Econometrics

Homework 4 : Basic Regression

1. Verify our line formula minimizes the sum of residuals squared.

→ We will use calculus to derive the formula which minimizes the squared residuals & then match that formula with our line formula.

Our estimate of Y_i is

$$\hat{Y}_i = \beta_0 + \beta_1 X_i \quad , \quad i = 1 \dots N$$

Sum of squared residuals $= \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2$

$$S = \sum_{i=1}^{N} (Y_i - \beta_0 - \beta_1 X_i)^2$$

we want to minimize it, so we substitute

$\frac{\partial S}{\partial \beta_0} = 0$ & $\frac{\partial S}{\partial \beta_1} = 0$

$$\frac{\partial S}{\partial \beta_0} = \sum_{i=1}^{N} 2\,(Y_i - \beta_0 - \beta_1 X_i) \times -1 = 0$$

$$\therefore \sum_{i=1}^{N} Y_i - N.\beta_0 - \beta_1 \sum_{i=1}^{N} X_i = 0$$

$$\therefore \frac{1}{N}\sum_{i=1}^{N} Y_i - \beta_0 - \beta_1 \frac{1}{N}\sum_{i=1}^{N} x_i = 0$$

$$\therefore \bar{Y} - \beta_0 - \beta_1 \bar{X} = 0$$

$$\therefore \bar{Y} = \beta_0 + \beta_1 \bar{X} \quad \text{— (1)}$$

Similarly,

$$\frac{\partial S}{\partial \beta_1} = \sum_{i=1}^{N} 2(Y_i - \beta_0 - \beta_1 X_i) \times -X_i = 0$$

$$\therefore \sum_{i=1}^{N} X_i Y_i - \beta_0 \sum X_i - \beta_1 \sum_{i=1}^{N} X_i^2 = 0$$

$$\therefore \frac{1}{N}\sum_{i=1}^{N} X_i Y_i - \beta_0 \frac{1}{N}\sum X_i - \beta_1 \frac{1}{N}\sum_{i=1}^{N} X_i^2 = 0$$

$$\therefore \bar{X}.\bar{Y} + \text{cov}(X,Y) - \beta_0 \bar{X} - \beta_1 \overline{X^2} = 0 \quad \text{— (2)}$$

from eq^n (1) $\beta_0 = \bar{Y} - \beta_1 \bar{X}$

$$\bar{X}.\bar{Y} + \text{cov}(X,Y) - \bar{X}\bar{Y} + \beta_1 (\bar{X})^2 - \beta_1 \overline{X^2} = 0$$

$$\therefore \beta_1 = \frac{\text{cov}(X,Y)}{\overline{X^2} - (\bar{X})^2} = \frac{\text{cov}(X,Y)}{\text{Var}(X)}$$

$$\therefore \beta_0 = \bar{Y} - \left(\frac{\text{cov}(X,Y)}{\text{Var}(X)}\right)\bar{X}$$

We can also write $\beta_1 = \dfrac{\frac{1}{N}\sum_{i=1}^{N} X_i Y_i - \left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)\left(\frac{1}{N}\sum_{i=1}^{N} Y_i\right)}{\frac{1}{N}\sum_{i=1}^{N} X_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)^2}$

2. By the methods in problem (1) if we want to force our line through the origin (0,0) how would we calculate the slope?

→ If we want to force our line through Origin we have to equate Y-intercept i.e. β_0 to 0 $\quad (\hat{Y}_i = \beta_0 + \beta_1 x_i)$

As $\beta_0 = \bar{Y} - \left(\frac{cov(X,Y)}{Var(X)}\right)\bar{X} = 0$

$\therefore \frac{cov(X,Y)}{Var(X)}\bar{X} = \bar{Y}$

$\therefore \beta_1 \bar{X} = \bar{Y}$

$\therefore \beta_1 = \frac{\bar{Y}}{\bar{X}}$

Which is the slope for our line formula. We will calculate it from the data.