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 Trail Blazer 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 mean starting salary for graduates of four-year business schools is $3,200 per month.
- 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 play the state lotteries regularly 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.

What Is a Hypothesis?

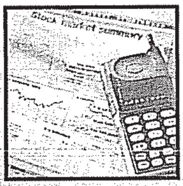
A hypothesis is a statement about a population parameter.

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 developed for the purpose of testing.

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.



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% of all surgeries involve complications such as glare, corneal haze, overcorrection or undercorrection 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% of all cases result in complications. The 5% 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% of future cases, this expectation will not be met. (Source: American Academy of Ophthalmology, San Francisco, Vol. 16, no. 43.)

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?

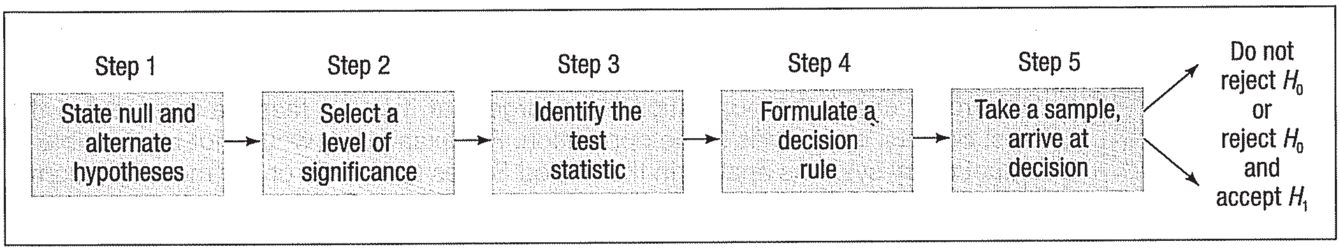
HYPOTHESIS TESTING A procedure based on sample evidence and probability theory to determine whether the hypothesis is a reasonable statement.

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 diagram at the bottom of this page. We will discuss in detail each of the steps.

Step 1: State the Null Hypothesis (H_0) and the Alternate Hypothesis (H_1)

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



Five-step systematic procedure.

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.

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 often called 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

Select a level of significance or risk.

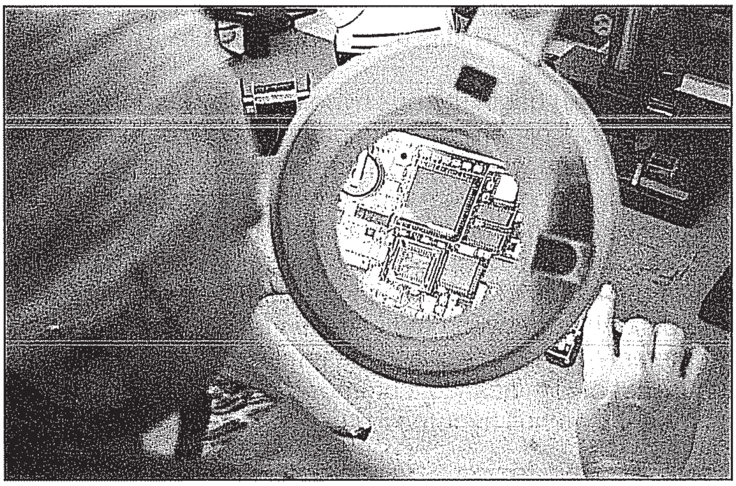
After establishing the null hypothesis and alternate hypothesis, the next step is to select the level of significance.

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

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 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.

	Researcher	
Null Hypotheses	**Accepts 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 or the sample size is large, the test statistic *z* is computed by:

z DISTRIBUTION AS A TEST STATISTIC $$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$ **[10–1]**

The *z* value is based on the sampling distribution of $\bar{X}$, which follows the normal distribution when the sample is reasonably large with a mean ($\mu_{\bar{X}}$) equal to μ, and a standard deviation $\sigma_{\bar{X}}$, which is equal to $\sigma/\sqrt{n}$. 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 of 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.

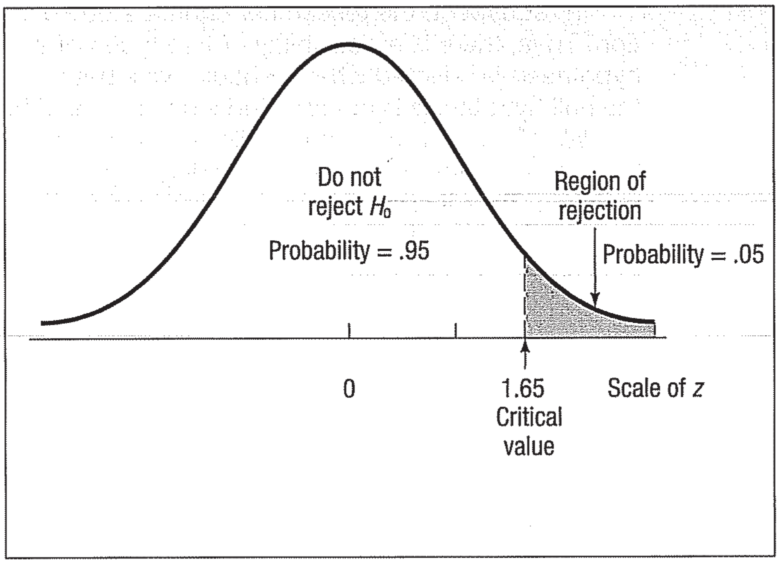


CHART 10–1 Sampling Distribution of the Statistic z, a Right-Tailed Test, .05 Level of Significance

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 is normally distributed.
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, the test statistic 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 test statistic 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."

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.

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

One-Tailed and Two-Tailed Tests of Significance

Refer to Chart 10–1 (previous page). 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 life of the tires is significantly below 60,000 miles on the average. They gladly

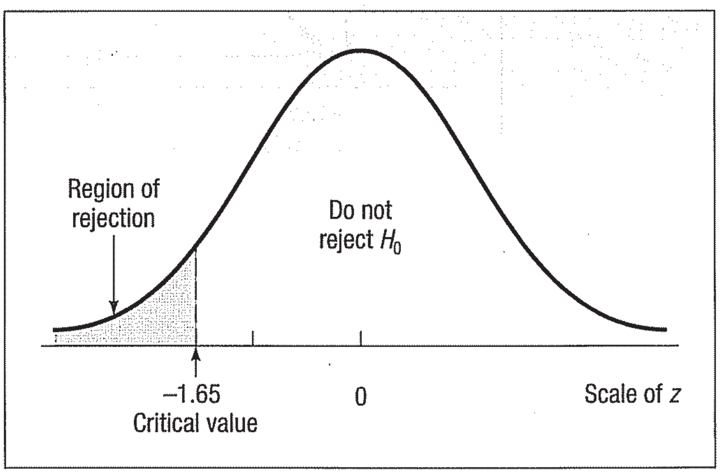


CHART 10–2 Sampling Distribution for the Statistic z, Left-Tailed Test, .05 Level of Significance

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

If H_1 states a direction, test is one-tailed.

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 less than 60,000 miles.* 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 financial planners is *less than or equal to* $65,000 per year.
H_1: The mean income of women financial planners 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 financial planners is $65,000 per year.
H_1: The mean income of women financial planners 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.

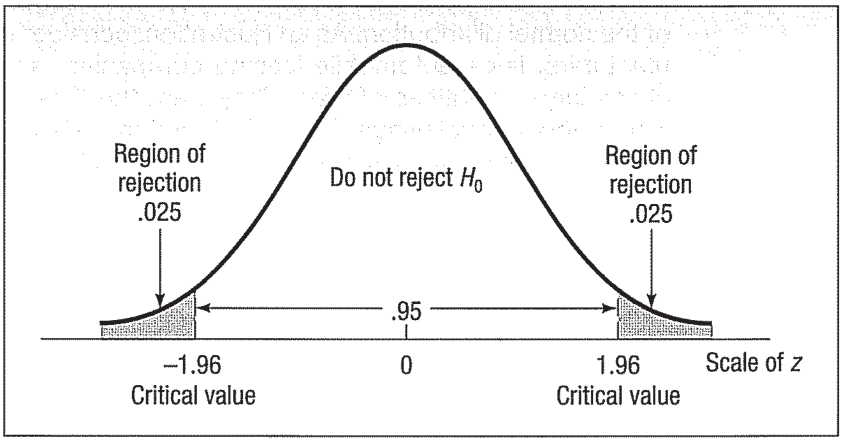


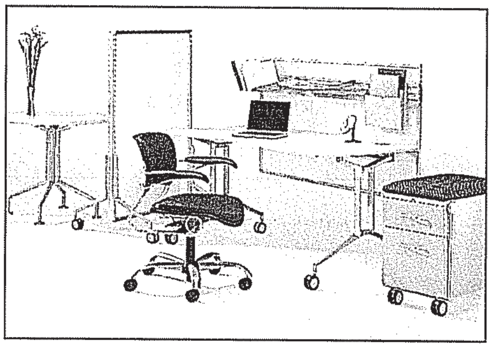
CHART 10–3 Regions of Nonrejection and Rejection for a Two-Tailed Test, .05 Level of Significance

Testing for a Population Mean with a 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.

EXAMPLE



The 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 the normal 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. To put it another way, 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:

$$H_0: \mu = 200$$
$$H_1: \mu \neq 200$$

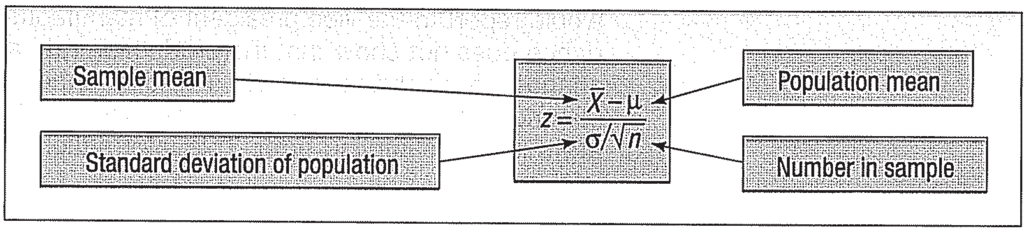
This is a *two-tailed* test because the alternate hypothesis does not state a direction. In other words, it does not state whether the mean production is greater than 200 or less than 200. The vice president wants only to find out whether the production rate is different from 200.

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. In this case, because we know that the population follows the normal distribution and we know σ, the population standard deviation, we use z as the test statistic. 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.

Formula for the test statistic

[10–1]



Step 4: Formulate the decision rule. The decision rule is formulated by finding the critical values of z from Appendix D. 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 D 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 D, 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.

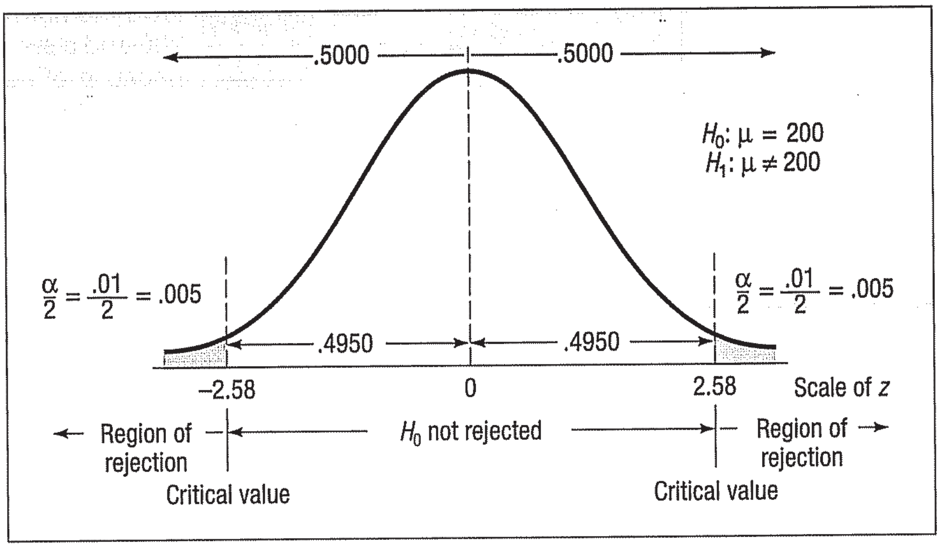


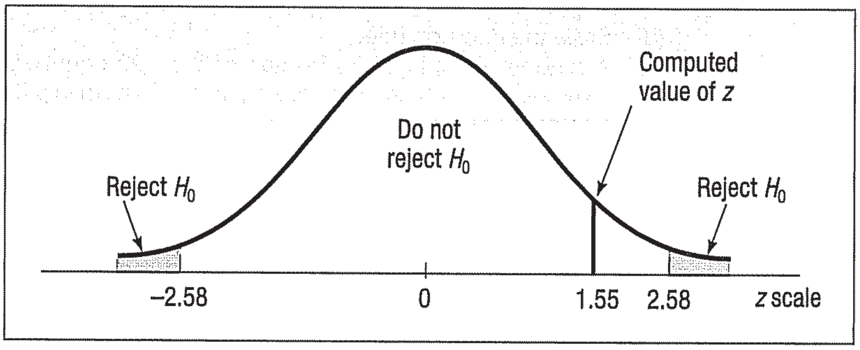
CHART 10–4 Decision Rule for the .01 Significance Level

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 last year's rate can reasonably be attributed to sampling error. This information is summarized in the following chart.



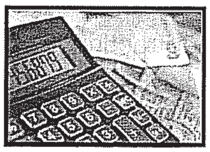
Did we prove that the assembly rate is still 200 per week? Not really. What we did, technically, was *fail to disprove the null hypothesis.* Failing to disprove the hypothesis that the population mean is 200 is not the same thing as proving it to be true. As we suggested in the chapter introduction, 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.

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 249. 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 it is for a test of hypothesis.

SELF-REVIEW 10–1



The annual turnover rate of the 200-count bottle of Bayer Aspirin follows the normal distribution with a mean of 6.0 and a standard deviation of 0.50. (This indicates that the stock of Bayer turns over on the pharmacy shelves an average of 6 times per year.) It is suspected that the mean turnover has changed and is not 6.0. Use the .05 significance level.

(a) State the null hypothesis and the alternate hypothesis.
(b) What is the probability of a Type I error?
(c) Give the formula for the test statistic.

(d) State the decision rule.
(e) A random sample of 64 bottles of the 200-count size Bayer Aspirin showed a mean turnover rate of 5.84. Shall we reject the hypothesis that the population mean is 6.0? Interpret the result.

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. To put it another way, 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:	A one-tailed test:
H_0: $\mu = 200$	H_0: $\mu \leq 200$
H_1: $\mu \neq 200$	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.

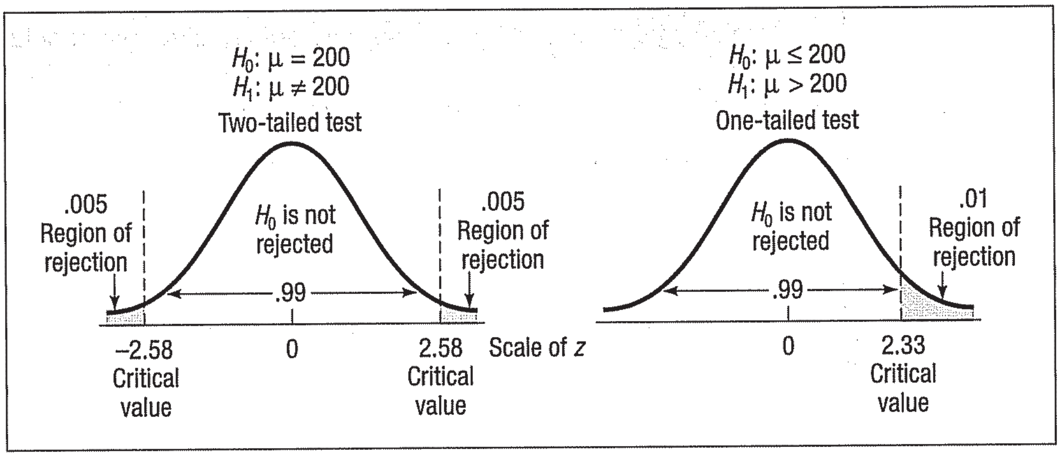
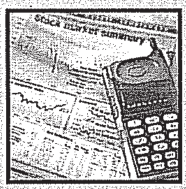


CHART 10-5 Rejection Regions for Two-Tailed and One-Tailed Tests, $\sigma = .01$

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.

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



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.

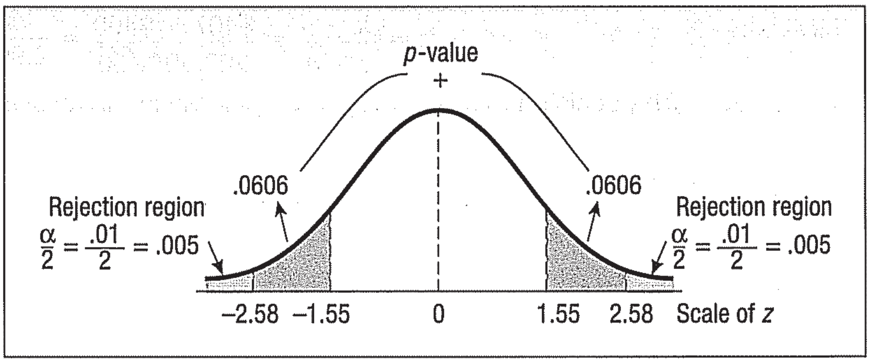
critical value is 1.96 and the computed value of the test statistic is 2.19, the decision is to reject the null hypothesis.

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 probability (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
> (a) .10, we have *some evidence* that H_0 is not true.
> (b) .05, we have *strong evidence* that H_0 is not true.
> (c) .01, we have *very strong evidence* that H_0 is not true.
> (d) .001, we have *extremely strong evidence* that H_0 is not true.

Testing for a Population Mean: Large Sample, Population Standard Deviation Unknown

In the preceding example, we knew that the population followed the normal distribution and σ, the population standard deviation. In most cases, however, we may not know for certain that the population follows the normal distribution or the population standard deviation. 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. As long as the sample size, n, is at least 30, s can be substituted for σ, as illustrated in the following formula:

z STATISTIC, σ UNKNOWN

$$z = \frac{\bar{X} - \mu}{s/\sqrt{n}} \qquad [10\text{–}2]$$

EXAMPLE

The Thompson's Discount Appliance Store issues its own credit card. The credit manager wants to find whether the mean monthly unpaid balance is more than \$400. The level of significance is set at .05. A random check of 60 unpaid balances revealed the sample mean is \$407 and the standard deviation of the sample is \$22.50. Should the credit manager conclude the population mean is greater than \$400, or is it reasonable that the difference of \$7 (\$407 − \$400 = \$7) is due to chance?

SOLUTION

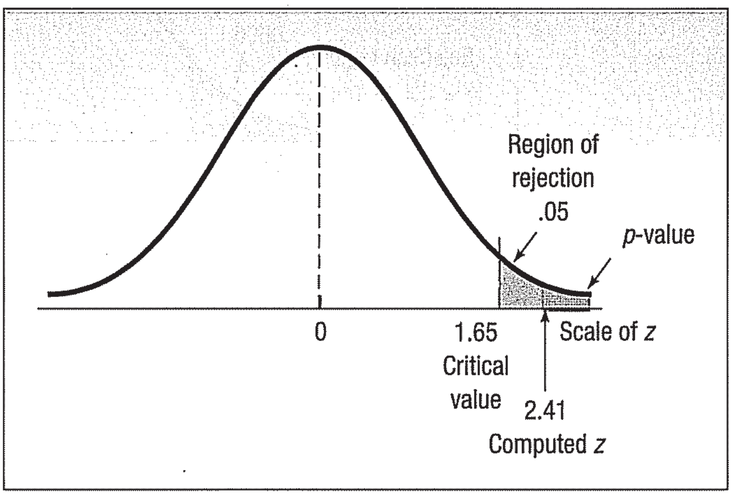
The null and alternate hypotheses are:

H_0: $\mu \leq \$400$
H_1: $\mu > \$400$

Because the alternate hypothesis states a direction, a one-tailed test is applied. The critical value of z is 1.65. The computed value of z is 2.41, found by using formula (10–2):

$$z = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{\$407 - \$400}{\$22.50/\sqrt{60}} = \frac{\$7}{2.9047} = 2.41$$

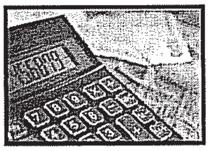
The decision rule is portrayed graphically in the following chart.



Because the computed value of the test statistic (2.41) is larger than the critical value (1.65), the null hypothesis is rejected. The credit manager can conclude the mean unpaid balance is greater than \$400.

The p-value provides additional insight into the decision. Recall the p-value is the probability of finding a test statistic as large as or larger than that obtained, when the null hypothesis is true. So we find the probability of a z value greater than 2.41. From Appendix D the probability of a z value between 0 and 2.41 is .4920. We want to determine the likelihood of a value *greater than* 2.41, so $.5000 - .4920 = .0080$. We conclude that the likelihood of finding a z value of 2.41 or larger when the null hypothesis is true is 0.80 percent. It is unlikely, therefore, that the null hypothesis is true.

SELF-REVIEW 10-2



According to recent information from the American Automobile Association, the mean age of passenger cars in the United States is 8.4 years. A sample of 40 cars in the student lots at the University of Tennessee showed the mean age to be 9.2 years. The standard deviation of this sample was 2.8 years. At the .01 significance level can we conclude the mean age is more than 8.4 years for the cars of Tennessee students?

(a) State the null hypothesis and the alternate hypothesis.
(b) Explain why z is the test statistic.
(c) What is the critical value of the test statistic?
(d) Compute the value of the test statistic.
(e) What is your decision regarding the null hypothesis?
(f) Interpret your decision from part e in a single sentence.
(g) What is the p-value?

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.

1. 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 follows the normal distribution and the standard deviation is 5. Use the .05 significance level.

2. 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 follows the normal distribution and the standard deviation is 3. Use the .02 significance level.

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

H_0: $\mu \leq 20$
H_1: $\mu > 20$

4. A sample of 64 observations is selected from a normal population. The sample mean is 215, and the sample 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.

5. 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 Crosset Truck Company bought 48 tires and found that the mean mileage for their trucks is 59,500 miles with a standard deviation of 5,000 miles. Is Crosset's experience different from that claimed by the manufacturer at the .05 significance level?

6. The MacBurger restaurant chain claims that the waiting time of customers for service is normally distributed, with a mean of 3 minutes and a 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?
7. A recent national survey found that high school students watched an average (mean) of 6.8 DVDs per month. A random sample of 36 college students revealed that the mean number of DVDs watched last month was 6.2, with a standard deviation of 0.5. 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." Over the first 35 days she was employed at the restaurant, the mean daily amount of her tips was $84.85, with a standard deviation of $11.38. At the .01 significance level, can Ms. Brigden conclude that she is earning an average of more than $80 in tips?

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 persons 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 that they are successful 90 percent of the time with this claim. A recent audit revealed they were successful 190 times out of 200. Can they conclude that their 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 π 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

Prior elections in Indiana indicate 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 Indiana.

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

SOLUTION

The following test of hypothesis can be conducted 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 Indiana, 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:

TEST OF HYPOTHESIS, ONE PROPORTION

$$z = \frac{p - \pi}{\sigma_p} \qquad \textbf{[10–3]}$$

where:

π is the population proportion.
p is the sample proportion.
n is the sample size.
σ_p is the standard error of the proportion. It is computed by $\sqrt{\pi(1 - \pi)/n}$, so the formula for z becomes:

TEST OF HYPOTHESIS, ONE PROPORTION

$$z = \frac{p - \pi}{\sqrt{\frac{\pi(1 - \pi)}{n}}} \qquad \textbf{[10–4]}$$

Finding the critical value

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 we use only the left side of the distribution. (See Chart 10–6.) 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 D 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.

Select a sample and make a decision regarding 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 Indiana revealed that 1,550 planned to vote for the

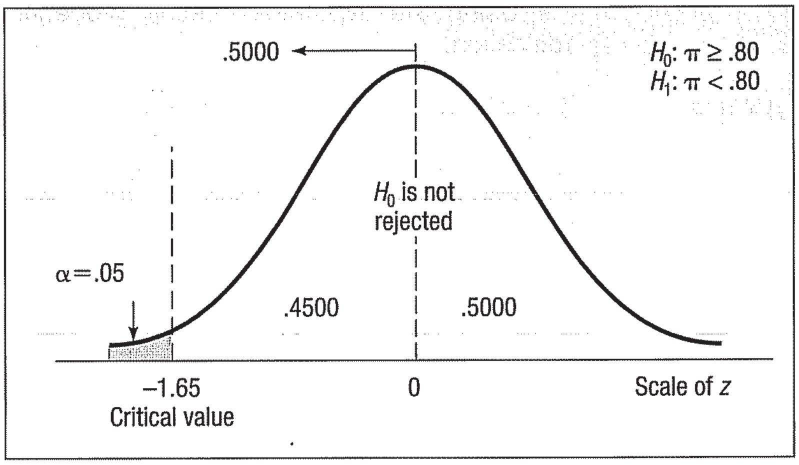


CHART 10–6 Rejection Region for the .05 Level of Significance, One-Tailed Test

incumbent governor. Is the sample proportion of .775 (found by 1,550/2,000) 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.

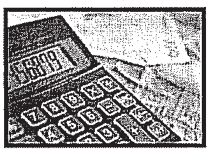
Using formula (10–4) and computing z gives

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

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 likely not due to sampling error. 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 D, 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 10–3



A recent insurance industry report indicated that 40 percent of those persons involved in minor traffic accidents this year have been involved in at 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.

(a) Can we use z as the test statistic? Tell why or why not.

(b) State the null hypothesis and the alternate hypothesis.
(c) Show the decision rule graphically.
(d) Compute the value of z and state your decision regarding the null hypothesis.
(e) Determine and interpret the p-value.

Exercises

9. 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?
a. State the decision rule.
b. Compute the value of the test statistic.
c. What is your decision regarding the null hypothesis?

10. 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?
a. State the decision rule.
b. Compute the value of the test statistic.
c. 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.

11. 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?

12. 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?

13. 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?

14. Research at the University of Toledo indicates that 50 percent of the 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.

Testing for a Population Mean: Small Sample, Population Standard Deviation Unknown

We are able to use the standard normal distribution, that is z, under two conditions:

1. The population is known to follow a normal distribution and the population standard deviation is known, or
2. The shape of the population is not known, but the number of observations in the sample is at least 30.

What do we do when the sample is less than 30 and the population standard deviation is not known? We encountered this same situation when constructing confidence intervals in the previous chapter. See pages 254–259 in Chapter 9. We summarized this problem in Chart 9–3 on page 256. 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.

To conduct a test of hypothesis using the *t* distribution, we adjust formula (10–2) as follows.

SMALL SAMPLE TEST FOR MEAN

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \qquad [10\text{–}5]$$

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

$\bar{X}$ is the mean of the sample.
μ is the hypothesized population mean.
s is the standard deviation of the sample.
n is the number of observations in the sample.

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 they 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 and determined the cost to process these selected claims. 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.

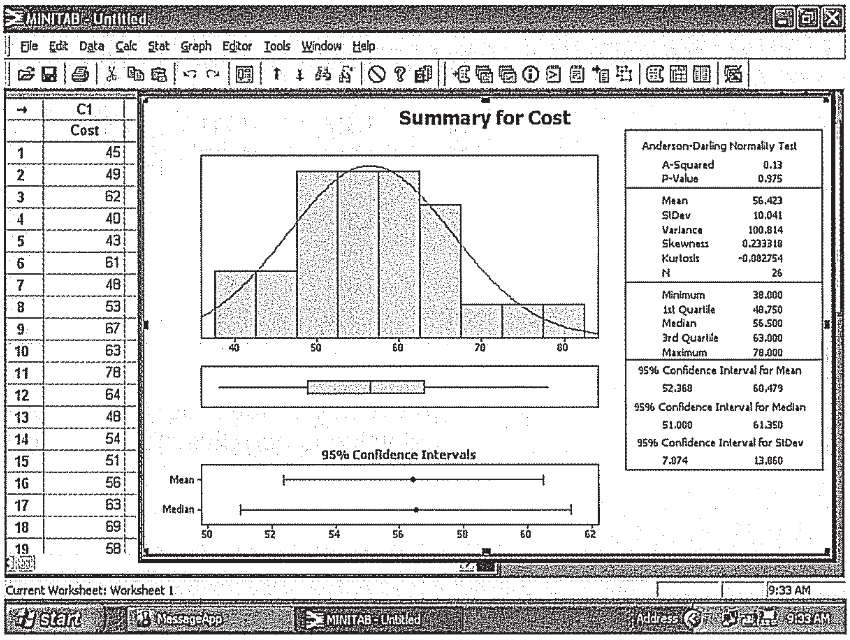
Step 1: State the null hypothesis and the alternate hypothesis. The null hypothesis is that the population mean is at least $60. The alternate hypothesis is that the population mean is less than $60. We can express the null and alternate hypotheses as follows:

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. When the sample is large we can make the substitution and still use the standard normal distribution. We usually define large as 30 or more observations. In this case there are only 26 observations. Consequently we cannot use the standard normal distribution. Instead, we use t. The value of the test statistic is computed by formula (10–5):

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

Step 4: Formulate the decision rule. The critical values of t are given in Appendix F, a portion of which is shown in Table 10–1. Appendix F 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 samples, written $n - 1$. In this case the number of observations in the sample is 26, 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 "0.010" until it intersects the row with 25 degrees of freedom. The value is 2.485. Because

TABLE 10–1 A Portion of the *t* Distribution Table

	Confidence Intervals					
	80%	90%	95%	98%	99%	99.9%
	Level of Significance for One-Tailed Test, α					
df	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

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.

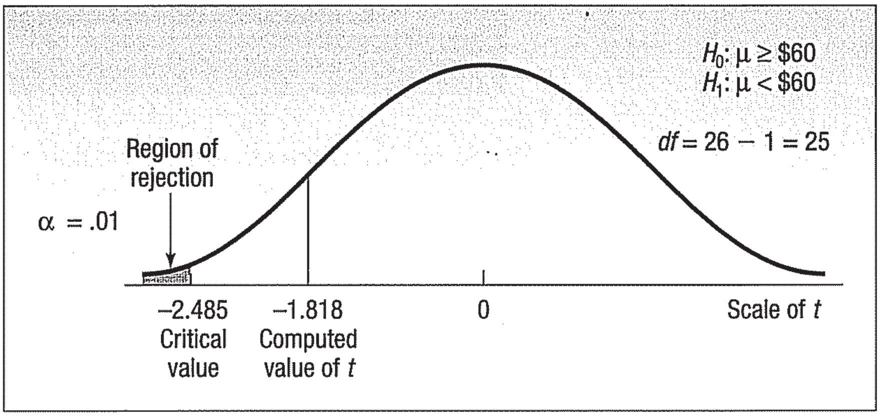


CHART 10-7 Rejection Region, *t* Distribution, .01 Significance Level

Step 5: Make a decision and interpret the result. From the MINITAB output on page 297, 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 (10–5) 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 10–7. It is in the region where the null hypothesis is *not* rejected.

In the previous example the mean and the standard deviation were calculated by MINITAB. The following example requires this information to be computed from the 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 F 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 test statistic, t, is to the left of −2.718 or to the right of 2.718. This information is summarized in Chart 10–8.

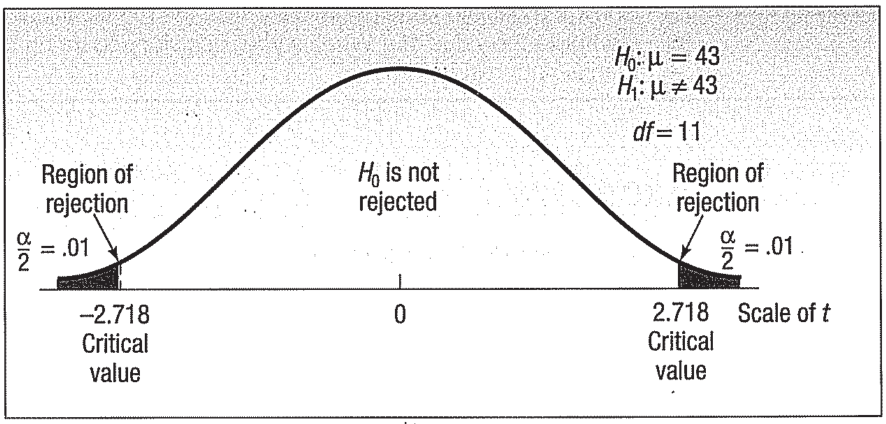


CHART 10–8 Regions of Rejection, Two-Tailed Test, Student's t Distribution, $\alpha = .02$

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

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

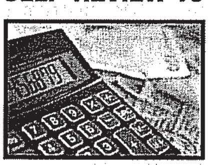
$$s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n - 1}} = \sqrt{\frac{35}{12 - 1}} = 1.784$$

Now we are ready to compute the value of t, using formula (10–5).

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

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 10–4



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 with the objective of making it 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?

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

Exercises

15. 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:

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

16. 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:
a. State the decision rule.
b. Compute the value of the test statistic.
c. What is your decision regarding the null hypothesis?

17. 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?

18. 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?

19. 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?

20. 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 a 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

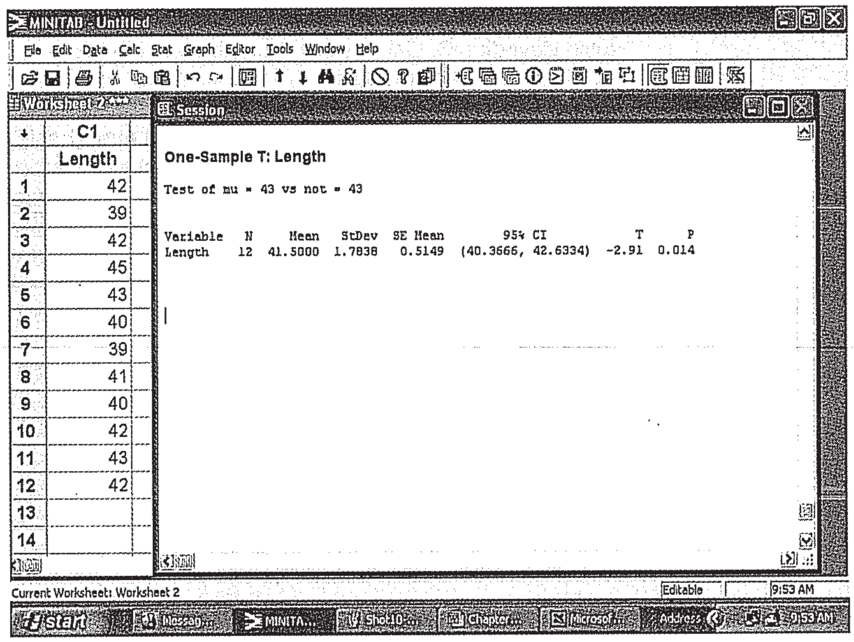
At the .05 significance level can the watchdog agency conclude the mean number of complaints per airport is less than 15 per month?
a. What assumption is necessary before conducting a test of hypothesis?
b. Plot the number of complaints per airport in a frequency distribution or a dot plot. Is it reasonable to conclude that the population follows a normal distribution?
c. Conduct a test of hypothesis and interpret the results.

A Software Solution

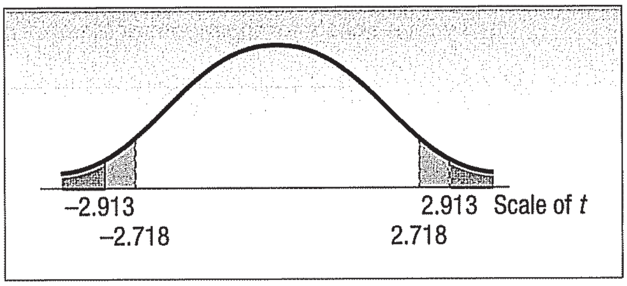
The MINITAB statistical software system, used in earlier chapters, provides an efficient way of conducting a one-sample test of hypothesis for a population mean. The steps to generate the following output are shown in the Software Commands section at the end of the chapter.

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 hypothesis is true. In the case of the Example on page 299, 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 on the next page, in which the p-value of .014 is shown in color and the significance level is the color area plus the grey 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 upper and the lower tails. 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.

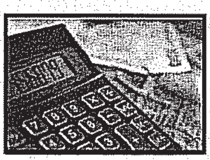
What if we were conducting a one-sided test, so that the entire rejection region would be in either the upper or the lower tail? In that case, we would report the area from only the one tail. In the counterbalance example, if H_1 were stated as $\mu < 43$, the inequality would point to the left. Thus, 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 F 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 F is reproduced as Table 10–3.) The two-tailed significance level corresponding to 2.718 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."

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

	Confidence Intervals					
	80%	90%	95%	98%	99%	99.9%
	Level of Significance for One-Tailed Test, α					
df	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
⋮	⋮	⋮	⋮	⋮	⋮	⋮
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

SELF-REVIEW 10–5 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?

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

Exercises

21. 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?
a. State the decision rule.
b. Compute the value of the test statistic.
c. What is your decision regarding the null hypothesis?
d. Estimate the p-value.

22. 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?
a. State the decision rule.
b. Compute the value of the test statistic.
c. What is your decision regarding the null hypothesis?
d. Estimate the p-value.

23. 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.

24. 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.

25. 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.

26. 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.

Chapter Outline

I. The objective of hypothesis testing is to check the validity of a statement about a population.

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 standard normal distribution is used when the population standard deviation is unknown, but the sample contains at least 30 observations.
   c. The *t* distribution is used when the population follows the normal distribution, the population standard deviation is unknown, and the sample contains fewer than 30 observations.

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-sample 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. Testing a hypothesis about a population mean.

A. If the population follows a normal distribution and the population standard deviation, σ, is known, the test statistic is the standard normal distribution and is determined from:

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

B. If the population standard deviation is not known, but there are at least 30 observations in the sample, s is substituted for σ. The test statistic is the standard normal distribution, and its value is determined from:

$$z = \frac{\bar{X} - \mu}{s/\sqrt{n}} \qquad \textbf{[10–2]}$$

C. If the population standard deviation is not known, but there are fewer than 30 observations in the sample, s is substituted for σ. The test statistic is the t distribution, and its value is determined from:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \qquad \textbf{[10–5]}$$

The major characteristics of the t distribution are:
   1. It is a continuous distribution.
   2. It is mound-shaped and symmetric.
   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. Testing about a population proportion.

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

B. The test statistic is

$$z = \frac{p - \pi}{\sqrt{\dfrac{\pi(1 - \pi)}{n}}} \qquad \textbf{[10–4]}$$

Pronunciation Key

SYMBOL	MEANING	PRONUNCIATION
H_0	Null hypothesis	*H sub zero*
H_1	Alternate hypothesis	*H sub one*
$\alpha/2$	Two-tailed significance level	*Alpha over 2*

Chapter Exercises

27. A new weight-watching company, Weight Reducers International, advertises that those who join will lose, on the average, 10 pounds the first two weeks. A random sample of 50 people who joined the new weight reduction program revealed the mean loss to be 9 pounds with a standard deviation of 2.8 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.

28. Dole Pineapple, Inc. is concerned that the 16-ounce can of sliced pineapple is being overfilled. The quality-control department took a random sample of 50 cans and found that the arithmetic mean weight was 16.05 ounces, with a sample standard deviation of 0.03

ounces. At the 5 percent level of significance, can we conclude that the mean weight is greater than 16 ounces? Determine the *p*-value.

29. 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.

30. A statewide real estate sales agency, Farm Associates, specializes in selling farm property in the state of Nebraska. Their records indicate that the mean selling time of farm property is 90 days. Because of recent drought conditions, they believe 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?

31. According to the local union president, the mean income of plumbers in the Salt Lake City follows the normal distribution. This normal distribution has 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.

32. 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.

33. NBC TV news, in a segment on the price of gasoline, reported last evening that the mean price nationwide is $2.10 per gallon for self-serve regular unleaded. A random sample of 35 stations in the Milwaukee, Wisconsin, area revealed that the mean price was $2.12 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.

34. The Rutter Nursery Company packages their 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.

35. 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?

36. 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.

37. 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.

38. 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.

39. 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 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.

40. 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?

41. 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 debit 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.

42. 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.

43. 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, reported in cups, of coffee 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?

44. 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.

45. egolf.com receives an average of 6.5 returns per day from online shoppers. For a sample of 12 days, they 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?

46. 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 as 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.

47. The Watch Corporation of Switzerland claims that their 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.

48. 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

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

49. Many grocery stores and large retailers such as Wal-Mart and K-Mart 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.

50. In the year 2003 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?

51. The 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.

exercises.com

52. The *USA Today* (http://www.usatoday.com/sports/baseball/front.htm) and Major League Baseball (http://www.majorleaguebaseball.com) websites regularly report information on individual player salaries. Go to one of these sites 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* $90.0 million? If you are more of a football, basketball, or hockey enthusiast, information is also available on their teams' salaries.

53. The Gallup Organization in Princeton, New Jersey, is one of the best-known polling organizations in the United States. They often combine with *USA Today* or CNN to conduct polls of current interest. They also maintain 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 Fast Facts. 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.

Dataset Exercises

54. Refer to the Real Estate data, which reports information on the homes sold in Denver, Colorado, last year.

- **a.** 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?
- **b.** 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?
- **c.** 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?
- **d.** 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?

55. Refer to the Baseball 2003 data, which reports information on the 30 Major League Baseball teams for the 2003 season.

- **a.** 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.
- **b.** Conduct a test of hypothesis to determine whether the mean attendance was more than 2,000,000 per team.

56. Refer to the Wage data, which reports 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.

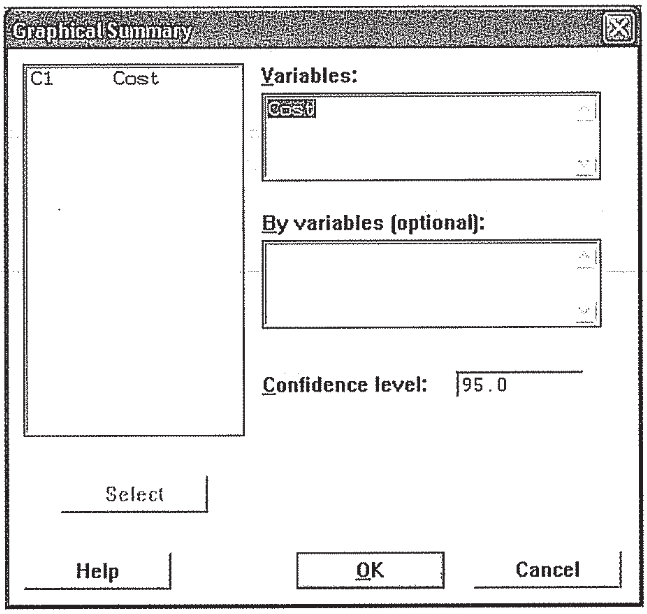
- **a.** Conduct a test of hypothesis to determine if the mean annual wage is greater than $30,000. Use the .05 significance level. Determine the p-value and interpret the result.
- **b.** Conduct a test of hypothesis to determine if the mean years of experience is different from 20. Use the .05 significance 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.

57. Refer to the CIA data, which reports demographic and economic information on 46 different countries.

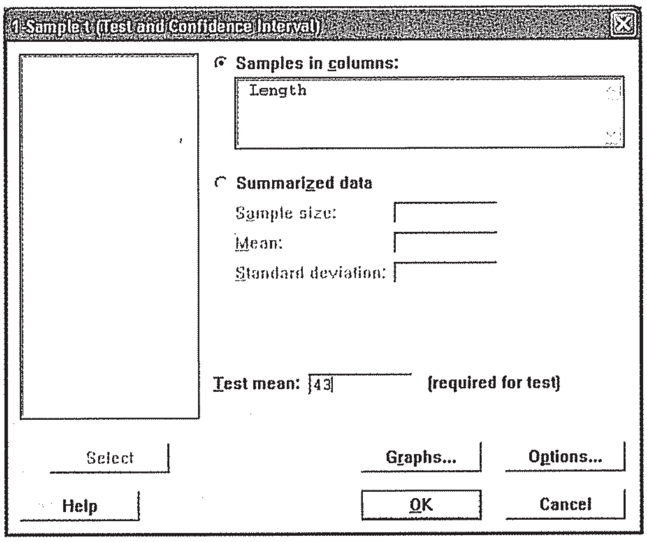
- **a.** Conduct a test of hypothesis to determine if the mean number of cell phones is greater than 4.0. Use the .05 significance level. What is the p-value?
- **b.** Conduct a test of hypothesis to determine if the mean size of the labor force is less than 50. Use the .05 significance level. What is the p-value?

Software Commands

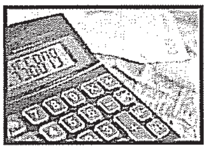
1. The MINITAB commands for the histogram and the descriptive statistics on page 297 are:
   a. Enter the 26 sample observations in column C1 and name the variable **Cost.**
   b. From the menu bar select **Stat, Basic Statistics,** and **Graphical Summary.** In the dialog box select **Cost** as the variable and click **OK.**



2. The MINITAB commands for the one-sample *t* test on page 302 are:
   a. Enter the sample data into column C1 and name the variable **Length.**
   b. From the menu bar select **Stat, Basic Statistics,** and **1-Sample t** and then hit **Enter.**
   c. Select **Length** as the variable, select **Test mean,** insert the number **43** and click **OK.**



Chapter 10 Answers to Self-Review



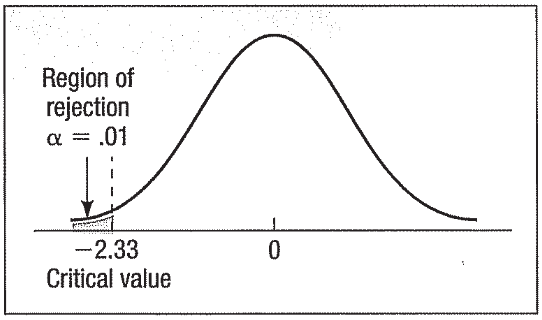
10–1 **a.** H_0: $\mu = 6.0$; H_1: $\mu \neq 6.0$
b. .05.
c. $z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$
d. Do not reject the null hypothesis if the computed z value falls between -1.96 and $+1.96$.
e. Yes. Computed $z = -2.56$, found by:

$$z = \frac{5.84 - 6.0}{0.5/\sqrt{64}} = \frac{-0.16}{.0625} = -2.56$$

Reject H_0 at the .05 level. Accept H_1. The mean turnover rate is not equal to 6.0.

10–2 **a.** H_0: $\mu \leq 8.4$; H_1: $\mu > 8.4$
b. Large sample, $n > 30$.
c. Reject H_0 if $z > 2.33$.
d. $z = \dfrac{9.2 - 8.4}{2.8/\sqrt{40}} = 1.81$
e. Do not reject H_0.
f. We cannot conclude that the mean age is greater than 8.4 years.
g. p-value $= .5000 - .4649 = .0351$

10–3 **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: $\pi < .40$
c.

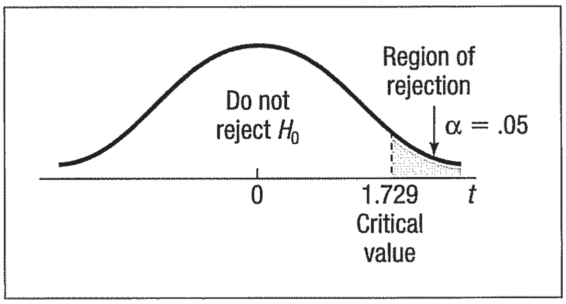


d. $z = -0.87$, found by:

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

Do not reject H_0.
e. The p-value is .1922, found by $.5000 - .3078$.

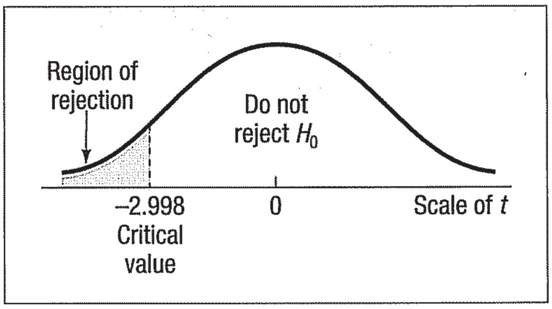
10–4 **a.** H_0: $\mu \leq 305$, H_1: $\mu > 305$.
b. $df = n - 1 = 20 - 1 = 19$



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

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

10–5 **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:

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

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

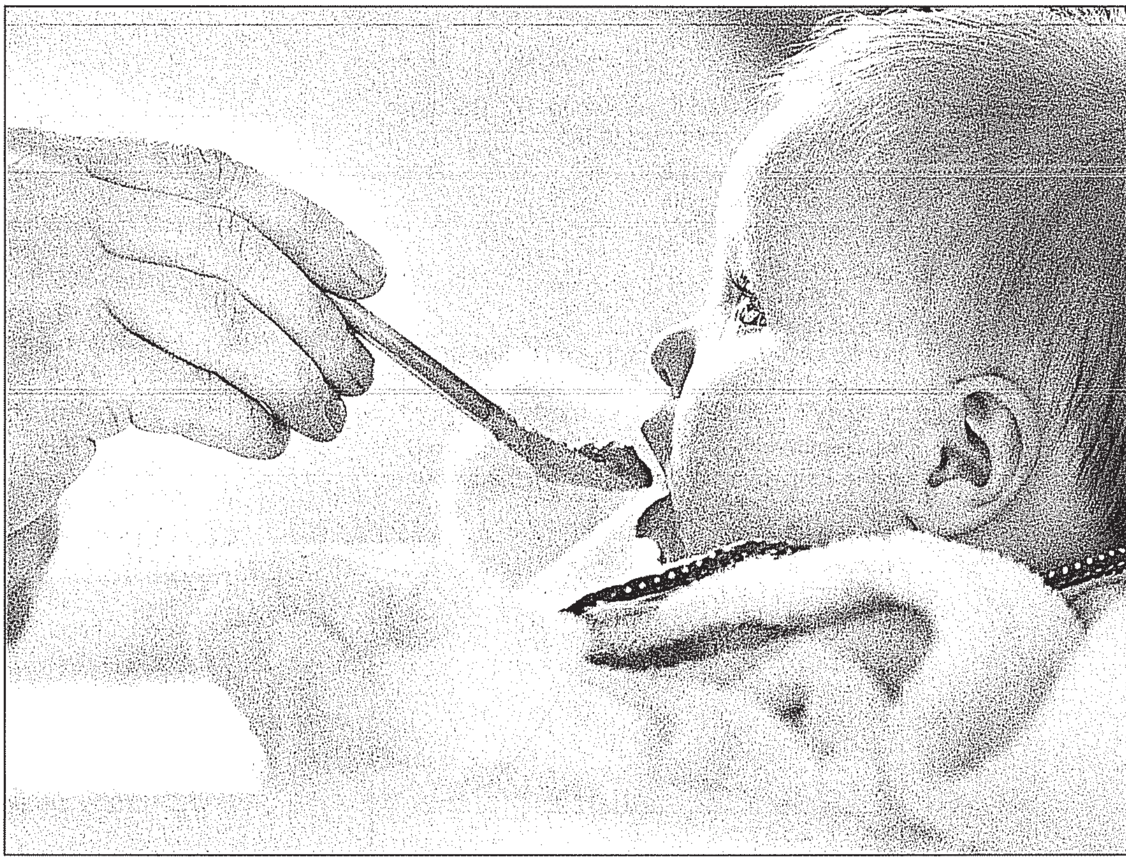
Then

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

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.

11

Two-Sample Tests of Hypothesis



The Gibbs Baby Food Company wishes to compare the weight gain of infants using their brand versus their competitor's. A sample of 40 babies using the Gibbs products revealed a mean weight gain of 7.6 pounds in the first three months after birth. The standard deviation of the sample was 2.3 pounds. A sample of 55 babies using the competitor's brand revealed a mean increase in weight of 8.1 pounds, with a standard deviation of 2.9 pounds. At the .05 significance level, can we conclude that babies using the Gibbs brand gained less weight? Compute the *p*-value and interpret it. (See Goal 1 and Exercise 3.)

GOALS

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

1 Conduct a test of a hypothesis about the difference between two independent population means.

2 Conduct a test of a hypothesis about the difference between two population proportions.

3 Conduct a test of a hypothesis about the mean difference between paired or dependent observations.

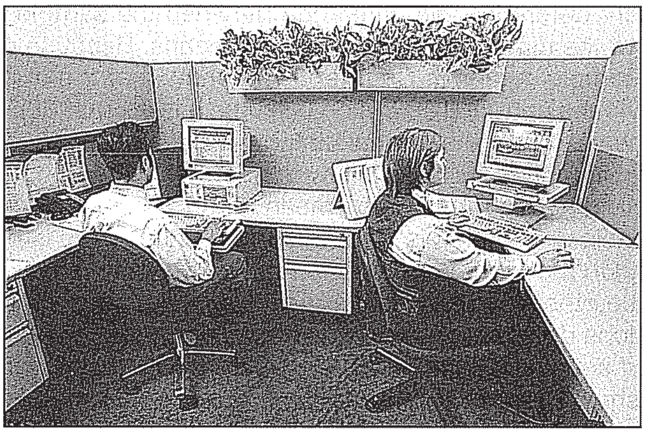
4 Understand the difference between dependent and independent samples.

Introduction

Chapter 10 began our study of hypothesis testing. We described the nature of hypothesis testing and conducted tests of a hypothesis in which we compared the results of a single sample to a population value. That is, we selected a single random sample from a population and conducted a test of whether the proposed population value was reasonable. Recall, in Chapter 10 we selected a sample of the number of desks assembled per week at the Jamestown Steel Company to determine whether there was a change in the production rate. Similarly, we sampled voters in one area of Indiana to determine whether the population proportion that would support the governor for reelection was less than .80. In both of these cases, we compared the results of a *single* sample statistic to a population parameter.

In this chapter we expand the idea of hypothesis testing to two samples. That is, we select random samples from two different populations to determine whether the population means or the population proportions are equal. Some questions we might want to test are:

1. Is there a difference in the mean value of residential real estate sold by male agents and female agents in south Florida?
2. Is there a difference in the mean number of defects produced on the day and the afternoon shifts at Kimble Products?
3. Is there a difference in the mean number of days absent between young workers (under 21 years of age) and older workers (more than 60 years of age) in the fast-food industry?
4. Is there is a difference in the proportion of Ohio State University graduates and University of Cincinnati graduates who pass the state Certified Public Accounting Examination on their first attempt?
5. Is there an increase in the production rate if music is piped into the production area?



We begin this chapter with the case in which we select random samples from two populations and wish to investigate whether these populations have the same mean.

Two-Sample Tests of Hypothesis: Independent Samples

A city planner in Florida wishes to know whether there is a difference in the mean hourly wage rate of plumbers and electricians in central Florida. A financial accountant wishes to know whether the mean rate of return for high yield mutual funds is different from the mean rate of return on global mutual funds. In each of these cases there are two independent populations. In the first case, the plumbers represent one population and the electricians the other. In the second case, high yield mutual funds are one population and global mutual funds the other.

In each of these cases, to investigate the question, we would select a random sample from each population and compute the mean of the two samples. If the two population means are the same, that is, the mean hourly rate is the same for the plumbers and the electricians, we would expect the *difference* between the two

Statistics in Action

The Election of 2000 turned out to be one of the closest in history. The news media were unable to project a winner, and the final decision, including recounts and court decisions, took more than five weeks. This was not the only election in which there was controversy. Shortly before the 1936 presidential election, the *New York Times* carried the headline: "*Digest* Poll Gives Landon 32 States: Landon Leads 4–3." However, Alfred Landon of Kansas was not elected President. In fact, Roosevelt won by more than 11 million votes and received 523 Electoral College votes. How could the headline have been so wrong?

The *Literary Digest* collected a sample of voters from lists of telephone numbers, automobile registrations, and *Digest* readers. In 1936 not many people could afford a telephone or an automobile. In addition those who read the *Digest* tended to be wealthier and vote

Republican. Thus, the population that was sampled did not represent the population of voters. A second problem was with the nonresponses. More than 10 million people were sent surveys, and more than 2.3 million responded. However, no attempt was made to see whether those responding represented a cross-section of all the voters.

With modern computers and survey methods, samples are carefully selected and checked to ensure they are representative. What happened to the *Literary Digest?* It went out of business shortly after the 1936 election.

sample means to be zero. But what if our sample results yield a difference other than zero? Is that difference due to chance or is it because there is a real difference in the hourly earnings? A two-sample test of means will help to answer this question.

We do need to return to the results of Chapter 8. Recall that we showed that a distribution of sample means would tend to approximate the normal distribution when the sample size is at least 30. We need to again assume that a distribution of sample means will follow the normal distribution. It can be shown mathematically that the distribution of the differences between sample means for two normal distributions is also normal.

We can illustrate this theory in terms of the city planner in Tampa, Florida. To begin, let's assume some information that is not usually available. Suppose that the population of plumbers has a mean of \$30.00 per hour and a standard deviation of \$5.00 per hour. The population of electricians has a mean of \$29.00 and a standard deviation of \$4.50. Now, from this information it is clear that the two population means are not the same. The plumbers actually earn \$1.00 per hour more than the electricians. But we cannot expect to uncover this difference each time we sample the two populations.

Suppose we select a random sample of 40 plumbers and a random sample of 35 electricians and compute the mean of each sample. Then, we determine the difference between the sample means. It is this difference between the sample means that holds our interest. If the populations have the same mean, then we would expect the difference between the two sample means to be zero. If there is a difference between the population means, then we expect to find a difference between the sample means.

To understand the theory, we need to take several pairs of samples, compute the mean of each, determine the difference between the sample means, and study the distribution of the differences in the sample means. Because of our study of the distribution of sample means in Chapter 8, we know that the distribution of the sample means follows the normal distribution (assume n is at least 30). If the two distributions of sample means follow the normal distribution, then we can reason that the distribution of their differences will also follow the normal distribution. This is the first hurdle.

The second hurdle refers to the mean of this distribution of differences. If we find the mean of this distribution is zero, that implies that there is no difference in the two populations. On the other hand, if the mean of the distribution of differences is equal to some value other than zero, either positive or negative, then we conclude that the two populations do not have the same mean.

To report some concrete results, let's return to the city planner in Tampa, Florida. Table 11–1 shows the result of selecting 20 different samples of 40 plumbers and 35 electricians, computing the mean of each sample, and finding the difference between the two sample means. In the first case the sample of 40 plumbers has a mean of \$29.80, and for the 35 electricians the mean is \$28.76. The difference between the sample means is \$1.04. This process was repeated 19 more times. Observe that in 17 of the 20 cases the mean of the plumbers is larger than the mean of the electricians.

Our final hurdle is that we need to know something about the *variability* of the distribution of differences. To put it another way, what is the standard deviation of this distribution of differences? Statistical theory shows that when we have independent populations, such as the case here, the distribution of the differences has a variance (standard deviation squared) equal to the sum of the two individual variances. This means that we can add the variances of the two sampling distributions.

VARIANCE OF THE DISTRIBUTION OF DIFFERENCES IN MEANS

$$s^2_{\bar{X}_1-\bar{X}_2} = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \qquad [11\text{–}1]$$

The term $s^2_{\bar{X}_1-\bar{X}_2}$ looks complex but need not be difficult to interpret. The s^2 portion reminds us that it is a sample variance, and the subscript $\bar{X}_1 - \bar{X}_2$ that it is a distribution of differences in the sample means.

TABLE 11–1 The Means of Random Samples of Plumbers and Electricians

Sample	Plumbers	Electricians	Difference
1	$29.80	$28.76	$1.04
2	30.32	29.40	0.92
3	30.57	29.94	0.63
4	30.04	28.93	1.11
5	30.09	29.78	0.31
6	30.02	28.66	1.36
7	29.60	29.13	0.47
8	29.63	29.42	0.21
9	30.17	29.29	0.88
10	30.81	29.75	1.06
11	30.09	28.05	2.04
12	29.35	29.07	0.28
13	29.42	28.79	0.63
14	29.78	29.54	0.24
15	29.60	29.60	0.00
16	30.60	30.19	0.41
17	30.79	28.65	2.14
18	29.14	29.95	−0.81
19	29.91	28.75	1.16
20	28.74	29.21	−0.47

We can put this equation in a more usable form by taking the square root, so that we have the standard deviation of the distribution of the differences. Finally, we standardize the distribution of the differences. The result is the following equation.

TEST STATISTIC FOR THE DIFFERENCE BETWEEN TWO MEANS

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad \textbf{[11–2]}$$

Before we present an example, let's review the assumptions necessary for using formula (11–2).

Assumptions for large sample test

1. The two samples must be unrelated, that is, independent.
2. The samples must be large enough that the distribution of the sample means follows the normal distribution. The usual practice is to require that both samples have at least 30 observations.

The following example shows the details of the two-sample test of hypothesis for two population means.

EXAMPLE

Customers at FoodTown Super Markets have a choice when paying for their groceries. They may check out and pay using the standard cashier assisted checkout, or they may use the new U-Scan procedure. In the standard procedure a FoodTown employee scans each item, puts it on a short conveyor where another employee puts it in a bag and then into the grocery cart. In the U-Scan procedure the customer scans each item, bags it, and places the bags in the cart themselves. The U-Scan procedure is designed to reduce the time a customer spends in the checkout line.

The U-Scan facility was recently installed at the Byrne Road FoodTown location. The store manager would like to know if the mean checkout time using the standard checkout method is longer than using the U-Scan. She gathered the following sample information. The time is measured from when the customer enters the line until their bags are in the cart. Hence the time includes both waiting in line and checking out.

Customer Type	Sample Mean	Sample Standard Deviation	Sample Size
Standard	5.50 minutes	0.40 minutes	50
U-Scan	5.30 minutes	0.30 minutes	100

SOLUTION

We use the five-step hypothesis testing procedure to investigate the question.

Step 1: State the null hypothesis and the alternate hypothesis. The null hypothesis is that there is no difference in the mean checkout times for the two groups. In other words, the difference of 0.20 minutes between the mean checkout time for the standard method and the mean checkout time for U-Scan is due to chance. The alternate hypothesis is that the mean checkout time is longer for those using the standard method. We will let μ_s refer to the mean checkout time for the population of standard customers and μ_u the mean checkout time for the U-Scan customers. The null and alternative hypotheses are:

$$H_0: \mu_s \leq \mu_u$$
$$H_1: \mu_s > \mu_u$$

Step 2: Select the level of significance. The significance level is the probability that we reject the null hypothesis when it is actually true. This likelihood is determined prior to selecting the sample or performing any calculations. The .05 and .01 significance levels are the most common, but other values, such as .02 and .10, are also used. In theory, we may select any value between 0 and 1 for the significance level. In this case we selected the .01 significance level.

Step 3: Determine the test statistic. In Chapter 10 we used the standard normal distribution (that is z) and t as test statistics. In this case, because the samples are large, we use the z distribution as the test statistic.

Step 4: Formulate a decision rule. The decision rule is based on the null and the alternate hypotheses (i.e., one-tailed or two-tailed test), the level of significance, and the test statistic used. We selected the .01 significance level, the z distribution as the test statistic, and we wish to determine whether the mean checkout time is longer using the standard method. We set the alternate hypothesis to indicate that the mean checkout time is longer for those using the standard method than the U-Scan method. Hence, the rejection region is in the upper tail of the standard normal distribution. To find the critical value, place .01 of the total area in the upper tail. This means that .4900 (.5000 − .0100) of the area is located between the z value of 0 and the critical value. Next, we search the body of Appendix D for a value located near .4900. It is 2.33, so our decision rule is to reject H_0 if the value computed from the test statistic exceeds 2.33. Chart 11–1 depicts the decision rule.

Step 5: Make the decision regarding H_0 and interpret the result. We use formula (11–2) to compute the value of the test statistic.

$$z = \frac{\bar{X}_s - \bar{X}_u}{\sqrt{\frac{s_s^2}{n_s} + \frac{s_u^2}{n_u}}} = \frac{5.5 - 5.3}{\sqrt{\frac{0.40^2}{50} + \frac{0.30^2}{100}}} = \frac{0.2}{0.064} = 3.13$$

Statistics in Action

Do you live to work or work to live? A recent poll of 802 working Americans revealed that, among those who considered their work as a career, the mean number of hours worked per day was 8.7. Among those who considered their work a job, the mean number of hours worked per day was 7.6.

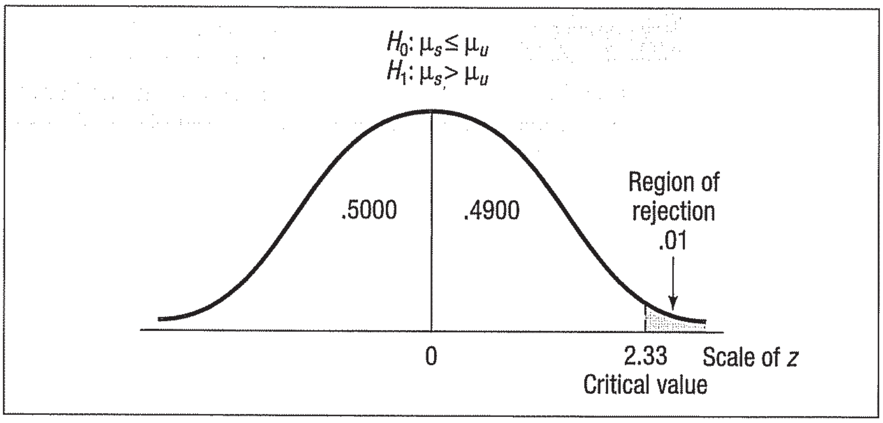


CHART 11–1 Decision Rule for One-Tailed Test at .01 Significance Level

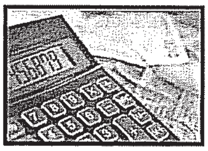
The computed value of 3.13 is larger than the critical value of 2.33. Our decision is to reject the null hypothesis and accept the alternate hypothesis. The difference of .20 minutes between the mean checkout time using the standard method is too large to have occurred by chance. To put it another way, we conclude the U-Scan method is faster.

What is the *p*-value for the test statistic? Recall that the *p*-value is the probability of finding a value of the test statistic this extreme when the null hypothesis is true. To calculate the *p*-value we need the probability of a *z* value larger than 3.13. From Appendix D we cannot find the probability associated with 3.13. The largest value available is 3.09. The area corresponding to 3.09 is .4990. In this case we can report that the *p*-value is less than .0010, found by .5000 − .4990. We conclude that there is very little likelihood that the null hypothesis is true!

In summary, the criteria for using the large sample test of means are:

1. *The samples are from independent populations*. This means, for example, that the sample checkout time for the U-Scan customers is unrelated to the checkout time for the other customers. If Mr. Smith is a FoodTown customer and his response time is sampled, that does not affect the checkout time for any other customers.
2. *Both sample sizes are at least 30*. In the FoodTown example, one sample was 50 and the other 100. Because both samples are considered large, we can substitute the sample standard deviations for the population standard deviations and use formula (11–2) to find the value of the test statistic.

Self-Review 11–1



Tom Sevits is the owner of the Appliance Patch. Recently Tom observed a difference in the dollar value of sales between the men and women he employs as sales associates. A sample of 40 days revealed the men sold a mean of $1,400 worth of appliances per day with a standard deviation of $200. For a sample of 50 days, the women sold a mean of $1,500 worth of appliances per day with a standard deviation of $250. At the .05 significance level can Mr. Sevits conclude that the mean amount sold per day is larger for the women?

(a) State the null hypothesis and the alternate hypothesis.
(b) What is the decision rule?
(c) What is the value of the test statistic?
(d) What is your decision regarding the null hypothesis?
(e) What is the *p*-value?
(f) Interpret the result.

Exercises

1. A sample of 40 observations is selected from one population. The sample mean is 102 and the sample standard deviation is 5. A sample of 50 observations is selected from a second population. The sample mean is 99 and the sample standard deviation is 6. Conduct the following test of hypothesis using the .04 significance level.

   $H_0: \mu_1 = \mu_2$
   $H_1: \mu_1 \neq \mu_2$

   a. Is this a one-tailed or a two-tailed test?
   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? Compute and interpret the *p*-value.
2. A sample of 65 observations is selected from one population. The sample mean is 2.67 and the sample standard deviation is 0.75. A sample of 50 observations is selected from a second population. The sample mean is 2.59 and the sample standard deviation is 0.66. Conduct the following test of hypothesis using the .08 significance level.

   $H_0: \mu_1 \leq \mu_2$
   $H_1: \mu_1 > \mu_2$

   a. Is this a one-tailed or a two-tailed test?
   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? Compute and interpret the *p*-value.

Note: Use the five-step hypothesis testing procedure to solve the following exercises.

3. The Gibbs Baby Food Company wishes to compare the weight gain of infants using their brand versus their competitor's. A sample of 40 babies using the Gibbs products revealed a mean weight gain of 7.6 pounds in the first three months after birth. The standard deviation of the sample was 2.3 pounds. A sample of 55 babies using the competitor's brand revealed a mean increase in weight of 8.1 pounds, with a standard deviation of 2.9 pounds. At the .05 significance level, can we conclude that babies using the Gibbs brand gained less weight? Compute the *p*-value and interpret it.
4. As part of a study of corporate employees, the Director of Human Resources for PNC, Inc. wants to compare the distance traveled to work by employees at their office in downtown Cincinnati with the distance for those in downtown Pittsburgh. A sample of 35 Cincinnati employees showed they travel a mean of 370 miles per month, with a standard deviation of 30 miles per month. A sample of 40 Pittsburgh employees showed they travel a mean of 380 miles per month, with a standard deviation of 26 miles per month. At the .05 significance level, is there a difference in the mean number of miles traveled per month between Cincinnati and Pittsburgh employees? Use the five-step hypothesis-testing procedure.
5. A financial analyst wants to compare the turnover rates, in percent, for shares of oil-related stocks versus other stocks, such as GE and IBM. She selected 32 oil-related stocks and 49 other stocks. The mean turnover rate of oil-related stocks is 31.4 percent and the standard deviation 5.1 percent. For the other stocks, the mean rate was computed to be 34.9 percent and the standard deviation 6.7 percent. Is there a significant difference in the turnover rates of the two types of stock? Use the .01 significance level.
6. Mary Jo Fitzpatrick is the Vice President for Nursing Services at St. Luke's Memorial Hospital. Recently she noticed that unionized jobs for nurses seem to offer higher wages. She decided to investigate and gathered the following sample information.

Group	Mean Wage	Sample Standard Deviation	Sample Size
Union	\$20.75	\$2.25	40
Nonunion	\$19.80	\$1.90	45

   Would it be reasonable for her to conclude that union nurses earn more? Use the .02 significance level. What is the *p*-value? Compute and interpret the *p*-value.

Two-Sample Tests about Proportions

In the previous section, we considered a test involving population means. However, we are often interested also in whether two sample proportions came from populations that are equal. Here are several examples.

- The Vice President of Human Resources wishes to know whether there is a difference in the proportion of hourly employees who miss more than 5 days of work per year at the Atlanta and the Houston plants.
- General Motors is considering a new design for the Pontiac Grand Am. The design is shown to a group of potential buyers under 30 years of age and another group over 60 years of age. Pontiac wishes to know whether there is a difference in the proportion of the two groups who like the new design.
- A consultant to the airline industry is investigating the fear of flying among adults. Specifically, they wish to know whether there is a difference in the proportion of men versus women who are fearful of flying.

In the above cases each sampled item or individual can be classified as a "success" or a "failure." That is, in the Pontiac Grand Am example each potential buyer is classified as "liking the new design" or "not liking the new design." We then compare the proportion in the under 30 group with the proportion in the over 60 group who indicated they liked the new design. Can we conclude that the differences are due to chance? In this study there is no measurement obtained, only classifying the individuals or objects. Then we assume the nominal scale of measurement.

To conduct the test, we assume each sample is large enough that the normal distribution will serve as a good approximation of the binomial distribution. The test statistic follows the standard normal distribution. We compute the value of z from the following formula:

TWO-SAMPLE TEST OF PROPORTIONS

$$z = \frac{p_1 - p_2}{\sqrt{\dfrac{p_c(1-p_c)}{n_1} + \dfrac{p_c(1-p_c)}{n_2}}} \quad \textbf{[11–3]}$$

Formula (11–3) is formula (11–2) with the respective sample proportions replacing the sample means and $p_c(1 - p_c)$ replacing the two sample standard deviations. In addition:

n_1 is the number of observations in the first sample.
n_2 is the number of observations in the second sample.
p_1 is the proportion in the first sample possessing the trait.
p_2 is the proportion in the second sample possessing the trait.
p_c is the pooled proportion possessing the trait in the combined samples. It is called the pooled estimate of the population proportion and is computed from the following formula.

POOLED PROPORTION

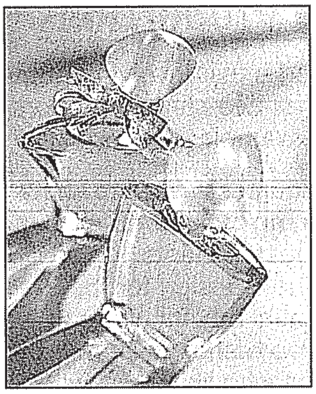
$$p_c = \frac{X_1 + X_2}{n_1 + n_2} \quad \textbf{[11–4]}$$

where:

X_1 is the number possessing the trait in the first sample.
X_2 is the number possessing the trait in the second sample.

The following example will illustrate the two-sample test of proportions.

EXAMPLE



The Manelli Perfume Company recently developed a new fragrance that they plan to market under the name "Heavenly." A number of market studies indicate that Heavenly has very good market potential. The Sales Department at Manelli is particularly interested in whether there is a difference in the proportions of younger and older women who would purchase Heavenly if it were marketed. There are two independent populations, a population consisting of the younger women and a population consisting of the older women. Each sampled woman will be asked to smell Heavenly and indicate whether she likes the fragrance well enough to purchase a bottle.

SOLUTION

We will use the usual five-step hypothesis-testing procedure.

Step 1: State H_0 and H_1. In this case the null hypothesis is: "There is no difference in the proportion of young women and older women who prefer Heavenly." We designate π_1 as the proportion of young women who would purchase Heavenly and π_2 as the proportion of older women who would purchase. The alternate hypothesis is that the two proportions are not equal.

$$H_0: \pi_1 = \pi_2$$
$$H_1: \pi_1 \neq \pi_2$$

Step 2: Select the level of significance. We selected the .05 significance level in this example.

Step 3: Determine the test statistic. If each sample is large enough, the test statistic follows the standard normal distribution. The value of the test statistic can be computed from formula (11–3).

Step 4: Formulate the decision rule. Recall that the alternate hypothesis from step 1 does not state a direction, so this is a two-tailed test. To determine the critical value, we divide the significance level in half and place this amount in each tail of the z distribution. Next, we subtract this amount from the total area to the right of zero. That is .5000 − .0250 = .4750. Finally, we search the body of the z table (Appendix D) for the closest value. It is 1.96. The critical values are −1.96 and +1.96. As before, if the computed z value falls in the region between +1.96 and −1.96, the null hypothesis is not rejected. If that does occur, it is assumed that any difference between the two sample proportions is due to chance variation. This information is summarized in Chart 11–2.

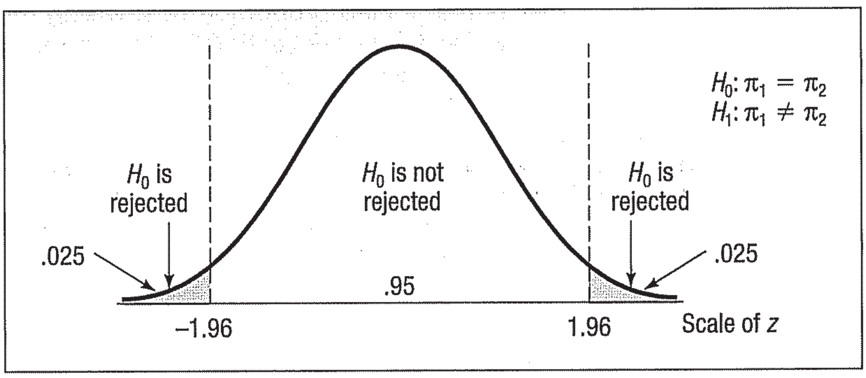


CHART 11–2 Decision Rules for Heavenly Fragrance Test, .05 Significance Level

Step 5: Select a sample and make a decision. A random sample of 100 young women revealed 20 liked the Heavenly fragrance well enough to purchase it. Similarly, a sample of 200 older women revealed 100 liked the fragrance well enough to make a purchase. We let p_1 refer to the young women and p_2 to the older women.

$$p_1 = \frac{X_1}{n_1} = \frac{20}{100} = .20 \qquad p_2 = \frac{X_2}{n_2} = \frac{100}{200} = .50$$

The research question is whether the difference of .30 in the two sample proportions is due to chance or whether there is a difference in the proportion of younger and older women who like the Heavenly fragrance.

Next, we combine or pool the sample proportions. We use formula (11–4).

$$p_c = \frac{X_1 + X_2}{n_1 + n_2} = \frac{20 + 100}{100 + 200} = .40$$

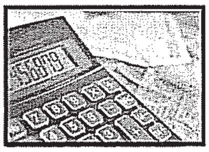
Note that the pooled proportion is closer to .50 than to .20 because more older women than younger women were sampled.

We use formula (11–3) to find the value of the test statistic.

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_c(1-p_c)}{n_1} + \frac{p_c(1-p_c)}{n_2}}} = \frac{.20 - .50}{\sqrt{\frac{.40(1-.40)}{100} + \frac{.40(1-.40)}{200}}} = -5.00$$

The computed value of −5.00 is in the area of rejection; that is, it is to the left of −1.96. Therefore, the null hypothesis is rejected at the .05 significance level. To put it another way, we reject the null hypothesis that the proportion of young women who would purchase Heavenly is equal to the proportion of older women who would purchase Heavenly. It is unlikely that the difference between the two sample proportions is due to chance. To find the p-value we go to Appendix D and look for the likelihood of finding a z value less than −5.00 or greater than 5.00. The largest value of z reported is 3.09, with a corresponding probability of .4990. So the probability of finding a z value greater than 5.00 or less than −5.00 is virtually zero. So we report zero as the p-value. There is very little likelihood the null hypothesis is true. We conclude that there is a difference in the proportion of younger and older women who would purchase Heavenly.

Self-Review 11–2



Of 150 adults who tried a new peach-flavored peppermint patty, 87 rated it excellent. Of 200 children sampled, 123 rated it excellent. Using the .10 level of significance, can we conclude that there is a significant difference in the proportion of adults and the proportion of children who rate the new flavor excellent?

(a) State the null hypothesis and the alternate hypothesis.
(b) What is the probability of a Type I error?
(c) Is this a one-tailed or a two-tailed test?
(d) What is the decision rule?
(e) What is the value of the test statistic?
(f) What is your decision regarding the null hypothesis?
(g) What is the p-value? Explain what it means in terms of this problem.

Exercises

7. The null and alternate hypotheses are:

H_0: $\pi_1 \leq \pi_2$
H_1: $\pi_1 > \pi_2$

A sample of 100 observations from the first population indicated that X_1 is 70. A sample of 150 observations from the second population revealed X_2 to be 90. Use the .05 significance level to test the hypothesis.

a. State the decision rule.
b. Compute the pooled proportion.
c. Compute the value of the test statistic.
d. What is your decision regarding the null hypothesis?

8. The null and alternate hypotheses are:

$$H_0: \pi_1 = \pi_2$$
$$H_1: \pi_1 \neq \pi_2$$

A sample of 200 observations from the first population indicated that X_1 is 170. A sample of 150 observations from the second population revealed X_2 to be 110. Use the .05 significance level to test the hypothesis.

a. State the decision rule.
b. Compute the pooled proportion.
c. Compute the value of the test statistic.
d. What is your decision regarding the null hypothesis?

Note: Use the five-step hypothesis-testing procedure in solving the following exercises.

9. The Damon family owns a large grape vineyard in western New York along Lake Erie. The grapevines must be sprayed at the beginning of the growing season to protect against various insects and diseases. Two new insecticides have just been marketed: Pernod 5 and Action. To test their effectiveness, three long rows were selected and sprayed with Pernod 5, and three others were sprayed with Action. When the grapes ripened, 400 of the vines treated with Pernod 5 were checked for infestation. Likewise, a sample of 400 vines sprayed with Action were checked. The results are:

Insecticide	Number of Vines Checked (sample size)	Number of Infested Vines
Pernod 5	400	24
Action	400	40

At the .05 significance level, can we conclude that there is a difference in the proportion of vines infested using Pernod 5 as opposed to Action?

10. The Roper Organization conducted identical surveys in 1995 and 2005. One question asked women was "Are most men basically kind, gentle, and thoughtful?" The 1995 survey revealed that, of the 3,000 women surveyed, 2,010 said that they were. In 2005, 1,530 of the 3,000 women surveyed thought that men were kind, gentle, and thoughtful. At the .05 level, can we conclude that women think men are less kind, gentle, and thoughtful in 2005 compared with 1995?

11. A nationwide sample of influential Republicans and Democrats was asked as a part of a comprehensive survey whether they favored lowering environmental standards so that high-sulfur coal could be burned in coal-fired power plants. The results were:

	Republicans	Democrats
Number sampled	1,000	800
Number in favor	200	168

At the .02 level of significance, can we conclude that there is a larger proportion of Democrats in favor of lowering the standards?

12. The research department at the home office of New Hampshire Insurance conducts ongoing research on the causes of automobile accidents, the characteristics of the drivers, and so on. A random sample of 400 policies written on single persons revealed 120 had at least one accident in the previous three-year period. Similarly, a sample of 600 policies written on married persons revealed that 150 had been in at least one accident. At the .05 significance level, is there a significant difference in the proportions of single and married persons having an accident during a three-year period?

Comparing Population Means with Small Samples

In an earlier section we assumed that the two population standard deviations were unknown but that we selected random samples containing 30 or more observations each. The large number of observations in our samples allowed us to use z as the test statistic. In this section we consider the case in which the population standard deviations are unknown and the number of observations in at least one of the samples is less than 30. We often refer to this as a "small sample test of means." The requirements for the small sample test are more stringent. The three required assumptions are:

Assumptions for small sample test of means

1. The sampled populations follow the normal distribution.
2. The two samples are from independent populations.
3. The standard deviations of the two populations are equal.

In this case, the t distribution is used to compare two population means. The formula for computing the test statistic t is similar to (11–2), but an additional calculation is necessary. The third assumption above indicates that the population standard deviations must be equal. The two sample standard deviations are pooled to form a single estimate of the unknown population standard deviation. In essence, we compute a weighted mean of the two sample standard deviations and use this as an estimate of the population standard deviation. The weights are the degrees of freedom that each sample provides. Why do we need to pool the standard deviations? In most cases when the samples each have fewer than 30 observations, the population standard deviations are not known. Thus, we calculate s, the sample standard deviation, and substitute it for σ, the population standard deviation. Because we assume that the two populations have equal standard deviations, the best estimate we can make of that value is to combine or pool all the information we have about the value of the population standard deviation.

The following formula is used to pool the sample standard deviations. Notice that two factors are involved: the number of observations in each sample and the sample standard deviations themselves.

POOLED VARIANCE

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \qquad \textbf{[11–5]}$$

where:

s_1^2 is the variance (standard deviation squared) of the first sample.
s_2^2 is the variance of the second sample.

The value of t is computed from the following equation.

TWO-SAMPLE TEST OF MEANS—SMALL SAMPLES

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \textbf{[11–6]}$$

where:

$\bar{X}_1$ is the mean of the first sample.
$\bar{X}_2$ is the mean of the second sample.
n_1 is the number of observations in the first sample.
n_2 is the number of observations in the second sample.
s_p^2 is the pooled estimate of the population variance.

The number of degrees of freedom in the test is the total number of items sampled minus the total number of samples. Because there are two samples, there are $n_1 + n_2 - 2$ degrees of freedom.

An example will help explain the details of the test.

EXAMPLE

Owens Lawn Care, Inc. manufactures and assembles lawnmowers that are shipped to dealers throughout the United States and Canada. Two different procedures have been proposed for mounting the engine on the frame of the lawnmower. The question is: Is there a difference in the mean time to mount the engines on the frames of the lawnmowers? The first procedure was developed by longtime Owens employee Herb Welles (designated as procedure 1), and the other procedure was developed by Owens Vice-President of Engineering William Atkins (designated as procedure 2). To evaluate the two methods, it was decided to conduct a time and motion study. A sample of five employees was timed using the Welles method and six using the Atkins method. The results, in minutes, are shown below. Is there a difference in the mean mounting times? Use the .10 significance level.

Welles (minutes)	Atkins (minutes)
2	3
4	7
9	5
3	8
2	4
	3

SOLUTION

Following the five steps to test a hypothesis, the null hypothesis states that there is no difference in mean mounting times between the two procedures. The alternate hypothesis indicates that there is a difference.

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

The required assumptions are:

1. The observations in the Welles sample are *independent* of the observations in the Atkins sample.
2. The two populations follow the normal distribution.
3. The two populations have equal standard deviations.

Is there a difference between the mean assembly times using the Welles and the Atkins methods? The degrees of freedom are equal to the total number of items sampled minus the number of samples. In this case that is $n_1 + n_2 - 2$. Five assemblers used the Welles method and six the Atkins method. Thus, there are 9 degrees of freedom, found by $5 + 6 - 2$. The critical values of t, from Appendix F for $df = 9$, a two-tailed test, and the .10 significance level, are −1.833 and 1.833. The decision rule is portrayed graphically in Chart 11–3. We do not reject the null hypothesis if the computed value of t falls between −1.833 and 1.833.

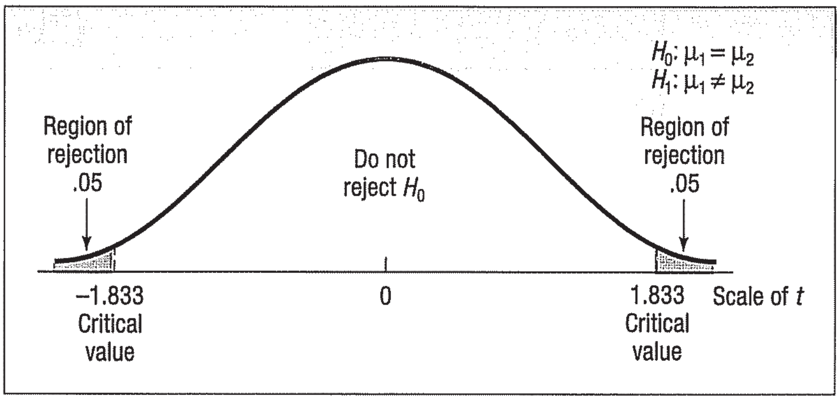


CHART 11–3 Regions of Rejection, Two-Tailed Test, $df = 9$, and .10 Significance Level

We use three steps to compute the value of t.

Step 1: Calculate the Sample Standard Deviations. See the details below.

Welles Method		Atkins Method	
X_1	$(X_1 - \bar{X}_1)^2$	X_2	$(X_2 - \bar{X}_2)^2$
2	$(2-4)^2 = 4$	3	$(3-5)^2 = 4$
4	$(4-4)^2 = 0$	7	$(7-5)^2 = 4$
9	$(9-4)^2 = 25$	5	$(5-5)^2 = 0$
3	$(3-4)^2 = 1$	8	$(8-5)^2 = 9$
2	$(2-4)^2 = 4$	4	$(4-5)^2 = 1$
20	34	3	$(3-5)^2 = 4$
		30	22

$$\bar{X}_1 = \frac{\Sigma X_1}{n_1} = \frac{20}{5} = 4 \qquad \bar{X}_2 = \frac{\Sigma X_2}{n_2} = \frac{30}{6} = 5$$

$$s_1 = \sqrt{\frac{\Sigma(X_1 - \bar{X}_1)^2}{n_1 - 1}} = \sqrt{\frac{34}{5-1}} = 2.9155 \qquad s_2 = \sqrt{\frac{\Sigma(X_2 - \bar{X}_2)^2}{n_2 - 1}} = \sqrt{\frac{22}{6-1}} = 2.0976$$

Step 2: Pool the Sample Variances. We use formula (11–5) to pool the sample variances (standard deviations squared).

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(5-1)(2.9155)^2 + (6-1)(2.0976)^2}{5 + 6 - 2} = 6.2222$$

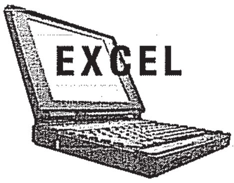
Step 3: Determine the value of t. The mean mounting time for the Welles method is 4.00 minutes, found by $\bar{X}_1 = 20/5$. The mean mounting time for the Atkins method is 5.00 minutes, found by $\bar{X}_2 = 30/6$. We use formula (11–6) to calculate the value of t.

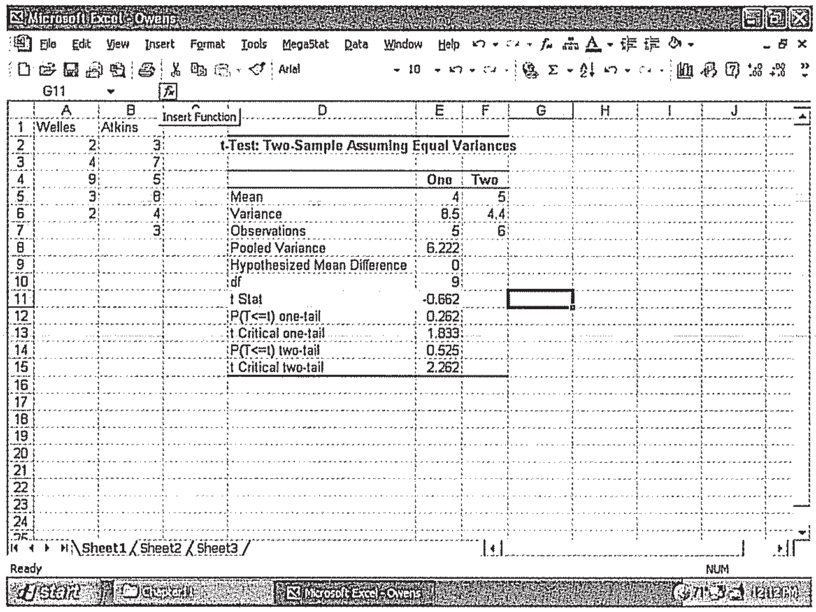
$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{4.00 - 5.00}{\sqrt{6.2222\left(\frac{1}{5} + \frac{1}{6}\right)}} = -0.662$$

The decision is not to reject the null hypothesis, because -0.662 falls in the region between -1.833 and 1.833. We conclude that there is no difference in the mean times to mount the engine on the frame using the two methods.

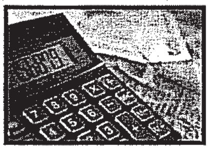
We can also estimate the p-value using Appendix F. Locate the row with 9 degrees of freedom, and use the two-tailed test column. Find the t value, without regard to the sign, which is closest to our computed value of 0.662. It is 1.383, corresponding to a significance level of .20. Thus, even had we used the 20 percent significance level, we would not have rejected the null hypothesis of equal means. We can report that the p-value is greater than .20.

Excel has a procedure called "t-Test: Two Sample Assuming Equal Variances" that will perform the calculations of formulas (11–5) and (11–6) as well as find the sample means and sample variances. The data are input in the first two columns of the Excel spreadsheet. They are labeled "Welles" and "Atkins." The output follows. The value of t, called the "t Stat," is -0.662, and the two-tailed p-value is .525. As we would expect, the p-value is larger than the significance level of .10. The conclusion is not to reject the null hypothesis.





Self-Review 11–3



The production manager at Bellevue Steel, a manufacturer of wheelchairs, wants to compare the number of defective wheelchairs produced on the day shift with the number on the afternoon shift. A sample of the production from 6 day shifts and 8 afternoon shifts revealed the following number of defects.

Day	5	8	7	6	9	7		
Afternoon	8	10	7	11	9	12	14	9

At the .05 significance level, is there a difference in the mean number of defects per shift?

(a) State the null hypothesis and the alternate hypothesis.
(b) What is the decision rule?
(c) What is the value of the test statistic?
(d) What is your decision regarding the null hypothesis?
(e) What is the *p*-value?
(f) Interpret the result.
(g) What are the assumptions necessary for this test?

Exercises

For Exercises 13 and 14: (a) state the decision rule, (b) compute the pooled estimate of the population variance, (c) compute the test statistic, (d) state your decision about the null hypothesis, and (e) estimate the *p*-value.

13. The null and alternate hypotheses are:

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

A random sample of 10 observations from one population revealed a sample mean of 23 and a sample deviation of 4. A random sample of 8 observations from another population revealed a sample mean of 26 and a sample standard deviation of 5. At the .05 significance level, is there a difference between the population means?

14. The null and alternate hypotheses are:

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

A random sample of 15 observations from the first population revealed a sample mean of 350 and a sample standard deviation of 12. A random sample of 17 observations from the second population revealed a sample mean of 342 and a sample standard deviation of 15. At the .10 significance level, is there a difference in the population means?

Note: Use the five-step hypothesis testing procedure for the following exercises.

15. A sample of scores on an examination given in Statistics 201 are:

Men	72	69	98	66	85	76	79	80	77
Women	81	67	90	78	81	80	76		

At the .01 significance level, is the mean grade of the women higher than that of the men?

16. A recent study compared the time spent together by single- and dual-earner couples. According to the records kept by the wives during the study, the mean amount of time spent together watching television among the single-earner couples was 61 minutes per day, with a standard deviation of 15.5 minutes. For the dual-earner couples, the mean number of minutes spent watching television was 48.4 minutes, with a standard deviation of 18.1 minutes. At the .01 significance level, can we conclude that the single-earner couples on average spend more time watching television together? There were 15 single-earner and 12 dual-earner couples studied.

17. Ms. Lisa Monnin is the budget director for Nexus Media, Inc. She would like to compare the daily travel expenses for the sales staff and the audit staff. She collected the following sample information.

Sales ($)	131	135	146	165	136	142	
Audit ($)	130	102	129	143	149	120	139

At the .10 significance level, can she conclude that the mean daily expenses are greater for the sales staff than the audit staff? What is the *p*-value?

18. The Tampa Bay (Florida) Area Chamber of Commerce wanted to know whether the mean weekly salary of nurses was larger than that of school teachers. To investigate, they collected the following information on the amounts earned last week by a sample of school teachers and nurses.

School teachers ($)	845	826	827	875	784	809	802	820	829	830	842	832
Nurses ($)	841	890	821	771	850	859	825	829				

Is it reasonable to conclude that the mean weekly salary of nurses is higher? Use the .01 significance level. What is the *p*-value?

Two-Sample Tests of Hypothesis: Dependent Samples

On page 324, we tested the difference between the means from two independent samples. We compared the mean time required to mount an engine using the Welles method to the time to mount the engine using the Atkins method. The samples were *independent,* meaning that the sample of assembly times using the Welles method was in no way related to the sample of assembly times using the Atkins method.

There are situations, however, in which the samples are not independent. To put it another way, the samples are **dependent** or related. As an example, Nickel Savings and Loan employs two firms, Schadek Appraisals and Bowyer Real Estate, to appraise the value of the real estate properties on which they make loans. It is important that these two firms be similar in their appraisal values. To review the consistency of

the two appraisal firms, Nickel Savings randomly selects 10 homes and has both Schadek Appraisals and Bowyer Real Estate appraise the value of the selected homes. For each home, there will be a pair of appraisal values. That is, for each home there will be an appraised value from both Schadek Appraisals and Bowyer Real Estate. The appraised values depend on, or are related to, the home selected. This is also referred to as a **paired sample.**

For hypothesis testing, we are interested in the distribution of the *differences* in the appraised value of each home. Hence, there is only one sample. To put it more formally, we are investigating whether the mean of the distribution of differences in the appraised values is 0. The sample is made up of the *differences* between the appraised values determined by Schadek Appraisals and the values from Bowyer Real Estate. If the two appraisal firms are reporting similar estimates, then sometimes Schadek Appraisals will be the higher value and sometimes Bowyer Real Estate will have the higher value. However, the mean of the distribution of differences will be 0. On the other hand, if one of the firms consistently reports the larger appraisal values, then the mean of the distribution of the differences will not be 0.

We will use the symbol μ_d to indicate the population mean of the distribution of differences. We assume the distribution of the population of differences follows the normal distribution. The test statistic follows the *t* distribution and we calculate its value from the following formula:

PAIRED *t* TEST

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \qquad \textbf{[11–7]}$$

There are $n - 1$ degrees of freedom and

$\bar{d}$ is the mean of the differences between the paired or related observations.
s_d is the standard deviation of the differences between the paired or related observations.
n is the number of paired observations.

The standard deviation of the differences is computed by the familiar formula for the standard deviation, except d is substituted for X. The formula is:

$$s_d = \sqrt{\frac{\Sigma(d - \bar{d})^2}{n - 1}}$$

The following example illustrates this test.

EXAMPLE

Recall that Nickel Savings and Loan wishes to compare the two companies they use to appraise the value of residential homes. Nickel Savings selected a sample of 10 residential properties and scheduled both firms for an appraisal. The results, reported in $000, are:

Home	Schadek	Bowyer
1	135	128
2	110	105
3	131	119
4	142	140
5	105	98
6	130	123
7	131	127
8	110	115
9	125	122
10	149	145

At the .05 significance level, can we conclude there is a difference in the mean appraised values of the homes?

SOLUTION

The first step is to state the null and the alternate hypotheses. In this case a two-tailed alternative is appropriate because we are interested in determining whether there is a *difference* in the appraised values. We are not interested in showing whether one particular firm appraises property at a higher value than the other. The question is whether the sample differences in the appraised values could have come from a population with a mean of 0. If the population mean of the differences is 0, then we conclude that there is no difference in the appraised values. The null and alternate hypotheses are:

H_0: $\mu_d = 0$
H_1: $\mu_d \neq 0$

There are 10 homes appraised by both firms, so $n = 10$, and $df = n - 1 = 10 - 1 = 9$. We have a two-tailed test, and the significance level is .05. To determine the critical value, go to Appendix F, move across the row with 9 degrees of freedom to the column for a two-tailed test and the .05 significance level. The value at the intersection is 2.262. This value appears in the box in Table 11–2 on page 330. The decision rule is to reject the null hypothesis if the computed value of t is less than −2.262 or greater than 2.262. Here are the computational details.

Home	Schadek	Bowyer	Difference, d	$(d - \bar{d})$	$(d - \bar{d})^2$
1	135	128	7	2.4	5.76
2	110	105	5	0.4	0.16
3	131	119	12	7.4	54.76
4	142	140	2	−2.6	6.76
5	105	98	7	2.4	5.76
6	130	123	7	2.4	5.76
7	131	127	4	−0.6	0.36
8	110	115	−5	−9.6	92.16
9	125	122	3	−1.6	2.56
10	149	145	4	−0.6	0.36
			46	0	174.40

$$\bar{d} = \frac{\Sigma d}{n} = \frac{46}{10} = 4.60$$

$$s_d = \sqrt{\frac{\Sigma(d - \bar{d})^2}{n - 1}} = \sqrt{\frac{174.4}{10 - 1}} = 4.402$$

Using formula (11–7), the value of the test statistic is 3.305, found by

$$t = \frac{\bar{d}}{s_d / \sqrt{n}} = \frac{4.6}{4.402 / \sqrt{10}} = \frac{4.6}{1.3920} = 3.305$$

Because the computed t falls in the rejection region, the null hypothesis is rejected. The population distribution of differences does not have a mean of 0. We conclude that there is a difference in the mean appraised values of the homes. The largest differences are for homes 3 and 8. Perhaps this would be an appropriate place to begin a more detailed review.

To find the p-value, we use Appendix F and the section for a two-tailed test. Move along the row with 9 degrees of freedom and find the values of t that are closest to our calculated value. For a .01 significance level, the value of t is 3.250. The computed value is larger than this value, but smaller than the value of 4.781 corresponding to the

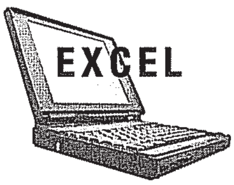
.001 significance level. Hence, the p-value is less than .01. This information is highlighted in Table 11–2.

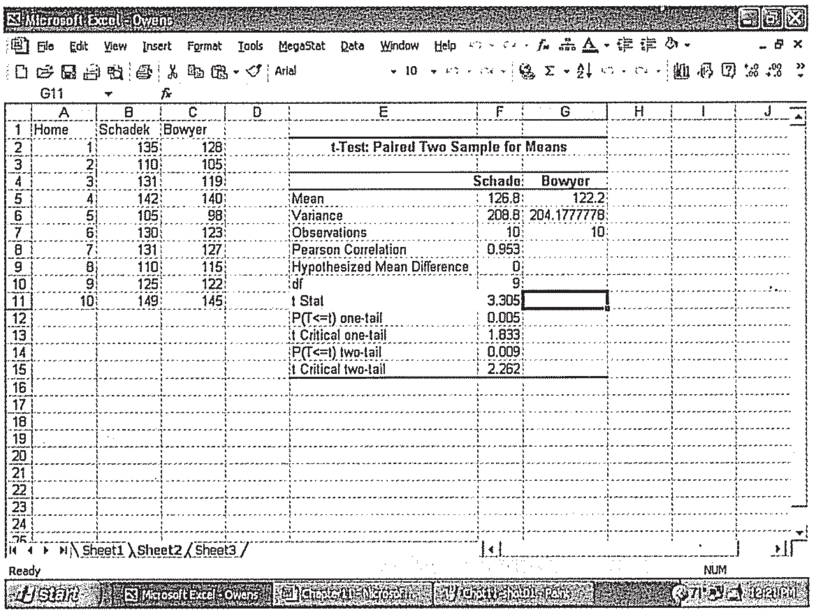
TABLE 11–2 A Portion of the t Distribution from Appendix F

	Confidence Intervals					
	80%	**90%**	**95%**	**98%**	**99%**	**99.9%**
	Level of Significance for One-Tailed Test, α					
df	**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**
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
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

Excel has a procedure called "t-Test: Paired Two-Sample for Means" that will perform the calculations of formula (11–7). The output from this procedure is given below.

The computed value of t is 3.3045, and the two-tailed p-value is .00916. Because the p-value is less than .05, we reject the hypothesis that the mean of the distribution of the differences between the appraised values is zero. In fact, this p-value is less than 1.0 percent. There is a small likelihood that the null hypothesis is true.





Home	Schadek	Bowyer
1	135	128
2	110	105
3	131	119
4	142	140
5	105	98
6	130	123
7	131	127
8	110	115
9	125	122
10	149	145

t-Test: Paired Two Sample for Means

	Schade	Bowyer
Mean	126.8	122.2
Variance	208.8	204.1777778
Observations	10	10
Pearson Correlation	0.953	
Hypothesized Mean Difference	0	
df	9	
t Stat	3.305	
P(T<=t) one-tail	0.005	
t Critical one-tail	1.833	
P(T<=t) two-tail	0.009	
t Critical two-tail	2.262	

Comparing Dependent and Independent Samples

Beginning students are often confused by the difference between tests for independent samples [formula (11–6)] and tests for dependent samples [formula (11–7)]. How do we tell the difference between dependent and independent samples? There are two types of dependent samples: (1) those characterized by a measurement, an intervention of some type, and then another measurement; and (2) a matching or pairing of the observations. To explain further:

1. The first type of dependent sample is characterized by a measurement followed by an intervention of some kind and then another measurement. This could be called a "before" and "after" study. Two examples will help to clarify. Suppose we want to show that, by placing speakers in the production area and playing soothing music, we are able to increase production. We begin by selecting a sample of workers and measuring their output under the current conditions. The speakers are then installed in the production area, and we again measure the output of the same workers. There are two measurements, before placing the speakers in the production area and after. The intervention is playing music in the production area.

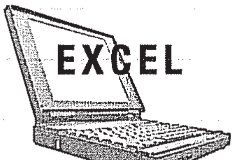
   A second example involves an educational firm that offers courses designed to increase test scores and reading ability. Suppose the firm wants to offer a course that will help high school juniors increase their SAT scores. To begin, each student takes the SAT in the junior year in high school. During the summer between the junior and senior year, they participate in the course that gives them tips on taking tests. Finally, during the fall of their senior year in high school, they retake the SAT. Again, the procedure is characterized by a measurement (taking the SAT as a junior), an intervention (the summer workshops), and another measurement (taking the SAT during their senior year).
2. The second type of dependent sample is characterized by matching or pairing observations. Nickel Savings in the previous example is a dependent sample of this type. They selected a property for appraisal and then had two appraisals on the same property. As a second example, suppose an industrial psychologist wishes to study the intellectual similarities of newly married couples. She selects a sample of newlyweds. Next, she administers a standard intelligence test to both the man and woman to determine the difference in the scores. Notice the matching that occurred: compare the scores that are paired or matched by marriage.

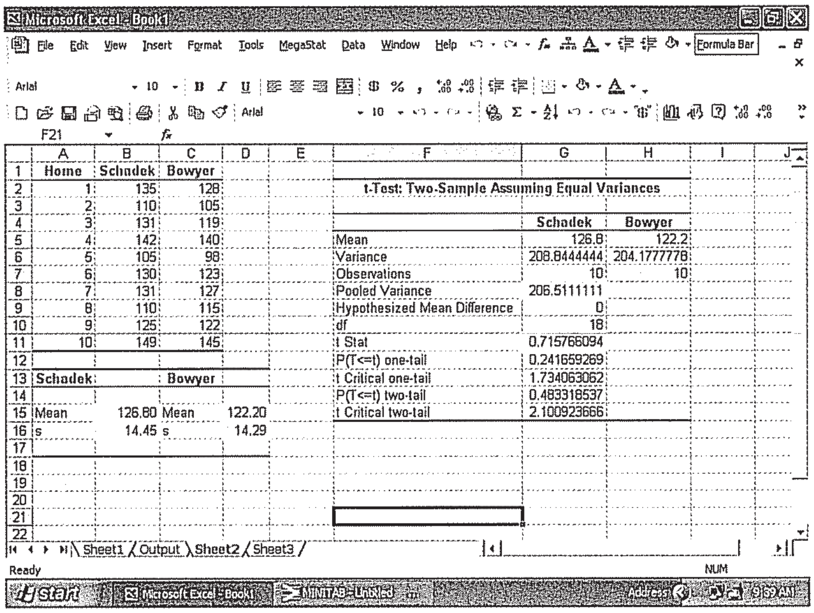
Why do we prefer dependent samples to independent samples? By using dependent samples, we are able to reduce the variation in the sampling distribution. To illustrate, we will use the Nickel Savings and Loan example just completed. Suppose we assume that we have two independent samples of real estate property for appraisal and conduct the following test of hypothesis, using formula (11–6). The null and alternate hypotheses are:

H_0: $\mu_1 = \mu_2$
H_1: $\mu_1 \neq \mu_2$

There are now two independent samples of 10 each. So the number of degrees of freedom is $10 + 10 - 2 = 18$. From Appendix D, for the .05 significance level, H_0 is rejected if t is less than -2.101 or greater than 2.101.

We use the same Excel commands as on page 90 in Chapter 3 to find the mean and the standard deviation of the two independent samples. We use the Excel commands on page 341 of this chapter to find the pooled variance and the value of the "t-Stat." These values are highlighted with shading.




Microsoft Excel - Book1

	A	B	C	D	E	F	G	H
1	Home	Schadek	Bowyer					
2	1	135	128			t-Test: Two-Sample Assuming Equal Variances		
3	2	110	105					
4	3	131	119				Schadek	Bowyer
5	4	142	140			Mean	126.8	122.2
6	5	105	98			Variance	208.8444444	204.1777778
7	6	130	123			Observations	10	10
8	7	131	127			Pooled Variance	206.5111111	
9	8	110	115			Hypothesized Mean Difference	0	
10	9	125	122			df	18	
11	10	149	145			t Stat	0.715766094	
12						P(T<=t) one-tail	0.241659269	
13	Schadek		Bowyer			t Critical one-tail	1.734063062	
14						P(T<=t) two-tail	0.483318537	
15	Mean	126.80	Mean	122.20		t Critical two-tail	2.100923666	
16	s	14.45	s	14.29				

Sheet1 / Output / Sheet2 / Sheet3

The mean of the appraised value of the 10 properties by Schadek is \$126,800, and the standard deviation is \$14,450. For Bowyer Real Estate the mean appraised value is \$122,200, and the standard deviation is \$14,290. To make the calculations easier, we use \$000 instead of \$. The value of the pooled estimate of the variance from formula (11–5) is

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(10 - 1)(14.45^2) + (10 - 1)(14.29)^2}{10 + 10 - 2} = 206.50$$

From formula (11–6), t is 0.716.

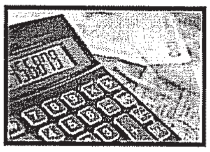
$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{126.8 - 122.2}{\sqrt{206.50\left(\frac{1}{10} + \frac{1}{10}\right)}} = \frac{4.6}{6.4265} = 0.716$$

The computed t (0.716) is less than 2.101, so the null hypothesis is not rejected. We cannot show that there is a difference in the mean appraisal value. That is not the same conclusion that we got before! Why does this happen? The numerator is the same in the paired observations test (4.6). However, the denominator is smaller. In the paired test the denominator is 1.3920 (see the calculations on page 329). In the case of the independent samples, the denominator is 6.4265. There is more variation or uncertainty. This accounts for the difference in the t values and the difference in the statistical decisions. The denominator measures the standard error of the statistic. When the samples are *not* paired, two kinds of variation are present: differences between the two appraisal firms and the difference in the value of the real estate. Properties numbered 4 and 10 have relatively high values, whereas number 5 is relatively low. These data show how different the values of the property are, but we are really interested in the difference between the two appraisal firms.

The trick is to pair the values to reduce the variation among the properties. The paired test uses only the difference between the two appraisal firms for the same property. Thus, the paired or dependent statistic focuses on the variation between Schadek Appraisals and Bowyer Real Estate. Thus, its standard error is always smaller. That, in turn, leads to a larger test statistic and a greater chance of rejecting the null hypothesis. So whenever possible you should pair the data.

There is a bit of bad news here. In the paired observations test, the degrees of freedom are half of what they are if the samples are not paired. For the real estate example, the degrees of freedom drop from 18 to 9 when the observations are paired. However, in most cases, this is a small price to pay for a better test.

Self-Review 11–4



Advertisements by Rivertown Fitness Center claim that completing their course will result in losing weight. A random sample of eight recent participants showed the following weights before and after completing the course. At the .01 significance level, can we conclude the students lost weight?

Name	Before	After
Hunter	155	154
Cashman	228	207
Mervine	141	147
Massa	162	157
Creola	211	196
Peterson	164	150
Redding	184	170
Poust	172	165

(a) State the null hypothesis and the alternate hypothesis.
(b) What is the critical value of t?
(c) What is the computed value of t?
(d) Interpret the result. What is the p-value?
(e) What assumption needs to be made about the distribution of the differences?

Exercises

19. The null and alternate hypotheses are:

H_0: $\mu_d \leq 0$
H_1: $\mu_d > 0$

The following sample information shows the number of defective units produced on the day shift and the afternoon shift for a sample of four days last month.

	Day			
	1	2	3	4
Day shift	10	12	15	19
Afternoon shift	8	9	12	15

At the .05 significance level, can we conclude there are more defects produced on the afternoon shift?

20. The null and alternate hypotheses are:

H_0: $\mu_d = 0$
H_1: $\mu_d \neq 0$

The following paired observations show the number of traffic citations given for speeding by Officer Dhondt and Officer Meredith of the South Carolina Highway Patrol for the last five months.

	Day				
	May	June	July	August	September
Officer Dhondt	30	22	25	19	26
Officer Meredith	26	19	20	15	19

At the .05 significance level, is there a difference in the mean number of citations given by the two officers?

Note: Use the five-step hypothesis testing procedure to solve the following exercises.

21. The management of Discount Furniture, a chain of discount furniture stores in the Northeast, designed an incentive plan for salespeople. To evaluate this innovative plan, 12 salespeople were selected at random, and their weekly incomes before and after the plan were recorded.

Salesperson	Before	After
Sid Mahone	$320	$340
Carol Quick	290	285
Tom Jackson	421	475
Andy Jones	510	510
Jean Sloan	210	210
Jack Walker	402	500
Peg Mancuso	625	631
Anita Loma	560	560
John Cuso	360	365
Carl Utz	431	431
A. S. Kushner	506	525
Fern Lawton	505	619

Was there a significant increase in the typical salesperson's weekly income due to the innovative incentive plan? Use the .05 significance level. Estimate the *p*-value, and interpret it.

22. The federal government recently granted funds for a special program designed to reduce crime in high-crime areas. A study of the results of the program in eight high-crime areas of Miami, Florida, yielded the following results.

	Number of Crimes by Area							
	A	B	C	D	E	F	G	H
Before	14	7	4	5	17	12	8	9
After	2	7	3	6	8	13	3	5

Has there been a decrease in the number of crimes since the inauguration of the program? Use the .01 significance level. Estimate the *p*-value.

Chapter Outline

I. In comparing two population means we wish to know whether they could be equal.

A. We are investigating whether the distribution of the difference between the means could have a mean of 0.

B. The test statistic is the standard normal *z* if the samples both contain at least 30 observations and the population standard deviations are unknown.

1. No assumption about the shape of either population is required.
2. The samples are from independent populations.
3. The formula to compute the value of *z* is

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad \textbf{[11-2]}$$

II. We can also test whether two samples come from populations with an equal proportion of successes.

A. The two sample proportions are pooled using the following formula:

$$p_c = \frac{X_1 + X_2}{n_1 + n_2} \qquad \textbf{[11-4]}$$

B. We compute the value of the test statistic from the following formula:

$$z = \frac{p_1 - p_2}{\sqrt{\dfrac{p_c(1-p_c)}{n_1} + \dfrac{p_c(1-p_c)}{n_2}}} \qquad \textbf{[11-3]}$$

III. The test statistic to compare two means is the t distribution if one or both of the samples contain fewer than 30 observations.

A. Both populations must follow the normal distribution.

B. The populations must have equal standard deviations.

C. The samples are independent.

D. Finding the value of t requires two steps.

1. The first step is to pool the standard deviations according to the following formula:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \qquad \textbf{[11-5]}$$

2. The value of t is computed from the following formula:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \qquad \textbf{[11-6]}$$

IV. For dependent samples, we assume the distribution of the paired differences between the populations has a mean of 0.

A. We first compute the mean and the standard deviation of the sample differences.

B. The value of the test statistic is computed from the following formula:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \qquad \textbf{[11-7]}$$

Pronunciation Key

SYMBOL	MEANING	PRONUNCIATION
p_c	Pooled proportion	*p sub c*
s_p^2	Pooled sample variance	*s sub p squared*
$\bar{X}_1$	Mean of the first sample	*X bar sub 1*
$\bar{X}_2$	Mean of the second sample	*X bar sub 2*
$\bar{d}$	Mean of the difference between dependent observations	*d bar*
s_d	Standard deviation of the difference between dependent observations	*s sub d*

Chapter Exercises

23. A recent study focused on the number of times men and women who live alone buy take-out dinners in a month. The information is summarized below.

Statistic	Men	Women
Mean	24.51	22.69
Standard deviation	4.48	3.86
Sample size	35	40

At the .01 significance level, is there a difference in the mean number of times men and women order takeout dinners in a month? What is the p-value?

24. Clark Heter is an industrial engineer at Lyons Products. He would like to determine whether there are more units produced on the afternoon shift than on the day shift. A sample of 54 day-shift workers showed that the mean number of units produced was 345, with a standard deviation of 21. A sample of 60 afternoon-shift workers showed that the mean number of units produced was 351, with a standard deviation of 28 units. At the .05 significance level, is the number of units produced on the afternoon shift larger?

25. Fry Brothers Heating and Air Conditioning, Inc. employs Larry Clark and George Murnen to make service calls to repair furnaces and air conditioning units in homes. Tom Fry, the owner, would like to know whether there is a difference in the mean number of service calls they make per day. A random sample of 40 days last year showed that Larry Clark made an average of 4.77 calls per day, with a standard deviation of 1.05 calls per day. For a sample of 50 days George Murnen made an average of 5.02 calls per day, with a standard deviation of 1.23 calls per day. At the .05 significance level, is there a difference in the mean number of calls per day between the two employees? What is the *p*-value?

26. A coffee manufacturer is interested in whether the mean daily consumption of regular coffee drinkers is less than that of decaffeinated-coffee drinkers. A random sample of 50 regular-coffee drinkers showed a mean of 4.35 cups per day, with a standard deviation of 1.20 cups per day. A sample of 40 decaffeinated-coffee drinkers showed a mean of 5.84 cups per day, with a standard deviation of 1.36 cups per day. Use the .01 significance level. Compute the *p*-value.

27. A cell phone company offers two plans to it subscribers. At the time new subscribers sign up, they are asked to provide some demographic information. The mean yearly income for a sample of 40 subscribers to Plan A is $57,000 with a standard deviation of $9,200. This distribution is positively skewed; the actual coefficient of skewness is 2.11. For a sample of 30 subscribers to Plan B the mean income is $61,000 with a standard deviation of $7,100. The distribution of Plan B subscribers is also positively skewed, but not as severely. The coefficient of skewness is 1.54. At the .05 significance level, is it reasonable to conclude the mean income of those selecting Plan B is larger? What is the *p*-value? Do the coefficients of skewness affect the results of the hypothesis test? Why?

28. A computer manufacturer offers a help line that purchasers can call for help 24 hours a day 7 days a week. Clearing these calls for help in a timely fashion is important to the company's image. After telling the caller that resolution of the problem is important the caller is asked whether the issue is "software" or "hardware" related. The mean time it takes a technician to resolve a software issue is 18 minutes with a standard deviation of 4.2 minutes. This information was obtained from a sample of 35 monitored calls. For a study of 45 hardware issues, the mean time for the technician to resolve the problem was 15.5 minutes with a standard deviation of 3.9 minutes. This information was also obtained from monitored calls. At the .05 significance level is it reasonable to conclude that it takes longer to resolve software issues? What is the *p*-value?

29. The manufacturer of Advil, a common headache remedy, recently developed a new formulation of the drug that is claimed to be more effective. To evaluate the new drug, a sample of 200 current users is asked to try it. After a one-month trial, 180 indicated the new drug was more effective in relieving a headache. At the same time a sample of 300 current Advil users is given the current drug but told it is the new formulation. From this group, 261 said it was an improvement. At the .05 significance level can we conclude that the new drug is more effective?

30. Each month the National Association of Purchasing Managers publishes the NAPM index. One of the questions asked on the survey to purchasing agents is: Do you think the economy is expanding? Last month, of the 300 responses 160 answered yes to the question. This month, 170 of the 290 responses indicated they felt the economy was expanding. At the .05 significance level, can we conclude that a larger proportion of the agents believe the economy is expanding this month?

31. As part of a recent survey among dual-wage-earner couples, an industrial psychologist found that 990 men out of the 1,500 surveyed believed the division of household duties was fair. A sample of 1,600 women found 970 believed the division of household duties was fair. At the .01 significance level, is it reasonable to conclude that the proportion of men who believe the division of household duties is fair is larger? What is the *p*-value?

32. There are two major Internet providers in the Colorado Springs, Colorado, area, one called HTC and the other Mountain Communications. We want to investigate whether there is a difference in the proportion of times a customer is able to access the Internet. During a one-week period, 500 calls were placed at random times throughout the day and night to HTC. A connection was made to the Internet on 450 occasions. A similar one-week study with Mountain Communications showed the Internet to be available on 352 of 400 trials. At the .01 significance level, is there a difference in the percent of time that access to the Internet is successful?

33. The owner of Bun 'N' Run Hamburger wishes to compare the sales per day at two locations. The mean number sold for 10 randomly selected days at the Northside site was

83.55, and the standard deviation was 10.50. For a random sample of 12 days at the Southside location, the mean number sold was 78.80 and the standard deviation was 14.25. At the .05 significance level, is there a difference in the mean number of hamburgers sold at the two locations? What is the *p*-value?

34. The Engineering Department at Sims Software, Inc., recently developed two chemical solutions designed to increase the usable life of computer disks. A sample of disks treated with the first solution lasted 86, 78, 66, 83, 84, 81, 84, 109, 65, and 102 hours. Those treated with the second solution lasted 91, 71, 75, 76, 87, 79, 73, 76, 79, 78, 87, 90, 76, and 72 hours. At the .10 significance level, can we conclude that there is a difference in the length of time the two types of treatment lasted?

35. The Willow Run Outlet Mall has two Gap Outlet Stores, one located on Peach Street and the other on Plum Street. The two stores are laid out differently, but both store managers claim their layout maximizes the amounts customers will purchase on impulse. A sample of 10 customers at the Peach Street store revealed they spent the following amounts more than planned: $17.58, $19.73, $12.61, $17.79, $16.22, $15.82, $15.40, $15.86, $11.82, and $15.85. A sample of 14 customers at the Plum Street store revealed they spent the following amounts more than they planned: $18.19, $20.22, $17.38, $17.96, $23.92, $15.87, $16.47, $15.96, $16.79, $16.74, $21.40, $20.57, $19.79, and $14.83. At the .01 significance level, is there a difference in the mean amounts purchased on impulse at the two stores?

36. The Grand Strand Family Medical Center is specifically set up to treat minor medical emergencies for visitors to the Myrtle Beach area. There are two facilities, one in the Little River Area and the other in Murrells Inlet. The Quality Assurance Department wishes to compare the mean waiting time for patients at the two locations. Samples of the waiting times, reported in minutes, follow:

Location	Waiting Time										
Little River	31.73	28.77	29.53	22.08	29.47	18.60	32.94	25.18	29.82	26.49	
Murrells Inlet	22.93	23.92	26.92	27.20	26.44	25.62	30.61	29.44	23.09	23.10	26.69	22.31

At the .05 significance level, is there a difference in the mean waiting time?

37. The Commercial Bank and Trust Company is studying the use of its automatic teller machines (ATMs). Of particular interest is whether young adults (under 25 years) use the machines more than senior citizens. To investigate further, samples of customers under 25 years of age and customers over 60 years of age were selected. The number of ATM transactions last month was determined for each selected individual, and the results are shown below. At the .01 significance level, can bank management conclude that younger customers use the ATMs more?

Under 25	10	10	11	15	7	11	10	9		
Over 60	4	8	7	7	4	5	1	7	4	10	5

38. Two boats, the *Prada* (Italy) and the *Oracle* (U.S.A.), are competing for a spot in the upcoming *America's Cup* race. They race over a part of the course several times. Below are the sample times in minutes. At the .05 significance level, can we conclude that there is a difference in their mean times?

Boat	Times (minutes)										
Prada (Italy)	12.9	12.5	11.0	13.3	11.2	11.4	11.6	12.3	14.2	11.3	
Oracle (U.S.A.)	14.1	14.1	14.2	17.4	15.8	16.7	16.1	13.3	13.4	13.6	10.8	19.0

39. The manufacturer of an MP3 player wanted to know whether a 10 percent reduction in price is enough to increase the sales of their product. To investigate, the owner randomly selected eight outlets and sold the MP3 player at the reduced price. At seven randomly selected outlets, the MP3 player was sold at the regular price. Reported below is the number of units sold last month at the sampled outlets. At the .01 significance level, can the manufacturer conclude that the price reduction resulted in an increase in sales?

Regular price	138	121	88	115	141	125	96	
Reduced price	128	134	152	135	114	106	112	120

40. A number of minor automobile accidents occur at various high-risk intersections in Teton County despite traffic lights. The Traffic Department claims that a modification in the type of light will reduce these accidents. The county commissioners have agreed to a proposed experiment. Eight intersections were chosen at random, and the lights at those intersections were modified. The numbers of minor accidents during a six-month period before and after the modifications were:

	Number of Accidents							
	A	B	C	D	E	F	G	H
Before modification	5	7	6	4	8	9	8	10
After modification	3	7	7	0	4	6	8	2

At the .01 significance level is it reasonable to conclude that the modification reduced the number of traffic accidents?

41. Lester Hollar is Vice President for Human Resources for a large manufacturing company. In recent years he has noticed an increase in absenteeism that he thinks is related to the general health of the employees. Four years ago, in an attempt to improve the situation, he began a fitness program in which employees exercise during their lunch hour. To evaluate the program, he selected a random sample of eight participants and found the number of days each was absent in the six months before the exercise program began and in the last six months. Below are the results. At the .05 significance level, can he conclude that the number of absences has declined? Estimate the *p*-value.

Employee	Before	After
1	6	5
2	6	2
3	7	1
4	7	3
5	4	3
6	3	6
7	5	3
8	6	7

42. The president of the American Insurance Institute wants to compare the yearly costs of auto insurance offered by two leading companies. He selects a sample of 15 families, some with only a single insured driver, others with several teenage drivers, and pays each family a stipend to contact the two companies and ask for a price quote. To make the data comparable, certain features, such as the amount deductible and limits of liability, are standardized. The sample information is reported below. At the .10 significance level, can we conclude that there is a difference in the amounts quoted?

Family	Progressive Car Insurance	GEICO Mutual Insurance
Becker	$2,090	$1,610
Berry	1,683	1,247
Wong	1,402	2,327
Debuck	1,830	1,367
DuBrul	930	1,461
Eckroate	697	1,789
German	1,741	1,621
Ruska	1,129	1,914
King	1,018	1,956
Kucic	1,881	1,772
Meredith	1,571	1,375
Obeid	874	1,527
Orlando	1,579	1,767
Phillips	1,577	1,636
Suzuki	860	1,188

43. Fairfield Homes is developing two parcels of land near Pigeon Fork, Tennessee. In order to test different advertising approaches, they use different media to reach potential buyers. The mean annual family income for 75 people making inquiries at the first development is $150,000, with a standard deviation of $40,000. A corresponding sample of 120 people at the second development had a mean of $180,000, with a standard deviation of $30,000. At the .05 significance level, can Fairfield conclude that the population means are different?

44. The following data resulted from a taste test of two different chocolate bars. The first number is a rating of the taste, which could range from 0 to 5, with a 5 indicating the person liked the taste. The second number indicates whether a "secret ingredient" was present. If the ingredient was present a code of "1" was used and a "0" otherwise. At the .05 significance level, does this data show a difference in the taste ratings?

Rating	"With/Without"	Rating	"With/Without"
3	1	1	1
1	1	4	0
0	0	4	0
2	1	2	1
3	1	3	0
1	1	4	0

45. An investigation of the effectiveness of an antibacterial soap in reducing operating room contamination resulted in the accompanying table. The new soap was tested in a sample of eight operating rooms in the greater Seattle area during the last year.

	Operating Room							
	A	B	C	D	E	F	G	H
Before	6.6	6.5	9.0	10.3	11.2	8.1	6.3	11.6
After	6.8	2.4	7.4	8.5	8.1	6.1	3.4	2.0

At the .05 significance level, can we conclude the contamination measurements are lower after use of the new soap?

46. The following data on annual rates of return were collected from five stocks listed on the New York Stock Exchange ("the big board") and five stocks listed on NASDAQ. At the .10 significance level, can we conclude that the annual rates of return are higher on the big board?

NYSE	NASDAQ
17.16	15.80
17.08	16.28
15.51	16.21
8.43	17.97
25.15	7.77

47. The city of Laguna Beach operates two public parking lots. The one on Ocean Drive can accommodate up to 125 cars and the one on Rio Rancho can accommodate up to 130 cars. City planners are considering both increasing the size of the lots and changing the fee structure. To begin, the Planning Office would like some information on the number of cars in the lots at various times of the day. A junior planner officer is assigned the task of visiting the two lots at random times of the day and evening and counting the number of cars in the lots. The study lasted over a period of one month. Below is the number of cars in the lots for 25 visits of the Ocean Drive lot and 28 visits of the Rio Rancho lot.

Ocean												
89	115	93	79	113	77	51	75	118	105	106	91	54
63	121	53	81	115	67	53	69	95	121	88	64	

Rio Rancho												
128	110	81	126	82	114	93	40	94	45	84	71	74
92	66	69	100	114	113	107	62	77	80	107	90	129
105	124											

Is it reasonable to conclude that there is a difference in the mean number of cars in the two lots? Use the .05 significance level.

48. The amount spent on housing is an important component of the cost of living. The total costs of housing for homeowners might include mortgage payments, property taxes, and utility costs (water, heat, electricity). An economist selected a sample of 20 homeowners in New England and then calculated these total housing costs as a percent of monthly income, five years ago and now. The information is reported below. Is it reasonable to conclude the percent is less now than five years ago?

Homeowner	Five Years Ago	Now	Homeowner	Five Years Ago	Now
1	17	10	11	35	32
2	20	39	12	16	32
3	29	37	13	23	21
4	43	27	14	33	12
5	36	12	15	44	40
6	43	41	16	44	42
7	45	24	17	28	22
8	19	26	18	29	19
9	49	28	19	39	35
10	49	26	20	22	12

exercises.com

WWW

49. Listed below are several prominent companies and their stock prices in June 2004. Go to the Web and look up today's price. There are many sources to find stock prices, such as Yahoo and CNNFI. The Yahoo address is http://www.finance.yahoo.com. Enter the symbol identification to find the current price. At the .05 significance level, can we conclude that the prices have changed?

Company	Symbol	Price
Coca-Cola	KO	$51.89
Walt Disney	DIS	24.06
Eastman Kodak	EK	25.31
Ford Motor Company	F	14.81
General Motors	GM	45.60
Goodyear Tire	GT	8.49
IBM	IBM	87.59
McDonald's	MCD	26.58
McGraw-Hill Publishing	MHP	77.18
Oracle	ORCL	11.01
Johnson and Johnson	JNJ	56.72
General Electric	GE	31.01
Home Depot	HD	35.60

50. The *USA Today* (http://www.usatoday.com/sports/baseball/front.htm) and Major League Baseball's (http://www.majorleaguebaseball.com) websites regularly report information on individual player salaries in the American League and the National League. Go to one of these sites and find the individual salaries for your favorite team in each league. Compute the mean and the standard deviation. Is it reasonable to conclude that there is a difference in the salaries of the two teams?

Dataset Exercises

51. Refer to the Real Estate data, which reports information on the homes sold in Denver, Colorado, last year.

a. At the .05 significance level, can we conclude that there is a difference in the mean selling price of homes with a pool and homes without a pool?

b. At the .05 significance level, can we conclude that there is a difference in the mean selling price of homes with an attached garage and homes without a garage?

c. At the .05 significance level, can we conclude that there is a difference in the mean selling price of homes in Township 1 and Township 2?

d. Find the median selling price of the homes. Divide the homes into two groups, those that sold for more than (or equal to) the median price and those that sold for less. Is there a difference in the proportion of homes with a pool for those that sold at or above the median price versus those that sold for less than the median price? Use the .05 significance level.

52. Refer to the Baseball 2003 data, which reports information on the 30 Major League Baseball teams for the 2003 season.

a. At the .05 significance level, can we conclude that there is a difference in the mean salary of teams in the American League versus teams in the National League?

b. At the .05 significance level, can we conclude that there is a difference in the mean home attendance of teams in the American League versus teams in the National League?

c. At the .05 significance level, can we conclude that there is a difference in the mean number of wins for teams that have artificial turf home fields versus teams that have grass home fields?

d. At the .05 significance level, can we conclude that there is a difference in the mean number of home runs for teams that have artificial turf home fields versus teams that have grass home fields?

53. Refer to the Wage data, which reports information on annual wages for a sample of 100 workers. Also included are variables relating to industry, years of education, and gender for each worker.

a. Conduct a test of hypothesis to determine if there is a difference in the mean annual wages of southern residents versus nonsouthern residents.

b. Conduct a test of hypothesis to determine if there is a difference in the mean annual wages of white and nonwhite wage earners.

c. Conduct a test of hypothesis to determine if there is a difference in the mean annual wages of Hispanic and non-Hispanic wage earners.

d. Conduct a test of hypothesis to determine if there is a difference in the mean annual wages of female and male wage earners.

e. Conduct a test of hypothesis to determine if there is a difference in the mean annual wages of married and nonmarried wage earners.

54. Refer to the CIA data, which reports demographic and economic information on 46 countries. Conduct a test of hypothesis to determine whether the mean percent of the population over 65 years of age in G20 countries is different from those that are not G20 members.

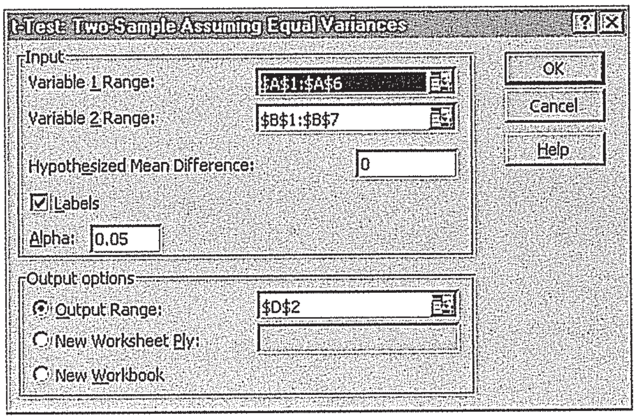
Software Commands

1. The Excel commands for the two-sample t-test on page 326 are:

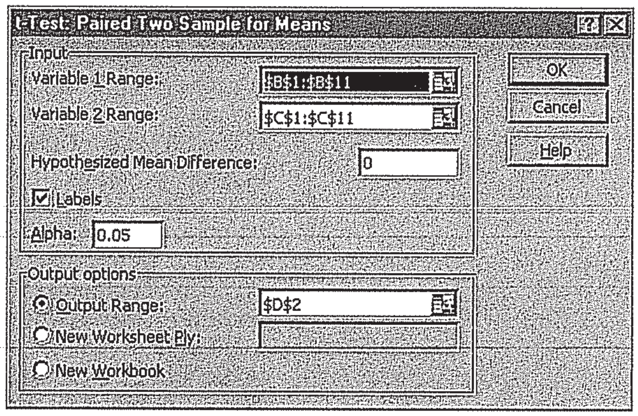
a. Enter the data into columns A and B (or any other columns) in the spreadsheet. Use the first row of each column to enter the variable name.

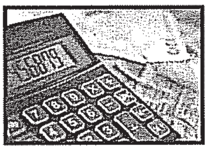
b. From the menu bar select **Tools** and **Data Analysis.** Select **t-Test: Two-Sample Assuming Equal Variances,** then click **OK.**

c. In the dialog box indicate that the range of **Variable 1** is from *A1* to *A6* and **Variable 2** from *B1* to *B7*, the **Hypothesized Mean Difference** is 0, the **Labels** are in the first row, **Alpha** is *0.05*, and the **Output Range** is *D2*. Click **OK.**



2. The Excel commands for the paired *t*-test on page 330 are:
   a. Enter the data into columns B and C (or any other two columns) in the spreadsheet, with the variable names in the first row.
   b. From the menu bar select **Tools** and **Data Analysis.** Select **t-Test: Paired Two Sample for Means,** then click **OK.**
   c. In the dialog box indicate that the range of **Variable 1** is from *B1* to *B11* and **Variable 2** from *C1* to *C11*, the **Hypothesized Mean Difference** is *0*, the **Labels** are in the first row, **Alpha** is *.05*, and the **Output Range** is *D2*. Click **OK.**





Chapter 11 Answers to Self-Review

11–1 **a.** H_0: $\mu_W \leq \mu_M$
H_1: $\mu_W > \mu_M$
The subscript W refers to the women and M to the men.

b. Reject H_0 if $z > 1.65$

c. $$z = \frac{\$1{,}500 - \$1{,}400}{\sqrt{\frac{(\$250)^2}{50} + \frac{(\$200)^2}{40}}} = 2.11$$

d. Reject the null hypothesis

e. p-value = .5000 − .4826 = .0174

f. The mean amount sold per day is larger for women.

11–2 **a.** H_0: $\pi_1 = \pi_2$
H_1: $\pi_1 \neq \pi_2$

b. .10

c. Two-tailed

d. Reject H_0 if z is less than −1.65 or greater than 1.65.

e. $$p_c = \frac{87 + 123}{150 + 200} = \frac{210}{350} = .60$$

$$p_1 = \frac{87}{150} = .58 \quad p_2 = \frac{123}{200} = .615$$

$$z = \frac{.58 - .615}{\sqrt{\frac{.60(.40)}{150} + \frac{.60(.40)}{200}}} = -0.66$$

f. Do not reject H_0.

g. p-value = 2(.5000 − .2454) = .5092
There is no difference in the proportion of adults and children that liked the proposed flavor.

11–3 **a.** H_0: $\mu_d = \mu_a$
H_1: $\mu_d \neq \mu_a$

b. $df = 6 + 8 - 2 = 12$
Reject H_0 if t is less than −2.179 or t is greater than 2.179.

c. $$\bar{X}_1 = \frac{42}{6} = 7.00 \quad s_1 = \sqrt{\frac{10}{6-1}} = 1.4142$$

$$\bar{X}_2 = \frac{80}{8} = 10.00 \quad s_2 = \sqrt{\frac{36}{8-1}} = 2.2678$$

$$s_p^2 = \frac{(6-1)(1.4142)^2 + (8-1)(2.2678)^2}{6+8-2}$$

$$= 3.8333$$

$$t = \frac{7.00 - 10.00}{\sqrt{3.8333\left(\frac{1}{6} + \frac{1}{8}\right)}} = -2.837$$

d. Reject H_0 because −2.837 is less than the critical value.

e. The p-value is less than .02.

f. The mean number of defects is not the same on the two shifts.

g. Independent populations, populations follow the normal distribution, populations have equal standard deviations.

11–4 **a.** H_0: $\mu_d \leq 0$, H_1: $\mu_d > 0$.

b. Reject H_0 if $t > 2.998$.

c.

Name	Before	After	d	$(d - \bar{d})$	$(d - \bar{d})^2$
Hunter	155	154	1	−7.875	62.0156
Cashman	228	207	21	12.125	147.0156
Mervine	141	147	−6	−14.875	221.2656
Massa	162	157	5	−3.875	15.0156
Creola	211	196	15	6.125	37.5156
Peterson	164	150	14	5.125	26.2656
Redding	184	170	14	5.125	26.2656
Poust	172	165	7	−1.875	3.5156
			71		538.8750

$$\bar{d} = \frac{71}{8} = 8.875$$

$$s_d = \sqrt{\frac{538.875}{8-1}} = 8.774$$

$$t = \frac{8.875}{8.774/\sqrt{8}} = 2.861$$

d. Do not reject H_0. We cannot conclude that the students lost weight. The p-value is less than .025 but larger than .01.

e. The distribution of the differences must follow a normal distribution.