

HYPOTHESIS TESTING

8.1 INTRODUCTION

As in the previous chapter, let us suppose that a random sample from a population distribution, specified except for a vector of unknown parameters, is to be observed. However, rather than wishing to explicitly estimate the unknown parameters, let us now suppose that we are primarily concerned with using the resulting sample to test some particular hypothesis concerning them. As an illustration, suppose that a construction firm has just purchased a large supply of cables that have been guaranteed to have an average breaking strength of at least 7,000 psi. To verify this claim, the firm has decided to take a random sample of 10 of these cables to determine their breaking strengths. They will then use the result of this experiment to ascertain whether or not they accept the cable manufacturer's hypothesis that the population mean is at least 7,000 pounds per square inch.

A statistical hypothesis is usually a statement about a set of parameters of a population distribution. It is called a hypothesis because it is not known whether or not it is true. A primary problem is to develop a procedure for determining whether or not the values of a random sample from this population are consistent with the hypothesis. For instance, consider a particular normally distributed population having an unknown mean value θ and known variance 1. The statement " θ is less than 1" is a statistical hypothesis that we could try to test by observing a random sample from this population. If the random sample is deemed to be consistent with the hypothesis under consideration, we say that the hypothesis has been "accepted"; otherwise we say that it has been "rejected."

Note that in accepting a given hypothesis we are not actually claiming that it is true but rather we are saying that the resulting data appear to be consistent with it. For instance, in the case of a normal $(\theta, 1)$ population, if a resulting sample of size 10 has an average value of 1.25, then although such a result cannot be regarded as being evidence in favor of the hypothesis " $\theta < 1$," it is not inconsistent with this hypothesis, which would thus be accepted. On the other hand, if the sample of size 10 has an average value of 3, then even though a sample value that large is possible when $\theta < 1$, it is so unlikely that it seems inconsistent with this hypothesis, which would thus be rejected.

8.2 SIGNIFICANCE LEVELS

Consider a population having distribution F_θ , where θ is unknown, and suppose we want to test a specific hypothesis about θ . We shall denote this hypothesis by H_0 and call it the *null hypothesis*. For example, if F_θ is a normal distribution function with mean θ and variance equal to 1, then two possible null hypotheses about θ are

$$(a) H_0 : \theta = 1$$

$$(b) H_0 : \theta \leq 1$$

Thus the first of these hypotheses states that the population is normal with mean 1 and variance 1, whereas the second states that it is normal with variance 1 and a mean less than or equal to 1. Note that the null hypothesis in (a), when true, completely specifies the population distribution; whereas the null hypothesis in (b) does not. A hypothesis that, when true, completely specifies the population distribution is called a *simple hypothesis*; one that does not is called a *composite hypothesis*.

Suppose now that in order to test a specific null hypothesis H_0 , a population sample of size n — say X_1, \dots, X_n — is to be observed. Based on these n values, we must decide whether or not to accept H_0 . A test for H_0 can be specified by defining a region C in n -dimensional space with the proviso that the hypothesis is to be rejected if the random sample X_1, \dots, X_n turns out to lie in C and accepted otherwise. The region C is called the *critical region*. In other words, the statistical test determined by the critical region C is the one that

$$\text{accepts } H_0 \quad \text{if } (X_1, X_2, \dots, X_n) \notin C$$

and

$$\text{rejects } H_0 \quad \text{if } (X_1, \dots, X_n) \in C$$

For instance, a common test of the hypothesis that θ , the mean of a normal population with variance 1, is equal to 1 has a critical region given by

$$C = \left\{ (X_1, \dots, X_n) : \left| \frac{\sum_{i=1}^n X_i}{n} - 1 \right| > \frac{1.96}{\sqrt{n}} \right\} \quad (8.2.1)$$

Thus, this test calls for rejection of the null hypothesis that $\theta = 1$ when the sample average differs from 1 by more than 1.96 divided by the square root of the sample size.

It is important to note when developing a procedure for testing a given null hypothesis H_0 that, in any test, two different types of errors can result. The first of these, called a *type I error*, is said to result if the test incorrectly calls for rejecting H_0 when it is indeed correct. The second, called a *type II error*, results if the test calls for accepting H_0 when it is false.

Now, as was previously mentioned, the objective of a statistical test of H_0 is not to explicitly determine whether or not H_0 is true but rather to determine if its validity is consistent with the resultant data. Hence, with this objective it seems reasonable that H_0 should only be rejected if the resultant data are very unlikely when H_0 is true. The classical way of accomplishing this is to specify a value α and then require the test to have the property that whenever H_0 is true its probability of being rejected is never greater than α . The value α , called the *level of significance of the test*, is usually set in advance, with commonly chosen values being $\alpha = .1, .05, .005$. In other words, the classical approach to testing H_0 is to fix a significance level α and then require that the test have the property that the probability of a type I error occurring can never be greater than α .

Suppose now that we are interested in testing a certain hypothesis concerning θ , an unknown parameter of the population. Specifically, for a given set of parameter values w , suppose we are interested in testing

$$H_0 : \theta \in w$$

A common approach to developing a test of H_0 , say at level of significance α , is to start by determining a point estimator of θ — say $d(\mathbf{X})$. The hypothesis is then rejected if $d(\mathbf{X})$ is “far away” from the region w . However, to determine how “far away” it need be to justify rejection of H_0 , we need to determine the probability distribution of $d(\mathbf{X})$ when H_0 is true since this will usually enable us to determine the appropriate critical region so as to make the test have the required significance level α . For example, the test of the hypothesis that the mean of a normal $(\theta, 1)$ population is equal to 1, given by Equation 8.2.1, calls for rejection when the point estimate of θ — that is, the sample average — is farther than $1.96/\sqrt{n}$ away from 1. As we will see in the next section, the value $1.96/\sqrt{n}$ was chosen to meet a level of significance of $\alpha = .05$.

8.3 TESTS CONCERNING THE MEAN OF A NORMAL POPULATION

8.3.1 CASE OF KNOWN VARIANCE

Suppose that X_1, \dots, X_n is a sample of size n from a normal distribution having an unknown mean μ and a known variance σ^2 and suppose we are interested in testing the null hypothesis

$$H_0 : \mu = \mu_0$$

against the alternative hypothesis

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

where μ_0 is some specified constant.

Since $\bar{X} = \sum_{i=1}^n X_i/n$ is a natural point estimator of μ , it seems reasonable to accept H_0 if \bar{X} is not too far from μ_0 . That is, the critical region of the test would be of the form

$$C = \{X_1, \dots, X_n : |\bar{X} - \mu_0| > c\} \quad (8.3.1)$$

for some suitably chosen value c .

If we desire that the test has significance level α , then we must determine the critical value c in Equation 8.3.1 that will make the type I error equal to α . That is, c must be such that

$$P_{\mu_0}\{|\bar{X} - \mu_0| > c\} = \alpha \quad (8.3.2)$$

where we write P_{μ_0} to mean that the preceding probability is to be computed under the assumption that $\mu = \mu_0$. However, when $\mu = \mu_0$, \bar{X} will be normally distributed with mean μ_0 and variance σ^2/n and so Z , defined by

$$Z \equiv \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

will have a standard normal distribution. Now Equation 8.3.2 is equivalent to

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

or, equivalently,

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

where Z is a standard normal random variable. However, we know that

$$P\{Z > z_{\alpha/2}\} = \alpha/2$$

and so

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

or

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

Thus, the significance level α test is to reject H_0 if $|\bar{X} - \mu_0| > z_{\alpha/2}\sigma/\sqrt{n}$ and accept otherwise; or, equivalently, to

$$\begin{aligned} \text{reject } H_0 & \text{ if } \frac{\sqrt{n}}{\sigma} |\bar{X} - \mu_0| > z_{\alpha/2} \\ \text{accept } H_0 & \text{ if } \frac{\sqrt{n}}{\sigma} |\bar{X} - \mu_0| \leq z_{\alpha/2} \end{aligned} \quad (8.3.3)$$

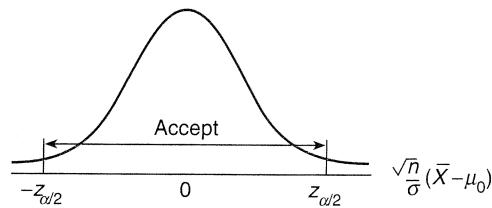


FIGURE 8.1

This can be pictorially represented as shown in Figure 8.1, where we have superimposed the standard normal density function [which is the density of the test statistic $\sqrt{n}(\bar{X} - \mu_0)/\sigma$ when H_0 is true].

EXAMPLE 8.3a It is known that if a signal of value μ is sent from location A, then the value received at location B is normally distributed with mean μ and standard deviation 2. That is, the random noise added to the signal is an $N(0, 4)$ random variable. There is reason for the people at location B to suspect that the signal value $\mu = 8$ will be sent today. Test this hypothesis if the same signal value is independently sent five times and the average value received at location B is $\bar{X} = 9.5$.

SOLUTION Suppose we are testing at the 5 percent level of significance. To begin, we compute the test statistic

$$\frac{\sqrt{n}}{\sigma} |\bar{X} - \mu_0| = \frac{\sqrt{5}}{2} (1.5) = 1.68$$

Since this value is less than $z_{.025} = 1.96$, the hypothesis is accepted. In other words, the data are not inconsistent with the null hypothesis in the sense that a sample average as far from the value 8 as observed would be expected, when the true mean is 8, over 5 percent of the time. Note, however, that if a less stringent significance level were chosen — say $\alpha = .1$ — then the null hypothesis would have been rejected. This follows since $z_{.05} = 1.645$, which is less than 1.68. Hence, if we would have chosen a test that had a 10 percent chance of rejecting H_0 when H_0 was true, then the null hypothesis would have been rejected.

The “correct” level of significance to use in a given situation depends on the individual circumstances involved in that situation. For instance, if rejecting a null hypothesis H_0 would result in large costs that would thus be lost if H_0 were indeed true, then we might elect to be quite conservative and so choose a significance level of .05 or .01. Also, if we initially feel strongly that H_0 was correct, then we would require very stringent data evidence to the contrary for us to reject H_0 . (That is, we would set a very low significance level in this situation.) ■

The test given by Equation 8.3.3 can be described as follows: For any observed value of the test statistic $\sqrt{n}|\bar{X} - \mu_0|/\sigma$, call it v , the test calls for rejection of the null hypothesis if the probability that the test statistic would be as large as v when H_0 is true is less than or equal to the significance level α . From this, it follows that we can determine whether or not to accept the null hypothesis by computing, first, the value of the test statistic and, second, the probability that a unit normal would (in absolute value) exceed that quantity. This probability — called the p -value of the test — gives the critical significance level in the sense that H_0 will be accepted if the significance level α is less than the p -value and rejected if it is greater than or equal.

In practice, the significance level is often not set in advance but rather the data are looked at to determine the resultant p -value. Sometimes, this critical significance level is clearly much larger than any we would want to use, and so the null hypothesis can be readily accepted. At other times the p -value is so small that it is clear that the hypothesis should be rejected.

EXAMPLE 8.3b In Example 8.3a, suppose that the average of the 5 values received is $\bar{X} = 8.5$. In this case,

$$\frac{\sqrt{n}}{\sigma} |\bar{X} - \mu_0| = \frac{\sqrt{5}}{4} = .559$$

Since

$$\begin{aligned} P\{|Z| > .559\} &= 2P\{Z > .559\} \\ &= 2 \times .288 = .576 \end{aligned}$$

it follows that the p -value is .576 and thus the null hypothesis H_0 that the signal sent has value 8 would be accepted at any significance level $\alpha < .576$. Since we would clearly never want to test a null hypothesis using a significance level as large as .576, H_0 would be accepted.

On the other hand, if the average of the data values were 11.5, then the p -value of the test that the mean is equal to 8 would be

$$\begin{aligned} P\{|Z| > 1.75\sqrt{5}\} &= P\{|Z| > 3.913\} \\ &\approx .00005 \end{aligned}$$

For such a small p -value, the hypothesis that the value 8 was sent is rejected. ■

We have not yet talked about the probability of a type II error — that is, the probability of accepting the null hypothesis when the true mean μ is unequal to μ_0 . This probability

will depend on the value of μ , and so let us define $\beta(\mu)$ by

$$\begin{aligned}\beta(\mu) &= P_\mu \{\text{acceptance of } H_0\} \\ &= P_\mu \left\{ \left| \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \right| \leq z_{\alpha/2} \right\} \\ &= P_\mu \left\{ -z_{\alpha/2} \leq \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq z_{\alpha/2} \right\}\end{aligned}$$

The function $\beta(\mu)$ is called the *operating characteristic* (or OC) *curve* and represents the probability that H_0 will be accepted when the true mean is μ .

To compute this probability, we use the fact that \bar{X} is normal with mean μ and variance σ^2/n and so

$$Z \equiv \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

Hence,

$$\begin{aligned}\beta(\mu) &= P_\mu \left\{ -z_{\alpha/2} \leq \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \leq z_{\alpha/2} \right\} \\ &= P_\mu \left\{ -z_{\alpha/2} - \frac{\mu}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu_0 - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2} - \frac{\mu}{\sigma/\sqrt{n}} \right\} \\ &= P_\mu \left\{ -z_{\alpha/2} - \frac{\mu}{\sigma/\sqrt{n}} \leq Z - \frac{\mu_0}{\sigma/\sqrt{n}} \leq z_{\alpha/2} - \frac{\mu}{\sigma/\sqrt{n}} \right\} \\ &= P \left\{ \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - z_{\alpha/2} \leq Z \leq \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\alpha/2} \right\} \\ &= \Phi \left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\alpha/2} \right) - \Phi \left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} - z_{\alpha/2} \right) \quad (8.3.4)\end{aligned}$$

where Φ is the standard normal distribution function.

For a fixed significance level α , the OC curve given by Equation 8.3.4 is symmetric about μ_0 and indeed will depend on μ only through $(\sqrt{n}/\sigma)|\mu - \mu_0|$. This curve with the abscissa changed from μ to $d = (\sqrt{n}/\sigma)|\mu - \mu_0|$ is presented in Figure 8.2 when $\alpha = .05$.

EXAMPLE 8.3c For the problem presented in Example 8.3a, let us determine the probability of accepting the null hypothesis that $\mu = 8$ when the actual value sent is 10. To do so, we compute

$$\frac{\sqrt{n}}{\sigma}(\mu_0 - \mu) = -\frac{\sqrt{5}}{2} \times 2 = -\sqrt{5}$$

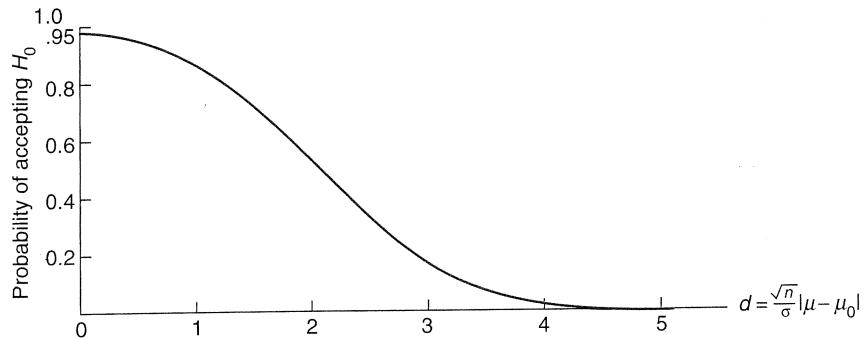


FIGURE 8.2 The OC curve for the two-sided normal test for significance level $\alpha = .05$.

As $z_{.025} = 1.96$, the desired probability is, from Equation 8.3.4,

$$\begin{aligned}
 & \Phi(-\sqrt{5} + 1.96) - \Phi(-\sqrt{5} - 1.96) \\
 &= 1 - \Phi(\sqrt{5} - 1.96) - [1 - \Phi(\sqrt{5} + 1.96)] \\
 &= \Phi(4.196) - \Phi(.276) \\
 &= .392 \quad \blacksquare
 \end{aligned}$$

REMARK

The function $1 - \beta(\mu)$ is called the *power-function* of the test. Thus, for a given value μ , the power of the test is equal to the probability of rejection when μ is the true value. ■

The operating characteristic function is useful in determining how large the random sample need be to meet certain specifications concerning type II errors. For instance, suppose that we desire to determine the sample size n necessary to ensure that the probability of accepting $H_0 : \mu = \mu_0$ when the true mean is actually μ_1 is approximately β . That is, we want n to be such that

$$\beta(\mu_1) \approx \beta$$

But from Equation 8.3.4, this is equivalent to

$$\Phi\left(\frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma} + z_{\alpha/2}\right) - \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma} - z_{\alpha/2}\right) \approx \beta \quad (8.3.5)$$

Although the foregoing cannot be analytically solved for n , a solution can be obtained by using the standard normal distribution table. In addition, an approximation for n can be derived from Equation 8.3.5 as follows. To start, suppose that $\mu_1 > \mu_0$. Then, because this implies that

$$\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2} \leq -z_{\alpha/2}$$

it follows, since Φ is an increasing function, that

$$\Phi\left(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2}\right) \leq \Phi(-z_{\alpha/2}) = P\{Z \leq -z_{\alpha/2}\} = P\{Z \geq z_{\alpha/2}\} = \alpha/2$$

Hence, we can take

$$\Phi\left(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - z_{\alpha/2}\right) \approx 0$$

and so from Equation 8.3.5

$$\beta \approx \Phi\left(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_{\alpha/2}\right) \quad (8.3.6)$$

or, since

$$\beta = P\{Z > z_\beta\} = P\{Z < -z_\beta\} = \Phi(-z_\beta)$$

we obtain from Equation 8.3.6 that

$$-z_\beta \approx (\mu_0 - \mu_1) \frac{\sqrt{n}}{\sigma} + z_{\alpha/2}$$

or

$$n \approx \frac{(z_{\alpha/2} + z_\beta)^2 \sigma^2}{(\mu_1 - \mu_0)^2} \quad (8.3.7)$$

In fact, the same approximation would result when $\mu_1 < \mu_0$ (the details are left as an exercise) and so Equation 8.3.7 is in all cases a reasonable approximation to the sample size necessary to ensure that the type II error at the value $\mu = \mu_1$ is approximately equal to β .

EXAMPLE 8.3d For the problem of Example 8.3a, how many signals need be sent so that the .05 level test of $H_0 : \mu = 8$ has at least a 75 percent probability of rejection when $\mu = 9.2$?

SOLUTION Since $z_{.025} = 1.96$, $z_{.25} = .67$, the approximation 8.3.7 yields

$$n \approx \frac{(1.96 + .67)^2}{(1.2)^2} 4 = 19.21$$

Hence a sample of size 20 is needed. From Equation 8.3.4, we see that with $n = 20$

$$\begin{aligned} \beta(9.2) &= \Phi\left(-\frac{1.2\sqrt{20}}{2} + 1.96\right) - \Phi\left(-\frac{1.2\sqrt{20}}{2} - 1.96\right) \\ &= \Phi(-.723) - \Phi(-4.643) \end{aligned}$$

$$\approx 1 - \Phi(.723)$$

$$\approx .235$$

Therefore, if the message is sent 20 times, then there is a 76.5 percent chance that the null hypothesis $\mu = 8$ will be rejected when the true mean is 9.2. ■

ONE-SIDED TESTS

In testing the null hypothesis that $\mu = \mu_0$, we have chosen a test that calls for rejection when \bar{X} is far from μ_0 . That is, a very small value of \bar{X} or a very large value appears to make it unlikely that μ (which \bar{X} is estimating) could equal μ_0 . However, what happens when the only alternative to μ being equal to μ_0 is for μ to be greater than μ_0 ? That is, what happens when the alternative hypothesis to $H_0 : \mu = \mu_0$ is $H_1 : \mu > \mu_0$? Clearly, in this latter case we would not want to reject H_0 when \bar{X} is small (since a small \bar{X} is more likely when H_0 is true than when H_1 is true). Thus, in testing

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0 \quad (8.3.8)$$

we should reject H_0 when \bar{X} , the point estimate of μ_0 , is much greater than μ_0 . That is, the critical region should be of the following form:

$$C = \{(X_1, \dots, X_n) : \bar{X} - \mu_0 > c\}$$

Since the probability of rejection should equal α when H_0 is true (that is, when $\mu = \mu_0$), we require that c be such that

$$P_{\mu_0}\{\bar{X} - \mu_0 > c\} = \alpha \quad (8.3.9)$$

But since

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

has a standard normal distribution when H_0 is true, Equation 8.3.9 is equivalent to

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

when Z is a standard normal. But since

$$P\{Z > z_\alpha\} = \alpha$$

we see that

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

Hence, the test of the hypothesis 8.3.8 is to reject H_0 if $\bar{X} - \mu_0 > z_\alpha \sigma / \sqrt{n}$, and accept otherwise; or, equivalently, to

$$\begin{aligned} \text{accept } H_0 & \text{ if } \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) \leq z_\alpha \\ \text{reject } H_0 & \text{ if } \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) > z_\alpha \end{aligned} \quad (8.3.10)$$

This is called a *one-sided* critical region (since it calls for rejection only when \bar{X} is large). Correspondingly, the hypothesis testing problem

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu > \mu_0$$

is called a one-sided testing problem (in contrast to the *two-sided* problem that results when the alternative hypothesis is $H_1 : \mu \neq \mu_0$).

To compute the *p*-value in the one-sided test, Equation 8.3.10, we first use the data to determine the value of the statistic $\sqrt{n}(\bar{X} - \mu_0)/\sigma$. The *p*-value is then equal to the probability that a standard normal would be at least as large as this value.

EXAMPLE 8.3e Suppose in Example 8.3a that we know in advance that the signal value is at least as large as 8. What can be concluded in this case?

SOLUTION To see if the data are consistent with the hypothesis that the mean is 8, we test

$$H_0 : \mu = 8$$

against the one-sided alternative

$$H_1 : \mu > 8$$

The value of the test statistic is $\sqrt{n}(\bar{X} - \mu_0)/\sigma = \sqrt{5}(9.5 - 8)/2 = 1.68$, and the *p*-value is the probability that a standard normal would exceed 1.68, namely,

$$p\text{-value} = 1 - \Phi(1.68) = .0465$$

Since the test would call for rejection at all significance levels greater than or equal to .0465, it would, for instance, reject the null hypothesis at the $\alpha = .05$ level of significance. ■

The operating characteristic function of the one-sided test, Equation 8.3.10,

$$\beta(\mu) = P_\mu\{\text{accepting } H_0\}$$

can be obtained as follows:

$$\begin{aligned}\beta(\mu) &= P_{\mu} \left\{ \bar{X} \leq \mu_0 + z_{\alpha} \frac{\sigma}{\sqrt{n}} \right\} \\ &= P \left\{ \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\alpha} \right\} \\ &= P \left\{ Z \leq \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\alpha} \right\}, \quad Z \sim \mathcal{N}(0, 1)\end{aligned}$$

where the last equation follows since $\sqrt{n}(\bar{X} - \mu)/\sigma$ has a standard normal distribution. Hence we can write

$$\beta(\mu) = \Phi \left(\frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_{\alpha} \right)$$

Since Φ , being a distribution function, is increasing in its argument, it follows that $\beta(\mu)$ decreases in μ ; which is intuitively pleasing since it certainly seems reasonable that the larger the true mean μ , the less likely it should be to conclude that $\mu \leq \mu_0$. Also since $\Phi(z_{\alpha}) = 1 - \alpha$, it follows that

$$\beta(\mu_0) = 1 - \alpha$$

The test given by Equation 8.3.10, which was designed to test $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$ can also be used to test, at level of significance α , the one-sided hypothesis

$$H_0 : \mu \leq \mu_0$$

versus

$$H_1 : \mu > \mu_0$$

To verify that it remains a level α test, we need show that the probability of rejection is never greater than α when H_0 is true. That is, we must verify that

$$1 - \beta(\mu) \leq \alpha \quad \text{for all } \mu \leq \mu_0$$

or

$$\beta(\mu) \geq 1 - \alpha \quad \text{for all } \mu \leq \mu_0$$

But it has previously been shown that for the test given by Equation 8.3.10, $\beta(\mu)$ decreases in μ and $\beta(\mu_0) = 1 - \alpha$. This gives that

$$\beta(\mu) \geq \beta(\mu_0) = 1 - \alpha \quad \text{for all } \mu \leq \mu_0$$

which shows that the test given by Equation 8.3.10 remains a level α test for $H_0 : \mu \leq \mu_0$ against the alternative hypothesis $H_1 : \mu > \mu_0$.

REMARK

We can also test the one-sided hypothesis

$$H_0 : \mu = \mu_0 \quad (\text{or } \mu \geq \mu_0) \quad \text{versus} \quad H_1 : \mu < \mu_0$$

at significance level α by

$$\begin{aligned} \text{accepting } H_0 & \text{ if } \frac{\sqrt{n}}{\sigma}(\bar{X} - \mu_0) \geq -z_\alpha \\ \text{rejecting } H_0 & \text{ otherwise} \end{aligned}$$

This test can alternatively be performed by first computing the value of the test statistic $\sqrt{n}(\bar{X} - \mu_0)/\sigma$. The p -value would then equal the probability that a standard normal would be less than this value, and the hypothesis would be rejected at any significance level greater than or equal to this p -value.

EXAMPLE 8.3f All cigarettes presently on the market have an average nicotine content of at least 1.6 mg per cigarette. A firm that produces cigarettes claims that it has discovered a new way to cure tobacco leaves that will result in the average nicotine content of a cigarette being less than 1.6 mg. To test this claim, a sample of 20 of the firm's cigarettes were analyzed. If it is known that the standard deviation of a cigarette's nicotine content is .8 mg, what conclusions can be drawn, at the 5 percent level of significance, if the average nicotine content of the 20 cigarettes is 1.54?

Note: The above raises the question of how we would know in advance that the standard deviation is .8. One possibility is that the variation in a cigarette's nicotine content is due to variability in the amount of tobacco in each cigarette and not on the method of curing that is used. Hence, the standard deviation can be known from previous experience.

SOLUTION We must first decide on the appropriate null hypothesis. As was previously noted, our approach to testing is not symmetric with respect to the null and the alternative hypotheses since we consider only tests having the property that their probability of rejecting the null hypothesis when it is true will never exceed the significance level α . Thus, whereas rejection of the null hypothesis is a strong statement about the data not being consistent with this hypothesis, an analogous statement cannot be made when the null hypothesis is accepted. Hence, since in the preceding example we would like to endorse the producer's claims only when there is substantial evidence for it, we should take this claim as the alternative hypothesis.

That is, we should test

$$H_0 : \mu \geq 1.6 \quad \text{versus} \quad H_1 : \mu < 1.6$$

Now, the value of the test statistic is

$$\sqrt{n}(\bar{X} - \mu_0)/\sigma = \sqrt{20}(1.54 - 1.6)/.8 = -.336$$

and so the p -value is given by

$$\begin{aligned} p\text{-value} &= P\{Z < -.336\}, \quad Z \sim N(0, 1) \\ &= .368 \end{aligned}$$

Since this value is greater than .05, the foregoing data do not enable us to reject, at the .05 percent level of significance, the hypothesis that the mean nicotine content exceeds 1.6 mg. In other words, the evidence, although supporting the cigarette producer's claim, is not strong enough to prove that claim. ■

REMARKS

(a) There is a direct analogy between confidence interval estimation and hypothesis testing. For instance, for a normal population having mean μ and known variance σ^2 , we have shown in Section 7.3 that a $100(1 - \alpha)$ percent confidence interval for μ is given by

$$\mu \in \left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

where \bar{x} is the observed sample mean. More formally, the preceding confidence interval statement is equivalent to

$$P \left\{ \mu \in \left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) \right\} = 1 - \alpha$$

Hence, if $\mu = \mu_0$, then the probability that μ_0 will fall in the interval

$$\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

is $1 - \alpha$, implying that a significance level α test of $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ is to reject H_0 when

$$\mu_0 \notin \left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

Similarly, since a $100(1 - \alpha)$ percent one-sided confidence interval for μ is given by

$$\mu \in \left(\bar{X} - z_{\alpha} \frac{\sigma}{\sqrt{n}}, \infty \right)$$

it follows that an α -level significance test of $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$ is to reject H_0 when $\mu_0 \notin (\bar{X} - z_{\alpha} \sigma / \sqrt{n}, \infty)$ — that is, when $\mu_0 < \bar{X} - z_{\alpha} \sigma / \sqrt{n}$.

TABLE 8.1 X_1, \dots, X_n Is a Sample from a $\mathcal{N}(\mu, \sigma^2)$ Population σ^2 Is Known $\bar{X} = \sum_{i=1}^n X_i/n$

H_0	H_1	Test Statistic TS	Significance Level α	p -Value if $TS = t$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\sqrt{n}(\bar{X} - \mu_0)/\sigma$	Reject if $ TS > z_{\alpha/2}$	$2P\{Z \geq t \}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\sqrt{n}(\bar{X} - \mu_0)/\sigma$	Reject if $TS > z_{\alpha}$	$P\{Z \geq t\}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\sqrt{n}(\bar{X} - \mu_0)/\sigma$	Reject if $TS < -z_{\alpha}$	$P\{Z \leq t\}$

Z is a standard normal random variable.

(b) A Remark on Robustness A test that performs well even when the underlying assumptions on which it is based are violated is said to be *robust*. For instance, the tests of Sections 8.3.1 and 8.3.1.1 were derived under the assumption that the underlying population distribution is normal with known variance σ^2 . However, in deriving these tests, this assumption was used only to conclude that \bar{X} also has a normal distribution. But, by the central limit theorem, it follows that for a reasonably large sample size, \bar{X} will approximately have a normal distribution no matter what the underlying distribution. Thus we can conclude that these tests will be relatively robust for any population distribution with variance σ^2 .

Table 8.1 summarizes the tests of this subsection.

8.3.2 CASE OF UNKNOWN VARIANCE: THE t -TEST

Up to now we have supposed that the only unknown parameter of the normal population distribution is its mean. However, the more common situation is one where the mean μ and variance σ^2 are both unknown. Let us suppose this to be the case and again consider a test of the hypothesis that the mean is equal to some specified value μ_0 . That is, consider a test of

$$H_0 : \mu = \mu_0$$

versus the alternative

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

It should be noted that the null hypothesis is not a simple hypothesis since it does not specify the value of σ^2 .

As before, it seems reasonable to reject H_0 when the sample mean \bar{X} is far from μ_0 . However, how far away it need be to justify rejection will depend on the variance σ^2 . Recall that when the value of σ^2 was known, the test called for rejecting H_0 when $|\bar{X} - \mu_0|$ exceeded $z_{\alpha/2}\sigma/\sqrt{n}$ or, equivalently, when

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

Now when σ^2 is no longer known, it seems reasonable to estimate it by

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

and then to reject H_0 when

$$\left| \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \right|$$

is large.

To determine how large a value of the statistic

$$\left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \right|$$

to require for rejection, in order that the resulting test have significance level α , we must determine the probability distribution of this statistic when H_0 is true. However, as shown in Section 6.5, the statistic T , defined by

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

has, when $\mu = \mu_0$, a t -distribution with $n-1$ degrees of freedom. Hence,

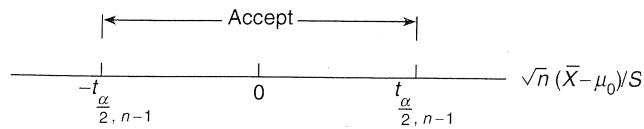
$$P_{\mu_0} \left\{ -t_{\alpha/2, n-1} \leq \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \leq t_{\alpha/2, n-1} \right\} = 1 - \alpha \quad (8.3.11)$$

where $t_{\alpha/2, n-1}$ is the 100 $\alpha/2$ upper percentile value of the t -distribution with $n-1$ degrees of freedom. (That is, $P\{T_{n-1} \geq t_{\alpha/2, n-1}\} = P\{T_{n-1} \leq -t_{\alpha/2, n-1}\} = \alpha/2$ when T_{n-1} has a t -distribution with $n-1$ degrees of freedom.) From Equation 8.3.11 we see that the appropriate significance level α test of

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

is, when σ^2 is unknown, to

$$\begin{aligned} \text{accept } H_0 & \text{ if } \left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \right| \leq t_{\alpha/2, n-1} \\ \text{reject } H_0 & \text{ if } \left| \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \right| > t_{\alpha/2, n-1} \end{aligned} \quad (8.3.12)$$

FIGURE 8.3 The two-sided *t*-test.

The test defined by Equation 8.3.12 is called a *two-sided t-test*. It is pictorially illustrated in Figure 8.3.

If we let t denote the observed value of the test statistic $T = \sqrt{n}(\bar{X} - \mu_0)/S$, then the p -value of the test is the probability that $|T|$ would exceed $|t|$ when H_0 is true. That is, the p -value is the probability that the absolute value of a t -random variable with $n - 1$ degrees of freedom would exceed $|t|$. The test then calls for rejection at all significance levels higher than the p -value and acceptance at all lower significance levels.

Program 8.3.2 computes the value of the test statistic and the corresponding p -value. It can be applied both for one- and two-sided tests. (The one-sided material will be presented shortly.)

EXAMPLE 8.3g Among a clinic's patients having blood cholesterol levels ranging in the medium to high range (at least 220 milliliters per deciliter of serum), volunteers were recruited to test a new drug designed to reduce blood cholesterol. A group of 50 volunteers was given the drug for 1 month and the changes in their blood cholesterol levels were noted. If the average change was a reduction of 14.8 with a sample standard deviation of 6.4, what conclusions can be drawn?

SOLUTION Let us start by testing the hypothesis that the change could be due solely to chance — that is, that the 50 changes constitute a normal sample with mean 0. Because the value of the t -statistic used to test the hypothesis that a normal mean is equal to 0 is

$$T = \sqrt{n} \bar{X}/S = \sqrt{50} 14.8/6.4 = 16.352$$

it is clear that we should reject the hypothesis that the changes were solely due to chance. Unfortunately, however, we are not justified at this point in concluding that the changes were 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 so-called placebo effect. Also, another possibility that may need to be taken into account would be the weather conditions during the month of testing, for it is certainly conceivable that this affects blood cholesterol level. Indeed, it must be concluded that the foregoing was a very poorly designed experiment, for in order to test whether a specific treatment has an effect on a disease that may be affected by many things, we should try to design the experiment so as to neutralize all other possible causes. The accepted approach for accomplishing this is to divide the volunteers at random into two

groups — one group to receive the drug and the other to receive a placebo (that is, a tablet that looks and tastes like the actual drug but has no physiological effect). The volunteers should not be told whether they are in the actual or control group, and indeed it is best if even the clinicians do not have this information (the so-called double-blind test) so as not to allow their own biases to play a role. Since the two groups are chosen at random from among the volunteers, we can now hope that on average all factors affecting the two groups will be the same except that one received the actual drug and the other a placebo. Hence, any difference in performance between the groups can be attributed to the drug. ■

EXAMPLE 8.3h A public health official claims that the mean home water use is 350 gallons a day. To verify this claim, a study of 20 randomly selected homes was instigated with the result that the average daily water uses of these 20 homes were as follows:

340	344	362	375
356	386	354	364
332	402	340	355
362	322	372	324
318	360	338	370

Do the data contradict the official's claim?

SOLUTION To determine if the data contradict the official's claim, we need to test

$$H_0 : \mu = 350 \quad \text{versus} \quad H_1 : \mu \neq 350$$

This can be accomplished by running Program 8.3.2 or, if it is inconvenient to utilize, by noting first that the sample mean and sample standard deviation of the preceding data set are

$$\bar{X} = 353.8, \quad S = 21.8478$$

Thus, the value of the test statistic is

$$T = \frac{\sqrt{20}(3.8)}{21.8478} = .7778$$

Because this is less than $t_{05,19} = 1.730$, the null hypothesis is accepted at the 10 percent level of significance. Indeed, the p -value of the test data is

$$p\text{-value} = P\{|T_{19}| > .7778\} = 2P\{T_{19} > .7778\} = .4462$$

indicating that the null hypothesis would be accepted at any reasonable significance level, and thus that the data are not inconsistent with the claim of the health official. ■

We can use a one-sided t -test to test the hypothesis

$$H_0 : \mu = \mu_0 \quad (\text{or } H_0 : \mu \leq \mu_0)$$

against the one-sided alternative

$$H_1 : \mu > \mu_0$$

The significance level α test is to

$$\begin{aligned} \text{accept } H_0 & \text{ if } \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \leq t_{\alpha, n-1} \\ \text{reject } H_0 & \text{ if } \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} > t_{\alpha, n-1} \end{aligned} \quad (8.3.13)$$

If $\sqrt{n}(\bar{X} - \mu_0)/S = v$, then the p -value of the test is the probability that a t -random variable with $n - 1$ degrees of freedom would be at least as large as v .

The significance level α test of

$$H_0 : \mu = \mu_0 \quad (\text{or } H_0 : \mu \geq \mu_0)$$

versus the alternative

$$H_1 : \mu < \mu_0$$

is to

$$\begin{aligned} \text{accept } H_0 & \text{ if } \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} \geq -t_{\alpha, n-1} \\ \text{reject } H_0 & \text{ if } \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} < -t_{\alpha, n-1} \end{aligned}$$

The p -value of this test is the probability that a t -random variable with $n - 1$ degrees of freedom would be less than or equal to the observed value of $\sqrt{n}(\bar{X} - \mu_0)/S$.

EXAMPLE 8.3i The manufacturer of a new fiberglass tire claims that its average life will be at least 40,000 miles. To verify this claim a sample of 12 tires is tested, with their lifetimes (in 1,000s of miles) being as follows:

Tire	1	2	3	4	5	6	7	8	9	10	11	12
Life	36.1	40.2	33.8	38.5	42	35.8	37	41	36.8	37.2	33	36

Test the manufacturer's claim at the 5 percent level of significance.

SOLUTION To determine whether the foregoing data are consistent with the hypothesis that the mean life is at least 40,000 miles, we will test

$$H_0 : \mu \geq 40,000 \quad \text{versus} \quad H_1 : \mu < 40,000$$

A computation gives that

$$\bar{X} = 37.2833, \quad S = 2.7319$$

and so the value of the test statistic is

$$T = \frac{\sqrt{12}(37.2833 - 40)}{2.7319} = -3.4448$$

Since this is less than $-t_{.05,11} = -1.796$, the null hypothesis is rejected at the 5 percent level of significance. Indeed, the p -value of the test data is

$$p\text{-value} = P\{T_{11} < -3.4448\} = P\{T_{11} > 3.4448\} = .0028$$

indicating that the manufacturer's claim would be rejected at any significance level greater than .003. ■

The preceding could also have been obtained by using Program 8.3.2, as illustrated in Figure 8.4.

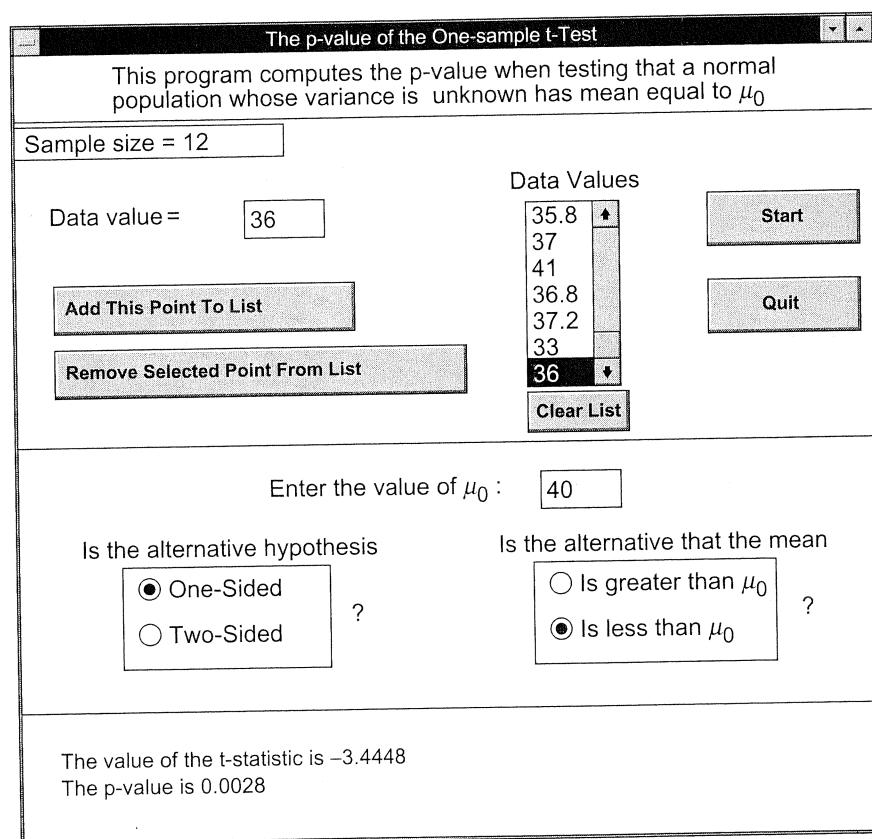


FIGURE 8.4

EXAMPLE 8.3j In a single-server queueing system in which customers arrive according to a Poisson process, the long-run average queueing delay per customer depends on the service distribution through its mean and variance. Indeed, if μ is the mean service time, and σ^2 is the variance of a service time, then the average amount of time that a customer spends waiting in queue is given by

$$\frac{\lambda(\mu^2 + \sigma^2)}{2(1 - \lambda\mu)}$$

provided that $\lambda\mu < 1$, where λ is the arrival rate. (The average delay is infinite if $\lambda\mu \geq 1$.) As can be seen by this formula, the average delay is quite large when μ is only slightly smaller than $1/\lambda$, where, since λ is the arrival rate, $1/\lambda$ is the average time between arrivals.

Suppose that the owner of a service station will hire a second server if it can be shown that the average service time exceeds 8 minutes. The following data give the service times (in minutes) of 28 customers of this queueing system. Do they indicate that the mean service time is greater than 8 minutes?

8.6, 9.4, 5.0, 4.4, 3.7, 11.4, 10.0, 7.6, 14.4, 12.2, 11.0, 14.4, 9.3, 10.5,
10.3, 7.7, 8.3, 6.4, 9.2, 5.7, 7.9, 9.4, 9.0, 13.3, 11.6, 10.0, 9.5, 6.6

SOLUTION Let us use the preceding data to test the null hypothesis that the mean service time is less than or equal to 8 minutes. A small p -value will then be strong evidence that the mean service time is greater than 8 minutes. Running Program 8.3.2 on these data shows that the value of the test statistic is 2.257, with a resulting p -value of .016. Such a small p -value is certainly strong evidence that the mean service time exceeds 8 minutes. ■

Table 8.2 summarizes the tests of this subsection.

TABLE 8.2 X_1, \dots, X_n Is a Sample from a $\mathcal{N}(\mu, \sigma^2)$ Population σ^2 Is Unknown $\bar{X} = \sum_{i=1}^n X_i/n$
 $S^2 = \sum_{i=1}^n (X_i - \bar{X})^2/(n - 1)$

H_0	H_1	Test Statistic TS	Significance Level α	p -Value if $TS = t$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\sqrt{n}(\bar{X} - \mu_0)/S$	Reject if $ TS > t_{\alpha/2, n-1}$	$2P\{T_{n-1} \geq t \}$
$\mu \leq \mu_0$	$\mu > \mu_0$	$\sqrt{n}(\bar{X} - \mu_0)/S$	Reject if $TS > t_{\alpha, n-1}$	$P\{T_{n-1} \geq t\}$
$\mu \geq \mu_0$	$\mu < \mu_0$	$\sqrt{n}(\bar{X} - \mu_0)/S$	Reject if $TS < -t_{\alpha, n-1}$	$P\{T_{n-1} \leq t\}$

T_{n-1} is a t -random variable with $n - 1$ degrees of freedom: $P\{T_{n-1} > t_{\alpha, n-1}\} = \alpha$.

8.4 TESTING THE EQUALITY OF MEANS OF TWO NORMAL POPULATIONS

A common situation faced by a practicing engineer is one in which she must decide whether two different approaches lead to the same solution. Often such a situation can be modeled as a test of the hypothesis that two normal populations have the same mean value.

8.4.1 CASE OF KNOWN VARIANCES

Suppose that X_1, \dots, X_n and Y_1, \dots, Y_m are independent samples from normal populations having unknown means μ_x and μ_y but known variances σ_x^2 and σ_y^2 . Let us consider the problem of testing the hypothesis

$$H_0 : \mu_x = \mu_y$$

versus the alternative

$$H_1 : \mu_x \neq \mu_y$$

Since \bar{X} is an estimate of μ_x and \bar{Y} of μ_y , it follows that $\bar{X} - \bar{Y}$ can be used to estimate $\mu_x - \mu_y$. Hence, because the null hypothesis can be written as $H_0 : \mu_x - \mu_y = 0$, it seems reasonable to reject it when $\bar{X} - \bar{Y}$ is far from zero. That is, the form of the test should be to

$$\begin{aligned} \text{reject } H_0 & \text{ if } |\bar{X} - \bar{Y}| > c \\ \text{accept } H_0 & \text{ if } |\bar{X} - \bar{Y}| \leq c \end{aligned} \tag{8.4.1}$$

for some suitably chosen value c .

To determine that value of c that would result in the test in Equations 8.4.1 having a significance level α , we need determine the distribution of $\bar{X} - \bar{Y}$ when H_0 is true. However, as was shown in Section 7.3.2,

$$\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_x - \mu_y, \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}\right)$$

which implies that

$$\frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \sim \mathcal{N}(0, 1) \tag{8.4.2}$$

Hence, when H_0 is true (and so $\mu_x - \mu_y = 0$), it follows that

$$(\bar{X} - \bar{Y}) \Bigg/ \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}$$

has a standard normal distribution; and thus

$$P_{H_0} \left\{ -z_{\alpha/2} \leq \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}} \leq z_{\alpha/2} \right\} = 1 - \alpha \quad (8.4.3)$$

From Equation 8.4.3, we obtain that the significance level α test of $H_0 : \mu_x = \mu_y$ versus $H_1 : \mu_x \neq \mu_y$ is

$$\begin{aligned} \text{accept } H_0 \quad \text{if } \frac{|\bar{X} - \bar{Y}|}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} \leq z_{\alpha/2} \\ \text{reject } H_0 \quad \text{if } \frac{|\bar{X} - \bar{Y}|}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} \geq z_{\alpha/2} \end{aligned}$$

Program 8.4.1 will compute the value of the test statistic $(\bar{X} - \bar{Y})/\sqrt{\sigma_x^2/n + \sigma_y^2/m}$.

EXAMPLE 8.4a Two new methods for producing a tire have been proposed. To ascertain which is superior, a tire manufacturer produces a sample of 10 tires using the first method and a sample of 8 using the second. The first set is to be road tested at location A and the second at location B. It is known from past experience that the lifetime of a tire that is road tested at one of these locations is normally distributed with a mean life due to the tire but with a variance due (for the most part) to the location. Specifically, it is known that the lifetimes of tires tested at location A are normal with standard deviation equal to 4,000 kilometers, whereas those tested at location B are normal with $\sigma = 6,000$ kilometers. If the manufacturer is interested in testing the hypothesis that there is no appreciable difference in the mean life of tires produced by either method, what conclusion should be drawn at the 5 percent level of significance if the resulting data are as given in Table 8.3?

TABLE 8.3 *Tire Lives in Units of 100 Kilometers*

Tires Tested at A	Tires Tested at B
61.1	62.2
58.2	56.6
62.3	66.4
64	56.2
59.7	57.4
66.2	58.4
57.8	57.6
61.4	65.4
62.2	
63.6	

SOLUTION A simple computation (or the use of Program 8.4.1) shows that the value of the test statistic is .066. For such a small value of the test statistic (which has a standard normal distribution when H_0 is true), it is clear that the null hypothesis is accepted. ■

It follows from Equation 8.4.1 that a test of the hypothesis $H_0 : \mu_x = \mu_y$ (or $H_0 : \mu_x \leq \mu_y$) against the one-sided alternative $H_1 : \mu_x > \mu_y$ would be to

$$\begin{aligned} \text{accept } H_0 \quad \text{if} \quad \bar{X} - \bar{Y} &\leq z_\alpha \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \\ \text{reject } H_0 \quad \text{if} \quad \bar{X} - \bar{Y} &> z_\alpha \sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}} \end{aligned}$$

8.4.2 CASE OF UNKNOWN VARIANCES

Suppose again that X_1, \dots, X_n and Y_1, \dots, Y_m are independent samples from normal populations having respective parameters (μ_x, σ_x^2) and (μ_y, σ_y^2) , but now suppose that all four parameters are unknown. We will once again consider a test of

$$H_0 : \mu_x = \mu_y \quad \text{versus} \quad H_1 : \mu_x \neq \mu_y$$

To determine a significance level α test of H_0 we will need to make the additional assumption that the unknown variances σ_x^2 and σ_y^2 are equal. Let σ^2 denote their value — that is,

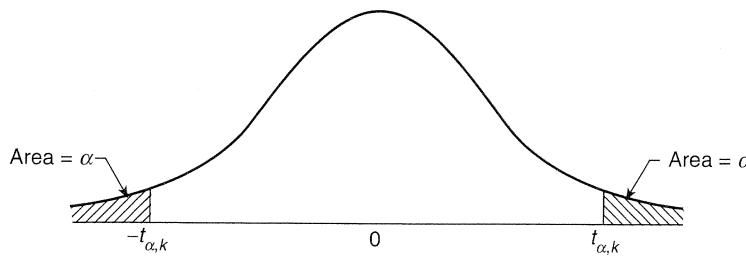
$$\sigma^2 = \sigma_x^2 = \sigma_y^2$$

As before, we would like to reject H_0 when $\bar{X} - \bar{Y}$ is “far” from zero. To determine how far from zero it need be, let

$$\begin{aligned} S_x^2 &= \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \\ S_y^2 &= \frac{\sum_{i=1}^m (Y_i - \bar{Y})^2}{m-1} \end{aligned}$$

denote the sample variances of the two samples. Then, as was shown in Section 7.3.2,

$$\frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{S_p^2(1/n + 1/m)}} \sim t_{n+m-2}$$

FIGURE 8.5 Density of a t -random variable with k degrees of freedom.

where S_p^2 , the *pooled* estimator of the common variance σ^2 , is given by

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$$

Hence, when H_0 is true, and so $\mu_x = \mu_y = 0$, the statistic

$$T \equiv \frac{\bar{X} - \bar{Y}}{\sqrt{S_p^2(1/n + 1/m)}}$$

has a t -distribution with $n+m-2$ degrees of freedom. From this, it follows that we can test the hypothesis that $\mu_x = \mu_y$ as follows:

$$\begin{aligned} \text{accept } H_0 & \text{ if } |T| \leq t_{\alpha/2, n+m-2} \\ \text{reject } H_0 & \text{ if } |T| > t_{\alpha/2, n+m-2} \end{aligned}$$

where $t_{\alpha/2, n+m-2}$ is the 100 $\alpha/2$ percentile point of a t -random variable with $n+m-2$ degrees of freedom (see Figure 8.5).

Alternatively, the test can be run by determining the p -value. If T is observed to equal v , then the resulting p -value of the test of H_0 against H_1 is given by

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

where T_{n+m-2} is a t -random variable having $n+m-2$ degrees of freedom.

If we are interested in testing the one-sided hypothesis

$$H_0 : \mu_x \leq \mu_y \quad \text{versus} \quad H_1 : \mu_x > \mu_y$$

then H_0 will be rejected at large values of T . Thus the significance level α test is to

$$\begin{aligned} \text{reject } H_0 & \text{ if } T \geq t_{\alpha, n+m-2} \\ \text{not reject } H_0 & \text{ otherwise} \end{aligned}$$

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

$$p\text{-value} = P\{T_{n+m-2} \geq v\}$$

Program 8.4.2 computes both the value of the test statistic and the corresponding p -value.

EXAMPLE 8.4b Twenty-two volunteers at a cold research institute caught a cold after having been exposed to various cold viruses. A random selection of 10 of these volunteers was given tablets containing 1 gram of vitamin C. These tablets were taken four times a day. The control group consisting of the other 12 volunteers was given placebo tablets that looked and tasted exactly the same as the vitamin C tablets. This was continued for each volunteer until a doctor, who did not know if the volunteer was receiving the vitamin C or the placebo tablets, decided that the volunteer was no longer suffering from the cold. The length of time the cold lasted was then recorded.

At the end of this experiment, the following data resulted.

Treated with Vitamin C	Treated with Placebo
5.5	6.5
6.0	6.0
7.0	8.5
6.0	7.0
7.5	6.5
6.0	8.0
7.5	7.5
5.5	6.5
7.0	7.5
6.5	6.0
	8.5
	7.0

Do the data listed prove that taking 4 grams daily of vitamin C reduces the mean length of time a cold lasts? At what level of significance?

SOLUTION To prove the above hypothesis, we would need to reject the null hypothesis in a test of

$$H_0 : \mu_p \leq \mu_c \quad \text{versus} \quad H_1 : \mu_p > \mu_c$$

where μ_c is the mean time a cold lasts when the vitamin C tablets are taken and μ_p is the mean time when the placebo is taken. Assuming that the variance of the length of the cold is the same for the vitamin C patients and the placebo patients, we test the above by running Program 8.4.2. This yields the information shown in Figure 8.6. Thus H_0 would be rejected at the 5 percent level of significance.

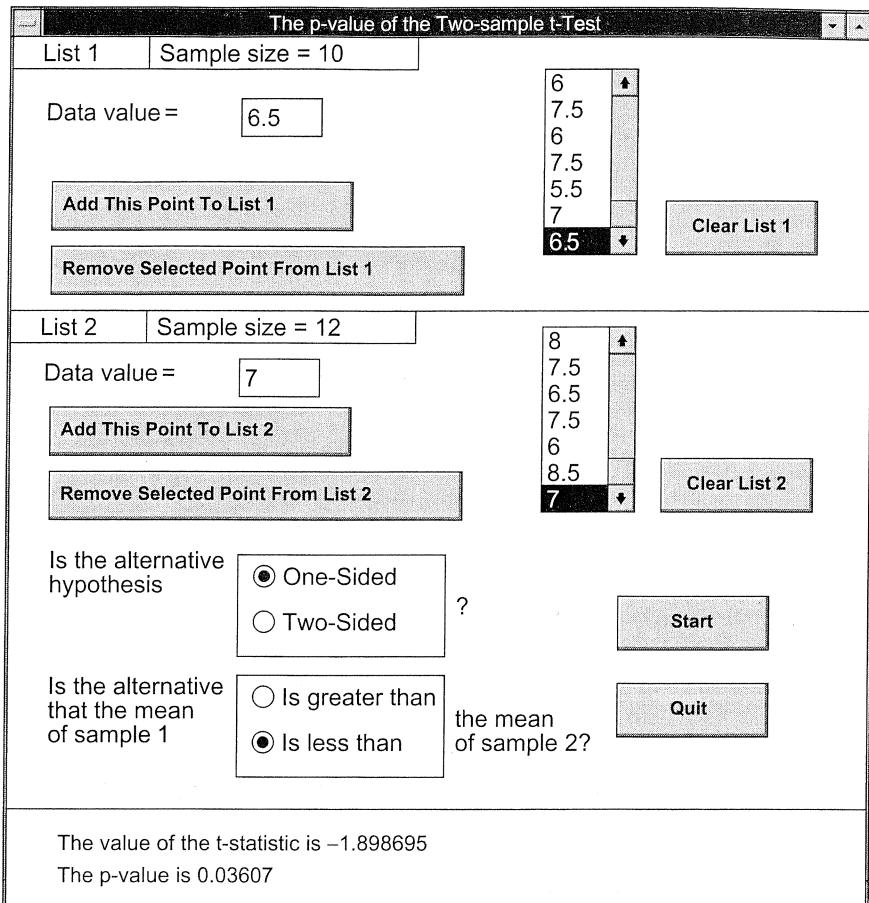


FIGURE 8.6

Of course, if it were not convenient to run Program 8.4.2 then we could have performed the test by first computing the values of the statistics \bar{X} , \bar{Y} , S_x^2 , S_y^2 , and S_p^2 , where the X sample corresponds to those receiving vitamin C and the Y sample to those receiving a placebo. These computations would give the values

$$\bar{X} = 6.450, \quad \bar{Y} = 7.125$$

$$S_x^2 = .581, \quad S_y^2 = .778$$

Therefore,

$$S_p^2 = \frac{9}{20}S_x^2 + \frac{11}{20}S_y^2 = .689$$

and the value of the test statistic is

$$TS = \frac{-0.675}{\sqrt{.689(1/10 + 1/12)}} = -1.90$$

Since $t_{0.5,20} = 1.725$, the null hypothesis is rejected at the 5 percent level of significance. That is, at the 5 percent level of significance the evidence is significant in establishing that vitamin C reduces the mean time that a cold persists. ■

EXAMPLE 8.4c Reconsider Example 8.4a, but now suppose that the population variances are unknown but equal.

SOLUTION Using Program 8.4.2 yields that the value of the test statistic is 1.028, and the resulting p -value is

$$p\text{-value} = P\{T_{16} > 1.028\} = .3192$$

Thus, the null hypothesis is accepted at any significance level less than .3192 ■

8.4.3 CASE OF UNKNOWN AND UNEQUAL VARIANCES

Let us now suppose that the population variances σ_x^2 and σ_y^2 are not only unknown but also cannot be considered to be equal. In this situation, since S_x^2 is the natural estimator of σ_x^2 and S_y^2 of σ_y^2 , it would seem reasonable to base our test of

$$H_0: \mu_x = \mu_y \quad \text{versus} \quad H_1: \mu_x \neq \mu_y$$

on the test statistic

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \quad (8.4.4)$$

However, the foregoing has a complicated distribution, which, even when H_0 is true, depends on the unknown parameters, and thus cannot be generally employed. The one situation in which we can utilize the statistic of Equation 8.4.4 is when n and m are both large. In such a case, it can be shown that when H_0 is true Equation 8.4.4 will have *approximately* a standard normal distribution. Hence, when n and m are large an *approximate* level α test of $H_0: \mu_x = \mu_y$ versus $H_1: \mu_x \neq \mu_y$ is to

$$\begin{aligned} \text{accept } H_0 & \text{ if } -z_{\alpha/2} \leq \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_x^2}{n} + \frac{S_y^2}{m}}} \leq z_{\alpha/2} \\ \text{reject } & \text{ otherwise} \end{aligned}$$

The problem of determining an exact level α test of the hypothesis that the means of two normal populations, having unknown and not necessarily equal variances, are equal is known as the Behrens-Fisher problem. There is no completely satisfactory solution known.

Table 8.4 presents the two-sided tests of this section.

TABLE 8.4 X_1, \dots, X_n Is a Sample from a $\mathcal{N}(\mu_1, \sigma_1^2)$ Population; Y_1, \dots, Y_m Is a Sample from a $\mathcal{N}(\mu_2, \sigma_2^2)$ Population

The Two Population Samples Are Independent
To Test

$$H_0: \mu_1 = \mu_2 \text{ versus } H_1: \mu_1 \neq \mu_2$$

Assumption	Test Statistic TS	Significance Level α	Test	p -Value if $TS = t$
σ_1, σ_2 known	$\frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}}$		Reject if $ TS > z_{\alpha/2}$	$2P\{Z \geq t \}$
$\sigma_1 = \sigma_2$	$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{(n-1)\bar{S}_1^2 + (m-1)\bar{S}_2^2}{n+m-2} \sqrt{1/n+1/m}}}$		Reject if $ TS > t_{\alpha/2, n+m-2}$	$2P\{T_{n+m-2} \geq t \}$
n, m large	$\frac{\bar{X} - \bar{Y}}{\sqrt{S_1^2/n + S_2^2/m}}$		Reject if $ TS > z_{\alpha/2}$	$2P\{Z \geq t \}$

8.4.4 THE PAIRED t -TEST

Suppose we are interested in determining whether the installation of a certain antipollution device will affect a car's mileage. To test this, a collection of n cars that do not have this device are gathered. Each car's mileage per gallon is then determined both before and after the device is installed. How can we test the hypothesis that the antipollution control has no effect on gas consumption?

The data can be described by the n pairs (X_i, Y_i) , $i = 1, \dots, n$, where X_i is the gas consumption of the i th car before installation of the pollution control device, and Y_i of the same car after installation. It is important to note that, since each of the n cars will be inherently different, we cannot treat X_1, \dots, X_n and Y_1, \dots, Y_n as being independent samples. For example, if we know that X_1 is large (say, 40 miles per gallon), we would certainly expect that Y_1 would also probably be large. Thus, we cannot employ the earlier methods presented in this section.

One way in which we can test the hypothesis that the antipollution device does not affect gas mileage is to let the data consist of each car's difference in gas mileage. That is, let $W_i = X_i - Y_i$, $i = 1, \dots, n$. Now, if there is no effect from the device, it should follow that the W_i would have mean 0. Hence, we can test the hypothesis of no effect by testing

$$H_0: \mu_w = 0 \quad \text{versus} \quad H_1: \mu_w \neq 0$$

where W_1, \dots, W_n are assumed to be a sample from a normal population having unknown mean μ_w and unknown variance σ_w^2 . But the t -test described in Section 8.3.2 shows that

this can be tested by

$$\begin{aligned} \text{accepting } H_0 \quad &\text{if } -t_{\alpha/2, n-1} < \sqrt{n} \frac{\bar{W}}{S_w} < t_{\alpha/2, n-1} \\ \text{rejecting } H_0 \quad &\text{otherwise} \end{aligned}$$

EXAMPLE 8.4d An industrial safety program was recently instituted in the computer chip industry. The average weekly loss (averaged over 1 month) in man-hours due to accidents in 10 similar plants both before and after the program are as follows:

Plant	Before	After	$A - B$
1	30.5	23	-7.5
2	18.5	21	2.5
3	24.5	22	-2.5
4	32	28.5	-3.5
5	16	14.5	-1.5
6	15	15.5	.5
7	23.5	24.5	1
8	25.5	21	-4.5
9	28	23.5	-4.5
10	18	16.5	-1.5

Determine, at the 5 percent level of significance, whether the safety program has been proven to be effective.

SOLUTION To determine this, we will test

$$H_0 : \mu_A - \mu_B \geq 0 \quad \text{versus} \quad H_1 : \mu_A - \mu_B < 0$$

because this will enable us to see whether the null hypothesis that the safety program has not had a beneficial effect is a reasonable possibility. To test this, we run Program 8.3.2, which gives the value of the test statistic as -2.266, with

$$p\text{-value} = P\{T_q \leq -2.266\} = .025$$

Since the *p*-value is less than .05, the hypothesis that the safety program has not been effective is rejected and so we can conclude that its effectiveness has been established (at least for any significance level greater than .025). ■

Note that the paired-sample *t*-test can be used even though the samples are not independent and the population variances are unequal.