In chapter 8, we discussed how to use sample data to make interval estimates, called confidence intervals, of population proportions and means. Along with using confidence intervals, many researchers also use a tool called hypothesis testing to assess the likelihood that a sample statistic reflects a population parameter. For instance, we could ask whether a particular segment of college students (such as students at your college) is similar to college students in general in the number of years it takes to complete their degree. Hypothesis tests use the logic of probability and sampling distributions. Tests of the relationship between a sample and a larger population are called one-sample tests. Researchers use the same procedures to test hypotheses about the difference between two samples, in two-sample tests, which we discuss in chapter 10.

The results of hypothesis tests are sometimes used to assess the effectiveness of programs meant to address some social problem. For example, researchers and policy-makers have long been concerned about the negative consequences of living in poor neighborhoods. Studies show that people living in high-poverty areas are more likely to struggle with health problems, less likely to do well in school, and less likely to experience a vibrant and cohesive community life. The Moving to Opportunity project (MTO), which took place during the 1990s in five large cities across the United States, randomly selected residents of high-poverty neighborhoods to receive housing vouchers to move to low-poverty areas. To monitor the success of the program, researchers conducted follow-up interviews with samples of participating families at various intervals of time after their relocations.

Studies have shown that MTO seems to “work” in some ways but not in others. For example, the proportion of MTO movers who have diabetes is lower than the known proportion for residents of poor neighborhoods in general, but mean educational test scores of students who moved to low-poverty areas through the program are not very different from those for students who continue to reside in poor neighborhoods.
A one-sample hypothesis test compares the Type 2 diabetes rate in a sample of MTO movers to the diabetes rate for the full population of residents of poor neighborhoods. The hypothesis test tells us the probability of obtaining our sample diabetes rate if the rate for all MTO movers were actually the same as the rate for poor residents overall. If that test determined that it was very unlikely that the diabetes rate among all MTO movers was the same as the diabetes rate for poor residents in general, this could be interpreted as evidence of the program’s success. Such results would indicate a low chance that the lower diabetes rate for the MTO sample was the result of sampling variation. We would conclude that the diabetes rate for the entire population of MTO movers (not just our one sample) is lower than the known diabetes rate for residents of high-poverty neighborhoods. On the other hand, if a one-sample hypothesis test indicated a high probability that mean test scores for a sample of students who had moved to low-poverty neighborhoods under MTO were no different from mean test scores for students who live in high-poverty areas, this evidence could undermine claims of MTO’s effectiveness.

It is likely that people’s lives improved in some ways after relocating to low-poverty neighborhoods through MTO yet remained unchanged in other ways. Should this relocation intervention be implemented on a larger scale? To make that decision, we need to know, first, whether some outcomes are more preferable (e.g., lower diabetes rates versus higher test scores) and, second, what magnitude of difference between movers and stayers is large enough to justify the costs of the intervention. Hypothesis tests cannot answer those questions. Both questions concern the values of a given society.

The Logic of Hypothesis Testing

Hypothesis testing relies on a counterfactual logic. Instead of trying to prove one thing, we generally try to disprove the opposite. Specifically, instead of trying to prove that our sample is different from a population of interest, we see whether we can disprove that our sample is the same as that population. We begin by generating a null hypothesis. The null hypothesis states that an unknown population parameter is equal to a specified value. Hypothesis testing determines how likely it is that we would obtain a sample statistic at least as far from the population parameter as our sample statistic, if that null hypothesis were correct. We often think about tests as providing us with decisive answers to questions. But hypothesis tests depart from our common understanding of how tests work because they do not give us definitive answers about whether our hypotheses are correct.

That may sound convoluted, but this kind of reasoning is something many of us use in daily life. For example, imagine that a sister and her brother are allowed to share one bag of M&Ms after dinner. Their parents leave the brother in charge of dividing them fairly. The sister has a strong preference for blue M&Ms and expects to get many in her half. After her brother divides the M&Ms and delivers his sister her half, she sees that she has just one blue one. Remembering the many blue M&Ms that she has seen in the past when she has had her own bag, she suspects her brother of hoarding the blue M&Ms for himself. She therefore believes that her “sample” is not representative...
of the “population” of M&Ms that made up the entire bag. The sister complains to her parents. Her parents, however, tell her not to assume that her brother has been unfair. Unlike their daughter, who wants to prove that her brother has treated her unfairly, her parents want to assume that their son has treated his sister fairly until they are convinced otherwise. Their null hypothesis is that the proportion of blue M&Ms in the population—that is, the full bag—is the same as the proportion in the sample—that is, the sister’s share. Casting doubt on the null hypothesis would require the parents to consider whether the proportion of blue M&Ms in their daughter’s half is sufficiently different from the proportion of M&Ms in the full bag to suggest that the difference is unlikely to have occurred due to random chance. Our conclusion at the end of a more precise statistical process would state whether there is a statistically significant difference between the proportion of blue M&Ms in the sample (i.e., the sister’s share) and the proportion of blue M&Ms in the population (i.e., the entire bag).

A hypothesis always makes a statement about a population with an unknown parameter. Hypothesis tests use the idea of a sampling distribution to imagine what would happen if we repeatedly collected samples. We know from chapter 7 that repeated samples will yield results that follow a normal distribution, with the true population value at the center of the distribution. One-sample hypothesis tests place the value specified in the null hypothesis at the center of that normal distribution. In other words, hypothesis tests utilize sampling distributions to tell us how likely it is that we would have observed a statistic as extreme as the one we saw in our sample (e.g., the average test score among former residents of high-poverty neighborhoods) if the null hypothesis were true.

**Null Hypotheses (H₀) and Alternative Hypotheses (Hₐ)**

The first step in hypothesis testing is to generate a null hypothesis, which is the hypothesis being tested. The null hypothesis, denoted by H₀, makes a statement about the value of a population parameter. We never hypothesize about a sample because we know the sample statistic. The null hypothesis will always state that the unknown parameter for the population represented by the sample is equal to a specific value. Where do researchers get the specific value for the null hypothesis? Sometimes researchers have data on a full population, such as an election result (which includes all voters), data from the U.S. Census (which includes all residents), or government crime reports (which include all reported crimes), and they want to know whether their sample resembles that known population. Sometimes researchers use an expected finding as the specific value for the population parameter. This expected finding might come from prior research or from a prediction or claim. For example, prior research might show that 40% of all college students engage in binge drinking; the corresponding null hypothesis would use 40% as the population parameter. A study of whether there is a gender difference in declared majors in science might use 56% women as the population parameter for the null hypothesis because 56% of all undergraduates are women. A researcher might wonder whether Americans eat the recommended five servings of vegetables a day; in
this case, the null hypothesis would use five servings as the population parameter. In all cases, the null hypothesis is a statement of no difference between population and sample, thus explaining why this statement is called the “null” hypothesis.

The next step is to formulate the alternative hypothesis ($H_a$). It, too, is a hypothesis about the population. The alternative hypothesis states that the true population value for the group being studied (i.e., the group from which we drew the sample) is different from the value established for the population parameter in the null hypothesis. The null and alternative hypotheses make competing and mutually exclusive claims about the value of a population parameter; they cannot both be true. The alternative hypothesis is sometimes referred to as the research hypothesis because it often refers to what the researcher expects to find, a difference between the hypothesized population parameter and the statistic from the sample being studied. Unlike the null hypothesis, the alternative hypothesis always states that the population parameter is equal to a range of values. For example, a researcher studying the gender composition of science majors knows that 56% of all undergraduate students are women and sets up the null hypothesis that the percentage of science majors who are women is equal to 56%. The researcher’s alternative hypothesis is that the percentage of women in science majors is not equal to 56%. Using the properties of the sampling distribution, the hypothesis test will tell us the probability of obtaining a result as extreme as our sample result if the null hypothesis were true.

**One-Tailed and Two-Tailed Tests**

There are two versions of hypothesis tests: one-tailed and two-tailed tests (sometimes called one-sided and two-sided). In a two-tailed test, the alternative hypothesis states that the population parameter is not equal to the value stated in the null hypothesis—it may be higher or lower than that value. Here, the researcher does not have a specific expectation of the direction of difference for the alternative hypothesis. In a one-tailed test, however, the researcher asks more specifically whether the population parameter is less than or greater than a specified value. The alternative hypothesis in a one-tailed test specifies that the population parameter is either higher or lower than the value specified by the null hypothesis. For example, the researcher studying gender and science majors might hypothesize that the proportion of women majoring in science is lower than .56, the proportion of women undergraduates, rather than hypothesizing that it is simply different. The other option for the alternative hypothesis would then be that the proportion of women in science majors is greater than .56.

**Hypothesis Tests for Proportions**

Next, we will conduct a hypothesis test for proportions. Let’s begin with an example. Imagine that your state is holding a referendum on the legalization of recreational marijuana. A poll of 200 randomly selected young citizens interviewed right before
the election shows that 59% of them support the legalization initiative. On Election Day, 54% of residents statewide voted to support the legalization initiative. Are young citizens really more supportive of the legalization of marijuana than voters in general?

Why do we even ask this question? At one level, it seems fairly obvious that young citizens are indeed more supportive of legalization. After all, if we just compare the percentages, it is clear that 59% of the young group support legalization, compared to 54% of all voters. That’s a difference of 5 percentage points.

But we cannot hastily jump to this conclusion. Fifty-nine percent is a sample statistic. We have seen in chapter 8 that we should not rely on a point estimate because there is a range of possible population values attached to any sample statistic. So if 59% of our sample of young people supports legalization, we know that there is a very high probability that the true proportion of young people in the population who support legalization will be 59% plus or minus a margin of error. That interval may include 54%, in which case we could not assert with confidence that young people are more likely than the general public to support legalization.

Hypothesis testing is a systematic way of addressing the question: Are young people more supportive of legalization than voters in general? We have two pieces of information at our disposal:

- In a random sample of 200 young people, 59% support legalization.
- In the full population, 54% of voters support legalization.

Let’s begin by assuming that—in reality—young people in the population support legalization at the same rate as all voters, which is 54%. Remember, this is an assumption, or a hypothesis. Where does it come from? We know that 54% of all voters support legalization. So, if young people support legalization at the same rate as everyone else, then 54% of young people in the population will support legalization.

This is our null hypothesis:

\[ H_0: \text{The proportion of young people in the population who support legalization is 54\%}. \]

In statistics, we use Greek letters to represent population parameters and Roman letters to represent sample statistics—in this case, \( \pi \) represents the proportion for the population (see Box 7.1 for a refresher on notation). Thus, the null hypothesis in statistical notation is:

\[ H_0: \pi = .54 \]

Next, we formulate the alternative hypothesis \((H_a)\). In this case, we are interested in the question of whether young people in the population are more supportive of legalization than the rest of the electorate. So our alternative hypothesis is:

\[ H_a: \text{The proportion of young people in the population who support legalization is greater than 54\%}. \]
In statistical notation:

\[ H_a: \pi > 0.54 \]

Unlike the null hypothesis, the alternative hypothesis always states that the population parameter is equal to a range of values. In this case, we are suggesting that the proportion of young people who support legalization is greater than 54%. This is a one-tailed test.

Let’s review what we have done. Using information about the full population (54% of them support legalization), we are testing the claim that young people in the population support legalization at the same rate as everyone else. Alternatively, young people may be more supportive of legalization.

The hypothesis test always proceeds under the assumption that the null hypothesis is true. We test the null hypothesis by seeing whether our sample data provide strong evidence against it. The evidence we use is the probability that we could obtain our sample statistic if the null hypothesis were true. If the evidence against the null hypothesis is strong enough, this leads us to support its opposite, the alternative hypothesis. In the context of this example, this means that we are going to answer the following question:

If 54% of all young people support the legalization of marijuana, what is the probability that, out of a randomly selected sample of 200 young people, at least 59% of them would support legalization?

Figure 9.1 graphically displays the question we are asking. The normal curve in Figure 9.1 is a theoretical distribution of sample proportions, centered on the hypothesized population parameter \( \pi = .54 \).
Using what we know from our discussion of the Central Limit Theorem and sampling distributions, we can calculate the probability of obtaining a sample proportion of .59 or higher under the assumption that the true population proportion is .54. Toward this end, we must calculate the z-score for \( p = .59 \) and use the normal table (see appendix A) to find the corresponding p-value, which is the probability associated with that z-score. In order to calculate z, we must calculate the standard error. You learned how to calculate standard errors and z-scores in chapter 8, in order to calculate confidence intervals. But there is one important difference between the construction of the standard error for a confidence interval and for a hypothesis test. Recall that when we construct a confidence interval, we calculate the standard error on the basis of the sample statistic:

\[
SE_p = \sqrt{\frac{p(1-p)}{N}}
\]

When we run a hypothesis test for the difference between a sample proportion and hypothesized population proportion, we use the population parameter (\( \pi \)) instead of the sample statistic (\( p \)).

\[
SE_p = \sqrt{\frac{\pi(1-\pi)}{N}}
\]

We use the population parameter because, in a hypothesis test, we proceed from the assumption that our guess about the population parameter (the \( H_0 \)) is true, and so we construct a distribution of sample proportions around that hypothesized value. In this example, we are operating from the assumption that the true population proportion for young people is .54.

The standard error is:

\[
SE_p = \sqrt{\frac{0.54(1-0.54)}{200}} = \sqrt{0.54(0.46)/200} = 0.035
\]

Next, we need to calculate a z-score for a sample proportion of at least .59. Or, in other words, how far away is our obtained sample proportion from the hypothesized population parameter, in standard errors? We calculate z using the same formula we used in chapter 8:

\[
z = \frac{p - \pi}{SE_p} = \frac{0.59 - 0.54}{0.035} = 1.43
\]

Figure 9.2 shows the normal curve, with the null hypothesis of \( \pi = .54 \) at the center. The sample proportion, \( p \), is marked on the right, and we have added its z-score of 1.43 and the p-value associated with that z-score, .076. Remember: “p” refers to the sample proportion (.59), and “p-value” refers to the probability of obtaining a given z-score or higher.

How do you find the p-value marked on Figure 9.2? By looking up your z-score in the normal table in appendix A. The normal table shows that the p-value corresponding to a z-score of 1.43, represented by the shaded area, is .076. In other words, if the
null hypothesis is true—that is, if 54% of all young people support legalization—then there is about a 7.6% chance that we would have obtained a sample statistic equal to 59% or higher.

Many analysts misinterpret the p-value. They may describe it as the probability that the null hypothesis is true. Or, they may characterize it as the probability of obtaining the specific sample percentage they got. These interpretations are incorrect. There is only one way to correctly interpret the p-value: It is the probability of obtaining a sample statistic at least as far from the population parameter as our sample statistic, if the null hypothesis were true. In the case of a sample statistic that is lower than the null hypothesis value, the p-value is the probability of obtaining the sample statistic or lower if the null hypothesis were true. Even some scientists who have spent their careers thinking carefully about p-values have difficulty translating the technical statement about what p-values tell us into an intuitive interpretation.

Some students find the interpretation of the p-value to be counterintuitive. This is because the lower the p-value, the more evidence the researcher has against the null hypothesis. In other words, lower p-values make it more likely that the researcher will reject the null hypothesis and accept the alternative hypothesis. A very low p-value means that, if our null hypothesis about the population were true, then it would have been very unlikely to obtain the sample statistic that we got. But since we did obtain that sample statistic, we conclude that our null hypothesis about the population is probably incorrect. So with a very low p-value, we will reject the null hypothesis. If we reject the null hypothesis, we can accept the alternative hypothesis.

How low is a low enough p-value? That is a subjective decision that the investigator makes before actually conducting the hypothesis test. Many researchers use a
threshold of 5% or 1%, meaning that the obtained p-value must be lower than one or the other of these thresholds in order to reject the null hypothesis. Statisticians call this threshold “\textit{alpha},” which is the Greek letter \(\alpha\). In social science, most researchers choose an \textbf{alpha-level} of .10, .05, or .01. Any p-value that is less than alpha allows the researcher to reject the null hypothesis and to declare that his or her finding is “\textit{statistically significant}.” That phrase simply means that the results of our hypothesis test have allowed us to reject the null hypothesis. If the p-value is above alpha, we cannot reject the null hypothesis and therefore must conclude that any difference between our sample statistic and the hypothesized population parameter could be due to the normal variation in results from different samples. In this case, we would say that the difference between our sample and the population is not statistically significant. Again, researchers generally set alpha before beginning to calculate sample z or its associated p-value.

Let’s return to our example. We found that if 54% of all young people supported legalization, then there’s a .076 chance that we would get a sample percentage of at least 59%. Should we reject our null hypothesis? The answer to that question depends on what alpha-level we established before conducting the hypothesis test. Table 9.1 summarizes three scenarios with different levels of alpha.

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|}
\hline
Scenario & Obtained p-Value & Alpha & Decision \\
\hline
1 & 0.076 & 0.10 & \textit{Reject null hypothesis} \\
2 & 0.076 & 0.05 & \textit{Do not reject null hypothesis} \\
3 & 0.076 & 0.01 & \textit{Do not reject null hypothesis} \\
\hline
\end{tabular}
\caption{Rejecting or Failing to Reject the Null Hypothesis for Different Levels of Alpha, Young People's Support for Legalization of Marijuana}
\end{table}

Only in Scenario 1, with an alpha-level of .10, would the researcher reject the null hypothesis and conclude that young people are indeed more supportive of legalization. Scenarios 2 and 3 have lower alpha-levels (.05 and .01, respectively), and therefore, using either of these thresholds, the p-value of .076 exceeds alpha. In Scenarios 2 and 3, the researcher could not conclude that young people support the legalization of marijuana at a higher rate than the rest of the population. Later in this chapter, we will explain how to choose an alpha-level.

\section*{The Steps of the Hypothesis Test}

We have discussed the process of conducting a hypothesis test for the difference between a sample proportion and a hypothesized population parameter in detail here. It can be condensed into five steps. First, set up null and alternative hypotheses. Second, set an alpha-level. Third, calculate the standard error. Fourth, calculate the z-score associated with the sample proportion. Fifth, find the p-value associated with the z-score in
One-Tailed and Two-Tailed Tests

Earlier, we introduced the distinction between one-tailed and two-tailed tests. Let’s return to our legalization of marijuana example and examine the difference between these tests in the context of this example. Here is the original set of hypotheses:

- \( H_0: 54\% \text{ of all young people support legalization.} \)
- \( H_a: \text{More than 54}\% \text{ of all young people support legalization.} \)

When we originally looked at this problem (see Figure 9.2), we found that the probability of obtaining our sample proportion, if the null hypothesis were true, is .076. Let’s be
more precise now about what this means. The p-value means that if 54% of all young people support legalization, there is a .076 chance of getting a sample proportion of at least 59%. But what if we offer a different alternative hypothesis? Consider this set of hypotheses:

\[ H_0: \text{54% of all young people support legalization.} \]

\[ H_a: \text{The percentage of all young people who support legalization is not equal to 54%.} \]

Notice that the alternative hypothesis is now two-tailed. In Figure 9.3, we depict what this new set of hypotheses is suggesting.

Notice that we have doubled our p-value. It is now .152 (.076 in each tail). This means that if the null hypothesis were true—that is, if 54% of all young people support legalization—there is a .152 chance of getting a sample proportion as far in either direction from the null hypothesis value as the one we actually obtained. In other words, assuming that the null hypothesis is true, there would be about a 15% chance of obtaining a sample percentage of 59% or higher or 49% or lower.

The mechanics of conducting a two-tailed test are identical to those of a one-tailed test, with one exception. For a two-tailed test, we double the p-value associated with the sample statistic. Naturally, this makes it harder to reject the null hypothesis. Another way to look at this is that a two-tailed test divides the alpha-level between two tails, making the threshold for rejecting the null hypothesis more stringent.

The decision of whether to employ a one-tailed or a two-tailed test is usually dictated by the question at hand. If the researcher is looking for the probability of a sample result
Hypothesis Tests for Means

367

either only above or only below the null hypothesis, then a one-tailed test is called for. A researcher would do so if she had a strong reason, such as findings from previous research, to expect a result that was either above or below the population parameter. On the other hand, a two-tailed test is required if the problem directs the researcher to look for an effect both below and above the null hypothesis—that is, a sample result that differs from the population in either direction. Alternatively, researchers may choose to employ a two-tailed test because it doubles the p-value and thus adds a higher degree of certainty if the null hypothesis is ultimately rejected.

Hypothesis Tests for Means

What if the research question you are interested in is about a mean, rather than a proportion or percentage? For example, you may want to know whether the mean GPA of sociology majors is different from the mean GPA of all students at a university. Or you may want to know whether the mean income for graduates of your college is higher than the mean income for the general population. Many questions that social scientists are interested in deal with whether the mean for a specific group differs from the mean for the general population. Just as we saw with proportions, we use a one-sample hypothesis test to answer these questions. The procedure is very similar to what we saw for proportions, with one key difference.

In chapter 8 you learned that we use a z-score when we either know the population standard deviation (\(\sigma\)) and can use it to estimate the standard error or can estimate the standard error without using \(\sigma\) (as in the case of proportions). You learned that we use a t-value when we do not know the population standard deviation and must estimate standard error using the sample standard deviation. In almost all cases, when testing the difference between a sample mean and a population mean, we do not know \(\sigma\) and therefore must use t.

Let’s look at income. Say we know that the mean income for young adults (age twenty-five to thirty-four) who are college graduates and are employed full-time is
Imagine that we draw a random sample of 200 graduates of your college who are employed full-time and ask them their income. We then calculate that the mean income for this sample is $51,750 and the standard deviation for the sample is $9,000. At first glance, it seems that graduates of your college earn more, on average, than college graduates overall. But we know that it is possible that this difference is due to sampling variation, rather than a real difference. As you know, whenever we draw a sample, our results are likely to vary somewhat from the true population mean, but most of the time the difference from the population will be small. (Remember, if you flip a coin twenty times, you will not always get ten heads and ten tails, but it would be unusual for you to get nineteen heads and one tail.)

Just as we did with proportions, we can determine how likely it is that we could get a sample mean as high as $51,750 if the mean income of graduates of your college is actually the same as that of college graduates overall. This is what a one-sample t-test does. The steps are similar to those for the hypothesis test for proportions. The main difference is that, instead of using the z-score associated with the sample statistic, we use the t-value. In both cases, we are using t or z to determine the probability of obtaining the sample result if the null hypothesis were true.

The steps for conducting the hypothesis test are shown in Table 9.2. We will go through each step in detail as we work through the example.

Table 9.2  Steps in Conducting One-Tailed Hypothesis Test for Mean Income of College Graduates

<table>
<thead>
<tr>
<th>Step 1: Set up hypotheses.</th>
<th>( H_0: \mu = 50,000; H_a: \mu &gt; 50,000 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 2: Set ( \alpha ), calculate DF, and find decision t.</td>
<td>( \alpha = 0.05; \text{DF} = N - 1 = 200 - 1 = 199; \text{decision t} = 1.645 )</td>
</tr>
<tr>
<td>Step 3: Estimate standard error.</td>
<td>( SE_y = \frac{s}{\sqrt{N}} = \frac{9,000}{\sqrt{200}} = \frac{9,000}{14.14} = 636.4 )</td>
</tr>
<tr>
<td>Step 4: Calculate sample t.</td>
<td>( t = \frac{\bar{y} - \mu}{SE_y} = \frac{51,750 - 50,000}{636.4} = \frac{1,750}{636.4} = 2.75 )</td>
</tr>
<tr>
<td>Step 5: Compare sample t and decision t; draw conclusion.</td>
<td>2.75 &gt; 1.645; therefore, reject ( H_0 )</td>
</tr>
</tbody>
</table>

Note: Because the normal table and t-table are set up differently (see appendices A and B), the steps in testing hypotheses about means and proportions differ slightly.

Step 1: Set up null and research hypotheses.

To conduct the hypothesis test, first, we set up hypotheses. In this case, our null hypothesis is that the mean income of all graduates of your college is the same as that of college graduates overall. In other words:

\[ H_0: \mu = 50,000 \]
(Remember, we use Greek letters to represent the population parameters and Roman letters to represent sample statistics—in this case, \( \mu \) and \( \bar{y} \) represent the means for the population and the sample, respectively.)

The alternative hypothesis is that the mean income of all graduates of your college is greater than that of college graduates overall. In other words:

\[ H_a : \mu > 50,000 \]

Drawing the t-curve to represent our null hypothesis, we place the known population mean, 50,000, at the location of the mean (Figure 9.4). We place the sample mean further out on the curve. (Until we calculate the t-value for the sample mean, we do not know exactly where on the curve the sample mean is located.) Recall from chapter 8 that a t-curve has the same general properties as the normal (z) curve and also represents a sampling distribution.

![Figure 9.4](image-url)  
**Figure 9.4** The Null Hypothesis and Sample Statistic on a Normal Curve, College Graduates’ Mean Income

As with the hypothesis test for a proportion, we will estimate the probability of obtaining a sample mean of 51,750 or higher if the true population mean is 50,000. However, as we will see, while we can look up the exact p-value associated with our sample statistic when that sample statistic is z, we must proceed a little differently when using t.

**Step 2: Set the alpha-level, find degrees of freedom (DF), and find decision t.**

We will set our alpha at .05. That means that we will reject our null hypothesis if the probability of obtaining our sample statistic or higher (if the null hypothesis were true) is below .05. Just like z, the sample t-value has a specific p-value associated with it, which is the exact probability of obtaining that score or higher if the null hypothesis were true. Unlike the normal table, the t-table does not tell us the specific p-value. It
just gives us a cutoff value associated with our alpha-level. If the t-value for our sample is beyond this cutoff point, that tells us that the probability of obtaining our sample statistic or higher is lower than .05 (our alpha-level). We call this cutoff point “decision t” because it is the criterion we use for making a decision about our null hypothesis. Statistical software, in contrast, will report the exact p-value for your sample t.

You learned how to find the t-value for a given alpha-level in chapter 8. To recap, you must first calculate degrees of freedom, because there are many t-curves, and you use the t-curve associated with the degrees of freedom (DF) in your sample.

\[ DF = N - 1 \]

For our example: \( DF = 200 - 1 = 199 \)

Looking in the t-table in appendix B, we see that the highest DF is 120. Because our DF is greater than 120, we use the last line of the table, labeled with the infinity symbol, \( \infty \). We see that a t-value of 1.96 is required for a two-tailed test at alpha = .05 and a t of 1.645 for a one-tailed test. Because our research hypothesis is that earnings for graduates of your college are greater than those of college graduates overall, we are conducting a one-tailed test. Our decision t is therefore 1.645.

Steps 3 and 4: Estimate the standard error and calculate the sample t.

Just as with the hypothesis test for a proportion, we calculate a standardized score for the sample value. Whereas for proportions we calculated a z-score, in this case, we are calculating a t-value. Remember from chapter 8 that both z-scores and t-values convert the distance between the sample score and the mean from the units of the variable (here, dollars of income) into standard error units. Instead of saying that the mean income for graduates of your college is $1,750 higher than the mean income for all college students, the t-value will tell us how many standard errors higher your college’s mean income is than the overall mean.

The formula for t, introduced in chapter 8, is:

\[ t = \frac{\bar{y} - \mu}{SE_{\bar{y}}} \]

In other words, we are calculating the difference between our sample mean and the population mean (\( \bar{y} - \mu \)) and then standardizing this in terms of the standard error (SE). In order to calculate the t-value, we must first estimate the standard error (SE). This is the third step of the hypothesis test. We do so using the standard deviation for our sample as a stand-in for the population standard deviation.

\[ SE_{\bar{y}} = \frac{s}{\sqrt{N}} \]

\[ SE_{\bar{y}} = \frac{9,000}{\sqrt{200}} = \frac{9,000}{14.14} = 636.4 \]
Now we can calculate $t$, Step 4 in the hypothesis test:

$$t = \frac{\bar{y} - \mu}{SE_{\bar{y}}} = \frac{51,750 - 50,000}{636.4} = \frac{1,750}{636.4} = 2.75$$

The $t$-value associated with our sample mean of 51,750 is 2.75.

*Step 5: Compare decision $t$ to sample $t$. If sample $t >$ decision $t$, reject the null hypothesis. State your conclusion.*

How do we know whether the probability of obtaining a sample with this $t$-value, if the true population mean is 50,000 (in other words, if the null hypothesis is correct), is below our alpha of .05? We know that a $t$-value of 1.645 (our decision $t$) is the cutoff point for an alpha of .05. If our sample $t$ is greater than decision $t$, the probability of obtaining our sample result if the null hypothesis were true is less than .05.

Figure 9.5 shows the sample mean, sample $t$-value, and decision $t$ on the curve. The sample $t$-value tells us that the sample mean is further out into the tail of the curve than our cutoff value, established by alpha.

We can now compare our sample $t$ of 2.75 with the decision $t$-value. In this case, 2.75 is greater than 1.645. We therefore reject the null hypothesis that earnings for graduates of your college are the same as those of college graduates overall. We conclude that earnings for graduates of your college are statistically significantly greater than those of college graduates overall.

![Figure 9.5](image-url)
Optional follow-up: Calculate a confidence interval to estimate the range within which the population mean is likely to fall.

If we want to estimate the actual mean earnings of graduates of your college, we would conduct a confidence interval for that mean. This is exactly what we did in chapter 8. If you need a refresher, see the summary on pp. 342–343 in chapter 8.

For our example:

\[ CI = \bar{y} \pm t(SE_{\bar{y}}), \text{ where } SE_{\bar{y}} = \frac{s}{\sqrt{N}} \text{ and } t \text{ at } DF = 199 \text{ and } 90\% \text{ confidence } = 1.645 \]

Notice that the t-value for a 90% confidence interval is the same as the t-value associated with an alpha of .05 for a one-tailed test. (Remember, this is not your sample t.) We calculated \( SE = 636.4 \) above. Plugging in the numbers:

\[ CI = 51,750 \pm 1.645(636.4) = 51,750 \pm 1,047 \rightarrow (50,703, 52,797) \]

We conclude that there is a 95% chance that the mean income for graduates of your college falls between $50,703 and $52,797. Notice that this confidence interval does not overlap with $50,000, the mean income for all college graduates.

**BOX 9.3: IN DEPTH**

Assumptions with Hypothesis Tests

Like confidence intervals, hypothesis tests assume that certain conditions hold. These assumptions are:

- Random sampling from a population.
- When we hypothesize about means, we use interval-ratio variables. When we construct hypotheses about proportions, we use nominal variables. As we have seen, ordinal variables can be treated as either interval or nominal, depending on the number of categories.
- If a population distribution is approximately normal, samples of any size are appropriate for hypothesis tests of means. Since in practice we do not observe the population distribution and thus cannot know if it is normal, a rule of thumb is that the sample size must be at least thirty. For proportions, the required sample size depends on the hypothesized proportion in the population. The further the hypothesized proportion in the population is from .5, the larger the sample size needs to be.*

* A rule of thumb for required sample size for proportions is that the product of the sample size (N) and the null hypothesis proportion (\( \pi \)) and the product of N and (1 − \( \pi \)) must both be greater than or equal to ten. In practice, the value of \( \pi \) in the null hypothesis is rarely close to zero or one, and thus an N of thirty is sufficient. With very small samples, a special probability distribution called the binomial distribution must be used instead of the normal distribution; this is beyond the scope of this book.
Example: Testing a Claim about a Population Mean

Let’s look at one more example. Imagine that a marketer claims that consumers give a new app, on average, 9 out of 10 stars. But when you sample consumers, you find that your sample rates the app, on average, 7 stars. (Let’s say the sample is eighty-five people and the standard deviation is 3.8.) You might be skeptical, rightly, of the marketer’s claim. However, as a statistics student, you know that the results from a sample can vary from the population. A one-sample t-test allows you to determine the probability of getting a mean rating of 7 stars if the population really gave the app an average of 9 stars. Instead of comparing your sample to actual data about the population, you are comparing it to a claim about the population. You do this in exactly the same way as in the previous example, since you are still comparing an observed sample mean to a (claimed) population mean.

In order to conduct the hypothesis test, you need four pieces of information: the hypothesized (claimed) population mean (µ), the sample mean (\( \bar{y} \)), the sample standard deviation (s), and the size of the sample (N). In this case, µ = 9, \( \bar{y} = 7 \), s = 3.8, and N = 85. The steps for the hypothesis test are as follows:

**Step 1: Set up null and research hypotheses.**

Note that we are setting this up as a two-tailed test.

\[ H_0: \mu = 9. \] In other words, the mean rating for the entire population of consumers who have used the product is 9.

\[ H_a: \mu \neq 9. \] In other words, the mean rating for the entire population of consumers who have used the product is different from 9.

Our sample mean of 7 is two stars away from the hypothesized population mean of 9 stars. Remember, the hypothesis test is testing the likelihood of getting a sample result at least that far away from 9. You are setting up the t-curve as if the population mean is 9 in order to find the probability of obtaining a sample mean at least as far from the population mean as 7. If you find that the probability of getting your sample mean is low enough (i.e., less than your alpha-level), you will conclude that the claimed population mean is probably false.

Figure 9.6 shows this setup. Because it is a two-tailed test, we have drawn in the sample mean of 7, along with another value, 11, which is the same distance above the hypothesized population mean as 7 is below it.

**Step 2: Set alpha-level, calculate degrees of freedom, and find your decision t in the t-table.**

\[ DF = N - 1 = 85 - 1 = 84 \]

Looking in the t-table, we see that 84 is not listed, so we use the next lowest degrees of freedom, 80. For an alpha of .01 and a two-tailed test, decision \( t = 2.639 \).
**Step 3: Estimate the standard error.**

\[ SE_y = \frac{s}{\sqrt{N}} \]

\[ SE_y = \frac{3.8}{\sqrt{85}} = \frac{3.8}{9.22} = 0.41 \]

**Step 4: Calculate sample \( t \).**

\[ t = \frac{\bar{y} - \mu}{SE_y} = \frac{7 - 9}{0.41} = \frac{-2}{0.41} = -4.88 \]

\( t \) is a negative number because it corresponds to a score that is below the hypothesized population mean. A \( t \)-value of positive 4.88 would correspond to the score that is the same distance above the mean, in this case 11.

**Step 5: Compare your sample \( t \) and your decision \( t \) and draw your conclusion.**

If sample \( t \) is greater than decision \( t \), reject the null hypothesis. (If you are conducting a \( t \)-test using statistical software, the software will compute the exact \( p \)-value for your sample \( t \)-value. In that case, you reject the null hypothesis if \( p < \alpha \).)

In this example, the absolute value of sample \( t \) is 4.88, well beyond decision \( t \) of 2.639. (Remember, you compare the absolute value of your sample \( t \) to the decision \( t \).) We reject the null hypothesis. We have shown that there is less than a 1% chance of obtaining a sample mean 2 or more points away from the claimed mean rating of 9.
In other words, it is highly improbable that the marketer is telling the truth about customer ratings of the app. Customer ratings are statistically significantly lower than the claimed rating of 9 stars.

Optional follow-up: Calculate a confidence interval for the sample mean.

If we want to estimate the actual customer ratings for the app, we would construct a confidence interval, as we did in chapter 8. In this case:

\[ CI = \bar{y} \pm t(SE_{\bar{y}}), \text{ where } SE_{\bar{y}} = \frac{s}{\sqrt{N}} \text{ and } t \text{ at } DF = 84 \text{ and } 99\% \text{ confidence} = 2.639 \]

We calculated \( SE_{\bar{y}} = .41 \) above. Note that \( t \) is the value found in the t-table for your DF and confidence level, not your sample t. Plugging in the numbers, we find:

\[ CI = 7 \pm 2.639(0.41) = 7 \pm 1.08 \rightarrow (5.92, 8.08) \]

There is a 99% probability that the mean rating for the app is between 5.92 and 8.08 stars. Notice that the confidence interval does not contain the marketer’s claimed rating of 9 stars.

---

**Error and Limitations: How Do We Know We Are Correct?**

Hypothesis tests are widely used in the social sciences; their simplicity is appealing. They allow the investigator to reduce a finding to either significant (reject \( H_0 \)) or not significant (do not reject \( H_0 \)). But one should always keep in mind that a hypothesis test is built around a single sample statistic, and that sample statistic is not especially stable. It is highly likely that if we were to draw another random sample, we would get a different sample statistic. It would probably be fairly close to the first one we obtained but not exactly the same.

Recall the example from the beginning of the chapter. A sample of 200 young people found that 59% supported legalization of marijuana. We obtained a \( z \)-score of 1.43 and a p-value of .076 for that sample. Imagine that we took a second random sample of 200 young people, and this time we find that 56% of them support legalization. This is depicted in Figure 9.7.

The \( z \)-score for a sample proportion of .56 is .57, and the corresponding p-value is .284. What is our conclusion now? Whereas for our earlier example the p-value of .076 allowed us to reject the null hypothesis at an alpha of .10, our p-value of .284 is well above even an alpha of .10.

In this example, we would not reject the null hypothesis no matter what alpha-level we had chosen. Although in practice we draw only one sample and conduct the hypothesis test on the basis of that single sample statistic, we must always be aware
that a different random sample—which would likely yield a slightly different sample statistic—might well produce a different decision about the null hypothesis.

There is an important caveat to keep in mind. In some cases, as we saw in Table 9.1, we would reject the null hypothesis with a larger alpha (Scenario 1) but not with a smaller alpha (Scenarios 2 and 3). However, the determination of alpha must always precede the execution of the hypothesis test. The researcher must choose a proper rejection threshold, and should make this decision on its merits, before calculating the p-value associated with the sample proportion. This decision depends on an understanding of the possible errors we may make.

**Type I and Type II Errors**

The null hypothesis cannot be directly observed. With it, we are making an educated guess about the value of a population parameter. That guess is either correct or incorrect. We make our determination by looking directly at the evidence we have—the sample statistic and sample standard deviation. Is the evidence consistent with our guess about the population parameter? If so, we do not reject the null hypothesis. If the observed evidence is inconsistent with our guess about the population parameter, then we reject the null hypothesis.

Once we have made a decision, we have to be aware of the possibility that we may have made the wrong decision. Why might this be? Mistaken decisions are the consequence of sampling variability. Any time we draw a random sample from a population, it is possible that the sample is not representative of that population. This is not likely, but it is possible. There are two types of error, “false positives,” in which we wrongly
reject the null hypothesis and support the research hypothesis, and “false negatives,” in which we wrongly fail to reject the null hypothesis, wrongly concluding that the research hypothesis is false.

Let’s return to the legalization of marijuana example. Imagine that we (somehow) are omniscient, and we know for a fact that 54% of all young people support legalization. We draw a random sample of 200 young people. As we saw when we studied sampling distributions, we know that most of those samples will produce percentages close to 54%. But a small number of them will be well below or well above 54%. We can be precise. Ninety-five percent of those sample proportions will be within 1.96 standard errors of 54%, and 5% will be more than 1.96 standard errors away from the true population parameter. If we were unlucky enough to get one of those samples far away from the center, we would mistakenly conclude that the null hypothesis was incorrect. We summarize the possible outcomes in Table 9.3.

<table>
<thead>
<tr>
<th></th>
<th>Reject $H_0$</th>
<th>Do Not Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ is true</td>
<td>(A) Type I error</td>
<td>(B) Correct decision</td>
</tr>
<tr>
<td>$H_0$ is false</td>
<td>(C) Correct decision</td>
<td>(D) Type II error</td>
</tr>
</tbody>
</table>

Let’s examine Scenarios A, B, C, and D, one at a time, using our legalization of marijuana example. Recall our hypotheses:

$H_0$: 54% of all young people support legalization.

$H_a$: More than 54% of all young people support legalization.

**Scenario A**: The null hypothesis is true, but we have rejected it. This is called a **Type I error** (also known as a false positive). Why would we have rejected it? Because, although in reality 54% of all young people support legalization, we had the bad luck of drawing a random sample in which the percentage supporting legalization was well below or well above 54%. This can happen because of sampling variability.

**Scenario B**: Here, we have made the correct call. The null hypothesis is true—54% of all young people do support legalization, and we obtained a sample proportion that was close enough to 54%. Our observed evidence—the sample proportion—is close enough to the null hypothesis proportion. So we do not reject the null hypothesis.

**Scenario C**: Here again, we have made the correct decision. In reality, the null hypothesis is incorrect. In the context of our example, this means that the true percentage of all young people in support of legalization is not equal to 54%. And because we obtained a sample statistic far enough away from 54%, we reject the null hypothesis.
Scenario D: Here, the null hypothesis is false, but we fail to reject it. In this case, we have committed a **Type II error** (also known as a false negative). What does this mean in context? It means that, in the population, the true percentage of young people who support legalization is not 54%. Yet we ended up with an unrepresentative sample—one that indicated that the sample proportion was indeed close to 54%.

Any time we conduct a hypothesis test, we are potentially committing a Type I or a Type II error. This is because of two factors: (1) we do not ever directly observe the population parameter; we are drawing a sample to assess the accuracy of our hypothesis; and (2) sampling variability means that there is always a chance that we end up with a sample that is not representative of our population. As a result, we should be careful in how we word our conclusions. We cannot definitively “prove” that the null hypothesis is true or false. We can only conclude, within a range of probability, that our sample findings likely reflect a true difference from the population parameter.

We mentioned earlier that, when conducting a hypothesis test, the researcher should choose an alpha-level based on the merits of the case. One of the most important factors that contributes to the decision about alpha-level is the gravity associated with committing a Type I or Type II error, in the context of the particular situation. When the consequences of a Type I error—a false positive—are more serious than the consequences of a Type II error—a false negative—the researcher should choose a low alpha-level. In fact, the probability of committing a Type I error is equal to the alpha-level. If alpha is set at 1%, the researcher must obtain a p-value less than 1% in order to reject the null hypothesis. This also means that the probability of committing a Type I error is only 1%.

The probability of committing a Type I error is inversely related to the probability of committing a Type II error. Calculating the probability of a Type II error is more complicated and depends upon the distance between the true population parameter and the hypothesized one. These calculations are beyond the scope of this book. Even so, because the probabilities of Type I and Type II errors are inversely related, we can minimize the probability of a Type II error by raising alpha. When the null hypothesis is false, as it is in Scenarios C and D, we want a high probability of rejecting it. The probability of rejecting a false null hypothesis is known as the **power** of a test. The more “powerful” a test, the less likely we are to commit a Type II error.

In our legalization of marijuana example, as we discussed above, a Type I error means that 54% of all young people support legalization, but we wrongly conclude that the true percentage is higher. A Type II error means that the percentage of young people who support legalization is higher than 54%, but we wrongly conclude that it is 54%. In the first case, we are overestimating support for legalization. In the second case, we believe that the population percentage for young people is 54%, but it isn’t. It’s not clear which error is a more serious one. Indeed, when we are simply measuring public opinion, the consequences of either a Type I or Type II error are usually not all that grave.
On the other hand, there are many examples where the trade-off between the two kinds of errors is very important. Our entire system of criminal justice is based upon the idea of minimizing a Type I error. Consider this set of hypotheses:

\[ H_0: \text{Defendant is not guilty.} \]
\[ H_a: \text{Defendant is guilty.} \]

In this context, a Type I error sends a defendant to prison even though that defendant is innocent. A Type II error frees a guilty defendant. Most people would agree that it is far worse to convict an innocent person than to free a guilty one. This is why, to convict a defendant, a jury must determine guilt “beyond a reasonable doubt.” This is the threshold our legal system uses to minimize Type I errors.

There are many cases in social science research where a Type I or Type II error can have substantial implications. Consider the example of Moving to Opportunity from the beginning of the chapter. Part of that study tested the effectiveness of moving residents out of low-income neighborhoods on diabetes rates. Researchers hoped that moving residents would decrease diabetes rates significantly. Here is the set of hypotheses:

\[ H_0: \text{Diabetes rates for those who moved will be the same as those for all residents of poor neighborhoods.} \]
\[ H_a: \text{Diabetes rates for those who moved will be lower than those for all residents of poor neighborhoods.} \]

The researchers randomly select a sample of residents who moved and find that their diabetes rate is quite close to the overall rate. The researchers conduct a hypothesis test and fail to reject the null hypothesis. In other words, they conclude that moving has no effect on diabetes rate. If this decision is incorrect, the researchers would have committed a Type II error.

What does a Type II error mean in this context? It means that the null hypothesis is false; diabetes rates among those in the population who moved are actually lower than those of residents of poor neighborhoods in general. The researchers might erroneously conclude that it is not worth devoting resources to moving people out of poor neighborhoods because it does not produce an effect on this health outcome. This is an example where making a Type II error has real-world consequences: You have lost the opportunity to reduce rates of diabetes and thus improve people’s health.

What Does Statistical Significance Really Tell Us? Statistical and Practical Significance

Testing hypotheses about the relationship of sample findings to population parameters can be very useful. However, they tell us only one thing. They tell us the probability that we would obtain our sample result if the null hypothesis were true. While this
allows us to support or reject the research hypothesis, as we have discussed there is always a known probability that we may be making a Type I or Type II error. In addition, our hypothesis test only shows whether our findings are statistically significant. That phrase simply means that our test has rejected the null hypothesis at our chosen alpha-level. In statistical terms only, there are significant differences between the sample and the population (or the hypothesized population parameter). Such differences are not always important on a practical or substantive level. For example, think about the research on the effects of moving away from a high-poverty neighborhood. If those who moved are 1 or 2 percentage points less likely to have an adverse health outcome, that difference might be statistically significant, but is it a large enough difference to be substantively important? Or, if graduates of your college make, on average, $150 more per year than college graduates overall, is that enough of a difference to be considered in enrollment choices or to be touted in your school’s advertising? If an educational program in prison reduces recidivism by 1%, is that a meaningful reduction? What if it reduces recidivism by 10%? The question of practical or substantive significance is one that the researcher—and those who read research results—must determine for themselves.

Moreover, the determination of statistical significance is extremely sensitive to sample size. It is important to understand that, if a sample size is large enough, even small differences can be statistically significant. Small samples are not necessarily bad samples, but they make it harder to reject a null hypothesis, all other factors being equal. Many students of introductory statistics will use their education to be better informed consumers of information. After reading this chapter, you know that, when research is reported, you should consider both whether the results are statistically significant and whether they are meaningful in other ways.

In fact, many statisticians argue that researchers should not rely on hypothesis testing and p-values. Their view is that even if Type I errors are relatively rare, they still occur often enough that many research findings claiming statistical significance are in fact false positives. Further, hypothesis testing gives us only one of two answers: either the null hypothesis is rejected, or it is not. In order to know more about the actual variable we are interested in—such as college graduates’ earnings—we need to estimate those earnings using a confidence interval. This is a useful addition to hypothesis testing. Finally, statisticians remind us of the inherent uncertainty and variability that occur in sampling. Although we know in theory that repeated samples will provide different results, in practice we are usually working with one sample. There is no solution to this problem, except to be cautious in our conclusions and, where possible, to seek to replicate research findings that have serious potential consequences.
One-sample hypothesis tests compare sample statistics to population parameters or to claims about a population.

- The steps for testing a one-sample hypothesis about the difference between an obtained sample statistic and a given population parameter are similar for proportions and for means. They are:
  
  1. Set up null and alternative hypotheses.
  2. Set the alpha-level. If you are doing a hypothesis test for means (using t), calculate DF for your sample and look in the t-table to determine the decision t.
  3. Estimate the standard error.
     - For proportions: $SE_p = \sqrt{\frac{\pi(1-\pi)}{N}}$
     - For means: $SE_y = \frac{s}{\sqrt{N}}$
  4. Calculate the sample statistic, either z (for proportions) or t (for means).
     - For proportions: $z = \frac{p - \pi}{SE_p}$
     - For means: $t = \frac{\bar{y} - \mu}{SE_y}$
  5. For proportions, look in the normal table to find the p associated with your sample z. For means, compare the sample t with the decision t. Determine whether you can reject the null hypothesis. If the p-value is lower than alpha (for proportions) or if sample t is higher than decision t (for means), you can reject the null hypothesis. State your conclusion.

- To estimate the range within which your population proportion or mean is likely to fall, follow the hypothesis test by calculating a confidence interval using the steps summarized in chapter 8.

- In a one-tailed test, the alternative hypothesis states that the population parameter is above or below a specified value but not both; in a two-tailed test, the alternative hypothesis is that the population parameter is either above or below a specified value.

- The researcher sets the alpha-level ($\alpha$), which is the threshold used to determine whether to reject the null hypothesis. If the p-value is lower than $\alpha$, the null hypothesis is rejected.

- The difference between a sample statistic and a hypothesized population parameter in a one-sample test is statistically significant if the p-value associated with the sample statistic (z or t) is less than alpha, allowing the researcher to reject the null hypothesis.

- Statistical significance means that the difference between the sample statistic and the hypothesized population parameter is unlikely to be due to sampling error; it is more likely reflective of a genuine difference between the two.

- A Type I error occurs when the null hypothesis is true but it is rejected. The probability of a Type I error is equal to alpha.
A Type II error occurs when the null hypothesis is false but it is not rejected.

The power of a hypothesis test is the probability of rejecting a false null hypothesis.

In this section, you will learn how to conduct one-sample hypothesis tests for means and proportions in Stata. We will use data from the Police Public Contact Survey (PPCS), conducted by the Bureau of Justice Statistics, which interviews a nationally representative sample of U.S. residents older than the age of sixteen. Participants were asked whether they had contact with the police in the last year and, if so, questions about the nature of that contact. We will use the following variables from the PPCS:

- Contact measures whether respondents had face-to-face contact with the police. There are two possible values for contact: 1 indicates that the respondent had contact with the police, and 0 means that the respondent had no contact.
- Force measures whether respondents who experienced police contact reported that the police used force (or threatened to use force) in their most recent police contact. A value of 1 means that the police either used or threatened to use force, and 0 means that there was no force (actual or threatened) during the interaction.
- Time measures the duration of traffic stops (in minutes) for those respondents whose most recent contact with the police was a traffic stop.

In recent years, the Black Lives Matter movement has drawn national attention to the issue of police mistreatment of African Americans. The questions in this section will focus on whether the rate of contact with police, rate of use of force by police, and mean duration of traffic stops for African Americans in the population are equal to the respective parameters in the overall population of U.S. residents. For this section, we limit the PPCS sample to only African American respondents. We will set an alpha-level of .05 for all three hypothesis tests in this section.

**Hypothesis Test for a Proportion**

We begin by conducting a set of hypothesis tests for proportions. The first question uses the contact variable to investigate the likelihood that the proportion of African Americans in the population who experienced contact with the police is equal to the known police contact rate in the overall population, 16.5% (or, as a proportion, .165). Our null hypothesis can be stated as:

\[ H_0: \text{The population rate of police contact for African Americans is equal to 0.165.} \quad (H_1: \pi = 0.165) \]

---

*a* We use data from the 2008 wave of the survey, which means that responses reflect respondents’ experiences with the police during the twelve-month period from January to December 2007.

† We use the arrest rate in the full PPCS sample as an estimate of the population parameter.
We have three options for alternative hypotheses: a two-tailed test, a right-tailed test, and a left-tailed test. Since the dependent variable is any kind of contact with police, not punitive contact specifically, we do not have a reason to expect that African Americans would have specifically more or less contact with police than occurs in the general population. For example, it may be that rates of certain kinds of contact with police (e.g., calling for help) are higher in the general population than among African Americans. Thus, we will conduct a two-tailed test, with the following alternative hypothesis:

\[ H_a: \text{The population rate of police contact for African Americans is not equal to 0.165.} \quad (H_a: \pi \neq 0.165). \]

Open **PPCS_CH9.dta**. To run the test, type the following command into Stata:

```
prtest contact = .165
```

"prtest" tells Stata to run the test for a proportion.* Setting the value of `contact` equal to a specific value tells Stata that this is a one-sample test. The “prtest” command produces the output shown in Figure 9.8.

The output yields all of the information we need to make a decision about our hypothesis test. We see that the mean for `contact` is .139. For variables that are coded 0/1, like `contact`, the mean is equal to the proportion of cases with a value of 1. Thus, 13.9% of African Americans in the PPCS sample had contact with the police. This is lower than the percentage in the general population, 16.5%. The output also provides us with the test statistic \( z = -5.4150 \).

Stata provides us with the p-values for all three possible hypothesis tests. At the bottom of the output, the results for these three hypotheses are, from left to right, for the one-tailed test that the proportion is less than .165, for the two-tailed test that the proportion is not equal to .165, and for the one-tailed test that the proportion is greater than .165. It is your job to pay attention to the results for the alternative hypothesis that you constructed prior to the test. In this case, our alternative hypothesis stated that the population proportion would not be equal to .165, corresponding to the middle test, circled in the output. The p-value associated with the z-statistic for the two-tailed test

![Figure 9.8](image)

0/1, like `contact`, the mean is equal to the proportion of cases with a value of 1. Thus, 13.9% of African Americans in the PPCS sample had contact with the police. This is lower than the percentage in the general population, 16.5%. The output also provides us with the test statistic \( z = -5.42 \).

Stata provides us with the p-values for all three possible hypothesis tests. At the bottom of the output, the results for these three hypotheses are, from left to right, for the one-tailed test that the proportion is less than .165, for the two-tailed test that the proportion is not equal to .165, and for the one-tailed test that the proportion is greater than .165. It is your job to pay attention to the results for the alternative hypothesis that you constructed prior to the test. In this case, our alternative hypothesis stated that the population proportion would not be equal to .165, corresponding to the middle test, circled in the output. The p-value associated with the z-statistic for the two-tailed test

---

* Note that Stata will conduct hypothesis tests for proportions only for variables that have two categories, which must be coded as 0 and 1.
is .0000, lower than our alpha-level of .05. Thus, we reject the null hypothesis that the rate of contact with the police for African Americans is equal to 16.5%.

We can also examine the results for the remaining two alternative hypotheses. The result on the left is for the one-tailed test with the alternative hypothesis that the population proportion for African Americans is lower than .165. For this left-tailed hypothesis test, the p-value is below .05; the evidence suggests that the parameter for African Americans is likely lower than .165. The result on the right is for the other one-tailed test, with the alternative hypothesis stating that the rate of contact with the police for African Americans is greater than .165. The p-value for this test is 1.0, indicating that, if the rate of contact for African Americans in the population were .165, there would be a nearly 100% chance of obtaining a sample proportion of .139 or higher.

What if discriminatory conduct by the police occurs not in the frequency of overall contact—after all, contact with the police includes many different situations—but in the ways in which people are treated once contact is made? We can use the variable `force`, whether respondents who experienced police contact reported that the police had used or threatened to use force in their most recent police contact, to test this idea.

We can run a hypothesis test for a proportion in which we test the null hypothesis that the population proportion for African Americans is equal to the proportion of the overall population who reported the police using or threatening to use force, which is 1.3%, or .013.* In this case, the null hypothesis is:

\[
H_0: \text{The proportion of African Americans in the population that experienced the use of force in their most recent police interaction is equal to } 0.013 (H_0: \pi = 0.013).
\]

As always, we have three options for our alternative hypothesis. Here, we predict that the population rate of encountering force in police interactions for African Americans is higher than 0.013:

\[
H_a: \pi > 0.013
\]

The syntax for the test is:

`prtest force = .013`

The output is shown in Figure 9.9.

This time, notice that the mean for `force` in the sample of African Americans is .029, higher than our null hypothesis value of .013, but is that difference statistically significant? (Remember, the mean is equal to the proportion of cases with a value of 1.) Our alternative hypothesis is that the population parameter is higher than the overall proportion of U.S. residents reporting the use of force by police ($H_a: \pi > .013$). In this case the p-value reported for that $H_a$ is .0001. This is lower than our alpha value of .05, indicating that, if the null hypothesis were true, we would be very unlikely to have observed a sample proportion of .029. In this case, there is evidence that the proportion of African Americans in the population who experience the use of force (or threat of force) by police is greater than in the overall population.

* We use the rate of force in the full PPCS sample as an estimate of the population parameter.
Hypothesis Test for a Mean

Next, we will conduct a one-sample hypothesis test for a mean, using the *time* variable, which measures how many minutes respondents’ most recent contact with police lasted.

We will test whether the mean duration of traffic stops for African Americans in the population is the same as the mean duration of traffic stops for the overall population, 11.83 minutes.* We use that value to generate our null hypothesis:

$$H_0: \text{The mean number of minutes that traffic stops last for African Americans in the population is equal to } 11.83 (H_0: \mu_y = 11.83).$$

The syntax for a test for means is similar to that for the test for proportions. Instead of asking for a “prtest,” we ask for a “ttest.” We specify in the syntax that we will be using the t-distribution (rather than the z-distribution) because we do not know the population standard deviation:

```stata
ttest time = 11.83
```

The output from the test is shown in Figure 9.10.

We learn from the output that the mean for the African American sample is 14.22, higher than the population mean from the null hypothesis, 11.83. But are we convinced by the results of our hypothesis test that the population mean for African Americans is significantly higher than 11.83? Our alternative hypothesis is that the population mean for African Americans is higher than 11.83, and the results are in the bottom-right corner of the output. The sample t, 3.3, is associated with a very low p-value, smaller than our alpha-level of .05. This indicates that we can reject the null hypothesis in favor of the alternative hypothesis. We can conclude that the mean duration of traffic stops for African Americans in the population is significantly higher than it is in the general population.

Overall, the results of these three hypothesis tests suggest that rate of police contact in the population of African Americans is actually statistically significantly lower

* We use the mean duration of traffic stops for the full PPCS sample as an estimate of the population parameter.
than in the general population. However, the results indicate that the rate of force and duration of contact in the African American population are statistically significantly higher than in the general population. This could suggest that discriminatory actions against African Americans by police occur less in the frequency of contact than in what occurs during that contact. In the next chapter, we will learn how to make explicit comparisons between the means and proportions for two groups by using two-sample hypothesis tests.

Review of Stata Commands

- Run a one-sample test for a proportion (0/1 variable only)
  \[ \text{prtest } \text{variable name}=\text{value} \]

- Run a one-sample test for a mean, standard deviation unknown
  \[ \text{ttest } \text{variable name}=\text{value} \]

In this section, you will learn how to conduct one-sample hypothesis tests for means and proportions in SPSS. We will use data from the Police Public Contact Survey (PPCS), conducted by the Bureau of Justice Statistics, which interviews a nationally representative sample of U.S. residents older than the age of sixteen. *Participants were asked whether they had contact with the police in the last year and, if so, questions about the nature of that contact. We will use the following variables from the PPCS:

- Contact measures whether respondents had face-to-face contact with the police. There are two possible values for contact: 1 indicates that the respondent had contact with the police, and 0 means that the respondent had no contact.
- Force measures whether respondents who experienced police contact reported that the police used force (or threatened to use force) in their most recent contact.

*We use data from the 2008 wave of the survey, which means that responses reflect respondents’ experiences with the police during the twelve-month period from January to December 2007.*
police contact. A value of 1 means that the police either used or threatened to use force, and 0 means that there was no force (actual or threatened) during the interaction.

- Time measures the duration of traffic stops (in minutes) for those respondents whose most recent contact with the police was a traffic stop.

In recent years, the Black Lives Matter movement has drawn national attention to the issue of police mistreatment of African Americans. The questions in this section will focus on whether the rate of contact with police, rate of use of force by police, and mean duration of traffic stops for African Americans in the population are equal to the respective parameters in the overall population of U.S. residents. For this section, we limit the PPCS sample to only African American respondents.

**Hypothesis Test for a Proportion**

We will begin by conducting a set of hypothesis tests for proportions. The first question uses the contact variable to investigate the likelihood that the proportion of African Americans in the population who experienced contact with the police is equal to the known police contact rate in the overall population, 16.5% (or, as a proportion, .165). Our null hypothesis can be stated as:

\[ H_0: \text{The population rate of police contact for African Americans is equal to 0.165} \ (H_0: \pi = 0.165). \]

We have three options for alternative hypotheses: a two-tailed test, a right-tailed test, and a left-tailed test. Since the dependent variable is any kind of contact with police, not punitive contact specifically, we do not have a reason to expect that African Americans would have specifically more or less contact with police than occurs in the general population. For example, it may be that rates of certain kinds of contact with police (e.g., calling for help) are higher in the general population than among African Americans. Thus, we will conduct a two-tailed test, with the following alternative hypothesis:

\[ H_a: \text{The population rate of police contact for African Americans is not equal to 0.165} \ (H_a: \pi \neq 0.165). \]

SPSS includes hypothesis tests only for means, not proportions. But since we are using a 0/1 variable, the proportion of cases scoring a 1 is equal to the mean of the variable. So, we can use the SPSS procedure that generates a hypothesis test on a mean. We will use an alpha-level of .05 for all hypothesis tests conducted in this section.

Open **PPCS_CH9.sav**. To run the hypothesis test, we use this SPSS procedure:

**Analyze → Compare Means → One Sample T Test**

* We use the arrest rate in the full PPCS sample as an estimate of the population parameter.
This opens the “One Sample T Test” dialog box. We move the contact variable into the “Test Variable(s)” box, and in the space next to “Test Value” we type in the null hypothesis value; in this case, .165 (circled), as shown in Figure 9.11.

![One-Sample T Test dialog box](image)

Figure 9.11

When we click on “OK,” SPSS produces the output shown in Figure 9.12.

![One-Sample Statistics table](image)

Figure 9.12

The output yields all of the information we need to make a decision about our hypothesis test. We see that the mean for contact is .14. For variables that are coded 0/1, like contact, the mean is equal to the proportion of cases with a value of 1. Thus, 14% of African Americans in the PPCS sample had contact with the police. This is lower than the percentage in the population, 16.5%. The output also provides us with the test statistic t: 5.82.*

SPSS provides us with the p-value only for the two-tailed hypothesis test. As we can see from the output, the p-value is .000 (at three decimal points). In other words, we

* SPSS automatically uses t for all hypothesis tests.
are highly confident that the proportion of African Americans who have had contact with the police is not equal to .165 (our null hypothesis). If we wish to conduct a one-tailed hypothesis test to determine whether the contact rate for African Americans is lower than it is for the full population (H₀: \( \pi < .165 \)), the p-value would be equal to the two-tailed p-value generated by SPSS, divided by 2. For this one-tailed hypothesis test, the p-value here would still be 0.

But what if discriminatory conduct by the police occurs not in the frequency of overall contact but in the ways in which people are treated once contact is made? We can use the variable force, whether respondents who experienced police contact reported that the police had used or threatened to use force in their most recent police contact, to test this idea.

Again, we can run a hypothesis test for a proportion in which we test the null hypothesis that the population proportion for African Americans is equal to the proportion of the overall population who reported the police using or threatening to use force, which is 1.3%, or .013. In this case, the null hypothesis is:

\[ H₀: \text{The proportion of African Americans in the population that experienced the use of force in their most recent police interaction is equal to 0.013 (H₀: } \pi = 0.013) \]

As always, we have three options for our alternative hypothesis. Here, we predict that the population rate of encountering force in police interactions for African Americans is higher than .013:

\[ Hₐ: \pi > 0.013 \]

We use the same procedures we did for the contact variable. The output is shown in Figure 9.13.

**Figure 9.13**

This time, notice that the mean for force in the sample of African Americans is .03, higher than our null hypothesis value of .013, but is that difference statistically significant? (Remember, the mean is equal to the proportion of cases with a value of 1.)

* We use the rate of force in the full PPCS sample as an estimate of the population parameter.
Our alternative hypothesis is that the population parameter is higher than the overall proportion of U.S. residents reporting the use of force by police ($H_a: \pi > .013$). In this case the two-tailed p-value reported is .008; the one-tailed p-value is equal to .004. This is lower than our alpha value of .05, indicating that, if the null hypothesis were true, we would be very unlikely to have observed a sample proportion of .03 or higher. In this case, there is evidence that the proportion of African Americans in the population who experience the use of force (or threat of force) by police is greater than in the overall population.

### Hypothesis Test for a Mean

Next, we will conduct a one-sample hypothesis test for a mean, using the *time* variable, which measures how many minutes respondents’ most recent contact with police lasted.

We will test whether the mean duration of traffic stops for African Americans in the population is the same as the mean duration of traffic stops for the overall population, 11.83 minutes. We use that value to generate our null hypothesis:

$$H_0: \text{The mean number of minutes that traffic stops last for African Americans in the population is equal to 11.83 (} H_0: \mu = 11.83).$$

The procedure we use for a test for means is the same as the one we used for the test of proportions.

**Analyze → Compare Means → One Sample T Test**

We indicate the variable we are using as well as the null hypothesis test value (11.83), as shown in Figure 9.14.

![Figure 9.14](image.png)

*We use the mean duration of traffic stops for the full PPCS sample as an estimate of the population parameter.*
The output is shown in Figure 9.15.

![One-Sample Statistics Table]

![One-Sample Test Table]

**Figure 9.15**

We learn from the output that the mean for the African American sample is 14.22, higher than the population mean from the null hypothesis, 11.83. But are we convinced by the results of our hypothesis test that the population mean for African Americans is significantly higher than 11.83? Our alternative hypothesis is that the population mean for African Americans is higher than 11.83, and the two-tailed p-value is contained in the output. The sample t, 3.3, is associated with a very low two-tailed p-value, .001. (We divide by 2 to obtain the one-tailed p-value, .0005.) This indicates that we can reject the null hypothesis in favor of the alternative hypothesis. We can conclude that the mean duration of traffic stops for African Americans in the population is significantly higher than it is in the general population.

Overall, the results of these three hypothesis tests suggest that rate of police contact in the population of African Americans is actually statistically significantly lower than in the general population. However, the results indicate that the rate of force and duration of contact in the African American population are statistically significantly higher than in the general population. This could suggest that discriminatory actions against African Americans by police occur less in the frequency of contact than in what occurs during that contact. In the next chapter, we will learn how to make explicit comparisons between the means and proportions for two groups by using two-sample hypothesis tests.

**Review of SPSS Procedures**

- Run a one-sample test for a proportion (0/1 variable only)
  
  Analyze → Compare Means → One Sample T Test

- Run a one-sample test for a mean
  
  Analyze → Compare Means → One Sample T Test
1. Prior to 2000, the U.S. Census did not allow people to choose more than one racial identity. Since 2000, people have been able to mark more than one box on the Census question about race. Since then, demographers have used Census data to study the multiracial population, as measured by people who mark more than one box on the Census race question. The 2010 Census showed that 2.9% of the entire U.S. population identified as multiracial. In a random sample of 350 people drawn from the western region of the United States, 4.5% identified as multiracial.

   a. State the null and alternative hypotheses to test whether the multiracial population in the West is larger than the overall rate in the United States.

   b. Explain whether the test for Part a would be a one- or two-tailed test.

2. Continuing with the example given in Problem 1, conduct a hypothesis test for proportions to assess whether the percentage of people identifying as multiracial in the West is equal to the overall population rate. Be sure to specify an alpha-level.

3. Researchers find that a preventative health care program implemented in Anytown reduces emergency room visits to a mean of .50 per household per year. You sample 105 low-income households in Anytown and find that their mean number of visits to an emergency room in the past year was .65, with a standard deviation of .41. You want to assess whether the preventative health care program is working as well for low-income residents as it is for residents overall.

   a. Conduct each of the steps of a hypothesis test (at an alpha-level of .05) for this question.

   b. Discuss the practical significance of the findings.

   c. Estimate a 95% confidence interval for the annual number of emergency room visits for low-income households. Explain what this interval means.

4. A researcher in California is investigating how gender is related to school discipline. She is studying a random sample of 500 students from a large high school in California with more than 5,000 students. In her sample, 83% of the students who have been suspended from school are boys. This is higher than for the state of California as a whole, where 79% of students who have been suspended are boys, according to the Department of Education’s Civil Rights Data Collection. After learning this, the principal of the school wants to know: Does this school really have a larger gender imbalance in its suspension rates than in California as a whole? The principal strongly believes that there is no reason to suspect that her high school is any harsher on boys than any other school in the state.

   a. Conduct a hypothesis test at the .01 alpha-level. Show all of the steps. In setting up the null and alternative hypotheses, keep in mind that the principal does not believe her school punishes boys more harshly than other schools do. Explain whether you can reject the null hypothesis.
b. How would you explain the p-value from the hypothesis test to the principal, who is unfamiliar with statistical analysis?

c. The principal reports to her staff that “there is only a 2.6% chance that the gender composition of the suspended population at their school is the same as it is across the state.” Is this correct? Explain your answer.

5. Virtually all politicians now use social media as a way of communicating with voters, and increasing numbers of people use social media as a way to engage in politics. A random sample of 300 social media users scores a mean of 6 on a political engagement scale that ranges from not at all politically active, 0, to extremely politically active, 10. Research suggests that the overall population score for political engagement is 4. Does this difference of 2 points on the political engagement scale indicate a real difference between social media users and the general population? A one-tailed t-test yields a t-statistic that is larger than the decision t at an alpha of .01. Therefore, we reject the null hypothesis, which states that the mean score for the population of social media users is equal to 4.

a. Which kind of error, Type I or Type II, is it possible that we have made in this example? Explain why.

b. What is the probability that we have committed the kind of error that you identified in Part a?

6. A state passes a law prohibiting advertising targeting children (e.g., the use of cartoon characters in television ads to sell cereal). After the law’s first year of implementation, a group of researchers wants to know if children have adopted less materialistic attitudes. They cannot survey all of the state’s children, so they draw a probability sample of 1,000 children and measure their materialistic attitudes on a scale ranging from 0 to 100. The sample mean is 68, with a standard deviation of 10. A meta-analysis of previous studies examining children’s materialistic attitudes suggests that American children overall have a mean score of 70 on the scale.

a. Conduct a hypothesis test at the .01 alpha-level to assess whether children from the state that has banned children’s advertising have the same mean score for materialistic attitudes as the nationwide mean.

b. Draw a t-distribution. Label the decision and sample t values on the x axis. Shade the appropriate area on the curve and label it.

c. Why is the sample t a negative value? Why is the decision t a negative value?

d. What would you conclude about the success of the elimination of advertising to children at reducing their materialistic values? Consider the statistical and practical significance of the result in your response.

e. In this example, which type of error—Type I or Type II—could we have made? Explain how you know and what the error would mean. Assess the consequences of making that particular error.
7. As the size of the gig economy (i.e., work arrangements that depart from traditional job arrangements and rely more on short-term contracts) increases, observers have debated whether this represents a turn toward more freedom and autonomy for workers or a dangerous loss of job security. A recent study estimates that 60% of adults see the growth of the gig economy as more negative than positive. Three colleagues want to know if gig economy workers themselves are just as likely as everyone else to see the gig economy as generally negative. They intend to conduct a one-sample hypothesis test to determine the answer to this question.

a. The null hypothesis for the test is that the proportion of gig economy workers who feel negatively about the gig economy is equal to the proportion of people overall who feel that way. But, the three colleagues disagree on what the alternative hypothesis should be. They find that their differences are irreconcilable. They go their own ways and conduct their own tests. The first colleague sets up the alternative hypothesis like this: \( H_0: \pi \neq .6 \). The second opts for \( H_0: \pi > .6 \). The third colleague prefers \( H_0: \pi < .6 \). Explain in words what each of these alternative hypotheses means.

b. Which of the three alternative hypotheses do you think makes the most sense? Explain your answer.

c. Use the normal table to find the z-score associated with an alpha-value of .05 for each of the three alternative hypotheses. Explain why the z-scores differ between the one-tailed tests. Explain why the z-score is different for the one-tailed tests compared to the two-tailed test.

8. Consider the following two hypotheses:

\[ H_0: \text{Casual users of e-cigarettes have no more health problems than the general population.} \]

\[ H_a: \text{Casual users of e-cigarettes have more health problems than the general population.} \]

a. Explain the consequences of making a Type I error in this case? A Type II error?

b. Which type of error do you think is a more serious one to make in this case? Explain your answer.

9. A marketing firm believes that its advertising campaign has increased the favorability of its client’s product among the coveted eighteen to thirty-four demographic. Using a probability sample of eighteen- to thirty-four-year-olds, the marketing firm found that 75% of them said that they liked the product. This compares to the less promising 50% favorability rating from this population in extensive previous market research. The firm conducts a one-sample test for proportions, with the null hypothesis being that the proportion of all eighteen- to thirty-four-year-olds who like the product is equal to .5. The test yields a p-value of .001. The firm declares victory! Based on the p-value, it
reports to its client that it is 99.9% sure that 75% of all eighteen- to thirty-four-
year-olds like the product after the ad campaign.

a. The firm has bungled its interpretation of the p-value. Explain why its inter-
pretation is incorrect. State in words what we can say about the p-value,
given the information provided above.

b. We learn later that the firm set up a two-tailed alternative hypothesis in this
analysis. Do you agree with that decision? Explain why.

c. If the firm had set up a one-tailed alternative hypothesis, which should it
have chosen: left- or right-tailed?

d. Would the choice of a one-tailed hypothesis test over a two-tailed test have
changed the general findings in this case? Explain your answer.

10. Chemists use litmus paper to determine whether solutions are basic or acidic.
Blue litmus paper turns red if the solution is acidic, and red litmus paper turns
blue if the solution is basic. Thus, the phrase “litmus test” generally means that
the result of the test adjudicates between two possible outcomes that are mu-
tually exclusive.

a. A litmus test decides between two mutually exclusive answers. Does a hy-
pothesis test do the same thing? Explain your answer.

b. A litmus test provides a definitive answer: Either the first or second possi-
ble answer is certainly correct. Does a hypothesis test do the same thing?
Explain your answer.

11. The cosmetics industry has been notorious for making products geared toward
people with lighter skin tones, excluding people with darker skin tones from
both marketing and product development. In September 2017 the pop star
Rihanna released an inclusive makeup line called Fenty, featuring forty shades
of foundation. She was convinced that there was a big market for makeup de-
signed for people with darker skin tones. Rihanna’s team wanted to know if
Fenty had led more people of color to shop at a high-end makeup retailer. The
team knew that, before Fenty was introduced, people of color made up just 6%
of the store’s clientele. A year after its launch, they drew a random sample of
one hundred of the store’s shoppers and found that eight people in the sample
identified as people of color.

a. Why shouldn’t Rihanna’s team make a conclusion about Fenty’s impact on
the retailer’s clientele based on the group of one hundred shoppers?

b. Set up the null and alternative hypotheses for a hypothesis test in this case.
State the hypotheses in words and statistical notation.

c. The team produces the sampling distribution shown in Figure 9.16 in the
process of conducting the hypothesis test, labeling five pieces of key in-
formation in the figure. Explain what each of the five labels in the figure
means. (In the case of figures that were arrived at by calculation, show how
they calculated the figure.)
d. Use the information provided on the curve in Figure 9.16 to state whether the null hypothesis should be rejected.

e. Based on the outcome of the hypothesis test, would you conclude that Fenty had definitively diversified, or not diversified, the store’s clientele? Explain why or why not you would draw a definitive conclusion.

12. During his presidential campaign, one candidate gave hundreds of speeches in which he mentioned the unemployment rate, promising to fix the “broken” economy if elected. One reporter noticed that the rates the candidate cited tended to be higher than the official unemployment rate released by the Federal Reserve, 4%. The reporter didn’t have the resources to analyze the hundreds of speeches that the candidate had delivered on the campaign trail, so she collaborated with a political science professor at a nearby university on analyzing a sample of the candidate’s speeches. After drawing a random sample of thirty speeches, they found that the unemployment rate stated by the candidate in the sample was, on average, 12%, with a standard deviation of 2%. They set up a hypothesis test to calculate the chance that sampling variability explained why this figure was so much higher than the Federal Reserve’s.

a. Which type of hypothesis test should they conduct: a test for a mean or proportion? Explain why.

b. Conduct the appropriate test, with an alpha-level of .05. Is the result of the test statistically significant? Explain why or why not.

c. Based on the results of the test, what is the probability that the sample of the candidate’s speeches is actually indicative of all of his campaign speeches in terms of the difference between the unemployment rates that he cites and the rate given by the Federal Reserve?

d. Do you find the difference between the candidate’s mean cited unemployment rate, 12%, and the one given by the Federal Reserve, 4%, to have practical significance? Explain your answer.
13. A researcher is investigating whether the cost of rental housing in San Francisco is the same as the cost of rental housing in the San Francisco Bay Area overall. The mean monthly rent for a representative sample of 1,000 rental units in San Francisco is $3,500, with a standard deviation of $250. Official statistics from the state indicate that monthly rent for the whole Bay Area averages $3,050. The null hypothesis is that the population mean for San Francisco’s monthly rents is $3,050, the same as in the Bay Area overall. The researcher conducts a one-tailed hypothesis test and fails to reject the null hypothesis. What alpha-level is the researcher using? Use the sample t-value to explain your answer.

**Stata Problems**

Open the World Values Survey (WVS) data file that includes data only for Chinese respondents, called `WVS_China.sav`. We will work with the same scitech variable that we saw in the practice problems for chapter 3 to ask whether people who live in China have the same feelings about science and technology as people across the globe. Scitech ranges from 1 (the world is a lot worse off because of science and technology) to 10 (the world is a lot better off because of science and technology). You will use the scitech mean from the full WVS data set, 7.26, as an estimate of the total population mean.

1. Use the “ttest” command to run the hypothesis test examining whether attitudes about science and technology for all people in China are the same as attitudes for people around the world.
2. Find the sample mean for Chinese respondents in the output.
3. What is the null hypothesis? What is the two-tailed alternative hypothesis?
4. What is the sample t-value?
5. Interpret the output for the two-tailed test. Does the evidence suggest that residents of China feel the same as the rest of the world about science and technology?

**SPSS Problems**

Open the World Values Survey (WVS) data file that includes data only for Chinese respondents, called `WVS_China.sav`. We will work with the same scitech variable that we saw in the practice problems for chapter 3 to ask whether people who live in China have the same feelings about science and technology as people across the globe. Scitech ranges from 1 (the world is a lot worse off because of science and technology) to 10 (the world is a lot better off because of science and technology). You will use the scitech mean from the full WVS data set, 7.26, as an estimate of the total population mean.

1. Use the “One Sample T Test” dialog box to run the hypothesis test examining whether attitudes about science and technology for all people in China are the same as attitudes for people around the world.
2. Find the sample mean for Chinese respondents in the output.
3. What is the null hypothesis? What is the two-tailed alternative hypothesis?

4. What is the sample t-value?

5. Interpret the output for the two-tailed test. Does the evidence suggest that residents of China feel the same as the rest of the world about science and technology?

---


