Chapter 8

Testing Hypotheses

A manufacturer of emergency equipment asserts that a respirator that it makes delivers pure air for 75 minutes on average. A government regulatory agency is charged with testing such claims, in this case to verify that the average time is not less than 75 minutes. To do so it would select a random sample of respirators, compute the mean time that they deliver pure air, and compare that mean to the asserted time 75 minutes.

In the sampling that we have studied so far the goal has been to estimate a population parameter. But the sampling done by the government agency has a somewhat different objective, not so much to *estimate* the population mean μ as to *test* an assertion—or a **hypothesis**[1]—about it, namely, whether it is as large as 75 or not. The agency is not necessarily interested in the actual value of μ, just whether it is as claimed. Their sampling is done to perform a test of hypotheses, the subject of this chapter.

1. A statement about a population parameter.

8.1 The Elements of Hypothesis Testing

Types of Hypotheses

A *hypothesis* about the value of a population parameter is an assertion about its value. As in the introductory example we will be concerned with testing the truth of two competing hypotheses, only one of which can be true.

Definition

The **null hypothesis**[2]*, denoted H_0, is the statement about the population parameter that is assumed to be true unless there is convincing evidence to the contrary.*

The **alternative hypothesis**[3]*, denoted H_a, is a statement about the population parameter that is contradictory to the null hypothesis, and is accepted as true only if there is convincing evidence in favor of it.*

Definition

Hypothesis testing[4] *is a statistical procedure in which a choice is made between a null hypothesis and an alternative hypothesis based on information in a sample.*

2. The statement that is assumed to be true unless there is convincing evidence to the contrary.

3. A statement that is accepted as true only if there is convincing evidence in favor of it.

4. A statistical procedure in which a choice is made between a null hypothesis and a specific alternative hypothesis based on information in a sample.

The end result of a hypotheses testing procedure is a choice of one of the following two possible conclusions:

1. Reject H_0 (and therefore accept H_a), or
2. Fail to reject H_0 (and therefore fail to accept H_a).

The null hypothesis typically represents the status quo, or what has historically been true. In the example of the respirators, we would believe the claim of the manufacturer unless there is reason not to do so, so the null hypotheses is $H_0 : \mu = 75.$ The alternative hypothesis in the example is the contradictory statement $H_a : \mu < 75.$ The null hypothesis will always be an assertion containing an equals sign, but depending on the situation the alternative hypothesis can have any one of three forms: with the symbol "<," as in the example just discussed, with the symbol ">," or with the symbol "≠" The following two examples illustrate the latter two cases.

EXAMPLE 1

A publisher of college textbooks claims that the average price of all hardbound college textbooks is $127.50. A student group believes that the actual mean is higher and wishes to test their belief. State the relevant null and alternative hypotheses.

Solution:

The default option is to accept the publisher's claim unless there is compelling evidence to the contrary. Thus the null hypothesis is $H_0 : \mu = 127.50.$ Since the student group thinks that the average textbook price is *greater* than the publisher's figure, the alternative hypothesis in this situation is $H_a : \mu > 127.50.$

> ### EXAMPLE 2
>
> The recipe for a bakery item is designed to result in a product that contains 8 grams of fat per serving. The quality control department samples the product periodically to insure that the production process is working as designed. State the relevant null and alternative hypotheses.
>
> Solution:
>
> The default option is to assume that the product contains the amount of fat it was formulated to contain unless there is compelling evidence to the contrary. Thus the null hypothesis is $H_0 : \mu = 8.0.$ Since to contain either more fat than desired or to contain less fat than desired are both an indication of a faulty production process, the alternative hypothesis in this situation is that the mean is *different* from 8.0, so $H_a : \mu \neq 8.0.$

In Note 8.8 "Example 1", the textbook example, it might seem more natural that the publisher's claim be that the average price is at most $127.50, not exactly $127.50. If the claim were made this way, then the null hypothesis would be $H_0 : \mu \leq 127.50$, and the value $127.50 given in the example would be the one that is least favorable to the publisher's claim, the null hypothesis. It is always true that if the null hypothesis is retained for its least favorable value, then it is retained for every other value.

Thus in order to make the null and alternative hypotheses easy for the student to distinguish, in every example and problem in this text we will always present one of the two competing claims about the value of a parameter with an equality. *The claim expressed with an equality is the null hypothesis.* This is the same as always stating the null hypothesis in the least favorable light. So in the introductory example about the respirators, we stated the manufacturer's claim as "the average is 75 minutes" instead of the perhaps more natural "the average is at least 75 minutes," essentially reducing the presentation of the null hypothesis to its worst case.

The first step in hypothesis testing is to identify the null and alternative hypotheses.

The Logic of Hypothesis Testing

Although we will study hypothesis testing in situations other than for a single population mean (for example, for a population proportion instead of a mean or in

comparing the means of two different populations), in this section the discussion will always be given in terms of a single population mean μ.

The null hypothesis always has the form $H_0 : \mu = \mu_0$ for a specific number μ_0 (in the respirator example $\mu_0 = 75$, in the textbook example $\mu_0 = 127.50$, and in the baked goods example $\mu_0 = 8.0$). Since the null hypothesis is accepted unless there is strong evidence to the contrary, the test procedure is based on the initial assumption that H_0 is true. This point is so important that we will repeat it in a display:

> *The test procedure is based on the initial assumption that H_0 is true.*

The criterion for judging between H_0 and H_a based on the sample data is: if the value of $\overline{X}$ would be highly unlikely to occur if H_0 were true, but favors the truth of H_a, then we reject H_0 in favor of H_a. Otherwise we do not reject H_0.

Supposing for now that $\overline{X}$ follows a normal distribution, when the null hypothesis is true the density function for the sample mean $\overline{X}$ must be as in <u>Figure 8.1 "The Density Curve for "</u>: a bell curve centered at μ_0. Thus if H_0 is true then $\overline{X}$ is likely to take a value near μ_0 and is unlikely to take values far away. Our decision procedure therefore reduces simply to:

1. if H_a has the form $H_a : \mu < \mu_0$ then reject H_0 if $\overline{x}$ is far to the left of μ_0;
2. if H_a has the form $H_a : \mu > \mu_0$ then reject H_0 if $\overline{x}$ is far to the right of μ_0;
3. if H_a has the form $H_a : \mu \neq \mu_0$ then reject H_0 if $\overline{x}$ is far away from μ_0 in either direction.

Figure 8.1 *The Density Curve for $\overline{X}$ if H_0 Is True*
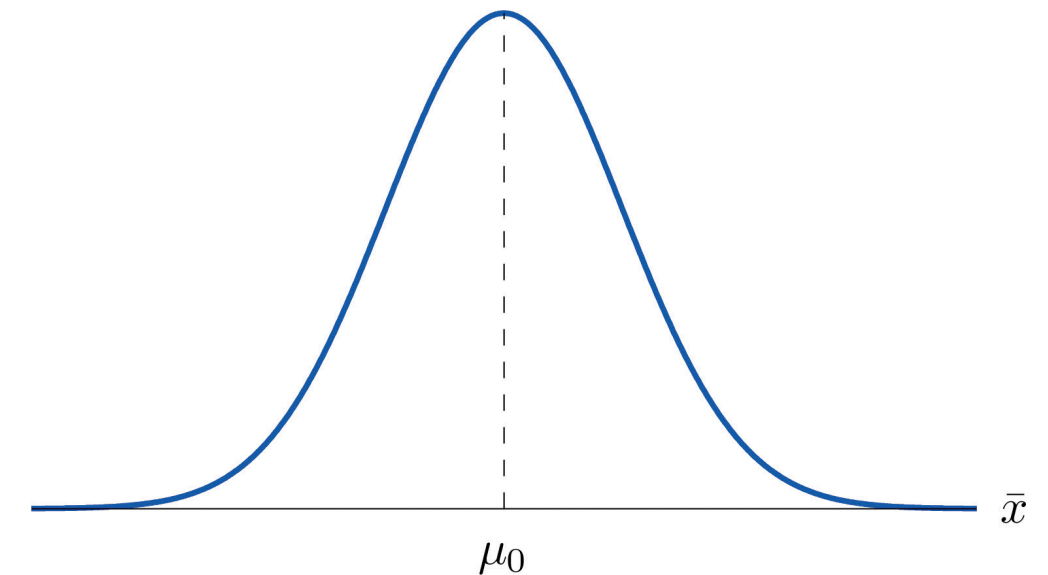
Think of the respirator example, for which the null hypothesis is $H_0 : \mu = 75$, the claim that the average time air is delivered for *all* respirators is 75 minutes. If the sample mean is 75 or greater then we certainly would not reject H_0 (since there is no issue with an emergency respirator delivering air even longer than claimed).

If the sample mean is slightly less than 75 then we would logically attribute the difference to sampling error and also not reject H_0 either.

Values of the sample mean that are smaller and smaller are less and less likely to come from a population for which the population mean is 75. Thus if the sample mean is far less than 75, say around 60 minutes or less, then we would certainly reject H_0, because we know that it is highly unlikely that the average of a sample would be so low if the population mean were 75. This is the *rare event criterion* for rejection: what we actually observed ($\overline{X} < 60$) would be so rare an event if $\mu = 75$ were true that we regard it as much more likely that the alternative hypothesis $\mu < 75$ holds.

In summary, to decide between H_0 and H_a in this example we would select a "**rejection region**[5]" of values sufficiently far to the left of 75, based on the rare event criterion, and reject H_0 if the sample mean $\overline{X}$ lies in the rejection region, but not reject H_0 if it does not.

5. An interval or union of intervals such that the null hypothesis is rejected if and only if the statistic of interest lies in this region.

The Rejection Region

Each different form of the alternative hypothesis H_a has its own kind of rejection region:

1. if (as in the respirator example) H_a has the form $H_a : \mu < \mu_0$, we reject H_0 if $\overline{x}$ is far to the left of μ_0, that is, to the left of some number C, so the rejection region has the form of an interval $(-\infty, C]$;
2. if (as in the textbook example) H_a has the form $H_a : \mu > \mu_0$, we reject H_0 if $\overline{x}$ is far to the right of μ_0, that is, to the right of some number C, so the rejection region has the form of an interval $[C, \infty)$;
3. if (as in the baked good example) H_a has the form $H_a : \mu \neq \mu_0$, we reject H_0 if $\overline{x}$ is far away from μ_0 in either direction, that is, either to the left of some number C or to the right of some other number C', so the rejection region has the form of the union of two intervals $(-\infty, C] \cup [C', \infty)$.

The key issue in our line of reasoning is the question of how to determine the number C or numbers C and C', called the *critical value* or *critical values* of the statistic, that determine the rejection region.

> ## Definition
>
> The **critical value**[6] or **critical values** *of a test of hypotheses are the number or numbers that determine the rejection region.*

Suppose the rejection region is a single interval, so we need to select a single number C. Here is the procedure for doing so. We select a small probability, denoted α, say 1%, which we take as our definition of "rare event:" an event is "rare" if its probability of occurrence is less than α. (In all the examples and problems in this text the value of α will be given already.) The probability that $\overline{X}$ takes a value in an interval is the area under its density curve and above that interval, so as shown in [Figure 8.2](#) (drawn under the assumption that H_0 is true, so that the curve centers at μ_0) the critical value C is the value of $\overline{X}$ that cuts off a tail area α in the probability density curve of $\overline{X}$. When the rejection region is in two pieces, that is, composed of two intervals, the total area above both of them must be α, so the area above each one is $\alpha / 2$, as also shown in [Figure 8.2](#).

6. The number or one of a pair of numbers that determines the rejection region.

Figure 8.2
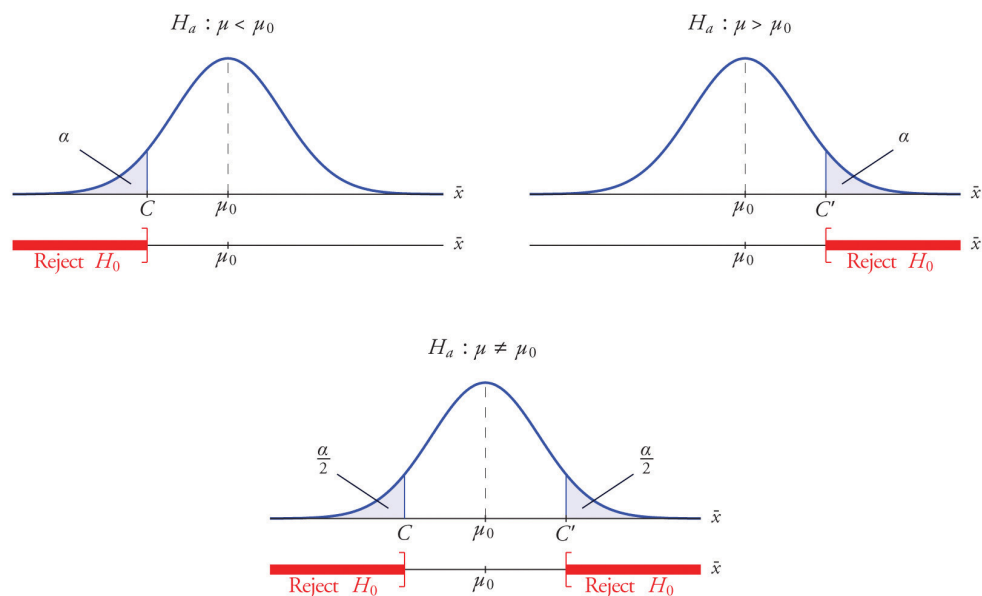
The number α is the total area of a tail or a pair of tails.

EXAMPLE 3

In the context of <u>Note 8.9 "Example 2"</u>, suppose that it is known that the population is normally distributed with standard deviation $\sigma = 0.15$ gram, and suppose that the test of hypotheses $H_0 : \mu = 8.0$ versus $H_a : \mu \neq 8.0$ will be performed with a sample of size 5. Construct the rejection region for the test for the choice $\alpha = 0.10.$ Explain the decision procedure and interpret it.

Solution:

If H_0 is true then the sample mean $\overline{X}$ is normally distributed with mean and standard deviation

$$\mu_{\overline{X}} = \mu = 8.0, \quad \sigma_{\overline{X}} = \sigma / \sqrt{n} = \frac{0.15}{\sqrt{5}} = 0.067$$

Since H_a contains the $\neq$ symbol the rejection region will be in two pieces, each one corresponding to a tail of area $\alpha / 2 = 0.10 / 2 = 0.05.$ From <u>Figure 12.3 "Critical Values of "</u>, $z_{0.05} = 1.645$, so C and C' are 1.645 standard deviations of $\overline{X}$ to the right and left of its mean 8.0:

$C = 8.0 - (1.645)(0.067) = 7.89$ and $C' = 8.0 + (1.645)(0.067) = 8.11$

The result is shown in <u>Figure 8.3 "Rejection Region for the Choice "</u>.

Figure 8.3
Rejection Region for
the Choice $\alpha = 0.10$
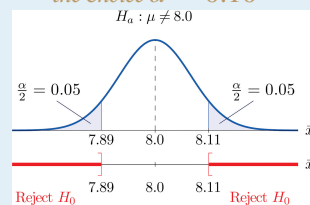
The decision procedure is: take a sample of size 5 and compute the sample mean $\overline{x}$. If $\overline{x}$ is either 7.89 grams or less or 8.11 grams or more then reject the hypothesis that the average amount of fat in *all* servings of the product

is 8.0 grams in favor of the alternative that it is different from 8.0 grams. Otherwise do not reject the hypothesis that the average amount is 8.0 grams.

The reasoning is that if the true average amount of fat per serving were 8.0 grams then there would be less than a 10% chance that a sample of size 5 would produce a mean of either 7.89 grams or less or 8.11 grams or more. Hence if that happened it would be more likely that the value 8.0 is incorrect (always assuming that the population standard deviation is 0.15 gram).

Because the rejection regions are computed based on areas in tails of distributions, as shown in Figure 8.2, hypothesis tests are classified according to the form of the alternative hypothesis in the following way.

Definition

If H_a has the form $\mu \neq \mu_0$ the test is called a **two-tailed test.**

If H_a has the form $\mu < \mu_0$ the test is called a **left-tailed test.**

If H_a has the form $\mu > \mu_0$ the test is called a **right-tailed test.**

Each of the last two forms is also called a **one-tailed test.**

Two Types of Errors

The format of the testing procedure in general terms is to take a sample and use the information it contains to come to a decision about the two hypotheses. As stated before our decision will always be either

1. reject the null hypothesis H_0 in favor of the alternative H_a presented, or
2. do not reject the null hypothesis H_0 in favor of the alternative H_a presented.

There are four possible outcomes of hypothesis testing procedure, as shown in the following table:

		True State of Nature	
		H_0 is true	H_0 is false
Our Decision	Do not reject H_0	Correct decision	Type II error
	Reject H_0	Type I error	Correct decision

As the table shows, there are two ways to be right and two ways to be wrong. Typically to reject H_0 when it is actually true is a more serious error than to fail to reject it when it is false, so the former error is labeled "Type I" and the latter error "Type II."

Definition

*In a test of hypotheses, a **Type I error**[7] is the decision to reject H_0 when it is in fact true. A **Type II error**[8] is the decision not to reject H_0 when it is in fact not true.*

Unless we perform a census we do not have certain knowledge, so we do not know whether our decision matches the true state of nature or if we have made an error. We reject H_0 if what we observe would be a "rare" event if H_0 were true. But rare events are not impossible: they occur with probability α. Thus when H_0 is true, a rare event will be observed in the proportion α of repeated similar tests, and H_0 will be erroneously rejected in those tests. Thus α is the probability that in following the testing procedure to decide between H_0 and H_a we will make a Type I error.

Definition

*The number α that is used to determine the rejection region is called the **level of significance of the test**[9]. It is the probability that the test procedure will result in a Type I error.*

---

7. Rejection of a true null hypothesis.

8. Failure to reject a false null hypothesis.

9. The probability α that defines an event as "rare;" the probability that the test procedure will lead to a Type I error.

The probability of making a Type II error is too complicated to discuss in a beginning text, so we will say no more about it than this: for a fixed sample size, choosing α smaller in order to reduce the chance of making a Type I error has the effect of increasing the chance of making a Type II error. The only way to simultaneously reduce the chances of making either kind of error is to increase the sample size.

Standardizing the Test Statistic

Hypotheses testing will be considered in a number of contexts, and great unification as well as simplification results when the relevant sample statistic is *standardized* by subtracting its mean from it and then dividing by its standard deviation. The resulting statistic is called a *standardized test statistic*. In every situation treated in this and the following two chapters the standardized test statistic will have either the standard normal distribution or Student's *t*-distribution.

Definition

A **standardized test statistic**[10] *for a hypothesis test is the statistic that is formed by subtracting from the statistic of interest its mean and dividing by its standard deviation.*

For example, reviewing Note 8.14 "Example 3", if instead of working with the sample mean $\overline{X}$ we instead work with the test statistic

$$\frac{\overline{X} - 8.0}{0.067}$$

then the distribution involved is standard normal and the critical values are just $\pm z_{0.05}$. The extra work that was done to find that $C = 7.89$ and $C' = 8.11$ is eliminated. In every hypothesis test in this book the standardized test statistic will be governed by either the standard normal distribution or Student's *t*-distribution. Information about rejection regions is summarized in the following tables:

When the test statistic has the standard normal distribution:		
Symbol in H_a	Terminology	Rejection Region

10. The standardized statistic used in performing the test.

<	Left-tailed test	$\left(-\infty, -z_\alpha\right]$
>	Right-tailed test	$\left[z_\alpha, \infty\right)$
≠	Two-tailed test	$\left(-\infty, -z_{\alpha/2}\right] \cup \left[z_{\alpha/2}, \infty\right)$

When the test statistic has Student's t-distribution:		
Symbol in H_a	Terminology	Rejection Region
<	Left-tailed test	$\left(-\infty, -t_\alpha\right]$
>	Right-tailed test	$\left[t_\alpha, \infty\right)$
≠	Two-tailed test	$\left(-\infty, -t_{\alpha/2}\right] \cup \left[t_{\alpha/2}, \infty\right)$

Every instance of hypothesis testing discussed in this and the following two chapters will have a rejection region like one of the six forms tabulated in the tables above.

No matter what the context a test of hypotheses can always be performed by applying the following systematic procedure, which will be illustrated in the examples in the succeeding sections.

Systematic Hypothesis Testing Procedure: Critical Value Approach

1. Identify the null and alternative hypotheses.
2. Identify the relevant test statistic and its distribution.
3. Compute from the data the value of the test statistic.
4. Construct the rejection region.
5. Compare the value computed in Step 3 to the rejection region constructed in Step 4 and make a decision. Formulate the decision in the context of the problem, if applicable.

The procedure that we have outlined in this section is called the "Critical Value Approach" to hypothesis testing to distinguish it from an alternative but equivalent approach that will be introduced at the end of Section 8.3 "The Observed Significance of a Test".

KEY TAKEAWAYS

- A test of hypotheses is a statistical process for deciding between two competing assertions about a population parameter.
- The testing procedure is formalized in a five-step procedure.

EXERCISES

1. State the null and alternative hypotheses for each of the following situations. (That is, identify the correct number μ_0 and write $H_0 : \mu = \mu_0$ and the appropriate analogous expression for H_a.)

   a. The average July temperature in a region historically has been 74.5°F. Perhaps it is higher now.
   b. The average weight of a female airline passenger with luggage was 145 pounds ten years ago. The FAA believes it to be higher now.
   c. The average stipend for doctoral students in a particular discipline at a state university is $14,756. The department chairman believes that the national average is higher.
   d. The average room rate in hotels in a certain region is $82.53. A travel agent believes that the average in a particular resort area is different.
   e. The average farm size in a predominately rural state was 69.4 acres. The secretary of agriculture of that state asserts that it is less today.

2. State the null and alternative hypotheses for each of the following situations. (That is, identify the correct number μ_0 and write $H_0 : \mu = \mu_0$ and the appropriate analogous expression for H_a.)

   a. The average time workers spent commuting to work in Verona five years ago was 38.2 minutes. The Verona Chamber of Commerce asserts that the average is less now.
   b. The mean salary for all men in a certain profession is $58,291. A special interest group thinks that the mean salary for women in the same profession is different.
   c. The accepted figure for the caffeine content of an 8-ounce cup of coffee is 133 mg. A dietitian believes that the average for coffee served in a local restaurants is higher.
   d. The average yield per acre for all types of corn in a recent year was 161.9 bushels. An economist believes that the average yield per acre is different this year.
   e. An industry association asserts that the average age of all self-described fly fishermen is 42.8 years. A sociologist suspects that it is higher.

3. Describe the two types of errors that can be made in a test of hypotheses.

4. Under what circumstance is a test of hypotheses certain to yield a correct decision?

ANSWERS

1.
   a. $H_0 : \mu = 74.5$ vs. $H_a : \mu > 74.5$
   b. $H_0 : \mu = 145$ vs. $H_a : \mu > 145$
   c. $H_0 : \mu = 14756$ vs. $H_a : \mu > 14756$
   d. $H_0 : \mu = 82.53$ vs. $H_a : \mu \neq 82.53$
   e. $H_0 : \mu = 69.4$ vs. $H_a : \mu < 69.4$

3. A Type I error is made when a true H_0 is rejected. A Type II error is made when a false H_0 is not rejected.

8.2 Large Sample Tests for a Population Mean

LEARNING OBJECTIVES

1. To learn how to apply the five-step test procedure for a test of hypotheses concerning a population mean when the sample size is large.
2. To learn how to interpret the result of a test of hypotheses in the context of the original narrated situation.

In this section we describe and demonstrate the procedure for conducting a test of hypotheses about the mean of a population in the case that the sample size n is at least 30. The Central Limit Theorem states that $\overline{X}$ is approximately normally distributed, and has mean $\mu_{\overline{X}} = \mu$ and standard deviation $\sigma_{\overline{X}} = \sigma / \sqrt{n}$, where μ and σ are the mean and the standard deviation of the population. This implies that the statistic

$$\frac{\overline{x} - \mu}{\sigma / \sqrt{n}}$$

has the standard normal distribution, which means that probabilities related to it are given in Figure 12.2 "Cumulative Normal Probability" and the last line in Figure 12.3 "Critical Values of ".

If we know σ then the statistic in the display is our test statistic. If, as is typically the case, we do not know σ, then we replace it by the sample standard deviation s. Since the sample is large the resulting test statistic still has a distribution that is approximately standard normal.

> ## Standardized Test Statistics for Large Sample Hypothesis Tests Concerning a Single Population Mean
>
> $$\text{If } \sigma \text{ is known: } \quad Z = \frac{\overline{x} - \mu_0}{\sigma \big/ \sqrt{n}}$$
>
> $$\text{If } \sigma \text{ is unknown: } \quad Z = \frac{\overline{x} - \mu_0}{s \big/ \sqrt{n}}$$
>
> The test statistic has the standard normal distribution.

The distribution of the standardized test statistic and the corresponding rejection region for each form of the alternative hypothesis (left-tailed, right-tailed, or two-tailed), is shown in Figure 8.4 "Distribution of the Standardized Test Statistic and the Rejection Region".

Figure 8.4 *Distribution of the Standardized Test Statistic and the Rejection Region*
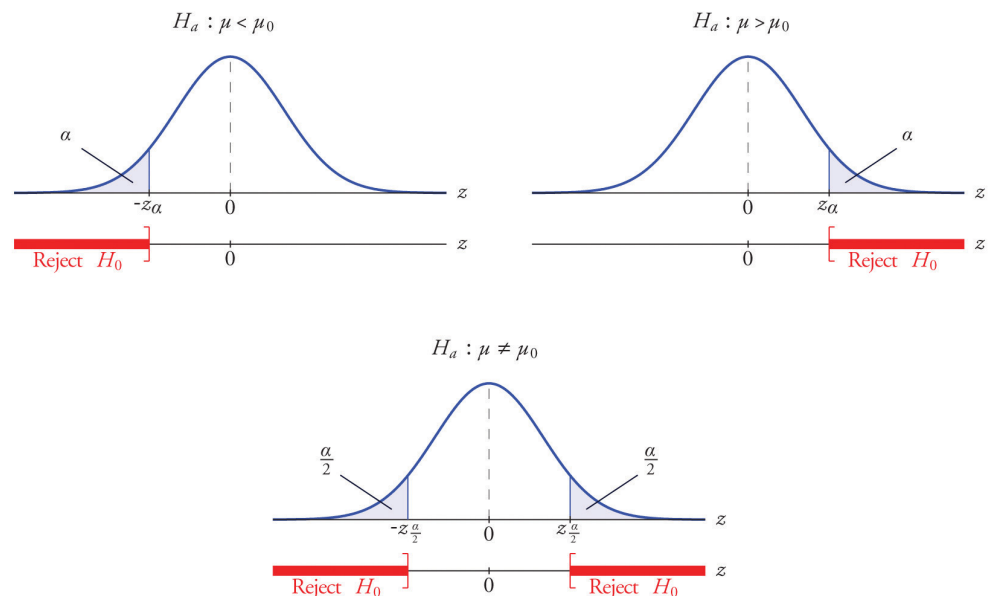
EXAMPLE 4

It is hoped that a newly developed pain reliever will more quickly produce perceptible reduction in pain to patients after minor surgeries than a standard pain reliever. The standard pain reliever is known to bring relief in an average of 3.5 minutes with standard deviation 2.1 minutes. To test whether the new pain reliever works more quickly than the standard one, 50 patients with minor surgeries were given the new pain reliever and their times to relief were recorded. The experiment yielded sample mean $\bar{x} = 3.1$ minutes and sample standard deviation $s = 1.5$ minutes. Is there sufficient evidence in the sample to indicate, at the 5% level of significance, that the newly developed pain reliever does deliver perceptible relief more quickly?

Solution:

We perform the test of hypotheses using the five-step procedure given at the end of Section 8.1 "The Elements of Hypothesis Testing".

- Step 1. The natural assumption is that the new drug is no better than the old one, but must be proved to be better. Thus if μ denotes the average time until all patients who are given the new drug experience pain relief, the hypothesis test is

$$H_0 : \mu = 3.5$$
$$\text{vs.} \, H_a : \mu < 3.5 \quad @ \, \alpha = 0.05$$

- Step 2. The sample is large, but the population standard deviation is unknown (the 2.1 minutes pertains to the old drug, not the new one). Thus the test statistic is

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

and has the standard normal distribution.

- Step 3. Inserting the data into the formula for the test statistic gives

$$Z = \frac{\overline{x} - \mu_0}{s / \sqrt{n}} = \frac{3.1 - 3.5}{1.5 / \sqrt{50}} = -1.886$$

- Step 4. Since the symbol in H_a is "<" this is a left-tailed test, so there is a single critical value, $-z_\alpha = -z_{0.05}$, which from the last line in <u>Figure 12.3 "Critical Values of "</u> we read off as -1.645. The rejection region is $\left(-\infty, -1.645\right]$.

  - Step 5. As shown in <u>Figure 8.5 "Rejection Region and Test Statistic for "</u> the test statistic falls in the rejection region. The decision is to reject H_0. In the context of the problem our conclusion is:

    The data provide sufficient evidence, at the 5% level of significance, to conclude that the average time until patients experience perceptible relief from pain using the new pain reliever is smaller than the average time for the standard pain reliever.

Figure 8.5
Rejection Region and Test Statistic for <u>Note 8.27 "Example 4"</u>

$H_a : \mu < 3.5$

$\alpha = 0.05$

$-z_\alpha = -1.645 \quad 0$

Reject H_0

$Z = -1.886$
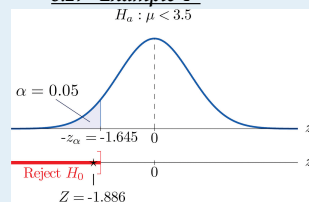
EXAMPLE 5

A cosmetics company fills its best-selling 8-ounce jars of facial cream by an automatic dispensing machine. The machine is set to dispense a mean of 8.1 ounces per jar. Uncontrollable factors in the process can shift the mean away from 8.1 and cause either underfill or overfill, both of which are undesirable. In such a case the dispensing machine is stopped and recalibrated. Regardless of the mean amount dispensed, the standard deviation of the amount dispensed always has value 0.22 ounce. A quality control engineer routinely selects 30 jars from the assembly line to check the amounts filled. On one occasion, the sample mean is $\bar{x} = 8.2$ ounces and the sample standard deviation is $s = 0.25$ ounce. Determine if there is sufficient evidence in the sample to indicate, at the 1% level of significance, that the machine should be recalibrated.

Solution:

- Step 1. The natural assumption is that the machine is working properly. Thus if μ denotes the mean amount of facial cream being dispensed, the hypothesis test is

$$H_0 : \mu = 8.1$$
$$\text{vs.} \, H_a : \mu \neq 8.1 \quad @ \, \alpha = 0.01$$

- Step 2. The sample is large and the population standard deviation is known. Thus the test statistic is

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$
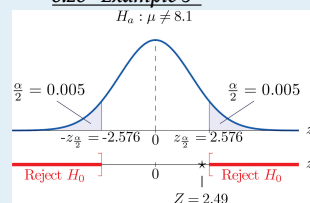
and has the standard normal distribution.

- Step 3. Inserting the data into the formula for the test statistic gives

$$Z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{8.2 - 8.1}{0.22 / \sqrt{30}} = 2.490$$

- Step 4. Since the symbol in H_a is "≠" this is a two-tailed test, so there are two critical values, $\pm z_{\alpha/2} = \pm z_{0.005}$, which from the last line in <u>Figure 12.3 "Critical Values of "</u> we read off as ± 2.576. The rejection region is $(-\infty, -2.576] \cup [2.576, \infty)$.

  - Step 5. As shown in <u>Figure 8.6 "Rejection Region and Test Statistic for "</u> the test statistic does not fall in the rejection region. The decision is not to reject H_0. In the context of the problem our conclusion is:

    The data do not provide sufficient evidence, at the 1% level of significance, to conclude that the average amount of product dispensed is different from 8.1 ounce. We conclude that the machine does not need to be recalibrated.

Figure 8.6

Rejection Region and Test Statistic for <u>*Note 8.28 "Example 5"*</u>

KEY TAKEAWAYS

- There are two formulas for the test statistic in testing hypotheses about a population mean with large samples. Both test statistics follow the standard normal distribution.
- The population standard deviation is used if it is known, otherwise the sample standard deviation is used.
- The same five-step procedure is used with either test statistic.

EXERCISES

BASIC

1. Find the rejection region (for the standardized test statistic) for each hypothesis test.

   a. $H_0 : \mu = 27$ vs. $H_a : \mu < 27$ @ $\alpha = 0.05$.
   b. $H_0 : \mu = 52$ vs. $H_a : \mu \neq 52$ @ $\alpha = 0.05$.
   c. $H_0 : \mu = -105$ vs. $H_a : \mu > -105$ @ $\alpha = 0.10$.
   d. $H_0 : \mu = 78.8$ vs. $H_a : \mu \neq 78.8$ @ $\alpha = 0.10$.

2. Find the rejection region (for the standardized test statistic) for each hypothesis test.

   a. $H_0 : \mu = 17$ vs. $H_a : \mu < 17$ @ $\alpha = 0.01$.
   b. $H_0 : \mu = 880$ vs. $H_a : \mu \neq 880$ @ $\alpha = 0.01$.
   c. $H_0 : \mu = -12$ vs. $H_a : \mu > -12$ @ $\alpha = 0.05$.
   d. $H_0 : \mu = 21.1$ vs. $H_a : \mu \neq 21.1$ @ $\alpha = 0.05$.

3. Find the rejection region (for the standardized test statistic) for each hypothesis test. Identify the test as left-tailed, right-tailed, or two-tailed.

   a. $H_0 : \mu = 141$ vs. $H_a : \mu < 141$ @ $\alpha = 0.20$.
   b. $H_0 : \mu = -54$ vs. $H_a : \mu < -54$ @ $\alpha = 0.05$.
   c. $H_0 : \mu = 98.6$ vs. $H_a : \mu \neq 98.6$ @ $\alpha = 0.05$.
   d. $H_0 : \mu = 3.8$ vs. $H_a : \mu > 3.8$ @ $\alpha = 0.001$.

4. Find the rejection region (for the standardized test statistic) for each hypothesis test. Identify the test as left-tailed, right-tailed, or two-tailed.

   a. $H_0 : \mu = -62$ vs. $H_a : \mu \neq -62$ @ $\alpha = 0.005$.
   b. $H_0 : \mu = 73$ vs. $H_a : \mu > 73$ @ $\alpha = 0.001$.
   c. $H_0 : \mu = 1124$ vs. $H_a : \mu < 1124$ @ $\alpha = 0.001$.
   d. $H_0 : \mu = 0.12$ vs. $H_a : \mu \neq 0.12$ @ $\alpha = 0.001$.

5. Compute the value of the test statistic for the indicated test, based on the information given.

   a. Testing $H_0 : \mu = 72.2$ vs. $H_a : \mu > 72.2$, σ unknown, $n = 55$, $\overline{x} = 75.1$, $s = 9.25$
   b. Testing $H_0 : \mu = 58$ vs. $H_a : \mu > 58$, $\sigma = 1.22$, $n = 40$, $\overline{x} = 58.5$, $s = 1.29$
   c. Testing $H_0 : \mu = -19.5$ vs. $H_a : \mu < -19.5$, σ unknown, $n = 30$, $\overline{x} = -23.2$, $s = 9.55$

    d. Testing $H_0 : \mu = 805$ vs. $H_a : \mu \neq 805$, $\sigma = 37.5$, $n = 75$, $\overline{x} = 818$, $s = 36.2$

6. Compute the value of the test statistic for the indicated test, based on the information given.

    a. Testing $H_0 : \mu = 342$ vs. $H_a : \mu < 342$, $\sigma = 11.2$, $n = 40$, $\overline{x} = 339$, $s = 10.3$

    b. Testing $H_0 : \mu = 105$ vs. $H_a : \mu > 105$, $\sigma = 5.3$, $n = 80$, $\overline{x} = 107$, $s = 5.1$

    c. Testing $H_0 : \mu = -13.5$ vs. $H_a : \mu \neq -13.5$, σ unknown, $n = 32$, $\overline{x} = -13.8$, $s = 1.5$

    d. Testing $H_0 : \mu = 28$ vs. $H_a : \mu \neq 28$, σ unknown, $n = 68$, $\overline{x} = 27.8$, $s = 1.3$

7. Perform the indicated test of hypotheses, based on the information given.

    a. Test $H_0 : \mu = 212$ vs. $H_a : \mu < 212$ @ $\alpha = 0.10$, σ unknown, $n = 36$, $\overline{x} = 211.2$, $s = 2.2$

    b. Test $H_0 : \mu = -18$ vs. $H_a : \mu > -18$ @ $\alpha = 0.05$, $\sigma = 3.3$, $n = 44$, $\overline{x} = -17.2$, $s = 3.1$

    c. Test $H_0 : \mu = 24$ vs. $H_a : \mu \neq 24$ @ $\alpha = 0.02$, σ unknown, $n = 50$, $\overline{x} = 22.8$, $s = 1.9$

8. Perform the indicated test of hypotheses, based on the information given.

    a. Test $H_0 : \mu = 105$ vs. $H_a : \mu > 105$ @ $\alpha = 0.05$, σ unknown, $n = 30$, $\overline{x} = 108$, $s = 7.2$

    b. Test $H_0 : \mu = 21.6$ vs. $H_a : \mu < 21.6$ @ $\alpha = 0.01$, σ unknown, $n = 78$, $\overline{x} = 20.5$, $s = 3.9$

    c. Test $H_0 : \mu = -375$ vs. $H_a : \mu \neq -375$ @ $\alpha = 0.01$, $\sigma = 18.5$, $n = 31$, $\overline{x} = -388$, $s = 18.0$

APPLICATIONS

9. In the past the average length of an outgoing telephone call from a business office has been 143 seconds. A manager wishes to check whether that average has decreased after the introduction of policy changes. A sample of 100 telephone calls produced a mean of 133 seconds, with a standard deviation of 35 seconds. Perform the relevant test at the 1% level of significance.

10. The government of an impoverished country reports the mean age at death among those who have survived to adulthood as 66.2 years. A relief agency examines 30 randomly selected deaths and obtains a mean of 62.3 years with

standard deviation 8.1 years. Test whether the agency's data support the alternative hypothesis, at the 1% level of significance, that the population mean is less than 66.2.

11. The average household size in a certain region several years ago was 3.14 persons. A sociologist wishes to test, at the 5% level of significance, whether it is different now. Perform the test using the information collected by the sociologist: in a random sample of 75 households, the average size was 2.98 persons, with sample standard deviation 0.82 person.

12. The recommended daily calorie intake for teenage girls is 2,200 calories/day. A nutritionist at a state university believes the average daily caloric intake of girls in that state to be lower. Test that hypothesis, at the 5% level of significance, against the null hypothesis that the population average is 2,200 calories/day using the following sample data: $n = 36, \overline{x} = 2{,}150, s = 203$.

13. An automobile manufacturer recommends oil change intervals of 3,000 miles. To compare actual intervals to the recommendation, the company randomly samples records of 50 oil changes at service facilities and obtains sample mean 3,752 miles with sample standard deviation 638 miles. Determine whether the data provide sufficient evidence, at the 5% level of significance, that the population mean interval between oil changes exceeds 3,000 miles.

14. A medical laboratory claims that the mean turn-around time for performance of a battery of tests on blood samples is 1.88 business days. The manager of a large medical practice believes that the actual mean is larger. A random sample of 45 blood samples yielded mean 2.09 and sample standard deviation 0.13 day. Perform the relevant test at the 10% level of significance, using these data.

15. A grocery store chain has as one standard of service that the mean time customers wait in line to begin checking out not exceed 2 minutes. To verify the performance of a store the company measures the waiting time in 30 instances, obtaining mean time 2.17 minutes with standard deviation 0.46 minute. Use these data to test the null hypothesis that the mean waiting time is 2 minutes versus the alternative that it exceeds 2 minutes, at the 10% level of significance.

16. A magazine publisher tells potential advertisers that the mean household income of its regular readership is $61,500. An advertising agency wishes to test this claim against the alternative that the mean is smaller. A sample of 40 randomly selected regular readers yields mean income $59,800 with standard deviation $5,850. Perform the relevant test at the 1% level of significance.

17. Authors of a computer algebra system wish to compare the speed of a new computational algorithm to the currently implemented algorithm. They apply

the new algorithm to 50 standard problems; it averages 8.16 seconds with standard deviation 0.17 second. The current algorithm averages 8.21 seconds on such problems. Test, at the 1% level of significance, the alternative hypothesis that the new algorithm has a lower average time than the current algorithm.

18. A random sample of the starting salaries of 35 randomly selected graduates with bachelor's degrees last year gave sample mean and standard deviation $41,202 and $7,621, respectively. Test whether the data provide sufficient evidence, at the 5% level of significance, to conclude that the mean starting salary of all graduates last year is less than the mean of all graduates two years before, $43,589.

ADDITIONAL EXERCISES

19. The mean household income in a region served by a chain of clothing stores is $48,750. In a sample of 40 customers taken at various stores the mean income of the customers was $51,505 with standard deviation $6,852.

   a. Test at the 10% level of significance the null hypothesis that the mean household income of customers of the chain is $48,750 against that alternative that it is different from $48,750.
   b. The sample mean is greater than $48,750, suggesting that the actual mean of people who patronize this store is greater than $48,750. Perform this test, also at the 10% level of significance. (The computation of the test statistic done in part (a) still applies here.)

20. The labor charge for repairs at an automobile service center are based on a standard time specified for each type of repair. The time specified for replacement of universal joint in a drive shaft is one hour. The manager reviews a sample of 30 such repairs. The average of the actual repair times is 0.86 hour with standard deviation 0.32 hour.

   a. Test at the 1% level of significance the null hypothesis that the actual mean time for this repair differs from one hour.
   b. The sample mean is less than one hour, suggesting that the mean actual time for this repair is less than one hour. Perform this test, also at the 1% level of significance. (The computation of the test statistic done in part (a) still applies here.)

LARGE DATA SET EXERCISES

21. Large Data Set 1 records the SAT scores of 1,000 students. Regarding it as a random sample of all high school students, use it to test the hypothesis that the population mean exceeds 1,510, at the 1% level of significance. (The null hypothesis is that $\mu = 1510$.)

    http://www.gone.2012books.lardbucket.org/sites/all/files/data1.xls

22. Large Data Set 1 records the GPAs of 1,000 college students. Regarding it as a random sample of all college students, use it to test the hypothesis that the population mean is less than 2.50, at the 10% level of significance. (The null hypothesis is that $\mu = 2.50$.)

    http://www.gone.2012books.lardbucket.org/sites/all/files/data1.xls

23. Large Data Set 1 lists the SAT scores of 1,000 students.

    http://www.gone.2012books.lardbucket.org/sites/all/files/data1.xls

    a. Regard the data as arising from a census of all students at a high school, in which the SAT score of every student was measured. Compute the population mean μ.
    b. Regard the first 50 students in the data set as a random sample drawn from the population of part (a) and use it to test the hypothesis that the population mean exceeds 1,510, at the 10% level of significance. (The null hypothesis is that $\mu = 1510$.)
    c. Is your conclusion in part (b) in agreement with the true state of nature (which by part (a) you know), or is your decision in error? If your decision is in error, is it a Type I error or a Type II error?

24. Large Data Set 1 lists the GPAs of 1,000 students.

    http://www.gone.2012books.lardbucket.org/sites/all/files/data1.xls

    a. Regard the data as arising from a census of all freshman at a small college at the end of their first academic year of college study, in which the GPA of every such person was measured. Compute the population mean μ.
    b. Regard the first 50 students in the data set as a random sample drawn from the population of part (a) and use it to test the hypothesis that the population mean is less than 2.50, at the 10% level of significance. (The null hypothesis is that $\mu = 2.50$.)
    c. Is your conclusion in part (b) in agreement with the true state of nature (which by part (a) you know), or is your decision in error? If your decision is in error, is it a Type I error or a Type II error?

ANSWERS

1.
   a. $Z \leq -1.645$
   b. $Z \leq -1.96$ or $Z \geq 1.96$
   c. $Z \geq 1.28$
   d. $Z \leq -1.645$ or $Z \geq 1.645$

3.
   a. $Z \leq -0.84$
   b. $Z \leq -1.645$
   c. $Z \leq -1.96$ or $Z \geq 1.96$
   d. $Z \geq 3.1$

5.
   a. $Z = 2.235$
   b. $Z = 2.592$
   c. $Z = -2.122$
   d. $Z = 3.002$

7.
   a. $Z = -2.18, -z_{0.10} = -1.28$, reject H_0.
   b. $Z = 1.61, z_{0.05} = 1.645$, do not reject H_0.
   c. $Z = -4.47, -z_{0.01} = -2.33$, reject H_0.

9. $Z = -2.86, -z_{0.01} = -2.33$, reject H_0.

11. $Z = -1.69, -z_{0.025} = -1.96$, do not reject H_0.

13. $Z = 8.33, z_{0.05} = 1.645$, reject H_0.

15. $Z = 2.02, z_{0.10} = 1.28$, reject H_0.

17. $Z = -2.08, -z_{0.01} = -2.33$, do not reject H_0.

19.
   a. $Z = 2.54, z_{0.05} = 1.645$, reject H_0;
   b. $Z = 2.54, z_{0.10} = 1.28$, reject H_0.

21. $H_0 : \mu = 1510$ vs. $H_a : \mu > 1510$. Test Statistic: $Z = 2.7882$. Rejection Region: $[2.33, \infty)$. Decision: Reject H_0.

23.
   a. $\mu_0 = 1528.74$
   b. $H_0 : \mu = 1510$ vs. $H_a : \mu > 1510$. Test Statistic: $Z = -1.41$. Rejection Region: $[1.28, \infty)$. Decision: Fail to reject H_0.
   c. No, it is a Type II error.

8.3 The Observed Significance of a Test

<div style="border:1px solid #ccc; background:#e9e7dd; padding:1em;">

LEARNING OBJECTIVES

1. To learn what the observed significance of a test is.
2. To learn how to compute the observed significance of a test.
3. To learn how to apply the *p*-value approach to hypothesis testing.

</div>

The Observed Significance

The conceptual basis of our testing procedure is that we reject H_0 only if the data that we obtained would constitute a rare event if H_0 were actually true. The level of significance α specifies what is meant by "rare." The *observed significance* of the test is a measure of how rare the value of the test statistic that we have just observed would be if the null hypothesis were true. That is, the *observed significance* of the test just performed is the probability that, if the test were repeated with a new sample, the result of the new test would be at least as contrary to H_0 and in support of H_a as what was observed in the original test.

<div style="border:1px solid #ccc; background:#f0f0f0; padding:1em;">

Definition

*The **observed significance or p-value**[11] of a specific test of hypotheses is the probability, on the supposition that H_0 is true, of obtaining a result at least as contrary to H_0 and in favor of H_a as the result actually observed in the sample data.*

</div>

Think back to Note 8.27 "Example 4" in Section 8.2 "Large Sample Tests for a Population Mean" concerning the effectiveness of a new pain reliever. This was a left-tailed test in which the value of the test statistic was –1.886. To be as contrary to H_0 and in support of H_a as the result $Z = -1.886$ actually observed means to obtain a value of the test statistic in the interval $\left(-\infty, -1.886\right]$. Rounding –1.886 to –1.89, we can read directly from Figure 12.2 "Cumulative Normal Probability" that $P\left(Z \leq -1.89\right) = 0.0294.$ Thus the *p*-value or observed significance of the test in Note 8.27 "Example 4" is 0.0294 or about 3%. Under repeated sampling from this population, if H_0 were true then only about 3% of all samples of size 50 would give a result as contrary to H_0 and in favor of H_a as the sample we observed. Note

11. The probability, if H_0 is true, of obtaining a result as contrary to H_0 and in favor of H_a as the result observed in the sample data.

that the probability 0.0294 is the area of the left tail cut off by the test statistic in this left-tailed test.

Analogous reasoning applies to a right-tailed or a two-tailed test, except that in the case of a two-tailed test being as far from 0 as the observed value of the test statistic but on the opposite side of 0 is just as contrary to H_0 as being the same distance away and on the same side of 0, hence the corresponding tail area is doubled.

Computational Definition of the Observed Significance of a Test of Hypotheses

The **observed significance** of a test of hypotheses is the area of the tail of the distribution cut off by the test statistic (times two in the case of a two-tailed test).
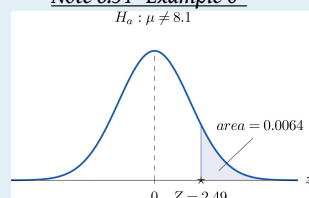
EXAMPLE 6

Compute the observed significance of the test performed in Note 8.28 "Example 5" in Section 8.2 "Large Sample Tests for a Population Mean".

Solution:

The value of the test statistic was $z = 2.490$, which by Figure 12.2 "Cumulative Normal Probability" cuts off a tail of area 0.0064, as shown in Figure 8.7 "Area of the Tail for ". Since the test was two-tailed, the observed significance is $2 \times 0.0064 = 0.0128$.

Figure 8.7
Area of the Tail for
Note 8.34 "Example 6"

$H_a : \mu \neq 8.1$

$area = 0.0064$

$0 \quad Z = 2.49$

The *p*-value Approach to Hypothesis Testing

In <u>Note 8.27 "Example 4"</u> in <u>Section 8.2 "Large Sample Tests for a Population Mean"</u> the test was performed at the 5% level of significance: the definition of "rare" event was probability $\alpha = 0.05$ or less. We saw above that the observed significance of the test was $p = 0.0294$ or about 3%. Since $p = 0.0294 < 0.05 = \alpha$ (or 3% is less than 5%), the decision turned out to be to reject: what was observed was sufficiently unlikely to qualify as an event so rare as to be regarded as (practically) incompatible with H_0.

In <u>Note 8.28 "Example 5"</u> in <u>Section 8.2 "Large Sample Tests for a Population Mean"</u> the test was performed at the 1% level of significance: the definition of "rare" event was probability $\alpha = 0.01$ or less. The observed significance of the test was computed in <u>Note 8.34 "Example 6"</u> as $p = 0.0128$ or about 1.3%. Since $p = 0.0128 > 0.01 = \alpha$ (or 1.3% is greater than 1%), the decision turned out to be not to reject. The event observed was unlikely, but not sufficiently unlikely to lead to rejection of the null hypothesis.

The reasoning just presented is the basis for a slightly different but equivalent formulation of the hypothesis testing process. The first three steps are the same as before, but instead of using α to compute critical values and construct a rejection region, one computes the p-value p of the test and compares it to α, rejecting H_0 if $p \leq \alpha$ and not rejecting if $p > \alpha$.

> ## Systematic Hypothesis Testing Procedure: *p*-Value Approach
>
> 1. Identify the null and alternative hypotheses.
> 2. Identify the relevant test statistic and its distribution.
> 3. Compute from the data the value of the test statistic.
> 4. Compute the p-value of the test.
> 5. Compare the value computed in Step 4 to significance level α and make a decision: reject H_0 if $p \leq \alpha$ and do not reject H_0 if $p > \alpha$.
>   Formulate the decision in the context of the problem, if applicable.

EXAMPLE 7

The total score in a professional basketball game is the sum of the scores of the two teams. An expert commentator claims that the average total score for NBA games is 202.5. A fan suspects that this is an overstatement and that the actual average is less than 202.5. He selects a random sample of 85 games and obtains a mean total score of 199.2 with standard deviation 19.63. Determine, at the 5% level of significance, whether there is sufficient evidence in the sample to reject the expert commentator's claim.

Solution:

- Step 1. Let μ be the true average total game score of all NBA games. The relevant test is

$$H_0 : \mu = 202.5$$
$$\text{vs.} H_a : \mu < 202.5 \quad @ \alpha = 0.05$$

- Step 2. The sample is large and the population standard deviation is unknown. Thus the test statistic is

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

and has the standard normal distribution.

- Step 3. Inserting the data into the formula for the test statistic gives

$$Z = \frac{\bar{x} - \mu_0}{s \big/ \sqrt{n}} = \frac{199.2 - 202.5}{19.63 \big/ \sqrt{85}} = -1.55$$

- Step 4. The area of the left tail cut off by $z = -1.55$ is, by <u>Figure 12.2 "Cumulative Normal Probability"</u>, 0.0606, as illustrated in <u>Figure 8.8 "Test Statistic for "</u>. Since the test is left-tailed, the p-value is just this number, $p = 0.0606$.

- Step 5. Since $p = 0.0606 > 0.05 = \alpha$, the decision is not to reject H_0. In the context of the problem our conclusion is:
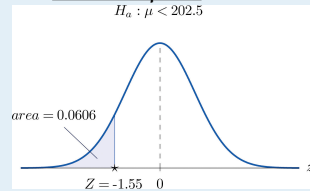
  The data do not provide sufficient evidence, at the 5% level of significance, to conclude that the average total score of NBA games is less than 202.5.

Figure 8.8
Test Statistic for Note
8.36 "Example 7"

EXAMPLE 8

Mr. Prospero has been teaching Algebra II from a particular textbook at Remote Isle High School for many years. Over the years students in his Algebra II classes have consistently scored an average of 67 on the end of course exam (EOC). This year Mr. Prospero used a new textbook in the hope that the average score on the EOC test would be higher. The average EOC test score of the 64 students who took Algebra II from Mr. Prospero this year had mean 69.4 and sample standard deviation 6.1. Determine whether these data provide sufficient evidence, at the 1% level of significance, to conclude that the average EOC test score is higher with the new textbook.

Solution:

- Step 1. Let μ be the true average score on the EOC exam of all Mr. Prospero's students who take the Algebra II course with the new textbook. The natural statement that would be assumed true unless there were strong evidence to the contrary is that the new book is about the same as the old one. The alternative, which it takes evidence to establish, is that the new book is better, which corresponds to a higher value of μ. Thus the relevant test is

$$H_0 : \mu = 67$$
$$\text{vs.} \, H_a : \mu > 67 \quad @ \, \alpha = 0.01$$

- Step 2. The sample is large and the population standard deviation is unknown. Thus the test statistic is

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

and has the standard normal distribution.

- Step 3. Inserting the data into the formula for the test statistic gives

$$Z = \frac{\bar{x} - \mu_0}{s \big/ \sqrt{n}} = \frac{69.4 - 67}{6.1 \big/ \sqrt{64}} = 3.15$$

- Step 4. The area of the right tail cut off by $z = 3.15$ is, by <u>Figure 12.2 "Cumulative Normal Probability"</u>, $1 - 0.9992 = 0.0008$, as shown in <u>Figure 8.9 "Test Statistic for "</u>. Since the test is right-tailed, the p-value is just this number, $p = 0.0008$.

  - Step 5. Since $p = 0.0008 < 0.01 = \alpha$, the decision is to reject H_0. In the context of the problem our conclusion is:

    The data provide sufficient evidence, at the 1% level of significance, to conclude that the average EOC exam score of students taking the Algebra II course from Mr. Prospero using the new book is higher than the average score of those taking the course from him but using the old book.

Figure 8.9
Test Statistic for <u>Note 8.37 "Example 8"</u>
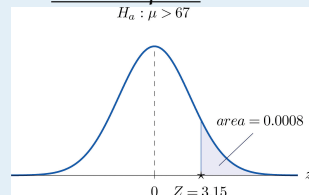
EXAMPLE 9

For the surface water in a particular lake, local environmental scientists would like to maintain an average pH level at 7.4. Water samples are routinely collected to monitor the average pH level. If there is evidence of a shift in pH value, in either direction, then remedial action will be taken. On a particular day 30 water samples are taken and yield average pH reading of 7.7 with sample standard deviation 0.5. Determine, at the 1% level of significance, whether there is sufficient evidence in the sample to indicate that remedial action should be taken.

Solution:

- Step 1. Let μ be the true average pH level at the time the samples were taken. The relevant test is

$$H_0 : \mu = 7.4$$
$$\text{vs.} \, H_a : \mu \neq 7.4 \quad @ \, \alpha = 0.01$$

- Step 2. The sample is large and the population standard deviation is unknown. Thus the test statistic is

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

and has the standard normal distribution.

- Step 3. Inserting the data into the formula for the test statistic gives

$$Z = \frac{\bar{x} - \mu_0}{s \big/ \sqrt{n}} = \frac{7.7 - 7.4}{0.5 \big/ \sqrt{30}} = 3.29$$

- Step 4. The area of the right tail cut off by $z = 3.29$ is, by <u>Figure 12.2</u> <u>"Cumulative Normal Probability"</u>, $1 - 0.9995 = 0.0005$, as illustrated in <u>Figure 8.10 "Test Statistic for "</u>. Since the test is two-tailed, the p-value is the double of this number,
$$p = 2 \times 0.0005 = 0.0010.$$

- Step 5. Since $p = 0.0010 < 0.01 = \alpha$, the decision is to reject H_0. In the context of the problem our conclusion is:
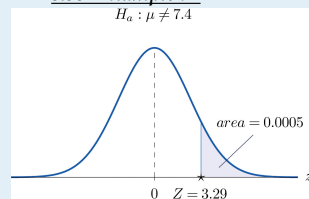
  The data provide sufficient evidence, at the 1% level of significance, to conclude that the average pH of surface water in the lake is different from 7.4. That is, remedial action is indicated.

Figure 8.10
Test Statistic for Note 8.38 "Example 9"

$H_a : \mu \neq 7.4$

$area = 0.0005$

$0 \quad Z = 3.29$

z

KEY TAKEAWAYS

- The observed significance or p-value of a test is a measure of how inconsistent the sample result is with H_0 and in favor of H_a.
- The p-value approach to hypothesis testing means that one merely compares the p-value to α instead of constructing a rejection region.
- There is a systematic five-step procedure for the p-value approach to hypothesis testing.

EXERCISES

BASIC

1. Compute the observed significance of each test.

   a. Testing $H_0 : \mu = 54.7$ vs. $H_a : \mu < 54.7$, test statistic $z = -1.72$.
   b. Testing $H_0 : \mu = 195$ vs. $H_a : \mu \neq 195$, test statistic $z = -2.07$.
   c. Testing $H_0 : \mu = -45$ vs. $H_a : \mu > -45$, test statistic $z = 2.54$.

2. Compute the observed significance of each test.

   a. Testing $H_0 : \mu = 0$ vs. $H_a : \mu \neq 0$, test statistic $z = 2.82$.
   b. Testing $H_0 : \mu = 18.4$ vs. $H_a : \mu < 18.4$, test statistic $z = -1.74$.
   c. Testing $H_0 : \mu = 63.85$ vs. $H_a : \mu > 63.85$, test statistic $z = 1.93$.

3. Compute the observed significance of each test. (Some of the information given might not be needed.)

   a. Testing $H_0 : \mu = 27.5$ vs. $H_a : \mu > 27.5$; $n = 49, \overline{x} = 28.9, s = 3.14$, test statistic $z = 3.12$.
   b. Testing $H_0 : \mu = 581$ vs. $H_a : \mu < 581$; $n = 32, \overline{x} = 560, s = 47.8$, test statistic $z = -2.49$.
   c. Testing $H_0 : \mu = 138.5$ vs. $H_a : \mu \neq 138.5$; $n = 44, \overline{x} = 137.6, s = 2.45$, test statistic $z = -2.44$.

4. Compute the observed significance of each test. (Some of the information given might not be needed.)

   a. Testing $H_0 : \mu = -17.9$ vs. $H_a : \mu < -17.9$; $n = 34, \overline{x} = -18.2, s = 0.87$, test statistic $z = -2.01$.
   b. Testing $H_0 : \mu = 5.5$ vs. $H_a : \mu \neq 5.5$; $n = 56, \overline{x} = 7.4, s = 4.82$, test statistic $z = 2.95$.
   c. Testing $H_0 : \mu = 1255$ vs. $H_a : \mu > 1255$; $n = 152, \overline{x} = 1257, s = 7.5$, test statistic $z = 3.29$.

5. Make the decision in each test, based on the information provided.

   a. Testing $H_0 : \mu = 82.9$ vs. $H_a : \mu < 82.9$ @ $\alpha = 0.05$, observed significance $p = 0.038$.
   b. Testing $H_0 : \mu = 213.5$ vs. $H_a : \mu \neq 213.5$ @ $\alpha = 0.01$, observed significance $p = 0.038$.

6. Make the decision in each test, based on the information provided.

a. Testing $H_0 : \mu = 31.4$ vs. $H_a : \mu > 31.4$ @ $\alpha = 0.10$, observed significance $p = 0.062$.

b. Testing $H_0 : \mu = -75.5$ vs. $H_a : \mu < -75.5$ @ $\alpha = 0.05$, observed significance $p = 0.062$.

APPLICATIONS

7. A lawyer believes that a certain judge imposes prison sentences for property crimes that are longer than the state average 11.7 months. He randomly selects 36 of the judge's sentences and obtains mean 13.8 and standard deviation 3.9 months.

    a. Perform the test at the 1% level of significance using the critical value approach.
    b. Compute the observed significance of the test.
    c. Perform the test at the 1% level of significance using the p-value approach. You need not repeat the first three steps, already done in part (a).

8. In a recent year the fuel economy of all passenger vehicles was 19.8 mpg. A trade organization sampled 50 passenger vehicles for fuel economy and obtained a sample mean of 20.1 mpg with standard deviation 2.45 mpg. The sample mean 20.1 exceeds 19.8, but perhaps the increase is only a result of sampling error.

    a. Perform the relevant test of hypotheses at the 20% level of significance using the critical value approach.
    b. Compute the observed significance of the test.
    c. Perform the test at the 20% level of significance using the p-value approach. You need not repeat the first three steps, already done in part (a).

9. The mean score on a 25-point placement exam in mathematics used for the past two years at a large state university is 14.3. The placement coordinator wishes to test whether the mean score on a revised version of the exam differs from 14.3. She gives the revised exam to 30 entering freshmen early in the summer; the mean score is 14.6 with standard deviation 2.4.

    a. Perform the test at the 10% level of significance using the critical value approach.
    b. Compute the observed significance of the test.
    c. Perform the test at the 10% level of significance using the p-value approach. You need not repeat the first three steps, already done in part (a).

10. The mean increase in word family vocabulary among students in a one-year foreign language course is 576 word families. In order to estimate the effect of a new type of class scheduling, an instructor monitors the progress of 60 students; the sample mean increase in word family vocabulary of these students is 542 word families with sample standard deviation 18 word families.

    a. Test at the 5% level of significance whether the mean increase with the new class scheduling is different from 576 word families, using the critical value approach.
    b. Compute the observed significance of the test.
    c. Perform the test at the 5% level of significance using the p-value approach. You need not repeat the first three steps, already done in part (a).

11. The mean yield for hard red winter wheat in a certain state is 44.8 bu/acre. In a pilot program a modified growing scheme was introduced on 35 independent plots. The result was a sample mean yield of 45.4 bu/acre with sample standard deviation 1.6 bu/acre, an apparent increase in yield.

    a. Test at the 5% level of significance whether the mean yield under the new scheme is greater than 44.8 bu/acre, using the critical value approach.
    b. Compute the observed significance of the test.
    c. Perform the test at the 5% level of significance using the p-value approach. You need not repeat the first three steps, already done in part (a).

12. The average amount of time that visitors spent looking at a retail company's old home page on the world wide web was 23.6 seconds. The company commissions a new home page. On its first day in place the mean time spent at the new page by 7,628 visitors was 23.5 seconds with standard deviation 5.1 seconds.

    a. Test at the 5% level of significance whether the mean visit time for the new page is less than the former mean of 23.6 seconds, using the critical value approach.
    b. Compute the observed significance of the test.
    c. Perform the test at the 5% level of significance using the p-value approach. You need not repeat the first three steps, already done in part (a).

ANSWERS

1.   a. p-value $= 0.0427$
         b. p-value $= 0.0384$
         c. p-value $= 0.0055$

3.   a. p-value $= 0.0009$
         b. p-value $= 0.0064$
         c. p-value $= 0.0146$

5.   a. reject H_0
         b. do not reject H_0

7.   a. $Z = 3.23, z_{0.01} = 2.33$, reject H_0
         b. p-value $= 0.0006$
         c. reject H_0

9.   a. $Z = 0.68, z_{0.05} = 1.645$, do not reject H_0
         b. p-value $= 0.4966$
         c. do not reject H_0

11.   a. $Z = 2.22, z_{0.05} = 1.645$, reject H_0
         b. p-value $= 0.0132$
         c. reject H_0

8.4 Small Sample Tests for a Population Mean

<div style="background:#ccc;padding:1em;">

LEARNING OBJECTIVE

1. To learn how to apply the five-step test procedure for test of hypotheses concerning a population mean when the sample size is small.

</div>

In the previous section hypotheses testing for population means was described in the case of large samples. The statistical validity of the tests was insured by the Central Limit Theorem, with essentially no assumptions on the distribution of the population. When sample sizes are small, as is often the case in practice, the Central Limit Theorem does not apply. One must then impose stricter assumptions on the population to give statistical validity to the test procedure. One common assumption is that the population from which the sample is taken has a normal probability distribution to begin with. Under such circumstances, if the population standard deviation is known, then the test statistic $\left(\bar{x} - \mu_0\right) \big/ \left(\sigma \big/ \sqrt{n}\right)$ still has the standard normal distribution, as in the previous two sections. If σ is unknown and is approximated by the sample standard deviation s, then the resulting test statistic $\left(\bar{x} - \mu_0\right) \big/ \left(s \big/ \sqrt{n}\right)$ follows Student's t-distribution with $n-1$ degrees of freedom.

> ### Standardized Test Statistics for Small Sample Hypothesis Tests Concerning a Single Population Mean
>
> $$\text{If } \sigma \text{ is known:} \quad Z = \frac{\overline{x} - \mu_0}{\sigma \big/ \sqrt{n}}$$
>
> $$\text{If } \sigma \text{ is unknown:} \quad T = \frac{\overline{x} - \mu_0}{s \big/ \sqrt{n}}$$
>
> The first test statistic (σ known) has the standard normal distribution.
>
> The second test statistic (σ unknown) has Student's t-distribution with $n-1$ degrees of freedom.
>
> The population must be normally distributed.

The distribution of the second standardized test statistic (the one containing s) and the corresponding rejection region for each form of the alternative hypothesis (left-tailed, right-tailed, or two-tailed), is shown in Figure 8.11 "Distribution of the Standardized Test Statistic and the Rejection Region". This is just like Figure 8.4 "Distribution of the Standardized Test Statistic and the Rejection Region", except that now the critical values are from the t-distribution. Figure 8.4 "Distribution of the Standardized Test Statistic and the Rejection Region" still applies to the first standardized test statistic (the one containing σ) since it follows the standard normal distribution.

Figure 8.11   *Distribution of the Standardized Test Statistic and the Rejection Region*
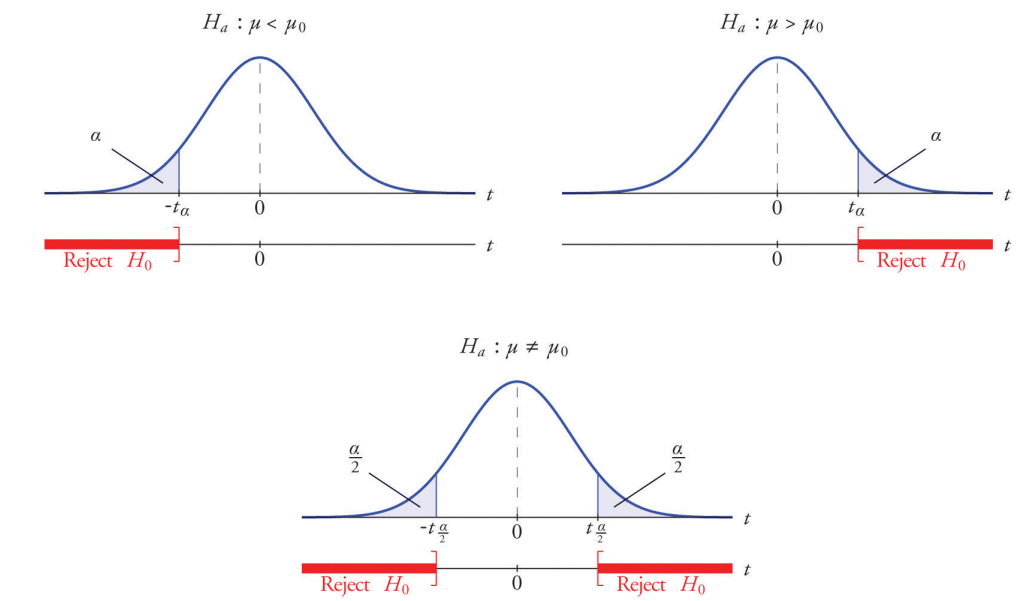
The p-value of a test of hypotheses for which the test statistic has Student's t-distribution can be computed using statistical software, but it is impractical to do so using tables, since that would require 30 tables analogous to Figure 12.2 "Cumulative Normal Probability", one for each degree of freedom from 1 to 30. Figure 12.3 "Critical Values of " can be used to approximate the p-value of such a test, and this is typically adequate for making a decision using the p-value approach to hypothesis testing, although not always. For this reason the tests in the two examples in this section will be made following the critical value approach to hypothesis testing summarized at the end of Section 8.1 "The Elements of Hypothesis Testing", but after each one we will show how the p-value approach could have been used.

EXAMPLE 10

The price of a popular tennis racket at a national chain store is $179. Portia bought five of the same racket at an online auction site for the following prices:

$$155 \quad 179 \quad 175 \quad 175 \quad 161$$

Assuming that the auction prices of rackets are normally distributed, determine whether there is sufficient evidence in the sample, at the 5% level of significance, to conclude that the average price of the racket is less than $179 if purchased at an online auction.

Solution:

- Step 1. The assertion for which evidence must be provided is that the average online price μ is less than the average price in retail stores, so the hypothesis test is

$$H_0 : \mu = 179$$
$$\text{vs.}\, H_a : \mu < 179 \quad @ \, \alpha = 0.05$$

- Step 2. The sample is small and the population standard deviation is unknown. Thus the test statistic is

$$T = \frac{\overline{x} - \mu_0}{s \big/ \sqrt{n}}$$

and has the Student t-distribution with $n-1 = 5 - 1 = 4$ degrees of freedom.

- Step 3. From the data we compute $\overline{x} = 169$ and $s = 10.39$. Inserting these values into the formula for the test statistic gives

$$T = \frac{\overline{x} - \mu_0}{s \big/ \sqrt{n}} = \frac{169 - 179}{10.39 \big/ \sqrt{5}} = -2.152$$

- Step 4. Since the symbol in H_a is "<" this is a left-tailed test, so there is a single critical value, $-t_\alpha = -t_{0.05} \left[df = 4\right]$. Reading from the row labeled $df = 4$ in <u>Figure 12.3 "Critical Values of "</u> its value is –2.132. The rejection region is $\left(-\infty, -2.132\right]$.

  - Step 5. As shown in <u>Figure 8.12 "Rejection Region and Test Statistic for "</u> the test statistic falls in the rejection region. The decision is to reject H_0. In the context of the problem our conclusion is:

    The data provide sufficient evidence, at the 5% level of significance, to conclude that the average price of such rackets purchased at online auctions is less than $179.

Figure 8.12
Rejection Region and Test Statistic for <u>Note 8.42 "Example 10"</u>
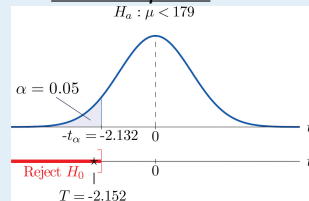
To perform the test in <u>Note 8.42 "Example 10"</u> using the *p*-value approach, look in the row in <u>Figure 12.3 "Critical Values of "</u> with the heading $df = 4$ and search for the two *t*-values that bracket the unsigned value 2.152 of the test statistic. They are 2.132 and 2.776, in the columns with headings $t_{0.050}$ and $t_{0.025}$. They cut off right tails of area 0.050 and 0.025, so because 2.152 is between them it must cut off a tail of area between 0.050 and 0.025. By symmetry –2.152 cuts off a left tail of area between 0.050 and 0.025, hence the *p*-value corresponding to $t = -2.152$ is between 0.025 and 0.05. Although its precise value is unknown, it must be less than $\alpha = 0.05$, so the decision is to reject H_0.

EXAMPLE 11

A small component in an electronic device has two small holes where another tiny part is fitted. In the manufacturing process the average distance between the two holes must be tightly controlled at 0.02 mm, else many units would be defective and wasted. Many times throughout the day quality control engineers take a small sample of the components from the production line, measure the distance between the two holes, and make adjustments if needed. Suppose at one time four units are taken and the distances are measured as

$$0.021 \quad 0.019 \quad 0.023 \quad 0.020$$

Determine, at the 1% level of significance, if there is sufficient evidence in the sample to conclude that an adjustment is needed. Assume the distances of interest are normally distributed.

Solution:

- Step 1. The assumption is that the process is under control unless there is strong evidence to the contrary. Since a deviation of the average distance to either side is undesirable, the relevant test is

$$H_0 : \mu = 0.02$$
$$\text{vs.} \, H_a : \mu \neq 0.02 \quad @ \, \alpha = 0.01$$

where μ denotes the mean distance between the holes.

- Step 2. The sample is small and the population standard deviation is unknown. Thus the test statistic is

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

and has the Student t-distribution with $n-1 = 4 - 1 = 3$ degrees of freedom.

- Step 3. From the data we compute $\bar{x} = 0.02075$ and $s =$ 0.00171. Inserting these values into the formula for the test statistic gives

$$T = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{0.02075 - 0.02}{0.00171 / \sqrt{4}} = 0.877$$

- Step 4. Since the symbol in H_a is "≠" this is a two-tailed test, so there are two critical values, $\pm t_{\alpha/2} = -t_{0.005} \left[df = 3 \right]$. Reading from the row in Figure 12.3 "Critical Values of " labeled $df = 3$ their values are ± 5.841. The rejection region is $(-\infty, -5.841] \cup [5.841, \infty)$.

- Step 5. As shown in Figure 8.13 "Rejection Region and Test Statistic for " the test statistic does not fall in the rejection region. The decision is not to reject H_0. In the context of the problem our conclusion is:

  The data do not provide sufficient evidence, at the 1% level of significance, to conclude that the mean distance between the holes in the component differs from 0.02 mm.

Figure 8.13

Rejection Region and Test Statistic for Note 8.43 "Example 11"
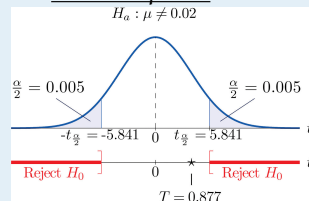
To perform the test in Note 8.43 "Example 11" using the p-value approach, look in the row in Figure 12.3 "Critical Values of " with the heading $df = 3$ and search for the two t-values that bracket the value 0.877 of the test statistic. Actually 0.877 is smaller than the smallest number in the row, which is 0.978, in the column with heading $t_{0.200}$. The value 0.978 cuts off a right tail of area 0.200, so because 0.877 is to its left it must cut off a tail of area greater than 0.200. Thus the p-value, which is the double of the area cut off (since the test is two-tailed), is greater than 0.400.

Although its precise value is unknown, it must be greater than $\alpha = 0.01$, so the decision is not to reject H_0.

KEY TAKEAWAYS

- There are two formulas for the test statistic in testing hypotheses about a population mean with small samples. One test statistic follows the standard normal distribution, the other Student's t-distribution.
- The population standard deviation is used if it is known, otherwise the sample standard deviation is used.
- Either five-step procedure, critical value or p-value approach, is used with either test statistic.

EXERCISES

BASIC

1. Find the rejection region (for the standardized test statistic) for each hypothesis test based on the information given. The population is normally distributed.

   a. $H_0 : \mu = 27$ vs. $H_a : \mu < 27$ @ $\alpha = 0.05$, $n = 12$, $\sigma = 2.2$.
   b. $H_0 : \mu = 52$ vs. $H_a : \mu \neq 52$ @ $\alpha = 0.05$, $n = 6$, σ unknown.
   c. $H_0 : \mu = -105$ vs. $H_a : \mu > -105$ @ $\alpha = 0.10$, $n = 24$, σ unknown.
   d. $H_0 : \mu = 78.8$ vs. $H_a : \mu \neq 78.8$ @ $\alpha = 0.10$, $n = 8$, $\sigma = 1.7$.

2. Find the rejection region (for the standardized test statistic) for each hypothesis test based on the information given. The population is normally distributed.

   a. $H_0 : \mu = 17$ vs. $H_a : \mu < 17$ @ $\alpha = 0.01$, $n = 26$, $\sigma = 0.94$.
   b. $H_0 : \mu = 880$ vs. $H_a : \mu \neq 880$ @ $\alpha = 0.01$, $n = 4$, σ unknown.
   c. $H_0 : \mu = -12$ vs. $H_a : \mu > -12$ @ $\alpha = 0.05$, $n = 18$, $\sigma = 1.1$.
   d. $H_0 : \mu = 21.1$ vs. $H_a : \mu \neq 21.1$ @ $\alpha = 0.05$, $n = 23$, σ unknown.

3. Find the rejection region (for the standardized test statistic) for each hypothesis test based on the information given. The population is normally distributed. Identify the test as left-tailed, right-tailed, or two-tailed.

   a. $H_0 : \mu = 141$ vs. $H_a : \mu < 141$ @ $\alpha = 0.20$, $n = 29$, σ unknown.
   b. $H_0 : \mu = -54$ vs. $H_a : \mu < -54$ @ $\alpha = 0.05$, $n = 15$, $\sigma = 1.9$.
   c. $H_0 : \mu = 98.6$ vs. $H_a : \mu \neq 98.6$ @ $\alpha = 0.05$, $n = 12$, σ unknown.
   d. $H_0 : \mu = 3.8$ vs. $H_a : \mu > 3.8$ @ $\alpha = 0.001$, $n = 27$, σ unknown.

4. Find the rejection region (for the standardized test statistic) for each hypothesis test based on the information given. The population is normally distributed. Identify the test as left-tailed, right-tailed, or two-tailed.

   a. $H_0 : \mu = -62$ vs. $H_a : \mu \neq -62$ @ $\alpha = 0.005$, $n = 8$, σ unknown.
   b. $H_0 : \mu = 73$ vs. $H_a : \mu > 73$ @ $\alpha = 0.001$, $n = 22$, σ unknown.
   c. $H_0 : \mu = 1124$ vs. $H_a : \mu < 1124$ @ $\alpha = 0.001$, $n = 21$, σ unknown.
   d. $H_0 : \mu = 0.12$ vs. $H_a : \mu \neq 0.12$ @ $\alpha = 0.001$, $n = 14$, $\sigma = 0.026$.

5. A random sample of size 20 drawn from a normal population yielded the following results: $\overline{x} = 49.2$, $s = 1.33$.

   a. Test $H_0 : \mu = 50$ vs. $H_a : \mu \neq 50$ @ $\alpha = 0.01$.
   b. Estimate the observed significance of the test in part (a) and state a decision based on the *p*-value approach to hypothesis testing.

6. A random sample of size 16 drawn from a normal population yielded the following results: $\overline{x} = -0.96$, s = 1.07.

   a. Test $H_0 : \mu = 0$ vs. $H_a : \mu < 0$ @ $\alpha = 0.001$.
   b. Estimate the observed significance of the test in part (a) and state a decision based on the *p*-value approach to hypothesis testing.

7. A random sample of size 8 drawn from a normal population yielded the following results: $\overline{x} = 289$, s = 46.

   a. Test $H_0 : \mu = 250$ vs. $H_a : \mu > 250$ @ $\alpha = 0.05$.
   b. Estimate the observed significance of the test in part (a) and state a decision based on the *p*-value approach to hypothesis testing.

8. A random sample of size 12 drawn from a normal population yielded the following results: $\overline{x} = 86.2$, s = 0.63.

   a. Test $H_0 : \mu = 85.5$ vs. $H_a : \mu \neq 85.5$ @ $\alpha = 0.01$.
   b. Estimate the observed significance of the test in part (a) and state a decision based on the *p*-value approach to hypothesis testing.

APPLICATIONS

9. Researchers wish to test the efficacy of a program intended to reduce the length of labor in childbirth. The accepted mean labor time in the birth of a first child is 15.3 hours. The mean length of the labors of 13 first-time mothers in a pilot program was 8.8 hours with standard deviation 3.1 hours. Assuming a normal distribution of times of labor, test at the 10% level of significance test whether the mean labor time for all women following this program is less than 15.3 hours.

10. A dairy farm uses the somatic cell count (SCC) report on the milk it provides to a processor as one way to monitor the health of its herd. The mean SCC from five samples of raw milk was 250,000 cells per milliliter with standard deviation 37,500 cell/ml. Test whether these data provide sufficient evidence, at the 10% level of significance, to conclude that the mean SCC of all milk produced at the dairy exceeds that in the previous report, 210,250 cell/ml. Assume a normal distribution of SCC.

11. Six coins of the same type are discovered at an archaeological site. If their weights on average are significantly different from 5.25 grams then it can be

assumed that their provenance is not the site itself. The coins are weighed and have mean 4.73 g with sample standard deviation 0.18 g. Perform the relevant test at the 0.1% (1/10th of 1%) level of significance, assuming a normal distribution of weights of all such coins.

12. An economist wishes to determine whether people are driving less than in the past. In one region of the country the number of miles driven per household per year in the past was 18.59 thousand miles. A sample of 15 households produced a sample mean of 16.23 thousand miles for the last year, with sample standard deviation 4.06 thousand miles. Assuming a normal distribution of household driving distances per year, perform the relevant test at the 5% level of significance.

13. The recommended daily allowance of iron for females aged 19–50 is 18 mg/day. A careful measurement of the daily iron intake of 15 women yielded a mean daily intake of 16.2 mg with sample standard deviation 4.7 mg.

    a. Assuming that daily iron intake in women is normally distributed, perform the test that the actual mean daily intake for all women is different from 18 mg/day, at the 10% level of significance.
    b. The sample mean is less than 18, suggesting that the actual population mean is less than 18 mg/day. Perform this test, also at the 10% level of significance. (The computation of the test statistic done in part (a) still applies here.)

14. The target temperature for a hot beverage the moment it is dispensed from a vending machine is 170°F. A sample of ten randomly selected servings from a new machine undergoing a pre-shipment inspection gave mean temperature 173°F with sample standard deviation 6.3°F.

    a. Assuming that temperature is normally distributed, perform the test that the mean temperature of dispensed beverages is different from 170°F, at the 10% level of significance.
    b. The sample mean is greater than 170, suggesting that the actual population mean is greater than 170°F. Perform this test, also at the 10% level of significance. (The computation of the test statistic done in part (a) still applies here.)

15. The average number of days to complete recovery from a particular type of knee operation is 123.7 days. From his experience a physician suspects that use of a topical pain medication might be lengthening the recovery time. He randomly selects the records of seven knee surgery patients who used the topical medication. The times to total recovery were:

$$128 \quad 135 \quad 121 \quad 142 \quad 126 \quad 151 \quad 123$$

   a. Assuming a normal distribution of recovery times, perform the relevant test of hypotheses at the 10% level of significance.

   b. Would the decision be the same at the 5% level of significance? Answer either by constructing a new rejection region (critical value approach) or by estimating the p-value of the test in part (a) and comparing it to α.

16. A 24-hour advance prediction of a day's high temperature is "unbiased" if the long-term average of the error in prediction (true high temperature minus predicted high temperature) is zero. The errors in predictions made by one meteorological station for 20 randomly selected days were:

$$
\begin{array}{rrrrr}
2 & 0 & -3 & 1 & -2 \\
1 & 0 & -1 & 1 & -1 \\
-4 & 1 & 1 & -4 & 0 \\
-4 & -3 & -4 & 2 & 2
\end{array}
$$

   a. Assuming a normal distribution of errors, test the null hypothesis that the predictions are unbiased (the mean of the population of all errors is 0) versus the alternative that it is biased (the population mean is not 0), at the 1% level of significance.

   b. Would the decision be the same at the 5% level of significance? The 10% level of significance? Answer either by constructing new rejection regions (critical value approach) or by estimating the p-value of the test in part (a) and comparing it to α.

17. Pasteurized milk may not have a standardized plate count (SPC) above 20,000 colony-forming bacteria per milliliter (cfu/ml). The mean SPC for five samples was 21,500 cfu/ml with sample standard deviation 750 cfu/ml. Test the null hypothesis that the mean SPC for this milk is 20,000 versus the alternative that it is greater than 20,000, at the 10% level of significance. Assume that the SPC follows a normal distribution.

18. One water quality standard for water that is discharged into a particular type of stream or pond is that the average daily water temperature be at most 18°C. Six samples taken throughout the day gave the data:

$$16.8 \quad 21.5 \quad 19.1 \quad 12.8 \quad 18.0 \quad 20.7$$

The sample mean $\overline{x} = 18.15$ exceeds 18, but perhaps this is only sampling error. Determine whether the data provide sufficient evidence, at the 10% level of significance, to conclude that the mean temperature for the entire day exceeds 18°C.

ADDITIONAL EXERCISES

19. A calculator has a built-in algorithm for generating a random number according to the standard normal distribution. Twenty-five numbers thus generated have mean 0.15 and sample standard deviation 0.94. Test the null hypothesis that the mean of all numbers so generated is 0 versus the alternative that it is different from 0, at the 20% level of significance. Assume that the numbers do follow a normal distribution.

20. At every setting a high-speed packing machine delivers a product in amounts that vary from container to container with a normal distribution of standard deviation 0.12 ounce. To compare the amount delivered at the current setting to the desired amount 64.1 ounce, a quality inspector randomly selects five containers and measures the contents of each, obtaining sample mean 63.9 ounces and sample standard deviation 0.10 ounce. Test whether the data provide sufficient evidence, at the 5% level of significance, to conclude that the mean of all containers at the current setting is less than 64.1 ounces.

21. A manufacturing company receives a shipment of 1,000 bolts of nominal shear strength 4,350 lb. A quality control inspector selects five bolts at random and measures the shear strength of each. The data are:

$$4{,}320 \quad 4{,}290 \quad 4{,}360 \quad 4{,}350 \quad 4{,}320$$

   a. Assuming a normal distribution of shear strengths, test the null hypothesis that the mean shear strength of all bolts in the shipment is 4,350 lb versus the alternative that it is less than 4,350 lb, at the 10% level of significance.
   b. Estimate the p-value (observed significance) of the test of part (a).
   c. Compare the p-value found in part (b) to $\alpha = 0.10$ and make a decision based on the p-value approach. Explain fully.

22. A literary historian examines a newly discovered document possibly written by Oberon Theseus. The mean average sentence length of the surviving undisputed works of Oberon Theseus is 48.72 words. The historian counts words in sentences between five successive 101 periods in the document in question to obtain a mean average sentence length of 39.46 words with standard deviation 7.45 words. (Thus the sample size is five.)

   a. Determine if these data provide sufficient evidence, at the 1% level of significance, to conclude that the mean average sentence length in the document is less than 48.72.
   b. Estimate the p-value of the test.
   c. Based on the answers to parts (a) and (b), state whether or not it is likely that the document was written by Oberon Theseus.

ANSWERS

1. a. $Z \leq -1.645$
    b. $T \leq -2.571$ or $T \geq 2.571$
    c. $T \geq 1.319$
    d. $Z \leq -1645$ or $Z \geq 1.645$

3. a. $T \leq -0.855$
    b. $Z \leq -1.645$
    c. $T \leq -2.201$ or $T \geq 2.201$
    d. $T \geq 3.435$

5. a. $T = -2.690$, $df = 19$, $-t_{0.005} = -2.861$, do not reject H_0.
    b. $0.01 < p\text{-value} < 0.02$, $\alpha = 0.01$, do not reject H_0.

7. a. $T = 2.398$, $df = 7$, $t_{0.05} = 1.895$, reject H_0.
    b. $0.01 < p\text{-value} < 0.025$, $\alpha = 0.05$, reject H_0.

9. $T = -7.560$, $df = 12$, $-t_{0.10} = -1.356$, reject H_0.

11. $T = -7.076$, $df = 5$, $-t_{0.0005} = -6.869$, reject H_0.

13. a. $T = -1.483$, $df = 14$, $-t_{0.05} = -1.761$, do not reject H_0;
    b. $T = -1.483$, $df = 14$, $-t_{0.10} = -1.345$, reject H_0;

15. a. $T = 2.069$, $df = 6$, $t_{0.10} = 1.44$, reject H_0;
    b. $T = 2.069$, $df = 6$, $t_{0.05} = 1.943$, reject H_0.

17. $T = 4.472$, $df = 4$, $t_{0.10} = 1.533$, reject H_0.

19. $T = 0.798$, $df = 24$, $t_{0.10} = 1.318$, do not reject H_0.

21. a. $T = -1.773$, $df = 4$, $-t_{0.05} = -2.132$, do not reject H_0.
    b. $0.05 < p\text{-value} < 0.10$
    c. $\alpha = 0.05$, do not reject H_0

8.5 Large Sample Tests for a Population Proportion

<div style="background-color:#d9d5c5;padding:1em;">

LEARNING OBJECTIVES

1. To learn how to apply the five-step critical value test procedure for test of hypotheses concerning a population proportion.
2. To learn how to apply the five-step p-value test procedure for test of hypotheses concerning a population proportion.

</div>

Both the critical value approach and the p-value approach can be applied to test hypotheses about a population proportion p. The null hypothesis will have the form $H_0 : p = p_0$ for some specific number p_0 between 0 and 1. The alternative hypothesis will be one of the three inequalities $p < p_0, p > p_0$, or $p \neq p_0$ for the same number p_0 that appears in the null hypothesis.

The information in <u>Section 6.3 "The Sample Proportion"</u> in <u>Chapter 6 "Sampling Distributions"</u> gives the following formula for the test statistic and its distribution. In the formula p_0 is the numerical value of p that appears in the two hypotheses, $q_0 = 1 - p_0, \widehat{p}$ is the sample proportion, and n is the sample size. Remember that the condition that the sample be large is not that n be at least 30 but that the interval

$$\left[\widehat{p} - 3 \sqrt{\frac{\widehat{p}\left(1 - \widehat{p}\right)}{n}}, \widehat{p} + 3 \sqrt{\frac{\widehat{p}\left(1 - \widehat{p}\right)}{n}} \right]$$

lie wholly within the interval $\left[0,1\right]$.

> ## Standardized Test Statistic for Large Sample Hypothesis Tests Concerning a Single Population Proportion
>
> $$Z = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0 q_0}{n}}}$$
>
> The test statistic has the standard normal distribution.

The distribution of the standardized test statistic and the corresponding rejection region for each form of the alternative hypothesis (left-tailed, right-tailed, or two-tailed), is shown in <u>Figure 8.14 "Distribution of the Standardized Test Statistic and the Rejection Region"</u>.

Figure 8.14 *Distribution of the Standardized Test Statistic and the Rejection Region*
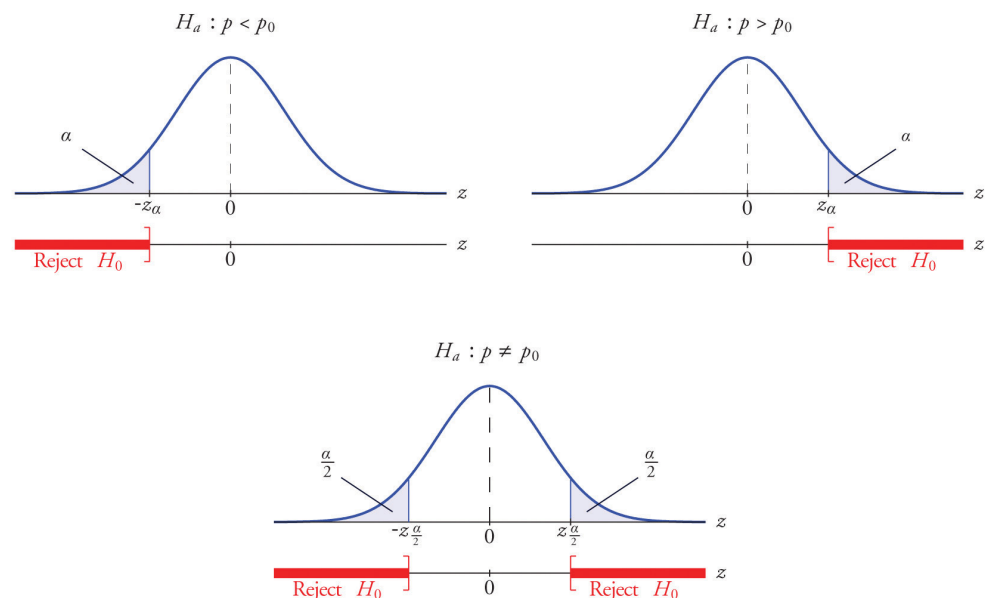
EXAMPLE 12

A soft drink maker claims that a majority of adults prefer its leading beverage over that of its main competitor's. To test this claim 500 randomly selected people were given the two beverages in random order to taste. Among them, 270 preferred the soft drink maker's brand, 211 preferred the competitor's brand, and 19 could not make up their minds. Determine whether there is sufficient evidence, at the 5% level of significance, to support the soft drink maker's claim against the default that the population is evenly split in its preference.

Solution:

We will use the critical value approach to perform the test. The same test will be performed using the p-value approach in .

We must check that the sample is sufficiently large to validly perform the test. Since $\widehat{p} = 270 \big/ 500 = 0.54$,

$$\sqrt{\frac{\widehat{p}\left(1-\widehat{p}\right)}{n}} = \sqrt{\frac{\left(0.54\right)\left(0.46\right)}{500}} \approx 0.02$$

hence

$$\left[\widehat{p} - 3\sqrt{\frac{\widehat{p}\left(1-\widehat{p}\right)}{n}}, \ \widehat{p} + 3\sqrt{\frac{\widehat{p}\left(1-\widehat{p}\right)}{n}}\right]$$

$$= \left[0.54 - \left(3\right)\left(0.02\right), 0.54 + \left(3\right)\left(0.02\right)\right]$$

$$= \left[0.48, 0.60\right] \subset \left[0,1\right]$$

so the sample is sufficiently large.

- Step 1. The relevant test is

$$H_0 : p = 0.50$$
$$\text{vs.} \ H_a : p > 0.50 \ \ @ \ \alpha = 0.05$$

where p denotes the proportion of all adults who prefer the company's beverage over that of its competitor's beverage.

- Step 2. The test statistic is

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

and has the standard normal distribution.

- Step 3. The value of the test statistic is

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.54 - 0.50}{\sqrt{\frac{(0.50)(0.50)}{500}}} = 1.789$$

- Step 4. Since the symbol in H_a is ">" this is a right-tailed test, so there is a single critical value, $z_\alpha = z_{0.05}$. Reading from the last line in <u>Figure 12.3 "Critical Values of "</u> its value is 1.645. The rejection region is $[1.645, \infty)$.

  - Step 5. As shown in <u>Figure 8.15 "Rejection Region and Test Statistic for "</u> the test statistic falls in the rejection region. The decision is to reject H_0. In the context of the problem our conclusion is:

    The data provide sufficient evidence, at the 5% level of significance, to conclude that a majority of adults prefer the company's beverage to that of their competitor's.

Figure 8.15
Rejection Region and Test Statistic for *Note 8.47 "Example 12"*
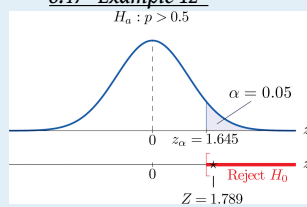
EXAMPLE 13

Globally the long-term proportion of newborns who are male is 51.46%. A researcher believes that the proportion of boys at birth changes under severe economic conditions. To test this belief randomly selected birth records of 5,000 babies born during a period of economic recession were examined. It was found in the sample that 52.55% of the newborns were boys. Determine whether there is sufficient evidence, at the 10% level of significance, to support the researcher's belief.

Solution:

We will use the critical value approach to perform the test. The same test will be performed using the p-value approach in <u>Note 8.50 "Example 15"</u>.

The sample is sufficiently large to validly perform the test since

$$\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} = \sqrt{\frac{(0.5255)\,(0.4745)}{5000}} \approx 0.01$$

hence

$$\left[\hat{p}-3\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}},\ \hat{p}+3\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}\right]$$

$$=\ \left[0.5255-0.03, 0.5255+0.03\right]$$

$$=\ \left[0.4955, 0.5555\right]\subset\left[0,1\right]$$

- Step 1. Let p be the true proportion of boys among all newborns during the recession period. The burden of proof is to show that severe economic conditions change it from the historic long-term value of 0.5146 rather than to show that it stays the same, so the hypothesis test is

$$H_0 : p\ =\ 0.5146$$

$$\text{vs.}\, H_a : p\ \neq\ 0.5146 \quad @\ \alpha = 0.10$$

- Step 2. The test statistic is

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

and has the standard normal distribution.

- Step 3. The value of the test statistic is

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{0.5255 - 0.5146}{\sqrt{\frac{(0.5146)(0.4854)}{5000}}} = 1.542$$

- Step 4. Since the symbol in H_a is "$\neq$" this is a two-tailed test, so there are a pair of critical values, $\pm z_{\alpha/2} = \pm z_{0.05} = \pm 1.645$. The rejection region is $(-\infty, -1.645] \cup [1.645, \infty)$.

  - Step 5. As shown in <u>Figure 8.16 "Rejection Region and Test Statistic for "</u> the test statistic does not fall in the rejection region. The decision is not to reject H_0. In the context of the problem our conclusion is:

    The data do not provide sufficient evidence, at the 10% level of significance, to conclude that the proportion of newborns who are male differs from the historic proportion in times of economic recession.

Figure 8.16
Rejection Region and
Test Statistic for Note
8.48 "Example 13"
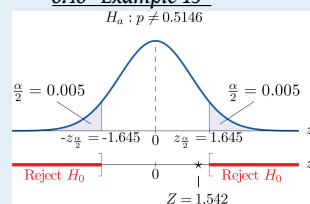
EXAMPLE 14

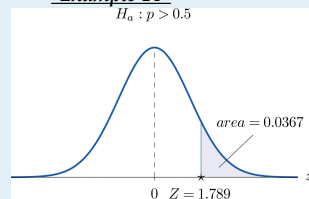Perform the test of Note 8.47 "Example 12" using the p-value approach.

Solution:

We already know that the sample size is sufficiently large to validly perform the test.

- Steps 1–3 of the five-step procedure described in Section 8.3.2 "The " have already been done in Note 8.47 "Example 12" so we will not repeat them here, but only say that we know that the test is right-tailed and that value of the test statistic is $Z = 1.789$.
- Step 4. Since the test is right-tailed the p-value is the area under the standard normal curve cut off by the observed test statistic, $z = 1.789$, as illustrated in Figure 8.17. By Figure 12.2 "Cumulative Normal Probability" that area and therefore the p-value is
  $$1 - 0.9633 = 0.0367.$$
- Step 5. Since the p-value is less than $\alpha = 0.05$ the decision is to reject H_0.

Figure 8.17
P-Value for Note 8.49 "Example 14"

EXAMPLE 15

Perform the test of Note 8.48 "Example 13" using the *p*-value approach.

Solution:

We already know that the sample size is sufficiently large to validly perform the test.

- Steps 1–3 of the five-step procedure described in Section 8.3.2 "The " have already been done in Note 8.48 "Example 13". They tell us that the test is two-tailed and that value of the test statistic is $Z = 1.542$.
- Step 4. Since the test is two-tailed the *p*-value is the double of the area under the standard normal curve cut off by the observed test statistic, $z = 1.542$. By Figure 12.2 "Cumulative Normal Probability" that area is $1 - 0.9382 = 0.0618$, as illustrated in Figure 8.18, hence the *p*-value is $2 \times 0.0618 = 0.1236$.
- Step 5. Since the *p*-value is greater than $\alpha = 0.10$ the decision is not to reject H_0.

Figure 8.18

P-Value for Note 8.50 "Example 15"

$H_a : p \neq 0.5146$

$area = 0.0618$

z

$0 \quad Z = 1.542$
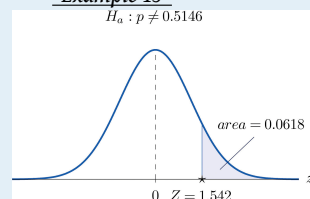
KEY TAKEAWAYS

- There is one formula for the test statistic in testing hypotheses about a population proportion. The test statistic follows the standard normal distribution.
- Either five-step procedure, critical value or *p*-value approach, can be used.

EXERCISES

BASIC

On all exercises for this section you may assume that the sample is sufficiently large for the relevant test to be validly performed.

1. Compute the value of the test statistic for each test using the information given.

   a. Testing $H_0 : p = 0.50$ vs. $H_a : p > 0.50$, $n = 360$, $\hat{p} = 0.56$.
   b. Testing $H_0 : p = 0.50$ vs. $H_a : p \neq 0.50$, $n = 360$, $\hat{p} = 0.56$.
   c. Testing $H_0 : p = 0.37$ vs. $H_a : p < 0.37$, $n = 1200$, $\hat{p} = 0.35$.

2. Compute the value of the test statistic for each test using the information given.

   a. Testing $H_0 : p = 0.72$ vs. $H_a : p < 0.72$, $n = 2100$, $\hat{p} = 0.71$.
   b. Testing $H_0 : p = 0.83$ vs. $H_a : p \neq 0.83$, $n = 500$, $\hat{p} = 0.86$.
   c. Testing $H_0 : p = 0.22$ vs. $H_a : p < 0.22$, $n = 750$, $\hat{p} = 0.18$.

3. For each part of Exercise 1 construct the rejection region for the test for $\alpha = 0.05$ and make the decision based on your answer to that part of the exercise.

4. For each part of Exercise 2 construct the rejection region for the test for $\alpha = 0.05$ and make the decision based on your answer to that part of the exercise.

5. For each part of Exercise 1 compute the observed significance (*p*-value) of the test and compare it to $\alpha = 0.05$ in order to make the decision by the *p*-value approach to hypothesis testing.

6. For each part of Exercise 2 compute the observed significance (*p*-value) of the test and compare it to $\alpha = 0.05$ in order to make the decision by the *p*-value approach to hypothesis testing.

7. Perform the indicated test of hypotheses using the critical value approach.

   a. Testing $H_0 : p = 0.55$ vs. $H_a : p > 0.55$ @ $\alpha = 0.05$, $n = 300$, $\hat{p} = 0.60$.
   b. Testing $H_0 : p = 0.47$ vs. $H_a : p \neq 0.47$ @ $\alpha = 0.01$, $n = 9750$, $\hat{p} = 0.46$.

8. Perform the indicated test of hypotheses using the critical value approach.

a. Testing $H_0 : p = 0.15$ vs. $H_a : p \neq 0.15$ @ $\alpha = 0.001$, $n = 1600$, $\hat{p} = 0.18$.

b. Testing $H_0 : p = 0.90$ vs. $H_a : p > 0.90$ @ $\alpha = 0.01$, $n = 1100$, $\hat{p} = 0.91$.

9. Perform the indicated test of hypotheses using the p-value approach.

a. Testing $H_0 : p = 0.37$ vs. $H_a : p \neq 0.37$ @ $\alpha = 0.005$, $n = 1300$, $\hat{p} = 0.40$.

b. Testing $H_0 : p = 0.94$ vs. $H_a : p > 0.94$ @ $\alpha = 0.05$, $n = 1200$, $\hat{p} = 0.96$.

10. Perform the indicated test of hypotheses using the p-value approach.

a. Testing $H_0 : p = 0.25$ vs. $H_a : p < 0.25$ @ $\alpha = 0.10$, $n = 850$, $\hat{p} = 0.23$.

b. Testing $H_0 : p = 0.33$ vs. $H_a : p \neq 0.33$ @ $\alpha = 0.05$, $n = 1100$, $\hat{p} = 0.30$.

APPLICATIONS

11. Five years ago 3.9% of children in a certain region lived with someone other than a parent. A sociologist wishes to test whether the current proportion is different. Perform the relevant test at the 5% level of significance using the following data: in a random sample of 2,759 children, 119 lived with someone other than a parent.

12. The government of a particular country reports its literacy rate as 52%. A nongovernmental organization believes it to be less. The organization takes a random sample of 600 inhabitants and obtains a literacy rate of 42%. Perform the relevant test at the 0.5% (one-half of 1%) level of significance.

13. Two years ago 72% of household in a certain county regularly participated in recycling household waste. The county government wishes to investigate whether that proportion has increased after an intensive campaign promoting recycling. In a survey of 900 households, 674 regularly participate in recycling. Perform the relevant test at the 10% level of significance.

14. Prior to a special advertising campaign, 23% of all adults recognized a particular company's logo. At the close of the campaign the marketing department commissioned a survey in which 311 of 1,200 randomly selected adults recognized the logo. Determine, at the 1% level of significance, whether the data provide sufficient evidence to conclude that more than 23% of all adults now recognize the company's logo.

15. A report five years ago stated that 35.5% of all state-owned bridges in a particular state were "deficient." An advocacy group took a random sample of 100 state-owned bridges in the state and found 33 to be currently rated as being "deficient." Test whether the current proportion of bridges in such condition is 35.5% versus the alternative that it is different from 35.5%, at the 10% level of significance.

16. In the previous year the proportion of deposits in checking accounts at a certain bank that were made electronically was 45%. The bank wishes to determine if the proportion is higher this year. It examined 20,000 deposit records and found that 9,217 were electronic. Determine, at the 1% level of significance, whether the data provide sufficient evidence to conclude that more than 45% of all deposits to checking accounts are now being made electronically.

17. According to the Federal Poverty Measure 12% of the U.S. population lives in poverty. The governor of a certain state believes that the proportion there is lower. In a sample of size 1,550, 163 were impoverished according to the federal measure.

    a. Test whether the true proportion of the state's population that is impoverished is less than 12%, at the 5% level of significance.
    b. Compute the observed significance of the test.

18. An insurance company states that it settles 85% of all life insurance claims within 30 days. A consumer group asks the state insurance commission to investigate. In a sample of 250 life insurance claims, 203 were settled within 30 days.

    a. Test whether the true proportion of all life insurance claims made to this company that are settled within 30 days is less than 85%, at the 5% level of significance.
    b. Compute the observed significance of the test.

19. A special interest group asserts that 90% of all smokers began smoking before age 18. In a sample of 850 smokers, 687 began smoking before age 18.

    a. Test whether the true proportion of all smokers who began smoking before age 18 is less than 90%, at the 1% level of significance.
    b. Compute the observed significance of the test.

20. In the past, 68% of a garage's business was with former patrons. The owner of the garage samples 200 repair invoices and finds that for only 114 of them the patron was a repeat customer.

a. Test whether the true proportion of all current business that is with repeat customers is less than 68%, at the 1% level of significance.
b. Compute the observed significance of the test.

ADDITIONAL EXERCISES

21. A rule of thumb is that for working individuals one-quarter of household income should be spent on housing. A financial advisor believes that the average proportion of income spent on housing is more than 0.25. In a sample of 30 households, the mean proportion of household income spent on housing was 0.285 with a standard deviation of 0.063. Perform the relevant test of hypotheses at the 1% level of significance. Hint: This exercise could have been presented in an earlier section.

22. Ice cream is legally required to contain at least 10% milk fat by weight. The manufacturer of an economy ice cream wishes to be close to the legal limit, hence produces its ice cream with a target proportion of 0.106 milk fat. A sample of five containers yielded a mean proportion of 0.094 milk fat with standard deviation 0.002. Test the null hypothesis that the mean proportion of milk fat in all containers is 0.106 against the alternative that it is less than 0.106, at the 10% level of significance. Assume that the proportion of milk fat in containers is normally distributed. Hint: This exercise could have been presented in an earlier section.

LARGE DATA SET EXERCISES

23. Large Data Sets 4 and 4A list the results of 500 tosses of a die. Let p denote the proportion of all tosses of this die that would result in a five. Use the sample data to test the hypothesis that p is different from 1/6, at the 20% level of significance.

http://www.gone.2012books.lardbucket.org/sites/all/files/data4.xls

http://www.gone.2012books.lardbucket.org/sites/all/files/data4A.xls

24. Large Data Set 6 records results of a random survey of 200 voters in each of two regions, in which they were asked to express whether they prefer Candidate A for a U.S. Senate seat or prefer some other candidate. Use the full data set (400 observations) to test the hypothesis that the proportion p of all voters who prefer Candidate A exceeds 0.35. Test at the 10% level of significance.

http://www.gone.2012books.lardbucket.org/sites/all/files/data6.xls

25. Lines 2 through 536 in Large Data Set 11 is a sample of 535 real estate sales in a certain region in 2008. Those that were foreclosure sales are identified with a 1 in the second column. Use these data to test, at the 10% level of significance, the hypothesis that the proportion p of all real estate sales in this region in 2008 that were foreclosure sales was less than 25%. (The null hypothesis is $H_0 : p = 0.25$.)

    http://www.gone.2012books.lardbucket.org/sites/all/files/data11.xls

26. Lines 537 through 1106 in Large Data Set 11 is a sample of 570 real estate sales in a certain region in 2010. Those that were foreclosure sales are identified with a 1 in the second column. Use these data to test, at the 5% level of significance, the hypothesis that the proportion p of all real estate sales in this region in 2010 that were foreclosure sales was greater than 23%. (The null hypothesis is $H_0 : p = 0.23$.)

    http://www.gone.2012books.lardbucket.org/sites/all/files/data11.xls

ANSWERS

1.
   a. $Z = 2.277$
   b. $Z = 2.277$
   c. $Z = -1.435$

3.
   a. $Z \geq 1.645$; reject H_0.
   b. $Z \leq -1.96$ or $Z \geq 1.96$; reject H_0.
   c. $Z \leq -1.645$; do not reject H_0.

5.
   a. p-value $= 0.0116$, $\alpha = 0.05$; reject H_0.
   b. p-value $= 0.0232$, $\alpha = 0.05$; reject H_0.
   c. p-value $= 0.0749$, $\alpha = 0.05$; do not reject H_0.

7.
   a. $Z = 1.74, z_{0.05} = 1.645$, reject H_0.
   b. $Z = -1.98, -z_{0.005} = -2.576$, do not reject H_0.

9.
   a. $Z = 2.24, p$-value $= 0.025$, $\alpha = 0.005$, do not reject H_0.
   b. $Z = 2.92, p$-value $= 0.0018$, $\alpha = 0.05$, reject H_0.

11. $Z = 1.11, z_{0.025} = 1.96$, do not reject H_0.

13. $Z = 1.93, z_{0.10} = 1.28$, reject H_0.

15. $Z = -0.523$, $\pm z_{0.05} = \pm 1.645$, do not reject H_0.

17.
   a. $Z = -1.798$, $-z_{0.05} = -1.645$, reject H_0;
   b. $p - $-value $= 0.0359$.

19.
   a. $Z = -8.92$, $-z_{0.01} = -2.33$, reject H_0;
   b. p-value ≈ 0.

21. $Z = 3.04, z_{0.01} = 2.33$, reject H_0.

23. $H_0 : p = 1/6$ vs. $H_a : p \neq 1/6$. Test Statistic: $Z = -0.76$.
   Rejection Region: $(-\infty, -1.28] \cup [1.28, \infty)$. Decision: Fail to reject H_0.

25. $H_0 : p = 0.25$ vs. $H_a : p < 0.25$. Test Statistic: $Z = -1.17$. Rejection
   Region: $(-\infty, -1.28]$. Decision: Fail to reject H_0.