9 Hypothesis Tests

CHAPTER CONTENTS

LEARNING OBJECTIVES

After studying this chapter and doing the exercises, you should be able to:

1 Set up appropriate null and alternative hypotheses for testing research hypotheses, and for testing the validity of a claim.

2 Give an account of the logical steps involved in a statistical hypothesis test.

3 Explain the meaning of the terms null hypothesis, alternative hypothesis, Type I error, Type II error, level of significance, *p*-value and critical value in statistical hypothesis testing.

4 Construct and interpret hypothesis tests for a population mean:

4.1 When the population standard deviation is known.

4.2 When the population standard deviation is unknown.

5 Construct and interpret hypothesis tests for a population proportion.

6 Explain the relationship between the construction of hypothesis tests and confidence intervals.

7 Calculate the probability of a Type II error for a hypothesis test of a population mean when the population standard deviation is known.

8 Estimate the sample size required for a hypothesis test of a population mean when the population standard deviation is known.

In Chapters 7 and 8 we showed how a sample could be used to construct point and interval estimates of population parameters. In this chapter we continue the discussion of statistical inference by showing how hypothesis testing can be used to determine whether a statement about the value of a population parameter should or should not be rejected.

In hypothesis testing we begin by making a tentative assumption about a population parameter. This tentative assumption is called the **null hypothesis** and is denoted by H_0. We then define another hypothesis, called the **alternative hypothesis**, which is the opposite of what is stated in the null hypothesis. We denote the alternative hypothesis by H_1. The hypothesis testing procedure uses data from a sample to assess the two competing statements indicated by H_0 and H_1.

This chapter shows how hypothesis tests can be conducted about a population mean and a population proportion. We begin by providing examples of approaches to formulating null and alternative hypotheses.

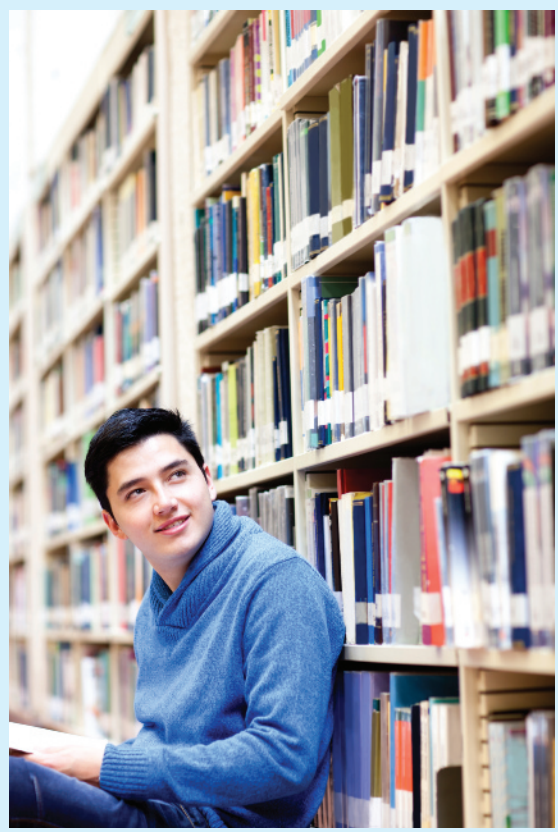
STATISTICS IN PRACTICE

Hypothesis testing in business research

The *British Journal of Management* (*BJM*) is one of the most highly rated academic journals globally in the field of management. It is published quarterly, and contains articles giving accounts of the latest research in the field. Any particular issue typically shows an authorship with wide geographic spread. For example, the June 2011 issue contained nine articles written by researchers from Germany, Switzerland, Italy, UK, Lebanon, Australia and Canada. The research topics addressed included links between work/home culture and employee well-being, attitudes towards corporate social responsibility, age discrimination in recruitment and assessment of research quality in UK universities.

Of the nine articles in the June 2011 *BJM* issue, seven reported research based on quantitative methodology. The other two featured qualitative research. All of the seven quantitative papers featured both descriptive statistics and extensive use of inferential statistics. The main tool in regard to the inferential results reported was the *statistical hypothesis test*. Between them, the seven articles reported a total of over 400 statistical hypothesis tests. In other words, most of these articles involved over 50 hypothesis tests per article. The *BJM* is not unusual in this respect. Similar results would be found if other academic journals in business were examined, as indeed academic journals in economics, finance, psychology and many other fields.

Many of the hypothesis tests in the *BJM* articles were those described in Chapters 10 to 18 of this book. In the present chapter, we set the scene by setting out the logic of statistical hypothesis testing, and illustrating the logic by describing several simple hypothesis tests.



9.1 DEVELOPING NULL AND ALTERNATIVE HYPOTHESES

It is not always obvious how the null and alternative hypotheses should be formulated. Care must be taken to structure the hypotheses appropriately so that the hypothesis testing conclusion provides the information the researcher or decision maker wants. The context of the situation is very important in determining how the hypotheses should be stated. All hypothesis testing applications involve collecting a sample and using the sample results to provide evidence for drawing a conclusion. Good questions to consider when formulating the null and alternative hypotheses are: What is the purpose of collecting the sample? What conclusions are we hoping to make?

In the chapter introduction, we stated that the null hypothesis H_0 is a tentative assumption about a population parameter such as a population mean or a population proportion. The alternative hypothesis H_1 states the opposite (or complement) of the null hypothesis. In some situations it is easier to identify the alternative hypothesis first and then develop the null hypothesis. In other situations it is easier to identify the null hypothesis first and then develop the alternative hypothesis. We shall illustrate these situations in the following examples.

The alternative hypothesis as a research hypothesis

Many applications of hypothesis testing involve an attempt to gather evidence in support of a research hypothesis. In these situations, it is often best to begin with the alternative hypothesis and make it the conclusion that the researcher hopes to support. Consider a particular model of car that currently attains an average fuel consumption of seven litres of fuel per 100 kilometres of driving. A product research group develops a new fuel injection system specifically designed to decrease the fuel consumption. To evaluate the new system, several will be manufactured, installed in cars and subjected to research-controlled driving tests. Here the product research group is looking for evidence to conclude that the new system *decreases* the mean fuel consumption. In this case, the research hypothesis is that the new fuel injection system will provide a mean litres-per-100 km rating below 7; that is, $\mu < 7$. As a general guideline, a research hypothesis should be stated as the *alternative hypothesis.* Hence, the appropriate null and alternative hypotheses for the study are:

$$H_0\colon \mu \geq 7$$

$$H_1\colon \mu < 7$$

If the sample results lead to the conclusion to reject H_0, the inference can be made that $H_1\colon \mu < 7$ is true. The researchers have the statistical support to state that the new fuel injection system decreases the mean litres of fuel consumed per 100km. The production of cars with the new fuel injection system should be considered. However, if the sample results lead to the conclusion that H_0 cannot be rejected, the researchers cannot conclude that the new fuel injection system is better than the current system. Production of cars with the new fuel injection system on the basis of improved fuel consumption cannot be justified. Perhaps more research and further testing can be conducted.

The conclusion that the research hypothesis is true is made if the sample data provide sufficient evidence to show that the null hypothesis can be rejected.

Successful companies stay competitive by developing new products, new methods, new systems and the like, that are better than those currently available. Before adopting something new, it is desirable to do research to determine if there is statistical support for the conclusion that the new approach is indeed better. In such cases, the research hypothesis is stated as the alternative hypothesis. For example, a new teaching method is developed that is believed to be better than the current method. The alternative hypothesis is that the new method is better. The null hypothesis is that the new method is no better than the old method. A new sales force bonus plan is developed in an attempt to increase sales. The alternative hypothesis is that the new bonus plan increases sales. The null hypothesis is that the new bonus plan does not increase sales. A new drug is developed with the goal of lowering blood pressure more than an existing drug. The alternative hypothesis is that the new drug lowers blood pressure more than the existing drug. The null hypothesis is that the new drug does not provide lower blood pressure than the existing drug. In each case, rejection of the null hypothesis H_0 provides statistical support for the research hypothesis. We will see many examples of hypothesis tests in research situations such as these throughout this chapter and in the remainder of the text.

The null hypothesis as an assumption to be challenged

Of course, not all hypothesis tests involve research hypotheses. In the following discussion we consider applications of hypothesis testing where we begin with a belief or an assumption that a statement about the value of a population parameter is true. We will then use a hypothesis test to challenge the assumption and determine if there is statistical evidence to conclude that the assumption is incorrect. In these situations, it is helpful to develop the null hypothesis first. The null hypothesis H_0 expresses the belief or assumption about the value of the population parameter. The alternative hypothesis H_1 is that the belief or assumption is incorrect.

As an example, consider the situation of a soft drinks manufacturer. The label on the bottle states that it contains 1.5 litres. We consider the label correct provided the population mean filling volume for the bottles is *at least* 1.5 litres. Without any reason to believe otherwise, we would give the manufacturer the benefit of the doubt and assume that the statement on the label is correct. So, in a hypothesis test about the population mean volume per bottle, we would begin with the assumption that the label is correct and state the null hypothesis as $\mu \geq 1.5$. The challenge to this assumption would imply that the label is incorrect and the bottles are being underfilled. This challenge would be stated as the alternative hypothesis $\mu < 1.5$. The null and alternative hypotheses are:

$$H_0\colon \mu \geq 1.5$$
$$H_1\colon \mu < 1.5$$

A trading standards office (TSO) with the responsibility for validating manufacturing labels could select a sample of soft drinks bottles, compute the sample mean filling weight and use the sample results to test the preceding hypotheses. If the sample results lead to the conclusion to reject H_0, the inference that $H_1\colon \mu < 1.5$ is true can be made. With this statistical support, the TSO is justified in concluding that the label is incorrect and underfilling of the bottles is occurring. Appropriate action to force the manufacturer to comply with labelling standards would be considered. However, if the sample results indicate H_0 cannot be rejected, the assumption that the manufacturer's labelling is correct cannot be rejected. With this conclusion, no action would be taken.

A manufacturer's product information is usually assumed to be true and stated as the null hypothesis. The conclusion that the information is incorrect can be made if the null hypothesis is rejected.

Let us now consider a variation of the soft drink bottle filling example by viewing the same situation from the manufacturer's point of view. The bottle-filling operation has been designed to fill soft drink bottles with 1.5 litres as stated on the label. The company does not want to underfill the containers because that could result in an underfilling complaint from customers or, perhaps, a TSO. However, the company does not want to overfill containers either because putting more soft drink than necessary into the containers would be an unnecessary cost. The company's goal would be to adjust the bottle-filling operation so that the population mean filling weight per bottle is 1.5 litres as specified on the label.

Although this is the company's goal, from time to time any production process can get out of adjustment. If this occurs in our example, underfilling or overfilling of the soft drink bottles will occur. In either case, the company would like to know about it in order to correct the situation by re-adjusting the bottle-filling operation to the designed 1.5 litres. In a hypothesis testing application, we would again begin with the assumption that the production process is operating correctly and state the null hypothesis as $\mu = 1.5$ litres. The alternative hypothesis that challenges this assumption is that $\mu \neq 1.5$, which indicates either overfilling or underfilling is occurring. The null and alternative hypotheses for the manufacturer's hypothesis test are:

$$H_0\colon \mu = 1.5$$
$$H_1\colon \mu \neq 1.5$$

Suppose that the soft drink manufacturer uses a quality control procedure to periodically select a sample of bottles from the filling operation and computes the sample mean filling volume per bottle. If the sample results lead to the conclusion to reject H_0, the inference is made that H_1: $\mu \neq 1.5$ is true. We conclude that the bottles are not being filled properly and the production process should be adjusted to restore the population mean to 1.5 litres per bottle. However, if the sample results indicate H_0 cannot be rejected, the assumption that the manufacturer's bottle filling operation is functioning properly cannot be rejected. In this case, no further action would be taken and the production operation would continue to run.

The two preceding forms of the soft drink manufacturing hypothesis test show that the null and alternative hypotheses may vary depending upon the point of view of the researcher or decision maker. To correctly formulate hypotheses it is important to understand the context of the situation and structure the hypotheses to provide the information the researcher or decision maker wants.

Summary of forms for null and alternative hypotheses

The hypothesis tests in this chapter involve one of two population parameters: the population mean and the population proportion. Depending on the situation, hypothesis tests about a population parameter may take one of three forms: two include inequalities in the null hypothesis, the third uses only an equality in the null hypothesis. For hypothesis tests involving a population mean, we let μ_0 denote the hypothesized value and choose one of the following three forms for the hypothesis test.

$$\begin{array}{lll} H_0\colon \mu \geq \mu_0 & \quad H_0\colon \mu \leq \mu_0 & \quad H_0\colon \mu = \mu_0 \\ H_1\colon \mu < \mu_0 & \quad H_1\colon \mu > \mu_0 & \quad H_1\colon \mu \neq \mu_0 \end{array}$$

For reasons that will be clear later, the first two forms are called one-tailed tests. The third form is called a two-tailed test.

In many situations, the choice of H_0 and H_1 is not obvious and judgement is necessary to select the proper form. However, as the preceding forms show, the equality part of the expression (either $\geq$, $\leq$ or =) *always* appears in the null hypothesis. In selecting the proper form of H_0 and H_1, keep in mind that the alternative hypothesis is often what the test is attempting to establish. Hence, asking whether the user is looking for evidence to support $\mu < \mu_0$, $\mu > \mu_0$ or $\mu \neq \mu_0$ will help determine H_1. The following exercises are designed to provide practice in choosing the proper form for a hypothesis test involving a population mean.

EXERCISES

1. The manager of the Costa Resort Hotel stated that the mean weekend guest bill is €600 or less. A member of the hotel's accounting staff noticed that the total charges for guest bills have been increasing in recent months. The accountant will use a sample of weekend guest bills to test the manager's claim.

a. Which form of the hypotheses should be used to test the manager's claim? Explain.

$$\begin{array}{lll} H_0\colon \mu \geq 600 & \quad H_0\colon \mu \leq 600 & \quad H_0\colon \mu = 600 \\ H_1\colon \mu < 600 & \quad H_1\colon \mu > 600 & \quad H_1\colon \mu \neq 600 \end{array}$$

b. What conclusion is appropriate when H_0 cannot be rejected?

c. What conclusion is appropriate when H_0 can be rejected?

COMPLETE SOLUTIONS

2. The manager of a car dealership is considering a new bonus plan designed to increase sales volume. Currently, the mean sales volume is 14 cars per month. The manager wants to conduct a research study to see whether the new bonus plan increases sales volume. To collect data on the plan, a sample of sales personnel will be allowed to sell under the new bonus plan for a one-month period.

a. Formulate the null and alternative hypotheses most appropriate for this research situation.

b. Comment on the conclusion when H_0 cannot be rejected.

c. Comment on the conclusion when H_0 can be rejected.

3. A production line operation is designed to fill cartons with laundry detergent to a mean weight of 0.75kg. A sample of cartons is periodically selected and weighed to determine whether underfilling or overfilling is occurring. If the sample data lead to a conclusion of underfilling or overfilling, the production line will be shut down and adjusted to obtain proper filling.

   a. Formulate the null and alternative hypotheses that will help in deciding whether to shut down and adjust the production line.
   b. Comment on the conclusion and the decision when H_0 cannot be rejected.
   c. Comment on the conclusion and the decision when H_0 can be rejected.

4. Because of high production-changeover time and costs, a director of manufacturing must convince management that a proposed manufacturing method reduces costs before the new method can be implemented. The current production method operates with a mean cost of €320 per hour. A research study will measure the cost of the new method over a sample production period.

   a. Formulate the null and alternative hypotheses most appropriate for this study.
   b. Comment on the conclusion when H_0 cannot be rejected.
   c. Comment on the conclusion when H_0 can be rejected.

9.2 TYPE I AND TYPE II ERRORS

The null and alternative hypotheses are competing statements about the population. Either the null hypothesis H_0 is true or the alternative hypothesis H_1 is true, but not both. Ideally the hypothesis testing procedure should lead to the acceptance of H_0 when H_0 is true and the rejection of H_0 when H_1 is true. Unfortunately, the correct conclusions are not always possible. Because hypothesis tests are based on sample information, we must allow for the possibility of errors. Table 9.1 illustrates the two kinds of errors that can be made in hypothesis testing.

The first row of Table 9.1 shows what can happen if the conclusion is to accept H_0. If H_0 is true, this conclusion is correct. However, if H_1 is true, we make a **Type II error**; that is, we accept H_0 when it is false. The second row of Table 9.1 shows what can happen if the conclusion is to reject H_0. If H_0 is true, we make a **Type I error**; that is, we reject H_0 when it is true. However, if H_1 is true, rejecting H_0 is correct.

Recall the hypothesis testing illustration discussed in Section 9.1 in which a product research group developed a new fuel injection system designed to decrease the fuel consumption of a particular car. With the current model achieving an average of seven litres of fuel per 100km, the hypothesis test was formulated as follows.

$$H_0: \mu \geq 7$$
$$H_1: \mu < 7$$

The alternative hypothesis, $H_1: \mu < 7$, indicates that the researchers are looking for sample evidence to support the conclusion that the population mean fuel consumption with the new fuel injection system is less than 7.

In this application, the Type I error of rejecting H_0 when it is true corresponds to the researchers claiming that the new system reduces fuel consumption ($\mu < 7$) when in fact the new system is no better than the current system.

TABLE 9.1 Errors and correct conclusions in hypothesis testing

		Population condition	
		H_0 true	**H_1 true**
Conclusion	**Accept H_0**	Correct conclusion	Type II error
	Reject H_0	Type I error	Correct conclusion

In contrast, the Type II error of accepting H_0 when it is false corresponds to the researchers concluding that the new system is no better than the current system ($\mu \geq 7$) when in fact the new system reduces fuel consumption.

For the fuel consumption hypothesis test, the null hypothesis is H_0: $\mu \geq 7$. Suppose the null hypothesis is true as an equality; that is, $\mu = 7$. The probability of making a Type I error when the null hypothesis is true as an equality is called the **level of significance**. This is an important concept. For the fuel efficiency hypothesis test, the level of significance is the probability of rejecting H_0: $\mu \geq 7$ when $\mu = 7$.

Level of significance

The level of significance is the probability of making a Type I error when the null hypothesis is true as an equality.

The Greek symbol α (alpha) is used to denote the level of significance. In practice, the person conducting the hypothesis test specifies the level of significance. By selecting α, that person is controlling the probability of making a Type I error. If the cost of making a Type I error is high, small values of α are preferred. If the cost of making a Type I error is not too high, larger values of α are typically used. Common choices for α are 0.05 and 0.01. Applications of hypothesis testing that only control for the Type I error are often called *significance tests*. Most applications of hypothesis testing are of this type.

Although most applications of hypothesis testing control for the probability of making a Type I error, they do not always control for the probability of making a Type II error. Hence, if we decide to accept H_0, we cannot determine how confident we can be with that decision. Because of the uncertainty associated with making a Type II error, statisticians often recommend that we use the statement 'do not reject H_0' instead of 'accept H_0'. Using the statement 'do not reject H_0' carries the recommendation to withhold both judgement and action. In effect, by not directly accepting H_0, the statistician avoids the risk of making a Type II error. Whenever the probability of making a Type II error has not been determined and controlled, we will not make the statement 'accept H_0'. In such cases, the two conclusions possible are: *do not reject* H_0 or *reject* H_0.

Although controlling for a Type II error in hypothesis testing is not common, it can be done. In Sections 9.7 and 9.8 we shall illustrate procedures for determining and controlling the probability of making a Type II error. If proper controls have been established for this error, action based on the 'accept H_0' conclusion can be appropriate.

EXERCISES

5. The label on a container of yoghurt claims that the yoghurt contains an average of one gram of fat or less. Answer the following questions for a hypothesis test that could be used to test the claim on the label.

a. Formulate the appropriate null and alternative hypotheses.
b. What is the Type I error in this situation? What are the consequences of making this error?
c. What is the Type II error in this situation? What are the consequences of making this error?

COMPLETE SOLUTIONS

6. Carpetland salespersons average €5000 per week in sales. The company's chief executive officer (CEO) proposes a remuneration plan with new selling incentives. The CEO hopes that the results of a trial selling period will enable them to conclude that the remuneration plan increases the average sales per salesperson.

a. Formulate the appropriate null and alternative hypotheses.
b. What is the Type I error in this situation? What are the consequences of making this error?
c. What is the Type II error in this situation? What are the consequences of making this error?

7. Suppose a new production method will be implemented if a hypothesis test supports the conclusion that the new method reduces the mean operating cost per hour.
   a. State the appropriate null and alternative hypotheses if the mean cost for the current production method is €320 per hour.
   b. What is the Type I error in this situation? What are the consequences of making this error?
   c. What is the Type II error in this situation? What are the consequences of making this error?

9.3 POPULATION MEAN: σ KNOWN

In this section we show how to conduct a hypothesis test about a population mean for the σ known case, i.e. where historical data and/or other information are available that enable us to obtain a good estimate of the population standard deviation prior to sampling. The methods presented in this section are exact if the sample is selected from a population that is normally distributed. In cases where it is not reasonable to assume the population is normally distributed, these methods are still applicable if the sample size is large enough. We provide some practical advice concerning the population distribution and the sample size at the end of this section.

One-tailed test

One-tailed tests about a population mean take one of the following two forms.

Lower-tail test	*Upper-tail test*
H_0: $\mu \geq \mu_0$	H_0: $\mu \leq \mu_0$
H_1: $\mu < \mu_0$	H_1: $\mu > \mu_0$

Consider an example. Trading Standards Offices (TSOs) periodically conduct statistical studies to test the claims that manufacturers make about their products. For example, suppose the label on a large bottle of Cola states that the bottle contains three litres of Cola. European legislation acknowledges that the bottling process cannot guarantee exactly three litres of Cola in each bottle, even if the mean filling volume for the population of all bottles filled is three litres per bottle. However, if the population mean filling volume is at least three litres per bottle, the rights of consumers will be protected. The legislation interprets the label information on a large bottle of Cola as a claim that the population mean filling weight is at least three litres per bottle. We shall show how a TSO can check the claim by conducting a lower-tail hypothesis test.

The first step is to formulate the null and alternative hypotheses for the test. If the population mean filling volume is at least three litres per bottle, the manufacturer's claim is correct. This establishes the null hypothesis for the test. However, if the population mean weight is less than three litres per bottle, the manufacturer's claim is incorrect. This establishes the alternative hypothesis. With μ denoting the population mean filling volume, the null and alternative hypotheses are as follows:

$$H_0\colon \mu \geq 3$$

$$H_1\colon \mu < 3$$

Note that the hypothesized value of the population mean is $\mu_0 \geq 3$. If the sample data indicate that H_0 cannot be rejected, the statistical evidence does not support the conclusion that a labelling violation has occurred. Hence, no action should be taken against the manufacturer. However, if the sample data indicate H_0 can be rejected, we shall conclude that the alternative hypothesis, H_1: $\mu < 3$, is true. In this case a conclusion of underfilling and a charge of a labelling violation against the manufacturer would be justified.

Suppose a sample of 36 bottles is selected and the sample mean is computed as an estimate of the population mean μ. If the value of the sample mean is less than three litres, the sample results will cast doubt on the null hypothesis. What we want to know is how much less than three litres the sample mean must be before we would be willing to declare the difference significant and risk making a Type I error by falsely accusing the manufacturer of a labelling violation. A key factor in addressing this issue is the value the decision-maker selects for the level of significance.

As noted in the preceding section, the level of significance, denoted by α, is the probability of making a Type I error by rejecting H_0 when the null hypothesis is true as an equality. The decision-maker must specify the level of significance. If the cost of making a Type I error is high, a small value should be chosen for the level of significance. If the cost is not high, a larger value is more appropriate. Suppose that in the Cola bottling study, the TSO made the following statement: 'If the manufacturer is meeting its weight specifications at $\mu = 3$, I would like a 99 per cent chance of not taking any action against the manufacturer. Although I do not want to accuse the manufacturer wrongly of underfilling, I am willing to risk a 1 per cent chance of making such an error.' From the TSO's statement, we set the level of significance for the hypothesis test at $\alpha = 0.01$. Hence, we must design the hypothesis test so that the probability of making a Type I error when $\mu = 3$ is 0.01.

For the Cola bottling study, by developing the null and alternative hypotheses and specifying the level of significance for the test, we carry out the first two steps required in conducting every hypothesis test. We are now ready to perform the third step of hypothesis testing: collect the sample data and compute the value of an appropriate test statistic.

Test statistic

For the Cola bottling study, previous Trading Standards tests show that the population standard deviation can be assumed known with a value of $\sigma = 0.18$. In addition, these tests also show that the population of filling weights can be assumed to have a normal distribution. From the study of sampling distributions in Chapter 7 we know that if the population from which we are sampling is normally distributed, the sampling distribution of the sample mean will also be normal in shape. Hence, for the Cola bottling study, the sampling distribution of $\overline{X}$ is normal. With a known value of $\sigma = 0.18$ and a sample size of $n = 36$, Figure 9.1 shows the sampling distribution of $\overline{X}$ when the null hypothesis is true as an equality; that is, when $\mu = \mu_0 = 3$. In constructing sampling distributions for hypothesis tests, it is assumed that H_0 is satisfied as an equality. Note that the standard error of is given by:

$$\sigma_{\overline{X}} = \sigma/\sqrt{n} = 0.18/\sqrt{36} = 0.03$$

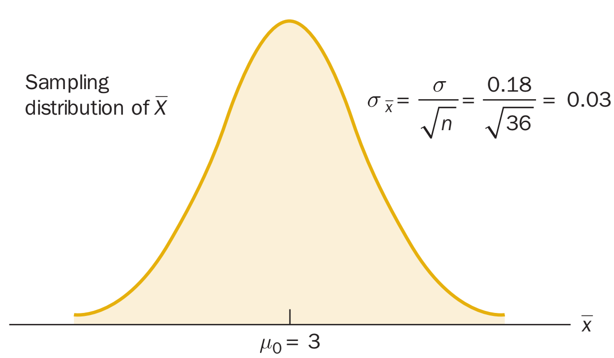
Because $\overline{X}$ has a normal sampling distribution, the sampling distribution of:

$$Z = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}} = \frac{\overline{X} - 3}{0.03}$$

is a standard normal distribution. A value $z = -1$ means that $\bar{x}$ is one standard error below the mean, a value $z = -2$ means that $\bar{x}$ is two standard errors below the mean, and so on. We can use the standard normal distribution table to find the lower-tail probability corresponding to any z value. For instance, the standard normal table shows that the cumulative probability for $z = -3.00$ is 0.0014.

FIGURE 9.1

Sampling distribution of $\overline{X}$ for the Cola bottling study when the null hypothesis is true as an equality ($\mu_0 = 3$)



This is the probability of obtaining a value that is three or more standard errors below the mean. As a result, the probability of obtaining a value $\bar{x}$ that is 3 or more standard errors below the hypothesized population mean $\mu_0 = 3$ is also 0.0014. Such a result is unlikely if the null hypothesis is true.

For hypothesis tests about a population mean for the σ known case, we use the standard normal random variable Z as a **test statistic** to determine whether $\bar{x}$ deviates from the hypothesized value μ_0 enough to justify rejecting the null hypothesis. The test statistic used in the σ known case is as follows (note that $\sigma_{\overline{X}} = \sigma/\sqrt{n}$).

Test statistic for hypothesis tests about a population mean: σ known

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \tag{9.1}$$

The key question for a lower-tail test is: How small must the test statistic z be before we choose to reject the null hypothesis? Two approaches can be used to answer this question.

The first approach uses the value z from expression (9.1) to compute a probability called a ***p*-value**. The p-value measures the support provided by the sample for the null hypothesis, and is the basis for determining whether the null hypothesis should be rejected given the level of significance. The second approach requires that we first determine a value for the test statistic called the **critical value**. For a lower-tail test, the critical value serves as a benchmark for determining whether the value of the test statistic is small enough to reject the null hypothesis. We begin with the p-value approach.

p-value approach

The p-value approach has become the preferred method of determining whether the null hypothesis can be rejected, especially when using computer software packages such as MINITAB, IBM SPSS and EXCEL. We begin with a formal definition for a p-value.

***p*-value**

The p-value is a probability, computed using the test statistic, that measures the degree to which the sample supports the null hypothesis.

Because a p-value is a probability, it ranges from 0 to 1. A small p-value indicates a sample result that is unusual given the assumption that H_0 is true. Small p-values lead to rejection of H_0, whereas large p-values indicate the null hypothesis should not be rejected.

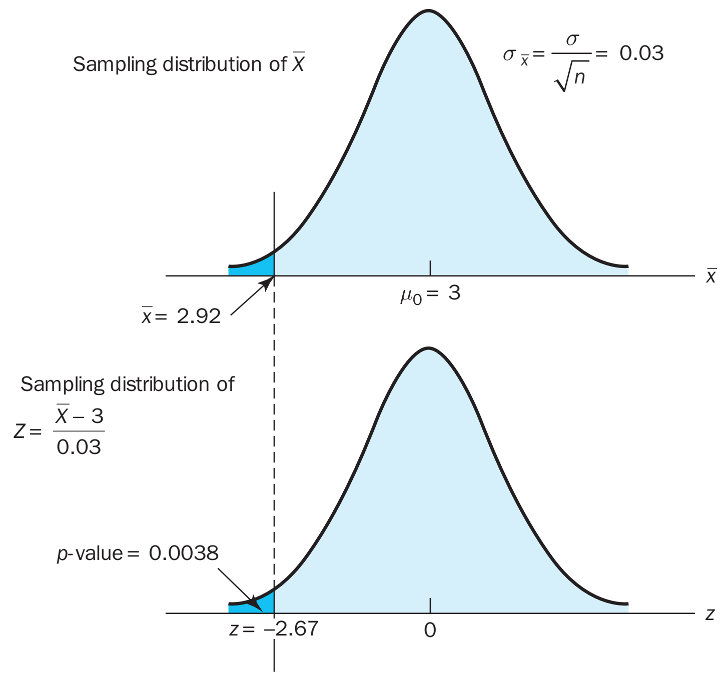
First, we use the value of the test statistic to compute the p-value. The method used to compute a p-value depends on whether the test is lower-tail, upper-tail, or a two-tailed test. For a lower tail test, the p-value is the probability of obtaining a value for the test statistic at least as small as that provided by the sample. To compute the p-value for the lower tail test in the σ known case, we find the area under the standard normal curve to the left of the test statistic. After computing the p-value, we then decide whether it is small enough to reject the null hypothesis. As we will show, this involves comparing it to the level of significance.

We now illustrate the p-value approach by computing the p-value for the Cola bottling lower-tail test. Suppose the sample of 36 Cola bottles provides a sample mean of $\bar{x} = 2.92$ litres. Is $\bar{x} = 2.92$ small enough to cause us to reject H_0? Because this test is a lower-tail test, the p-value is the area under the standard normal curve to the left of the test statistic. Using $\bar{x} = 2.92$, $\sigma = 0.18$ and $n = 36$, we compute the value z of the test statistic:

COLA

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{2.92 - 3}{0.18/\sqrt{36}} = -2.67$$

FIGURE 9.2
p-value for the Cola bottling study when $\bar{x} = 2.92$ and $z = -2.67$



The p-value is the probability that the test statistic Z is less than or equal to -2.67 (the area under the standard normal curve to the left of $z = -2.67$).

Using the standard normal distribution table, we find that the cumulative probability for $z = -2.67$ is 0.00382. Figure 9.2 shows that $\bar{x} = 2.92$ corresponds to $z = -2.67$ and a p-value $= 0.0038$. This p-value indicates a small probability of obtaining a sample mean of $\bar{x} = 2.92$ or smaller when sampling from a population with $\mu = 3$. This p-value does not provide much support for the null hypothesis, but is it small enough to cause us to reject H_0? The answer depends upon the level of significance for the test.

As noted previously, the TSO selected a value of 0.01 for the level of significance. The selection of $\alpha = 0.01$ means that the TSO is willing to accept a probability of 0.01 of rejecting the null hypothesis when it is true as an equality ($\mu_0 = 3$). The sample of 36 bottles in the Cola bottling study resulted in a p-value $= 0.0038$, which means that the probability of obtaining a value of $\bar{x} = 2.92$ or less when the null hypothesis is true as an equality is 0.0038. Because 0.0038 is less than $\alpha = 0.01$ we reject H_0. Therefore, we find sufficient statistical evidence to reject the null hypothesis at the 0.01 level of significance.

We can now state the general rule for determining whether the null hypothesis can be rejected when using the p-value approach. For a level of significance α, the rejection rule using the p-value approach is as follows:

Rejection rule using *p*-value

$$\text{Reject } H_0 \text{ if } p\text{-value} \leq \alpha$$

In the Cola bottling test, the p-value of 0.0038 resulted in the rejection of the null hypothesis. The basis for rejecting H_0 is a comparison of the p-value to the level of significance ($\alpha = 0.01$) specified by the TSO. However, the observed p-value of 0.0038 means that we would reject H_0 for any value $\alpha \geq 0.0038$. For this reason, the p-value is also called the *observed level of significance* or the *attained level of significance.*

Different decision-makers may express different opinions concerning the cost of making a Type I error and may choose a different level of significance. By providing the p-value as part of the hypothesis testing results, another decision-maker can compare the reported p-value to their own level of significance and possibly make a different decision with respect to rejecting H_0. The smaller the p-value, the greater the evidence against H_0, and the more the evidence in favour of H_1. Here are some guidelines statisticians suggest for interpreting small p-values:

- Less than 0.01 – Very strong evidence to conclude H_1 is true.
- Between 0.01 and 0.05 – Moderately strong evidence to conclude H_1 is true.
- Between 0.05 and 0.10 – Weak evidence to conclude H_1 is true.
- Greater than 0.10 – Insufficient evidence to conclude H_1 is true.

Critical value approach

For a lower-tail test, the critical value is the value of the test statistic that corresponds to an area of α (the level of significance) in the lower tail of the sampling distribution of the test statistic. In other words, the critical value is the largest value of the test statistic that will result in the rejection of the null hypothesis. Let us return to the Cola bottling example and see how this approach works.

In the σ known case, the sampling distribution for the test statistic Z is a standard normal distribution. Therefore, the critical value is the value of the test statistic that corresponds to an area of $\alpha = 0.01$ in the lower tail of a standard normal distribution. Using the standard normal distribution table, we find that $z = -2.33$ gives an area of 0.01 in the lower tail (see Figure 9.3). So if the sample results in a value of the test statistic that is less than or equal to -2.33, the corresponding p-value will be less than or equal to 0.01; in this case, we should reject the null hypothesis. Hence, for the Cola bottling study the critical value rejection rule for a level of significance of 0.01 is:

$$\text{Reject } H_0 \text{ if } z \leq -2.33$$

In the Cola bottling example, $\bar{x} = 2.92$ and the test statistic is $z = -2.67$. Because $z = -2.67 < -2.33$, we can reject H_0 and conclude that the Cola manufacturer is under-filling bottles.

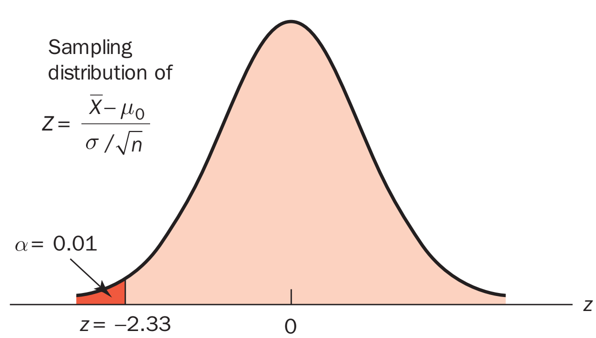
We can generalize the rejection rule for the critical value approach to handle any level of significance. The rejection rule for a lower-tail test follows.

Rejection rule for a lower-tail test: critical value approach

$$\text{Reject } H_0 \text{ if } z \leq -z_\alpha$$

where $-z_\alpha$ is the critical value; that is, the z value that provides an area of α in the lower tail of the standard normal distribution.

FIGURE 9.3
Critical value for the Cola bottling hypothesis test



The p-value approach and the critical value approach will always lead to the same rejection decision. That is, whenever the p-value is less than or equal to α, the value of the test statistic will be less than or equal to the critical value. The advantage of the p-value approach is that the p-value tells us *how* statistically significant the results are (the observed level of significance). If we use the critical value approach, we only know that the results are significant at the stated level of significance α.

Computer procedures for hypothesis testing provide the p-value, so it is rapidly becoming the preferred method of doing hypothesis tests. If you do not have access to a computer, you may prefer to use the critical value approach. For some probability distributions it is easier to use statistical tables to find a critical value than to use the tables to compute the p-value. This topic is discussed further in the next section.

At the beginning of this section, we said that one-tailed tests about a population mean take one of the following two forms:

Lower-tail test	*Upper-tail test*
H_0: $\mu \geq \mu_0$	H_0: $\mu \leq \mu_0$
H_1: $\mu < \mu_0$	H_1: $\mu > \mu_0$

We used the Cola bottling study to illustrate how to conduct a lower-tail test. We can use the same general approach to conduct an upper-tail test. The test statistic is still computed using equation (9.1). But, for an upper-tail test, the p-value is the probability of obtaining a value for the test statistic at least as large as that provided by the sample. To compute the p-value for the upper-tail test in the σ known case, we must find the area under the standard normal curve to the right of the test statistic. Using the critical value approach causes us to reject the null hypothesis if the value of the test statistic is greater than or equal to the critical value z_α. In other words, we reject H_0 if $z \geq z_\alpha$.

Two-tailed test

In hypothesis testing, the general form for a **two-tailed test** about a population mean is as follows:

$$H_0: \mu = \mu_0$$

$$H_1: \mu \neq \mu_0$$

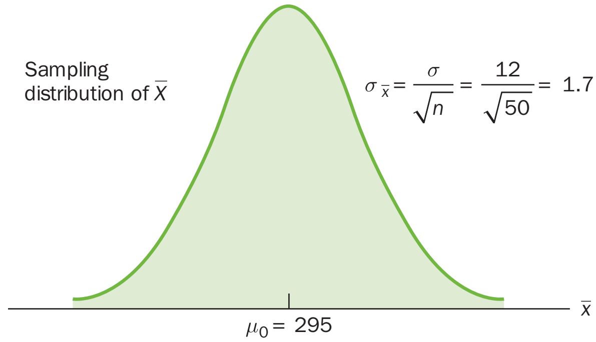
In this subsection we show how to conduct a two-tailed test about a population mean for the σ known case. As an illustration, we consider the hypothesis testing situation facing MaxFlight, a manufacturer of golf equipment who use a high technology manufacturing process to produce golf balls with an average driving distance of 295 metres. Sometimes the process gets out of adjustment and produces golf balls with average distances different from 295 metres. When the average distance falls below 295 metres, the company worries about losing sales because the golf balls do not provide as much distance as advertised. However, some of the national golfing associations impose equipment standards for professional competition and when the average driving distance exceeds 295 metres, MaxFlight's golf balls may be rejected for exceeding the overall distance standard concerning carry and roll.

MaxFlight's quality control programme involves taking periodic samples of 50 golf balls to monitor the manufacturing process. For each sample, a hypothesis test is done to determine whether the process has fallen out of adjustment. Let us formulate the null and alternative hypotheses. We begin by assuming that the process is functioning correctly; that is, the golf balls being produced have a mean driving distance of 295 metres. This assumption establishes the null hypothesis. The alternative hypothesis is that the mean driving distance is not equal to 295 yards. The null and alternative hypotheses for the MaxFlight hypothesis test are as follows:

$$H_0: \mu = 295$$

$$H_1: \mu \neq 295$$

FIGURE 9.4
Sampling distribution of $\bar{X}$ for the MaxFlight hypothesis test



If the sample mean is significantly less than 295 metres or significantly greater than 295 metres, we will reject H_0. In this case, corrective action will be taken to adjust the manufacturing process. On the other hand, if $\bar{X}$ does not deviate from the hypothesized mean $\mu_0 = 295$ by a significant amount, H_0 will not be rejected and no action will be taken to adjust the manufacturing process.

The quality control team selected $\alpha = 0.05$ as the level of significance for the test. Data from previous tests conducted when the process was known to be in adjustment show that the population standard deviation can be assumed known with a value of $\sigma = 12$. With a sample size of $n = 50$, the standard error of the sample mean is:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{50}} = 1.7$$

Because the sample size is large, the central limit theorem (see Chapter 7) allows us to conclude that the sampling distribution of $\bar{X}$ can be approximated by a normal distribution. Figure 9.4 shows the sampling distribution of $\bar{X}$ for the MaxFlight hypothesis test with a hypothesized population mean of $\mu_0 = 295$.

GOLFTEST

Suppose that a sample of 50 golf balls is selected and that the sample mean is 297.6 metres. This sample mean suggests that the population mean may be larger than 295 metres. Is this value $\bar{x} = 297.6$ sufficiently larger than 295 to cause us to reject H_0 at the 0.05 level of significance? In the previous section we described two approaches that can be used to answer this question: the p-value approach and the critical value approach.

p-value approach

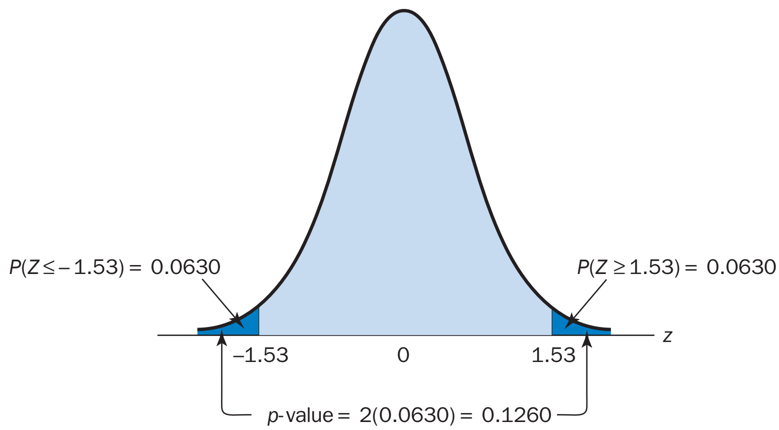
The p-value is a probability that measures the degree of support provided by the sample for the null hypothesis. For a two-tailed test, values of the test statistic in *either* tail show a lack of support for the null hypothesis. For a two-tailed test, the p-value is the probability of obtaining a value for the test statistic *at least as unlikely* as that provided by the sample. Let us compute the p-value for the MaxFlight hypothesis test.

First, we compute the value of the test statistic. For the σ known case, the test statistic Z is a standard normal random variable. Using equation (9.1) with $\bar{x} = 297.6$, the value of the test statistic is:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{297.6 - 295}{12/\sqrt{50}} = 1.53$$

Now we find the probability of obtaining a value for the test statistic *at least as unlikely* as $z = 1.53$. Clearly values ≥ 1.53 are *at least as unlikely*. But, because this is a two-tailed test, values ≤ -1.53 are also *at least as unlikely* as the value of the test statistic provided by the sample. Referring to Figure 9.5, we see that the two-tailed p-value in this case is given by $P(Z \leq -1.53) + P(Z \geq 1.53)$. Because the normal curve is symmetrical, we can compute this probability by finding the area under the standard normal curve to the left of $z = -1.53$ and doubling it. The table of cumulative probabilities for the standard normal distribution shows that the area to the left of $z = -1.53$ is 0.0630. Doubling this, we find the p-value for the MaxFlight two-tailed hypothesis test is $2(0.0630) = 0.126$.

FIGURE 9.5
p-value for the MaxFlight hypothesis text



Next we compare the p-value to the level of significance α. With $\alpha = 0.05$, we do not reject H_0 because the p-value $= 0.126 > 0.05$. Because the null hypothesis is not rejected, no action will be taken to adjust the MaxFlight manufacturing process.

The computation of the p-value for a two-tailed test may seem a bit confusing as compared to the computation of the p-value for a one-tailed test. But it can be simplified by following these three steps:

1. Compute the value of the test statistic z.
2. If the value of the test statistic is in the upper tail ($z > 0$), find the area under the standard normal curve to the right of z. If the value of the test statistic is in the lower tail, find the area under the standard normal curve to the left of z.
3. Double the tail area, or probability, obtained in step 2 to obtain the p-value.

In practice, the computation of the p-value is done automatically when using computer software such as MINITAB, IBM SPSS and EXCEL.

Critical value approach

Now let us see how the test statistic can be compared to a critical value to make the hypothesis testing decision for a two-tailed test. Figure 9.6 shows that the critical values for the test will occur in both the lower and upper tails of the standard normal distribution. With a level of significance of $\alpha = 0.05$, the area in each tail beyond the critical values is $\alpha/2 = 0.05/2 = 0.025$. Using the table of probabilities for the standard normal distribution, we find the critical values for the test statistic are $-z_{0.025} = -1.96$ and $z_{0.025} = 1.96$. Using the critical value approach, the two-tailed rejection rule is:

$$\text{Reject } H_0 \text{ if } z \leq -1.96 \text{ or if } z \geq 1.96$$

Because the value of the test statistic for the MaxFlight study is $z = 1.53$, the statistical evidence will not permit us to reject the null hypothesis at the 0.05 level of significance.

FIGURE 9.6
Critical values for the MaxFlight hypothesis test

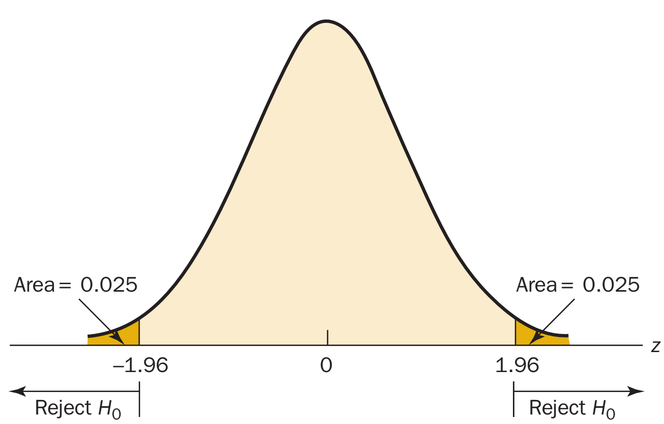


TABLE 9.2 Summary of hypothesis tests about a population mean: σ known case

	Lower-tail test	Upper-tail test	Two-tailed test
Hypotheses	H_0: $\mu \geq \mu_0$ H_1: $\mu < \mu_0$	H_0: $\mu \leq \mu_0$ H_1: $\mu > \mu_0$	H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$
Test statistic	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$	$z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
Rejection rule: ***p*-value approach**	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection rule: **critical value approach**	Reject H_0 if $z \leq -z_\alpha$	Reject H_0 if $z \geq z_\alpha$	Reject H_0 if $z \leq -z_{\alpha/2}$ or if $z \geq z_{\alpha/2}$

Summary and practical advice

We presented examples of a lower-tail test and a two-tailed test about a population mean. Based upon these examples, we can now summarize the hypothesis testing procedures about a population mean for the σ known case as shown in Table 9.2. Note that μ_0 is the hypothesized value of the population mean. The hypothesis testing steps followed in the two examples presented in this section are common to every hypothesis test.

Steps of hypothesis testing

Step 1 Formulate the null and alternative hypotheses.
Step 2 Specify the level of significance α.
Step 3 Collect the sample data and compute the value of the test statistic.

p-value approach

Step 4 Use the value of the test statistic to compute the p-value.
Step 5 Reject H_0 if the p-value $\leq \alpha$.

Critical value approach

Step 4 Use the level of significance α to determine the critical value and the rejection rule.
Step 5 Use the value of the test statistic and the rejection rule to determine whether to reject H_0.

Practical advice about the sample size for hypothesis tests is similar to the advice we provided about the sample size for interval estimation in Chapter 8. In most applications, a sample size of $n \geq 30$ is adequate when using the hypothesis testing procedure described in this section. In cases where the sample size is less than 30, the distribution of the population from which we are sampling becomes an important consideration. If the population is normally distributed, the hypothesis testing procedure that we described is exact and can be used for any sample size. If the population is not normally distributed but is at least roughly symmetrical, sample sizes as small as 15 can be expected to provide acceptable results. With smaller sample sizes, the hypothesis testing procedure presented in this section should only be used if the analyst believes, or is willing to assume, that the population is at least approximately normally distributed.

Relationship between interval estimation and hypothesis testing

We close this section by discussing the relationship between interval estimation and hypothesis testing. In Chapter 8 we showed how to construct a confidence interval estimate of a population mean. For the σ

known case, the confidence interval estimate of a population mean corresponding to a $1 - \alpha$ confidence coefficient is given by:

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{9.2}$$

Doing a hypothesis test requires us first to formulate the hypotheses about the value of a population parameter. In the case of the population mean, the two-tailed test takes the form:

$$H_0\colon \mu = \mu_0$$
$$H_1\colon \mu \neq \mu_0$$

where μ_0 is the hypothesized value for the population mean. Using the two-tailed critical value approach, we do not reject H_0 for values of the sample mean that are within $-z_{\alpha/2}$ and $+z_{\alpha/2}$ standard errors of μ_0. Hence, the do-not-reject region for the sample mean in a two-tailed hypothesis test with a level of significance of α is given by:

$$\mu_0 \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{9.3}$$

A close look at expression (9.2) and expression (9.3) provides insight about the relationship between the estimation and hypothesis testing approaches to statistical inference. Both procedures require the computation of the values $z_{\alpha/2}$ and $\sigma/\sqrt{n}$. Focusing on α, we see that a confidence coefficient of $(1 - \alpha)$ for interval estimation corresponds to a level of significance of α in hypothesis testing. For example, a 95 per cent confidence interval corresponds to a 0.05 level of significance for hypothesis testing. Furthermore, expressions (9.2) and (9.3) show that, because $z_{\alpha/2}\sigma/\sqrt{n}$ is the plus or minus value for both expressions, if $\bar{x}$ is in the do-not-reject region defined by (9.3), the hypothesized value μ_0 will be in the confidence interval defined by (9.2). Conversely, if the hypothesized value μ_0 is in the confidence interval defined by (9.2), the sample mean will be in the do-not-reject region for the hypothesis $H_0\colon \mu = \mu_0$ as defined by (9.3). These observations lead to the following procedure for using a confidence interval to conduct a two-tailed hypothesis test.

A confidence interval approach to testing a hypothesis of the form

$$H_0\colon \mu = \mu_0$$
$$H_1\colon \mu \neq \mu_0$$

1. Select a simple random sample from the population and use the value of the sample mean to construct the confidence interval for the population mean μ.

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

2. If the confidence interval contains the hypothesized value μ_0, do not reject H_0. Otherwise, reject H_0.

We return to the MaxFlight hypothesis test, which resulted in the following two-tailed test:

$$H_0\colon \mu = 295$$

$$H_1\colon \mu \neq 295$$

To test this hypothesis with a level of significance of $\alpha = 0.05$, we sampled 50 golf balls and found a sample mean distance of $\bar{x} = 297.6$ yards. The population standard deviation is $\sigma = 12$. Using these results with $z_{0.025} = 1.96$, we find that the 95 per cent confidence interval estimate of the population mean is:

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 297.6 \pm 1.96\frac{12}{\sqrt{50}} = 297.6 \pm 3.3$$

This finding enables the quality control manager to conclude with 95 per cent confidence that the mean distance for the population of golf balls is between 294.3 and 300.9 metres. Because the hypothesized value for the population mean, $\mu_0 = 295$, is in this interval, the conclusion from the hypothesis test is that the null hypothesis, H_0: $\mu = 295$, cannot be rejected.

Note that this discussion and example pertain to two-tailed hypothesis tests about a population mean. The same confidence interval and two-tailed hypothesis testing relationship exists for other population parameters. The relationship can also be extended to one-tailed tests about population parameters. Doing so, however, requires the construction of one-sided confidence intervals.

EXERCISES

Note to student: Some of the exercises ask you to use the *p*-value approach and others ask you to use the critical value approach. Both methods will provide the same hypothesis testing conclusion. We provide exercises with both methods to give you practice using both. In later sections and in following chapters, we will generally emphasize the *p*-value approach as the preferred method, but you may select either based on personal preference.

Methods

8. Consider the following hypothesis test:

$$H_0: \mu \geq 20$$
$$H_1: \mu < 20$$

COMPLETE SOLUTIONS

A sample of 50 gave a sample mean of 19.4. The population standard deviation is 2.

a. Compute the value of the test statistic.
b. What is the *p*-value?
c. Using $\alpha = 0.05$, what is your conclusion?
d. What is the rejection rule using the critical value? What is your conclusion?

9. Consider the following hypothesis test:

$$H_0: \mu = 15$$
$$H_1: \mu \neq 15$$

A sample of 50 provided a sample mean of 14.15. The population standard deviation is 3.

a. Compute the value of the test statistic.
b. What is the *p*-value?
c. At $\alpha = 0.05$, what is your conclusion?
d. What is the rejection rule using the critical value? What is your conclusion?

10. Consider the following hypothesis test:

$$H_0: \mu \leq 50$$
$$H_1: \mu > 50$$

A sample of 60 is used and the population standard deviation is 8. Use the critical value approach to state your conclusion for each of the following sample results. Use $\alpha = 0.05$.

a. $\bar{x} = 52.5$.
b. $\bar{x} = 51.0$.
c. $\bar{x} = 51.8$.

Applications

11. Suppose that the mean length of the working week for a population of workers has been previously reported as 39.2 hours. We would like to take a current sample of workers to see whether the mean length of a working week has changed from the previously reported 39.2 hours.

a. State the hypotheses that will help us determine whether a change occurred in the mean length of a working week.
b. Suppose a current sample of 112 workers provided a sample mean of 38.5 hours. Use a population standard deviation σ = 4.8 hours. What is the *p*-value?
c. At α = 0.05, can the null hypothesis be rejected? What is your conclusion?
d. Repeat the preceding hypothesis test using the critical value approach.

12. Suppose the national mean sales price for new two-bedroom houses is £181 900. A sample of 40 new two-bedroom house sales in the north-east of England showed a sample mean of £166 400. Use a population standard deviation of £33 500.

a. Formulate the null and alternative hypotheses that can be used to determine whether the sample data support the conclusion that the population mean sales price for new two-bedroom houses in the north-east is less than the national mean of £181 900.
b. What is the value of the test statistic?
c. What is the *p*-value?
d. At α = 0.01, what is your conclusion?

13. Fowler Marketing Research bases charges to a client on the assumption that telephone surveys can be completed in a mean time of 15 minutes or less per interview. If a longer mean interview time is necessary, a premium rate is charged. Suppose a sample of 35 interviews shows a sample mean of 17 minutes. Use σ = 4 minutes. Is the premium rate justified?

a. Formulate the null and alternative hypotheses for this application.
b. Compute the value of the test statistic.
c. What is the *p*-value?
d. At α = 0.01, what is your conclusion?

COMPLETE SOLUTIONS

14. CCN and ActMedia provided a television channel targeted to individuals waiting in supermarket checkout lines. The channel showed news, short features and advertisements. The length of the programme was based on the assumption that the population mean time a shopper stands in a supermarket checkout line is eight minutes. A sample of actual waiting times will be used to test this assumption and determine whether actual mean waiting time differs from this standard.

a. Formulate the hypotheses for this application.
b. A sample of 120 shoppers showed a sample mean waiting time of eight and a half minutes. Assume a population standard deviation σ = 3.2 minutes. What is the *p*-value?
c. At α = 0.05, what is your conclusion?
d. Compute a 95 per cent confidence interval for the population mean. Does it support your conclusion?

15. During the global economic upheavals in late 2008, research companies affiliated to the Worldwide Independent Network of Market Research carried out polls in 17 countries to assess people's views on the economic outlook. One of the questions asked respondents to rate their trust in their government's management of the financial situation, on a 0 to 10 scale (10 being maximum trust). Suppose the worldwide population mean on this trust question was 5.2, and we are interested in the question of whether the population mean in Germany was different from this worldwide mean.

a. State the hypotheses that could be used to address this question.

b. In the Germany survey, respondents gave the government a mean trust score of 4.0. Suppose the sample size in Germany was 1050, and the population standard deviation score was $\sigma = 2.9$. What is the 95 per cent confidence interval estimate of the population mean trust score for Germany?

c. Use the confidence interval to conduct a hypothesis test. Using $\alpha = 0.05$, what is your conclusion?

16. A production line operates with a mean filling weight of 500 grams per container. Overfilling or underfilling presents a serious problem and when detected requires the operator to shut down the production line to readjust the filling mechanism. From past data, a population standard deviation $\sigma = 25$ grams is assumed. A quality control inspector selects a sample of 30 items every hour and at that time makes the decision of whether to shut down the line for readjustment. The level of significance is $\alpha = 0.05$.

a. State the hypotheses in the hypothesis test for this quality control application.

b. If a sample mean of 510 grams were found, what is the p-value? What action would you recommend?

c. If a sample mean of 495 grams were found, what is the p-value? What action would you recommend?

d. Use the critical value approach. What is the rejection rule for the preceding hypothesis testing procedure? Repeat parts (b) and (c). Do you reach the same conclusion?

9.4 POPULATION MEAN: σ UNKNOWN

In this section we describe how to do hypothesis tests about a population mean for the σ unknown case. In this case, the sample must be used to compute estimates of both μ (estimated by $\bar{x}$) and σ (estimated by s). The steps of the hypothesis testing procedure are the same as those for the σ known case described in Section 9.3. But, with σ unknown, the computation of the test statistic and p-value are a little different. For the σ known case, the sampling distribution of the test statistic has a standard normal distribution. For the σ unknown case, the sampling distribution of the test statistic has slightly more variability because the sample is used to compute estimates of both μ and σ.

In Chapter 8, Section 8.2 we showed that an interval estimate of a population mean for the σ unknown case is based on a probability distribution known as the t distribution. Hypothesis tests about a population mean for the σ unknown case are also based on the t distribution. The test statistic has a t distribution with $n - 1$ degrees of freedom.

Test statistic for hypothesis tests about a population mean: σ unknown

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \tag{9.4}$$

In Chapter 8 we said that the t distribution is based on an assumption that the population from which we are sampling has a normal distribution. However, research shows that this assumption can be relaxed considerably when the sample size is large enough. We provide some practical advice concerning the population distribution and sample size at the end of the section.

One-tailed test

Consider an example of a one-tailed test about a population mean for the σ unknown case. A travel magazine wants to classify international airports according to the mean rating given by business travellers. A rating scale from 0 to 10 will be used, and airports with a population mean rating greater than seven will be designated as superior service airports. The magazine staff surveyed a sample of 60 business travellers at each airport. Suppose the sample for Abu Dhabi International Airport provided a sample mean rating of $\overline{x} = 7.25$ and a sample standard deviation of $s = 1.052$. Do the data indicate that Abu Dhabi should be designated as a superior service airport?

AIRRATING

We want to construct a hypothesis test for which the decision to reject H_0 will lead to the conclusion that the population mean rating for Abu Dhabi International Airport is *greater* than seven. Accordingly, an upper-tail test with H_1: $\mu > 7$ is required. The null and alternative hypotheses for this upper-tail test are as follows:

$$H_0: \mu \leq 7$$

$$H_1: \mu > 7$$

We will use $\alpha = 0.05$ as the level of significance for the test.

Using expression (9.4) with $\overline{x} = 7.25$, $s = 1.052$ and $n = 60$, the value of the test statistic is:

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{7.25 - 7}{1.052/\sqrt{60}} = 1.84$$

The sampling distribution of t has $n - 1 = 60 - 1 = 59$ degrees of freedom. Because the test is an upper-tail test, the p-value is the area under the curve of the t distribution to the right of $t = 1.84$.

The t distribution table in most textbooks will not contain sufficient detail to determine the exact p-value, such as the p-value corresponding to $t = 1.84$. For instance, using Table 2 in Appendix B, the t distribution with 59 degrees of freedom provides the following information.

Area in upper tail	0.20	0.10	0.05	0.025	0.01	0.005
t value (59 df)	0.848	1.296	1.671	2.001	2.391	2.662

$t = 1.84$

We see that $t = 1.84$ is between 1.671 and 2.001. Although the table does not provide the exact p-value, the values in the 'Area in upper tail' row show that the p-value must be less than 0.05 and greater than 0.025. With a level of significance of $\alpha = 0.05$, this placement is all we need to know to make the decision to reject the null hypothesis and conclude that Abu Dhabi should be classified as a superior service airport. Computer packages such as MINITAB, IBM SPSS and EXCEL can easily determine the exact p-value associated with the test statistic $t = 1.84$. Each of these packages will show that the p-value is 0.035 for this example. A p-value $= 0.035 < 0.05$ leads to the rejection of the null hypothesis and to the conclusion Abu Dhabi should be classified as a superior service airport.

The critical value approach can also be used to make the rejection decision. With $\alpha = 0.05$ and the t distribution with 59 degrees of freedom, $t_{0.05} = 1.671$ is the critical value for the test. The rejection rule is therefore:

$$\text{Reject } H_0 \text{ if } t \geq 1.671$$

With the test statistic $t = 1.84 > 1.671$, H_0 is rejected and we can conclude that Abu Dhabi can be classified as a superior service airport.

Two-tailed test

To illustrate a two-tailed test about a population mean for the σ unknown case, consider the hypothesis testing situation facing Mega Toys. The company manufactures and distributes its products through more than 1000 retail outlets. In planning production levels for the coming winter season, Mega Toys must decide how many units of each product to produce prior to knowing the actual demand at the retail level. For this year's most important new toy, Mega Toys' marketing director is expecting demand to average 40 units per retail outlet. Prior to making the final production decision based on this estimate, Mega Toys decided to survey a sample of 25 retailers to gather more information about the demand for the new product. Each retailer was provided with information about the features of the new toy along with the cost and the suggested selling price. Then each retailer was asked to specify an anticipated order quantity.

With μ denoting the population mean order quantity per retail outlet, the sample data will be used to conduct the following two-tailed hypothesis test:

$$H_0: \mu = 40$$

$$H_1: \mu \neq 40$$

If H_0 cannot be rejected, Mega Toys will continue its production planning based on the marketing director's estimate that the population mean order quantity per retail outlet will be $\mu = 40$ units. However, if H_0 is rejected, Mega Toys will immediately re-evaluate its production plan for the product. A two-tailed hypothesis test is used because Mega Toys wants to re-evaluate the production plan if the population mean quantity per retail outlet is less than anticipated or greater than anticipated. Because no historical data are available (it is a new product), the population mean and the population standard deviation must both be estimated using $\bar{x}$ and s from the sample data.

The sample of 25 retailers provided a mean of $\bar{x} = 37.4$ and a standard deviation of $s = 11.79$ units. Before going ahead with the use of the t distribution, the analyst constructed a histogram of the sample data in order to check on the form of the population distribution. The histogram of the sample data showed no evidence of skewness or any extreme outliers, so the analyst concluded that the use of the t distribution with $n - 1 = 24$ degrees of freedom was appropriate. Using equation (9.4) with $\bar{x} = 37.4$ $\mu_0 = 40$, $s = 11.79$, and $n = 25$, the value of the test statistic is:

ORDERS

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{37.4 - 40}{11.79/\sqrt{25}} = -1.10$$

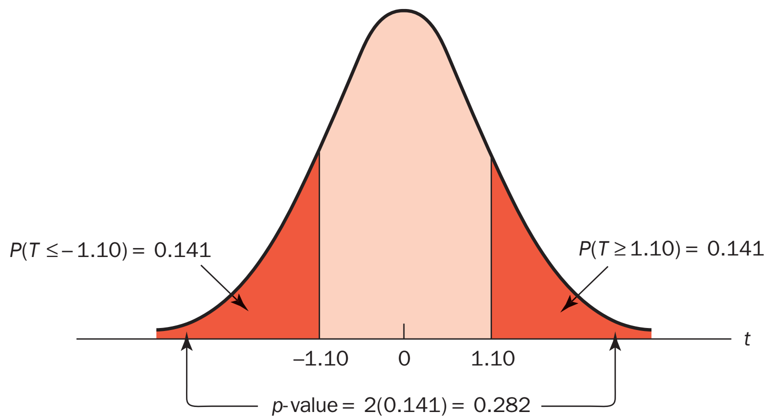
Because this is a two-tailed test, the p-value is two times the area under the curve for the t distribution to the left of $t = -1.10$. Using Table 2 in Appendix B, the t distribution table for 24 degrees of freedom provides the following information.

Area in upper tail	0.20	0.10	0.05	0.025	0.01	0.005
t value (24 df)	0.858	1.318	1.711	2.064	2.492	2.797

$t = 1.10$

The t distribution table only contains positive t values. Because the t distribution is symmetrical, however, we can find the area under the curve to the right of $t = 1.10$ and double it to find the p-value. We see that $t = 1.10$ is between 0.858 and 1.318. From the 'Area in upper tail' row, we see that the area in the tail to the right of $t = 1.10$ is between 0.20 and 0.10. Doubling these amounts, we see that the p-value must be between 0.40 and 0.20. With a level of significance of $\alpha = 0.05$, we now know that the p-value is greater than α. Therefore, H_0 cannot be rejected. There is insufficient evidence to conclude that Mega Toys should change its production plan for the coming season. Using MINITAB, IBM SPSS or EXCEL, we find that the exact p-value is 0.282. Figure 9.7 shows the two areas under the curve of the t distribution corresponding to the exact p-value.

FIGURE 9.7
Area under the curve in both tails provides the *p*-value



The test statistic can also be compared to the critical value to make the two-tailed hypothesis testing decision. With $\alpha = 0.05$ and the t distribution with 24 degrees of freedom, $-t_{0.025} = -2.064$ and $t_{0.025} = 2.064$ are the critical values for the two-tailed test. The rejection rule using the test statistic is:

$$\text{Reject } H_0 \text{ if } t \leq -2.064 \text{ or if } t \geq 2.064$$

Based on the test statistic $t = -1.10$, H_0 cannot be rejected. This result indicates that Mega Toys should continue its production planning for the coming season based on the expectation that $\mu = 40$ or do further investigation amongst its retailers.

Summary and practical advice

Table 9.3 provides a summary of the hypothesis testing procedures about a population mean for the σ unknown case. The key difference between these procedures and the ones for the σ known case are that s is used, instead of σ, in the computation of the test statistic. For this reason, the test statistic follows the t distribution.

The applicability of the hypothesis testing procedures of this section is dependent on the distribution of the population being sampled and the sample size. When the population is normally distributed, the hypothesis tests described in this section provide exact results for any sample size. When the population is not normally distributed, the procedures are approximations. Nonetheless, we find that sample sizes greater than 50 will provide good results in almost all cases. If the population is approximately normal, small sample sizes (e.g. $n < 15$) can provide acceptable results. In situations where the population cannot be approximated by a normal distribution, sample sizes of $n \geq 15$ will provide acceptable results as long as the population is not significantly skewed and does not contain outliers. If the population is significantly skewed or contains outliers, samples sizes approaching 50 are a good idea.

TABLE 9.3 Summary of hypothesis tests about a population mean: σ unknown case

	Lower-tail test	Upper-tail test	Two-tailed test
Hypotheses	H_0: $\mu \geq \mu_0$ H_1: $\mu < \mu_0$	H_0: $\mu \leq \mu_0$ H_1: $\mu > \mu_0$	H_0: $\mu = \mu_0$ H_1: $\mu \neq \mu_0$
Test statistic	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$	$t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$
Rejection rule: ***p*-value approach**	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$	Reject H_0 if p-value $\leq \alpha$
Rejection rule: **critical value approach**	Reject H_0 if $t \leq -t_\alpha$	Reject H_0 if $t \geq t_\alpha$	Reject H_0 if $t \leq -t_{\alpha/2}$ or if $t \geq t_{\alpha/2}$

EXERCISES

Methods

17. Consider the following hypothesis test:

$$H_0: \mu \leq 12$$
$$H_1: \mu > 12$$

A sample of 25 provided a sample mean $\bar{x} = 14$ and a sample standard deviation $s = 4.32$.

a. Compute the value of the test statistic.
b. What does the t distribution table (Table 2 in Appendix B) tell you about the p-value?
c. At $\alpha = 0.05$, what is your conclusion?
d. What is the rejection rule using the critical value? What is your conclusion?

18. Consider the following hypothesis test:

COMPLETE SOLUTIONS

$$H_0: \mu = 18$$
$$H_1: \mu \neq 18$$

A sample of 48 provided a sample mean $\bar{x} = 17$ and a sample standard deviation $s = 4.5$.

a. Compute the value of the test statistic.
b. What does the t distribution table (Table 2 in Appendix B) tell you about the p-value?
c. At $\alpha = 0.05$, what is your conclusion?
d. What is the rejection rule using the critical value? What is your conclusion?

19. Consider the following hypothesis test:

$$H_0: \mu \geq 45$$
$$H_1: \mu < 45$$

A sample of size 36 is used. Using $\alpha = 0.01$, identify the p-value and state your conclusion for each of the following sample results.

a. $\bar{x} = 44$ and $s = 5.2$.
b. $\bar{x} = 43$ and $s = 4.6$.
c. $\bar{x} = 46$ and $s = 5.0$.

Applications

20. Grolsch lager, like some of its competitors, can be bought in handy 300ml bottles. If a bottle such as Grolsch is marked as containing 300ml, legislation requires that the production batch from which the bottle came must have a mean fill volume of at least 300ml.

COMPLETE SOLUTIONS

a. Formulate hypotheses that could be used to determine whether the mean fill volume for a production batch satisfies the legal requirement of being at least 300ml.
b. Suppose you take a random sample of 30 bottles from a lager-bottling production line and find that the mean fill for the sample of 30 bottles is 299.5ml, with a sample standard deviation of 1.9ml. What is the p-value?
c. At $\alpha = 0.01$, what is your conclusion?

21. Consider a daily TV programme – like the 10 o'clock news – that over the last calendar year had a mean daily audience of 4.0 million viewers. Assume that for a sample of 40 days during the current year, the daily audience was 4.15 million viewers with a sample standard deviation of 0.45 million viewers.

a. If the TV management company would like to test for a change in mean viewing audience, what statistical hypotheses should be set up?
b. What is the *p*-value?
c. Select your own level of significance. What is your conclusion?

22. A popular pastime amongst football fans is participation in 'fantasy football' competitions. Participants choose a squad of players and a manager, with the objective of increasing the valuation of the squad over the season. Suppose that at the start of the competition, the mean valuation of all available strikers was £4.7 million per player.

a. Formulate the null and alternative hypotheses that could be used by a football pundit to determine whether mid-fielders have a higher mean valuation than strikers.
b. Suppose a random sample of 30 mid-fielders from the available list had a mean valuation at the start of the competition of £5.80 million with a sample standard deviation of £2.46 million. On average, by how much did the valuation of mid-fielders exceed that of strikers?
c. At α = 0.05, what is your conclusion?

23. Most new models of car sold in the European Union have to undergo an official test for fuel consumption. The test is in two parts: an urban cycle and an extra-urban cycle. The urban cycle is carried out under laboratory conditions, over a total distance of 4km at an average speed of 19km per hour. Consider a new car model for which the official fuel consumption figure for the urban cycle is published as 11.8 litres of fuel per 100km. A consumer affairs organization is interested in examining whether this published figure is truly indicative of urban driving.

a. State the hypotheses that would enable the consumer affairs organization to conclude that the model's fuel consumption is more than the published 11.8 litres per 100km.
b. A sample of 50 mileage tests with the new model of car showed a sample mean of 12.10 litres per 100km and a sample standard deviation of 0.92 litre per 100km. What is the *p*-value?
c. What conclusion should be drawn from the sample results? Use α = 0.01.
d. Repeat the preceding hypothesis test using the critical value approach.

24. SuperScapes specializes in custom-designed landscaping for residential areas. The estimated labour cost associated with a particular landscaping proposal is based on the number of plantings of trees, shrubs and so on to be used for the project. For cost-estimating purposes, managers use two hours of labour time for the planting of a medium-sized tree. Actual times from a sample of ten plantings during the past month follow (times in hours).

1.7	1.5	2.6	2.2	2.4	2.3	2.6	3.0	1.4	2.3

With a 0.05 level of significance, test to see whether the mean tree-planting time differs from two hours.

a. State the null and alternative hypotheses.
b. Compute the sample mean.
c. Compute the sample standard deviation.
d. What is the *p*-value?
e. What is your conclusion?

9.5 POPULATION PROPORTION

In this section we show how to do a hypothesis test about a population proportion π. Using π_0 to denote the hypothesized value for π, the three forms for a hypothesis test are as follows.

$$H_0\colon \pi \geq \pi_0 \qquad H_0\colon \pi \leq \pi_0 \qquad H_0\colon \pi = \pi_0$$
$$H_1\colon \pi < \pi_0 \qquad H_1\colon \pi > \pi_0 \qquad H_1\colon \pi \neq \pi_0$$

The first form is a lower-tail test, the second form is an upper-tail test, and the third form is a two-tailed test.

Hypothesis tests about a population proportion are based on the difference between the sample proportion p and the hypothesized population proportion π_0. The methods used to do the hypothesis test are similar to those used for hypothesis tests about a population mean. The only difference is that we use the sample proportion and its standard error to compute the test statistic. The p-value approach or the critical value approach is then used to determine whether the null hypothesis should be rejected.

Consider an example involving a situation faced by Aspire gymnasium. Over the past year, 20 per cent of the users of Aspire were women. In an effort to increase the proportion of women users, Aspire implemented a special promotion designed to attract women. One month afterwards, the gym manager requested a statistical study to determine whether the proportion of women users at Aspire had increased. An upper-tail test with H_1: $\pi > 0.20$ is appropriate, because the objective of the study is to determine whether the proportion of women users increased The null and alternative hypotheses for the Aspire hypothesis test are as follows:

WOMENGYM

$$H_0\colon \pi \leq 0.20$$

$$H_1\colon \pi > 0.20$$

If H_0 can be rejected, the test results will give statistical support for the conclusion that the proportion of women users increased and the promotion was beneficial. The gym manager specified that a level of significance of $\alpha = 0.05$ be used in carrying out this hypothesis test.

The next step of the hypothesis testing procedure is to select a sample and compute the value of an appropriate test statistic. We begin with a general discussion of how to compute the value of the test statistic for any form of a hypothesis test about a population proportion. The sampling distribution of P, the point estimator of the population parameter π, is the basis for constructing the test statistic.

When the null hypothesis is true as an equality, the expected value of P equals the hypothesized value π_0; that is, $E(P) = \pi_0$. The standard error of P is given by:

$$\sigma_P = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}$$

In Chapter 7 we said that if $n\pi \geq 5$ and $n(1-\pi) \geq 5$, the sampling distribution of P can be approximated by a normal distribution.* Under these conditions, which usually apply in practice, the quantity:

$$Z = \frac{P - \pi_0}{\sigma_P} \tag{9.5}$$

has a standard normal probability distribution, with $\sigma_P = \sqrt{\pi_0(1-\pi_0)/n}$. Expression (9.5) gives the test statistic used to do hypothesis tests about a population proportion.

Test statistic for hypothesis tests about a population proportion

$$z = \frac{p - \pi_0}{\sqrt{\dfrac{\pi_0(1-\pi_0)}{n}}} \tag{9.6}$$

We can now compute the test statistic for the Aspire hypothesis test. Suppose a random sample of 400 gym users was selected and that 100 of the users were women.

*In most applications involving hypothesis tests of a population proportion, sample sizes are large enough to use the normal approximation. The exact sampling distribution of P is discrete with the probability for each value of P given by the binomial distribution. So hypothesis testing is more complicated for small samples when the normal approximation cannot be used.

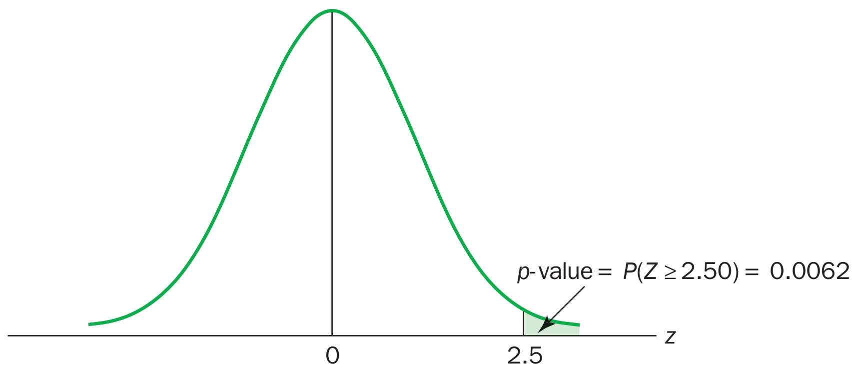


FIGURE 9.8
Calculation of the *p*-value for the Aspire hypothesis

The proportion of women users in the sample is $p = 100/400 = 0.25$. Using expression (9.6), the value of the test statistic is:

$$z = \frac{p - \pi_0}{\sqrt{\dfrac{\pi_0(1 - \pi_0)}{n}}} = \frac{0.25 - 0.20}{\sqrt{\dfrac{0.20(1 - 0.20)}{400}}} = \frac{0.05}{0.02} = 2.50$$

Because the Aspire hypothesis test is an upper-tail test, the p-value is the probability that Z is greater than or equal to $z = 2.50$. That is, it is the area under the standard normal curve to the right of $z = 2.50$. Using the table of cumulative probabilities for the standard normal distribution, we find that the p-value for the Aspire test is therefore $(1 - 0.9938) = 0.0062$. Figure 9.8 shows this p-value calculation.

Recall that the gym manager specified a level of significance of $\alpha = 0.05$. A p-value $= 0.0062 < 0.05$ gives sufficient statistical evidence to reject H_0 at the 0.05 level of significance. The test provides statistical support for the conclusion that the special promotion increased the proportion of women users at the Aspire gymnasium.

The decision whether to reject the null hypothesis can also be made using the critical value approach. The critical value corresponding to an area of 0.05 in the upper tail of a standard normal distribution is $z_{0.05} = 1.645$. Hence, the rejection rule using the critical value approach is to reject H_0 if $z \geq 1.645$. Because $z = 2.50 > 1.645$, H_0 is rejected.

Again, we see that the p-value approach and the critical value approach lead to the same hypothesis testing conclusion, but the p-value approach provides more information. With a p-value $= 0.0062$, the null hypothesis would be rejected for any level of significance greater than or equal to 0.0062.

Summary of hypothesis tests about a population proportion

The procedure used to conduct a hypothesis test about a population proportion is similar to the procedure used to conduct a hypothesis test about a population mean. Although we only illustrated how to conduct a hypothesis test about a population proportion for an upper-tail test, similar procedures can be used for lower-tail and two-tailed tests. Table 9.4 provides a summary of the hypothesis tests about a population proportion.

TABLE 9.4 Summary of hypothesis tests about a population proportion

	Lower-tail test	Upper-tail test	Two-tailed test
Hypotheses	$H_0: \pi \geq \pi_0$ $H_1: \pi < \pi_0$	$H_0: \pi \leq \pi_0$ $H_1: \pi > \pi_0$	$H_0: \pi = \pi_0$ $H_1: \pi \neq \pi_0$
Test statistic	$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$	$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$	$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$
Rejection rule: *p*-value approach	Reject H_0 if *p*-value $\leq \alpha$	Reject H_0 if *p*-value $\leq \alpha$	Reject H_0 if *p*-value $\leq \alpha$
Rejection rule: critical value approach	Reject H_0 if $z \leq -z_\alpha$	Reject H_0 if $z \geq z_\alpha$	Reject H_0 if $z \leq -z_{\alpha/2}$ or if $z \geq z_{\alpha/2}$

EXERCISES

Methods

25. Consider the following hypothesis test:

$$H_0\text{: } \pi = 0.20$$
$$H_1\text{: } \pi \neq 0.20$$

A sample of 400 provided a sample proportion $p = 0.175$.

a. Compute the value of the test statistic.
b. What is the p-value?
c. At $\alpha = 0.05$, what is your conclusion?
d. What is the rejection rule using the critical value? What is your conclusion?

26. Consider the following hypothesis test:

$$H_0\text{: } \pi \geq 0.75$$
$$H_1\text{: } \pi < 0.75$$

A sample of 300 items was selected. At $\alpha = 0.05$, compute the p-value and state your conclusion for each of the following sample results.

a. $p = 0.68$.
b. $p = 0.72$.
c. $p = 0.70$.
d. $p = 0.77$.

Applications

27. An airline promotion to business travellers is based on the assumption that at least two-thirds of business travellers use a laptop computer on overnight business trips.

a. State the hypotheses that can be used to test the assumption.
b. What is the sample proportion from an American Express sponsored survey that found 355 of 546 business travellers use a laptop computer on overnight business trips?
c. What is the p-value?
d. Use $\alpha = 0.05$. What is your conclusion?

28. Eagle Outfitters is a chain of stores specializing in outdoor clothing and camping gear. It is considering a promotion that involves sending discount coupons to all their credit card customers by direct mail. This promotion will be considered a success if more than 10 per cent of those receiving the coupons use them. Before going nationwide with the promotion, coupons were sent to a sample of 100 credit card customers.

EAGLE

a. Formulate hypotheses that can be used to test whether the population proportion of those who will use the coupons is sufficient to go national.
b. The file 'Eagle' contains the sample data. Compute a point estimate of the population proportion.
c. Use $\alpha = 0.05$ to conduct your hypothesis test. Should Eagle go national with the promotion?

29. In an IPSOS South Africa opinion poll in May 2012, a sample of adult South Africans were asked their opinions about the performance of the president, Jacob Zuma. One of the response options was the view that the president was performing 'well'.

a. Formulate the hypotheses that can be used to help determine whether more than 50 per cent of the adult population believe the president was performing well.
b. Suppose that, of the 3565 respondents to the poll, 2140 expressed the view that the president was performing well. What is the sample proportion? What is the p-value?
c. At $\alpha = 0.01$, what is your conclusion?

COMPLETE SOLUTIONS

30. A study by *Consumer Reports* showed that 64 per cent of supermarket shoppers believe supermarket brands to be as good as national name brands. To investigate whether this result applies to its own product, the manufacturer of a national name-brand ketchup asked a sample of shoppers whether they believed that supermarket ketchup was as good as the national brand ketchup.

a. Formulate the hypotheses that could be used to determine whether the percentage of supermarket shoppers who believe that the supermarket ketchup was as good as the national brand ketchup differed from 64 per cent.
b. If a sample of 100 shoppers showed 52 stating that the supermarket brand was as good as the national brand, what is the *p*-value?
c. At $\alpha = 0.05$, what is your conclusion?
d. Should the national brand ketchup manufacturer be pleased with this conclusion? Explain.

31. Microsoft Outlook is the most widely used email manager. A Microsoft executive claims that Microsoft Outlook is used by at least 75 per cent of Internet users. A sample of Internet users will be used to test this claim.

a. Formulate the hypotheses that can be used to test the claim.
b. A Merrill Lynch study reported that Microsoft Outlook is used by 72 per cent of Internet users. Assume that the report was based on a sample size of 300 Internet users. What is the *p*-value?
c. At $\alpha = 0.05$, should the executive's claim of at least 75 per cent be rejected?

32. In the elections in Greece in mid-June 2012, the centre-right New Democracy party polled 29.66 per cent of the vote. About a month before the election, a Public Issue opinion poll had estimated the proportion of support for each party. Did New Democracy's support change during the last month of the election campaign?

a. Formulate the null and alternative hypotheses.
b. Suppose the Public Issue opinion poll in May had a random sample of 1200 potential voters, and that 26.0 per cent expressed support for New Democracy. What is the *p*-value?
c. Using $\alpha = 0.05$, what is your conclusion?

9.6 HYPOTHESIS TESTING AND DECISION-MAKING

In the previous sections of this chapter we have illustrated hypothesis testing applications that are considered significance tests. After formulating the null and alternative hypotheses, we selected a sample and computed the value of a test statistic and the associated p-value. We then compared the p-value to a controlled probability of Type I error, α, which is called the level of significance for the test. If p-value $\leq \alpha$, we concluded 'reject H_0' and declared the results significant; otherwise, we made the conclusion 'do not reject H_0'. With a significance test, we control the probability of making the Type I error, but not the Type II error. Consequently, we recommended the conclusion 'do not reject H_0' rather than 'accept H_0' because the latter puts us at risk of making the Type II error of accepting H_0 when it is false. With the conclusion 'do not reject H_0', the statistical evidence is considered inconclusive and is usually an indication to postpone a decision or action until further research and testing can be undertaken.

However, if the purpose of a hypothesis test is to make a decision when H_0 is true and a different decision when H_1 is true, the decision-maker may want to, and in some cases be forced to, take action with both the conclusion *do not reject* H_0 and the conclusion *reject* H_0. If this situation occurs, statisticians generally recommend controlling the probability of making a Type II error. With the probabilities of both the Type I and Type II error controlled, the conclusion from the hypothesis test is either to *accept* H_0 or *reject* H_0. In the first case, H_0 is concluded to be true, while in the second case, H_1 is concluded true. Thus, a decision and appropriate action can be taken when either conclusion is reached.

A good illustration of this situation is lot-acceptance sampling, a topic we will discuss in more depth in Chapter 20 (on the online platform). For example, a quality control manager must decide to accept a shipment of batteries from a supplier or to return the shipment because of poor quality. Assume that design specifications require batteries from the supplier to have a mean useful life of at least 120 hours. To evaluate the quality of an incoming shipment, a sample of 36 batteries will be selected and tested. On the basis of the sample, a decision must be made to accept the shipment of batteries or to return it to the supplier because of poor quality. Let μ denote the mean number of hours of useful life for batteries in the shipment. The null and alternative hypotheses about the population mean follow.

$$H_0: \mu \geq 120$$

$$H_1: \mu < 120$$

If H_0 is rejected, the alternative hypothesis is concluded to be true. This conclusion indicates that the appropriate action is to return the shipment to the supplier. However, if H_0 is not rejected, the decision-maker must still determine what action should be taken. Therefore, without directly concluding that H_0 is true, but merely by not rejecting it, the decision-maker will have made the decision to accept the shipment as being of satisfactory quality.

In such decision-making situations, it is recommended that the hypothesis testing procedure be extended to control the probability of making a Type II error. Knowledge of the probability of making a Type II error will be helpful because a decision will be made and action taken when we do not reject H_0, In Sections 9.7 and 9.8 we explain how to compute the probability of making a Type II error and how the sample size can be adjusted to help control the probability of making a Type II error.

9.7 CALCULATING THE PROBABILITY OF TYPE II ERRORS

In this section we show how to calculate the probability of making a Type II error for a hypothesis test about a population mean. We illustrate the procedure by using the lot-acceptance example described in Section 9.6. The null and alternative hypotheses about the mean number of hours of useful life for a shipment of batteries are H_0: $\mu \geq 120$ and H_1: $\mu < 120$. If H_0 is rejected, the decision will be to return the shipment to the supplier because the mean hours of useful life are less than the specified 120 hours. If H_0 is not rejected, the decision will be to accept the shipment.

Suppose a level of significance of $\alpha = 0.05$ is used to conduct the hypothesis test. The test statistic in the σ known case is:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 120}{\sigma/\sqrt{n}}$$

Based on the critical value approach and $z_{0.05} = 1.645$, the rejection rule for the lower-tail test is to reject H_0 if $z \leq -1.645$. Suppose a sample of 36 batteries will be selected and based upon previous testing the population standard deviation can be assumed known with a value of $\sigma = 12$ hours. The rejection rule indicates that we will reject H_0 if:

$$z = \frac{\bar{x} - 120}{12/\sqrt{36}} \leq -1.645$$

Solving for $\bar{x}$ in the preceding expression indicates that we will reject H_0 if:

$$\bar{x} \leq 120 - 1.645\left(\frac{12}{\sqrt{36}}\right) = 116.71$$

Rejecting H_0 when $\bar{x} \leq 116.71$ means we will accept the shipment whenever $\bar{x} > 116.71$. We are now ready to compute probabilities associated with making a Type II error. We make a Type II error whenever the true shipment mean is less than 120 hours and we decide to accept H_0: $\mu \geq 120$. To compute the probability of making a Type II error, we must therefore select a value of μ less than 120 hours. For example, suppose the shipment is considered to be of poor quality if the batteries have a mean life of $\mu = 112$ hours. If $\mu = 112$, what is the probability of accepting H_0: $\mu \geq 120$ and hence committing a Type II error? This probability is the probability that the sample mean $\bar{x}$ is greater than 116.71 when $\mu = 112$.

FIGURE 9.9
Probability of a Type II error when $\mu = 112$

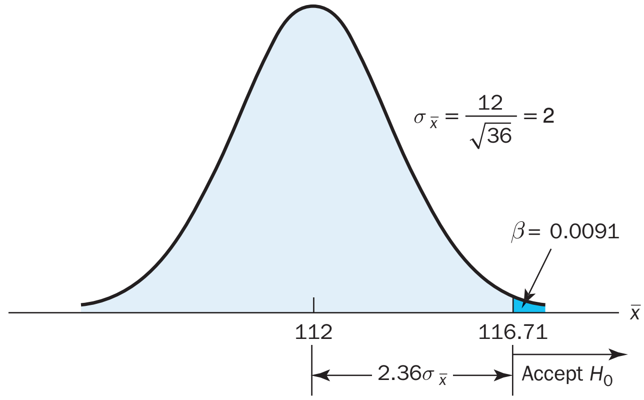


Figure 9.9 shows the sampling distribution of the sample mean when the mean is $\mu = 112$. The shaded area in the upper tail gives the probability of obtaining $\bar{x} > 116.71$. Using the standard normal distribution, we see that at $\bar{x} = 116.71$:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{116.71 - 112}{12/\sqrt{36}} = 2.36$$

The standard normal distribution table shows that with $z = 2.36$, the area in the upper tail is $1 - 0.0909 = 0.0091$. Denoting the probability of making a Type II error as β, we see if $\mu = 112$, $\beta = 0.0091$. If the mean of the population is 112 hours, the probability of making a Type II error is only 0.0091.

We can repeat these calculations for other values of μ less than 120. Doing so will show a different probability of making a Type II error for each value of μ. For example, suppose the shipment of batteries has a mean useful life of $\mu = 115$ hours. Because we will accept H_0 whenever $\bar{x} > 116.71$ the z value for $\mu = 115$ is given by:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{116.71 - 115}{12/\sqrt{36}} = 0.86$$

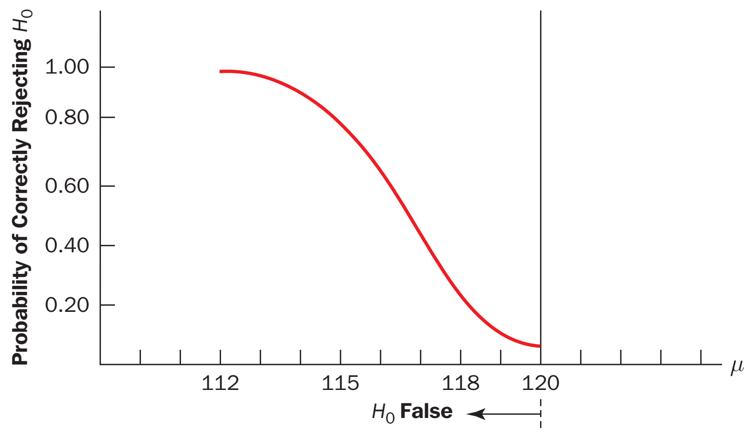
From the standard normal distribution table, we find that the area in the upper tail of the standard normal distribution for $z = 0.86$ is $1 - 0.8051 = 0.1949$. The probability of making a Type II error is $\beta = 0.1949$ when the true mean is $\mu = 115$.

In Table 9.5 we show the probability of making a Type II error for a number of values of μ less than 120. Note that as μ increases towards 120, the probability of making a Type II error increases towards an upper bound of 0.95. However, as μ decreases to values further below 120, the probability of making a Type II error diminishes. This pattern is what we should expect. When the true population mean μ is close to the null hypothesis value of $\mu = 120$, the probability is high that we will make a Type II error. However, when the true population mean μ is far below the null hypothesis value of 120, the probability is low that we will make a Type II error.

TABLE 9.5 Probability of making a Type II error for the lot-acceptance hypothesis test

Value of μ	$z = \frac{116.71 - \mu}{12/\sqrt{36}}$	Probability of a Type II error (β)	Power ($1 - \beta$)
112	2.36	0.0091	0.9909
114	1.36	0.0869	0.9131
115	0.86	0.1949	0.8051
116.71	0.00	0.5000	0.5000
117	−0.15	0.5596	0.4404
118	−0.65	0.7422	0.2578
119.999	−1.645	0.9500	0.0500

FIGURE 9.10
Power curve for the lot-acceptance hypothesis test



The probability of correctly rejecting H_0 when it is false is called the **power** of the test. For any particular value of μ, the power is $1 - \beta$; that is, the probability of correctly rejecting the null hypothesis is 1 minus the probability of making a Type II error. Values of power are also listed in Table 9.5. On the basis of these values, the power associated with each value of μ is shown graphically in Figure 9.10. Such a graph is called a **power curve**. Note that the power curve extends over the values of μ for which the null hypothesis is false. The height of the power curve at any value of μ indicates the probability of correctly rejecting H_0 when H_0 is false. Another graph, called the *operating characteristic curve,* is sometimes used to provide information about the probability of making a Type II error. The operating characteristic curve shows the probability of accepting H_0 and thus provides β for the values of μ where the null hypothesis is false. The probability of making a Type II error can be read directly from this graph.

In summary, the following step-by-step procedure can be used to compute the probability of making a Type II error in hypothesis tests about a population mean.

1. Formulate the null and alternative hypotheses.
2. Use the level of significance α and the critical value approach to determine the critical value and the rejection rule for the test.
3. Use the rejection rule to solve for the value of the sample mean corresponding to the critical value of the test statistic.
4. Use the results from step 3 to state the values of the sample mean that lead to the acceptance of H_0. These values define the acceptance region for the test.
5. Use the sampling distribution of $\overline{X}$ for a value of μ satisfying the alternative hypothesis, and the acceptance region from step 4, to compute the probability that the sample mean will be in the acceptance region. This probability is the probability of making a Type II error at the chosen value of μ.

EXERCISES

Methods

33. Consider the following hypothesis test.

$$H_0: \mu \geq 10$$
$$H_1: \mu < 10$$

The sample size is 120 and the population standard deviation is assumed known, $\sigma = 5$. Use $\alpha = 0.05$.

a. If the population mean is 9, what is the probability that the sample mean leads to the conclusion *do not reject* H_0?

b. What type of error would be made if the actual population mean is 9 and we conclude that H_0: $\mu \geq 10$ is true?

c. What is the probability of making a Type II error if the actual population mean is 8?

34. Consider the following hypothesis test.

COMPLETE SOLUTIONS

$$H_0: \mu = 20$$
$$H_1: \mu \neq 20$$

A sample of 200 items will be taken and the population standard deviation is $\sigma = 10$. Use $\alpha = 0.05$. Compute the probability of making a Type II error if the population mean is:

a. $\mu = 18.0$.

b. $\mu = 22.5$.

c. $\mu = 21.0$.

Applications

35. Fowler Marketing Research bases charges to a client on the assumption that telephone survey interviews can be completed within 15 minutes or less. If more time is required, a premium rate is charged. With a sample of 35 interviews, a population standard deviation of four minutes, and a level of significance of 0.01, the sample mean will be used to test the null hypothesis H_0: $\mu \leq 15$.

a. What is your interpretation of the Type II error for this problem? What is its impact on the firm?

b. What is the probability of making a Type II error when the actual mean time is $\mu = 17$ minutes?

c. What is the probability of making a Type II error when the actual mean time is $\mu = 18$ minutes?

d. Sketch the general shape of the power curve for this test.

36. Refer to Exercise 35. Assume the firm selects a sample of 50 interviews and repeat parts (b) and (c). What observation can you make about how increasing the sample size affects the probability of making a Type II error?

37. *Young Adult* magazine states the following hypotheses about the mean age of its subscribers.

$$H_0: \mu = 28$$
$$H_1: \mu \neq 28$$

a. What would it mean to make a Type II error in this situation?

b. The population standard deviation is assumed known at $\sigma = 6$ years and the sample size is 100. With $\alpha = 0.05$, what is the probability of accepting H_0 for μ equal to 26, 27, 29 and 30?

c. What is the power at $\mu = 26$? What does this result tell you?

38. Sparr Investments specializes in tax-deferred investment opportunities for its clients. Recently Sparr offered a payroll deduction investment scheme for the employees of a particular company. Sparr estimates that the employees are currently averaging €100 or less per month in tax-deferred investments. A sample of 40 employees will be used to test Sparr's hypothesis about the current level of investment activity among the population of employees. Assume the employee monthly tax-deferred investment amounts have a standard deviation of €75 and that a 0.05 level of significance will be used in the hypothesis test.

a. What would it mean to make a Type II error in this situation?

b. What is the probability of the Type II error if the actual mean employee monthly investment is €120?

c. What is the probability of the Type II error if the actual mean employee monthly investment is €130?

d. Assume a sample size of 80 employees is used and repeat parts (b) and (c).

9.8 DETERMINING THE SAMPLE SIZE FOR HYPOTHESIS TESTS ABOUT A POPULATION MEAN

Consider a hypothesis test about the value of a population mean. The level of significance specified by the user determines the probability of making a Type I error for the test. By controlling the sample size, the user can also control the probability of making a Type II error. Let us show how a sample size can be determined for the following lower-tail test about a population mean.

$$H_0: \mu \geq \mu_0$$

$$H_1: \mu < \mu_0$$

The upper panel of Figure 9.11 is the sampling distribution of $\bar{x}$ when H_0 is true with $\mu = \mu_0$. For a lower-tail test, the critical value of the test statistic is denoted $-z_\alpha$. In the upper panel of the figure the vertical line, labelled c, is the corresponding value of $\bar{x}$. Note that, if we reject H_0 when $\bar{x} = c$ the probability of a Type I error will be α. With z_α representing the z value corresponding to an area of α in the upper tail of the standard normal distribution, we compute c using the following formula:

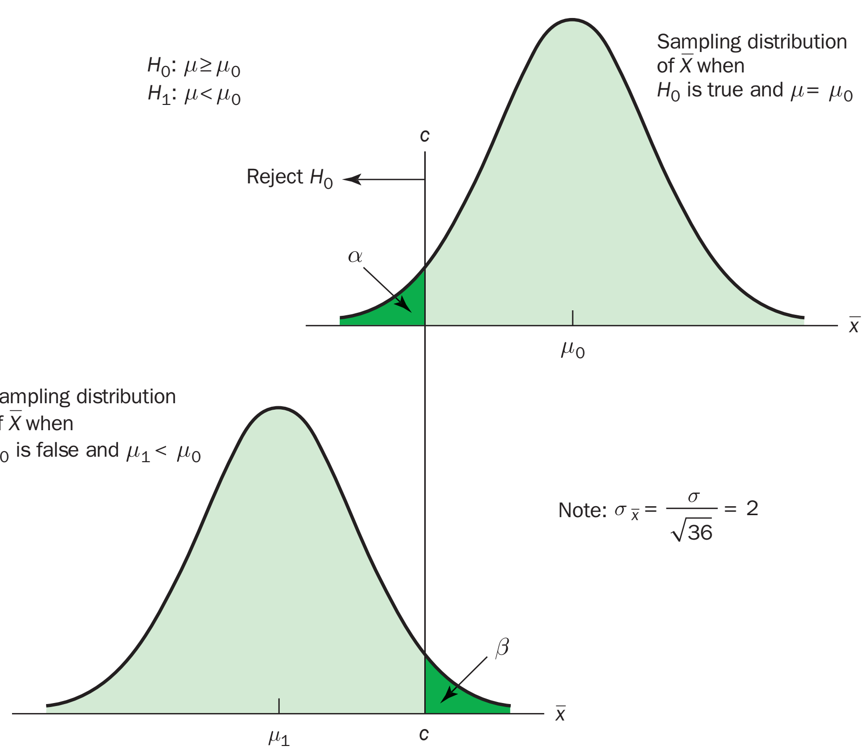
$$c = \mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} \quad \textbf{(9.7)}$$

The lower panel of Figure 9.11 is the sampling distribution of $\bar{X}$ when the alternative hypothesis is true with $\mu = \mu_1 < \mu_0$. The shaded region shows β, the probability of a Type II error if the null hypothesis is accepted when $\bar{x} > c$. With z_β representing the z value corresponding to an area of β in the upper tail of the standard normal distribution, we compute c using the following formula:

$$c = \mu_1 + z_\beta \frac{\sigma}{\sqrt{n}} \quad \textbf{(9.8)}$$

We wish to select a value for c so that when we reject or do not reject H_0, the probability of a Type I error is equal to the chosen value of α and the probability of a Type II error is equal to the chosen value of β.

FIGURE 9.11
Determining the sample size for specified levels of the Type I (α) and Type II (β) errors



Therefore, both equations (9.7) and (9.8) must provide the same value for *c*. Hence, the following equation must be true.

$$\mu_0 - z_\alpha \frac{\sigma}{\sqrt{n}} = \mu_1 + z_\beta \frac{\sigma}{\sqrt{n}}$$

To determine the required sample size, we first solve for $\sqrt{n}$ as follows.

$$\mu_0 - \mu_1 = z_\alpha \frac{\sigma}{\sqrt{n}} + z_\beta \frac{\sigma}{\sqrt{n}} = \frac{(z_\alpha + z_\beta)\sigma}{\sqrt{n}}$$

and:

$$\sqrt{n} = \frac{(z_\alpha + z_\beta)\sigma}{(\mu_0 - \mu_1)}$$

Squaring both sides of the expression provides the following sample size formula for a one-tailed hypothesis test about a population mean.

Sample size for a one-tailed hypothesis test about a population mean

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{(\mu_0 - \mu_1)^2} \tag{9.9}$$

z_α = z value giving an area of α in the upper tail of a standard normal distribution.
z_β = z value giving an area of β in the upper tail of a standard normal distribution.
σ = the population standard deviation.
μ_0 = the value of the population mean in the null hypothesis.
μ_1 = the value of the population mean used for the Type II error.

Although the logic of equation (9.9) was developed for the hypothesis test shown in Figure 9.11, it holds for any one-tailed test about a population mean. Note that in a two-tailed hypothesis test about a population mean, $z_{\alpha/2}$ is used instead of z_α in equation (9.9).

Let us return to the lot-acceptance example from Sections 9.6 and 9.7. The design specification for the shipment of batteries indicated a mean useful life of at least 120 hours for the batteries. Shipments were rejected if H_0: $\mu \geq 120$ was rejected. Let us assume that the quality control manager makes the following statements about the allowable probabilities for the Type I and Type II errors:

Type I error statement: If the mean life of the batteries in the shipment is $\mu = 120$, I am willing to risk an $\alpha = 0.05$ probability of rejecting the shipment.

Type II error statement: If the mean life of the batteries in the shipment is five hours under the specification (i.e. $\mu = 115$), I am willing to risk a $\beta = 0.10$ probability of accepting the shipment.

Statements about the allowable probabilities of both errors must be made before the sample size can be determined. These statements are based on the judgement of the manager. Someone else might specify different restrictions on the probabilities.

In the example, $\alpha = 0.05$ and $\beta = 0.10$. Using the standard normal probability distribution, we have $z_{0.05} = 1.645$ and $z_{0.10} = 1.28$. From the statements about the error probabilities, we note that $\mu_0 = 120$ and $\mu_1 = 115$. The population standard deviation was assumed known at $\sigma = 12$. By using equation (9.9), we find that the recommended sample size for the lot-acceptance example is:

$$n = \frac{(1.645 + 1.28)^2 (12)^2}{(120 - 115)^2} = 49.3$$

Rounding up, we recommend a sample size of 50.

Because both the Type I and Type II error probabilities have been controlled at allowable levels with $n = 50$, the quality control manager is now justified in using the *accept* H_0 and *reject* H_0 statements for the hypothesis test. The accompanying inferences are made with allowable probabilities of making Type I and Type II errors.

We can make three observations about the relationship among α, β and the sample size n.

1 Once two of the three values are known, the other can be computed.

2 For a given level of significance α, increasing the sample size will reduce β.

3 For a given sample size, decreasing α will increase β, whereas increasing α will decrease β.

The third observation should be kept in mind when the probability of a Type II error is not being controlled. It suggests that one should not choose unnecessarily small values for the level of significance α. For a given sample size, choosing a smaller level of significance means more exposure to a Type II error. Inexperienced users of hypothesis testing often think that smaller values of α are always better. They are better if we are concerned only about making a Type I error. However, smaller values of α have the disadvantage of increasing the probability of making a Type II error.

EXERCISES

Methods

39. Consider the following hypothesis test.

$$H_0\text{: } \mu \geq 10$$
$$H_1\text{: } \mu < 10$$

The sample size is 120 and the population standard deviation is 5. Use $\alpha = 0.05$. If the actual population mean is 9, the probability of a Type II error is 0.2912. Suppose the researcher wants to reduce the probability of a Type II error to 0.10 when the actual population mean is 9. What sample size is recommended?

40. Consider the following hypothesis test.

$$H_0\text{: } \mu = 20$$
$$H_1\text{: } \mu \neq 20$$

The population standard deviation is 10. Use $\alpha = 0.05$. How large a sample should be taken if the researcher is willing to accept a 0.05 probability of making a Type II error when the actual population mean is 22?

COMPLETE SOLUTIONS

Applications

41. A special industrial battery must have a mean life of at least 400 hours. A hypothesis test is to be conducted with a 0.02 level of significance. If the batteries from a particular production run have an actual mean use life of 385 hours, the production manager wants a sampling procedure that only 10 per cent of the time would show erroneously that the batch is acceptable. What sample size is recommended for the hypothesis test? Use 30 hours as an estimate of the population standard deviation.

42. *Young Adult* magazine states the following hypotheses about the mean age of its subscribers.

$$H_0\text{: } \mu = 28$$
$$H_1\text{: } \mu \neq 28$$

If the manager conducting the test will permit a 0.15 probability of making a Type II error when the true mean age is 29, what sample size should be selected? Assume $\sigma = 6$ and a 0.05 level of significance.

43. H_0: $\mu = 120$ and H_1: $\mu \neq 120$ are used to test whether a bath soap production process is meeting the standard output of 120 bars per batch. Use a 0.05 level of significance for the test and a planning value of 5 for the standard deviation.

a. If the mean output drops to 117 bars per batch, the firm wants to have a 98 per cent chance of concluding that the standard production output is not being met. How large a sample should be selected?

b. With your sample size from part (a), what is the probability of concluding that the process is operating satisfactorily for each of the following actual mean outputs: 117, 118, 119, 121, 122 and 123 bars per batch? That is, what is the probability of a Type II error in each case?

ONLINE RESOURCES

For the data files, online summary, additional questions and answers, and software section for Chapter 9, go to the online platform.

SUMMARY

Hypothesis testing uses sample data to determine whether a statement about the value of a population parameter should or should not be rejected. The hypotheses are two competing statements about a population parameter. One is called the null hypothesis (H_0), and the other is called the alternative hypothesis (H_1). In Section 9.1 we provided guidelines for formulating hypotheses for situations frequently encountered in practice.

In all hypothesis tests, a relevant test statistic is calculated using sample data. The test statistic can be used to compute a *p*-value for the test. A *p*-value is a probability that measures the degree of support provided by the sample for the null hypothesis. If the *p*-value is less than or equal to the level of significance α, the null hypothesis can be rejected.

Conclusions can also be drawn by comparing the value of the test statistic to a critical value. For lower-tail tests, the null hypothesis is rejected if the value of the test statistic is less than or equal to the critical value. For upper-tail tests, the null hypothesis is rejected if the value of the test statistic is greater than or equal to the critical value. Two-tailed tests consist of two critical values: one in the lower tail of the sampling distribution and one in the upper tail. In this case, the null hypothesis is rejected if the value of the test statistic is less than or equal to the critical value in the lower tail or greater than or equal to the critical value in the upper tail.

We illustrated the relationship between hypothesis testing and interval construction in Section 9.3.

When historical data or other information provides a basis for assuming that the population standard deviation is known, the hypothesis testing procedure is based on the standard normal distribution. When σ is unknown, the sample standard deviation s is used to estimate σ and the hypothesis testing procedure is based on the t distribution.

In the case of hypothesis tests about a population proportion, the hypothesis testing procedure uses a test statistic based on the standard normal distribution.

Extensions of hypothesis testing procedures to include an analysis of the Type II error were also presented. In Section 9.7 we showed how to compute the probability of making a Type II error. In Section 9.8 we showed how to determine a sample size that will control for both the probability of making a Type I error and a Type II error.

KEY TERMS

Alternative hypothesis
Critical value
Level of significance
Null hypothesis
One-tailed test
***p*-value**
Power
Power curve
Test statistic
Two-tailed test
Type I error
Type II error

KEY FORMULAE

Test statistic for hypothesis tests about a population mean: σ known

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \tag{9.1}$$

Test statistic for hypothesis tests about a population mean: σ unknown

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \tag{9.4}$$

Test statistic for hypothesis tests about a population proportion

$$z = \frac{p - \pi_0}{\sqrt{\dfrac{\pi_0(1 - \pi_0)}{n}}} \tag{9.6}$$

Sample size for a one-tailed hypothesis test about a population mean

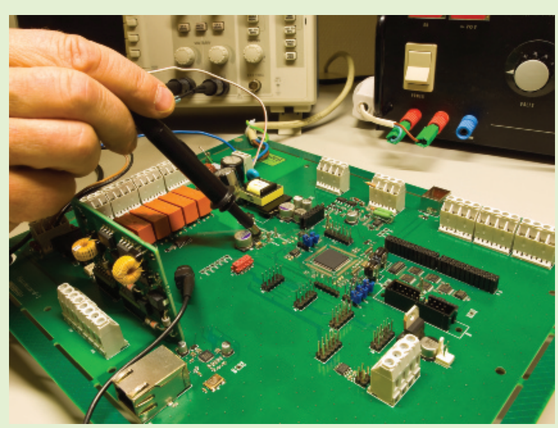
$$n = \frac{(z_\alpha + z_\beta)^2\sigma^2}{(\mu_0 - \mu_1)^2} \tag{9.9}$$

In a two-tailed test, replace z_α with $z_{\alpha/2}$.

CASE PROBLEM 1

Quality Associates

Quality Associates, a consulting firm, advises its clients about sampling and statistical procedures that can be used to control their manufacturing processes. In one particular application, a client gave Quality Associates a sample of 800 observations taken during a time in which that client's process was operating satisfactorily. The sample standard deviation for these data was 0.21; hence, with so



The components of an electronic product are tested much data, the population standard deviation was assumed to be 0.21. Quality Associates then suggested that random samples of size 30 be taken periodically to monitor the process on an ongoing basis. By analyzing the new samples, the client could quickly learn whether the process was operating satisfactorily. When the process was not operating satisfactorily, corrective action could be taken to eliminate the problem. The design specification indicated the mean for the process should be 12. The hypothesis test suggested by Quality Associates follows.

$$H_0: \mu = 12$$

$$H_1: \mu \neq 12$$

Corrective action will be taken any time H_0 is rejected.

QUALITY

The data set 'Quality' on the online platform contains data from four samples, each of size 30, collected at hourly intervals during the first day of operation of the new statistical control procedure.

Managerial report

1. Do a hypothesis test for each sample at the 0.01 level of significance and determine what action, if any, should be taken. Provide the test statistic and *p*-value for each test.
2. Compute the standard deviation for each of the four samples. Does the assumption of 0.21 for the population standard deviation appear reasonable?
3. Compute limits for the sample mean $\bar{X}$ around $\mu = 12$ such that, as long as a new sample mean is within those limits, the process will be considered to be operating satisfactorily. If $\bar{X}$ exceeds the upper limit or if is below the lower limit, corrective action will be taken. These limits are referred to as upper and lower control limits for quality control purposes.
4. Discuss the implications of changing the level of significance to a larger value. What mistake or error could increase if the level of significance is increased?

CASE PROBLEM 2

Ethical behaviour of business students at the World Academy

During the global recession of 2008 and 2009, there were many accusations of unethical behaviour by bank directors, financial managers and other corporate officers. At that time, an article appeared that suggested part of the reason for such unethical business behaviour may stem from the fact that cheating has become more prevalent among business students (*Chronicle of Higher Education*, February 10, 2009). The article reported that 56 per cent of business students admitted to cheating at some time during their academic career as compared to 47 per cent of non-business students.

Cheating has been a concern of the dean of the Faculty of Business at the World Academy for several years. Some faculty members believe that cheating is more widespread at the World Academy than at other universities, while other faculty members think that cheating is not a major problem in the Academy. To resolve some of these issues, the dean commissioned a study to assess the current ethical behaviour of business students at the World Academy. As part of this study, an anonymous exit survey was administered to a sample of 90 business students from this year's graduating class. Responses to the following questions were used to obtain data regarding three types of cheating.

During your time at the World Academy, did you ever present work copied off the Internet as your own?

Yes _____ No _____

During your time at the World Academy, did you ever copy answers off another student's exam?

Yes _____ No _____

During your time at the World Academy, did you ever collaborate with other students on projects that were supposed to be completed individually?

Yes _____ No _____

Any student who answered Yes to one or more of these questions was considered to have been involved in some type of cheating. A portion of the data collected follows. The complete data set is in the file named 'World Academy' on the accompanying online platform.

WORLD ACADEMY

Student	*Copied from Internet*	*Copied on exam*	*Collaborated on Individual project*	*Gender*
1	No	No	No	Female
2	No	No	No	Male
3	Yes	No	Yes	Male
4	Yes	Yes	No	Male
5	No	No	Yes	Male
6	Yes	No	No	Female
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
88	No	No	No	Male
89	No	Yes	Yes	Male
90	No	No	No	Female

Managerial report

Prepare a report for the dean of the faculty that summarizes your assessment of the nature of cheating by business students at the World Academy. Be sure to include the following items in your report.

1. Use descriptive statistics to summarize the data and comment on your findings.
2. Develop 95 per cent confidence intervals for the proportion of all students, the proportion of male students and the proportion of female students who were involved in some type of cheating.
3. Conduct a hypothesis test to determine if the proportion of business students at the World Academy who were involved in some type of cheating is less than that of business students at other institutions as reported by the *Chronicle of Higher Education*.
4. Conduct a hypothesis test to determine if the proportion of business students at the World Academy who were involved in some form of cheating is less than that of non-business students at other institutions as reported by the *Chronicle of Higher Education*.
5. What advice would you give to the dean based up-on your analysis of the data?