4.2

A random sample of 100 20-year-old men is selected from a population and these men’s height and weight are recorded. A regression of weight on height yields *Weight* = -79.24 + 4.16 \* *Height*, *R*2 = 0.72, *SER* = 12.6, where *Weight* is measured in pounds and *Height* is measured in inches.

**What is the regression’s weight prediction for someone who is 64 inches tall? 68 inches tall? 72 inches tall?**

**A man has a late growth spurt and grows 2 inches over the course of a year. What is the regression’s prediction for the increase in this man’s weight?**

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

5.4

Read the box “The Economic Value of a Year of Education: Homoskedasticity or Heteroskedasticity?” in Section 5.4. Use the regression reported in Equation (5.23) to answer the following.

A randomly selected 30-year-old worker reports an education level of 16 years. What is the worker’s expected average hourly earnings?

**A high school graduate (12 years of education) is contemplating going to a community college for a 2-year degree. How much are this worker’s average hourly earnings expected to increase?**

**A high school counselor tells a student that, on average, college graduates earn $10 per hour more than high school graduates. Is this statement consistent with the regression evidence? What range of values is consistent with the regression evidence?**

6.2-6.4 – 6.6, 6.7

**6.2** Using the regression results in column (1):

**Do workers with college degrees earn more, on average, than workers with only high school diplomas? How much more?**

**Do men earn more than women, on average? How much more?**

**6.3 Using** the regression results in column (2):

**Is age an important determinant of earnings? Explain.**

**Sally is a 29-year-old female college graduate. Betsy is a 34-year-old female college graduate. Predict Sally’s and Betsy’s earnings.**

**6.4 Using** the regression results in column (3):

**Do there appear to be important regional differences?**

**Why is the regressor *West* omitted from the regression? What would happen if it were included?**

**Juanita is a 28-year-old female college graduate from the South. Jennifer is a 28-year-old female college graduate from the Midwest. Calculate the expected difference in earnings between Juanita and Jennifer.**

A researcher plans to study the causal effect of a strong legal system on the number of scandals in a country, using data from a random sample of countries in Asia. The researcher plans to regress the number of scandals on how strong a legal system is in the countries (an indicator variable taking the value 1 or 0, based on expert opinion).

**Do you think this regression suffers from omitted variable bias? Explain why. Which variables would you add to the regression?**

**Using the expression for omitted variable bias given in Equation (6.1), assess whether the regression will likely over- or underestimate the effect of a strong legal system on the number of scandals in a country. That is, do you think that bn1 7 b1 or bn1 6 b1?**

**6.7**

Critique each of the following proposed research plans. Your critique should explain any problems with the proposed research and describe how the research plan might be improved. Include a discussion of any additional data that need to be collected and the appropriate statistical techniques for analyzing those data.

**A researcher wants to determine whether a leading global university  
is guilty of racial bias in admissions. To determine potential bias, the researcher collects data on the race of all applicants to the university  
for a given year. The researcher plans to conduct a difference-in-means test to determine whether the proportion of acceptances among Black candidates is systematically different from the proportion of acceptances among other candidates.**

**A researcher is interested in identifying the impact of a mother’s education on the educational attainment of her child. She collects data on a random sample of individuals aged between 25 and 40 years who are out of the schooling system. The data set contains information on each person’s level of schooling, the type of school attended, gender and ethnicity, as well as information on the schooling of their parents and the demographic characteristics of the household in which they grew up. The researcher plans to regress years of schooling achieved by an individual on the years of schooling of their mother, including in the regression  
the other potential determinants of schooling (number of siblings and whether parents lived together or are separated) as controls.**

**7.2**

1. **7.2 Using** the regression results in column (1):

**Is the high school earnings difference estimated from this regression statistically significant at the 5% level? Construct a 95% confidence interval of the difference.**

**Is the male–female earnings difference estimated from this regression statistically significant at the 5% level? Construct a 95% confidence interval for the difference.**

1. **7.3 Using** the regression results in column (2):

**Is age an important determinant of earnings? Use an appropriate statistical test and/or confidence interval to explain your answer.**

**Suppose Alvo is a 30-year-old male college graduate, and Kal is a 40-year-old male college graduate. Construct a 95% confidence interval for the expected difference between their earnings.**

1. **7.4 Using** the regression results in column (3):

**Are there any important regional differences? Use an appropriate hypothesis test to explain your answer.**

**Juan is a 32-year-old male high school graduate from the North. Mel is a 32-year-old male college graduate from the West. Ari is a 32-year-old male college graduate from the East. Construct a 95% confidence interval for the difference in expected earnings between Juan and Mel. Explain how you would construct a 95% confidence interval for the difference in expected earnings between Juan and Ari. (*Hint:* What would happen if you included *West* and excluded *East* from the regression?)**

**9.1 and 9.3**

**9.1 Explain** the difference between internal validity and external validity. Is it possible for an econometric study to have internal validity but not external validity?

**9.3 What** is the effect of measurement error in *Y*? How is this different from the effect of measurement error in *X*?

**Chapter 8 Review the Concepts**

**8.2 A** Cobb–Douglas production function relates production (*Q*) to factors of production—capital (*K*), labor (*L*), and raw materials (*M*)—and an error term *u* using the equation *Q* = l*K*b1*L*b2*M*b3*eu*, where l, b1, b2, and b3 are production parameters. Suppose you have data on production and the factors of production from a random sample of firms with the same Cobb–Douglas production function. How would you use regression analysis to estimate the production parameters?

**8.3 How** is the slope coefficient interpreted in a log-linear model, where the independent variable is in logarithms, but the dependent variable is not? In a linear-log model? In a log-log model?

**Chapter 8 Exercises**

**8.2 Suppose** a researcher collects data on houses that have sold in a particular neighborhood over the past year and obtains the regression results in the following table.

**Using the results in column (1), what is the expected change in price of building a 1500-square-foot addition to a house? Construct a 99% confidence interval for the percentage change in price.**

**How is the coefficient on ln (*Size*) interpreted in column (2)? What is the effect of a doubling of the size of a house on its price?**

**Using column (2), what is the estimated effect of view on price? Con- struct a 99% confidence interval for this effect. Is the effect statistically different from 0?**

**Using the results from the regression in column (3), calculate the effect of adding two bedrooms to a house. Is the effect statistically significant? Which of the two variables—size or number of bedrooms—do you think is relatively more important in determining the price of a house?**

**Is the coefficient on condition significant in column (4)?**

**Is the interaction term between *Pool* and *View* statistically significant in column (5)? Find the effect of adding a view on the price of a house with a pool, as well as a house without a pool.**

**8.4** Read the box “The Effect of Ageing on Healthcare Expenditures: A Red Herring?” in Section 8.3.

**Consider a male aged 60 years. Use the results from column (1) of Table 8.1 and the method in Key Concept 8.1 to estimate the expected change in the logarithm of health care expenditures (*HCE*) associated with an additional year of age.**

**Repeat (a), assuming a man aged 70 years.**

**Explain why the answers to (a) and (b) are different.**

**Is the difference in the answers to (a) and (b) statistically significant at the 5% level? Explain.**

**How would you change the regression if you suspected that the effect of age on HCE was different for men than for women?**

**8.6**Refer to Table 8.3.

**A researcher suspects that the effect of *%Eligible for subsidized lunch* has a nonlinear effect on test scores. In particular, he conjectures that increases in this variable from 10% to 20% have little effect on test scores but that changes from 50% to 60% have a much larger effect.**

**Describe a nonlinear specification that can be used to model this form of nonlinearity.**

**How would you test whether the researcher’s conjecture was better than the linear specification in column (7) of Table 8.3?**

**A researcher suspects that the effect of income on test scores is different in districts with small classes than in districts with large classes.**

**Describe a nonlinear specification that can be used to model this form of nonlinearity.**

**How would you test whether the researcher’s conjecture was better than the linear specification in column (7) of Table 8.3?**

**Review The Concept Chapter 12**

**12.2 Describe** the key characteristics of a valid instrument. If you were a researcher, how would you determine if the variable you have selected for an endogenous regressor is a valid instrument or not?

**12.3 In** their study of the effect of institutions on economic development, sup- pose Acemoglu et al. had used the prevalence of malaria as an instrument. Would this instrument be relevant? Would it be exogenous? Would it be a valid instrument?

**12.5** Consider the IV regression model

*Yi* =b0 +b1*Xi* +b2*Wi* +*ui*,

where *Xi* is correlated with *ui* and *Zi* is an instrument. Suppose that the first three assumptions in Key Concept 12.4 are satisfied. Which IV assumption is not satisfied when?

***Zi* is independent of (*Yi*, *Xi*, *Wi*)?**

***Zi* = *Wi*?**

***Wi* = 1 for all *i*?**

***Zi* = *Xi*?**

**12.6**  Suppose a researcher is considering developing an IV regression model with one regressor ,*Xi* ,and one instrument ,*Zi* .If she has a sample of *n* = 113,what range must the correlation coefficient be between *Xi* and *Zi* in order for *Zi* to be considered a strong instrument? [*Hint:* See Equation (7.14).]

**12.9 A** researcher is interested in the effect of more secure property rights on income across countries. He collects recent data from 60 countries and runs the OLS regression *Yi* = b0 + b1*Xi* + *ui*, where *Yi* is a country’s GDP per capita and *Xi* is an index taking values between 0 and 10 reflecting the protection against expropriation where a higher value indicates greater protection against expropriation, that is, more secure property rights.

**a.** Explain why the OLS estimates are likely to be unreliable and indicate in which direction they might be biased. (*Hint:* In which direction does causality run in this example?)

**b.** All the countries in the researcher’s sample were former colonies. Institutions securing property rights could originate from early institutions established alongside European settlements. The decision for Europeans to settle or otherwise could reflect concerns for mortality among settlers. Explain how settler mortality might be used as an instrument to estimate the effect of more secure property rights on income across countries.

Chapter 13 Review

**13.2 A** clinical trial is carried out for a new cholesterol-lowering drug. The drug is given to 500 patients, and a placebo is given to another 500 patients, using random assignment of the patients. How would you estimate the treatment effect of the drug? Suppose you had data on the weight, age, and sex of each patient. Could you use these data to improve your estimate? Explain. Suppose you had data on the cholesterol level of each patient before he or she entered the experiment. Could you use these data to improve your estimate? Explain.

**13.5 Consider** the quasi-experiment described in Section 13.4 involving the draft lottery, military service, and civilian earnings. Explain why there might be heterogeneous effects of military service on civilian earnings; that is, explain why b1*i* in Equation (13.9) depends on *i*. Explain why there might be heterogeneous effects of the lottery outcome on the probability of military service; that is, explain why p1*i* in Equation (13.11) depends on *i*. If there are heterogeneous responses of the sort you described

Exercises

**13.1 How** would you calculate the small class treatment effect from the results in Table 13.1? Can you distinguish this treatment effect from the aide treatment effect? How would you have to change the program to correctly estimate both effects?

**13.2 For** the following calculations, use the results in column (3) of Table 13.2. Consider two classrooms, A and B, which have identical values of the regressors in column (3) of Table 13.2, except that:

**Classroom A is a small class, and classroom B is a regular-sized class. Construct a 90% confidence interval for the expected difference in average test scores.**

**Classroom A has a teacher with 6 years of experience, and classroom B has a teacher with 12 years of experience. Construct a 95% confidence interval for the expected difference in average test scores.**

**Classroom A is a small-sized class with a teacher with 6 years of experience, and classroom B is a regular-sized class with a teacher with 12 years of experience. Construct a 95% confidence interval for the expected difference in average test scores. (*Hint:* In STAR, the teachers were randomly assigned to the different types of classrooms.)**

**Why is the intercept missing from column (4)?**

**13.5 Consider** a study to evaluate the effect on college student grades of dorm room Internet connections. In a large dorm, half the rooms are randomly wired for high-speed Internet connections (the treatment group), and final course grades are collected for all residents. Which of the following pose threats to internal validity, and why?

**Midway through the year all the male athletes move into a fraternity and drop out of the study. (Their final grades are not observed.)**

**Engineering students assigned to the control group put together a local area network so that they can share a private wireless Internet connection that they pay for jointly.**

**The art majors in the treatment group never learn how to access their Internet accounts.**

**The economics majors in the treatment group provide access to their Internet connection to those in the control group, for a fee.**