Module 8: Linear least squares regression practice problems

Ordinary least squares estimators from intermediate values

(The attached PDF file has better formatting.)

** Exercise 1.1: Ordinary Least Squares Estimates

We have 4 pairs of points

i	X_i	Y_i
1	0.5	1.5
2	1.0	2.5
3	1.5	3.5
4	2.0	5.5

Note that $\sum X_i$ = 5.0, $\sum Y_i$ = 13.0, $\sum X_i Y_i$ = 19.5, and $\sum X_i^2$ = 7.5

A. What is $\sum(x_i - \overline{x})^2$?
B. What is $\sum(x_i - \overline{x})(y_i - \overline{y})$?
C. What is B, the ordinary least squares estimate of β?
D. What is A, the ordinary least squares estimate of α?

Part A: Derive $\sum x_i^2$ from $\sum X_i$ and $\sum X_i^2$ as

$$\sum x_i^2 = \sum X_i^2 - (\sum X_i)^2 / N = 7.5 - 5.0^2 / 4 = 7.5 - 25/4 = 1.250$$

Part B: Derive $\sum x_i y_i$ from $\sum X_i$, $\sum Y_i$, and $\sum X_i Y_i$ .as

$$\sum x_i y_i = \sum X_i Y_i - \sum X_i \times \sum Y_i / N = 19.5 - 5 \times 13 / 4 = 3.250$$

Part C: $B = \dfrac{\sum x_i y_i}{\sum x_i^2}$ = 3.250 / 1.250 = 2.600

Alternatively, we use the formula involving the expressions above**:**

$$B = \frac{\sum X_i Y_i - N\overline{X}\overline{Y}}{\sum X_i^2 - N\overline{X}^2} = \frac{\sum X_i Y_i - \dfrac{\sum X_i \sum Y_i}{N}}{\sum X_i^2 - \dfrac{\left(\sum X_i\right)^2}{N}}$$

= [19.5 – 4 × 5/4 × 13/4] / (7.5 – 4 × (5/4)^2] = 2.600

Part D: B = 2.600, so A = $\overline{Y}$ – B× $\overline{X}$ = ¼ × 13 – 2.6 × ¼ × 5 = 0.000.

Jacob: Could we also solve this problem by regressing the four y values on the four x values?

Rachel: This exercise is deliberately simple, so you can verity the computation using the intermediate values with a regression analysis (using Excel or R or any statistical package) of the four observations. Final exam problems may give just the intermediate values (not the raw observations), with an N or 100 or 200.