CHAPTER 9

Testing Statistical Hypotheses

The great tragedy of science—the slaying of a beautiful hypothesis by an ugly set of data.

Thomas H. Huxley, English biologist (Biogenesis and Abiogenesis)

We all learn by experience, and the lesson this time is that you should never lose sight of the alternative.

Sherlock Holmes, in *The Adventures of Black Peter* by Sir Arthur Conan Doyle

CONTENTS

Introductory Statistics, DOI: 10.1016/B978-0-12-374388-6.00009-0

We explain what a statistical hypothesis is and show how sample data can be used to test it. We distinguish between the null hypothesis and the alternative hypothesis. We explain the significance of rejecting a null hypothesis and of not rejecting it. We introduce the concept of the p value that results from a test.

Tests concerning the mean of a normal population are studied, when the population variance is both known and unknown. One-sided and two-sided tests are considered. Tests concerning a population proportion are presented.

9.1 INTRODUCTION

There has been a great deal of controversy in recent years over the possible dangers of living near a high-level electromagnetic field (EMF). One researcher, after hearing many anecdotal tales concerning the large increases in cancers, especially among children, in communities living near an EMF, decided to study the possible dangers. To do so, she first studied maps giving the locations of electric power lines and then used them to select a fairly large community that was located in a high-level EMF area. She spent time interviewing people in the local schools, hospitals, and public health facilities in order to discover the number of children who had been afflicted with (any type of) cancer in the previous 3 years, and she found that there had been 32 such cases.

She then visited a government public health library to learn about the number of cases of childhood cancer that could be expected in a community the size of the one she was considering. She learned that the average number of cases of childhood cancer over a 3-year period in such a community was 16.2, with a standard deviation of 4.7.

Is the discovery of 32 cases of childhood cancers significantly large enough, in comparison with the average number of 16.2, for the researcher to conclude that there is some special factor in the community being studied that increases the chance for children to contract cancer? Or is it possible that there is nothing special about the community and that the greater number of cancers is due solely to chance? In this chapter we will show how such questions can be answered.

9.2 HYPOTHESIS TESTS AND SIGNIFICANCE LEVELS

Statistical inference is the science of drawing conclusions about a population based on information contained in a sample. A particular type of inference is involved with the testing of hypotheses concerning some of the parameters of the population distribution. These hypotheses will usually specify that a population parameter, such as the population mean or variance, has a value that lies in a particular region. We must then decide whether this hypothesis is consistent with data obtained in a sample.

Definition *A* statistical hypothesis *is a statement about the nature of a population. It is often stated in terms of a population parameter.*

To test a statistical hypothesis, we must decide whether that hypothesis appears to be consistent with the data of the sample. For instance, suppose that a tobacco firm claims that it has discovered a new way of curing tobacco leaves that will result in a mean nicotine content of a cigarette of 1.5 milligrams or less. Suppose that a researcher is skeptical of this claim and indeed believes that the mean will exceed 1.5 milligrams. To disprove the claim of the tobacco firm, the researcher has decided to test its hypothesis that the mean is less than or equal to 1.5 milligrams. The statistical hypothesis to be tested, which is called the *null hypothesis* and is denoted by H_0, is thus that the mean nicotine content is less than or equal to 1.5 milligrams. Symbolically, if we let μ denote this mean nicotine content per cigarette, then we can express the null hypothesis as

$$H_0\colon \mu \le 1.5$$

The alternative to the null hypothesis, which the tester is actually trying to establish, is called the *alternative hypothesis* and is designated by H_1. For our example, H_1 is the hypothesis that the mean nicotine content exceeds 1.5 milligrams, which can be written symbolically as

$$H_1\colon \mu > 1.5$$

The *null hypothesis,* denoted by H_0, is a statement about a population parameter. The alternative hypothesis is denoted by H_1. The null hypothesis will be rejected if it appears to be inconsistent with the sample data and will not be rejected otherwise.

To test the null hypothesis that the mean nicotine content per cigarette is less than or equal to 1.5 milligrams, a random sample of cigarettes cured by the new method should be chosen and their nicotine content measured. If the resulting sample data are not "consistent" with the null hypothesis, then we say that the null hypothesis is rejected; if they are "consistent" with the null hypothesis, then the null hypothesis is not rejected.

The decision of whether to reject the null hypothesis is based on the value of a test statistic.

Definition *A* test statistic *is a statistic whose value is determined from the sample data. Depending on the value of this test statistic, the null hypothesis will be rejected or not.*

In the cigarette example being considered, the test statistic might be the average nicotine content of the sample of cigarettes. The statistical test would then reject the null hypothesis when this test statistic was sufficiently larger than 1.5. In general, if we let TS denote the test statistic, then to complete our specifications of the test, we must designate the set of values of TS for which the null hypothesis will be rejected.

Definition *The* critical region, *also called the* rejection region, *is that set of values of the test statistic for which the null hypothesis is rejected.*

The statistical test of the null hypothesis H_0 is completely specified once the test statistic and the critical region are specified. If TS denotes the test statistic and C denotes the critical region, then the statistical test of the null hypothesis H_0 is as follows:

Reject H_0	if TS is in C
Do not reject H_0	if TS is not in C

For instance, in the nicotine example we have been considering, if it were known that the standard deviation of a cigarette's nicotine content was 0.8 milligrams, then one possible test of the null hypothesis is to use the test statistic $\overline{X}$, equal to the sample mean nicotine level, along with the critical region

$$C = \left\{\overline{X} \geq 1.5 + \frac{1.312}{\sqrt{n}}\right\}$$

That is, the null hypothesis is to be

Rejected	if $\overline{X} \geq 1.5 + \frac{1.312}{\sqrt{n}}$
Not rejected	otherwise

where n is the sample size. (The rationale behind the choice of this particular critical region will become apparent in the next section.)

For instance, if the foregoing test is employed and if the sample size is 36, then the null hypothesis that the population mean is less than or equal to 1.5 will be rejected if $\overline{X} \geq 1.719$ and will not be rejected if $\overline{X} < 1.719$. It is important to note that even when the estimate of μ—namely, the value of the sample mean $\overline{X}$—exceeds 1.5, the null hypothesis may still not be rejected. Indeed, when $n = 36$, a sample mean value of 1.7 will not result in rejection of the null hypothesis. This is true even though such a large value of the sample mean is certainly not evidence in support of the null hypothesis. Nevertheless, it is consistent with the null hypothesis in that if the population mean is 1.5, then there is a reasonable probability that the average of a sample of size 36 will be as large as 1.7. On the other hand, a value of the sample mean as large as 1.9 is so unlikely if the population mean is less than or equal to 1.5 that it will lead to rejection of this hypothesis.

The rejection of the null hypothesis H_0 is a strong statement that H_0 does not appear to be consistent with the observed data. The result that H_0 is not rejected is a weak statement that should be interpreted to mean that H_0 is consistent with the data.

Thus, in any procedure for testing a given null hypothesis, two different types of errors can result. The first, called a *type I error*, is said to result if the test rejects H_0 when H_0 is true. The second, called a *type II error*, is said to occur if the test does not reject H_0 when H_0 is false. Now, it must be understood that the objective of a statistical test of the null hypothesis H_0 is not to determine whether H_0 is true, but rather to determine if its truth is consistent with the resultant data. Therefore, given this objective, it is reasonable that H_0 should be rejected only if the sample data are very unlikely when H_0 is true. The classical way of accomplishing this is to specify a small value α and then require that the test have the property that whenever H_0 is true, its probability of being rejected is less than or equal to α. The value α, called the *level of significance* of the test, is usually set in advance, with commonly chosen values being $\alpha = 0.10, 0.05$, and 0.01.

The classical procedure for testing a null hypothesis is to fix a small significance level α and then require that the probability of rejecting H_0 when H_0 is true is less than or equal to α.

Because of the asymmetry in the test regarding the null and alternative hypotheses, it follows that the only time in which an hypothesis can be regarded as having been "proved" by the data is when the null hypothesis is rejected (thus "proving" that the alternative is true). For this reason the following rule should be noted.

If you are trying to establish a certain hypothesis, then that hypothesis should be designated as the alternative hypothesis. Similarly, if you are trying to discredit a hypothesis, that hypothesis should be designated the null hypothesis.

Thus, for instance, if the tobacco company is running the experiment to prove that the mean nicotine level of its cigarettes is less than 1.5, then it should choose for the null hypothesis

$$H_0\colon \mu \geq 1.5$$

and for the alternative hypothesis

$$H_1\colon \mu > 1.5$$

Then the company could use a rejection of the null hypothesis as "proof" of its claim that the mean nicotine content was less than 1.5 milligrams.

Suppose now that we are interested in developing a test of a certain hypothesis regarding θ, a parameter of the population distribution. Specifically, suppose that for a given region R we are trying to test the null hypothesis that θ lies in the region R.

That is, we want to test

$$\mathrm{H}_0\text{: } \theta \text{ lies in } R$$

against the alternative

$$\mathrm{H}_1\text{: } \theta \text{ does not lie in } R$$

An approach to developing a test of H_0, at level of significance α, is to start by determining a point estimator of θ. The test will reject H_0 when this point estimator is "far away" from the region R. However, to determine how "far away" it needs to be to justify rejection of H_0, first we need to determine the probability distribution of the point estimator when H_0 is true. This will enable us to specify the appropriate critical region so that the probability that the estimator will fall in that region when H_0 is true is less than or equal to α. In the following section we will illustrate this approach by considering tests concerning the mean of a normal population.

PROBLEMS

1. Consider a trial in which a jury must decide between hypothesis A that the defendant is guilty and hypothesis B that he or she is innocent.
   (a) In the framework of hypothesis testing and the U.S. legal system, which of the hypotheses should be the null hypothesis?
   (b) What do you think would be the appropriate significance level in this situation?
2. A British pharmaceutical company, Glaxo Holdings, has recently developed a new drug for migraine headaches. Among the claims Glaxo made for its drug, called *somatriptan*, was that the mean time needed for it to enter the bloodstream is less than 10 minutes. To convince the Food and Drug Administration of the validity of this claim, Glaxo conducted an experiment on a randomly chosen set of migraine sufferers. To prove the company's claim, what should Glaxo have taken as the null and the alternative hypotheses?
3. Suppose a test of

$$\mathrm{H}_0\text{: } \mu = 0 \quad \text{against} \quad \mathrm{H}_1\text{: } \mu \neq 0$$

   resulted in rejection of H_0 at the 5 percent level of significance. Which of the following statements is (are) accurate?
   (a) The data proved that μ is significantly different from 0, meaning that it is far away from 0.
   (b) The data were significantly strong enough to conclude that μ is not equal to 0.

(c) The probability that μ is equal to 0 is less than 0.05.
(d) The hypothesis that μ is equal to 0 was rejected by a procedure that would have resulted in rejection only 5 percent of the time when μ is equal to 0.

Historical Perspective

The concept of significance level was originated by the English statistician Ronald A. Fisher. Fisher also formulated the concept of the null hypothesis as the hypothesis that one is trying to disprove. In Fisher's words, "Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis." The idea of an alternative hypothesis was due to the joint efforts of the Polish-born statistician Jerzy Neyman and his longtime collaborator Egon (son of Karl) Pearson. Fisher, however, did not accept the idea of an alternative hypothesis, arguing that in most scientific applications it was not possible to specify such alternatives, and a great feud ensued between Fisher on one side and Neyman and Pearson on the other. Due to both Fisher's temperament, which was contentious to say the least, and the fact that he was already involved in a controversy with Neyman over the relative benefits of confidence interval estimates, which were originated by Neyman, and Fisher's own fiducial interval estimates (which are not much used today), the argument became extremely personal and vitriolic. At one point Fisher called Neyman's position "horrifying for intellectual freedom in the West."


Jerzy Neyman
(Public Information Office, University of California, Berkeley)

Fisher is famous for his scientific feuds. Aside from the one just mentioned, he carried on a most heated debate with Karl Pearson over the relative merits of two different general approaches for obtaining point estimators, called the *method of moments* and the *method of maximum likelihood*. Fisher, who was a founder of the field of population genetics, also carried out a long-term feud with Sewell Wright, another influential population geneticist, over the role played by chance in the determination of future gene frequencies. (Curiously enough, it was the biologist Wright and not the statistician Fisher who championed cause as a key factor in long-term evolutionary developments.)

4. Let μ denote the mean value of some population. Suppose that in order to test

$$H_0: \mu \leq 1.5$$

against the alternative hypothesis

$$H_1: \mu > 1.5$$

a sample is chosen from the population.

(a) Suppose for this sample that H_0 was not rejected. Does this imply that the sample data would have resulted in rejection of the null hypothesis if we had been testing the following?

$$H_0: \mu > 1.5 \quad \text{against} \quad H_1: \mu \leq 1.5$$

(b) Suppose this sample resulted in the rejection of H_0. Does this imply that the same sample data would have resulted in not rejecting the null hypothesis if we had been testing the following?

$$H_0: \mu > 1.5 \quad \text{against} \quad H_1: \mu \leq 1.5$$

Assume that all tests are at the 5 percent level of significance, and explain your answers!

9.3 TESTS CONCERNING THE MEAN OF A NORMAL POPULATION: CASE OF KNOWN VARIANCE

Suppose that $X_1, \ldots, X_n$ are a sample from a normal distribution having an unknown mean μ and a known variance σ^2, and suppose we want to test the null hypothesis that the mean μ is equal to some specified value against the alternative that it is not. That is, we want to test

$$H_0: \mu = \mu_0$$

against the alternative hypothesis

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

for a specified value μ_0.

Since the natural point estimator of the population mean μ is the sample mean

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

it would seem reasonable to reject the hypothesis that the population mean is equal to μ_0 when $\overline{X}$ is far away from μ_0. That is, the critical region of the test should be of the form

$$C = \left\{X_1, \ldots, X_n: \quad \left|\overline{X} - \mu_0\right| \geq c\right\}$$

for a suitable value of c.

Suppose we want the test to have significance level α. Then c must be chosen so that the probability, when μ_0 is the population mean, that $\overline{X}$ differs from μ_0 by c

or more is equal to α. That is, c should be such that

$$P\{|\overline{X} - \mu_0| \geq c\} = \alpha \quad \text{when} \quad \mu = \mu_0 \tag{9.1}$$

However, when μ is equal to μ_0, $\overline{X}$ is normally distributed with mean μ_0 and standard deviation $\sigma/\sqrt{n}$, and so the standardized variable Z, defined by

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\sqrt{n}}{\sigma}(\overline{X} - \mu_0)$$

will have a standard normal distribution. Now, since the inequality

$$|\overline{X} - \mu_0| \geq c$$

is equivalent to

$$\frac{\sqrt{n}}{\sigma}|\overline{X} - \mu_0| \geq \frac{\sqrt{n}}{\sigma}c$$

it follows that probability statement (9.1) is equivalent to

$$P\left\{|Z| \geq \sqrt{n}\frac{c}{\sigma}\right\} = \alpha$$

Since the probability that the absolute value of a standard normal exceeds some value is equal to twice the probability that a standard normal exceeds that value (see Fig. 9.1), we see from the preceding that

$$2P\left\{Z \geq \sqrt{n}\frac{c}{\alpha}\right\} = \alpha$$

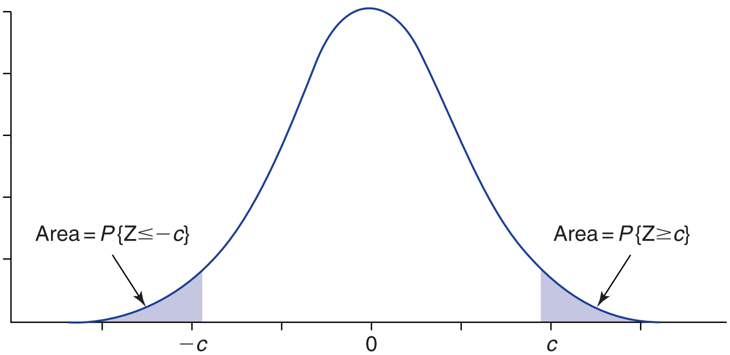


FIGURE 9.1
$P\{|Z| \geq c\} = P\{Z \geq c\} + P\{Z \leq -c\} = 2P\{Z \geq c\}$.

or

$$P\left\{Z \geq \sqrt{n}\frac{c}{\sigma}\right\} = \frac{\alpha}{2}$$

Since $z_{\alpha/2}$ is defined to be such that

$$P\{Z \geq z_{\alpha/2}\} = \frac{\alpha}{2}$$

it follows from the preceding that

$$\sqrt{n}\frac{c}{\sigma} = z_{\alpha/2}$$

or

$$c = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Therefore, the significance-level-α test of the null hypothesis that the population mean is equal to the specified value μ_0 against the alternative that it is not equal to μ_0 is to reject the null hypothesis if

$$\left|\overline{X} - \mu_0\right| \geq z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

or, equivalently, to

$$\begin{array}{ll} \text{Reject } H_0 & \text{if } \dfrac{\sqrt{n}}{\sigma}\left|\overline{X} - \mu_0\right| \geq z_{\alpha/2} \\ \text{Not reject } H_0 & \text{otherwise} \end{array}$$

This test is pictorially depicted in Fig. 9.2. Note that in Fig. 9.2 we have superimposed the standard normal density function over the real line, since that is the density of the test statistic $\sqrt{n}\left(\overline{X} - \mu_0\right)/\sigma$ when H_0 is true. Also, because of this fact, the preceding test is often called the *Z test*.

■ Example 9.1

Suppose that if a signal of intensity μ is emitted from a particular star, then the value received at an observatory on earth is a normal random variable with mean μ and standard deviation 4. In other words, the value of the signal emitted is altered by *random noise*, which is normally distributed with mean 0 and standard deviation 4. It is suspected that the intensity of the signal is equal to 10. Test whether this hypothesis is plausible if the same signal is independently

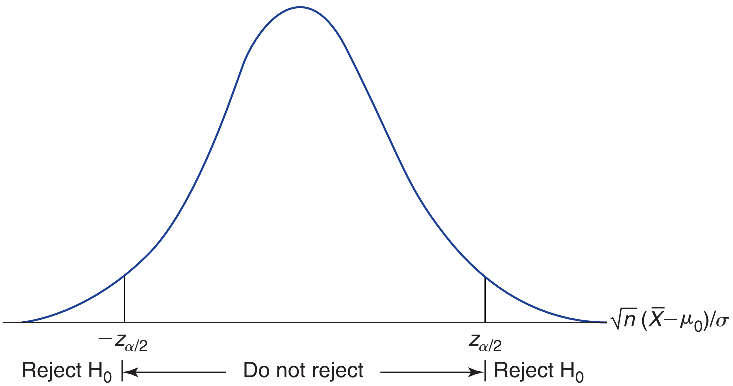


FIGURE 9.2
Test of $H_0: \mu = \mu_0$ *against* $H_1: \mu \neq \mu_0$.

received 20 times and the average of the 20 values received is 11.6. Use the 5 percent level of significance.

Solution

If μ represents the actual intensity of the signal emitted, then the null hypothesis we want to test is

$$H_0: \mu = 10$$

against the alternative

$$H_1: \mu \neq 10$$

Suppose we are interested in testing this at significance level 0.05. To begin, we compute the value of the statistic

$$\frac{\sqrt{n}}{\sigma}\left|\bar{X} - \mu_0\right| = \frac{\sqrt{20}}{4}|11.6 - 10| = 1.79$$

Since this value is less than $z_{0.025} = 1.96$, the null hypothesis is not rejected. In other words, we conclude that the data are not inconsistent with the null hypothesis that the value of the signal is equal to 10. The reason for this is that a sample mean as far from the value 10 as the one observed would occur, when H_0 is true, over 5 percent of the time. Note, however, that if the significance level were chosen to be $\alpha = 0.1$, as opposed to $\alpha = 0.05$, then the null hypothesis would be rejected (since $z_{\alpha/2} = z_{0.05} = 1.645$). ■

It is important to note that the "correct" level of significance to use in any given hypothesis-testing situation depends on the individual circumstances of that situation. If rejecting H_0 resulted in a large cost that would be wasted if H_0 were indeed true, then we would probably elect to be conservative and choose a small significance level. For instance, suppose that H_1 is the hypothesis that a new method of production is superior to the one presently in use. Since a rejection of H_0 would result in a change of methods, we would want to make certain that the probability of rejection when H_0 is true is quite small; that is, we would want a small value of α. Also, if we initially felt quite strongly that the null hypothesis was true, then we would require very strong data evidence to the contrary for us to reject H_0, and so we would again choose a very small significance level.

The hypothesis test just given can be described as follows: The value, call it v, of the test statistic $\sqrt{n}\left(\overline{X} - \mu_0\right)/\sigma$ is determined. The test now calls for rejection of H_0 if the probability that the absolute value of the test statistic will be as large as $|v|$ is, when H_0 is true, less than or equal to α. It therefore follows that the test can be performed by computing, first, the value v of the test statistic and, second, the probability that the absolute value of a standard normal will exceed $|v|$. This probability, called the *p value*, gives the critical significance level, in the sense that H_0 will be rejected if the p value is less than or equal to the significance level α and will not be rejected otherwise.

The *p value* is the smallest significance level at which the data lead to rejection of the null hypothesis. It gives the probability that data as unsupportive of H_0 as those observed will occur when H_0 is true. A small p value (say, 0.05 or less) is a strong indicator that the null hypothesis is not true. The smaller the p value, the greater the evidence for the falsity of H_0.

In practice, the significance level is often not set in advance; rather, the data are used to determine the p value. This value is often either so large that it is clear that the null hypothesis should not be rejected or so small that it is clear that the null hypothesis should be rejected.

■ Example 9.2

Suppose that the average of the 20 values in Example 9.1 is equal to 10.8. In this case the absolute value of the test statistic is

$$\frac{\sqrt{n}}{\sigma}|\overline{X} - \mu_0| = \frac{\sqrt{20}}{4}|10.8 - 10| = 0.894$$

Since

$$\begin{aligned} P\{|Z| \geq 0.894\} &= 2P\{Z \geq 0.894\} \\ &= 0.371 \quad \text{(from Table D.1)} \end{aligned}$$

it follows that the p value is 0.371. Therefore, the null hypothesis that the signal value is 10 will not be rejected at any significance level less than 0.371. Since we never want to use a significance level as high as that, H_0 will not be rejected.

On the other hand, if the value of the sample mean were 7.8, then the absolute value of the test statistic would be

$$\frac{\sqrt{20}}{4}(2.2) = 2.46$$

and so the p value would be

$$\begin{aligned} p \text{ value} &= P\{|Z| \geq 2.46\} \\ &= 2P\{Z \geq 2.46\} \\ &= 0.014 \end{aligned}$$

Thus, H_0 would be rejected at all significance levels above 0.014 and would not be rejected for lower significance levels. ■

The next example is concerned with determining the probability of not rejecting the null hypothesis when it is false.

■ Example 9.3

In Example 9.1, assuming a 0.05 significance level, what is the probability that the null hypothesis (that the signal intensity is equal to 10) will not be rejected when the actual signal value is 9.2?

Solution

In Example 9.1, $\sigma = 4$ and $n = 20$. Therefore, the significance-level-0.05 test of

$$H_0\colon \mu = 10 \quad \text{against} \quad H_1\colon \mu \neq 10$$

is to reject H_0 if

$$\frac{\sqrt{20}}{4}|\overline{X} - 10| \geq z_{0.025}$$

or, equivalently, if

$$|\overline{X} - 10| \geq \frac{4z_{0.025}}{\sqrt{20}}$$

Since $4z_{0.025}\sqrt{20} = 4 \times 1.96/\sqrt{20} = 1.753$, this means that H_0 is to be rejected if the distance between $\overline{X}$ and 10 is at least 1.753. That is, H_0 will be rejected if either

$$\overline{X} \geq 10 + 1.753$$

or

$$\overline{X} \leq 10 - 1.753$$

That is, if

$$\overline{X} \geq 11.753 \quad \text{or} \quad \overline{X} \leq 8.247$$

then H_0 will be rejected.

Now, if the population mean is 9.2, then $\overline{X}$ will be normal with mean 9.2 and standard deviation $4/\sqrt{20} = 0.894$; and so the standardized variable

$$Z = \frac{\overline{X} - 9.2}{0.894}$$

will be a standard normal random variable. Thus, when the true value of the signal is 9.2, we see that

$$\begin{aligned} P\{\text{rejection of } H_0\} &= P\{\overline{X} \geq 11.753\} + P\{\overline{X} \leq 8.247\} \\ &= P\left\{\frac{\overline{X} - 9.2}{0.894} \geq \frac{11.753 - 9.2}{0.894}\right\} + P\left\{\frac{\overline{X} - 9.2}{0.894} \leq \frac{8.247 - 9.2}{0.894}\right\} \\ &= P\{Z \geq 2.856\} + P\{Z \leq -1.066\} \\ &= 0.0021 + 0.1432 \\ &= 0.1453 \end{aligned}$$

That is, when the true signal value is 9.2, there is an 85.47 percent chance that the 0.05 significance level test will not reject the null hypothesis that the signal value is equal to 10. ■

PROBLEMS

In all problems, assume that the relevant distribution is normal.

1. The device that an astronomer utilizes to measure distances results in measurements that have a mean value equal to the actual distance of the object being surveyed and a standard deviation of 0.5 light-years. Present theory indicates that the actual distance from Earth to the asteroid Phyla is 14.4 light-years. Test this hypothesis, at the 5 percent level of significance, if six independent measurements yielded the data

   15.1, 14.8, 14.0, 15.2, 14.7, 14.5

2. A previous sample of fish in Lake Michigan indicated that the mean polychlorinated biphenyl (PCB) concentration per fish was 11.2 parts per million with a standard deviation of 2 parts per million. Suppose a new random sample of 10 fish has the following concentrations:

   11.5, 12.0, 11.6, 11.8, 10.4, 10.8, 12.2, 11.9, 12.4, 12.6

   Assume that the standard deviation has remained equal to 2 parts per million, and test the hypothesis that the mean PCB concentration has also remained unchanged at 11.2 parts per million. Use the 5 percent level of significance.
3. To test the hypothesis

$$H_0\colon \mu = 105 \quad \text{against} \quad H_1\colon \mu \neq 105$$

   a sample of size 9 is chosen. If the sample mean is $\overline{X} = 100$, find the p value if the population standard deviation is known to be
   (a) $\sigma = 5$
   (b) $\sigma = 10$
   (c) $\sigma = 15$
   In which cases would the null hypothesis be rejected at the 5 percent level of significance? What about at the 1 percent level?
4. Repeat Prob. 3 for a sample mean that is the same but for a sample size of 36.
5. A colony of laboratory mice consists of several thousand mice. The average weight of all the mice is 32 grams with a standard deviation of 4 grams. A laboratory assistant was asked by a scientist to select 25 mice for an experiment. However, before performing the experiment, the scientist decided to weigh the mice as an indicator of whether the assistant's selection constituted a random sample or whether it was made with some unconscious bias (perhaps the mice selected were the ones that were slowest in avoiding the assistant, which might indicate some inferiority about this group). If the sample mean of the 25 mice was 30.4, would this be significant evidence, at the 5 percent level of significance, against the hypothesis that the selection constituted a random sample?
6. It is known that the value received at a local receiving station is equal to the value sent plus a random error that is normal with mean 0 and standard deviation 2. If the same value is sent 7 times, compute the p value for the test of the null hypothesis that the value sent is equal to 14, if the values received are

   14.6, 14.8, 15.1, 13.2, 12.4, 16.8, 16.3

7. Historical data indicate that household water use tends to be normally distributed with a mean of 360 gallons and a standard deviation of 40 gallons per day. To see if this is still the situation, a random sample of 200 households was chosen. The average daily water use in these households was then seen to equal 374 gallons per day.
   (a) Are these data consistent with the historical distribution? Use the 5 percent level of significance.
   (b) What is the p value?
8. When a certain production process is operating properly, it produces items that each have a measurable characteristic with mean 122 and standard deviation 9. However, occasionally the process goes out of control, and this results in a change in the mean of the items produced. Test the hypothesis that the process is presently in control if a random sample of 10 recently produced items had the following values:

   123, 120, 115, 125, 131, 127, 130, 118, 125, 128

   Specify the null and alternative hypotheses, and find the p value.
9. A leasing firm operates on the assumption that the annual number of miles driven in its leased cars is normally distributed with mean 13,500 and standard deviation 4000 miles. To see whether this assumption is valid, a random sample of 36 one-year-old cars has been checked. What conclusion can you draw if the average mileage on these 36 cars is 15,233?
10. A population distribution is known to have standard deviation 20. Determine the p value of a test of the hypothesis that the population mean is equal to 50, if the average of a sample of 64 observations is
    (a) 52.5
    (b) 55.0
    (c) 57.5
11. Traffic authorities claim that traffic lights are red for a time that is normal with mean 30 seconds and standard deviation 1.4 seconds. To test this claim, a sample of 40 traffic lights was checked. If the average time of the 40 red lights observed was 32.2 seconds, can we conclude, at the 5 percent level of significance, that the authorities are incorrect? What about at the 1 percent level of significance?
12. The number of cases of childhood cancer occurring within a 3-year span in communities of a specified size has an approximately normal distribution with mean 16.2 and standard deviation 4.7. To see whether this distribution changes when the community is situated near a high-level electromagnetic field, a researcher chose such a community and subsequently discovered that there had been a total of 32 cases of childhood cancers within the last 3 years. Using these

data, find the p value of the test of the hypothesis that the distribution of the number of childhood cancers in communities near high-level electromagnetic fields remains normal with mean 16.2 and standard deviation 4.7.

13. The following data are known to come from a normal population having standard deviation 2. Use them to test the hypothesis that the population mean is equal to 15. Determine the significance levels at which the test would reject and those at which it would not reject this hypothesis.

$$15.6, 16.4, 14.8, 17.2, 16.9, 15.3, 14.0, 15.9$$

14. Suppose, in Prob. 1, that current theory is wrong and that the actual distance to the asteroid Phyla is 14.8 light-years. What is the probability that a series of 10 readings, each of which has a mean equal to the actual distance and a standard deviation of 0.8 light-years, will result in a rejection of the null hypothesis that the distance is 14 light-years? Use a 1 percent level of significance.

15. In Prob. 6 compute the probability that the null hypothesis that the value 14 is sent will be rejected, at the 5 percent level of significance, when the actual value sent is
    (a) 15
    (b) 13
    (c) 16

9.3.1 One-Sided Tests

So far we have been considering two-sided hypothesis-testing problems in which the null hypothesis is that μ is equal to a specified value μ_0 and the test is to reject this hypothesis if $\overline{X}$ is either too much larger or too much smaller than μ_0. However, in many situations, the hypothesis we are interested in testing is that the mean is less than or equal to some specified value μ_0 versus the alternative that it is greater than that value. That is, we are often interested in testing

$$\mathrm{H}_0\colon \mu \leq \mu_0$$

against the alternative

$$\mathrm{H}_1\colon \mu > \mu_0$$

Since we would want to reject H_0 only when the sample mean $\overline{X}$ is much larger than μ_0 (and no longer when it is much smaller), it can be shown, in exactly the same fashion as was done in the two-sided case, that the significance-level-α test

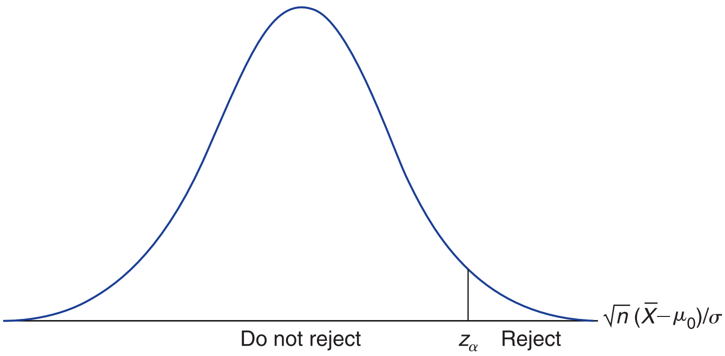


FIGURE 9.3
Testing $H_0\colon \mu \le \mu_0$ *against* $H_1\colon \mu > \mu_0$.

is to

$$\begin{array}{ll} \text{Reject } H_0 & \text{if } \sqrt{n}\,\dfrac{\overline{X}-\mu_0}{\sigma} \ge z_\alpha \\ \text{Not reject } H_0 & \text{otherwise} \end{array}$$

A pictorial depiction of this test is shown in Fig. 9.3.

This test can be carried out alternatively by first computing the value of the test statistic $\sqrt{n}\,(\overline{X}-\mu_0)/\sigma$. The p value is then equal to the probability that a standard normal random variable is at least as large as this value. That is, if the value of the test statistic is v, then

$$p \text{ value } = P\{Z \ge v\}$$

The null hypothesis is then rejected at any significance level greater than or equal to the p value.

In similar fashion, we can test the null hypothesis

$$H_0\colon \mu \ge \mu_0$$

against the alternative

$$H_1\colon \mu < \mu_0$$

by first computing the value of the test statistic $\sqrt{n}\,(\overline{X}-\mu_0)/\sigma$. The p value then equals the probability that a standard normal is less than or equal to this value, and the null hypothesis is rejected if the significance level is at least as large as the p value.

■ Example 9.4

All cigarettes presently being sold have an average nicotine content of at least 1.5 milligrams per cigarette. A firm that produces cigarettes claims that it has discovered a new technique for curing tobacco leaves that results in an average nicotine content of a cigarette of less than 1.5 milligrams. To test this claim, a sample of 20 of the firm's cigarettes was analyzed. If it were known that the standard deviation of a cigarette's nicotine content was 0.7 milligrams, what conclusions could be drawn, at the 5 percent level of significance, if the average nicotine content of these 20 cigarettes were 1.42 milligrams?

Solution

To see if the results establish the firm's claim, let us see if they would lead to rejection of the hypothesis that the firm's cigarettes do not have an average nicotine content lower than 1.5 milligrams. That is, we should test

$$H_0\colon \mu \geq 1.5$$

against the firm's claim of

$$H_1\colon \mu < 1.5$$

Since the value of the test statistic is

$$\sqrt{n}\,\frac{\overline{X} - \mu_0}{\sigma} = \sqrt{20}\,\frac{1.42 - 1.5}{0.7} = -0.511$$

it follows that the p value is

$$p \text{ value} = P\{Z \leq -0.511\} = 0.305$$

Since the p value exceeds 0.05, the foregoing data do not enable us to reject the null hypothesis and conclude that the mean content per cigarette is less than 1.5 milligrams. In other words, even though the evidence supports the cigarette producer's claim (since the average nicotine content of those cigarettes tested was indeed less than 1.5 milligrams), that evidence is not strong enough to *prove* the claim. This is because a result at least as supportive of the alternative hypothesis H_1 as that obtained would be expected to occur 30.5 percent of the time when the mean nicotine content was 1.5 milligrams per cigarette. ■

Statistical hypothesis tests in which either the null or the alternative hypothesis states that a parameter is greater (or less) than a certain value are called *one-sided* tests.

Table 9.1 Hypothesis Tests Concerning the Mean μ of a Normal Population with Known Variance σ^2.

$X_1, \ldots, X_n$ are sample data, and $\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$

H_0	H_1	Test statistic TS	Significance-level-α test	p Value if TS $= v$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\sqrt{n}\,\frac{\overline{X}-\mu_0}{\sigma}$	Reject H_0 if $\lvert TS \rvert \geq z_{\alpha/2}$ Do not reject H_0 otherwise	$2P\{Z \geq \lvert v \rvert\}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\sqrt{n}\,\frac{\overline{X}-\mu_0}{\sigma}$	Reject H_0 if $TS \geq z_\alpha$ Do not reject H_0 otherwise	$P\{Z \geq v\}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\sqrt{n}\,\frac{\overline{X}-\mu_0}{\sigma}$	Reject H_0 if $TS \leq -z_\alpha$ Do not reject H_0 otherwise	$P\{Z \leq v\}$

We have assumed so far that the underlying population distribution is the normal distribution. However, we have only used this assumption to conclude that $\sqrt{n}\left(\overline{X} - \mu\right)/\sigma$ has a standard normal distribution. But by the central limit theorem this result will approximately hold, no matter what the underlying population distribution, as long as n is reasonably large. A rule of thumb is that a sample size of $n \geq 30$ will almost always suffice. Indeed, for many population distributions, a value of n as small as 4 or 5 will result in a good approximation. Thus, all the hypothesis tests developed so far can often be used even when the underlying population distribution is not normal.

Table 9.1 summarizes the tests presented in this section.

PROBLEMS

1. The weights of salmon grown at a commercial hatchery are normally distributed with a standard deviation of 1.2 pounds. The hatchery claims that the mean weight of this year's crop is at least 7.6 pounds. Suppose a random sample of 16 fish yielded an average weight of 7.2 pounds. Is this strong enough evidence to reject the hatchery's claims at the
   (a) 5 percent level of significance?
   (b) 1 percent level of significance?
   (c) What is the p value?
2. Consider a test of H_0: $\mu \leq 100$ versus H_1: $\mu > 100$. Suppose that a sample of size 20 has a sample mean of $\overline{X} = 105$. Determine the p value of this outcome if the population standard deviation is known to equal
   (a) 5
   (b) 10
   (c) 15

Statistics in Perspective

Three Mile Island

A still-unsettled question is whether the nuclear accident at Three Mile Island, which released low-level nuclear radiation into the areas surrounding it, is responsible for an increase in the number of cases of hyperthyroidism. Hyperthyroidism, which results when the thyroid gland is malfunctioning, can lead to mental retardation if it is not treated quickly. It has been reported that 11 babies suffering from hyperthyroidism were born in the surrounding areas between March 28, 1979 (the day of the accident), and December 28, 1979 (nine months later). In addition, it was reported that the normal number of such babies to be born in the surrounding areas over a 9-month period is approximately normally distributed with a mean approximately equal to 3 and a standard deviation approximately equal to 2. Given this information, let us start by determining the probability that such a large number of cases of hyperthyroidism as 11 could have occurred by chance.

To begin, note that if the accident did not have any health effects and if the 9 months following the accident were ordinary months, then the number of newborn babies suffering from hyperthyroidism should have an approximately normal distribution with mean 3 and standard deviation 2. On the other hand, if the accident had a deleterious effect on hyperthyroidism, then the mean of the distribution would be larger than 3. Hence, let us suppose that the data come from a normal distribution with standard deviation 2 and use them to test

$$H_0\colon \mu \le 3 \quad \text{against} \quad H_1\colon \mu > 3$$

where μ is the mean number of newborns who suffer from hyperthyroidism.

Since the observed number is 11, the p value of these data is

$$\begin{aligned} p \text{ value} &= P\{X \ge 11\} \\ &= P\{X \ge 10.5\} \quad \text{continuity correction} \\ &= P\left\{\frac{X-3}{2} \ge \frac{10.5-3}{2}\right\} \\ &\approx P\{Z \ge 3.75\} \\ &< 0.0001 \end{aligned}$$

Thus the null hypothesis would be rejected at the 1 percent (or even at the 0.1 percent) level of significance.

It is important to note that this test does *not* prove that the nuclear accident was the cause of the increase in hyperthyroidism; and in fact it *does not even prove that there was an increase in this disease*. Indeed, it is hard to know what can be concluded from this test without having a great deal more information. For instance, one difficulty results from our not knowing why the particular hypothesis considered was chosen to be studied. That is, was there some prior scientific reason for believing that a release of nuclear radiation might result in increased hyperthyroidism in newborns, or did someone just check all possible diseases he could think of (and possibly for a variety of age groups) and then test whether there was a significant change in its incidence after the accident? The trouble with such an approach (which is often called

Statistics in Perspective (continued)

data mining, or *going on a fishing expedition*) is that even if no real changes resulted from the accident, just by chance some of the many tests might yield a significant result. (For instance, if 20 independent hypothesis tests are run, then even if all the null hypotheses are true, at least one of them will be rejected at the 1 percent level of significance with probability $1 - (0.99)^{20} = 0.18$.)

Another difficulty in interpreting the results of our hypothesis test concerns the confidence we have in the numbers given to us. For instance, can we really be certain that under normal conditions the mean number of newborns suffering from hyperthyroidism is equal to 3? Is it not more likely that whereas on average 3 newborns would normally be diagnosed to be suffering from this disease, other newborn sufferers may go undetected? Would there not be a much smaller chance that a sufferer would fail to be diagnosed as being such in the period following the accident, given that everyone was alert for such increases in that period? Also, perhaps there are degrees of hyperthyroidism, and a newborn diagnosed as being a sufferer in the tense months following the accident would not have been so diagnosed in normal times.

Note that we are not trying to argue that there was not a real increase in hyperthyroidism following the accident at Three Mile Island. Rather, we are trying to make the reader aware of the potential difficulties in correctly evaluating a statistical study.

3. Repeat Prob. 2, this time supposing that the value of the sample mean is 108.
4. It is extremely important in a certain chemical process that a solution to be used as a reactant have a pH level greater than 8.40. A method for determining pH that is available for solutions of this type is known to give measurements that are normally distributed with a mean equal to the actual pH and with a standard deviation of 0.05. Suppose 10 independent measurements yielded the following pH values:

   $$8.30, 8.42, 8.44, 8.32, 8.43, 8.41, 8.42, 8.46, 8.37, 8.42$$

   Suppose it is a very serious mistake to run the process with a reactant having a pH level less than or equal to 8.40.
   (a) What null hypothesis should be tested?
   (b) What is the alternative hypothesis?
   (c) Using the 5 percent level of significance, what would you advise—to use or not to use the solution?
   (d) What is the p value of the hypothesis test?
5. An advertisement for a toothpaste claims that use of the product significantly reduces the number of cavities of children in their cavity-prone years. Cavities per year for this age group are normal with mean 3 and standard deviation 1. A study of 2500 children who used this toothpaste found an average of 2.95 cavities per child. Assume that the standard deviation of the number of cavities of a child using this new toothpaste remains equal to 1.

(a) Are these data strong enough, at the 5 percent level of significance, to establish the claim of the toothpaste advertisement?
(b) Is this a significant enough reason for your children to switch to this toothpaste?

6. A farmer claims to be able to produce larger tomatoes. To test this claim, a tomato variety that has a mean diameter size of 8.2 centimeters with a standard deviation of 2.4 centimeters is used. If a sample of 36 tomatoes yielded a sample mean of 9.1 centimeters, does this prove that the mean size is indeed larger? Assume that the population standard deviation remains equal to 2.4, and use the 5 percent level of significance.
7. Suppose that the cigarette firm is now, after the test described in Example 9.4, even more convinced about its claim that the mean nicotine content of its cigarettes is less than 1.5 milligrams per cigarette. Would you suggest another test? With the same sample size?
8. The following data come from a normal population having standard deviation 4:

   105, 108, 112, 121, 100, 105, 99, 107, 112, 122, 118, 105

   Use them to test the null hypothesis that the population mean is less than or equal to 100 at the
   (a) 5 percent level of significance
   (b) 1 percent level of significance
   (c) What is the p value?
9. A soft drink company claims that its machines dispense, on average, 6 ounces per cup with a standard deviation of 0.14 ounces. A consumer advocate is skeptical of this claim, believing that the mean amount dispensed is less than 6 ounces. To gain information, a sample of size 100 is chosen. If the average amount per cup is 5.6 ounces, what conclusions can be drawn? State the null and alternative hypotheses, and give the p value.
10. The significance-level-α test of

$$H_0: \mu = \mu_0 \quad \text{against} \quad H_1: \mu > \mu_0$$

   is the same as the one for testing

$$H_0: \mu \leq \mu_0 \quad \text{against} \quad H_1: \mu > \mu_0$$

   Does this seem reasonable to you? Explain!

9.4 THE t TEST FOR THE MEAN OF A NORMAL POPULATION: CASE OF UNKNOWN VARIANCE

We have previously assumed that the only unknown parameter of the normal population distribution is its mean. However, by far the more common case is when the standard deviation σ is also unknown. In this section we will show how to perform hypothesis tests of the mean in this situation.

To begin, suppose that we are about to observe the results of a sample of size n from a normal population having an unknown mean μ and an unknown standard deviation σ, and suppose that we are interested in using the data to test the null hypothesis

$$H_0\colon \mu = \mu_0$$

against the alternative

$$H_1\colon \mu \neq \mu_0$$

As in the previous section, it again seems reasonable to reject H_0 when the point estimator of the population mean μ—that is, the sample mean $\overline{X}$—is far from μ_0. However, how far away it needs to be to justify rejection of H_0 was shown in Sec. 9.3 to depend on the standard deviation σ. Specifically, we showed that a significance-level-α test called for rejecting H_0 when $|\overline{X} - \mu_0|$ was at least $z_{\alpha/2}\sigma/\sqrt{n}$ or, equivalently, when

$$\frac{\sqrt{n}|\overline{X} - \mu_0|}{\sigma} \geq z_{\alpha/2}$$

Now, when σ is no longer assumed to be known, it is reasonable to estimate it by the sample standard deviation S, given by

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

and to employ a test that calls for rejecting H_0 when the absolute value of the test statistic T is large, where

$$T = \sqrt{n}\frac{\overline{X} - \mu_0}{S}$$

To determine how large $|T|$ needs to be to justify rejection at the α level of significance, we need to know its probability distribution when H_0 is true. However, as noted in Sec. 8.6, when $\mu = \mu_0$, the statistic T has a t distribution with $n-1$ degrees of freedom. Since the absolute value of such a random variable will exceed $t_{n-1,\alpha/2}$ with probability α (see Fig. 9.4), it follows that a significance-level-α test of

$$H_0\colon \mu = \mu_0 \quad \text{versus} \quad H_1\colon \mu \neq \mu_0$$

is, when σ is unknown, to

$$\begin{array}{ll} \text{Reject } H_0 & \text{if } |T| \geq t_{n-1,\alpha/2} \\ \text{Not reject } H_0 & \text{otherwise} \end{array}$$

This test, which is pictorially illustrated in Fig. 9.5, is called a *two-sided t test*.

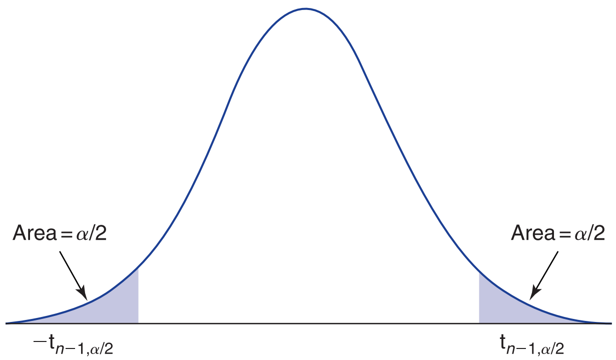


FIGURE 9.4
$P\{|T_{n-1}| \geq t_{n-1,\alpha/2}\} = \alpha$.

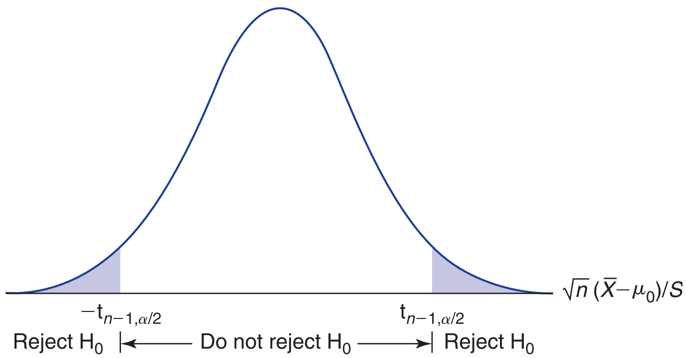


FIGURE 9.5
The significance-level-α two-sided t test.

If we let v denote the value of the test statistic $T = \sqrt{n}\,(\bar{X} - \mu_0)/S$, then the p value of the data is the probability that the absolute value of a t random variable having $n - 1$ degrees of freedom will be as large as $|v|$, which is equal to twice the probability that a t random variable with $n - 1$ degrees of freedom will be as large as $|v|$. (That is, the p value is the probability that a value of the test statistic at least as large as the one obtained would have occurred if the null hypothesis were true.) The test then calls for rejection at all significance levels that are at least as large as the p value.

If the value of the test statistic is v, then

$$\begin{aligned} p \text{ value} &= P\{|T_{n-1}| \geq |v|\} \\ &= 2P\{T_{n-1} \geq |v|\} \end{aligned}$$

where T_{n-1} is a t random variable with $n - 1$ degrees of freedom.

■ Example 9.5

Among a clinic's patients having high blood cholesterol levels of at least 240 milliliters per deciliter of blood serum, volunteers were recruited to test a new drug designed to reduce blood cholesterol. A group of 40 volunteers were given the drug for 60 days, and the changes in their blood cholesterol levels were noted. If the average change was a decrease of 6.8 with a sample standard deviation of 12.1, what conclusions can we draw? Use the 5 percent level of significance.

Solution

Let us begin by testing the hypothesis that any changes in blood cholesterol levels were due purely to chance. That is, let us use the data to test the null hypothesis

$$H_0\colon \mu = 0 \quad \text{versus} \quad H_1\colon \mu \neq 0$$

where μ is the mean decrease in cholesterol. The value of the test statistic T is

$$T = \frac{\sqrt{n}(\overline{X} - \mu_0)}{S} = \frac{\sqrt{40}(6.8)}{12.1} = 3.554$$

Since, from Table D.2, $t_{39,0.025} = 2.02$, the null hypothesis is rejected at the 5 percent level of significance. In fact the p value of the data is given by

$$\begin{aligned} p \text{ value} &= 2P\{T_{39} > 3.554\} \\ &= 0.0001 \qquad \text{from Program 8-2} \end{aligned}$$

Thus, at any significance level greater than 0.0001, we reject the hypothesis that the change in levels is due solely to chance.

However, note that we would not be justified at this point in concluding that the changes in cholesterol levels are due to the specific drug used and not to some other possibility. For instance, it is well known that any medication received by a patient (whether or not this medication is directly relevant to the patient's suffering) often leads to an improvement in the patient's condition (*the placebo effect*). Also other factors might be involved that could have caused the reduction in blood cholesterol levels; for instance, weather conditions during the testing period might conceivably have affected these levels.

Indeed, it must be concluded that the preceding testing scheme was very poorly designed for learning about the effectiveness of the drug, for in order to test whether a particular treatment has an effect on a disease that may be affected by many things, it is necessary to design an experiment that neutralizes all other possible causes of change except for the drug. The accepted approach for accomplishing this is to divide the volunteers at random into two groups: One group is to receive the drug, and the other group (the *control* group) is to receive a placebo (that is, a tablet, that looks and tastes like the actual drug but that has no physiological effect). The volunteers should not be told whether they are in the actual group or in the control group. Indeed, it is best if even the clinicians do not have this information (such tests are called *double-blind*) so as not to allow their own hopes and biases to play a role in their before-and-after evaluations of the patients. Since the two groups are chosen at random from the volunteers, we can now hope that on average all factors affecting the two groups will be the same except that one group received the actual drug and the other received a placebo. Hence, any difference in performance between the two groups can be attributed to the drug. ■

Program 9-1 computes the value of the test statistic T and the corresponding p value. It can be applied for both one- and two-sided tests. (The one-sided tests will be presented shortly.)

■ Example 9.6

Historical data indicate that the mean acidity (pH) level of rain in a certain industrial region in West Virginia is 5.2. To see whether there has been any recent change in this value, the acidity levels of 12 rainstorms over the past year have been measured, with the following results:

6.1, 5.4, 4.8, 5.8, 6.6, 5.3, 6.1, 4.4, 3.9, 6.8, 6.5, 6.3

Are these data strong enough, at the 5 percent level of significance, for us to conclude that the acidity of the rain has changed from its historical value?

Solution

To test the hypothesis of no change in acidity, that is, to test

$$H_0\colon \mu = 5.2 \quad \text{versus} \quad H_1\colon \mu \neq 5.2$$

first we compute the value of the test statistic T. Now, a simple calculation using the given data yields for the values of the sample mean and sample standard deviation,

$$\overline{X} = 5.667 \quad \text{and} \quad S = 0.921$$

Thus, the value of the test statistic is

$$T = \sqrt{12}\,\frac{5.667 - 5.2}{0.921} = 1.76$$

Since, from Table D.2 of App. D, $t_{11,0.025} = 2.20$, the null hypothesis is not rejected at the 5 percent level of significance. That is, the data are not strong enough to enable us to conclude, at the 5 percent level of significance, that the acidity of the rain has changed.

We could also have solved this problem by computing the p value by running Program 9-1 as follows:

The value of mu-zero is 5.2
The sample size is 12
The data values are 6.1, 5.4, 4.8, 5.8, 6.6, 5.3, 6.1, 4.4, 3.9, 6.8, 6.5, and 6.3
The program computes the value of the t-statistic as 1.755621
The p-value is 0.1069365

Thus, the p value is 0.107, and so the null hypothesis would not be rejected even at the 10 percent level of significance. ■

Suppose now that we want to test the null hypothesis

$$H_0\colon \mu \le \mu_0$$

against the alternative

$$H_1\colon \mu > \mu_0$$

In this situation, we want to reject the null hypothesis that the population mean is less than or equal to μ_0 only when the test statistic

$$T = \sqrt{n}\,\frac{\overline{X} - \mu_0}{S}$$

is significantly large (for this will tend to occur when the sample mean is significantly larger than μ_0). Therefore, we obtain the following significance-level-α test:

$$\begin{array}{ll} \text{Reject } H_0 & \text{if } T \ge t_{n-1,\alpha} \\ \text{Do not reject } H_0 & \text{otherwise} \end{array}$$

A pictorial depiction of the test is shown in Fig. 9.6.

Equivalently, the preceding test can be performed by first computing the value of the test statistic T, say its value is v, and then computing the p value, which is equal to the probability that a t random variable with $n-1$ degrees of freedom will be at least as large as v. That is, if $T = v$, then

$$p \text{ value} = P\{T_{n-1} \ge v\}$$

If we want to test the hypothesis

$$H_0\colon \mu \ge \mu_0 \quad \text{versus} \quad H_1\colon \mu < \mu_0$$

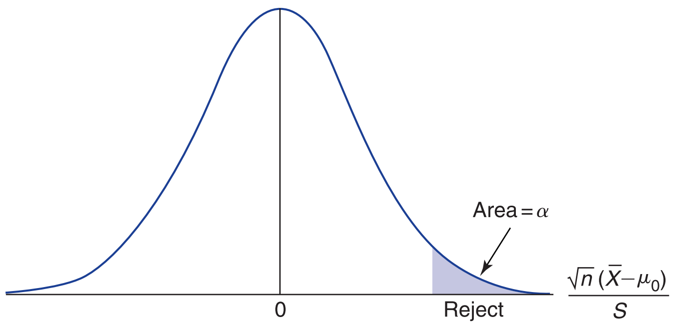


FIGURE 9.6
Testing $H_0: \mu \leq \mu_0$ *against* $H_1: \mu > \mu_0$.

then the test is analogous. The significance-level-α test is again based on the test statistic

$$T = \sqrt{n}\frac{\overline{X} - \mu_0}{S}$$

and the test is as follows:

Reject H_0	if $T \leq -t_{n-1,\alpha}$
Do not reject H_0	otherwise

In addition, the p value equals the probability that a t random variable with $n - 1$ degrees of freedom is less than or equal to the observed value of T.

Program 9-1 will compute the value of the test statistic T and the resulting p value. If only summary data are provided, then Program 8-2, which computes probabilities concerning t random variables, can be employed.

■ Example 9.7

The manufacturer of a new fiberglass tire claims that the average life of a set of its tires is at least 50,000 miles. To verify this claim, a sample of 8 sets of tires was chosen, and the tires subsequently were tested by a consumer agency. If the resulting values of the sample mean and sample variance were, respectively, 47.2 and 3.1 (in 1000 miles), test the manufacturer's claim.

Solution

To determine whether the foregoing data are consistent with the hypothesis that the mean life is at least 50,000 miles, we will test

$$H_0: \mu \geq 50 \quad \text{versus} \quad H_1: \mu < 50$$

Statistics in Perspective

What is the Appropriate Null Hypothesis?

Suppose that the television tubes produced by a certain manufacturer are known to have mean lifetimes of 3000 hours of use. An outside consultant claims that a new production method will lead to a greater mean life. To check this, a pilot program is designed to produce a sample of tubes by the newly suggested approach. How should the manufacturer use the resulting data?

At first glance, it might appear that the data should be used to test

$$H_0: \mu \leq 3000 \quad \text{against} \quad H_1: \mu > 3000$$

Then a rejection of H_0 would be strong evidence that the newly proposed approach resulted in an improved tube. However, the trouble with testing this hypothesis is that if the sample size is large enough, then there is a reasonable chance of rejecting H_0 even in cases where the new mean life is only, say, 3001 hours, and it might not be economically feasible to make the changeover for such a small increase in mean life. Indeed, the data should be used to test

$$H_0: \mu \leq 3000 + c$$

against

$$H_1: \mu > 3000 + c$$

where c is the smallest increase in mean life that would make it economically feasible to make the production change. ■

A rejection of the null hypothesis H_0 would then discredit the claim of the manufacturer. The value of the test statistic T is

$$T = \sqrt{8}\,\frac{47.2 - 50}{3.1} = -2.55$$

Since $t_{7,0.05} = 1.895$ and the test calls for rejecting H_0 when T is less than or equal to $-t_{7,\alpha}$, it follows that H_0 is rejected at the 5 percent level of significance. On the other hand, since $t_{7,0.01} = 2.998$, H_0 would not be rejected at the 1 percent level. Running Program 8-2 shows that the p value is equal to 0.019, illustrating that the data strongly indicate that the manufacturer's claim is invalid. ■

The t test can be used even when the underlying distribution is not normal, provided the sample size is reasonably large. This is true because, by the central limit theorem, the sample mean $\overline{X}$ will be approximately normal no matter what the population distribution and because the sample standard deviation S will approximately equal σ. Indeed, since for large n the t distribution with $n - 1$ degrees of freedom is almost identical to the standard normal, the foregoing is equivalent to noting that $\sqrt{n}\,(\overline{X} - \mu_0)/S$ will have an approximately standard normal distribution when μ_0 is the population mean and the sample size n is large.

Table 9.2 summarizes the tests presented in this section.

Table 9.2 Hypothesis Tests Concerning the Mean μ of a Normal Population with Unknown Variance σ^2.

$X_1, \ldots, X_n$ are sample data; $\overline{X} = \frac{\sum_{i=1}^n X_i}{n}$ $S^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{n-1}$

H_0	H_1	Test statistic TS	Significance-level-α test	p Value if TS $= v$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\sqrt{n}\,\frac{\overline{X}-\mu_0}{S}$	Reject H_0 if $\lvert TS \rvert \geq t_{n-1,\alpha/2}$ Do not reject otherwise	$2P\{T_{n-1} = \lvert v \rvert\}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\sqrt{n}\,\frac{\overline{X}-\mu_0}{S}$	Reject H_0 if $TS \geq t_{n-1,\alpha}$ Do not reject otherwise	$P\{T_{n-1} \geq v\}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\sqrt{n}\,\frac{\overline{X}-\mu_0}{S}$	Reject H_0 if $TS \leq -t_{n-1,\alpha}$ Do not reject H_0 otherwise	$P\{T_{n-1} \leq v\}$

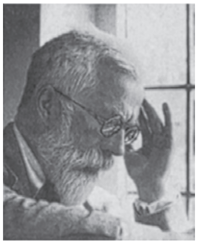
T_{n-1} is a t random variable with $n-1$ degrees of freedom, and $t_{n-1,\alpha}$ and $t_{n-1,\alpha/2}$ are such that $P\{T_{n-1} \geq t_{n-1,\alpha}\} = \alpha$ and $P\{T_{n-1} \geq t_{n-1,\alpha/2}\} = \alpha/2$.

PROBLEMS

1. There is some variability in the amount of phenobarbital in each capsule sold by a manufacturer. However, the manufacturer claims that the mean value is 20.0 milligrams. To test this, a sample of 25 pills yielded a sample mean of 19.7 with a sample standard deviation of 1.3. What inference would you draw from these data? In particular, are the data strong enough evidence to discredit the claim of the manufacturer? Use the 5 percent level of significance.
2. A fast-food establishment has been averaging about \$2000 of business per weekday. To see whether business is changing due to a deteriorating economy (which may or may not be good for the fast-food industry), management has decided to carefully study the figures for the next 8 days. Suppose the figures are

   2050, 2212, 1880, 2121, 2205, 2018, 1980, 2188

   (a) What are the null and the alternative hypotheses?
   (b) Are the data significant enough, at the 5 percent level, to prove that a change has occurred?
   (c) What about at the 1 percent level?
   (d) If you can run Program 9-1 or some equivalent software, find the p value.
3. To test the hypothesis that a normal population has mean 100, a random sample of size 10 is chosen. If the sample mean is 110, will you reject the null hypothesis if the following is known?



(A. Barrington/Science Photo Library/Photo Researchers)

Ronald A. Fisher

William S. Gosset

Historical Perspective

In 1908, William Sealy Gosset, writing under the name *Student*, published the distribution of the t statistic $\sqrt{n}\,(\overline{X} - \mu)/S$. It was an important result, for it enabled one to make tests of population means when only small samples were available, as was often the case at the Guinness brewery, where Gosset was employed. Its importance, however, was not noted, and it was mainly ignored by the statistical community at the time. This was primarily because the idea of learning from small samples went against the prevailing scientific beliefs, which were that "if your sample was sufficiently large, then substitute S for σ and use the normal distribution, and if your sample was not sufficiently large, then do not apply statistics." One of the few to realize its importance was R. A. Fisher, who in a later paper refined and fixed some technical errors in Gosset's work. However, it was not until Fisher's book *Statistical Methods for Research Workers* appeared in 1925 that the t test became widely used and appreciated. Fisher's book was a tremendous success, and it went through 11 editions in its first 25 years. While it was extremely influential, it was, like Fisher's other writings, not easy to read. Indeed, it was said by a coworker at the time that "No student should attempt to read it who has not read it before."

(Note: Photograph of Gosset from Student: A Statistical Biography of William Sealy Gosset. Based on writings by E. S. Pearson, edited and augmented by R. L. Plackett with the assistance of G. A. Barnard. Clarendon Press, Oxford, 1990. Photograph from Annals of Eugenics, 1939, vol. 9.)

(a) The population standard deviation is known to equal 15.
(b) The population standard deviation is unknown, and the sample standard deviation is 15.

Use the 5 percent level of significance.

4. The number of lunches served daily at a school cafeteria last year was normally distributed with mean 300. The menu has been changed this year to healthier foods, and the administration wants to test the hypothesis that the mean number of lunches sold is unchanged. A sample of 12 days yielded the following number of lunches sold:

312, 284, 281, 295, 306, 273, 264, 258, 301, 277, 280, 275

Is the hypothesis that the mean is equal to 300 rejected at the
(a) 10 percent
(b) 5 percent
(c) 1 percent
level of significance?

5. An oceanographer wants to check whether the average depth of the ocean in a certain region is 55 fathoms, as had been previously reported. He took soundings at 36 randomly chosen locations in the region and obtained a sample mean of 56.4 fathoms with a sample standard deviation of 5.1 fathoms. Are these data significant enough to reject the null hypothesis that the mean depth is 55 fathoms, at the
   (a) 10 percent
   (b) 5 percent
   (c) 1 percent

   level of significance?
6. Twenty years ago, entering first-year high school students could do an average of 24 push-ups in 60 seconds. To see whether this remains true today, a random sample of 36 first-year students was chosen. If their average was 22.5 with a sample standard deviation of 3.1, can we conclude that the mean is no longer equal to 24? Use the 5 percent level of significance.
7. The mean response time of a species of pigs to a stimulus is 0.8 seconds. Twenty-eight pigs were given 2 ounces of alcohol and then tested. If their average response time was 1.0 seconds with a standard deviation of 0.3 seconds, can we conclude that alcohol affects the mean response time? Use the 5 percent level of significance.
8. Previous studies have shown that mice gain, on average, 5 grams of weight during their first 10 days after being weaned. A group of 36 mice had the artificial sweetener aspartame added to their food. Their average gain was 4.5 grams, with a sample standard deviation of 0.9 grams. Can we conclude, at the 1 percent level, that the addition of aspartame had an effect?
9. Use the results of last Sunday's National Football League (NFL) professional games to test the hypothesis that the average number of points scored by winning teams is 26.2. Use the 5 percent level of significance.
10. Use the results of last Sunday's major league baseball scores to test the hypothesis that the average number of runs scored by winning teams is 5.6. Use the 5 percent level of significance.
11. A bakery was taken to court for selling loaves of bread that were underweight. These loaves were advertised as weighing 24 ounces. In its defense, the bakery claimed that the advertised weight was meant to imply not that each loaf weighed exactly 24 ounces, but rather that the average value over all loaves was 24 ounces. The prosecution in a rebuttal produced evidence that a randomly chosen sample of 20 loaves had an average weight of 22.8 ounces with a sample standard deviation of 1.4 ounces. In her ruling, the judge stated that advertising a weight of 24 ounces would be acceptable if the mean weight were at least 23 ounces.

(a) What hypothesis should be tested?
(b) For the 5 percent level of significance, what should the judge rule?

12. A recently published study claimed that the average academic year salary of full professors at colleges and universities in the United States is \$87,800. Students at a certain private school guess that the average salary of their professors is higher than this figure and so have decided to test the null hypothesis

$$H_0: \mu \leq 87,800 \quad \text{against} \quad H_1: \mu > 87,800$$

where μ is the average salary of full professors at their school. A random sample of 10 professors elicited the following salaries (in units of \$1000):

91.0, 79.8, 102.0, 93.5, 82.0, 88.6, 90.0, 98.6, 101.0, 84.0

(a) Is the null hypothesis rejected at the 10 percent level of significance?
(b) What about at the 5 percent level?

13. A car is advertised as getting at least 31 miles per gallon in highway driving on trips of at least 100 miles. Suppose the miles per gallon obtained in 8 independent experiments (each consisting of a nonstop highway trip of 100 miles) are

28, 29, 31, 27, 30, 35, 25, 29

(a) If we want to check if these data disprove the advertising claim, what should we take as the null hypothesis?
(b) What is the alternative hypothesis?
(c) Is the claim disproved at the 5 percent level of significance?
(d) What about at the 1 percent level?

14. A manufacturer claims that the mean lifetime of the batteries it produces is at least 250 hours of use. A sample of 20 batteries yielded the following data:

237, 254, 255, 239, 244, 248, 252, 255, 233, 259, 236,
232, 243, 261, 255, 245, 248, 243, 238, 246

(a) Are these data consistent, at the 5 percent level, with the claim of the manufacturer?
(b) What about at the 1 percent level?

15. A water official insists that the average daily household water use in a certain county is at least 400 gallons. To check this claim, a random sample of 25 households was checked. The average of those sampled was 367 with a sample standard deviation of 62. Is this consistent with the official's claim?

16. A company supplies plastic sheets for industrial use. A new type of plastic has been produced, and the company would like to prove to an independent assessor that the average stress resistance of this new product is greater than 30.0, where stress resistance is measured in pounds per square inch necessary to crack the sheet. A random sample of size 12 yielded the following values of stress resistance:

30.1, 27.8, 32.2, 29.4, 24.8, 31.6, 28.8, 29.4, 30.5, 27.6, 33.9, 31.4

(a) Do these data establish that the mean stress resistance is greater than 30.0 pounds per square inch, at the 5 percent level of significance?
(b) What was the null hypothesis in part (a)?
(c) If the answer to (a) is no, do the data establish that the mean stress resistance is less than 30 pounds per square inch?

17. A medical scientist believes that the average basal temperature of (outwardly) healthy individuals has increased over time and is now greater than 98.6° F (37° C). To prove this, she has randomly selected 100 healthy individuals. If their mean temperature is 98.74° F with a sample standard deviation of 1.1° F, does this prove her claim at the 5 percent level? What about at the 1 percent level?

18. In 2001, entering students at a certain university had an average score of 542 on the verbal part of the SAT. A random sample of the scores of 20 students in the 2003 class resulted in these scores:

542, 490, 582, 511, 515, 564, 500, 602, 488, 512, 518,
522, 505, 569, 575, 515, 520, 528, 533, 515

Do the given data prove that the average score has decreased to below 542? Use the 5 percent level of significance.

9.5 HYPOTHESIS TESTS CONCERNING POPULATION PROPORTIONS

In this section we will consider tests concerning the proportion of members of a population that possess a certain characteristic. We suppose that the population is very large (in theory, of infinite size), and we let p denote the unknown proportion of the population with the characteristic. We will be interested in testing the null hypothesis

$$H_0\colon p \le p_0$$

against the alternative

$$H_1\colon p \ge p_0$$

for a specified value p_0.

If a random selection of n elements of the population is made, then X, the number with the characteristic, will have a binomial distribution with parameters n and p. Now it should be clear that we want to reject the null hypothesis that the proportion is less than or equal to p_0 only when X is sufficiently large. Hence, if the observed value of X is x, then the p value of these data will equal the probability that at least as large a value would have been obtained if p had been equal to p_0 (which is the largest possible value of p under the null hypothesis). That is, if we observe that X is equal to x, then

$$p \text{ value} = P\{X \ge x\}$$

where X is a binomial random variable with parameters n and p_0.

The p value can now be computed either by using the normal approximation or by running Program 5-1, which computes the binomial probabilities. The null hypothesis should then be rejected at any significance level that is greater than or equal to the p value.

■ Example 9.8

A noted educator claims that over half the adult U.S. population is concerned about the lack of educational programs shown on television. To gather data about this issue, a national polling service randomly chose and questioned 920 individuals. If 478 (52 percent) of those surveyed stated that they are concerned at the lack of educational programs on television, does this prove the claim of the educator?

Solution

To prove the educator's claim, we must show that the data are strong enough to reject the hypothesis that at most 50 percent of the population is concerned about the lack of educational programs on television. That is, if we let p denote the proportion of the population that is concerned about this issue, then we should use the data to test

$$H_0\colon p \le 0.50 \quad \text{versus} \quad H_1\colon p > 0.50$$

Since 478 people in the sample were concerned, it follows that the p value of these data is

$$\begin{aligned} p \text{ value} &= P\{X \geq 478\} \quad \text{when } X \text{ is binomial } (920, 0.50) \\ &= 0.1243 \qquad \text{from Program 5-1} \end{aligned}$$

For such a large p value we cannot conclude that the educator's claim has been proved. Although the data are certainly in support of that claim, since 52 percent of those surveyed were concerned by the lack of educational programs on television, such a result would have had a reasonable chance of occurring even if the claim were incorrect, and so the null hypothesis is not rejected.

If Program 5-1 were not available to us, then we could have approximated the p value by using the normal approximation to binomial probabilities. Since $np = 920(0.50) = 460$ and $np(1-p) = 460(0.5) = 230$, this would have yielded the following:

$$\begin{aligned} p \text{ value} &= P\{X \geq 478\} \\ &= P\{X \geq 477.5\} \quad \text{continuity correction} \\ &= P\left\{\frac{X - 460}{\sqrt{230}} \geq \frac{477.5 - 460}{\sqrt{230}}\right\} \\ &\approx P\{Z \geq 1.154\} = 0.1242 \end{aligned}$$

Thus the p value obtained by the normal approximation is quite close to the exact p value obtained by running Program 5-1. ■

For another type of example in which we are interested in a hypothesis test of a binomial parameter, consider a process that produces items that are classified as being either acceptable or defective. A common assumption is that each item produced is independently defective with a certain probability p, and so the number of defective items in a batch of size n will have a binomial distribution with parameters n and p.

■ Example 9.9

A computer chip manufacturer claims that at most 2 percent of the chips it produces are defective. An electronics company, impressed by that claim, has purchased a large quantity of chips. To determine if the manufacturer's claim is plausible, the company has decided to test a sample of 400 of these chips. If there are 13 defective chips (3.25 percent) among these 400, does this disprove (at the 5 percent level of significance) the manufacturer's claim?

Solution

If p is the probability that a chip is defective, then we should test the null hypothesis

$$H_0\colon p \le 0.02 \quad \text{against} \quad H_1\colon p > 0.02$$

That is, to see if the data disprove the manufacturer's claim, we must take that claim as the null hypothesis. Since 13 of the 400 chips were observed to be defective, the p value is equal to the probability that such a large number of defectives would have occurred if p were equal to 0.02 (its largest possible value under H_0). Therefore,

$$\begin{aligned} p \text{ value} &= P\{X \ge 13\} \quad && \text{where } X \text{ is binomial } (400, 0.02) \\ &= 0.0619 && \text{from Program 5-1} \end{aligned}$$

and so the data, though clearly not in favor of the manufacturer's claim, are not quite strong enough to reject that claim at the 5 percent level of significance.

If we had used the normal approximation, then we would have obtained the following result for the p value:

$$\begin{aligned} p \text{ value} &= P\{X \ge 13\} && \text{where } X \text{ is binomial } (400, 0.02) \\ &= P\{X \ge 12.5\} && \text{continuity correction} \\ &= P\left\{\frac{X-8}{\sqrt{8\,(0.98)}} \ge \frac{12.5-8}{\sqrt{8\,(0.98)}}\right\} \\ &\approx P\{Z \ge 1.607\} && \text{where } Z \text{ is standard normal} \\ &= 0.054 \end{aligned}$$

Thus, the approximate p value obtained by using the normal approximation, though not as close to the actual p value of 0.062 as we might have liked, is still accurate enough to lead to the correct conclusion that the data are not quite strong enough to reject the null hypothesis at the 5 percent level of significance. ■

Once again, let p denote the proportion of members of a large population who possess a certain characteristic, but suppose that we now want to test

$$H_0\colon p \ge p_0$$

against

$$H_0\colon p < p_0$$

for some specified value p_0. That is, we want to test the null hypothesis that the proportion of the population with the characteristic is at least p_0 against the alternative that it is less than p_0. If a random sample of n members of the population results in x of them having the characteristic, then the p value of these data is given by

$$p \text{ value} = P\{X \leq x\}$$

where X is a binomial random variable with parameters n and p_0.

That is, when the null hypothesis is that p is at least as large as p_0, then the p value is equal to the probability that a value as small as or smaller than the one observed would have occurred if p were equal to p_0.

9.5.1 Two-Sided Tests of p

Computation of the p value of the test data becomes slightly more involved when we are interested in testing the hypothesis

$$\mathrm{H}_0: p = p_0$$

against the two-sided alternative

$$\mathrm{H}_1: p \neq p_0$$

for a specified value p_0.

Again suppose that a sample of size n is chosen, and let X denote the number of members of the sample who possess the characteristic of interest. We will want to reject H_0 when X/n, the proportion of the sample with the characteristic, is either much smaller or much larger than p_0 or, equivalently, when X is either very small or very large in relation to np_0. Since we want the total probability of rejection to be less than or equal to α when p_0 is indeed the true proportion, we can attain these objectives by rejecting for both large and small values of X with probability, when H_0 is true, $\alpha/2$. That is, if we observe a value such that the probability is less than or equal to $\alpha/2$ that X would be either that large or that small when H_0 is true, then H_0 should be rejected.

Therefore, if the observed value of X is x, then H_0 will be rejected if either

$$P\{X \leq x\} \leq \frac{\alpha}{2}$$

or

$$P\{X \geq x\} \leq \frac{\alpha}{2}$$

when X is a binomial random variable with parameters n and p_0. Hence, the significance-level-α test will reject H_0 if

$$\mathrm{Min}\{P\{X \leq x\}, P\{X \geq x\}\} \leq \frac{\alpha}{2}$$

or, equivalently, if

$$2\,\mathrm{Min}\{P\{X \le x\}, P\{X \ge x\}\} \le \alpha$$

where X is binomial (n, p_0). From this, it follows that if x members of a random sample of size n have the characteristic, then the p value for the test of

$$\mathrm{H}_0\colon p = p_0 \quad \text{versus} \quad \mathrm{H}_1\colon p \ne p_0$$

is as follows:

$$p \text{ value} = 2\,\mathrm{Min}\{P\{X \le x\}, P\{X \ge x\}\}$$

where X is a binomial random variable with parameters n and p_0.

Since it will usually be evident which of the two probabilities in the expression for the p value will be smaller (if $x \le np_0$, then it will almost always be the first, and otherwise the second, probability), Program 5-1 or the normal approximation is needed only once to obtain the p value.

■ Example 9.10

Historical data indicate that 4 percent of the components produced at a certain manufacturing facility are defective. A particularly acrimonious labor dispute has recently been concluded, and management is curious about whether it will result in any change in this figure of 4 percent. If a random sample of 500 items indicated 16 defectives (3.2 percent), is this significant evidence, at the 5 percent level of significance, to conclude that a change has occurred?

Solution

To be able to conclude that a change has occurred, the data need to be strong enough to reject the null hypothesis when you are testing

$$\mathrm{H}_0\colon p = 0.04 \quad \text{versus} \quad \mathrm{H}_1\colon p \ne 0.04$$

where p is the probability that an item is defective. The p value of the observed data of 16 defectives in 500 items is

$$p \text{ value} = 2\,\mathrm{Min}\{P\{X \le 16\}, P\{X \ge 16\}\}$$

where X is a binomial (500, 0.04) random variable. Since $500 \times 0.04 = 20$, we see that

$$p \text{ value} = 2P\{X \le 16\}$$

Since X has mean 20 and standard deviation $\sqrt{20\,(0.96)} = 4.38$, it is clear that twice the probability that X will be less than or equal to 16—a value less than 1 standard deviation lower than the mean—is not going to be small enough to justify rejection. Indeed, it can be shown that

$$p \text{ value} = 2P\{X \leq 16\} = 0.432$$

and so there is not sufficient evidence to reject the hypothesis that the probability of a defective item has remained unchanged. ■

■ Example 9.11

Identical, also called monozygotic, twins form when a single fertilized egg splits into two genetically identical parts. The twins share the same DNA set, thus they may share many similar attributes. However, since physical appearance is influenced by environmental factors and not just genetics, identical twins can actually look very different. Fraternal, also called dizygotic, twins develop when two separate eggs are fertilized and implant in the uterus. The genetic connection of fraternal twins is no more nor less the same as siblings born at separate times. The literature states that 28 percent of all twin pairs are identical twins.

Suppose that a hypothetical doctor interested in testing whether 28 percent was accurate has decided to gather data on twins born in the hospital in which the doctor works. However, in obtaining permission to run such a study she discovers that finding out whether a gender similar twin pair is monozygotic requires a DNA test, which is both expensive and requires the permission of the twin-bearing parents. To avoid this expense she reasons that if p is the probability that twins are identical, then the probability that they will be of the same sex can be easily derived. Letting SS be the event that a twin pair is of the same sex, then conditioning on whether the pair is identical or not gives

$$\begin{aligned} P(SS) &= P(SS|\text{identical})P(\text{identical}) + P(SS|\text{fraternal})P(\text{fraternal}) \\ &= 1(p) + \frac{1}{2}(1-p) \\ &= \frac{1+p}{2} \end{aligned}$$

where the preceding used that fraternal twins, being genetically the same as any pair of siblings, would have one chance in two of being of the same sex. Thus, if $p = 0.28$, then

$$P(SS) = \frac{1.28}{2} = 0.64$$

Based on the preceding analysis the doctor has decided to test the hypothesis that the probability that a twin pair will be identical is 0.28 by testing whether the probability that a twin pair is of the same sex is 0.64. Assuming that data collected over one year by the researcher showed that 36 of 74 twin pairs were of the same sex, what conclusion can be drawn?

Solution

Let q be the probability that a twin pair is of the same sex. Then to test the hypothesis that 28 percent of all twin pairs are identical twins, the researcher will test the null hypothesis

$$H_0: q = 0.64 \quad \text{versus} \quad H_1: q \neq 0.64$$

Now, the number of the 74 twin pairs that are of the same sex has a binomial distribution with parameters 74 and q. Hence, the p value of the test of H_0 that results when 36 of 74 twin pairs are of the same sex is

$$p \text{ value} = 2 \min\{P\{X \leq 36\}, P\{X \geq 36\}\}$$

where X is a binomial (74, 0.64) random variable. Because $74 \times 0.64 = 47.36$, we see that

$$p \text{ value} = 2P\{X \leq 36\}$$

Using the normal approximation yields

$$\begin{aligned}
p \text{ value} &= 2P\{X \leq 36.5\} \\
&= 2P\left\{\frac{X - 74(0.64)}{\sqrt{74(0.64)(0.36)}} \leq \frac{36.5 - 74(0.64)}{\sqrt{74(0.64)(0.36)}}\right\} \\
&\approx 2P\left\{Z \leq \frac{36.5 - 74(0.64)}{\sqrt{74(0.64)(0.36)}}\right\} \\
&= 2P\{Z \leq -2.630\} \\
&= 2(1 - P\{Z \leq 2.630\}) \\
&= 0.0086
\end{aligned}$$

Thus the null hypothesis would be rejected at even the 1 percent level of significance. ■

Table 9.3 sums up the tests concerning the population proportion p.

Table 9.3 Hypothesis Tests Concerning p, the Proportion of a Large Population that Has a Certain Characteristic

The number of population members in a sample of size n that have the characteristic is X, and B is a binomial random variable with parameters n and p_0.

H_0	H_1	Test statistic TS	p Value if TS $= x$
$P \leq p_0$	$p > p_0$	X	$P\{B \geq x\}$
$P \geq p_0$	$p < p_0$	X	$P\{B \leq x\}$
$P = p_0$	$p \neq p_0$	X	$2\,\text{Min}\{P\{B \leq x\}, P\{B \geq x\}\}$

PROBLEMS

In solving the following problems, either make use of Program 5-1 or equivalent software to compute the relevant binomial probabilities, or use the normal approximation.

1. A standard drug is known to be effective in 72 percent of cases in which it is used to treat a certain infection. A new drug has been developed, and testing has found it to be effective in 42 cases out of 50. Is this strong enough evidence to prove that the new drug is more effective than the old one? Find the relevant p value.
2. An economist thinks that at least 60 percent of recently arrived immigrants who have been working in the health profession in the United States for more than 1 year feel that they are underemployed with respect to their training. Suppose a random sample of size 450 indicated that 294 individuals (65.3 percent) felt they were underemployed. Is this strong enough evidence, at the 5 percent level of significance, to prove that the economist is correct? What about at the 1 percent level of significance?
3. Shoplifting is a serious problem for retailers. In the past, a large department store found that 1 out of every 14 people entering the store engaged in some form of shoplifting. To help alleviate this problem, 3 months ago the store hired additional security guards. This additional hiring was widely publicized. To assess its effect, the store recently chose 300 shoppers at random and closely followed their movements by camera. If 18 of these 300 shoppers were involved in shoplifting, does this prove, at the 5 percent level of significance, that the new policy is working?
4. Let p denote the proportion of voters in a large city who are in favor of restructuring the city government, and consider a test of the hypothesis

$$H_0: p \geq 0.60 \quad \text{against} \quad H_1: p < 0.60$$

A random sample of n voters indicated that x are in favor of restructuring. In each of the following cases, would a significance-level-α test result in rejection of H_0?

(a) $n = 100$, $x = 50$, $\alpha = 0.10$
(b) $n = 100$, $x = 50$, $\alpha = 0.05$
(c) $n = 100$, $x = 50$, $\alpha = 0.01$
(d) $n = 200$, $x = 100$, $\alpha = 0.01$

Historical Perspective

The First Published Hypothesis Test "Proved" the Existence of God

Remarkably enough, the first published paper in which a statistical test was made of a null hypothesis was used to claim the existence of God. In a paper published in the *Philosophical Transactions of the Royal Society* in 1710, John Arbuthnot looked at the number of males and females born in each of the 82 years from 1629 to 1710, and he discovered that in each of these years the number of male births exceeded the number of female births. Arbuthnot argued that this could not have been due solely to chance, for if each birth were equally likely to be either a boy or a girl (and so each year would be equally likely to have either more male births or more female births), then the probability of the observed outcome would equal $(1/2)^{82}$. Thus, he argued, the hypothesis that the event occurred solely by chance must be rejected (in our language, the p value of the test of H_0: $p = 1/2$ versus H_1: $p \neq 1/2$ was $2(1/2)^{82}$). Arbuthnot then argued that the result must have been due to planning, and as he believed it was beneficial to initially have an excess of male babies, since males tend to do more hazardous work than females and thus tend to die earlier, he concluded that it was the work of God. (For reasons not totally understood, it appears that the probability of a newborn's being male is closer to 0.51 than it is to 0.50.)

5. A politician claims that over 50 percent of the population is in favor of her candidacy. To prove this claim, she has commissioned a polling organization to do a study. This organization chose a random sample of individuals in the population and asked each member of the sample if he or she was in favor of the politician's candidacy.

(a) To prove the politician's claim, what should be the null and alternative hypotheses?
Consider the following three alternative results, and give the relevant p values for each one.
(b) A random sample of 100 voters indicated that 56 (56 percent) are in favor of her candidacy.
(c) A random sample of 200 voters indicated that 112 (56 percent) are in favor of her candidacy.

(d) A random sample of 500 indicated that 280 (56 percent) are in favor of her candidacy.
Give an intuitive explanation for the discrepancy in results, if there are any, even though in each of cases (b), (c), and (d) the same percentage of the sample was in favor.

6. A revamped television news program has claimed to its advertisers that at least 24 percent of all television sets that are on when the program runs are tuned in to it. This figure of 24 percent is particularly important because the advertising rate increases at that level of viewers. Suppose a random sample indicated that 50 out of 200 televisions were indeed tuned in to the program.
   (a) Is this strong enough evidence, at the 5 percent significance level, to establish the accuracy of the claim?
   (b) Is this strong enough evidence, at the 5 percent significance level, to prove that the claim is unfounded?
   (c) Would you say that the results of this sample are evidence for or against the claim of the news program?
   (d) What do you think should be done next?

7. Three independent news services are running a poll to determine if over half the population supports an initiative concerning limitations on driving automobiles in the downtown area. Each news service wants to see if the evidence indicates that over half the population is in favor. As a result, all three services will be testing

$$H_0: p \leq 0.5 \quad \text{against} \quad H_1: p > 0.5$$

where p is the proportion of the population in favor of the initiative.
   (a) Suppose the first news organization samples 100 people, of whom 56 are in favor of the initiative. Is this strong enough evidence, at the 5 percent level of significance, to reject the null hypothesis and in doing so establish that over half the population favors the initiative?
   (b) Suppose the second news organization samples 120 people, of whom 68 are in favor of the initiative. Is this strong enough evidence, at the 5 percent level of significance, to reject the null hypothesis?
   (c) Suppose the third news organization samples 110 people, of whom 62 are in favor of the initiative. Is this strong enough evidence, at the 5 percent level of significance, to reject the null hypothesis?
   (d) Suppose the news organizations combine their samples, to come up with a sample of 330 people, of whom 186 support the initiative. Is this strong enough evidence, at the 5 percent level of significance, to reject the null hypothesis?

8. An ambulance service claims that at least 45 percent of its calls involve life-threatening emergencies. To check this claim, a random sample of 200 calls was selected from the service's files. If 70 of these calls involved life-threatening emergencies, is the service's claim believable
   (a) at the 5 percent
   (b) at the 1 percent

   level of significance?
9. A retailer has received a large shipment of items of a certain type. If it can be established that over 4 percent of the items in the shipment are defective, then the shipment will be returned. Suppose that 5 defectives are found in a random sample of 90 items. Should the shipment be returned to its sender? Use the 10 percent level of significance. What about at the 5 percent level?
10. A campus newspaper editorial claims that at least 75 percent of the students favor traditional course grades rather than a pass/fail option. To gain information, a dean randomly sampled 50 students and learned that 32 of them favor traditional grades. Are these data consistent with the claim made in the editorial? Use the 5 percent level of significance.
11. A recent survey published by the Higher Educational Research Institute stated that 22 percent of entering college students classified themselves as politically liberal. If 65 out of a random sample of 264 entering students at the University of California at Berkeley classified themselves as liberals, does this establish, at the 5 percent level of significance, that the percentage at Berkeley is higher than the national figure?
12. It has been "common wisdom" for some time that 22 percent of the population have a firearm at home. In a recently concluded poll, 54 out of 200 randomly chosen people were found to have a firearm in their homes. Is this strong enough evidence, at the 5 percent level of significance, to disprove common wisdom?
13. The average length of a red light is 30 seconds. Because of this, a certain individual feels lucky whenever he has to wait less than 15 seconds when encountering a red light. This individual assumes that the probability that he is lucky is 0.5. To test this hypothesis, he timed himself at 30 red lights. If he had to wait more than 15 seconds a total of 19 times, should he reject the hypothesis that p is equal to 0.5?
    (a) Use the 10 percent level of significance.
    (b) Use the 5 percent level of significance.
    (c) What is the p value?
14. A statistics student wants to test the hypothesis that a certain coin is equally likely to land on either heads or tails when it is flipped. The student flips the coin 200 times, obtaining 116 heads and 84 tails.

(a) For the 5 percent level of significance, what conclusion should be drawn?
(b) What are the null and the alternative hypotheses?
(c) What is the p value?

15. Twenty-five percent of women of child-bearing age smoke. A scientist wanted to test the hypothesis that this is also the proportion of smokers in the population of women who suffer ectopic pregnancies. To do so, the scientist chose a random sample of 120 women who had recently suffered an ectopic pregnancy. If 48 of these women turn out to be smokers, what is the p value of the test of the hypothesis

$$H_0: p = 0.25 \quad \text{against} \quad H_1: p \neq 0.25$$

where p is the proportion of smokers in the population of women who have suffered an ectopic pregnancy?

KEY TERMS

Statistical hypothesis: A statement about the nature of a population. It is often stated in terms of a population parameter.

Null hypothesis: A statistical hypothesis that is to be tested.

Alternative hypothesis: The alternative to the null hypothesis.

Test statistic: A function of the sample data. Depending on its value, the null hypothesis will be either rejected or not rejected.

Critical region: If the value of the test statistic falls in this region, then the null hypothesis is rejected.

Significance level: A small value set in advance of the testing. It represents the maximal probability of rejecting the null hypothesis when it is true.

Z test: A test of the null hypothesis that the mean of a normal population having a known variance is equal to a specified value.

***p* value**: The smallest significance level at which the null hypothesis is rejected.

One-sided tests: Statistical hypothesis tests in which either the null or the alternative hypothesis is that a population parameter is less than or equal to (or greater than or equal to) some specified value.

***t* test**: A test of the null hypothesis that the mean of a normal population having an unknown variance is equal to a specified value.

SUMMARY

A *statistical hypothesis* is a statement about the parameters of a population distribution.

The hypothesis to be tested is called the *null hypothesis* and is denoted by H_0. The *alternative hypothesis* is denoted by H_1.

A hypothesis test is defined by a *test statistic*, which is a function of the sample data, and a *critical region*. The null hypothesis is rejected if the value of the test statistic falls within the critical region and is not rejected otherwise. The critical region is chosen so that the probability of rejecting the null hypothesis, when it is true, is no greater than a predetermined value α, called the *significance level* of the test. The significance level is typically set equal to such values as 0.10, 0.05, and 0.01. The 5 percent level of significance, that is, $\alpha = 0.05$, has become the most common in practice.

Since the significance level is set to equal some small value, there is only a small chance of rejecting H_0 when it is true. Thus a statistical hypothesis test is basically trying to determine whether the data are consistent with a given null hypothesis. Therefore, rejecting H_0 is a strong statement that the null hypothesis does not appear to be consistent with the data, whereas not rejecting H_0 is a much weaker statement to the effect that H_0 is not inconsistent with the data. For this reason, the hypothesis that one is trying to establish should generally be designated as the alternative hypothesis so that it can be "statistically proved" by a rejection of the null hypothesis.

Often in practice a significance level is not set in advance, but rather the test statistic is observed to determine the minimal significance level that would result in a rejection of the null hypothesis. This minimal significance level is called the *p value*. Thus, once the p value is determined, the null hypothesis will be rejected at any significance level that is at least as large as the p value. The following rules of thumb concerning the p value are in rough use:

p value > 0.1	Data provide weak evidence against H_0.
p value ≈ 0.05	Data provide moderate evidence against H_0.
p value < 0.01	Data provide strong evidence against H_0.

1. *Testing* H_0: $\mu = \mu_0$ *against* H_1: $\mu \neq \mu_0$ *in a normal population having known standard deviation* σ: The significance-level-α test is based on the test statistic

$$\sqrt{n}\frac{\overline{X} - \mu_0}{\sigma}$$

and it is to

$$\begin{array}{ll} \text{Reject } H_0 & \text{if } \sqrt{n}\,\dfrac{|\overline{X} - \mu_0|}{\sigma} \geq z_{\alpha/2} \\ \text{Not reject } H_0 & \text{otherwise} \end{array}$$

If the observed value of the test statistic is v, then the p value is given by

$$\begin{aligned} p \text{ value} &= P\{|Z| \geq |v|\} \\ &= 2P\{Z \geq |v|\} \end{aligned}$$

where Z is a standard normal random variable.

2. *Testing*

(1) $H_0 : \mu \leq \mu_0$ *against* $H_1 : \mu > \mu_0$

or

(2) $H_0 : \mu \geq \mu_0$ *against* $H_1 : \mu < \mu_0$

in a normal population having known standard deviation σ: These are called *one-sided tests*. The significance-level-α test in both situations is based on the test statistic $\sqrt{n}\,(\overline{X} - \mu_0)/\sigma$. The test in situation (1) is to

$$\begin{array}{ll} \text{Reject } H_0 & \text{if } \sqrt{n}\,\dfrac{(\overline{X} - \mu_0)}{\sigma} \geq z_\alpha \\ \text{Not reject } H_0 & \text{otherwise} \end{array}$$

Alternatively the test in (1) can be performed by first determining the p value of the data. If the value of the test statistic is v, then the p value is

$$p \text{ value} = P\{Z \geq v\}$$

where Z is a standard normal random variable. The null hypothesis will now be rejected at any significance level at least as large as the p value.

In situation (2), the significance-level-α test is to

$$\begin{array}{ll} \text{Reject } H_0 & \text{if } \sqrt{n}\,\dfrac{(\overline{X} - \mu_0)}{\sigma} \leq -z_\alpha \\ \text{Not reject } H_0 & \text{otherwise} \end{array}$$

Alternatively, if the value of the test statistic $\sqrt{n}\,(\overline{X} - \mu_0)/\sigma$ is v, then the p value is given by

$$p \text{ value} = P\{Z \leq v\}$$

where Z is a standard normal random variable.

3. *Two-sided t test of*

$$H_0 : \mu = \mu_0 \quad \textit{against} \quad H_1 : \mu \neq \mu_0$$

in a normal population whose variance is unknown: This test is based on the test statistic

$$T = \frac{\sqrt{n}\,(\overline{X} - \mu_0)}{S}$$

where n is the sample size and S is the sample standard deviation. The significance-level-α test is to

$$\begin{array}{ll} \text{Reject } \mathrm{H}_0 & \text{if } |T| \geq t_{n-1,\alpha/2} \\ \text{Not reject } \mathrm{H}_0 & \text{otherwise} \end{array}$$

The value $t_{n-1,\alpha/2}$ is such that

$$P\{T_{n-1} > t_{n-1,\alpha/2}\} = \frac{\alpha}{2}$$

when T_{n-1} is a t random variable having $n-1$ degrees of freedom. This is called the *t test*.

This t test can be alternatively run by first calculating the value of the test statistic T. If it is equal to v, then the p value is given by

$$\begin{aligned} p \text{ value} &= P\{|T_{n-1}| \geq |v|\} \\ &= 2P\{T_{n-1} \geq |v|\} \end{aligned}$$

where T_{n-1} is a t random variable with $n-1$ degrees of freedom.

4. *One-sided t tests of*

(1) $\mathrm{H}_0\colon \mu \leq \mu_0$ *against* $\mathrm{H}_1\colon \mu > \mu_0$

or

(2) $\mathrm{H}_0\colon \mu \geq \mu_0$ *against* $\mathrm{H}_1\colon \mu < \mu_0$

in a normal population having an unknown variance: These tests are again based on the test statistic

$$T = \sqrt{n}\frac{\overline{X} - \mu_0}{S}$$

where n is the sample size and S is the sample standard deviation.

The significance-*level-α test* of (1) is to

$$\begin{array}{ll} \text{Reject } \mathrm{H}_0 & \text{if } T \geq t_{n-1,\alpha} \\ \text{Not reject } \mathrm{H}_0 & \text{otherwise} \end{array}$$

Alternatively, the p value may be derived. If the value of the test statistic T is v, the p value is obtained from

$$p \text{ value } = P\{T_{n-1} \geq v\}$$

where T_{n-1} is a t random variable having $n-1$ degrees of freedom.

The significance-level-α test of (2) is to

$$\begin{array}{ll} \text{Reject } \mathrm{H}_0 & \text{if } T \leq -t_{n-1,\alpha} \\ \text{Not reject } \mathrm{H}_0 & \text{otherwise} \end{array}$$

If the value of T is v, then the p value of the test of (2) is

$$p \text{ value} = P\{T_{n-1} \leq v\}$$

5. *Hypothesis tests concerning proportions*: If p is the proportion of a large population that has a certain characteristic, then to test

$$H_0\colon p \leq p_0 \quad \text{versus} \quad H_1\colon p > p_0$$

a random sample of n elements of the population should be drawn. The test statistic is X, the number of members of the sample with the characteristic. If the value of X is x, then the p value is given by

$$p \text{ value} = P\{B \geq x\}$$

where B is a binomial random variable with parameters n and p_0.
Suppose we had wanted to test

$$H_0\colon p \geq p_0 \quad \text{versus} \quad H_1\colon p < p_0$$

If the observed value of the test statistic is x, then the p value is given by

$$p \text{ value} = P\{B \leq x\}$$

where again B is binomial with parameters n and p_0.
The binomial probabilities can be calculated by using Program 5-1 or can be approximated by making use of the normal approximation to the binomial.
Suppose now that the desired test is two-sided; that is, we want to test

$$H_0\colon p = p_0 \quad \text{versus} \quad H_1\colon p \neq p_0$$

If the number of members of the sample with the characteristic is x, then the p value is

$$p \text{ value} = 2\,\text{Min}\{P\{B \leq x\}, P\{B \geq x\}\}$$

where B is binomial with parameters n and p_0.

REVIEW PROBLEMS AND PROPOSED CASE STUDIES

1. Suppose you were to explain to a person who has not yet studied statistics that a statistical test has just resulted in the rejection of the null hypothesis that a population mean μ is equal to 0. That is, $H_0\colon \mu = 0$

has been rejected, say, at the 5 percent level of significance. Which of the following is a more accurate statement?

(a) The evidence of the data indicated that the population mean differs significantly from 0.

(b) The evidence of the data was significant enough to indicate that the population mean differs from 0.

What is misleading about the less accurate of these two statements?

2. Suppose that the result of a statistical test was that the p value was equal to 0.11.

   (a) Would the null hypothesis be rejected at the 5 percent level of significance?

   (b) Would you say that this test provided evidence for the truth of the null hypothesis? Briefly explain your answer.

3. Suppose you happened to read the following statement in your local newspaper. "A recent study provided significant evidence that the mean heights of women have increased over the past twenty years."

   (a) Do you regard this as a precise statement?

   (b) What interpretation would you give to the statement?

4. A fact that has been long known but little understood is that in their early years twins tend to have lower IQ levels and tend to be slower in picking up language skills than nontwins. Recently, some psychologists have speculated that this may be due to the fact that parents spend less time with a twin child than they do with a single child. The reason for this is possibly that a twin always has to share the parent's attention with her or his sibling. The reason is also possibly economic in nature, since twins place a greater economic burden on parents than do single children, and so parents of twins may have less time in general to spend with their offspring.

   Devise a study that could be used to test the hypothesis that twins obtain less parental time than single children.

   Assuming that this hypothesis is correct, devise a study that might enable you to conclude that this is the reason for the long known but little understood fact.

5. An individual's present route to work results in, on average, 40 minutes of travel time per trip. An alternate route has been suggested by a friend, who claims that it will reduce the travel time. Suppose that the new route was tried on 10 randomly chosen occasions with the following times resulting:

$$44, 38.5, 37.5, 39, 38.2, 36, 42, 36.5, 36, 34$$

Do these data establish the claim that the new route is shorter, at the

(a) 1 percent
(b) 5 percent
(c) 10 percent

level of significance?

6. To test the null hypothesis

$$H_0: \mu = 15 \quad \text{versus} \quad H_1: \mu \neq 15$$

a sample of size 12 is taken. If the sample mean is 14.4, find the p value if the population standard deviation is known to equal

(a) 0.5
(b) 1.0
(c) 2.0

7. It has been claimed that over 30 percent of entering college students have blood cholesterol levels of at least 200. Use the last 20 students in the list in App. A to test this hypothesis. What conclusion do you draw at the 5 percent level of significance?

8. Psychologists who consider themselves disciples of Alfred Adler believe that birth order has a strong effect on personality. Adler believed that firstborn (including only) children tend to be more self-confident and success-oriented than later-born children. For instance, of the first 102 appointments to the U.S. Supreme Court, 55 percent have been firstborn children, whereas only 37 percent of the population at large are firstborn.

(a) Using these data about the Supreme Court, test the hypothesis that the belief of Adlerians is wrong and being firstborn does not have a statistical effect on one's personality.
(b) Is the result of (a) a convincing proof of the validity of the Adlerian position? (*Hint*: Recall *data mining*.)
(c) Construct your own study to try to prove or disprove Adler's belief. Choose some sample of successful people (perhaps sample 200 major league baseball players), and find out what percentage of them are firstborn.

9. An individual named Nicholas Caputo was the clerk of Essex County, New Jersey, for an extended period. One of his duties as clerk was to hold a drawing to determine whether Democratic or Republican candidates would be listed first on county ballots. During his reign as clerk, the Democrats won the drawing on 40 of 41 occasions. As a result, Caputo, a Democrat, acquired the nickname *the man with the golden arm*. In 1985 Essex County Republicans sued Caputo, claiming that he

was discriminating against them. If you were the judge, how would you rule? Explain!

10. A recent theory claims that famous people are more likely to die in the 6-month period after their birthday than in the 6-month period preceding it. That is, the claim is that a famous person born on July 1 would be more likely to die between July 1 and December 31 than between January 1 and July 1. The reasoning is that a famous person would probably look forward to all the attention and affection lavished on the birthday, and this anticipation would strengthen the person's "will to live." A countertheory is that famous people are less likely to die in the 6-month period following their birthdays due to their increased strength resulting from their birthday celebration. Still others assert that both theories are wrong.

    Let p denote the probability that a famous person will die within a 6-month period following his or her birthday, and consider a test of

$$H_0\colon p = \frac{1}{2} \quad \text{versus} \quad H_1\colon p \neq \frac{1}{2}$$

    (a) Suppose someone compiled a list of 200 famous dead people in each of 25 separate fields and then ran 25 separate tests of the stated null hypothesis.

    Even if H_0 is always true, what is the probability that at least one of the tests will result in a rejection of H_0 at the 5 percent level of significance?

    (b) Compile a list of between 100 and 200 famous dead people, and use it to test the stated hypothesis.

11. Choose a random sample of 16 women from the list provided in App. A, and use their weights to test the null hypothesis that the average weight of all the women on the list is not greater than 110 pounds. Use the 5 percent level of significance.

12. Suppose that team A and team B are to play a National Football League game and team A is favored by f points. Let $S(A)$ and $S(B)$ denote, respectively, the scores of teams A and B, and let $X = S(A) - S(B) - f$. That is, X is the amount by which team A beats the point spread. It has been claimed that the distribution of X is normal with mean 0 and standard deviation 14. Use data concerning randomly chosen football games to test this hypothesis.

13. The random walk model for the price of a stock or commodity assumes that the successive differences in the logarithms of the closing prices of a given commodity constitute a random sample from a normal population. The following data give the closing prices of gold on 17

consecutive trading days in 1994. Use it to test the hypothesis that the mean daily change is equal to 0.

Closing prices					
387.10	391.00	389.50	391.00	395.00	396.25
388.00	391.95	390.25	390.50	393.50	395.45
389.65	391.05	388.00	394.00	396.25	

Note: The data are ordered by columns. The first value is 387.10, the second 388.00, the third 389.65, the fourth 391.00, and so on.

14. A null hypothesis will be rejected when the value of the test statistic TS is large. The observed value of TS is 1.3. Suppose that when the null hypothesis is true, the probability that TS is at least as large as 1.3 is 0.063.

(a) Will the null hypothesis be rejected at the 5 percent level of significance?

(b) Will the null hypothesis be rejected at the 10 percent level of significance?

(c) What is the p value?

15. It has been a long-time belief that the proportion of California births of African American mothers that result in twins is about 1.32 percent. (The twinning rate appears to be influenced by the ethnicity of the mother: claims are that it is 1.05 for Caucasian Americans, and 0.72 percent for Asian Americans.) A scientist believes that this number is no longer correct and that the actual percentage is about 1.8 percent. Consequently, she has decided to test the null hypothesis that the proportion is at most 1.32 percent by gathering data on the next 1,000 recorded birthing events in California.

(a) What is the minimal number of twin births needed to be able to reject, at the 5 percent level of significance, the null hypothesis that the probability that the twinning rate in African American mothers is no greater than 1.32 per hundred births?

(b) What is the probability that the preceding null hypothesis will be rejected, at the 5 percent level of significance, if the actual twinning rate is 1.80?

16. In 1995, the Fermi Laboratory announced the discovery of the top quark, the last of six quarks predicted by the "standard model of physics." The evidence for its existence was statistical in nature and involved signals created when antiprotons and protons were forced to collide. In a *Physical Review Letters* paper documenting the evidence, Abe, Akimoto, and Akopian (known in physics circle as the three A's) based their conclusion on a theoretical analysis that indicated that the number of decay events in a certain time interval would have a Poisson distribution with a mean equal to 6.7 if a top quark did not exist and with a larger mean if it did exist. In a careful analysis of the data the three A's showed that the actual count was 27. Is this strong enough evidence to prove the hypothesis that the mean of the Poisson distribution was greater than 6.7?