The proof of (4.1) is not that difficult, given the properties of normally distributed random variables in Appendix B. Each $\hat{\beta}_j$ can be written as $\hat{\beta}_j = \beta_j + \sum_{i=1}^{n} w_{ij}u_i$, where $w_{ij} = \hat{r}_{ij}/\text{SSR}_j$, $\hat{r}_{ij}$ is the i^{th} residual from the regression of the x_j on all the other independent variables, and SSR_j is the sum of squared residuals from this regression [see equation (3.62)]. Since the w_{ij} depend only on the independent variables, they can be treated as nonrandom. Thus, $\hat{\beta}_j$ is just a linear combination of the errors in the sample, $\{u_i: i = 1, 2, \ldots, n\}$. Under Assumption MLR.6 (and the random sampling Assumption MLR.2), the errors are independent, identically distributed Normal$(0,\sigma^2)$ random variables. An important fact about independent normal random variables is that a linear combination of such random variables is normally distributed (see Appendix B). This basically completes the proof. In Section 3.3, we showed that $\text{E}(\hat{\beta}_j) = \beta_j$, and we derived $\text{Var}(\hat{\beta}_j)$ in Section 3.4; there is no need to re-derive these facts.

EXPLORING FURTHER 4.1

Suppose that u is independent of the explanatory variables, and it takes on the values $-2, -1, 0, 1$, and 2 with equal probability of 1/5. Does this violate the Gauss-Markov assumptions? Does this violate the CLM assumptions?

The second part of this theorem follows immediately from the fact that when we standardize a normal random variable by subtracting off its mean and dividing by its standard deviation, we end up with a standard normal random variable.

The conclusions of Theorem 4.1 can be strengthened. In addition to (4.1), any linear combination of the $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k$ is also normally distributed, and any subset of the $\hat{\beta}_j$ has a *joint* normal distribution. These facts underlie the testing results in the remainder of this chapter. In Chapter 5, we will show that the normality of the OLS estimators is still *approximately* true in large samples even without normality of the errors.

4.2 Testing Hypotheses about a Single Population Parameter: The *t* Test

This section covers the very important topic of testing hypotheses about any single parameter in the population regression function. The population model can be written as

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u, \tag{4.2}$$

and we assume that it satisfies the CLM assumptions. We know that OLS produces unbiased estimators of the β_j. In this section, we study how to test hypotheses about a particular β_j. For a full understanding of hypothesis testing, one must remember that the β_j are unknown features of the population, and we will never know them with certainty. Nevertheless, we can *hypothesize* about the value of β_j and then use statistical inference to test our hypothesis.

In order to construct hypotheses tests, we need the following result:

THEOREM 4.2

***t* DISTRIBUTION FOR THE STANDARDIZED ESTIMATORS**

Under the CLM assumptions MLR.1 through MLR.6,

$$(\hat{\beta}_j - \beta_j)/\text{se}(\hat{\beta}_j) \sim t_{n-k-1} = t_{df}, \tag{4.3}$$

where $k + 1$ is the number of unknown parameters in the population model $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$ (k slope parameters and the intercept β_0) and $n - k - 1$ is the degrees of freedom (df).

This result differs from Theorem 4.1 in some notable respects. Theorem 4.1 showed that, under the CLM assumptions, $(\hat{\beta}_j - \beta_j)/\mathrm{sd}(\hat{\beta}_j) \sim \mathrm{Normal}(0,1)$. The t distribution in (4.3) comes from the fact that the constant σ in $\mathrm{sd}(\hat{\beta}_j)$ has been replaced with the random variable $\hat{\sigma}$. The proof that this leads to a t distribution with $n - k - 1$ degrees of freedom is difficult and not especially instructive. Essentially, the proof shows that (4.3) can be written as the ratio of the standard normal random variable $(\hat{\beta}_j - \beta_j)/\mathrm{sd}(\hat{\beta}_j)$ over the square root of $\hat{\sigma}^2/\sigma^2$. These random variables can be shown to be independent, and $(n - k - 1)\,\hat{\sigma}^2/\sigma^2 \sim \chi^2_{n-k-1}$. The result then follows from the definition of a t random variable (see Section B.5).

Theorem 4.2 is important in that it allows us to test hypotheses involving the β_j. In most applications, our primary interest lies in testing the **null hypothesis**

$$\mathrm{H}_0\colon \beta_j = 0, \tag{4.4}$$

where j corresponds to any of the k independent variables. It is important to understand what (4.4) means and to be able to describe this hypothesis in simple language for a particular application. Since β_j measures the partial effect of x_j on (the expected value of) y, after controlling for all other independent variables, (4.4) means that, once $x_1, x_2, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k$ have been accounted for, x_j has *no effect* on the expected value of y. We cannot state the null hypothesis as "x_j does have a partial effect on y" because this is true for any value of β_j other than zero. Classical testing is suited for testing *simple hypotheses* like (4.4).

As an example, consider the wage equation

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u.$$

The null hypothesis $\mathrm{H}_0\colon \beta_2 = 0$ means that, once education and tenure have been accounted for, the number of years in the workforce (*exper*) has no effect on hourly wage. This is an economically interesting hypothesis. If it is true, it implies that a person's work history prior to the current employment does not affect wage. If $\beta_2 > 0$, then prior work experience contributes to productivity, and hence to wage.

You probably remember from your statistics course the rudiments of hypothesis testing for the mean from a normal population. (This is reviewed in Appendix C.) The mechanics of testing (4.4) in the multiple regression context are very similar. The hard part is obtaining the coefficient estimates, the standard errors, and the critical values, but most of this work is done automatically by econometrics software. Our job is to learn how regression output can be used to test hypotheses of interest.

The statistic we use to test (4.4) (against any alternative) is called "the" ***t* statistic** or "the" ***t* ratio** of $\hat{\beta}_j$ and is defined as

$$t_{\hat{\beta}_j} \equiv \hat{\beta}_j/\mathrm{se}(\hat{\beta}_j). \tag{4.5}$$

We have put "the" in quotation marks because, as we will see shortly, a more general form of the t statistic is needed for testing other hypotheses about β_j. For now, it is important to know that (4.5) is suitable only for testing (4.4). For particular applications, it is helpful to index t statistics using the name of the independent variable; for example, t_{educ} would be the t statistic for $\hat{\beta}_{educ}$.

The t statistic for $\hat{\beta}_j$ is simple to compute given $\hat{\beta}_j$ and its standard error. In fact, most regression packages do the division for you and report the t statistic along with each coefficient and its standard error.

Before discussing how to use (4.5) formally to test H_0: $\beta_j = 0$, it is useful to see why $t_{\hat{\beta}_j}$ has features that make it reasonable as a test statistic to detect $\beta_j \neq 0$. First, since $se(\hat{\beta}_j)$ is always positive, $t_{\hat{\beta}_j}$ has the same sign as $\hat{\beta}_j$: if $\hat{\beta}_j$ is positive, then so is $t_{\hat{\beta}_j}$, and if $\hat{\beta}_j$ is negative, so is $t_{\hat{\beta}_j}$. Second, for a given value of $se(\hat{\beta}_j)$, a larger value of $\hat{\beta}_j$ leads to larger values of $t_{\hat{\beta}_j}$. If $\hat{\beta}_j$ becomes more negative, so does $t_{\hat{\beta}_j}$.

Since we are testing H_0: $\beta_j = 0$, it is only natural to look at our unbiased estimator of β_j, $\hat{\beta}_j$, for guidance. In any interesting application, the point estimate $\hat{\beta}_j$ will *never* exactly be zero, whether or not H_0 is true. The question is: How far is $\hat{\beta}_j$ from zero? A sample value of $\hat{\beta}_j$ very far from zero provides evidence against H_0: $\beta_j = 0$. However, we must recognize that there is a sampling error in our estimate $\hat{\beta}_j$, so the size of $\hat{\beta}_j$ must be weighed against its sampling error. Since the standard error of $\hat{\beta}_j$ is an estimate of the standard deviation of $\hat{\beta}_j$, $t_{\hat{\beta}_j}$ measures how many estimated standard deviations $\hat{\beta}_j$ is away from zero. This is precisely what we do in testing whether the mean of a population is zero, using the standard *t* statistic from introductory statistics. Values of $t_{\hat{\beta}_j}$ sufficiently far from zero will result in a rejection of H_0. The precise rejection rule depends on the alternative hypothesis and the chosen significance level of the test.

Determining a rule for rejecting (4.4) at a given significance level—that is, the probability of rejecting H_0 when it is true—requires knowing the sampling distribution of $t_{\hat{\beta}_j}$ when H_0 is true. From Theorem 4.2, we know this to be t_{n-k-1}. This is the key theoretical result needed for testing (4.4).

Before proceeding, it is important to remember that we are testing hypotheses about the *population* parameters. We are *not* testing hypotheses about the estimates from a particular sample. Thus, it never makes sense to state a null hypothesis as "H_0: $\hat{\beta}_1 = 0$" or, even worse, as "H_0: $.237 = 0$" when the estimate of a parameter is .237 in the sample. We are testing whether the unknown population value, β_1, is zero.

Some treatments of regression analysis define the *t* statistic as the *absolute value* of (4.5), so that the *t* statistic is always positive. This practice has the drawback of making testing against one-sided alternatives clumsy. Throughout this text, the *t* statistic always has the same sign as the corresponding OLS coefficient estimate.

Testing against One-Sided Alternatives

To determine a rule for rejecting H_0, we need to decide on the relevant **alternative hypothesis**. First, consider a **one-sided alternative** of the form

$$H_1: \beta_j > 0. \qquad [4.6]$$

When we state the alternative as in equation (4.6), we are really saying that the null hypothesis is H_0: $\beta_j \leq 0$. For example, if β_j is the coefficient on education in a wage regression, we only care about detecting that β_j is different from zero when β_j is actually positive. You may remember from introductory statistics that the null value that is hardest to reject in favor of (4.6) is $\beta_j = 0$. In other words, if we reject the null $\beta_j = 0$ then we automatically reject $\beta_j < 0$. Therefore, it suffices to act as if we are testing H_0: $\beta_j = 0$ against H_1: $\beta_j > 0$, effectively ignoring $\beta_j < 0$, and that is the approach we take in this book.

How should we choose a rejection rule? We must first decide on a **significance level** ("level" for short) or the probability of rejecting H_0 when it is in fact true. For concreteness, suppose we have decided on a 5% significance level, as this is the most popular choice. Thus, we are willing to mistakenly reject H_0 when it is true 5% of the time. Now, while $t_{\hat{\beta}_j}$ has a t distribution under H_0—so that it has zero mean—under the alternative $\beta_j > 0$, the expected value of $t_{\hat{\beta}_j}$ is positive. Thus, we are looking for a "sufficiently large" positive value of $t_{\hat{\beta}_j}$ in order to reject H_0: $\beta_j = 0$ in favor of H_1: $\beta_j > 0$. Negative values of $t_{\hat{\beta}_j}$ provide no evidence in favor of H_1.

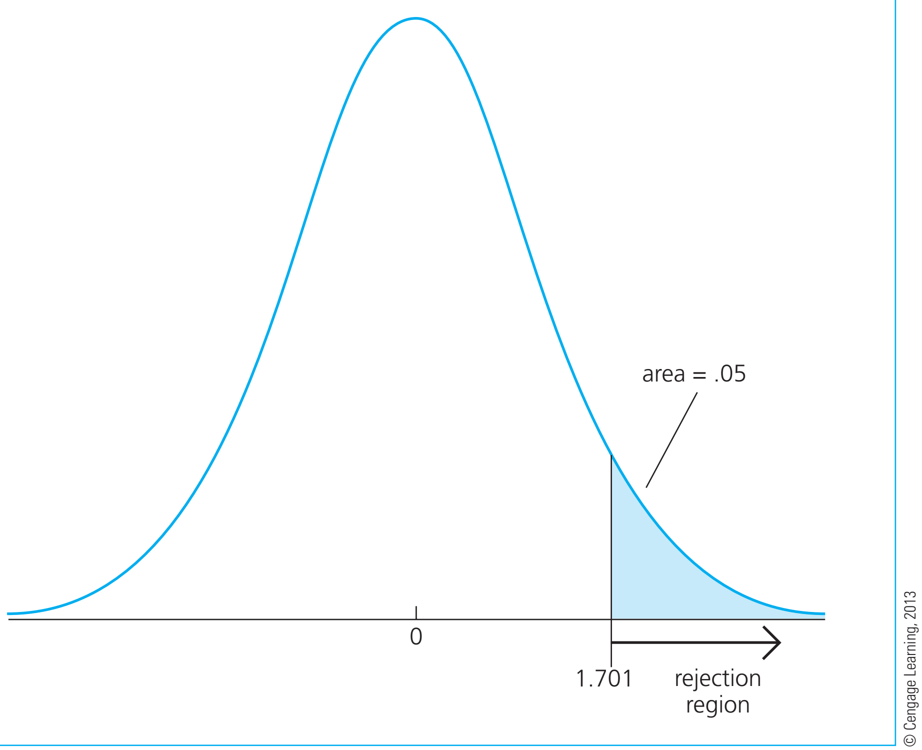
The definition of "sufficiently large," with a 5% significance level, is the 95th percentile in a t distribution with $n - k - 1$ degrees of freedom; denote this by c. In other words, the **rejection rule** is that H_0 is rejected in favor of H_1 at the 5% significance level if

$$t_{\hat{\beta}_j} > c. \qquad [4.7]$$

By our choice of the **critical value** c, rejection of H_0 will occur for 5% of all random samples when H_0 is true.

The rejection rule in (4.7) is an example of a **one-tailed test**. To obtain c, we only need the significance level and the degrees of freedom. For example, for a 5% level test and with $n - k - 1 = 28$ degrees of freedom, the critical value is $c = 1.701$. If $t_{\hat{\beta}_j} \leq 1.701$, then we fail to reject H_0 in favor of (4.6) at the 5% level. Note that a negative value for $t_{\hat{\beta}_j}$, no matter how large in absolute value, leads to a failure in rejecting H_0 in favor of 4.6). (See Figure 4.2.)

FIGURE 4.2 5% rejection rule for the alternative H_1: $\beta_j > 0$ with 28 *df*.



The same procedure can be used with other significance levels. For a 10% level test and if $df = 21$, the critical value is $c = 1.323$. For a 1% significance level and if $df = 21$, $c = 2.518$. All of these critical values are obtained directly from Table G.2. You should note a pattern in the critical values: As the significance level falls, the critical value increases, so that we require a larger and larger value of $t_{\hat{\beta}_j}$ in order to reject H_0. Thus, if H_0 is rejected at, say, the 5% level, then it is automatically rejected at the 10% level as well. It makes no sense to reject the null hypothesis at, say, the 5% level and then to redo the test to determine the outcome at the 10% level.

As the degrees of freedom in the t distribution get large, the t distribution approaches the standard normal distribution. For example, when $n - k - 1 = 120$, the 5% critical value for the one-sided alternative (4.7) is 1.658, compared with the standard normal value of 1.645. These are close enough for practical purposes; for degrees of freedom greater than 120, one can use the standard normal critical values.

EXAMPLE 4.1 HOURLY WAGE EQUATION

Using the data in WAGE1.RAW gives the estimated equation

$$\widehat{\log(wage)} = \underset{(.104)}{.284} + \underset{(.007)}{.092}\, educ + \underset{(.0017)}{.0041}\, exper + \underset{(.003)}{.022}\, tenure$$
$$n = 526,\ R^2 = .316,$$

where standard errors appear in parentheses below the estimated coefficients. We will follow this convention throughout the text. This equation can be used to test whether the return to *exper*, controlling for *educ* and *tenure*, is zero in the population, against the alternative that it is positive. Write this as $H_0\colon \beta_{exper} = 0$ versus $H_1\colon \beta_{exper} > 0$. (In applications, indexing a parameter by its associated variable name is a nice way to label parameters, since the numerical indices that we use in the general model are arbitrary and can cause confusion.) Remember that β_{exper} denotes the unknown population parameter. It is nonsense to write "H_0: .0041 = 0" or "$H_0\colon \hat{\beta}_{exper} = 0$."

Since we have 522 degrees of freedom, we can use the standard normal critical values. The 5% critical value is 1.645, and the 1% critical value is 2.326. The t statistic for $\hat{\beta}_{exper}$ is

$$t_{exper} = .0041/.0017 \approx 2.41,$$

and so $\hat{\beta}_{exper}$, or *exper*, is statistically significant even at the 1% level. We also say that "$\hat{\beta}_{exper}$ is statistically greater than zero at the 1% significance level."

The estimated return for another year of experience, holding tenure and education fixed, is not especially large. For example, adding three more years increases log(*wage*) by 3(.0041) = .0123, so wage is only about 1.2% higher. Nevertheless, we have persuasively shown that the partial effect of experience *is* positive in the population.

The one-sided alternative that the parameter is less than zero,

$$H_1\colon \beta_j < 0, \tag{4.8}$$

also arises in applications. The rejection rule for alternative (4.8) is just the mirror image of the previous case. Now, the critical value comes from the left tail of the t distribution. In practice, it is easiest to think of the rejection rule as

$$t_{\hat{\beta}_j} < -c, \qquad [4.9]$$

where c is the critical value for the alternative H_1: $\beta_j > 0$. For simplicity, we always assume c is positive, since this is how critical values are reported in t tables, and so the critical value $-c$ is a negative number.

For example, if the significance level is 5% and the degrees of freedom is 18, then $c = 1.734$, and so H_0: $\beta_j = 0$ is rejected in favor of H_1: $\beta_j < 0$ at the 5% level if $t_{\hat{\beta}_j} < -1.734$. It is important to remember that, to reject H_0 against the negative alternative (4.8), we must get a negative t statistic. A positive t ratio, no matter how large, provides no evidence in favor of (4.8). The rejection rule is illustrated in Figure 4.3.

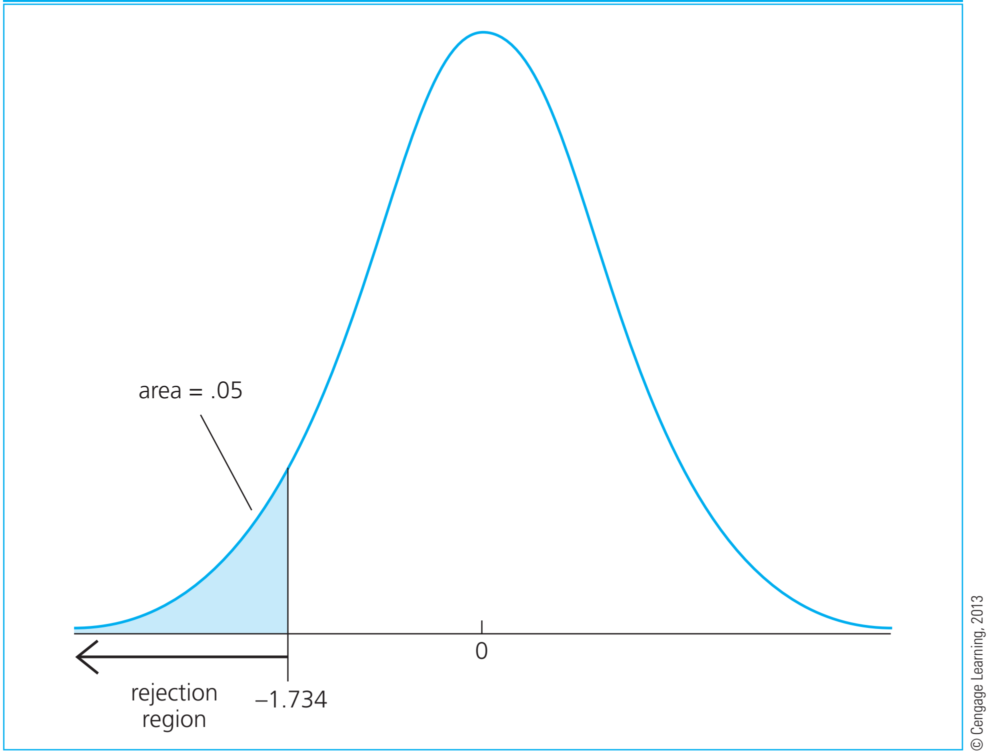
EXPLORING FURTHER 4.2

Let community loan approval rates be determined by

$$apprate = \beta_0 + \beta_1 percmin + \beta_2 avginc + \beta_3 avgwlth + \beta_4 avgdebt + u,$$

where *percmin* is the percentage minority in the community, *avginc* is average income, *avgwlth* is average wealth, and *avgdebt* is some measure of average debt obligations. How do you state the null hypothesis that there is *no* difference in loan rates across neighborhoods due to racial and ethnic composition, when average income, average wealth, and average debt have been controlled for? How do you state the alternative that there is discrimination against minorities in loan approval rates?

FIGURE 4.3 5% rejection rule for the alternative H_1: $\beta_j < 0$ with 18 *df*.



EXAMPLE 4.2 STUDENT PERFORMANCE AND SCHOOL SIZE

There is much interest in the effect of school size on student performance. (See, for example, *The New York Times Magazine*, 5/28/95.) One claim is that, everything else being equal, students at smaller schools fare better than those at larger schools. This hypothesis is assumed to be true even after accounting for differences in class sizes across schools.

The file MEAP93.RAW contains data on 408 high schools in Michigan for the year 1993. We can use these data to test the null hypothesis that school size has no effect on standardized test scores against the alternative that size has a negative effect. Performance is measured by the percentage of students receiving a passing score on the Michigan Educational Assessment Program (MEAP) standardized tenth-grade math test (*math*10). School size is measured by student enrollment (*enroll*). The null hypothesis is H_0: $\beta_{enroll} = 0$, and the alternative is H_1: $\beta_{enroll} < 0$. For now, we will control for two other factors, average annual teacher compensation (*totcomp*) and the number of staff per one thousand students (*staff*). Teacher compensation is a measure of teacher quality, and staff size is a rough measure of how much attention students receive.

The estimated equation, with standard errors in parentheses, is

$$\begin{aligned}
\widehat{math10} &= \underset{(6.113)}{2.274} + \underset{(.00010)}{.00046}\, totcomp + \underset{(.040)}{.048}\, staff - \underset{(.00022)}{.00020}\, enroll \\
n &= 408,\ R^2 = .0541.
\end{aligned}$$

The coefficient on *enroll*, $-.00020$, is in accordance with the conjecture that larger schools hamper performance: higher enrollment leads to a lower percentage of students with a passing tenth-grade math score. (The coefficients on *totcomp* and *staff* also have the signs we expect.) The fact that *enroll* has an estimated coefficient different from zero could just be due to sampling error; to be convinced of an effect, we need to conduct a t test.

Since $n - k - 1 = 408 - 4 = 404$, we use the standard normal critical value. At the 5% level, the critical value is -1.65; the t statistic on *enroll* must be *less* than -1.65 to reject H_0 at the 5% level.

The t statistic on *enroll* is $-.00020/.00022 \approx -.91$, which is larger than -1.65: we *fail* to reject H_0 in favor of H_1 at the 5% level. In fact, the 15% critical value is -1.04, and since $-.91 > -1.04$, we fail to reject H_0 even at the 15% level. We conclude that *enroll* is not statistically significant at the 15% level.

The variable *totcomp* is statistically significant even at the 1% significance level because its t statistic is 4.6. On the other hand, the t statistic for *staff* is 1.2, and so we cannot reject H_0: $\beta_{staff} = 0$ against H_1: $\beta_{staff} > 0$ even at the 10% significance level. (The critical value is $c = 1.28$ from the standard normal distribution.)

To illustrate how changing functional form can affect our conclusions, we also estimate the model with all independent variables in logarithmic form. This allows, for example, the school size effect to diminish as school size increases. The estimated equation is

$$\begin{aligned}
\widehat{math10} &= -\underset{(48.70)}{207.66} + \underset{(4.06)}{21.16}\log(totcomp) + \underset{(4.19)}{3.98}\log(staff) - \underset{(0.69)}{1.29}\log(enroll) \\
n &= 408,\ R^2 = .0654.
\end{aligned}$$

The t statistic on log(*enroll*) is about -1.87; since this is below the 5% critical value -1.65, we reject H_0: $\beta_{\log(enroll)} = 0$ in favor of H_1: $\beta_{\log(enroll)} < 0$ at the 5% level.

In Chapter 2, we encountered a model where the dependent variable appeared in its original form (called *level* form), while the independent variable appeared in log form (called *level-log* model). The interpretation of the parameters is the same in the multiple regression context, except, of course, that we can give the parameters a ceteris paribus interpretation. Holding *totcomp* and *staff* fixed, we have $\Delta\widehat{math10} = -1.29[\Delta\log(enroll)]$, so that

$$\Delta\widehat{math10} \approx -(1.29/100)(\%\Delta enroll) \approx -.013(\%\Delta enroll).$$

Once again, we have used the fact that the change in $\log(enroll)$, when multiplied by 100, is approximately the percentage change in *enroll*. Thus, if enrollment is 10% higher at a school, $\widehat{math10}$ is predicted to be $.013(10) = 0.13$ percentage points lower (*math*10 is measured as a percentage).

Which model do we prefer: the one using the level of *enroll* or the one using $\log(enroll)$? In the level-level model, enrollment does not have a statistically significant effect, but in the level-log model it does. This translates into a higher R-squared for the level-log model, which means we explain more of the variation in *math*10 by using *enroll* in logarithmic form (6.5% to 5.4%). The level-log model is preferred because it more closely captures the relationship between *math*10 and *enroll*. We will say more about using R-squared to choose functional form in Chapter 6.

Two-Sided Alternatives

In applications, it is common to test the null hypothesis H_0: $\beta_j = 0$ against a **two-sided alternative**; that is,

$$H_1: \beta_j \neq 0. \quad [4.10]$$

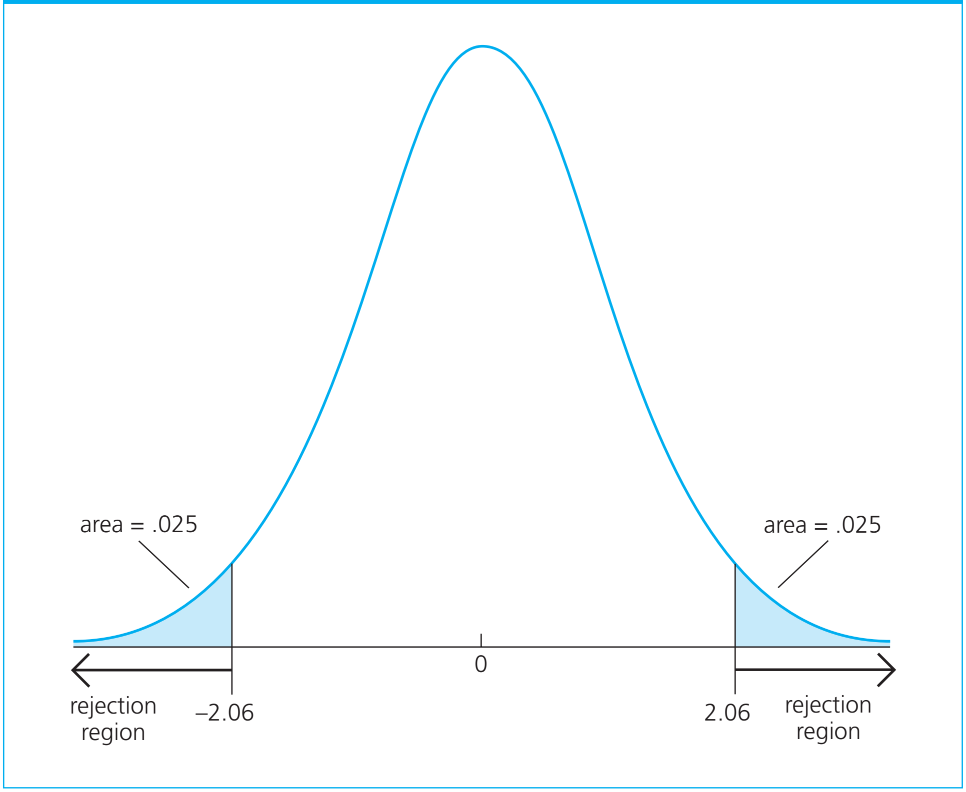
Under this alternative, x_j has a ceteris paribus effect on y without specifying whether the effect is positive or negative. This is the relevant alternative when the sign of β_j is not well determined by theory (or common sense). Even when we know whether β_j is positive or negative under the alternative, a two-sided test is often prudent. At a minimum, using a two-sided alternative prevents us from looking at the estimated equation and then basing the alternative on whether $\hat{\beta}_j$ is positive or negative. Using the regression estimates to help us formulate the null or alternative hypotheses is not allowed because classical statistical inference presumes that we state the null and alternative about the population before looking at the data. For example, we should not first estimate the equation relating math performance to enrollment, note that the estimated effect is negative, and then decide the relevant alternative is H_1: $\beta_{enroll} < 0$.

When the alternative is two-sided, we are interested in the *absolute value* of the t statistic. The rejection rule for H_0: $\beta_j = 0$ against (4.10) is

$$|t_{\hat{\beta}_j}| > c, \quad [4.11]$$

where $|\cdot|$ denotes absolute value and c is an appropriately chosen critical value. To find c, we again specify a significance level, say 5%. For a **two-tailed test**, c is chosen to make the area in each tail of the t distribution equal 2.5%. In other words, c is the 97.5th percentile in the t distribution with $n - k - 1$ degrees of freedom. When $n - k - 1 = 25$, the 5% critical value for a two-sided test is $c = 2.060$. Figure 4.4 provides an illustration of this distribution.

FIGURE 4.4 5% rejection rule for the alternative H_1: $\beta_j \neq 0$ with 25 *df*.



When a specific alternative is not stated, it is usually considered to be two-sided. In the remainder of this text, the default will be a two-sided alternative, and 5% will be the default significance level. When carrying out empirical econometric analysis, it is always a good idea to be explicit about the alternative and the significance level. If H_0 is rejected in favor of (4.10) at the 5% level, we usually say that "x_j is **statistically significant**, or statistically different from zero, at the 5% level." If H_0 is not rejected, we say that "x_j is **statistically insignificant** at the 5% level."

EXAMPLE 4.3 DETERMINANTS OF COLLEGE GPA

We use GPA1.RAW to estimate a model explaining college GPA (*colGPA*), with the average number of lectures missed per week (*skipped*) as an additional explanatory variable. The estimated model is

$$\begin{aligned}\widehat{colGPA} &= \underset{(.33)}{1.39} + \underset{(.094)}{.412}\, hsGPA + \underset{(.011)}{.015}\, ACT - \underset{(.026)}{.083}\, skipped \\ n &= 141,\ R^2 = .234.\end{aligned}$$

We can easily compute t statistics to see which variables are statistically significant, using a two-sided alternative in each case. The 5% critical value is about 1.96, since the degrees of freedom ($141 - 4 = 137$) is large enough to use the standard normal approximation. The 1% critical value is about 2.58.

The t statistic on *hsGPA* is 4.38, which is significant at very small significance levels. Thus, we say that "*hsGPA* is statistically significant at any *conventional* significance level." The t statistic on *ACT* is 1.36, which is not statistically significant at the 10% level against a two-sided alternative. The coefficient on *ACT* is also practically small: a 10-point increase in *ACT*, which is large, is predicted to increase *colGPA* by only .15 points. Thus, the variable *ACT* is practically, as well as statistically, insignificant.

The coefficient on *skipped* has a t statistic of $-.083/.026 = -3.19$, so *skipped* is statistically significant at the 1% significance level ($3.19 > 2.58$). This coefficient means that another lecture missed per week lowers predicted *colGPA* by about .083. Thus, holding *hsGPA* and *ACT* fixed, the predicted difference in *colGPA* between a student who misses no lectures per week and a student who misses five lectures per week is about .42. Remember that this says nothing about specific students; rather, .42 is the estimated average across a subpopulation of students.

In this example, for each variable in the model, we could argue that a one-sided alternative is appropriate. The variables *hsGPA* and *skipped* are very significant using a two-tailed test and have the signs that we expect, so there is no reason to do a one-tailed test. On the other hand, against a one-sided alternative ($\beta_3 > 0$), *ACT* is significant at the 10% level but not at the 5% level. This does not change the fact that the coefficient on *ACT* is pretty small.

Testing Other Hypotheses about β_j

Although H_0: $\beta_j = 0$ is the most common hypothesis, we sometimes want to test whether β_j is equal to some other given constant. Two common examples are $\beta_j = 1$ and $\beta_j = -1$. Generally, if the null is stated as

$$H_0: \beta_j = a_j, \tag{4.12}$$

where a_j is our hypothesized value of β_j, then the appropriate t statistic is

$$t = (\hat{\beta}_j - a_j)/\mathrm{se}(\hat{\beta}_j).$$

As before, t measures how many estimated standard deviations $\hat{\beta}_j$ is away from the hypothesized value of β_j. The general t statistic is usefully written as

$$t = \frac{(estimate - hypothesized\ value)}{standard\ error}. \tag{4.13}$$

Under (4.12), this t statistic is distributed as t_{n-k-1} from Theorem 4.2. The usual t statistic is obtained when $a_j = 0$.

We can use the general t statistic to test against one-sided or two-sided alternatives. For example, if the null and alternative hypotheses are H_0: $\beta_j = 1$ and H_1: $\beta_j > 1$, then we find the critical value for a one-sided alternative *exactly* as before: the difference is in how we compute the t statistic, not in how we obtain the appropriate c. We reject H_0 in favor of H_1 if $t > c$. In this case, we would say that "$\hat{\beta}_j$ is statistically greater than one" at the appropriate significance level.

EXAMPLE 4.4 CAMPUS CRIME AND ENROLLMENT

Consider a simple model relating the annual number of crimes on college campuses (*crime*) to student enrollment (*enroll*):

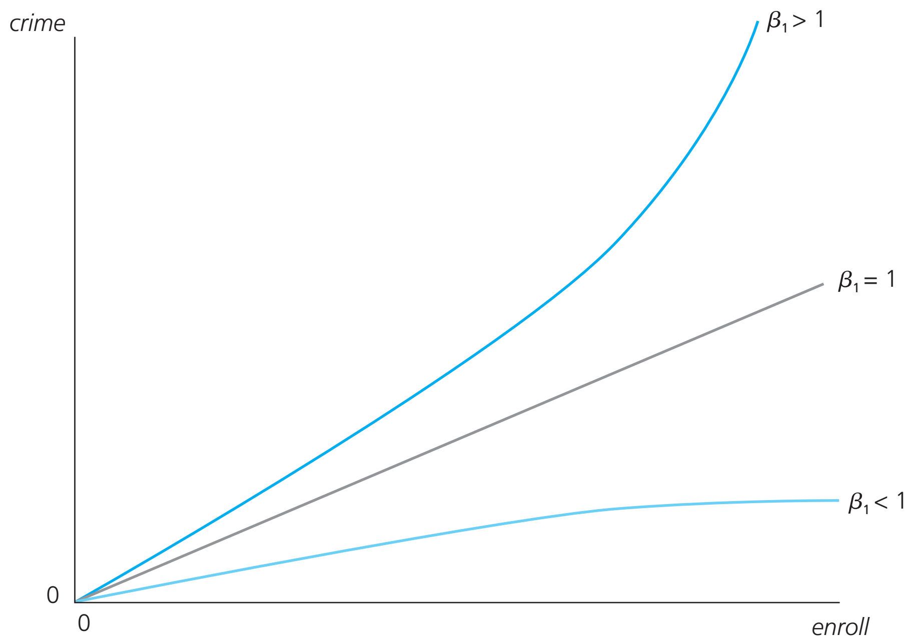
$$\log(crime) = \beta_0 + \beta_1 \log(enroll) + u.$$

This is a constant elasticity model, where β_1 is the elasticity of crime with respect to enrollment. It is not much use to test H_0: $\beta_1 = 0$, as we expect the total number of crimes to increase as the size of the campus increases. A more interesting hypothesis to test would be that the elasticity of crime with respect to enrollment is one: H_0: $\beta_1 = 1$. This means that a 1% increase in enrollment leads to, on average, a 1% increase in crime. A noteworthy alternative is H_1: $\beta_1 > 1$, which implies that a 1% increase in enrollment increases campus crime by *more* than 1%. If $\beta_1 > 1$, then, in a relative sense—not just an absolute sense—crime is more of a problem on larger campuses. One way to see this is to take the exponential of the equation:

$$crime = \exp(\beta_0) enroll^{\beta_1} \exp(u).$$

(See Appendix A for properties of the natural logarithm and exponential functions.) For $\beta_0 = 0$ and $u = 0$, this equation is graphed in Figure 4.5 for $\beta_1 < 1$, $\beta_1 = 1$, and $\beta_1 > 1$.

FIGURE 4.5 Graph of $crime = enroll^{\beta_1}$ for $\beta_1 < 1$, $\beta_1 = 1$, and $\beta_1 > 1$.



We test $\beta_1 = 1$ against $\beta_1 > 1$ using data on 97 colleges and universities in the United States for the year 1992, contained in the data file CAMPUS.RAW. The data come from the FBI's *Uniform Crime Reports*, and the average number of campus crimes in the sample is about 394, while the average enrollment is about 16,076. The estimated equation (with estimates and standard errors rounded to two decimal places) is

$$\begin{aligned} \widehat{\log(crime)} &= -6.63 + 1.27 \log(enroll) \\ &\quad\ \ (1.03) \quad (0.11) \\ n &= 97, R^2 = .585. \end{aligned} \qquad [4.14]$$

The estimated elasticity of *crime* with respect to *enroll*, 1.27, is in the direction of the alternative $\beta_1 > 1$. But is there enough evidence to conclude that $\beta_1 > 1$? We need to be careful in testing this hypothesis, especially because the statistical output of standard regression packages is much more complex than the simplified output reported in equation (4.14). Our first instinct might be to construct "the" t statistic by taking the coefficient on log(*enroll*) and dividing it by its standard error, which is the t statistic reported by a regression package. But this is the *wrong* statistic for testing H_0: $\beta_1 = 1$. The correct t statistic is obtained from (4.13): we subtract the hypothesized value, unity, from the estimate and divide the result by the standard error of $\hat{\beta}_1$: $t = (1.27 - 1)/.11 = .27/.11 \approx 2.45$. The one-sided 5% critical value for a t distribution with $97 - 2 = 95$ df is about 1.66 (using $df = 120$), so we clearly reject $\beta_1 = 1$ in favor of $\beta_1 > 1$ at the 5% level. In fact, the 1% critical value is about 2.37, and so we reject the null in favor of the alternative at even the 1% level.

We should keep in mind that this analysis holds no other factors constant, so the elasticity of 1.27 is not necessarily a good estimate of ceteris paribus effect. It could be that larger enrollments are correlated with other factors that cause higher crime: larger schools might be located in higher crime areas. We could control for this by collecting data on crime rates in the local city.

For a two-sided alternative, for example H_0: $\beta_j = -1$, H_1: $\beta_j \neq -1$, we still compute the t statistic as in (4.13): $t = (\hat{\beta}_j + 1)/\text{se}(\hat{\beta}_j)$ (notice how subtracting -1 means adding 1). The rejection rule is the usual one for a two-sided test: reject H_0 if $|t| > c$, where c is a two-tailed critical value. If H_0 is rejected, we say that "$\hat{\beta}_j$ is statistically different from negative one" at the appropriate significance level.

EXAMPLE 4.5 HOUSING PRICES AND AIR POLLUTION

For a sample of 506 communities in the Boston area, we estimate a model relating median housing price (*price*) in the community to various community characteristics: *nox* is the amount of nitrogen oxide in the air, in parts per million; *dist* is a weighted distance of the community from five employment centers, in miles; *rooms* is the average number of rooms in houses in the community; and *stratio* is the average student-teacher ratio of schools in the community. The population model is

$$\log(price) = \beta_0 + \beta_1 \log(nox) + \beta_2 \log(dist) + \beta_3 rooms + \beta_4 stratio + u.$$

Thus, β_1 is the elasticity of *price* with respect to *nox*. We wish to test H_0: $\beta_1 = -1$ against the alternative H_1: $\beta_1 \neq -1$. The t statistic for doing this test is $t = (\hat{\beta}_1 + 1)/\text{se}(\hat{\beta}_1)$.

Using the data in HPRICE2.RAW, the estimated model is

$$\widehat{\log(price)} = \underset{(0.32)}{11.08} - \underset{(.117)}{.954}\log(nox) - \underset{(.043)}{.134}\log(dist) + \underset{(.019)}{.255}\,rooms - \underset{(.006)}{.052}\,stratio$$

$$n = 506,\ R^2 = .581.$$

The slope estimates all have the anticipated signs. Each coefficient is statistically different from zero at very small significance levels, including the coefficient on log(nox). But we do not want to test that $\beta_1 = 0$. The null hypothesis of interest is H_0: $\beta_1 = -1$, with corresponding t statistic $(-.954 + 1)/.117 = .393$. There is little need to look in the t table for a critical value when the t statistic is this small: the estimated elasticity is not statistically different from -1 even at very large significance levels. Controlling for the factors we have included, there is little evidence that the elasticity is different from -1.

Computing *p*-Values for *t* Tests

So far, we have talked about how to test hypotheses using a classical approach: after stating the alternative hypothesis, we choose a significance level, which then determines a critical value. Once the critical value has been identified, the value of the t statistic is compared with the critical value, and the null is either rejected or not rejected at the given significance level.

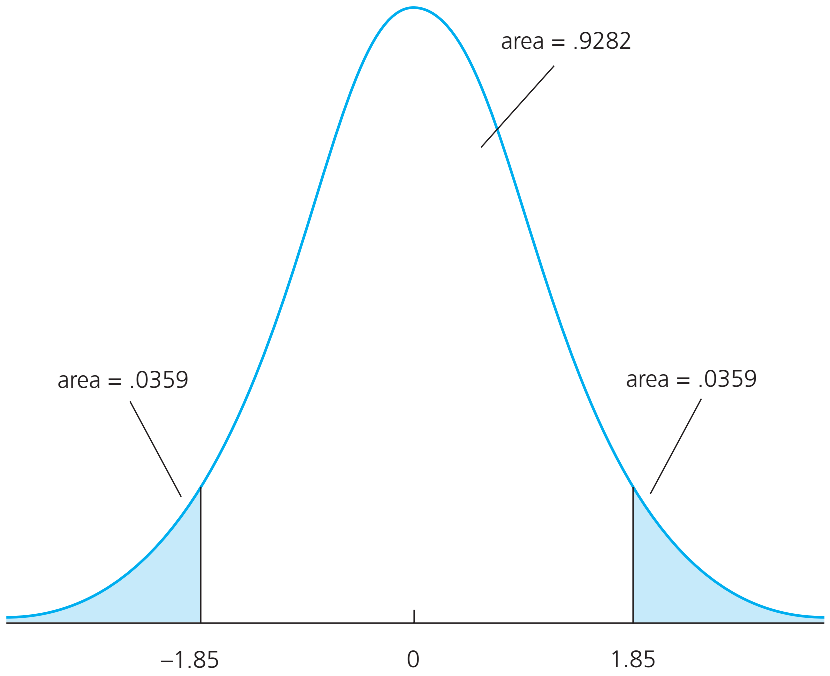
Even after deciding on the appropriate alternative, there is a component of arbitrariness to the classical approach, which results from having to choose a significance level ahead of time. Different researchers prefer different significance levels, depending on the particular application. There is no "correct" significance level.

Committing to a significance level ahead of time can hide useful information about the outcome of a hypothesis test. For example, suppose that we wish to test the null hypothesis that a parameter is zero against a two-sided alternative, and with 40 degrees of freedom we obtain a t statistic equal to 1.85. The null hypothesis is not rejected at the 5% level, since the t statistic is less than the two-tailed critical value of $c = 2.021$. A researcher whose agenda is not to reject the null could simply report this outcome along with the estimate: the null hypothesis is not rejected at the 5% level. Of course, if the t statistic, or the coefficient and its standard error, are reported, then we can also determine that the null hypothesis would be rejected at the 10% level, since the 10% critical value is $c = 1.684$.

Rather than testing at different significance levels, it is more informative to answer the following question: Given the observed value of the t statistic, what is the *smallest* significance level at which the null hypothesis would be rejected? This level is known as the ***p*-value** for the test (see Appendix C). In the previous example, we know the p-value is greater than .05, since the null is not rejected at the 5% level, and we know that the p-value is less than .10, since the null is rejected at the 10% level. We obtain the actual p-value by computing the probability that a t random variable, with 40 df, is larger than 1.85 in absolute value. That is, the p-value is the significance level of the test when we use the value of the test statistic, 1.85 in the above example, as the critical value for the test. This p-value is shown in Figure 4.6.

Because a p-value is a probability, its value is always between zero and one. In order to compute p-values, we either need extremely detailed printed tables of the

FIGURE 4.6 Obtaining the *p*-value against a two-sided alternative, when $t = 1.85$ and $df = 40$.



t distribution—which is not very practical—or a computer program that computes areas under the probability density function of the *t* distribution. Most modern regression packages have this capability. Some packages compute *p*-values routinely with each OLS regression, but only for certain hypotheses. If a regression package reports a *p*-value along with the standard OLS output, it is almost certainly the *p*-value for testing the null hypothesis H_0: $\beta_j = 0$ against the two-sided alternative. The *p*-value in this case is

$$P(|T| > |t|), \tag{4.15}$$

where, for clarity, we let T denote a *t* distributed random variable with $n - k - 1$ degrees of freedom and let *t* denote the numerical value of the test statistic.

The *p*-value nicely summarizes the strength or weakness of the empirical evidence against the null hypothesis. Perhaps its most useful interpretation is the following: the *p*-value is the probability of observing a *t* statistic as extreme as we did *if the null hypothesis is true*. This means that *small* *p*-values are evidence *against* the null; large *p*-values provide little evidence against H_0. For example, if the *p*-value = .50 (reported always as a decimal, not a percentage), then we would observe a value of the *t* statistic as extreme as we did in 50% of all random samples when the null hypothesis is true; this is pretty weak evidence against H_0.

In the example with $df = 40$ and $t = 1.85$, the *p*-value is computed as

$$p\text{-value} = P(|T| > 1.85) = 2P(T > 1.85) = 2(.0359) = .0718,$$

where P($T > 1.85$) is the area to the right of 1.85 in a t distribution with 40 *df*. (This value was computed using the econometrics package Stata; it is not available in Table G.2.) This means that, if the null hypothesis is true, we would observe an absolute value of the t statistic as large as 1.85 about 7.2 percent of the time. This provides some evidence against the null hypothesis, but we would not reject the null at the 5% significance level.

The previous example illustrates that once the p-value has been computed, a classical test can be carried out at any desired level. If α denotes the significance level of the test (in decimal form), then H_0 is rejected if p-value $< \alpha$; otherwise, H_0 is not rejected at the $100 \cdot \alpha\%$ level.

Computing p-values for one-sided alternatives is also quite simple. Suppose, for example, that we test H_0: $\beta_j = 0$ against H_1: $\beta_j > 0$. If $\hat{\beta}_j < 0$, then computing a p-value is not important: we know that the p-value is greater than .50, which will never cause us to reject H_0 in favor of H_1. If $\hat{\beta}_j > 0$, then $t > 0$ and the p-value is just the probability that a t random variable with the appropriate *df* exceeds the value t. Some regression packages only compute p-values for two-sided alternatives. But it is simple to obtain the one-sided p-value: just divide the two-sided p-value by 2.

If the alternative is H_1: $\beta_j < 0$, it makes sense to compute a p-value if $\hat{\beta}_j < 0$ (and hence $t < 0$): p-value $= P(T < t) = P(T > |t|)$ because the t distribution is symmetric about zero. Again, this can be obtained as one-half of the p-value for the two-tailed test.

EXPLORING FURTHER 4.3

Suppose you estimate a regression model and obtain $\hat{\beta}_1 = .56$ and p-value $= .086$ for testing H_0: $\beta_1 = 0$ against H_1: $\beta_1 \neq 0$. What is the p-value for testing H_0: $\beta_1 = 0$ against H_1: $\beta_1 > 0$?

Because you will quickly become familiar with the magnitudes of t statistics that lead to statistical significance, especially for large sample sizes, it is not always crucial to report p-values for t statistics. But it does not hurt to report them. Further, when we discuss F testing in Section 4.5, we will see that it is important to compute p-values, because critical values for F tests are not so easily memorized.

A Reminder on the Language of Classical Hypothesis Testing

When H_0 is not rejected, we prefer to use the language "we fail to reject H_0 at the x% level," rather than "H_0 is accepted at the x% level." We can use Example 4.5 to illustrate why the former statement is preferred. In this example, the estimated elasticity of *price* with respect to *nox* is $-.954$, and the t statistic for testing H_0: $\beta_{nox} = -1$ is $t = .393$; therefore, we cannot reject H_0. But there are many other values for β_{nox} (more than we can count) that cannot be rejected. For example, the t statistic for H_0: $\beta_{nox} = -.9$ is $(-.954 + .9)/.117 = -.462$, and so this null is not rejected either. Clearly $\beta_{nox} = -1$ and $\beta_{nox} = -.9$ cannot both be true, so it makes no sense to say that we "accept" either of these hypotheses. All we can say is that the data do not allow us to reject either of these hypotheses at the 5% significance level.

Economic, or Practical, versus Statistical Significance

Because we have emphasized *statistical significance* throughout this section, now is a good time to remember that we should pay attention to the magnitude of the *coefficient* estimates in addition to the size of the t statistics. The statistical significance of a variable x_j is determined entirely by the size of $t_{\hat{\beta}_j}$, whereas the **economic significance** or **practical significance** of a variable is related to the size (and sign) of $\hat{\beta}_j$.