Chapter 5

QUESTIONS

5.1. Explain briefly what is meant by
   a. Log-log model
   b. Log-lin model
   c. Lin-log model
   d. Elasticity coefficient
   e. Elasticity at mean value

5.2. What is meant by a slope coefficient and an elasticity coefficient? What is the relationship between the two?

5.3. Fill in the blanks in Table 5-12.

TABLE 5-12 FUNCTIONAL FORMS OF REGRESSION MODELS

Model	When appropriate
$\ln Y_i = B_1 + B_2 \ln X_i$	—
$\ln Y_i = B_1 + B_2 X_i$	—
$Y_i = B_1 + B_2 \ln X_i$	—
$Y_i = B_1 + B_2 \left(\frac{1}{X_i}\right)$	—

5.4. Complete the following sentences:
   a. In the double-log model the slope coefficient measures . . .
   b. In the lin-log model the slope coefficient measures . . .
   c. In the log-lin model the slope coefficient measures . . .
   d. Elasticity of Y with respect to X is defined as . . .
   e. Price elasticity is defined as . . .
   f. Demand is said to be elastic if the absolute value of the price elasticity is . . . , but demand is said to be inelastic if it is . . .

5.5. State with reason whether the following statements are true (T) or false (F):
   a. For the double-log model, the slope and elasticity coefficients are the same.
   b. For the linear-in-variable (LIV) model, the slope coefficient is constant but the elasticity coefficient is variable, whereas for the log-log model, the elasticity coefficient is constant but the slope is variable.
   c. The R^2 of a log-log model can be compared with that of a log-lin model but not with that of a lin-log model.
   d. The R^2 of a lin-log model can be compared with that of a linear (in variables) model but not with that of a double-log or log-lin model.
   e. Model A: $\ln Y = -0.6 + 0.4X$; $r^2 = 0.85$
   Model B: $\hat{Y} = 1.3 + 2.2X$; $r^2 = 0.73$
   Model A is a better model because its r^2 is higher.

5.6. The *Engel expenditure curve* relates a consumer's expenditure on a commodity to his or her total income. Letting Y = the consumption expenditure on a commodity and X = the consumer income, consider the following models:
   a. $Y_i = B_1 + B_2 X_i + u_i$
   b. $Y_i = B_1 + B_2(1/X_i) + u_i$
   c. $\ln Y_i = B_1 + B_2 \ln X_i + u_i$

d. $\ln Y_i = B_1 + B_2(1/X_i) + u_i$
e. $Y_i = B_1 + B_2 \ln X_i + u_i$
f. $\ln(Y) = B_1 - B_2\left(\frac{1}{X}\right)$. This model is known as the log-inverse model.

Which of these models would you choose for the Engel curve and why? (*Hint:* Interpret the various slope coefficients, find out the expressions for elasticity of expenditure with respect to income, etc.)

5.7. The growth model Eq. (5.18) was fitted to several U.S. economic time series and the following results were obtained:

Time series and period	B_1	B_2	r^2
Real GNP (1954–1987) (1982 dollars)	7.2492 $t = (529.29)$	0.0302 (44.318)	0.9839
Labor force participation rate (1973–1987)	4.1056 $t = (1290.8)$	0.053 (15.149)	0.9464
S&P 500 index (1954–1987)	3.6960 $t = (57.408)$	0.0456 (14.219)	0.8633
S&P 500 index (1954–1987 quarterly data)	3.7115 $t = (114.615)$	0.0114 (27.819)	0.8524

a. In each case find out the instantaneous rate of growth.
b. What is the compound rate of growth in each case?
c. For the S&P data, why is there a difference in the two slope coefficients? How would you reconcile the difference?

PROBLEMS

5.8. Refer to the cubic total cost (TC) function given in Eq. (5.32).

a. The marginal cost (MC) is the change in the TC for a unit change in output; that is, it is the rate of change of the TC with respect to output. (Technically, it is the derivative of the TC with respect to X, the output.) Derive this function from regression (5.32).
b. The average variable cost (AVC) is the total variable cost (TVC) divided by the total output. Derive the AVC function from regression (5.32).
c. The average cost (AC) of production is the TC of production divided by total output. For the function given in regression (5.32), derive the AC function.
d. Plot the various cost curves previously derived and confirm that they resemble the stylized textbook cost curves.

5.9. Are the following models linear in the parameters? If not, is there any way to make them linear-in-parameter (LIP) models?

a. $Y_i = \dfrac{1}{B_1 + B_2 X_i}$
b. $Y_i = \dfrac{X_i}{B_1 + B_2 X_i^2}$

5.10. Based on 11 annual observations, the following regressions were obtained:

Model A: $\hat{Y}_t = 2.6911 - 0.4795X_t$
$\text{se} = (0.1216)\ (0.1140) \quad r^2 = 0.6628$

$$\text{Model B: } \ln \hat{Y}_t = 0.7774 - 0.2530 \ln X_t$$
$$\text{se} = (0.0152)\ (0.0494) \quad r^2 = 0.7448$$

where Y = the cups of coffee consumed per person per day and X = the price of coffee in dollars per pound.

a. Interpret the slope coefficients in the two models.

b. You are told that $\overline{Y} = 2.43$ and $\overline{X} = 1.11$. At these mean values, estimate the price elasticity for Model A.

c. What is the price elasticity for Model B?

d. From the estimated elasticities, can you say that the demand for coffee is price inelastic?

e. How would you interpret the intercept in Model B? (*Hint:* Take the antilog.)

f. Since the r^2 of Model B is larger than that of Model A, Model B is preferable to Model A. Comment on this statement.

5.11. Refer to the Cobb-Douglas production function given in regression (5.11).

a. Interpret the coefficient of the labor input X_2. Is it statistically different from 1?

b. Interpret the coefficient of the capital input X_3. Is it statistically different from zero? And from 1?

c. What is the interpretation of the intercept value of -1.6524?

d. Test the hypothesis that $B_2 = B_3 = 0$.

5.12. In their study of the demand for international reserves (i.e., foreign reserve currency such as the dollar or International Monetary Fund [IMF] drawing rights), Mohsen Bahami-Oskooee and Margaret Malixi[31] obtained the following regression results for a sample of 28 less developed countries (LDC):

$$\ln(R/P) = 0.1223 + 0.4079 \ln(Y/P) + 0.5040 \ln \sigma_{BP} - 0.0918 \ln \sigma_{EX}$$
$$t = (2.5128)\ (17.6377)\qquad (15.2437)\ (-2.7449)$$
$$R^2 = 0.8268$$
$$F = 1151$$
$$n = 1120$$

where R = the level of nominal reserves in U.S. dollars
P = U.S. implicit price deflator for GNP
Y = the nominal GNP in U.S. dollars
σ_{BP} = the variability measure of balance of payments
σ_{EX} = the variability measure of exchange rates

(*Notes:* The figures in parentheses are t ratios. This regression was based on quarterly data from 1976 to 1985 (40 quarters) for each of the 28 countries, giving a total sample size of 1120.)

a. A priori, what are the expected signs of the various coefficients? Are the results in accord with these expectations?

b. What is the interpretation of the various partial slope coefficients?

[31] See Mohsen Bahami-Oskooee and Margaret Malixi, "Exchange Rate Flexibility and the LDCs Demand for International Reserves," *Journal of Quantitative Economics,* vol. 4, no. 2, July 1988, pp. 317–328.

c. Test the statistical significance of each estimated partial regression coefficient (i.e., the null hypothesis is that *individually* each true or population regression coefficient is equal to zero).

d. How would you test the hypothesis that all partial slope coefficients are simultaneously zero?

5.13. Based on the U.K. data on annual percentage change in wages (Y) and the percent annual unemployment rate (X) for the years 1950 to 1966, the following regression results were obtained:

$$\hat{Y}_t = -1.4282 + 8.7243\left(\frac{1}{X_t}\right)$$

$$\text{se} = (2.0675) \qquad (2.8478) \qquad r^2 = 0.3849$$

$$F(1,15) = 9.39$$

a. What is the interpretation of 8.7243?

b. Test the hypothesis that the estimated slope coefficient is not different from zero. Which test will you use?

c. How would you use the F test to test the preceding hypothesis?

d. Given that $\overline{Y} = 4.8$ percent and $\overline{X} = 1.5$ percent, what is the rate of change of Y at these mean values?

e. What is the elasticity of Y with respect to X at the mean values?

f. How would you test the hypothesis that the true $r^2 = 0$?

5.14. Table 5-13 gives data on the Consumer Price Index, $Y(1980 = 100)$, and the money supply, X (billions of German marks), for Germany for the years 1971 to 1987.

TABLE 5-13 CONSUMER PRICE INDEX (Y) (1980 = 100) AND THE MONEY SUPPLY (X) (MARKS, IN BILLIONS), GERMANY, 1971–1987

Year	Y	X
1971	64.1	110.02
1972	67.7	125.02
1973	72.4	132.27
1974	77.5	137.17
1975	82.0	159.51
1976	85.6	176.16
1977	88.7	190.80
1978	91.1	216.20
1979	94.9	232.41
1980	100.0	237.97
1981	106.3	240.77
1982	111.9	249.25
1983	115.6	275.08
1984	118.4	283.89
1985	121.0	296.05
1986	120.7	325.73
1987	121.1	354.93

Source: International Economic Conditions, annual ed., June 1988, The Federal Reserve Bank of St. Louis, p. 24.

a. Regress the following:
1. Y on X
2. $\ln Y$ on $\ln X$
3. $\ln Y$ on X
4. Y on $\ln X$
b. Interpret each estimated regression.
c. For each model, find the rate of change of Y with respect to X.
d. For each model, find the elasticity of Y with respect to X. For some of these models, the elasticity is to be computed at the mean values of Y and X.
e. Based on all these regression results, which model would you choose and why?

5.15. Based on the following data, estimate the model:

$$\left(\frac{1}{Y_i}\right) = B_1 + B_2X_i + u_i$$

Y	86	79	76	69	65	62	52	51	51	48
X	3	7	12	17	25	35	45	55	70	120

a. What is the interpretation of B_2?
b. What is the rate of change of Y with respect to X?
c. What is the elasticity of Y with respect to X?
d. For the same data, run the regression

$$Y_i = B_1 + B_2\left(\frac{1}{X_i}\right) + u_i$$

e. Can you compare the r^2s of the two models? Why or why not?
f. How do you decide which is a better model?

5.16. *Comparing two r^2s when dependent variables are different.*[32] Suppose you want to compare the r^2 values of the growth model (5.19) with the linear trend model (5.23) of the consumer credit outstanding regressions given in the text. Proceed as follows:
a. Obtain $\ln \hat{Y}_t$, that is, the estimated log value of each observation from model (5.19).
b. Obtain the antilog values of the values obtained in (*a*).
c. Compute r^2 between the values obtained in (*b*) and the actual Y values using the definition of r^2 given in Question 3.5.
d. This r^2 value is comparable with the r^2 value obtained from linear model (5.23).

Use the preceding steps to compare the r^2 values of models (5.19) and (5.23).

5.17. Based on the GNP/money supply data given in Table 5-14 (found on the textbook's Web site), the following regression results were obtained (Y = GNP, X = M2):

[32]For additional details and numerical computation, see Gujarati and Porter, *Basic Econometrics*, 5th ed., McGraw-Hill, New York, 2009, pp. 203–205.

Model	Intercept	Slope	r^2
Log-linear	0.7826 $t = 11.40$	0.8539 $t = 108.93$	0.997
Log-lin (growth model)	7.2392 $t = 80.85$	0.0001 $t = 14.07$	0.832
Lin-log	−24299 $t = -15.45$	3382.4 $t = 18.84$	0.899
Linear (LIV model)	703.28 $t = 8.04$	0.4718 $t = 65.58$	0.991

a. For each model, interpret the slope coefficient.
b. For each model, estimate the elasticity of the GNP with respect to money supply and interpret it.
c. Are all r^2 values directly comparable? If not, which ones are?
d. Which model will you choose? What criteria did you consider in your choice?
e. According to the monetarists, there is a one-to-one relationship between the rate of changes in the money supply and the GDP. Do the preceding regressions support this view? How would you test this formally?

5.18. Refer to the energy demand data given in Table 5-3. Instead of fitting the log-linear model to the data, fit the following linear model:

$$Y_t = B_1 + B_2X_{2t} + B_3X_{3t} + u_t$$

a. Estimate the regression coefficients, their standard errors, and obtain R^2 and adjusted R^2.
b. Interpret the various regression coefficients
c. Are the estimated partial regression coefficients *individually* statistically significant? Use the p values to answer the question.
d. Set up the ANOVA table and test the hypothesis that $B_2 = B_3 = 0$.
e. Compute the income and price elasticities at the mean values of Y, X_2, and X_3. How do these elasticities compare with those given in regression (5.12)?
f. Using the procedure described in Problem 5.16, compare the R^2 values of the linear and log-linear regressions. What conclusion do you draw from these computations?
g. Obtain the normal probability plot for the residuals obtained from the linear-in-variable regression above. What conclusions do you draw?
h. Obtain the normal probability plot for the residuals obtained from the log-linear regression (5.12) and decide whether the residuals are approximately normally distributed.
i. If the conclusions in (g) and (h) are different, which regression would you choose and why?

5.19. To explain the behavior of business loan activity at large commercial banks, Bruce J. Summers used the following model:[33]

$$Y_t = \frac{1}{A + Bt} \qquad \textbf{(A)}$$

[33] See his article, "A Time Series Analysis of Business Loans at Large Commercial Banks," *Economic Review*, Federal Reserve Bank of St. Louis, May/June, 1975, pp. 8–14.

where Y is commercial and industrial (C&I) loans in millions of dollars, and t is time, measured in months. The data used in the analysis was collected monthly for the years 1966 to 1967, a total of 24 observations.

For estimation purposes, however, the author used the following model:

$$\frac{1}{Y_t} = A + Bt \qquad \textbf{(B)}$$

The regression results based on this model for banks including New York City banks and excluding New York City banks are given in Equations (1) and (2), respectively:

$$\frac{1}{Y_t} = 52.00 - 0.2t$$

$$t = (96.13)\,(-24.52) \qquad \bar{R}^2 = 0.84 \qquad \textbf{(1)}$$

$$\frac{\hat{1}}{Y_t} = 26.79 - 0.14t \qquad \text{DW} = 0.04^*$$

$$t = (196.70)\,(-66.52) \qquad \bar{R}^2 = 0.97 \qquad \textbf{(2)}$$

$$\text{DW} = 0.03^*$$

*Durbin-Watson (DW) statistic (see Chapter 10).

a. Why did the author use Model (B) rather than Model (A)?
b. What are the properties of the two models?
c. Interpret the slope coefficients in Models (1) and (2). Are the two slope coefficients statistically significant?
d. How would you find out the standard errors of the intercept and slope coefficients in the two regressions?
e. Is there a difference in the behavior of New York City and the non–New York City banks in their C&I activity? How would you go about testing the difference, if any, formally?

5.20. Refer to regression (5.31).
a. Interpret the slope coefficient.
b. Using Table 5-11, compute the elasticity for this model. Is this elasticity constant or variable?

5.21. Refer to the data given in Table 5-5 (found on the textbook's Web site). Fit an appropriate Engle curve to the various expenditure categories in relation to total personal consumption expenditure and comment on the statistical results.

5.22. Table 5-15 gives data on the annual rate of return Y (%) on Afuture mutual fund and a return on a market portfolio as represented by the Fisher Index, X (%). Now consider the following model, which is known in the finance literature as the *characteristic line*.

$$Y_t = B_1 + B_2X_i + u_i \qquad \textbf{(1)}$$

TABLE 5-15 ANNUAL RATES OF RETURN (%) ON AFUTURE FUND (Y) AND ON THE FISHER INDEX (X), 1971–1980

Year	Y	X
1971	67.5	19.5
1972	19.2	8.5
1973	−35.2	−29.3
1974	−42.0	−26.5
1975	63.7	61.9
1976	19.3	45.5
1977	3.6	9.5
1978	20.0	14.0
1979	40.3	35.3
1980	37.5	31.0

Source: Haim Levy and Marshall Sarnat, *Portfolio and Investment Selection: Theory and Practice*, Prentice-Hall International, Englewood Cliffs, N.J., 1984, pp. 730, 738.

In the literature there is no consensus about the prior value of B_1. Some studies have shown it to be positive and statistically significant and some have shown it to be statistically insignificant. In the latter case, Model (1) becomes a regression-through-the-origin model, which can be written as

$$Y_t = B_2 X_t + u_t \tag{2}$$

Using the data given in Table 5-15, to estimate both these models and decide which model fits the data better.

5.23. *Raw R^2 for the regression-through-the-origin model.* As noted earlier, for the regression-through-the-origin regression model the conventionally computed R^2 may not be meaningful. One suggested alternative for such models is the so-called "raw" R^2, which is defined (for the two-variable case) as follows:

$$\text{Raw } r^2 = \frac{\left(\sum X_i Y_i\right)^2}{\sum X_i^2 \sum Y_i^2}$$

If you compare the raw R^2 with the traditional r^2 computed from Eq. (3.43), you will see that the sums of squares and cross-products in the raw r^2 are not mean-corrected.

For model (2) in Problem 5.22 compute the raw r^2. Compare this with the r^2 value that you obtained for Model (1) in Problem (5.22). What general conclusion do you draw?

5.24. For regression (5.39) compute the raw r^2 value and compare it with that given in Eq. (5.40).

5.25. Consider data on the weekly stock prices of Qualcomm, Inc., a digital wireless telecommunications designer and manufacturer, over the time period of 1995 to 2000. The complete data can be found in Table 5-16 on the textbook's Web site.

a. Create a scattergram of the closing stock price over time. What kind of pattern is evident in the plot?
b. Estimate a linear model to predict the closing stock price based on time. Does this model seem to fit the data well?
c. Now estimate a squared model by using both time and time-squared. Is this a better fit than in part (*b*)?
d. Now attempt to fit a cubic or third-degree polynomial to the data as follows:

$$Y_i = B_0 + B_1X_i + B_2X_i^2 + B_3X_i^3 + u_i$$

where Y = stock price and X = time. Which model seems to be the best estimator for the stock prices?

5.26. Table 5-17 on the textbook's Web site contains data about several magazines. The variables are: magazine name, cost of a full-page ad, circulation (projected, in thousands), percent male among the predicted readership, and median household income of readership. The goal is to predict the advertisement cost.
a. Create scattergrams of the cost variable versus each of the three other variables. What types of relationships do you see?
b. Estimate a linear regression equation with all the variables and create a residuals versus fitted values plot. Does the plot exhibit constant variance from left to right?
c. Now estimate the following mixed model:

$$\ln Y_i = B_0 + B_1 \ln Circ + B_2\, PercMale + B_3\, MedIncome + u_i$$

and create another residual plot. Does this model fit better than the one in part (*b*)?

5.27. Refer to Example 4.5 (Table 4-6) about education, GDP, and population for 38 countries.
a. Estimate a linear (LIV) model for the data. What are the resulting equation and relevant output values (i.e., F statistic, t values, and R^2)?
b. Now attempt to estimate a log-linear model (where both of the independent variables are also in the natural log format).
c. With the log-linear model, what does the coefficient of the GDP variable indicate about education? What about the population variable?
d. Which model is more appropriate?

5.28. Table 5-18 on the textbook's Web site contains data on average life expectancy for 40 countries. It comes from the *World Almanac and Book of Facts, 1993*, by Pharos Books. The independent variables are the ratio of the number of people per television set and the ratio of number of people per physician.
a. Try fitting a linear (LIV) model to the data. Does this model seem to fit well?
b. Create two scattergrams, one of the natural log of life expectancy versus the natural log of people per television, and one of the natural log of life expectancy versus the natural log of people per physician. Do the graphs appear linear?
c. Estimate the equation for a log-linear model. Does this model fit well?

d. What do the coefficients of the log-linear model indicate about the relationships of the variables to life expectancy? Does this seem reasonable?

5.29 Refer to Example 5.6 in the chapter. It was shown that the percentage change in the index of hourly earnings and the unemployment rate from 1958–1969 followed the traditional Phillips curve model. An updated version of the data, from 1965–2007, can be found in Table 5-19 on the textbook's Web site.

a. Create a scattergram using the percentage change in hourly earnings as the Y variable and the unemployment rate as the X variable. Does the graph appear linear?

b. Now create a scattergram as above, but use $1/X$ as the independent variable. Does this seem better than the graph in part (*a*)?

c. Fit Eq. (5.29) to the new data. Does this model seem to fit well? Also create a regular linear (LIV) model as in Eq. (5.30). Which model is better? Why?

Chapter 7

7.2 The June 1997 issue of *Management Accounting* gave the following rule for predicting your current salary if you are a managerial accountant. Take \$31,865. Next add \$20,811 if you are top management, add \$3,604 if you are senior management, or subtract \$11,419 if you are entry management. Then add \$1,105 for every year you have been a managerial accountant. Add \$7,600 if you have a master's degree or subtract \$12,467 if you have no college degree. Add \$11,527 if you have a professional certification. Finally, add \$8,667 if you are male.

a. How do you think the journal derived this method of estimating an accountant's current salary? Be specific.

b. How could a managerial accountant use this information to determine whether he or she is significantly underpaid? Be as specific as possible.

7.3 Management proposed the following regression model to predict sales at a fast-food outlet:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \varepsilon_i$$

where y_i = sales (\$1000s)
$x_{1,i}$ = number of competitors within one mile
$x_{2,i}$ = population within one mile (1000s)
$x_{3,i}$ = 1 if a drive up window is present and 0 otherwise

The following estimated regression equation was developed after 20 outlets were surveyed:

$$\hat{y}_i = 10.1 - 4.2x_{1,i} + 6.8x_{2,i} + 15.3x_{3,i}$$

a. What is the expected amount of sales attributable to the drive-up window?

b. Predict sales for a store with two competitors, a population of 8,000 within one mile, and no drive-up window.

c. Predict sales for a store with one competitor, a population of 3,000 within one mile, and a drive-up window.

7.4 The white male professors at the University of Shucks Corner (USC) are threatening to sue the university for sexual discrimination. The kingpin behind the threatened suit regressed annual salaries (in thousands of dollars) of all the university professors (S) on the following variables:

D_{MA} = 1 if the professor is male, 0 otherwise
D_A = 1 if the professor is an associate professor, 0 otherwise
D_F = 1 if the professor is a full professor, 0 otherwise

And they obtained the following results (standard errors are in parentheses):

$$\underset{(10.0)}{S = 44.5} - \underset{(1.0)}{2.2D_{MA}} + \underset{(3.0)}{3.2D_A} + \underset{(5.2)}{18.4D_F}$$

where n = 120
$R^2 = 0.82$

Note: There are three ranks of professors (assistant, associate, and full professors listed from junior to senior) at USC.

a. Formally test the hypothesis that there is no sexual discrimination at USC. State your hypothesis clearly, show all work, etc. Use the tables provided if needed. Based on this formal test, what conclusion would you draw?

b. What assumptions are you implicitly making when you conducted the test in part (a)? Offer all assumptions you are making.

c. What, if anything, can you say about the average salary of male assistant professors relative to a male full professor based on the regression results?

d. Another fast-thinking professor thinks that the preceding model is not the most appropriate model to estimate. She thinks that either one of the following models will be much easier to interpret and help get at the issue more directly.

(1) $\hat{S}_i = \beta_0 + \beta_1 D_{MA,i} + \beta_2 D_{FE,i} + \beta_3 D_{A,i} + \beta_4 D_{F,i} + \beta_5 D_{AT,i} + \varepsilon_i$

(2) $\hat{S}_i = \beta_1 D_{MA,i} + \beta_2 D_{FE,i} + \beta_3 D_{F,i} + \beta_4 D_{AT,i} + \varepsilon_i$

where all variables are as defined earlier and

$D_{FE} = 1$ if the professor is female, 0 otherwise

$D_{AT} = 1$ if the professor is an assistant professor, 0 otherwise

Discuss the feasibility of the two proposed options and their advantages and disadvantages over the original model (if any).

e. Suppose you have been hired by the university to defend USC. What would be your approach in terms of trying to discredit the analysis of those suing the university?

7.5 The following table contains the OLS coefficient estimates of the *Income* model, where $Income_i$ is a dependent variable. Estimated standard errors are reported in parentheses below the coefficient estimates.

$Income_i$ = Annual income of individual i, measured in thousands of dollars

Age_i = Age of individual i, measured in years

$Experience_i$ = Work experience of individual i, measured in years

$Female_i$ = 1 if individual i is female, 0 otherwise

	(1)	(2)
Intercept	9.6849	45.667
	(3.826)	(16.220)
Age	0.40762	−1.4581
	(.103)	(.938)
Age Squared	—	0.0118
		(.012)
Experience	—	1.1994
		(.454)
Female	—	−57.737
		(30.090)
Female × Age	—	2.9376
		(1.679)
Female × Age Squared	—	−0.0317
		(.022)
Female × Experience	—	−0.967
		(.803)
R-squared	0.168	0.5926
SSR	5,207.7	2,549.9
Observations	80	80

a. According to regression (1), explain how age is related to income.

b. Is age statistically significant in regression (1)? Be as specific as possible when answering this question.

c. The results of regression (2) conflict with these results. Why?

d. There is one formal test to see whether regression (2) or regression (1) is preferable. State your null and alternative hypotheses. Calculate the required test statistic, state the decision rule you use, and the inference you draw from the test. Test the hypothesis at a 5 percent significance level.

e. In regression (2), what is the marginal effect of being female on income?

7.6 Suppose that we are trying to explain wage using gender and years of education.

a. Write out a model that allows the intercept between males and females to differ. Define your dummy variable, and explain what the coefficients mean.

b. Now write out a model that allows the slopes to differ as well. Now explain what the coefficients mean.

c. If women with no education started out at a lower salary than men with no education, but then had larger returns to education, what sign would you expect your coefficients from part (b) to have? Draw the relevant picture for the above scenario.

7.7 Suppose you are interested in estimating how a faculty member's salary is related to whether they have moved to a different university during their career and whether they moved to or from California. Regressions are given in the accompanying table. The dependent variable is the natural log of annual salary. The independent variables are a 1 if the person moved and 0 otherwise; number of years since last moved is a 1 if the person moved to California and 0 otherwise; number of years since moved to California is 1 if the person moved from California and 0 otherwise; number of years since moved from California, number of years of experience, number of years of experience squared, Male is 1 if the person is male and 0 otherwise, the number of top 5 articles the faculty member has published, the number of top 36 articles the faculty member has published, and number of other articles published. Standard errors are in the parentheses below the coefficient estimates.

	(1) All Faculty	(2) All Faculty
Moved	0.1134	0.0490
	(0.0203)	(0.0215)
Years since last move	−0.0063	−0.0058
	(0.0010)	(0.0010)
Moved to CA	—	0.1271
	—	(0.0562)
Years since move to CA	—	−0.0022
	—	(0.0031)
Moved from CA	—	0.1265
	—	(0.0399)
Years since move from CA	—	−0.0017
	—	(0.0020)
Experience	0.0113	0.0115
	.0029)	(0.0028)
Experience squared	−0.00015	−0.00015
	(0.00007)	(0.00006)
Male	0.0066	0.0141
	(0.0190)	(0.0190)
Top 5 articles	0.0357	0.0308
	(0.0034)	(0.0033)
Top 36 articles	0.0056	0.0063
	(0.0029)	(0.0028)
Other articles	0.0035	0.0037
	(0.0008)	(0.0008)
R-square	0.4794	0.5025
Observations	1024	1024

a. Consider the regression results contained in column 1. Explain the marginal effects of moved, experience, male, and top 5 articles, and comment if the coefficient estimates are statistically significant at the 5 percent level.

b. Now consider the results contained in column 2. Explain what the marginal effects on moved to California and moved from California mean, and comment if each of these are statistically significant.

c. Is regression (1) or regression (2) statistically preferred at the 5 percent level? Provide the hypothesis, test statistic, rejection rule, and your decision.

d. Where does the experience term reach a maximum? What does this number mean to a faculty member?

e. If you were going to give advice to a faculty member on how to maximize their salary using your preferred model from part (d), what would you tell them?

Exercises

E7.1 Suppose you think that there are diminishing returns to studying. Test this hypothesis by opening up the file survey data.xls and creating a new variable *Hours study squared.*

a. Regress *GPA* on *Hours studied* and *Hours studied squared.* Explain how *Hours studied* affects *GPA*, and comment on the statistical significance of the independent variables at the 5 percent level.

b. Where does *Hours studied* reach a maximum?

c. Now estimate a more fully specified model by regressing *GPA* on *Hours studied, Hours studied squared, Work, Video game, Texts,* and *Male.* Interpret the marginal effect of *Hours studied,* and comment on the statistical significance of all of the independent variables at the 5 percent level.

E7.2 A human resources manager at ABC Inc. is trying to determine the variable that best explains the variation of employee salaries using the sample of 52 full-time employees in the file ABC.xls. Create a table of correlation coefficients to help this manager identify whether the employee's (a) gender, (b) age, (c) number of years of relevant work experience prior to employment at Beta, (d) the number of years of employment at Beta, or (e) the number of years of postsecondary education has the strongest linear relationship with annual salary. Which of these variables has the weakest linear relationship with annual salary?

E7.3 The National Basketball Association (NBA) records a variety of statistics for each team. Four of these statistics are the proportion of games won (*PCT*), the proportion of field goals made by the team (*FG%*), the proportion of three-point shots made by the team's opponent (*OPP 3 PT%*), and the number of turnovers committed by the team's opponent (*Opp TO*). The data contained in NBA.xls show the values of these statistics for the 29 teams in the NBA for a portion of the 2004 season (www.nba.com, January 3, 2004).

a. Determine the estimated regression equation that can be used to predict the proportion of games won given the proportion of field goals made by the team.

b. Provide an interpretation for the slope of the estimated regression equation developed in part (a).

c. Determine the estimated regression equation that can be used to predict the proportion of games won given the proportion of field goals made by the team, the proportion of three-point shots made by the team's opponent, and the number of turnovers committed by the team's opponent.

d. Discuss the implications of the estimated regression equations developed in part (c).

e. Estimate the proportion of games won for a team with the following values for the three independent variables: *FG%* = 0.45, *Opp 3 Pt%* = 0.34, and *Opp TO* = 17.

f. Does the estimated regression equation provide a good fit for the data? Explain.

g. Use the *F*-test to determine the overall significance of the relationship. What is your conclusion at the 0.05 level of significance?

h. Use the *t*-test to determine the significance of each independent variable. What is your conclusion at the 0.05 level of significance?

E7.4 The National Football League rates prospects by position on a scale that ranges from 5 to 9. The ratings are interpreted as follows: 8–9 should start the first year; 7–7.9

should start; 6–6.9 will make the team as backup; and 5–5.9 can make the club and contribute. The file Football.xls shows the position, weight, speed (for 40 yards), and ratings for 25 NFL prospects (*USA Today*, April 14, 2000).

a. Develop a dummy variable that will account for the player's position.

b. Develop an estimated regression equation to show how rating is related to position, weight, and speed.

c. At the 0.05 level of significance, test whether the estimated regression equation indicates a significant relationship between the independent variables and the dependent variable.

d. Does the estimated regression equation provide a good fit for the data? Explain.

e. Is position a significant factor in the payer's rating? Use a significance level of 0.05. Explain.

f. Suppose that a new offensive tackle prospect who weighs 300 pounds ran the 40 yards in 5.1 seconds. Use the estimated regression equation developed in part (b) to estimate the ratings for this player. What will this player's role be on the team next year?

E7.5 Open file Housing prices.xls.

a. Regress housing price on square feet, bedrooms, lot size, and pool. Explain what each of the coefficients mean, and comment on the statistical significance for each coefficient at a 5 percent significance level.

b. Now suppose that you think that an interaction effect exists between bedrooms and lot size. Create a new variable in Excel to reflect this, and add this variable to your regression you estimated in part (a). Comment on the statistical significance of all of the variables at a 5 percent level, and explain the marginal effects of both bedrooms and square feet.

E7.6 Open file Housing prices.xls.

a. Create a new variable log housing price by using the ln function in Excel. Regress log housing price on square feet, bedrooms, lot size, and pool. Explain what each of the coefficients mean, and comment on the statistical significance for each coefficient at a 5 percent significance level.

b. Create another new variable log square feet. Regress log housing price on log square feet, bedrooms, lot size, and pool. Explain what each of the coefficients mean, and comment on the statistical significance for each coefficient at a 5 percent significance level.

E7.7 Open file Video game.xls. Regress units sold on time since announcement, PS3 consoles sold, average rating, online MP, sequel, part of series, and multi-platform. Comment on the statistical significance of each variable at the 5 percent level. What variables would you suggest keeping, and which ones would you not include? Why? Using an F-test, test whether the variables you would omit are jointly significantly significant at a 5 percent level.