Part III

Drawing Conclusions from Samples

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Electrica	4	9.2	2.3	0.446667
Readyforever	4	7.4	1.85	0.043333
Voltagenow	4	8.6	2.15	0.016667

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.45	2	0.21	1.243421	0.333571	4.256495
Within Groups	1.52	9	0.168889			
Total	1.94	11				

In this part...

- ✔ Use confidence intervals to provide a range of possible values for a population parameter; these can be constructed for any population parameter: mean, variance, standard deviation, and so on.

- ✔ Use t-distribution to describe the statistical properties of sample means that are estimated from *small* samples; use standard normal distribution is for *large* samples.

- ✔ Draw conclusions about population properties — from a single population variance to multiple population variances — with hypothesis testing

- ✔ Test multiple population means with a special new technique called Analysis of Variance (ANOVA). This technique can be used to evaluate claims about the effectiveness of competing products, or identify the most profitable products to produce.

Chapter 11

Confidence Intervals and the Student's t-Distribution

. . . .

In This Chapter

▶ Getting familiar with the t-distribution

▶ Developing techniques for constructing confidence intervals

. . . .

A *confidence interval* is a range of numbers that's likely to contain the true value of an unknown population *parameter*, such as the population mean. (Parameters are numerical values that describe the properties of a population; they are discussed in Chapter 10.)

Here's an example. Suppose you are asked to estimate how long it takes to commute to work each day. You respond by saying, "On average, it takes about 20 minutes to get to work." This estimate may be useful, but it doesn't give any indication how much your commuting time may vary from one day to the next.

Suppose instead you respond by saying "Most days, it takes between 15 and 25 minutes to get to work." This range of values is more meaningful than the estimated average time of 20 minutes. With this interval, it's clear that the average commute time is 20 minutes (because this is halfway between 15 and 25 minutes.) In addition, the numbers tell you that it'll be an unusual day if your commuting time is more than 25 minutes or fewer than 15 minutes.

This range of estimated values is known as a *confidence interval*. The starting point in constructing a confidence interval is the estimated mean or average, which is 20 minutes in this example. The next step is to construct a *margin of error*, which represents the degree of uncertainty associated with the estimated mean. In this example, the margin of error is five minutes.

Confidence intervals may be constructed for any population parameter: mean, variance, standard deviation, etc. This chapter covers the techniques that are used to estimate confidence intervals for the population mean. These techniques

are based on one of two probability distributions. One of these is the standard normal distribution (which I cover in detail in Chapter 9). The other is known as the *Student's t-distribution* (also known simply as the *t-distribution*) — which I introduce in this chapter.

Almost Normal: The Student's t-Distribution

The purpose of the t-distribution is to describe the statistical properties of sample means that are estimated from *small* samples; the standard normal distribution is used for *large* samples.

The means of small samples are likely to vary more dramatically from one sample to the next than the means of large samples. (In a small sample, a single observation that is unusually large or small will have a greater impact on the sample mean than it would in a larger sample.) It therefore makes sense that different probability distributions should be used to describe the properties of small and large sample means.

Properties of the t-distribution

The t-distribution shares a few key properties with the standard normal distribution (which is discussed in Chapter 9).

Properties shared by the t-distribution and the standard normal distribution

The properties shared by the t-distribution and the standard normal distribution are as follows:

- They have a mean of 0.
- They're *symmetric* about the mean (that is, the area below the mean is a mirror image of the area above the mean).
- They can be described graphically with a bell-shaped curve.

Several key differences also exist between the two distributions, including the following:

- The t-distribution has more area in the "tails," and less area near the mean than the standard normal distribution.
- The variance and standard deviation of the t-distribution are larger than those of the standard normal distribution.

The larger variance and standard deviation in the t-distribution reflect that much more variability occurs among the means of small samples than among the means of large samples.

Degrees of freedom

As with the normal distribution, the t-distribution is an infinite family of distributions. Whereas the mean and standard deviation uniquely identify each normal distribution, each t-distribution is characterized by a value known as *degrees of freedom* (df).

When you're estimating the sample mean, the number of degrees of freedom for the t-distribution equals the number of sample members that can *vary*. For example, if you choose a sample of size n to estimate the sample mean $\overline{X}$, the corresponding t-distribution has $n - 1$ degrees of freedom because the combination of $n - 1$ elements in the sample plus the sample mean uniquely identify the last element in the sample. Therefore, you have only $n - 1$ independent variables in the sample.

Suppose you choose a sample of three students to estimate the mean GPA of a university. If the sample mean, $\overline{X}$, equals 3.0, the first student's GPA is 2.5, the second student's GPA is 3.5, and the third student's GPA must be 3.0 because the sum of the GPAs must be 9.0 for the sample mean to be 3.0. As a result, the GPAs of any two students in this sample, along with the value of $\overline{X}$, uniquely determine the value of the third student's GPA. Therefore, the corresponding t-distribution has two degrees of freedom.

Moments of the t-distribution

A *moment* is a summary measure of a probability distribution (see Chapter 7 for a detailed explanation on moments). Probability distributions, including the t-distribution, have several moments, including:

- ✔ The first moment of a distribution is the expected value, $E(X)$, which represents the mean or average value of the distribution.

  For the t-distribution with ν degrees of freedom, the mean (or expected value) equals $\mu = E(X) = 0$. μ represents the mean of a population or a probability distribution, and ν commonly designates the number of degrees of freedom of a distribution.

- ✔ The second central moment is the variance (σ^2), and it measures the spread of the distribution about the expected value. The more spread out a distribution is, the more "stretched out" is the graph of the distribution. In other words, the tails will be further from the mean, and the area near the mean will be smaller. For example, based on Figures 11-1 and 11-3, it can be seen that the t-distribution with 2 degrees of freedom is far more spread out than the t-distribution with 30 degrees of freedom.

  You use the formula $\sigma^2 = \dfrac{\nu}{\nu - 2}$ to calculate the variance of the t-distribution.

As an example, with 10 degrees of freedom, the variance of the t-distribution is computed by substituting 10 for v in the variance formula:

$$\sigma^2 = \frac{v}{v-2}$$
$$= \frac{10}{10-2}$$
$$= \frac{10}{8}$$
$$= 1.25$$

With 30 degrees of freedom, the variance of the t-distribution equals

$$\sigma^2 = \frac{v}{v-2}$$
$$= \frac{30}{30-2}$$
$$= \frac{30}{28}$$
$$= 1.07$$

These calculations show that as the degrees of freedom increases, the variance of the t-distribution declines, getting progressively closer to 1.

✔ The standard deviation is the square root of the variance (σ). (It is not a separate moment.)

For the t-distribution, you find the standard deviation with this formula:

$$\sigma = \sqrt{\frac{v}{v-2}}$$

TIP

For most applications, the standard deviation is a more useful measure than the variance because the standard deviation and expected value are measured in the same units while the variance is measured in *squared* units. For example, suppose you assume that the returns on a portfolio follow the t-distribution. You measure both the expected value of the returns and the standard deviation as a percentage; you measure the variance as a *squared* percentage, which is a difficult concept to interpret.
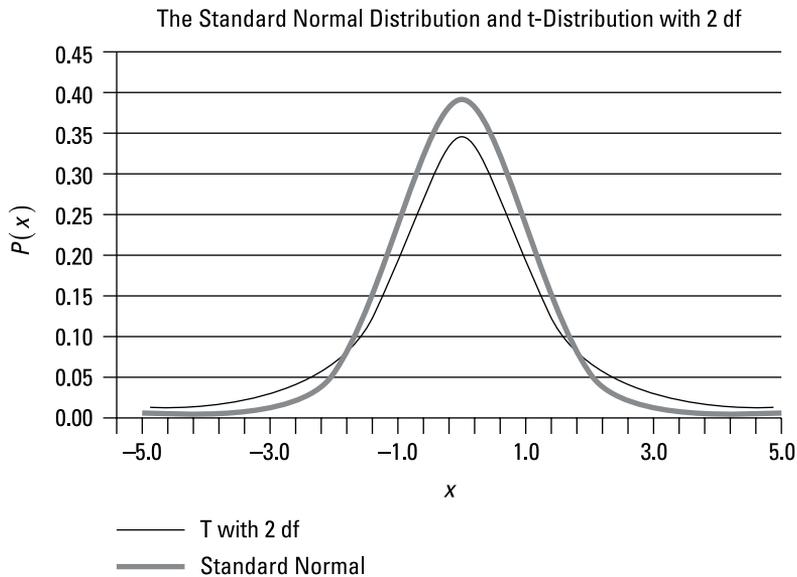
Graphing the t-distribution

One of the interesting properties of the t-distribution is that the greater the degrees of freedom, the more closely the t-distribution resembles the standard normal distribution. As the degrees of freedom increases, the area in the tails of the t-distribution decreases while the area near the center

increases. (The tails consists of the extreme values of the distribution, both negative and positive.) Eventually, when the degrees of freedom reaches 30 or more, the t-distribution and the standard normal distribution are extremely similar.

Figures 11-1, 11-2, and 11-3 illustrate the relationship between the t-distribution with different degrees of freedom and the standard normal distribution. Figure 11-1 shows the standard normal and the t-distribution with two degrees of freedom (df). Notice how the t-distribution is significantly more spread out than the standard normal distribution.

The graph in Figure 11-1 shows that the t-distribution has more area in the tails and less area around the mean than the standard normal distribution. (The standard normal distribution curve is shown with square markers.) As a result, more extreme observations (positive and negative) are likely to occur under the t-distribution than under the standard normal distribution.

The Standard Normal Distribution and t-Distribution with 2 df

Figure 11-1:
The standard normal and t-distribution with two degrees of freedom.

Figure 11-2 compares the standard normal distribution with the t-distribution with ten degrees of freedom. The two are much closer to each other here than in Figure 11-1.

The Standard Normal Distribution and t-Distribution with 10 df

Figure 11-2:
The standard normal
and
t-distribution
with ten
degrees of
freedom.
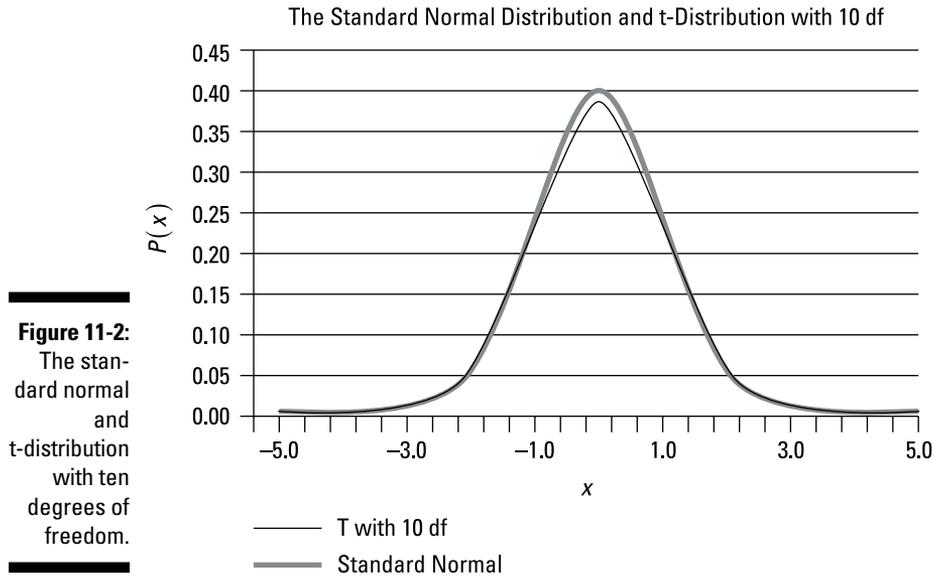
As you can see in Figure 11-3, with 30 degrees of freedom, the t-distribution and the standard normal distribution are almost indistinguishable.

The Standard Normal Distribution and t-Distribution with 30 df

Figure 11-3:
The standard normal
and
t-distribution
with 30
degrees of
freedom.
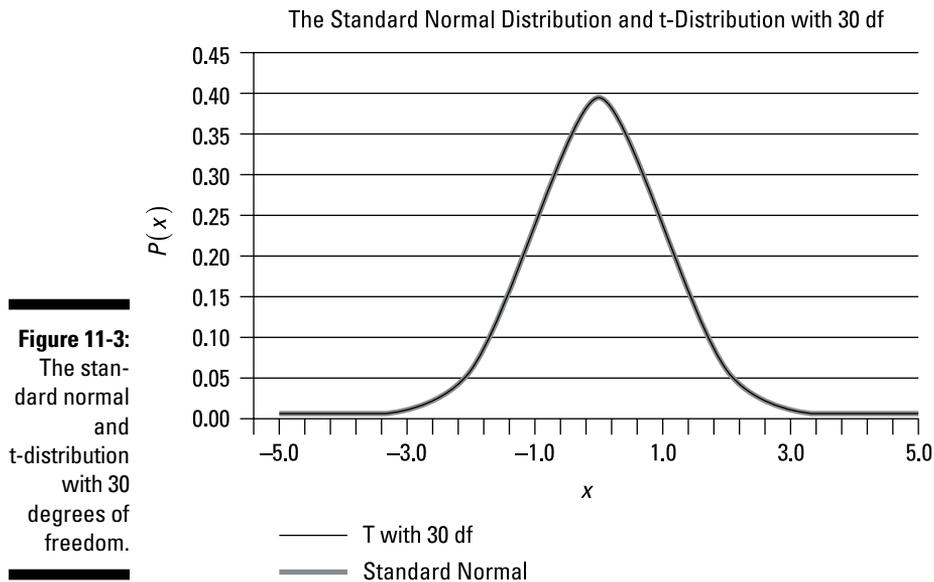
Probabilities and the t-table

The t-table is used to show probabilities for the t-distribution. The top row of the t-table lists different values of $t\alpha$, where the right tail of the t-distribution has a probability (area) equal to α ("alpha"). Table 11-1 is an excerpt from the full t-table.

Table 11-1		The t-Table			
Degrees of Freedom	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169

Table 11-1 shows that with ten degrees of freedom and with $\alpha = 0.05$, $t\alpha = 1.812$. So the right 5 percent tail of the distribution is located 1.812 standard deviations above the mean.

Alternatively, assume X is a *random variable* that follows the t-distribution with 10 degrees of freedom. (Random variables are discussed in Chapter 7.) In this case, $P(X \geq 1.812) = 0.05$. This is equivalent to saying the area under the curve to the right of 1.812 is 0.05, or 5 percent of the total area (see Figure 11-4).
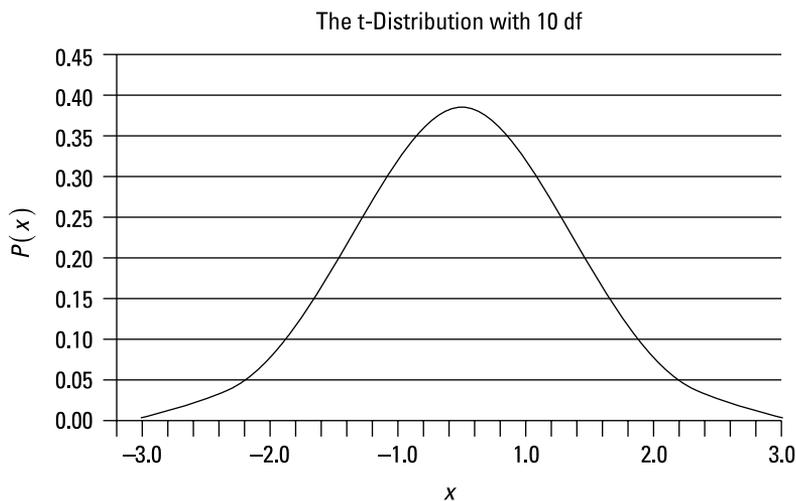
Figure 11-4: The t-distribution with 10 degrees of freedom.

The shaded region starts at 1.812, which represents 1.812 standard deviations above the mean. The total area of the shaded region is 0.05 or 5 percent; therefore, the probability that a t-distributed random variable with 10 degrees of freedom exceeds 1.812 is 5 percent.

Point estimates vs. interval estimates

When you don't know the mean, standard deviation, variance, and other summary measures of a population, you need to estimate them from a sample.

To estimate the mean of a population, you use the mean of a sample drawn from the population. You express the sample mean as $\bar{X}$ ("X bar"). In a similar manner, to estimate the variance of a population, you use the sample variance, s^2. And you estimate the standard deviation of a population with the sample standard deviation, *s.* (I cover techniques for estimating the sample variance and standard deviation in Chapter 4.)

These sample measures are formally known as *point estimators* — formulas that help estimate a population measure. For example, $\bar{X}$ is a point estimator of the population mean μ. The numerical value of $\bar{X}$ is a *point estimate.*

The distinction between *estimator* and *estimate* seems very subtle — an estimator is a formula, and an estimate is a numerical value.

The usefulness of a point estimator (formula) is limited by the fact that it produces only a single number. Suppose a portfolio manager wants to estimate the mean annual return of a stock he holds by choosing a sample of historical returns and calculating the sample mean. Say the sample mean turns out to be 8 percent. This info is useful, but it's difficult to judge how much the stock's returns may fluctuate from one year to the next based on this result.

Instead, suppose that the portfolio manager can estimate, with 95 percent certainty, that the return on the stock is between 6 and 10 percent, showing the stock's returns are relatively stable over time — the stock isn't extremely risky. The estimated range from 6 to 10 percent is an *interval estimate.*

In general, you compute an interval estimate with this formula:

> point estimate ± margin of error

This can be written as:

> (point estimate – margin of error, point estimate + margin of error)

The symbol ± indicates that two values exist: point estimate – margin of error, and point estimate + margin of error.

The margin of error depends on several factors, such as the type of point estimate being used, the size of the sample being used to construct the point estimate, and so forth. The margin of error is a measure of the degree of uncertainty associated with the point estimate.

Calculate an interval estimate of the population mean (μ) with this formula:

$$\bar{X} \pm \text{margin of error}$$

The margin of error is a measure of how much uncertainty is associated with the value of $\bar{X}$. Its value is closely related to the standard deviation of the underlying population and the size of the sample used to estimate $\bar{X}$.

Estimating confidence intervals for the population mean

A confidence interval is a specific type of interval estimate characterized by:

- ✔ A confidence coefficient, expressed as $(1 - \alpha)$

  α is known as the *level of significance*. For example, if you choose the level of significance to be 0.05, then the corresponding confidence coefficient equals $(1 - \alpha) = (1 - 0.05) = 0.95$.

- ✔ A confidence level, expressed as $100(1 - \alpha)$

  For example, if the confidence coefficient equals 0.95, then the corresponding confidence level equals $100(0.95) = 0.95 = 95$ percent.

Suppose that a 95 percent confidence interval is constructed for the population mean age in the United States based on the ages of people randomly chosen throughout the country. If this process is repeated 100 times (for example, 100 samples are drawn and a new confidence interval is estimated in each case), then you would expect that the true population mean age is contained in 95 of these 100 confidence intervals.

Two possible situations may arise when constructing a confidence interval for the population mean: A known population standard deviation and an unknown population standard deviation that you must estimate with the sample standard deviation(s). I discuss these situations in the following sections.

Known population standard deviation

If you know the population standard deviation, then the confidence interval is based on the *standard normal* distribution (which I discuss in detail in Chapter 9). Here's the formula for constructing this confidence interval:

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

where:

$\bar{X}$ is the sample mean.

σ is the population standard deviation.

n is the sample size.

α is the level of significance.

$Z_{\alpha/2}$ is a *quantile* or *critical value,* which represents the location of the right tail of the standard normal distribution with an area of $\alpha/2$.

$Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ is the margin of error.

The confidence interval can also be written as:

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

The two values contained in this interval are known as:

✔ The *lower limit* of the confidence interval: $\bar{X} - Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$

✔ The *upper limit* of the confidence interval: $\bar{X} + Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$

For example, suppose you want to construct a 95 percent confidence interval. This implies that $\alpha = 0.05$ (or 5 percent) so that $\alpha/2 = 0.025$ or 2.5 percent.

You can find the value of $Z_{\alpha/2}$ from a standard normal probability table, such as shown in Table 11-2. The standard normal table shows probabilities below of a specific value. Because the area above $Z_{\alpha/2} = 0.025$, the area below $Z_{\alpha/2} = 1 - 0.025 = 0.975$ (due to the *symmetry* of the standard normal distribution).

By searching in the body of the standard normal table for the area 0.9750, you get the appropriate value of $Z_{\alpha/2}$. See Table 11-2 for this result.

Table 11-2	The Standard Normal Table		
Z	**0.05**	**0.06**	**0.07**
1.7	0.9599	0.9608	0.9616
1.8	0.9678	0.9686	0.9693
1.9	0.9744	0.9750	0.9756
2.0	0.9798	0.9803	0.9812

Table 11-2 shows that the appropriate value of $Z_{\alpha/2}$ is 1.96. (You find the value 1.9 in the first [Z] column and the value 0.06 in the third [0.06] column.)

You construct many confidence intervals using a 90 percent confidence level, a 95 percent confidence level, or a 99 percent confidence level. In these three cases, the value of $Z_{\alpha/2}$ is as follows:

Confidence Level	$Z_{\alpha/2}$
90 percent	1.645
95 percent	1.960
99 percent	2.576

These results indicate that for a standard normal random variable Z, the following expressions are true:

$P(Z \leq 1.645) = 0.9500$

$P(Z \geq 1.645) = 0.0500$

$P(Z \leq 1.960) = 0.9750$

$P(Z \geq 1.960) = 0.0250$

$P(Z \leq 2.576) = 0.9950$

$P(Z \geq 2.576) = 0.0050$

The resulting confidence interval may then be expressed as follows:

$$P\left(\overline{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

This expression shows that the population mean is contained within this interval with a level of confidence equal to $100(1 - \alpha)$.

For example, suppose that a hedge fund holds a portfolio consisting of 500 stocks. The standard deviation is 20 percent. If you choose a sample of 10 stocks and determine the sample mean to be 8 percent, you construct a 90 percent confidence interval by following these steps:

1. **Figure $\alpha/2$.**

   $100(1 - \alpha) = 90$ percent

   $\alpha = 0.10$

   $\alpha/2 = 0.05$

2. **Use the standard normal table (Table 11-2) to find the critical value: $Z_{\alpha/2} = 1.645$.**

3. **Compute the confidence interval.**

> sample size: $n = 10$
>
> population standard deviation: $\sigma = 0.20$
>
> sample mean: $\overline{X} = 0.08$

Therefore, the appropriate confidence interval is

$$\overline{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 0.08 \pm 1.645 \frac{0.20}{\sqrt{10}}$$
$$= 0.08 \pm 0.104$$
$$= (-0.024,\ 0.184)$$
$$= (-2.4\%,\ 18.4\%)$$

For a 95 percent confidence interval, the only change you need to make is to the critical value, which you determine as follows:

> $100(1 - \alpha) = 95$ percent
>
> $\alpha = 0.05$
>
> $\alpha/2 = 0.025$
>
> $Z_{\alpha/2} = Z_{0.025} = 1.96$

The 95 percent confidence interval is

$$\overline{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 0.08 \pm 1.96 \frac{0.20}{\sqrt{10}}$$
$$= 0.08 \pm 0.124$$
$$= (-0.044,\ 0.204)$$
$$= (-4.4\%,\ 20.4\%)$$

Finally, you determine a 99 percent confidence interval with these adjustments:

> $100(1 - \alpha) = 99$ percent
>
> $\alpha = 0.01$
>
> $\alpha/2 = 0.005$
>
> $Z_{\alpha/2} = Z_{0.025} = 2.576$

The 99 percent confidence interval is

$$\overline{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = 0.08 \pm 2.576 \frac{0.20}{\sqrt{10}}$$
$$= 0.08 \pm 0.163$$
$$= (-0.083,\ 0.243)$$
$$= (-8.3\%,\ 24.3\%)$$

TIP

As the level of confidence increases so does the *width* of the confidence interval because the only way to have more confidence that the interval actually contains the population mean is to include more values.

Unknown population standard deviation

If the population standard deviation is *not* known, then you compute an interval estimate for the population mean as follows:

$$\bar{X} \pm t_{\alpha/2}^{n-1} \frac{s}{\sqrt{n}}$$

where:

$t_{\alpha/2}^{n-1}$ is a quantile (critical value) which represents the location of the right tail of the t-distribution with $n-1$ degrees of freedom with an area of $\alpha/2$

s is the sample standard deviation

In this case, you make the following changes to the formula:

✔ You use the *sample* standard deviation (*s*) rather than the population standard deviation.

✔ You use the t-distribution rather than the standard normal distribution because of the greater uncertainty associated with the sample standard deviation. $t_{\alpha/2}^{n-1}$ is a quantile or critical value taken from the t-distribution and represents the location of the right tail of the t-distribution with $n-1$ degrees of freedom whose area equals $\alpha/2$.

As an example, suppose that $\alpha = 0.05$ so that $\alpha/2 = 0.025$. Also assume that the appropriate number of degrees of freedom is 9. You can get the value of $t_{\alpha/2}^{n-1}$ from a t-table, as in Table 11-1.

The appropriate column heading is $t_{0.025}$; with nine degrees of freedom, the value of $t_{\alpha/2}^{n-1}$ is 2.262.

For example, a university has 10,000 students and wants to estimate the average GPA of the entire student body. It picks a sample of ten students, and the sample mean GPA is 3.10. The sample standard deviation is 0.25. You construct confidence intervals for the population mean as follows:

✔ For a 90 percent confidence interval, the value of $\alpha/2$ is 0.05:

$100(1 - \alpha) = 90$ percent

$\alpha = 0.10$

$\alpha/2 = 0.05$

With $n - 1 = 9$ degrees of freedom, based on the t-table (Table 11-1), $t_{\alpha/2}^{n-1} = 1.833$.

The sample size is $n = 10$, the population standard deviation is $s = 0.25$, and the sample mean is $\overline{X} = 3.10$. Therefore, the appropriate confidence interval is

$$\overline{X} \pm t_{\alpha/2}^{n-1} \frac{s}{\sqrt{n}} = 3.10 + 1.833 \frac{0.25}{\sqrt{10}}$$

$$= 3.10 \pm 0.1449$$

$$= (2.955,\ 3.2449)$$

✔ For a 95 percent confidence interval, you follow similar calculations but change the critical value:

$$100(1 - \alpha) = 95 \text{ percent}$$

$$\alpha = 0.05$$

$$\alpha/2 = 0.025$$

With $n - 1 = 9$ degrees of freedom, based on the t-table (Table 11-1), $t_{\alpha/2}^{n-1} = 2.262$. Therefore, the appropriate confidence interval is

$$\overline{X} \pm t_{\alpha/2}^{n-1} \frac{s}{\sqrt{n}} = 3.10 + 2.262 \frac{0.25}{\sqrt{10}}$$

$$= 3.10 \pm 0.1788$$

$$= (2.9212,\ 3.2788)$$

✔ For a 99 percent confidence interval, you again change the critical value to

$$100(1 - \alpha) = 99 \text{ percent}$$

$$\alpha = 0.01$$

$$\alpha/2 = 0.005$$

With $n - 1 = 9$ degrees of freedom, based on the t-table (Table 11-1), $t_{\alpha/2}^{n-1} = 3.250$. Therefore, the appropriate confidence interval is

$$\overline{X} \pm t_{\alpha/2}^{n-1} \frac{s}{\sqrt{n}} = 3.10 + 3.250 \frac{0.25}{\sqrt{10}}$$

$$= 3.10 \pm 0.2569$$

$$= (2.8431,\ 3.3569)$$

In each case, the confidence interval is wider than it would be when using the standard normal distribution.

Chapter 12

Testing Hypotheses about the Population Mean

In This Chapter

▶ Understanding the hypothesis testing process

▶ Testing hypotheses about two population means

*H*ypothesis testing is a multi-step statistical process which is used to test claims about a population measure, such as the mean. For example, you can use hypothesis testing on the following statements to determine whether they're true:

✔ Mean income in the United States has risen over the past 25 years.

✔ The average age of the population of Egypt is above 30.

✔ The average return to the stocks in a portfolio is 10 percent.

✔ The United States and Canada have average work weeks identical in length.

✔ The average lifetime of brandy drinkers is 90.

You test hypotheses with a series of steps designed to show whether you can justify a claim. These steps apply to a lot of situations; for example, you can test claims about a population's mean, a population's variance, whether a population is normally distributed, and so forth.

This chapter focuses on testing hypotheses about the mean value of a single population and the equality of the means of two different populations.

Applying the Key Steps in Hypothesis Testing for a Single Population Mean

Hypothesis testing requires sample data to draw conclusions about the characterstics of the underlying population. The necessary steps for any type of hypothesis test are outlined in the following sections.

Writing the null hypothesis

The *null hypothesis* is a statement that's assumed to be true unless strong contrary evidence exists. The null hypothesis can take several forms. You can use it to test statements about population measures, such as means and standard deviations, and to test statements about the relationship between two populations. An example of a null hypothesis is the mean IQ of Star Trek fans is higher than the mean IQ of Star Wars fans.

You write the null hypothesis for testing the value of a single population mean as

$$H_0: \mu = \mu_0$$

where H_0 stands for the null hypotheses, μ is the true population mean (whose value we do not know,) and μ_0 is the hypothesized value of the population, or the value that you *think* is true.

For example, if you want to test the hypothesis that the mean number of runs scored per game in the American League is 4; you write the null hypothesis as $H_0: \mu = 4.0$.

If actual data shows that this is false, you *reject* the null hypothesis; otherwise, you *don't* reject the null hypothesis. (You never *accept* the null hypothesis; instead, you fail to reject it if there is not enough evidence against it.)

Coming up with an alternative hypothesis

Suppose that the null hypothesis is false. For example, you are testing the null hypothesis that the mean number of runs scored per game in the American League is 4. If data taken from actual games shows that this is false, it must be true that:

The number of runs scored is *more than* 4

The number of runs scored is *less than* 4

Prior to testing the null hypothesis, you must specify what alternative you accept if the null hypothesis is false. It turns out that there are actually three ways to express the alternative hypothesis:

The number of runs scored is *more than* 4

The number of runs scored is *less than* 4

The number of runs is *different from* 4

The alternative that you choose depends on what type of action is taken as a result of the hypothesis test. For example, suppose that the commissioner decides that if the number of runs scored is *less than* 4, the league encourages teams to shorten the distance to their outfield fences (which encourages more home runs.) You therefore use "the number of runs scored is *less than* 4" as your alternative hypothesis. This ensures that no action is taken unless it's extremely clear that the number of runs is less than 4.

There are special names associated with the three types of alternative hypotheses:

✔ Right-tailed test

✔ Left-tailed test

✔ Two-tailed test

A right-tailed test indicates that the actual population mean is *greater than* the hypothesized mean. A left-tailed test indicates that the actual population mean is *less than* the hypothesized population mean. A two-tailed test is a combination of the right-tailed and left-tailed tests; it indicates that the actual population mean is *different than* the hypothesized mean. (This combines the two alternative hypotheses that the actual population mean is *greater than* the hypothesized mean and the actual population mean is *less than* the hypothesized mean.)

Right-tailed test

A *right-tailed test* is a test to determine if the actual value of the population mean is *greater than* the hypothesized value.

Suppose you're testing a hypothesis about the mean of a population, and you're interested in only strong evidence that the mean is *greater than* a specified value. In this case, you set up a right-tailed test. ("Right tail" refers to the largest values in a probability distribution.)

As an example of a right-tailed test, suppose that a department store wants to know whether the mean length of time its merchandise remains in inventory is 30 days. If the mean time is greater than 30 days, the store will overhaul its ordering procedures; if the mean is equal to or less than 30 days, the store will do nothing.

In this case, it's extremely important for the store to know whether the mean exceeds 30 days because a key decision depends on this information. The store doesn't want to spend time overhauling its procedures unless strong evidence shows that it's necessary; therefore, the most appropriate choice is a right-tailed test that shows the mean is greater than 30 days.

In general, you write the alternative hypothesis with a right-tailed test as

$$H_1: \mu > \mu_0$$

Here, H_1 represents the alternative hypothesis. In this example, you'd write the alternative hypothesis as $H_1: \mu > 30$.

Left-tailed test

A *left-tailed test* is a test to determine if the actual value of the population mean is *less than* the hypothesized value. ("Left tail" refers to the smallest values in a probability distribution.)

Suppose that you're testing a hypothesis about the mean of a population, and you're interested only in strong evidence that the mean is *less than* a specified value. In this case, you set up a left-tailed test.

For example, a pension fund wants to know whether any of its portfolio managers are earning an average return that falls short of the return to the Standard & Poor's 500 (S&P) stock index. (Assume this return is currently 8 percent.) If so, these managers won't receive the company's annual Christmas bonus.

The S&P is an index that represents the values of the 500 largest publicly traded U.S. stocks; it's often used as a benchmark for comparing the returns of portfolio managers.

In this situation, the fund is interested in knowing only which managers don't qualify for the Christmas bonus. As a result, the most appropriate choice for the alternative hypothesis a left-tailed test that shows the mean return is less than 8 percent.

In general, you write the alternative hypothesis for a left-tailed test as:

$$H_1: \mu < \mu_0$$

In this example, you'd write the alternative hypothesis as $H_1: \mu < 0.08$.

Two-tailed test

Building on the right-tailed test and the left-tailed test, consider the *two-tailed test,* which is used to determine if the actual value of the population mean is *different than* the hypothesized value; for example, greater than or less than. (A two-tailed test uses both the right tail and left tail of a probability distribution.)

Suppose you're testing a hypothesis about the mean of a population, and you need to know whether the mean is different from a specified value.

For example, a bottling company wants to be sure that the mean volume of its 1-liter bottles is actually 1 liter. Any value less than or more than this measurement could lead to significant problems. So the most appropriate choice is a two-tailed test that shows the mean volume is *not equal to* 1.

In general, you express the alternative hypothesis for a two-tailed test as

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

In this example, you'd write the alternative hypothesis as $H_1: \mu \neq 1$. This expression indicates that if the null hypothesis is false then either $H_1: \mu > 1$ or $H_1: \mu < 1$ will be accepted in its place, depending on the value of the test statistic relative to the critical values.

In this case, a two-tailed test was conducted due to the extreme importance of determining immediately if the mean content of the bottles is *either* less than 1 or greater than 1. If overfilled bottles are a problem, but not underfilled bottles, you would use a right-tailed test. If underfilled bottles are a problem, but not overfilled bottles, you would use a left-tailed test.

Choosing a level of significance

To test a hypothesis, you must specify a *level of significance* — the probability of rejecting the null hypothesis when it's actually true.

Rejecting the null hypothesis when it is actually true is is known as a *Type I error*. By contrast, a *Type II error* occurs when you fail to reject the null hypothesis when it's actually false. The level of significance of a hypothesis test equals the probability of committing a Type I error. A Type I error is sometimes known as a "false positive"; a Type II error is sometimes known as a "false negative."

In the process of testing a hypothesis, the following four results can take place. The two possible correct decisions are:

- ✔ Rejecting the null hypothesis when it's false
- ✔ Failing to reject the null hypothesis when it's true

The two possible incorrect decisions are:

- ✔ Rejecting the null hypothesis when it's true
- ✔ Failing to reject the null hypothesis when it's false

The probability of committing a Type I error is often designated with the letter α ("alpha"), and the probability of committing a Type II error is often designated with the letter β ("beta").

The larger is the probability of a Type I error that you choose for a hypothesis test, the smaller will be the probability of a Type II error, and vice versa. (One way to reduce both is to increase the sample size used for the hypothesis test.)

Note: The probabilities of Type I and Type II errors do *not* add up to 1; they are *not* complementary events. (Complementary events are discussed in Chapter 6.)

When you're conducting a hypothesis test, you choose the value of α to find the right balance between avoiding Type I and Type II errors. In some types of applications, avoiding Type I errors is critically important; in other cases, Type I errors may not be as serious.

In many hypothesis tests of a population value (such as the mean), the level of significance is often 0.01, 0.05, or 0.10, with 0.05 being most common.

Although both Type I and Type II errors represent serious mistakes, in some situations, one mistake is far more important to avoid than the other. For example, in a jury trial, the null hypothesis is "the defendant is innocent," which is assumed to be true unless strong contrary evidence suggests otherwise. The alternative hypothesis is that "the defendant is guilty."

In this situation, four outcomes can occur:

- ✔ The jury reaches a correct decision by acquitting an innocent defendant.
- ✔ The jury reaches a correct decision by convicting a guilty defendant.
- ✔ The jury commits a Type I error by wrongly convicting an innocent defendant. (In this situation, the null hypothesis of innocence has been incorrectly rejected.)
- ✔ The jury commits a Type II error by acquitting a guilty defendant (because the null hypothesis of innocence hasn't been rejected when it's actually false).

For a jury trial, avoiding a Type I error is far more important than avoiding a Type II error; as such, you set α equal to a very small value, which would

imply a much larger value for β. (α would never be set equal to 0 because that would ensure that no one is ever convicted!)

Because a Type I error in this case indicates that an innocent person has been convicted, and a Type II error indicates that a guilty person walks free, it's clearly imperative to avoid Type I errors even if it means more Type II errors.

TIP

Sir William Blackstone (1723–1780), the famous English judge and politician, once wrote that "It is better that ten guilty persons escape than that one inno-cent suffer." A statistician might rephrase this in slightly less elegant terms: "It is extremely important to avoid Type I errors in a jury trial."

Computing the test statistic

A *test statistic* is a numerical measure you construct to determine whether you should reject the null hypothesis. It shows how far the sample mean is from the hypothesized value of the population mean in terms of standard deviations. You calculate this value from a sample drawn from the underlying population.

For example, say you're testing a hypothesis about the mean age of the residents in a city. The city government wants to know whether the mean age is 40. You choose a sample of city residents, and you compute the mean age of the sample members. If the sample mean age is substantially different from 40, the null hypothesis will likely be rejected.

If you conduct a hypothesis test of the value of a single population mean, the form of the test statistic depends on two key details: the size of the sample chosen from the population and whether the population standard deviation is known.

REMEMBER

When you're testing hypotheses about the population mean, the cutoff point between a small sample and a large sample is 30. Any sample size less than 30 is small; a sample size of 30 or more is large.

When you're conducting hypothesis tests of the mean with a small sample, the test statistic follows the Student's t-distribution. With a large sample, the test statistic follows the standard normal distribution. (See Chapters 9 and 11 for discussions on the normal distribution and Student's t-distribution.)

For a small sample (less than 30), the test statistic is

$$t = \frac{\overline{X} - \mu_0}{s / \sqrt{n}}$$

In this formula,

- ✔ *t* indicates that this test statistic follows the Student's t-distribution
- ✔ $\bar{X}$ is the sample mean
- ✔ *s* is the sample standard deviation
- ✔ *n* is the sample size
- ✔ $\frac{s}{\sqrt{n}}$ is the *standard error* of the sample mean

For a large sample (30 or more), two possibilities exist:

- ✔ In the unlikely case that you don't know the population mean but do know the population standard deviation (σ), the appropriate test statistic is

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

- ✔ If the population standard deviation is unknown, the appropriate test statistic is

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

The letter *Z* indicates that these test statistics follow the standard normal distribution. The standard normal distribution (see Chapter 9) is the special case of the normal distribution with mean (μ) of 0 and a standard deviation (σ) of 1.

Comparing the critical value(s)

After you calculate a test statistic, you compare it to one or two *critical values,* depending on the alternative hypothesis, to determine whether you should reject the null hypothesis.

A critical value shows the number of standard deviations away from the mean of a distribution where a specified percentage of the distribution is above the critical value and the remainder of the distribution is below the critical value.

For example, based on the standard normal table (see Chapter 9), the probability that a standard normal random variable Z is less than 1.645 equals 0.95 or 95 percent. As a result, the probability that Z is greater than 1.645 is 0.05 or 5 percent. 1.645 is the critical value that divides the lower 95 percent of the distribution from the upper 5 percent of the distribution. Due to the symmetry of the standard normal distribution, –1.645 is the critical value that divides the lower 5 percent of the distribution from the upper 95 percent of the distribution.

This is shown in Figure 12-1. The shaded region is the upper 5 percent of the standard normal distribution, and the unshaded region is the lower 95 percent of the distribution.

The Standard Normal Distribution
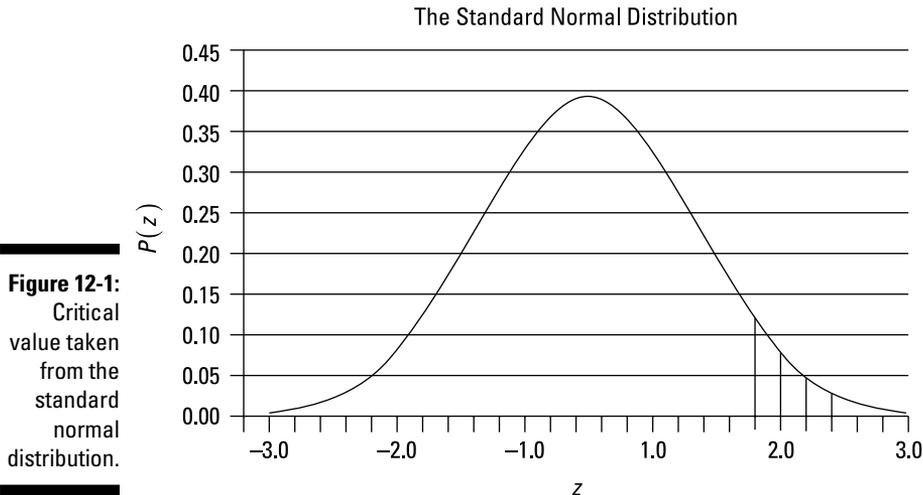
The appropriate critical value depends on whether you are conducting a right-tailed test, a left-tailed test, or a two-tailed test, as follows:

✔ A right-tailed test has one positive critical value.

✔ A left-tailed test has one negative critical value.

✔ A two-tailed test has two critical values, one positive and one negative.

The appropriate critical value also depends on the sample size and whether or not the population standard deviation is known. In the following sections, I show you how to determine the critical values for a hypothesis test for the value of the population mean, for both small and large samples.

Small sample (n < 30)

As I mention earlier, a small sample is less than 30. When you use a small sample to test a hypothesis about the population mean, you take the resulting critical value or values from the Student's t-distribution, as follows:

✔ Right-tailed test: critical value = t_α^{n-1}

✔ Left-tailed test: critical value = $-t_\alpha^{n-1}$

✔ Two-tailed test: critical value = $\pm t_{\alpha/2}^{n-1}$

Note: α is the level of significance, and *n* represents the sample size.

TIP

You draw these critical values from the Student's t-distribution with $n - 1$ *degrees of freedom* (*df*). (See the Student's t-table in Chapter 11.)

TECHNICAL STUFF

The number of degrees of freedom refers to the number of *independent* elements in a sample. When testing hypotheses about a single population mean, the degrees of freedom equals the sample size (n) minus 1. This is because the sample data is used to estimate one value: the sample mean. For any given set of $n - 1$ sample elements and the sample mean, the remaining sample element is a known value. For example, if a sample contains the elements 1, 2, 3 and 4, the sample mean equals $(1 + 2 + 3 + 4) / 4 = 2.5$. If the sample elements 1, 2, 3 are chosen, the sample mean of 2.5 implies that the missing element is 4. (In other words, the sample mean of 2.5 indicates that the sum of the sample elements is 10. Because the sample elements 1, 2, and 3 sum to 6, the remaining element must be 4.) Therefore, one sample element is uniquely determined, while the remaining $n - 1$ sample elements are completely variable. As a result, the degrees of freedom equal $n - 1$.

The number of degrees of freedom used with the t-distribution depends on the particular application. For testing hypotheses about the population mean, the appropriate number of degrees of freedom is one less than the sample size (that is, $n - 1$). (See Chapter 11 for details on the Student's t-distribution.)

The critical value or values are used to locate the areas under the curve of a distribution that are too extreme to be consistent with the null hypothesis. For a right-tailed test, these are the large positive values, which are collectively known as the right tail of the distribution. For a left-tailed test, these are the large negative values, which are collectively known as the left tail of the distribution. In either case, the area in the tail equals the level of significance of the hypothesis test. For a two-tailed test, the value of the level of significance (α) is split in half; the area in the right tail equals $\alpha/2$, and the area in left tail equals $\alpha/2$, for a total of α.

Right-tailed test with a small sample

As an example of a right-tailed test, suppose the level of significance is 0.05 and the sample size is 10; then you get a single positive critical value:

$$t_\alpha^{n-1} = t_{0.05}^9$$

Refer to Table 12-1 to find the intersection of the row representing 9 degrees of freedom and the column headed $t_{0.05}$.

| Table 12-1	The Student's t-distribution				
Degrees of Freedom (df)	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947

The critical value is 1.833, or $t_{\alpha}^{n-1} = t_{0.05}^{9} = 1.833$, as shown in Figure 12-2.

The shaded region in the right tail represents the *rejection region*; if the test statistic falls in this area, the null hypothesis will be rejected.
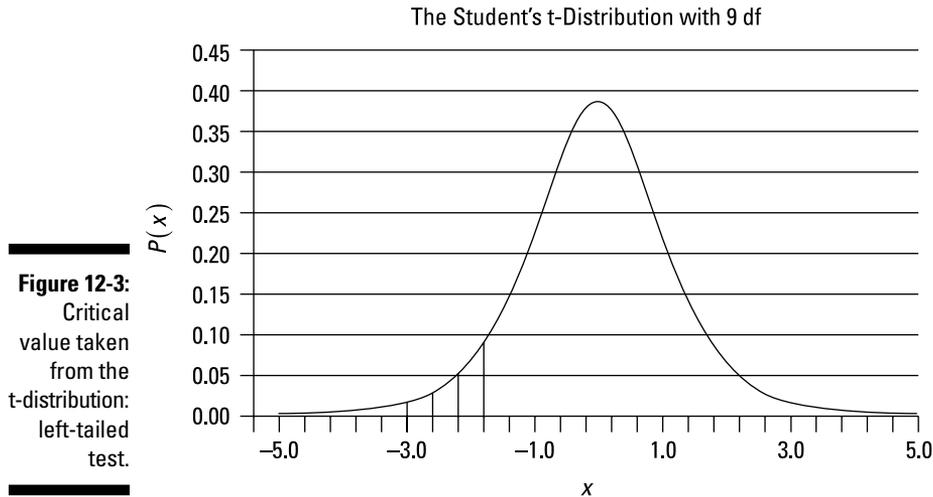
Left-tailed test with a small sample

As an example of a left-tailed test, suppose the level of significance is 0.05 and the sample size is 10; then you get a single negative critical value:

$$-t_{\alpha}^{n-1} = -t_{0.05}^{9}$$

The Student's t-Distribution with 9 df
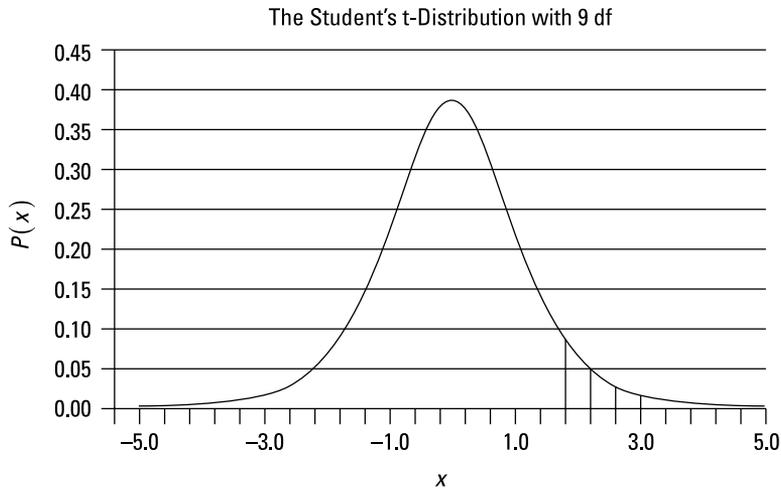
Figure 12-2: Critical value taken from the t-distribution: right-tailed test.

You get this number from the t-table (Table 12-1) at the intersection of the row representing 9 degrees of freedom and the $t_{0.05}$ column; the critical value is –1.833, or $-t_\alpha^{n-1} = -t_{0.05}^9 = -1.833$, as shown in Figure 12-3.

The Student's t-Distribution with 9 df

Figure 12-3:
Critical value taken from the t-distribution: left-tailed test.

The shaded region in the left tail represents the *rejection region*; if the test statistic falls in this area, the null hypothesis will be rejected.

Two-tailed test with a small sample

As an example of a two-tailed test, suppose the level of significance is 0.05 and the sample size is 10; then you get a positive and a negative critical value:
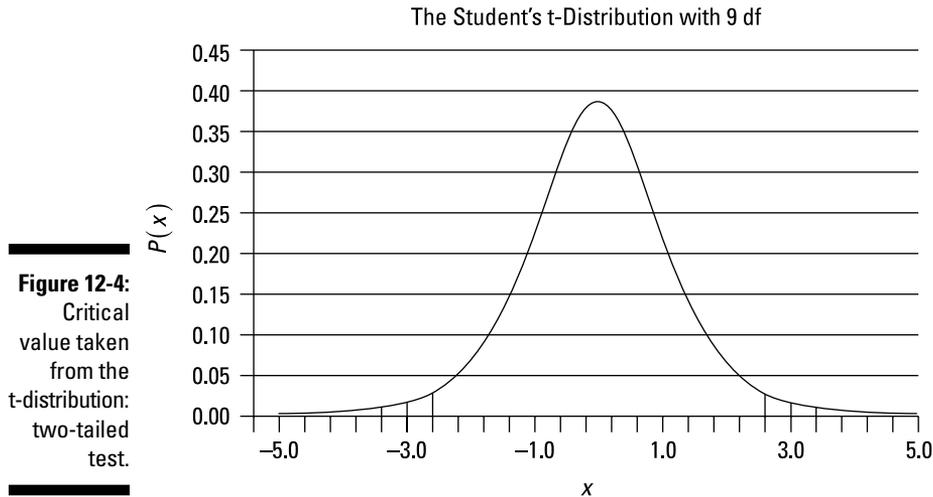
$$\pm t_{\alpha/2}^{n-1} = \pm t_{0.025}^9$$

You can find the value of the positive critical value $\pm t_{0.025}^9$ directly from Table 12-1.

In this case, you find the positive critical value $t_{0.025}^9$ at the intersection of the row representing 9 degrees in the *Degrees of Freedom (df)* column and the $t_{0.025}$ column. The positive critical value is 2.262; therefore, the negative critical value is –2.262. You represent these two values like so (as Figure 12-4 illustrates):

$$\pm t_{\alpha/2}^{n-1} = \pm 2.262$$

The shaded region in the two tails represents the *rejection region*; if the test statistic falls in either tail, the null hypothesis will be rejected.

The Student's t-Distribution with 9 df

Figure 12-4:
Critical value taken from the t-distribution: two-tailed test.

Large sample (n ≥ 30)

A large sample has a size greater than or equal to 30. When you use a large sample to test a hypothesis about the population mean, you take the resulting critical value or values from the standard normal distribution, as follows:

- ✔ Right-tailed test: critical value = Z_α
- ✔ Left-tailed test: critical value = $-Z_\alpha$
- ✔ Two-tailed test: critical values = $\pm Z_{\alpha/2}$

Because you draw these critical values from the standard normal distribution, you don't have to calculate degrees of freedom. Unlike the Student's t-distribution, the standard normal distribution isn't based on degrees of freedom. I walk you through how to find these critical values in the following sections.

For hypothesis testing applications, the critical values listed in Table 12-2 are used frequently; you may want to memorize them.

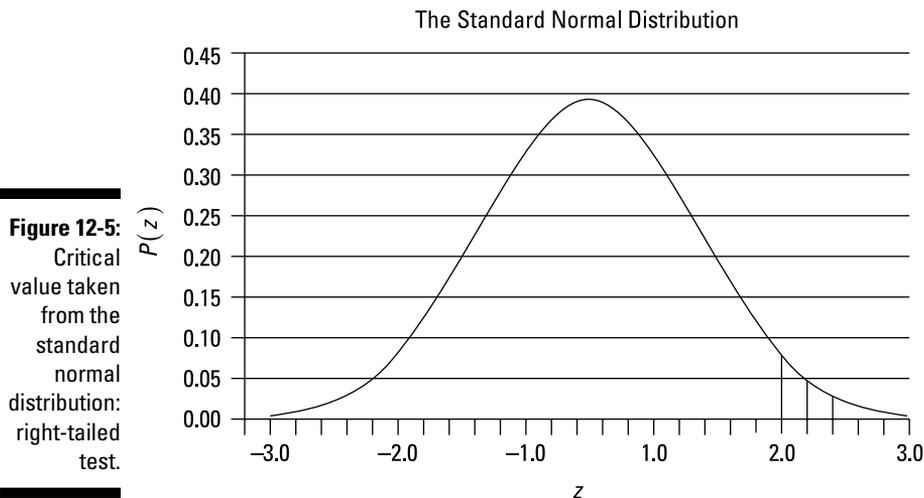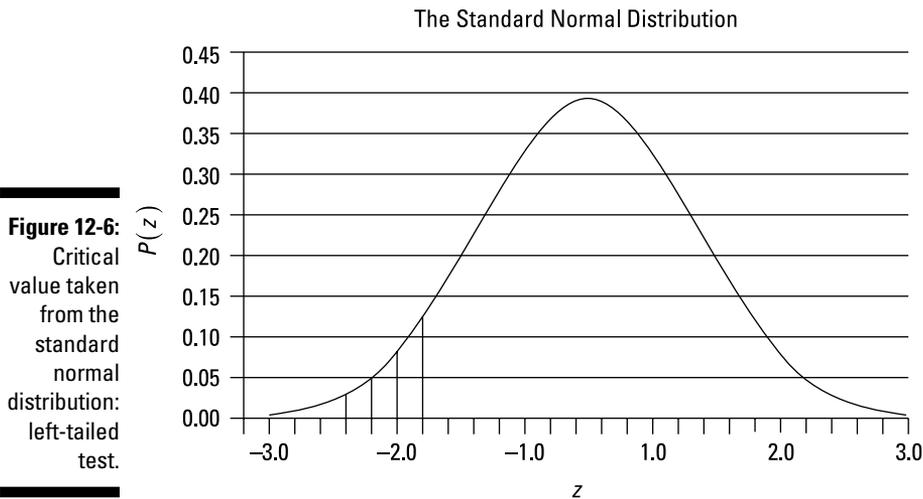| Table 12-2	Common Critical Values of the Standard Normal Distribution		
α	*Right-Tailed Test*	*Left-Tailed Test*	*Two-Tailed Test*
0.01	2.328	−2.328	±2.576
0.05	1.645	−1.645	±1.960
0.10	1.282	−1.282	±1.645

Right-tailed test with a large sample

A *right-tailed* hypothesis test of the population mean with a level of significance of 0.05 has a single positive critical value: $Z_\alpha = Z_{0.05}$. You find the value by checking the body of Table 12-3 for a probability of $1 - \alpha$, which is 0.9500.

Table 12-3	Standard Normal Table — Positive Values			
Z	*0.04*	*0.05*	*0.06*	*0.07*
1.5	0.9382	0.9394	0.9406	0.9418
1.6	0.9495	0.9505	0.9515	0.9525
1.7	0.9591	0.9599	0.9608	0.9616
1.8	0.9671	0.9678	0.9686	0.9693
1.9	0.9738	0.9744	0.9750	0.9756
2.0	0.9793	0.9798	0.9803	0.9808

Unfortunately, this exact value isn't in the table. The two closest values are 0.9495 and 0.9505, which you can find at the intersections of row *1.6* under the Z column and the *0.04* and *0.05* columns.. The critical value is, therefore, halfway between 1.64 and 1.65; average it out to get 1.645, or $Z_\alpha = 1.645$, and see Figure 12-5 for a graphical depiction.

A left-tailed hypothesis test with a level of significance of 0.05 has a single negative critical value $-Z_\alpha = -Z_{0.05}$, or simply -1.645: $-Z_\alpha = -1.645$. Figure 12-6 represents this critical value graphically.

Figure 12-5: Critical value taken from the standard normal distribution: right-tailed test.

The Standard Normal Distribution

Two-tailed test with a large sample

For a two-tailed hypothesis test of the population mean with a level of significance of 0.05, the two critical values are $\pm Z_{\alpha/2} = \pm Z_{0.025}$.

You can find the positive critical value in a standard normal table, like Table 12-3.

Finding critical values in a standard normal table is more complicated than finding critical values in a t-table. The body of the standard normal table contains probabilities, unlike in the t-table where the probabilities are contained in the column headings.

In this example, you find the positive critical value $Z_{\alpha/2} = Z_{0.025}$ by checking the body of the table for a probability of
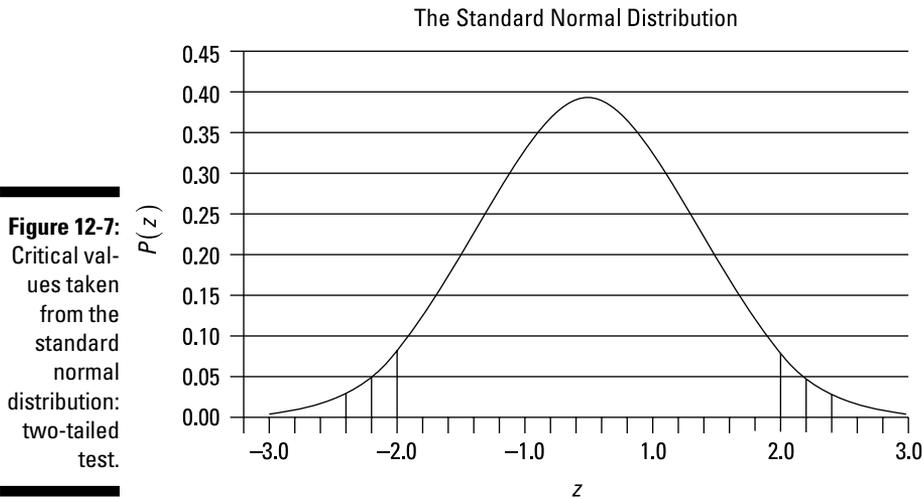
$$(1 - \alpha / 2) = (1 - 0.05 / 2) = (1 - 0.025) = 0.9750.$$

In other words, the positive critical value represents the number of standard deviations above the mean at which

✔ 2.5 percent of the area under the standard normal curve is to the right of this point.

✔ 97.5 percent of the area under the standard normal curve is to the left of this point.

Because the standard normal table shows areas to the left of specified values, you can find the positive critical value by locating the probability 0.9750, not 0.0250, in the body of the table (Table 12-3). You find this probability by

following the row 1.9 under the *Z* column to the *0.06* column. Therefore, the critical value $Z_{\alpha/2} = Z_{0.025} = 1.96$. The corresponding negative critical value is –1.96. You can write these critical values as $\pm Z_{\alpha/2} = \pm 1.96$. Figure 12-7 shows these values graphically.

The Standard Normal Distribution

Figure 12-7: Critical values taken from the standard normal distribution: two-tailed test.

Using the decision rule

You make the decision to reject the null hypothesis by looking at the relationship between the test statistic and the critical value(s), as follows:

- ✔ **Right-tailed test:** If the test statistic is *greater than* the critical value, reject the null hypothesis H_0: $\mu = \mu_0$ in favor of the alternative hypothesis H_1: $\mu > \mu_0$; otherwise, don't reject the null hypothesis. There is insufficient evidence to show that the null hypothesis is false.

- ✔ **Left-tailed test:** If the test statistic is *less than* the critical value, reject the null hypothesis H_0: $\mu = \mu_0$ in favor of the alternative hypothesis H_1: $\mu < \mu_0$; otherwise, don't reject the null hypothesis. There is insufficient evidence to show that the null hypothesis is false.

- ✔ **Two-tailed test;** If the test statistic is *less than* the negative critical value, reject the null hypothesis H_0: $\mu = \mu_0$ in favor of the alternative hypothesis H_1: $\mu < \mu_0$ If the test statistic is *greater than* the positive critical value, reject the null hypothesis H_0: $\mu = \mu_0$ in favor of the alternative hypothesis H_1: $\mu > \mu_0$. Otherwise, don't reject the null hypothesis. There is insufficient evidence to show that the null hypothesis is false.

As an example, suppose that the government of a small country is interested in studying the characteristics of household incomes in the country. The government wants to know whether the mean household income is greater than $25,000 per year. If so, the government will propose new types of taxes; otherwise, no new taxes will occur. The appropriate steps for testing the null hypothesis that the mean household income equals $25,000 at the 5 percent level of significance are given as follows:.

The null and alternative hypotheses are

H_0: μ = 25,000

H_1: μ > 25,000

In this example, the government uses a right-tailed test because it's looking for strong evidence that the mean household incomes are *greater than* $25,000 per year. If true, the government will take an important action.

Assume that the level of significance is 0.05. The government's chief statistician selects a sample of 100 households and computes the sample mean household income to be $27,200 per year. The population standard deviation is unknown; instead, the government statistician computes the sample standard deviation, and it turns out to be $8,400.

Because the government statistician chose a large sample (greater than or equal to 30), he uses the standard normal distribution to test this hypothesis. Because the population standard deviation is unknown, the appropriate test statistic is

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

The value of the test statistic is, therefore,

$$
\begin{aligned}
Z &= \frac{\bar{X} - \mu_0}{s / \sqrt{n}} \\
&= \frac{27,200 - 25,000}{8,400 / \sqrt{100}} \\
&= \frac{2,200}{840} \\
&= 2.62
\end{aligned}
$$

The critical value is $Z\alpha$ = $Z_{0.05}$ = 1.645 (see Table 12-3).

Because the test statistic of 2.62 exceeds the critical value of 1.645, the government statistician rejects the null hypothesis in favor of the alternative

hypothesis that the population mean exceeds $25,000. As a result, there are new taxes. Figure 12-8 shows this result graphically.

As another example, say the same government wants to study the average crop yields of its wheat farmers. The government wants to know whether the mean yield is equal to 10,000 bushels per year.

If the mean yield is below 10,000, the government will provide cash assistance to the farmers. If the mean yield is above 10,000, the government will export some of the surplus wheat to foreign countries. The government's chief statistician can test the null hypothesis that the mean crop yield equals 10,000 bushels. (Assume that he chooses a 5 percent level of significance.)

The null and alternative hypotheses are

$H_0: \mu = 10,000$

$H_1: \mu \neq 10,000$

The Standard Normal Distribution
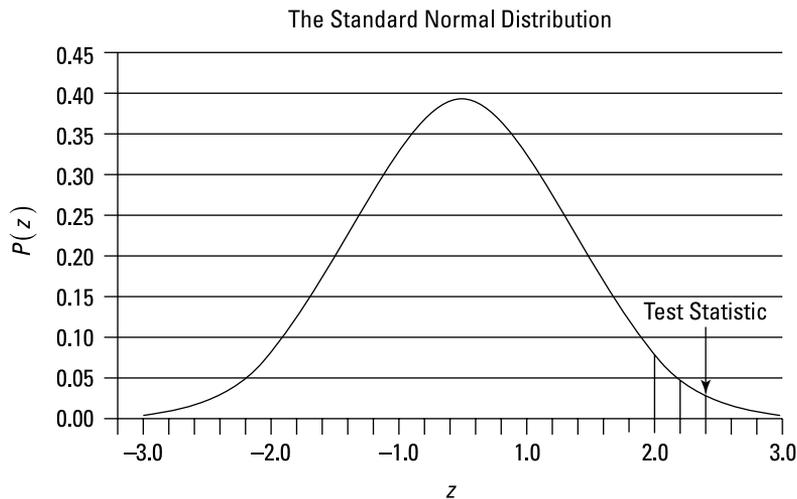
Figure 12-8:
Standard normal distribution: The null hypothesis is rejected.

This example requires a two-tailed test, because the government is looking for strong evidence that mean crop yields are either *less than* or *greater than* 10,000 bushels per year. The government will undertake an important action in either case.

Assume that the level of significance is 0.05. The government statistician selects a sample of eight farms, and estimates the sample mean and standard deviation. The mean crop yield turns out to be 9,400 bushels. The sample standard deviation is 420 bushels.

Due to the small number of farms in the country, the government statistician chose a small sample (less than 30); therefore, the hypothesis test is based on the Student's t-distribution. The appropriate test statistic is, therefore,

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$$

If the population standard deviation (σ) is known, you use the standard normal distribution, regardless of the sample size. (See Chapter 11 for details.)

The value of the test statistic is

$$t = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$$
$$= \frac{10,200 - 10,000}{420 / \sqrt{8}}$$
$$= \frac{200}{148.49}$$
$$= 1.35$$

The critical values are $\pm t_{\alpha/2}^{n-1}$.

With a sample size of $n = 8$, the appropriate number of degrees of freedom is $n - 1 = 7$. With a level of significance of 0.05, the value of $\alpha/2$ is 0.025. Therefore, you find the critical values in the t-table (Table 12-1) as follows:

$$\pm t_{\alpha/2}^{n-1} = \pm t_{0.025}^{7} = \pm 2.365$$

With a two-tailed test, the decision rule is to

✔ Reject the null hypothesis H_0: $\mu = \mu_0$ in favor of the alternative hypothesis H_1: $\mu < \mu_0$ if the test statistic is less than the negative critical value (–2.365).

✔ Reject the null hypothesis H_0: $\mu = \mu_0$ in favor of the alternative hypothesis H_1: $\mu > \mu_0$ if the test statistic is greater than the positive critical value (2.365).

✔ Not reject the null hypothesis if the test statistic is between the negative and positive critical values (–2.365 and 2.365).

Because the test statistic is 1.35, it's *greater than* the negative critical value of –2.365, and *less than* the positive critical value of 2.365. In other words, the test statistic is *not* in the rejection region, as shown in Figure 12-9.

The Student's t-Distribution with 9 df
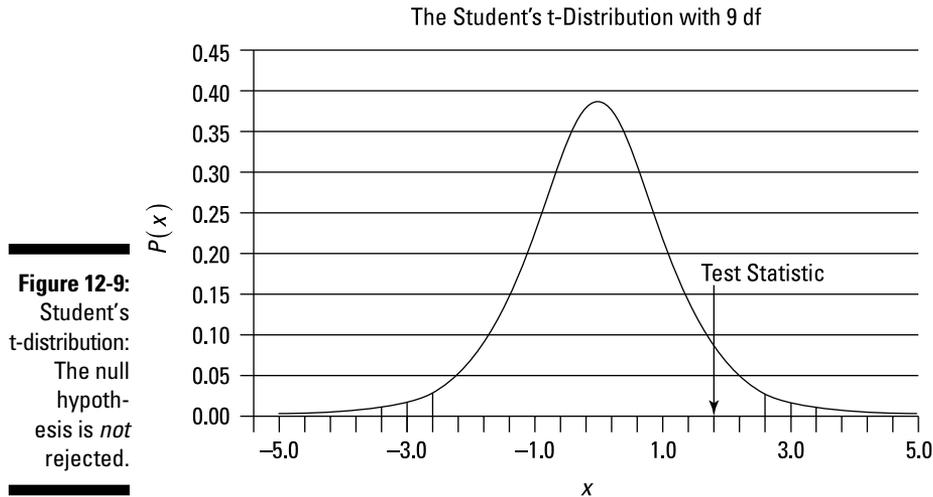
In this example, you do *not* reject the null hypothesis H_0: $\mu = \mu_0$. As a result, the government takes no action.

Testing Hypotheses About Two Population Means

In addition to testing claims about the mean of a population, hypothesis testing can be used to compare the equality of two different population means. For example, you can use hypothesis testing on the following statements to determine whether they're true:

✔ The mean price of gasoline per gallon is equal in New York and New Jersey.

✔ The average life expectancy of men is the same in the United States and Canada.

✔ The mean annual rainfall is equal in Washington and Oregon.

✔ The length of the average flight delay is the same at Kennedy Airport and LaGuardia Airport.

The basic procedure for testing hypotheses about two population means is similar to the procedure for a single population mean (which I discuss in the

section "Applying the Key Steps in Hypothesis Testing for a Single Population Mean"). The most important differences are the form of the test statistics you use for two population means and the calculation of the critical values. I outline the differences in the following sections.

Writing the null hypothesis for two population means

To test the equality of two population means, you write the null hypothesis as

$$H_0: \mu_1 = \mu_2$$

In this formula, H_0 is the null hypothesis, μ_1 is the mean of population 1, and μ_2 is the mean of population 2.

Note that when testing hypotheses about two population means, one population is arbitrarily chosen to be "population 1" and the other becomes "population 2."

Defining the alternative hypotheses for two population means

Just as you have an alternative hypothesis for testing a single population mean, when you test two population means, you also need an alternative hypothesis. If the null hypothesis is rejected, you must specify what other result will be accepted instead. This is the role of the alternative hypothesis.

The alternative hypothesis can take one of three forms:

- ✔ **Right-tailed test:** $H_1: \mu_1 > \mu_2$
- ✔ **Left-tailed test:** $H_1: \mu_1 < \mu_2$
- ✔ **Two-tailed test:** $H_1: \mu_1 \neq \mu_2$

A right-tailed test is used to indicate if the mean of population 1 is *greater than* the mean of population 2. Similarly, a left-tailed test is used to show if the mean if population 1 is *less than* the mean of population 2. A two-tailed test is used to show if the mean of population 1 is *different than* the mean of population 2.

Determining the test statistics for two population means

When you're testing hypotheses about two population means, you can choose from several test statistics. The choice depends on:

✔ whether the samples drawn from the two populations are independent of each other

✔ whether the variances of the two populations are equal

✔ whether the samples chosen from the two populations are large (at least 30) or small (less than 30)

Samples are *independent* if they're not related to each other. For example, samples of GPAs at two universities are independent samples, because none of the students in these samples attend both universities.

If you choose independent samples from two populations, you choose the test statistic and critical values based on the following questions:

✔ Are the variances of the two populations equal?

✔ If the variances are unequal, are the sample sizes large (at least 30)?

If the samples are *dependent,* the choice for test statistics and critical values are different. For example, suppose that medical researchers are conducting a study to determine whether a new cholesterol drug is effective in reducing LDL (bad cholesterol) in patients. If you chose a sample of LDL readings chosen from a set of patients prior to taking the drug and a sample of LDL readings among the same patients after taking the drug, these two samples would be closely related and, therefore, *dependent.* This type of hypothesis test requires a different procedure for constructing the test statistic and critical values than for independent samples. I explore using independent and dependent, or *paired,* samples in the following sections.

Using independent samples

When using independent samples, you first have to decide whether the populations being tested have equal variances (or if you have reason to believe that they're equal).

With equal population variances, the test statistic requires the calculation of a pooled variance — this is the variance that the two populations have in common. You use the Student's t-distribution to find the test statistic and critical values.

With unequal population variances, there are two possibilities.

- ✔ You use the standard normal distribution for the test statistic and critical values if the samples are large (at least 30).

- ✔ You use the t-distribution if at least one of the samples is small (less than 30).

The choice of distribution for the hypothesis test based on independent samples is summarized in Table 12-4:

Table 12-4	Choice of Probability Distribution for Independent Samples
Condition	*Distribution*
Equal variances	Student's t
Unequal variances: large samples	Standard Normal (Z)
Unequal variances: at least one small sample	Student's t

Equal population variances

If the variances of two populations are equal (or are assumed to be equal) the appropriate test statistic is based on the Student's t-distribution:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Here's what each term means:

- ✔ $\bar{x}_1$ is the mean of the sample chosen from population 1.

- ✔ $\bar{x}_2$ is the mean of the sample chosen from population 2.

- ✔ μ_1 is the mean of population 1.

- ✔ μ_2 is the mean of population 2.

- ✔ $(\mu_1 - \mu_2)_0$ is the hypothesized difference between populations 1 and 2, which is 0 when the population means are hypothesized to be equal.

- ✔ n_1 is the size of the sample chosen from population 1.

- ✔ n_2 is the size of the sample chosen from population 2.

- ✔ s^2_1 is the variance of the sample chosen from population 1.

- s^2_2 is the variance of the sample chosen from population 2.
- s^2_p is the estimated common pooled variance of the two populations, or in mathematical terms:

$$s^2_p = \frac{(n_1 - 1)s^2_1 + (n_2 - 1)s^2_2}{n_1 + n_2 - 2}$$

If you are conducting a hypothesis test of two population means with equal population variances, you take the critical values from the Student's t-distribution with $n_1 + n_2 - 2$ degrees of freedom, which gives you the following critical values:

- Right-tailed test: $t_\alpha^{n_1 + n_2 - 2}$
- Left-tailed test: $-t_\alpha^{n_1 + n_2 - 2}$
- Two-tailed test: $\pm t_{\alpha/2}^{n_1 + n_2 - 2}$

As an example, say a marketing company is interested in determining whether men and women are equally likely to buy a new product. The company randomly chooses samples of men and women and asks them to assign a numerical value to their likelihood of buying the product (1 being the least likely, and 10 being the most likely).

Based on past experience, the population variances are assumed to be equal. The first step is to assign one group to be the first population ("population 1") and the other group to be the second population ("population 2"). The company designates men as population 1 and women as population 2.

The next step is to choose samples from both populations. (The sizes of these samples do not have to be equal.) Suppose that the company chooses samples of 21 men and 21 women. These samples are used to compute the sample mean and sample standard deviation for both men and women. (Sample means are covered in Chapter 3; sample standard deviations are covered in Chapter 4.)

Assume that the sample mean score of the men is 7.2; the sample mean score of the women is 6.7. Also assume that the sample standard deviation of the men is 0.4, and the sample standard deviation of the women is 0.3. With this data in place, the null hypothesis that the population mean scores are equal is tested by the marketing company at the 5 percent level of significance.

You can summarize the sample data like so:

$\bar{x}_1 = 7.2$ and $\bar{x}_2 = 6.7$

$s_1 = 0.4$ and $s_2 = 0.3$

$n_1 = 21$ and $n_2 = 21$

The null hypothesis is $H_0: \mu_1 = \mu_2.$ The alternative hypothesis is $H_1: \mu_1 \neq \mu_2$.

To compute the test statistic, you first calculate the pooled variance:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$
$$= \frac{(21 - 1)(0.4)^2 + (21 - 1)(0.3)^2}{21 + 21 - 2}$$
$$= \frac{32 + 18}{40}$$
$$= 0.125$$

You then substitute this result into the test statistic formula:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$
$$= \frac{(7.2 - 6.7) - 0}{\sqrt{0.125\left(\frac{1}{21} + \frac{1}{21}\right)}}$$
$$= \frac{0.5}{0.109}$$
$$= 4.587$$

You can find the appropriate critical values from Table 12-5 (which is an excerpt from the Student's t-table, covered in Chapter 11). These are found as follows. The top row of the Student's t-table lists different values of tα, where the right tail of the Student's t-distribution has a probability (area) equal to α ("alpha").

In this case, alpha (α) is 0.05; using a tail area of 0.025 (α/2) and 40 degrees of freedom, you find that the critical values are:

$$\pm t_{\alpha/2}^{n_1 + n_2 - 2} = \pm t_{0.05/2}^{21 + 21 - 2} = \pm t_{0.025}^{40} = \pm 2.021$$

Table 12-5   **The Student's t-Distribution with a Large Number of Degrees of Freedom**

Degrees of Freedom (df)	$t_{0.10}$	$t_{0.05}$	$t_{0.025}$	$t_{0.01}$	$t_{0.005}$
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660

TIP

Note that with a large number of degrees of freedom, the Student's t-distribution closely resembles the standard normal distribution (see Chapter 9 for more discussion of the normal distribution). For example, if you perform a two-tailed hypothesis test with $\alpha = 0.05$, the critical values drawn from the standard normal distribution are ± 1.96, compared with ± 2.000 for the Student's t-distribution with 60 degrees of freedom.

Because the test statistic (4.587) exceeds the positive critical value (2.021), the null hypothesis H_0: $\mu_1 = \mu_2$ is rejected.

With a two-tailed test, there are actually two alternatives available to the null hypothesis: H_1: $\mu_1 > \mu_2$ (that is, the mean rating among men is greater than the mean rating among women) or H_1: $\mu_1 < \mu_2$ (that is, the mean rating among men is less than the mean rating among women). In this case, the test statistic is large and positive, which suggests that the mean for men is greater than the mean for women. A large and positive test statistic indicates that the sample mean for men is significantly greater than the sample mean for women. In other words, men are more likely to buy the new product than women.

Unequal population variances: At least one sample is small

If the variances of two populations *aren't* equal (or you don't have any reason to believe that they're equal) and at least one sample is small (less than 30), the appropriate test statistic is

$$t = \frac{\left(\bar{x}_1 - \bar{x}_2 \right) - \left(\mu_1 - \mu_2 \right)_0}{\sqrt{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2} \right)}}$$

In this case, you get the critical values from the t-distribution with degrees of freedom equal to

$$df = \frac{\left[\left(s_1^2 / n_1 \right) + \left(s_2^2 / n_2 \right) \right]^2}{\left[\dfrac{\left(s_1^2 / n_1 \right)^2}{\left(n_1 - 1 \right)} + \dfrac{\left(s_2^2 / n_2 \right)^2}{\left(n_2 - 1 \right)} \right]}$$

TIP

This value isn't necessarily equal to a whole number; if the resulting value contains a fractional part, you must round it to the next closest whole number.

For example, assume that Major League Baseball (MLB) is interested in determining whether the mean number of runs scored per game is higher in the American League (AL) than in the National League (NL). The population variances are assumed to be unequal.

The first step is to assign one group to represent the first population ("population 1") and the other group to represent the second population ("population 2"). MLB designates the American League as population 1 and the National League as population 2.

The next step is to choose samples from both populations. Suppose that MLB choose a sample of 10 American League and 12 National League teams. The results are used to compute the sample mean and sample standard deviation for both leagues. Assume that the sample mean for runs scored among the AL games is 8.1, whereas the sample mean for the NL games is 7.9. The sample standard deviation is 0.5 for AL games and 0.3 for NL games.

MLB tests the null hypothesis that the population mean scores are equal at the 5 percent level of significance.

Here's a summary of the sample data:

$\bar{x}_1 = 8.1$ and $\bar{x}_2 = 7.9$

$s_1 = 0.5$ and $s_2 = 0.3$

$n_1 = 10$ and $n_2 = 12$

The null hypothesis is

$H_0: \mu_1 = \mu_2$

Because MLB is interested in determining whether the mean number of runs scored per game is higher in the American League than in the National League, you use a right-tailed test. The alternative hypothesis is $H_1: \mu_1 > \mu_2$.

In other words, the test is designed to find strong evidence that the mean of population 1 is *greater than* the mean of population 2. You then solve the test statistic as follows:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)}}$$

$$= \frac{(8.1 - 7.9) - 0}{\sqrt{\left(\dfrac{0.5^2}{10}\right) + \left(\dfrac{0.3^2}{12}\right)}}$$

$$= \frac{0.2}{0.1803}$$

$$= 1.109$$

And you find the degrees of freedom like so:

$$df = \frac{\left[\left[\left(s_1^2 / n_1\right) + \left(s_2^2 / n_2\right)\right]\right]^2}{\dfrac{\left(s_1^2 / n_1\right)^2}{\left(n_1 - 1\right)} + \dfrac{\left(s_2^2 / n_2\right)^2}{\left(n_2 - 1\right)}}$$

$$= \frac{\left[\left(0.5^2 / 10\right) + \left(0.3^2 / 1\right)\right]^2}{\dfrac{\left(0.5^2 / 10\right)^2}{\left(10 - 1\right)} + \dfrac{\left(0.3^2 / 12\right)^2}{\left(12 - 1\right)}}$$

$$= \frac{\left[0.025 + 0.0075\right]^2}{0.00006944 + 0.0000051136}$$

$$= \frac{0.00105625}{0.00007455536}$$

$$= 14.167$$

You round down the value of 14.167 to 14 because the degrees of freedom must be a whole number (or *integer*). With 14 degrees of freedom and a 5 percent level of significance, the critical value is $t_{0.05}^{14} = 1.761$.

This result is obtained from Table 12-1 by finding the column headed $t_{0.05}$ and the row corresponding to 14 degrees of freedom.

Because the test statistic (1.109) is below the critical value (1.761), the null hypothesis that H_0: $\mu_1 = \mu_2$ fails to be rejected. There's insufficient evidence to conclude that more runs are scored during American League games than National League games.

Unequal population variances: Both sample sizes are large

If the variances of two populations *aren't* equal, and the size of both samples is 30 or greater, the appropriate test statistic is

$$Z = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)_0}{\sqrt{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)}}$$

This test statistic is based on the standard normal distribution.

As an example, say that a restaurant chain is interested in finding out whether the average sale per customer is the same in its domestic and foreign restaurants. The population variances are assumed to be unequal. The restaurant chooses a random sample of 40 domestic and 50 foreign restaurants, designating domestic restaurants as population 1 and foreign restaurants as population 2.

The sample mean spending per customer is $5.14 in the domestic market and $4.59 in the foreign market. The sample standard deviation is $0.54 in the domestic market and $0.38 in the foreign market. The null hypothesis that the population mean spending is equal in the two markets is tested at the 5 percent level of significance.

Here's a summary of this data:

$$\bar{x}_1 = 5.14 \text{ and } \bar{x}_2 = 4.59$$
$$s_1 = 0.54 \text{ and } s_2 = 0.38$$
$$n_1 = 40 \text{ and } n_2 = 50$$

The null hypothesis is $H_0: \mu_1 = \mu_2$.

Because example requires a two-tailed test, the alternative hypothesis is $H_1: \mu_1 \neq \mu_2$.

You find the test statistic like so:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

$$= \frac{(5.14 - 4.59) - 0}{\sqrt{\left(\frac{0.54^2}{40}\right) + \left(\frac{0.38^2}{50}\right)}}$$

$$= \frac{0.055}{\sqrt{0.00729 + 0.002888}}$$

$$= 5.452$$

The critical values are then $\pm Z_{\alpha/2} = \pm Z_{0.025} = \pm 1.96$ (see Table 12-3).

Because the test statistic (5.452) is greater than the positive critical value (1.96), the null hypothesis $H_0: \mu_1 = \mu_2$ is rejected.

Because this is a two-tailed test, you may reject the null hypothesis in favor of the alternative $H_1: \mu_1 > \mu_2$ (that is, mean spending per customer is greater in the domestic market than the foreign market) or $H_1: \mu_1 < \mu_2$ (that is, mean spending per customer is lower in the domestic market than the foreign market.) Because the test statistic is large and positive, the alternative $H_1: \mu_1 > \mu_2$ is chosen. In other words, mean spending per customer in the domestic market is greater than in the foreign market.

Working with dependent samples

You can choose samples to compare the mean of a population before and after a given event. In this case, the samples aren't independent; instead, they're dependent, or *paired* samples. Examples of paired samples include:

> The cholesterol readings of randomly selected patients before taking a new drug and the cholesterol readings of the same patients after taking the drug

> The grade point averages of randomly chosen students before being tutored and the grade point averages of the same students after being tutored

> The productivity of a randomly selected group of employees prior to taking a new training course and the productivity of the same employees after taking the training course

With paired samples, the null hypothesis is based on the *differences* between the sample elements. Instead of stating that the population means are equal, the null hypothesis is that the difference between the population means equals 0.

When you're testing hypotheses about the equality of two population means with paired samples, you write the null hypothesis as

$$H_0: \mu_d = 0$$

where μ_d represents the mean difference between the two populations; it equals $\mu_d = \mu_1 - \mu_2$.

The three possible alternative hypotheses are

> ✔ **Right-tailed test:** $H_1: \mu_d > 0$. In this case, the alternative hypothesis is that the mean of population 1 is *greater than* the mean of population 2.

> ✔ **Left-tailed test:** $H_1: \mu_d < 0$. In this case, the alternative hypothesis is that the mean of population 1 is *less than* the mean of population 2.

> ✔ **Two-tailed test:** $H_1: \mu_d \neq 0$. In this case, the alternative hypothesis is that the means of populations 1 and 2 *aren't* equal.

For paired samples, the test statistic is always based on the Student's t-distribution:

$$t = \frac{\overline{d}}{s_d / \sqrt{n}}$$

Here, $\overline{d}$ is the average difference between paired samples, and s_d is the standard deviation of the sample differences.

Compute the mean of the differences like this:

$$\bar{d} = \frac{\sum\limits_{i=1}^{n} d_i}{n}$$

This formula indicates that you calculate the average difference between the paired samples by adding up all the individual differences and then dividing by the total number of elements in each sample..

Compute the standard deviation of the differences like this:

$$s_d = \sqrt{\frac{\sum\limits_{i=1}^{n}\left(d_i - \bar{d}\right)^2}{n-1}}$$

Note that this is the sample standard deviation formula (covered in Chapter 4.)

With paired samples, you take the critical values from the Student's t-distribution with $n - 1$ degrees of freedom, where n is the number of paired observations.

For example, say a pharmaceutical company is testing a new diet pill to determine whether taking it leads to weight loss. The company chooses a sample of eight volunteers. Table 12-6 shows the mean weights of these individuals before and after using the diet pill, along with the necessary calculations for computing the sample standard deviation:

Table 12-6   Paired Differences Between Two Samples

Subject	Weight Prior to Taking Diet Pill (x_1)	Weight After Taking Diet Pill (x_2)	$d_i = x_1 - x_2$	$\left(d_i - \bar{d}\right)^2$
1	192	190	2	1.891
2	189	185	4	0.391
3	204	199	5	2.641
4	177	177	0	11.391
5	156	151	5	2.641
6	228	224	4	0.391
7	244	239	5	2.641
8	201	199	2	1.891
		Sum	27	23.875
		Mean	3.375	

The company tests the null hypothesis that weight remains unchanged after taking the diet pill at the 5 percent level of significance. The null hypothesis is H_0: $\mu_1 = \mu_2$.

Because the pharmaceutical company is looking for strong evidence that the weights of the volunteers *dropped* after taking the pill, it uses a right-tailed test. (In other words, the mean weights of the volunteers before taking the pill is *greater than* the mean weights of the volunteers after taking the pill.)

The alternative hypothesis is H_1: $\mu_1 > \mu_2$.

You work through the test statistic as follows:

1. The first step is to compute the mean of the differences:

$$\bar{d} = \frac{27}{8} = 3.375$$

2. The next step is to compute the sample standard deviation of the differences:

$$s_d = \sqrt{\frac{\sum\limits_{i=1}^{n}\left(d_I - \bar{d}\right)^2}{n-1}}$$
$$= \sqrt{\frac{23.875}{8-1}}$$
$$= 1.847$$

3. These results are used to compute the test statistic:

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$
$$= \frac{3.375}{1.847 / \sqrt{8}}$$
$$= \frac{3.375}{0.653}$$
$$= 5.168$$

The critical value is found in Table 12-1: $t_\alpha^{n-1} = t_{0.05}^7 = 1.895$

Because the test statistic (5.168) exceeds the critical value (1.895), the null hypothesis is rejected in favor of the alternative hypothesis, which states that the difference between the weights prior to taking the pill and after taking the pill is positive. The results show that the pills are contributing to weight loss.

Chapter 13

Testing Hypotheses about Multiple Population Means

*T*he *analysis of variance* (ANOVA) methodology allows you to directly compare the means of two or more populations. In Chapter 12, I show you how to test hypotheses about the equality of two population means, but with ANOVA, you can test hypotheses about the equality of any number of population means. You can use ANOVA for a wide variety of applications, such as evaluating claims about the effectiveness of competing products, determining whether a new production process reduces costs, identifying the most profitable products to produce, and so forth.

The ANOVA methodology is based on a continuous distribution known as the *F-distribution*. I cover the properties of the F-distribution in depth in this chapter, as well as techniques for computing probabilities under this distribution. I also show you how to calculate the moments of the F-distribution. The F-distribution reappears in later chapters, including Chapters 14 and 16.

Getting to Know the F-Distribution

The F-distribution is a *continuous* probability distribution, which means that it is defined for an *infinite* number of different values. (Continuous probability distributions, such as the normal distribution, are introduced in Chapter 9.)

The F-distribution (named after the statistician Sir Ronald Fisher) can be used for several types of applications, including testing hypotheses about the equality of two population variances and testing the validity of a multiple regression equation. (Testing hypotheses about the equality of two population variances is covered in Chapter 14; multiple regression analysis is covered in Chapter 16.)

The F-distribution shares one important property with the Student's t-distribution (introduced in Chapter 11): Probabilities are determined by a concept known as *degrees of freedom*. Unlike the Student's t-distribution, the F-distribution is characterized by *two* different types of degrees of freedom — *numerator* and *denominator* degrees of freedom.

The F-distribution has two extremely important properties:

- ✔ It's defined only for positive values.
- ✔ It's *not* symmetrical about its mean; instead, it's *positively skewed*.

A distribution is positively skewed if the mean is greater than the median. (The mean and the median are introduced in Chapter 3. The mean is the average value of a distribution, and the median is the midpoint; half of the values in the distribution are below the median, and half are above.)

A good example of a positively skewed distribution is household incomes. Suppose that half of the households in a country have incomes below $50,000 and half have incomes above $50,000; this indicates that the median household income is $50,000. Among households with incomes below $50,000, the smallest possible value is $0. Among households with incomes above $50,000, there may be incomes of several million dollars per year. This imbalance between incomes below the median and above the median causes the mean to be substantially higher than the median. Suppose for example that the mean income in this case is $120,000. This shows that the distribution of household incomes is positively skewed.

Another key property of the F-distribution is that it's uniquely characterized by two values, or parameters, known as *degrees of freedom (df)*. These are known as numerator degrees of freedom and denominator degrees of freedom.

Figure 13-1 shows a graph of the F-distribution for different combinations of numerator and denominator degrees of freedom. In each case, numerator degrees of freedom are listed first, and denominator degrees of freedom are listed second (for example, *1,5* indicates 1 numerator degree of freedom, and 5 denominator degrees of freedom). The *level of significance* in each case is 0.05.

A level of significance is used to test a *hypothesis*. (Hypothesis testing is covered in detail in Chapter 12.) A hypothesis test begins with a *null hypothesis;*

this is a statement that's assumed to be true unless there is *strong* contrary evidence. There is also an *alternative* hypothesis; this is a statement that is accepted in place of the null hypothesis if there's sufficient evidence to reject the null hypothesis.

The level of significance, designated α (alpha), refers to the probability of incorrectly rejecting the null hypothesis when it is actually true. This is known as a *Type I error*. By contrast, a *Type II error* occurs when you fail to reject the null hypothesis when it's actually false. Therefore, with a level of significance of 0.05, there is a 5 percent chance of committing a Type I error.

Figure 13-1 shows that the distribution isn't defined for negative values (as you can see, no negative values appear along the horizontal axis). Additionally, as the number of degrees of freedom increases, the shape of the distribution shifts to the right. The distribution has a long right tail (more formally, it's *skewed to the right,* or *positively skewed*).
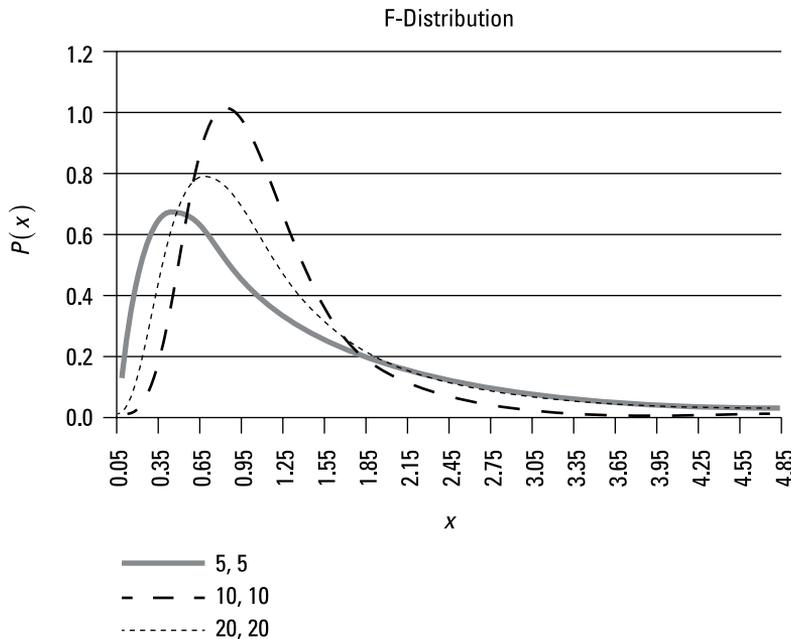
Figure 13-1:
The shape of the F-distribution varies with its degrees of freedom (df).

In the following sections, I go into even more detail about the F-distribution, such as the properties of the F random variable and show you how to compute the moments of the F-distribution.

Defining an F random variable

The F-distribution is defined in terms of the chi-square (χ^2) distribution (see Chapter 14 for details). The chi-square distribution is a continuous distribution that is characterized by its *degrees of freedom*. Like the F-distribution, the chi-square distribution is only defined for positive values and is positively skewed.

The chi-square distribution has several different applications, including testing hypotheses about the variance of a population and testing hypotheses about the probability distribution followed by a population.

The following equation shows that an F random variable is the ratio of two *independent* chi-square random variables (χ_1^2 and χ_2^2) and their respective degrees of freedom (v_1 and v_2):

$$F^{v_1, v_2} = \frac{\chi_1^2 / v_1}{\chi_2^2 / v_2}$$

F^{v_1, v_2} is a random variable that follows the F-distribution and has v_1 numerator degrees of freedom and v_2 denominator degrees of freedom.

Measuring the moments of the F-distribution

Moments are summary measures of a probability distribution and include the following:

- ✔ The *expected value* is known as the first moment of a probability distribution and represents the mean or average value of a distribution.

- ✔ The *variance* is the second central moment and shows how spread out or scattered the values of a distribution are around the expected value.

- ✔ The *standard deviation* isn't a separate moment but is the square root of the variance.

For most applications, the standard deviation is more useful than the variance (because the standard deviation is measured in the same units as the expected value whereas the variance is not). For the F-distribution, you use this formula to determine the expected value:

$$E(X) = \frac{v_2}{v_2 - 2}$$

$E(X)$ represents the expected value, and v_2 represents the denominator degrees of freedom (defined in the previous section).

REMEMBER

The expected value formula requires the denominator degrees of freedom to be greater than 2. Otherwise, the expected value becomes negative or undefined.

The expected value represents the *average* value of the F-distribution. For example, Figure 13-1 shows a graph of the F-distribution with 5 numerator degrees of freedom and 5 denominator degrees of freedom. The expected value equals:

$$E(X) = \frac{v_2}{v_2 - 2}$$
$$= \frac{5}{5 - 2}$$
$$= \frac{5}{3}$$
$$= 1.67$$

Figure 13-1 also shows a graph of the F-distribution with 20 numerator degrees of freedom and 20 denominator degrees of freedom. The expected value equals:

$$E(X) = \frac{v_2}{v_2 - 2}$$
$$= \frac{20}{20 - 2}$$
$$= \frac{20}{18}$$
$$= 1.11$$

This shows that the average value of the F-distribution with 20 numerator degrees of freedom and 20 denominator degrees of freedom is *less than* the average value of the F-distribution with 5 numerator degrees of freedom and 5 denominator degrees of freedom.

To compute the variance, you use this formula:

$$\sigma^2 = \frac{2v_2^2(v_1 + v_2 - 2)}{v_1(v_2 - 2)^2(v_2 - 4)}$$

REMEMBER

The variance formula requires the denominator degrees of freedom to be greater than 4; otherwise, the variance becomes negative or undefined.

The standard deviation is the *square root* of the variance:

$$\sigma = \sqrt{\frac{2v_2^2 (v_1 + v_2 - 2)}{v_1 (v_2 - 2)^2 (v_2 - 4)}}$$

The variance and the standard deviation are used as measures of how spread out the values of the F-distribution are compared with the expected value.

For example, for the F-distribution with 5 numerator degrees of freedom and 5 denominator degrees of freedom, the variance equals

$$\begin{aligned}
\sigma^2 &= \frac{2v_2^2 (v_1 + v_2 - 2)}{v_1 (v_2 - 2)^2 (v_2 - 4)} \\
&= \frac{2(5^2)(5 + 5 - 2)}{5(5 - 2)^2 (5 - 4)} \\
&= \frac{2(25)(8)}{5(9)(1)} \\
&= \frac{400}{45} \\
&= 8.89
\end{aligned}$$

The standard deviation equals the square root of 8.89, or 2.98.

For the F-distribution with 20 numerator degrees of freedom and 20 denominator degrees of freedom, the variance equals

$$\begin{aligned}
\sigma^2 &= \frac{2v_2^2 (v_1 + v_2 - 2)}{v_1 (v_2 - 2)^2 (v_2 - 4)} \\
&= \frac{2(20^2)(20 + 20 - 2)}{20(20 - 2)^2 (20 - 4)} \\
&= \frac{2(400)(38)}{20(324)(16)} \\
&= \frac{30,400}{103,680} \\
&= 0.29
\end{aligned}$$

The standard deviation equals the square root of 0.29, or 0.54.

In Figure 13-1, the F-distribution with 20 numerator degrees of freedom and 20 denominator degrees of freedom has a tail that falls off very rapidly (so that the distribution is less spread out) compared with the F-distribution with 5 numerator degrees of freedom and 5 denominator degrees of freedom;

therefore, the distribution with 20 numerator and denominator degrees of freedom has a lower variance and standard deviation.

Using ANOVA to Test Hypotheses

You use analysis of variance (ANOVA) to test hypotheses about the equality of two or more population means. ANOVA is based on experiments performed on subjects that are *independent* of each other; in other words, they are not related to each other. For example, suppose that a department store chain wants to compare the mean sales of household appliances at its stores in New York, Boston, and Philadelphia. Because these stores are in different geographical locations, the sales at one store don't influence sales at the other stores (they're independent of each other.) And because the sales at these stores are independent, ANOVA can be used to test the hypothesis that mean sales are equal at all three stores.

As an example of the ANOVA process, suppose a manufacturer is considering releasing one of three new types of batteries and wants to determine whether one of these batteries has a longer mean lifetime than the others. If so, it will manufacture this battery exclusively. Otherwise, it will randomly pick one of the three to be manufactured. The proposed names for the three battery types are Electrica, Readyforever, and Voltagenow.

In this experiment, battery lifetime is referred to as the *dependent variable*. The hypothesis tested is that the mean battery lifetime is the same for Electrica, Readyforever, and Voltagenow. The three battery types are *treatments*.

If the mean lifetimes of the Electrica, Readyforever, and Voltagenow batteries are different, this may be due to two different sources. These are known as variation *between* groups (battery types) and variation *within* groups (variation among batteries of the same type).

The process used to test whether the mean battery lifetimes are equal for each type is known as *one-way* ANOVA. If the manufacturer wants to compare the mean lifetimes according to type *and* determine if there can be substantial differences within each type, a more complex version of ANOVA is used. This is known as *two-way* ANOVA.

In the following sections, I walk you through the necessary steps for testing the hypothesis that multiple population means are equal. First, the null and alternative hypotheses are explained, followed by a discussion of a concept known as the level of significance. I show you how to construct a test statistic from the F-distribution and how to find the critical values that the test statistic is compared with. I also explain how the final conclusion is reached.

The following procedure is based on the assumptions that

✔ The samples chosen are independent of each other

✔ The underlying populations have equal variances

✔ The populations are normally distributed

Writing the null and alternative hypotheses

The null hypothesis is a statement that's assumed to be true unless you find strong contrary evidence. For testing the hypothesis that three population means are equal, you write:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

In this expression,

H_0 represents the null hypothesis

μ_1, μ_2, μ_3 represent the means of population 1, 2 and 3

The alternative hypothesis is a statement that you accept in the event that the null hypothesis is rejected (for example, there's strong evidence against it). When testing the hypothesis that three population means are equal, the alternative hypothesis is simply that the three population means are *not* equal.

This alternative hypothesis can be expressed in different ways, such as:

H_1: The three means are *not* all equal

H_1: At least one of the three means is different from the others

In these examples, H_1 represents the alternative hypothesis.

Choosing the level of significance

To test a hypothesis, you have to choose a *level of significance*. The level of significance, designated with α, equals the probability of incorrectly rejecting the null hypothesis when it's actually true. This is called a *Type I error*. A Type II error occurs when you fail to reject the null hypothesis when it's not

true. (Check out Chapter 12 for details on Type I and Type II errors in hypothesis testing.)

For many business applications, the level of significance is chosen as 0.05 (5 percent). Other frequent choices include 0.001, 0.01, and 0.10.

Computing the test statistic

The test statistic is a numerical value that is used to determine if the null hypothesis should be rejected. The form of the test statistic depends on the type of hypothesis being tested. If the test statistic has an extremely large positive or negative value, this may be a sign that the null hypothesis is incorrect and should be rejected.

Constructing the test statistic for ANOVA is quite complex, compared with other types of hypothesis tests (see Chapter 12 for a discussion of the steps required for hypothesis testing).

Referring to the battery example, assume that the manufacturer randomly chooses a sample of four Electrica batteries, four Readyforever batteries, and four Voltagenow batteries and then tests their lifetimes. Table 13-1 lists the results (in hundreds of hours).

Table 13-1	Battery Lifetimes (in Hundreds of Hours)		
	Electrica	*Readyforever*	*Voltagenow*
Battery 1	2.4	1.9	2.0
Battery 2	1.7	2.1	2.3
Battery 3	3.2	1.8	2.1
Battery 4	1.9	1.6	2.2

Each element in this table can be represented as a variable with two *indexes*, one for the row and one for the column. In general, this is written as X_{ij}. The *subscript i* represents the row index and *j,* represents the column index. For example, X_{23} represents the element found in the second row and third column. (In Table 13-1, this is 2.3.) X_{31} represents the element found in the third row and the first column. (In Table 13-1, this is 3.2.) Table 13-2 shows the appropriate indexes for all the elements in Table 13-1.

Table 13-2	Battery Lifetimes Shown with Subscripts		
	Electrica	*Readyforever*	*Voltagenow*
Battery 1	X_{11}	X_{12}	X_{13}
Battery 2	X_{21}	X_{22}	X_{23}
Battery 3	X_{31}	X_{32}	X_{33}
Battery 4	X_{41}	X_{42}	X_{43}

The data in Table 13-1 is used to construct the test statistic. The first step in constructing the test statistic is to calculate the following three measures:

Error sum of squares (SSE)

Treatment sum of squares (SSTR)

Total sum of squares (SST)

The calculations are detailed in the following sections.

Finding the error sum of squares (SSE)

The error sum of squares (abbreviated SSE) is obtained by first computing the mean lifetime of each battery type. For each battery of a specified type, the mean is subtracted from each individual battery's lifetime and then squared. The sum of these squared terms for all battery types equals the SSE.

SSE is a measure of *sampling error*. This refers to the fact that the values computed from a sample will be somewhat different from one sample to the next.

To compute the SSE for this example, the first step is to find the mean for each column. So, for example, you find the mean of column 1, with this formula:

$$\bar{X}_{\bullet 1} = \frac{\sum_{i=1}^{n} X_{i1}}{n_1}$$

Here's what each term means:

- $\bar{X}_{\bullet 1}$ is the mean of column 1 (the bar indicates that this is a mean). The subscripts indicate that this average is computed from all elements within column 1.

- X_{i1} is the value of X in row i and column 1.

- n_1 is the number of elements in column 1.

So, using the values in Table 13-1, you find the mean of column 1 like so:

$$\bar{X}_{\bullet 1} = \frac{\sum_{i=1}^{n} X_{i1}}{n_1}$$
$$= \frac{2.4 + 1.7 + 3.2 + 1.9}{4}$$
$$= \frac{9.2}{4}$$
$$= 2.3$$

In other words, you sum the lifetimes of the four Electrica batteries and divide by 4. The mean lifetime of the Electrica batteries in this sample is 2.3.

Similarly, you find the mean of column 2 (the Readyforever batteries) as

$$\bar{X}_{\bullet 2} = \frac{\sum_{i=1}^{n} X_{i2}}{n_2}$$
$$= \frac{1.9 + 2.1 + 1.8 + 1.6}{4}$$
$$= \frac{7.4}{4}$$
$$= 1.85$$

And column 3 (the Voltagenow batteries) as

$$\bar{X}_{\bullet 3} = \frac{\sum_{i=1}^{n} X_{i3}}{n_3}$$
$$= \frac{2.0 + 2.3 + 2.1 + 2.2}{4}$$
$$= \frac{8.6}{4}$$
$$= 2.15$$

The next step is to subtract the mean of each column from each element within that column, then square the result. I set up the calculations in Table 13-3.

Table 13-3	Battery Lifetimes: Squared Differences from the Column Means		
	Electrica	*Readyforever*	*Voltagenow*
Battery 1	$(2.4 - 2.3)^2 = 0.01$	$(1.9 - 1.85)^2 = 0.0025$	$(2.0 - 2.15)^2 = 0.0225$
Battery 2	$(1.7 - 2.3)^2 = 0.36$	$(2.1 - 1.85)^2 = 0.0625$	$(2.3 - 2.15)^2 = 0.0225$
Battery 3	$(3.2 - 2.3)^2 = 0.81$	$(1.8 - 1.85)^2 = 0.0025$	$(2.1 - 2.15)^2 = 0.0025$
Battery 4	$(1.9 - 2.3)^2 = 0.16$	$(1.6 - 1.85)^2 = 0.0625$	$(2.2 - 2.15)^2 = 0.0025$
Sum	**1.34**	**0.13**	**0.05**

For example, because 2.3 is the mean of column 1, you subtract 2.3 from each element in column 1. You square the result in each row, and the sum of these squared values is 1.34. Repeat the process for columns 2 and 3 to get sums of 0.13 and 0.05, respectively. Add up the sums to get the error sum of squares (SSE): 1.34 + 0.13 + 0.05 = 1.52.

The error sum of squares shows how much variation there is among the lifetimes of the batteries of a given type. The smaller the SSE, the more uniform the lifetimes of the different battery types.

Calculating the treatment sum of squares (SSTR)

After you find the SSE, your next step is to compute the *treatment sum of squares (SSTR).* This is a measure of how much variation there is among the mean lifetimes of the battery types. With a low SSTR, the mean lifetimes of the different battery types are similar to each other.

First, you need to calculate the overall average for the sample, known as the *overall mean* or *grand mean*. In the battery example from the previous sections, you have 12 total observations (four batteries chosen from each of three battery types; the data are in Table 13-1). You may obtain the overall mean by adding up the 12 sample values and dividing by 12:

$$\frac{2.4 + 1.7 + 3.2 + 1.9 + 1.9 + 2.1 + 1.8 + 1.6 + 2.0 + 2.3 + 2.1 + 2.2}{12} = \frac{25.2}{12} = 2.1$$

You then compute the SSTR with the following steps for each column:

1. **Compute the squared difference between the column mean and the overall mean.**

2. **Multiply the result by the number of elements in the column.**

So in this example, SSTR equals

$$SSTR = 4(2.3 - 2.1)^2 + 4(1.85 - 2.1)^2 + 4(2.15 - 2.1)^2 = 0.42.$$

The calculations are based on the following results obtained in previous sections:

✔ There are four observations in each column.

✔ The overall mean is 2.1.

✔ The column means are 2.3 for column 1, 1.85 for column 2 and 2.15 for column 3.

After you compute SSE and SSTR, the sum of these terms is calculated, giving the total sum of squares (SST). This is shown in the next section.

Computing the total sum of squares (SST)

The *total sum of squares (SST)* equals the sum of the SSTR and the SSE (see the preceding sections). So using the battery example, you get

$$
\begin{aligned}
SST &= SSTR + SSE \\
&= 0.42 + 1.52 \\
&= 1.94
\end{aligned}
$$

When you compute SSE, SSTR, and SST, you're ready to proceed to the next step in computing the test statistic. The test statistic is computed from the mean (average) of SSE and SSTR; these are known as:

Error mean square (MSE)

Treatment mean square (MSTR)

The calculations are detailed in the following sections.

Getting the error mean square (MSE)

After you find the sums of squares (see sections "Calculating the treatment sum of squares (SSTR)" and "Computing the total sum of squares (SST)"), you compute the means of the SSE and SSTR. These are known as *error mean square (MSE)* and *treatment mean square (MSTR)*.

You find the MSE by dividing the SSE by N (total number of observations) minus t (total number of treatments) as shown in this formula:

$$MSE = \frac{SSE}{N - t}$$

In the battery lifetimes example (based on Table 13-1), there are a total of 12 observations or elements in the sample data, so N = 12; there are also three battery types, so the number of treatments is t = 3. Therefore, the MSE is

$$MSE = \frac{SSE}{N-t}$$
$$= \frac{1.52}{12-3} = 0.1689$$

MSE measures the average variation *within* the treatments; for example, how different the battery means are within the same type.

Getting the treatment mean square (MSTR)

The MSTR equals the SSTR divided by the number of treatments, minus 1 (t – 1), which you can write mathematically as:

$$MSTR = \frac{SSTR}{t-1}$$

So you find the MSTR for the battery example, (here, *t* is the number of battery types) as follows:

$$MSTR = \frac{SSTR}{t-1}$$
$$= \frac{0.42}{3-1} = 0.21$$

MSTR measures the average variation *among* the treatment means, such as how different the means of the battery types are from each other.

Solving for the F-statistic

The test statistic for the ANOVA process follows the F-distribution, and it's often called the *F-statistic*. The test statistic is computed as follows:

$$F = \frac{MSTR}{MSE}$$
$$= \frac{0.21}{0.1689} = 1.24$$

The test statistic shows the ratio of the treatment mean square (MSTR) to the error mean square (MSE). The greater is this value, the more unlikely it is that the means of the three batteries are equal to each other. As a result, a sufficiently large value of this test statistic results in the null hypothesis being rejected.

Finding the critical values using the F-table

Because the F-distribution is based on two types of degrees of freedom, there's one table for each possible value of α (the level of significance). Table 13-4 shows the different values of the F-distribution corresponding to a 0.05 (5 percent) level of significance.

Table 13-4			The F-Distribution with $\alpha = 0.05$					
$v2\backslash v1$	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
2	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18

The numbers across the top row of the table represent the numerator degrees of freedom (v_1). You read across this top row to find the appropriate numerator degrees of freedom. The first column represents the denominator degrees of freedom (v_2); you read down this column to find the appropriate denominator degrees of freedom. The critical value is found at the intersection of the row and column you choose. For example, suppose that the numerator degrees of freedom is 5 and the denominator degrees of freedom is 7. The appropriate test statistic is 3.97.

For the one-way ANOVA process, you compute the numerator and denominator degrees of freedom as follows:

Numerator degrees of freedom = treatments $- 1 = t - 1 = 3 - 1 = 2$

Denominator degrees of freedom = total observations minus treatments = $N - t = 12 - 3 = 9$

In this example, you're looking for a right-tail area of 5 percent under the F-distribution with numerator degrees of freedom = $v_1 = t - 1 = 2$ and denominator degrees of freedom = $v_2 = N - t = 9$. You find this critical value at the intersection of the *2 degrees of freedom* column and the *9 degrees of freedom* row. The critical value equals 4.26, which you can write as

$$F_\alpha^{v_1, v_2} = F_{0.05}^{2,9} = 4.26$$

The superscripts represent the numerator and denominator degrees of freedom, respectively. The subscript represents the level of significance.

Coming to the decision

The one-way ANOVA hypothesis test is a right-tailed test. This type of test leads to the rejection of the null hypothesis if the value of the test statistic is *too large* to be consistent with the null hypothesis; in other words, if the test statistic is *greater than* the critical value, the null hypothesis is rejected. (Otherwise, the null hypothesis is *not* rejected; there is not enough evidence against it.)

In the battery lifetime example introduced in the section "Using ANOVA to Test Hypotheses," the test statistic equals 1.24, whereas the critical value equals 4.26. Because the test statistic does *not* exceed the critical value, the null hypothesis that the three population means are equal is *not* rejected. This indicates that there is not enough evidence against the hypothesis of equal means to reject it.

Figure 13-2 shows the F-distribution with 2 numerator degrees of freedom and 9 denominator degrees of freedom and a level of significance of 0.05. The test statistic is 1.24 and the critical value is 4.26.

The area to the right of the critical value is the *rejection region*. This is the area under the F-distribution, which is too far away from the critical value to be consistent with the null hypothesis. Because the test statistic doesn't fall within the rejection region, the null hypothesis fails to be rejected. The result indicates that there is not enough evidence against the assumed equality of the mean lifetimes of Electrica, Readyforever, and Voltagenow. Because the manufacturer wants to produce the battery with the longest mean lifetime, it can choose any of the three.

F-Distribution
Degrees of Freedom = 2, 9

Figure 13-2:
The null
hypoth-
esis is *not*
rejected.
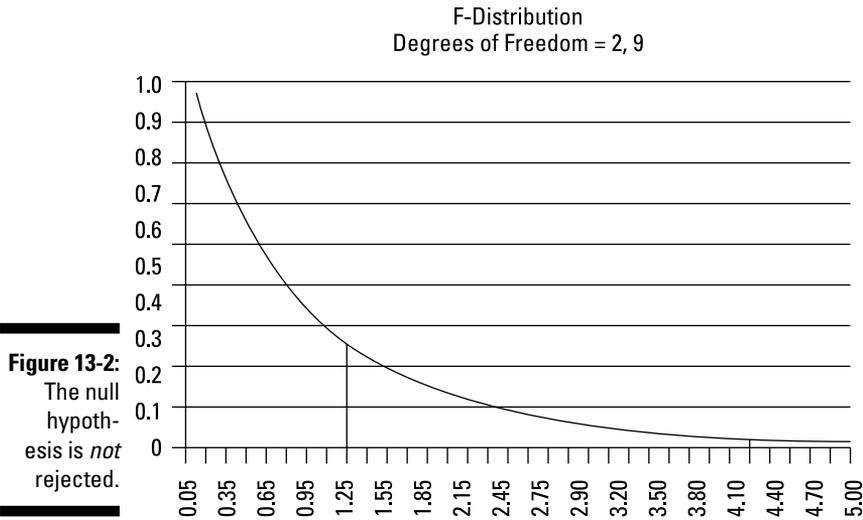
Using a spreadsheet

As an alternative to computing the test statistic and the critical value, you can use a spreadsheet program, such as Excel, to test the null hypothesis that the three population means are equal. If you choose to use a spreadsheet, you'd skip the steps in the sections "Computing the test statistic" and "Finding the critical values using the F-table."

The output of the spreadsheet program is obtained as follows. The first step is to enter the sample data:

	Electrica	*ReadyForever*	*Voltagenow*
Battery 1	2.4	1.9	2.0
Battery 2	1.7	2.1	2.3
Battery 3	3.2	1.8	2.1
Battery 4	1.9	1.6	2.2

If you are using Excel, the next step is to choose the Data tab, then select Data Analysis. This opens up a dialog box containing several statistical procedures that may be performed. The appropriate choice for this example is ANOVA: Single Factor. Click this choice to open a new dialog box; enter the range of cells containing the data into the box Input Range. Be sure to check the box Labels in First Row. Click the OK button to produce the output shown in Figure 13-3.

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Electrica	4	9.2	2.3	0.446667
Readyforever	4	7.4	1.85	0.043333
Voltagenow	4	8.6	2.15	0.016667

Figure 13-3:
The results
of a one-
way ANOVA
test.

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.45	2	0.21	1.243421	0.333571	4.256495
Within Groups	1.52	9	0.168889			
Total	1.94	11				

Figure 13-3 shows the following key results:

✔ The column means in the *Average* column of the *Summary* section (2.3, 1.85 and 2.15)

✔ The error sum of squares (SSE) in the *Within Groups* row of the *ANOVA: Source of Variation* section (1.52)

✔ The treatment sum of squares (SSTR) in the *Between Groups* row of the *ANOVA: Source of Variation* section (0.42)

✔ The total sum of squares (SST) or *Total* (1.94)

✔ The treatment mean square (MSTR) at the intersection *Between Groups* and the *MS* column under the *ANOVA: Source of Variation* section (0.21)

✔ The error mean square (MSE) at the intersection of *Within Groups* and *MS* under the *ANOVA: Source of Variation* section (0.168889)

✔ The F-statistic at the intersection of *Between Groups* and the *F* column (1.243421)

✔ The critical value at the intersection of *Between Groups* and *F crit* Column (4.256495)

The printout confirms the results of the previous sections. Because the test statistic is less than the critical value, the null hypothesis that the mean battery lifetimes are equal should *not* be rejected.

Chapter 14

Testing Hypotheses about the Population Mean

. . . .

In This Chapter

▶ Introducing the chi-square distribution

▶ Testing hypotheses about the variance of a single population

▶ Testing hypotheses about the equality of two population variances

▶ Implementing goodness of fit tests with the chi-square distribution

. . . .

*T*his chapter covers two types of hypothesis tests: tests about the population variance, and *goodness of fit* tests. Goodness of fit tests determine whether a population follows a specified distribution, such as the normal distribution (for a thorough introduction to the normal distribution, see Chapter 9). Because many business applications rely on the assumption of normality, goodness of fit tests are particularly valuable.

To implement a goodness of fit test, you use a continuous distribution known as the *chi-square distribution.* This distribution has many interesting features, which I explain in detail and illustrate throughout this chapter; its properties are quite different from the normal distribution.

I also explain how to use a chi-square table to compute probabilities under the chi-square distribution, and I show you how to compute moments for the chi-square distribution. (*Moments* are summary measures of a probability distribution that provide a great deal of useful information in a very compact form.)

You can use the chi-square distribution to test hypotheses about the variance of a population. For example, you can use the chi-square distribution to determine the level of risk contained in a stock portfolio. (The process of testing a hypothesis about a population variance is closely related to other types of hypothesis tests, which I cover in Chapter 12.)

What about two populations, you ask? Testing hypotheses about the equality of two population variances requires yet another continuous distribution: the F-distribution. (I provide an overview of the F-distribution in Chapter 13.) The F-distribution also plays a key role in multiple regression analysis (see Chapter 16).

Staying Positive with the Chi-Square Distribution

The chi-square distribution (χ^2) is a *continuous* probability distribution, which means that it's defined for an infinite number of values. I introduce continuous probability distributions, including the normal, Student's t-, and F-distributions, in Chapters 9, 11, and 13. (To read about discrete probability distributions, check out Chapter 8.)

The chi-square distribution has several different applications. This section shows you how to use the chi-square distribution to:

- ✔ Test hypotheses about the variance of a population
- ✔ Carry out "goodness of fit" tests

Portfolio managers, financial analysts, traders, and so on regularly use continuous distributions in business applications to analyze the properties of financial variables. Two of the more widely used continuous distributions are the normal and Student's t-distributions (see Chapters 9 and 11, respectively). Many business situations can be described with the normal distribution, such as returns to stocks, corporate profits, and so on. The normal and Student's t-distribution can also be used to construct confidence intervals (described in Chapter 11) and test hypotheses about population means (described in Chapter 12.)

As with the Student's t-distribution, the chi-square distribution is uniquely characterized by a value known as *degrees of freedom (df)*. The number of degrees of freedom is based on the sizes of samples used to estimate population parameters, such as the mean or the variance.

Here are two important features of the chi-square distribution:

- ✔ It's defined only for positive values.
- ✔ It's *not* symmetrical about its mean; instead, it's positively skewed.

A distribution may be symmetrical about its mean, in which case the area below the mean is a mirror image of the area above the mean. For a symmetric distribution, the mean equals the median. (I discuss symmetry in Chapter 3.) A distribution may also be negatively skewed, where the mean is less than the median, or positively skewed, where the mean is greater than the median.

The chi-square distribution is positively skewed; graphically, it has a long right tail. The next section shows several graphs of the chi-square distribution with different numbers of degrees of freedom. The smaller the degrees of freedom, the more skewed the distribution is; with a larger number of degrees of freedom, the distribution becomes more symmetrical and begins to resemble the normal distribution.

Following the graphs of the chi-square distribution is a discussion of how to compute the moments of the chi-square distribution.

Representing the chi-square distribution graphically

Figures 14-1, 14-2, and 14-3 show the chi-square distribution with 5, 10, and 30 degrees of freedom. In each case, the horizontal axis represents different possible values of the chi-square distribution; the vertical axis represents the corresponding probabilities. With a continuous distribution such as the chi-square, probabilities correspond to areas under the curve.
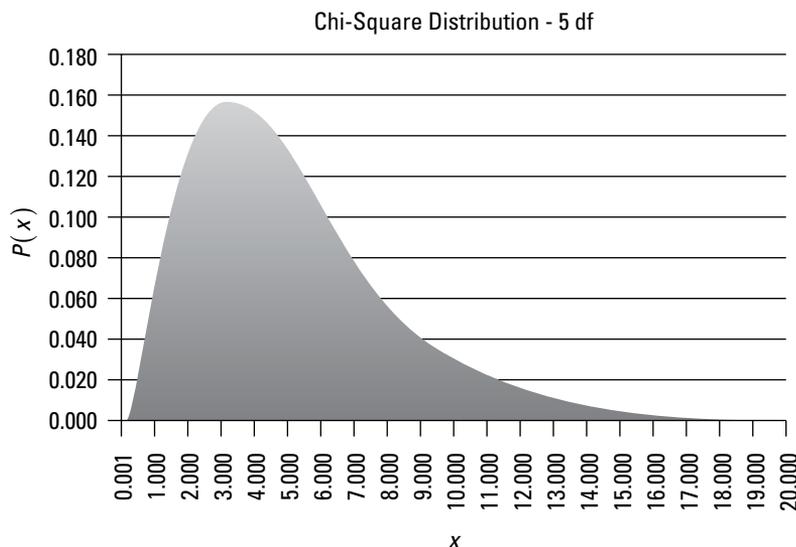
Figure 14-1:
The chi-square distribution with 5 degrees of freedom.
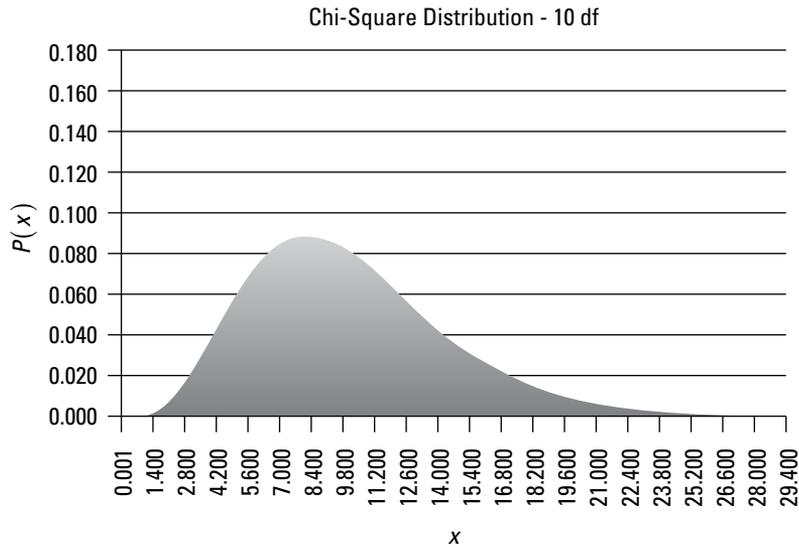
Chi-Square Distribution - 10 df

Chi-Square Distribution - 30 df
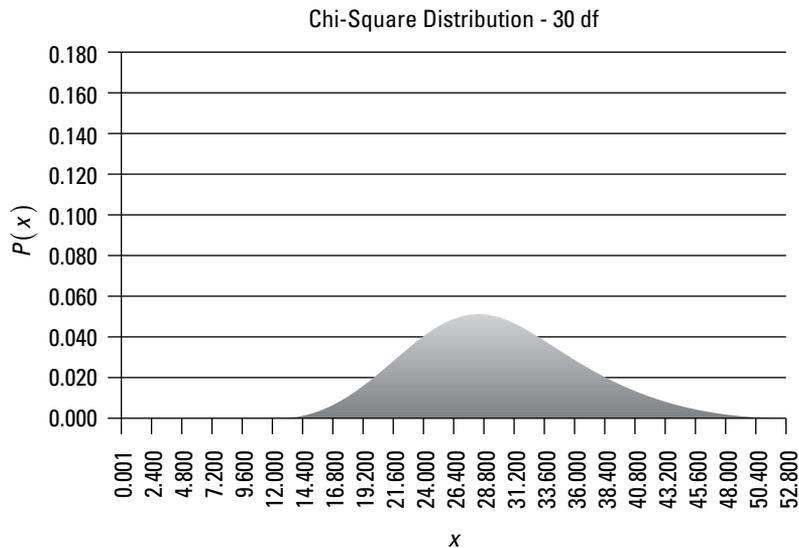
As you can see in each figure, the distribution isn't defined for negative values — that is, no negative values appear along the horizontal axis. Additionally, as the number of degrees of freedom increases, the distribution shifts to the right and begins to resemble the normal distribution (it has a long right tail and is skewed to the right).

Defining a chi-square random variable

A *chi-square random variable* is composed of a sum of independent, squared standard normal random variables (Z^2) (see Chapter 7 for details). The standard normal distribution is the special case of the normal distribution where the mean (μ) equals 0 and the standard deviation (σ) equals 1. You can write the definition of a chi-square random variable mathematically as

$$\chi_v^2 = Z_1^2 + Z_2^2 + Z_3^2 + \ldots + Z_v^2$$

Because each standard normal random variable is squared, the sum of these terms is guaranteed to be positive (which is why the chi-square distribution isn't defined for negative values).

The letter v (or "nu") represents the number of terms in this expression; here, v is the number of degrees of freedom of the distribution. For example, the chi-square distribution with 5 degrees of freedom is defined as follows:

$$\chi_5^2 = Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2 + Z_5^2$$

Checking out the moments of the chi-square distribution

Moments are summary measures of a probability distribution (see Chapter 8 for details) and include the *expected value* (or mean) and the *variance* (how spread out the values are). The *standard deviation* is the square root of the variance.

Each probability distribution has its own unique set of formulas for computing the expected value, variance, and standard deviation. For the chi-square distribution, these are given as follows:

- ✔ The expected value equals the number of degrees of freedom (v) of the distribution:

  $$E(X) = v$$

  For example, in a chi-square distribution with 5 degrees of freedom, the expected value is 5.

- ✔ The variance equals two times the number of degrees of freedom:

  $$\sigma^2 = 2v$$

  For example, for the chi-square distribution with 5 degrees of freedom, the variance is $2 \times 5 = 10$.

✔ The standard deviation is the square root of the variance:

$$\sigma = \sqrt{2\upsilon}$$

For example, for the chi-square distribution with 5 degrees of freedom, the standard deviation is the square root of 10, which is approximately 3.16.

REMEMBER

Moments capture the key properties of a probability distribution. The expected value is another name for the average; the variance and standard deviation show how "spread out" the values of the distribution are relative to the expected value.

Testing Hypotheses about the Population Variance

In business, one of the most widely used applications of the chi-square distribution is to determine whether the variance of a population equals a specified value. The basic approach to testing a hypothesis about the population variance exactly mirrors the approach used for the population mean (which I cover in Chapter 12). The most important changes take place in the test statistic and critical values you use.

In the following sections, I walk you through the steps to testing hypotheses about the population variance.

Defining what you assume to be true: The null hypothesis

The first step in the hypothesis testing procedure is writing the *null hypothesis,* which is a statement that's assumed to be true unless strong contrary evidence exists against it. For example, suppose that a manufacturer is concerned that the variance of the computer chips that it produces exceeds 0.001, which would indicate that there's a problem with the production process. The manufacturer could test this hypothesis by selecting a sample of computer chips and computing their sample variance.

The manufacturer may not want to make any changes to the production process unless clear evidence shows that it's necessary. Therefore, it uses the null hypothesis that the variance equals 0.001. If this hypothesis is rejected, the alternative that the variance exceeds 0.001 is accepted instead. Unless the null hypothesis can be disproved with strong evidence, no changes are made to the production process. (Hypothesis testing is introduced in Chapter 12.)

For testing hypotheses about the population variance, the null hypothesis statement is based on the assumption that the population variance equals the *hypothesized value* of the population (σ_0^2). This assumption isn't abandoned without strong contradictory evidence.

Mathematically, you write the null hypothesis as

$$H_0 : \sigma^2 = \sigma_0^2$$

The variance with a subscript of 0 (σ_0^2) is the hypothesized value of the variance. This is the value that you believe the population variance is equal to. The hypothesis test shows whether this belief is backed up by actual data.

For example, suppose that an economist wants to determine whether the variance of the inflation rate over the past 20 years equals 0.0001, in which case, you write the null hypothesis as $H_0 : \sigma^2 = 0.0001$. The economist continues to assume that this is the correct variance unless the hypothesis test provides strong evidence against this claim.

Stating the alternative hypothesis

Your second step in a hypothesis test is to specify the *alternative hypothesis*. If the statistical evidence against the null hypothesis is strong enough to reject it, you need an alternative statement to accept in its place.

The alternative hypothesis is a statement of what you accept to be true if the null hypothesis is rejected. For example, the economist in the previous section may want to know whether the actual variance is less than 0.0001, greater than 0.0001, or simply different from 0.0001 if the null hypothesis is rejected.

You can express the alternative hypothesis in three ways: as right-tailed, left-tailed, and two-tailed tests.

- ✔ With a right-tailed test, you look for evidence that the actual population variance is *greater than* the hypothesized value.
- ✔ With a left-tailed test, you look for evidence that the population variance is *less than* the hypothesized value.
- ✔ With a two-tailed test, you look for evidence that the population variance is *either* less than or greater than the hypothesized value.

I explore each option in the following sections. (Right-tailed tests, left-tailed tests, and two-tailed tests are introduced in Chapter 12.)

Right-tailed test: Determining whether the hypothesized variance is too low

If you're interested in knowing only whether the population variance is greater than the hypothesized value, you use a right-tailed test. In this case, you express the alternative hypothesis (H_1) as

$$H_1 : \sigma^2 > \sigma_0^2$$

For example, suppose that a manufacturing company wants to keep the weights of its computer chips as uniform as possible. It determined from experience that the maximum variance the chips can tolerate is 0.0006 milligrams squared. (Variances are measured in terms of squared units, as I discuss in Chapter 4.) The manufacturing company can test the variance by choosing a sample off of the assembly line and computing the sample variance. In this case, the company can test the hypothesis that the variance equals 0.0006 ($H_0 : \sigma^2 = 0.0006$); the alternative hypothesis is that the variance exceeds (or is greater than) 0.0006 ($H_1 : \sigma^2 > 0.0006$).

The results of this test show whether the manufacturing process is working correctly or whether it needs to be adjusted.

Left-tailed test: Determining whether the hypothesized variance is too high

If you're interested in knowing only whether the population variance is less than the hypothesized value, you use a left-tailed test. In this case, you express the alternative hypothesis as

$$H_1 : \sigma^2 < \sigma_0^2$$

For example, suppose that an equity analyst is studying the pattern of returns to U.S. stocks since the outbreak of the last financial crisis. The analyst wants to determine whether markets have begun to stabilize since the crisis began, which is indicated by a drop in the variances of the returns to U.S. stocks. The analyst believes that one stock is particularly representative of the performance of the overall economy. He wants to see whether the variance of its returns has remained at 0.0004 or whether it's fallen below this level. In this case, the analyst can test the null hypothesis that the variance equals 0.0004 ($H_0 : \sigma^2 = 0.0004$); the alternative hypothesis is that the variance is less than 0.0004 ($H_1 : \sigma^2 < 0.0004$).

The results of this test show whether the variance of this stock has fallen below 0.0004. If so, the markets have stabilized since the outbreak of the financial crisis.

Two-tailed test: Determining whether the hypothesized variance is too low or too high

In some situations, it's extremely important for you to know whether the population variance is greater than or less than the hypothesized value. In this case, you use the two-tailed test, and write the alternative hypothesis as

$$H_1 : \sigma^2 \neq \sigma_0^2$$

For example, suppose that the variance of the returns to an investor's portfolio has historically been 0.0009; the investor wants to determine whether this number has increased or decreased over the past year. In this case, the investor can use a two-tailed hypothesis test. The null hypothesis is that the variance equals 0.0009 ($H_0 : \sigma^2 = 0.0009$), and the alternative hypothesis is that the variance doesn't equal 0.0009 ($H_1 : \sigma^2 \neq 0.0009$).

Choosing the level of significance

To test a hypothesis, you have to choose a *level of significance*. The level of significance, designated with α, refers to the probability of rejecting the null hypothesis when it's actually true, called a *Type I error*. (Chapter 12 provides details on Type I and Type II errors in hypothesis testing.)

You must choose the level of significance carefully. The greater the level of significance, the greater the likelihood of rejecting the null hypothesis when it's true — and the lower the likelihood of failing to reject the null hypothesis when it's false.

You choose the level of significance based on the relative importance of avoiding these errors. For many business applications, the level of significance is set to 0.05 (or 5 percent.) Other commonly used values are 0.01 and 0.10.

Calculating the test statistic

To test hypotheses about the population variance, you must draw a sample from the underlying population so you can compute the sample variance. The sample variance is required to compute the test statistic:

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

This equation shows that the test statistic follows the chi-square distribution, with $n - 1$ degrees of freedom (n is the sample size); s^2 is the sample variance, and σ_0^2 is the hypothesized value of the population variance.

This expression is used as a test statistic because it can be shown to follow the chi-square distribution with $n - 1$ degrees of freedom.

The purpose of the test statistic is to determine how extreme a sample statistic is (in this case, the sample variance) compared with the hypothesized value of the corresponding population parameter (here, the population variance.) If the test statistic is too extreme (the value is an extremely large positive or negative number), it's highly unlikely that the null hypothesis is true, and it will be rejected. Otherwise, the null hypothesis won't be rejected.

To determine how extreme the test statistic is, you compare its value to one or two numbers known as *critical values,* depending on the alternative hypothesis. When testing hypotheses about the population variance, critical values are taken from the chi-square distribution. They represent the cutoff point between a specified area under the chi-square distribution.

For example, for the chi-square distribution with 10 degrees of freedom, a critical value of 18.30 is the cutoff point between the upper 5 percent of the chi-square distribution and the lower 95 percent of the chi-square distribution.

In other words, for a chi-square random variable X,

$$P(X \geq 18.30) = 0.05$$
$$P(X \leq 18.30) = 0.95$$

Determining the critical value(s)

To test a hypothesis about the variance of a population, the critical value(s) depends on the alternative hypothesis. Unlike critical values drawn from the standard normal distribution or the Student's t-distribution, the chi-square distribution has no negative critical values. Instead, you determine the critical values with the alternative hypothesis tests as explained in the following sections.

Right-tailed test: Testing hypotheses about the population variance

A right-tailed test has a single critical value because you're looking only for evidence that the test statistic is *too large* to be consistent with the null hypothesis. If you don't find this evidence, you won't reject the null hypothesis. The form of the critical value is

$$\chi^2_{\alpha, n-1}$$

In this expression,

χ^2 = a value chosen from the chi-square distribution

α = the level of significance of the hypothesis test (for example, 0.01, 0.05, 0.10, and so on)

n = the sample size

The values of α and n uniquely identify the appropriate test statistic drawn from the chi-square distribution. This value represents the threshold of the right tail of the chi-square distribution with area α and $n - 1$ degrees of freedom. The area in the right tail is α. You can find this critical value in a chi-square table, such as Table 14-1.

Table 14-1				**The Chi-Square Table**			
df\Right-Rail Area	**0.99**	**0.975**	**0.95**	**0.90**	**0.10**	**0.05**	**0.025**	**0.01**
1	0.00	0.00	0.00	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209

For example, suppose you conduct a right-tailed test with a level of significance of 0.05 (5 percent). You draw a sample of size 10. Plugging those numbers into the critical value, you get

$$\chi^2_{\alpha,\,n-1} = \chi^2_{0.05,\,9}$$

You then look at the chi-square table (Table 14-1). The top row represents areas in the right tail of the chi-square distribution. The first column represents the number of degrees of freedom.

In this example, you're looking for a right-tail area of 0.05 with 9 degrees of freedom ($n - 1 = 10 - 1 = 9$). By looking in the row corresponding to 9 degrees of freedom and the column corresponding to a right-tail area of 0.05, you see that the critical value is 16.919. Therefore,

$$\chi^2_{\alpha,\,n-1} = \chi^2_{0.05,\,9} = 16.919$$

As a result, if the test statistic is greater than 16.919, you reject the null hypothesis; otherwise, you don't reject the null hypothesis.

Left-tailed test: Testing hypotheses about the population variance

A left-tailed test has a single critical value because you're looking only for evidence that the test statistic is *too small* to be consistent with the null hypothesis. If you don't find this evidence, you won't reject the null hypothesis. The form of the test statistic is

$$\chi^2_{(1-\alpha),\, n-1}$$

This value represents the threshold of the left tail of the chi-square distribution with area α and $n - 1$ degrees of freedom. The area in the right tail is, therefore, $1 - \alpha$.

Using the example in the previous section and referring to Table 14-1, if you do a left-tailed test with a level of significance of 0.05 and a sample of size 10, you find the appropriate critical value in the row with 9 degrees of freedom but a right-tail area of 0.95, which is

$$\chi^2_{(1-\alpha),\, n-1} = \chi^2_{0.95,\, 9} = 3.325$$

As a result, if the test statistic is less than 3.325, you reject the null hypothesis; otherwise, you don't reject the null hypothesis.

Two-tailed test: Testing hypotheses about the population variance

A two-tailed test has two critical values. You're looking for evidence that the test statistic is too large *or* too small to be consistent with the null hypothesis. If you don't find this evidence, you won't reject the null hypothesis. The form of the critical values are

$$\chi^2_{(1-\alpha/2),\, n-1}$$

$$\chi^2_{(\alpha/2),\, n-1}$$

The two-tailed test has a right tail and a left tail. Each has an area equal to $\alpha/2$. So, for example, if you do a two-tailed test with a level of significance of 0.05 and a sample of size 10, the appropriate critical values are 2.700 and 19.023 (see Table 14-1).

$$\chi^2_{(1-\alpha/2),\, n-1} = \chi^2_{0.975,\, 9} = 2.700$$

$$\chi^2_{(\alpha/2),\, n-1} = \chi^2_{0.025,\, 9} = 19.023$$

The boundary of the left 2.5 percent tail of the chi-square distribution is 2.700, and the boundary of the right 2.5 percent tail of the chi-square distribution is 19.023. Note that with the right-tailed test, the right tail has

an area of 5 percent; with a left-tailed test, the left tail has an area of 5 percent. With a two-tailed test, the 5 percent area is split between the left and right tails; therefore, each has an area of 2.5 percent.

As a result, if the test statistic is less than 2.700 or greater than 19.023, you reject the null hypothesis; otherwise, you don't reject the null hypothesis.

Making the decision

You decide whether to reject the null hypothesis by looking at the relationship between the test statistic and the critical value(s). There are three possible cases: a right-tailed test, a left-tailed test, and a two-tailed test.

- ✔ **Right-tailed test:** If the test statistic is greater than the critical value $\chi^2_{\alpha,\,n-1}$, you reject the null hypothesis H_0: $\sigma^2 = \sigma_0^2$ in favor of the alternative hypothesis H_1: $\sigma^2 > \sigma_0^2$. Otherwise, you don't reject the null hypothesis.

- ✔ **Left-tailed test:** If the test statistic is less than the critical value $\chi^2_{(1-\alpha),\,n-1}$, you reject the null hypothesis H_0: $\sigma^2 = \sigma_0^2$ in favor of the alternative hypothesis H_1: $\sigma^2 < \sigma_0^2$. Otherwise, you don't reject the null hypothesis.

- ✔ **Two-tailed test:** If the test statistic is less than the critical value $\chi^2_{(1-\alpha/2),\,n-1}$, you reject the null hypothesis H_0: $\sigma^2 = \sigma_0^2$ in favor of the alternative hypothesis H_1: $\sigma^2 < \sigma_0^2$.

  If the test statistic is greater than the critical value $\chi^2_{(\alpha/2),\,n-1}$, you reject the null hypothesis H_0: $\sigma^2 = \sigma_0^2$ in favor of the alternative hypothesis H_1: $\sigma^2 > \sigma_0^2$. Otherwise, you don't reject the null hypothesis.

As an example of the complete process used to test hypotheses about the population variance, suppose that an investor chooses a sample of 30 stocks from her portfolio. She calculates the standard deviation of the returns on these stocks (that is, their *volatility*) to be 23 percent on an annual basis. The investor wants to know whether the volatility of the entire portfolio is less than 25 percent on an annual basis at the 5 percent level of significance. (A volatility of 25 percent [0.25] translates into a variance of $0.25^2 = 0.0625$.) So the null hypothesis is $H_0 : \sigma^2 = 0.0625$, and alternative hypotheses is $H_1 : \sigma^2 < 0.0625$.

Because the investor wants to know only whether the variance is less than 0.0625, you use a left-tailed test. The level of significance is $\alpha = 0.05$.

With a sample size of 30 and a sample variance of 0.23, the test statistic is

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

$$= \frac{(30-1)(0.23)^2}{(0.25)^2}$$

$$= 24.546$$

Because this is a left-tailed test with $\alpha = 0.05$ and sample size = 30, the number of degrees of freedom = 29 (30 – 1). The critical value is, therefore, $\chi^2_{(1-\alpha),\,n-1} = \chi^2_{0.95,\,29}$.

You can find the result in a chi-square table, such as Table 14-2.

Table 14-2	**The Chi-Square Table with Larger Numbers of Degrees of Freedom**							
df\Right-Tail Area	*0.99*	*0.975*	*0.95*	*0.90*	*0.10*	*0.05*	*0.025*	*0.01*
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892

You find the critical value in the row corresponding to 29 degrees of freedom ($n - 1 = 30 - 1 = 29$) and the column with a right-tail area of 0.095 ($[1 - \alpha] = [1 - 0.05] = 0.095$). The result is 17.708.

To reject this hypothesis, the test statistic must be less than the critical value. In this case, the critical value is 24.55, and the test statistic is 17.708; therefore, the null hypothesis isn't rejected. There isn't enough evidence to conclude that the portfolio volatility is less than 25 percent.

Practicing the Goodness of Fit Tests

One of the most important applications of the chi-square distribution is to test whether a population conforms to a specific probability distribution. This type of test is called a *goodness of fit test*.

In this section, I show you examples of how to use sample data from a population to determine whether the population follows the Poisson distribution (covered in Chapter 8) or the normal distribution (discussed in Chapter 9). Note that these aren't the only possible applications of goodness of fit tests; in principle, you can compare any population to any probability distribution.

Comparing a population to the Poisson distribution

You use the Poisson distribution to describe the distribution of events occurring over a given interval of time. To test the hypothesis that a population follows the Poisson distribution, you express the null and alternative hypotheses as follows:

- ✔ H_0: The population follows the Poisson distribution.
- ✔ H_1: The population doesn't follow the Poisson distribution.

Alternatively, the null and alternative hypotheses may include an assumption about the parameter λ, which represents the expected number of events that occur during a given time frame.

For example, the null and alternative hypotheses could be

- ✔ H_0: The population follows the Poisson distribution with $\lambda = 1$.
- ✔ H_1: The population doesn't follow the Poisson distribution with $\lambda = 1$.

Use this approach if you have reason to believe that the value of $\lambda = 1$. In this case, the interpretation of the results is slightly different. If the null hypothesis that the population follows the Poisson distribution is rejected, the population actually follows a different distribution. If the null hypothesis that the population follows the Poisson distribution with $\lambda = 1$ is rejected, the population either doesn't follow the Poisson distribution or it follows the Poisson distribution but with a different value of λ.

One of the unusual features of a goodness of fit test is that you always implement the alternative hypothesis as a right-tailed test. Based on the construction of the test statistic, the null hypothesis that a population follows a specified distribution is rejected only if the test statistic is *too large;* therefore, a goodness of fit test is always right-tailed.

And you construct the test statistic in such a way as to see how closely the elements in a sample match up with the assumed probability distribution. To construct the test statistic, you choose sample data and arrange them into categories.

For example, suppose that a bank manager wants to determine whether the distribution of customers that enters the bank during lunch hour (12 noon to 1 p.m.) follows the Poisson distribution. This information helps the manager determine the optimal number of tellers to use during this time period.

In this case, the population consists of the number of customers that enter the bank during lunch hour. Suppose that the manager chooses a random sample of 100 lunch hours from the past year and counts the number of customer that enters during each of those 100 hours. He then organizes the results as shown here:

Number of Customers per Hour	Number of Hours
0	9
1	12
2	15
3	20
4	27
5	12
6	5

According to these results, during each hour in the sample, the number of customers ranged from 0 to 6, so the manager organizes the data into a total of seven categories. The number of customers in each category is known as the *observed frequency* of the category. You must compare these numbers with the *expected frequencies* — the number of customers expected if the distribution of customers per hour really does follow the Poisson distribution.

In this example, you can find the expected frequencies for each category by computing the Poisson probabilities for each category and multiplying the result by the sample size. For example, suppose that the probability of three customers entering the bank each hour under the Poisson distribution is 0.2240, indicating that in a sample of 100 hours, the expected number of customers (or the expected frequency) is $0.2240 \times 100 = 22.40$ customers. (Of course, it's impossible for 22.40 customers to show up during lunch hour! This is simply an average.)

After you determine the expected frequency of each category, you compute the test statistic with this formula:

$$\chi^2 = \sum_{j=1}^{k} \frac{\left(O_j - E_j\right)^2}{E_j}$$

REMEMBER

Here, j is an index for the category being tested, k is the number of total categories, O_j is the observed frequency in category j, and E_j is the expected frequency in category j.

The closer the observed frequencies are to the expected frequencies, the smaller the value of the test statistic. A small value for this statistic indicates that the null hypothesis (which states that the population follows the Poisson distribution) should *not* be rejected.

Because the goodness of fit test is always right-tailed, it has a single critical value:

$$\chi^2_{\alpha,\,k-1-m}$$

Note that m is a parameter whose value equals 0 if the null hypothesis specifies a value of λ and 1 if the null hypothesis doesn't specify a value of λ.

Unlike hypothesis tests of the population variance, where the appropriate number of degrees of freedom is $n-1$, with a goodness of fit test, the appropriate number of degrees of freedom is $k-1-m$.

When you determine the values of the test statistic and the critical value, the decision rule is to reject the null hypothesis if the test statistic exceeds the critical value; otherwise, don't reject the null hypothesis.

To test the hypothesis that the distribution of customers that enters the bank during lunch hour follows the Poisson distribution, the first step is to specify the null and alternative hypotheses:

✔ H_0: The population follows the Poisson distribution.

✔ H_1: The population doesn't follow the Poisson distribution.

Assume that the level of significance is 0.05 (5 percent).

Before you construct the table of observed and expected frequencies, you must estimate the value of λ from the sample data, because it isn't specified in the null hypothesis. In this case, λ represents the average number of bank customers per hour.

Because each possible number of bank customers is repeated many times in the sample, the average number of bank customers per hour can be computed as a weighted average (see Chapter 3). The formula is

$$\bar{X} = \frac{\sum\limits_{i=1}^{n} w_i X_i}{\sum\limits_{i=1}^{n} w_i}$$

In this formula,

$\bar{X}$ is the sample mean

X_i is a single sample element

w_i is the *weight* associated with element X_i, which equals the number of times that the element appears in the sample

To compute the numerator of this formula, you multiply each number of customers per hour in the sample by the actual number of hours in which this number occurred. This is shown as follows:

Number of Customers per Hour	Number of Hours	Customers per Hour × Number of Hours
0	9	(0)(9) = 0
1	12	(1)(12) = 12
2	15	(2)(15) = 30
3	20	(3)(20) = 60
4	27	(4)(27) = 108
5	12	(5)(12) = 60
6	5	(6)(5) = 30
SUM		300

This results in a sum of 300. The denominator is the sum of the weights:

$$9 + 12 + 15 + 20 + 27 + 12 + 5 = 100$$

The average number of customers is

$$\bar{X} = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i} = \frac{300}{100} = 3$$

You use this result as the value of lambda: $\lambda = 3$.

The next step is to compute the expected frequencies for each category. You find the probability of no customers entering the bank during the next hour when $\lambda = 3$ from the Poisson distribution with this formula:

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

The key terms in this formula are

X = a Poisson random variable

x = number of events (phone calls) that occur

λ = the average number of events that occur per time (hour)

e = a constant equal to approximately 2.71828

! = the "factorial" operator (introduced in Chapter 8)

The factorial operator can only be applied to positive whole numbers and zero. So 0! equals 1, as does 1!, and 2! equals (2)(1) = 2; in other words, 2! equals itself times all smaller positive whole numbers. Based on this pattern,

3! equals (3)(2)(1) = 6, and 4! equals (4)(3)(2)(1) = 24. All remaining factorials are computed in the same way. The factorial operator may be used for several applications; one of these is to count the number of *arrangements* that may be formed from a collection of objects. For example, if three paintings are hung next to each other in the reading room of a library, the number of ways the paintings may be arranged equals 3! = 6.

For the bank customer case, the probability of no customers entering the bank during the lunch hour is computed with the Poisson formula as follows:

$$P(X=0) = e^{-\lambda} \frac{\lambda^x}{x!}$$
$$= e^{-3} \frac{3^0}{0!}$$
$$= 0.0498$$

You do the same calculations with the probabilities for X = 1, X = 2 all the way up to X = 6. The probability that X = 1 is computed as follows:

$$P(X=1) = e^{-\lambda} \frac{\lambda^x}{x!}$$
$$= e^{-3} \frac{3^1}{1!}$$
$$= 0.1494$$

The probability that X = 2 is computed as follows:

$$P(X=2) = e^{-\lambda} \frac{\lambda^x}{x!}$$
$$= e^{-3} \frac{3^2}{2!}$$
$$= 0.2240$$

The probabilities for X = 3, 4, 5, and 6 are computed in a similar manner:

$P(X = 3) = 0.2240$

$P(X = 4) = 0.1680$

$P(X = 5) = 0.1008$

$P(X = 6) = 0.0504$

Because the sample size is 100, you multiply the probabilities by 100 to get the expected frequencies, as shown here:

$X = 0$: expected frequency = 0.0498(100) = 4.98

$X = 1$: expected frequency = 0.1494(100) = 14.94

$X = 2$: expected frequency = 02240(100) = 22.40

X = 3: expected frequency = 0.2240(100) = 22.40

$X = 4$: expected frequency = 0.1680(100) = 16.80

$X = 5$: expected frequency = 0.1008(100) = 10.08

$X = 6$: expected frequency = 0.0504(100) = 5.04

Substitute these values into the test statistic formula:

$$\chi^2 = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j}$$

$$= \frac{(0 - 4.98)^2}{4.98} + \frac{(12 - 14.94)^2}{14.94} + \frac{(30 - 22.40)^2}{22.40} + \frac{(60 - 22.40)^2}{22.40} +$$

$$\frac{(108 - 16.80)^2}{16.80} + \frac{(60 - 10.08)^2}{10.08} + \frac{(30 - 5.04)^2}{5.04}$$

$$= 71.22$$

Then, you determine the critical value as follows:

$$\chi^2_{\alpha,\, k-1-m}$$

The first step is to identify the values of α, k, and m:

✔ $\alpha = 0.05$ because you're using a level of significance of 0.05 (5 percent).

✔ $k = 7$ because there are seven categories (the number of customers that enter the bank during lunch hour is 0, 1, 2, 3, 4, 5, or 6, for a total of 7 possibilities).

✔ $m = 1$ because the null hypothesis doesn't specify a value of λ. (In other words, you computed the value of λ from the sample data.)

Therefore, $k - 1 - m = 7 - 1 - 1 = 5$.

You can find the critical value in Table 14-1 by finding the intersection of the 0.05 right-tail area column and the 5 df row:

$$\chi^2_{\alpha,\, k-1-m} = \chi^2_{0.05,\, 7-1-1} = \chi^2_{0.05,\, 5} = 11.070$$

The test statistic doesn't exceed the critical value. Because this is a right-tailed test, the correct conclusion is that the null hypothesis can't be rejected. In other words, the number of cellphone calls per hour does follow the Poisson distribution.

Comparing a population to the normal distribution

Testing the hypothesis that a population follows the normal distribution is similar to testing the hypothesis that a population follows the Poisson distribution (see the previous section). The two most important differences are that you compute the expected frequencies from the normal distribution, and the definition of m is slightly different for the critical value. In this case, m is defined as follows:

- ✔ $m = 0$ if the value of the mean (μ) and standard deviation (σ) are both specified in the null hypothesis.

- ✔ $m = 1$ if the value of the mean or the standard deviation (but not both) is specified in the null hypothesis.

- ✔ $m = 2$ if the value of neither the mean nor the standard deviation are specified in the null hypothesis.

As an example, suppose that a portfolio manager wants to determine whether the returns to a portfolio are normally distributed, with a mean of 5 percent and a standard deviation of 10 percent.

The observed frequencies are 22 for –15 to –5 percent returns, 29 for –5 to 5 percent returns, 37 for 5 to 15 percent returns, and 12 for 15 to 25 percent returns. The null and alternative hypotheses are

- ✔ H_0: The population is normally distributed with a mean of 5 percent and standard deviation of 10 percent.

- ✔ H_1: The population isn't normally distributed with mean of 5 percent and standard deviation of 10 percent.

Assume that the level of significance is 0.05 (5 percent).

You determine the expected frequencies from the standard normal distribution by following these steps:

1. **Define X to be the return to a portfolio.**

    The mean return is 5 percent and the standard deviation of the return is 10 percent.

2. **Assume that X is normally distributed.**

    To compute probabilities for X using the normal table, you must first convert it into a standard normal random variable (I show you how to do so in Chapter 9).

In this example, the returns are normally distributed with a mean of 5 percent and a standard deviation of 10 percent. Next, you compute the probability that X is between –15 percent and –5 percent.

Because X is a normal random variable but not *standard normal,* you must convert X into the equivalent standard normal form. (Recall that the standard normal distribution has a mean of 0 and a standard deviation of 1, as discussed in Chapter 9.) The appropriate formula is

$$Z = \frac{X - \mu}{\sigma}$$

In this formula,

μ is the mean of X.

σ is the standard deviation of X.

By converting X into a standard normal random variable, it is now possible to compute probabilities for X, using the standard normal tables.

$$P\left(-15 \leq X \leq -5\right) = P\left(\frac{-15 - 5}{10} \leq Z \leq \frac{-5 - 5}{10}\right) = P\left(-2.00 \leq Z \leq -1.00\right)$$

The standard normal tables are set up to compute *cumulative* probabilities; in other words, the probability that Z is less than or equal to a specified value.

In this example, you're looking for the probability that Z is between –2.00 and –1.00. This can be computed from the standard normal tables by rewriting the expression in the equivalent form:

$$P\left(Z \leq -1.00\right) - P\left(Z \leq -2.00\right)$$

You can get these probabilities from the standard normal table. See Table 14-3 for a selection of probabilities associated with negative Z values.

Table 14-3	Selected Standard Normal Probabilities for Negative Z Values			
Z	*0.00*	*0.01*	*0.02*	*0.03*
–2.0	0.0228	0.0222	0.0217	0.0212
–1.5	0.0668	0.0655	0.0643	0.0630
–1.0	0.1587	0.1562	0.1539	0.1515

You find the probability that Z is less than or equal to –1.00 at the intersection of the row for –1.0 under the Z column and the *0.00* column, which is 0.1587.

Likewise, you find the probability that Z is less than or equal to –2.00 at the intersection of the –2.0 row and the *0.00* column, which is 0.0228.

Combining these values gives you $P(Z \leq -1.00) - P(Z \leq -2.00) = 0.1587 - 0.0228 = 0.1359$.

You determine the probability that X is between –5 percent and +5 percent as follows.

$$P(-5 \leq X \leq 5) = P\left(\frac{-5-5}{10} \leq Z \leq \frac{5-5}{10}\right) = P(-1.00 \leq Z \leq 0.00)$$

Algebraically, this is equivalent to

$$P(Z \leq 0.00) - P(Z \leq -1.00)$$

One of the properties of the standard normal distribution is that the probability that Z is less than or equal to 0 is 0.5 because the entire area under the standard normal curve equals 1 and because the distribution is symmetrical about the mean of 0. These statements imply the following:

$$P(Z \leq 0.00) = 0.5$$

$$P(Z \geq 0.00) = 0.5$$

Based on Table 14-3, the probability that Z is less than or equal to –1.00 = 0.1587. Therefore, $P(Z \leq 0.00) - P(Z \leq -1.00) = 0.5 - 0.1587 = 0.3413$.

You compute the probability that X is between +5 percent and +15 percent as follows:

$$P(5 \leq X \leq 15) = P\left(\frac{5-5}{10} \leq Z \leq \frac{15-5}{10}\right) = P(0.00 \leq Z \leq 1.00)$$

You can rewrite this as $P(Z \leq 1.00) - P(Z \leq 0.00)$.

You can find the probability that Z is less than or equal to 1.00 in the standard normal table. Take a look at Table 14-4 to see a section of this table for positive Z values.

Table 14-4   **Selected Standard Normal Probabilities for Positive Z Values**

Z	0.00	0.01	0.02	0.03
1.0	0.8413	0.8438	0.8461	0.8485
1.5	0.9332	0.9345	0.9357	0.9370
2.0	0.9772	0.9778	0.9783	0.9788

You have already determined that the probability that Z is less than or equal to 0 equals 0.5. You can see the probability that Z is less than or equal to 1.00 by intersecting the row for *1.0* and the *0.00* column, which is 0.8413. Therefore, $P(Z \leq 1.00) - P(Z \leq 0.00) = 0.8413 - 0.5000 = 0.3413$.

You determine the probability that X is between +15 percent and +25 percent in a similar manner:

$$P(15 \leq X \leq 25) = P\left(\frac{15-5}{10} \leq Z \leq \frac{25-5}{10}\right) = P(1.00 \leq Z \leq 2.00) \text{ or}$$

$$P(Z \leq 2.00) - P(Z \leq 1.00) = 0.9772 - 0.8413 = 0.1359$$

Because the sample size equals 100, the expected frequency of each category equals the probability of each category times 100.

$P(-15\% \leq X \leq -5\%): 0.1359(100) = 13.59$

$P(-5\% \leq X \leq 5\%): 0.3413(100) = 34.13$

$P(5\% \leq X \leq 15\%): 0.3413(100) = 34.13$

$P(15\% \leq X \leq 25\%): 0.1359(100) = 13.59$

You can then combine the observed and expected returns into a single table, as Table 14-5 shows.

Table 14-5	Observed and Expected Frequencies			
Returns	*−15% to −5%*	*−5% to 5%*	*5% to 15%*	*15% to 25%*
Observed frequency	22	29	37	12
Expected frequency	13.59	34.13	34.13	13.59

Based on this table, the test statistic is computed as follows:

$$\chi^2 = \sum_{j=1}^{k} \frac{\left(O_j - E_j\right)^2}{E_j}$$

$$= \frac{\left(22 - 13.59\right)^2}{13.59} + \frac{\left(29 - 34.13\right)^2}{34.13} + \frac{\left(37 - 34.13\right)^2}{34.13} + \frac{\left(12 - 13.59\right)^2}{13.59}$$

$$= 6.40$$

The critical value is determined as follows:

$$\chi^2_{\alpha,\, k-1-m}$$

The first step is to identify the values of α, λ, and μ.

- ✔ α = 0.05 because you're using a level of significance of 0.05 (5 percent).
- ✔ λ = 4 because there are four categories of returns: –15 percent to –5 percent, –5 percent to +5 percent, +5 percent to +15 percent, and +15 percent to +25 percent.
- ✔ *m* = 0, because the value of the mean (μ) and standard deviation (σ) are both specified in the null hypothesis.

Therefore, λ – 1 – μ = 3.

You can find the critical value in Table 14-1 by finding the intersection of the 0.05 right-tail area column and the 3 df row:

$$\chi^2_{\alpha,\, k-1-m} = \chi_{0.05,\, 3} = 7.815$$

Because this is a right-tailed test, the test statistic must exceed the critical value to reject the null hypothesis that the population is normal with a mean of 5 percent and a standard deviation of 10 percent. Because the test statistic is 6.40 and the critical value is 7.815, you don't reject the null hypothesis. This indicates that the population is, in fact, normally distributed with a mean of 5 percent and a standard deviation of 10 percent.

Testing Hypotheses about the Equality of Two Population Variances

Hypothesis testing for the equality of two population variances is based on the F-distribution (covered in Chapter 13). One of the unique features of the F-distribution is that it's characterized by two types of degrees of freedom, known as *numerator* degrees of freedom and *denominator* degrees of freedom.

The degrees of freedom are called numerator and denominator because an F random variable is actually the ratio of two chi-square random variables, each of which has its own number of degrees of freedom. This is shown in the following equation:

$$F = \frac{\chi^2_1 / \upsilon_1}{\chi^2_2 / \upsilon_2}$$

In this expression,

χ_1^2, χ_2^2 = two chi-square random variables

$\upsilon_1, \upsilon_2,$ = the degrees of freedom corresponding to χ_1^2 and χ_2^2

v_1 = the *numerator* degrees of freedom of F

v_2 = the *denominator* degrees of freedom of F

Just like the chi-square distribution, discussed earlier in this chapter, the F-distribution isn't defined for negative values and is skewed to the right.

The basic six-step process you use to test hypotheses about the equality of two population variances is the same as for testing hypotheses about a single population variance (which I explain in detail in the section "Testing Hypotheses about the Population Variance"). The main differences are the form of the null and alternative hypotheses and the calculation of the test statistic and critical values, which are based on the F-distribution instead of the chi-square distribution.

In the following sections, I walk you through testing hypotheses for two population variances.

The null hypothesis: Equal variances

The first step in the hypothesis testing procedure is writing the *null hypothesis,* which is a statement that's assumed to be true unless strong contrary evidence exists against it.

In this case, the null hypothesis is written as follows:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

σ_1^2 is the variance of population 1, and σ_2^2 is the variance of population 2.

The null hypothesis is that the two population variances are equal. This is accepted unless strong evidence indicates otherwise.

The alternative hypothesis: Unequal variances

The alternative hypothesis is a statement of what you will accept to be true if the null hypothesis is rejected. The alternative hypothesis can take one of three forms:

✔ **Right-tailed test:** You use a right-tailed test if you're interested only in knowing whether the variance of population 1 is greater than the variance of population 2. In this case, the alternative hypothesis is

$$H_1 : \sigma_1^2 > \sigma_2^2$$

✔ **Left-tailed test:** You use a left-tailed test if you're interested only in knowing whether the variance of population 1 is less than the variance of population 2. In this case, the alternative hypothesis is

$$H_1 : \sigma_1^2 < \sigma_2^2$$

✔ **Two-tailed test:** You use a two-tailed test to determine whether the variances of population 1 and 2 are different. In this case, the alternative hypothesis is

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

The test statistic

For testing hypotheses about the equality of two population variances, the appropriate test statistic is

$$F = \frac{s_1^2}{s_2^2}$$

Here, F indicates that the test statistic follows the F-distribution, s_1^2 is the variance of the sample drawn from population 1, and s_2^2 is the variance of the sample drawn from population 2. Note that the test statistic requires that s_1^2 be greater than or equal to s_2^2.

The critical value(s)

To test a hypothesis, you have to choose a *level of significance.* The level of significance, designated with α, refers to the probability of rejecting the null hypothesis when it's actually true.

To test a hypothesis about the equality of two population variances, you use the following critical values.

Right-tailed test for the F-distribution

A right-tailed test has a single critical value:

$$F_\alpha^{\upsilon_1, \upsilon_2}$$

TIP

υ_1 is the numerator degrees of freedom of the F-distribution and equals $n_1 - 1$, where n_1 is the size of the sample drawn from population 1. υ_2 is the denominator degrees of freedom of the F-distribution and equals $n_2 - 1$, where n_2 is the size of the sample drawn from population 2.

This critical value represents the threshold of the right tail of the F-distribution with υ_{21} and υ_{22} degrees of freedom; the area in the right tail is

α. You can find this critical value in an F-table. Because each critical F-value requires two types of degrees of freedom, it's impossible to show both degrees of freedom and the level of significance together in the same table. Instead, you must dedicate an entire table to a single value of the level of significance. (You can see an excerpt of the F-table for a value of α equal to 0.05 in Table 14-6.)

For example, say you conduct a right-tail test with a level of significance of 0.05 (5 percent). You draw a sample size of 5 from the first population and a sample size of 4 from the second population.

You compute the numerator degrees of freedom by subtracting 1 from the size of the sample drawn from population 1:

$$v_1 = n_1 - 1 = 5 - 1 = 4$$

You find the denominator degrees of freedom by subtracting 1 from the size of the sample drawn from population 2:

$$v_2 = n_2 - 1 = 4 - 1 = 3$$

You can find the appropriate critical value in Table 14-6.

Table 14-6	A Section of the F-Table with $\alpha = 0.05$						
$v_2 \backslash v_1$	**3**	**4**	**5**	**6**	**7**	**8**	**9**
3	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	3.86	3.63	3.48	3.37	3.29	3.23	3.18

The top row represents the numerator degrees of freedom (v_1). The first column represents the denominator degrees of freedom (v_2). In this example, you're looking for a right-tail area of 5 percent with $v_1 = n_1 - 1 = 5 - 1$, which is 4 numerator degrees of freedom, and $v_2 = n_2 - 1 = 4 - 1$, which is 3 denominator degrees of freedom.

You find this critical value at the intersection of the *4* column and the row labeled *3* under the v_2/v_1 heading; it equals 9.12.

Left-tailed test for the F-distribution

A left-tailed test also has a single critical value, represented as

$$F_\alpha^{v_1, v_2}$$

This is a very unusual result. The critical value is the same for a right-tailed or a left-tailed test because the F-distribution is undefined for negative values. Also, the test statistic is set up with the larger sample variance in the numerator. The null hypothesis is rejected when the ratio of the sample variances is substantially greater than 1. The test statistic can't be negative.

Two-tailed test for the F-distribution

A two-tailed test has a *single* critical value:

$$F_{\alpha/2}^{v_1, v_2}$$

The decision about the equality of two population variances

You make the decision whether to reject the null hypothesis by looking at the relationship between the test statistic and the critical value(s). Here, I break down the results of the three alternative hypothesis tests:

- ✔ **Right-tailed test:** If the test statistic is greater than the critical value $F_\alpha^{v_1, v_2}$, you reject the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ in favor of the alternative hypothesis $H_1 : \sigma_1^2 > \sigma_2^2$; otherwise, you don't reject the null hypothesis.

- ✔ **Left-tailed test:** If the test statistic is greater than the critical value $F_\alpha^{v_1, v_2}$, you reject the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ in favor of the alternative hypothesis $H_1 : \sigma_1^2 < \sigma_2^2$; otherwise, you don't reject the null hypothesis.

- ✔ **Two-tailed test:** If the test statistic is greater than the critical value $F_{\alpha/2}^{v_1, v_2}$, you reject the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ in favor of the alternative hypothesis $H_1 : \sigma_1^2 \neq \sigma_2^2$; otherwise, you don't reject the null hypothesis.

As an example, suppose that an investor wants to determine whether two portfolios have the same volatility (that is, standard deviation.) She takes a sample of ten stocks from each portfolio. The sample standard deviation of portfolio 1 is 26 percent, and the sample standard deviation of portfolio 2 is 24 percent.

The null hypothesis is $H_0 : \sigma_1^2 = \sigma_2^2$, and the alternative hypothesis is $H_1 : \sigma_1^2 \neq \sigma_2^2$.

Assume that the level of significance is $\alpha = 0.05$ (5 percent).

The test statistic is

$$F = \frac{s_1^2}{s_2^2}$$

with s_1^2 greater than or equal to s_2^2.

Plugging in the numbers, you get the following result:

$$F = \frac{s_1^2}{s_2^2}$$
$$= \frac{(0.26)^2}{(0.24)^2}$$
$$= 1.174$$

Because this is a two-tailed test with a 5 percent level of significance, with both samples having size 10, the numerator and denominator degrees of freedom both equal 9. The critical value is $= F^{9,9}_{(0.025)}$ (that is, 4.03), as you find from the F-table with $\alpha = 0.025$ (see Table 14-7).

Table 14-7	**A Section of the F Table with $\alpha = 0.025$.**			
$\upsilon_2 \backslash \upsilon_1$	*7*	*8*	*9*	*10*
7	4.99	4.90	4.82	4.76
8	4.53	4.43	4.36	4.30
9	4.20	4.10	4.03	3.96
100	3.95	3.85	3.78	3.72

Because the test statistic is 1.174, which is well below the critical value of 4.03, you don't reject the null hypothesis. The investor concludes that the volatilities of the two portfolios are equal.