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,

$$\text{Earnings}_i = \beta 0 + \beta_1 \text{Education}_i \ + \ \beta_2 \text{Experience}_i + e_i$$

- **What we observe:** $\boldsymbol{x}$ and y.
- **What we don't observe, but should estimate**: $\boldsymbol{\beta}$ 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 $\beta 1$ times $x1$ and $\beta 2$ times $x2$.
- $\beta 0$, $\beta 1$, $\beta 2$ and ϵ_i unknown quantities. Hence, we should infer/estimate that (this is *inferential* statistics!).
- $\beta 0$, $\beta 1$ and $\beta 2$ are scalars (single numbers and constants), where the vector containing all estimations is defined as, $\boldsymbol{\beta} \;=\; \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}_{\boldsymbol{\beta}}$
- 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 "Aldred." If,

  $\beta 0 \;=\; -3$,

  $\beta 1 \;=\; 1$, and

  $\beta 2 \;=\; 2$, then,

  we have that,

  $\hat{y}_{\text{Alfred}} - 3 \;+\; 1(2) \;+\; 2(2) = 3$. Then if $y \;=\; 3$ and $\hat{y} \;=\; 3$, **by how much did I miss my prediction?** Thus:

$$\hat{y}_{\text{Alfred}} \;=\; 3 \;=\; -3 \;+1(2) \;+2(2) \;+\; 0$$

Here, ϵ_{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):

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

Let's do this by hand...:

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

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

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

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

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

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

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

This means that $\beta 0 = 1$, $\beta 1 = 0$ and that $\beta 2 = 1$. Let's re-write our formula:

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

The formula we had before: $\begin{bmatrix}3\\5\\7\end{bmatrix}_y \;=\; \beta 0 + \beta 1 \begin{bmatrix}2\\7\\3\end{bmatrix}_{x1} \;+\; \beta 2 \begin{bmatrix}2\\4\\6\end{bmatrix}_{x2} + e_i$

Results we have now: $\begin{bmatrix}3\\5\\7\end{bmatrix}_y \;=\; 1 + 0 \begin{bmatrix}2\\7\\3\end{bmatrix}_{x1} \;+\; 1 \begin{bmatrix}2\\4\\6\end{bmatrix}_{x2} + e_i$

Ok, let's now turn to `R`.