



## CHAPTER

# 8

## Hypothesis Testing

### Objectives

After completing this chapter, you should be able to

- 1** Understand the definitions used in hypothesis testing.
- 2** State the null and alternative hypotheses.
- 3** Find critical values for the  $z$  test.
- 4** State the five steps used in hypothesis testing.
- 5** Test means when  $\sigma$  is known, using the  $z$  test.
- 6** Test means when  $\sigma$  is unknown, using the  $t$  test.
- 7** Test proportions, using the  $z$  test.
- 8** Test variances or standard deviations, using the chi-square test.
- 9** Test hypotheses, using confidence intervals.
- 10** Explain the relationship between type I and type II errors and the power of a test.

### Outline

#### Introduction

#### **8-1** Steps in Hypothesis Testing—Traditional Method

#### **8-2** $z$ Test for a Mean

#### **8-3** $t$ Test for a Mean

#### **8-4** $z$ Test for a Proportion

#### **8-5** $\chi^2$ Test for a Variance or Standard Deviation

#### **8-6** Additional Topics Regarding Hypothesis Testing

#### Summary



### Statistics Today

#### How Much Better Is Better?

Suppose a school superintendent reads an article which states that the overall mean score for the SAT is 910. Furthermore, suppose that, for a sample of students, the average of the SAT scores in the superintendent's school district is 960. Can the superintendent conclude that the students in his school district scored higher on average? At first glance, you might be inclined to say yes, since 960 is higher than 910. But recall that the means of samples vary about the population mean when samples are selected from a specific population. So the question arises, Is there a real difference in the means, or is the difference simply due to chance (i.e., sampling error)? In this chapter, you will learn how to answer that question by using statistics that explain hypothesis testing. See *Statistics Today—Revisited* for the answer. In this chapter, you will learn how to answer many questions of this type by using statistics that are explained in the theory of hypothesis testing.

#### Introduction

Researchers are interested in answering many types of questions. For example, a scientist might want to know whether the earth is warming up. A physician might want to know whether a new medication will lower a person's blood pressure. An educator might wish to see whether a new teaching technique is better than a traditional one. A retail merchant might want to know whether the public prefers a certain color in a new line of fashion. Automobile manufacturers are interested in determining whether seat belts will reduce the severity of injuries caused by accidents. These types of questions can be addressed through statistical **hypothesis testing**, which is a decision-making process for evaluating claims about a population. In hypothesis testing, the researcher must define the population under study, state the particular hypotheses that will be investigated, give the significance level, select a sample from the population, collect the data, perform the calculations required for the statistical test, and reach a conclusion.

Hypotheses concerning parameters such as means and proportions can be investigated. There are two specific statistical tests used for hypotheses concerning means: the *z test*

and the  $t$  test. This chapter will explain in detail the hypothesis-testing procedure along with the  $z$  test and the  $t$  test. In addition, a hypothesis-testing procedure for testing a single variance or standard deviation using the chi-square distribution is explained in Section 8-5.

The three methods used to test hypotheses are

1. The traditional method
2. The  $P$ -value method
3. The confidence interval method

The *traditional method* will be explained first. It has been used since the hypothesis-testing method was formulated. A newer method, called the  *$P$ -value method*, has become popular with the advent of modern computers and high-powered statistical calculators. It will be explained at the end of Section 8-2. The third method, the *confidence interval method*, is explained in Section 8-6 and illustrates the relationship between hypothesis testing and confidence intervals.

## 8-1

### Steps in Hypothesis Testing—Traditional Method

Every hypothesis-testing situation begins with the statement of a hypothesis.

A **statistical hypothesis** is a conjecture about a population parameter. This conjecture may or may not be true.

#### Objective 1

Understand the definitions used in hypothesis testing.

There are two types of statistical hypotheses for each situation: the null hypothesis and the alternative hypothesis.

The **null hypothesis**, symbolized by  $H_0$ , is a statistical hypothesis that states that there is no difference between a parameter and a specific value, or that there is no difference between two parameters.

The **alternative hypothesis**, symbolized by  $H_1$ , is a statistical hypothesis that states the existence of a difference between a parameter and a specific value, or states that there is a difference between two parameters.

(Note: Although the definitions of null and alternative hypotheses given here use the word *parameter*, these definitions can be extended to include other terms such as *distributions* and *randomness*. This is explained in later chapters.)

As an illustration of how hypotheses should be stated, three different statistical studies will be used as examples.

**Situation A** A medical researcher is interested in finding out whether a new medication will have any undesirable side effects. The researcher is particularly concerned with the pulse rate of the patients who take the medication. Will the pulse rate increase, decrease, or remain unchanged after a patient takes the medication?

Since the researcher knows that the mean pulse rate for the population under study is 82 beats per minute, the hypotheses for this situation are

$$H_0: \mu = 82 \quad \text{and} \quad H_1: \mu \neq 82$$

The null hypothesis specifies that the mean will remain unchanged, and the alternative hypothesis states that it will be different. This test is called a *two-tailed test* (a term that will be formally defined later in this section), since the possible side effects of the medicine could be to raise or lower the pulse rate.

#### Objective 2

State the null and alternative hypotheses.

**Situation B** A chemist invents an additive to increase the life of an automobile battery. If the mean lifetime of the automobile battery without the additive is 36 months, then her hypotheses are

$$H_0: \mu = 36 \quad \text{and} \quad H_1: \mu > 36$$

In this situation, the chemist is interested only in increasing the lifetime of the batteries, so her alternative hypothesis is that the mean is greater than 36 months. The null hypothesis is that the mean is equal to 36 months. This test is called *right-tailed*, since the interest is in an increase only.

*Unusual Stat*

Sixty-three percent of people would rather hear bad news before hearing the good news.

**Situation C** A contractor wishes to lower heating bills by using a special type of insulation in houses. If the average of the monthly heating bills is \$78, her hypotheses about heating costs with the use of insulation are

$$H_0: \mu = \$78 \quad \text{and} \quad H_1: \mu < \$78$$

This test is a *left-tailed test*, since the contractor is interested only in lowering heating costs. To state hypotheses correctly, researchers must translate the *conjecture* or *claim* from words into mathematical symbols. The basic symbols used are as follows:

Equal to	=	Greater than	>
Not equal to	≠	Less than	<

The null and alternative hypotheses are stated together, and the null hypothesis contains the equals sign, as shown (where  $k$  represents a specified number).

Two-tailed test	Right-tailed test	Left-tailed test
$H_0: \mu = k$	$H_0: \mu = k$	$H_0: \mu = k$
$H_1: \mu \neq k$	$H_1: \mu > k$	$H_1: \mu < k$

The formal definitions of the different types of tests are given later in this section. In this book, the null hypothesis is always stated using the equals sign. This is done because in most professional journals, and when we test the null hypothesis, the assumption is that the mean, proportion, or standard deviation is equal to a given specific value. Also, when a researcher conducts a study, he or she is generally looking for evidence to support a claim. Therefore, the claim should be stated as the alternative hypothesis, i.e., using  $<$  or  $>$  or  $\neq$ . Because of this, the alternative hypothesis is sometimes called the **research hypothesis**.

Table 8-1 Hypothesis-Testing Common Phrases	
>	<
Is greater than	Is less than
Is above	Is below
Is higher than	Is lower than
Is longer than	Is shorter than
Is bigger than	Is smaller than
Is increased	Is decreased or reduced from
=	≠
Is equal to	Is not equal to
Is the same as	Is different from
Has not changed from	Has changed from
Is the same as	Is not the same as



A claim, though, can be stated as either the null hypothesis or the alternative hypothesis; however, the statistical evidence can only *support* the claim if it is the alternative hypothesis. Statistical evidence can be used to *reject* the claim if the claim is the null hypothesis. These facts are important when you are stating the conclusion of a statistical study.

Table 8-1 shows some common phrases that are used in hypotheses and conjectures, and the corresponding symbols. This table should be helpful in translating verbal conjectures into mathematical symbols.

**Example 8-1**

State the null and alternative hypotheses for each conjecture.

- A researcher thinks that if expectant mothers use vitamin pills, the birth weight of the babies will increase. The average birth weight of the population is 8.6 pounds.
- An engineer hypothesizes that the mean number of defects can be decreased in a manufacturing process of compact disks by using robots instead of humans for certain tasks. The mean number of defective disks per 1000 is 18.
- A psychologist feels that playing soft music during a test will change the results of the test. The psychologist is not sure whether the grades will be higher or lower. In the past, the mean of the scores was 73.

**Solution**

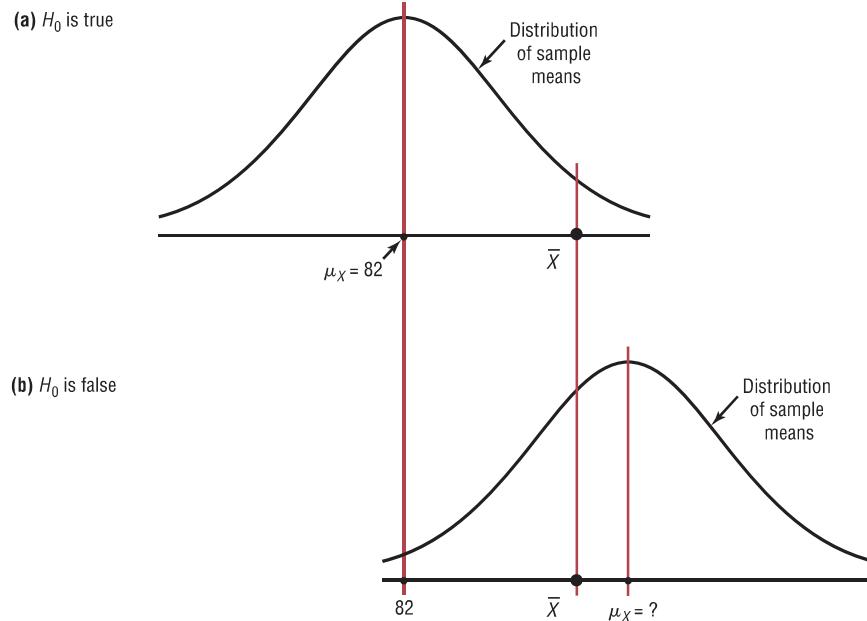
- $H_0: \mu = 8.6$  and  $H_1: \mu > 8.6$
- $H_0: \mu = 18$  and  $H_1: \mu < 18$
- $H_0: \mu = 73$  and  $H_1: \mu \neq 73$

After stating the hypothesis, the researcher designs the study. The researcher selects the correct *statistical test*, chooses an appropriate *level of significance*, and formulates a plan for conducting the study. In situation A, for instance, the researcher will select a sample of patients who will be given the drug. After allowing a suitable time for the drug to be absorbed, the researcher will measure each person's pulse rate.

Recall that when samples of a specific size are selected from a population, the means of these samples will vary about the population mean, and the distribution of the sample means will be approximately normal when the sample size is 30 or more. (See Section 6-3.) So even if the null hypothesis is true, the mean of the pulse rates of the sample of patients will not, in most cases, be exactly equal to the population mean of 82 beats per minute. There are two possibilities. Either the null hypothesis is true, and the difference between the sample mean and the population mean is due to chance; *or* the null hypothesis is false, and the sample came from a population whose mean is not 82 beats per minute but is some other value that is not known. These situations are shown in Figure 8-1.

The farther away the sample mean is from the population mean, the more evidence there would be for rejecting the null hypothesis. The probability that the sample came from a population whose mean is 82 decreases as the distance or absolute value of the difference between the means increases.

If the mean pulse rate of the sample were, say, 83, the researcher would probably conclude that this difference was due to chance and would not reject the null hypothesis. But if the sample mean were, say, 90, then in all likelihood the researcher would conclude that the medication increased the pulse rate of the users and would reject the null hypothesis. The question is, Where does the researcher draw the line? This decision is not made on feelings or intuition; it is made statistically. That is, the difference must be significant and in all likelihood not due to chance. Here is where the concepts of statistical test and level of significance are used.

**Figure 8–1****Situations in Hypothesis Testing**

A **statistical test** uses the data obtained from a sample to make a decision about whether the null hypothesis should be rejected.

The numerical value obtained from a statistical test is called the **test value**.

In this type of statistical test, the mean is computed for the data obtained from the sample and is compared with the population mean. Then a decision is made to reject or not reject the null hypothesis on the basis of the value obtained from the statistical test. If the difference is significant, the null hypothesis is rejected. If it is not, then the null hypothesis is not rejected.

In the hypothesis-testing situation, there are four possible outcomes. In reality, the null hypothesis may or may not be true, and a decision is made to reject or not reject it on the basis of the data obtained from a sample. The four possible outcomes are shown in Figure 8–2. Notice that there are two possibilities for a correct decision and two possibilities for an incorrect decision.

**Figure 8–2****Possible Outcomes of a Hypothesis Test**

	$H_0$ true	$H_0$ false
Reject $H_0$	<b>Error Type I</b>	Correct decision
Do not reject $H_0$	Correct decision	<b>Error Type II</b>

If a null hypothesis is true and it is rejected, then a *type I error* is made. In situation A, for instance, the medication might not significantly change the pulse rate of all the users in the population; but it might change the rate, by chance, of the subjects in the sample. In this case, the researcher will reject the null hypothesis when it is really true, thus committing a type I error.

On the other hand, the medication might not change the pulse rate of the subjects in the sample, but when it is given to the general population, it might cause a significant increase or decrease in the pulse rate of users. The researcher, on the basis of the data obtained from the sample, will not reject the null hypothesis, thus committing a *type II error*.

In situation B, the additive might not significantly increase the lifetimes of automobile batteries in the population, but it might increase the lifetimes of the batteries in the sample. In this case, the null hypothesis would be rejected when it was really true. This would be a type I error. On the other hand, the additive might not work on the batteries selected for the sample, but if it were to be used in the general population of batteries, it might significantly increase their lifetimes. The researcher, on the basis of information obtained from the sample, would not reject the null hypothesis, thus committing a type II error.

A **type I error** occurs if you reject the null hypothesis when it is true.

A **type II error** occurs if you do not reject the null hypothesis when it is false.

The hypothesis-testing situation can be likened to a jury trial. In a jury trial, there are four possible outcomes. The defendant is either guilty or innocent, and he or she will be convicted or acquitted. See Figure 8–3.

Now the hypotheses are

$H_0$ : The defendant is innocent

$H_1$ : The defendant is not innocent (i.e., guilty)

Next, the evidence is presented in court by the prosecutor, and based on this evidence, the jury decides the verdict, innocent or guilty.

If the defendant is convicted but he or she did not commit the crime, then a type I error has been committed. See block 1 of Figure 8–3. On the other hand, if the defendant is convicted and he or she has committed the crime, then a correct decision has been made. See block 2.

If the defendant is acquitted and he or she did not commit the crime, a correct decision has been made by the jury. See block 3. However, if the defendant is acquitted and he or she did commit the crime, then a type II error has been made. See block 4.

**Figure 8–3**  
Hypothesis Testing and  
a Jury Trial

$H_0$ : The defendant is innocent.  
 $H_1$ : The defendant is not innocent.  
The results of a trial can be shown as follows:

	$H_0$ true (innocent)	$H_0$ false (not innocent)
Reject $H_0$ (convict)	1. Type I error	2. Correct decision
Do not reject $H_0$ (acquitt)	3. Correct decision	4. Type II error

The decision of the jury does not prove that the defendant did or did not commit the crime. The decision is based on the evidence presented. If the evidence is strong enough, the defendant will be convicted in most cases. If the evidence is weak, the defendant will be acquitted in most cases. Nothing is proved absolutely. Likewise, the decision to reject or not reject the null hypothesis does not prove anything. *The only way to prove anything statistically is to use the entire population*, which, in most cases, is not possible. The decision, then, is made on the basis of probabilities. That is, when there is a large difference between the mean obtained from the sample and the hypothesized mean, the null hypothesis is probably not true. The question is, How large a difference is necessary to reject the null hypothesis? Here is where the level of significance is used.

### Unusual Stats

Of workers in the United States, 64% drive to work alone and 6% of workers walk to work.

The **level of significance** is the maximum probability of committing a type I error. This probability is symbolized by  $\alpha$  (Greek letter **alpha**). That is,  $P(\text{type I error}) = \alpha$ .

The probability of a type II error is symbolized by  $\beta$ , the Greek letter **beta**. That is,  $P(\text{type II error}) = \beta$ . In most hypothesis-testing situations,  $\beta$  cannot be easily computed; however,  $\alpha$  and  $\beta$  are related in that decreasing one increases the other.

Statisticians generally agree on using three arbitrary significance levels: the 0.10, 0.05, and 0.01 levels. That is, if the null hypothesis is rejected, the probability of a type I error will be 10%, 5%, or 1%, depending on which level of significance is used. Here is another way of putting it: When  $\alpha = 0.10$ , there is a 10% chance of rejecting a true null hypothesis; when  $\alpha = 0.05$ , there is a 5% chance of rejecting a true null hypothesis; and when  $\alpha = 0.01$ , there is a 1% chance of rejecting a true null hypothesis.

In a hypothesis-testing situation, the researcher decides what level of significance to use. It does not have to be the 0.10, 0.05, or 0.01 level. It can be any level, depending on the seriousness of the type I error. After a significance level is chosen, a *critical value* is selected from a table for the appropriate test. If a  $z$  test is used, for example, the  $z$  table (Table E in Appendix C) is consulted to find the critical value. The critical value determines the critical and noncritical regions.

The **critical value** separates the critical region from the noncritical region. The symbol for critical value is C.V.

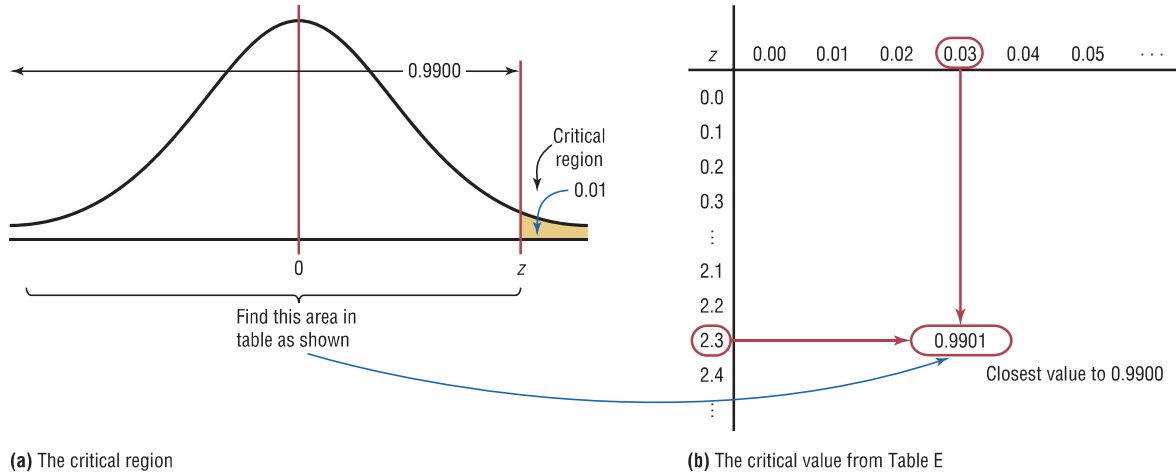
The **critical** or **rejection region** is the range of values of the test value that indicates that there is a significant difference and that the null hypothesis should be rejected.

The **noncritical** or **nonrejection region** is the range of values of the test value that indicates that the difference was probably due to chance and that the null hypothesis should not be rejected.

The critical value can be on the right side of the mean or on the left side of the mean for a one-tailed test. Its location depends on the inequality sign of the alternative hypothesis. For example, in situation B, where the chemist is interested in increasing the average lifetime of automobile batteries, the alternative hypothesis is  $H_1: \mu > 36$ . Since the inequality sign is  $>$ , the null hypothesis will be rejected only when the sample mean is significantly greater than 36. Hence, the critical value must be on the right side of the mean. Therefore, this test is called a right-tailed test.

A **one-tailed test** indicates that the null hypothesis should be rejected when the test value is in the critical region on one side of the mean. A one-tailed test is either a **right-tailed test** or **left-tailed test**, depending on the direction of the inequality of the alternative hypothesis.



**Figure 8-4****Finding the Critical Value for  $\alpha = 0.01$  (Right-Tailed Test)**

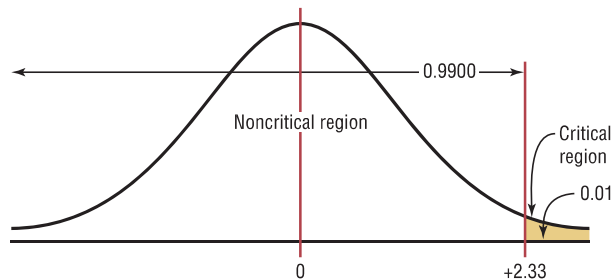
To obtain the critical value, the researcher must choose an alpha level. In situation B, suppose the researcher chose  $\alpha = 0.01$ . Then the researcher must find a  $z$  value such that 1% of the area falls to the right of the  $z$  value and 99% falls to the left of the  $z$  value, as shown in Figure 8-4(a).

**Objective 3**

Find critical values for the  $z$  test.

Next, the researcher must find the area value in Table E closest to 0.9900. The critical  $z$  value is 2.33, since that value gives the area closest to 0.9900 (that is, 0.9901), as shown in Figure 8-4(b).

The critical and noncritical regions and the critical value are shown in Figure 8-5.

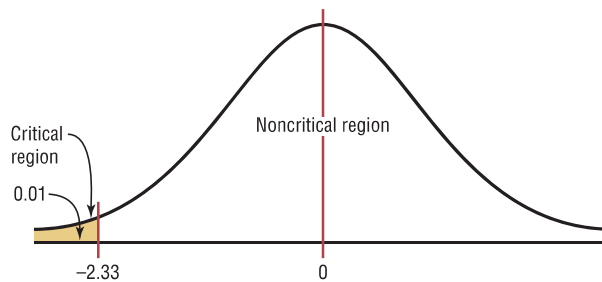
**Figure 8-5****Critical and Noncritical Regions for  $\alpha = 0.01$  (Right-Tailed Test)**

Now, move on to situation C, where the contractor is interested in lowering the heating bills. The alternative hypothesis is  $H_1: \mu < \$78$ . Hence, the critical value falls to the left of the mean. This test is thus a left-tailed test. At  $\alpha = 0.01$ , the critical value is  $-2.33$ , since 0.0099 is the closest value to 0.01. This is shown in Figure 8-6.

When a researcher conducts a two-tailed test, as in situation A, the null hypothesis can be rejected when there is a significant difference in either direction, above or below the mean.

**Figure 8-6**

**Critical and Noncritical Regions for  $\alpha = 0.01$  (Left-Tailed Test)**



In a **two-tailed test**, the null hypothesis should be rejected when the test value is in either of the two critical regions.

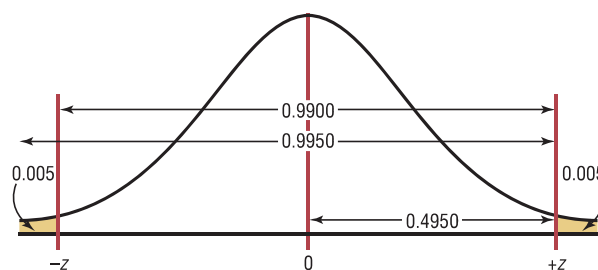
For a two-tailed test, then, the critical region must be split into two equal parts. If  $\alpha = 0.01$ , then one-half of the area, or 0.005, must be to the right of the mean and one-half must be to the left of the mean, as shown in Figure 8-7.

In this case, the  $z$  value on the left side is found by looking up the  $z$  value corresponding to an area of 0.0050. The  $z$  value falls about halfway between  $-2.57$  and  $-2.58$  corresponding to the areas 0.0049 and 0.0051. The average of  $-2.57$  and  $-2.58$  is  $[(-2.57) + (-2.58)] \div 2 = -2.575$  so if the  $z$  value is needed to three decimal places,  $-2.575$  is used; however, if the  $z$  value is rounded to two decimal places,  $-2.58$  is used.

On the right side, it is necessary to find the  $z$  value corresponding to  $0.99 + 0.005$ , or 0.9950. Again, the value falls between 0.9949 and 0.9951, so  $+2.575$  or 2.58 can be used. See Figure 8-7.

**Figure 8-7**

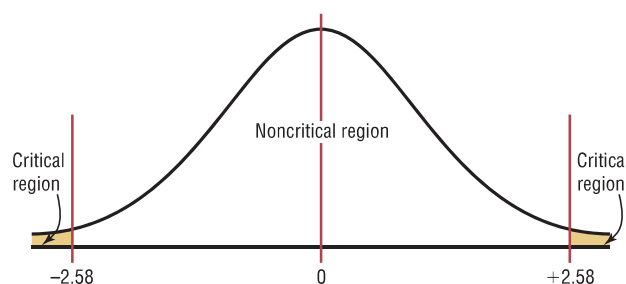
**Finding the Critical Values for  $\alpha = 0.01$  (Two-Tailed Test)**



The critical values are  $+2.58$  and  $-2.58$ , as shown in Figure 8-8.

**Figure 8-8**

**Critical and Noncritical Regions for  $\alpha = 0.01$  (Two-Tailed Test)**



Similar procedures are used to find other values of  $\alpha$ .

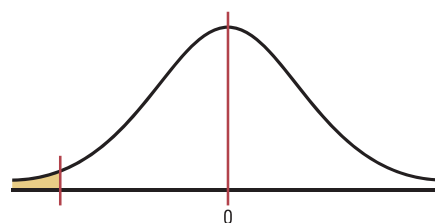
Figure 8-9 with rejection regions shaded shows the critical value (C.V.) for the three situations discussed in this section for values of  $\alpha = 0.10$ ,  $\alpha = 0.05$ , and  $\alpha = 0.01$ . The procedure for finding critical values is outlined next (where  $k$  is a specified number).

**Figure 8-9**

**Summary of  
Hypothesis Testing  
and Critical Values**

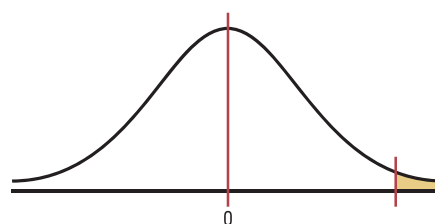
$$\begin{aligned} H_0: \mu &= k \\ H_1: \mu &< k \end{aligned} \begin{cases} \alpha = 0.10, \text{ C.V.} = -1.28 \\ \alpha = 0.05, \text{ C.V.} = -1.65 \\ \alpha = 0.01, \text{ C.V.} = -2.33 \end{cases}$$

**(a) Left-tailed**



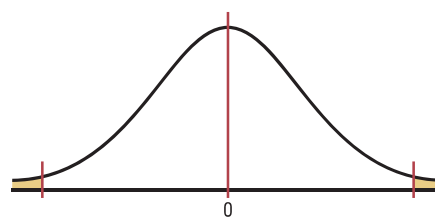
$$\begin{aligned} H_0: \mu &= k \\ H_1: \mu &> k \end{aligned} \begin{cases} \alpha = 0.10, \text{ C.V.} = +1.28 \\ \alpha = 0.05, \text{ C.V.} = +1.65 \\ \alpha = 0.01, \text{ C.V.} = +2.33 \end{cases}$$

**(b) Right-tailed**



$$\begin{aligned} H_0: \mu &= k \\ H_1: \mu &\neq k \end{aligned} \begin{cases} \alpha = 0.10, \text{ C.V.} = \pm 1.65 \\ \alpha = 0.05, \text{ C.V.} = \pm 1.96 \\ \alpha = 0.01, \text{ C.V.} = \pm 2.58 \end{cases}$$

**(c) Two-tailed**



## Procedure Table

### Finding the Critical Values for Specific $\alpha$ Values, Using Table E

- Step 1** Draw the figure and indicate the appropriate area.
- If the test is left-tailed, the critical region, with an area equal to  $\alpha$ , will be on the left side of the mean.
  - If the test is right-tailed, the critical region, with an area equal to  $\alpha$ , will be on the right side of the mean.
  - If the test is two-tailed,  $\alpha$  must be divided by 2; one-half of the area will be to the right of the mean, and one-half will be to the left of the mean.
- Step 2**
- For a left-tailed test, use the  $z$  value that corresponds to the area equivalent to  $\alpha$  in Table E.
  - For a right-tailed test, use the  $z$  value that corresponds to the area equivalent to  $1 - \alpha$ .
  - For a two-tailed test, use the  $z$  value that corresponds to  $\alpha/2$  for the left value. It will be negative. For the right value, use the  $z$  value that corresponds to the area equivalent to  $1 - \alpha/2$ . It will be positive.