A

Presentation of regression results

The results of a regression analysis can be presented in various different ways. However, the most common way is to write the estimated equation with standard errors of coefficients written in brackets below the estimated coefficients and include some more statistics below the equation. For the consumption function that will be presented in Computer Example 2, the results are summarized as shown below:

$$\hat{C}_t = \underset{(6.565)}{15.116} + \underset{(0.038)}{0.160}Y_t^d \quad (4.73)$$

$$R^2 = 0.932 \quad n = 20 \quad \hat{\sigma} = 6.879 \quad (4.74)$$

From this summary we can (a) read estimated effects of changes in the explanatory variables on the dependent variable, (b) predict values of the dependent variable for given values of the explanatory variable, (c) perform hypothesis testing for the estimated coefficients, and (d) construct confidence intervals for the estimated coefficients.

Applications

Application 1: the demand function

From economic theory we know that the demand for a commodity depends basically on the price of that commodity (the law of demand). Other possible determinants can include prices of other competing goods (close substitutes) or those that complement that commodity (close complements), and of course the level of income of the consumer. In order to include all those determinants we need to employ a multiple regression analysis. However, for pedagogical purposes we have to restrict ourselves to one explanatory variable. Therefore, we can assume a partial demand function where the quantity demanded is affected only by the price of the product. (Another way of doing this is to assume a *ceteris paribus* (other things remaining the same) demand function, in which we simply assume that the other variables entering the relationship remain constant, and thus do not affect the quantity demanded.) The population regression function will have the form:

$$q_t = a_0 + a_1 p_t + u_t \quad (4.75)$$

where the standard notation is used with q_t denoting quantity demanded and p_t the price of the product. From economic theory we expect a_1 to be negative reflecting the law of demand (the higher the price the less the quantity demanded). We can collect time series data for sales of a product and the price level of this product and estimate the above specification. The interpretation of the obtained results will be as follows. For a_1: if the price of the product will be increased by one unit of measurement (i.e. if measured in £ an increase of £1.00), the consumption of this product will be decreased (because a_1 will be negative) by $\hat{a}_1$ units. For a_0: if the price of the product is zero consumers will consume $\hat{a}_0$ quantity of this product. R^2 is expected to be somehow

low (lets say 0.6) suggesting that there are additional variables that affect the quantity demanded, that we did not include in our equation, while it is also possible to obtain the price elasticity of this product for a given year (lets say 1999) from the equation:

$$\frac{p_{99}}{\hat{q}_{99}} \frac{\Delta q}{\Delta p} = \frac{p_{99}}{\hat{q}_{99}} \hat{a}_1 \quad (4.76)$$

Application 2: a production function

One of the most basic relationships in economic theory is the production function, that, usually, relates output (denoted by Y) to the possible factor inputs affecting production, such as labour (L) and capital (K). The general form of this relationship can be expressed by:

$$Y_t = f(K_t, L_t) \quad (4.77)$$

A frequently utilized form of this function – due to its properties that we will see later – is the well-known Cobb–Douglas production function:

$$Y_t = AK_t^a L_t^\beta \quad (4.78)$$

where a and β are constant terms that express the responsiveness of output to capital and labour respectively. A can be regarded as an exogenous efficiency/technology parameter. Obviously the greater is A, the higher is maximum output keeping labour and capital constant. In the short run we can assume that the stock of capital is fixed (short-run can be viewed here as a period that once the decision about capital has been made it cannot be changed by the producer until the next period). Then, in the short run, maximum output depends only on the labour input, and the production function becomes:

$$Y_t = g(L_t) \quad (4.79)$$

Using the Cobb–Douglas form of function (and for K_t constant and equal to K_0) we will have:

$$Y_t = (AK_0^a)L_t^\beta = A^* L_t^\beta \quad (4.80)$$

where $A^* = (AK_0^a)$. This short-run production function is now a bivariate model, and after applying a logarithmic transformation can be estimated with the OLS method. Taking the natural logarithm of both sides and adding an error term we have:

$$\begin{aligned} \ln Y_t &= \ln(A^*) + \beta \ln(L_t) + u_t \\ &= c + \beta \ln(L_t) + u_t \quad (4.81) \end{aligned}$$

where $c = \ln(A^*)$, and β is the elasticity of output with respect to labour (one of the properties of the Cobb–Douglas production function). This elasticity denotes the percentage change in output that results from a 1 per cent change in the labour input.

(A)

We may use time series data on production and employment for the manufacturing sector of a country (or aggregate GDP and employment data) to obtain estimates of c and β for the above model.

✓Application 3: Okun's law

Okun (1962) developed an empirical relationship, using quarterly data from 1947:2 to 1960:4, between changes in the state of the economy (captured by changes in GNP) and changes in the unemployment rate, known as Okun's law. His results provide an important insight into the sensitivity of the unemployment rate to economic growth. The basic relationship is that of connecting the growth rate of unemployment (UNEMP) (which constitutes the dependent variable) to a constant and the growth rate of GNP (the independent variable) as follows:

$$\Delta UNEMP_t = a + b\Delta GNP_t + u_t \tag{4.82}$$

Applying OLS the sample regression equation that Okun obtained was:

$$\begin{aligned} \Delta \widehat{UNEMP}_t &= 0.3 - 0.3\Delta GNP_t \\ R^2 &= 0.63 \end{aligned} \tag{4.83}$$

The constant in this equation shows the mean change in the unemployment rate when the growth rate of the economy is equal to zero, so from the obtained results we conclude that when the economy does not grow the unemployment rate rises by 0.3 per cent. The negative b coefficient suggests that when the state of the economy improves, the unemployment rate falls. The relationship, though, is less than one to one. A 1 per cent increase in GNP is connected with only a 0.3 per cent decrease in the unemployment rate. This result is called Okun's law. It is easy to collect data on GNP and unemployment, calculate their respective growth rates and check whether Okun's law is valid for different countries and different time periods.

✓Application 4: the Keynesian consumption function

Another basic relationship in economic theory is the Keynesian consumption function that simply states that consumption (C_t) is a positive linear function of disposable (after tax) income (Y_t^d). The relationship is as follows:

$$C_t = a + \delta Y_t^d \tag{4.84}$$

where a is the autonomous consumption (consumption even when disposable income is zero) and δ is the marginal propensity to consume. In this function we expect $a > 0$ and $0 > \delta > 1$. A $\hat{\delta} = 0.7$ means that the marginal propensity to consume is 0.7. A Keynesian consumption function is estimated below as a worked-out computer exercise example.

on the EViews command line, or by clicking on **Quick/Estimate equation** and then writing the equation (i.e. y c x) in the new window. Note that the option for **OLS (LS – Least Squares (NLS and ARMA))** is automatically chosen by EViews and the sample is automatically chosen to be from 1 to 20.

Either way the output shown in Table 4.7 is shown in a new window which provides estimates for alpha (the coefficient of the constant term) and beta (the coefficient of X).

Questions and exercises

Questions

1 An outlier is an observation that is very far from the sample regression function. Suppose the equation is initially estimated using all observations and then reestimated omitting outliers. How will the estimated slope coefficient change? How will R^2 change? Explain.

2 Regression equations are sometimes estimated using an explanatory variable that is a deviation from some value of interest. An example is a capacity utilization rate–unemployment rate equation, such as:

$$u_t = a_0 + a_1(CAP_t - CAP_t^f) + e_t$$

where CAP_t^f is a single value representing the capacity utilization rate corresponding to full employment (the value of 87.5% is sometimes used for this value).

(a) Will the estimated intercept from this equation differ from that in the equation with only CAP_t as an explanatory variable? Explain.

(b) Will the estimated slope coefficient from this equation differ from that in the equation with only CAP_t as an explanatory variable? Explain.

3 Prove that the OLS coefficient for the slope parameter in the simple linear regression model is unbiased.

4 Prove that the OLS coefficient for the slope parameter in the simple linear regression model is BLUE.

5 State the assumptions of the simple linear regression model and explain why they are necessary.

Exercise 4.1

The following data refer to the quantity sold for a good Y (measured in kg), and the price of that good X (measured in pence per kg), for 10 different market locations:

Y:	198	181	170	179	163	145	167	203	251	147
X:	23	24.5	24	27.2	27	24.4	24.7	22.1	21	25

(a) Assuming a linear relationship among the two variables, obtain the OLS estimators of a and β.

(b) On a scatter diagram of the data, draw in your OLS sample regression line.

(c) Estimate the elasticity of demand for this good at the point of sample means (i.e. when $Y = \bar{Y}$ and $X = \bar{X}$).

Exercise 4.2

The table below shows the average growth rates of GDP and employment for 25 OECD countries for the period 1988–97.

Countries	*Empl.*	*GDP*	*Countries*	*Empl.*	*GDP*
Australia	1.68	3.04	Korea	2.57	7.73
Austria	0.65	2.55	Luxembourg	3.02	5.64
Belgium	0.34	2.16	Netherlands	1.88	2.86
Canada	1.17	2.03	New Zealand	0.91	2.01
Denmark	0.02	2.02	Norway	0.36	2.98
Finland	−1.06	1.78	Portugal	0.33	2.79
France	0.28	2.08	Spain	0.89	2.60
Germany	0.08	2.71	Sweden	−0.94	1.17
Greece	0.87	2.08	Switzerland	0.79	1.15
Iceland	−0.13	1.54	Turkey	2.02	4.18
Ireland	2.16	6.40	United Kingdom	0.66	1.97
Italy	−0.30	1.68	United States	1.53	2.46
Japan	1.06	2.81			

(a) Assuming a linear relationship obtain the OLS estimators.

(b) Provide an interpretation of the coefficients.

Exercise 4.3

In the Keynesian consumption function:

$$C_t = a + \delta Y_t^d$$

the estimated marginal propensity to consume is simply $\hat{\delta}$ while the average propensity to consume is $C/Y^d = \hat{a}/Y^d + \hat{\delta}$. Using data from 200 UK households on annual income and consumption (both of which were measured in UK£) we found the following regression equation:

$$C_t = 138.52 + 0.725Y_t^d \quad R^2 = 0.862$$

(a) Provide an interpretation of the constant in this equation and comment about its sign and magnitude.

(b) Calculate the predicted consumption of a hypothetical household with annual income £40,000.

A

(c) With Y_t^d on the x-axis draw a graph of the estimated MPC and APC.

Exercise 4.4

Obtain annual data for the inflation rate and the unemployment rate of a country.

(a) Estimate the following regression which is known as the Phillips curve:

$$\pi_t = a_0 + a_1 UNEMP_t + u_t$$

where π_t is inflation and $UNEMP_t$ is unemployment. Present the results in the usual way.

(b) Estimate the alternative model:

$$\pi_t - \pi_{t-1} = a_0 + a_1 UNEMP_{t-1} + u_t$$

and calculate the NAIRU (i.e. when $\pi_t - \pi_{t-1} = 0$).

(c) Reestimate the above equations splitting your sample into different decades. What factors account for differences in the results? Which period has the 'best-fitting' equation? State the criteria you have used.

Exercise 4.5

The following equation has been estimated by OLS:

$$\hat{R}_t = \underset{(0.33)}{0.567} + \underset{(0.066)}{1.045} R_{mt} \quad n = 250$$

where R_t and R_{mt} denote the excess return of a stock and the excess return of the market index for the London Stock Exchange.

(a) Derive a 95% confidence interval for each coefficient.

(b) Are these coefficients statistically significant? Explain what is the meaning of your findings regarding the CAPM theory.

(c) Test the hypothesis H_0: $\beta = 1$ and H_a: $\beta < 1$ at the 1% level of significance. If you reject H_0 what does this indicate about this stock?

Exercise 4.6

Obtain time series data on real business fixed investment (I) and an appropriate rate of interest (r). Consider the following population regression function:

$$I_t = \alpha_0 + \alpha_1 r_t + e_t$$

(a) What are the expected signs of the coefficients in this equation?

(b) Explain the rationale for each of these signs.

(c) How can you use this equation to estimate the interest elasticity of investment?

(d) Estimate the population regression function.

(e) Which coefficients are statistically significant? Are the signs those expected?

(f) Construct a 99% confidence interval for the coefficient of r_t.

(g) Estimate the log-linear version of the population regression function:

$$\ln I_t = a_0 + a_1 \ln r_t + u_t$$

(h) Is the estimated interest rate elasticity of investment significant?

(i) Do you expect this elasticity to be elastic or inelastic and why?

(j) Perform a hypothesis test of whether investment is interest-elastic.

Exercise 4.7

The file salaries_01.wf1 contains data for senior officers from a large number of UK firms. The variable *salary* is the salary that each one of them gets, measured in thousand pounds. The variable *years_senior* measures the number of years for which they are senior officers, while the variable *years_comp* measures the number of years for which they have worked in the company at the time of the research.

(a) Find summary statistics for the three above-mentioned variables and discuss them.

(b) Estimate a simple regression that explains whether and how salary level is affected by the years for which they are senior officers. Estimate another regression that now explains whether and how salary level is affected by the years for which they have worked in the same company. Report your results and comment on them. Which relationship seems to be more robust and why?

estimating the mean in expression (3.2). In fact, if there is no regressor, then b_1 does not enter expression (4.6) and the two problems are identical except for the different notation (m in expression (3.2), b_0 in expression (4.6)). Just as there is a unique estimator, $\overline{Y}$, that minimizes the expression (3.2), so is there a unique pair of estimators of β_0 and β_1 that minimize expression (4.6).

The estimators of the intercept and slope that minimize the sum of squared mistakes in expression (4.6) are called the **ordinary least squares (OLS) estimators** of β_0 and β_1.

OLS has its own special notation and terminology. The OLS estimator of β_0 is denoted $\hat{\beta}_0$, and the OLS estimator of β_1 is denoted $\hat{\beta}_1$. The **OLS regression line** is the straight line constructed using the OLS estimators, that is, $\hat{\beta}_0 + \hat{\beta}_1X$. The **predicted value** of Y_i given X_i, based on the OLS regression line, is $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1X_i$. The **residual** for the i^{th} observation is the difference between Y_i and its predicted value, that is, the residual is $\hat{u}_i = Y_i - \hat{Y}_i$.

You could compute the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ by trying different values of b_0 and b_1 repeatedly until you find those that minimize the total squared mistakes in expression (4.6); they are the least squares estimates. This method would be quite tedious, however. Fortunately there are formulas, derived by minimizing expression (4.6) using calculus, that streamline the calculation of the OLS estimators.

The OLS formulas and terminology are collected in Key Concept 4.2. These formulas are implemented in virtually all statistical and spreadsheet programs. These formulas are derived in Appendix 4.2.

OLS Estimates of the Relationship Between Test Scores and the Student-Teacher Ratio

When OLS is used to estimate a line relating the student-teacher ratio to test scores using the 420 observations in Figure 4.2, the estimated slope is −2.28 and the estimated intercept is 698.9. Accordingly, the OLS regression line for these 420 observations is

$$\widehat{TestScore} = 698.9 - 2.28 \times STR, \tag{4.7}$$

where *TestScore* is the average test score in the district and *STR* is the student-teacher ratio. The symbol "^" over *TestScore* in Equation (4.7) indicates that this is the predicted value based on the OLS regression line. Figure 4.3 plots this OLS regression line superimposed over the scatterplot of the data previously shown in Figure 4.2.

A small value of the p-value, say less than 5%, provides evidence against the null hypothesis in the sense that the chance of obtaining a value of $\hat{\beta}_1$ by pure random variation from one sample to the next is less than 5% if in fact the null hypothesis is correct. If so, the null hypothesis is rejected at the 5% significance level.

Alternatively, the hypothesis can be tested at the 5% significance level simply by comparing the value of the t-statistic to ± 1.96, the critical value for a two-sided test, and rejecting the null hypothesis at the 5% level if $|t^{act}| > 1.96$.

These steps are summarized in Key Concept 4.6.

Application to test scores. The OLS estimator of the slope coefficient, estimated using the 420 observations in Figure 4.2 and reported in Equation (4.7), is -2.28. Its standard error is 0.52, that is, $SE(\hat{\beta}_1) = 0.52$. Thus, to test the null hypothesis that $\beta_{ClassSize} = 0$, we construct the t-statistic using Equation (4.20); accordingly, $t^{act} = (-2.28 - 0)/0.52 = -4.38$.

This t-statistic exceeds the 1% two-sided critical value of 2.58, so the null hypothesis is rejected in favor of the two-sided alternative at the 1% significance level. Alternatively, we can compute the p-value associated with $t = -4.38$. This probability is the area in the tails of standard normal distribution, as shown in Figure 4.6. This probability is extremely small, approximately .00001, or .001%. That is, if the null hypothesis $\beta_{ClassSize} = 0$ is true, the probability of obtaining a value of $\hat{\beta}_1$ as far from the null as the value we actually obtained is extremely small, less than .001%. Because this event is so unlikely, it is reasonable to conclude that the null hypothesis is false.

One-Sided Hypotheses Concerning β_1

The discussion so far has focused on testing the hypothesis that $\beta_1 = \beta_{1,0}$ against the hypothesis that $\beta_1 \neq \beta_{1,0}$. This is a two-sided hypothesis test, because under the alternative β_1 could be either larger or smaller than $\beta_{1,0}$. Sometimes, however, it is appropriate to use a one-sided hypothesis test. For example, in the student-teacher ratio/test score problem, many people think that smaller classes provide a better learning environment. Under that hypothesis, β_1 is negative: smaller classes lead to higher scores. It might make sense, therefore, to test the null hypothesis that $\beta_1 = 0$ (no effect) against the one-sided alternative that $\beta_1 < 0$.

For a one-sided test, the null hypothesis and the one-sided alternative hypothesis are

$$H_0: \beta_1 = \beta_{1,0} \text{ vs. } H_1: \beta_1 < \beta_{1,0}, \text{ (one-sided alternative).} \quad (4.23)$$

Interpretation of the Regression Coefficients

The mechanics of regression with a binary regressor are the same as if it is con-tinuous. The interpretation of β_1, however, is different, and it turns out that regres-sion with a binary variable is equivalent to performing a difference of means analysis, as described in Section 3.4.

To see this, suppose you have a variable D_i that equals either 0 or 1, depend-ing on whether the student-teacher ratio is less than 20:

$$D_i = \begin{cases} 1 \text{ if the student-teacher ratio in } i^{\text{th}} \text{ district} < 20 \\ 0 \text{ if the student-teacher ratio in } i^{\text{th}} \text{ district} \geq 20. \end{cases} \tag{4.29}$$

The population regression model with D_i as the regressor is

$$Y_i = \beta_0 + \beta_1 D_i + u_i,\ i = 1, \ldots, n. \tag{4.30}$$

This is the same as the regression model with the continuous regressor X_i, except that now the regressor is the binary variable D_i. Because D_i is not continuous, it is not useful to think of β_1 as a slope; indeed, because D_i can take on only two values, there is no "line" so it makes no sense to talk about a slope. Thus we will not refer to β_1 as the slope in Equation (4.30); instead we will simply refer to β_1 as the **coefficient multiplying D_i** in this regression or, more compactly, the **coefficient on D_i**.

If β_1 in Equation (4.30) is not a slope, then what is it? The best way to inter-pret β_0 and β_1 in a regression with a binary regressor is to consider, one at a time, the two possible cases, $D_i = 0$ and $D_i = 1$. If the student-teacher ratio is high, then $D_i = 0$ and Equation (4.30) becomes

$$Y_i = \beta_0 + u_i \quad (D_i = 0). \tag{4.31}$$

Because $E(u_i \mid D_i) = 0$, the conditional expectation of Y_i when $D_i = 0$ is $E(Y_i \mid D_i = 0) = \beta_0$, that is, β_0 is the population mean value of test scores when the student-teacher ratio is high. Similarly, when $D_i = 1$,

$$Y_i = \beta_0 + \beta_1 + u_i \quad (D_i = 1). \tag{4.32}$$

Thus, when $D_i = 1$, $E(Y_i \mid D_i = 1) = \beta_0 + \beta_1$; that is, $\beta_0 + \beta_1$ is the population mean value of test scores when the student-teacher ratio is low.

Because $\beta_0 + \beta_1$ is the population mean of Y_i when $D_i = 1$ and β_0 is the pop-ulation mean of Y_i when $D_i = 0$, the difference $(\beta_0 + \beta_1) - \beta_0 = \beta_1$ is the differ-

ence between these two means. In other words, β_1 is the difference between the conditional expectation of Y_i when $D_i = 1$ and when $D_i = 0$, or $\beta_1 = E(Y_i \mid D_i = 1) - E(Y_i \mid D_i = 0)$. In the test score example, β_1 is the difference between mean test score in districts with low student-teacher ratios and the mean test score in districts with high student-teacher ratios.

Because β_1 is the difference in the population means, it makes sense that the OLS estimator β_1 is the difference between the sample averages of Y_i in the two groups, and in fact this is the case.

Hypothesis tests and confidence intervals. If the two population means are the same, then β_1 in Equation (4.30) is zero. Thus, the null hypothesis that the two population means are the same can be tested against the alternative hypoth-esis that they differ by testing the null hypothesis $\beta_1 = 0$ against the alternative $\beta_1 \neq 0$. This hypothesis can be tested using the procedure outlined in Section 4.5. Specifically, the null hypothesis can be rejected at the 5% level against the two-sided alternative when the OLS t-statistic $t = \hat{\beta}_1 / SE(\hat{\beta}_1)$ exceeds 1.96 in absolute value. Similarly, a 95% confidence interval for β_1, constructed as $\hat{\beta}_1 \pm 1.96SE(\hat{\beta}_1)$ as described in Section 4.6, provides a 95% confidence interval for the difference between the two population means.

Application to Test Scores. As an example, a regression of the test score against the student-teacher ratio binary variable D defined in Equation (4.29) esti-mated by OLS using the 420 observations in Figure 4.2, yields

$$\underset{ \quad\quad (1.3) \quad (1.8)}{\widehat{TestScore} = 650.0 + 7.4D} \tag{4.33}$$

where the standard errors of the OLS estimates of the coefficients β_0 and β_1 are given in parentheses below the OLS estimates. Thus the average test score for the subsample with student-teacher ratios greater than or equal to 20 (that is, for which $D = 0$) is 650.0, and the average test score for the subsample with student-teacher ratios less than 20 (so $D = 1$) is $650.0 + 7.4 = 657.4$. Thus the difference between the sample average test scores for the two groups is 7.4. This is the OLS estimate of β_1, the coefficient on the student-teacher ratio binary variable D.

Is the difference in the population mean test scores in the two groups statis-tically significantly different from zero at the 5% level? To find out, construct the t-statistic on β_1: $t = 7.4/1.8 = 4.04$. This exceeds 1.96 in absolute value, so the

In contrast, Figure 4.7 illustrates a case in which the conditional distribution of u_i spreads out as x increases. For small values of x, this distribution is tight, but for larger values of x, it has a greater spread. Thus, in Figure 4.7 the variance of u_i given $X_i = x$ increases with x, so that the errors in Figure 4.7 are heteroskedastic.

The definitions of heteroskedasticity and homoskedasticity are summarized in Key Concept 4.8.

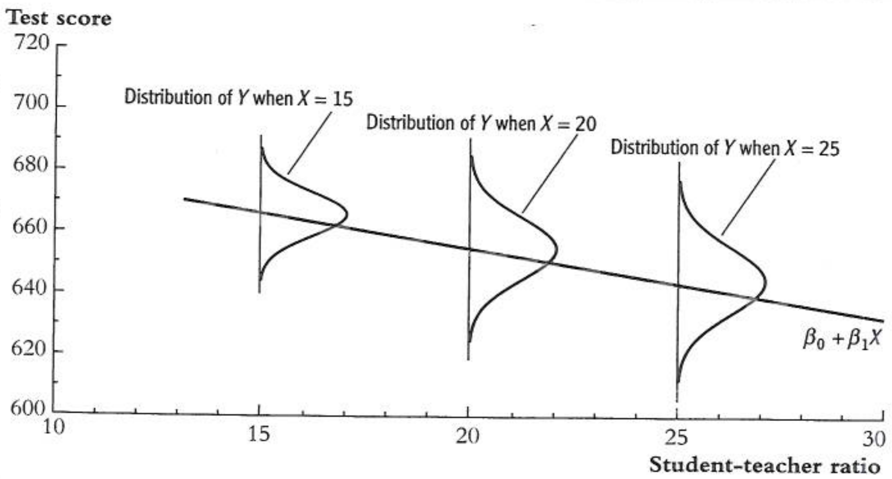
Example. These terms are a mouthful and the definitions might seem abstract. To help clarify them with an example, we digress from the student-teacher ratio/test score problem and instead return to the example of earnings of male versus female college graduates considered in Section 3.5. Let $MALE_i$ be a binary variable that equals 1 for male college graduates and equals 0 for female graduates. The binary variable regression model relating someone's earnings to his or her gender is

$$Earnings_i = \beta_0 + \beta_1 MALE_i + u_i \tag{4.41}$$

for $i = 1, \ldots, n$. Because the regressor is binary, β_1 is the difference in the population means of the two groups, in this case, the difference in mean earnings between men and women who graduated from college.

FIGURE 4.7 An Example of Heteroskedasticity

Like Figure 4.4, this shows the conditional distribution of test scores for three different class sizes. Unlike Figure 4.4, these distributions become more spread out (have a larger variance) for larger class sizes. Because the variance of the distribution of u given X, $\text{var}(u|X)$, depends on X, u is heteroskedastic.



Stw

Key Concept 4.8

Heteroskedasticity and Homoskedasticity

The error term u_i is homoskedastic if the variance of the conditional distribution of u_i given X_i, $\mathrm{var}(u_i | X_i = x)$, is constant for $i = 1, \ldots, n$, and in particular does not depend on x; otherwise, the error term is heteroskedastic.

The definition of homoskedasticity states that the variance of u_i does not depend on the regressor. Here the regressor is $MALE_i$, so at issue is whether the variance of the error term depends on $MALE_i$. In other words, is the variance of the error term the same for men and for women? If so, the error is homoskedastic; if not, it is heteroskedastic.

Deciding whether the variance of u_i depends on $MALE_i$ requires thinking hard about what the error term actually is. In this regard, it is useful to write Equation (4.41) as two separate equations, one for men and one for women:

$$Earnings_i = \beta_0 + u_i \text{ (women) and} \tag{4.42}$$

$$Earnings_i = \beta_0 + \beta_1 + u_i \text{ (men).} \tag{4.43}$$

Thus, for women, u_i is the deviation of the i^{th} woman's earnings from the population mean earnings for women (β_0), and for men, u_i is the deviation of the i^{th} man's earnings from the population mean earnings for men ($\beta_0 + \beta_1$). It follows that the statement, "the variance of u_i does not depend on $MALE$," is equivalent to the statement, "the variance of earnings is the same for men as it is for women." In other words, in this example, the error term is homoskedastic if the variance of the population distribution of earnings is the same for men and women; if these variances differ, the error term is heteroskedastic.

Mathematical Implications of Homoskedasticity

The OLS estimators remain unbiased and asymptotically normal. Because the least squares assumptions in Key Concept 4.3 place no restrictions on the

5tw

predicted value (99)
residual (99)
least squares assumptions (103)
standard error of $\hat{\beta}_1$ (112)
t-statistic (113)
p-value (113)
confidence interval for β_1 (117)
confidence level (117)
indicator variable (119)
dummy variable (119)
coeffecient multiplying variable D_1 (120)
coeffecient on D_1 (120)
regression R^2 (122)
explained sum of squares (ESS) (122)
total sum of squares (TSS) (122)
sum of squared residuals (SSR) (123)
standard error of the regression (SER) (123)
heteroskedasticity and homoskedasticity (124)
best linear unbiased estimator (BLUE) (127)
weighted least squares (127)
homoskedasticity-only standard errors (128)
heteroskedasticity-robust standard error (128)

Review the Concepts

4.1 Explain the difference between $\hat{\beta}_1$ and β_1; between the residual $\hat{u}_i$ and the regression error u_i; and between the OLS predicted value $\hat{Y}_i$ and $E(Y_i|X_i)$.

4.2 Outline the procedures for computing the p-value of a two-sided test of H_0: $\mu_Y = 0$ using an i.i.d. set of observations Y_i, $i = 1, \ldots, n$. Outline the procedures for computing the p-value of a two-sided test of H_0: $\beta_1 = 0$ in a regression model using an i.i.d. set of observations (Y_i, X_i), $i = 1, \ldots, n$.

4.3 Explain how you could use a regression model to estimate the wage gender gap using the data from Section 3.5. What are the dependent and independent variables?

4.4 Sketch a hypothetical scatterplot of data for an estimated regression with $R^2 = 0.9$. Sketch a hypothetical scatterplot of data for a regression with $R^2 = 0.5$.

Exercises

Solutions to exercises denoted by * can be found on the text Web site at **www.aw.com/stock_watson.**

$\boxed{\text{*4.1}}$ Suppose that a researcher, using data on class size (CS) and average test scores from 100 third-grade classes, estimates the OLS regression,

$$\widehat{TestScore} = \underset{(20.4)}{520.4} - \underset{(2.21)}{5.82} \times CS,\ R^2 = 0.08,\ SER = 11.5.$$

a. A classroom has 22 students. What is the regression's prediction for that classroom's average test score?

b. Last year a classroom had 19 students, and this year it has 23 students. What is the regression's prediction for the change in the classroom average test score?

c. Construct a 95% confidence interval for β_1, the regression slope coefficient.

$\boxed{\text{d.}}$ Calculate the p-value for the two-sided test of the null hypothesis H_0: $\beta_1 = 0$. Do you reject the null hypothesis at the 5% level? At the 1% level?

e. The sample average class size across the 100 classrooms is 21.4. What is the sample average of the test scores across the 100 classrooms? (*Hint:* Review the formulas for the OLS estimators.)

f. What is the sample standard deviation of test scores across the 100 classrooms? (*Hint:* Review the formulas for the R^2 and SER.)

$\boxed{\text{4.2}}$ Suppose that a researcher, using wage data on 250 randomly selected male workers and 280 female workers, estimates the OLS regression,

$$\widehat{Wage} = \underset{(0.18)}{12.68} + \underset{(0.84)}{2.79}\ Male,\ R^2 = 0.06,\ SER = 3.10$$

where $Wage$ is measured in \$/hour and $Male$ is a binary variable that is equal to one if the person is a male and 0 if the person is a female. Define the wage gender gap as the difference in mean earnings between men and women.

a. What is the estimated gender gap?

b. Is the estimated gender gap significantly different from zero? (Compute the p-value for testing the null hypothesis that there is no gender gap.)

c. Construct a 95% confidence interval for the gender gap.

d. In the sample, what is the mean wage of women? Of men?

e. Another researcher uses these same data, but regresses $Wages$ on $Female$, a variable that is equal to one if the person is female and zero if the person a male. What are the regression estimates calculated from this regression?

$$\widehat{Wage} = \_\_\_\_ + \_\_\_\_\ Female,\ R^2 = \_\_\_\_,\ SER = \_\_\_\_.$$

***4.3** Show that the first least squares assumption, $E(u_i|X_i) = 0$, implies that $E(Y_i|X_i) = \beta_0 + \beta_1 X_i$.

Stw

4.4 Show t hat $\hat{\beta}_0$ is an unbiased estimator of β_0. (Hint: use the fact that $\hat{\beta}_1$ is unbiased, which is shown in Appendix 4.3).

4.5 Suppose that a random sample of 200 20-year-old men is selected from a population and their height and weight is recorded. A regression of weight on height yields:

$$\widehat{Weight} = \underset{(2.15)}{-99.41} + \underset{(0.31)}{3.94}\, Height,\ R^2 = 0.81,\ SER = 10.2,$$

where *Weight* is measured in pounds and *Height* is measured in inches.

a. What is the regression's weight prediction for someone who is 70 inches tall? 65 inches tall? 74 inches tall?

b. A person has a late growth spurt and grows 1.5 inches over the course of a year. What is the regression's prediction for the increase in the person's weight?

c. Construct a 99% confidence interval for the weight gain in (b).

d. Suppose that instead of measuring weight and height in pounds and inches, they are measured in kilograms and centimeters. What are the regression estimates from this new kilogram-centimeter regression? (Give all results, estimated coefficients, standard errors, R^2, and *SER*.)

4.6 Starting from Equation (4.15), derive the variance of $\hat{\beta}_0$ under homoskedasticity given in Equation (4.61) in Appendix 4.4.

APPENDIX 4.1 The California Test Score Data Set

The California Standardized Testing and Reporting data set contains data on test performance, school characteristics, and student demographic backgrounds. The data used here are from all 420 K–6 and K–8 districts in California with data available for 1998 and 1999.

W

EXAMPLE 2.1

(Soybean Yield and Fertilizer)

Suppose that soybean yield is determined by the model

$$yield = \beta_0 + \beta_1 fertilizer + u, \tag{2.3}$$

so that $y = yield$ and $x = fertilizer$. The agricultural researcher is interested in the effect of fertilizer on yield, holding other factors fixed. This effect is given by β_1. The error term u contains factors such as land quality, rainfall, and so on. The coefficient β_1 measures the effect of fertilizer on yield, holding other factors fixed: $\Delta yield = \beta_1 \Delta fertilizer$.

EXAMPLE 2.2

(A Simple Wage Equation)

A model relating a person's wage to observed education and other unobserved factors is

$$wage = \beta_0 + \beta_1 educ + u. \tag{2.4}$$

If *wage* is measured in dollars per hour and *educ* is years of education, then β_1 measures the change in hourly wage given another year of education, holding all other factors fixed. Some of those factors include labor force experience, innate ability, tenure with current employer, work ethic, and innumerable other things.

The linearity of (2.1) implies that a one-unit change in x has the *same* effect on y, regardless of the initial value of x. This is unrealistic for many economic applications. For example, in the wage-education example, we might want to allow for *increasing* returns: the next year of education has a *larger* effect on wages than did the previous year. We will see how to allow for such possibilities in Section 2.4.

The most difficult issue to address is whether model (2.1) really allows us to draw ceteris paribus conclusions about how x affects y. We just saw in equation (2.2) that β_1 *does* measure the effect of x on y, holding all other factors (in u) fixed. Is this the end of the causality issue? Unfortunately, no. How can we hope to learn in general about the ceteris paribus effect of x on y, holding other factors fixed, when we are ignoring all those other factors?

Section 2.5 will show that we are only able to get reliable estimators of β_0 and β_1 from a random sample of data when we make an assumption restricting how the unobservable u is related to the explanatory variable x. Without such a restriction, we will not be able to estimate the ceteris paribus effect, β_1. Because u and x are random variables, we need a concept grounded in probability.

Before we state the key assumption about how x and u are related, we can always make one assumption about u. As long as the intercept β_0 is included in the equation, nothing is lost by assuming that the average value of u in the population is zero.

W

we consider statistical properties after we explicitly impose assumptions on the population model equation (2.1).

EXAMPLE 2.3

(CEO Salary and Return on Equity)

For the population of chief executive officers, let y be annual salary (*salary*) in thousands of dollars. Thus, $y = 856.3$ indicates an annual salary of \$856,300, and $y = 1452.6$ indicates a salary of \$1,452,600. Let x be the average return on equity (*roe*) for the CEO's firm for the previous three years. (Return on equity is defined in terms of net income as a percentage of common equity.) For example, if $roe = 10$, then average return on equity is 10 percent.

To study the relationship between this measure of firm performance and CEO compensation, we postulate the simple model

$$salary = \beta_0 + \beta_1 roe + u.$$

The slope parameter β_1 measures the change in annual salary, in thousands of dollars, when return on equity increases by one percentage point. Because a higher *roe* is good for the company, we think $\beta_1 > 0$.

The data set CEOSAL1.RAW contains information on 209 CEOs for the year 1990; these data were obtained from *Business Week* (5/6/91). In this sample, the average annual salary is \$1,281,120, with the smallest and largest being \$223,000 and \$14,822,000, respectively. The average return on equity for the years 1988, 1989, and 1990 is 17.18 percent, with the smallest and largest values being 0.5 and 56.3 percent, respectively.

Using the data in CEOSAL1.RAW, the OLS regression line relating *salary* to *roe* is

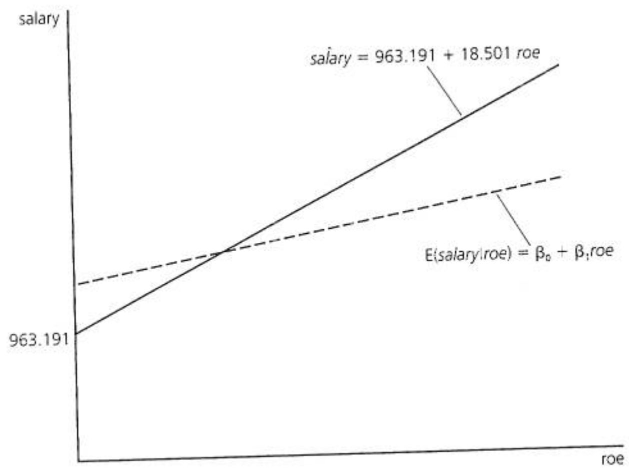
$$\widehat{salary} = 963.191 + 18.501\, roe, \tag{2.26}$$

where the intercept and slope estimates have been rounded to three decimal places; we use "*salary* hat" to indicate that this is an estimated equation. How do we interpret the equation? First, if the return on equity is zero, $roe = 0$, then the predicted *salary* is the intercept, 963.191, which equals \$963,191 since *salary* is measured in thousands. Next, we can write the predicted change in salary as a function of the change in *roe*: $\Delta\widehat{salary} = 18.501\,(\Delta roe)$. This means that if the return on equity increases by one percentage point, $\Delta roe = 1$, then *salary* is predicted to change by about 18.5, or \$18,500. Because (2.26) is a linear equation, this is the estimated change regardless of the initial salary.

We can easily use (2.26) to compare predicted salaries at different values of *roe*. Suppose $roe = 30$. Then $\widehat{salary} = 963.191 + 18.501(30) = 1518.221$, which is just over \$1.5 million. However, this does *not* mean that a particular CEO whose firm had a $roe = 30$ earns \$1,518,221. Many other factors affect salary. This is just our prediction from the OLS regression line (2.26). The estimated line is graphed in Figure 2.5, along with the population regression function E(*salary*|*roe*). We will never know the PRF, so we cannot tell how close the SRF is to the PRF. Another sample of data will give a different regression line, which may or may not be closer to the population regression line.

W

Figure 2.5
The OLS regression line $\hat{salary} = 963.191 + 18.501\ roe$ and the (unknown) population regression function.



EXAMPLE 2.4

(Wage and Education)

For the population of people in the workforce in 1976, let $y = wage$, where *wage* is measured in dollars per hour. Thus, for a particular person, if $wage = 6.75$, the hourly *wage* is \$6.75. Let $x = educ$ denote years of schooling; for example, $educ = 12$ corresponds to a complete high school education. Since the average wage in the sample is \$5.90, the consumer price index indicates that this amount is equivalent to \$16.64 in 1997 dollars.
Using the data in WAGE1.RAW where $n = 526$ individuals, we obtain the following OLS regression line (or sample regression function):

$$\hat{wage} = -0.90 + 0.54\ educ. \tag{2.27}$$

We must interpret this equation with caution. The intercept of −0.90 literally means that a person with no education has a predicted hourly wage of −90 cents an hour. This, of course, is silly. It turns out that only 18 people in the sample of 526 have less than eight years of education. Consequently, it is not surprising that the regression line does poorly at very low levels of education. For a person with eight years of education, the predicted wage is $\hat{wage} = -0.90 + 0.54(8) = 3.42$, or \$3.42 per hour (in 1976 dollars).

3

QUESTION 2.2

The estimated wage from (2.27), when $educ = 8$, is \$3.42 in 1976 dollars. What is this value in 1997 dollars? (*Hint:* You have enough information in Example 2.4 to answer this question.)

The slope estimate in (2.27) implies that one more year of education increases hourly wage by 54 cents an hour. Therefore, four more years of education increase the predicted wage by $4(0.54) = 2.16$, or \$2.16 per hour. These are fairly large effects. Because of the linear nature of (2.27), another year of education increases the wage by the same amount, regardless of the initial level of education. In Section 2.4, we discuss some methods that allow for nonconstant marginal effects of our explanatory variables.

EXAMPLE 2.5

(Voting Outcomes and Campaign Expenditures)

The file VOTE1.RAW contains data on election outcomes and campaign expenditures for 173 two-party races for the U.S. House of Representatives in 1988. There are two candidates in each race, A and B. Let *voteA* be the percentage of the vote received by Candidate A and *shareA* be the the percentage of total campaign expenditures accounted for by Candidate A. Many factors other than *shareA* affect the election outcome (including the quality of the candidates and possibly the dollar amounts spent by A and B). Nevertheless, we can estimate a simple regression model to find out whether spending more relative to one's challenger implies a higher percentage of the vote.
The estimated equation using the 173 observations is

$$\hat{voteA} = 26.81 + 0.464\ shareA. \tag{2.28}$$

This means that if the share of Candidate A's spending increases by one percentage point, Candidate A receives almost one-half a percentage point (0.464) more of the total vote. Whether or not this is a causal effect is unclear, but it is not unbelievable. If $shareA = 50$, *voteA* is predicted to be about 50, or half the vote.

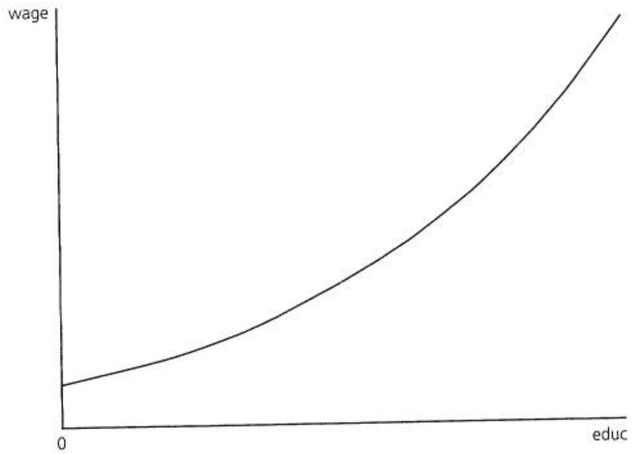
In some cases, regression analysis is not used to determine causality but to simply look at whether two variables are positively or negatively related, much like a standard correlation analysis. An example of this occurs in Problem 2.12, where you are asked to use data from Biddle and Hamermesh (1990) on time spent sleeping and working to investigate the tradeoff between these two factors.

QUESTION 2.3

In Example 2.5, what is the predicted vote for Candidate A if $shareA = 60$ (which means 60 percent)? Does this answer seem reasonable?

W

Figure 2.6
wage = exp($\beta_0 + \beta_1 educ$), with $\beta_1 > 0$.



Notice how we multiply β_1 by 100 to get the percentage change in *wage* given one additional year of education. Since the percentage change in *wage* is the same for each additional year of education, the change in *wage* for an extra year of education *increases* as education increases; in other words, (2.42) implies an *increasing* return to education. By exponentiating (2.42), we can write $wage = \exp(\beta_0 + \beta_1 educ + u)$. This equation is graphed in Figure 2.6, with $u = 0$.

Estimating a model such as (2.42) is straightforward when using simple regression. Just define the dependent variable, y, to be $y = \log(wage)$. The independent variable is represented by $x = educ$. The mechanics of OLS are the same as before: the intercept and slope estimates are given by the formulas (2.17) and (2.19). In other words, we obtain $\hat{\beta}_0$ and $\hat{\beta}_1$ from the OLS regression of log(*wage*) on *educ*.

EXAMPLE 2.10

(A Log Wage Equation)

Using the same data as in Example 2.4, but using log(*wage*) as the dependent variable, we obtain the following relationship:

$$\widehat{\log(wage)} = 0.584 + 0.083\ educ \tag{2.44}$$

$$n = 526,\ R^2 = 0.186.$$

The coefficient on *educ* has a percentage interpretation when it is multiplied by 100: *wage* increases by 8.3 percent for every additional year of education. This is what economists mean when they refer to the "return to another year of education."

It is important to remember that the main reason for using the log of *wage* in (2.42) is to impose a constant percentage effect of education on *wage*. Once equation (2.42) is obtained, the natural log of *wage* is rarely mentioned. In particular, it is *not* correct to say that another year of education increases log(*wage*) by 8.3 percent.

The intercept in (2.42) is not very meaningful, as it gives the predicted log(*wage*), when $educ = 0$. The *R*-squared shows that *educ* explains about 18.6 percent of the variation in log(*wage*) (*not wage*). Finally, equation (2.44) might not capture all of the nonlinearity in the relationship between wage and schooling. If there are "diploma effects," then the twelfth year of education—graduation from high school—could be worth much more than the eleventh year. We will learn how to allow for this kind of nonlinearity in Chapter 7.

Another important use of the natural log is in obtaining a **constant elasticity model**.

EXAMPLE 2.11

(CEO Salary and Firm Sales)

We can estimate a constant elasticity model relating CEO salary to firm sales. The data set is the same one used in Example 2.3, except we now relate *salary* to *sales*. Let *sales* be annual firm sales, measured in millions of dollars. A constant elasticity model is

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + u, \tag{2.45}$$

where β_1 is the elasticity of *salary* with respect to *sales*. This model falls under the simple regression model by defining the dependent variable to be $y = \log(salary)$ and the independent variable to be $x = \log(sales)$. Estimating this equation by OLS gives

$$\widehat{\log(salary)} = 4.822 + 0.257 \log(sales) \tag{2.46}$$

$$n = 209,\ R^2 = 0.211.$$

The coefficient of log(*sales*) is the estimated elasticity of *salary* with respect to *sales*. It implies that a 1 percent increase in firm sales increases CEO salary by about 0.257 percent—the usual interpretation of an elasticity.

The two functional forms covered in this section will often arise in the remainder of this text. We have covered models containing natural logarithms here because they appear so frequently in applied work. The interpretation of such models will not be much different in the multiple regression case.

W

$$kids = \beta_0 + \beta_1 educ + u,$$

where u is the unobserved error.

(i) What kinds of factors are contained in u? Are these likely to be correlated with level of education?

(ii) Will a simple regression analysis uncover the ceteris paribus effect of education on fertility? Explain.

2.2 In the simple linear regression model $y = \beta_0 + \beta_1 x + u$, suppose that $E(u) \neq 0$. Letting $\alpha_0 = E(u)$, show that the model can always be rewritten with the same slope, but a new intercept and error, where the new error has a zero expected value.

$\boxed{\textbf{2.3}}$ The following table contains the *ACT* scores and the *GPA* (grade point average) for 8 college students. Grade point average is based on a four-point scale and has been rounded to one digit after the decimal.

Student	*GPA*	*ACT*
1	2.8	21
2	3.4	24
3	3.0	26
4	3.5	27
5	3.6	29
6	3.0	25
7	2.7	25
8	3.7	30

(i) Estimate the relationship between *GPA* and *ACT* using OLS; that is, obtain the intercept and slope estimates in the equation

$$\hat{GPA} = \hat{\beta}_0 + \hat{\beta}_1 ACT.$$

Comment on the direction of the relationship. Does the intercept have a useful interpretation here? Explain. How much higher is the *GPA* predicted to be if the *ACT* score is increased by 5 points?

(ii) Compute the fitted values and residuals for each observation, and verify that the residuals (approximately) sum to zero.

(iii) What is the predicted value of *GPA* when $ACT = 20$?

(iv) How much of the variation in *GPA* for these 8 students is explained by *ACT*? Explain.

2.4 The data set BWGHT.RAW contains data on births to women in the United States. Two variables of interest are the dependent variable, infant birth weight in ounces (*bwght*), and an explanatory variable, average number of cigarettes the mother smoked per day during pregnancy (*cigs*). The following simple regression was estimated using data on $n = 1388$ births:

$$\hat{bwght} = 119.77 - 0.514\, cigs$$

(i) What is the predicted birth weight when $cigs = 0$? What about when $cigs = 20$ (one pack per day)? Comment on the difference.

(ii) Does this simple regression necessarily capture a causal relationship between the child's birth weight and the mother's smoking habits? Explain.

(iii) To predict a birth weight of 125 ounces, what would *cigs* have to be? Comment.

(iv) What fraction of the women in the sample do not smoke while pregnant? Does this help reconcile your finding from part (iii)?

2.5 In the linear consumption function

$$\hat{cons} = \hat{\beta}_0 + \hat{\beta}_1 inc,$$

the (estimated) *marginal propensity to consume* (MPC) out of income is simply the slope, $\hat{\beta}_1$, while the *average propensity to consume* (APC) is $\hat{cons}/inc = \hat{\beta}_0/inc + \hat{\beta}_1$. Using observations for 100 families on annual income and consumption (both measured in dollars), the following equation is obtained:

$$\hat{cons} = -124.84 + 0.853\, inc$$

$$n = 100,\ R^2 = 0.692.$$

(i) Interpret the intercept in this equation, and comment on its sign and magnitude.

(ii) What is the predicted consumption when family income is \$30,000?

(iii) With *inc* on the *x*-axis, draw a graph of the estimated MPC and APC.

2.6 Using data from 1988 for houses sold in Andover, Massachusetts, from Kiel and McClain (1995), the following equation relates housing price (*price*) to the distance from a recently built garbage incinerator (*dist*):

$$\log(\hat{price}) = 9.40 + 0.312 \log(dist)$$

$$n = 135,\ R^2 = 0.162.$$

(i) Interpret the coefficient on $\log(dist)$. Is the sign of this estimate what you expect it to be?

(ii) Do you think simple regression provides an unbiased estimator of the ceteris paribus elasticity of *price* with respect to *dist*? (Think about the city's decision on where to put the incinerator.)

(iii) What other factors about a house affect its price? Might these be correlated with distance from the incinerator?

2.7 Consider the savings function

$$sav = \beta_0 + \beta_1 inc + u,\ u = \sqrt{inc}\cdot e,$$