UNIVERSITY OF ESSEX
Session 2011–12

DEPARTMENT OF ECONOMICS
Professor M. J. Chambers

EC114 Introduction to Quantitative Economics

Problem Set 12: Sketch Solutions

Ordinary Least Squares Estimation

1. (a) The estimates a and b are given by the following equations:

$$b = \frac{\sum P_i Q_i - \dfrac{\sum P_i \sum Q_i}{n}}{\sum P_i^2 - \dfrac{(\sum P_i)^2}{n}}, \quad a = \bar{Q} - b\bar{P}.$$

Here, the sample size $n = 10$, and the other quantities are given in the question. We obtain

$$b = \frac{7825 - \dfrac{141 \times 583}{10}}{2045 - \dfrac{(141)^2}{10}} = \frac{-395.3}{56.9} = -6.9473,$$

while $a = (583/10) - (-6.9473)(141/10) = 156.2569$. Hence the sample regression line is

$$\hat{Q} = 156.2569 - 6.9473P.$$

(b) The price elasticity of demand is given by

$$\eta = \frac{dQ}{dP}\frac{P}{Q}.$$

We shall estimate dQ/dP by the slope coefficient estimate b and evaluate at $\bar{P}$, $\bar{Q}$:

$$\eta = b\frac{\bar{P}}{\bar{Q}} = -6.9473\frac{14.1}{58.3} = -1.6802.$$

Hence the demand for cod appears to be price-elastic i.e. the demand falls by more than 1% (in fact, by 1.68%) when price increases by 1% (at least at the sample means).

(c) We shall use the formula

$$R^2 = \frac{b^2 \sum p_i^2}{\sum q_i^2},$$

where $p_i = P_i - \bar{P}$ and $q_i = Q_i - \bar{Q}$. We have already computed $\sum p_i^2 = 56.9$ in the denominator of b. We also have

$$\sum q_i^2 = \sum Q_i^2 - \frac{(\sum Q_i)^2}{n} = 37,421 - \frac{(583)^2}{10} = 3432.1.$$

Hence we find that

$$R^2 = \frac{(-6.9473)^2 \times 56.9}{3432.1} = 0.8002.$$

Hence 80% of the variation in the demand for cod can be attributed to the price of cod.

2. (a) We shall use the formulae $b = \sum x_i y_i / \sum x_i^2$ and $a = \bar{Y} - b\bar{X}$. We obtain

$$b = \frac{-311}{964.92} = -0.3223, \quad a = 16 - (-0.3223)20.583 = 22.6339.$$

The sample regression line is therefore

$$\hat{Y} = 22.6339 - 0.3223X.$$

(b) Assuming that $X_{13} = 24$ we obtain

$$\hat{Y}_{13} = 22.6339 - (0.3223 \times 24) = 14.8987,$$

so the prediction is that 14.9% of the population will be below the poverty line in this region. The reliability of this prediction will depend on a number of factors, such as how well the sample regression equation fits the data for values around $X = 24$, and the extent to which other factors (in addition to the availability of medical care) affect poverty.

(c) The coefficient of determination is given by

$$R^2 = \frac{b^2 \sum x_i^2}{\sum y_i^2} = \frac{(-0.3223)^2 \times 964.92}{240} = 0.4176.$$

Thus only around 42% of the variation in poverty can be explained by the variations in the number of doctors. It is unclear whether the number of doctors has any significant effect on poverty levels.

3. *(Optional extension question)*

(a) Divide both sides of each equation by -2 and re-arrange to get:

$$\begin{aligned} \sum Y_i &= na + b\sum X_i, \\ \sum X_i Y_i &= a\sum X_i + b\sum X_i^2. \end{aligned}$$

From the first of these equations we obtain

$$na = \sum Y_i - b\sum X_i \quad \Rightarrow \quad a = \frac{\sum Y_i}{n} - b\frac{\sum X_i}{n}$$

or $a = \bar{Y} - b\bar{X}$ as required. Now substitute this expression for a into the second equation:

$$\sum X_i Y_i = \left(\frac{\sum Y_i}{n} - b\frac{\sum X_i}{n}\right)\sum X_i + b\sum X_i^2,$$

which can be re-written as

$$\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n} = b\left(\sum X_i^2 - \frac{(\sum X_i)^2}{n}\right).$$

Solving for b yields

$$b = \frac{\sum X_i Y_i - \dfrac{\sum X_i \sum Y_i}{n}}{\sum X_i^2 - \dfrac{(\sum X_i)^2}{n}}$$

as required. Finally, noting that

$$\sum x_i^2 = \sum X_i^2 - \frac{(\sum X_i)^2}{n}, \quad \sum x_i y_i = \sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}$$

results straightforwardly in $b = \sum x_i y_i / \sum x_i^2$.

(b) We shall begin by expanding SST by noting that $Y_i = \hat{Y}_i + e_i$:

$$\begin{aligned} \text{SST} &= \sum (Y_i - \bar{Y})^2 \\ &= \sum (\hat{Y}_i - \bar{Y} + e_i)^2 \\ &= \sum \left((\hat{Y}_i - \bar{Y})^2 + 2e_i(\hat{Y}_i - \bar{Y}) + e_i^2 \right) \\ &= \text{SSE} + 2\sum e_i(\hat{Y}_i - \bar{Y}) + \text{SSR}. \end{aligned}$$

We therefore need to show that $\sum e_i(\hat{Y}_i - \bar{Y}) = 0$. Recall that $\hat{Y}_i = a + bX_i$. Making this substitution we obtain

$$\begin{aligned} \sum e_i(\hat{Y}_i - \bar{Y}) &= \sum e_i(a + bX_i - \bar{Y}) \\ &= a\sum e_i + b\sum X_i e_i - \bar{Y}\sum e_i. \end{aligned}$$

But, from the normal equations (first-order conditions for OLS), we know that $\sum e_i = 0$ and $\sum X_i e_i = 0$. Hence SST=SSE+SSR as required.

(c) We begin with the definition

$$R^2 = \frac{\text{SSE}}{\text{SST}}.$$

From (b) we know that SST=SSE+SSR implying that SSE=SST−SSR. Hence

$$R^2 = \frac{\text{SST}-\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSR}}{\text{SST}}.$$

For the second expression it is clear that the denominator is equal to SST, and so it remains to be shown that SSE=$b^2 \sum x_i y_i$. To do this we can expand SSE as follows:

$$\text{SSE} = \sum (\hat{Y}_i - \bar{Y})^2 = \sum (a + bX_i - \bar{Y})^2.$$

But $a = \bar{Y} - b\bar{X}$ and so

$$\text{SSE} = \sum (\bar{Y} - b\bar{X} + bX_i - \bar{Y})^2 = \sum (bX_i - b\bar{X})^2 = b^2 \sum x_i^2$$

as required.