

Unit 20

Hypotheses Testing

Objectives:

- To understand how to formulate a null hypothesis and an alternative hypothesis about a population proportion, and how to choose a significance level
- To understand how to collect data and calculate the value of an appropriate test statistic in a hypothesis test about a population proportion
- To understand how to define a rejection region and obtain the p -value in a hypothesis test about a population proportion
- To understand how to summarize the results of a hypothesis test about a population proportion

We have introduced several terms and concepts important in hypothesis testing. We now want to define a hypothesis test formally and begin to apply hypothesis testing in a variety of different situations. We can formally define a hypothesis test as the following four steps:

- Step 1: State the null and alternative hypotheses, and choose a significance level.
- Step 2: Collect data, and calculate the value of an appropriate test statistic.
- Step 3: Define the rejection region, decide whether or not to reject the null hypothesis, and obtain the p -value (probability value) of the test.
- Step 4: State the results, and perform any further analysis which may be required.

We shall begin to illustrate the application of these four steps by considering hypothesis tests about a population proportion.

When we first defined the null hypothesis and the alternative hypothesis, we considered a hypothesis test to decide whether or not to believe the claim of a manufacturer that a lighter will ignite on the first try 75% of the time. Recall that a hypothesis test is very much analogous to a court trial, where we assume that the null hypothesis, which states "The defendant is innocent," is true, until and unless there is sufficient evidence to believe the alternative hypothesis, which states "The defendant is guilty." In a hypothesis test concerning a population parameter, a null hypothesis is a statement about equality, and an alternative hypothesis is a statement about inequality. In a hypothesis test about a population proportion, the statement of the null hypothesis contains a hypothesized value for the population proportion, and the alternative hypothesis is a statement that the hypothesized value is not correct. One of the reasons the null hypothesis is so named is because the meaning of the word *null* suggests no difference, no change, no effect, etc. By imagining that the manufacturer's claim is on trial, our null hypothesis is "The lighter will ignite on the first try 75% of the time," and the alternative hypothesis is "The lighter will not ignite on the first try 75% of the time."

We shall find it convenient to introduce some widely-used abbreviations in hypothesis testing. First, we use H_0 to represent a null hypothesis and H_1 to represent an alternative hypothesis. Also, when writing a hypothesis, we may choose to use symbols in place of some words. The population proportion λ is the focus of the hypothesis test concerning the lighter, and the hypothesized value for λ is 0.75. A much shorter way of stating the null hypothesis "The lighter will ignite on the first try 75% of the time" is to simply say " $\lambda = 0.75$," and a much shorter way of stating the alternative hypothesis "The lighter will not ignite on the first try 75% of the time" is to simply say " $\lambda \neq 0.75$." As part of the first step of our hypothesis test, we need to choose a significance level; we shall choose $\alpha = 0.10$, which is what we originally chose when we first considered the illustration concerning the lighter. This enables us to complete the first step of our hypotheses test by writing

$$H_0: \lambda = 0.75 \quad \text{vs.} \quad H_1: \lambda \neq 0.75 \quad (\alpha = 0.10) .$$

Having now completed the first step in our hypothesis test, we move on to the second step. The second step in a hypothesis test is to collect data and to calculate the value of an appropriate test statistic on which our conclusion will be based. In a court trial, this second step corresponds to collecting and presenting evidence. In

the hypothesis test concerning the lighter, this second step corresponds to obtaining a simple random sample of \bar{p} observations and basing our conclusions on a statistic that we obtain from our sample; the sample proportion \bar{p} of times the lighter will ignite on the first try certainly seems a reasonable statistic on which to base our conclusions about the lighter.

In hypothesis testing, a *test statistic* is a statistic which is used to decide whether to believe the H_0 or to believe the H_1 . Remember that in hypothesis testing, we assume the null hypothesis to be true unless the evidence is sufficiently strong to suggest that we should believe the H_1 . Deciding whether or not the evidence in a court trial is sufficiently strong to return a verdict of guilty is ultimately a subjective decision, since there is no mathematical formula which will weigh all the evidence presented in a typical court trial. In hypothesis testing, however, deciding whether or not the evidence is sufficiently strong to believe the H_1 is based on doing an appropriate probability calculation. Figures 19-1 and 19-2 are illustrations of the probability calculation we did with each of two different data sets in order to decide whether or not the evidence in the data was strong enough to convince us to believe the H_1 .

The probability calculations illustrated in Figures 19-1 and 19-2 were done under the assumption that $H_0: \lambda = 0.75$ is true. Under this assumption, with a sufficiently large sample size n , we are able to treat the sampling distribution of \bar{p} as a normal distribution. The z -score of \bar{p} , which is

$$z = \frac{\bar{p} - \mu_{\bar{p}}}{\sigma_{\bar{p}}} = \frac{\bar{p} - \lambda}{\sqrt{\frac{\lambda(1-\lambda)}{n}}} = \frac{\bar{p} - 0.75}{\sqrt{\frac{0.75(1-0.75)}{n}}}$$

tells us how many standard deviations of difference there is between \bar{p} and the hypothesized 0.75. This is exactly the z -score we needed to calculate to do the probability calculations illustrated in Figures 19-1 and 19-2. In general, whenever we have a null hypothesis $H_0: \lambda = \lambda_0$, where we use λ_0 to represent the hypothesized value for λ , the test statistic is

$$z = \frac{\bar{p} - \lambda_0}{\sqrt{\frac{\lambda_0(1-\lambda_0)}{n}}}$$

which we shall refer to as the z statistic about a population proportion.

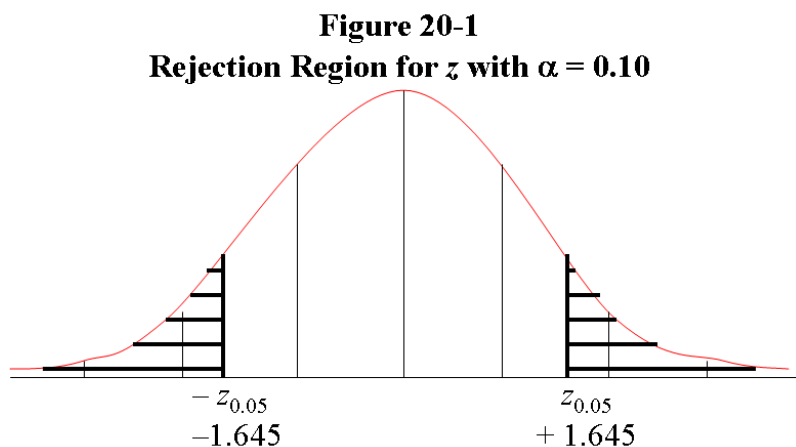
To complete the second step in our hypothesis test about the lighter, we must collect data, and calculate the value of our test statistic. To collect data, we observe a simple random sample of attempts at igniting the lighter. Let us imagine that we observe a simple random sample of $n = 300$ attempts, and we find that the lighter ignites on the first try in 214 of these attempts (which is exactly the same data we used in Figure 19-1). The sample proportion of successes is $\bar{p} = 214/300 = 0.7133$, from which we calculate the value of our z statistic as follows:

$$z = \frac{\bar{p} - \lambda_0}{\sqrt{\frac{\lambda_0(1-\lambda_0)}{n}}} = \frac{0.7133 - 0.75}{\sqrt{\frac{0.75(1-0.75)}{300}}} = -1.468$$

The third step in a hypothesis test is where we decide whether to believe H_0 or H_1 . In our hypothesis test concerning the lighter, we are basing this decision on whether the number of standard deviations of difference between \bar{p} and the hypothesized 0.75, as measured by the z statistic, is beyond what we would expect from sampling error. Our chosen significance level α determines what values of the z statistic are beyond what we should expect from sampling error. In our hypothesis test about the lighter, we chose $\alpha = 0.10$, which implies that values of the z statistic which have less than a 0.10 probability of occurring (in either direction) provide us with sufficient evidence to believe H_1 .

Figure 20-1 displays a standard normal curve with two shaded areas: one containing the lower 5% (0.05) of the area under the curve and one containing the upper 5% (0.05) of the area under the curve. Together, these areas define a set of values for the z statistic which has a 0.10 (0.05+0.05) probability of occurring if $H_0: \lambda = 0.75$ is true. Notice that in Figure 20-1 the z -score above which 0.05 of the area lies has been designated by $z_{0.05}$; also, since the standard normal curve is symmetric, we know that the z -score below which 0.05 of the area lies must be $-z_{0.05}$ (the negative of $z_{0.05}$). If our z statistic is either larger than $z_{0.05}$ or smaller than $-z_{0.05}$, then we have sufficient evidence to believe that $H_1: \lambda \neq 0.75$ is correct. In hypothesis testing, a *rejection region* (sometimes also called a *critical region*) is a set of test statistic values which lead to rejecting the null hypothesis in favor of the alternative hypothesis. In order for us to define our rejection region in the hypothesis test about the lighter, we must find what the value of $z_{0.05}$ is.

From Table A.2, you will find that the area under the standard normal curve above the z -score +1.64 is 0.0505, and that the area under the standard normal curve above the z -score +1.65 is 0.0495. From this, we deduce that the z -score above which lies 0.05 of the area must be between +1.64 and +1.65. At the bottom of Table B.2, you will find that $z_{0.05} = 1.645$. It is convenient to use the notation z_A to represent the z -score above which lies whatever area A is the subscript. For instance, you will find at the bottom of Table A.2 that $z_{0.025} = 1.960$ and $z_{0.005} = 2.576$, which tell us that 0.025 of the area lies above the z -score +1.960 and 0.005 of the area lies above the z -score +2.576. In general, the rejection for the hypothesis test



$$H_0: \lambda = \lambda_0 \quad \text{vs.} \quad H_1: \lambda \neq \lambda_0$$

with significance level α will be defined by

$$z \geq z_{\alpha/2} \quad \text{or} \quad z \leq -z_{\alpha/2} \quad (\text{i.e., } |z| \geq z_{\alpha/2}).$$

Figure 20-1 graphically displays the rejection region in our hypothesis test about the lighter, and we can define this rejection region algebraically as

$$z \geq +1.645 \quad \text{or} \quad z \leq -1.645 \quad (\text{i.e., } |z| \geq 1.645).$$

If the value of the z statistic is in the rejection region, then we would consider our data unlikely enough to make us doubt that $H_0: \lambda = 0.75$ is true; however, if the value of the z statistic is not in the rejection region, then we have no reason to doubt that $H_0: \lambda = 0.75$ is true.

Since the H_0 is assumed to be true unless sufficient evidence is found against it, the custom in statistical terminology has been to state the decision in terms of the H_0 . If we decide that there is sufficient evidence against the H_0 , we say that we reject the H_0 , and, of course, saying that we reject H_0 is equivalent to saying that we accept H_1 .

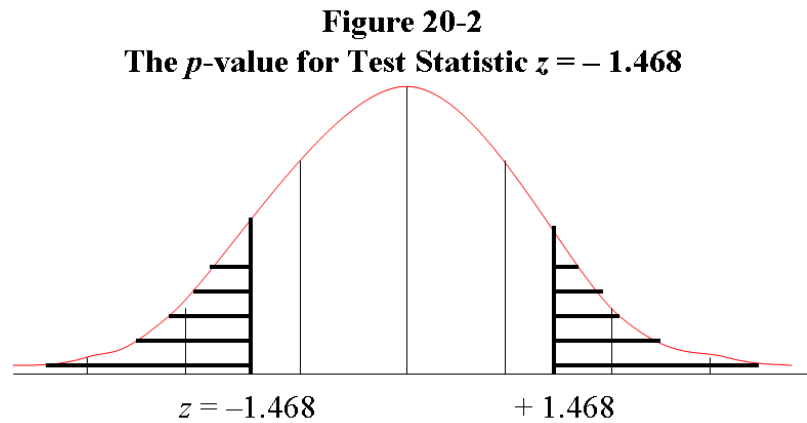
On the other hand, if we decide that there is not sufficient evidence against the H_0 , then one might think that we could say that we accept the H_0 ; however, instead of saying that we accept the H_0 , it has become customary in statistical terminology to say that we do not reject H_0 . The reason for saying "do not reject the H_0 " instead of saying "accept the H_0 ," even though these statements can be considered equivalent, is to emphasize that our decision to believe the H_0 is not based on evidence suggesting that H_0 is true. When we decide to believe H_0 , it is because there is not sufficient evidence against H_0 and not because there is evidence to support H_0 . Remember that we assume H_0 is true at the outset of a hypothesis test, and the purpose of the hypothesis test is to look for evidence indicating that H_1 is true.

Once again, we see the parallels between a hypothesis test and a court trial. The two possible verdicts in a court trial are stated as "guilty" and "not guilty." The reason for saying "not guilty" instead of saying

"innocent," even though these statements can be considered grammatically equivalent, is to emphasize that the verdict of "not guilty" is based on a lack of sufficient evidence suggesting guilt and not based necessarily on evidence indicating innocence.

As part of the third step in our hypothesis test concerning the lighter, we must decide whether or not to reject the null hypothesis. Since our test statistic, calculated in the second step, was found to be $z = -1.468$, which is not in the rejection region, our decision is not to reject $H_0: \lambda = 0.75$; in other words, our data does not provide sufficient evidence against $H_0: \lambda = 0.75$.

In order to complete the third step in a hypothesis test, we must obtain what is called a probability value for the test. The *probability value*, usually just called the *p-value*, of a hypothesis test is the probability of obtaining a test statistic value more supportive of the alternative hypothesis than the test statistic value actually observed, under the assumption H_0 is true. In our hypothesis test about the lighter, the *p-value* is the probability of obtaining a sample proportion farther away from (below or above) the



hypothesized 0.75 than that actually observed, under the assumption $H_0: \lambda = 0.75$. This is the probability of obtaining a z statistic which represents a larger distance away from the hypothesized proportion than does the observed $z = -1.468$, and this probability is represented by the shaded region in Figure 20-2. In Figure 20-2, the observed test statistic value $z = -1.468$ has been located on the horizontal axis, and, in addition, the value $+1.468$ has been located; the shaded area below $z = -1.468$ together with the shaded area above $+1.468$ correspond to test statistic values which represent a larger distance away from the hypothesized proportion than does $z = -1.468$. We have actually already calculated this *p*-value, since this is exactly the probability represented in Figure 19-1.

To obtain the shaded area above $+1.468$ in Figure 20-2 from Table A.2, we must first round $+1.468$ to $+1.47$. Then, we look for the table entry corresponding to the row labeled 1.4 and the column labeled 0.07. This table entry is 0.0708, and from the symmetry of a normal distribution we know that this is the area above $+1.468$ and also the area below -1.468 . Consequently, the total shaded area in Figure 20-2 must be $0.0708 + 0.0708 = 0.1416$. The *p*-value of this hypothesis test then is 0.1416, which we can designate by writing *p*-value = 0.1416. This is exactly the same probability we obtained with Figure 19-1.

When we found this probability to be 0.1416 with Figure 19-1, we then knew that we did not have sufficient evidence to reject H_0 , because 0.1416 was not less than the chosen significance level $\alpha = 0.10$. When the *p*-value is larger than the chosen significance level, then we have not found sufficient evidence to reject H_0 ; when the *p*-value is smaller than the chosen significance level, then we have found sufficient evidence to reject H_0 . You should now realize, however, that the *p*-value can give us more information than simply whether or not we should reject H_0 with the chosen significance level. The *p*-value 0.1416 just obtained tells us not only that H_0 is not rejected with a 0.10 significance level, but also tells us that H_0 would not be rejected with any of the commonly chosen significance levels, 0.01, 0.05, and 0.10 (because 0.1416 is larger than each of these). Suppose for the sake of argument, though, that we had obtained *p*-value = 0.0454; then we would see that H_0 would have been rejected with a significance level of 0.05 or 0.10, but H_0 would not have been rejected with a significance level of 0.01. Remember that the choice of significance level α is made by the person(s) performing the hypothesis test. Those who read the results of a hypothesis test may wish to know whether or not the conclusion would have been the same with a different significance level.

The *p*-value of a hypothesis test is often treated as a measure of how strongly the observed data supports the H_1 . For instance, if we had obtained *p*-value = 0.002, then we would see that not only do we have sufficient evidence to reject H_0 with $\alpha = 0.10$, but also with $\alpha = 0.05$ and with $\alpha = 0.01$. The smaller the *p*-value is, the stronger we might consider the evidence against H_0 to be.

After having defined our rejection region, having decided not to reject H_0 , and having obtained the *p*-value, we have now completed the first three steps in our hypothesis test about the lighter. The fourth step

consists of stating results and performing any further analysis which may be required. A clear, concise statement of the results of a hypothesis test is certainly important. The results of hypothesis testing will be of no value to anyone unless others can read and understand the results. We suggest that every hypothesis test can and should be summarized in a few sentences which include the observed test statistic value, the tabled value which defines the rejection region, the conclusion, and the p -value; also, even though we may use symbols as convenient abbreviations in our statements of H_0 and H_1 , for the sake of clarity it is preferable to summarize the results of a hypothesis test with complete sentences and words instead of symbols. For instance, we can summarize the results of the hypothesis test concerning the lighter as follows:

Since $z = -1.468$ and $z_{0.05} = 1.645$, we do not have sufficient evidence to reject H_0 . We conclude that the proportion of times the lighter will ignite on the first try is not different from 0.75 (p -value = 0.1416).

In the statement of results, we concluded that the population proportion "is not different from 0.75." This appears to be a rather negative way of saying that the population proportion "is equal to 0.75." Even though these two statements are grammatically equivalent, we chose to state the negative version in our results in order to emphasize that concluding H_0 is true is based on a lack of sufficient evidence to support H_1 and not based necessarily on evidence suggesting that H_0 is true. It is correct to state the conclusion either way, however.

The use of the test statistic z in a hypothesis test about λ is based on the assumption that we may treat the sampling distribution of \bar{p} as a normal distribution. The Central Limit Theorem for sample proportions says that this assumption is reasonable when $n\lambda \geq 5$ and $n(1 - \lambda) \geq 5$. In the illustration just completed, the sample size was $n = 300$, and we can use the hypothesized value of λ (0.75) to verify that this sample size was sufficiently large to treat the sampling distribution of \bar{p} as a normal distribution. Since $(300)(0.75) = 225$ and $(300)(1 - 0.25) = 75$ are both considerably larger than 5, the use of the test statistic z was appropriate.

As another illustration of a hypothesis test, we shall have you perform the four steps of the hypothesis test described in Table 20-1. The first step is to state H_0 and H_1 , and choose a significance level α . Remember that H_0 is what we assume to be true until and unless there is sufficient evidence against it, and H_1 is the statement we are looking for evidence to support; also, H_0 is generally a statement involving equality, and H_1 is generally a statement involving inequality. Using the appropriate symbols, write H_0 and H_1 under Step 1 in Table 20-1 to complete the first step of the hypothesis test. (As the first step of the hypothesis test, you should have $H_0: \lambda = 0.07$, $H_1: \lambda \neq 0.07$, $\alpha = 0.05$.)

The second step is to collect data and calculate the value of the test statistic. We can verify that the use of the z statistic is appropriate by observing that $(250)(0.07) = 17.5$ and $(250)(0.93) = 232.5$ are both larger than 5. Find the sample proportion \bar{p} , and calculate the value of the z statistic under Step 2 in Table 20-1 to complete the second step of the hypothesis test. (You should find that $\bar{p} = 27/250 = 0.108$ and $z = +2.355$.)

The third step is to define the rejection region, decide whether or not to reject H_0 , and obtain the p -value of the test. Under Step 3 in Table 20-1, draw a graph to display the rejection region, write the rejection region algebraically, indicate whether or not to reject H_0 , and write the p -value of the test. (Figure 20-3 displays the rejection region graphically where $z_{0.025} = 1.960$, and this rejection can be written algebraically as $z \geq +1.960$ or $z \leq -1.960$, or as $|z| \geq 1.960$; you should find that H_0 is rejected and that p -value = 0.0185 which is obtained by averaging the entries for 2.35 and 2.36 in Table A.2 and doubling this result.)

Complete the fourth step of the hypothesis test by summarizing the results under Step 4 in Table 20-1. Recall that in our earlier illustration concerning the lighter, we did not reject the H_0 , thus concluding that the population proportion was equal to the hypothesized value. In the illustration of Table 20-1 concerning underweight packages, we reject the H_0 , thus concluding that the population proportion was not equal to the hypothesized value. Rejecting the H_0 and concluding that λ is not equal to the hypothesized value leaves us with the question of whether λ is less than or greater than the hypothesized value.

Answers to Self-Test Problems

- 20-1** (a) Step 1: $H_0: \lambda = 0.4$, $H_1: \lambda \neq 0.4$, $\alpha = 0.01$
Step 2: $\bar{p} = 165/500 = 0.33$ and $z = -3.195$
Step 3: The rejection is $z \geq +2.576$ or $z \leq -2.576$ (which can be written as $|z| \geq 1.960$).
 H_0 is rejected; p -value < 0.002 .
Step 4: Since $z = -3.195$ and $z_{0.005} = 2.576$, we have sufficient evidence to reject H_0 . We conclude that the proportion of registered voters in the state intending to vote for the candidate is different from 0.4 (p -value < 0.002). The data suggest that the proportion is less than 0.4.
- (b) $500(0.4) = 200$ and $500(1 - 0.4) = 300$ are both larger than 5 implying n is sufficiently large.
- (c) Since H_0 is rejected, a Type I error is possible, which is concluding that $\lambda \neq 0.4$ when really $\lambda = 0.4$.
- (d) H_0 would have been rejected with $\alpha = 0.05$ and with $\alpha = 0.10$.
- (e) a bar chart or pie chart
- 20-2** (a) $H_0: \mu = 20,000$, $H_1: \mu \neq 20,000$, $\alpha = 0.05$ (b) b) We do not know $\sigma_x = \sigma / \sqrt{n} = \sigma / \sqrt{45}$.

Summary

The hypothesis which is assumed to be true at the outset of a hypothesis test is called the *null hypothesis*, abbreviated H_0 , and is generally a statement of equality. The hypothesis for which sufficient evidence is required before it will be believed is called the *alternative hypothesis* (or sometimes also called the *research hypothesis*), abbreviated H_1 , and is generally a statement of inequality. The first step in a hypothesis test is stating H_0 and H_1 , and choosing a significance level α . The hypotheses in a test to see if there is evidence that a population proportion is different from a hypothesized value can be written as $H_0: \lambda = \lambda_0$ vs. $H_1: \lambda \neq \lambda_0$.

A *test statistic* in hypothesis testing is a statistic which is used to decide whether to believe the H_0 or to believe the H_1 . With a null hypothesis $H_0: \lambda = \lambda_0$, the z statistic

$$z = \frac{\bar{p} - \lambda_0}{\sqrt{\frac{\lambda_0(1 - \lambda_0)}{n}}}$$

can be used shall if the simple random sample size is large enough so that $n\lambda_0 \geq 5$ and $n(1 - \lambda_0) \geq 5$. The second step of a hypothesis test is to collect data and calculate the value of the test statistic.

A *rejection region* in hypothesis testing, sometimes also called a *critical region*, is a set of test statistic values which lead to rejecting the null hypothesis in favor of the alternative hypothesis. In general, the rejection for the hypothesis test $H_0: \lambda = \lambda_0$ vs. $H_1: \lambda \neq \lambda_0$ with significance level α will be defined by $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ (i.e., $|z| \geq z_{\alpha/2}$). Since the H_0 is assumed to be true unless sufficient evidence is found against it, the custom in statistical terminology has been to state the decision in terms of the H_0 . If we decide that there is sufficient evidence against the H_0 , we say that we reject the H_0 , and saying that we reject H_0 is equivalent to saying that we accept H_1 . However, if we decide that there is not sufficient evidence against the H_0 , it has become customary in statistical terminology to say that we do not reject H_0 instead of saying that we accept the H_0 ; this is to emphasize that we decided to believe H_0 because there is not sufficient evidence against H_0 and not because there is evidence to support H_0 . The *probability value*, usually just called the p -value, of a hypothesis test is the probability of obtaining a test statistic value more supportive of the alternative hypothesis than the test statistic value actually observed, under the assumption H_0 is true. The third step of a hypothesis test is to define the rejection region, decide whether or not to reject H_0 , and obtain the p -value of the test.

The fourth step of a hypothesis test consists of stating results and performing any further analysis which may be required. A hypothesis test should be summarized in a few sentences which make clear the results to the reader. It is suggested that the summary of results include the observed test statistic value, the tabled value

which defines the rejection region, the conclusion, and the p -value. When the null hypothesis $H_0: \lambda = \lambda_0$ is rejected, the statement of the results of the test should also include an indication of whether the population proportion appears to be less than or greater than the hypothesized proportion.