3 PROPERTIES OF THE REGRESSION COEFFICIENTS AND HYPOTHESIS TESTING

With the aid of regression analysis we can obtain estimates of the parameters of a relationship. However, they are only estimates. The next question to ask is, how reliable are they? We shall answer this first in general terms, investigating the conditions for unbiasedness and the factors governing their variance. Building on this, we shall develop a means of testing whether a regression estimate is compatible with a specific prior hypothesis concerning the true value of a parameter, and hence we shall derive a confidence interval for the true value, that is, the set of all hypothetical values not contradicted by the experimental result. We shall also see how to test whether the goodness of fit of a regression equation is better than might be expected on the basis of pure chance.

3.1 The Random Components of the Regression Coefficients

A least squares regression coefficient is a special form of random variable whose properties depend on those of the disturbance term in the equation. This will be demonstrated first theoretically and then by means of a controlled experiment. In particular, we will investigate the implications for the regression coefficients of certain assumptions concerning the disturbance term.

Throughout the discussion we shall continue to work with the simple regression model where $y$ depends on $x$ according to the relationship

$$y = \alpha + \beta x + u \quad (3.1)$$

and we are fitting the regression equation

$$\hat{y} = a + bx \quad (3.2)$$

given a sample of $n$ observations. We shall also continue to assume that $x$ is a nonstochastic exogenous variable; that is, its value in each observation may be considered to be predetermined by factors unconnected with the present relationship.

First, note that $y$ has two components. It has a nonrandom component ($\alpha + \beta x$), which owes nothing to the laws of chance ($\alpha$ and $\beta$ may be unknown, but nevertheless they are fixed constants), and it has the random component $u$.

This implies that, when we calculate $b$ according to the usual formula
\[ b = \frac{\text{Cov}(x, y)}{\text{Var}(x)} \]  

(3.3)

\( b \) also has a random component. \( \text{Cov}(x, y) \) depends on the values of \( y \), and the values of \( y \) depend on the values of \( u \). If the values of the disturbance term had been different in the \( n \) observations, we would have obtained different values of \( y \), hence of \( \text{Cov}(x, y) \), and hence of \( b \).

We can in theory decompose \( b \) into its nonrandom and random components. In view of (3.1),

\[ \text{Cov}(x, y) = \text{Cov}(x, [\alpha + \beta x + u]) = \text{Cov}(x, \alpha) + \text{Cov}(x, \beta x) + \text{Cov}(x, u) \]  

(3.4)

using Covariance Rule 1 in Section 1.2. By Covariance Rule 3, \( \text{Cov}(x, \alpha) \) must be equal to zero. By Covariance Rule 2, \( \text{Cov}(x, \beta x) \) is equal to \( \beta \text{Cov}(x, x) \). \( \text{Cov}(x, x) \) is the same thing as \( \text{Var}(x) \). Hence we can write

\[ \text{Cov}(x, y) = \beta \text{Var}(x) + \text{Cov}(x, u) \]  

(3.5)

and so

\[ b = \frac{\text{Cov}(x, y)}{\text{Var}(x)} = \beta + \frac{\text{Cov}(x, u)}{\text{Var}(x)} \]  

(3.6)

Thus we have shown that the regression coefficient \( b \) obtained from any sample consists of (1) a fixed component, equal to the true value, \( \beta \), and (2) a random component dependent on \( \text{Cov}(x, u) \), which is responsible for its variations around this central tendency. Similarly, one may easily show that \( a \) has a fixed component equal to the true value, \( \alpha \), plus a random component that depends on the random factor \( u \).

Note that you are not able to make these decompositions in practice because you do not know the true values of \( \alpha \) and \( \beta \) or the actual values of \( u \) in the sample. We are interested in them because they enable us to say something about the theoretical properties of \( a \) and \( b \), given certain assumptions.

### 3.2 A Monte Carlo Experiment

Nobody seems to know for certain how the Monte Carlo experiment got its name. Probably it has something to do with the famous casino, as a symbol of the laws of chance.

The basic concept will be explained by means of an analogy. Suppose you have trained a pig to find truffles for you. These fungi grow wild in the ground in France and Italy and are considered to be delicious. They are expensive because they are hard to find, and a good truffle pig is highly valued. The question is, how do you know if your pig is any good at truffle hunting? It may find them from time to time, but for all you know it may miss a lot as well. If you were really interested you could evaluate your pig by taking a piece of land, burying truffles in several places, letting the pig loose, and seeing how many it located. By means of this controlled experiment, you would have a direct measure
What has this got to do with regression analysis? The problem is that we never know the true values of $\alpha$ and $\beta$ (otherwise, why should we use regression analysis to estimate them?), so we have no means of telling whether the technique is giving us good or bad estimates. A Monte Carlo experiment is an artificial, controlled experiment that allows us to check.

The simplest possible Monte Carlo experiment has three parts. First,

1. you choose the true values of $\alpha$ and $\beta$,
2. you choose the value of $x$ in each observation, and
3. you use some random number generating process to provide the random factor $u$ in each observation.

Second, you *generate* the value of $y$ in each observation, using the relationship (3.1) and the values of $\alpha$, $\beta$, $x$ and $u$. Third, using only the values of $y$ thus generated and the data for $x$, you use regression analysis to obtain estimates $a$ and $b$. You can then see if $a$ is a good estimator of $\alpha$ and if $b$ is a good estimator of $\beta$, and this will give you some idea of whether the regression technique is working properly.

In the first two steps you are preparing a challenge for the regression technique. You are in complete control of the model that you are constructing and you *know* the true values of the parameters because you yourself have determined them. In the third step you see whether the regression technique can meet your challenge and provide good estimates of $\alpha$ and $\beta$ using only the data on $y$ and $x$. Note that the inclusion of a stochastic term in the generation of $y$ is responsible for the element of challenge. If you did not include it, the observations would lie exactly on the straight line (3.1), and it would be a trivial matter to determine the exact values of $\alpha$ and $\beta$ from the data on $y$ and $x$.

Quite arbitrarily, let us put $\alpha$ equal to 2 and $\beta$ equal to 0.5, so the true relationship is

$$y = 2 + 0.5x + u \quad \text{(3.7)}$$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$u$</th>
<th>$y$</th>
<th>$x$</th>
<th>$u$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.59</td>
<td>1.91</td>
<td>11</td>
<td>1.59</td>
<td>9.09</td>
</tr>
<tr>
<td>2</td>
<td>-0.24</td>
<td>2.76</td>
<td>12</td>
<td>-0.92</td>
<td>7.08</td>
</tr>
<tr>
<td>3</td>
<td>-0.83</td>
<td>2.67</td>
<td>13</td>
<td>-0.71</td>
<td>7.79</td>
</tr>
<tr>
<td>4</td>
<td>0.03</td>
<td>4.03</td>
<td>14</td>
<td>-0.25</td>
<td>8.75</td>
</tr>
<tr>
<td>5</td>
<td>-0.38</td>
<td>4.12</td>
<td>15</td>
<td>1.69</td>
<td>11.19</td>
</tr>
<tr>
<td>6</td>
<td>-2.19</td>
<td>2.81</td>
<td>16</td>
<td>0.15</td>
<td>10.15</td>
</tr>
<tr>
<td>7</td>
<td>1.03</td>
<td>6.53</td>
<td>17</td>
<td>0.02</td>
<td>10.52</td>
</tr>
<tr>
<td>8</td>
<td>0.24</td>
<td>6.24</td>
<td>18</td>
<td>-0.11</td>
<td>10.89</td>
</tr>
<tr>
<td>9</td>
<td>2.53</td>
<td>9.03</td>
<td>19</td>
<td>-0.91</td>
<td>10.59</td>
</tr>
<tr>
<td>10</td>
<td>-0.13</td>
<td>6.87</td>
<td>20</td>
<td>1.42</td>
<td>13.42</td>
</tr>
</tbody>
</table>
Table 3.2

<table>
<thead>
<tr>
<th>Sample</th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.63</td>
<td>0.54</td>
</tr>
<tr>
<td>2</td>
<td>2.52</td>
<td>0.48</td>
</tr>
<tr>
<td>3</td>
<td>2.13</td>
<td>0.45</td>
</tr>
<tr>
<td>4</td>
<td>2.14</td>
<td>0.50</td>
</tr>
<tr>
<td>5</td>
<td>1.71</td>
<td>0.56</td>
</tr>
<tr>
<td>6</td>
<td>1.81</td>
<td>0.51</td>
</tr>
<tr>
<td>7</td>
<td>1.72</td>
<td>0.56</td>
</tr>
<tr>
<td>8</td>
<td>3.18</td>
<td>0.41</td>
</tr>
<tr>
<td>9</td>
<td>1.26</td>
<td>0.58</td>
</tr>
<tr>
<td>10</td>
<td>1.94</td>
<td>0.52</td>
</tr>
</tbody>
</table>

To keep things simple, we will assume that we have 20 observations and that the values of $x$ go from 1 to 20. For $u$, the disturbance term, we will use random numbers drawn from a normally distributed population with zero mean and unit variance. We will need a set of 20 and will denote them $r_{m1}$ to $r_{m20}$, $u_1$, the disturbance term in the first observation, is simply equal to $r_{m1}$, $u_2$ to $m_2$, etc.

Given the value of $x$ and $u$ in each observation, it is possible to calculate the value of $y$ using (3.7), and this is done in Table 3.1.

If you now regress $y$ on $x$, you obtain

$$\hat{y} = 1.63 + 0.54x$$ (3.8)

In this case $a$ is an underestimate of $\alpha$ (1.63 as opposed to 2.00) and $b$ is a slight overestimate of $\beta$ (0.54 as opposed to 0.50). The discrepancies are caused by the collective effects of the disturbance terms in the 20 observations.

Of course, one experiment such as this is hardly enough to allow us to evaluate the regression technique. It gave quite good results, but perhaps this was a fluke. To check further, we will repeat the experiment, keeping the same true equation (3.7) and the same values of $x$, but using a new set of random numbers for the disturbance term drawn from the same distribution (zero mean and unit variance). From these, and the values of $x$, we generate a new set of values for $y$.

To save space, the table giving the new values of $u$ and $y$ is omitted. The result when the new values of $y$ are regressed on $x$ is

$$\hat{y} = 2.52 + 0.48x$$ (3.9)

This second experiment also turned out quite well. Now $a$ is an overestimate of $\alpha$ and $b$ is a slight underestimate of $\beta$. Table 3.2 gives the estimates $a$ and $b$ with the experiment repeated 10 times, using a different set of random numbers for the disturbance term in each case.

You can see that, although you sometimes get overestimates and sometimes underestimates, on the whole $a$ and $b$ are clustered around the true values of 2.00 and 0.50, respectively. And there are more good estimates than bad ones. Taking $b$, for example, if you repeated the experiment a very large number of times and constructed a frequency table, you would obtain an approximation to the probability density function shown in Figure 3.1. It is a normal distribution with mean 0.50 and standard deviation 0.0388.
It has been asserted that the discrepancies between the regression coefficients and the true values of the parameters are caused by the disturbance term $u$. A consequence of this is that the bigger is the random element, the more accurate will be the estimate, in general.

This will be illustrated with a second set of Monte Carlo experiments related to the first. We shall use the same values for $\alpha$ and $\beta$ as before, and the same values of $x$, and the same source of random numbers for the disturbance term, but we will now make the disturbance term in each observation, which will be denoted $u'$, equal to twice the random number drawn: $u'_{1}=2r_{1}$, $u'_{2}=2r_{2}$, etc. In fact, we will use exactly the same sample of random numbers as before, but double them. Corresponding to Table 3.1, we now have Table 3.3.

Regressing $y$ on $x$, we now obtain the equation

$$\hat{y} = 1.26 + 0.58x$$ (3.10)

This is much less accurate than its counterpart, equation (3.8).
Table 3.4 gives the results for all 10 experiments, putting \( u' = 2u \). We will call this set of experiments II and the original set, summarized in Table 3.2, I. Comparing Tables 3.2 and 3.4, you can see that the values of \( a \) and \( b \) are much more erratic in the latter, although there is still no systematic tendency either to underestimate or to overestimate.

Detailed inspection reveals an important feature. In Set I, the value of \( b \) in sample 1 was 0.54, an overestimate of 0.04. In Set II, the value of \( b \) in sample 1 was 0.58, an overestimate of 0.08. Exactly twice as much as before. The same is true for each of the other nine samples, and also for the regression coefficient \( a \) in each sample. Doubling the disturbance term in each observation causes a doubling of the errors in the regression coefficients.

This result follows directly from the decomposition of \( b \) given by (3.6). In Set I the error component of \( b \) is given by \( \text{Cov}(x, u)/\text{Var}(x) \). In Set II it is given by \( \text{Cov}(x, u')/\text{Var}(x) \), and

\[
\frac{\text{Cov}(x, u')}{\text{Var}(x)} = \frac{\text{Cov}(x, 2u) - \text{Cov}(x, u)}{\text{Var}(x)} = 2 \frac{\text{Cov}(x, u)}{\text{Var}(x)}
\]

(3.11)

\begin{table}
\centering
\begin{tabular}{llll}
\hline
Sample & \( a \) & \( b \) \\
\hline
1 & 1.26 & 0.58 \\
2 & 3.05 & 0.45 \\
3 & 2.26 & 0.39 \\
4 & 2.28 & 0.50 \\
5 & 1.42 & 0.61 \\
6 & 1.61 & 0.52 \\
7 & 1.44 & 0.63 \\
8 & 4.37 & 0.33 \\
9 & 0.52 & 0.65 \\
10 & 1.88 & 0.55 \\
\hline
\end{tabular}
\end{table}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{b_distributions}
\caption{Distribution of \( b \) when the standard deviation of \( u \) is doubled}
\end{figure}
PROPERTIES OF THE REGRESSION COEFFICIENTS

The increase in inaccuracy is reflected in the probability density function for $b$ in Set II, shown as the solid curve in Figure 3.2. This is still centered over the true value, 0.50, but, if you compare it with that for Set I, the dotted curve, you will see that it is flatter and wider. Doubling the values of $u$ has caused a doubling of the standard deviation of the distribution.

3.3 Assumptions Concerning the Disturbance Term

It is thus obvious that the properties of the regression coefficients depend critically on the properties of the disturbance term. Indeed the latter has to satisfy four conditions, known as the Gauss-Markov conditions, if ordinary least squares regression analysis is to give the best possible results. It is not an exaggeration to say that one of the things that distinguishes a competent user of regression analysis from an incompetent one is an awareness of the importance of these conditions. If they are not satisfied, the user should be aware of the fact. If remedial action is possible, he or she should be capable of taking it. If it is not possible, he or she should be able to judge how seriously the results may have been affected. We shall list the conditions one by one, explaining briefly why they are important. The last three will be treated in detail in later chapters.

**Gauss-Markov Condition 1: $E(u_i) = 0$ for All Observations**

The first condition is that the expected value of the disturbance term in any observations should be zero. Sometimes it will be positive, sometimes negative, but it should not have a systematic tendency in either direction.

Actually, if a constant term is included in the regression equation, it is usually reasonable to assume that this condition is satisfied automatically since the role of the constant term is to pick up any systematic tendency in $y$ not accounted for by the explanatory variables included in the regression equation.

**Gauss-Markov Condition 2: pop.var($u_i$) Constant for All Observations**

The second condition is that the variance of the disturbance term should be constant for all observations. Sometimes the disturbance term will be greater, sometimes smaller, but there should not be any a priori reason for it to be more erratic in some observations than in others. The constant is usually denoted $\sigma_u^2$, often abbreviated to $\sigma^2$, and the condition is written

$$\text{pop.var}(u_i) = \sigma_u^2 \text{ for all } i$$ (3.12)

Since $E(u_i)$ is zero, pop.var($u_i$) is equal to $E(u_i^2)$, so the condition can also be written

$$E(u_i^2) = \sigma_u^2 \text{ for all } i$$ (3.13)
\( \sigma_u \), of course, is unknown. One of the tasks of regression analysis is to estimate the standard deviation of the disturbance term.

If this condition is not satisfied, the OLS regression coefficients will be inefficient, and you should be able to obtain more reliable results by using a modification of the regression technique. This will be discussed in Chapter **.

**Gauss-Markov Condition 3: \( u_i \) distributed independently of \( u_j (i \neq j) \)**

This condition states that there should be no systematic association between the values of the disturbance term in any two observations. For example, just because the disturbance term is large and positive in one observation, there should be no tendency for it to be large and positive in the next (or large and negative, for that matter, or small and positive, or small and negative). The values of the disturbance term should be absolutely independent of one another.

The condition implies that the population covariance between \( u_i \) and \( u_j \) is zero, because

\[
\text{pop.cov}(u_i, u_j) = E[(u_i - \mu_u)(u_j - \mu_u)] = E(u_i u_j) = E(u_i)E(u_j) = 0 \quad (3.14)
\]

(Note that the population means of \( u_i \) and \( u_j \) are zero, by virtue of the first Gauss-Markov condition, and that rewriting \( E(u_i u_j) \) as \( E(u_i)E(u_j) \) requires \( u_i \) and \( u_j \) to be independent – see the Review chapter.)

If this condition is not satisfied, OLS will again give inefficient estimates. Chapter ** discusses the problems that arise and ways of getting around them.

**Gauss-Markov Condition 4: The Disturbance Term Should Be Distributed Independently of the Explanatory Variables**

The final condition comes in two versions, weak and strong. The strong version is that the explanatory variables should be non-stochastic, that is, not have random components. This is actually very unrealistic for economic variables and we will eventually switch to the weak version of the condition, where the explanatory variables are allowed to have random components provided that they are distributed independently of the disturbance term. However for the time being we will use the strong version because it simplifies the analysis of the properties of the estimators.

It is not easy to think of truly nonstochastic variables, other than time, so the following example is a little artificial. Suppose that we are relating earnings to schooling in terms of highest grade completed (HGC). Suppose that we know from the national census that 1% of the population have HGC=8, 3% have HGC=9, 5% have HGC=10, 7% have HGC=11, 43% have HGC=12 (graduation from high school), and so on. Suppose that we have decided to undertake a survey with sample size 1,000 and we want the sample to match the population as far as possible. We might then select what is known as a stratified random sample, designed so that it includes 10 individuals with HGC=8, 30 individuals with HGC=9, and so on. The values of HGC in the sample would then be predetermined and therefore nonstochastic. Schooling and other demographic variables in large surveys drawn in such a way as to be representative of the population as a whole, like the National Longitudinal Survey
of Youth, probably approximate this condition quite well.

If this condition is satisfied, it follows that the population covariance between the explanatory variable and the disturbance term is zero. Since $E(u_i)$ is zero, and the term involving $x$ is nonstochastic,

$$\text{pop.cov}(x_i, u_i) = E[(x_i - \bar{x})(u_i - \mu_u)]$$
$$= (x_i - \bar{x}) E(u_i) = 0 \quad (3.15)$$

Chapters ** and ** discuss two important cases in which this condition is unlikely to be satisfied, and the consequences.

**The Normality Assumption**

In addition to the Gauss-Markov conditions, one usually assumes that the disturbance term is normally distributed. You should know all about the normal distribution from your introductory statistics course. The reason is that if $u$ is normally distributed, so will be the regression coefficients, and this will be useful to us later in the chapter when we come to the business of performing tests of hypotheses and constructing confidence intervals for $\alpha$ and $\beta$ using the regression results.

The justification for the assumption depends on the Central Limit Theorem. In essence, this states that, if a random variable is the composite result of the effects of a large number of other random variables, it will have an approximately normal distribution even if its components do not, provided that none of them is dominant. The disturbance term $u$ is composed of a number of factors not appearing explicitly in the regression equation, so, even if we know nothing about the distribution of these factors (or even their identity), we are entitled to assume that they are normally distributed. At any rate, you are very unlikely to be challenged on this point.

3.4 Unbiasedness of the Regression Coefficients

From (3.6) we can show that $b$ must be an unbiased estimator of $\beta$ if the fourth Gauss-Markov condition is satisfied:

$$E(b) = E\left\{ \beta + \frac{\text{Cov}(x,u)}{\text{Var}(x)} \right\} = \beta + E\left\{ \frac{\text{Cov}(x,u)}{\text{Var}(x)} \right\} \quad (3.16)$$

since $\beta$ is a constant. If we adopt the strong version of the fourth Gauss-Markov condition and assume that $x$ is nonrandom, we may also take $\text{Var}(x)$ as a given constant, and so

$$E(b) = \beta + \frac{1}{\text{Var}(x)} E[\text{Cov}(x,u)] \quad (3.17)$$

We will demonstrate that $E[\text{Cov}(x, u)]$ is zero:
\[ E(\text{Cov}(x, u)) = E\left( \frac{1}{n} \sum (x_i - \bar{x})(u_i - \bar{u}) \right) \]
\[ = \frac{1}{n} \sum E((x_i - \bar{x})(u_i - \bar{u})) \]
\[ = \frac{1}{n} \sum (x_i - \bar{x})E(u_i - \bar{u}) = 0 \] (3.18)

In the second line, the second expected value rule has been used to bring \((1/n)\) out of the expression as a common factor, and the first rule has been used to break up the expectation of the sum into the sum of the expectations. In the third line, the term involving \(x\) has been brought out because \(x\) is nonstochastic. By virtue of the first Gauss-Markov condition, \(E(u_i)\) is zero, and hence \(E(\bar{u})\) is also zero. Therefore

\[ E(b) = \beta \] (3.19)

In other words, \(b\) is an unbiased estimator of \(\beta\). We can obtain the same result with the weak version of the fourth Gauss-Markov condition (allowing \(x\) to have a random component but assuming that it is distributed independently of \(u\)); this is demonstrated in Chapter **.

Unless the random factor in the \(n\) observations happens to cancel out exactly, which can happen only by coincidence, \(b\) will be different from \(\beta\) in any particular experiment, but in view of (3.19) there will be no systematic tendency for it to be either higher or lower. The same is true for the regression coefficient \(a\). Using equation (2.31),

\[ a = \bar{y} - bx \] (3.20)

Hence

\[ E(a) = E(\bar{y}) - \bar{x}E(b) \] (3.21)

Since \(y_i\) is determined by

\[ y_i = \alpha + \beta x_i + u_i \] (3.22)

we have

\[ E(y_i) = \alpha + \beta x_i + E(u_i) \]
\[ = \alpha + \beta x_i \] (3.23)

because \(E(u_i)\) is zero if the first Gauss-Markov condition is satisfied. Hence

\[ E(\bar{y}) = \alpha + \beta \bar{x} \] (3.24)

Substituting this into (3.21), and using the result that \(E(b) = \beta\),

\[ E(a) = (\alpha + \beta \bar{x}) - \bar{x}\beta = \alpha \] (3.25)

Thus \(a\) is an unbiased estimator of \(\alpha\) provided that the Gauss-Markov conditions 1 and 4 are satisfied. Of course in any given sample the random factor will cause a discrepancy.
3.5 Precision of the Regression Coefficients

Now we shall consider the population variances of \( a \) and \( b \) about their population means. These are given by the following expressions (proofs for equivalent expressions can be found in Thomas, 1983, Section 8.3.3):

\[
\text{pop. var}(a) = \frac{\sigma_u^2}{n} \left( 1 + \frac{\bar{x}^2}{\text{Var}(x)} \right) \quad \text{and} \quad \text{pop. var}(b) = \frac{\sigma_u^2}{n \text{Var}(x)}
\]  

(3.26)

Equation (3.26) has three obvious implications. First, the variances of both \( a \) and \( b \) are directly inversely proportional to the number of observations in the sample. This makes good sense. The more information you have, the more accurate your estimates are likely to be.

Second, the variances are proportional to the variance of the disturbance term. The bigger the luck factor, the worse the estimates are likely to be, other things being equal. This is illustrated graphically in Figures 3.3a and 3.3b. In both diagrams the nonstochastic component of the relationship between \( y \) and \( x \), depicted by the dotted line, is given by

\[ y = 3.0 + 0.8x \]  

(3.27)

There are 20 observations, with the values of \( x \) being the integers from 1 to 20. The same random numbers are used to generate the values of the disturbance term, but those in the Figure 3.3b have been multiplied by a factor of 5. As a consequence the regression line, depicted by the solid line, is a much poorer approximation to the nonstochastic relationship in Figure 3.3b than in Figure 3.3a.

Third, the variance of the regression coefficients is inversely related to the variance of \( x \). What is the reason for this? Remember that (1) the regression coefficients are calculated on the assumption that the observed variations in \( y \) are due to variations in \( x \), but (2) they are in reality partly due to variations in \( x \) and partly to variations in \( u \). The smaller the variance of \( x \), the greater is likely to be the relative influence of the luck factor in determining the variations in \( y \) and the more likely is regression analysis to be led astray. This is illustrated by Figures 3.4a and 3.4b. The nonstochastic component of the relationship is given by (3.26), and the disturbance terms are identical. In Figure 3.4a the values of \( x \) are the integers from 1 to 20. In Figure 3.4b, the values of \( x \) are the numbers 9.1, 9.2, ..., 10.9, 11. In Figure 3.4b the effect of the variance of \( x \) on the values of \( y \) is completely overshadowed by the effect of the much larger random variance of \( u \).

Of course, Figures 3.3 and 3.4 make the same point in different ways. As can be seen from (3.26), it is the relative size of \( \sigma_u^2 \) and \( \text{Var}(x) \) that is important, rather than the actual size of either.
Figure 3.3a. Disturbance term with relatively small variance

Figure 3.3b. Disturbance term with relatively large variance
Figure 3.4a. $x$ with relatively large variance

Figure 3.4b. $x$ with relatively small variance
Unfortunately, in practice, one cannot calculate the population variances of either \( a \) or \( b \) because \( \sigma^2_u \) is unknown. However, we can derive an estimator of \( \sigma^2_u \) from the residuals. Clearly the scatter of the residuals around the regression line will reflect the unseen scatter of \( u \) about the line \( y = \alpha + \beta x \), although in general the residual and the value of the disturbance term in any given observation are not equal to one another. Hence the sample variance of the residuals, \( \text{Var}(e) \), which we can measure, will be a guide to \( \sigma^2_u \), which we cannot.

Before going any further, ask yourself the following question. Which line is likely to be closer to the points representing the sample of observations on \( x \) and \( y \), the true line \( y = \alpha + \beta x \) or the regression line \( \hat{y} = a + bx \)? The answer is the regression line, because by definition it is drawn in such a way as to minimize the sum of the squares of the distances between it and the observations. Hence the spread of the residuals will tend to be smaller than the spread of the values of \( u \), and \( \text{Var}(e) \) will tend to underestimate \( \sigma^2_u \). Indeed, it can be shown that the expected value of \( \text{Var}(e) \), when there is just one explanatory variable, is \( [(n-2)/n] \sigma^2_u \). However, it follows that, if one defines \( s^2_u \) by

\[
 s^2_u = \frac{n}{n-2} \text{Var}(e) \tag{3.28}
\]

\( s^2_u \) will be an unbiased estimator of \( \sigma^2_u \) (for a proof, see Thomas).

Using (3.26) and (3.28), one can obtain estimates of the population variances of \( a \) and \( b \) and, by taking square roots, estimates of their standard deviations. Rather than talk about the “estimate of the standard deviation of the probability density function” of a regression coefficient, which is a bit cumbersome, one uses the term “standard error” of a regression coefficient, which in this text will frequently be abbreviated to s.e. For simple regression analysis, therefore, one has

\[
 \text{s.e.}(a) = \sqrt{\frac{s^2_u}{n} \left\{ 1 + \frac{\bar{x}^2}{\text{Var}(x)} \right\}} \quad \text{and} \quad \text{s.e.}(b) = \sqrt{\frac{s^2_u}{n \text{Var}(x)}} \tag{3.29}
\]

The standard errors of the regression coefficient will automatically be calculated for you as part of the computer output.

These relationships will be illustrated with the Monte Carlo experiment described in Section 3.2. In Set I, \( u \) was determined by random numbers drawn from a population with zero mean and unit variance, so \( \sigma^2_u = 1 \). \( x \) was the set of numbers from 1 to 20, and one can easily calculate \( \text{Var}(x) \), which is 33.25. Hence

\[
 \text{pop.var}(a) = \frac{1}{20} \left\{ 1 + \frac{10.5^2}{33.25} \right\} = 0.2158, \tag{3.30}
\]

and

\[
 \text{pop.var}(b) = \frac{1}{20 \times 33.25} = 0.001504. \tag{3.31}
\]
Therefore, the true standard deviation of \( b \) is \( \sqrt{0.001504} = 0.039 \). What did the computer make of it in the 10 samples in Set I? It has to calculate the standard error using (3.29), with the results shown in Table 3.5 in the 10 samples. As you can see, most of the estimates are quite good.

One fundamental point must be emphasised. The standard error gives only a general guide to the likely accuracy of a regression coefficient. It enables you to obtain some idea of the width, or narrowness, of its probability density function as represented in Figure 3.1, but it does not tell you whether your regression estimate comes from the middle of the function, and is therefore accurate, or from the tails, and is therefore relatively inaccurate.

The higher the variance of the disturbance term, the higher the sample variance of the residuals is likely to be, and hence the higher will be the standard errors of the coefficients in the regression equation, reflecting the risk that the coefficients are inaccurate. However, it is only a risk. It is possible that in any particular sample the effects of the disturbance term in the different observations will cancel each other out and the regression coefficients will be accurate after all. The trouble is that in general there is no way of telling whether you happen to be in this fortunate position or not.

**Exercises**

Where performance on a game of skill is measured numerically, the improvement that comes with practice is called a learning curve. This is especially obvious with arcade-type video games. The first time players try a new one, they are likely to score very little. With more attempts, their scores should gradually improve as they become accustomed to the game, although obviously as there will be variations caused by the luck factor. Suppose that their scores are determined by the learning curve

\[
y = 500 + 100x + u,
\]

where \( y \) is the score, \( x \) is the number of times that they have played before, and \( u \) is a disturbance term.

The following table gives the results of the first 20 games of a new player: \( x \) automatically goes from 0 to 19; \( u \) was set equal to 400 times the numbers generated by a normally distributed random variable with zero mean and unit variance; and \( y \) was been determined by \( x \) and \( u \) according to the learning curve.

<table>
<thead>
<tr>
<th>Sample</th>
<th>s.e.(b)</th>
<th>Sample</th>
<th>s.e.(b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.043</td>
<td>6</td>
<td>0.044</td>
</tr>
<tr>
<td>2</td>
<td>0.041</td>
<td>7</td>
<td>0.039</td>
</tr>
<tr>
<td>3</td>
<td>0.038</td>
<td>8</td>
<td>0.040</td>
</tr>
<tr>
<td>4</td>
<td>0.035</td>
<td>9</td>
<td>0.033</td>
</tr>
<tr>
<td>5</td>
<td>0.027</td>
<td>10</td>
<td>0.033</td>
</tr>
</tbody>
</table>
Regressing $y$ on $x$, one obtains the equation (standard errors in parentheses):

$$\hat{y} = 369 + 116.8 \, x$$

(190) (17.1)

3.1 Why is the constant in this equation not equal to 500 and the coefficient of $x$ not equal to 100?

3.2 What is the meaning of the standard errors?

3.3 The experiment is repeated with nine other new players (the disturbance term being generated by 400 times a different set of 20 random numbers in each case), and the regression results for all ten players are shown in the following table. Why do the constant, the coefficient of $x$, and the standard errors vary from sample to sample?

<table>
<thead>
<tr>
<th>Player</th>
<th>Constant</th>
<th>Standard error of constant</th>
<th>Coefficient of $x$</th>
<th>Standard error of coefficient of $x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>369</td>
<td>190</td>
<td>116.8</td>
<td>17.1</td>
</tr>
<tr>
<td>2</td>
<td>699</td>
<td>184</td>
<td>90.1</td>
<td>16.5</td>
</tr>
<tr>
<td>3</td>
<td>531</td>
<td>169</td>
<td>78.5</td>
<td>15.2</td>
</tr>
<tr>
<td>4</td>
<td>555</td>
<td>158</td>
<td>99.5</td>
<td>14.2</td>
</tr>
<tr>
<td>5</td>
<td>407</td>
<td>120</td>
<td>122.6</td>
<td>10.8</td>
</tr>
<tr>
<td>6</td>
<td>427</td>
<td>194</td>
<td>104.3</td>
<td>17.5</td>
</tr>
<tr>
<td>7</td>
<td>412</td>
<td>175</td>
<td>123.8</td>
<td>15.8</td>
</tr>
<tr>
<td>8</td>
<td>613</td>
<td>192</td>
<td>95.8</td>
<td>17.3</td>
</tr>
<tr>
<td>9</td>
<td>234</td>
<td>146</td>
<td>130.1</td>
<td>13.1</td>
</tr>
<tr>
<td>10</td>
<td>485</td>
<td>146</td>
<td>109.6</td>
<td>13.1</td>
</tr>
</tbody>
</table>
3.4 The variance of $x$ is equal to $33.25$ and the population variance of $u$ is equal to $160,000$. Show that the standard deviation of the probability density function of the coefficient of $x$ is equal to $15.5$ using equation (3.29). Are the standard errors in the table good estimates of this standard deviation?

3.6 The Gauss-Markov Theorem

In the Review, we considered estimators of the unknown population mean $\mu$ of a random variable $x$, given a sample of observations. Although we instinctively use the sample mean $\bar{x}$ as our estimator, we saw that it was only one of an infinite number of possible unbiased estimators of $\mu$. The reason that the sample mean is preferred to any other estimator is that, under certain assumptions, it is the most efficient.

Similar considerations apply to regression coefficients. We shall see that the OLS estimators are not the only unbiased estimators of the regression coefficients, but, provided that the Gauss-Markov conditions are satisfied, they are the most efficient. The other side of the coin is that, if the Gauss-Markov conditions are not satisfied, it will in general be possible to find estimators for that are more efficient than OLS.

We will not attempt a general discussion of these issues here. We will instead give an illustration. We shall assume that we have a relationship given by

$$y = \alpha + \beta x + u,$$  \hfill (3.32)

and we shall confine our attention to estimators of $\beta$. Someone who had never heard of regression analysis, on seeing a scatter diagram of a sample of observations, might be tempted to obtain an estimate of the slope merely by joining the first and the last observations, and by dividing the increase in the height by the horizontal distance between them, as in Figure 3.5. The estimator $b$ would then be given by

$$b = \frac{y_n - y_1}{x_n - x_1},$$  \hfill (3.33)

and we shall confine our attention to estimators of $\beta$. Someone who had never heard of regression analysis, on seeing a scatter diagram of a sample of observations, might be tempted to obtain an estimate of the slope merely by joining the first and the last observations, and by dividing the increase in the height by the horizontal distance between them, as in Figure 3.5. The estimator $b$ would then be given by

\begin{align*}
\text{Figure 3.5. Naïve estimation of } b
\end{align*}
What are the properties of this estimator? First, we will investigate whether it is biased or unbiased. Applying (3.32) to the first and last observations, we have

\[ y_1 = \alpha + \beta x_1 + u_1 \]  
\[ y_n = \alpha + \beta x_n + u_n \]

Hence

\[ b = \frac{\beta x_n + u_n - \beta x_1 + u_1}{x_n - x_1} = \beta + \frac{u_n - u_1}{x_n - x_1} \]

Thus we have decomposed this naïve estimator into two components, the true value and an error term. This decomposition is parallel to that for the OLS estimator in Section 3.1, but the error term is different. The expected value of the estimator is given by

\[ E(b) = E(\beta) + E\left(\frac{u_n - u_1}{x_n - x_1}\right) \]

\[ = \beta + \frac{1}{x_n - x_1} E(u_n - u_1) \]

since \( b \) is a constant and \( x_1 \) and \( x_n \) are nonstochastic. If the first Gauss-Markov condition is satisfied,

\[ E(u_n - u_1) = E(u_n) - E(u_1) = 0 \]

Therefore, despite being naïve, this estimator is unbiased.

This is not by any means the only estimator besides OLS that is unbiased. You could derive one by joining any two arbitrarily selected observations, and in fact the possibilities are infinite if you are willing to consider less naïve procedures.

It is intuitively easy to see that we would not prefer a naïve estimator such as (3.33) to OLS. Unlike OLS, which takes account of every observation, it employs only the first and the last and is wasting most of the information in the sample. The naïve estimator will be sensitive to the value of the disturbance term \( u \) in those two observations, whereas the OLS estimator combines all that the values of the disturbance term and takes greater advantage of the possibility that to some extent cancel each other out. More rigorously, it can be shown that the population variance of the naïve estimator is greater than that of the OLS estimator, and that the naïve estimator is therefore less efficient.

With less naive estimators the superior efficiency of OLS may not be so obvious. Nevertheless, provided that the Gauss-Markov conditions for the disturbance term are satisfied, the OLS regression coefficients will be best linear unbiased estimators (BLUE): unbiased, as has already been demonstrated; linear, because they are linear functions of the values of \( y \); and best because they are the most efficient of the class of unbiased linear estimators. This is proved by the Gauss-Markov theorem (for a concise treatment not using matrix algebra, see Thomas, 1983, Section 8.3).
Exercises

3.5 An investigator correctly believes that the relationship between two variables $x$ and $y$ is given by

$$y = \alpha + \beta x + u,$$

Given a sample of $n$ observations, the investigator estimates $\beta$ by calculating it as the average value of $y$ divided by the average value of $x$. Discuss the properties of this estimator. What difference would it make if it could be assumed that $\alpha$ is equal to zero?

3.6* An investigator correctly believes that the relationship between two variables $x$ and $y$ is given by

$$y = \alpha + \beta x + u,$$

Given a sample of observations on $y$, $x$ and a third variable $z$ (which is not a determinant of $y$), the investigator estimates $\beta$ as $\text{Cov}(y, z)/\text{Cov}(x, z)$. Discuss the properties of this estimator. (It can be shown that its population variance is equal to the population variance of the corresponding OLS estimator divided by the square of $r_{x,z}$, where $r_{x,z}$ is the correlation coefficient for $x$ and $z$.)

3.7 Testing Hypotheses Relating to the Regression Coefficients

Which comes first, theoretical hypothesizing or empirical research? There is a bit like asking which came first, the chicken or the egg. In practice, theorizing and experimentation feed on each other, and questions of this type cannot be answered. For this reason, we will approach the topic of hypothesis testing from both directions. On the one hand, we may suppose that the theory has come first and that the purpose of the experiment is to evaluate its plausibility. This will lead to the execution of significance tests. Alternatively, we may perform the experiment first and then consider what theoretical hypotheses would be consistent with the results. This will lead to the construction of confidence intervals.

You will already have encountered the logic underlying significance tests and confidence intervals in an introductory statistics course. You will thus be familiar with most of the concepts in the following applications to regression analysis. There is, however, one topic that may be new: the use of one-tailed tests. Such tests are used very frequently in regression analysis. Indeed, they are, or they ought to be, more common than the traditional textbook two-tailed tests. It is therefore important that you understand the rationale for their use, and this involves a sequence of small analytical steps. None of this should present any difficulty, but be warned that, if you attempt to use a short cut or, worse, try to reduce the whole business to the mechanical use of a few formulae, you will be asking for trouble.
Formulation of a Null Hypothesis

We will start by assuming that the theory precedes the experiment and that you have some hypothetical relationship in your mind. For example, you may believe that the percentage rate of price inflation in an economy, \( p \), depends on the percentage rate of wage inflation, \( w \), according to the linear equation

\[
p = \alpha + \beta w + u \tag{3.39}
\]

where \( \alpha \) and \( \beta \) are parameters and \( u \) is a disturbance term. You might further hypothesize that, apart from the effects of the disturbance term, price inflation is equal to wage inflation. Under these circumstances you would say that the hypothesis that you are going to test, known as your null hypothesis and denoted \( H_0 \), is that \( \beta \) is equal to 1. We also define an alternative hypothesis, denoted \( H_1 \), which represents your conclusion if the experimental test indicates that \( H_0 \) is false. In the present case \( H_1 \) is simply that \( \beta \) is not equal to 1. The two hypotheses are stated using the notation

\[
H_0: \beta = 1 \\
H_1: \beta \neq 1
\]

In this particular case, if we really believe that price inflation is equal to wage inflation, we are trying to establish the credibility of \( H_0 \) by subjecting it to the strictest possible test and hoping that it emerges unscathed. In practice, however, it is more usual to set up a null hypothesis and attack it with the objective of establishing the alternative hypothesis as the correct conclusion. For example, consider the simple earnings function

\[
EARNINGS = \alpha + \beta HGC + u \tag{3.40}
\]

where \( EARNINGS \) is hourly earnings in dollars and \( HGC \) is highest grade completed. On very reasonable theoretical grounds, you expect earnings to be dependent on schooling, but your theory is not strong enough to enable you to specify a particular value for \( \beta \). You can nevertheless establish the dependence of earnings on schooling by the inverse procedure in which you take as your null hypothesis the assertion that earnings does not depend on schooling, that is, that \( \beta \) is zero. Your alternative hypothesis is that \( \beta \) is not equal to zero, that is, that schooling does affect earnings. If you can reject the null hypothesis, you have established the relationship, at least in general terms. Using the conventional notation, your null and alternative hypotheses are \( H_0: \beta = 0 \) and \( H_1: \beta \neq 0 \), respectively.

The following discussion uses the simple regression model

\[
y = \alpha + \beta x + u, \tag{3.41}
\]

It will be confined to the slope coefficient, \( \beta \), but exactly the same procedures are applied to the constant term, \( \alpha \). We will take the general case, where you have defined a null hypothesis that \( \beta \) is equal to some specific value, say \( \beta_0 \), and the alternative hypothesis is that \( \beta \) is not equal to this value (\( H_0: \beta = \beta_0, H_1: \beta \neq \beta_0 \)), and you may be attempting to attack or defend the null hypothesis as it suits your purpose. We will assume that the four Gauss-Markov conditions are satisfied.
Developing the Implications of a Hypothesis

If $H_0$ is correct, estimates of $\beta$ obtained using regression analysis will be distributed with mean $\beta_0$ and variance $\sigma_u^2/[n \text{ Var}(x)]$ (see 3.26). We will now introduce the assumption that $u$ has a normal distribution. If this is the case, $b$ will also be normally distributed, as shown in Figure 3.6. "sd" in the figure refers to the standard deviation of $b$, that is $\sqrt{\sigma_u^2/[n \text{ Var}(x)]}$. In view of the structure of the normal distribution, most estimates of $\beta$ will lie within two standard deviations of $\beta_0$ (if $H_0$: $\beta = \beta_0$ is true).

Initially we will assume that we know the standard deviation of $b$. This is a most unreasonable assumption, and we will drop it later. In practice we have to estimate it, along with $\alpha$ and $\beta$, but it will simplify the discussion if for the time being we suppose that we know it exactly, and hence are in a position to draw Figure 3.6.

We will illustrate this with the price inflation model (3.39). Suppose that for some reason we know that the standard deviation of $b$ is equal to 0.1. Then, if our null hypothesis $H_0$: $\beta = 1$ is correct,
Figure 3.7  Example distribution of b (price inflation/wage inflation model)

regression estimates would be distributed as shown in Figure 3.7. You can see that, provided that the null hypothesis is correct, the estimates will generally lie between 0.8 and 1.2.

Compatibility, Freakiness and the Significance Level

Now we come to the crunch. Suppose that we take an actual sample of observations on average rates of price inflation and wage inflation over the past five years for a sample of OECD countries and estimate $\beta$ using regression analysis. If the estimate is close to 1.0, we should almost certainly be satisfied with the null hypothesis, since it and the sample result are compatible with one another, but suppose, on the other hand, that the estimate is a long way from 1.0. Suppose that it is equal to 0.7. This is three standard deviations below 1.0. The probability of being three standard deviations away from the mean, positive or negative, is only 0.0027, which is very low. You could come to either of two conclusions about this worrisome result:

1. You could continue to maintain that your null hypothesis $\beta = 1$ is correct, and that the experiment has given a freak result. You concede that the probability of such a low value of $b$ is very small, but nevertheless it does occur 0.27 percent of the time and you reckon that this is one of those times.

2. You could conclude that the hypothesis is contradicted by the regression result. You are not convinced by the explanation in (1) because the probability is so small and you think that a much more likely explanation is that $\beta$ is not really equal to 1. In other words, you adopt the alternative hypothesis $H_0: \beta \neq 1$ instead.

How do you decide when to choose (1) and when to choose (2)? Obviously, the smaller the probability of obtaining a regression estimate like the one you have obtained, given your hypothesis, the more likely you are to abandon the hypothesis and choose (2). How small should the probability be before choosing (2)?

There is, and there can be, no definite answer to this question. In most applied work in economics either 5 percent or 1 percent is taken as the critical limit. If 5 percent is taken, the switch to (2) is made when the null hypothesis implies that the probability of obtaining such an extreme value of $b$ is less than 5 percent. The null hypothesis is then said to be rejected at the 5 percent significance level.

This occurs when $b$ is more than 1.96 standard deviations from $\beta_0$. If you look up the normal distribution table, Table A.1 at the end of the text, you will see that the probability of $b$ being more than 1.96 standard deviations above its mean is 2.5 percent, and similarly the probability of it being more than 1.96 standard deviations below its mean is 2.5 percent. The total probability of it being more than 1.96 standard deviations away is thus 5 percent.

We can summarize this decision rule mathematically by saying that we will reject the null hypothesis if
where $Z$ is the number of standard deviations between the regression estimate and the hypothetical value of $\beta$:

$$Z = \frac{\text{distance between regression estimate and hypothetical value}}{\text{standard deviation of } b} = \frac{b - \beta_0}{\text{s.d.}(b)}$$  \hspace{1cm} (3.43)

The null hypothesis will not be rejected if

$$-1.96 \leq Z \leq 1.96$$  \hspace{1cm} (3.44)

This condition can be expressed in terms of $b$ and $\beta_0$ by substituting for $Z$ from (3.43):

$$-1.96 \leq \frac{b - \beta_0}{\text{s.d.}(b)} \leq 1.96$$  \hspace{1cm} (3.45)

Multiplying through by the standard deviation of $b$, one obtains

$$-1.96 \text{s.d.}(b) \leq b - \beta_0 \leq 1.96 \text{s.d.}(b)$$  \hspace{1cm} (3.46)

from which one obtains

$$\beta_0 - 1.96 \text{s.d.}(b) \leq b \leq \beta_0 + 1.96 \text{s.d.}(b)$$  \hspace{1cm} (3.47)

Equation (3.47) gives the set of values of $b$ which will not lead to the rejection of a specific null hypothesis $\beta = \beta_0$. It is known as the acceptance region for $b$, at the 5 percent significance level.

In the case of the price inflation/wage inflation example, where $\text{s.d.}(b)$ is equal to 0.1, you would reject at the 5 percent level if $b$ lies more than 0.196 above or below the hypothetical mean, that is, above 1.196 or below 0.804. The acceptance region is therefore those values of $b$ from 0.804 to 1.196. This is illustrated by the unshaded area in Figure 3.8.
Similarly, the null hypothesis is said to be rejected at the 1 percent significance level if the hypothesis implies that the probability of obtaining such an extreme value of $b$ is less than 1 percent. This occurs when $b$ is more than 2.58 standard deviations above or below the hypothetical value of $\beta$, that is, when

$$Z > 2.58 \text{ or } Z < -2.58 \quad (3.48)$$

Looking at the normal distribution table again, you will see that the probability of $b$ being more than 2.58 standard deviations above its mean is 0.5 percent, and there is the same probability of it being more than 2.58 standard deviations below it, so the combined probability of such an extreme value is 1 percent. In the case of our example, you would reject the null hypothesis $\beta = 1$ if the regression estimate lay above 1.258 or below 0.742.

You may ask, why do people usually report, or at least consider reporting, the results at both the 5 percent and the 1 percent significance levels? Why not just one? The answer is that they are trying to strike a balance between the risks of making Type I errors and Type II errors. A Type I error occurs when you reject a true null hypothesis. A Type II error occurs when you do not reject a false one.

Obviously, the lower your critical probability, the smaller is the risk of a Type I error. If your significance level is 5 percent, you will reject a true hypothesis 5 percent of the time. If it is 1 percent, you will make a Type I error 1 percent of the time. Thus the 1 percent significance level is safer in this respect. If you reject the hypothesis at this level, you are almost certainly right to do so. For this reason the 1 percent significance level is described as higher than the 5 percent.

At the same time, if the null hypothesis happens to be false, the higher your significance level, the wider is your acceptance region, the greater is your chance of not rejecting it, and so the greater is the risk of a Type II error. Thus you are caught between the devil and the deep blue sea. If you insist on a very high significance level, you incur a relatively high risk of a Type II error if the hypothesis happens to be false. If you choose a low significance level, you run a relatively high risk of making a Type I error if the hypothesis happens to be true.
Most people take out a crude form of insurance policy and perform the test at both these levels, being prepared to quote the results of each. Actually, it is frequently superfluous to quote both results explicitly. Since $b$ has to be more extreme for the hypothesis to be rejected at the 1 percent level than at the 5 percent level, if you reject at the 1 percent level it automatically follows that you reject at the 5 percent level, and there is no need to say so. Indeed, you look ignorant if you do. And if you do not reject at the 5 percent level, it automatically follows that you will not reject at the 1 percent level, and again you would look ignorant if you said so. The only time when you should quote both results is when you reject the null hypothesis at the 5 percent level but not at the 1 percent level.

What Happens if the Standard Deviation of $b$ is Not Known

So far we have assumed that the standard deviation of $b$ is known, which is most unlikely in practice. It has to be estimated by the standard error of $b$, given by (3.29). This causes two modifications to the test procedure. First, $Z$ is now defined using $s.e.(b)$ instead of $s.d.(b)$, and it is referred to as the $t$ statistic:

$$ t = \frac{b - \beta_0}{s.e.(b)} $$

(3.49)

Second, the critical levels of $t$ depend upon what is known as a $t$ distribution instead of a normal distribution. We will not go into the reasons for this, or even describe the $t$ distribution mathematically. Suffice to say that it is a cousin of the normal distribution, its exact shape depending on the number of degrees of freedom in the regression, and that it approximates the normal distribution increasingly closely as the number of degrees of freedom increases. You will certainly have encountered the $t$ distribution in your introductory statistics course. Table A.2 at the end of the text gives the critical values of $t$ cross-classified by significance level and the number of degrees of freedom.

The estimation of each parameter in a regression equation consumes one degree of freedom in the sample. Hence the number of degrees of freedom is equal to the number of observations in the sample minus the number of parameters estimated. The parameters are the constant (assuming that this is specified in the regression model) and the coefficients of the explanatory variables. In the present case of simple regression analysis, only two parameters, $\alpha$ and $\beta$, are estimated and hence the number of degrees of freedom is $n-2$. It should be emphasized that when we come to multiple regression analysis a more general expression will be required.

The critical value of $t$, which we will denote $t_{\text{crit}}$, replaces the number 1.96 in (3.45), so the condition that a regression estimate should not lead to the rejection of a null hypothesis $H_0: \beta = \beta_0$ is

$$ -t_{\text{crit}} \leq \frac{b - \beta_0}{s.e.(b)} \leq t_{\text{crit}} $$

(3.50)
Hence we have the decision rule: reject $H_0$ if $\left| \frac{b - \beta_0}{s.e.(b)} \right| > t_{crit}$, do not reject if $\left| \frac{b - \beta_0}{s.e.(b)} \right| \leq t_{crit}$, where $\left| \frac{b - \beta_0}{s.e.(b)} \right|$ is the absolute value (numerical value, neglecting the sign) of $t$.

**Examples**

```
. reg earnings hgc
```

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>Number of obs = 570</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>3977.38016</td>
<td>1</td>
<td>3977.38016</td>
<td>F( 1, 568) = 65.64</td>
</tr>
<tr>
<td>Residual</td>
<td>34419.6569</td>
<td>568</td>
<td>60.5979875</td>
<td>Prob &gt; F = 0.0000</td>
</tr>
<tr>
<td>Total</td>
<td>38397.0371</td>
<td>569</td>
<td>67.4816117</td>
<td>R-squared = 0.1036</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adj R-squared = 0.1020</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Root MSE = 7.7845</td>
</tr>
</tbody>
</table>

| earnings | Coef. | Std. Err. | t       | P>|t| | [95% Conf. Interval] |
|----------|-------|-----------|---------|-----|---------------------|
|          |       |           |         |     |                     |
| hgc      | 1.073055 | .1324501 | 8.102   | 0.000 | .8129028   1.333206 |
| _cons   | -1.391004 | 1.820305 | -0.764 | 0.445 | -4.966354   2.184347 |

In Section 2.6 hourly earnings were regressed on years of schooling using data from the United States National longitudinal Survey of Youth with the output shown above. The first two columns give the names of the variables, here just HGC and the intercept (Stata denotes this as _cons) and the estimates of their coefficients. The third column gives the corresponding standard errors. Let us suppose that one of the purposes of the regression was to confirm our intuition that earnings are affected by education. Accordingly, we set up the null hypothesis that $\beta$ is equal to zero and try to refute it. The corresponding $t$ statistic, using (3.49), is simply the estimate of the coefficient divided by its standard error:

$$
 t = \frac{b - \beta_0}{s.e.(b)} = \frac{b}{s.e.(b)} - \frac{0}{s.e.(b)} = 1.073 \div 0.132 = 8.13
$$

Since there are 570 observations in the sample and we have estimated two parameters, the number of degrees of freedom is 568. Table A.2 does not give the critical values of $t$ for 568 degrees of freedom, but we know that they must be lower than the corresponding critical values for 120, since the critical value is inversely related to the number of degrees of freedom. The critical value with 120 degrees of freedom at the 5 percent level is 1.980. Hence we can be sure that we would reject $H_0$ at the 5 percent level with 568 degrees of freedom and we conclude that schooling does affect earnings.

To put this test into words, with 568 degrees of freedom the upper and lower 2.5 percent tails of the $t$ distribution start approximately 1.980 standard deviations above and below its mean of zero. A regression coefficient which is estimated to lie within 1.980 standard deviations of zero would not cause the null hypothesis to be rejected. In this case, however, the discrepancy is equivalent to 8.13 estimated standard deviations and we come to the conclusion that the regression result contradicts the null hypothesis.

Of course since we are using the 5 percent significance level as the basis for the test there is in principle a 5 percent risk of a Type I error, if the null hypothesis is true. In this case we could reduce
the risk to 1 percent by using the 1 percent significance level instead. The critical value of $t$ at the 1 percent significance level with 120 degrees of freedom is 2.617. Since the $t$ statistic is greater than this, we see that we can easily reject the null hypothesis at this level as well.

Note that when the 5 percent and 1 percent tests lead to the same conclusion, there is no need to report both, and indeed you would look ignorant if you did. Read carefully the box on reporting test results.

This procedure of establishing a relationship between a dependent and an explanatory variable by setting up, and then refuting, a null hypothesis that $\beta$ is equal to zero is used very frequently indeed. Consequently all serious regression packages automatically print out the $t$ statistic for this special case, that is, the coefficient divided by its standard error. The ratio is often denoted "the" $t$-statistic. In the regression output, the $t$ statistics for the constant and slope coefficient appear in the middle column. (You will note that the $t$ statistic for the slope coefficient was actually 8.10, not 8.13. The calculation above was subject to rounding error.)

However, if the null hypothesis specifies some non-zero value of $\beta$, the more general expression (3.50) has to be used and the $t$ statistic has to be calculated by hand. For example, consider again the price inflation/wage inflation model (3.39) and suppose that the fitted model is (standard errors in parentheses):

$$
\hat{p} = -1.21 + 0.82 w
$$

(0.05) (0.10) (3.52)

If we now investigate the hypothesis that price inflation is equal to wage inflation, our null hypothesis is that the coefficient of $w$ is equal to 1.0. The corresponding $t$ statistic is

$$
t = \frac{b - \beta_0}{s.e.(b)} = \frac{0.82 - 1.00}{0.10} = -1.8
$$

(3.53)

If there are, say, 20 observations in the sample, the number of degrees of freedom is 18 and the critical value of $t$ at the 5 percent significance level is 2.101. The absolute value of our $t$ statistic is less than this, so on this occasion we do not reject the null hypothesis. The estimate 0.82 is below our hypothesized value 1.00, but not so far below as to exclude the possibility that the null hypothesis is correct. One final note on reporting regression results: some writers place the $t$ statistic in parentheses under a coefficient instead of the standard error. You should be careful to check, and when you are presenting results yourself, you should make it clear which you are giving.

---

**The Reject/Fail-to-Reject Terminology**

In this section it has been shown that you should reject the null hypothesis if the absolute value of the $t$ statistic is greater than $t_{crit}$, and that you fail to reject it otherwise. Why "fail to reject", which is a clumsy expression? Would it not be better just to say that you accept the hypothesis if the absolute value of the $t$ statistic is less than $t_{crit}$?

The argument against using the term accept is that you might find yourself "accepting" several mutually exclusive hypotheses at the same time. For instance, in the price-inflation/wage inflation example, you would not reject a null hypothesis $H_0: \beta = 0.9$, or a null hypothesis $H_0: \beta = 0.8$. It is logical to say that you would not reject these null hypotheses, as well as the null hypothesis $H_0: \beta = 1$ discussed in the text, but it makes little sense to say that you simultaneously accept the three hypotheses. In the next section you will see that one can define a whole range of hypotheses which would not be rejected by a given experimental result, so it would be incautious to pick out one as being "accepted".
Reporting the Results of $t$ tests

Suppose you have a theoretical relationship

$$y = \alpha + \beta x + u$$

and your null and alternative hypotheses are $H_0: \beta = \beta_0$, $H_0: \beta \neq \beta_0$. Given an experimental estimate $b$ of $\beta_0$, the acceptance and rejection regions for the hypothesis for the 5 percent and 1 percent significance levels can be represented in general terms by the left part of Figure 3.9.

The right side of the figure gives the same regions for a specific example, the price inflation/wage inflation model, the null hypothesis being that $\beta$ is equal to 1. The null hypothesis will not be rejected at the 5 percent level if $b$ lies within 2.101 standard errors of unity, that is, in the range 0.79 to 1.21, and it will not be rejected at the 1 percent level if $b$ lies within 2.878 standard deviations of unity, that is, in the range 0.71 to 1.29.

From Figure 3.9 it can be seen that there are three types of decision zone:

1. where $b$ is so far from the hypothetical $\beta$ that the null hypothesis is rejected at both the 5 percent and the 1 percent levels.
2. where $b$ is far enough from the hypothetical $\beta$ for the null hypothesis to be rejected at the 5 percent but not the 1 percent level
3. where $b$ is close enough to the hypothetical $\beta$ for the null hypothesis not to be rejected at either level.

From the diagram it can be verified that if the null hypothesis is rejected at the 1 percent level, it is automatically rejected at the 5 percent level. Hence in case (1) it is only necessary to report the rejection of the hypothesis at the 1 percent level. To report that it is rejected also at the 5 percent level is superfluous and suggests that you are not aware of this. It would be a bit like reporting that a certain high-jumper can clear two metres, and then adding that the athlete can also clear one and a half metres.

In case (3), likewise, you only need to make one statement, in this case that the hypothesis is not rejected at the 5 percent level. It automatically follows that it is not rejected at the 1 percent level, and to add a statement to this effect as well would be like saying that the high-jumper cannot clear one and a half metres, and also reporting that the athlete cannot clear two metres either.

Only in case (2) is it necessary (and desirable) to report the results of both tests.

Note that if you find that you can reject the null hypothesis at the 5% level, you should not stop there. You have established that the null hypothesis can be rejected at that level, but there remains a 5 percent chance of a Type I error. You should also perform the test at the 1% level. If you find that you can reject the null hypothesis at this level, this is the outcome that you should report. The risk of a Type I error is now only 1 percent and your conclusion is much more convincing. This is case (1) above. If you cannot reject at the 1 percent level, you have reached case (2) and you should report the results of both tests.
The fifth column of the output above, headed $p \, |t|$, provides an alternative approach to reporting the significance of regression coefficients. The figures in this column give the probability of obtaining the corresponding $t$ statistic, if the null hypothesis $H_0: \beta = 0$ were true. A $p$ value of less than 0.01 means that the probability is less than 1 percent, which in turn means that the null hypothesis would be rejected at the 1 percent level; a $p$ value between 0.01 and 0.05 means that the null hypothesis would be rejected at the 5 percent, but not the 1 percent level; and a $p$ value of 0.05 or more means that it would not be rejected at the 5 percent level.

The $p$ value approach is more informative than the 5 percent/1 percent approach, in that it gives the exact probability of a Type I error, if the null hypothesis is true. For example, in the earnings function output above, the $p$ values for the intercept is 0.445, meaning that the probability of obtaining a $t$ statistic of 0.764 or greater, in absolute terms, on a pure chance basis is in this case 44.5 percent. Hence the null hypothesis that the intercept is zero would not be rejected at any sensible significance level. In the case of the slope coefficient, the $p$ value is 0.0000, meaning that the probability of obtaining a $t$ statistic as large as 8.102, or larger, is less than 0.005 percent. Hence we would reject the null hypothesis that the slope coefficient is zero at the 1% level. Indeed we would reject it at the 0.1% level – see below. Choice between the $p$ value approach and the 5 percent/1 percent approach appears to be entirely conventional. The medical literature uses $p$ values, but the economics literature generally uses 5 percent/1 percent.
0.1 percent tests

If the $t$ statistic is very high, you should check whether you can reject the null hypothesis at the 0.1% level. If you can, you should report it, because it demonstrates that you are able to reject the null hypothesis with minimal risk of a Type I error.

Exercises

3.7 Give more examples of daily instances in which decisions involving possible Type I and Type II errors may arise.

3.8 Before beginning a certain course, 36 students are given an aptitude test. The scores, and the course results (pass/fail) are given below:

<table>
<thead>
<tr>
<th>student</th>
<th>test score</th>
<th>course result</th>
<th>student</th>
<th>test score</th>
<th>course result</th>
<th>student</th>
<th>test score</th>
<th>course result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>fail</td>
<td>13</td>
<td>26</td>
<td>fail</td>
<td>25</td>
<td>9</td>
<td>fail</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>pass</td>
<td>14</td>
<td>43</td>
<td>pass</td>
<td>26</td>
<td>36</td>
<td>pass</td>
</tr>
<tr>
<td>3</td>
<td>33</td>
<td>fail</td>
<td>15</td>
<td>43</td>
<td>fail</td>
<td>27</td>
<td>61</td>
<td>pass</td>
</tr>
<tr>
<td>4</td>
<td>62</td>
<td>pass</td>
<td>16</td>
<td>68</td>
<td>pass</td>
<td>28</td>
<td>79</td>
<td>fail</td>
</tr>
<tr>
<td>5</td>
<td>59</td>
<td>fail</td>
<td>17</td>
<td>63</td>
<td>pass</td>
<td>29</td>
<td>57</td>
<td>fail</td>
</tr>
<tr>
<td>6</td>
<td>63</td>
<td>pass</td>
<td>18</td>
<td>42</td>
<td>fail</td>
<td>30</td>
<td>46</td>
<td>pass</td>
</tr>
<tr>
<td>7</td>
<td>80</td>
<td>pass</td>
<td>19</td>
<td>51</td>
<td>fail</td>
<td>31</td>
<td>70</td>
<td>fail</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>fail</td>
<td>20</td>
<td>45</td>
<td>fail</td>
<td>32</td>
<td>31</td>
<td>pass</td>
</tr>
<tr>
<td>9</td>
<td>60</td>
<td>pass</td>
<td>21</td>
<td>22</td>
<td>fail</td>
<td>33</td>
<td>68</td>
<td>pass</td>
</tr>
<tr>
<td>10</td>
<td>76</td>
<td>pass</td>
<td>22</td>
<td>30</td>
<td>pass</td>
<td>34</td>
<td>62</td>
<td>pass</td>
</tr>
<tr>
<td>11</td>
<td>13</td>
<td>fail</td>
<td>23</td>
<td>40</td>
<td>fail</td>
<td>35</td>
<td>56</td>
<td>pass</td>
</tr>
<tr>
<td>12</td>
<td>41</td>
<td>pass</td>
<td>24</td>
<td>26</td>
<td>fail</td>
<td>36</td>
<td>36</td>
<td>pass</td>
</tr>
</tbody>
</table>

Do you think that the aptitude test is useful for selecting students for admission to the course, and if so, how would you determine the pass mark? (Discuss the trade-off between Type I and Type II errors associated with the choice of pass-mark.)

3.9 A researcher hypothesizes that years of schooling, $HGC$, may be related to the number of siblings (brothers and sisters), $SIBLINGS$, according to the relationship

$$HGC = \alpha + \beta SIBLINGS + u$$

She is prepared to test the null hypothesis $H_0: \beta = 0$ against the alternative hypothesis $H_1: \beta \neq 0$ at the 5 percent and 1 percent levels. She has a sample of 60 observations. What should she report

1. if $b = 0.20$, s.e.($b$) = 0.07?
2. if $b = -0.12$, s.e.($b$) = 0.07?
3. if $b = 0.06$, s.e.($b$) = 0.07?
4. if $b = 0.20$, s.e.($b$) = 0.07?
A researcher with a sample of 50 individuals with similar education but differing amounts of training hypothesizes that hourly earnings, \( EARNINGS \), may be related to hours of training, \( TRAINING \), according to the relationship

\[
EARNINGS = \alpha + \beta \cdot TRAINING + u
\]

He is prepared to test the null hypothesis \( H_0 : \beta = 0 \) against the alternative hypothesis \( H_1 : \beta \neq 0 \) at the 5 percent and 1 percent levels. What should he report?

1. if \( b = 0.30 \), s.e.(\( b \)) = 0.12?
2. if \( b = 0.55 \), s.e.(\( b \)) = 0.12?
3. if \( b = 0.10 \), s.e.(\( b \)) = 0.12?
4. if \( b = -0.27 \), s.e.(\( b \)) = 0.12?

3.11 Perform a \( t \) test on the slope coefficient and the intercept of the educational attainment function fitted using your EAEF data set, and state your conclusions.

3.12 Perform a \( t \) test on the slope coefficient and the intercept of the earnings function fitted using your EAEF data set, and state your conclusions.

3.13* In Exercise 2.7, the growth rate of employment was regressed on the growth rate of GDP for a sample of 25 OECD countries. Perform \( t \) tests on the slope coefficient and the intercept and state your conclusions.

3.8 Confidence Intervals

Thus far we have been assuming that the hypothesis preceded the empirical investigation. This is not necessarily the case. Usually theory and experimentation are interactive, and the earnings function regression provides a typical example. We ran the regression in the first place because economic theory tells us to expect earnings to be affected by schooling. The regression result confirmed this intuition since we rejected the null hypothesis \( \beta = 0 \) but we were then left with something of a vacuum, since our theory is not strong enough to suggest that the true value of \( \beta \) is equal to some specific number. However, we can now move in the opposite direction and ask ourselves the following question: given our regression result, what hypotheses would be compatible with it?

Obviously a hypothesis \( \beta = 1.073 \) would be compatible, because then hypothesis and experimental result coincide. Also \( \beta = 1.072 \) and \( \beta = 1.074 \) would be compatible, because the difference between hypothesis and experimental result would be so small. The question is, how far can a hypothetical value differ from our experimental result before they become incompatible and we have to reject the null hypothesis?

We can answer this question by exploiting the previous analysis. From (3.50), we can see that regression coefficient \( b \) and hypothetical value \( \beta \) are incompatible if either
That is, if either

\[ b - \beta > \text{s.e.}(b) \times t_{\text{crit}} \quad \text{or} \quad b - \beta < -\text{s.e.}(b) \times t_{\text{crit}} \]  

(3.55)

that is, if either

\[ b - \text{s.e.}(b) \times t_{\text{crit}} > \beta \quad \text{or} \quad b + \text{s.e.}(b) \times t_{\text{crit}} < \beta \]  

(3.56)

It therefore follows that a hypothetical \( \beta \) is compatible with the regression result if both

\[ b - \text{s.e.}(b) \times t_{\text{crit}} \leq \beta \quad \text{or} \quad b + \text{s.e.}(b) \times t_{\text{crit}} \geq \beta \]  

(3.57)

that is, if \( \beta \) satisfies the double inequality

\[ b - \text{s.e.}(b) \times t_{\text{crit}} \leq \beta \leq b + \text{s.e.}(b) \times t_{\text{crit}} \]  

(3.58)

Any hypothetical value of \( \beta \) which satisfies (3.58) will therefore automatically be compatible with the estimate \( b \), i.e., will not be rejected by it. The set of all such values, given by the interval between the lower and upper limits of the inequality, is known as the confidence interval for \( \beta \).

Note that the center of the confidence interval is \( b \) itself. The limits are equidistant on either side. Note also that, since the value of \( t_{\text{crit}} \) depends upon the choice of significance level, the limits will also depend on this choice. If the 5 percent significance level is adopted, the corresponding confidence interval is known as the 95 percent confidence interval. If the 1 percent level is chosen, one obtains the 99 percent confidence interval, and so on.

Since \( t_{\text{crit}} \) is greater for the 1 percent level than for the 5 percent level, for any given number of degrees of freedom, it follows that the 99 percent interval is wider than the 95 percent interval. Since they are both centered on \( b \), the 99 percent interval encompasses all the hypothetical values of \( \beta \) in the 95 percent confidence interval and some more on either side as well.

**Example**

In the earnings function output above, the coefficient of \( HGC \) was 1.073, its standard error was 0.132, and the critical value of \( t \) at the 5 percent significance level was about 1.97. The corresponding 95 percent confidence interval is therefore

\[ 1.073 - 0.132 \times 1.97 \leq \beta \leq 1.073 + 0.132 \times 1.97 \]  

(3.59)

that is,

\[ 0.813 \leq \beta \leq 1.333 \]  

(3.60)
We would therefore reject hypothetical values above 1.333 and below 0.813. Any hypotheses within these limits would not be rejected, given the regression result. This confidence interval actually appears as the final column in the output above. However, this is not a standard feature of a regression package, so you usually have to calculate the interval yourself.

Exercises

3.14 Calculate the 99 percent confidence interval for β in the preceding example, and explain why it includes some values not included in the 95 percent confidence interval.

3.15 Calculate the 95 percent confidence interval for the slope coefficient in the earnings function fitted with your EAEF data set.
3.16 Calculate the 95 percent confidence interval for $\beta$ in the price inflation/wage inflation example:

$$\hat{\beta} = -1.21 + 0.82 w$$

(0.05) (0.10)

What can you conclude from this calculation?

3.9 One-Tailed t tests

In our discussion of $t$ tests, we started out with our null hypothesis $H_0: \beta = \beta_0$ and tested to see whether we should reject it or not, given the regression coefficient $b$. If we did reject it, then by implication we accepted the alternative hypothesis $H_1: \beta \neq \beta_0$.

Thus far the alternative hypothesis has been merely the negation of the null hypothesis. However, if we are able to be more specific about the alternative hypothesis, we should be able to improve the testing procedure. We will investigate three cases: first, the very special case where there is only one conceivable other true value of $\beta$, which we will denote $\beta_1$; second, where, if $\beta$ is not equal to $\beta_0$, it must be greater than $\beta_0$; and third, where, if $\beta$ is not equal to $\beta_0$, it must be less than $\beta_0$.

$H_0: \beta = \beta_0, H_1: \beta = \beta_1$

In this case, for some reason, there are only two possible values of the true coefficient of $x$, $\beta_0$ and $\beta_1$. For sake of argument we will assume for the time being that $\beta_1$ is greater than $\beta_0$.

Suppose that we wish to test $H_0$ at the 5 percent significance level, and we follow the usual procedure discussed earlier in the chapter. We locate the limits of the upper and lower 2.5 percent tails under the assumption that $H_0$ is true, labeled $A$ and $B$ in Figure 3.8, and we reject $H_0$ if the regression coefficient $b$ lies to the left of $A$ or to the right of $B$.

![Figure 3.10. Distribution of $b$ under $H_0$ and $H_1$](image-url)
Now, if $b$ does lie to the right of $B$, it is much more compatible with $H_1$ than with $H_0$; the probability of it lying to the right of $B$ is much greater if $H_1$ is true than if $H_0$ is true. We should have no hesitation in rejecting $H_0$ in favor of $H_1$.

However, if $b$ lies to the left of $A$, the test procedure will lead us to a perverse conclusion. It tells us to reject $H_0$ in favor of $H_1$, even though the probability of $b$ lying to the left of $A$ is negligible if $H_1$ is true. We have not even drawn the probability density function that far for $H_1$. If such a value of $b$ occurs only once in a million times when $H_1$ is true, but 2.5 percent of the time when $H_0$ is true, it is much more logical to assume that $H_0$ is true. Of course once in a million times you will make a mistake, but the rest of the time you will be right.

Hence we will reject $H_0$ only if $b$ lies in the upper 2.5 percent tail, that is, to the right of $B$. We are now performing a one-tailed test, and we have reduced the probability of making a Type I error to 2.5 percent. Since the significance level is defined to be the probability of making a Type I error, it is now also 2.5 percent.

As we have seen, economists usually prefer 5 percent and 1 percent significance tests, rather than 2.5 percent tests. If you want to perform a 5 percent test, you move $B$ to the left so that you have 5 percent of the probability in the tail and the probability of making a Type I error is increased to 5 percent. *(Question: why would you deliberately choose to increase the probability of making a Type I error? Answer, because at the same time you are reducing the probability of making a Type II error, that is, of not rejecting the null hypothesis when it is false. Most of the time your null hypothesis is that the coefficient is zero, and you are trying to disprove this, demonstrating that the variable in question does have an effect. In such a situation, by using a one tailed test, you reduce the risk of not rejecting a false null hypothesis, while holding the risk of a Type I error at 5 percent.)*

If the standard deviation of $b$ is known (most unlikely in practice), so that the distribution is normal, $B$ will be $Z$ standard deviations to the right of $\beta_0$, where $Z$ is given by $A(Z) = 0.9500$ in Table A.1. The appropriate value of $Z$ is 1.64. If the standard deviation is unknown and has been estimated as the standard error of $b$, you have to use a $t$ distribution: you look up the critical value of $t$ in Table A.2 for the appropriate number of degrees of freedom in the column headed 5 percent.

Similarly, if you want to perform a 1 percent test, you move $B$ to the right to the point where the tail contains 1 percent of the probability. Assuming that you have had to calculate the standard error of $b$ from the sample data, you look up the critical value of $t$ in the column headed 1 percent.

We have assumed in this discussion that $\beta_1$ is greater than $\beta_0$. Obviously, if it is less than $\beta_0$, we should use the same logic to construct a one-tailed test, but now we should use the left tail as the rejection region for $H_0$ and drop the right tail.

**The Power of a Test**

In this particular case we can calculate the probability of making a Type II error, that is, of accepting a false hypothesis. Suppose that we have adopted a false hypothesis $H_0$: $\beta = \beta_0$ and that an alternative hypothesis $H_1$: $\beta = \beta_1$ is in fact true. Referring to Figure 3.8 again, we will accept $H_0$ if the sample regression coefficient $b$ lies to the left of $B$. Since $H_1$ is true, the probability of $b$ lying to the left of $B$ is given by the area under the curve for $H_1$ to the left of $B$. All we have to do is to calculate $t$ for the point $B$, assuming $\beta = \beta_0$, and use a $t$ distribution table to find the probability of $b$ being more than $t$ standard errors to the left of $\beta_1$. 
If this probability is denoted $\gamma$, the power of the test, defined to be the probability of not making a Type II error, is $(1 - \gamma)$. Obviously, you have a trade-off between the power of the test and the significance level. The higher the significance level, the further $B$ will be to the right, and so the larger $\gamma$ will be, so the lower the power of the test will be.

In using a one-tailed instead of a two-tailed test, you are able to obtain greater power for any level of significance. As we have seen, you would move $B$ in Figure 3.8 to the left if you were performing a one-tailed test at the 5 percent significance level, thereby reducing the probability of accepting $H_0$ if it happened to be false. However, you must remember that this gain in power is conditional on the validity of your reasons for using a one-tailed test.

$$H_0: \beta = \beta_0, \ H_1: \beta > \beta_0$$

We have discussed the case in which the alternative hypothesis involved a specific hypothetical value $\beta_i$, with $\beta_i$ greater than $\beta_0$. Clearly, the logic that led us to use a one-tailed test would still apply even if $H_1$ were more general and merely asserted that $\beta_i > \beta_0$, without stating any particular value.

We would still wish to eliminate the left tail from the rejection region because a low value of $b$ is more probable under $H_0: \beta = \beta_0$ than under $H_1: \beta > \beta_0$, and this would be evidence in favor of $H_0$, not against it. Therefore, we would still prefer a one-tailed $t$ test, using the right tail as the rejection region, to a two-tailed test. Note that, since $\beta_i$ is not defined, we now have no way of calculating the power of such a test. However, we can still be sure that, for any given significance level, the power of a one-tailed test will be greater than that of the corresponding two-tailed test.

$$H_0: \beta = \beta_0, \ H_1: \beta < \beta_0$$

Similarly if the alternative hypothesis were $H_0: \beta < \beta_0$, we would prefer a one-tailed test using the left tail as the rejection region.

**Justification of the Use of a One-Tailed Test**

The use of a one-tailed test has to be justified beforehand on the grounds of theory, common sense, or previous experience. When stating the justification, you should be careful not to exclude the possibility that the null hypothesis is true. For example, suppose that you are relating household expenditure on clothing to household income. You would of course expect a significant positive effect, given a large sample. But you justification should not be that, on the basis of theory and commonsense, that the coefficient should be positive. This is too strong, for it rules out the null hypothesis of no effect and there is nothing to test. Instead, you should say that, on the basis of theory and common sense, you would exclude the possibility that income has a negative effect. This then leaves the possibility that the effect is zero and the alternative that it is positive.

One-tailed tests are very important in practice in econometrics. As we have seen, the usual way of establishing that an explanatory variable really does influence a dependent variable is to set up the null hypothesis $H_0: \beta = 0$ and try to refute it. Very frequently, our theory is strong enough to tell us
that, if \( x \) does influence \( y \), its effect will be in a given direction. If we have good reason to believe that the effect is not negative, we are in a position to use the alternative hypothesis \( H_1: \beta > 0 \) instead of the more general \( H_1: \beta \neq 0 \). This is an advantage because the critical value of \( t \) for rejecting \( H_0 \) is lower for the one-tailed test, so it is easier to refute the null hypothesis and establish the relationship.

**Examples**

In the earnings function regression, there were 568 degrees of freedom and the critical value of \( t \), using the 1 percent significance level and a two-tailed test, is approximately 2.59. If we take advantage of the fact that it is reasonable to expect income not to have a negative coefficient, we could use a one-tailed test and the critical value is reduced to approximately 2.33. The \( t \) statistic is in fact equal to 8.10, so in this case the refinement makes no difference. The estimated coefficient is so large relative to its standard error that we reject the null hypothesis regardless of whether we use a two-tailed or a one-tailed test.

In the price inflation/wage inflation example, exploiting the possibility of using a one-tailed test does make a difference. The null hypothesis was that wage inflation is reflected fully in price inflation and we have \( H_0: \beta = 1 \). The main reason why the types of inflation may be different is that improvements in labor productivity may cause price inflation to be lower than wage inflation. Certainly this will not cause price inflation to be greater than wage inflation and so in this case we are justified in ruling out \( \beta > 1 \). We are left with \( H_0: \beta = 1 \) and \( H_1: \beta < 1 \). Given a regression coefficient 0.82 and a standard error 0.10, the \( t \) statistic for the null hypothesis is -1.80. This was not high enough, in absolute terms, to cause \( H_0 \) to be rejected at the 5 percent level using a two-tailed test (critical value 2.10). However, if we use a one-tailed test, as we are entitled to, the critical value falls to 1.73 and we can reject the null hypothesis. In other words, we can conclude that price inflation is significantly lower than wage inflation.

**Exercises**

3.17 Explain whether it would have been possible to perform one-tailed tests instead of two-tailed tests in Exercise 3.9. If you think that one-tailed tests are justified, perform them and state whether the use of a one-tailed test makes any difference.

3.18* Explain whether it would have been possible to perform one-tailed tests instead of two-tailed tests in Exercise 3.10. If you think that one-tailed tests are justified, perform them and state whether the use of a one-tailed test makes any difference.

3.19* Explain whether it would have been possible to perform one-tailed tests instead of two-tailed tests in Exercise 3.11. If you think that one-tailed tests are justified, perform them and state whether the use of a one-tailed test makes any difference.

**3.10 The F Test of Goodness of Fit**
Even if there is no relationship between $y$ and $x$, in any given sample of observations there may appear to be one, if only a faint one. Only by coincidence will the sample covariance be exactly equal to zero. Accordingly, only by coincidence will the correlation coefficient and $R^2$ be exactly equal to zero.

This presents us with a problem. How do we know if the value of $R^2$ for the regression reflects a true relationship or if it has arisen as a matter of chance?

We could in principle adopt the following procedure. We take as our null hypothesis that there is no relationship between $y$ and $x$, and calculate the value that would be exceeded by $R^2$ as a matter of chance, 5 percent of the time. We then take this figure as the critical level of the 5 percent significance test. If it is exceeded, we reject the null hypothesis.

Such a test, like the $t$ test on a coefficient, would not be foolproof. Indeed, at the 5 percent significance level, one would risk making a Type I error (rejecting the null hypothesis when it is in fact true) 5 percent of the time. Of course you could cut down on this risk by using a higher significance level, for example, the 1 percent level. The critical level of $R^2$ would then be that which would be exceeded by chance only 1 percent of the time, so it would be higher than the critical level for the 5 percent test.

How does one find the critical level of $R^2$ at either significance level? Well, there is a slight problem. There is no such thing as a table of critical levels of $R^2$. The traditional procedure is to use an indirect approach and perform what is known as an $F$ test based on analysis of variance. (For an explanation of the theory behind it, see, for example, Mood and Graybill, 1963.)

Suppose that, as in this case, you can decompose the variance of the dependent variable into "explained" and "unexplained" components using (2.46):

$$\text{Var}(y) = \text{Var}(\hat{y}) + \text{Var}(e)$$

Using the definition of sample variance, and multiplying through by $n$, we can rewrite the decomposition as

$$\sum (y - \bar{y})^2 = \sum (\hat{y} - \bar{y})^2 + \sum e^2$$

(Remember that $\bar{\sigma}$ is zero and that the sample mean of $\hat{y}$ is equal to the sample mean of $y$.)

The left side is $TSS$, the total sum of squares of the values of the dependent variable about its sample mean. The first term on the right side is $ESS$, the explained sum of squares, and the second term is $RSS$, the unexplained, residual sum of squares which we initially simply called $S$:

$$TSS = ESS + RSS$$

The $F$ statistic for the goodness of fit of a regression is written as the explained sum of squares, per explanatory variable, divided by the residual sum of squares, per degree of freedom remaining:

$$F = \frac{ESS/k}{RSS/(n-k-1)}$$

where $k$ is the number of explanatory variables.
By dividing both the numerator and the denominator of the ratio by \( TSS \), this \( F \) statistic may equivalently be expressed in terms of \( R^2 \):

\[
F = \frac{(ESS/TSS)/k}{(RSS/TSS)/(n-k-1)} = \frac{R^2/k}{(1-R^2)/(n-k-1)}
\]

(3.65)

In the present context, \( k \) is 1, so (3.65) becomes

\[
F = \frac{R^2}{(1-R^2)/(n-2)}
\]

(3.66)

Having calculated \( F \) from your value of \( R^2 \), you look up \( F_{\text{crit}} \), the critical level of \( F \), in the appropriate table. If \( F \) is greater than \( F_{\text{crit}} \), you reject the null hypothesis and conclude that the "explanation" of \( y \) is better than is likely to have arisen by chance.

Table A.3 gives the critical levels of \( F \) at the 5 percent, 1 percent and 0.1 percent significance levels. In each case the critical level depends on the number of explanatory variables, \( k \), which is read from along the top of the table, and the number of degrees of freedom, \((n-k-1)\), which is read off down the side. In the present context, we are concerned with simple regression analysis, \( k \) is 1, and we should use the first column of the table.

In the earnings function example, \( R^2 \) was 0.1036. Since there were 570 observations, the \( F \) statistic is equal to \( R^2/\{(1 - R^2)/568\} = 0.1036 / (0.8964/568) = 65.65 \). At the 0.1 percent significance level, the critical level of \( F \) for 1 and 120 degrees of freedom (looking at the first column, row 120) is 11.38. The critical value for 1 and 568 degrees of freedom must be lower, so we have no hesitation in rejecting the null hypothesis in this particular example. In other words, the underlying value of \( R^2 \) is so high that we reject the suggestion that it could have arisen by chance. In practice the \( F \) statistic is always computed for you, along with \( R^2 \), so you never actually have to use (3.65) yourself.

Why do people bother with this indirect approach? Why not have a table of critical levels of \( R^2 \)? The answer is that the \( F \) table is useful for testing many forms of analysis of variance, of which \( R^2 \) is only one. Rather than have a specialized table for each application, it is more convenient (or, at least, it saves a lot of paper) to have just one general table, and make transformations like (3.65) when necessary.

Of course you could derive critical levels of \( R^2 \) if you were sufficiently interested. The critical level of \( R^2 \) would be related to the critical level of \( F \) by

\[
F_{\text{crit}} = \frac{R^2_{\text{crit}}/k}{(1-R^2_{\text{crit}})/(n-k-1)}
\]

(3.67)

which yields

\[
R^2_{\text{crit}} = \frac{kF_{\text{crit}}}{kF_{\text{crit}} + (n-k-1)}
\]

(3.68)
In the earnings function example, the critical value of $F$ at the 1 percent significance level was approximately 11.38. Hence in this case, with $k = 1$,

$$R^2_{crit} = \frac{11.38}{11.38 + 568} = 0.020$$

(3.69)

Although it is low, our $R^2$ is greater than 0.020, so a direct comparison of $R^2$ with its critical value confirms the conclusion of the $F$ test that we should reject the null hypothesis.

**Exercises**

3.20 In Exercise 2.6, in the regression of employment growth rates on growth rates of GDP using a sample of 25 OECD countries, $R^2$ was 0.5909. Calculate the corresponding $F$ statistic and check that it is equal to 33.22, the value printed in the output. Perform the $F$ test at the 5 percent and 1 percent significance levels. Is it necessary to report the results of the tests at both levels?

3.21 Similarly, calculate the $F$ statistic from the value of $R^2$ obtained in the earnings function fitted using your EAEF data set and check that it is equal to the value printed in the ouput. Perform an appropriate $F$ test.

**3.11 Relationship between the $F$ Test of Goodness of Fit and the $t$ test on the Slope Coefficient in Simple Regression Analysis**

In the context of simple regression analysis (and only simple regression analysis) the $F$ test on $R^2$ and the two-tailed $t$ test on the slope coefficient both have $H_0: \beta = 0$ as their null hypothesis and $H_1: \beta \neq 0$ as their alternative hypothesis. This gives rise to the possibility that they might lead to different conclusions. Fortunately, they are in fact equivalent. The $F$ statistic is equal to the square of the $t$ statistic, and the critical value of $F$, at any given significance level, is equal to the square of the critical value of $t$. Starting with the definition of $F$ in (3.66),
The proof that the critical value of $F$ is equal to the critical value of $t$ is a little more complicated and will be omitted. When we come to multiple regression analysis, we will see that the $F$ test and the $t$ tests have different roles and different null hypotheses. However, in simple regression analysis the fact that they are equivalent means that there is no point in performing both. Indeed, you would look ignorant if you did. Obviously, provided that it is justifiable, a one-tailed $t$ test would be preferable to either.

Exercises

3.22 Verify that the $F$ statistic in the earnings function regression run by you using your EAEF data set is equal to the square of the $t$ statistic for the slope coefficient, and that the critical value of $F$ at the 1 percent significance level is equal to the square of the critical value of $t$.

3.23 In Exercise 2.7 both researchers obtained values of $R^2$ equal to 0.79 in their regressions. Was this a coincidence?