1 Using the data in GPA2.RAW on 4,137 college students, the following equation was estimated by OLS:

\[ \overline{\text{colgpa}} = 1.392 - .0135 \text{hsperc} + .00148 \text{sat} \]
\[ n = 4,137, R^2 = .273, \]

where \( \overline{\text{colgpa}} \) is measured on a four-point scale, \( \text{hsperc} \) is the percentile in the high school graduating class (defined so that, for example, \( \text{hsperc} = 5 \) means the top 5% of the class), and \( \text{sat} \) is the combined math and verbal scores on the student achievement test.

(i) Why does it make sense for the coefficient on \( \text{hsperc} \) to be negative?
(ii) What is the predicted college GPA when \( \text{hsperc} = 20 \) and \( \text{sat} = 1,050 \)?
(iii) Suppose that two high school graduates, A and B, graduated in the same percentile from high school, but Student A’s SAT score was 140 points higher (about one standard deviation in the sample). What is the predicted difference in college GPA for these two students? Is the difference large?
(iv) Holding \( \text{hsperc} \) fixed, what difference in SAT scores leads to a predicted \( \overline{\text{colgpa}} \) difference of .50, or one-half of a grade point? Comment on your answer.

2 The data in WAGE2.RAW on working men was used to estimate the following equation:

\[ \overline{\text{educ}} = 10.36 - .094 \text{sibs} + .131 \text{meduc} + .210 \text{feduc} \]
\[ n = 722, R^2 = .214, \]

where \( \text{educ} \) is years of schooling, \( \text{sibs} \) is number of siblings, \( \text{meduc} \) is mother’s years of schooling, and \( \text{feduc} \) is father’s years of schooling.

(i) Does \( \text{sibs} \) have the expected effect? Explain. Holding \( \text{meduc} \) and \( \text{feduc} \) fixed, by how much does \( \text{sibs} \) have to increase to reduce predicted years of education by one year? (A noninteger answer is acceptable here.)
(ii) Discuss the interpretation of the coefficient on \( \text{meduc} \).
(iii) Suppose that Man A has no siblings, and his mother and father each have 12 years of education. Man B has no siblings, and his mother and father each have 16 years of education. What is the predicted difference in years of education between B and A?

3 The following model is a simplified version of the multiple regression model used by Biddle and Hamermesh (1990) to study the tradeoff between time spent sleeping and working and to look at other factors affecting sleep:

\[ \text{sleep} = \beta_0 + \beta_1 \text{totwrk} + \beta_2 \text{educ} + \beta_3 \text{age} + u, \]

where \( \text{sleep} \) and \( \text{totwrk} \) (total work) are measured in minutes per week and \( \text{educ} \) and \( \text{age} \) are measured in years. (See also Computer Exercise C3 in Chapter 2.)

(i) If adults trade off sleep for work, what is the sign of \( \beta_1 \)?
(ii) What signs do you think \( \beta_2 \) and \( \beta_3 \) will have?
(iii) Using the data in SLEEP75.RAW, the estimated equation is

\[
\bar{\text{sleep}} = 3.638.25 - .148 \text{totwrk} - 11.13 \text{educ} + 2.20 \text{age}
\]

\[n = 706, \, R^2 = .113.\]

If someone works five more hours per week, by how many minutes is \(\text{sleep}\) predicted to fall? Is this a large tradeoff?

(iv) Discuss the sign and magnitude of the estimated coefficient on \(\text{educ}\).

(v) Would you say \(\text{totwrk}, \text{educ},\) and \(\text{age}\) explain much of the variation in \(\text{sleep}\)? What other factors might affect the time spent sleeping? Are these likely to be correlated with \(\text{totwrk}\)?

9 The following equation describes the median housing price in a community in terms of amount of pollution (\(\text{nox}\) for nitrous oxide) and the average number of rooms in houses in the community (\(\text{rooms}\)):

\[
\log(\text{price}) = \beta_0 + \beta_1 \log(\text{nox}) + \beta_2 \text{rooms} + u.
\]

(i) What are the probable signs of \(\beta_1\) and \(\beta_2\)? What is the interpretation of \(\beta_1\)? Explain.

(ii) Why might \(\text{nox}\) [or more precisely, \(\log(\text{nox})\)] and \(\text{rooms}\) be negatively correlated? If this is the case, does the simple regression of \(\log(\text{price})\) on \(\log(\text{nox})\) produce an upward or a downward biased estimator of \(\beta_1\)?

(iii) Using the data in HPRICE2.RAW, the following equations were estimated:

\[
\log(\text{price}) = 11.71 - 1.043 \log(\text{nox}), \, n = 506, \, R^2 = .264.
\]

\[
\log(\text{price}) = 9.23 - .718 \log(\text{nox}) + .306 \text{rooms}, \, n = 506, \, R^2 = .514.
\]

Is the relationship between the simple and multiple regression estimates of the elasticity of \(\text{price}\) with respect to \(\text{nox}\) what you would have predicted, given your answer in part? (ii) Does this mean that \(-.718\) is definitely closer to the true elasticity than \(-1.043\)?
C2 Use the data in HPRICE1.RAW to estimate the model

\[ price = \beta_0 + \beta_sqrft + \beta_{bdrms} + u, \]

where \( price \) is the house price measured in thousands of dollars.

(i) Write out the results in equation form.
(ii) What is the estimated increase in price for a house with one more bedroom, holding square footage constant?
(iii) What is the estimated increase in price for a house with an additional bedroom that is 140 square feet in size? Compare this to your answer in part (ii).
(iv) What percentage of the variation in price is explained by square footage and number of bedrooms?
(v) The first house in the sample has \( sqrft = 2,438 \) and \( bdrms = 4 \). Find the predicted selling price for this house from the OLS regression line.
(vi) The actual selling price of the first house in the sample was $300,000 (so \( price = 300 \)). Find the residual for this house. Does it suggest that the buyer underpaid or overpaid for the house?

C3 The file CEOSAL2.RAW contains data on 177 chief executive officers and can be used to examine the effects of firm performance on CEO salary.

(i) Estimate a model relating annual salary to firm sales and market value. Make the model of the constant elasticity variety for both independent variables. Write the results out in equation form.
(ii) Add \( profits \) to the model from part (i). Why can this variable not be included in logarithmic form? Would you say that these firm performance variables explain most of the variation in CEO salaries?
(iii) Add the variable \( ceoten \) to the model in part (ii). What is the estimated percentage return for another year of CEO tenure, holding other factors fixed?
(iv) Find the sample correlation coefficient between the variables \( \log(mkval) \) and \( profits \). Are these variables highly correlated? What does this say about the OLS estimators?
Use the data in DISCRIM.RAW to answer this question. These are ZIP code–level data on prices for various items at fast-food restaurants, along with characteristics of the zip code population, in New Jersey and Pennsylvania. The idea is to see whether fast-food restaurants charge higher prices in areas with a larger concentration of blacks.

(i) Find the average values of prpblick and income in the sample, along with their standard deviations. What are the units of measurement of prpblick and income?

(ii) Consider a model to explain the price of soda, psoda, in terms of the proportion of the population that is black and median income:

\[ psoda = \beta_0 + \beta_1prpblick + \beta_2income + u. \]

Estimate this model by OLS and report the results in equation form, including the sample size and R-squared. (Do not use scientific notation when reporting the estimates.) Interpret the coefficient on prpblick. Do you think it is economically large?

(iii) Compare the estimate from part (ii) with the simple regression estimate from psoda on prpblick. Is the discrimination effect larger or smaller when you control for income?

(iv) A model with a constant price elasticity with respect to income may be more appropriate. Report estimates of the model

\[ \log(psoda) = \beta_0 + \beta_1prpblick + \beta_2\log(income) + u. \]

If prpblick increases by .20 (20 percentage points), what is the estimated percentage change in psoda? (Hint: The answer is 2.xx, where you fill in the “xx.”)

(v) Now add the variable prppov to the regression in part (iv). What happens to \( \hat{\beta}_{prpblick} \)?

(vi) Find the correlation between \( \log(income) \) and prppov. Is it roughly what you expected?

(vii) Evaluate the following statement: “Because \( \log(income) \) and prppov are so highly correlated, they have no business being in the same regression.”