proportion of Democrats and Republicans surveyed who supported the new health care law is illustrated in Figure 8.7.

Figure 8.7 Ninety-Five Percent Confidence Interval for the Proportion of Democrats and Republicans Who Supported the New Health Care Law



Source: Data from Frank Newport, "Americans' Views on Healthcare Law Remain Stable," *Gallup*, April 15, 2010.

Interpreting the Results

We are 95% confident that the true population proportion supporting the new health care law in April 2010 was between 0.77 and 0.85 (or between 77% and 85%) for Democrats, and somewhere between 0.08 and 0.12 (or between 8% and 12%) for Republicans. Based on the sample, it is clear that among Americans there were partisan differences in support for the new health care law. Democrats were much more likely than Republicans to support the health care reform.

A Closer Look 8.3
A Cautionary Note: The Margin of Error

The most common application of estimation using confidence intervals (also called the margin of error) is demonstrated in opinion and election polls. Pollsters usually interview a random sample representative of a defined population to assess their opinion on a certain issue or their voting preference in a particular election. For example, a 2009 *New York Times* Survey[6] of 708 unemployed adults reported that 69% are more stressed than usual. Also reported was the poll's margin of error of plus or minus 4 percentage points. The margin of error in a poll tells us how well a randomly selected sample represents the population from which it was selected. The *New York Times* poll results indicate that we can be 95% confident that the true percentage of unemployed who feel stressed is somewhere between 65% (69% – 4%) and 73% (69% + 4%).

A margin of error of 4% means that with a sample of 708 individuals, we can be 95% confident that the true value of 69% is within ±4 percentage points of what it would be if the entire unemployed adult population had been polled.

It should be emphasized that the margin of errors only measures the extent of random sampling errors in samples that were randomly selected. Unfortunately, there are several other possible sources of errors in all polls or surveys. They include refusals to be interviewed (nonresponse), question wording, or interviewer bias. Such errors may arise even when the poll's sampling design follows sound methodological principles. Such systematic error may affect the accuracy of polls, yet it is difficult or impossible to quantify the errors that may result from these factors.

A classic example of the failure to predict election results due to systematic and unquantifiable error is the 2008 New Hampshire presidential primaries, when the then Senator Barack Obama ran against Senator Hillary Clinton, in a race where all the published polls predicted victory for Obama, forecasting that he would have a 10 to 15 points lead over Clinton. When the votes were counted, contrary to the polls' outcome, Clinton ended up beating Obama 39.4% to 36.8%.

The error was one of the most significant in modern polling history. Pollsters, social scientists, and sampling experts have examined their data and sampling technique in an attempt to figure out what went wrong.

According to several polling experts,[7] the main problem in the race was the division among voters along socioeconomic lines. Clinton beat Obama by 12 points (47% to 35%) among those with family income below $50,000. In contrast, Obama beat Clinton by five points (40% to 35% among those earning more than $50,000). There was an education gap too. College graduates voted for Obama 39% to 34%. Clinton won among those who had never attended college, 43% to 35%. Another problem was the long-standing pattern in preelection polls for poor white voters to overstate their support for black candidates.[8]

MAIN POINTS

- The goal of most research is to find population parameters. The major objective of sampling theory and statistical inference is to provide estimates of unknown parameters from sample statistics.

- Researchers make point estimates and interval estimates. Point estimates are sample statistics used to estimate the exact value of a population parameter. Interval estimates are ranges of values within which the population parameter may fall.

- Confidence intervals can be used to estimate population parameters such as means or proportions. Their accuracy is defined with the confidence level. The most common confidence levels are 90%, 95%, and 99%.

- To establish a confidence interval for a mean or a proportion, add or subtract from the mean or the proportion the product of the standard error and the Z value corresponding to the confidence level.

KEY TERMS

Confidence interval (interval estimate)
Confidence level
Estimation
Margin of error
Point estimate

ON YOUR OWN

Log on to the web-based student study site at **www.pineforge.com/ssds6e** for additional study questions, quizzes, web resources, and links to social science journal articles reflecting the statistics used in this chapter.

SPSS DEMONSTRATION

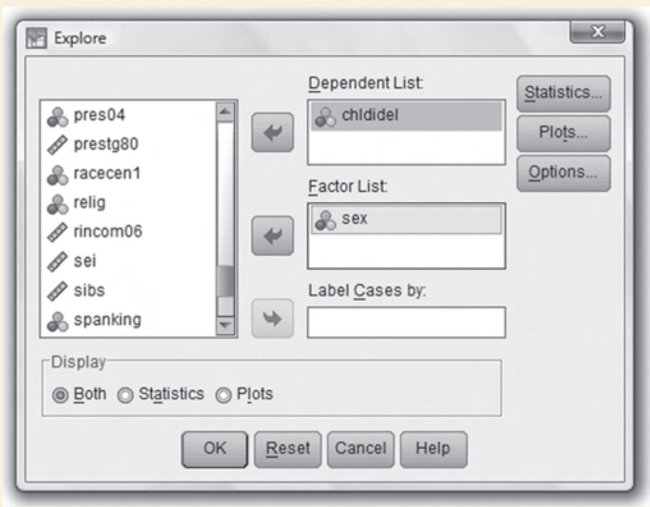
[GSS08PFP-A]

Producing Confidence Intervals Around a Mean

SPSS calculates confidence intervals around a sample mean or proportion with the Explore procedure. Let's investigate the average ideal number of children for men and women.

Activate the Explore procedure by selecting the *Analyze* menu, *Descriptive Statistics*, then *Explore*. The opening dialog box has spaces for both dependent and independent variables. Before running the Explore procedure, we defined the category of the variable CHILDIDEL "as many as want" (coded 98) as missing. To do that, go to the *Data* menu, *Define Variable Properties*, find the variable CHILDIDEL in the variable list, place it in the box "Variables to Scan," and click on *Continue*; in the new dialog box that pops up, select the small square in the *Missing* column corresponding with category "as many as want," coded as 98 (the procedure is not visually represented here)." Place CHLDIDEL in the Dependent List box, and SEX in the Factor List box, as shown in Figure 8.8. (For this analysis, we've eliminated response Categories 8 [as many as I want] and 9 [don't know or no answer].)

Figure 8.8 Explore Dialog Box



Click on the *Statistics* button. Note that the Descriptives choice also includes the confidence interval for the mean, which by default is calculated at the 95% confidence level. Let's change that to the 99% level by erasing the "95" and substituting "99" (Figure 8.9).

Click on *Continue* to return to the main dialog box. Recall that Explore produces several statistics and plots by default. For this example, we don't need to view the graphics, so click on the *Statistics* button in the *Display* section. Your screen should now look like Figure 8.10.

Click on *OK* to run the procedure.

The output from the Explore procedure (Figure 8.11) is divided into two parts, one for males and one for females. The mean ideal number of children for males is 2.40; for females, it's 2.52. Our data indicate that, on average, the ideal number of children is slightly greater for women than it is for men.

Figure 8.9 Setting the Confidence Interval

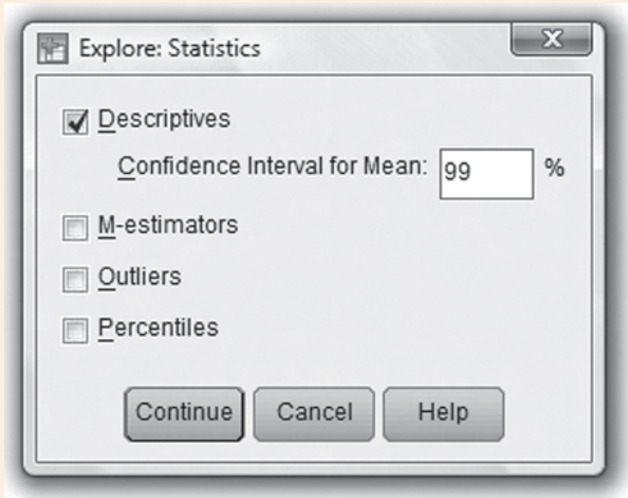
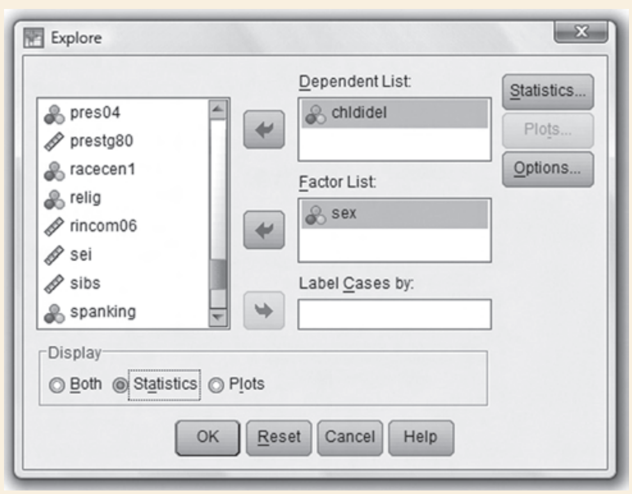


Figure 8.10 Explore Dialog Box



The 99% confidence interval for males runs from about 2.29 to 2.51 children. One way to interpret this result is to state that, in 100 samples each of size 398 males from the U.S. adult population, we would expect the confidence interval to include the true population value for the mean ideal number of children 99 times out of those 100. We can never be sure that in this particular sample the confidence interval includes the

Figure 8.11 SPSS Output for CHLDIDEL by Sex

Case Processing Summary

RESPONDENTS SEX		Cases					
		Valid		Missing		Total	
		N	Percent	N	Percent	N	Percent
IDEAL NUMBER OF CHILDREN	MALE	398	58.1%	287	41.9%	685	100.0%
	FEMALE	454	55.7%	361	44.3%	815	100.0%

Descriptives

RESPONDENTS SEX				Statistic	Std. Error
IDEAL NUMBER OF CHILDREN	MALE	Mean		2.40	.042
		99% Confidence Interval for Mean	Lower Bound	2.29	
			Upper Bound	2.51	
		5% Trimmed Mean		2.38	
		Median		2.00	
		Variance		.719	
		Std. Deviation		.848	
		Minimum		0	
		Maximum		7	
		Range		7	
		Interquartile Range		1	
		Skewness		.887	.122
		Kurtosis		3.720	.244
	FEMALE	Mean		2.52	.040
		99% Confidence Interval for Mean	Lower Bound	2.42	
			Upper Bound	2.63	
		5% Trimmed Mean		2.50	
		Median		2.00	
		Variance		.731	
		Std. Deviation		.855	
		Minimum		0	
		Maximum		7	
		Range		7	
		Interquartile Range		1	
		Skewness		.744	.115
		Kurtosis		2.502	.229

population mean. As explained in this chapter, the confidence interval of any one sample either does or does not contain the (unknown) population mean, so no probability value can be associated with a particular confidence interval. Still, our best estimate for the mean ideal number of children falls within a narrow range of only about 0.22. For females, the 99% confidence interval is almost the same, varying from about 2.42 to 2.63 children, or only 0.21.

SPSS PROBLEMS

[GSS08PFP-A]

1. Recall that the GSS sample includes men and women from 18 to 89 years of age. Does it matter that we may have responses from men and women of diverse ages? Would our results change if we selected a younger sample of men and women?
   a. To take the SPSS demonstration one step further, use the Select Cases procedure to select respondents based on the variable AGE who are less than or equal to 35 years old. Do this by selecting *Data* and then *Select Cases*. Next, select *If Condition is satisfied* and then click on *If*. Find and highlight the variable AGE in the scroll-down box on the left of your screen. Click the arrow next to the scroll-down box. AGE will now appear in the box on the right. Now, tell SPSS that you want to select respondents who are 35 years of age or less. The box on the right should now read AGE $\leq$ 35. Click *Continue* and then *OK*.
   b. Using this younger sample, repeat the Explore procedure that we just completed in the demonstration. What differences exist between men and women in this younger sample on the ideal number of children? How do these results compare with those based on the entire sample?

2. Calculate the 90% confidence interval for the following variables, comparing lower, working, middle, and upper classes (CLASS) in the GSS sample. First, tell SPSS that we want to select all cases in the sample by selecting *Data*, *Select Cases*, and then *All Cases*, and then *OK*. Then, use the Explore procedure using CLASS as your factor variable (*Analyze*, *Descriptive Statistics*, *Explore*). Make a summary statement of your findings.
   a. CHILDS (Number of children in the household)
   b. EDUC (Respondent's highest year of school completed)
   c. PAEDUC (Father's highest year of school completed)
   d. PRESTG80 (Respondent's occupational prestige)
   e. MAEDUC (Mother's highest year of school completed)

CHAPTER EXERCISES

1. In a study of crime, the FBI found that 13.2% of all Americans had been victims of crime during a 1-year period. This result was based on a sample of 1,105 adults.
   a. Estimate the percentage of U.S. adults who were victims at the 90% confidence level. State in words the meaning of the result.
   b. Estimate the percentage of victims at the 99% confidence level.
   c. Imagine that the FBI doubles the sample size in a new sample but finds the same value of 13.2% for the percentage of victims in the second sample. By how much would the 90% confidence interval shrink? By how much would the 99% confidence interval shrink?
   d. Considering your answers to (a), (b), and (c), can you suggest why national surveys, such as those by Gallup, Roper, or *The New York Times*, typically take samples of size 1,000 to 1,500?

2. Use the data on education from Chapter 6, Exercise 5.

	Mean	*Standard Deviation*	N
Lower class	11.36	2.96	121
Working class	12.73	2.79	676
Middle class	14.40	3.04	636
Upper class	15.49	2.95	53

a. Construct the 95% confidence interval for the mean number of years of education for lower-class and middle-class respondents.
b. Construct the 99% confidence interval for the mean number of years of education for lower-class and middle-class respondents.
c. As our confidence in the result increases, how does the size of the confidence interval change? Explain why this is so.

3. There has been a great deal of discussion about global warming in recent years. In 2006,[9] the Pew Research Center conducted a survey of 1,501 Americans to assess their opinion of global warming. The data show that 615 respondents of the 1,501 surveyed felt global warming is a very serious problem.
   a. Estimate the proportion of all adult Americans who felt global warming is a very serious problem at the 95% confidence interval.
   b. Estimate the proportion of all adult Americans who felt global warming is a very serious problem at the 99% confidence interval.
   c. If you were going to write a report on this poll result, would you prefer to use the 99% or 95% confidence interval? Explain why.

4. Use the data in Chapter 5, Exercise 6, about occupational prestige and education.

PRESTG80			*Statistic*
RS OCCUPATIONAL PRESTIGE SCORE (1980)	High school diploma	Mean	40.41
		Median	39.00
		Standard deviation	11.756
		Minimum	17
		Maximum	75
		Range	58
		Interquartile range	18
	Bachelor's degree	Mean	51.93
		Median	51.00
		Standard deviation	12.934
		Minimum	17
		Maximum	86
		Range	69
		Interquartile range	21

a. Construct the 90% confidence interval for occupational prestige for respondents with only a high school diploma ($N = 706$).
b. Construct the 90% confidence interval for occupational prestige for respondents with a bachelor's degree ($N = 254$). State in words the meaning of the result.
c. Use these statistics to discuss differences in occupational prestige scores by educational attainment.

5. Gallup conducted a survey[10] in April 1 to 25, 2010, to determine the congressional vote preference of the American voters. They found that 51% of the male voters preferred a Republican candidate over a Democratic candidate in a sample of 5,490 registered voters. Gallup asks you, their statistical consultant, to tell them whether you could declare the Republican candidate as the likely winner of the votes coming from men if there was an election today. What is your advice? Why?

6. You have been doing research for your statistics class on how nervous the American adults are in general. You have decided to use HINTS 2007 data set that has a scale (going from 0 to 24) measuring the psychological distress of the respondents.
   a. According to HINTS 2007 data, the average psychological distress score, for this sample of size 1,390, is 4.75, with a standard deviation of 4.53. Construct the 95% confidence interval for the true average psychological distress score.
   b. One of your classmates, who claims to be good at statistics, complains about your confidence interval calculation. She or he asserts that the psychological distress scores are not normally distributed, which in turn makes the confidence interval calculation meaningless. Assume that she or he is correct about the distribution of psychological distress scores. Does that imply that the calculation of a confidence interval is not appropriate? Why or why not?

7. From the 2008 GSS subsample, we find that 78.6% of respondents believe in some form of life after death ($N = 1{,}303$).
   a. What is the 95% confidence interval for the percentage of the U.S. population who believe in life after death?
   b. Without doing any calculations, make an educated guess at the lower and upper bounds of 90% and 99% confidence intervals.

8. A social service agency plans to conduct a survey to determine the mean income of its clients. The director of the agency prefers that you measure the mean income very accurately, to within ±\$500. From a sample taken 2 years ago, you estimate that the standard deviation of income for this population is about \$5,000. Your job is to figure out the necessary sample size to reduce sampling error to ±\$500.
   a. Do you need to have an estimate of the current mean income to answer this question? Why or why not?
   b. What sample size should be drawn to meet the director's requirement at the 95% level of confidence? (*Hint:* Use the formula for a confidence interval and solve for N, the sample size.)
   c. What sample size should be drawn to meet the director's requirement at the 99% level of confidence?

9. Data from a 2008 GSS subsample show that the mean number of children per respondent was 1.94, with a standard deviation of 1.70. A total of 2,020 people answered this question. Estimate the population mean number of children per adult using a 90% confidence interval.

10. A subsample of the 2008 MTF survey suggests that adolescents are generally concerned with social issues. In fact, 79.9% of the 1,488 respondents who answered the question reported that they either sometimes or often/all the time think about social issues. Estimate at the 95% and 99% confidence levels the proportion of all adolescents who sometimes or often think about social issues.

11. According to a 2008 survey[11] by the Pew Research Center, two out of three Americans aged between 18 and 29 years use social networking websites such as MySpace and Facebook. Interestingly, 27% of the two hundred and twenty-five 18- to 29-year-olds surveyed say that they have received information about the 2008 presidential political candidates from social networking websites. What is the 95% confidence interval for the percentage of Americans aged between 18 and 29 years that use social networking websites to obtain information about political candidates?

12. According to a report[12] published by the Pew Research Center in February 2010, 61% of Millenials (Americans in their teens and 20s) think that their generation has a unique and distinctive identity ($N = 527$).
    a. Calculate the 95% confidence interval to estimate the percentage of Millenials who believe that their generation has a distinctive identity as compared with the other generations (the Generation X, Baby Boomers, or the Silent Generation).
    b. Calculate the 99% confidence interval.
    c. Are both these results compatible with the conclusion that the majority of Millenials believe that they have a unique identity that separates them from the previous generations?

13. Whether one views homosexual relations as wrong is closely related to whether one views homosexuality as a biological trait or the outcome of one's environment and/or socialization. Thus, it is not surprising that several religious groups that condemn homosexual relations have proclaimed their ability to "cure" gays of their sexual orientation. After all, their assumption is that homosexuality is not a trait that a person is born with. In 2008, GSS respondents ($N = 1{,}269$) were asked what they thought about homosexual relations. The data show that 51.5% believed that homosexual relations were always wrong, while 38.1% believed that homosexual relations were not wrong at all.
    a. For each reported percentage, calculate the 95% confidence interval.
    b. About 10% of GSS respondents were in the middle, some saying that homosexual relations were almost always wrong or sometimes wrong. Calculate the 95% confidence interval.
    c. What conclusions can you draw about the public's opinions of homosexual behavior based on your calculations?

14. A subsample of the 2008 MTF survey suggests that black adolescents are more accepting of having friends of a different race than white adolescents. Of the 184 black adolescents who answered the question about race and friendship, 44% reported that having friends of a different race is "desirable." However, of the 867 white adolescents who answered the question about race and friendship, only 39% reported that having friends of a different race is "desirable."
    a. Calculate the 95% confidence interval for each proportion reported.
    b. Write a summary based on your calculation suggesting at least two reasons for the observed difference.

15. Many women are raising their children while also working outside the home. Many have studied the impact this trend has on heterosexual families.
    a. How do you think Americans feel about this trend? Do you think there are differences among certain subgroups of the population? If yes, which subgroup(s) do you believe are most likely to view this trend favorably? If no, explain why you do not think there are differences among subgroups of the population.
    b. Researchers collecting data for the 2008 GSS offer insight into public attitudes toward this issue. They asked respondents to share their opinion about working mothers. Of the 623 male respondents who answered the question, 17% strongly agreed that a working mother does not hurt children. Construct a 90% confidence interval for this statistic.
    c. Of the 697 female respondents who answered the question, 35.6% strongly agreed that a working mother does not hurt children. Construct a 90% confidence interval for this statistic.
    d. Offer an explanation for why there is a difference between men and women on this issue.