a. How would you interpret the coefficients of *P*, *I*, and *P′*?
b. Is the demand for coffee price elastic?
c. Are coffee and tea substitute or complementary products?
d. How would you interpret the coefficient of *t*?
e. What is the trend rate of growth or decline in coffee consumption in the United States? If there is a decline in coffee consumption, what accounts for it?
f. What is the income elasticity of demand for coffee?
g. How would you test the hypothesis that the income elasticity of demand for coffee is not significantly different from 1?
h. What do the dummy variables represent in this case?
i. How do you interpret the dummies in this model?
j. Which of the dummies are statistically significant?
k. Is there a pronounced seasonal pattern in coffee consumption in the United States? If so, what accounts for it?
l. Which is the benchmark quarter in this example? Would the results change if we chose another quarter as the base quarter?
m. The preceding model only introduces the *differential intercept* dummies. What implicit assumption is made here?
n. Suppose someone contends that this model is *misspecified* because it assumes that the slopes of the various variables remain constant between quarters. How would you rewrite the model to take into account *differential slope* dummies?
o. If you had the data, how would you go about reformulating the demand function for coffee?

6.9. In a study of the determinants of direct airfares to Cleveland, Paul W. Bauer and Thomas J. Zlatoper obtained the following regression results (in tabular form) to explain one-way airfare for first class, coach, and discount airfares. (The dependent variable is one-way airfare in dollars).
The explanatory variables are defined as follows:

Carriers = the number of carriers
Pass = the total number of passengers flown on route (all carriers)
Miles = the mileage from the origin city to Cleveland
Pop = the population of the origin city
Inc = per capita income of the origin city
Corp = the proxy for potential business traffic from the origin city
Slot = the dummy variable equaling 1 if the origin city has a slot-restricted airport
= 0 if otherwise
Stop = the number of on-flight stops
Meal = the dummy variable equaling 1 if a meal is served
= 0 if otherwise
Hub = the dummy variable equaling 1 if the origin city has a hub airline
= 0 if otherwise
EA = the dummy variable equaling 1 if the carrier is Eastern Airlines
= 0 if otherwise
CO = the dummy variable equaling 1 if the carrier is Continental Airlines
= 0 if otherwise

The results are given in Table 6-11.

a. What is the rationale for introducing both carriers and squared carriers as explanatory variables in the model? What does the negative sign for carriers and the positive sign for carriers squared suggest?

b. As in part (*a*), what is the rationale for the introduction of miles and squared miles as explanatory variables? Do the observed signs of these variables make economic sense?

TABLE 6-11 DETERMINANTS OF DIRECT AIR FARES TO CLEVELAND

Explanatory variable	First class	Coach	Discount
Carriers	−19. 50 *t = (−0.878)	−23.00 (−1.99)	−17.50 (−3.67)
Carriers^2	2.79 (0.632)	4.00 (1.83)	2.19 (2.42)
Miles	0.233 (5.13)	0.277 (12.00)	0.0791 (8.24)
Miles^2	−0.0000097 (−0.495)	−0.000052 (−4.98)	−0.000014 (−3.23)
Pop	−0.00598 (−1.67)	−0.00114 (−4.98)	−0.000868 (−1.05)
Inc	−0.00195 (−0.686)	−0.00178 (−1.06)	−0.00411 (−6.05)
Corp	3.62 (3.45)	1.22 (2.51)	−1.06 (−5.22)
Pass	−0.000818 (−0.771)	−0.000275 (−0.527)	0.853 (3.93)
Stop	12.50 (1.36)	7.64 (2.13)	−3.85 (−2.60)
Slot	7.13 (0.299)	−0.746 (−0.067)	17.70 (3.82)
Hub	11.30 (0.90)	4.18 (0.81)	−3.50 (−1.62)
Meal	11.20 (1.07)	0.945 (0.177)	1.80 (0.813)
EA	−18.30 (−1.60)	5.80 (0.775)	−10.60 (−3.49)
CO	−66.40 (−5.72)	−56.50 (−7.61)	−4.17 (−1.35)
Constant term	212.00 (5.21)	126.00 (5.75)	113.00 (12.40)
R^2	0.863	0.871	0.799
Number of observations	163	323	323

Note: *Figures in parentheses represent t values.

Source: Paul W. Bauer and Thomas J. Zlatoper, *Economic Review*, Federal Reserve Bank of Cleveland, vol. 25, no. 1, 1989, Tables 2, 3, and 4, pp. 6–7.

c. The population variable is observed to have a negative sign. What is the implication here?

d. Why is the coefficient of the per capita income variable negative in all the regressions?

e. Why does the stop variable have a positive sign for first-class and coach fares but a negative sign for discount fares? Which makes economic sense?

f. The dummy for Continental Airlines consistently has a negative sign. What does this suggest?

g. Assess the statistical significance of each estimated coefficient. *Note:* Since the number of observations is sufficiently large, use the normal approximation to the t distribution at the 5% level of significance. Justify your use of one-tailed or two-tailed tests.

h. Why is the slot dummy significant only for discount fares?

i. Since the number of observations for coach and discount fare regressions is the same, 323 each, would you pull all 646 observations and run a regression similar to the ones shown in the preceding table? If you do that, how would you distinguish between coach and discount fare observations? (*Hint:* dummy variables.)

j. Comment on the overall quality of the regression results given in the preceding table.

6.10. In a regression of weight on height involving 51 students, 36 males and 15 females, the following regression results were obtained:[15]

1. $$\widehat{\text{Weight}}_i = -232.06551 + 5.5662\text{height}_i$$
$$t = (-5.2066) \qquad (8.6246)$$

2. $$\widehat{\text{Weight}}_i = -122.9621 + 23.8238\text{dumsex}_\text{i} + 3.7402\text{height}_i$$
$$t = (-2.5884) \qquad (4.0149) \qquad (5.1613)$$

3. $$\widehat{\text{Weight}}_i = -107.9508 + 3.5105\text{height}_i + 2.0073\text{dumsex}_i + 0.3263\text{dumht.}$$
$$t = (-1.2266) \quad (2.6087) \qquad (0.0187) \qquad (0.2035)$$

where weight is in pounds, height is in inches, and where

$$\begin{aligned} \text{Dumsex} &= 1 \text{ if male} \\ &= 0 \text{ if otherwise} \end{aligned}$$

Dumht. = the interactive or differential slope dummy

a. Which regression would you choose, 1 or 2? Why?

b. If 2 is in fact preferable but you choose 1, what kind of error are you committing?

c. What does the dumsex coefficient in 2 suggest?

d. In Model 2 the differential intercept dummy is statistically significant whereas in Model 3 it is statistically insignificant. What accounts for this change?

e. Between Models 2 and 3, which would you choose? Why?

f. In Models 2 and 3 the coefficient of the height variable is about the same, but the coefficient of the dummy variable for sex changes dramatically. Do you have any idea what is going on?

[15] A former colleague, Albert Zucker, collected these data and estimated the various regressions.

To answer questions (*d*), (*e*), and (*f*) you are given the following *correlation matrix*.

	Height	Dumsex	Dumht.
Height	1	0.6276	0.6752
Dumsex	0.6276	1	0.9971
Dumht.	0.6752	0.9971	1

The interpretation of this table is that the coefficient of correlation between height and dumsex is 0.6276 and that between dumsex and dumht. is 0.9971.

6.11. Table 6-12 on the textbook's Web site gives *nonseasonally* adjusted quarterly data on the retail sales of hobby, toy, and game stores (in millions) for the period 1992: I to 2008: II.

Consider the following model:

$$\text{Sales}_t = B_1 + B_2D_{2t} + B_3D_{3t} + B_4D_{4t} + u_t$$

where $D_2 = 1$ in the second quarter, $= 0$ if otherwise
$D_3 = 1$ in the third quarter, $= 0$ if otherwise
$D_4 = 1$ in the fourth quarter, $= 0$ if otherwise

a. Estimate the preceding regression.
b. What is the interpretation of the various coefficients?
c. Give a logical reason for why the results are this way.
***d.** How would you use the estimated regression to deseasonalize the data?

6.12. Use the data of Problem 6.11 but estimate the following model:

$$\text{Sales}_t = B_1D_{1t} + B_2D_{2t} + B_3D_{3t} + B_4D_{4t} + u_t$$

In this model there is a dummy assigned to each quarter.

a. How does this model differ from the one given in Problem 6.11?
b. To estimate this model, will you have to use a regression program that suppresses the intercept term? In other words, will you have to run a regression through the origin?
c. Compare the results of this model with the previous one and determine which model you prefer and why.

6.13. Refer to Eq. (6.17) in the text. How would you modify this equation to allow for the possibility that the coefficient of *Tuition* also differs from region to region? Present your results.

6.14. How would you check that in Eq. (6.19) the slope coefficient of X varies by sex as well as race?

6.15. Reestimate Eq. (6.30) by assigning a dummy for each quarter and compare your results with those given in Eq. (6.30). In estimating such an equation, what precaution must you take?

*Optional.

Problems

12.1 Draw a picture that explains why OLS is not the most appropriate estimator when the dependent variable only takes on two values. List four other reasons the Logit or Probit is preferred to OLS in this circumstance.

12.2 How are the multinomial logit and the ordered probit different? Why are the logit or probit not appropriate models in this circumstance?

12.3 A study tried to find the determinants of the increase in the number of households headed by a female. Using 1940 and 1960 historical census data, a logit model was estimated to predict whether a woman is the head of a household (living on her own) or whether she is living within another's household. The limited dependent variable takes on a value of 1 if the female lives on her own and is 0 if she shares housing. The results for 1960 using 6,051 observations on prime-age whites and 1,294 on nonwhites were as shown in the accompanying table:

Regression	**(1) White**	**(2) Nonwhite**
Regression Model	Logit	Logit
Constant	1.459	−2.874
	(0.685)	(1.423)
Age	−0.275	0.084
	(0.037)	(0.068)
Age Squared	0.00463	0.00021
	(0.00044)	(0.00081)
Education	−0.171	−0.127
	(0.026)	(0.038)
Farm Status	−0.687	−0.498
	(0.173)	(0.346)
South	0.376	0.520
	(0.098)	(0.180)
Expected Family Earnings	0.0018	0.0011
	(0.00019)	(0.00024)
Family Composition	4.123	2.751
	(0.294)	(0.345)
Pseudo-R^2	0.266	0.189
Percent Correctly Predicted	82.0	83.4

where *Age* is measured in years, *Education* is years of schooling of the family head, *Farm status* is a binary variable taking the value of one if the family head lived on a farm, *South* is a binary variable for living in a certain region of the country, *Expected family earnings* was generated from a separate OLS regression to predict earnings from a set of regressors, and *Family composition* refers to the number of family members under the age of 18 divided by the total number in the family.

The mean values for the variables were as shown in the next table.

Variable	**(1) White Mean**	**(2) Nonwhite Mean**
Age	46.1	42.9
Age squared	2,263.5	1,965.6
Education	12.6	10.4
Farm status	0.03	0.02
South	0.3	0.5
Expected family earnings	2,336.4	1,507.3
Family composition	0.2	0.3

a. Interpret the results. Do the coefficients have the expected signs? Why do you think age was entered both in levels and in squares?

b. Calculate the difference in the predicted probability between whites and nonwhites and the sample mean values of the explanatory variables. Why do you think the study did not combine the observations and allow for a nonwhite binary variable to enter?

c. What would be the effect on the probability of a nonwhite woman living on her own, if *Education* and *Family composition* were changed from their current mean to the mean of whites, while all other variables were left unchanged at the nonwhite mean values?

d. How would you calculate marginal effects in a more advanced statistical package? Why are marginal effects more informative than the coefficient estimates printed in this problem?

e. What are the primary reasons probit or logit models are used instead of the linear probability model?

12.4 Suppose you were hired by the state of California to conduct a study of voter willingness to support a tax increase to fund new school spending. Define V_i such that:

$V_i = 1$ if voter i supports the tax increase

$V_i = 0$ if voter i does not support the tax increase

Thus, V_i is a binomial random variable with population mean, μ, and population variance, σ^2. You think that voter support of the tax increase is related to the income of each individual.

a. Suppose that you regress V_i on $Income_i$ using a linear probability model. What properties would this model have, assuming that income is the only factor that affects voting behavior? Is the estimate of the slope reliable?

b. How would your answer change in part (a) if you instead used a logit or probit model?

Exercises

E12.1 Use the file **Hunting Data.xls**, which is collected from the 2006 National Survey of Fishing, Hunting, and Wildlife Associated Recreation, for the following questions.

a. Use OLS to regress hunt on male married, divorced, somecol, hsonly, ba, somepost, postgrad, black, other, income, and south central. Comment on the results. Why is OLS not the appropriate estimation strategy in this circumstance?

b. Estimate the same model as in part (a) with a probit model. Comment on the results. Make sure to estimate the marginal effects.

c. Estimate the same model as in part (a) with a logit model. Comment on the results. Make sure to estimate the marginal effects

d. Which model do you think is the most appropriate to use and why?

E12.2 Using the file **Cigarette Data.xls**, which is collected from the 2003 National Health Interview Survey,

a. Using OLS, regress smoke on black, Hispanic, male, age, family size, married, and divorced. Comment on the results.

b. Estimate the same model as in part (a) with a probit model. Comment on the results. Make sure to estimate the marginal effects.

c. Estimate the same model as in part (a) with a logit model. Comment on the results. Make sure to estimate the marginal effects.

d. Which model do you think is the most appropriate to use and why?

E12.3 Using the file **Publication.xls**, estimate a multinomial logit to investigate the factors that lead a researcher to either not publish (publish = 0), be a moderate publisher (publish = 1), or be a prolific publisher (publish = 2). Make sure to obtain the marginal effects and comment on your results.

E12.4 Using the file **Publication.xls**, estimate an ordered probit to investigate the factors that lead a researcher to either not publish (publish = 0), be a moderate publisher (publish = 1), or be a prolific publisher (publish = 2). Make sure to obtain the marginal effects and comment on your results.