9

# Large-Sample Tests of Hypotheses


© Scott Olson/Getty Images

## GENERAL OBJECTIVE

In this chapter, the concept of a statistical test of hypothesis is formally introduced. The sampling distributions of statistics presented in Chapters 7 and 8 are used to construct large-sample tests concerning the values of population parameters of interest to the experimenter.

## CHAPTER INDEX



## MY PERSONAL TRAINER

**Rejection Regions, *p*-Values, and Conclusions**
**How Do I Calculate $\beta$?**

## An Aspirin a Day . . . ?

Will an aspirin a day reduce the risk of heart attack? A very large study of U.S. physicians showed that a single aspirin taken every other day reduced the risk of heart attack in men by one-half. However, three days later, a British study reported a completely opposite conclusion. How could this be? The case study at the end of this chapter looks at how the studies were conducted, and you will analyze the data using large-sample techniques.

## 9.1 TESTING HYPOTHESES ABOUT POPULATION PARAMETERS

In practical situations, statistical inference can involve either estimating a population parameter or making decisions about the value of the parameter. For example, if a pharmaceutical company is fermenting a vat of antibiotic, samples from the vat can be used to *estimate* the mean potency $\mu$ for all of the antibiotic in the vat. In contrast, suppose that the company is not concerned about the exact mean potency of the antibiotic, but is concerned only that it meet the minimum government potency standards. Then the company can use samples from the vat to decide between these two possibilities:

- The mean potency $\mu$ does not exceed the minimum allowable potency.
- The mean potency $\mu$ exceeds the minimum allowable potency.

The pharmaceutical company's problem illustrates a **statistical test of hypothesis.**

The reasoning used in a statistical test of hypothesis is similar to the process in a court trial. In trying a person for theft, the court must decide between innocence and guilt. As the trial begins, the accused person is assumed to be *innocent.* The prosecution collects and presents all available evidence in an attempt to contradict the innocent hypothesis and hence obtain a conviction. If there is enough evidence against innocence, the court will reject the innocence hypothesis and declare the defendant *guilty.* If the prosecution does not present enough evidence to prove the defendant guilty, the court will find him *not guilty.* Notice that this does not prove that the defendant is innocent, but merely that there was not enough evidence to conclude that the defendant was guilty.

We use this same type of reasoning to explain the basic concepts of hypothesis testing. These concepts are used to test the four population parameters discussed in Chapter 8: a single population mean or proportion ($\mu$ or $p$) and the difference between two population means or proportions ($\mu_1 - \mu_2$ or $p_1 - p_2$). When the sample sizes are large, the point estimators for each of these four parameters have normal sampling distributions, so that all four large-sample statistical tests follow the same general pattern.

## 9.2 A STATISTICAL TEST OF HYPOTHESIS

A statistical test of hypothesis consists of five parts:

1. The null hypothesis, denoted by $H_0$
2. The alternative hypothesis, denoted by $H_a$
3. The test statistic and its $p$-value
4. The rejection region
5. The conclusion

When you specify these five elements, you define a particular test; changing one or more of the parts creates a new test. Let's look at each part of the statistical test of hypothesis in more detail.

**Definition 1–2** The two competing hypotheses are the **alternative hypothesis $H_a$,** generally the hypothesis that the researcher wishes to support, and the **null hypothesis $H_0$,** a contradiction of the alternative hypothesis.

As you will soon see, it is easier to show support for the alternative hypothesis by proving that the null hypothesis is false. Hence, the statistical researcher always begins by assuming that the null hypothesis $H_0$ is true. The researcher then uses the sample data to decide whether the evidence favors $H_a$ rather than $H_0$ and draws one of these two **conclusions:**

- Reject $H_0$ and conclude that $H_a$ is true.
- Accept (do not reject) $H_0$ as true.

**EXAMPLE 9.1**

You wish to show that the average hourly wage of carpenters in the state of California is different from \$14, which is the national average. This is the alternative hypothesis, written as

2 $H_a : \mu \neq 14$

The null hypothesis is

1 $H_0 : \mu = 14$

You would like to reject the null hypothesis, thus concluding that the California mean is not equal to \$14.

---

**EXAMPLE 9.2**

A milling process currently produces an average of 3% defectives. You are interested in showing that a simple adjustment on a machine will decrease $p$, the proportion of defectives produced in the milling process. Thus, the alternative hypothesis is

2 $H_a : p < .03$

and the null hypothesis is

1 $H_0 : p = .03$

If you can reject $H_0$, you can conclude that the adjusted process produces fewer than 3% defectives.

---

MY TIP

Two-tailed ⇔ Look for a $\neq$ sign in $H_a$.

One-tailed ⇔ Look for a $>$ or $<$ sign in $H_a$.

There is a difference in the forms of the alternative hypotheses given in Examples 9.1 and 9.2. In Example 9.1, no directional difference is suggested for the value of $\mu$; that is, $\mu$ might be either larger or smaller than \$14 if $H_a$ is true. This type of test is called a **two-tailed test of hypothesis.** In Example 9.2, however, you are specifically interested in detecting a directional difference in the value of $p$; that is, if $H_a$ is true, the value of $p$ is less than .03. This type of test is called a **one-tailed test of hypothesis.**

The decision to reject or accept the null hypothesis is based on information contained in a sample drawn from the population of interest. This information takes these forms:

3

- **Test statistic:** a single number calculated from the sample data
- ***p*-value:** a probability calculated using the test statistic

Either or both of these measures act as decision makers for the researcher in deciding whether to reject or accept $H_0$.

**EXAMPLE 9.3**

For the test of hypothesis in Example 9.1, the average hourly wage $\bar{x}$ for a random sample of 100 California carpenters might provide a good *test statistic* for testing

$$H_0 : \mu = 14 \quad \text{versus} \quad H_a : \mu \neq 14$$

If the null hypothesis $H_0$ is true, then the sample mean should not be too far from the population mean $\mu = 14$. Suppose that this sample produces a sample mean $\bar{x} = 15$ with standard deviation $s = 2$. Is this sample evidence likely or unlikely to occur, if in fact $H_0$ is true? You can use two measures to find out. Since the sample size is large, the sampling distribution of $\bar{x}$ is approximately normal with mean $\mu = 14$ and standard error $\sigma/\sqrt{n}$, estimated as

$$\text{SE} = \frac{s}{\sqrt{n}} = \frac{2}{\sqrt{100}} = .2$$

- The **test statistic** $\bar{x} = 15$ lies

**3**

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \approx \frac{15 - 14}{.2} = 5$$

standard deviations from the population mean $\mu$.

- The ***p*-value** is the probability of observing a test statistic as extreme as or more extreme than the observed value, if in fact $H_0$ is true. For this example, we define "extreme" as far below or far above what we would have expected. That is,

$$p\text{-value} = P(z > 5) + P(z < -5) \approx 0$$

The *large value of the test statistic* and the *small p-value* mean that you have observed a very unlikely event, if indeed $H_0$ is true and $\mu = 14$.

---

**4**

How do you decide whether to reject or accept $H_0$? The entire set of values that the test statistic may assume is divided into two sets, or regions. One set, consisting of values that support the alternative hypothesis and lead to rejecting $H_0$, is called the **rejection region.** The other, consisting of values that support the null hypothesis, is called the **acceptance region.**

For example, in Example 9.1, you would be inclined to believe that California's average hourly wage was different from \$14 if the sample mean is either much less than \$14 or much greater than \$14. The two-tailed rejection region consists of very small and very large values of $\bar{x}$, as shown in Figure 9.1. In Example 9.2, since you want to prove that the percentage of defectives has *decreased,* you would be inclined to reject $H_0$ for values of $\hat{p}$ that are much smaller than .03. Only *small* values of $\hat{p}$ belong in the left-tailed rejection region shown in Figure 9.2. When the rejection region is in the left tail of the distribution, the test is called a **left-tailed test.** A test with its rejection region in the right tail is called a **right-tailed test.**

**FIGURE 9.1**
Rejection and acceptance regions for Example 9.1

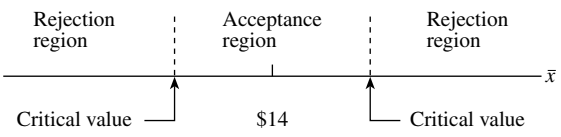
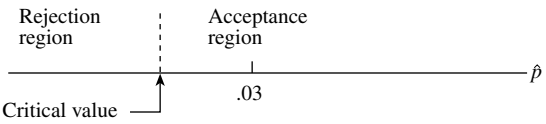


**FIGURE 9.2**
Rejection and acceptance regions for Example 9.2




5

If the test statistic falls into the rejection region, then the null hypothesis is rejected. If the test statistic falls into the acceptance region, then either the null hypothesis is accepted or the test is judged to be inconclusive. We will clarify the different types of conclusions that are appropriate as we consider several practical examples of hypothesis tests.

Finally, how do you decide on the **critical values** that separate the acceptance and rejection regions? That is, how do you decide how much statistical evidence you need before you can reject $H_0$? This depends on the amount of confidence that you, the researcher, want to attach to the test conclusions and the **significance level $\alpha$,** the risk you are willing to take of making an incorrect decision.

**Definition** A **Type I error** for a statistical test is the error of rejecting the null hypothesis when it is true. The **level of significance (significance level)** for a statistical test of hypothesis is

$$\alpha = P(\text{Type I error}) = P(\text{falsely rejecting } H_0) = P(\text{rejecting } H_0 \text{ when it is true})$$

This value $\alpha$ represents the *maximum tolerable risk* of incorrectly rejecting $H_0$. Once this significance level is fixed, the rejection region can be set to allow the researcher to reject $H_0$ with a fixed degree of confidence in the decision.

In the next section, we will show you how to use a test of hypothesis to test the value of a population mean $\mu$. As we continue, we will clarify some of the computational details and add some additional concepts to complete your understanding of hypothesis testing.

## 9.3 A LARGE-SAMPLE TEST ABOUT A POPULATION MEAN

Consider a random sample of $n$ measurements drawn from a population that has mean $\mu$ and standard deviation $\sigma$. You want to test a hypothesis of the form†

1 $H_0 : \mu = \mu_0$

MY TIP
The null hypothesis will *always* have an "equals" sign attached.

where $\mu_0$ is some hypothesized value for $\mu$, versus a one-tailed alternative hypothesis:

2 $H_a : \mu > \mu_0$

The subscript zero indicates the value of the parameter specified by $H_0$. Notice that $H_0$ provides an exact value for the parameter to be tested, whereas $H_a$ gives a range of possible values for $\mu$.

†Note that if the test rejects the null hypothesis $\mu = \mu_0$ in favor of the alternative hypothesis $\mu > \mu_0$, then it will certainly reject a null hypothesis that includes $\mu < \mu_0$, since this is even more contradictory to the alternative hypothesis. For this reason, in this text we state the null hypothesis for a one-tailed test as $\mu = \mu_0$ rather than $\mu \leq \mu_0$.

## The Essentials of the Test

The sample mean $\bar{x}$ is the best estimate of the actual value of $\mu$, which is presently in question. What values of $\bar{x}$ would lead you to believe that $H_0$ is false and $\mu$ is, in fact, greater than the hypothesized value? The values of $\bar{x}$ that are extremely *large* would imply that $\mu$ is larger than hypothesized. Hence, you should reject $H_0$ if $\bar{x}$ is too large.

The next problem is to define what is meant by "too large." Values of $\bar{x}$ that lie too many standard deviations to the right of the mean are not very likely to occur. Those values have very little area to their right. Hence, you can define "too large" as being too many standard deviations away from $\mu_0$. But what is "too many"? This question can be answered using the *significance level* $\alpha$, the probability of rejecting $H_0$ when $H_0$ is *true.*

Remember that the standard error of $\bar{x}$ is estimated as

$$\text{SE} = \frac{s}{\sqrt{n}}$$

Since the sampling distribution of the sample mean $\bar{x}$ is approximately normal when $n$ **is large,** the number of standard deviations that $\bar{x}$ lies from $\mu_0$ can be measured using the **standardized test statistic,**

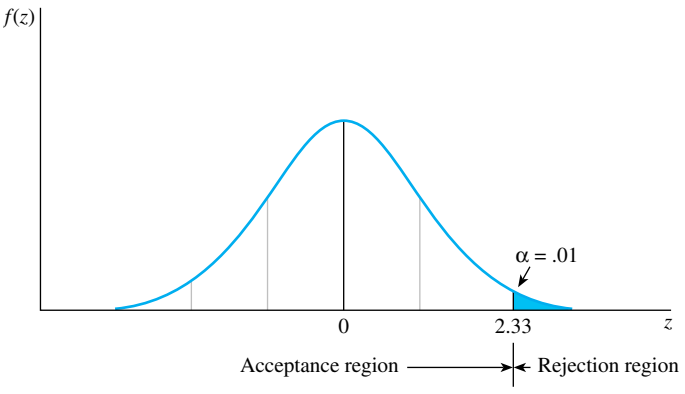
3

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

which has an approximate standard normal distribution when $H_0$ is true and $\mu = \mu_0$. The significance level $\alpha$ is equal to the area under the normal curve lying above the rejection region. Thus, if you want $\alpha = .01$, you will reject $H_0$ when $\bar{x}$ is more than 2.33 standard deviations to the right of $\mu_0$. Equivalently, you will reject $H_0$ if the standardized test statistic $z$ is greater than 2.33 (see Figure 9.3).

4

**FIGURE 9.3**
The rejection region for a right-tailed test with $\alpha = .01$




**EXAMPLE 9.4**

The average weekly earnings for female social workers is $670. Do men in the same positions have average weekly earnings that are higher than those for women? A random sample of $n = 40$ male social workers showed $\bar{x} = \$725$ and $s = \$102$. Test the appropriate hypothesis using $\alpha = .01$.

MY TIP
For one-tailed tests, look for directional words like "greater," "less than," "higher," "lower," etc.

**Solution** You would like to show that the average weekly earnings for men are higher than \$670, the women's average. Hence, if $\mu$ is the average weekly earnings for male social workers, you can set out the formal test of hypothesis in steps:

**1–2** **Null and alternative hypotheses:**

$H_0$: $\mu = 670$ versus $H_a$: $\mu > 670$

**3** **Test statistic:** Using the sample information, with $s$ as an estimate of the population standard deviation, calculate

$$z \approx \frac{\bar{x} - 670}{s/\sqrt{n}} = \frac{725 - 670}{102/\sqrt{40}} = 3.41$$

**4** **Rejection region:** For this one-tailed test, values of $\bar{x}$ much larger than 670 would lead you to reject $H_0$; or, equivalently, values of the *standardized test statistic* $z$ in the right tail of the standard normal distribution. To control the risk of making an incorrect decision as $\alpha = .01$, you must set the **critical value** separating the rejection and acceptance regions so that the area in the right tail is exactly $\alpha = .01$. This value is found in Table 3 of Appendix I to be $z = 2.33$, as shown in Figure 9.3. The null hypothesis will be rejected if the observed value of the test statistic, $z$, is greater than 2.33.

**5** **Conclusion:** Compare the observed value of the test statistic, $z = 3.41$, with the critical value necessary for rejection, $z = 2.33$. Since the observed value of the test statistic falls in the rejection region, you can reject $H_0$ and conclude that the average weekly earnings for male social workers are higher than the average for female social workers. The probability that you have made an incorrect decision is $\alpha = .01$.

---

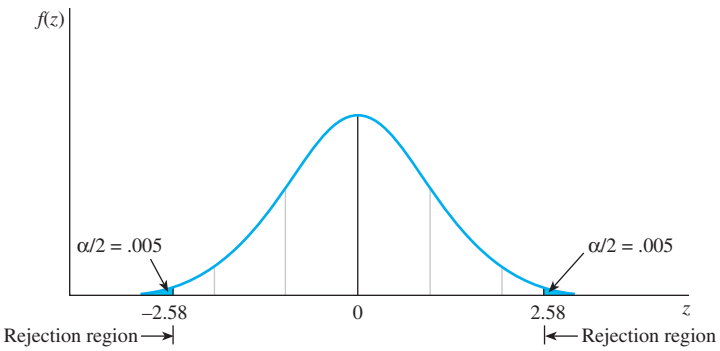
MY TIP
If the test is two-tailed, you will not see any directional words. The experimenter is only looking for a "difference" from the hypothesized value.

If you wish to detect departures either greater or less than $\mu_0$, then the alternative hypothesis is *two-tailed,* written as

$$H_a : \mu \neq \mu_0$$

which implies either $\mu > \mu_0$ or $\mu < \mu_0$. Values of $\bar{x}$ that are either "too large" or "too small" in terms of their distance from $\mu_0$ are placed in the rejection region. If you choose $\alpha = .01$, the area in the rejection region is equally divided between the two tails of the normal distribution, as shown in Figure 9.4. Using the standardized test statistic $z$, you can reject $H_0$ if $z > 2.58$ or $z < -2.58$. For different values of $\alpha$, the critical values of $z$ that separate the rejection and acceptance regions will change accordingly.

**FIGURE 9.4**
The rejection region for a two-tailed test with $\alpha = .01$



**EXAMPLE 9.5**

The daily yield for a local chemical plant has averaged 880 tons for the last several years. The quality control manager would like to know whether this average has changed in recent months. She randomly selects 50 days from the computer database and computes the average and standard deviation of the $n = 50$ yields as $\bar{x} = 871$ tons and $s = 21$ tons, respectively. Test the appropriate hypothesis using $\alpha = .05$.

**Solution**

**1–2** **Null and alternative hypotheses:**

$H_0$: $\mu = 880$ versus $H_a$: $\mu \neq 880$

**3** **Test statistic:** The point estimate for $\mu$ is $\bar{x}$. Therefore, the test statistic is

$$z \approx \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{871 - 880}{21/\sqrt{50}} = -3.03$$

**4** **Rejection region:** For this two-tailed test, you use values of $z$ in both the right and left tails of the standard normal distribution. Using $\alpha = .05$, the **critical values** separating the rejection and acceptance regions cut off areas of $\alpha/2 = .025$ in the right and left tails. These values are $z = \pm 1.96$ and the null hypothesis will be rejected if $z > 1.96$ or $z < -1.96$.

**5** **Conclusion:** Since $z = -3.03$ and the calculated value of $z$ falls in the rejection region, the manager can reject the null hypothesis that $\mu = 880$ tons and conclude that it has changed. The probability of rejecting $H_0$ when $H_0$ is true and $\alpha = .05$, a fairly small probability. Hence, she is reasonably confident that the decision is correct.

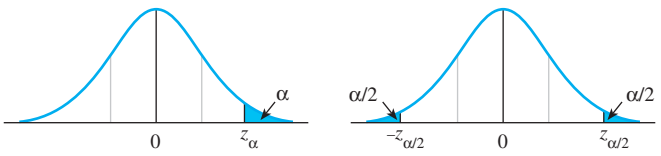
## LARGE-SAMPLE STATISTICAL TEST FOR $\mu$

1. Null hypothesis: $H_0 : \mu = \mu_0$
2. Alternative hypothesis:

| One-Tailed Test | Two-Tailed Test |
|---|---|
| $H_a : \mu > \mu_0$ (or, $H_a : \mu < \mu_0$) | $H_a : \mu \neq \mu_0$ |

3. Test statistic: $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ estimated as $z = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$
4. Rejection region: Reject $H_0$ when

| One-Tailed Test | Two-Tailed Test |
|---|---|
| $z > z_\alpha$ (or $z < -z_\alpha$ when the alternative hypothesis is $H_a : \mu < \mu_0$) | $z > z_{\alpha/2}$ or $z < -z_{\alpha/2}$ |



**Assumptions:** The $n$ observations in the sample are randomly selected from the population and $n$ is large—say, $n \geq 30$.

## Calculating the *p*-Value

In the previous examples, the decision to reject or accept $H_0$ was made by comparing the calculated value of the test statistic with a critical value of $z$ based on the significance level $\alpha$ of the test. However, different significance levels may lead to different conclusions. For example, if in a right-tailed test, the test statistic is $z = 2.03$, you can reject $H_0$ at the 5% level of significance because the test statistic exceeds $z = 1.645$. However, you cannot reject $H_0$ at the 1% level of significance, because the test statistic is less than $z = 2.33$ (see Figure 9.5). To avoid any ambiguity in their conclusions, some experimenters prefer to use a variable level of significance called the ***p*-value** for the test.

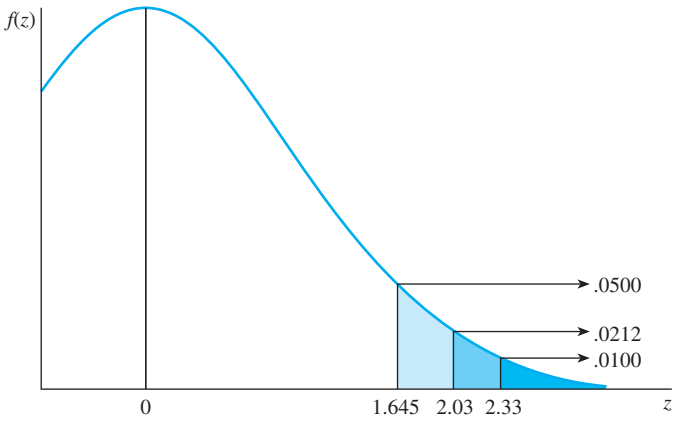
**Definition** The ***p*-value** or observed significance level of a statistical test is the smallest value of $\alpha$ for which $H_0$ can be rejected. It is the *actual risk* of committing a Type I error, if $H_0$ is rejected based on the observed value of the test statistic. The $p$-value measures the strength of the evidence against $H_0$.

In the right-tailed test with observed test statistic $z = 2.03$, the smallest critical value you can use and still reject $H_0$ is $z = 2.03$. For this critical value, the risk of an incorrect decision is

$$P(z \geq 2.03) = 1 - .9788 = .0212$$

This probability is the *p-value* for the test. *Notice that it is actually the area to the right of the calculated value of the test statistic.*

**FIGURE 9.5**
Variable rejection regions




**TIP**
*p*-value = tail area (one or two tails) "beyond" the observed value of the test statistic

A *small p-value* indicates that the observed value of the test statistic lies far away from the hypothesized value of $\mu$. This presents strong evidence that $H_0$ is false and should be rejected. *Large p-values* indicate that the observed test statistic is not far from the hypothesized mean and does not support rejection of $H_0$. How small does the $p$-value need to be before $H_0$ can be rejected?

**Definition** If the $p$-value is less than or equal to a preassigned significance level $\alpha$, then the null hypothesis can be rejected, and you can report that the results are **statistically significant** at level $\alpha$.

In the previous instance, if you choose $\alpha = .05$ as your significance level, $H_0$ can be rejected because the $p$-value is less than .05. However, if you choose $\alpha = .01$ as your significance level, the $p$-value (.0212) is not small enough to allow rejection of $H_0$. The results are significant at the 5% level, but not at the 1% level. You might see these results reported in professional journals as *significant* ($p < .05$).[†]

**EXAMPLE 9.6**

Refer to Example 9.5. The quality control manager wants to know whether the daily yield at a local chemical plant—which has averaged 880 tons for the last several years—has changed in recent months. A random sample of 50 days gives an average yield of 871 tons with a standard deviation of 21 tons. Calculate the $p$-value for this two-tailed test of hypothesis. Use the $p$-value to draw conclusions regarding the statistical test.

**Solution** The rejection region for this two-tailed test of hypothesis is found in both tails of the normal probability distribution. Since the observed value of the test statistic is $z = -3.03$, the smallest rejection region that you can use and still reject $H_0$ is $|z| > 3.03$. For this rejection region, the value of $\alpha$ is the $p$-value:

$$p\text{-value} = P(z > 3.03) + P(z < -3.03) = (1 - .9988) + .0012 = .0024$$

*Notice that the two-tailed p-value is actually twice the tail area corresponding to the calculated value of the test statistic.* If this $p$-value $= .0024$ is less than or equal to the preassigned level of significance $\alpha$, $H_0$ can be rejected. For this test, you can reject $H_0$ at either the 1% or the 5% level of significance.

If you are reading a research report, how small should the $p$-value be before you decide to reject $H_0$? Many researchers use a "sliding scale" to classify their results.

- If the $p$-value is less than .01, $H_0$ is rejected. The results are **highly significant.**
- If the $p$-value is between .01 and .05, $H_0$ is rejected. The results are **statistically significant.**
- If the $p$-value is between .05 and .10, $H_0$ is usually not rejected. The results are only **tending toward statistical significance.**
- If the $p$-value is greater than .10, $H_0$ is not rejected. The results are **not statistically significant.**

**EXAMPLE 9.7**

Standards set by government agencies indicate that Americans should not exceed an average daily sodium intake of 3300 milligrams (mg). To find out whether Americans are exceeding this limit, a sample of 100 Americans is selected, and the mean and standard deviation of daily sodium intake are found to be 3400 mg and 1100 mg, respectively. Use $\alpha = .05$ to conduct a test of hypothesis.

[†]In reporting statistical significance, many researchers write ($p < .05$) or ($P < .05$) to mean that the $p$-value of the test was smaller than .05, making the results significant at the 5% level. The symbol $p$ or $P$ in the expression has no connection with our notation for probability or with the binomial parameter $p$.

**Solution** The hypotheses to be tested are

$$H_0 : \mu = 3300 \quad \text{versus} \quad H_a : \mu > 3300$$

and the test statistic is

$$z \approx \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{3400 - 3300}{1100/\sqrt{100}} = .91$$

The two approaches developed in this section yield the same conclusions.

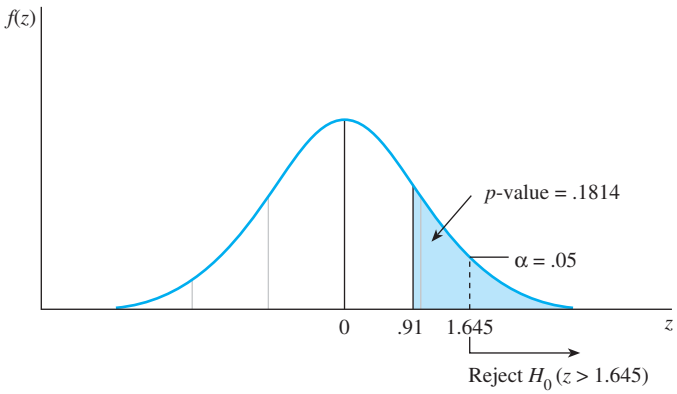
MY TIP
small $p$-value ⇔ large $z$-value
small $p$-value ⇒ reject $H_0$
How small? $p$-value ≤ $\alpha$

- **The critical value approach:** Since the significance level is $\alpha = .05$ and the test is one-tailed, the rejection region is determined by a critical value with tail area equal to $\alpha = .05$; that is, $H_0$ can be rejected if $z > 1.645$. Since $z = .91$ is not greater than the critical value, $H_0$ is not rejected (see Figure 9.6).
- **The *p*-value approach:** Calculate the $p$-value, the probability that $z$ is greater than or equal to $z = .91$:

$$p\text{-value} = P(z > .91) = 1 - .8186 = .1814$$

The null hypothesis can be rejected only if the *p-value is less than or equal to the specified 5% significance level.* Therefore, $H_0$ is not rejected and the results are *not statistically significant* (see Figure 9.6). There is not enough evidence to indicate that the average daily sodium intake exceeds 3300 mg.

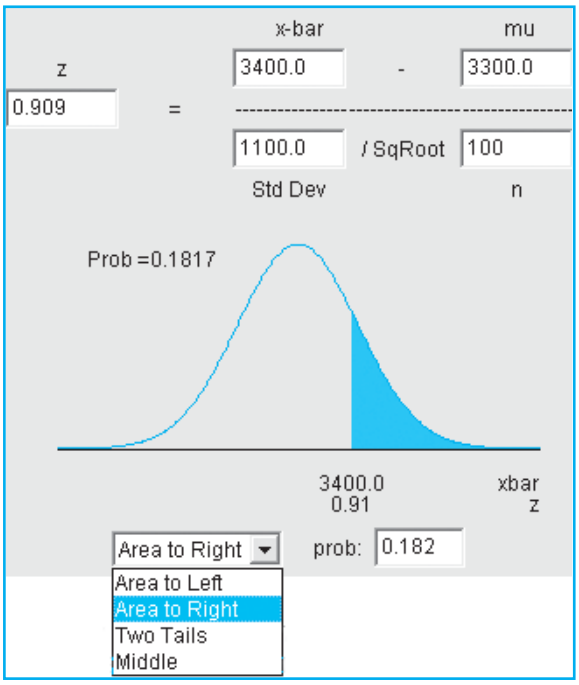
**FIGURE 9.6**
Rejection region and $p$-value for Example 9.7




## MY APPLET

You can use the **Large-Sample Test of a Population Mean** applet to visualize the $p$-values for either one- or two-tailed tests of the population mean $\mu$ (Figure 9.7). Remember, however, that these large-sample $z$-tests are restricted to samples of size $n \geq 30$. The applet does not prohibit you from entering a value of $n < 30$; you'll have to be careful to check the sample size before you start! The procedure follows the same pattern as with previous applets. You enter the values of $\bar{x}$, $n$, and $s$—remember to press "Enter" after each entry to record the changes. The applet will calculate $z$ (using full accuracy) and give you the option of choosing one- or two-tailed $p$-values (*Area to Left, Area to Right,* or *Two Tails*), as well as a *Middle* area that you will not need.

**FIGURE 9.7**
Large-Sample Test of a Population Mean applet




For the data of Example 9.7, the $p$-value is the one-tailed area to the right of $z = .909$. Do the results shown in the applet confirm our conclusions in Example 9.7? Remember that the applet uses full accuracy for the calculation of $z$ and its corresponding probability. This means that the probability we calculate using Table 3 in Appendix I may be slightly different from the probability shown in the applet.

Notice that these two approaches are actually the same, as shown in Figure 9.6. As soon as the calculated value of the test statistic $z$ becomes *larger than* the critical value, $z_\alpha$, the $p$-value becomes *smaller than* the significance level $\alpha$. You can use the most convenient of the two methods; the conclusions you reach will always be the same! The $p$-value approach does have two advantages, however:

- Statistical output from packages such as *MINITAB* usually reports the $p$-value of the test.
- Based on the $p$-value, your test results can be evaluated using any significance level you wish to use. Many researchers report the smallest possible significance level for which their results are *statistically significant.*

Sometimes it is easy to confuse the significance level $\alpha$ with the $p$-value (or observed significance level). They are both probabilities calculated as areas in the tails of the sampling distribution of the test statistic. However, the significance level $\alpha$ is preset by the experimenter before collecting the data. The $p$-value is linked directly to the data and actually describes how likely or unlikely the sample results are, assuming that $H_0$ is true. *The smaller the p-value, the more unlikely it is that $H_0$ is true!*