CHAPTER 5

Exercise Solutions

EXERCISE 5.1

(a) $\bar{y} = 1,\ \ \bar{x}_2 = 0,\ \ \bar{x}_3 = 0$

x_{i2}^*	x_{i3}^*	y_i^*
0	1	0
1	−2	1
2	1	2
−2	0	−2
1	−1	−1
−2	−1	−2
0	1	1
−1	1	0
1	0	1

(b) $\sum y_i^* x_{i2}^* = 13, \qquad \sum x_{i2}^{*2} = 16, \qquad \sum y_i^* x_{i3}^* = 4, \qquad \sum x_{i3}^{*2} = 10$

(c)
$$b_2 = \frac{\left(\sum y_i^* x_{i2}^*\right)\left(\sum x_{i3}^{*2}\right) - \left(\sum y_i^* x_{i3}^*\right)\left(\sum x_{i2}^* x_{i3}^*\right)}{\left(\sum x_{i2}^{*2}\right)\left(\sum x_{i3}^{*2}\right) - \left(\sum x_{i2}^* x_{i3}^*\right)^2} = \frac{13\times 10 - 4\times 0}{16\times 10 - 0^2} = 0.8125$$

$$b_3 = \frac{\left(\sum y_i^* x_{i3}^*\right)\left(\sum x_{i2}^{*2}\right) - \left(\sum y_i^* x_{i2}^*\right)\left(\sum x_{i2}^* x_{i3}^*\right)}{\left(\sum x_{i2}^{*2}\right)\left(\sum x_{i3}^{*2}\right) - \left(\sum x_{i2}^* x_{i3}^*\right)^2} = \frac{4\times 16 - 13\times 0}{16\times 10 - 0^2} = 0.4$$

$$b_1 = \bar{y} - b_2\bar{x}_2 - b_3\bar{x}_3 = 1$$

(d) $\hat{e} = \left(-0.4,\ 0.9875,\ -0.025,\ -0.375,\ -1.4125,\ 0.025,\ 0.6,\ 0.4125,\ 0.1875\right)$

(e)
$$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{N-K} = \frac{3.8375}{9-3} = 0.6396$$

(f)
$$r_{23} = \frac{\sum (x_{i2}-\bar{x}_2)(x_{i3}-\bar{x}_3)}{\sqrt{\sum (x_{i2}-\bar{x}_2)^2 \sum (x_{i3}-\bar{x}_3)^2}} = \frac{\sum x_{i2}^* x_{i3}^*}{\sqrt{\sum x_{i2}^{*\,2} \sum x_{i3}^{*\,2}}} = 0$$

(g)
$$\text{se}(b_2) = \sqrt{\text{var}(b_2)} = \sqrt{\frac{\hat{\sigma}^2}{\sum (x_{i2}-\bar{x}_2)^2 (1-r_{23}^2)}} = \sqrt{\frac{0.6396}{16}} = 0.1999$$

(h)
$$SSE = \sum \hat{e}_i^2 = 3.8375 \qquad SST = \sum (y_i - \bar{y})^2 = 16,$$

$$SSR = SST - SSE = 12.1625 \qquad R^2 = \frac{SSR}{SST} = \frac{12.1625}{16} = 0.7602$$

EXERCISE 5.2

(a) A 95% confidence interval for β_2 is

$$b_2 \pm t_{(0.975,6)}\,\mathrm{se}(b_2) = 0.8125 \pm 2.447 \times 0.1999 = (0.3233, 1.3017)$$

(b) The null and alternative hypotheses are

$$H_0:\beta_2 = 1, \quad H_1:\beta_2 \neq 1$$

The calculated *t*-value is

$$t = \frac{b_2 - 1}{\mathrm{se}(b_2)} = \frac{0.8125 - 1}{0.1999} = -0.9377$$

At a 5% significance level, we reject H_0 if $|t| > t_{(0.975,\,6)} = 2.447$. Since $|-0.9377| < 2.447$, we do not reject H_0.

EXERCISE 5.3

(a) (i) The t-statistic for b_1 is $\dfrac{b_1}{\text{se}(b_1)} = \dfrac{0.0091}{0.0191} = 0.476$.

(ii) The standard error for b_2 is $\text{se}(b_2) = \dfrac{0.0276}{6.6086} = 0.00418$.

(iii) The estimate for β_3 is $b_3 = 0.0002 \times (-6.9624) = -0.0014$.

(iv) To compute R^2, we need *SSE* and *SST*. From the output, $SSE = 5.752896$. To find *SST*, we use the result

$$\hat{\sigma}_y = \sqrt{\frac{SST}{N-1}} = 0.0633$$

which gives $SST = 1518 \times (0.0633)^2 = 6.08246$. Thus,

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{5.75290}{6.08246} = 0.054$$

(v) The estimated error standard deviation is $\hat{\sigma} = \sqrt{\dfrac{SSE}{(N-K)}} = \sqrt{\dfrac{5.752896}{1519-4}} = 0.061622$

(b) The value $b_2 = 0.0276$ implies that if $\ln(TOTEXP)$ increases by 1 unit the alcohol share will increase by 0.0276. The change in the alcohol share from a 1-unit change in total expenditure depends on the level of total expenditure. Specifically, $d(WALC)/d(TOTEXP) = 0.0276/TOTEXP$. A 1% increase in total expenditure leads to a 0.000276 increase in the alcohol share of expenditure.

The value $b_3 = -0.0014$ suggests that if the age of the household head increases by 1 year the share of alcohol expenditure of that household decreases by 0.0014.

The value $b_4 = -0.0133$ suggests that if the household has one more child the share of the alcohol expenditure decreases by 0.0133.

(c) A 95% confidence interval for β_3 is

$$b_3 \pm t_{0.975,1515)}\,\text{se}(b_3) = -0.0014 \pm 1.96 \times 0.0002 = (-0.0018, -0.0010)$$

This interval tells us that, if the age of the household head increases by 1 year, the share of the alcohol expenditure is estimated to decrease by an amount between 0.0018 and 0.001.

Exercise 5.3 (Continued)

(d) The null and alternative hypotheses are $H_0:\beta_4=0,\ H_1:\beta_4\neq 0$.

The calculated *t*-value is $t=\dfrac{b_4}{\text{se}(b_4)}=-4.075$

At a 5% significance level, we reject H_0 if $|t|>t_{(0.975,\,1515)}=1.96$. Since $|-4.075|>1.96$, we reject H_0 and conclude that the number of children in the household influences the budget proportion on alcohol. Having an additional child is likely to lead to a smaller budget share for alcohol because of the non-alcohol expenditure demands of that child. Also, perhaps households with more children prefer to drink less, believing that drinking may be a bad example for their children.

EXERCISE 5.4

(a) The regression results are:

$$\begin{array}{l} \widehat{WTRANS} = -0.0315 + 0.0414\ln\left(TOTEXP\right) - 0.0001AGE - 0.0130\,NK \qquad R^2 = 0.0247 \\ \text{(se)} \qquad (0.0322)\ (0.0071) \qquad\qquad (0.0004) \qquad (0.0055) \end{array}$$

(b) The value $b_2 = 0.0414$ suggests that as $\ln\left(TOTEXP\right)$ increases by 1 unit the budget proportion for transport increases by 0.0414. Alternatively, one can say that a 10% increase in total expenditure will increase the budget proportion for transportation by 0.004. (See Chapter 4.3.3.) The positive sign of b_2 is according to our expectation because as households become richer they tend to use more luxurious forms of transport and the proportion of the budget for transport increases.

The value $b_3 = -0.0001$ implies that as the age of the head of the household increases by 1 year the budget share for transport decreases by 0.0001. The expected sign for b_3 is not clear. For a given level of total expenditure and a given number of children, it is difficult to predict the effect of age on transport share.

The value $b_4 = -0.0130$ implies that an additional child decreases the budget share for transport by 0.013. The negative sign means that adding children to a household increases expenditure on other items (such as food and clothing) more than it does on transportation. Alternatively, having more children may lead a household to turn to cheaper forms of transport.

(c) The *p*-value for testing $H_0:\beta_3 = 0$ against the alternative $H_1:\beta_3 \neq 0$ where β_3 is the coefficient of *AGE* is 0.869, suggesting that *AGE* could be excluded from the equation. Similar tests for the coefficients of the other two variables yield *p*-values less than 0.05.

(d) The proportion of variation in the budget proportion allocated to transport explained by this equation is 0.0247.

(e) For a one-child household:

$$\begin{aligned} \widehat{WTRANS}_0 &= -0.0315 + 0.0414\ln(TOTEXP_0) - 0.0001AGE_0 - 0.013NK_0 \\ &= -0.0315 + 0.0414\times\ln(98.7) - 0.0001\times 36 - 0.013\times 1 \\ &= 0.1420 \end{aligned}$$

For a two-child household:

$$\begin{aligned} \widehat{WTRANS}_0 &= -0.0315 + 0.0414\ln(TOTEXP_0) - 0.0001AGE_0 - 0.013NK_0 \\ &= -0.0315 + 0.0414\times\ln(98.7) - 0.0001\times 36 - 0.013\times 2 \\ &= 0.1290 \end{aligned}$$

EXERCISE 5.5

(a) The estimated equation is

$$\begin{array}{l} VALUE = 28.4067 - 0.1834 CRIME - 22.8109 NITOX + 6.3715 ROOMS - 0.0478 AGE \\ \text{(se)} \quad (5.3659) \quad (0.0365) \qquad (4.1607) \qquad (0.3924) \qquad (0.0141) \\ \qquad -1.3353 DIST + 0.2723 ACCESS - 0.0126 TAX - 1.1768 PTRATIO \\ \qquad (0.2001) \qquad (0.0723) \qquad (0.0038) \qquad (0.1394) \end{array}$$

The estimated equation suggests that as the per capita crime rate increases by 1 unit the home value decreases by \$183.4. The higher the level of air pollution the lower the value of the home; a one unit increase in the nitric oxide concentration leads to a decline in value of \$22,811. Increasing the average number of rooms leads to an increase in the home value; an increase in one room leads to an increase of \$6,372. An increase in the proportion of owner-occupied units built prior to 1940 leads to a decline in the home value. The further the weighted distances to the five Boston employment centers the lower the home value by \$1,335 for every unit of weighted distance. The higher the tax rate per \$10,000 the lower the home value. Finally, the higher the pupil-teacher ratio, the lower the home value.

(b) A 95% confidence interval for the coefficient of *CRIME* is

$$b_2 \pm t_{(0.975,497)}\text{se}(b_2) = -0.1834 \pm 1.965 \times 0.0365 = (-0.255, -0.112).$$

A 95% confidence interval for the coefficient of *ACCESS* is

$$b_7 \pm t_{(0.975,497)}\text{se}(b_7) = 0.2723 \pm 1.965 \times 0.0723 = (0.130, 0.414)$$

(c) We want to test $H_0: \beta_{room} = 7$ against $H_1: \beta_{room} \neq 7$. The value of the *t* statistic is

$$t = \frac{b_{rooms} - 7}{\text{se}(b_{rooms})} = \frac{6.3715 - 7}{0.3924} = -1.6017$$

At $\alpha = 0.05$, we reject H_0 if the absolute calculated *t* is greater than 1.965. Since $|-1.6017| < 1.965$, we do not reject H_0. The data is consistent with the hypothesis that increasing the number of rooms by one increases the value of a house by \$7000.

(d) We want to test $H_0: \beta_{ptratio} \geq -1$ against $H_1: \beta_{ptratio} < -1$. The value of the *t* statistic is

$$t = \frac{-1.1768 + 1}{0.1394} = -1.2683$$

At a significance level of $\alpha = 0.05$, we reject H_0 if the calculated *t* is less than the critical value $t_{(0.05,497)} = -1.648$. Since $-1.2683 > -1.648$, we do not reject H_0. We cannot conclude that reducing the pupil-teacher ratio by 10 will increase the value of a house by more than \$10,000.

EXERCISE 5.6

In each case we use a two-tail test with a 5% significance level. The critical values are given by $t_{(0.025,60)} = -2.000$ and $t_{(0.975,60)} = 2.000$. The rejection region is $t < -2$ or $t > 2$.

(a) The value of the t statistic for testing the null hypothesis $H_0 : \beta_2 = 0$ against the alternative $H_1 : \beta_2 \neq 0$ is

$$t = \frac{b_2}{\text{se}(b_2)} = \frac{3}{\sqrt{4}} = 1.5$$

Since $-2 < 1.5 < 2$, we fail to reject H_0 and conclude that there is no sample evidence to suggest that $\beta_2 \neq 0$.

(b) For testing H_0: $\beta_1 + 2\beta_2 = 5$ against the alternative H_1: $\beta_1 + 2\beta_2 \neq 5$, we use the statistic

$$t = \frac{\left(b_1 + 2b_2\right) - 5}{\text{se}\left(b_1 + 2b_2\right)}$$

For the numerator of the t-value, we have $b_1 + 2b_2 - 5 = 2 + 2\times 3 - 5 = 3$

The denominator is given by

$$\begin{aligned}\text{se}(b_1 + b_2) &= \sqrt{\text{var}(b_1 + 2b_2)} = \sqrt{\text{var}(b_1) + 4\times\text{var}(b_2) + 4\times\text{cov}(b_1, b_2)} \\ &= \sqrt{3 + 4\times 4 - 4\times 2} = \sqrt{11} = 3.3166\end{aligned}$$

Therefore, $t = \dfrac{3}{3.3166} = 0.9045$

Since $-2 < 0.9045 < 2$, we fail to reject H_0. There is no sample evidence to suggest that $\beta_1 + 2\beta_2 \neq 5$.

(c) For testing $H_0 : \beta_1 - \beta_2 + \beta_3 = 4$ against the alternative $H_1 : \beta_1 - \beta_2 + \beta_3 \neq 4$, we use

$$t = \frac{(b_1 - b_2 + b_3) - 4}{\text{se}(b_1 - b_2 + b_3)}$$

Now, $(b_1 - b_2 + b_3) - 4 = 2 - 3 - 1 - 4 = -6$, and

$$\begin{aligned}\text{se}(b_1 - b_2 + b_3) &= \sqrt{\text{var}(b_1 - b_2 + b_3)} \\ &= \sqrt{\text{var}(b_1) + \text{var}(b_2) + \text{var}(b_3) - 2\text{cov}(b_1, b_2) + 2\text{cov}(b_1, b_3) - 2\text{cov}(b_2, b_3)} \\ &= \sqrt{3 + 4 + 3 + 2\times 2 + 2\times 1 - 0} = 4\end{aligned}$$

Thus, $t = \dfrac{-6}{4} = -1.5$

Since $-2 < -1.5 < 2$, we fail to reject H_0 and conclude that there is insufficient sample evidence to suggest that $\beta_1 - \beta_2 + \beta_3 = 4$ is incorrect.

EXERCISE 5.7

The variance of the error term is given by:

$$\hat{\sigma}^2 = \frac{SSE}{N-K} = \frac{11.12389}{202-3} = 0.05590$$

Thus, the standard errors of the least square estimates, b_2 and b_3 are :

$$\text{se}(b_2) = \sqrt{\text{var}(b_2)} = \sqrt{\frac{\hat{\sigma}^2}{\left(1-r_{23}^2\right)\sum(x_{i2}-\bar{x}_2)^2}} = \sqrt{\frac{0.05590}{\left(1-(-0.114255)^2\right)\times 1210.178}} = 0.00684$$

$$\text{se}(b_3) = \sqrt{\text{var}(b_3)} = \sqrt{\frac{\hat{\sigma}^2}{\left(1-r_{23}^2\right)\sum(x_{i3}-\bar{x}_3)^2}} = \sqrt{\frac{0.05590}{\left(1-(-0.114255)^2\right)\times 30307.57}} = 0.00137$$

EXERCISE 5.8

(a) Equations describing the marginal effects of nitrogen and phosphorus on yield are

$$\frac{\partial E(YIELD)}{\partial(NITRO)} = 8.011 - 2\times 1.944\times NITRO - 0.567\times PHOS$$
$$= 8.011 - 3.888NITRO - 0.567PHOS$$

$$\frac{\partial E(YIELD)}{\partial(PHOS)} = 4.800 - 2\times 0.778\times PHOS - 0.567\times NITRO$$
$$= 4.800 - 1.556PHOS - 0.567NITRO$$

These equations indicate that the marginal effect of both fertilizers declines – we have diminishing marginal products – and these marginal effects eventually become negative. Also, the marginal effect of one fertilizer is smaller, the larger is the amount of the other fertilizer that is applied.

(b) (i) The marginal effects when $NITRO = 1$ and $PHOS = 1$ are

$$\frac{\partial E(YIELD)}{\partial(NITRO)} = 8.011 - 3.888 - 0.567 = 3.556$$

$$\frac{\partial E(YIELD)}{\partial(PHOS)} = 4.800 - 1.556 - 0.567 = 2.677$$

(ii) The marginal effects when $NITRO = 2$ and $PHOS = 2$ are

$$\frac{\partial E(YIELD)}{\partial(NITRO)} = 8.011 - 3.888\times 2 - 0.567\times 2 = -0.899$$

$$\frac{\partial E(YIELD)}{\partial(PHOS)} = 4.800 - 1.556\times 2 - 0.567\times 2 = 0.554$$

When $NITRO = 1$ and $PHOS = 1$, the marginal products of both fertilizers are positive. Increasing the fertilizer applications to $NITRO = 2$ and $PHOS = 2$ reduces the marginal effects of both fertilizers, with that for nitrogen becoming negative.

(c) To test these hypotheses, the coefficients are defined according to the following equation

$$YIELD = \beta_1 + \beta_2 NITRO + \beta_3 PHOS + \beta_4 NITRO^2 + \beta_5 PHOS^2 + \beta_6 NITRO\times PHOS + e$$

(i) The settings $NITRO = 1$ and $PHOS = 1$ will yield a zero marginal effect for nitrogen if $\beta_2 + 2\beta_4 + \beta_6 = 0$. Thus, we test $H_0: \beta_2 + 2\beta_4 + \beta_6 = 0$ against the alternative $H_1: \beta_2 + 2\beta_4 + \beta_6 \neq 0$. The value of the test statistic is

$$t = \frac{b_2 + 2b_4 + b_6}{\text{se}(b_2 + 2b_4 + b_6)} = \frac{8.011 - 2\times 1.944 - 0.567}{\sqrt{0.233}} = 7.367$$

Exercise 5.8(c)(i) (Continued)

Since $t >\ t_c = t_{(0.975,\,21)} = 2.080$, we reject the null hypothesis and conclude that the marginal effect of nitrogen on yield is not zero when *NITRO* = 1 and *PHOS* = 1.

(ii) To test whether the marginal effect of nitrogen is zero when $NITRO = 2$ and $PHOS = 1$, we test $H_0: \beta_2 + 4\beta_4 + \beta_6 = 0$ against $H_1: \beta_2 + 4\beta_4 + \beta_6 \neq 0$. The value of the test statistic is

$$t = \frac{b_2 + 4b_4 + b_6}{\text{se}\left(b_2 + 4b_4 + b_6\right)} = \frac{8.011 - 4\times 1.944 - 0.567}{\sqrt{0.040}} = -1.660$$

Since $|t| < 2.080\ = t_{(0.975,\,21)}$, we do not reject the null hypothesis. A zero marginal yield with respect to nitrogen cannot be rejected when *NITRO* = 1 and *PHOS* = 2.

(iii) To test whether the marginal effect of nitrogen is zero when $NITRO = 3$ and $PHOS = 1$, we test $H_0: \beta_2 + 6\beta_4 + \beta_6 = 0$ against the alternative $H_1: \beta_2 + 6\beta_4 + \beta_6 \neq 0$. The value of the test statistic is

$$t = \frac{b_2 + 6b_4 + b_6}{\text{se}\left(b_2 + 6b_4 + b_6\right)} = \frac{8.011 - 6\times 1.944 - 0.567}{\sqrt{0.233}} = -8.742$$

Since $|t| > 2.080\ = t_{(0.975,\,21)}$, we reject the null hypothesis and conclude that the marginal product of yield to nitrogen is not zero when *NITRO* = 3 and *PHOS* = 1.

(d) The maximizing levels $NITRO^*$ and $PHOS^*$ are those values for *NITRO* and *PHOS* such that the first-order partial derivatives are equal to zero.

$$\frac{\partial E\left(YIELD\right)}{\partial\left(PHOS\right)} = \beta_3 + 2\beta_5 PHOS^* + \beta_6 NITRO^* = 0$$

$$\frac{\partial E\left(YIELD\right)}{\partial\left(NITRO\right)} = \beta_2 + 2\beta_4 NITRO^* + \beta_6 PHOS^* = 0$$

The solutions and their estimates are

$$NITRO^* = \frac{2\beta_2\beta_5 - \beta_3\beta_6}{\beta_6^2 - 4\beta_4\beta_5} = \frac{2\times 8.011\times(-0.778) - 4.800\times(-0.567)}{(-0.567)^2 - 4\times(-1.944)(-0.778)} = 1.701$$

$$PHOS^* = \frac{2\beta_3\beta_4 - \beta_2\beta_6}{\beta_6^2 - 4\beta_4\beta_5} = \frac{2\times 4.800\times(-1.944) - 8.011\times(-0.567)}{(-0.567)^2 - 4\times(-1.944)(-0.778)} = 2.465$$

The yield maximizing levels of fertilizer are not necessarily the optimal levels. The optimal levels are those where the marginal cost of the inputs is equal to the marginal value product of those inputs. Thus, the optimal levels are those for which

$$\frac{\partial E\left(YIELD\right)}{\partial\left(PHOS\right)} = \frac{PRICE_{PHOS}}{PRICE_{PEANUTS}} \quad \text{and} \quad \frac{\partial E\left(YIELD\right)}{\partial\left(NITRO\right)} = \frac{PRICE_{NITRO}}{PRICE_{PEANUTS}}$$

EXERCISE 5.9

(a) The marginal effect of experience on wages is

$$\frac{\partial WAGE}{\partial EXPER} = \beta_3 + 2\beta_4 EXPER$$

(b) We expect β_2 to be positive as workers with a higher level of education should receive higher wages. Also, we expect β_3 and β_4 to be positive and negative, respectively. When workers are relatively inexperienced, additional experience leads to a larger increase in their wages than it does after they become relatively experienced. Also, eventually we expect wages to decline with experience as a worker gets older and their productivity declines. A negative β_3 and a positive β_4 gives a quadratic function with these properties.

(c) Wages start to decline at the point where the quadratic curve reaches a maximum. The maximum is reached when the first derivative is zero. Thus, the number of years of experience at which wages start to decline, $EXPER^*$, is such that

$$\beta_3 + 2\beta_4 EXPER^* = 0$$

$$EXPER^* = -\frac{\beta_3}{2\beta_4}$$

(d) (i) A point estimate of the marginal effect of education on wages is

$$\frac{\partial WAGE}{\partial EDUC} = b_2 = 2.2774$$

A 95% interval estimate is given by

$$b_2 \pm t_{(0.975,998)}\text{se}(b_2) = 2.2774 \pm 1.962 \times 0.1394 = (2.0039, 2.5509)$$

(ii) A point estimate of the marginal effect of experience on wages when $EXPER = 4$ is

$$\frac{\partial WAGE}{\partial EXPER} = b_3 + 2b_4 \times (4) = 0.6821 - 8 \times 0.0101 = 0.6013$$

To compute an interval estimate, we need the standard error of this quantity which is given by

$$\begin{aligned}\text{se}(b_3 + 8b_4) &= \sqrt{\text{var}(b_3) + 8^2\,\text{var}(b_4) + 2\times 8 \times \text{cov}(b_3, b_4)}\\ &= \sqrt{0.010987185 + 64 \times 0.000003476 - 16 \times 0.000189259}\\ &= 0.09045\end{aligned}$$

A 95% interval estimate is given by

$$\begin{aligned}(b_3 + 8b_4) \pm t_{(0.975,998)}\text{se}(b_3 + 8b_4) &= 0.6013 \pm 1.962 \times 0.09045\\ &= (0.4238, 0.7788)\end{aligned}$$

Exercise 5.9(d) (continued)

(iii) A point estimate of the marginal effect of experience on wages when $EXPER = 25$ is

$$\frac{\partial WAGE}{\partial EXPER} = b_3 + 2b_4 \times (25) = 0.6821 - 50 \times 0.0101 = 0.1771$$

To compute an interval estimate, we need the standard error of this quantity which is given by

$$\begin{aligned}\mathrm{se}\left(b_3 + 50b_4\right) &= \sqrt{\mathrm{var}\left(b_3\right) + 50^2\,\mathrm{var}\left(b_4\right) + 2\times 50\times \mathrm{cov}\left(b_3, b_4\right)} \\ &= \sqrt{0.010987185 + 2500\times 0.000003476 - 100\times 0.000189259} \\ &= 0.02741\end{aligned}$$

A 95% interval estimate is given by

$$\begin{aligned}\left(b_3 + 50b_4\right) \pm t_{(0.975,998)}\mathrm{se}\left(b_3 + 50b_4\right) &= 0.1771 \pm 1.962 \times 0.02741 \\ &= (0.1233, 0.2309)\end{aligned}$$

(iv) Using the equation derived in part (c), we find:

$$EXPER^* = -\frac{b_3}{2b_4} = \frac{0.6821}{2\times 0.0101} = 33.77$$

We estimate that wages will decline after approximately 34 years of experience.

To obtain an interval estimate for $EXPER^*$, we require $\mathrm{se}\left(-b_3/2b_4\right)$ which in turn requires the derivatives

$$\frac{\partial EXPER^*}{\partial \beta_3} = -\frac{1}{2\beta_4} \qquad\qquad \frac{\partial EXPER^*}{\partial \beta_4} = \frac{\beta_3}{2\beta_4^2}$$

Then,

$$\begin{aligned}\mathrm{var}\left(EXPER^*\right) = &\left(\frac{\partial EXPER^*}{\partial \beta_3}\right)^2 \mathrm{var}\left(b_3\right) + \left(\frac{\partial EXPER^*}{\partial \beta_4}\right)^2 \mathrm{var}\left(b_4\right) \\ &+ 2\left(\frac{\partial EXPER^*}{\partial \beta_3}\right)\left(\frac{\partial EXPER^*}{\partial \beta_4}\right)\mathrm{cov}\left(b_3, b_4\right)\end{aligned}$$

and

$$\mathrm{var}\left(EXPER^*\right) = \left(-\frac{1}{2b_4}\right)^2 \mathrm{var}\left(b_3\right) + \left(\frac{b_3}{2b_4^2}\right)^2 \mathrm{var}\left(b_4\right) + 2\left(-\frac{1}{2b_4}\right)\left(\frac{b_3}{2b_4^2}\right)\mathrm{cov}\left(b_3, b_4\right)$$

Substituting into this expression yields

Exercise 5.9(d)(iv) (continued)

$$\operatorname{var}\left(EXPER^*\right) = \left(\frac{1}{2\times 0.0101}\right)^2 \times 0.010987185 + \left(\frac{0.6821}{2\times 0.0101^2}\right)^2 \times 0.000003476$$

$$-2\times\left(\frac{1}{2\times 0.0101}\right)\times\left(\frac{0.6821}{2\times 0.0101^2}\right)\times 0.000189259$$

$$= 3.131785$$

$$\operatorname{se}\left(EXPER^*\right) = \sqrt{3.131785} = 1.770$$

A 95% interval estimate for $EXPER^*$ is

$$EXPER^* \pm t_{(0.975,998)}\operatorname{se}\left(EXPER^*\right) = 33.77 \pm 1.962\times 1.77 = (30.3, 37.2)$$

<u>Note</u>: The above answers to part (d) are based on hand calculations using the estimates and covariance matrix values reported in Table 5.9 of the text. If the computations are made using software and the file *cps4c_small.dat*, slightly different results are obtained. These results do not suffer from the rounding error caused by truncating the number of digits reported in Table 5.9. The answers obtained using software for parts (d)(ii) ,(iii), and (iv) are:

(d) (ii)

$$\left(b_3 + 8b_4\right) \pm t_{(0.975,998)}\operatorname{se}\left(b_3 + 8b_4\right) = 0.60137 \pm 1.962\times 0.090418$$

$$= (0.4239, 0.7789)$$

(iii)

$$\left(b_3 + 50b_4\right) \pm t_{(0.975,998)}\operatorname{se}\left(b_3 + 50b_4\right) = 0.17756 \pm 1.962\times 0.027425$$

$$= (0.1237, 0.2314)$$

(iv)

$$EXPER^* \pm t_{(0.975,998)}\operatorname{se}\left(EXPER^*\right) = 33.798 \pm 1.962\times 1.7762 = (30.3, 37.3)$$

EXERCISE 5.10

The EViews output for verifying the answers to Exercise 5.1 is given in the following table.

Method: Least Squares
Dependent Variable: Y
Method: Least Squares
Included observations: 9

	Coefficient	Std. Error	t-Statistic	Prob.
X1	1.000000	0.266580	3.751221	0.0095
X2	0.812500	0.199935	4.063823	0.0066
X3	0.400000	0.252900	1.581654	0.1648

R-squared	0.760156	Mean dependent var	1.000000
Adjusted R-squared	0.680208	S.D. dependent var	1.414214
S.E. of regression	0.799740	Akaike info criterion	2.652140
Sum squared resid	3.837500	Schwarz criterion	2.717882
Log likelihood	–8.934631	Hannan-Quinn criter.	1.728217

(c) The least squares estimates can be read directly from the table.

(d) The residuals from the estimated equation are:

–0.4000	0.9875	–0.0250	–0.3750	–1.4125	0.0250	0.6000	0.4125	0.1875

(e) The estimate $\hat{\sigma}^2$ is given by the square of "S.E. of regression". That is,

$$\hat{\sigma}^2 = 0.79974^2 = 0.639584$$

(f) The correlation matrix for the three variables is

	X2	X3	Y
X2	1.000000	0.000000	0.812500
X3	0.000000	1.000000	0.316228
Y	0.812500	0.316228	1.000000

The correlation between x_2 and x_3 is zero.

(g) The standard error for b_2 can be read directly from the EViews output.

(h) From the EViews output, SSE = "Sum squared resid" = 3.8375, and $R^2 = 0.760156$.

To obtain SST note that $s_y^2 = 1.414214^2 = 2$. Then,

$$SST = \sum (y_i - \bar{y})^2 = (n-1)s_y^2 = 8 \times 2 = 16$$

$$SSR = SST - SSE = 16 - 3.8375 = 12.1625$$

EXERCISE 5.11

(a) Estimates, standard errors and *p*-values for each of the coefficients in each of the estimated share equations are given in the following table.

Explanatory Variables		Dependent Variable					
		Food	Fuel	Clothing	Alcohol	Transport	Other
Constant	Estimate	0.8798	0.3179	−0.2816	0.0149	−0.0191	0.0881
	Std Error	0.0512	0.0265	0.0510	0.0370	0.0572	0.0536
	p-value	0.0000	0.0000	0.0000	0.6878	0.7382	0.1006
ln(*TOTEXP*)	Estimate	−0.1477	−0.0560	0.0929	0.0327	0.0321	0.0459
	Std Error	0.0113	0.0058	0.0112	0.0082	0.0126	0.0118
	p-value	0.0000	0.0000	0.0000	0.0001	0.0111	0.0001
AGE	Estimate	0.00227	0.00044	−0.00056	−0.00220	0.00077	−0.00071
	Std Error	0.00055	0.00029	0.00055	0.00040	0.00062	0.00058
	p-value	0.0000	0.1245	0.3062	0.0000	0.2167	0.2242
NK	Estimate	0.0397	0.0062	−0.0048	−0.0148	−0.0123	−0.0139
	Std Error	0.0084	0.0044	0.0084	0.0061	0.0094	0.0088
	p-value	0.0000	0.1587	0.5658	0.0152	0.1921	0.1157

An increase in total expenditure leads to decreases in the budget shares allocated to food and fuel and increases in the budget shares of the commodity groups clothing, alcohol, transport and other. Households with an older household head devote a higher proportion of their budget to food, fuel and transport and a lower proportion to clothing, alcohol and other. Having more children means a higher proportion spent on food and fuel and lower proportions spent on the other commodities.

The coefficients of $\ln(TOTEXP)$ are significantly different from zero for all commodity groups. At a 5% significance level, age has a significant effect on the shares of food and alcohol, but its impact on the other budget shares is measured less precisely. Significance tests for the coefficients of the number of children yield a similar result. NK has an impact on the food and alcohol shares, but we can be less certain about the effect on the other groups. To summarize, $\ln(TOTEXP)$ has a clear impact in all equations, but the effect of *AGE* and *NK* is only significant in the food and alcohol equations.

Exercise 5.11 (continued)

(b) The *t*-values and *p*-values for testing $H_0: \beta_2 \le 0$ against $H_1: \beta_2 > 0$ are reported in the table below. Using a 5% level of significance, the critical value for each test is $t_{(0.95, 496)} = 1.648$.

	t-value	*p*-value	decision
WFOOD	−13.083	1.0000	Do not reject H_0
WFUEL	−9.569	1.0000	Do not reject H_0
WCLOTH	8.266	0.0000	Reject H_0
WALC	4.012	0.0000	Reject H_0
WTRANS	2.548	0.0056	Reject H_0
WOTHER	3.884	0.0001	Reject H_0

Those commodities which are regarded as necessities ($b_2 < 0$) are food and fuel. The tests suggest the rest are luxuries. While alcohol, transportation and other might be luxuries, it is difficult to see clothing categorized as a luxury. Perhaps a finer classification is necessary to distinguish between basic and luxury clothing.

EXERCISE 5.12

(a) The expected sign for β_2 is negative because, as the number of grams in a given sale increases, the price per gram should decrease, implying a discount for larger sales. We expect β_3 to be positive; the purer the cocaine, the higher the price. The sign for β_4 will depend on how demand and supply are changing over time. For example, a fixed demand and an increasing supply will lead to a fall in price. A fixed supply and increased demand would lead to a rise in price.

(b) The estimated equation is:

$$\begin{array}{llllll} PRICE = & 90.8467 - & 0.0600\,QUANT + & 0.1162\,QUAL - & 2.3546\,TREND & R^2 = 0.5097 \\ \text{(se)} & (8.5803) & (0.0102) & (0.2033) & (1.3861) & \\ (t) & (10.588) & (-5.892) & (0.5717) & (-1.6987) & \end{array}$$

The estimated values for β_2, β_3 and β_4 are -0.0600, 0.1162 and -2.3546, respectively. They imply that as quantity (number of grams in one sale) increases by 1 unit, the price will go down by 0.0600. Also, as the quality increases by 1 unit the price goes up by 0.1162. As time increases by 1 year, the price decreases by 2.3546. All the signs turn out according to our expectations, with β_4 implying supply has been increasing faster than demand.

(c) The proportion of variation in cocaine price explained by the variation in quantity, quality and time is 0.5097.

(d) For this hypothesis we test $H_0:\beta_2 \ge 0$ against $H_1:\beta_2 < 0$. The calculated t-value is -5.892. We reject H_0 if the calculated t is less than the critical $t_{(0.95,52)} = -1.675$. Since the calculated t is less than the critical t value, we reject H_0 and conclude that sellers are willing to accept a lower price if they can make sales in larger quantities.

(e) We want to test $H_0:\beta_3 \le 0$ against $H_1:\beta_3 > 0$. The calculated t-value is 0.5717. At $\alpha = 0.05$ we reject H_0 if the calculated t is greater than 1.675. Since for this case, the calculated t is not greater than the critical t, we do not reject H_0. We cannot conclude that a premium is paid for better quality cocaine.

(f) The average annual change in the cocaine price is given by the value of $b_4 = -2.3546$. It has a negative sign suggesting that the price decreases over time. A possible reason for a decreasing price is the development of improved technology for producing cocaine, such that suppliers can produce more at the same cost.

EXERCISE 5.13

(a) The estimated regression is

$$\begin{array}{l} PRICE = -41948 + 90.970SQFT - 755.04AGE \\ \text{(se)} \qquad (6990) \quad (2.403) \qquad\quad (140.89) \end{array}$$

(i) The estimate $b_2 = 90.97$ implies that holding age constant, on average, a one square foot increase in the size of the house increases the selling price by 90.97 dollars.

The estimate $b_3 = -755.04$ implies that holding *SQFT* constant, on average, an increase in the age of the house by one year decreases the selling price by 755.04 dollars.

The estimate b_1 could be interpreted as the average price of land if its value was meaningful. Since a negative price is unrealistic, we view the equation as a poor model for data values in the vicinity of $SQFT = 0$ and $AGE = 0$.

(ii) A point estimate for the price increase is $\frac{\partial PRICE}{\partial SQFT} = b_2 = 90.9698$

A 95% interval estimate for β_2, given that $t_c = t_{(0.975,1077)} = 1.962$ is

$$b_2 \pm t_c \,\text{se}(b_2) = 90.9698 \pm 1.962 \times 2.4031 = (86.25, 95.69)$$

(iii) The *t*-value for testing $H_0: \beta_3 \geq -1000$ against $H_1: \beta_3 < -1000$ is

$$t = \frac{b_3 - (-1000)}{\text{se}(b_3)} = \frac{-755.0414 - (-1000)}{140.8936} = 1.7386$$

The corresponding *p*-value is $P\left(t_{(1077)} < 1.7386\right) = 0.959$. The critical value for a 5% significance level is $t_{(0.05,1077)} = -1.646$. The rejection region is $t \leq -1.646$. Since the *t*-value is greater than the critical value and the *p*-value is greater than 0.05, we fail to reject the null hypothesis. We conclude that the estimated equation is compatible with the hypothesis that an extra year of age decreases the price by \$1000 or less.

(b) The estimated regression is:

$$\begin{array}{l} PRICE = 170150 - 55.784SQFT + 0.023153SQFT^2 - 2797.8AGE + 30.160AGE^2 \\ \text{(se)} \qquad (10432) \quad (6.389) \qquad\quad (0.000964) \qquad\quad (305.1) \qquad\quad (5.071) \end{array}$$

For the remainder of part (b), we refer to these estimates as b_1, b_2, b_3, b_4, b_5 in the same order as they appear in the equation, with corresponding parameters $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$.

(i) The marginal effect of *SQFT* on *PRICE* is given by

$$\frac{\partial PRICE}{\partial SQFT} = \beta_2 + 2\beta_3 SQFT$$

Exercise 5.13(b)(i) (continued)

The estimated marginal effect of *SQFT* on *PRICE* for the smallest house where $SQFT = 662$ is

$$\frac{\partial PRICE}{\partial SQFT} = -55.7842 + 2 \times 0.023153 \times 662 = -25.13$$

The estimated marginal effect of *SQFT* on *PRICE* for a house with *SQFT* = 2300 is

$$\frac{\partial PRICE}{\partial SQFT} = -55.7842 + 2 \times 0.023153 \times 2300 = 50.72$$

The estimated marginal effect of *SQFT* on *PRICE* for the largest house where $SQFT = 7897$ is

$$\frac{\partial PRICE}{\partial SQFT} = -55.7842 + 2 \times 0.023153 \times 7897 = 309.89$$

These values suggest that as the size of the house gets larger the price or cost for extra square feet gets larger, and that, for small houses, extra space leads to a decline in price. The result for small houses is unrealistic. However, it is possible that additional square feet leads to a higher price increase in larger houses than it does in smaller houses.

(ii) The marginal effect of *AGE* on *PRICE* is given by

$$\frac{\partial PRICE}{\partial AGE} = \beta_4 + 2\beta_5 AGE$$

The estimated marginal effect of *AGE* on *PRICE* for the oldest house $(AGE = 80)$ is

$$\frac{\partial PRICE}{\partial AGE} = -2797.788 + 2 \times 30.16033 \times 80 = 2027.86$$

The estimated marginal effect of *AGE* on *PRICE* for a house when $AGE = 20$ is

$$\frac{\partial PRICE}{\partial AGE} = -2797.788 + 2 \times 30.16033 \times 20 = -1591.38$$

The estimated marginal effect of *AGE* on *PRICE* for the newest house $(AGE = 1)$ is

$$\frac{\partial PRICE}{\partial AGE} = -2797.788 + 2 \times 30.16033 \times 1 = -2737.47$$

When a house is new, extra years of age have the greatest negative effect on price. Aging has a smaller and smaller negative effect as the house gets older. This result is as expected. However, unless a house has some kind of heritage value, it is unrealistic for the oldest houses to increase in price as they continue to age, as is suggested by the marginal effect for $AGE = 80$. The quadratic function has a minimum at an earlier age than is desirable.

Exercise 5.13(b) (continued)

(iii) A 95% interval for the marginal effect of *SQFT* on *PRICE* when $SQFT = 2300$, and using $t_c = t_{(0.975,1075)}$ = 1.962, is:

$$\text{me} \pm t_c \,\text{se}\left(\text{me}\right) = 50.719 \pm 1.962 \times 2.5472 = (45.72, 55.72)$$

The standard error for me can be found using software or from

$$\begin{aligned}\text{se}\left(\text{me}\right) &= \sqrt{\text{var}\left(b_2\right) + 4600^2\,\text{var}\left(b_3\right) + 2\times 4600\,\text{cov}\left(b_2, b_3\right)} \\ &= \sqrt{40.82499 + 4600^2 \times 9.296015\times 10^{-7} + 9200 \times (-0.005870334)} \\ &= 2.5472\end{aligned}$$

(iv) The null and alternative hypotheses are

$$H_0: \beta_4 + 40\beta_5 \geq -1000 \qquad\qquad H_1: \beta_4 + 40\beta_5 < -1000$$

The *t*-value for the test is

$$t = \frac{b_4 + 40b_5 - (-1000)}{\text{se}\left(b_4 + 40b_5\right)} = \frac{-591.375}{139.554} = -4.238$$

The corresponding *p*-value is $P\left(t_{(1075)} < -4.238\right) = 0.0000$. The critical value for a 5% significance level is $t_{(0.05,1075)} = -1.646$. The rejection region is $t \leq -1.646$. Since the *t*-value is less than the critical value and the *p*-value is less than 0.05, we reject the null hypothesis. We conclude that, for a 20-year old house, an extra year of age decreases the price by more than \$1000.

The standard error $\text{se}\left(b_4 + 40b_5\right)$ can be found using software or from

$$\begin{aligned}\text{se}\left(b_4 + 40b_5\right) &= \sqrt{\text{var}\left(b_4\right) + 40^2\,\text{var}\left(b_5\right) + 2\times 40\,\text{cov}\left(b_4, b_5\right)} \\ &= \sqrt{93095.48 + 1600\times 25.71554 + 80\times(-1434.561)} \\ &= 139.55\end{aligned}$$

(c) The estimated regression is:

$$\begin{aligned}&PRICE = \underset{(12143)}{114597} - \underset{(6.898)}{30.729}SQFT + \underset{(0.000943)}{0.022185}SQFT^2 \\ &\qquad - \underset{(410.61)}{442.03}AGE + \underset{(4.939)}{26.519}AGE^2 - \underset{(0.11244)}{0.93062}SQFT\times AGE\end{aligned}$$

(se)

For the remainder of part (c), we refer to these estimates as $b_1, b_2, b_3, b_4, b_5, b_6$ in the same order as they appear in the equation, with corresponding parameters $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$.

Exercise 5.13(c) (continued)

(i) The marginal effect of *SQFT* on *PRICE* is given by

$$\frac{\partial PRICE}{\partial SQFT} = \beta_2 + 2\beta_3 SQFT + \beta_6 AGE$$

When $AGE = 20$, the estimated marginal effect of *SQFT* on *PRICE* for the smallest house where $SQFT = 662$ is

$$\frac{\partial PRICE}{\partial SQFT} = -30.7289 + 2\times 0.022185\times 662 - 0.93062\times 20 = -19.97$$

When $AGE = 20$ the estimated marginal effect of *SQFT* on *PRICE* for a house with *SQFT* = 2300 is

$$\frac{\partial PRICE}{\partial SQFT} = -30.7289 + 2\times 0.022185\times 2300 - 0.93062\times 20 = 52.71$$

When $AGE = 20$, the estimated marginal effect of *SQFT* on *PRICE* for the largest house where $SQFT = 7897$ is

$$\frac{\partial PRICE}{\partial SQFT} = -30.7289 + 2\times 0.0221846\times 7897 - 0.930621\times 20 = 301.04$$

These values lead to similar conclusions to those obtained in part (b). As the size of the house gets larger the price or cost for extra square feet gets larger. For small houses, extra space appears to lead to a decline in price. This result for small houses is unrealistic. It would be more realistic if the quadratic reached a minimum before the smallest house in the sample.

(ii) The marginal effect of *AGE* on *PRICE* is given by

$$\frac{\partial PRICE}{\partial AGE} = \beta_4 + 2\beta_5 AGE + \beta_6 SQFT$$

When $SQFT = 2300$, the estimated marginal effect of *AGE* on *PRICE* for the oldest house $(AGE = 80)$ is

$$\frac{\partial PRICE}{\partial AGE} = -442.0336 + 2\times 26.519\times 80 - 0.93062\times 2300 = 1660.6$$

When $SQFT = 2300$, the estimated marginal effect of *AGE* on *PRICE* for a house of $AGE = 20$ is

$$\frac{\partial PRICE}{\partial AGE} = -442.0336 + 2\times 26.519\times 20 - 0.93062\times 2300 = -1521.7$$

Exercise 5.13(c)(ii) (continued)

When $SQFT = 2300$, the estimated marginal effect of *AGE* on *PRICE* for the newest house $(AGE = 1)$ is

$$\frac{\partial PRICE}{\partial AGE} = -442.0336 + 2 \times 26.519 \times 1 - 0.93062 \times 2300 = -2529.4$$

These results lead to similar conclusions to those reached in part (b).When a house is new, extra years of age have the greatest negative effect on price. Aging has a smaller and smaller negative effect as the house gets older. This result is as expected. However, unless a house has some kind of heritage value, the positive marginal effect for $AGE = 80$ is unrealistic. We do not expect the oldest houses to increase in price as they continue to age.

(iii) A 95% interval for the marginal effect of *SQFT* on *PRICE* when $SQFT = 2300$ and $AGE = 20$, and using $t_c = t_{(0.975,1074)}$ = 1.962, is:

$$\text{me} \pm t_c \,\text{se}\left(\text{me}\right) = 52.708 \pm 1.962 \times 2.4825 = (47.84, 57.58)$$

The standard error for me was found using software.

(iv) The null and alternative hypotheses are

$$H_0 : \beta_4 + 40\beta_5 + 2300\beta_6 \geq -1000 \qquad H_1 : \beta_4 + 40\beta_5 + 2300\beta_6 < -1000$$

The *t*-value for the test is

$$t = \frac{b_4 + 40b_5 + 2300b_6 - (-1000)}{\text{se}\left(b_4 + 40b_5 + 2300b_6\right)} = \frac{-521.701}{135.630} = -3.847$$

The corresponding *p*-value is $P\left(t_{(1074)} < -3.847\right) = 0.0001$. The critical value for a 5% significance level is $t_{(0.05,1074)} = -1.646$. The rejection region is $t \leq -1.646$. Since the *t*-value is less than the critical value and the *p*-value is less than 0.05, we reject the null hypothesis. We conclude that, for a 20-year old house with $SQFT = 2300$, an extra year of age decreases the price by more than $1000.

(d) The results from the two quadratic specifications in parts (c) and (d) are similar, but they are vastly different from those from the linear model in part (a). In part (a) the marginal effect of *SQFT* is constant at 91, whereas in parts (b) and (c), it varies from approximately -20 to +300. The marginal effect of *AGE* is constant at -755 in part (a) but varies from approximately -2600 to +1800 in parts (b) and (c), with a similar pattern in (b) and (c), but some noticeable differences in magnitudes. These differences carry over to the interval estimates for the marginal effect of *SQFT* and to the hypothesis tests on the marginal effect of *AGE*. The marginal effects are clearly not constant and so the linear function is inadequate. Both quadratic functions are an improvement, but they do give some counterintuitive results for old houses and small houses. It is interesting that the intercept is positive in the quadratic equations, and hence has the potential to be interpreted as the average price of the land. Both estimates seem large however, relative to house prices.

EXERCISE 5.14

(a) The estimated regression is:

$$\ln(PRICE) = 11.1196 - 0.038762\,SQFT100 - 0.017555\,AGE + 0.00017336\,AGE^2$$
$$\text{(se)} \quad (0.0274) \quad (0.000869) \qquad (0.001356) \qquad (0.00002266)$$

(b) The estimate $\hat{\alpha}_2 = 0.03876$ suggests that, holding age constant, an increase in the size of the house by one hundred square feet increases the price by 3.88% on average.

(c) The required derivative is given by

$$\frac{\partial \ln(PRICE)}{\partial AGE} = \alpha_3 + 2\alpha_4 AGE$$

When $AGE = 5$, $\dfrac{\partial \ln(PRICE)}{\partial AGE} = -0.017555 + 2 \times 0.00017336 \times 5 = -0.01582$

This estimate implies that, holding $SQFT$ constant, the price of a 5-year old house will decrease at a rate of 1.58% per year.

When $AGE = 20$, $\dfrac{\partial \ln(PRICE)}{\partial AGE} = -0.017555 + 2 \times 0.00017336 \times 20 = -0.01062$

This estimate implies that, holding $SQFT$ constant, the price of a 20-year old house will decrease at a rate of 1.06% per year.

(d) The required derivatives are given by

$$\begin{aligned}\frac{\partial PRICE}{\partial AGE} &= \left(\alpha_3 + 2\alpha_4 AGE\right) \times PRICE \\ &= \left(\alpha_3 + 2\alpha_4 AGE\right) \times \exp\left\{\alpha_1 + \alpha_2 SQFT100 + \alpha_3 AGE + \alpha_4 AGE^2\right\}\end{aligned}$$

$$\begin{aligned}\frac{\partial PRICE}{\partial SQFT100} &= \alpha_2 PRICE \\ &= \alpha_2 \times \exp\left\{\alpha_1 + \alpha_2 SQFT100 + \alpha_3 AGE + \alpha_4 AGE^2\right\}\end{aligned}$$

where $\exp\{x\}$ is notation for the exponential function e^x.

(e) To estimate these marginal effects we first find

$$\begin{aligned}PRICE_0 &= \exp\left\{\hat{\alpha}_1 + \hat{\alpha}_2 SQFT100 + \hat{\alpha}_3 AGE + \hat{\alpha}_4 AGE^2\right\} \\ &= \exp\left\{11.11959 + 0.0387624 \times 23 - 0.017555 \times 20 + 0.00017336 \times 20^2\right\} \\ &= 124165\end{aligned}$$

Then,

Exercise 5.14(e) (continued)

$$\frac{\partial PRICE}{\partial AGE} = \left(-0.017555 + 2 \times 0.00017336 \times 20\right) \times 124165 = -1318.7$$

$$\frac{\partial PRICE}{\partial SQFT100} = 0.0387624 \times 124165 = 4813$$

(f) We require the standard errors of

$$\frac{\partial PRICE}{\partial AGE} = \left(\hat{\alpha}_3 + 40\hat{\alpha}_4\right) \times \exp\left\{\hat{\alpha}_1 + 23\hat{\alpha}_2 + 20\hat{\alpha}_3 + 400\hat{\alpha}_4\right\}$$

$$\frac{\partial PRICE}{\partial SQFT100} = \hat{\alpha}_2 \times \exp\left\{\hat{\alpha}_1 + 23\hat{\alpha}_2 + 20\hat{\alpha}_3 + 400\hat{\alpha}_4\right\}$$

These expressions are nonlinear functions of the least squares estimators for the α 's. To compute their standard errors, we need the delta method introduced on pages 193-4 of the text. Using computer software, we find the standard errors are

$$\text{se}\left(\frac{\partial PRICE}{\partial AGE}\right) = 72.671 \qquad \text{se}\left(\frac{\partial PRICE}{\partial SQFT100}\right) = 121.637$$

(g) A 95% interval estimate for the marginal effect of *SQFT100* is

$$\text{me} \pm t_{(0.975,1076)}\,\text{se}\left(\text{me}\right) = 4812.9 \pm 1.962 \times 121.637 = (4574, 5052)$$

(h) The null and alternative hypotheses are

$$H_0 : \left(\alpha_3 + 40\alpha_4\right) \times \exp\left\{\alpha_1 + 23\alpha_2 + 20\alpha_3 + 400\alpha_4\right\} \geq -1000$$

$$H_1 : \left(\alpha_3 + 40\alpha_4\right) \times \exp\left\{\alpha_1 + 23\alpha_2 + 20\alpha_3 + 400\alpha_4\right\} < -1000$$

The calculated value of the *t*-statistic is

$$t = \frac{-1318.7 - (-1000)}{72.671} = -4.386$$

The corresponding *p*-value is $P\left(t_{(1076)} < -4.386\right) = 0.0000$. The critical value for a 5% significance level is $t_{(0.05,1076)} = -1.646$. The rejection region is $t \leq -1.646$. Since the *t*-value is less than the critical value and the *p*-value is less than 0.05, we reject the null hypothesis. We conclude that, for a 20-year old house with $SQFT = 2300$, an extra year of age decreases the price by more than \$1000.

Remark: A comparison of the results in parts (g) and (h) with those from the quadratic function with the interaction term in Exercise 5.13(c) shows that similar conclusions are reached, although the interval estimate in (g) is narrower, and the estimated marginal effect is smaller. Similarly, the marginal effect in (h) is smaller (in absolute value) and estimated more precisely than its counterpart in Exercise 5.13(c).

EXERCISE 5.15

(a) The estimated regression model is:

$$\begin{array}{l} VOTE = 52.16 + 0.6434 GROWTH - 0.1721 INFLATION \\ \text{(se)} \quad (1.46) \quad (0.1656) \qquad\qquad (0.4290) \end{array}$$

The hypothesis test results on the significance of the coefficients are:

$H_0:\beta_2 = 0 \quad H_1:\beta_2 > 0$ *p*-value = 0.0003 significant at 10% level

$H_0:\beta_3 = 0 \quad H_1:\beta_3 < 0$ *p*-value = 0.3456 not significant at 10% level

One-tail tests were used because more growth is considered favorable, and more inflation is considered not favorable, for re-election of the incumbent party.

(b) (i) For $INFLATION = 4$ and $GROWTH = -3$, the predicted percentage vote is

$$VOTE_0 = 52.1565 + 0.64342 \times (-3) - 0.172076 \times 4 = 49.54$$

(ii) For $INFLATION = 4$ and $GROWTH = 0$, the predicted percentage vote is

$$VOTE_0 = 52.1565 + 0.64342 \times (0) - 0.172076 \times 4 = 51.47$$

(iii) For $INFLATION = 4$ and $GROWTH = 3$, the predicted percentage vote is

$$VOTE_0 = 52.1565 + 0.64342 \times 3 - 0.172076 \times 4 = 53.40$$

(c) Ignoring the error term, the incumbent party will get the majority of the vote when

$$\beta_1 + \beta_2 GROWTH + \beta_3 INFLATION > 50$$

When $INFLATION = 4$, this requirement becomes

$$\beta_1 + \beta_2 GROWTH + 4\beta_3 > 50$$

(i) When $GROWTH = -3$, the hypotheses are

$$H_0:\beta_1 - 3\beta_2 + 4\beta_3 \le 50 \qquad H_1:\beta_1 - 3\beta_2 + 4\beta_3 > 50$$

Given that $t_{(0.99,30)} = 2.457$, we reject H_0 when

$$t = \frac{b_1 - 3b_2 + 4b_3 - 50}{\text{se}\left(b_1 - 3b_2 + 4b_3\right)} > 2.457$$

Now,

$$\begin{aligned} \text{var}\left(b_1 - 3b_2 + 4b_3\right) &= \text{var}\left(b_1\right) + 3^2\,\text{var}\left(b_2\right) + 4^2\,\text{var}\left(b_3\right) - 2 \times 3\text{cov}\left(b_1, b_2\right) \\ &\quad + 2 \times 4\text{cov}\left(b_1, b_3\right) - 2 \times 3 \times 4\text{cov}\left(b_2, b_3\right) \\ &= 2.127815 + 9 \times 0.027433 + 16 \times 0.184003 + 6 \times 0.048748 \\ &\quad - 8 \times 0.498011 - 24 \times 0.011860 \\ &= 1.34252 \end{aligned}$$

Exercise 5.15(c)(i) (continued)

The calculated *t*-value is

$$t = \frac{b_1 - 3b_2 + 4b_3 - 50}{\mathrm{se}\left(b_1 - 3b_2 + 4b_3\right)} = \frac{49.538 - 50}{\sqrt{1.34252}} = -0.399$$

Since $-0.399 < 2.457$, we do not reject H_0. There is no evidence to suggest that the incumbent part will get the majority of the vote when $INFLATION = 4$ and $GROWTH = -3$.

(ii) When $GROWTH = 0$, the hypotheses are

$$H_0: \beta_1 + 4\beta_3 \le 50 \qquad H_1: \beta_1 + 4\beta_3 > 50$$

We reject H_0 when $t = \dfrac{b_1 + 4b_3 - 50}{\mathrm{se}\left(b_1 + 4b_3\right)} > 2.457$.

The standard error can be calculated from a similar expression to that given in (c)(i). Using computer software, we find $\mathrm{se}\left(b_1 + 4b_3\right) = 1.04296$.

The calculated *t*-value is

$$t = \frac{b_1 + 4b_3 - 50}{\mathrm{se}\left(b_1 + 4b_3\right)} = \frac{51.4682 - 50}{1.04296} = 1.408$$

Since $1.408 < 2.457$, we do not reject H_0. There is insufficient evidence to suggest that the incumbent part will get the majority of the vote when $INFLATION = 4$ and $GROWTH = 0$.

(iii) When $GROWTH = 3$, the hypotheses are

$$H_0: \beta_1 + 3\beta_2 + 4\beta_3 \le 50 \qquad H_1: \beta_1 + 3\beta_2 + 4\beta_3 > 50$$

We reject H_0 when $t = \dfrac{b_1 + 3b_2 + 4b_3 - 50}{\mathrm{se}\left(b_1 + 3b_2 + 4b_3\right)} > 2.457$.

The standard error can be calculated from a similar expression to that given in (c)(i). Using computer software, we find $\mathrm{se}\left(b_1 + 3b_2 + 4b_3\right) = 1.15188$.

The calculated *t*-value is

$$t = \frac{b_1 + 3b_2 + 4b_3 - 50}{\mathrm{se}\left(b_1 + 3b_2 + 4b_3\right)} = \frac{53.3985 - 50}{1.15188} = 2.950$$

Since $2.950 > 2.457$, we reject H_0. We conclude that the incumbent part will get the majority of the vote when $INFLATION = 4$ and $GROWTH = 3$.

As a president seeking re-election, you would not want to conclude that you would be re-elected without strong evidence to support such a conclusion. Setting up re-election as the alternative hypothesis with a 1% significance level reflects this scenario.

EXERCISE 5.16

(a) The estimated regression is:

$$\widehat{SAL1} = 22963 - 470.845PR1 + 92.990PR2 + 165.113PR3 \qquad R^2 = 0.443$$
$$\text{(se)} \quad (9806) \quad (79.578) \quad (70.013) \quad (93.670)$$

(b) The estimate $b_2 = -470.845$ suggests that, holding *PR2* and *PR3* constant, a one cent increase in the price of brand 1 leads to a decrease in the sales of brand 1 by 471 units.

The estimate $b_3 = 92.990$ suggests that, holding *PR1* and *PR3* constant, a one cent increase in the price of brand 2 leads to an increase in the sales of brand 1 by 93 units.

The estimate $b_4 = 165.113$ suggests that, holding *PR*1 and *PR*2 constant, a one cent increase in the price of brand 3 leads to an increase in the sales of brand 1 by 165 units.

The estimates of β_2, β_3 and β_4 have the expected signs. The sign of β_2 is negative, reflecting the fact that quantity demanded will fall as price rises, while the signs of the other two coefficients are positive, reflecting the fact that brands 2 and 3 are substitutes. Increases in their prices will increase the demand for brand 1.

(c) The hypothesis test results on the significance of the coefficients are:

$H_0: \beta_2 = 0$	$H_1: \beta_2 < 0$	*p*-value = 0.0000	significant at 5% level
$H_0: \beta_3 = 0$	$H_1: \beta_3 > 0$	*p*-value = 0.0952	not significant at 5% level
$H_0: \beta_4 = 0$	$H_1: \beta_4 > 0$	*p*-value = 0.0422	significant at 5% level

(d) (i) The hypotheses are

$$H_0: \beta_2 = -300 \qquad H_1: \beta_2 \neq -300$$

Since $t_{(0.975,48)} = 2.011$, we reject H_0 if $t = (b_2 + 300)/\text{se}(b_2) > 2.011$ or $t < -2.011$.

The *t*-value is

$$t = \frac{b_2 + 300}{\text{se}(b_2)} = \frac{-470.845 + 300}{79.578} = -2.147$$

Since $-2.147 < -2.011$, we reject H_0 and conclude that a 1-cent increase in the price of brand 1 does not reduce its sales by 300 cans.

Exercise 5.16(d) (continued)

(ii) The hypotheses are

$$H_0:\beta_3 = 300 \qquad H_1:\beta_3 \neq 300$$

Since $t_{(0.975,48)} = 2.011$, we reject H_0 if $t = (b_3 - 300)/\text{se}(b_3) > 2.011$ or $t < -2.011$.

The t-value is

$$t = \frac{b_3 + 300}{\text{se}(b_3)} = \frac{92.990 - 300}{70.013} = -2.957$$

Since $-2.957 < -2.011$, we reject H_0 and conclude that a 1-cent increase in the price of brand 2 does not increase sales of brand 1 by 300 cans.

(iii) The hypotheses are

$$H_0:\beta_4 = 300 \qquad H_1:\beta_4 \neq 300$$

Since $t_{(0.975,48)} = 2.011$, we reject H_0 if $t = (b_4 - 300)/\text{se}(b_4) > 2.011$ or $t < -2.011$.

The t-value is

$$t = \frac{b_4 - 300}{\text{se}(b_4)} = \frac{165.113 - 300}{93.670} = -1.440$$

Since $-2.011 < -1.440 < 2.011$, we do not reject H_0. There is no evidence to suggest that the increase in sales of brand 1 from a 1-cent increase in the price of brand 3 is different from 300 cans.

(iv) Price changes in brands 2 and 3 will have the same effect on sales of brand 1 if $\beta_3 = \beta_4$.

Thus we test $H_0:\beta_3 = \beta_4$ against the alternative $H_1:\beta_3 \neq \beta_4$ and we reject H_0 if $t > 2.011$ or $t < -2.011$. The t-statistic is calculated as follows:

$$t = \frac{b_3 - b_4}{\text{se}(b_3 - b_4)} = \frac{92.990 - 165.113}{123.118} = -0.586$$

The standard error $\text{se}(b_3 - b_4) = 123.118$ can be calculated using computer software or from the coefficient covariance matrix as follows

$$\begin{aligned}\text{se}(b_3 - b_4) &= \sqrt{\text{var}(b_3) + \text{var}(b_4) - 2\text{cov}(b_3, b_4)} \\ &= \sqrt{4901.763 + 8774.127 - 2 \times (-741.048)} \\ &= 123.118\end{aligned}$$

Since $-2.011 < -0.586 < 2.011$, we fail to reject H_0. There is no evidence to suggest that price changes in brands 2 and 3 have different effects on sales of brand 1.

Exercise 5.16(d)(iv) (continued)

In part (ii) we concluded that the effect of a price increase in brand 2 was not 300 cans. In part (iii) we concluded that the effect of a price increase in brand 3 could be 300 cans. And in part (iv) we concluded that the effect of increases in prices for brands 2 and 3 could be equal. On the surface, this may seem like a contradiction: the results from parts (ii) and (iii) suggest the effects are different and the part (iv) result suggests they are the same. To appreciate that the hypothesis-test conclusions are indeed compatible, it must be appreciated that we never conclude null hypotheses are true, only that we have insufficient evidence to reject them. Thus, in part (iii), the effect of a price increase in brand 3 could be 300 cans, but it also could be something else. And in part (iv) it could be true that $\beta_3 = \beta_4$, but it could also be true that they are not equal.

(v) Suppose that prices are set at $PR1_0, PR2_0$ and $PR3_0$ and that average sales are $SAL1_0$. That is,

$$SAL1_0 = \beta_1 + \beta_2 PR1_0 + \beta_3 PR2_0 + \beta_4 PR3_0$$

(Strickly speaking, we are looking at no change in *average* sales so we can ignore the error term.)

Now suppose that all prices go up by 1 cent and that average sales do not change. That is,

$$\begin{aligned} SAL1_0 &= \beta_1 + \beta_2\left(PR1_0 + 1\right) + \beta_3\left(PR2_0 + 1\right) + \beta_4\left(PR3_0 + 1\right) \\ &= \beta_1 + \beta_2 PR1_0 + \beta_3 PR2_0 + \beta_4 PR3_0 + \left(\beta_2 + \beta_3 + \beta_4\right) \end{aligned}$$

For $SAL1_0$ to be the same in these two equations we require $\beta_2 + \beta_3 + \beta_4 = 0$. Thus, we test

$$H_0 : \beta_2 + \beta_3 + \beta_4 = 0 \qquad H_1 : \beta_2 + \beta_3 + \beta_4 \neq 0$$

The *t*-value is calculated as follows:

$$t = \frac{b_2 + b_3 + b_4}{\text{se}(b_2 + b_3 + b_4)} = \frac{-470.845 + 92.990 + 165.113}{123.416} = -1.724$$

Since $-2.011 < -1.724 < 2.011$, we fail to reject H_0. The results are compatible with the hypothesis that sales remain unchanged if all 3 prices go up by 1 cent.

For calculation of $\text{se}(b_2 + b_3 + b_4) = 123.416$, we can use computer software or

$$\begin{aligned} \text{se}\left(b_2 + b_3 - b_4\right) &= \sqrt{\text{var}\left(b_2\right) + \text{var}\left(b_3\right) + \text{var}\left(b_4\right) + 2\text{cov}\left(b_2, b_3\right) + 2\text{cov}\left(b_2, b_4\right) + 2\text{cov}\left(b_3, b_4\right)} \\ &= \sqrt{6332.635 + 4901.763 + 8774.127 - 2\times 1642.598 - 2\times 4.815 - 2\times 741.048} \\ &= 123.416 \end{aligned}$$

EXERCISE 5.17

(a) The estimated linear regression from Exercise 5.16 is

$$\widehat{SAL1} = 22963 - 470.845 PR1 + 92.990 PR2 + 165.113 PR3 \qquad R^2 = 0.443$$
$$\text{(se)} \quad (9806) \quad (79.578) \quad (70.013) \quad (93.670)$$

A point estimate for expected sales when $PR1 = 90$, $PR2 = 75$ and $PR3 = 75$ is

$$\widehat{SAL1} = 22963.43 - 470.8447 \times (90) + 92.9900 \times (75) + 165.1129 \times (75) = -54.88$$

Using $t_c = t_{(0.975,48)} = 2.011$, a 95% interval estimate is given by

$$\widehat{SAL1} \pm t_c \,\text{se}(\widehat{SAL1}) = -54.88 \pm 2.011 \times 1385.523 = (-2841, 2731)$$

with $\text{se}(\widehat{SAL1}) = \text{se}\left(b_1 + 90b_2 + 75b_3 + 75b_4\right) = 1385.523$ found using computer software.

The interval estimate contains a wide range of negative values which are clearly infeasible. Sales cannot be negative. The values $PR1 = 90$, $PR2 = 75$ and $PR3 = 75$ are unfavorable ones for sales of brand 1, but they are nevertheless within the ranges of the sample data. Thus, the linear model is not a good one for forecasting.

(b) The estimated log-linear regression is

$$\widehat{\ln(SAL1)} = 10.45595 - 0.062176 PR1 + 0.014174 PR2 + 0.021472 PR3$$
$$\text{(se)} \quad (1.03046) \quad (0.008362) \quad (0.007357) \quad (0.009843)$$

A point estimate for expected log-sales when $PR1 = 90$, $PR2 = 75$ and $PR3 = 75$ is

$$\widehat{\ln(SAL1)} = 10.45595 - 0.062176 \times 90 + 0.014174 \times 75 + 0.021472 \times 75 = 7.53356$$

Using $t_c = t_{(0.975,48)} = 2.010635$, a 95% interval estimate for expected log-sales is given by

$$\widehat{\ln(SAL1)} \pm t_c \,\text{se}\left(\widehat{\ln(SAL1)}\right) = 7.53356 \pm 2.010635 \times 0.145589 = (7.24083, 7.82629)$$

Converting this interval into one for sales using the exponential function, we have

$$\left(\exp(7.24083), \exp(7.82629)\right) = (1395, 2506)$$

Comparing this interval with the one obtained from the linear function, we find that the two upper bounds of the intervals are of similar magnitude, but the lower bound for the interval from the log-linear model is positive and much larger than that from the linear model. Also, the width of the interval from the log-linear model is much narrower, suggesting more accurate estimation of expected sales.

(c) When *SAL1* is the dependent variable the coefficients show the change in number of cans sold from a 1-cent change in price. When $\ln(SAL1)$ is the dependent variable, by multiplying the coefficients by 100, we get the the percentage change in number of cans sold from a 1-cent change in price.

EXERCISE 5.18

The estimated regression is

$$\begin{aligned}
LCRMRTE = & -3.482 - 2.433PRBARR - 0.8077PRBCONV \qquad R^2 = 0.601 \\
\text{(se)} \quad & (0.351) \quad (0.320) \qquad\qquad (0.1110) \\
& + 0.3338PRBPRIS + 200.6POLPC + 0.002187WCON \\
& \quad (0.4700) \qquad\qquad (43.6) \qquad\quad (0.000834)
\end{aligned}$$

All five variables are expected to have negative effects on the crime rate. We expect each of them to act as a deterrent to crime. In the estimated equation the probability of an arrest and the probability of conviction have negative signs as expected, and both coefficients are significantly less than zero with *p*-values of 0.0000. On the other hand, the coefficients of the other three variables, the probability of a prison sentence, the number of police and the weekly wage in construction have positive signs, which is contrary to our expectations. Of these three variables, the coefficient of $PRBARR$ is not significantly different from zero, but the other two, $POLPC$ and $WCON$, are significantly different from zero, and have unexpected positive signs. Thus, it appears that the variables, $PRBARR$ and $PRBCONV$ are the most important for crime deterrence. The positive sign for the coefficient of $POLPC$ may have been caused by endogeneity, a concept considered in Chapter 10. In the context of this example, high crime rates may be more likely to exist in counties with greater numbers of police because more police are employed to counter high crime rates. It is less clear why *WCON* should have a positive sign. Perhaps construction companies have to pay higher wages to attract workers to counties with higher crime rates.

EXERCISE 5.19

(a) The estimated regression is:

$$\begin{array}{l} \ln(WAGE) = 1.1005 + 0.09031 EDUC + 0.005776 EXPER + 0.008941 HRSWK \\ \text{(se)} \quad (0.1095) \quad (0.00608) \qquad (0.001275) \qquad (0.001581) \end{array}$$

$$R^2 = 0.2197$$

The estimate $b_2 = 0.0903$ implies that holding other variables constant, an additional year of education increases wage by 9.03% on average.

The estimate $b_3 = 0.005776$ implies that holding other variables constant, an extra year of related work experience increases wage on average by 0.58%.

The estimate $b_4 = 0.008941$ implies that holding other variables constant, working an extra hour per week increases wage by 0.89% on average.

All coefficient estimates are significantly different from zero, with *p*-values of 0.0000.

(b) The null and alternative hypotheses are

$$H_0 : \beta_2 \ge 0.1 \qquad H_1 : \beta_2 < 0.1$$

The critical value for a 5% significance level is $t_{(0.05,996)} = -1.646$. We reject H_0 when $t = (b_2 - 0.1)/\text{se}(b_2) < -1.646$.

The value of the *t*-statistic is

$$t = \frac{b_2 - 0.1}{\text{se}(b_2)} = \frac{0.09031 - 0.1}{0.00608} = -1.595$$

The corresponding *p*-value is 0.0555. Since $-1.565 > -1.646$, we do not reject H_0. There is not sufficient evidence to show that the return to another year of education is less than 10%.

(c) A 90% confidence interval for $100\beta_4$ is given by

$$(100 \times b_4) \pm t_{(0.95,996)} \text{se}(100 \times b_4) = 0.8941 \pm 1.646 \times 0.1581 = (0.634, 1.154)$$

We estimate with 90% confidence that the wage return to working an extra hour per week lies between 0.63% and 1.15%.

(d) The estimates with quadratic terms and interaction term for *EDUC* and *EXPER* are given in the table on page 165.

The coefficient estimates for variables $EDUC^2$, *EXPER*, $EXPER^2$ and *HRSWK* are significantly different from zero at a 5% level of significance. That for $EDUC \times EXPER$ is significant at a 10% level. The coefficient of the remaining variable *EDUC* is not significantly different from zero.

Exercise 5.19(d) (continued)

Estimates of wage equation with quadratic and interaction terms included

Variable	Coefficient	Estimate	Std. Error	*t*-value	*p*-value
C	β_1	0.9266081	0.3404072	2.722	0.0066
$EDUC$	β_2	0.0490281	0.0366258	1.339	0.1810
$EDUC^2$	β_3	0.0023649	0.0011048	2.141	0.0325
$EXPER$	β_4	0.0527446	0.0097493	5.410	0.0000
$EXPER^2$	β_5	–0.0006287	0.0000888	–7.080	0.0000
$EDUC \times EXPER$	β_6	–0.0009238	0.0005054	–1.828	0.0679
$HRSWK$	β_7	0.0066930	0.0015681	4.268	0.0000

(e) Defining the coefficients as they appear in the above table, the marginal effects on $\ln(WAGE)$ are

$$\frac{\partial \ln(WAGE)}{\partial EDUC} = \beta_2 + 2\beta_3 EDUC + \beta_6 EXPER$$

$$\frac{\partial \ln(WAGE)}{\partial EXPER} = \beta_4 + 2\beta_5 EXPER + \beta_6 EDUC$$

(f) For Jill,

$$\begin{aligned}\frac{\partial \ln(WAGE)}{\partial EDUC} &= b_2 + 32b_3 + 10b_6 \\ &= 0.049028 + 32 \times 0.0023649 - 10 \times 0.0009238 = 0.115\end{aligned}$$

For Wendy,

$$\begin{aligned}\frac{\partial \ln(WAGE)}{\partial EDUC} &= b_2 + 24b_3 + 10b_6 \\ &= 0.049028 + 24 \times 0.0023649 - 10 \times 0.0009238 = 0.097\end{aligned}$$

We estimate that Jill has a greater marginal effect of education than Wendy. As education increases, the marginal effect of education increases. There are “increasing returns” to education.

(g) Jill’s marginal effect of education will be greater than that of Wendy if

$$\beta_2 + 32\beta_3 + 10\beta_6 > \beta_2 + 24\beta_3 + 10\beta_6$$

which will be true if and only if $32\beta_3 > 24\beta_3$. Now the inequality $32\beta_3 > 24\beta_3$ holds if $\beta_3 > 0$ and does not hold if $\beta_3 \le 0$. Thus a suitable test is $H_0 : \beta_3 \le 0$ against $H_1 : \beta_3 > 0$. From the above table, the *p*-value for this test is $0.0325/2 = 0.0163$. Thus, we reject H_0 and conclude that Jill’s marginal effect of education is greater than that of Wendy.

Exercise 5.19 (continued)

(h) For Chris,

$$\frac{\partial \ln(WAGE)}{\partial EXPER} = b_4 + 40b_5 + 16b_6$$
$$= 0.052745 - 40\times 0.0006287 - 16\times 0.0009238 = 0.0128$$

For Dave,

$$\frac{\partial \ln(WAGE)}{\partial EXPER} = b_4 + 60b_5 + 16b_6$$
$$= 0.052745 - 60\times 0.0006287 - 16\times 0.0009238 = 0.0002$$

We estimate that Chris has a greater marginal effect of exerience than Dave. As experience increases, the marginal effect of experience decreases. There are “decreasing returns” to experience.

(i) For someone with 16 years of education, the marginal effect of experience is

$$\frac{\partial \ln(WAGE)}{\partial EXPER} = \beta_4 + 2\beta_5 EXPER + 16\beta_6 .$$

Assuming $\beta_5 < 0$, the marginal effect of experience will be negative when

$$EXPER > \frac{-\beta_4 - 16\beta_6}{2\beta_5} = EXPER^*$$

A point estimate for $EXPER^*$ is

$$EXPER^* = \frac{-b_4 - 16b_6}{2b_5} = \frac{-0.0527446 + 16\times 0.0009238}{-2\times 0.0006287} = 30.19$$

The delta method is required to get the standard error

$$\text{se}\left(EXPER^*\right) = \text{se}\left(\frac{-b_4 - 16b_6}{2b_5}\right) = 1.5163$$

A 95% interval estimate is given by

$$EXPER^* \pm t_{(0.975,993)}\text{se}\left(EXPER^*\right) = 30.191 \pm 1.962\times 1.5163 = (27.22, 33.17)$$

We estimate with 95% confidence that the number of years of experience after which the marginal return to experience becomes negative is between 27.2 and 33.2 years.

EXERCISE 5.20

(a) $ADVERT_0 = 1.75$ will be optimal if $\beta_3 + 2\times 1.75\beta_4 = 1$. Thus the null and alternative hypotheses are $H_0: \beta_3 + 3.5\beta_4 = 1$ and $H_1: \beta_3 + 3.5\beta_4 \neq 1$. The *t*-value is

$$t = \frac{b_3 + 3.5b_4 - 1}{\text{se}\left(b_3 + 3.5b_4\right)} = \frac{12.1512 + 3.5\times(-2.76796) - 1}{0.68085} = 2.149$$

and the corresponding *p*-value is 0.0350. Thus we reject H_0 and conclude that $ADVERT_0 = 1.75$ is not optimal.

(b) $ADVERT_0 = 1.9$ will be optimal if $\beta_3 + 2\times 1.9\beta_4 = 1$. Thus the null and alternative hypotheses are $H_0: \beta_3 + 3.8\beta_4 = 1$ and $H_1: \beta_3 + 3.8\beta_4 \neq 1$. The *t*-value is

$$t = \frac{b_3 + 3.8b_4 - 1}{\text{se}\left(b_3 + 3.8b_4\right)} = \frac{12.1512 + 3.8\times(-2.76796) - 1}{0.65419} = 0.968$$

and the corresponding *p*-value is 0.3365. Thus we fail to reject H_0 and conclude that $ADVERT_0 = 1.9$ could be optimal.

(c) $ADVERT_0 = 2.3$ will be optimal if $\beta_3 + 2\times 2.3\beta_4 = 1$. Thus the null and alternative hypotheses are $H_0: \beta_3 + 4.6\beta_4 = 1$ and $H_1: \beta_3 + 4.6\beta_4 \neq 1$. The *t*-value is

$$t = \frac{b_3 + 4.6b_4 - 1}{\text{se}\left(b_3 + 4.6b_4\right)} = \frac{12.1512 + 4.6\times(-2.76796) - 1}{1.05435} = -1.500$$

and the corresponding *p*-value is 0.1381. Thus we fail to reject H_0 and conclude that $ADVERT_0 = 2.3$ could be optimal.

Note that we have found that both 1.9 and 2.3 could be optimal values for advertising expenditure. A null hypothesis that used any value for $ADVERT_0$ in between these two values would also not be rejected. This outcome illustrates why we never accept null hypotheses as the truth. The best we can do is to say there is insufficient evidence to conclude a null hypothesis is not true.

You might be surprised by the fact that 2.3 lies outside the 95% interval estimate for $ADVERT_0$ found on page 195 of the text. To appreciate how the difference can arise, note that for part (c) we could also have set up the hypothesis

$$H_0: ADVERT_0 = \frac{1-\beta_3}{2\beta_4} = 2.3$$

which is identical algebraically to $H_0: \beta_3 + 4.6\beta_4 = 1$. In this case the t value is

Exercise 5.20 (continued)

$$t = \frac{\left(\dfrac{1-b_3}{2b_4}\right) - 2.3}{\text{se}\left(\dfrac{1-b_3}{2b_4}\right)} = \frac{\left(\dfrac{1-12.1512}{2\times(-2.76796)}\right) - 2.3}{0.12872} = -2.219$$

The p-value is 0.0297, and H_0 is rejected. The different outcome arises because the delta method used to find $\text{se}\left((1-b_3)/2b_4\right)$ is a large sample approximation needed for nonlinear functions of the b's, whereas $\text{se}\left(b_3+4.6b_4\right)$ involves getting the standard error for a linear function of the b's, something we can do exactly without a large sample approximation.

EXERCISE 5.21

(a) The estimated equation is

$$\widehat{TIME} = 19.9166 + 0.36923DEPART + 1.3353REDS + 2.7548TRAINS$$
$$\text{(se)} \quad (1.2548)\ (0.01553) \qquad (0.1390) \qquad (0.3038)$$

Interpretations of each of the coefficients are:

β_1: The estimated time it takes Bill to get to work when he leaves Carnegie at 6:30AM and encounters no red lights and no trains is 19.92 minutes.

β_2: If Bill leaves later than 6:30AM, his traveling time increases by 3.7 minutes for every 10 minutes that his departure time is later than 6:30AM (assuming the number of red lights and trains are constant).

β_3: Each red light increases traveling time by 1.34 minutes.

β_4: Each train increases traveling time by 2.75 minutes.

(b) The 95% confidence intervals for the coefficients are:

$$\beta_1: \quad b_1 \pm t_{(0.975,227)}\text{se}(b_1) = 19.9166 \pm 1.970 \times 1.2548 = (17.44, 22.39)$$

$$\beta_2: \quad b_2 \pm t_{(0.975,227)}\text{se}(b_2) = 0.36923 \pm 1.970 \times 0.01553 = (0.339, 0.400)$$

$$\beta_3: \quad b_3 \pm t_{(0.975,227)}\text{se}(b_3) = 1.3353 \pm 1.970 \times 0.1390 = (1.06, 1.61)$$

$$\beta_4: \quad b_4 \pm t_{(0.975,227)}\text{se}(b_4) = 2.7548 \pm 1.970 \times 0.3038 = (2.16, 3.35)$$

In the context of driving time, these intervals are relatively narrow ones. We have obtained precise estimates of each of the coefficients.

(c) The hypotheses are $H_0: \beta_3 \geq 2$ and $H_1: \beta_3 < 2$. The critical value is $t_{(0.05,227)} = -1.652$. We reject H_0 when the calculated *t*-value is less than -1.652. This *t*-value is

$$t = \frac{1.3353 - 2}{0.1390} = -4.78$$

Since $-4.78 < -1.652$, we reject H_0. We conclude that the delay from each red light is less than 2 minutes.

(d) The hypotheses are $H_0: \beta_4 = 3$ and $H_1: \beta_4 \neq 3$. The critical values are $t_{(0.05,227)} = -1.652$ and $t_{(0.95,227)} = 1.652$. We reject H_0 when the calculated *t*-value is such that $t < -1.652$ or $t > 1.652$. This *t*-value is

$$t = \frac{2.7548 - 3}{0.3038} = -0.807$$

Since $-1.652 < -0.807 < 1.652$, we do not reject H_0. The data are consistent with the hypothesis that each train delays Bill by 3 minutes.

Exercise 5.21 (continued)

(e) Delaying the departure time by 30 minutes, increases travel time by $30\beta_2$. Thus, the null hypothesis is $H_0: 30\beta_2 \geq 10$, or $H_0: \beta_2 \geq 1/3$, and the alternative is $H_1: \beta_2 < 1/3$. We reject H_0 if $t \leq t_{(0.05,227)} = -1.652$, where the calculated t-value is

$$t = \frac{0.36923 - 0.33333}{0.01553} = 2.31$$

Since $2.31 > -1.652$, we do not reject H_0. The data are consistent with the hypothesis that delaying departure time by 30 minutes increases travel time by at least 10 minutes.

(f) If we assume that β_2, β_3 and β_4 are all non-negative, then the minimum time it takes Bill to travel to work is β_1. Thus, the hypotheses are $H_0: \beta_1 \leq 20$ and $H_1: \beta_1 > 20$. We reject H_0 if $t \geq t_{(0.95,227)} = 1.652$, where the calculated t-value is

$$t = \frac{19.9166 - 20}{1.2548} = -0.066$$

Since $-0.066 < 1.652$, we do not reject H_0. The data support the null hypothesis that the minimum travel time is less than or equal to 20 minutes. It was necessary to assume that β_2, β_3 and β_4 are all positive or zero, otherwise increasing one of the other variables will lower the travel time and the hypothesis would need to be framed in terms of more coefficients than β_1.

EXERCISE 5.22

The estimated equation is

$$\begin{array}{l} TIME = 19.9166 + 0.36923 DEPART + 1.3353 REDS + 2.7548 TRAINS \\ \text{(se)} \quad (1.2548)\ (0.01553) \qquad\qquad (0.1390) \qquad\qquad (0.3038) \end{array}$$

(a) The delay from a train is β_4 and the delay from a red light is β_3. Thus, the null and alternative hypotheses are

$$H_0: 3\beta_3 = \beta_4 \qquad \text{and} \qquad H_1: 3\beta_3 \neq \beta_4$$

The critical values for the *t*-test are $t_{(0.975,227)} = -1.970$ and $t_{(0.975,227)} = 1.970$. The rejection region is $t < -1.970$ or $t > 1.970$. The calculated value of the *t*-test statistic is

$$t = \frac{3b_3 - b_4}{\text{se}(3b_3 - b_4)} = \frac{3 \times 1.3353 - 2.7548}{0.5205} = 2.404$$

where the standard error is computed from

$$\begin{aligned} \text{se}(3b_3 - b_4) &= \sqrt{9 \times \text{var}(b_3) + \text{var}(b_4) - 2 \times 3 \times \text{cov}(b_2, b_3)} \\ &= \sqrt{9 \times 0.019311 + 0.092298 + 6 \times 0.00081} \\ &= 0.5205 \end{aligned}$$

The null hypothesis is rejected because $2.404 > 1.970$. The *p*-value is 0.017. The delay from a train is not equal to three times the delay from a red light.

(b) This test is similar to that in part (a), but it is a one-tail test rather than a two-tail test. The hypotheses are

$$H_0: \beta_4 \geq 3\beta_3 \qquad \text{and} \qquad H_1: \beta_4 < 3\beta_3$$

The rejection region for the *t*-test is $t < t_{(0.05,227)} = -1.652$, and the calculated *t*-value is

$$t = \frac{b_4 - 3b_3}{\text{se}(b_4 - 3b_3)} = \frac{2.7548 - 3 \times 1.3353}{0.5205} = -2.404$$

Since $-2.404 < -1.652$, we reject H_0. The delay from a train is less than three times the delay from a red light.

(c) The delay from 3 trains is $3\beta_4$. The extra time gained by leaving 5 minutes earlier is $5 + 5\beta_2$. Thus, the hypotheses are

$$H_0: 3\beta_4 \leq 5 + 5\beta_2 \qquad \text{and} \qquad H_0: 3\beta_4 > 5 + 5\beta_2$$

The rejection region for the *t*-test is $t > t_{(0.95,227)} = 1.652$, where the *t*-value is calculated as

$$t = \frac{3b_4 - 5b_2 - 5}{\text{se}(3b_4 - 5b_2)} = \frac{3 \times 2.7548 - 5 \times 0.36923 - 5}{0.9174} = 1.546$$

Exercise 5.22(c) (continued)

and the standard error is computed from

$$\begin{aligned}\text{se}(3b_4 - 5b_2) &= \sqrt{9\times\text{var}(b_4) + 25\times\text{var}(b_2) - 30\times\text{cov}(b_2, b_4)} \\ &= \sqrt{9\times 0.092298 + 25\times 0.000241 + 30\times 0.000165} \\ &= 0.9174\end{aligned}$$

Since $1.546 < 1.652$, we do not reject H_0 at a 5% significance level. Alternatively, we do not reject H_0 because the *p*-value = 0.0617, which is greater than 0.05. There is insufficient evidence to conclude that leaving 5 minutes earlier is not enough time.

(d) The expected time taken when the departure time is 7:15AM, and no red lights or trains are encountered, is $\beta_1 + 45\beta_2$. Thus, the null and alternative hypotheses are

$$H_0: \beta_1 + 45\beta_2 \le 45 \qquad \text{and} \qquad H_1: \beta_1 + 45\beta_2 > 45$$

The rejection region for the *t*-test is $t > t_{(0.95,227)} = 1.652$, where the *t*-value is calculated as

$$t = \frac{b_1 + 45b_2 - 45}{\text{se}(b_1 + 45b_2)} = \frac{19.9166 + 45\times 0.36923 - 45}{1.1377} = -7.44$$

and the standard error is computed from

$$\begin{aligned}\text{se}(b_1 + 45b_2) &= \sqrt{\text{var}(b_1) + 45^2\times\text{var}(b_2) + 90\times\text{cov}(b_1, b_2)} \\ &= \sqrt{1.574617 + 2025\times 0.00024121 - 90\times 0.00854061} \\ &= 1.1377\end{aligned}$$

Since $-7.44 < 1.652$, we do not reject H_0 at a 5% significance level. Alternatively, we do not reject H_0 because the *p*-value = 1.000, which is greater than 0.05. There is insufficient evidence to conclude that Bill will not get to the University before 8:00AM.

EXERCISE 5.23

The estimated model is

$$\widehat{SCORE} = -39.594 + 47.024 \times AGE - 20.222 \times AGE^2 + 2.749 \times AGE^3$$
$$\text{(se)} \quad (28.153) \quad (27.810) \qquad (8.901) \qquad (0.925)$$

The within sample predictions, with age expressed in terms of years (not units of 10 years) are graphed in the following figure. They are also given in a table on page 176.

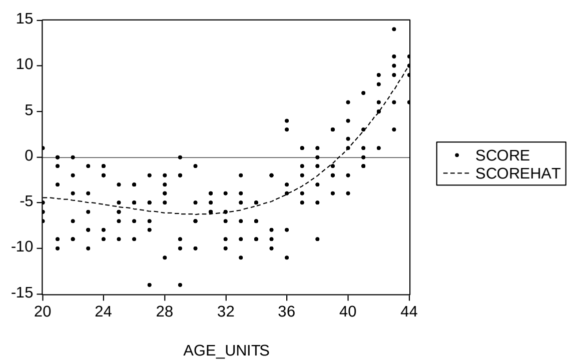

Figure xr5.23 Fitted line and observations

(a) To test the hypothesis that a quadratic function is adequate we test $H_0 : \beta_4 = 0$. The *t*-value is 2.972, with corresponding *p*-value 0.0035. We therefore reject H_0 and conclude that the quadratic function is not adequate. For suitable values of β_2, β_3 and β_4, the cubic function can decrease at an increasing rate, then go past a point of inflection after which it decreases at a decreasing rate, and then it can reach a minimum and increase. These are characteristics worth considering for a golfer. That is, the golfer improves at an increasing rate, then at a decreasing rate, and then declines in ability. These characteristics are displayed in Figure xr5.23.

(b) (i) Using the predictions in the table on page 176, we find the predicted score is lowest (−6.29) at the age of 30. Thus, we predict that Lion was at the peak of his career at age 30.

Mathematically, we can find the value for *AGE* at which $E(SCORE)$ is a minimum by considering the derivative

$$\frac{dE(SCORE)}{dAGE} = \beta_2 + 2\beta_3 AGE + 3\beta_4 AGE^2$$

Setting this derivative equal to zero and solving for age yields

$$AGE^* = \frac{-2\beta_3 \pm \sqrt{4\beta_3^2 - 12\beta_2\beta_4}}{6\beta_4}$$

Exercise 5.23(b)(i) (continued)

Replacing $\beta_2, \beta_3, \beta_4$ by their estimates b_2, b_3, b_4 gives the two solutions

$$AGE_1^* = \frac{-2\times(-20.2222)+\sqrt{4\times(-20.2222)^2-12\times 47.02386\times 2.74934}}{6\times 2.74934} = 3.008$$

$$AGE_2^* = \frac{-2\times(-20.2222)-\sqrt{4\times(-20.2222)^2-12\times 47.02386\times 2.74934}}{6\times 2.74934} = 1.895$$

The second derivative

$$\frac{d^2E(SCORE)}{dAGE^2} = 2b_3AGE + 6b_4AGE$$

is positive when $AGE = AGE_1^*$ and negative when $AGE = AGE_2^*$. Thus, the expected score $E(SCORE)$ is a minimum when $AGE = 3.008$, which is equivalent to 30.08 years.

(ii) Lion’s game is improving at an increasing rate between the ages of 20 and 25, where the differences between the predictions are increasing.

(iii) Lion’s game is improving at a decreasing rate between the ages of 25 and 30, where the differences between the predictions are declining.

We can consider (ii) and (iii) mathematically in the following way. When Lion’s game is improving the first derivative will be negative. It can be verified that the estimated first derivative will be negative for values of *AGE* between 2 and 3. If Lion’s game is improving at an increasing rate, the second derivative will also be negative; it will be positive when Lion’s game is improving at a decreasing rate. Thus, to find the age at which Lion’s improvement changes from an increasing rate to a decreasing rate we find that *AGE* for which the second derivative is zero, namely

$$AGE_3^* = \frac{-2b_3}{6b_4} = \frac{-2\times(-20.2222)}{6\times 2.74934} = 2.452$$

which is equivalent to 24.52 years.

(iv) At the age of 20, Lion’s predicted score is –4.4403. His predicted score then declines and rises again, reaching –4.1145 at age 36. Thus, our estimates suggest that, when he reaches the age of 36, Lion will play worse than he did at age 20.

(v) At the age of 40 Lion’s predicted score becomes positive implying that he can no longer score less than par.

(c) At the age of 70, the predicted score (relative to par) for Lion Forrest is 241.71. To break 100 it would need to be less than 28 $(=100-72)$. Thus, he will not be able to break 100 when he is 70.

Exercise 5.23 (continued)

Predicted scores at different ages

age	predicted scores
20	– 4.4403
21	– 4.5621
22	– 4.7420
23	– 4.9633
24	– 5.2097
25	– 5.4646
26	– 5.7116
27	– 5.9341
28	– 6.1157
29	– 6.2398
30	– 6.2900
31	– 6.2497
32	– 6.1025
33	– 5.8319
34	– 5.4213
35	– 4.8544
36	– 4.1145
37	– 3.1852
38	– 2.0500
39	– 0.6923
40	0.9042
41	2.7561
42	4.8799
43	7.2921
44	10.0092

EXERCISE 5.24

(a) The coefficient estimates, standard errors, t-values and p-values are in the following table.

Dependent Variable: ln(*PROD*)

	Coeff	Std. Error	t-value	p-value
C	-1.5468	0.2557	-6.0503	0.0000
ln(*AREA*)	0.3617	0.0640	5.6550	0.0000
ln(*LABOR*)	0.4328	0.0669	6.4718	0.0000
ln(*FERT*)	0.2095	0.0383	5.4750	0.0000

All estimates have elasticity interpretations. For example, a 1% increase in labor will lead to a 0.4328% increase in rice output. A 1% increase in fertilizer will lead to a 0.2095% increase in rice output. All p-values are less than 0.0001 implying all estimates are significantly different from zero at conventional significance levels.

(b) The null and alternative hypotheses are $H_0:\beta_2 = 0.5$ and $H_1:\beta_2 \neq 0.5$. The 1% critical values are $t_{(0.995,348)} = 2.59$ and $t_{(0.005,348)} = -2.59$. Thus, the rejection region is $t \geq 2.59$ or $t \leq -2.59$. The calculated value of the test statistic is

$$t = \frac{0.3617 - 0.5}{0.064} = -2.16$$

Since $-2.59 < -2.16 < 2.59$, we do not reject H_0. The data are compatible with the hypothesis that the elasticity of production with respect to land is 0.5.

(c) A 95% interval estimate of the elasticity of production with respect to fertilizer is given by

$$b_4 \pm t_{(0.975,348)} \times \text{se}(b_4) = 0.2095 \pm 1.967 \times 0.03826 = (0.134,\ 0.285)$$

This relatively narrow interval implies the fertilizer elasticity has been precisely measured.

(d) This hypothesis test is a test of $H_0:\beta_3 \leq 0.3$ against $H_1:\beta_3 > 0.3$. The rejection region is $t \geq t_{(0.95,348)} = 1.649$. The calculated value of the test statistic is

$$t = \frac{0.433 - 0.3}{0.067} = 1.99$$

We reject H_0 because $1.99 > 1.649$. There is evidence to conclude that the elasticity of production with respect to labor is greater than 0.3. Reversing the hypotheses and testing $H_0:\beta_3 \geq 0.3$ against $H_1:\beta_3 < 0.3$, leads to a rejection region of $t \leq -1.649$. The calculated t-value is $t = 1.99$. The null hypothesis is not rejected because $1.99 > -1.649$.

EXERCISE 5.25

(a) Taking logarithms yields the equation

$$\ln(Y) = \beta_1 + \beta_2 \ln(K) + \beta_3 \ln(L) + \beta_4 \ln(E) + \beta_5 \ln(M) + e$$

where $\beta_1 = \ln(\alpha)$. This form of the production function is linear in the coefficients β_1, β_2, β_3, β_4 and β_5, and hence is suitable for least squares estimation.

(b) Coefficient estimates and their standard errors are given in the following table.

	Estimated coefficient	Standard error
β_2	0.05607	0.25927
β_3	0.22631	0.44269
β_4	0.04358	0.38989
β_5	0.66962	0.36106

(c) The estimated coefficients show the proportional change in output that results from proportional changes in *K*, *L*, *E* and *M*. All these estimated coefficients have positive signs, and lie between zero and one, as is required for profit maximization to be realistic. Furthermore, they sum to approximately one, indicating that the production function has constant returns to scale. However, from a statistical point of view, all the estimated coefficients are not significantly different from zero; the large standard errors suggest the estimates are not reliable.