EXERCISE 5.5

(a) The estimated equation is

$$\widehat{VALUE} = 28.4067 - 0.1834CRIME - 22.8109NITOX + 6.3715ROOMS - 0.0478AGE$$
$$\text{(se)} \quad (5.3659) \quad (0.0365) \quad (4.1607) \quad (0.3924) \quad (0.0141)$$

$$-1.3353DIST + 0.2723ACCESS - 0.0126TAX - 1.1768PTRATIO$$
$$(0.2001) \quad (0.0723) \quad (0.0038) \quad (0.1394)$$

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 $\left|-1.6017\right| < 1.965$, 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 5.9

(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

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

A 95% interval estimate is given by

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

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

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

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

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

A 95% interval estimate is given by

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

Exercise 5.9(d) (continued)

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

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

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

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

A 95% interval estimate is given by

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

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

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

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

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

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

Then,

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

and

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

Substituting into this expression yields

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:

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

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 $H_0: \beta_2 \geq 0$ 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 $H_0: \beta_3 \leq 0$ 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

$$\widehat{PRICE} = -41948 + 90.970 SQFT - 755.04 AGE$$
$$\text{(se)} \quad (6990) \quad (2.403) \quad (140.89)$$

(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 $\dfrac{\widehat{\partial PRICE}}{\partial SQFT} = b_2 = 90.9698$

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:

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

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

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

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

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

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

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

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

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

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

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

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

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

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:

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

The standard error for $\widehat{\text{me}}$ can be found using software or from

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

(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:

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

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

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

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

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

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

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

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

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

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

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

Exercise 5.13(c)(ii) (continued)

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

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

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:

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

The standard error for $\widehat{\text{me}}$ was found using software.

(iv) The null and alternative hypotheses are

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

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 $t \le -1.646$. 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.