

10

GOALS

When you have completed this chapter you will be able to:

- 1 Define a *hypothesis* and *hypothesis testing*.
- 2 Describe the five-step hypothesis-testing procedure.
- 3 Distinguish between a *one-tailed* and a *two-tailed test of hypothesis*.
- 4 Conduct a test of hypothesis about a population mean.
- 5 Conduct a test of hypothesis about a population proportion.
- 6 Define *Type I* and *Type II errors*.
- 7 Compute the probability of a Type II error.

One-Sample Tests of Hypothesis



According to the Coffee Research Organization the typical American coffee drinker consumes an average of 3.1 cups per day. A sample of 12 senior citizens reported the amounts of coffee in cups consumed in a particular day. At the .05 significance level does the sample data provided in the exercise suggest a difference between the national average and the sample mean from senior citizens. (See Exercise 39, Goal 4.)

One-Sample Tests of Hypothesis

331

Introduction

Chapter 8 began our study of statistical inference. We described how we could select a random sample and from this sample estimate the value of a population parameter. For example, we selected a sample of 5 employees at Spence Sprockets, found the number of years of service for each sampled employee, computed the mean years of service, and used the sample mean to estimate the mean years of service for all employees. In other words, we estimated a population parameter from a sample statistic.

Chapter 9 continued the study of statistical inference by developing a confidence interval. A confidence interval is a range of values within which we expect the population parameter to occur. In this chapter, rather than develop a range of values within which we expect the population parameter to occur, we develop a procedure to test the validity of a statement about a population parameter. Some examples of statements we might want to test are:



- The mean speed of automobiles passing milepost 150 on the West Virginia Turnpike is 68 miles per hour.
- The mean number of miles driven by those leasing a Chevy TrailBlazer for three years is 32,000 miles.
- The mean time an American family lives in a particular single-family dwelling is 11.8 years.
- The 2005 mean starting salary in sales for a graduate of a four-year college is \$37,130.
- Thirty-five percent of retirees in the upper Midwest sell their home and move to a warm climate within 1 year of their retirement.
- Eighty percent of those who regularly play the state lotteries never win more than \$100 in any one play.

This chapter and several of the following chapters are concerned with statistical hypothesis testing. We begin by defining what we mean by a statistical hypothesis and statistical hypothesis testing. Next, we outline the steps in statistical hypothesis testing. Then we conduct tests of hypothesis for means and proportions. In the last section of the chapter, we describe possible errors due to sampling in hypothesis testing.

What Is a Hypothesis?

A hypothesis is a statement about a population parameter.

A hypothesis is a statement about a population. Data are then used to check the reasonableness of the statement. To begin we need to define the word *hypothesis*. In the United States legal system, a person is innocent until proven guilty. A jury hypothesizes that a person charged with a crime is innocent and subjects this hypothesis to verification by reviewing the evidence and hearing testimony before reaching a verdict. In a similar sense, a patient goes to a physician and reports various symptoms. On the basis of the symptoms, the physician will order certain diagnostic tests, then, according to the symptoms and the test results, determine the treatment to be followed.

In statistical analysis we make a claim, that is, state a hypothesis, collect data, then use the data to test the assertion. We define a statistical hypothesis as follows.

HYPOTHESIS A statement about a population parameter subject to verification.



Statistics in Action

LASIK is a 15-minute surgical procedure that uses a laser to reshape an eye's cornea with the goal of improving eyesight. Research shows that about 5 percent of all surgeries involve complications such as glare, corneal haze, over-correction or under-correction of vision, and loss of vision. In a statistical sense, the research tests a null hypothesis that the surgery will not improve eyesight with the alternative hypothesis that the surgery will improve eyesight. The sample data of LASIK surgery shows that 5 percent of all cases result in complications. The 5 percent represents a Type I error rate. When a person decides to have the surgery, he or she expects to reject the null hypothesis. In 5 percent of future cases, this expectation will not be met. (Source: American Academy of Ophthalmology Journal, San Francisco, Vol. 16, no. 43.)

In most cases the population is so large that it is not feasible to study all the items, objects, or persons in the population. For example, it would not be possible to contact every systems analyst in the United States to find his or her monthly income. Likewise, the quality assurance department at Cooper Tire cannot check each tire produced to determine whether it will last more than 60,000 miles.

As noted in Chapter 8, an alternative to measuring or interviewing the entire population is to take a sample from the population. We can, therefore, test a statement to determine whether the sample does or does not support the statement concerning the population.

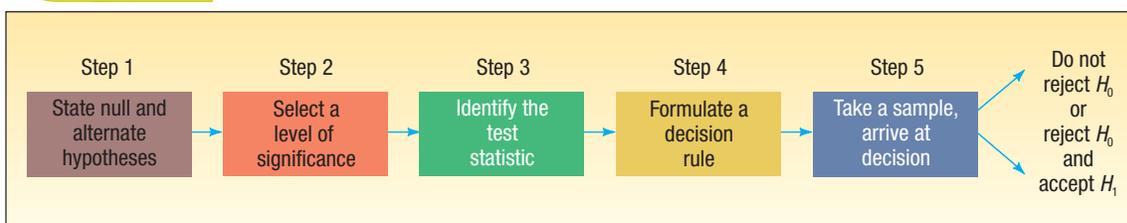
What Is Hypothesis Testing?

The terms *hypothesis testing* and *testing a hypothesis* are used interchangeably. Hypothesis testing starts with a statement, or assumption, about a population parameter—such as the population mean. As noted, this statement is referred to as a *hypothesis*. A hypothesis might be that the mean monthly commission of sales associates in retail electronics stores, such as Circuit City, is \$2,000. We cannot contact all these sales associates to ascertain that the mean is in fact \$2,000. The cost of locating and interviewing every electronics sales associate in the United States would be exorbitant. To test the validity of the assumption ($\mu = \$2,000$), we must select a sample from the population of all electronics sales associates, calculate sample statistics, and based on certain decision rules accept or reject the hypothesis. A sample mean of \$1,000 for the electronics sales associates would certainly cause rejection of the hypothesis. However, suppose the sample mean is \$1,995. Is that close enough to \$2,000 for us to accept the assumption that the population mean is \$2,000? Can we attribute the difference of \$5 between the two means to sampling error, or is that difference statistically significant?

HYPOTHESIS TESTING A procedure based on sample evidence and probability theory to determine whether the hypothesis is a reasonable statement.

Five-Step Procedure for Testing a Hypothesis

There is a five-step procedure that systematizes hypothesis testing; when we get to step 5, we are ready to reject or not reject the hypothesis. However, hypothesis testing as used by statisticians does not provide proof that something is true, in the manner in which a mathematician “proves” a statement. It does provide a kind of “proof beyond a reasonable doubt,” in the manner of the court system. Hence, there are specific rules of evidence, or procedures, that are followed. The steps are shown in the following diagram. We will discuss in detail each of the steps.



One-Sample Tests of Hypothesis

333

Step 1: State the Null Hypothesis (H_0) and the
Alternate Hypothesis (H_1)Five-step systematic
procedure

The first step is to state the hypothesis being tested. It is called the **null hypothesis**, designated H_0 , and read “*H sub zero*.” The capital letter H stands for hypothesis, and the subscript zero implies “no difference.” There is usually a “not” or a “no” term in the null hypothesis, meaning that there is “no change.” For example, the null hypothesis is that the mean number of miles driven on the steel-belted tire is not different from 60,000. The null hypothesis would be written $H_0: \mu = 60,000$. Generally speaking, the null hypothesis is developed for the purpose of testing. We either reject or fail to reject the null hypothesis. The null hypothesis is a statement that is not rejected unless our sample data provide convincing evidence that it is false.

We should emphasize that, if the null hypothesis is not rejected on the basis of the sample data, we cannot say that the null hypothesis is true. To put it another way, failing to reject the null hypothesis does not prove that H_0 is true, it means we have *failed to disprove* H_0 . To prove without any doubt the null hypothesis is true, the population parameter would have to be known. To actually determine it, we would have to test, survey, or count every item in the population. This is usually not feasible. The alternative is to take a sample from the population.

State the null hypothesis
and the alternative
hypothesis.

It should also be noted that we often begin the null hypothesis by stating, “There is no *significant* difference between . . .,” or “The mean impact strength of the glass is not *significantly* different from . . .” When we select a sample from a population, the sample statistic is usually numerically different from the hypothesized population parameter. As an illustration, suppose the hypothesized impact strength of a glass plate is 70 psi, and the mean impact strength of a sample of 12 glass plates is 69.5 psi. We must make a decision about the difference of 0.5 psi. Is it a true difference, that is, a significant difference, or is the difference between the sample statistic (69.5) and the hypothesized population parameter (70.0) due to chance (sampling)? As noted, to answer this question we conduct a test of significance, commonly referred to as a test of hypothesis. To define what is meant by a null hypothesis:

NULL HYPOTHESIS A statement about the value of a population parameter developed for the purpose of testing numerical evidence.

The **alternate hypothesis** describes what you will conclude if you reject the null hypothesis. It is written H_1 and is read “*H sub one*.” It is also referred to as the research hypothesis. The alternate hypothesis is accepted if the sample data provide us with enough statistical evidence that the null hypothesis is false.

ALTERNATE HYPOTHESIS A statement that is accepted if the sample data provide sufficient evidence that the null hypothesis is false.

The following example will help clarify what is meant by the null hypothesis and the alternate hypothesis. A recent article indicated the mean age of U.S. commercial aircraft is 15 years. To conduct a statistical test regarding this statement, the first step is to determine the null and the alternate hypotheses. The null hypothesis represents the current or reported condition. It is written $H_0: \mu = 15$. The alternate hypothesis is that the statement is not true, that is, $H_1: \mu \neq 15$. It is important to remember that no matter how the problem is stated, *the null hypothesis will always contain the equal sign*. The equal sign (=) will never appear in the alternate hypothesis. Why? Because the null hypothesis is the statement being tested, and we need a specific value to include in our calculations. We turn to the alternate hypothesis only if the data suggests the null hypothesis is untrue.

Step 2: Select a Level of Significance

After setting up the null hypothesis and alternate hypothesis, the next step is to state the level of significance.

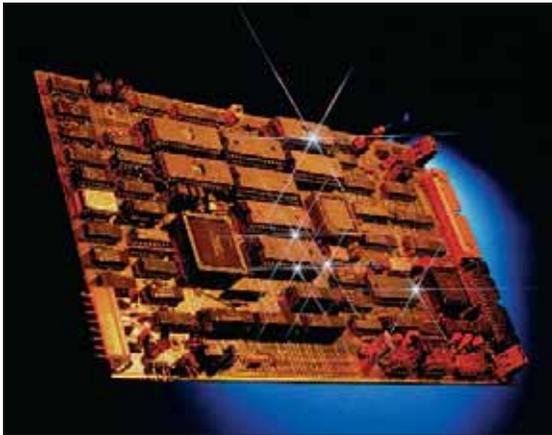
LEVEL OF SIGNIFICANCE The probability of rejecting the null hypothesis when it is true.

Select a level of
significance or risk.

The level of significance is designated α , the Greek letter alpha. It is also sometimes called the level of risk. This may be a more appropriate term because it is the risk you take of rejecting the null hypothesis when it is really true.

There is no one level of significance that is applied to all tests. A decision is made to use the .05 level (often stated as the 5 percent level), the .01 level, the .10 level, or any other level between 0 and 1. Traditionally, the .05 level is selected for consumer research projects, .01 for quality assurance, and .10 for political polling. You, the researcher, must decide on the level of significance *before* formulating a decision rule and collecting sample data.

To illustrate how it is possible to reject a true hypothesis, suppose a firm manufacturing personal computers uses a large number of printed circuit boards. Suppliers



bid on the boards, and the one with the lowest bid is awarded a sizable contract. Suppose the contract specifies that the computer manufacturer's quality-assurance department will sample all incoming shipments of circuit boards. If more than 6 percent of the boards sampled are substandard, the shipment will be rejected. The null hypothesis is that the incoming shipment of boards contains 6 percent or less substandard boards. The alternate hypothesis is that more than 6 percent of the boards are defective.

A sample of 50 circuit boards received July 21 from Allied Electronics revealed that 4 boards, or 8 percent, were substandard. The shipment was rejected because it exceeded the maximum of 6 percent substandard printed circuit boards. If the shipment was actually substandard, then the decision to return the boards to the supplier was correct. However, suppose

the 4 substandard printed circuit boards selected in the sample of 50 were the only substandard boards in the shipment of 4,000 boards. Then only $\frac{1}{10}$ of 1 percent were defective ($4/4,000 = .001$). In that case, less than 6 percent of the entire shipment was substandard and rejecting the shipment was an error. In terms of hypothesis testing, we rejected the null hypothesis that the shipment was not substandard when we should have accepted the null hypothesis. By rejecting a true null hypothesis, we committed a Type I error. The probability of committing a Type I error is α .

TYPE I ERROR Rejecting the null hypothesis, H_0 , when it is true.

The probability of committing another type of error, called a Type II error, is designated by the Greek letter beta (β).

TYPE II ERROR Accepting the null hypothesis when it is false.

The firm manufacturing personal computers would commit a Type II error if, unknown to the manufacturer, an incoming shipment of printed circuit boards from Allied Electronics contained 15 percent substandard boards, yet the shipment

One-Sample Tests of Hypothesis

was accepted. How could this happen? Suppose 2 of the 50 boards in the sample (4 percent) tested were substandard, and 48 of the 50 were good boards. According to the stated procedure, because the sample contained less than 6 percent substandard boards, the shipment was accepted. It could be that *by chance* the 48 good boards selected in the sample were the only acceptable ones in the entire shipment consisting of thousands of boards!

In retrospect, the researcher cannot study every item or individual in the population. Thus, there is a possibility of two types of error—a Type I error, wherein the null hypothesis is rejected when it should have been accepted, and a Type II error, wherein the null hypothesis is not rejected when it should have been rejected.

We often refer to the probability of these two possible errors as *alpha*, α , and *beta*, β . Alpha (α) is the probability of making a Type I error, and beta (β) is the probability of making a Type II error.

The following table summarizes the decisions the researcher could make and the possible consequences.

Null Hypothesis	Researcher	
	Does Not Reject H_0	Rejects H_0
H_0 is true	Correct decision	Type I error
H_0 is false	Type II error	Correct decision

Step 3: Select the Test Statistic

There are many test statistics. In this chapter we use both z and t as the test statistic. In other chapters we will use such test statistics as F and χ^2 , called chi-square.

TEST STATISTIC A value, determined from sample information, used to determine whether to reject the null hypothesis.

In hypothesis testing for the mean (μ) when σ is known, the test statistic z is computed by:

TESTING A MEAN, σ KNOWN

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

[10–1]

The z value is based on the sampling distribution of \bar{X} , which follows the normal distribution with a mean ($\mu_{\bar{X}}$) equal to μ , and a standard deviation $\sigma_{\bar{X}}$, which is equal to σ/\sqrt{n} . We can thus determine whether the difference between \bar{X} and μ is statistically significant by finding the number of standard deviations \bar{X} is from μ , using formula (10–1).

Step 4: Formulate the Decision Rule

The decision rule states the conditions when H_0 is rejected.

A decision rule is a statement of the specific conditions under which the null hypothesis is rejected and the conditions under which it is not rejected. The region or area of rejection defines the location of all those values that are so large or so small that the probability of their occurrence under a true null hypothesis is rather remote.

Chart 10–1 portrays the rejection region for a test of significance that will be conducted later in the chapter.



Statistics in Action

During World War II, allied military planners needed estimates of the number of German tanks. The information provided by traditional spying methods was not reliable, but statistical methods proved to be valuable. For example, espionage and reconnaissance led analysts to estimate that 1,550 tanks were produced during June 1941. However, using the serial numbers of captured tanks and statistical analysis, military planners estimated 244. The actual number produced, as determined from German production records, was 271. The estimate using statistical analysis turned out to be much more accurate. A similar type of analysis was used to estimate the number of Iraqi tanks destroyed during Desert Storm.

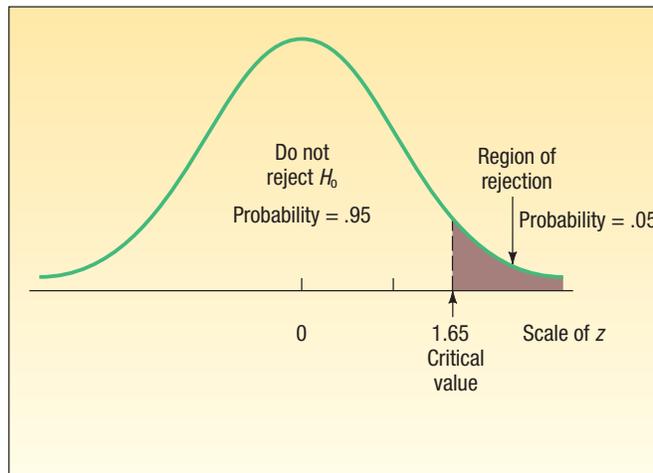


CHART 10–1 Sampling Distribution of the Statistic z , a Right-Tailed Test, .05 Level of Significance

Note in the chart that:

1. The area where the null hypothesis is not rejected is to the left of 1.65. We will explain how to get the 1.65 value shortly.
2. The area of rejection is to the right of 1.65.
3. A one-tailed test is being applied. (This will also be explained later.)
4. The .05 level of significance was chosen.
5. The sampling distribution of the statistic z follows the normal probability distribution.
6. The value 1.65 separates the regions where the null hypothesis is rejected and where it is not rejected.
7. The value 1.65 is the **critical value**.

CRITICAL VALUE The dividing point between the region where the null hypothesis is rejected and the region where it is not rejected.

Step 5: Make a Decision

The fifth and final step in hypothesis testing is computing the test statistic, comparing it to the critical value, and making a decision to reject or not to reject the null hypothesis. Referring to Chart 10–1, if, based on sample information, z is computed to be 2.34, the null hypothesis is rejected at the .05 level of significance. The decision to reject H_0 was made because 2.34 lies in the region of rejection, that is, beyond 1.65. We would reject the null hypothesis, reasoning that it is highly improbable that a computed z value this large is due to sampling error (chance).

Had the computed value been 1.65 or less, say 0.71, the null hypothesis would not be rejected. It would be reasoned that such a small computed value could be attributed to chance, that is, sampling error.

As noted, only one of two decisions is possible in hypothesis testing—either accept or reject the null hypothesis. Instead of “accepting” the null hypothesis, H_0 , some researchers prefer to phrase the decision as: “Do not reject H_0 ,” “We fail to reject H_0 ,” or “The sample results do not allow us to reject H_0 .”

It should be reemphasized that there is always a possibility that the null hypothesis is rejected when it should not be rejected (a Type I error). Also, there is a definable chance that the null hypothesis is accepted when it should be rejected (a Type II error).

One-Sample Tests of Hypothesis

337

Before actually conducting a test of hypothesis, we will differentiate between a one-tailed test of significance and a two-tailed test.

SUMMARY OF THE STEPS IN HYPOTHESIS TESTING

1. Establish the null hypothesis (H_0) and the alternate hypothesis (H_1).
2. Select the level of significance, that is α .
3. Select an appropriate test statistic.
4. Formulate a decision rule based on steps 1, 2, and 3 above.
5. Make a decision regarding the null hypothesis based on the sample information. Interpret the results of the test.

One-Tailed and Two-Tailed Tests of Significance

Refer to Chart 10–1. It depicts a one-tailed test. The region of rejection is only in the right (upper) tail of the curve. To illustrate, suppose that the packaging department at General Foods Corporation is concerned that some boxes of Grape Nuts are significantly overweight. The cereal is packaged in 453-gram boxes, so the null hypothesis is $H_0: \mu \leq 453$. This is read, “the population mean (μ) is equal to or less than 453.” The alternate hypothesis is, therefore, $H_1: \mu > 453$. This is read, “ μ is greater than 453.” Note that the inequality sign in the alternate hypothesis ($>$) points to the region of rejection in the upper tail. (See Chart 10–1.) Also note that the null hypothesis includes the equal sign. That is, $H_0: \mu \leq 453$. The equality condition always appears in H_0 , never in H_1 .

Chart 10–2 portrays a situation where the rejection region is in the left (lower) tail of the normal distribution. As an illustration, consider the problem of automobile manufacturers, large automobile leasing companies, and other organizations that purchase large quantities of tires. They want the tires to average, say, 60,000 miles of wear under normal usage. They will, therefore, reject a shipment of tires if tests reveal that the mean life of the tires is significantly below 60,000 miles. They gladly accept a shipment if the mean life is greater than 60,000 miles! They are not concerned with this possibility, however. They are concerned only if they have sample evidence to conclude that the tires will average less than 60,000 miles of useful life. Thus, the test is set up to satisfy the concern of the automobile manufacturers that *the mean life of the tires is*

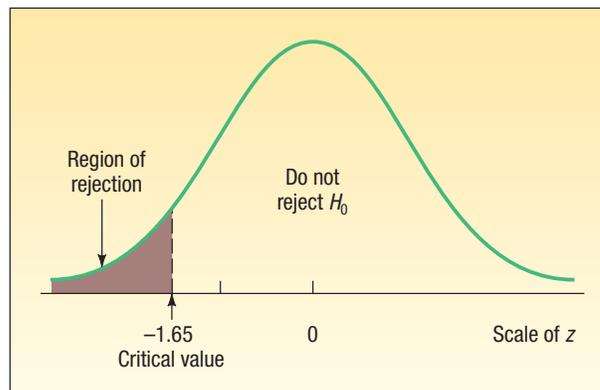


CHART 10–2 Sampling Distribution for the Statistic z , Left-Tailed Test,
.05 Level of Significance

Test is one-tailed if H_1 states $\mu >$ or $\mu <$

If H_1 states a direction, test is one-tailed

no less than 60,000 miles. This statement appears in the alternate hypothesis. The null and alternate hypotheses in this case are written $H_0: \mu \geq 60,000$ and $H_1: \mu < 60,000$.

One way to determine the location of the rejection region is to look at the direction in which the inequality sign in the alternate hypothesis is pointing (either $<$ or $>$). In this problem it is pointing to the left, and the rejection region is therefore in the left tail.

In summary, a test is *one-tailed* when the alternate hypothesis, H_1 , states a direction, such as:

H_0 : The mean income of women stockbrokers is *less than or equal to* \$65,000 per year.

H_1 : The mean income of women stockbrokers is *greater than* \$65,000 per year.

If no direction is specified in the alternate hypothesis, we use a *two-tailed* test. Changing the previous problem to illustrate, we can say:

H_0 : The mean income of women stockbrokers is \$65,000 per year.

H_1 : The mean income of women stockbrokers is *not equal to* \$65,000 per year.

If the null hypothesis is rejected and H_1 accepted in the two-tailed case, the mean income could be significantly greater than \$65,000 per year or it could be significantly less than \$65,000 per year. To accommodate these two possibilities, the 5 percent area of rejection is divided equally into the two tails of the sampling distribution (2.5 percent each). Chart 10–3 shows the two areas and the critical values. Note that the total area in the normal distribution is 1.0000, found by $.9500 + .0250 + .0250$.

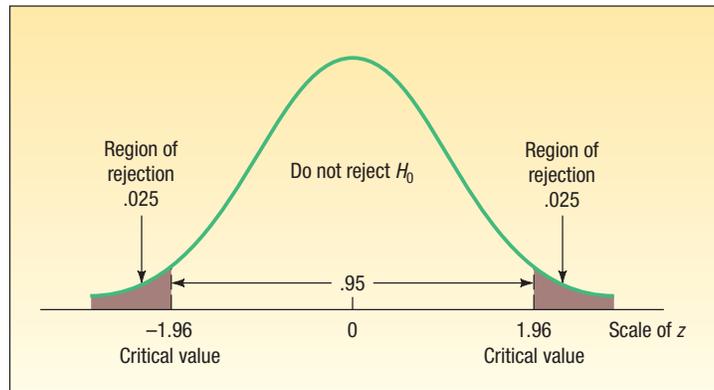


CHART 10–3 Regions of Nonrejection and Rejection for a Two-Tailed Test, .05 Level of Significance

Testing for a Population Mean: Known Population Standard Deviation

A Two-Tailed Test

An example will show the details of the five-step hypothesis testing procedure. We also wish to use a two-tailed test. That is, we are *not* concerned whether the sample results are larger or smaller than the proposed population mean. Rather, we are interested in whether it is *different from* the proposed value for the population mean. We begin, as we did in the previous chapter, with a situation in which we have historical information about the population and in fact know its standard deviation.

One-Sample Tests of Hypothesis

Example



Jamestown Steel Company manufactures and assembles desks and other office equipment at several plants in western New York State. The weekly production of the Model A325 desk at the Fredonia Plant follows a normal probability distribution with a mean of 200 and a standard deviation of 16. Recently, because of market expansion, new production methods have been introduced and new employees hired. The vice president of manufacturing would like to investigate whether there has been a *change* in the weekly production of the Model A325 desk. Is the mean number of desks produced at the Fredonia Plant *different from* 200 at the .01 significance level?

Solution

We use the statistical hypothesis testing procedure to investigate whether the production rate has changed from 200 per week.

Step 1: State the null hypothesis and the alternate hypothesis. The null hypothesis is “The population mean is 200.” The alternate hypothesis is “The mean is different from 200” or “The mean is not 200.” These two hypotheses are written:

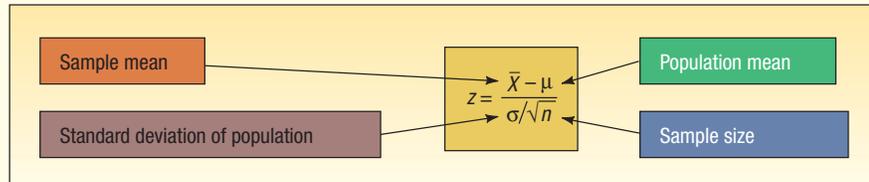
$$H_0: \mu = 200$$

$$H_1: \mu \neq 200$$

This is a *two-tailed test* because the alternate hypothesis does not state a direction. In other words, it does not state whether the mean production is greater than 200 or less than 200. The vice president wants only to find out whether the production rate is different from 200.

Step 2: Select the level of significance. As noted, the .01 level of significance is used. This is α , the probability of committing a Type I error, and it is the probability of rejecting a true null hypothesis.

Step 3: Select the test statistic. The test statistic for a mean when σ is known is z . It was discussed at length in Chapter 7. Transforming the production data to standard units (z values) permits their use not only in this problem but also in other hypothesis-testing problems. Formula (10–1) for z is repeated below with the various letters identified.



[10–1]

Formula for the test
statistic

Step 4: Formulate the decision rule. The decision rule is formulated by finding the critical values of z from Appendix B.1. Since this is a two-tailed test, half of .01, or .005, is placed in each tail. The area where H_0 is not rejected, located between the two tails, is therefore .99. Appendix B.1 is based on half of the area under the curve, or .5000. Then, .5000 – .0050 is .4950, so .4950 is the area between 0 and the critical value. Locate .4950 in the body of the table. The value nearest to .4950 is .4951. Then read the critical value in the row and column corresponding to .4951. It is 2.58. For your convenience, Appendix B.1, Areas under the Normal Curve, is repeated in the inside back cover.

All the facets of this problem are shown in the diagram in Chart 10–4.

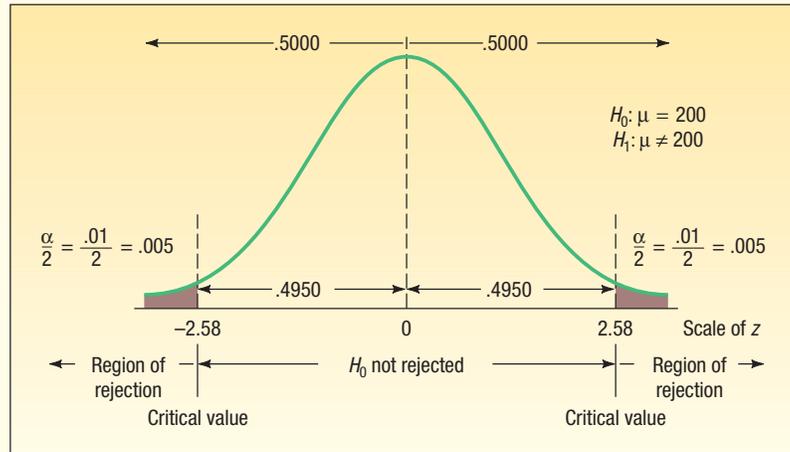


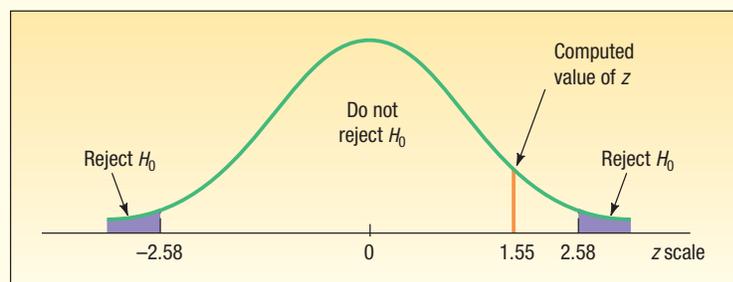
CHART 10–4 Decision Rule for the .01 Significance Level

The decision rule is, therefore: Reject the null hypothesis and accept the alternate hypothesis (which states that the population mean is not 200) if the computed value of z is not between -2.58 and $+2.58$. Do not reject the null hypothesis if z falls between -2.58 and $+2.58$.

Step 5: Make a decision and interpret the result. Take a sample from the population (weekly production), compute z , apply the decision rule, and arrive at a decision to reject H_0 or not to reject H_0 . The mean number of desks produced last year (50 weeks, because the plant was shut down 2 weeks for vacation) is 203.5. The standard deviation of the population is 16 desks per week. Computing the z value from formula (10–1):

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{203.5 - 200}{16/\sqrt{50}} = 1.55$$

Because 1.55 does not fall in the rejection region, H_0 is not rejected. We conclude that the population mean is *not* different from 200. So we would report to the vice president of manufacturing that the sample evidence does not show that the production rate at the Fredonia Plant has changed from 200 per week. The difference of 3.5 units between the historical weekly production rate and that last year can reasonably be attributed to sampling error. This information is summarized in the following chart.



One-Sample Tests of Hypothesis

341

Did we prove that the assembly rate is still 200 per week? Not really. What we did, technically, was *fail to disprove the null hypothesis*. Failing to disprove the hypothesis that the population mean is 200 is not the same thing as proving it to be true. As we suggested in the chapter introduction, the conclusion is analogous to the American judicial system. To explain, suppose a person is accused of a crime but is acquitted by a jury. If a person is acquitted of a crime, the conclusion is that there was not enough evidence to prove the person guilty. The trial did not prove that the individual was innocent, only that there was not enough evidence to prove the defendant guilty. That is what we do in statistical hypothesis testing when we do not reject the null hypothesis. The correct interpretation is that we have failed to disprove the null hypothesis.

We selected the significance level, .01 in this case, before setting up the decision rule and sampling the population. This is the appropriate strategy. The significance level should be set by the investigator, but it should be determined *before* gathering the sample evidence and not changed based on the sample evidence.

Comparing confidence intervals and hypothesis testing.

How does the hypothesis testing procedure just described compare with that of confidence intervals discussed in the previous chapter? When we conducted the test of hypothesis regarding the production of desks we changed the units from desks per week to a z value. Then we compared the computed value of the test statistic (1.55) to that of the critical values (-2.58 and 2.58). Because the computed value was in the region where the null hypothesis was not rejected, we concluded that the population mean could be 200. To use the confidence interval approach, on the other hand, we would develop a confidence interval, based on formula (9–1). See page 298. The interval would be from 197.66 to 209.34, found by $203.5 \pm 2.58(16/\sqrt{50})$. Note that the proposed population value, 200, is within this interval. Hence, we would conclude that the population mean could reasonably be 200.

In general, H_0 is rejected if the confidence interval does not include the hypothesized value. If the confidence interval includes the hypothesized value, then H_0 is not rejected. So the “do not reject region” for a test of hypothesis is equivalent to the proposed population value occurring in the confidence interval. The primary difference between a confidence interval and the “do not reject” region for a hypothesis test is whether the interval is centered around the sample statistic, such as \bar{X} , as in the confidence interval, or around 0, as in the test of hypothesis.

Self-Review 10–1



Heinz, a manufacturer of ketchup, uses a particular machine to dispense 16 ounces of its ketchup into containers. From many years of experience with the particular dispensing machine, Heinz knows the amount of product in each container follows a normal distribution with a mean of 16 ounces and a standard deviation of 0.15 ounce. A sample of 50 containers filled last hour revealed the mean amount per container was 16.017 ounces. Does this evidence suggest that the mean amount dispensed is different

from 16 ounces? Use the .05 significance level.

- State the null hypothesis and the alternate hypothesis.
- What is the probability of a Type I error?
- Give the formula for the test statistic.
- State the decision rule.
- Determine the value of the test statistic.
- What is your decision regarding the null hypothesis?
- Interpret, in a single sentence, the result of the statistical test.

A One-Tailed Test

In the previous example, we emphasized that we were concerned only with reporting to the vice president whether there had been a change in the mean number of desks assembled at the Fredonia Plant. We were not concerned with whether the change was an increase or a decrease in the production.

To illustrate a one-tailed test, let's change the problem. Suppose the vice president wants to know whether there has been an *increase* in the number of units assembled. Can we conclude, because of the improved production methods, that the mean number of desks assembled in the last 50 weeks was more than 200? Look at the difference in the way the problem is formulated. In the first case we wanted to know whether there was a *difference* in the mean number assembled, but now we want to know whether there has been an *increase*. Because we are investigating different questions, we will set our hypotheses differently. The biggest difference occurs in the alternate hypothesis. Before, we stated the alternate hypothesis as "different from"; now we want to state it as "greater than." In symbols:

A two-tailed test:

$$H_0: \mu = 200$$

$$H_1: \mu \neq 200$$

A one-tailed test:

$$H_0: \mu \leq 200$$

$$H_1: \mu > 200$$

The critical values for a one-tailed test are different from a two-tailed test at the same significance level. In the previous example, we split the significance level in half and put half in the lower tail and half in the upper tail. In a one-tailed test we put all the rejection region in one tail. See Chart 10–5.

For the one-tailed test, the critical value is 2.33, found by: (1) subtracting .01 from .5000 and (2) finding the z value corresponding to .4900.

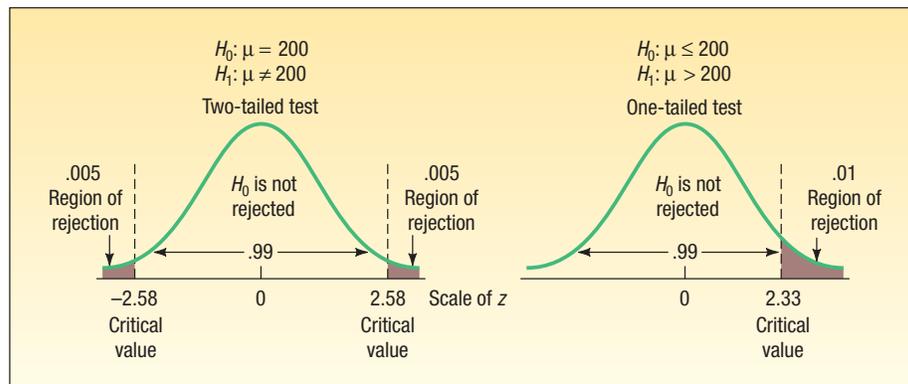


CHART 10–5 Rejection Regions for Two-Tailed and One-Tailed Tests, $\alpha = .01$

p-Value in Hypothesis Testing

In testing a hypothesis, we compare the test statistic to a critical value. A decision is made to either reject the null hypothesis or not to reject it. So, for example, if the critical value is 1.96 and the computed value of the test statistic is 2.19, the decision is to reject the null hypothesis.

In recent years, spurred by the availability of computer software, additional information is often reported on the strength of the rejection or acceptance. That is, how confident are we in rejecting the null hypothesis? This approach reports the prob-

One-Sample Tests of Hypothesis

343



Statistics in Action

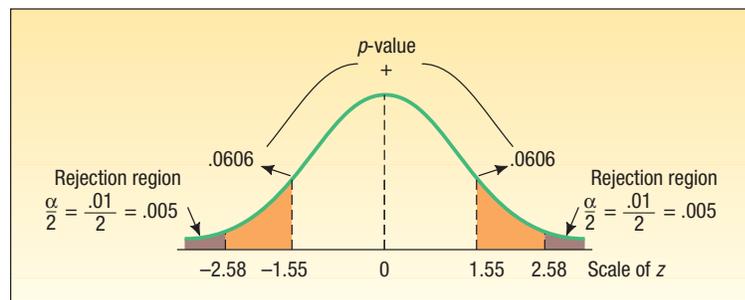
There is a difference between *statistically significant* and *practically significant*. To explain, suppose we develop a new diet pill and test it on 100,000 people. We conclude that the typical person taking the pill for two years lost one pound. Do you think many people would be interested in taking the pill to lose one pound? The results of using the new pill were statistically significant but not practically significant.

ability (assuming that the null hypothesis is true) of getting a value of the test statistic at least as extreme as the value actually obtained. This process compares the probability, called the ***p*-value**, with the significance level. If the *p*-value is smaller than the significance level, H_0 is rejected. If it is larger than the significance level, H_0 is not rejected.

***p*-VALUE** The probability of observing a sample value as extreme as, or more extreme than, the value observed, given that the null hypothesis is true.

Determining the *p*-value not only results in a decision regarding H_0 , but it gives us additional insight into the strength of the decision. A very small *p*-value, such as .0001, indicates that there is little likelihood the H_0 is true. On the other hand, a *p*-value of .2033 means that H_0 is not rejected, and there is little likelihood that it is false.

How do we compute the *p*-value? To illustrate we will use the example in which we tested the null hypothesis that the mean number of desks produced per week at Fredonia was 200. We did not reject the null hypothesis, because the *z* value of 1.55 fell in the region between -2.58 and 2.58 . We agreed not to reject the null hypothesis if the computed value of *z* fell in this region. The probability of finding a *z* value of 1.55 or more is .0606, found by $.5000 - .4394$. To put it another way, the probability of obtaining an \bar{X} greater than 203.5 if $\mu = 200$ is .0606. To compute the *p*-value, we need to be concerned with the region less than -1.55 as well as the values greater than 1.55 (because the rejection region is in both tails). The two-tailed *p*-value is .1212, found by $2(.0606)$. The *p*-value of .1212 is greater than the significance level of .01 decided upon initially, so H_0 is not rejected. The details are shown in the following graph. In general, the area is doubled in a two-sided test. Then the *p*-value can easily be compared with the significance level. The same decision rule is used as in the one-sided test.



A *p*-value is a way to express the likelihood that H_0 is false. But how do we interpret a *p*-value? We have already said that if the *p*-value is less than the significance level, then we reject H_0 ; if it is greater than the significance level, then we do not reject H_0 . Also, if the *p*-value is very large, then it is likely that H_0 is true. If the *p*-value is small, then it is likely that H_0 is not true. The following box will help to interpret *p*-values.

INTERPRETING THE WEIGHT OF EVIDENCE AGAINST H_0 If the *p*-value is less than

- .10, we have *some* evidence that H_0 is not true.
- .05, we have *strong* evidence that H_0 is not true.
- .01, we have *very strong* evidence that H_0 is not true.
- .001, we have *extremely strong* evidence that H_0 is not true.

Self-Review 10–2



Refer to Self-Review 10–1.

- Suppose the next to the last sentence is changed to read: Does this evidence suggest that the mean amount dispensed is *more than* 16 ounces? State the null hypothesis and the alternate hypothesis under these conditions.
- What is the decision rule under the new conditions stated in part (a)?
- A second sample of 50 filled containers revealed the mean to be 16.040 ounces. What is the value of the test statistic for this sample?
- What is your decision regarding the null hypothesis?
- Interpret, in a single sentence, the result of the statistical test.
- What is the p -value? What is your decision regarding the null hypothesis based on the p -value? Is this the same conclusion reached in part (d)?

Exercises

For Exercises 1–4 answer the questions: (a) Is this a one- or two-tailed test? (b) What is the decision rule? (c) What is the value of the test statistic? (d) What is your decision regarding H_0 ? (e) What is the p -value? Interpret it.

- The following information is available.

$$H_0: \mu = 50$$

$$H_1: \mu \neq 50$$

The sample mean is 49, and the sample size is 36. The population standard deviation is 5. Use the .05 significance level.

- The following information is available.

$$H_0: \mu \leq 10$$

$$H_1: \mu > 10$$

The sample mean is 12 for a sample of 36. The population standard deviation is 3. Use the .02 significance level.

- A sample of 36 observations is selected from a normal population. The sample mean is 21, and the population standard deviation is 5. Conduct the following test of hypothesis using the .05 significance level.

$$H_0: \mu \leq 20$$

$$H_1: \mu > 20$$

- A sample of 64 observations is selected from a normal population. The sample mean is 215, and the population standard deviation is 15. Conduct the following test of hypothesis using the .03 significance level.

$$H_0: \mu \geq 220$$

$$H_1: \mu < 220$$

For Exercises 5–8: (a) State the null hypothesis and the alternate hypothesis. (b) State the decision rule. (c) Compute the value of the test statistic. (d) What is your decision regarding H_0 ? (e) What is the p -value? Interpret it.

- The manufacturer of the X-15 steel-belted radial truck tire claims that the mean mileage the tire can be driven before the tread wears out is 60,000 miles. The population standard deviation of the mileage is 5,000 miles. Crosset Truck Company bought 48 tires and found that the mean mileage for its trucks is 59,500 miles. Is Crosset's experience different from that claimed by the manufacturer at the .05 significance level?
- The MacBurger restaurant chain claims that the mean waiting time of customers is 3 minutes with a population standard deviation of 1 minute. The quality-assurance department found in a sample of 50 customers at the Warren Road MacBurger that the mean waiting time was 2.75 minutes. At the .05 significance level, can we conclude that the mean waiting time is less than 3 minutes?
- A recent national survey found that high school students watched an average (mean) of 6.8 DVDs per month with a population standard deviation of 0.5 hours. A random sample

One-Sample Tests of Hypothesis

345

of 36 college students revealed that the mean number of DVDs watched last month was 6.2. At the .05 significance level, can we conclude that college students watch fewer DVDs a month than high school students?

8. At the time she was hired as a server at the Grumney Family Restaurant, Beth Brigden was told, “You can average more than \$80 a day in tips.” Assume the standard deviation of the population distribution is \$3.24. Over the first 35 days she was employed at the restaurant, the mean daily amount of her tips was \$84.85. At the .01 significance level, can Ms. Brigden conclude that she is earning an average of more than \$80 in tips?

Testing for a Population Mean: Population Standard Deviation Unknown

In the preceding example, we knew σ , the population standard deviation. In most cases, however, the population standard deviation is unknown. Thus, σ must be based on prior studies or estimated by the sample standard deviation, s . The population standard deviation in the following example is not known, so the sample standard deviation is used to estimate σ .

To find the value of the test statistic we use the t distribution and revise formula [10–1] as follows:

TESTING A MEAN, σ UNKNOWN

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

[10–2]

with $n - 1$ degrees of freedom, where:

\bar{X} is the sample mean.

μ is the hypothesized population mean.

s is the sample standard deviation.

n is the number of observations in the sample.

We encountered this same situation when constructing confidence intervals in the previous chapter. See pages 302–304 in Chapter 9. We summarized this problem in Chart 9–3 on page 305. Under these conditions the correct statistical procedure is to replace the standard normal distribution with the t distribution. To review, the major characteristics of the t distribution are:

1. It is a continuous distribution.
2. It is bell-shaped and symmetrical.
3. There is a family of t distributions. Each time the degrees of freedom change, a new distribution is created.
4. As the number of degrees of freedom increases the shape of the t distribution approaches that of the standard normal distribution.
5. The t distribution is flatter, or more spread out, than the standard normal distribution.

The following example shows the details.

Example

The McFarland Insurance Company Claims Department reports the mean cost to process a claim is \$60. An industry comparison showed this amount to be larger than most other insurance companies, so the company instituted cost-cutting measures. To evaluate the effect of the cost-cutting measures, the Supervisor of the Claims Department selected a random sample of 26 claims processed last month. The sample information is reported below.

\$45	\$49	\$62	\$40	\$43	\$61
48	53	67	63	78	64
48	54	51	56	63	69
58	51	58	59	56	57
38	76				

At the .01 significance level is it reasonable to conclude that mean cost to process a claim is now less than \$60?

Solution

We will use the five-step hypothesis testing procedure.

Step 1: State the null hypothesis and the alternate hypothesis. The null hypothesis is that the population mean is at least \$60. The alternate hypothesis is that the population mean is less than \$60. We can express the null and alternate hypotheses as follows:

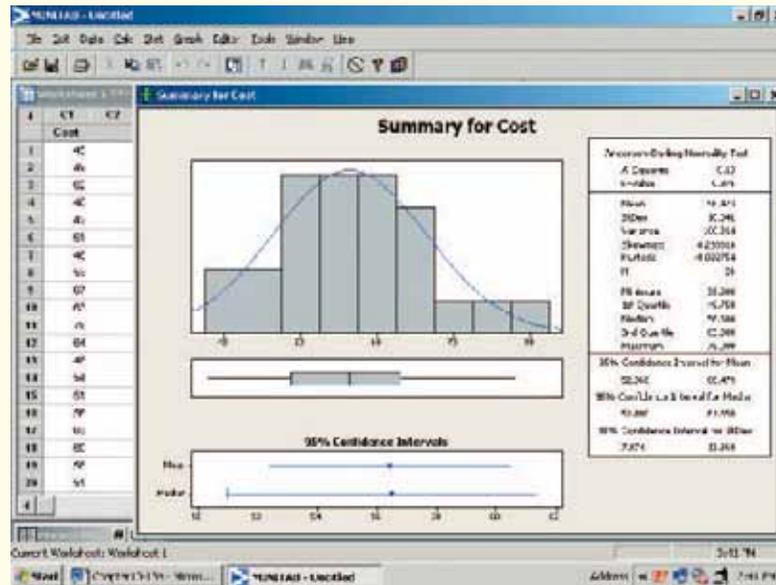
$$H_0: \mu \geq \$60$$

$$H_1: \mu < \$60$$

The test is *one-tailed* because we want to determine whether there has been a *reduction* in the cost. The inequality in the alternate hypothesis points to the region of rejection in the left tail of the distribution.

Step 2: Select the level of significance. We decided on the .01 significance level.

Step 3: Select the test statistic. The test statistic in this situation is the *t* distribution. Why? First it is reasonable to conclude that the distribution of the cost per claim follows the normal distribution. We can confirm this from the histogram on the right-hand side of the following MINITAB output. Observe the normal distribution superimposed on the frequency distribution.



We do not know the standard deviation of the population. So we substitute the sample standard deviation. The value of the test statistic is computed by formula (10–2):

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

Step 4: Formulate the decision rule. The critical values of t are given in Appendix B.2, a portion of which is shown in Table 10–1. Appendix B.2 is also repeated in the back inside cover of the text. The far left column of the table is labeled “ df ” for degrees of freedom. The number of degrees of freedom is the total number of observations in the sample minus the number of populations sampled, written $n - 1$. In this case the number of observations in the sample is 26, and we sampled 1 population, so there are $26 - 1 = 25$ degrees of freedom. To find the critical value, first locate the row with the appropriate degrees of freedom. This row is shaded in Table 10–1. Next, determine whether the test is one-tailed or two-tailed. In this case, we have a one-tailed test, so find the portion of the table that is labeled “one-tailed.” Locate the column with the selected significance level. In this example, the significance level is .01. Move down the column labeled “0.01” until it intersects the row with 25 degrees of freedom. The value is 2.485. Because this is a one-sided test and the rejection region is in the left tail, the critical value is negative. The decision rule is to reject H_0 if the value of t is less than -2.485 .

TABLE 10–1 A Portion of the t Distribution Table

Confidence Intervals						
	80%	90%	95%	98%	99%	99.9%
df	Level of Significance for One-Tailed Test, α					
	0.100	0.050	0.025	0.010	0.005	0.0005
	Level of Significance for Two-Tailed Test, α					
	0.20	0.10	0.05	0.02	0.01	0.001
∴	∴	∴	∴	∴	∴	∴
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.768
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646

Step 5: Make a decision and interpret the result. From the MINITAB output, next to the histogram, the mean cost per claim for the sample of 26 observations is \$56.42. The standard deviation of this sample is \$10.04. We insert these values in formula (10–2) and compute the value of t :

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{\$56.42 - \$60}{\$10.04/\sqrt{26}} = -1.818$$

Because -1.818 lies in the region to the right of the critical value of -2.485 , the null hypothesis is not rejected at the .01 significance level. We have not demonstrated that the cost-cutting measures reduced the mean cost per claim to less than \$60. To put it another way, the difference of \$3.58 ($\$56.42 - \60) between the sample mean and the population mean could be due to sampling error. The computed value of t is shown in Chart 10–6. It is in the region where the null hypothesis is *not* rejected.

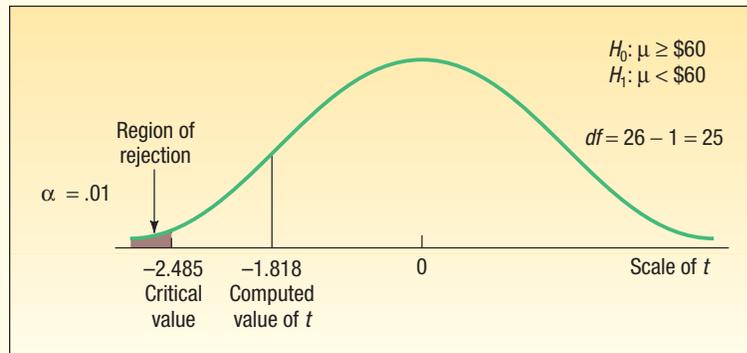


CHART 10-6 Rejection Region, t Distribution, .01 Significance Level

In the previous example the mean and the standard deviation were computed using MINITAB. The following example shows the details when the sample mean and sample standard deviation are calculated from sample data.

Example

The mean length of a small counterbalance bar is 43 millimeters. The production supervisor is concerned that the adjustments of the machine producing the bars have changed. He asks the Engineering Department to investigate. Engineering selects a random sample of 12 bars and measures each. The results are reported below in millimeters.

42	39	42	45	43	40	39	41	40	42	43	42
----	----	----	----	----	----	----	----	----	----	----	----

Is it reasonable to conclude that there has been a change in the mean length of the bars? Use the .02 significance level.

Solution

We begin by stating the null hypothesis and the alternate hypothesis.

$$H_0: \mu = 43$$

$$H_1: \mu \neq 43$$

The alternate hypothesis does not state a direction, so this is a two-tailed test. There are 11 degrees of freedom, found by $n - 1 = 12 - 1 = 11$. The t value is 2.718, found by referring to Appendix B.2 for a two-tailed test, using the .02 significance level, with 11 degrees of freedom. The decision rule is: Reject the null hypothesis if the computed t is to the left of -2.718 or to the right of 2.718 . This information is summarized in Chart 10-7.

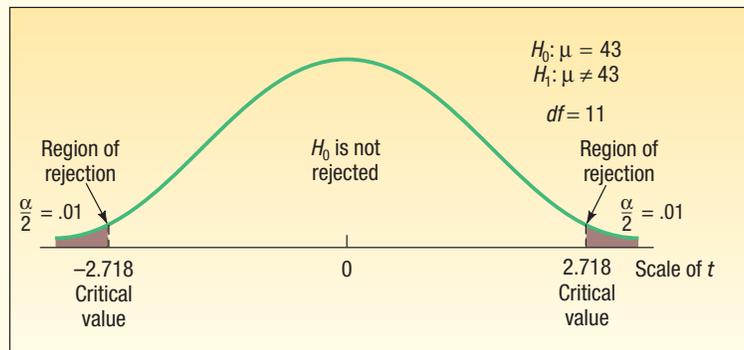


CHART 10-7 Regions of Rejection, Two-Tailed Test, Student's t Distribution, $\alpha = .02$

One-Sample Tests of Hypothesis

We calculate the standard deviation of the sample using formula (3–11). The mean, \bar{X} , is 41.5 millimeters, and the standard deviation, s , is 1.784 millimeters. The details are shown in Table 10–2.

TABLE 10–2 Calculations of the Sample Standard Deviation

X (mm)	$X - \bar{X}$	$(X - \bar{X})^2$
42	0.5	0.25
39	-2.5	6.25
42	0.5	0.25
45	3.5	12.25
43	1.5	2.25
40	-1.5	2.25
39	-2.5	6.25
41	-0.5	0.25
40	-1.5	2.25
42	0.5	0.25
43	1.5	2.25
42	0.5	0.25
498	0	35.00

$$\bar{X} = \frac{498}{12} = 41.5 \text{ mm}$$

$$s = \sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}} = \sqrt{\frac{35}{12 - 1}} = 1.784$$

Now we are ready to compute the value of t , using formula (10–2).

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{41.5 - 43.0}{1.784/\sqrt{12}} = -2.913$$

The null hypothesis that the population mean is 43 millimeters is rejected because the computed t of -2.913 lies in the area to the left of -2.718 . We accept the alternate hypothesis and conclude that the population mean is not 43 millimeters. The machine is out of control and needs adjustment.

Self-Review 10–3



The mean life of a battery used in a digital clock is 305 days. The lives of the batteries follow the normal distribution. The battery was recently modified to last longer. A sample of 20 of the modified batteries had a mean life of 311 days with a standard deviation of 12 days. Did the modification increase the mean life of the battery?

- State the null hypothesis and the alternate hypothesis.
- Show the decision rule graphically. Use the .05 significance level.
- Compute the value of t . What is your decision regarding the null hypothesis? Briefly summarize your results.

Exercises

9. Given the following hypothesis:

$$H_0: \mu \leq 10$$

$$H_1: \mu > 10$$

For a random sample of 10 observations, the sample mean was 12 and the sample standard deviation 3. Using the .05 significance level:

- State the decision rule.
- Compute the value of the test statistic.
- What is your decision regarding the null hypothesis?

10. Given the following hypothesis:

$$H_0: \mu = 400$$

$$H_1: \mu \neq 400$$

For a random sample of 12 observations, the sample mean was 407 and the sample standard deviation 6. Using the .01 significance level:

- State the decision rule.
- Compute the value of the test statistic.
- What is your decision regarding the null hypothesis?

11. The Rocky Mountain district sales manager of Rath Publishing, Inc., a college textbook publishing company, claims that the sales representatives make an average of 40 sales calls per week on professors. Several reps say that this estimate is too low. To investigate, a random sample of 28 sales representatives reveals that the mean number of calls made last week was 42. The standard deviation of the sample is 2.1 calls. Using the .05 significance level, can we conclude that the mean number of calls per salesperson per week is more than 40?
12. The management of White Industries is considering a new method of assembling its golf cart. The present method requires 42.3 minutes, on the average, to assemble a cart. The mean assembly time for a random sample of 24 carts, using the new method, was 40.6 minutes, and the standard deviation of the sample was 2.7 minutes. Using the .10 level of significance, can we conclude that the assembly time using the new method is faster?
13. A spark plug manufacturer claimed that its plugs have a mean life in excess of 22,100 miles. Assume the life of the spark plugs follows the normal distribution. A fleet owner purchased a large number of sets. A sample of 18 sets revealed that the mean life was 23,400 miles and the standard deviation was 1,500 miles. Is there enough evidence to substantiate the manufacturer's claim at the .05 significance level?
14. Most air travelers now use e-tickets. Electronic ticketing allows passengers to not worry about a paper ticket, and it costs the airline companies less to handle than paper ticketing. However, in recent times the airlines have received complaints from passengers regarding their e-tickets, particularly when connecting flights and a change of airlines were involved. To investigate the problem an independent watchdog agency contacted a random sample of 20 airports and collected information on the number of complaints the airport had with e-tickets for the month of March. The information is reported below.

14	14	16	12	12	14	13	16	15	14
12	15	15	14	13	13	12	13	10	13

At the .05 significance level can the watchdog agency conclude the mean number of complaints per airport is less than 15 per month?

- a. What assumption is necessary before conducting a test of hypothesis?
- b. Plot the number of complaints per airport in a frequency distribution or a dot plot. Is it reasonable to conclude that the population follows a normal distribution?
- c. Conduct a test of hypothesis and interpret the results.

A Software Solution



The MINITAB statistical software system, used in earlier chapters and the previous section, provides an efficient way of conducting a one-sample test of hypothesis for a population mean. The steps to generate the following output are shown in the **Software Commands** section at the end of the chapter.

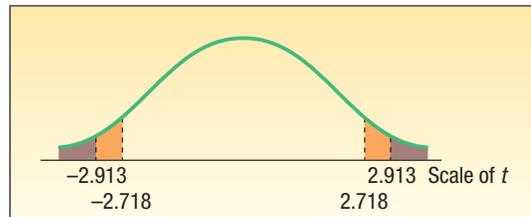
Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Length	20	41.5000	1.5000	0.37500	(40.3560, 42.6000)	-2.91	0.014

An additional feature of most statistical software packages is to report the p -value, which gives additional information on the null hypothesis. The p -value is the probability of a t value as extreme as that computed, given that the null hypoth-

One-Sample Tests of Hypothesis

esis is true. Using the data from the previous counterbalance bar example, the p -value of .014 is the likelihood of a t value of -2.91 or less plus the likelihood of a t value of 2.91 or larger, given a population mean of 43 . Thus, comparing the p -value to the significance level tells us whether the null hypothesis was close to being rejected, barely rejected, and so on.

To explain further, refer to the diagram below. The p -value of .014 is the darker or shaded area and the significance level is the total shaded area. Because the p -value of .014 is less than the significance level of .02, the null hypothesis is rejected. Had the p -value been larger than the significance level—say, .06, .19, or .57—the null hypothesis would not be rejected. If the significance level had initially been selected as .01, the null hypothesis would not be rejected.



In the preceding example the alternate hypothesis was two-sided, so there were rejection areas in both the lower (left) tail and the upper (right) tail. To determine the p -value, it was necessary to determine the area to the left of -2.913 for a t distribution with 11 degrees of freedom and add to it the value of the area to the right of 2.913 , also with 11 degrees of freedom.

What if we were conducting a one-sided test, so that the entire rejection region would be in either the upper or the lower tail? In that case, we would report the area from only the one tail. In the counterbalance example, if H_1 were stated as $\mu < 43$, the inequality would point to the left. Thus, we would have reported the p -value as the area to the left of -2.913 . This value is .007, found by $.014/2$. Thus, the p -value for a one-tailed test would be .007.

How can we estimate a p -value without a computer? To illustrate, recall that, in the example regarding the length of a counterbalance, we rejected the null hypothesis that $\mu = 43$ and accepted the alternate hypothesis that $\mu \neq 43$. The significance level was .02, so logically the p -value is less than .02. To estimate the p -value more accurately, go to Appendix B.2 and find the row with 11 degrees of freedom. The computed t value of 2.913 is between 2.718 and 3.106. (A portion of Appendix B.2 is reproduced as Table 10–3.) The two-tailed significance level corresponding to 2.718

TABLE 10–3 A Portion of Student’s t Distribution

Confidence Intervals						
	80%	90%	95%	98%	99%	99.9%
df	Level of Significance for One-Tailed Test, α					
	0.100	0.050	.0025	0.010	0.005	0.0005
	Level of Significance for Two-Tailed Test, α					
	0.20	0.10	0.05	0.02	0.01	0.001
∴	∴	∴	∴	∴	∴	∴
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073

is .02, and for 3.106 it is .01. Therefore, the p -value is between .01 and .02. The usual practice is to report that the p -value is *less* than the larger of the two significance levels. So we would report, “the p -value is less than .02.”

Self-Review 10–4



A machine is set to fill a small bottle with 9.0 grams of medicine. A sample of eight bottles revealed the following amounts (grams) in each bottle.

9.2	8.7	8.9	8.6	8.8	8.5	8.7	9.0
-----	-----	-----	-----	-----	-----	-----	-----

At the .01 significance level, can we conclude that the mean weight is less than 9.0 grams?

- State the null hypothesis and the alternate hypothesis.
- How many degrees of freedom are there?
- Give the decision rule.
- Compute the value of t . What is your decision regarding the null hypothesis?
- Estimate the p -value.

Exercises

15. Given the following hypothesis:

$$H_0: \mu \geq 20$$

$$H_1: \mu < 20$$

A random sample of five resulted in the following values: 18, 15, 12, 19, and 21. Using the .01 significance level, can we conclude the population mean is less than 20?

- State the decision rule.
 - Compute the value of the test statistic.
 - What is your decision regarding the null hypothesis?
 - Estimate the p -value.
16. Given the following hypothesis:

$$H_0: \mu = 100$$

$$H_1: \mu \neq 100$$

A random sample of six resulted in the following values: 118, 105, 112, 119, 105, and 111. Using the .05 significance level, can we conclude the mean is different from 100?

- State the decision rule.
 - Compute the value of the test statistic.
 - What is your decision regarding the null hypothesis?
 - Estimate the p -value.
17. Experience raising New Jersey Red chickens revealed the mean weight of the chickens at five months is 4.35 pounds. The weights follow the normal distribution. In an effort to increase their weight, a special additive is added to the chicken feed. The subsequent weights of a sample of five-month-old chickens were (in pounds):

4.41	4.37	4.33	4.35	4.30	4.39	4.36	4.38	4.40	4.39
------	------	------	------	------	------	------	------	------	------

At the .01 level, has the special additive increased the mean weight of the chickens? Estimate the p -value.

18. The liquid chlorine added to swimming pools to combat algae has a relatively short shelf life before it loses its effectiveness. Records indicate that the mean shelf life of a 5-gallon jug of chlorine is 2,160 hours (90 days). As an experiment, Holdlonger was added to the chlorine to find whether it would increase the shelf life. A sample of nine jugs of chlorine had these shelf lives (in hours):

2,159	2,170	2,180	2,179	2,160	2,167	2,171	2,181	2,185
-------	-------	-------	-------	-------	-------	-------	-------	-------

At the .025 level, has Holdlonger increased the shelf life of the chlorine? Estimate the p -value.

One-Sample Tests of Hypothesis

353

19. Wyoming fisheries contend that the mean number of cutthroat trout caught during a full day of fly-fishing on the Snake, Buffalo, and other rivers and streams in the Jackson Hole area is 4.0. To make their yearly update, the fishery personnel asked a sample of fly fishermen to keep a count of the number caught during the day. The numbers were: 4, 4, 3, 2, 6, 8, 7, 1, 9, 3, 1, and 6. At the .05 level, can we conclude that the mean number caught is greater than 4.0? Estimate the p -value.
20. Hugger Polls contends that an agent conducts a mean of 53 in-depth home surveys every week. A streamlined survey form has been introduced, and Hugger wants to evaluate its effectiveness. The number of in-depth surveys conducted during a week by a random sample of agents are:

53	57	50	55	58	54	60	52	59	62	60	60	51	59	56
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

At the .05 level of significance, can we conclude that the mean number of interviews conducted by the agents is more than 53 per week? Estimate the p -value.

Tests Concerning Proportions

In the previous chapter we discussed confidence intervals for proportions. We can also conduct a test of hypothesis for a proportion. Recall that a proportion is the ratio of the number of successes to the number of observations. We let X refer to the number of successes and n the number of observations, so the proportion of successes in a fixed number of trials is X/n . Thus, the formula for computing a sample proportion, p , is $p = X/n$. Consider the following potential hypothesis-testing situations.

- Historically, General Motors reports that 70 percent of leased vehicles are returned with less than 36,000 miles. A recent sample of 200 vehicles returned at the end of their lease showed 158 had less than 36,000 miles. Has the proportion increased?
- The American Association of Retired Persons (AARP) reports that 60 percent of retired people under the age of 65 would return to work on a full-time basis if a suitable job were available. A sample of 500 retirees under 65 revealed 315 would return to work. Can we conclude that more than 60 percent would return to work?
- Able Moving and Storage, Inc., advises its clients for long-distance residential moves that their household goods will be delivered in 3 to 5 days from the time they are picked up. Able's records show it is successful 90 percent of the time with this claim. A recent audit revealed it was successful 190 times out of 200. Can the company conclude its success rate has increased?

Some assumptions must be made and conditions met before testing a population proportion. To test a hypothesis about a population proportion, a random sample is chosen from the population. It is assumed that the binomial assumptions discussed in Chapter 6 are met: (1) the sample data collected are the result of counts; (2) the outcome of an experiment is classified into one of two mutually exclusive categories—a “success” or a “failure”; (3) the probability of a success is the same for each trial; and (4) the trials are independent, meaning the outcome of one trial does not affect the outcome of any other trial. The test we will conduct shortly is appropriate when both $n\pi$ and $n(1 - \pi)$ are at least 5. n is the sample size, and p is the population proportion. It takes advantage of the fact that a binomial distribution can be approximated by the normal distribution.

$n\pi$ and $n(1 - \pi)$ must
be at least 5.

Example

Suppose prior elections in a certain state indicated it is necessary for a candidate for governor to receive at least 80 percent of the vote in the northern section of the state to be elected. The incumbent governor is interested in assessing his chances of returning to office and plans to conduct a survey of 2,000 registered voters in the northern section of the state.

Using the hypothesis-testing procedure, assess the governor's chances of reelection.

Solution

This situation regarding the governor's reelection meets the binomial conditions.

- There are only two possible outcomes. That is, a sampled voter will either vote or not vote for the governor.
- The probability of a success is the same for each trial. In this case the likelihood a particular sampled voter will support reelection is .80.
- The trials are independent. This means, for example, the likelihood the 23rd voter sampled will support reelection is not affected by what the 24th or 52nd voter does.
- The sample data is the result of counts. We are going to count the number of voters who support reelection in the sample of 2,000.

We can use the normal approximation to the binomial distribution, discussed in Chapter 7, because both $n\pi$ and $n(1 - \pi)$ exceed 5. In this case, $n = 2,000$ and $\pi = .80$ (π is the proportion of the vote in the northern part of the state, or 80 percent, needed to be elected). Thus, $n\pi = 2,000(.80) = 1,600$ and $n(1 - \pi) = 2,000(1 - .80) = 400$. Both 1,600 and 400 are greater than 5.

Step 1: State the null hypothesis and the alternate hypothesis. The null hypothesis, H_0 , is that the population proportion π is .80 or larger. The alternate hypothesis, H_1 , is that the proportion is less than .80. From a practical standpoint, the incumbent governor is concerned only when the proportion is less than .80. If it is equal to or greater than .80, he will have no problem; that is, the sample data would indicate he will probably be reelected. These hypotheses are written symbolically as:

$$H_0: \pi \geq .80$$

$$H_1: \pi < .80$$

H_1 states a direction. Thus, as noted previously, the test is one-tailed with the inequality sign pointing to the tail of the distribution containing the region of rejection.

Step 2: Select the level of significance. The level of significance is .05. This is the likelihood that a true hypothesis will be rejected.

Step 3: Select the test statistic. z is the appropriate statistic, found by:

$$\text{TEST OF HYPOTHESIS, ONE PROPORTION} \quad z = \frac{p - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}} \quad [10-3]$$

where:

- π is the population proportion.
- p is the sample proportion.
- n is the sample size.

Step 4: Formulate the decision rule. The critical value or values of z form the dividing point or points between the regions where H_0 is rejected and where it is not rejected. Since the alternate hypothesis states a direction, this is a one-tailed test. The sign of the inequality points to the left, so only the left side of the curve is used. (See Chart 10–8.) The significance level was given as .05 in step 2. This probability is in the left tail and determines the region of rejection. The area between zero and the critical value is .4500, found by $.5000 - .0500$. Referring to Appendix B.1 and searching for .4500, we find the critical value of z is 1.65. The decision rule is, therefore: Reject the null hypothesis and accept the alternate hypothesis if the computed value of z falls to the left of -1.65 ; otherwise do not reject H_0 .

Step 5: Make a decision and interpret the result. Select a sample and make a decision about H_0 . A sample survey of 2,000 potential voters in the northern part of the state revealed that 1,550 planned to vote for the incumbent governor. Is the sample proportion of .775 (found by $1,550/2,000$)

Finding the critical
value

Select a sample and
make a decision
regarding H_0 .

One-Sample Tests of Hypothesis

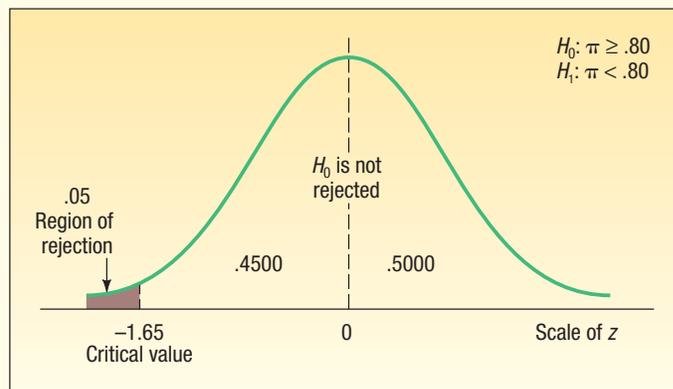


CHART 10–8 Rejection Region for the .05 Level of Significance, One-Tailed Test

close enough to .80 to conclude that the difference is due to sampling error? In this case:

p is .775, the proportion in the sample who plan to vote for the governor.

n is 2,000, the number of voters surveyed.

π is .80, the hypothesized population proportion.

z is a normally distributed test statistic when the hypothesis is true and the other assumptions are true.

Using formula (10–3) and computing z gives

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}} = \frac{\frac{1,550}{2,000} - .80}{\sqrt{\frac{.80(1 - .80)}{2,000}}} = \frac{.775 - .80}{\sqrt{.00008}} = -2.80$$

The computed value of z (-2.80) is in the rejection region, so the null hypothesis is rejected at the .05 level. The difference of 2.5 percentage points between the sample percent (77.5 percent) and the hypothesized population percent in the northern part of the state necessary to carry the state (80 percent) is statistically significant. It is probably not due to sampling variation. To put it another way, the evidence at this point does not support the claim that the incumbent governor will return to the governor’s mansion for another four years.

The p -value is the probability of finding a z value less than -2.80 . From Appendix B.1, the probability of a z value between zero and -2.80 is .4974. So the p -value is .0026, found by $.5000 - .4974$. The governor cannot be confident of reelection because the p -value is less than the significance level.

Self-Review 10–5



A recent insurance industry report indicated that 40 percent of those persons involved in minor traffic accidents this year have been involved in a least one other traffic accident in the last five years. An advisory group decided to investigate this claim, believing it was too large. A sample of 200 traffic accidents this year showed 74 persons were also involved in another accident within the last five years. Use the .01 significance level.

- Can we use z as the test statistic? Tell why or why not.
- State the null hypothesis and the alternate hypothesis.
- Show the decision rule graphically.
- Compute the value of z and state your decision regarding the null hypothesis.
- Determine and interpret the p -value.

Exercises

21. The following hypotheses are given.

$$H_0: \pi \leq .70$$

$$H_1: \pi > .70$$

A sample of 100 observations revealed that $p = .75$. At the .05 significance level, can the null hypothesis be rejected?

- State the decision rule.
 - Compute the value of the test statistic.
 - What is your decision regarding the null hypothesis?
22. The following hypotheses are given.

$$H_0: \pi = .40$$

$$H_1: \pi \neq .40$$

A sample of 120 observations revealed that $p = .30$. At the .05 significance level, can the null hypothesis be rejected?

- State the decision rule.
- Compute the value of the test statistic.
- What is your decision regarding the null hypothesis?

Note: It is recommended that you use the five-step hypothesis-testing procedure in solving the following problems.

23. The National Safety Council reported that 52 percent of American turnpike drivers are men. A sample of 300 cars traveling southbound on the New Jersey Turnpike yesterday revealed that 170 were driven by men. At the .01 significance level, can we conclude that a larger proportion of men were driving on the New Jersey Turnpike than the national statistics indicate?
24. A recent article in *USA Today* reported that a job awaits only one in three new college graduates. The major reasons given were an overabundance of college graduates and a weak economy. A survey of 200 recent graduates from your school revealed that 80 students had jobs. At the .02 significance level, can we conclude that a larger proportion of students at your school have jobs?
25. Chicken Delight claims that 90 percent of its orders are delivered within 10 minutes of the time the order is placed. A sample of 100 orders revealed that 82 were delivered within the promised time. At the .10 significance level, can we conclude that less than 90 percent of the orders are delivered in less than 10 minutes?
26. Research at the University of Toledo indicates that 50 percent of students change their major area of study after their first year in a program. A random sample of 100 students in the College of Business revealed that 48 had changed their major area of study after their first year of the program. Has there been a significant decrease in the proportion of students who change their major after the first year in this program? Test at the .05 level of significance.

Type II Error

Recall that the level of significance, identified by the symbol α , is the probability that the null hypothesis is rejected when it is true. This is called a Type I error. The most common levels of significance are .05 and .01 and are set by the researcher at the outset of the test.

In a hypothesis-testing situation there is also the possibility that a null hypothesis is not rejected when it is actually false. That is, we accept a false null hypothesis. This is called a Type II error. The probability of a Type II error is identified by the Greek letter beta (β). The following examples illustrate the details of determining the value of β .

Example

A manufacturer purchases steel bars to make cotter pins. Past experience indicates that the mean tensile strength of all incoming shipments is 10,000 psi and that the standard deviation, σ , is 400 psi.

One-Sample Tests of Hypothesis

In order to make a decision about incoming shipments of steel bars, the manufacturer set up this rule for the quality-control inspector to follow: “Take a sample of 100 steel bars. At the .05 significance level if the sample mean (\bar{X}) strength falls between 9,922 psi and 10,078 psi, accept the lot. Otherwise the lot is to be rejected.” Refer to Chart 10–9, Region A. It shows the region where each lot is rejected and where it is not rejected. The mean of this distribution is designated μ_0 . The tails of the curve represent the probability of making a Type I error, that is, rejecting the incoming lot of steel bars when in fact it is a good lot, with a mean of 10,000 psi.

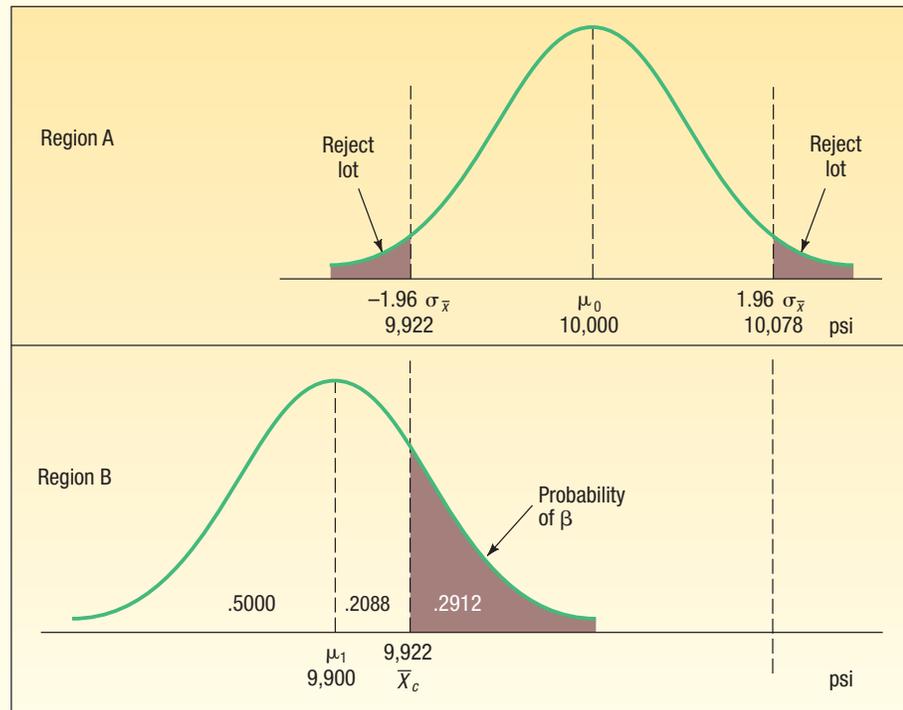


CHART 10–9 Charts Showing Type I and Type II Errors

Suppose the unknown population mean of an incoming lot, designated μ_1 , is really 9,900 psi. What is the probability that the quality-control inspector will fail to reject the shipment (a Type II error)?

Solution

The probability of committing a Type II error, as represented by the shaded area in Chart 10–9, Region B, can be computed by determining the area under the normal curve that lies above 9,922 pounds. The calculation of the areas under the normal curve was discussed in Chapter 7. Reviewing briefly, it is necessary first to determine the probability of the sample mean falling between 9,900 and 9,922. Then this probability is subtracted from .5000 (which represents all the area beyond the mean of 9,900) to arrive at the probability of making a Type II error in this case.

The number of standard units (z value) between the mean of the incoming lot (9,900), designated by μ_1 , and \bar{X}_c , representing the critical value for 9,922, is computed by:

TYPE II ERROR

$$z = \frac{\bar{X}_c - \mu_1}{\sigma/\sqrt{n}}$$

[10–4]

With $n = 100$ and $\sigma = 400$, the value of z is 0.55:

$$z = \frac{\bar{X}_c - \mu_1}{\sigma/\sqrt{n}} = \frac{9,922 - 9,900}{400/\sqrt{100}} = \frac{22}{40} = 0.55$$

The area under the curve between 9,900 and 9,922 (a z value of 0.55) is .2088. The area under the curve beyond 9,922 pounds is $.5000 - .2088$, or .2912; this is the probability of making a Type II error—that is, accepting an incoming lot of steel bars when the population mean is 9,900 psi.

Another illustration, in Chart 10–10, Region C, depicts the probability of accepting a lot when the population mean is 10,120. To find the probability:

$$z = \frac{\bar{X}_c - \mu_1}{\sigma/\sqrt{n}} = \frac{10,078 - 10,120}{400/\sqrt{100}} = -1.05$$

The probability that z is less than -1.05 is .1469, found by $.5000 - .3531$. Therefore, β , or the probability of a Type II error, is .1469.

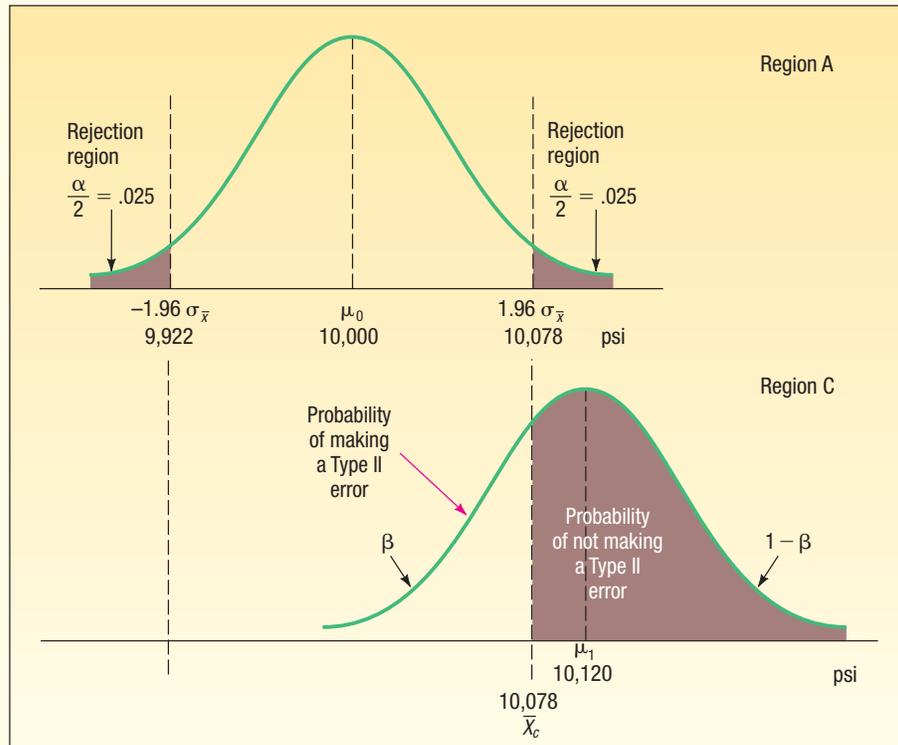


CHART 10–10 Type I and Type II Errors (Another Example)

Using the methods illustrated by Charts 10–9 Region B and 10–10 Region C, the probability of accepting a hypothesis as true when it is actually false can be determined for any value of μ_1 .

Type II error probabilities are shown in the center column of Table 10–4 for selected values of μ_1 , given in the left column. The right column gives the probability of not making a Type II error, which is also known as the power of a test.

One-Sample Tests of Hypothesis

TABLE 10–4 Probabilities of a Type II Error for $\mu_0 = 10,000$ Pounds and Selected Alternative Means, .05 Level of Significance

Selected Alternative Mean (pounds)	Probability of Type II Error (β)	Probability of Not Making a Type II Error ($1 - \beta$)
9,820	.0054	.9946
9,880	.1469	.8531
9,900	.2912	.7088
9,940	.6736	.3264
9,980	.9265	.0735
10,000	— *	—
10,020	.9265	.0735
10,060	.6736	.3264
10,100	.2912	.7088
10,120	.1469	.8531
10,180	.0054	.9946

*It is not possible to make a Type II error when $\mu = \mu_0$.

Self-Review 10–6



Refer to the previous Example. Suppose the true mean of an incoming lot of steel bars is 10,180 psi. What is the probability that the quality control inspector will accept the bars as having a mean of 10,000 psi? (It sounds implausible that steel bars will be rejected if the tensile strength is higher than specified. However, it may be that the cotter pin has a dual function in an outboard motor. It may be designed not to shear off if the motor hits a small object, but to shear off if it hits a rock. Therefore, the steel should not be *too* strong.)

The light area in Chart 10–10, Region C, represents the probability of falsely accepting the hypothesis that the mean tensile strength of the incoming steel is 10,000 psi. What is the probability of committing a Type II error?

Exercises

- 27. Refer to Table 10–4 and the example just completed. With $n = 100$, $\sigma = 400$, $\bar{X}_c = 9,922$, and $\mu_1 = 9,880$, verify that the probability of a Type II error is .1469.
- 28. Refer to Table 10–4 and the example just completed. With $n = 100$, $\sigma = 400$, $\bar{X}_c = 9,922$, and $\mu_1 = 9,940$, verify that the probability of a Type II error is .6736.

Chapter Summary

- I. The objective of hypothesis testing is to check the validity of a statement about a population parameter.
- II. The steps in conducting a test of hypothesis are:
 - A. State the null hypothesis (H_0) and the alternate hypothesis (H_1).
 - B. Select the level of significance.
 - 1. The level of significance is the likelihood of rejecting a true null hypothesis.
 - 2. The most frequently used significance levels are .01, .05, and .10, but any value between 0 and 1.00 is possible.
 - C. Select the test statistic.
 - 1. A test statistic is a value calculated from sample information used to determine whether to reject the null hypothesis.
 - 2. Two test statistics were considered in this chapter.
 - a. The standard normal distribution is used when the population follows the normal distribution and the population standard deviation is known.

- b. The t distribution is used when the population follows the normal distribution and the population standard deviation is unknown.
- D. State the decision rule.
1. The decision rule indicates the condition or conditions when the null hypothesis is rejected.
 2. In a two-tailed test, the rejection region is evenly split between the upper and lower tails.
 3. In a one-tailed test, all of the rejection region is in either the upper or the lower tail.
- E. Select a sample, compute the value of the test statistic, make a decision regarding the null hypothesis, and interpret the results.
- III. A p -value is the probability that the value of the test statistic is as extreme as the value computed, when the null hypothesis is true.
- IV. When testing a hypothesis about a population mean:
- A. If the population standard deviation, σ , is known, the test statistic is the standard normal distribution and is determined from:

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \quad [10-1]$$

- B. If the population standard deviation is not known, s is substituted for σ . The test statistic is the t distribution, and its value is determined from:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \quad [10-2]$$

The major characteristics of the t distribution are:

1. It is a continuous distribution.
 2. It is mound-shaped and symmetrical.
 3. It is flatter, or more spread out, than the standard normal distribution.
 4. There is a family of t distributions, depending on the number of degrees of freedom.
- V. When testing about a population proportion:
- A. The binomial conditions must be met.
- B. Both $n\pi$ and $n(1 - \pi)$ must be at least 5.
- C. The test statistic is

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}} \quad [10-3]$$

- VI. There are two types of errors that can occur in a test of hypothesis.
- A. A Type I error occurs when a true null hypothesis is rejected.
1. The probability of making a Type I error is equal to the level of significance.
 2. This probability is designated by the Greek letter α .
- B. A Type II error occurs when a false null hypothesis is not rejected.
1. The probability of making a Type II error is designated by the Greek letter β .
 2. The likelihood of a Type II error is found by

$$z = \frac{\bar{X}_c - \mu_1}{\sigma/\sqrt{n}} \quad [10-4]$$

Pronunciation Key

SYMBOL	MEANING	PRONUNCIATION
H_0	Null hypothesis	<i>H sub zero</i>
H_1	Alternate hypothesis	<i>H sub one</i>
$\alpha/2$	Two-tailed significance level	<i>Alpha over 2</i>
\bar{X}_c	Limit of the sample mean	<i>X bar sub c</i>
μ_0	Assumed population mean	<i>mu sub zero</i>

One-Sample Tests of Hypothesis

361

Chapter Exercises

29. According to the local union president, the mean gross income of plumbers in the Salt Lake City area follows the normal probability distribution with a mean of \$45,000 and a standard deviation of \$3,000. A recent investigative reporter for KYAK TV found, for a sample of 120 plumbers, the mean gross income was \$45,500. At the .10 significance level, is it reasonable to conclude that the mean income is not equal to \$45,000? Determine the p -value.
30. Rutter Nursery Company packages its pine bark mulch in 50-pound bags. From a long history, the production department reports that the distribution of the bag weights follows the normal distribution and the standard deviation of this process is 3 pounds per bag. At the end of each day, Jeff Rutter, the production manager, weighs 10 bags and computes the mean weight of the sample. Below are the weights of 10 bags from today's production.

45.6	47.7	47.6	46.3	46.2	47.4	49.2	55.8	47.5	48.5
------	------	------	------	------	------	------	------	------	------

- a. Can Mr. Rutter conclude that the mean weight of the bags is less than 50 pounds? Use the .01 significance level.
- b. In a brief report, tell why Mr. Rutter can use the z distribution as the test statistic.
- c. Compute the p -value.
31. A new weight-watching company, Weight Reducers International, advertises that those who join will lose, on the average, 10 pounds the first two weeks with a standard deviation of 2.8 pounds. A random sample of 50 people who joined the new weight reduction program revealed the mean loss to be 9 pounds. At the .05 level of significance, can we conclude that those joining Weight Reducers on average will lose less than 10 pounds? Determine the p -value.
32. Dole Pineapple, Inc., is concerned that the 16-ounce can of sliced pineapple is being overfilled. Assume the standard deviation of the process is .03 ounces. The quality-control department took a random sample of 50 cans and found that the arithmetic mean weight was 16.05 ounces. At the 5 percent level of significance, can we conclude that the mean weight is greater than 16 ounces? Determine the p -value.
33. According to a recent survey, Americans get a mean of 7 hours of sleep per night. A random sample of 50 students at West Virginia University revealed the mean number of hours slept last night was 6 hours and 48 minutes (6.8 hours). The standard deviation of the sample was 0.9 hours. Is it reasonable to conclude that students at West Virginia sleep less than the typical American? Compute the p -value.
34. A statewide real estate sales agency, Farm Associates, specializes in selling farm property in the state of Nebraska. Its records indicate that the mean selling time of farm property is 90 days. Because of recent drought conditions, the agency believes that the mean selling time is now greater than 90 days. A statewide survey of 100 farms sold recently revealed that the mean selling time was 94 days, with a standard deviation of 22 days. At the .10 significance level, has there been an increase in selling time?
35. NBC TV news, in a segment on the price of gasoline, reported last evening that the mean price nationwide is \$2.50 per gallon for self-serve regular unleaded. A random sample of 35 stations in the Milwaukee, Wisconsin, area revealed that the mean price was \$2.52 per gallon and that the standard deviation was \$0.05 per gallon. At the .05 significance level, can we conclude that the price of gasoline is higher in the Milwaukee area? Determine the p -value.
36. A recent article in *Vitality* magazine reported that the mean amount of leisure time per week for American men is 40.0 hours. You believe this figure is too large and decide to conduct your own test. In a random sample of 60 men, you find that the mean is 37.8 hours of leisure per week and that the standard deviation of the sample is 12.2 hours. Can you conclude that the information in the article is untrue? Use the .05 significance level. Determine the p -value and explain its meaning.
37. In recent years the interest rate on home mortgages has declined to less than 6.0 percent. However, according to a study by the Federal Reserve Board the rate charged on credit card debt is more than 14 percent. Listed below is the interest rate charged on a sample of 10 credit cards.

14.6	16.7	17.4	17.0	17.8	15.4	13.1	15.8	14.3	14.5
------	------	------	------	------	------	------	------	------	------

Is it reasonable to conclude the mean rate charged is greater than 14 percent? Use the .01 significance level.

38. A recent article in *The Wall Street Journal* reported that the 30-year mortgage rate is now less than 6 percent. A sample of eight small banks in the Midwest revealed the following 30-year rates (in percent):

4.8	5.3	6.5	4.8	6.1	5.8	6.2	5.6
-----	-----	-----	-----	-----	-----	-----	-----

At the .01 significance level, can we conclude that the 30-year mortgage rate for small banks is less than 6 percent? Estimate the p -value.

39. According to the Coffee Research Organization (<http://www.coffeeresearch.org>) the typical American coffee drinker consumes an average of 3.1 cups per day. A sample of 12 senior citizens revealed they consumed the following amounts of coffee, reported in cups, yesterday.

3.1	3.3	3.5	2.6	2.6	4.3	4.4	3.8	3.1	4.1	3.1	3.2
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

At the .05 significance level does this sample data suggest there is a difference between the national average and the sample mean from senior citizens?

40. The postanesthesia care area (recovery room) at St. Luke's Hospital in Maumee, Ohio, was recently enlarged. The hope was that with the enlargement the mean number of patients per day would be more than 25. A random sample of 15 days revealed the following numbers of patients.

25	27	25	26	25	28	28	27	24	26	25	29	25	27	24
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

At the .01 significance level, can we conclude that the mean number of patients per day is more than 25? Estimate the p -value and interpret it.

41. eGolf.com receives an average of 6.5 returns per day from online shoppers. For a sample of 12 days, it received the following number of returns.

0	4	3	4	9	4	5	9	1	6	7	10
---	---	---	---	---	---	---	---	---	---	---	----

At the .01 significance level, can we conclude the mean number of returns is less than 6.5?

42. During recent seasons, Major League Baseball has been criticized for the length of the games. A report indicated that the average game lasts 3 hours and 30 minutes. A sample of 17 games revealed the following times to completion. (Note that the minutes have been changed to fractions of hours, so that a game that lasted 2 hours and 24 minutes is reported at 2.40 hours.)

2.98	2.40	2.70	2.25	3.23	3.17	2.93	3.18	2.80
2.38	3.75	3.20	3.27	2.52	2.58	4.45	2.45	

Can we conclude that the mean time for a game is less than 3.50 hours? Use the .05 significance level.

43. Watch Corporation of Switzerland claims that its watches on average will neither gain nor lose time during a week. A sample of 18 watches provided the following gains (+) or losses (–) in seconds per week.

–0.38	–0.20	–0.38	–0.32	+0.32	–0.23	+0.30	+0.25	–0.10
–0.37	–0.61	–0.48	–0.47	–0.64	–0.04	–0.20	–0.68	+0.05

Is it reasonable to conclude that the mean gain or loss in time for the watches is 0? Use the .05 significance level. Estimate the p -value.

44. Listed below is the rate of return for one year (reported in percent) for a sample of 12 mutual funds that are classified as taxable money market funds.

4.63	4.15	4.76	4.70	4.65	4.52	4.70	5.06	4.42	4.51	4.24	4.52
------	------	------	------	------	------	------	------	------	------	------	------

One-Sample Tests of Hypothesis

363

Using the .05 significance level is it reasonable to conclude that the mean rate of return is more than 4.50 percent?

45. Many grocery stores and large retailers such as Wal-Mart and Kmart have installed self-checkout systems so shoppers can scan their own items and cash out themselves. How do customers like this service and how often do they use it? Listed below is the number of customers using the service for a sample of 15 days at the Wal-Mart on Highway 544 in Surfside, South Carolina.

120	108	120	114	118	91	118	92	104	104
112	97	118	108	117					

Is it reasonable to conclude that the mean number of customers using the self-checkout system is more than 100 per day? Use the .05 significance level.

46. In the year 2006 the mean fare to fly from Charlotte, North Carolina, to Seattle, Washington, on a discount ticket was \$267. A random sample of round-trip discount fares on this route last month gives:

\$321	\$286	\$290	\$330	\$310	\$250	\$270	\$280	\$299	\$265	\$291	\$275	\$281
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

At the .01 significance level can we conclude that the mean fare has increased? What is the p -value?

47. The publisher of *Celebrity Living* claims that the mean sales for personality magazines that feature people such as Angelina Jolie or Paris Hilton are 1.5 million copies per week. A sample of 10 comparable titles shows a mean weekly sales last week of 1.3 million copies with a standard deviation of 0.9 million copies. Does this data contradict the publisher's claim? Use the 0.01 significance level.
48. A United Nations report shows the mean family income for Mexican migrants to the United States is \$27,000 per year. A FLOC (Farm Labor Organizing Committee) evaluation of 25 Mexican family units reveals a mean to be \$30,000 with a sample standard deviation of \$10,000. Does this information disagree with the United Nations report? Apply the 0.01 significance level.
49. Traditionally, two percent of the citizens of the United States live in a foreign country because they are disenchanted with U.S. politics or social attitudes. In order to test if this proportion has increased since the September 11, 2001, terror attacks, U.S. consulates contacted a random sample of 400 of these expatriates. The sample yields 12 people who report they are living overseas because of political or social attitudes. Can you conclude this data shows the proportion of politically motivated expatriates has increased? Use the 0.05 significance level.
50. According to a study by the American Pet Food Dealers Association, 63 percent of U.S. households own pets. A report is being prepared for an editorial in the *San Francisco Chronicle*. As a part of the editorial a random sample of 300 households showed 210 own pets. Does this data disagree with the Pet Food Dealers Association data? Use a .05 level of significance.
51. Tina Dennis is the comptroller for Meek Industries. She believes that the current cash-flow problem at Meek is due to the slow collection of accounts receivable. She believes that more than 60 percent of the accounts are in arrears more than three months. A random sample of 200 accounts showed that 140 were more than three months old. At the .01 significance level, can she conclude that more than 60 percent of the accounts are in arrears for more than three months?
52. The policy of the Suburban Transit Authority is to add a bus route if more than 55 percent of the potential commuters indicate they would use the particular route. A sample of 70 commuters revealed that 42 would use a proposed route from Bowman Park to the downtown area. Does the Bowman-to-downtown route meet the STA criterion? Use the .05 significance level.
53. Past experience at the Crowder Travel Agency indicated that 44 percent of those persons who wanted the agency to plan a vacation for them wanted to go to Europe. During the most recent busy season, a sampling of 1,000 plans was selected at random from the files. It was found that 480 persons wanted to go to Europe on vacation. Has there been a significant shift upward in the percentage of persons who want to go to Europe? Test at the .05 significance level.

54. From past experience a television manufacturer found that 10 percent or less of its sets needed any type of repair in the first two years of operation. In a sample of 50 sets manufactured two years ago, 9 needed repair. At the .05 significance level, has the percent of sets needing repair increased? Determine the p -value.
55. An urban planner claims that, nationally, 20 percent of all families renting condominiums move during a given year. A random sample of 200 families renting condominiums in the Dallas Metroplex revealed that 56 had moved during the past year. At the .01 significance level, does this evidence suggest that a larger proportion of condominium owners moved in the Dallas area? Determine the p -value.
56. The cost of weddings in the United States has skyrocketed in recent years. As a result many couples are opting to have their weddings in the Caribbean. A Caribbean vacation resort recently advertised in *Bride Magazine* that the cost of a Caribbean wedding was less than \$10,000. Listed below is a total cost in \$000 for a sample of 8 Caribbean weddings.

9.7	9.4	11.7	9.0	9.1	10.5	9.1	9.8
-----	-----	------	-----	-----	------	-----	-----

- At the .05 significance level is it reasonable to conclude the mean wedding cost is less than \$10,000 as advertised?
57. The U.S. president's call for designing and building a missile defense system that ignores restrictions of the Anti-Ballistic Missile Defense System treaty (ABM) is supported by 483 of the respondents in a nationwide poll of 1,002 adults. Is it reasonable to conclude that the nation is evenly divided on the issue? Use the .05 significance level.
58. One of the major U.S. automakers wishes to review its warranty. The warranty covers the engine, transmission, and drivetrain of all new cars for up to two years or 24,000 miles, whichever comes first. The manufacturer's quality-assurance department believes that the mean number of miles driven by owners is more than 24,000. A sample of 35 cars revealed that the mean number of miles was 24,421, with a standard deviation of 1,944 miles.
- Conduct the following hypothesis test. Use the .05 significance level.

$$H_0: \mu \leq 24,000$$

$$H_1: \mu > 24,000$$
 - What is the largest value for the sample mean for which H_0 is not rejected?
 - Suppose the population mean shifts to 25,000 miles. What is the probability this change will not be detected?
59. A cola-dispensing machine is set to dispense 9.00 ounces of cola per cup, with a standard deviation of 1.00 ounces. The manufacturer of the machine would like to set the control limit in such a way that, for samples of 36, 5 percent of the sample means will be greater than the upper control limit, and 5 percent of the sample means will be less than the lower control limit.
- At what value should the control limit be set?
 - What is the probability that if the population mean shifts to 8.9, this change will not be detected?
 - What is the probability that if the population mean shifts to 9.3, this change will not be detected?
60. The owners of the Franklin Park Mall wished to study customer shopping habits. From earlier studies the owners were under the impression that a typical shopper spends 0.75 hours at the mall, with a standard deviation of 0.10 hours. Recently the mall owners added some specialty restaurants designed to keep shoppers in the mall longer. The consulting firm, Brunner and Swanson Marketing Enterprises, was hired to evaluate the effects of the restaurants. A sample of 45 shoppers by Brunner and Swanson revealed that the mean time spent in the mall had increased to 0.80 hours.
- Develop a test of hypothesis to determine if the mean time spent in the mall is more than 0.75 hours. Use the .05 significance level.
 - Suppose the mean shopping time actually increased from 0.75 hours to 0.77 hours. What is the probability this increase would not be detected?
 - When Brunner and Swanson reported the information in part (b) to the mall owners, the owners were upset with the statement that a survey could not detect a change from 0.75 to 0.77 hours of shopping time. How could this probability be reduced?

One-Sample Tests of Hypothesis

365

61. The following null and alternate hypotheses are given.

$$H_0: \mu \leq 50$$

$$H_1: \mu > 50$$

Suppose the population standard deviation is 10. The probability of a Type I error is set at .01 and the probability of a Type II error at .30. Assume that the population mean shifts from 50 to 55. How large a sample is necessary to meet these requirements?

62. An insurance company, based on past experience, estimates the mean damage for a natural disaster in its area is \$5,000. After introducing several plans to prevent loss, it randomly samples 200 policyholders and finds the mean amount per claim was \$4,800 with a standard deviation of \$1,300. Does it appear the prevention plans were effective in reducing the mean amount of a claim? Use the .05 significance level.
63. A national grocer's magazine reports the typical shopper spends eight minutes in line waiting to check out. A sample of 24 shoppers at the local Farmer Jack's showed a mean of 7.5 minutes with a standard deviation of 3.2 minutes. Is the waiting time at the local Farmer Jack's less than that reported in the national magazine? Use the .05 significance level.

exercises.com



64. *USA Today* (<http://www.usatoday.com/sports/baseball/salaries/default.aspx>) reports information on individual player salaries. Go to this site and find the individual salaries for your favorite team. Compute the mean and the standard deviation. Is it reasonable to conclude that the mean salary on your favorite team is *different from* \$3.20 million? If you are more of a football, basketball, or hockey enthusiast, information is also available on their teams' salaries.
65. The Gallup Organization in Princeton, New Jersey, is one of the best-known polling organizations in the United States. It often combines with *USA Today* or CNN to conduct polls of current interest. It also maintains a website at: <http://www.gallup.com/>. Consult this website to find the most recent polling results on presidential approval ratings. You may need to click on **Gallup Poll**. Test whether the majority (more than 50 percent) approve of the president's performance. If the article does not report the number of respondents included in the survey, assume that it is 1,000, a number that is typically used.

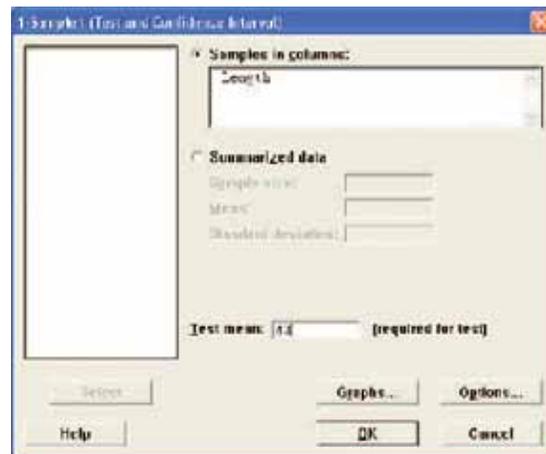
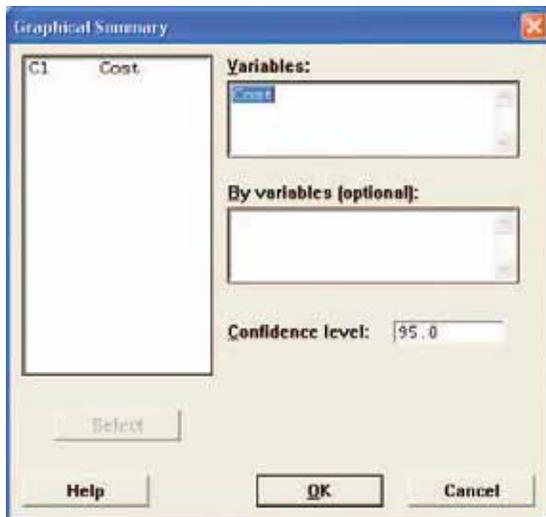
Data Set Exercises

66. Refer to the Real Estate data, which report information on the homes sold in Denver, Colorado, last year.
- A recent article in the *Denver Post* indicated that the mean selling price of the homes in the area is more than \$220,000. Can we conclude that the mean selling price in the Denver area is more than \$220,000? Use the .01 significance level. What is the p -value?
 - The same article reported the mean size was more than 2,100 square feet. Can we conclude that the mean size of homes sold in the Denver area is more than 2,100 square feet? Use the .01 significance level. What is the p -value?
 - Determine the proportion of homes that have an attached garage. At the .05 significance level can we conclude that more than 60 percent of the homes sold in the Denver area had an attached garage? What is the p -value?
 - Determine the proportion of homes that have a pool. At the .05 significance level, can we conclude that less than 40 percent of the homes sold in the Denver area had a pool? What is the p -value?
67. Refer to the Baseball 2005 data, which report information on the 30 Major League Baseball teams for the 2005 season.
- Conduct a test of hypothesis to determine whether the mean salary of the teams was different from \$80.0 million. Use the .05 significance level.
 - Conduct a test of hypothesis to determine whether the mean attendance was more than 2,000,000 per team.
68. Refer to the Wage data, which report information on the annual wages for a sample of 100 workers. Also included are variables relating to the industry, years of education, and gender for each worker.
- Conduct a test of hypothesis to determine if the mean annual wage is greater than \$30,000. Use the .05 significance level. Determine the p -value and interpret the result.

- b. Conduct a test of hypothesis to determine if the mean years of experience is different from 20. Use the .05 significance level. Determine the p -value and interpret the result.
 - c. Conduct a test of hypothesis to determine if the mean age is less than 40. Use the .05 significance level. Determine the p -value and interpret the result.
 - d. Conduct a test of hypothesis to determine if the proportion of union workers is greater than 15 percent. Use the .05 significance level and report the p -value.
69. Refer to the CIA data, which report demographic and economic information on 46 different countries.
- a. Conduct a test of hypothesis to determine if the mean number of cell phones is greater than 4.0. Use the .05 significance level. What is the p -value?
 - b. Conduct a test of hypothesis to determine if the mean size of the labor force is less than 50. Use the .05 significance level. What is the p -value?

Software Commands

- The MINITAB commands for the histogram and the descriptive statistics on page 346 are:
 - Enter the 26 sample observations in column C1 and name the variable *Cost*.
 - From the menu bar select **Stat, Basic Statistics, and Graphical Summary**. In the dialog box select **Cost** as the variable and click **OK**.
- The MINITAB commands for the one-sample t test on page 350 are:
 - Enter the sample data into column C1 and name the variable *Length*.
 - From the menu bar select **Stat, Basic Statistics, and 1-Sample t** and then hit **Enter**.
 - Select **Length** as the variable, select **Test mean**, insert the number 43 and click **OK**.



One-Sample Tests of Hypothesis

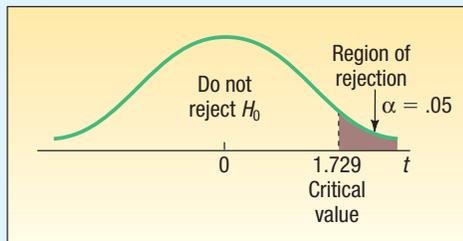
Chapter 10 Answers to Self-Review



- 10-1** a. $H_0: \mu = 16.0; H_1: \mu \neq 16.0$
 b. .05
 c. $z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$
 d. Reject H_0 if $z < -1.96$ or $z > 1.96$
 e. $z = \frac{16.017 - 16.0}{0.15/\sqrt{50}} = \frac{0.0170}{0.0212} = 0.80$
 f. Do not reject H_0
 g. We cannot conclude the mean amount dispensed is different from 16.0 ounces.

- 10-2** a. $H_0: \mu \leq 16.0; H_1: \mu > 16.0$
 b. Reject H_0 if $z > 1.65$
 c. $z = \frac{16.040 - 16.0}{0.15/\sqrt{50}} = \frac{.0400}{.0212} = 1.89$
 d. Reject H_0
 e. The mean amount dispensed is more than 16.0 ounces.
 f. $p\text{-value} = .5000 - .4706 = .0294$. The p -value is less than α (.05), so H_0 is rejected. It is the same conclusion as in part (d).

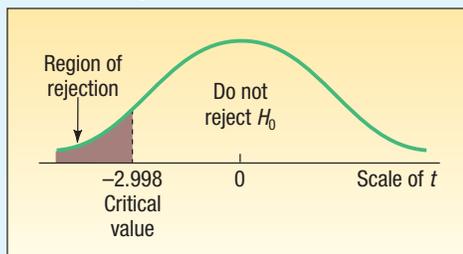
- 10-3** a. $H_0: \mu \leq 305; H_1: \mu > 305$.
 b. $df = n - 1 = 20 - 1 = 19$
 The decision rule is to reject H_0 if $t > 1.729$.



c. $t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{311 - 305}{12/\sqrt{20}} = 2.236$

Reject H_0 because $2.236 > 1.729$. The modification increased the mean battery life to more than 305 days.

- 10-4** a. $H_0: \mu \geq 9.0; H_1: \mu < 9.0$.
 b. 7, found by $n - 1 = 8 - 1 = 7$.
 c. Reject H_0 if $t < -2.998$.



- d. $t = -2.494$, found by:

$$s = \sqrt{\frac{0.36}{8-1}} = 0.2268$$

$$\bar{X} = \frac{70.4}{8} = 8.8$$

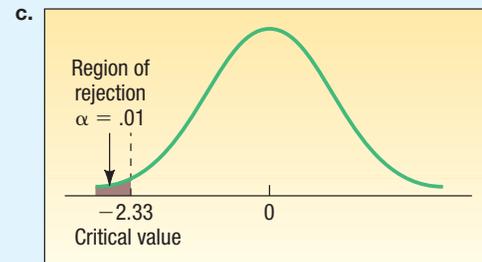
Then

$$t = \frac{8.8 - 9.0}{0.2268/\sqrt{8}} = -2.494$$

Since -2.494 lies to the right of -2.998 , H_0 is not rejected. We have not shown that the mean is less than 9.0.

- e. The p -value is between .025 and .010.

- 10-5** a. Yes, because both $n\pi$ and $n(1 - \pi)$ exceed 5:
 $n\pi = 200(.40) = 80$, and
 $n(1 - \pi) = 200(.60) = 120$.
 b. $H_0: \pi \geq .40$
 $H_1: \pi < .40$
 Reject H_0 if $z < -2.33$



- d. $z = -0.87$, found by:

$$z = \frac{.37 - .40}{\sqrt{\frac{.40(1 - .40)}{200}}} = \frac{-.03}{\sqrt{.0012}} = -0.87$$

Do not reject H_0 .

- e. The p -value is .1922, found by $.5000 - .3078$.
10-6 .0054, found by determining the area under the curve between 10,078 and 10,180 (Chart 10-10C).

$$z = \frac{\bar{X}_c - \mu_1}{\sigma/\sqrt{n}} = \frac{10,078 - 10,180}{400/\sqrt{100}} = -2.55$$

The area under the curve for a z of -2.55 is .4946 (Appendix B.1), and $.5000 - .4946 = .0054$.