CHAPTER 2

Exercise Solutions

EXERCISE 2.1

(a)

x	y	$x-\bar{x}$	$(x-\bar{x})^2$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$
3	5	2	4	3	6
2	2	1	1	0	0
1	3	0	0	1	0
−1	2	−2	4	0	0
0	−2	−1	1	−4	4
$\sum x_i =$ 5	$\sum y_i =$ 10	$\sum(x_i-\bar{x}) =$ 0	$\sum(x_i-\bar{x})^2 =$ 10	$\sum(y-\bar{y}) =$ 0	$\sum(x-\bar{x})(y-\bar{y}) =$ 10

$\bar{x} = 1, \quad \bar{y} = 2$

(b) $b_2 = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sum(x-\bar{x})^2} = \frac{10}{10} = 1.$ b_2 is the estimated slope of the fitted line.

$b_1 = \bar{y} - b_2\bar{x} = 2 - 1\times 1 = 1.$ b_1 is the estimated value of *y* when *x* = 0, it is the estimated intercept of the fitted line.

(c)

$$\sum_{i=1}^{5} x_i^2 = 3^2 + 2^2 + 1^2 + (-1)^2 + 0^2 = 15$$

$$\sum_{i=1}^{5} x_i y_i = 3\times 5 + 2\times 2 + 1\times 3 + (-1)\times 2 + 0\times(-2) = 20$$

$$\sum_{i=1}^{5} x_i^2 - N\bar{x}^2 = 15 - 5\times 1^2 = 10 = \sum_{i=1}^{5}(x_i-\bar{x})^2$$

$$\sum_{i=1}^{5} x_i y_i - N\bar{x}\bar{y} = 20 - 5\times 1\times 2 = 10 = \sum_{i=1}^{5}(x_i-\bar{x})(y_i-\bar{y})$$

(d)

x_i	y_i	$\hat{y}_i$	$\hat{e}_i$	$\hat{e}_i^2$	$x_i\hat{e}_i$
3	5	4	1	1	3
2	2	3	−1	1	−2
1	3	2	1	1	1
−1	2	0	2	4	−2
0	−2	1	−3	9	0
$\sum x_i =$ 5	$\sum y_i =$ 10	$\sum \hat{y}_i =$ 10	$\sum \hat{e}_i =$ 0	$\sum \hat{e}_i^2 =$ 16	$\sum x_i\hat{e}_i =$ 0

(e) Refer to Figure xr2.1 below.

Exercise 2.1 (continued)

(f)

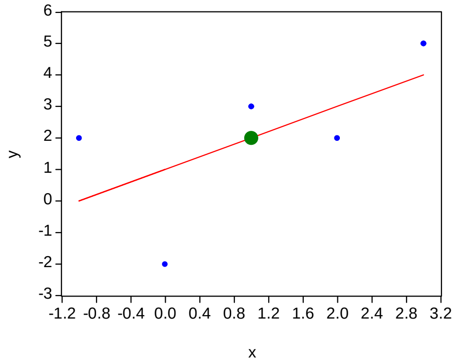

Figure xr2.1 Fitted line, mean and observations

(g) $\bar{y} = b_1 + b_2\bar{x}$ $\quad \bar{y} = 2, \bar{x} = 1, b_1 = 1, b_2 = 1$

Therefore: $2 = 1 + 1 \times 1$

(h) $\bar{\hat{y}} = \sum \hat{y}_i / N = (4 + 3 + 2 + 0 + 1)/5 = 2 = \bar{y}$

(i) $\hat{\sigma}^2 = \dfrac{\sum \hat{e}_i^2}{N-2} = \dfrac{16}{3} = 5.3333$

(j) $\widehat{\operatorname{var}(b_2)} = \dfrac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2} = \dfrac{5.3333}{10} = .53333$

EXERCISE 2.2

(a) Using equation (B.30),

$$P(110 < X < 140) = P\left(\frac{110-\mu_{y|x=\$1000}}{\sqrt{\sigma^2_{y|x=\$1000}}} < \frac{X-\mu_{y|x=\$1000}}{\sqrt{\sigma^2_{y|x=\$1000}}} < \frac{140-\mu_{y|x=\$1000}}{\sqrt{\sigma^2_{y|x=\$1000}}}\right)$$

$$= P\left(\frac{110-125}{\sqrt{49}} < Z < \frac{140-125}{\sqrt{49}}\right) = P(-2.1429 < Z < 2.1429) = 0.9679$$

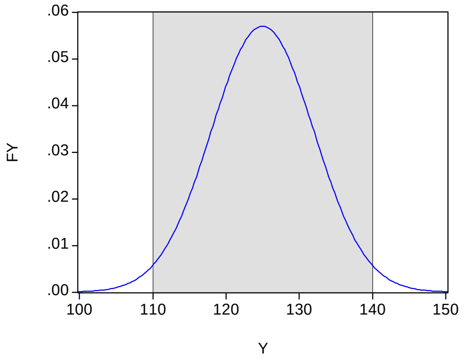

Figure xr2.2 Sketch of PDF

(b) Using the same formula as above:

$$P(110 < X < 140) = P\left(\frac{110-125}{\sqrt{81}} < Z < \frac{140-125}{\sqrt{81}}\right) = P(-1.6667 < Z < 1.6667) = 0.9044$$

EXERCISE 2.3

(a) The observations on y and x and the estimated least-squares line are graphed in part (b). The line drawn for part (a) will depend on each student's subjective choice about the position of the line. For this reason, it has been omitted.

(b) Preliminary calculations yield:

$$\sum x_i = 21 \qquad \sum y_i = 44 \qquad \sum (x_i - \bar{x})(y_i - \bar{y}) = 22 \qquad \sum (x_i - \bar{x})^2 = 17.5$$

$$\bar{y} = 7.3333 \qquad \bar{x} = 3.5$$

The least squares estimates are

$$b_2 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} = \frac{22}{17.5} = 1.257$$

$$b_1 = \bar{y} - b_2\bar{x} = 7.3333 - 1.2571 \times 3.5 = 2.9333$$

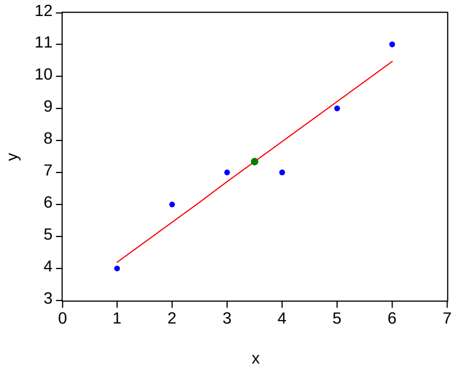

Figure xr2.3 Observations and fitted line

(c) $\bar{y} = \sum y_i / N = 44/6 = 7.3333$

$\bar{x} = \sum x_i / N = 21/6 = 3.5$

The predicted value for y at $x = \bar{x}$ is

$$\hat{y} = b_1 + b_2\bar{x} = 2.9333 + 1.2571 \times 3.5 = 7.3333$$

We observe that $\hat{y} = b_1 + b_2\bar{x} = \bar{y}$. That is, the predicted value at the sample mean $\bar{x}$ is the sample mean of the dependent variable $\bar{y}$. This implies that the least-squares estimated line passes through the point $(\bar{x}, \bar{y})$.

Exercise 2.3 (continued)

(d) The values of the least squares residuals, computed from $\hat{e}_i = y_i - b_1 - b_2 x_i$, are:

$$\hat{e}_1 = -0.19048 \qquad \hat{e}_2 = 0.55238 \qquad \hat{e}_3 = 0.29524$$

$$\hat{e}_4 = -0.96190 \qquad \hat{e}_5 = -0.21905 \qquad \hat{e}_6 = 0.52381$$

Their sum is $\sum \hat{e}_i = 0$.

(e) $\sum x_i \hat{e}_i = 1 \times -0.190 + 2 \times 0.552 + 3 \times 0.295 + 4 \times -0.962 + 5 \times -0.291 + 6 \times 0.524 = 0$

EXERCISE 2.4

(a) If $\beta_1 = 0,$ the simple linear regression model becomes

$$y_i = \beta_2 x_i + e_i$$

(b) Graphically, setting $\beta_1 = 0$ implies the mean of the simple linear regression model $E(y_i) = \beta_2 x_i$ passes through the origin (0, 0).

(c) To save on subscript notation we set $\beta_2 = \beta.$ The sum of squares function becomes

$$S(\beta) = \sum_{i=1}^{N}(y_i - \beta x_i)^2 = \sum_{i=1}^{N}(y_i^2 - 2\beta x_i y_i + \beta^2 x_i^2) = \sum y_i^2 - 2\beta \sum x_i y_i + \beta^2 \sum x_i^2$$

$$= 352 - 2\times 176\beta + 91\beta^2 = 352 - 352\beta + 91\beta^2$$

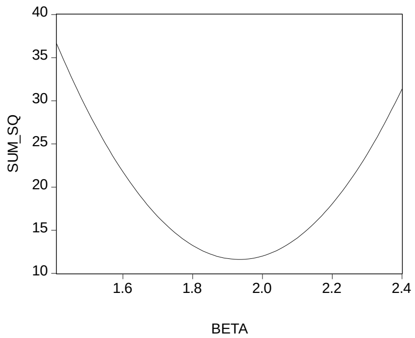

Figure xr2.4(a) Sum of squares for β_2

The minimum of this function is approximately 12 and occurs at approximately $\beta_2 = 1.95.$ The significance of this value is that it is the least-squares estimate.

(d) To find the value of β that minimizes $S(\beta)$ we obtain

$$\frac{dS}{d\beta} = -2\sum x_i y_i + 2\beta \sum x_i^2$$

Setting this derivative equal to zero, we have

$$b\sum x_i^2 = \sum x_i y_i \quad \text{or} \quad b = \frac{\sum x_i y_i}{\sum x_i^2}$$

Thus, the least-squares estimate is

$$b_2 = \frac{176}{91} = 1.9341$$

which agrees with the approximate value of 1.95 that we obtained geometrically.

Exercise 2.4 (Continued)

(e)

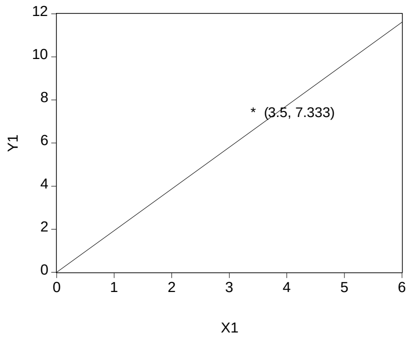

Figure xr2.4(b) Fitted regression line and mean

The fitted regression line is plotted in Figure xr2.4 (b). Note that the point $(\bar{x}, \bar{y})$ does not lie on the fitted line in this instance.

(f) The least squares residuals, obtained from $\hat{e}_i = y_i - b_2 x_i$ are:

$\hat{e}_1 = 2.0659$ $\quad \hat{e}_2 = 2.1319$ $\quad \hat{e}_3 = 1.1978$

$\hat{e}_4 = -0.7363$ $\quad \hat{e}_5 = -0.6703$ $\quad \hat{e}_6 = -0.6044$

Their sum is $\sum \hat{e}_i = 3.3846.$ Note this value is not equal to zero as it was for $\beta_1 \neq 0$.

(g) $$\sum x_i \hat{e}_i = 2.0659 \times 1 + 2.1319 \times 2 + 1.1978 \times 3$$
$$- 0.7363 \times 4 - 0.6703 \times 5 - 0.6044 \times 6 = 0$$

EXERCISE 2.5

(a) The consultant's report implies that the least squares estimates satisfy the following two equations

$$b_1 + 450b_2 = 7500$$

$$b_1 + 600b_2 = 8500$$

Solving these two equations yields

$$b_2 = \frac{1000}{150} = 6.6667 \qquad b_1 = 4500$$

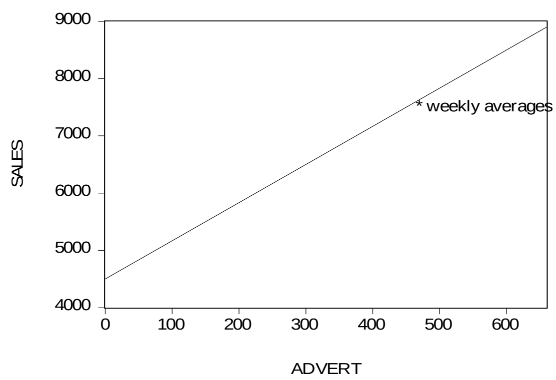

Figure xr2.5 Graph of sales-advertising regression line

EXERCISE 2.6

(a) The intercept estimate $b_1 = -240$ is an estimate of the number of sodas sold when the temperature is 0 degrees Fahrenheit. A common problem when interpreting the estimated intercept is that we often do not have any data points near $X = 0$. If we have no observations in the region where temperature is 0, then the estimated relationship may not be a good approximation to reality in that region. Clearly, it is impossible to sell –240 sodas and so this estimate should not be accepted as a sensible one.

The slope estimate $b_2 = 6$ is an estimate of the increase in sodas sold when temperature increases by 1 Fahrenheit degree. This estimate does make sense. One would expect the number of sodas sold to increase as temperature increases.

(b) If temperature is 80 degrees, the predicted number of sodas sold is

$$\hat{y} = -240 + 6 \times 80 = 240$$

(c) If no sodas are sold, $y = 0,$ and

$$0 = -240 + 6 \times x \quad \text{or} \quad x = 40$$

Thus, she predicts no sodas will be sold below 40°F.

(d) A graph of the estimated regression line:

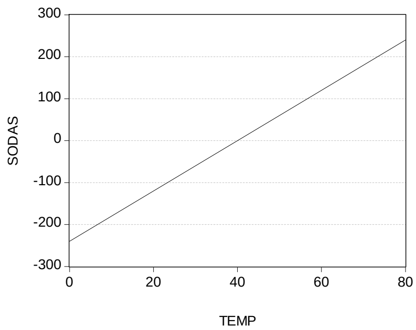

Figure xr2.6 Graph of regression line for soda sales and temperature

EXERCISE 2.7

(a) Since

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{N-2} = 2.04672$$

it follows that

$$\sum \hat{e}_i^2 = 2.04672(N-2) = 2.04672 \times 49 = 100.29$$

(b) The standard error for b_2 is

$$\text{se}(b_2) = \sqrt{\widehat{\text{var}(b_2)}} = \sqrt{0.00098} = 0.031305$$

Also,

$$\widehat{\text{var}(b_2)} = \frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}$$

Thus,

$$\sum \left(x_i - \bar{x}\right)^2 = \frac{\hat{\sigma}^2}{\widehat{\text{var}\left(b_2\right)}} = \frac{2.04672}{0.00098} = 2088.5$$

(c) The value $b_2 = 0.18$ suggests that a 1% increase in the percentage of males 18 years or older who are high school graduates will lead to an increase of \$180 in the mean income of males who are 18 years or older.

(d) $b_1 = \bar{y} - b_2\bar{x} = 15.187 - 0.18 \times 69.139 = 2.742$

(e) Since $\sum \left(x_i - \bar{x}\right)^2 = \sum x_i^2 - N\bar{x}^2$, we have

$$\sum x_i^2 = \sum \left(x_i - \bar{x}\right)^2 + N\bar{x}^2 = 2088.5 + 51 \times 69.139^2 = 245{,}879$$

(f) For Arkansas

$$\hat{e}_i = y_i - \hat{y}_i = y_i - b_1 - b_2 x_i = 12.274 - 2.742 - 0.18 \times 58.3 = -0.962$$

EXERCISE 2.8

(a) The EZ estimator can be written as

$$b_{EZ} = \frac{y_2 - y_1}{x_2 - x_1} = \left(\frac{1}{x_2 - x_1}\right) y_2 - \left(\frac{1}{x_2 - x_1}\right) y_1 = \sum k_i y_i$$

where

$$k_1 = \frac{-1}{x_2 - x_1}, \quad k_2 = \frac{1}{x_2 - x_1}, \quad \text{and} \quad k_3 = k_4 = \ldots = k_N = 0$$

Thus, b_{EZ} is a linear estimator.

(b) Taking expectations yields

$$E(b_{EZ}) = E\left[\frac{y_2 - y_1}{x_2 - x_1}\right] = \frac{1}{x_2 - x_1} E(y_2) - \frac{1}{x_2 - x_1} E(y_1)$$

$$= \frac{1}{x_2 - x_1}(\beta_1 + \beta_2 x_2) - \frac{1}{x_2 - x_1}(\beta_1 + \beta_2 x_1)$$

$$= \frac{\beta_2 x_2}{x_2 - x_1} - \frac{\beta_2 x_1}{x_2 - x_1} = \beta_2 \left(\frac{x_2}{x_2 - x_1} - \frac{x_1}{x_2 - x_1}\right) = \beta_2$$

Thus, b_{EZ} is an unbiased estimator.

(c) The variance is given by

$$\operatorname{var}(b_{EZ}) = \operatorname{var}(\sum k_i y_i) = \sum k_i^2 \operatorname{var}(e_i) = \sigma^2 \sum k_i^2$$

$$= \sigma^2 \left(\frac{1}{(x_2 - x_1)^2} + \frac{1}{(x_2 - x_1)^2}\right) = \frac{2\sigma^2}{(x_2 - x_1)^2}$$

(d) If $e_i \sim N(0, \sigma^2)$, then $b_{EZ} \sim N\left[\beta_2, \frac{2\sigma^2}{(x_2 - x_1)^2}\right]$

Exercise 2.8 (continued)

(e) To convince E.Z. Stuff that var(b_2) < var(b_{EZ}), we need to show that

$$\frac{2\sigma^2}{(x_2 - x_1)^2} > \frac{\sigma^2}{\sum(x_i - \bar{x})^2} \quad \text{or that} \quad \sum(x_i - \bar{x})^2 > \frac{(x_2 - x_1)^2}{2}$$

Consider

$$\frac{(x_2 - x_1)^2}{2} = \frac{\left[(x_2 - \bar{x}) - (x_1 - \bar{x})\right]^2}{2} = \frac{(x_2 - \bar{x})^2 + (x_1 - \bar{x})^2 - 2(x_2 - \bar{x})(x_1 - \bar{x})}{2}$$

Thus, we need to show that

$$2\sum_{i=1}^{N}(x_i - \bar{x})^2 > (x_2 - \bar{x})^2 + (x_1 - \bar{x})^2 - 2(x_2 - \bar{x})(x_1 - \bar{x})$$

or that

$$(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + 2(x_2 - \bar{x})(x_1 - \bar{x}) + 2\sum_{i=3}^{N}(x_i - \bar{x})^2 > 0$$

or that

$$\left[(x_1 - \bar{x}) + (x_2 - \bar{x})\right]^2 + 2\sum_{i=3}^{N}(x_i - \bar{x})^2 > 0.$$

This last inequality clearly holds. Thus, b_{EZ} is not as good as the least squares estimator.

Rather than prove the result directly, as we have done above, we could also refer Professor E.Z. Stuff to the Gauss Markov theorem.

EXERCISE 2.9

(a) Plots of $UNITCOST_t$ against $CUMPROD_t$ and $\ln(UNITCOST_t)$ against $\ln(CUMPROD_t)$ appear in Figure xr2.9(a) & (b). The two plots are quite similar in nature.

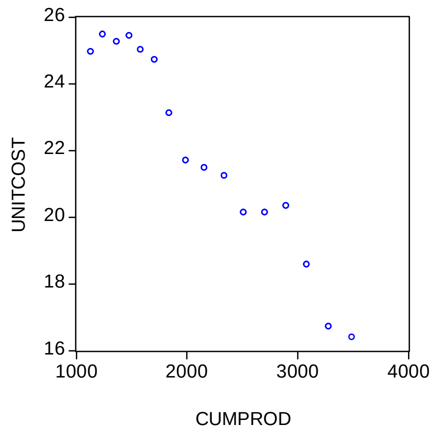

Figure xr2.9(a) The learning curve data

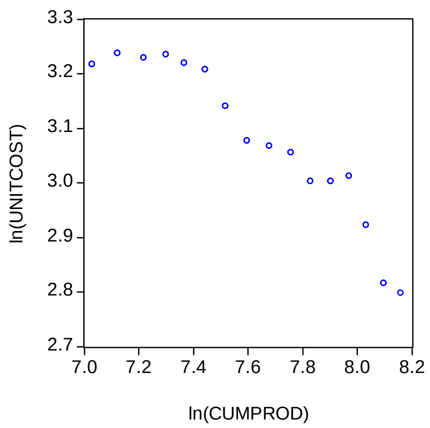

Figure xr2.9(b) Learning curve data with logs

Exercise 2.9 (continued)

(b) The least squares estimates are

$$b_1 = 6.0191 \qquad b_2 = -0.3857$$

Since $\ln(UNITCOST_1) = \beta_1$, an estimate of u_1 is

$$\widehat{UNITCOST_1} = \exp(b_1) = \exp(6.0191) = 411.208$$

This result suggests that 411.2 was the cost of producing the first unit. The estimate $b_2 = -0.3857$ suggests that a 1% increase in cumulative production will decrease costs by 0.386%. The numbers seem sensible.

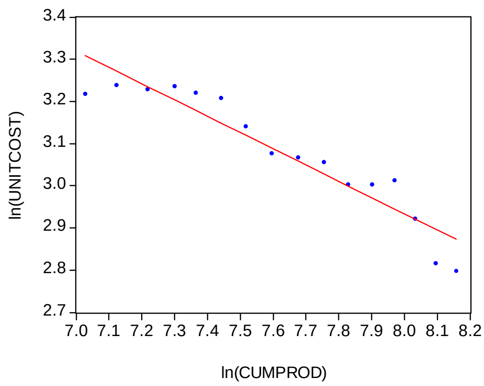

Figure xr2.9(c) Observations and fitted line

(c) The coefficient covariance matrix has the elements

$$\widehat{\text{var}(b_1)} = 0.075553 \qquad \widehat{\text{var}(b_2)} = 0.001297 \qquad \widehat{\text{cov}(b_1, b_2)} = -0.009888$$

(d) The error variance estimate is

$$\hat{\sigma}^2 = 0.049930^2 = 0.002493.$$

(e) When $CUMPROD_0 = 2000$, the predicted unit cost is

$$\widehat{UNITCOST_0} = \exp\left(6.01909 - 0.385696 \ln(2000)\right) = 21.921$$

EXERCISE 2.10

(a) The model is a simple regression model because it can be written as $y = \beta_1 + \beta_2 x + e$ where $y = r_j - r_f$, $x = r_m - r_f$, $\beta_1 = \alpha_j$ and $\beta_2 = \beta_j$.

(b)

Firm	Microsoft	General Electric	General Motors	IBM	Disney	Exxon-Mobil
$b_2 = \hat{\beta}_j$	1.430	0.983	1.074	1.268	0.959	0.403

The stocks Microsoft, General Motors and IBM are aggressive with Microsoft being the most aggressive with a beta value of $\hat{\beta}_2 = 1.430$. General Electric, Disney and Exxon-Mobil are defensive with Exxon-Mobil being the most defensive since it has a beta value of $\hat{\beta}_2 = 0.403$.

(c)

Firm	Microsoft	General Electric	General Motors	IBM	Disney	Exxon-Mobil
$b_1 = \hat{\alpha}_j$	0.010	0.006	-0.002	0.007	-0.001	0.007

All the estimates of $\hat{\alpha}_j$ are close to zero and are therefore consistent with finance theory. In the case of Microsoft, Figure xr2.10 illustrates how close the fitted line is to passing through the origin.

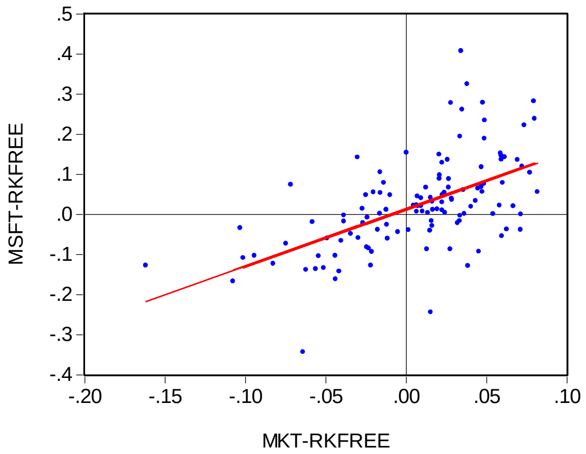

Figure xr2.10 Observations and fitted line for microsoft

Exercise 2.10 (continued)

(d) The estimates for β_j given $\alpha_j = 0$ are as follows.

Firm	Microsoft	General Electric	General Motors	IBM	Disney	Exxon-Mobil
$\hat{\beta}_j$	1.464	1.003	1.067	1.291	0.956	0.427

The restriction $\alpha_j = 0$ has led to only small changes in the $\hat{\beta}_j$.

EXERCISE 2.11

(a)

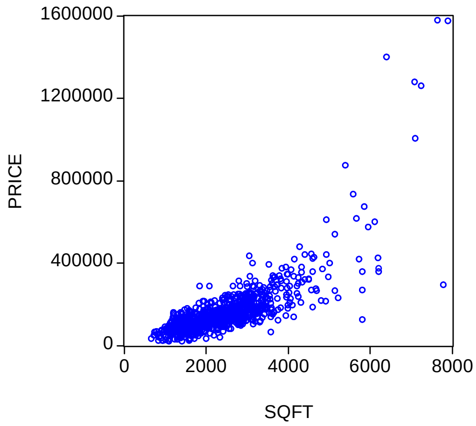

Figure xr2.11(a) Price against square feet – all houses

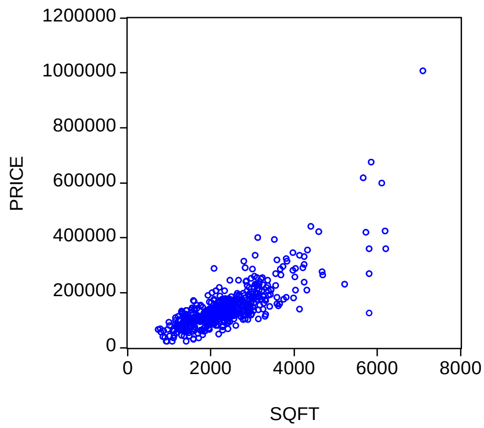

Figure xr2.11(b) Price against square feet for houses of traditional style

Exercise 2.11 (continued)

(b) The estimated equation for all houses is

$$\widehat{PRICE} = -60,861 + 92.747\, SQFT$$

The coefficient 92.747 suggests house price increases by approximately \$92.75 for each additional square foot of house size. The intercept, if taken literally, suggests a house with zero square feet would cost $-$\$60,861, a meaningless value. The model should not be accepted as a serious one in the region of zero square feet.

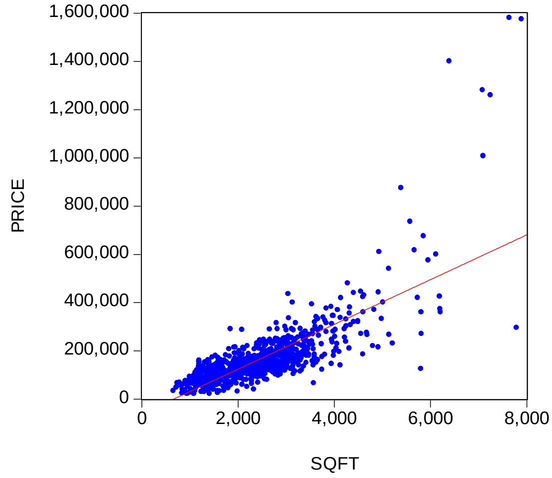

Figure xr2.11(c) Fitted line for Exercise 2.11(b)

(c) The estimated equation for traditional style houses is

$$\widehat{PRICE} = -28,408 + 73.772\, SQFT$$

The slope of 73.772 suggests that house price increases by approximately \$73.77 for each additional square foot of house size. The intercept term is $-$28,408 which would be interpreted as the dollar price of a traditional house of zero square feet. Once again, this estimate should not be accepted as a serious one. A negative value is meaningless and there is no data in the region of zero square feet.

Comparing the estimates to those in part (b), we see that extra square feet are not worth as much in traditional style houses as they are for houses in general (\$77.77 < \$92.75). A comparison of intercepts is not meaningful, but we can compare equations to see which type of house is more expensive. The prices are equal when

$$-28,408 + 73.772\, SQFT = -60,861 + 92.747\, SQFT$$

Solving for *SQFT* yields

$$SQFT = \frac{60861 - 28408}{92.747 - 73.772} = 1710$$

Exercise 2.11(c) (continued)

(c) Thus, we predict that the price of traditional style houses is greater than the price of houses in general when $SQFT < 1710$. Traditional style houses are cheaper when $SQFT > 1710$.

(d) Residual plots:

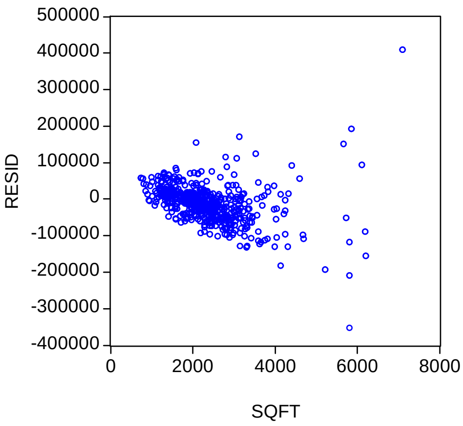

Figure xr2.11(d) Residuals against square feet – all houses

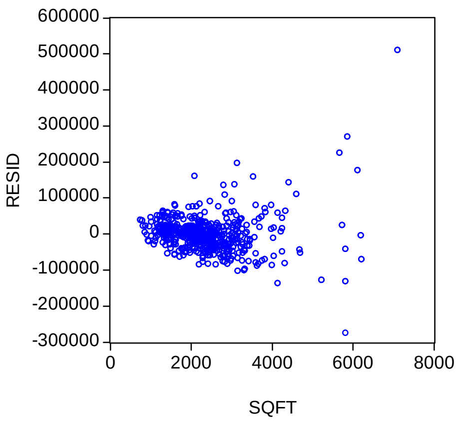

Figure xr2.11(e) Residuals against square feet for houses of traditional style

The magnitude of the residuals tends to increase as housing size increases suggesting that SR3 $\text{var}(e \mid x_i) = \sigma^2$ [homoskedasticity] could be violated. The larger residuals for larger houses imply the spread or variance of the errors is larger as *SQFT* increases. Or, in other words, there is not a constant variance of the error term for all house sizes.

EXERCISE 2.12

(a) We can see a positive relationship between price and house size.

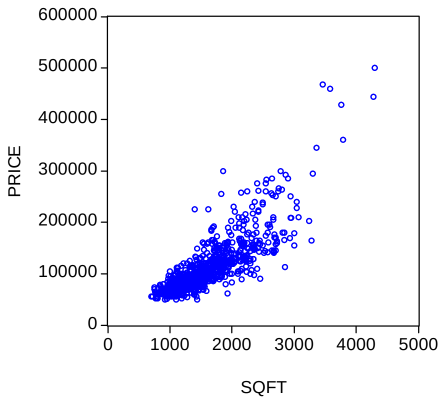

Figure xr2.12(a) Price against square feet

(b) The estimated equation for all houses in the sample is

$$\widehat{PRICE} = -18,386 + 81.389\, SQFT$$

The coefficient 81.389 suggests house price increases by approximately \$81 for each additional square foot in size. The intercept, if taken literally, suggests a house with zero square feet would cost $-$\$18,386, a meaningless value. The model should not be accepted as a serious one in the region of zero square feet.

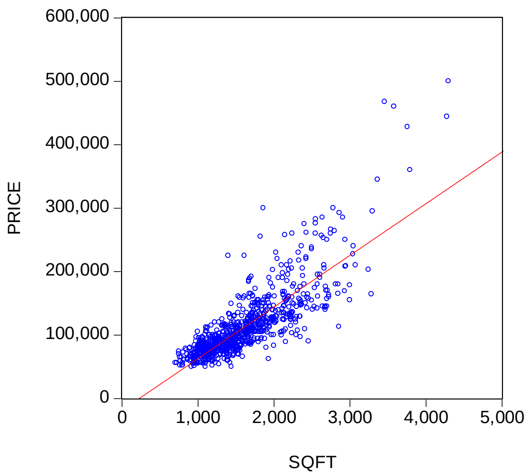

Figure xr2.12(b) Fitted regression line

Exercise 2.12 (continued)

(c) The estimated equation when a house is vacant at the time of sale is

$$\widehat{PRICE} = -4792.70 + 69.908SQFT$$

For houses that are occupied it is

$$\widehat{PRICE} = -27,169 + 89.259SQFT$$

These results suggest that price increases by \$69.91 for each additional square foot in size for vacant houses and by \$89.26 for each additional square foot of house size for houses that are occupied. Also, the two estimated lines will cross such that vacant houses will have a lower price than occupied houses when the house size is large, and occupied houses will be cheaper for small houses. To obtain the break-even size where prices are equal we write

$$-4792.70 + 69.908SQFT = -27,169 + 89.259SQFT$$

Solving for *SQFT* yields

$$SQFT = \frac{27169 - 4792.7}{89.259 - 69.908} = 1156$$

Thus, we estimate that occupied houses have a lower price per square foot when $SQFT < 1156$ and a higher price per square foot when $SQFT > 1156$.

(d) Residual plots

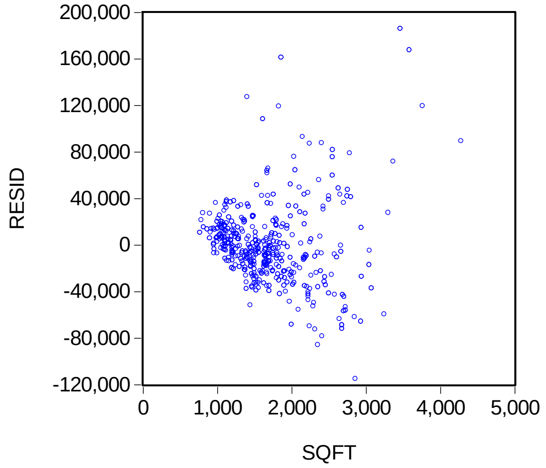

Figure xr2.12(c) Residuals against square feet for occupied houses

Exercise 2.12(d) (continued)

(d)

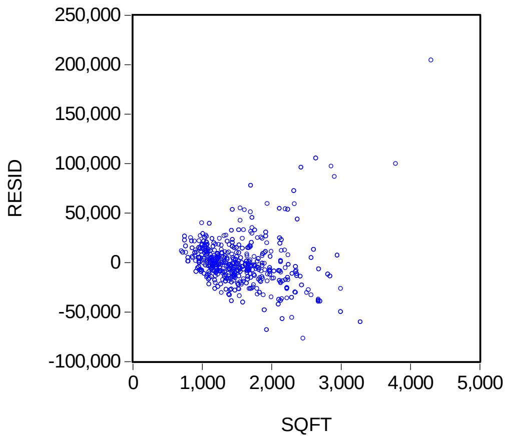

Figure xr2.12(d) Residuals against square feet for vacant houses

The magnitude of the residuals tends to be larger for larger-sized houses suggesting that SR3 $\text{var}(e \mid x_i) = \sigma^2$ [the homoskedasticity assumption of the model] could be violated. As the size of the house increases, the spread of distribution of residuals increases, indicating that there is not a constant variance of the error term with respect to house size.

(e) Using the model estimated in part (b), the predicted price when $SQFT = 2000$ is

$$\widehat{PRICE} = -18,386 + 81.389 \times 2000 = \$144,392$$

EXERCISE 2.13

(a)

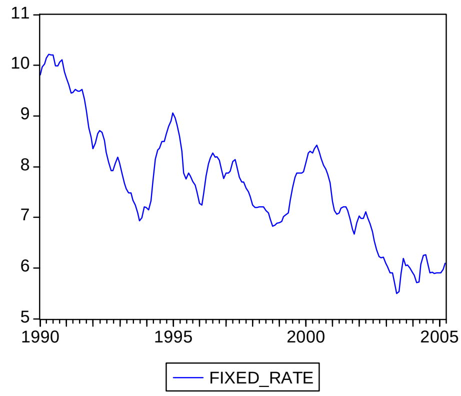

Figure xr2.13(a) Fixed rate against time

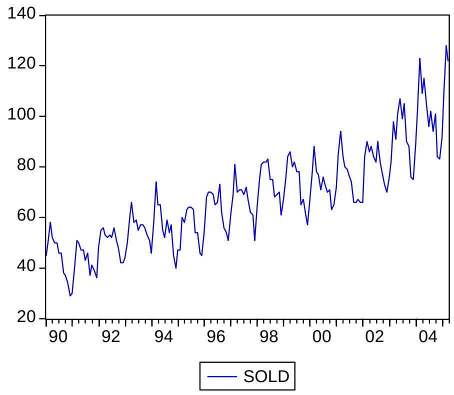

Figure xr2.13(b) Houses sold (1000's) against time

Exercise 2.13(a) (continued)

(a)

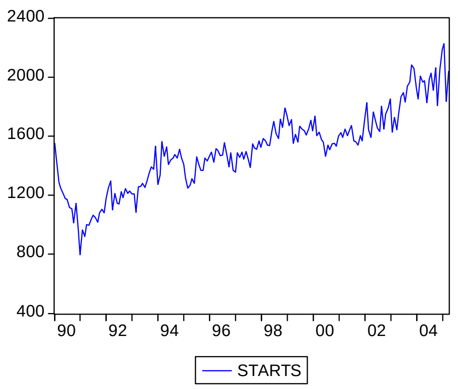

Figure xr2.13(c) New privately owned houses started against time

(b) Refer to Figure xr2.13(d).

(c) The estimated model is

$$\widehat{STARTS} = 2992.739 - 194.233 FIXED\_RATE$$

The coefficient −194.233 suggests that the number of new privately owned housing starts decreases by 194,233 for a 1% increase in the 30 year fixed interest rate for home loans. The intercept suggests that when the 30 year fixed interest rate is 0%, 2,992,739 will be started. Caution should be exercise with this interpretation, however, because it is beyond the range of the data.

Figure xr2.13(d) shows us where the fitted line lies among the data points. The fitted line appears to go evenly through the centre of data and the residuals appear be of relatively equal magnitude as we move along the fitted line.

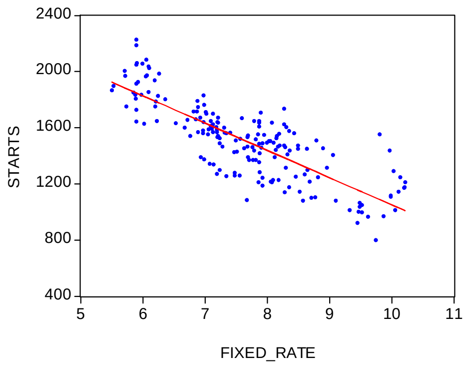

Figure xr2.13 (d) Fitted line and observations for housing starts against the fixed rate

Exercise 2.13 (continued)

(d) Refer to Figure xr2.13(e).

(e) The estimated model is

$$\widehat{SOLD} = 167.548 - 13.034 FIXED\_RATE$$

The coefficient −13.034 suggests that a 1% increase in the 30 year fixed interest rate for home loans is associated with a decrease of around 13,034 houses sold. The intercept suggests that when the 30 year fixed interest rate is 0%, 167,548 houses will be sold over a period of 1 month. Caution should be exercise with this interpretation, however, because it is beyond the range of the data.

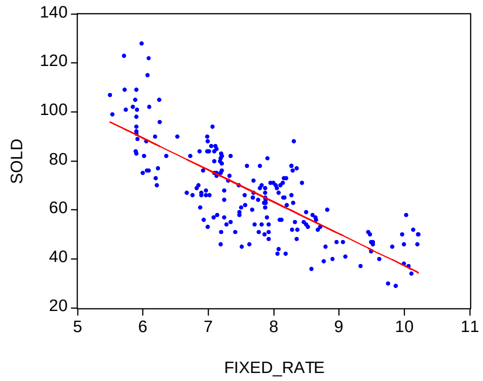

Figure xr2.13(e) Fitted line and observations for houses sold against fixed rate

Figure xr2.13(e) shows us where the fitted line lies amongst the data points. From this figure we can see that the data appear slightly convex relative to the fitted line suggesting that a different functional form might be suitable. A plot of the residuals against the fixed rate might shed more light oin this question. We can see also that the residuals appear to have a constant distribution over the majority of fixed rates.

(f) Using the model estimated in part (c), the predicted number of monthly housing starts for $FIXED\_RATE = 6$ is

$$\left(\widehat{STARTS}\right) \times 1000 = \left(2992.739 - 194.233 \times 6\right) \times 1000 = 1827.34 \times 1000 = 1{,}827{,}340$$

There will be 1,827,340 new privately owned houses started at a 30 year fixed interest rate of 6%. This is a seasonally adjusted annual rate. On a monthly basis we estimate 155,278 starts.

EXERCISE 2.14

(a)

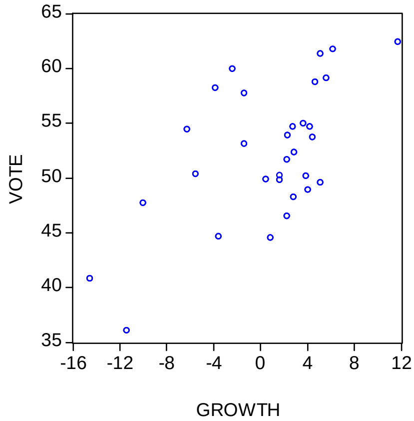

Figure xr2.14(a) Incumbent share against growth rate of real GDP per capita

There appears to be a positive association between *VOTE* and *GROWTH*.

(b) The estimated equation is

$$\widehat{VOTE} = 51.939 + 0.660 GROWTH$$

The coefficient 0.660 suggests that for an increase in 1% of the annual growth rate of GDP per capita, there is an associated increase in the share of votes of the incumbent party of 0.660.

The coefficient 51.939 indicates that the incumbent party receives 51.9% of the votes on average, when the growth rate in real GDP is zero. This suggests that when there is no real GDP growth, the incumbent party will still maintain the majority vote.

A graph of the fitted line and data is shown in Figure xr2.14(b).

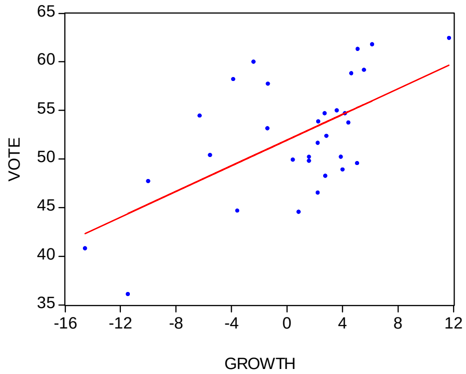

Figure xr2.14(b) Graph of vote-growth regression

Exercise 2.14 (continued)

(c) Figure xr2.14(c) shows a plot of *VOTE* against *INFLATION*. It shows a negative correlation between the two variables.

The estimated equation is:

$$\widehat{VOTE} = 53.496 - 0.445 INFLATION$$

The coefficient −0.445 indicates that a 1% increase in inflation, the GDP deflator, during the incumbent party's first 15 quarters, is associated with a 0.445 drop in the share of votes.

The coefficient 53.496 suggest that on average, when inflation is at 0% for that party's first 15 quarters, the associated share of votes won by the incumbent party is 53.496%; the incumbent party maintains the majority vote when inflation, during their first 15 quarters, is at 0%.

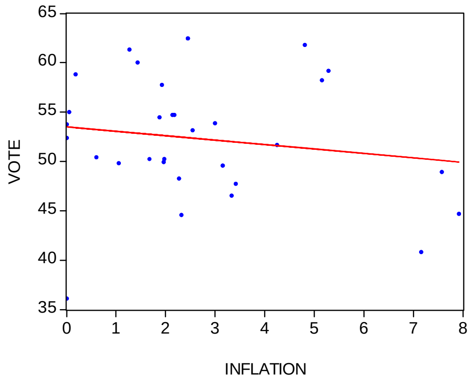

Figure xr2.14(c) Graph of vote-inflation regression line and observations

EXERCISE 2.15

(a)

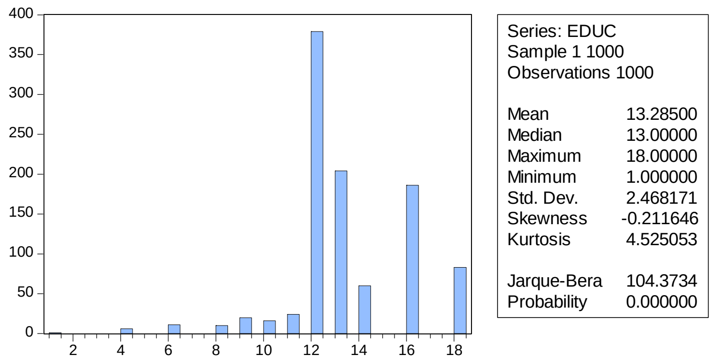

Figure xr2.15(a) Histogram and statistics for *EDUC*

From Figure xr2.15 we can see that the observations of *EDUC* are skewed to the left indicating that there are few observations with less than 12 years of education. Half of the sample has more than 13 years of education, with the average being 13.29 years of education. The maximum year of education received is 18 years, and the lowest level of education achieved is 1 year.

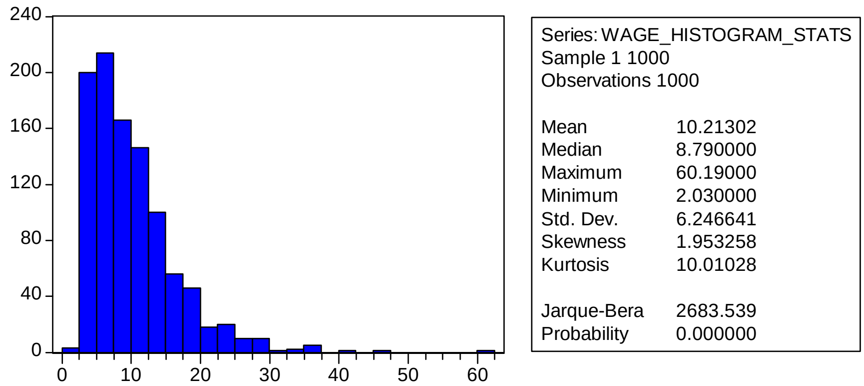

Figure xr2.15(b) Histogram and statistics for *WAGE*

Figure xr2.15(b) shows us that the observations for *WAGE* are skewed to the right indicating that most of the observations lie between the hourly wages of 5 to 20, and that there are few observations with an hourly wage greater than 20. Half of the sample earns an hourly wage of more than 8.79 dollars an hour, with the average being 10.21 dollars an hour. The maximum earned in this sample is 60.19 dollars an hour and the least earned in this sample is 2.03 dollars an hour.

Exercise 2.15 (continued)

(b) The estimated equation is

$$\widehat{WAGE} = -4.912 + 1.139 EDUC$$

The coefficient 1.139 represents the associated increase in the hourly wage rate for an extra year of education. The coefficient −4.912 represents the estimated wage rate of a worker with no years of education. It should not be considered meaningful as it is not possible to have a negative hourly wage rate. Also, as shown in the histogram, there are no data points at or close to the region *EDUC* = 0.

(c) The residuals are plotted against education in Figure xr2.15(c). There is a pattern evident; as *EDUC* increases, the magnitude of the residuals also increases. If the assumptions SR1-SR5 hold, there should not be any patterns evident in the least squares residuals.

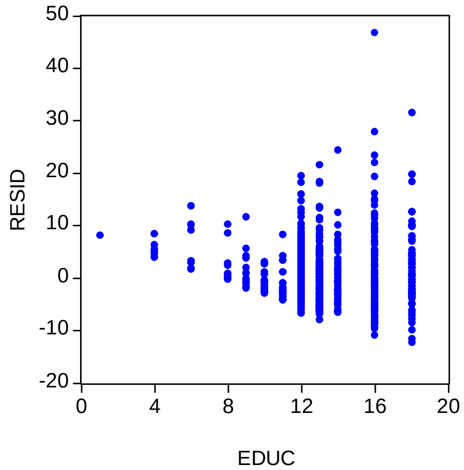

Figure xr2.15(c) Residuals against education

(d) The estimated regressions are

If female: $\widehat{WAGE} = -5.963 + 1.121 EDUC$

If male: $\widehat{WAGE} = -3.562 + 1.131 EDUC$

If black: $\widehat{WAGE} = 0.653 + 0.590 EDUC$

If white: $\widehat{WAGE} = -5.151 + 1.167 EDUC$

From these regression results we can see that the hourly wage of a white worker increases significantly more, per additional year of education, compared to that of a black worker. Similarly, the hourly wage of a male worker increases more per additional year of education than that of a female worker; although this difference is relatively small.