According to economist Ethan Harris, “People may not remember too many numbers about the economy, but there are certain signposts they do pay attention to. As a short hard way to assess how the economy is doing, everybody notices the price of gas.”

The impact of high and volatile fuel prices is felt across the nation, affecting consumer spending and the economy, but the burden remains greater among distinct social economic groups and geographic areas. Lower-income Americans spend eight times more of their disposable income on gasoline than wealthier Americans do. For example, in Wilcox, Alabama, individuals spend 12.72% of their income to fuel one vehicle, while in Hunterdon Co., New Jersey, people spend 1.52%. Nationally, Americans spend 3.8% of their income fueling one vehicle.

The first state to reach the $5 per gallon milestone was California in 2012. California’s drivers were especially hit hard by the rising price of gas, due in part to their reliance on automobiles, especially for work commuters. In 2016, gasoline prices remained higher for states along the West Coast, particularly in Alaska, Hawaii, and California.

Let’s say we drew a random sample of California gas stations ($N = 100$) and calculated the mean price for a gallon of regular gas, $\bar{p} = 2.78$. Based on consumer information, we also know that nationally the mean price of a gallon was $p = 2.32$ with a standard deviation of $\sigma = 0.17$ for the same week. We can thus compare the mean price of gas in California with the mean price of all gas stations in May 2016. By comparing these means, we are asking whether it is reasonable to consider our random sample of California gas as representative of the population of gas stations in the United States. Actually, we expect to find that the
average price of gas from a sample of California gas stations will be unrepresentative of the population of gas stations because we assume higher gas prices in California.

The sample mean of $2.78 is higher than the population mean of $2.32, but it is an estimate based on a single sample. Thus, it could mean one of two things: (1) the average price of gas in California is indeed higher than the national average or (2) the average price of gas in California is about the same as the national average, and this sample happens to show a particularly high mean.

How can we decide which explanation makes more sense? Because most estimates are based on single samples and different samples may result in different estimates, sampling results cannot be used directly to make statements about a population. We need a procedure that allows us to evaluate hypotheses about population parameters based on sample statistics. In Chapter 7 (“Estimation”), we saw that population parameters can be estimated from sample statistics. In this chapter, we will learn how to use sample statistics to make decisions about population parameters. This procedure is called statistical hypothesis testing.

ASSUMPTIONS OF STATISTICAL HYPOTHESIS TESTING

Statistical hypothesis testing requires several assumptions. These assumptions include considerations of the level of measurement of the variable, the method of sampling, the shape of the population distribution, and the sample size. The specific assumptions may vary, depending on the test or the conditions of testing. However, without exception, all statistical tests assume random sampling. Tests of hypotheses about means also assume interval-ratio level of measurement and require that the population under consideration be normally distributed or that the sample size be larger than 50.

Based on our data, we can test the hypothesis that the average price of gas in California is higher than the average national price of gas. The test we are considering meets these conditions:

1. The sample of California gas stations was randomly selected.
2. The variable *price per gallon* is measured at the interval-ratio level.
3. We cannot assume that the population is normally distributed. However, because our sample size is sufficiently large (*N* > 50), we know, based on the central limit theorem, that the sampling distribution of the mean will be approximately normal.

STATING THE RESEARCH AND NULL HYPOTHESES

Hypotheses are usually defined in terms of interrelations between variables and are often based on a substantive theory. Earlier, we defined hypotheses as tentative answers to
research questions. They are tentative because we can find evidence for them only after being empirically tested. The testing of hypotheses is an important step in this evidence-gathering process.

The Research Hypothesis ($H_1$)

Our first step is to formally express the hypothesis in a way that makes it amenable to a statistical test. The substantive hypothesis is called the research hypothesis and is symbolized as $H_1$. Research hypotheses are always expressed in terms of population parameters because we are interested in making statements about population parameters based on our sample statistics.

In our research hypothesis ($H_1$), we state that the average price of gas in California is higher than the average price of gas nationally. Symbolically, we use $\mu$ to represent the population mean; our hypothesis can be expressed as

$$H_1: \mu > \$2.32$$

In general, the research hypothesis ($H_1$) specifies that the population parameter is one of the following:

1. Not equal to some specified value: $\mu \neq$ some specified value
2. Greater than some specified value: $\mu >$ some specified value
3. Less than some specified value: $\mu <$ some specified value

In a one-tailed test, the research hypothesis is directional; that is, it specifies that a population mean is either less than ($<$) or greater than ($>$) some specified value. We can express our research hypothesis as either

$$H_1: \mu < \text{some specified value}$$

or

$$H_1: \mu > \text{some specified value}$$

The research hypothesis we’ve stated for the average price of a gallon of regular gas in California is a one-tailed test.

When a one-tailed test specifies that the population mean is greater than some specified value, we call it a right-tailed test because we will evaluate the outcome at the right tail of the sampling distribution. If the research hypothesis specifies that the population mean is less than some specified value, it is called a left-tailed test because the outcome will be evaluated at the left tail of the sampling distribution. Our example is a right-tailed test because the research hypothesis states that the mean gas prices in California are higher than $\$2.32$. (Refer to Figure 8.1 on page 204.)

Sometimes, we have some theoretical basis to believe that there is a difference between groups, but we cannot anticipate the direction of that difference. For example, we may...
have reason to believe that the average price of California gas is different from that of the general population, but we may not have enough research or support to predict whether it is higher or lower. When we have no theoretical reason for specifying a direction in the research hypothesis, we conduct a two-tailed test. The research hypothesis specifies that the population mean is not equal to some specified value. For example, we can express the research hypothesis about the mean price of gas as

\[ H_1: \mu \neq 2.32 \]

With both one- and two-tailed tests, our null hypothesis of no difference remains the same. It can be expressed as

\[ H_0: \mu = \text{some specified value} \]

### The Null Hypothesis (\( H_0 \))

Is it possible that in the population there is no real difference between the mean price of gas in California and the mean price of gas in the nation and that the observed difference of 0.46 is actually due to the fact that this particular sample happened to contain California gas stations with higher prices? Since statistical inference is based on probability theory, it is not possible to prove or disprove the research hypothesis directly. We can, at best, estimate the likelihood that it is true or false.
To assess this likelihood, statisticians set up a hypothesis that is counter to the research hypothesis. The null hypothesis, symbolized as $H_0$, contradicts the research hypothesis and states that there is no difference between the population mean and some specified value. It is also referred to as the hypothesis of “no difference.” Our null hypothesis can be stated symbolically as

$$H_0: \mu = 2.32$$

Rather than directly testing the substantive hypothesis ($H_1$) that there is a difference between the mean price of gas in California and the mean price nationally, we test the null hypothesis ($H_0$) that there is no difference in prices. In hypothesis testing, we hope to reject the null hypothesis to provide indirect support for the research hypothesis.

**PROBABILITY VALUES AND ALPHA**

Now let’s put all our information together. We’re assuming that our null hypothesis ($\mu = 2.32$) is true, and we want to determine whether our sample evidence casts doubt on that assumption, suggesting that there is evidence for research hypothesis, $\mu > 2.32$. What are the chances that we would have randomly selected a sample of California gas stations such that the average price per gallon is higher than $2.32$, the average for the nation? We can determine the chances or probability because of what we know about the sampling distribution and its properties. We know, based on the central limit theorem, that if our sample size is larger than 50, the sampling distribution of the mean is approximately normal, with a mean and a standard deviation (standard error) of

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}$$

We are going to assume that the null hypothesis is true and then see if our sample evidence casts doubt on that assumption. We have a population mean $\mu = 2.32$ and a standard deviation $\sigma = 0.17$. Our sample size is $N = 100$, and the sample mean is $2.78$. We can assume that the distribution of means of all possible samples of size $N = 100$ drawn from this distribution would be approximately normal, with a mean of $2.32$ and a standard deviation of

$$\sigma_{\bar{X}} = \frac{0.17}{\sqrt{100}} = 0.02$$

This sampling distribution is shown in Figure 8.1. Also shown in Figure 8.1 is the mean gas price we observed for our sample of California gas stations.

Because this distribution of sample means is normal, we can use Appendix B to determine the probability of drawing a sample mean of $2.78$ or higher from this population. We will translate our sample mean into a $Z$ score so that we can determine its location relative to the population mean. In Chapter 5 (“The Normal Distribution”), we learned how to translate a raw score into a $Z$ score by using Formula 5.1:

$$Z = \frac{Y - \bar{Y}}{s}$$
Because we are dealing with a sampling distribution in which our raw score is \( \bar{Y} \) (the mean), and the standard deviation (standard error) is \( \sigma / \sqrt{N} \), we need to modify the formula somewhat:

\[
Z = \frac{\bar{Y} - \mu_r}{\sigma / \sqrt{N}}
\]

Converting the sample mean to a Z-score equivalent is called computing the test statistic. The Z value we obtain is called the Z statistic (obtained). The obtained Z gives us the number of standard deviations (standard errors) that our sample is from the hypothesized value (\( \mu \) or \( \mu_r \)), assuming the null hypothesis is true. For our example, the obtained Z is

\[
Z = \frac{2.78 - 2.32}{0.17 / \sqrt{100}} = \frac{0.46}{0.02} = 23.00
\]

Before we determine the probability of our obtained Z statistic, let’s determine whether it is consistent with our research hypothesis. Recall that we defined our research hypothesis as a right-tailed test (\( \mu > \mu_r \)), predicting that the difference would be assessed on the right tail of the sampling distribution. The positive value of our obtained Z statistic confirms that we will be evaluating the difference on the right tail. (If we had a negative obtained Z, it would mean the difference would have to be evaluated at the left tail of the distribution, contrary to our research hypothesis.)

To determine the probability of observing a Z value of 23.00, assuming that the null hypothesis is true, look up the value in Appendix B to find the area to the right of (above) the Z of 23.00. Our calculated Z value is not listed in Appendix B, so we’ll need to rely on the last Z value reported in the table, 4.00. Recall from Chapter 5, where we calculated Z scores and their probability, that the Z values are located in Column A. The \( p \) value is the probability to the right of the obtained Z, or the “area beyond Z” in Column C. This area includes the proportion of all sample means that are \( \geq 2.78 \) or higher. The proportion is less than .0001 (Figure 8.2). This value is the probability of getting a result as extreme as the sample result if the null hypothesis is true; it is symbolized as \( p \). Thus, for our example, \( p \leq .0001 \).

A \( p \) value can be defined as the probability associated with the obtained value of Z. It is a measure of how unusual or rare our obtained statistic is compared with what is stated in our null hypothesis. The smaller the \( p \) value, the more evidence we have that the null hypothesis should be rejected in favor of the research hypothesis. The larger the \( p \) value, we can assume that the null hypothesis is true and fail to reject it. Based on the \( p \) value, we can also make a statement regarding the significance of the results. A result is deemed “statistically significant” if the probability is less than or equal to the alpha level.

Researchers usually define in advance what a sufficiently improbable Z value is by specifying a cutoff point below which \( p \) must fall to reject the null hypothesis. This cutoff point, called alpha and denoted by the Greek letter \( \alpha \), is customarily set at the .05, .01, or .001 level. Let’s say that we decide to reject the null hypothesis if \( p \leq .05 \). The
value .05 is referred to as alpha (\( \alpha \)); it defines for us what result is sufficiently improbable to allow us to take the risk and reject the null hypothesis. An alpha (\( \alpha \)) of .05 means that even if the obtained Z statistic is due to sampling error, so that the null hypothesis is true, we would allow a 5% risk of rejecting it. Alpha values of .01 and .001 are more cautionary levels of risk. The difference between \( p \) and alpha is that \( p \) is the actual probability associated with the obtained value of \( Z \), whereas alpha is the level of probability determined in advance at which the null hypothesis is rejected. The null hypothesis is rejected when \( p \leq \alpha \).

We have already determined that our obtained \( Z \) has a probability value less than .0001. Since our observed \( p \) is less than .05 (\( p = .0001 < \alpha = .05 \)), we reject the null hypothesis. The value of .0001 means that fewer than 1 out of 10,000 samples drawn from this population are likely to have a mean that is 23.00 \( Z \) scores above the hypothesized mean of $2.78. Another way to say it is as follows: There is only 1 chance out of 10,000 (or .0001%) that we would draw a random sample with a \( Z \geq 23.00 \) if the mean price of California gas were equal to the national mean price. We can state that the difference between the average price of gas in California and nationally is statistically significant at the .05 level, or specify the level of significance by saying that the level of significance is less than .0001. For more about significance, refer to A Closer Look 8.1.

Recall that our hypothesis was a one-tailed test (\( \mu > $2.32 \)). In a two-tailed test, sample outcomes may be located at both the higher and the lower ends of the sampling distribution. Thus, the null hypothesis will be rejected if our sample outcome falls either at the left or right tail of the sampling distribution. For instance, a .05 alpha or \( p \) level means that \( H_0 \) will be rejected if our sample outcome falls among either the lowest or the highest 5% of the sampling distribution.
MORE ABOUT SIGNIFICANCE

Just because a relationship between two variables is statistically significant does not mean that the relationship is important theoretically or practically. Recall that we are relying on information from a sample to infer characteristics about the population. If you decide to reject the null hypothesis, you must still determine what inferences you can make about the population. Ronald Wasserstein and Nicole Lazar (2016) advise, “Researchers should bring many contextual factors into play to derive scientific inferences, including the design of a study, the quality of the measurements, the external evidence for the phenomenon under study, and the validity of assumptions that underlies the data analysis.” Indeed, determining significance is just one part of the research process.

The application of hypothesis testing and significance presented in this text reflects how our discipline currently utilizes and reports hypothesis testing. Yet scholars and statisticians have expressed concern about reducing scientific inquiry to the pursuit of single measure; that is to say, the only result that matters is when \( p < .05 \) or some arbitrary level of significance. According to demographer Jan Hoem (2008), “The scientific importance of an empirical finding depends much more on its contribution to the development or falsification of a substantive theory than on the values of indicators of statistical significance.”

Many have argued how hypothesis testing is problematic because it fails to provide definitive evidence about the existence of real relationships in the data. Despite these criticisms, hypothesis testing remains the primary model by which we derive statistical inference. Several academic journals have adopted new standards for data (e.g., eliminating \( p \) values, reporting nonsignificant findings along with significant ones), in hopes of improving the quality and integrity of research.

Suppose we had expressed our research hypothesis about the mean price of gas as

\[
H_1: \mu \neq \$2.32
\]

The null hypothesis to be directly tested still takes the form \( H_0: \mu = \$2.32 \) and our obtained \( Z \) is calculated using the same formula (Formula 8.1) as was used with a one-tailed test. To find \( p \) for a two-tailed test, look up the area in Column C of Appendix B that corresponds to your obtained \( Z \) (as we did earlier) and then multiply it by 2 to obtain the two-tailed probability. Thus, the two-tailed \( p \) value for \( Z = 23.00 \) is \( .0001 \times 2 = .0002 \). This probability is less than our stated alpha (.05), and thus, we reject the null hypothesis.

THE FIVE STEPS IN HYPOTHESIS TESTING: A SUMMARY

Statistical hypothesis testing can be organized into five basic steps. Let’s summarize these steps:

1. Making assumptions
2. Stating the research and null hypotheses and selecting alpha
3. Selecting the sampling distribution and specifying the test statistic
4. Computing the test statistic
5. Making a decision and interpreting the results

1. Making Assumptions: Statistical hypothesis testing involves making several assumptions regarding the level of measurement of the variable, the method of sampling, the shape of the population distribution, and the sample size. In our example, we made the following assumptions:

   - A random sample was used.
   - The variable price per gallon is measured on an interval-ratio level of measurement.
   - Because \( N > 50 \), the assumption of normal population is not required.

2. Stating the Research and Null Hypotheses and Selecting Alpha: The substantive hypothesis is called the research hypothesis and is symbolized as \( H_1 \). Research hypotheses are always expressed in terms of population parameters because we are interested in making statements about population parameters based on sample statistics. Our research hypothesis was

   \[
   H_1: \mu > $2.32
   \]

   The null hypothesis, symbolized as \( H_0 \), contradicts the research hypothesis in a statement of no difference between the population mean and our hypothesized value. For our example, the null hypothesis was stated symbolically as

   \[
   H_0: \mu = $2.32
   \]

   We set alpha at \(.05\), meaning that we would reject the null hypothesis if the probability of our obtained \( Z \) was less than or equal to \(.05\).

3. Selecting the Sampling Distribution and Specifying the Test Statistic: The normal distribution and the \( Z \) statistic are used to test the null hypothesis.

4. Computing the Test Statistic: Based on Formula 8.1, our \( Z \) statistic is \(23.00\).

5. Making a Decision and Interpreting the Results: We confirm that our obtained \( Z \) is on the right tail of the distribution, consistent with our research hypothesis. We determine that the \( p \) value of \(23.00\) is less than \(.0001\), less than our \(.05\) alpha level. We have evidence to reject the null hypothesis of no difference between the mean price of California gas and the mean price of gas nationally. Based on these data, we conclude that the average price of California gas is significantly higher than the national average.

**ERRORS IN HYPOTHESIS TESTING**

We should emphasize that because our conclusion is based on sample data, we will never really know if the null hypothesis is true or false. In fact, as we have seen, there
Table 8.1  Type I and Type II Errors

<table>
<thead>
<tr>
<th>Decision Made</th>
<th>True State of Affairs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( H_0 ) is True</td>
<td>( H_0 ) is False</td>
</tr>
<tr>
<td>Reject ( H_0 )</td>
<td>Type I error (( \alpha ))</td>
<td>Correct decision</td>
</tr>
<tr>
<td>Do not reject ( H_0 )</td>
<td>Correct decision</td>
<td>Type II error</td>
</tr>
</tbody>
</table>

is a 0.01% chance that the null hypothesis is true and that we are making an error by rejecting it.

The null hypothesis can be either true or false, and in either case, it can be rejected or not rejected. If the null hypothesis is true and we reject it nonetheless, we are making an incorrect decision. This type of error is called a **Type I error**. Conversely, if the null hypothesis is false but we fail to reject it, this incorrect decision is a **Type II error**.

In Table 8.1, we show the relationship between the two types of errors and the decisions we make regarding the null hypothesis. The probability of a Type I error—rejecting a true hypothesis—is equal to the chosen alpha level. For example, when we set alpha at the .05 level, we know that the probability that the null hypothesis is in fact true is .05 (or 5%).

We can control the risk of rejecting a true hypothesis by manipulating alpha. For example, by setting alpha at .01, we are reducing the risk of making a Type I error to 1%. Unfortunately, however, Type I and Type II errors are inversely related; thus, by reducing alpha and lowering the risk of making a Type I error, we are increasing the risk of making a Type II error (Table 8.1).

As long as we base our decisions on sample statistics and not population parameters, we have to accept a degree of uncertainty as part of the process of statistical inference.

**LEARNING CHECK 8.1**

The implications of research findings are not created equal. For example, researchers might hypothesize that eating spinach increases the strength of weight lifters. Little harm will be done if the null hypothesis that eating spinach has no effect on the strength of weight lifters is rejected in error. The researchers would most likely be willing to risk a high probability of a Type I error, and all weight lifters would eat spinach. However, when the implications of research have important consequences (funding of social programs or medical testing), the balancing act between Type I and Type II errors becomes more important. Can you think of some examples where researchers would want to minimize Type I errors? When might they want to minimize Type II errors?
The t Statistic and Estimating the Standard Error

The Z statistic we have calculated (Formula 8.1) to test the hypothesis involving a sample of California gas stations assumes that the population standard deviation (σ) is known. The value of σ is required to calculate the standard error

\[ \sigma / \sqrt{N} \]

In most situations, σ will not be known, and we will need to estimate it using the sample standard deviation s. We then use the t statistic instead of the Z statistic to test the null hypothesis. The formula for computing the t statistic is

\[ t = \frac{\overline{Y} - \mu}{s / \sqrt{N}} \]  

(8.2)

The t value we calculate is called the t statistic (obtained). The obtained t represents the number of standard deviation units (or standard error units) that our sample mean is from the hypothesized value of μ, assuming that the null hypothesis is true.

The t Distribution and Degrees of Freedom

To understand the t statistic, we should first be familiar with its distribution. The t distribution is actually a family of curves, each determined by its degrees of freedom. The concept of degrees of freedom is used in calculating several statistics, including the t statistic. The degrees of freedom (df) represent the number of scores that are free to vary in calculating each statistic.

To calculate the degrees of freedom, we must know the sample size and whether there are any restrictions in calculating that statistic. The number of restrictions is then subtracted from the sample size to determine the degrees of freedom. When calculating the t statistic for a one-sample test, we start with the sample size N and lose 1 degree of freedom we estimate.\(^6\) Note that the degrees of freedom will increase as the sample size increases. In the case of a single-sample mean, the df is calculated as follows:

\[ df = N - 1 \]  

(8.3)

Comparing the t and Z Statistics

Notice the similarities between the formulas for the t and Z statistics. The only apparent difference is in the denominator. The denominator of Z is the standard error based on the population standard deviation σ. For the denominator of t, we replace σ / \sqrt{N} with s / \sqrt{N}, the estimated standard error based on the sample standard deviation.

However, there is another important difference between the Z and t statistics: Because it is estimated from sample data, the denominator of the t statistic is subject to sampling
error. The sampling distribution of the test statistic is not normal, and the standard normal distribution cannot be used to determine probabilities associated with it.

In Figure 8.3, we present the $t$ distribution for several $df$s. Like the standard normal distribution, the $t$ distribution is bell shaped. The $t$ statistic, similar to the $Z$ statistic, can have positive and negative values. A positive $t$ statistic corresponds to the right tail of the distribution; a negative value corresponds to the left tail. Note that when the $df$ is small, the $t$ distribution is much flatter than the normal curve. But as the degrees of freedom increases, the shape of the $t$ distribution gets closer to the normal distribution, until the two are almost identical when $df$ is greater than 120.

Appendix C summarizes the $t$ distribution. Note that the $t$ table differs from the normal ($Z$) table in several ways. First, the column on the left side of the table shows the degrees of freedom. The $t$ statistic will vary depending on the degrees of freedom, which must first be computed ($df = N - 1$). Second, the probabilities or alpha, denoted as significance levels, are arrayed across the top of the table in two rows, the first for a one-tailed and the second for a two-tailed test. Finally, the values of $t$, listed as the entries of this table, are a function of (a) the degrees of freedom, (b) the level of significance (or probability), and (c) whether the test is a one- or a two-tailed test.

To illustrate the use of this table, let's determine the probability of observing a $t$ value of 2.021 with 40 degrees of freedom and a two-tailed test. Locating the proper row ($df = 40$) and column (two-tailed test), we find the $t$ statistic of 2.021 corresponding to the .05 level of significance. Restated, we can say that the probability of obtaining a $t$ statistic of...
2.021 is .05, or that there are less than 5 chances out of 100 that we would have drawn a random sample with an obtained $t$ of 2.021 if the null hypothesis were correct.

**HYPOTHESIS TESTING WITH ONE SAMPLE AND POPULATION VARIANCE UNKNOWN**

To illustrate the application of the $t$ statistic, let’s test a two-tailed hypothesis about a population mean $\mu$. Let’s say we drew a random sample of 280 white females who worked full time in 2014. We found their mean earnings to be $41,653, with a standard deviation, $s = 29,563. Based on data from the U.S. Census Bureau, we also know that the 2014 mean earnings nationally for all women was $\mu = 39,621$. However, we do not know the value of the population standard deviation. We want to determine whether the sample of white women was representative of the population of all full-time women workers in 2014. Although we suspect that white American women experienced a relative advantage in earnings, we are not sure enough to predict that their earnings were indeed higher than the earnings of all women nationally. Therefore, the statistical test is two tailed.

Let’s apply the five-step model to test the hypothesis that the average earnings of white women differed from the average earnings of all women working full-time in the United States in 2014.

1. **Making Assumptions:** Our assumptions are as follows:
   - A random sample is selected.
   - Because $N > 50$, the assumption of normal population is not required.
   - The level of measurement of the variable *income* is interval ratio.

2. **Stating the Research and the Null Hypotheses and Selecting Alpha:** The research hypothesis is

   $H_1: \mu \neq 39,621$

   and the null hypothesis is

   $H_0: \mu = 39,621$

   We’ll set alpha at .05, meaning that we will reject the null hypothesis if the probability of our obtained statistic is less than or equal to .05.

3. **Selecting the Sampling Distribution and Specifying the Test Statistic:** We use the $t$ distribution and the $t$ statistic to test the null hypothesis.

4. **Computing the Test Statistic:** We first calculate the $df$ associated with our test:

   $df = (N - 1) = (280 - 1) = 279$
To evaluate the probability of obtaining a sample mean of $41,653, assuming the average earnings of white women were equal to the national average of $39,621, we need to calculate the obtained $t$ statistic by using Formula 8.2:

$$
 t = \frac{\bar{Y} - \mu}{s / \sqrt{N}} = \frac{41,653 - 39,621}{29,563 / \sqrt{280}} = \frac{2,032}{1766.73} = 1.15
$$

5. **Making a Decision and Interpreting the Results:** Given our research hypothesis, we will conduct a two-tailed test. To determine the probability of observing a $t$ value of 1.15 with 279 degrees of freedom, let’s refer to Appendix C. From the first column, we can see that 279 degrees of freedom is not listed, so we’ll have to use the last row, $df = \infty$, to assess the significance of our obtained $t$ statistic.

Our obtained $t$ statistic of 1.15 is not listed in the last row. It is less than 1.645 ($t$ critical for .05 one-tailed test) and 1.282 ($t$ critical for .10 one-tailed test). The probability of 1.15 can be estimated as $p > .10$, leading to the conclusion that we fail to reject the null hypothesis. We do not have sufficient evidence to reject the null hypothesis.

## HYPOTHESIS TESTING WITH TWO SAMPLE MEANS

The two examples that we reviewed at the beginning of this chapter dealt with data from one sample compared with data from the population. In practice, social scientists are often more interested in situations involving two (sample) parameters than those involving one, such as the differences between men and women, Democrats and Republicans, whites and nonwhites, or high school or college graduates. Specifically, we may be interested in finding out whether the average years of education for one racial/ethnic group is the same, lower, or higher than another group.

U.S. data on educational attainment reveal that Asian and Pacific Islanders have more years of education than any other racial/ethnic groups; this includes the percentage of those earning a high school degree or higher or a college degree or higher. Though years of education have steadily increased for blacks and Hispanics since 1990, their numbers remain behind Asian and Pacific Islanders and whites.

Using data from the 2014 General Social Survey (GSS), we examine the difference in white and black educational attainment. From the GSS sample, white respondents reported an average of 13.88 years of education and blacks, an average of 13.00 years as shown in Table 8.2. These sample averages could mean either (a) the average number of years of education for whites is higher than the average for blacks or (b) the average for whites is actually about the same as for blacks, but our sample just happens to indicate a higher average for whites. What we are applying here is a bivariate analysis (for more information, refer to Chapters 2 and 9), a method to detect and describe the relationship between two variables—race/ethnicity and educational attainment.

The statistical procedures discussed in the following sections allow us to test whether the differences that we observe between two samples are large enough for us to conclude that
the populations from which these samples are drawn are different as well. We present tests for the significance of the differences between two groups. Primarily, we consider differences between sample means and differences between sample proportions.

Hypothesis testing with two samples follows the same structure as for one-sample tests: The assumptions of the test are stated, the research and null hypotheses are formulated and the alpha level selected, the sampling distribution and the test statistic are specified, the test statistic is computed, and a decision is made whether or not to reject the null hypothesis.

The Assumption of Independent Samples

One important difference between one- and two-sample hypothesis testing involves sampling procedures. With a two-sample case, we assume that the samples are independent of each other. The choice of sample members from one population has no effect on the choice of sample members from the second population. In our comparison of whites and blacks, we are assuming that the selection of whites is independent of the selection of black individuals. (The requirement of independence is also satisfied by selecting one sample randomly, then dividing the sample into appropriate subgroups. For example, we could randomly select a sample and then divide it into groups based on gender, religion, income, or any other attribute that we are interested in.)

Stating the Research and Null Hypotheses

The second difference between one- and two-sample tests is in the form taken by the research and the null hypotheses. In one-sample tests, both the null and the research hypotheses are statements about a single population parameter, \( \mu \). In contrast, with two-sample tests, we compare two population parameters.

Our research hypothesis \( (H_1) \) is that the average years of education for whites is not equal to the average years of education for black respondents. We are stating a hypothesis about the relationship between race/ethnicity and education in the general population by

Table 8.2  Years of Education for White and Black Men and Women, GSS 2014

<table>
<thead>
<tr>
<th></th>
<th>Whites (Sample 1)</th>
<th>Blacks (Sample 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>13.88</td>
<td>13.00</td>
</tr>
<tr>
<td><strong>Standard deviation</strong></td>
<td>2.99</td>
<td>2.28</td>
</tr>
<tr>
<td><strong>Variance</strong></td>
<td>8.94</td>
<td>5.20</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>445</td>
<td>88</td>
</tr>
</tbody>
</table>
comparing the mean educational attainment of whites with the mean educational attainment of blacks. Symbolically, we use \( \mu \) to represent the population mean; the subscript 1 refers to our first sample (whites) and subscript 2 to our second sample (blacks). Our research hypothesis can then be expressed as

\[ H_1: \mu_1 \neq \mu_2 \]

Because \( H_1 \) specifies that the mean education for whites is not equal to the mean education for blacks, it is a nondirectional hypothesis. Thus, our test will be a two-tailed test. Alternatively, if there were sufficient basis for deciding which population mean score is larger (or smaller), the research hypothesis for our test would be a one-tailed test:

\[ H_1: \mu_1 < \mu_2 \text{ or } H_1: \mu_1 > \mu_2 \]

In either case, the null hypothesis states that there are no differences between the two population means:

\[ H_0: \mu_1 = \mu_2 \]

We are interested in finding evidence to reject the null hypothesis of no difference so that we have sufficient support for our research hypothesis.

**LEARNING CHECK 8.2**

For the following research situations, state your research and null hypotheses:

- There is a difference between the mean statistics grades of social science majors and the mean statistics grades of business majors.
- The average number of children in two-parent black families is lower than the average number of children in two-parent nonblack families.
- Grade point averages are higher among girls who participate in organized sports than among girls who do not.

**THE SAMPLING DISTRIBUTION OF THE DIFFERENCE BETWEEN MEANS**

The sampling distribution allows us to compare our sample results with all possible sample outcomes and estimate the likelihood of their occurrence. Tests about differences between two sample means are based on the sampling distribution of the difference between means. The sampling distribution of the difference between two sample
means is a theoretical probability distribution that would be obtained by calculating all the possible mean differences by drawing all possible independent random samples of size \( N_1 \) and \( N_2 \) from two populations.

The properties of the sampling distribution of the difference between two sample means are determined by a corollary to the central limit theorem. This theorem assumes that our samples are independently drawn from normal populations, but that with sufficient sample size \((N_1 > 50, N_2 > 50)\) the sampling distribution of the difference between means will be approximately normal, even if the original populations are not normal. This sampling distribution has a mean \( \mu_{1-2} \) and a standard deviation (standard error)

\[
\sigma_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}
\]

which is based on the variances in each of the two populations \((\sigma_1^2 \text{ and } \sigma_2^2)\).

**Estimating the Standard Error**

Formula 8.4 assumes that the population variances are known and that we can calculate the standard error \( \sigma_{\bar{Y}_1 - \bar{Y}_2} \) (the standard deviation of the sampling distribution). However, in most situations, the only data we have are based on sample data, and we do not know the true value of the population variances, \( \sigma_1^2 \) and \( \sigma_2^2 \). Thus, we need to estimate the standard error from the sample variances, \( s_1^2 \) and \( s_2^2 \). The estimated standard error of the difference between means is symbolized as \( S_{\bar{Y}_1 - \bar{Y}_2} \) (instead of \( \sigma_{\bar{Y}_1 - \bar{Y}_2} \)).

**Calculating the Estimated Standard Error**

When we can assume that the two population variances are equal, we combine information from the two sample variances to calculate the estimated standard error.

\[
S_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}} \sqrt{\frac{N_1 + N_2}{N_1N_2}}
\]

where \( S_{\bar{Y}_1 - \bar{Y}_2} \) is the estimated standard error of the difference between means, and \( s_1^2 \) and \( s_2^2 \) are the variances of the two samples. As a rule of thumb, when either sample variance is more than twice as large as the other, we can no longer assume that the two population variances are equal and would need to use Formula 8.8 in A Closer Look 8.2.

**The t Statistic**

As with single sample means, we use the \( t \) distribution and the \( t \) statistic whenever we estimate the standard error for a difference between means test. The \( t \) value we calculate is the obtained \( t \). It represents the number of standard deviation units (or standard error units) that our mean difference \( \bar{Y}_1 - \bar{Y}_2 \) is from the hypothesized value of \( \mu_1 - \mu_2 \), assuming that the null hypothesis is true.
The formula for computing the \( t \) statistic for a difference between means test is

\[
t = \frac{\overline{Y}_1 - \overline{Y}_2}{S_{\overline{Y}_1 - \overline{Y}_2}}
\]  

(8.6)

where \( S_{\overline{Y}_1 - \overline{Y}_2} \) is the estimated standard error.

**Calculating the Degrees of Freedom for a Difference Between Means Test**

To use the \( t \) distribution for testing the difference between two sample means, we need to calculate the degrees of freedom. As we saw earlier, the degrees of freedom (\( df \)) represent the number of scores that are free to vary in calculating each statistic. When calculating the \( t \) statistic for the two-sample test, we lose 2 degrees of freedom, one for every population variance we estimate. When population variances are assumed to be equal or if the size of both samples is greater than 50, the \( df \) is calculated as follows:

\[
df = (N_1 + N_2) - 2
\]  

(8.7)

When we cannot assume that the population variances are equal and when the size of one or both samples is equal to or less than 50, we use Formula 8.9 in A Closer Look 8.2 to calculate the degrees of freedom.

**THE FIVE STEPS IN HYPOTHESIS TESTING ABOUT DIFFERENCE BETWEEN MEANS: A SUMMARY**

As with single-sample tests, statistical hypothesis testing involving two sample means can be organized into five steps.

1. **Making Assumptions:** In our example, we made the following assumptions:
   - Independent random samples are used.
   - The variable *years of education* is measured at an interval-ratio level of measurement.
   - Because \( N_1 > 50 \) and \( N_2 > 50 \), the assumption of normal population is not required.
   - The population variances are assumed to be equal.

2. **Stating the Research and Null Hypotheses and Selecting Alpha:** Our research hypothesis is that the mean education of whites is different from the mean education of blacks, indicating a two-tailed test. Symbolically, the research hypothesis is expressed as

\[
H_1: \mu_1 \neq \mu_2
\]

with \( \mu_1 \) representing the mean education of whites and \( \mu_2 \) the mean education of blacks.
CALCULATING THE ESTIMATED STANDARD ERROR AND THE DEGREES OF FREEDOM (df) WHEN THE POPULATION VARIANCES ARE ASSUMED TO BE UNEQUAL

If the variances of the two samples ($s_1^2$ and $s_2^2$) are very different (one variance is twice as large as the other), the formula for the estimated standard error becomes

$$S_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}$$

(8.8)

When the population variances are unequal and the size of one or both samples is equal to or less than 50, we use another formula to calculate the degrees of freedom associated with the $t$ statistic:

$$df = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\left(\frac{s_1^2}{N_1}/(N_1 - 1) + \frac{s_2^2}{N_2}/(N_2 - 1)\right)}$$

(8.9)

The null hypothesis states that there are no differences between the two population means, or

$$H_0: \mu_1 = \mu_2$$

We are interested in finding evidence to reject the null hypothesis of no difference so that we have sufficient support for our research hypothesis. We will reject the null hypothesis if the probability of $t$ (obtained) is less than or equal to .05 (our alpha value).

3. **Selecting the Sampling Distribution and Specifying the Test Statistic:** The $t$ distribution and the $t$ statistic are used to test the significance of the difference between the two sample means.

4. **Computing the Test Statistic:** To test the null hypothesis about the differences between the mean education of whites and blacks, we need to translate the ratio of the observed differences to its standard error into a $t$ statistic (based on data presented in Table 8.2). The obtained $t$ statistic is calculated using Formula 8.6:

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{S_{\bar{Y}_1 - \bar{Y}_2}}$$

where $S_{\bar{Y}_1 - \bar{Y}_2}$ is the estimated standard error of the sampling distribution. Because the population variances are assumed to be equal, $df$ is $(N_1 + N_2) - 2 = (445 + 88) - 2 = 531$ and we can combine information from the two sample variances to estimate the standard error (Formula 8.5):

$$S_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{\frac{(445 - 1)(2.99)^2 + (88 - 1)(2.28)^2}{(445 + 88) - 2}} = \sqrt{\frac{445 + 88}{445(88)}} = (2.89)(0.12) = 0.35$$
We substitute this value into the denominator for the \( t \) statistic (Formula 8.6):

\[
 t = \frac{13.88 - 13.00}{0.35} = 2.51
\]

5. **Making a Decision and Interpreting the Results:** We confirm that our obtained \( t \) is on the right tail of the distribution. Since our obtained \( t \) statistic of 2.51 is greater than \( t = 1.96 \) \((df = \infty, \text{two tailed; see Appendix C})\), we can state that its probability is less than .05. We can reject the null hypothesis of no difference between the educational attainment of whites and blacks. We conclude that based on our sample data, white men and women, on average, have significantly higher years of education than black men and women do.

**LEARNING CHECK 8.3**

Would you change your decision in the previous example if alpha was .01? Why or why not?

**STATISTICS IN PRACTICE:**

**CIGARETTE USE AMONG TEENS**

Administered annually since 1975, the Monitoring the Future (MTF) survey measures the extent of and beliefs regarding drug use among 8th, 10th, and 12th graders. In recent years, data collected from the MTF surveys revealed decreases or stability in drug use among youths, particularly for cigarettes, alcohol, marijuana, cocaine, and methamphetamine.9

Let’s examine data from the MTF 2014 survey, comparing first-time cigarette use between black and white students.

We will rely on SPSS to calculate the \( t \) obtained for the data. We will not present the complete five-step model and \( t \)-test calculation because we want to focus here on interpreting the SPSS output. However, we will need a research hypothesis and an alpha level to guide our interpretation. SPSS always estimates a two-tailed test, namely, does the gap of 1.47 (7.2844 - 5.8095) indicate a difference in when black and white adolescents first smoke cigarettes? We’ll set alpha at .05.

The output includes two tables. The Group Statistics table (Figure 8.4) presents descriptive statistics for each group. The survey results indicate that black students are more likely to smoke cigarettes later (in earlier grades) than white students. The mean grade of first use of cigarettes is 5.8095 for black students and 7.2844 for white students.

In the second table (Figure 8.5), labeled Independent Samples Test, \( t \) statistics are presented for equal variances assumed (\(-3.444\)) and equal variances not assumed (\(-3.665\)). Both \( t \)-obtained statistics are negative, indicating that the average grade for black students is lower than the average grade for white students. In order to determine which \( t \) statistic to use, review the results of the Levene’s test for Equality of Variances. The Levene’s test (a calculation that we will not cover in this text) tests the null hypothesis that
the population variances are equal. If the significance of the reported \( F \) statistic is equal to or less than .05 (the baseline alpha for the Levene's test), we can reject the null hypothesis that the variances are equal; if the significance is greater than .05, we fail to reject the null hypothesis. (In other words, if the significance for the Levene’s test is greater than .05, refer to the \( t \) obtained for equal variances assumed; if the significance is less than .05, refer to the \( t \) obtained for equal variances not assumed.) Since the significance of \( F \) is .782 > .05, we fail to reject the null hypothesis and conclude that the variances are equal. Thus, the \( t \) obtained that we will use for this model is -3.444 (the one corresponding to equal variances assumed).

SPSS calculates the exact probability of the \( t \) obtained for a two-tailed test. There is no need to estimate it based on Appendix C (as we did in our previous example). The significance of -3.444 is .001, which is less than our alpha level of .05. We reject the null hypothesis of no difference for grade of first-time cigarette use between white and black students. On average, black students first use cigarettes at a later grade (1.47 grades earlier) than white students.

**LEARNING CHECK 8.4**

State the null and research hypothesis for this SPSS example.

Would you change your decision in the previous example if alpha was .01? Why or why not?
HYPOTHESIS TESTING WITH TWO SAMPLE PROPORTIONS

In the preceding sections, we have learned how to test for the significance of the difference between two population means when the variable is measured at an interval-ratio level. Yet numerous variables in the social sciences are measured at a nominal or an ordinal level. These variables are often described in terms of proportions or percentages. For example, we might be interested in comparing the proportion of those who support immigrant policy reform among Hispanics and non-Hispanics or the proportion of men and women who supported the Democratic candidate during the last presidential election. In this section, we present statistical inference techniques to test for significant differences between two sample proportions.

Hypothesis testing with two sample proportions follows the same structure as the statistical tests presented earlier: The assumptions of the test are stated, the research and null hypotheses are formulated, the sampling distribution and the test statistic are specified, the test statistic is calculated, and a decision is made whether or not to reject the null hypothesis.

In 2013, the Pew Research Center presented a comparison of first-generation Americans (immigrants who were foreign born) and second-generation Americans (adults who have at least one immigrant parent) on several key demographic variables. Based on several measures of success, the Center documented social mobility between the generations, confirming that second-generation Americans were doing better than the first-generation Americans. The statistical question we examine here is whether the difference between the generations is significant.

For example, according to the Center’s report, the proportion of first-generation Hispanic Americans who earned a bachelor’s degree or higher was 0.11 ($p_1$); the proportion of second-generation Hispanic Americans with the same response was 0.21 ($p_2$). A total of 899 first-generation Hispanic Americans ($N_1$) and 351 second-generation Hispanic Americans ($N_2$) answered this question. We use the five-step model to determine whether the difference between the two proportions is significant.

1. Making Assumptions: Our assumptions are as follows:
   - Independent random samples of $N_1 > 50$ and $N_2 > 50$ are used.
   - The level of measurement of the variable is nominal.

2. Stating the Research and Null Hypotheses and Selecting Alpha: We propose a two-tailed test that the population proportions for first-generation and second-generation Hispanic Americans are not equal.

   $H_1: \pi_1 \neq \pi_2$

   $H_0: \pi_1 = \pi_2$

We decide to set alpha at .05.
3. **Selecting the Sampling Distribution and Specifying the Test Statistic:** The population distributions of dichotomies are not normal. However, based on the central limit theorem, we know that the sampling distribution of the difference between sample proportions is normally distributed when the sample size is large (when \(N_1 > 50\) and \(N_2 > 50\)), with mean \(\mu_{p_1 - p_2}\) and the estimated standard error \(S_{p_1 - p_2}\). Therefore, we can use the normal distribution as the sampling distribution, and we can calculate \(Z\) as the test statistic.\(^{11}\)

The formula for computing the \(Z\) statistic for a difference between proportions test is

\[
Z = \frac{p_1 - p_2}{S_{p_1 - p_2}}
\]

where \(p_1\) and \(p_2\) are the sample proportions for first- and second-generation Hispanic Americans, and \(S_{p_1 - p_2}\) is the estimated standard error of the sampling distribution of the difference between sample proportions.

The estimated standard error is calculated using the following formula:

\[
S_{p_1 - p_2} = \sqrt{\frac{p_1(1-p_1)}{N_1} + \frac{p_2(1-p_2)}{N_2}}
\]

4. **Calculating the Test Statistic:** We calculate the standard error using Formula 8.11:

\[
S_{p_1 - p_2} = \sqrt{\frac{0.11(1-0.11)}{899} + \frac{0.21(1-0.21)}{351}} = \sqrt{0.000581547} = 0.02
\]

Substituting this value into the denominator of Formula 8.10, we get

\[
Z = \frac{0.11 - 0.21}{0.02} = -5.00
\]

5. **Making a Decision and Interpreting the Results:** Our obtained \(Z\) of \(-5.00\) indicates that the difference between the two proportions will be evaluated at the left tail (the negative side) of the \(Z\) distribution. To determine the probability of observing a \(Z\) value of \(-5.00\) if the null hypothesis is true, look up the value in Appendix B (Column C) to find the area to the right of (above) the obtained \(Z\).

Note that a \(Z\) score of 5.00 is not listed in Appendix B; however, the value exceeds the largest \(Z\) reported in the table, 4.00. The \(p\) value corresponding to a \(Z\) score of \(-5.00\) would be less than \(.0001\). For a two-tailed test, we’ll have to multiply \(p\) by 2 \((.0001 \times 2 = .0002)\). If this were a one-tailed test, we would not have to multiply the \(p\) value by 2. The probability of \(-5.00\) for a two-tailed test is less than our alpha level of \(.05\) \((.0002 < .05)\).

Thus, we reject the null hypothesis of no difference. Based on the Pew Research data, we conclude that there is a significantly higher proportion of college graduates among
second-generation Hispanic Americans compared with first-generation Hispanic Americans.

**LEARNING CHECK 8.5**

If alpha was changed to .01, two-tailed test, would your final decision change? Explain.

We continue our analysis of the 2013 Pew Research Center data, this time examining the difference in educational attainment between first- and second-generation Asian Americans presented in Table 8.3. Our research hypothesis is whether there is a lower proportion of college graduates among first-generation Asian Americans than second-generation Asian Americans, indicating a one-tailed test. We'll set alpha at .05.

The final calculation for $Z$ is

$$Z = \frac{0.50 - 0.55}{0.02} = -2.50$$

The one-tailed probability of $-2.50$ is .0062. Comparing .0062 to our alpha, we reject the null hypothesis of no difference. We conclude that a significantly higher proportion of second-generation Asian Americans (55%) have a bachelor’s degree or higher compared with first-generation Asian Americans (50%). The 5% difference is significant at the .05 level.

**LEARNING CHECK 8.6**

If alpha was changed to .01, one-tailed test, would your final decision change? Explain.

<table>
<thead>
<tr>
<th>First-Generation Asian Americans</th>
<th>Second-Generation Asian Americans</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_1 = .50$</td>
<td>$\rho_2 = .55$</td>
</tr>
<tr>
<td>$N_1 = 2,684$</td>
<td>$N_2 = 566$</td>
</tr>
</tbody>
</table>

Let’s conclude with an example of how the results of statistical hypothesis testing are presented in the social science research literature. Keep in mind that the research literature does not follow the same format or the degree of detail that we’ve presented in this chapter. For example, most research articles do not include a formal discussion of the null hypothesis or the sampling distribution. The presentation of statistical analyses and detail will vary according to the journal’s editorial policy or the standard format for the discipline.

It is not uncommon for a single research article to include the results of 10 to 20 statistical tests. Results have to be presented succinctly and in summary form. An author’s findings are usually presented in a summary table that may include the sample statistics (e.g., the sample means), the obtained test statistics (t or Z), the p level, and an indication of whether or not the results are statistically significant.

Robert Emmet Jones and Shirley A. Rainey (2006) examined the relationship between race, environmental attitudes, and perceptions about environmental health and justice. Researchers have documented how people of color and the poor are more likely than whites and more affluent groups to live in areas with poor environmental quality and protection, exposing them to greater health risks. Yet little is known about how this disproportional exposure and risk are perceived by those affected. Jones and Rainey studied black and white residents from the Red River community in Tennessee, collecting data from interviews and a mail survey during 2001 to 2003.

They created a series of index scales measuring residents’ attitudes pertaining to environmental problems and issues. The Environmental Concern (EC) Index measures public concern for specific environmental problems in the neighborhood. It includes questions on drinking water quality, landfills, loss of trees, lead paint and poisoning, the condition of green areas, and stream and river conditions. EC-II measures public concern (very unconcerned to very concerned) for the overall environmental quality in the neighborhood. EC-III measures the seriousness (not serious at all to very serious) of environmental problems in the neighborhood. Higher scores on all EC indicators indicate greater concern for environmental problems in their neighborhood. The Environmental Health (EH) Index measures public perceptions of certain physical side effects, such as headaches, nervous disorders, significant weight loss or gain, skin rashes, and breathing problems. The EH Index measures the likelihood (very unlikely to very likely) that the person believes that he or she or a household member experienced health problems due to exposure to environmental contaminants in his or her neighborhood. Higher EH scores reflect a greater likelihood that respondents believe that they have experienced health problems from exposure to environmental contaminants. Finally, the Environmental Justice (EJ) Index measures public perceptions about environmental justice, measuring the extent to which they agreed (or disagreed) that public officials had informed residents about environmental problems, enforced environmental laws, or held meetings to address residents’ concerns. A higher mean EJ score indicates a greater
Table 8.4  Environmental Concern (EC), Environmental Health (EH), and Environmental Justice (EJ)

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Group</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>t</th>
<th>Significance (One Tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC Index</td>
<td>Blacks</td>
<td>56.2</td>
<td>13.7</td>
<td>6.2</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Whites</td>
<td>42.6</td>
<td>15.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EC-II</td>
<td>Blacks</td>
<td>4.4</td>
<td>1.0</td>
<td>5.6</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Whites</td>
<td>3.5</td>
<td>1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EC-III</td>
<td>Blacks</td>
<td>3.4</td>
<td>1.1</td>
<td>6.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Whites</td>
<td>2.3</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EH Index</td>
<td>Blacks</td>
<td>23.0</td>
<td>10.5</td>
<td>5.1</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Whites</td>
<td>16.0</td>
<td>7.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EJ Index</td>
<td>Blacks</td>
<td>31.0</td>
<td>7.3</td>
<td>3.8</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Whites</td>
<td>27.2</td>
<td>6.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


*Note:* N = 78 blacks, 113 whites.

The results presented [in Table 1] suggest that as a group, Blacks are significantly more concerned than Whites about local environmental issues.
environmental conditions (EC Index). . . . The results . . . also indicate that as a group, Blacks believe they have suffered more health problems from exposure to poor environmental conditions in their neighborhood than Whites (EH Index). . . . [T]here is greater likelihood that Blacks feel local public agencies and officials failed to deal with environmental problems in their neighborhood in a fair, just, and effective manner (EJ Index). (p. 485)
MAIN POINTS

• Statistical hypothesis testing is a decision-making process that enables us to determine whether a particular sample result falls within a range that can occur by an acceptable level of chance. The process of statistical hypothesis testing consists of five steps: (1) making assumptions, (2) stating the research and null hypotheses and selecting alpha, (3) selecting a sampling distribution and a test statistic, (4) computing the test statistic, and (5) making a decision and interpreting the results.

• Statistical hypothesis testing may involve a comparison between a sample mean and a population mean or a comparison between two sample means. If we know the population variance(s) when testing for differences between means, we can use the $Z$ statistic and the normal distribution. However, in practice, we are unlikely to have this information.

• When testing for differences between means when the population variance(s) are unknown, we use the $t$ statistic and the $t$ distribution.

• Tests involving differences between proportions follow the same procedure as tests for differences between means when population variances are known. The test statistic is $Z$, and the sampling distribution is approximated by the normal distribution.

KEY TERMS

alpha ($\alpha$) 206
degrees of freedom ($df$) 211
left-tailed test 203
null hypothesis ($H_0$) 205
one-tailed test 203
$p$ value 206
research hypothesis ($H_1$) 203
right-tailed test 203
sampling distribution of the difference between means 216

statistical hypothesis testing 202
$t$ distribution 211
t statistic (obtained) 211
two-tailed test 204
Type I error 210
Type II error 210
$Z$ statistic (obtained) 206

DIGITAL RESOURCES

Get the edge on your studies. https://edge.sagepub.com/ssdsess3e

Take a quiz to find out what you’ve learned.
Review key terms with eFlashcards.
Dive into real research with SAGE journal articles.
CHAPTER EXERCISES

1. It is known that, nationally, doctors working for health maintenance organizations (HMOs) average 13.5 years of experience in their specialties, with a standard deviation of 7.6 years. The executive director of an HMO in a Western state is interested in determining whether or not its doctors have less experience than the national average. A random sample of 150 doctors from HMOs shows a mean of only 10.9 years of experience.

   a. State the research and the null hypotheses to test whether or not doctors in this HMO have less experience than the national average.

   b. Using an alpha level of .01, calculate this test.

2. In this chapter, we examined the difference in educational attainment between first- and second-generation Hispanic and Asian Americans based on the proportion of each group with a bachelor’s degree. We present additional data from the Pew Research Center’s 2013 report, measuring the percentage of each group that owns a home.

<table>
<thead>
<tr>
<th>Percentage Owning a Home</th>
</tr>
</thead>
<tbody>
<tr>
<td>First-Generation Hispanic Americans N = 899</td>
</tr>
<tr>
<td>Second-Generation Hispanic Americans N = 351</td>
</tr>
<tr>
<td>First-Generation Asian Americans N = 2,684</td>
</tr>
<tr>
<td>Second-Generation Asian Americans N = 566</td>
</tr>
</tbody>
</table>


   a. Test whether there is a significant difference in the proportion of homeowners between first- and second-generation Hispanic Americans. Set alpha at .05.

   b. Test whether there is a significant difference in the proportion of homeowners between first- and second-generation Asian Americans. Set alpha at .01.

3. For each of the following situations determine whether a one- or a two-tailed test is appropriate. Also, state the research and the null hypotheses.

   a. You are interested in finding out if the average household income of residents in your state is different from the national average household. According to the U.S. Census, for 2014, the national average household income is $53,657.15

   b. You believe that students in small liberal arts colleges attend more parties per month than students nationwide. It is known that nationally undergraduate students attend an average of 3.2 parties per month. The average number of parties per month will be calculated from a random sample of students from small liberal arts colleges.
c. A sociologist believes that the average income of elderly women is lower than the average income of elderly men.

d. Is there a difference in the amount of study time on-campus and off-campus students devote to their schoolwork during an average week? You prepare a survey to determine the average number of study hours for each group of students.

e. Reading scores for a group of third graders enrolled in an accelerated reading program are predicted to be higher than the scores for nonenrolled third graders.

f. Stress (measured on an ordinal scale) is predicted to be lower for adults who own dogs (or other pets) than for non–pet owners.

4. In 2016, the Pew Research Center surveyed 1,799 white and 1,001 black Americans about their views on race and inequality. Pew researchers found “profound differences between black and white adults in their views on racial discrimination, barriers to black progress and the prospects for change.” White and black respondents also disagreed about the best methods to achieve racial equality. For example, 34% of whites and 41% of blacks said that “bringing people of different racial backgrounds together to talk about race” would be a very effective tactic for groups striving to help blacks achieve equality. Test whether the proportion of white respondents who support this tactic is significantly less than the proportion of black respondents.

a. State the null and research hypotheses.

b. Calculate the Z statistic and test the hypothesis at the .05 level. What is your Step 5 decision?

5. One way to check on how representative a survey is of the population from which it was drawn is to compare various characteristics of the sample with the population characteristics. A typical variable used for this purpose is age. The GSS 2014 found a mean age of 50.12 and a standard deviation of 17.07 for its sample of 1,490 American adults. Assume that we know from census data that the mean age of all American adults is 37.20. Use this information to answer the following questions.

a. State the research and the null hypotheses for a two-tailed test of means.

b. Calculate the t statistic and test the null hypothesis at the .001 significance level. What did you find?

c. What is your decision about the null hypothesis? What does this tell us about how representative the sample is of the American adult population?

6. Is there a significant difference in the level of community service participation between college and high school graduates? According to the GSS 2014, 44% of 88 college graduates reported volunteering in the previous month compared with 30% of 245 high school graduates.

a. What is the research hypothesis? Should you conduct a one- or a two-tailed test? Why?

b. Present the five-step model, testing your hypothesis at the .05 level. What do you conclude?
7. GSS 2014 respondents were asked, “Some people say the following things are important for being truly American. Others say they are not important. How important do you think each of the following is—to be a Christian?” Responses were measured on a 4-point scale: 1 = very important, 2 = fairly important, 3 = not very important, and 4 = not important at all. Those with a high school degree had an average score of 2.35 (s = 1.21, N = 189) while respondents with a bachelor’s degree had an average score of 3.05 (s = 1.05, N = 61).

a. What is the appropriate test statistic? Why?

b. Test the null hypothesis with a one-tailed test (high school graduates are less likely to believe being Christian is not important at all); \( \alpha = .05 \). What do you conclude about the difference in attitudes between high school and bachelor’s degree graduates?

c. If you conducted a two-tailed test with \( \alpha = .05 \), would your decision have been different?

8. We compare the proportion who indicated that it was “very important” to be born in this country to be American for two GSS 2014 groups: respondents (1) born in the United States (native born) and (2) not born in the United States (foreign born). Test the research hypothesis that a higher proportion of native-born respondents than foreign-born respondents indicated being born in the United States was “very important” to be American. Set alpha at .05.

<table>
<thead>
<tr>
<th></th>
<th>Native Born</th>
<th>Foreign Born</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion</td>
<td>.48</td>
<td>.36</td>
</tr>
<tr>
<td>( N )</td>
<td>338</td>
<td>59</td>
</tr>
</tbody>
</table>

9. In surveys conducted during August 2016 (months before the election), the Pew Research Center reported that among 752 men, 55% indicated that regardless of how they felt about Hillary Clinton, the election of a woman as president would be very important historically. Among 815 women, 65% reported the same. Do these differences reflect a significant gender gap?

a. If you wanted to test the research hypothesis that the proportion of male voters who believe in the historical importance of the election of a woman as president is less than the proportion of female voters who believe the same, would you conduct a one- or a two-tailed test?

b. Test the research hypothesis at the .05 alpha level. What do you conclude?

c. If alpha were changed to .01, would your decision remain the same?

10. In the chapter, we examined the difference between white and black students in the timing of when they first tried cigarettes. In this SPSS output, we examine the grade when MTF 2014 white and black students first tried alcohol. Present Step 5 (final decision) for these data. Assume alpha = .05, two-tailed test.
11. Research indicates that charitable giving is more common among older adults, though increased giving by Millennials is part of a growing trend. We examine charitable giving (measured in dollars) for two age groups: (1) 30–39 years and (2) 50–59 years of age, based on data from the GSS 2014. Assume alpha = .05 for a two-tailed test. What can you conclude about the difference in giving between the two age groups?

<table>
<thead>
<tr>
<th>Race</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>49</td>
<td>7.0408</td>
<td>1.70733</td>
<td>.24390</td>
</tr>
<tr>
<td>Black</td>
<td>224</td>
<td>7.8125</td>
<td>1.67027</td>
<td>.11160</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NGradeAlcohol</th>
<th>F</th>
<th>Sig.</th>
<th>T</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
<th>Mean Difference</th>
<th>Std. Error Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal variances assumed</td>
<td>.009</td>
<td>.925</td>
<td>-2.918</td>
<td>271</td>
<td>.004</td>
<td>-.77168</td>
<td>.26446</td>
</tr>
<tr>
<td>Equal variances not assumed</td>
<td>-2.877</td>
<td>69.546</td>
<td>.005</td>
<td>77168</td>
<td>.26822</td>
<td>-1.30570</td>
<td>-2.3567</td>
</tr>
</tbody>
</table>

12. Based on the 2014 MTF survey, we compare social network media usage (measured by an ordinal scale: 1 = none, 2 = less than an hour, 3 = 1–2 hours, 4 = 3–5 hours, 5 = 6–9 hours, 6 = 10–19 hours, 7 = 20–29 hours, 8 = 30–39 hours, and 9 = 40+ hours) between males and females. (Recall that the MTF 2014 is a survey of teens in the 8th, 10th, and 12th grades.)

a. Interpret the group means for males and females. Which group spends more time on social media?
b. Assume alpha = .05 for a two-tailed test. What can you conclude about the difference in social network usage between the two groups?

c. If alpha were changed to .01, would your Step 5 decision change? Explain.

<table>
<thead>
<tr>
<th>sex 2014 R01</th>
</tr>
</thead>
<tbody>
<tr>
<td>K'S.SEX F1234</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>SocialNetHR 2014 C09 PHB/W SOCIAL NET WEB F11</td>
</tr>
<tr>
<td>1 MALE: (1)</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>184</td>
</tr>
<tr>
<td>2 FEMALE: (2)</td>
</tr>
<tr>
<td>220</td>
</tr>
</tbody>
</table>

### Independent Samples Test

| Levene's Test for Equality of Variances | t-test for Equality of Means |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference |
| SocialNetHR 2014 C09 PHB/W SOCIAL NET WEB F11 | Equal variances assumed | 2.866 | .091 | -5.828 | 402 | .000 | -.1447 | .248 | -.1935 | -.959 |
| | Equal variances not assumed | -5.867 | .000 | 397.706 | .000 | -.1447 | .247 | -.1932 | -.962 |

13. The GSS 2014 asked respondents “after an average work day, about how many hours do you have to relax or pursue the activities you enjoy”? In this exercise, we selected married GSS respondents and calculated the t test for HRSRELAX.

a. Is there a significant difference between married men and married women in the number of hours they have to relax during the day? Set alpha at .05.

b. If alpha was changed to .01, would your Step 5 decision change? Explain.

<table>
<thead>
<tr>
<th>GROUPS</th>
<th>N</th>
<th>MEAN</th>
<th>SD</th>
<th>SE MEAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>MALE</td>
<td>126</td>
<td>3.56</td>
<td>2.566</td>
<td>.229</td>
</tr>
<tr>
<td>FEMALE</td>
<td>113</td>
<td>2.88</td>
<td>2.164</td>
<td>.204</td>
</tr>
</tbody>
</table>

### Independent Samples Test

| Levene's Test for Equality of Variances | t-test for Equality of Means |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | F | Sig. | t | df | Sig. (2-tailed) | Mean Difference | Std. Error Difference |
| HOURS PER DAY & HAVE TO RELAX | Equal variances assumed | 2.596 | .108 | 2.225 | 237 | .027 | .687 | .309 | .079 | 1.296 |
| | Equal variances not assumed | 2.246 | .026 | 236.126 | .026 | .687 | .306 | .084 | 1.290 |

Copyright ©2018 by SAGE Publications, Inc.  
This work may not be reproduced or distributed in any form or by any means without express written permission of the publisher.