

Professor: Héctor Bahamonde.

e: hibano@utu.fi

w: www.hectorbahamonde.com

Course: OLS.

The “Mechanic” behind OLS Let’s think about the relationship *schooling* and *earnings*, *controlling* for *experience*. What does it mean “to control for” something in this context?

Now, let’s suppose we have the following data,

Name (i)	Earnings (Y)	Education (x1)	Experience (x2)
Alfred	3	2	2
Brandon	5	7	4
Charly	7	3	6

Hypothesis: “The more education, the higher earnings,” for the average level of experience (And what do I mean by “for the average level of experience” and *Why does it matter?*)

I. BY HOW MUCH DO MY EARNINGS RISE IF MY SCHOOLING GOES UP?

The linear model is given by the next formula,

$$\begin{aligned} \text{Earnings}_i &= \beta_0 + \beta_1 \text{Education}_i + \beta_2 \text{Experience}_i + e_i \\ Y_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i \end{aligned}$$

- **What we observe:** x and y .
- **What we don’t observe, but should estimate:** β and ϵ .

Let’s revisit the formula, but this time in matrix form:

Let’s define what we know:

$$Y = \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}$$

$$X = \begin{bmatrix} 2 & 2 \\ 7 & 4 \\ 3 & 6 \end{bmatrix}$$

...and see how OLS looks like but in matrix form:

$$\begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}_y = \beta_0 + \beta_1 \begin{bmatrix} 2 \\ 7 \\ 3 \end{bmatrix}_{x1} + \beta_2 \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}_{x2} + e_i$$

It's easy to see that:

- We should multiply β_1 times x_1 and β_2 times x_2 .
- β_0 , β_1 , β_2 and ϵ_i unknown quantities. Hence, we should infer/estimate that (this is *inferential* statistics!).
- β_0 , β_1 and β_2 are scalars (single numbers and constants), where the vector containing all estimations is defined as, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}_\beta$
- The only parameter that's indexed (i.e. one row per observation or "individual," hence the tiny "i") is ϵ_i . We will address this in another class. But basically, it's the "error" or "residual."

Let's think about the case of "Alfred." If $\beta_0 = -3$, $\beta_1 = 1$ and $\beta_2 = 2$, we have that $y_{\text{Alfred}} - 3 + 1(2) + 2(2) = 3$. Thus:

$$y_{\text{Alfred}} = 3 = -3 + 1(2) + 2(2) + 0$$

Here, $\epsilon_{\text{Alfred}}=0$. In this sense, e_i is just the difference between what we observe and what estimate in the model. "Philosophically" it means more than that, but we'll talk about this soon.

Let's continue...

Deriving β (quantitative effect of X on Y):

$$\beta = \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T y$$

Let's do this by hand...:

$$y = \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 2 \\ 1 & 7 & 4 \\ 1 & 3 & 6 \end{bmatrix}$$

$$\mathbf{X}^T = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 7 & 3 \\ 2 & 4 & 6 \end{bmatrix}$$

$$\mathbf{X}^T \times \mathbf{X} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 7 & 3 \\ 2 & 4 & 6 \end{bmatrix}_{\mathbf{X}^T} \times \begin{bmatrix} 1 & 2 & 2 \\ 1 & 7 & 4 \\ 1 & 3 & 6 \end{bmatrix}_{\mathbf{X}} = \begin{bmatrix} 3 & 12 & 12 \\ 12 & 62 & 50 \\ 12 & 50 & 56 \end{bmatrix}$$

$$\left(\mathbf{X}^T \times \mathbf{X}\right)^{-1} = \frac{1}{\left(\mathbf{X}^T \mathbf{X}\right)} = \frac{1}{\det\left(\mathbf{X}^T \mathbf{X}\right)} \times \text{Adj}\left(\mathbf{X}^T \mathbf{X}\right) = \begin{bmatrix} 3 & -0.22 & -0.44 \\ -0.22 & 0.074 & -0.0185 \\ -0.44 & -0.0185 & 0.129 \end{bmatrix}$$

$$\boldsymbol{\beta} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T y = \begin{bmatrix} 3 & -0.22 & -0.44 \\ -0.22 & 0.074 & -0.0185 \\ -0.44 & -0.0185 & 0.129 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 1 \\ 2 & 7 & 3 \\ 2 & 4 & 6 \end{bmatrix} \times \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}_{\boldsymbol{\beta}}$$

This means that $\beta_0 = 1$, $\beta_1 = 0$ and that $\beta_2 = 1$. Let's re-write our formula:

$$\boldsymbol{\beta} = \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{y} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}_{\boldsymbol{\beta}}$$

The formula we had before: $\begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}_y = \beta_0 + \beta_1 \begin{bmatrix} 2 \\ 7 \\ 3 \end{bmatrix}_{x_1} + \beta_2 \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}_{x_2} + e_i$

Results we have now: $\begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}_y = 1 + 0 \begin{bmatrix} 2 \\ 7 \\ 3 \end{bmatrix}_{x_1} + 1 \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}_{x_2} + e_i$