5.3 Consider the following model that relates the proportion of a household's budget spent on alcohol *WALC* to total expenditure *TOTEXP*, age of the household head *AGE*, and the number of children in the household *NK*.

$$WALC = \beta_1 + \beta_2 \ln(TOTEXP) + \beta_3 AGE + \beta_4 NK + e$$

The data in the file *london.dat* were used to estimate this model. See Exercise 4.10 for more details about the data. Note that only households with one or two children are being considered. Thus, *NK* takes only the values one or two. Output from estimating this equation appears in Table 5.6.

Table 5.6 Output for Exercise 5.3

Dependent Variable: *WALC*
Included observations: 1519

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.0091	0.0190		0.6347
ln(*TOTEXP*)	0.0276		6.6086	0.0000
AGE		0.0002	−6.9624	0.0000
NK	−0.0133	0.0033	−4.0750	0.0000

R-squared		Mean dependent var	0.0606
S.E. of regression		S.D. dependent var	0.0633
Sum squared resid	5.752896		

(a) Fill in the following blank spaces that appear in this table.

  (i) The t-statistic for b_1
  (ii) The standard error for b_2
  (iii) The estimate b_3
  (iv) R^2
  (v) $\hat{\sigma}$

(b) Interpret each of the estimates b_2, b_3, and b_4.

(c) Compute a 95% interval estimate for β_3. What does this interval tell you?

(d) Test the hypothesis that the budget proportion for alcohol does not depend on the number of children in the household. Can you suggest a reason for the test outcome?

5.8* An agricultural economist carries out an experiment to study the production relationship between the dependent variable $YIELD$ = peanut yield (pounds per acre) and the production inputs

$NITRO$ = amount of nitrogen applied (hundreds of pounds per acre)
$PHOS$ = amount of phosphorus fertilizer (hundreds of pounds per acre)

A total $N = 27$ observations were obtained using different test fields. The estimated quadratic model, with an interaction term, is

$$\widehat{YIELD} = 1.385 + 8.011NITRO + 4.800PHOS - 1.944NITRO^2 - 0.778PHOS^2 - 0.567NITRO \times PHOS$$

(a) Find equations describing the marginal effect of nitrogen on yield and the marginal effect of phosporus on yield. What do these equations tell you?

(b) What are the marginal effects of nitrogen and of phosphorus when (i) $NITRO$ and $PHOS = 1$ and (ii) when $NITRO = 2$ and $PHOS = 2$? Comment on your findings.

(c) Test the hypothesis that the marginal effect of nitrogen is zero, when
(iv) $PHOS = 1$ and $NITRO = 1$
(v) $PHOS = 1$ and $NITRO = 2$
(vi) $PHOS = 1$ and $NITRO = 3$
Note: The following information may be useful:

$$\widehat{\text{var}(b_2 + 2b_4 + b_6)} = 0.233$$

$$\widehat{\text{var}(b_2 + 4b_4 + b_6)} = 0.040$$

$$\widehat{\text{var}(b_2 + 6b_4 + b_6)} = 0.233$$

(d) ◆[This part requires the use of calculus] For the function estimated, what levels of nitrogen and phosphorus give maximum yield? Are these levels the optimal fertilizer applications for the peanut producer?

5.13 The file *br2.dat* contains data on 1,080 houses sold in Baton Rouge, Louisiana, during mid-2005. We will be concerned with the selling price ($PRICE$), the size of the house in square feet ($SQFT$), and the age of the house in years (AGE).

(a) Use all observations to estimate the following regression model and report the results

$$PRICE = \beta_1 + \beta_2 SQFT + \beta_3 AGE + e$$

(i) Interpret the coefficient estimates.

(ii) Find a 95% interval estimate for the price increase for an extra square foot of living space—that is, $\partial PRICE/\partial SQFT$.

(iii) Test the hypothesis that having a house a year older decreases price by 1000 or less ($H_0 : \beta_3 \geq -1000$) against the alternative that it decreases price by more than 1000 ($H_1 : \beta_3 < -1000$).

(b) Add the variables $SQFT^2$ and AGE^2 to the model in part (a) and re-estimate the equation. Report the results.

(i) Find estimates of the marginal effect $\partial PRICE/\partial SQFT$ for the smallest house in the sample, the largest house in the sample, and a house with 2300 $SQFT$. Comment on these values. Are they realistic?

(ii) Find estimates of the marginal effect $\partial PRICE/\partial AGE$ for the oldest house in the sample, the newest house in the sample, and a house that is 20 years old. Comment on these values. Are they realistic?

(iii) Find a 95% interval estimate for the marginal effect $\partial PRICE/\partial SQFT$ for a house with 2300 square feet.

(iv) For a house that is 20 years old, test the hypothesis

$$H_0 : \frac{\partial PRICE}{\partial AGE} \geq -1000 \text{ against } H_1 : \frac{\partial PRICE}{\partial AGE} < -1000$$

(c) Add the interaction variable $SQFT \times AGE$ to the model in part (b) and re-estimate the equation. Report the results. Repeat parts (i), (ii), (iii), and (iv) from part (b) for this new model. Use $SQFT = 2300$ and $AGE = 20$.

(d) From your answers to parts (a), (b), and (c), comment on the sensitivity of the results to the model specification.

5.16 Data on the weekly sales of a major brand of canned tuna by a supermarket chain in a large midwestern U.S. city during a mid-1990's calendar year are contained in the file *tuna.dat*. There are 52 observations on the variables. $SAL1$ = unit sales of brand no. 1 canned tuna; $APR1$ = price per can of brand no. 1 canned tuna; $APR2$, $APR3$ = price per can of brands no. 2 and 3 of canned tuna.

(a) The prices $APR1$, $APR2$, and $APR3$ are expressed in dollars. Multiply the observations on each of these variables by 100 to express them in terms of cents; call the new variables $PR1$, $PR2$, and $PR3$. Estimate the following regression model and report the results:

$$SAL1 = \beta_1 + \beta_2 PR1 + \beta_3 PR2 + \beta_4 PR3 + e$$

(b) Interpret the estimates b_2, b_3, and b_4. Do they have the expected signs?

(c) Using suitable one-tail tests and a 5% significance level, test whether each of the coefficients b_2, b_3, and b_4 are significantly different from zero.

(d) Using a 5% significance level, test the following hypotheses:

(i) A 1-cent increase in the price of brand one reduces its sales by 300 cans.

(ii) A 1-cent increase in the price of brand two increases the sales of brand one by 300 cans.

(iii) A 1-cent increase in the price of brand three increases the sales of brand one by 300 cans.

(iv) The effect of a price increase in brand two on sales of brand one is the same as the effect of a price increase in brand three on sales of brand one. Does the outcome of this test contradict your findings from parts (ii) and (iii)?

(v) If prices of all 3 brands go up by 1 cent, there is no change in sales.

5.17 (a) Reconsider the model $SAL1 = \beta_1 + \beta_2 PR1 + \beta_3 PR2 + \beta_4 PR3 + e$ from Exercise 5.16. Estimate this model if you have not already done so, and find a 95% interval estimate for expected sales when $PR1 = 90$, $PR2 = 75$, and $PR3 = 75$. What is wrong with this interval?

(b) Estimate the alternative model $\ln(SAL1) = \alpha_1 + \alpha_2 PR1 + \alpha_3 PR2 + \alpha_4 PR3 + e$, and find a 95% interval estimate for expected log of sales when $PR1 = 90$, $PR2 = 75$, and $PR3 = 75$. Convert this interval into one for sales, and compare it with what you got in part (a).

(c) How does the interpretation of the coefficients in the model with $\ln(SAL1)$ as the dependent variable differ from that for the coefficients in the model with $SAL1$ as the dependent variable?

5.25 Consider the following aggregate production function for the U.S. manufacturing sector:

$$Y = \alpha K^{\beta_2} L^{\beta_3} E^{\beta_4} M^{\beta_5} \exp\{e\}$$

where Y is gross output, K is capital, L is labor, E is energy, and M denotes other intermediate materials. The data underlying these variables are given in index form in the file *manuf.dat*.

(a) Show that taking logarithms of the production function puts it in a form suitable for least squares estimation.

(b) Estimate the unknown parameters of the production function and find the corresponding standard errors.

(c) Discuss the economic and statistical implications of these results.

6.4 Consider the wage equation

$$\ln(WAGE) = \beta_1 + \beta_2 EDUC + \beta_3 EDUC^2 + \beta_4 EXPER + \beta_5 EXPER^2 + \beta_6 (EDUC \times EXPER) + \beta_7 HRSWK + e$$

where the explanatory variables are years of education, years of experience and hours worked per week. Estimation results for this equation, and for modified versions of it obtained by dropping some of the variables, are displayed in Table 6.4. These results are from the 1000 observations in the file *cps4c_small.dat*.

(a) Using an approximate 5% critical value of $t_c = 2$, what coefficient estimates are not significantly different from zero?

(b) What restriction on the coefficients of Eqn (A) gives Eqn (B)? Use an F-test to test this restriction. Show how the same result can be obtained using a t-test.

(c) What restrictions on the coefficients of Eqn (A) give Eqn (C)? Use an *F*-test to test these restrictions. What question would you be trying to answer by performing this test?

(d) What restrictions on the coefficients of Eqn (B) give Eqn (D)? Use an *F*-test to test these restrictions. What question would you be trying to answer by performing this test?

(e) What restrictions on the coefficients of Eqn (A) give Eqn (E)? Use an *F*-test to test these restrictions. What question would you be trying to answer by performing this test?

(f) Based on your answers to parts (a) to (e), which model would you prefer? Why?

(g) Compute the missing AIC value for Eqn (D) and the missing SC value for Eqn (A). Which model is favored by the AIC? Which model is favored by the SC?

6.9 In Exercise 5.25 we expressed the model

$$Y = \alpha K^{\beta_2} L^{\beta_3} E^{\beta_4} M^{\beta_5} \exp\{e\}$$

in terms of logarithms and estimated it using data in the file *manuf.dat*. Use the data and results from Exercise 5.25 to test the following hypotheses:

(a) $H_0: \beta_2 = 0$ against $H_1: \beta_2 \neq 0$.
(b) $H_0: \beta_2 = 0, \beta_3 = 0$ against $H_1: \beta_2 \neq 0$ and/or $\beta_3 \neq 0$.
(c) $H_0: \beta_2 = 0, \beta_4 = 0$ against $H_1: \beta_2 \neq 0$ and/or $\beta_4 \neq 0$.
(d) $H_0: \beta_2 = 0, \beta_3 = 0, \ \beta_4 = 0$ against $H_1: \beta_2 \neq 0$ and/or $\beta_3 \neq 0$ and/or $\beta_4 \neq 0$.
(e) $H_0: \beta_2 + \beta_3 + \beta_4 + \beta_5 = 1$ against $H_1: \beta_2 + \beta_3 + \beta_4 + \beta_5 \neq 1$.
(f) Analyze the impact of collinearity on this model.

6.13 The file *toodyay.dat* contains 48 annual observations on a number of variables related to wheat yield in the Toodyay Shire of Western Australia, for the period 1950–1997. Those variables are

$$
\begin{aligned}
Y &= \text{wheat yield in tonnes per hectare,} \\
t &= \text{trend term to allow for technological change,} \\
RG &= \text{rainfall at germination (May–June),} \\
RD &= \text{rainfall at development stage (July–August), and} \\
RF &= \text{rainfall at flowering (September–October).}
\end{aligned}
$$

The unit of measurement for rainfall is centimeters. A model that allows for the yield response to rainfall to be different for the three different periods is

$$Y = \beta_1 + \beta_2 t + \beta_3 RG + \beta_4 RD + \beta_5 RF + e$$

(a) Estimate this model. Report the results and comment on the signs and significance of the estimated coefficients.

(b) Test the hypothesis that the response of yield to rainfall is the same irrespective of whether the rain falls during germination, development, or flowering.

(c) Estimate the model under the restriction that the three responses to rainfall are the same. Comment on the results.

6.22* In Chapter 5.7 we used the data in file *pizza4.dat* to estimate the model

$$PIZZA = \beta_1 + \beta_2 AGE + \beta_3 INCOME + \beta_4(AGE \times INCOME) + e$$

(a) Test the hypothesis that age does not affect pizza expenditure—that is, test the joint hypothesis $H_0{:}\beta_2 = 0,\ \beta_4 = 0$. What do you conclude?
(b) Construct point estimates and 95% interval estimates of the marginal propensity to spend on pizza for individuals of ages 20, 30, 40, 50, and 55. Comment on these estimates.
(c) Modify the equation to permit a "life-cycle" effect in which the marginal effect of income on pizza expenditure increases with age, up to a point, and then falls. Do so by adding the term $(AGE^2 \times INC)$ to the model. What sign do you anticipate on this term? Estimate the model and test the significance of the coefficient for this variable. Did the estimate have the expected sign?
(d) Using the model in (c), construct point estimates and 95% interval estimates of the marginal propensity to spend on pizza for individuals of ages 20, 30, 40, 50 and 55. Comment on these estimates. In light of these values, and of the range of age in the sample data, what can you say about the quadratic function of age that describes the marginal propensity to spend on pizza?
(e) For the model in part (c), are each of the coefficient estimates for AGE, $(AGE \times INC)$ and $(AGE^2 \times INC)$ significantly different from zero at a 5% significance level? Carry out a joint test for the significance of these variables. Comment on your results.
(f) Check the model used in part (c) for collinearity. Add the term $(AGE^3 \times INC)$ to the model in (c) and check the resulting model for collinearity.

7.3 Henry Saffer and Frank Chaloupka ("The Demand for Illicit Drugs," *Economic Inquiry*, 37(3), 1999, 401–411) estimate demand equations for alcohol, marijuana, cocaine, and heroin using a sample of size $N = 44{,}889$. The estimated equation for alcohol use after omitting a few control variables is shown in the chart at the top of page 289.

The variable definitions (sample means in parentheses) are as follows:

The dependent variable is the number of days alcohol was used in the past 31 days (3.49)

ALCOHOL PRICE—price of a liter of pure alcohol in 1983 dollars (24.78)
INCOME—total personal income in 1983 dollars (12,425)
GENDER—a binary variable = 1 if male (0.479)
MARITAL STATUS—a binary variable = 1 if married (0.569)
AGE 12–20—a binary variable = 1 if individual is 12–20 years of age (0.155)
AGE 21–30—a binary variable = 1 if individual is 21–30 years of age (0.197)
BLACK—a binary variable = 1 if individual is black (0.116)
HISPANIC—a binary variable = 1 if individual is Hispanic (0.078)

Demand for Illicit Drugs

Variable	Coefficient	*t*-statistic
C	4.099	17.98
ALCOHOL PRICE	−0.045	5.93
INCOME	0.000057	17.45
GENDER	1.637	29.23
MARITAL STATUS	−0.807	12.13
AGE 12–20	−1.531	17.97
AGE 21–30	0.035	0.51
BLACK	−0.580	8.84
HISPANIC	−0.564	6.03

(a) Interpret the coefficient of alcohol price.
(b) Compute the price elasticity at the means of the variables.
(c) Compute the price elasticity at the means of alcohol price and income, for a married black male, age 21–30.
(d) Interpret the coefficient of income. If we measured income in \$1,000 units, what would the estimated coefficient be?
(e) Interpret the coefficients of the indicator variables, as well as their significance.

7.9* In the STAR experiment (Section 7.5.3), children were randomly assigned within schools into three types of classes: small classes with 13 to 17 students, regular-sized classes with 22–25 students, and regular-sized classes with a full-time teacher aide to assist the teacher. Student scores on achievement tests were recorded, as was some information about the students, teachers, and schools. Data for the kindergarten classes is contained in the data file *star.dat*.

(a) Calculate the average of *TOTALSCORE* for (i) students in regular-sized classrooms with full time teachers, but no aide; (ii) students in regular-sized classrooms with full time teachers, and an aide; and (iii) students in small classrooms. What do you observe about test scores in these three types of learning environments?

(b) Estimate the regression model $TOTALSCORE_i = \beta_1 + \beta_2 SMALL_i + \beta_3 AIDE_i + e_i$, where *AIDE* is a indicator variable equaling one for classes taught by a teacher and an aide and zero otherwise. What is the relation of the estimated coefficients from this regression to the sample means in part (a)? Test the statistical significance of β_3 at the 5% level of significance.

(c) To the regression in (b) add the additional explanatory variable *TCHEXPER*. Is this variable statistically significant? Does its addition to the model affect the estimates of β_2 and β_3?

(d) To the regression in (c) add the additional explanatory variables *BOY*, *FREELUNCH*, and *WHITE_ASIAN*. Are any of these variables statistically significant? Does their addition to the model affect the estimates of β_2 and β_3?

(e) To the regression in (d) add the additional explanatory variables *TCHWHITE*, *TCHMASTERS*, *SCHURBAN*, and *SCHRURAL*. Are any of these variables statistically significant? Does their addition to the model affect the estimates of β_2 and β_3?

(f) Discuss the importance of parts (c), (d), and (e) to our estimation of the "treatment" effects in part (b).

(g) Add to the models in (b) through (e) indicator variables for each school

$$SCHOOL\_j = \begin{cases} 1 & \text{if student is in school } j \\ 0 & \text{otherwise} \end{cases}$$

Test the joint significance of these school "fixed effects." Does the inclusion of these fixed effect indicator variables substantially alter the estimates of β_2 and β_3?