Example 1: Consistency of OLS with sample size

x: study hours , y: quiz scores

Suppose the true regression is

$$y_i = 0 + 1x_i + \varepsilon_i$$

If you study 0 hours, you get a 0

For every hour you study, you get an extra point.

So to get a 10 you need at least 10 hours of studying

$x \sim \text{Unif}(0,10)$, $\varepsilon \sim \mathcal{N}(0,1)$

In matrix form

$$Y = X\beta + \varepsilon \implies \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

We know $\hat\beta = (X^TX)^{-1}X^Ty$

Run the numbers

Let's say we have a sample of three students $x = [2\ 5\ 8]$, $y = [2.5\ 4\ 9]$

$$X^TX = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 5 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 3 & 15 \\ 15 & 93 \end{bmatrix}$$

$$(X^TX)^{-1} = \frac{1}{(3 \cdot 93) - (15 \cdot 15)} \begin{bmatrix} 93 & -15 \\ -15 & 3 \end{bmatrix} = \frac{1}{54} \begin{bmatrix} 93 & -15 \\ -15 & 3 \end{bmatrix}$$

$$X^Ty = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 5 & 8 \end{bmatrix} \begin{bmatrix} 2.5 \\ 4 \\ 9 \end{bmatrix} = \begin{bmatrix} 15.5 \\ 97 \end{bmatrix}$$

$$\hat\beta = \underbrace{\frac{1}{54}}_{\text{determ.}} \overbrace{\begin{bmatrix} 93 & -15 \\ -15 & 3 \end{bmatrix}}^{\text{Adjoint}} \begin{bmatrix} 15.5 \\ 97 \end{bmatrix} = \begin{bmatrix} -0.25 \\ 1.083 \end{bmatrix}$$

$$\text{Var}(\hat\beta) = (X^TX)^{-1}X^T \underbrace{\sigma^2 I}_{\text{Var}(\varepsilon)} X(X^TX)^{-1}$$

? Tell me the dimensions of $(X^TX)^{-1}X^T$, σ^2, $\sigma^2 I$ and $X(X^TX)^{-1}$

$$= \sigma^2 (X^TX)^{-1}X^TX(X^TX)^{-1}$$

$$= \sigma^2 (X^TX)^{-1} I_2$$

$$= \sigma^2 \frac{1}{54} \begin{bmatrix} 93 & -15 \\ -15 & 3 \end{bmatrix}$$

Since $\varepsilon \sim \mathcal{N}(0,1) \implies \sigma^2 = 1$

$$\text{Var}(\hat\beta) = \frac{1}{54} \begin{bmatrix} 93 & -15 \\ -15 & 3 \end{bmatrix} = \begin{bmatrix} \frac{31}{18} & -\frac{5}{18} \\ -\frac{5}{18} & \frac{1}{18} \end{bmatrix}$$

$\implies \text{Var}(\hat\beta_0) = \frac{31}{18} \approx 1.72$

$\text{Var}(\hat\beta_1) = \frac{1}{18} \approx 0.06$

$\implies SE(\hat\beta_0) = \sqrt{1.72} = 1.31$

$SE(\hat\beta_1) = \sqrt{0.06} \approx 0.24$

Example 2: Heteroskedasticity

$$y_i = 0 + 1x_i + \varepsilon$$

$x \sim \text{Unif}(0,10)$, but!

$\varepsilon \sim \mathcal{N}(0, x^2)$

↳ Variance grows with x

$$\hat\beta = (X^TX)^{-1}X^Ty$$

$$\text{Var}(\hat\beta) = E[(\hat\beta - \beta)(\hat\beta - \beta)^T]$$

$$= (X^TX)^{-1}X^T E[\varepsilon\varepsilon^T] X (X^TX)^{-1}$$

$$E[\varepsilon\varepsilon^T] \neq \sigma^2 I$$

We need to estimate the var-cov matrix of the residuals. In this case,

$$E[\varepsilon\varepsilon^T] = \begin{bmatrix} x_1^2 & & 0 \\ & x_2^2 & \\ 0 & & x_3^2 \end{bmatrix}$$ "No autocorrelation still holds"

$$= \begin{bmatrix} 4 & & 0 \\ & 25 & \\ 0 & & 64 \end{bmatrix} = \Omega$$

Since it's not a constant diagonal, there's no common factor to pull out and hence it does not simplify nicely

$$\text{Var}(\hat\beta) = (X^TX)^{-1}X^T \Omega X (X^TX)^{-1}$$

$$= \begin{bmatrix} \frac{31}{18} & -\frac{5}{18} \\ -\frac{5}{18} & \frac{1}{18} \end{bmatrix} \begin{bmatrix} 93 & 645 \\ 645 & 4737 \end{bmatrix} \begin{bmatrix} \frac{31}{18} & -\frac{5}{18} \\ -\frac{5}{18} & \frac{1}{18} \end{bmatrix}$$

$$= \begin{bmatrix} -19 & -205 \\ 10 & 84 \end{bmatrix} \begin{bmatrix} \frac{31}{18} & -\frac{5}{18} \\ -\frac{5}{18} & \frac{1}{18} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{213}{9} & -\frac{55}{9} \\ -\frac{55}{9} & \frac{17}{9} \end{bmatrix}$$

$\implies \text{Var}(\hat\beta_0) = \frac{213}{9}$, $SE(\hat\beta_0) = 4.92$

$\text{Var}(\hat\beta_1) = \frac{17}{9}$, $SE(\hat\beta_1) = 1.37$

Compare with homoskedastic SEs.

Major implication of heteroskedasticity:

OLS underestimates SE

We use robust SE to correct them. We'll see more details in the next lecture.

Example 3: Multicollinearity

x_1 := Study hours , x_2 := Sleep hours , y := Quiz score

$x_1 \longleftrightarrow x_2$, $x_1\uparrow \to x_2\downarrow$ and $x_2\uparrow \to x_1\downarrow$

The regressors are not independent

Suppose the true model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$\beta_0 = 1$ "Baseline score for existing"

$\beta_1 = 2$, $\beta_2 = 1.5$, $\varepsilon \sim \mathcal{N}(0,1)$

Suppose $x_1 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$ and $x_2 = 10 - x_1 + u$

$\implies u \sim \mathcal{N}(0, \frac{1}{4})$

The more you study the less you sleep + some random noise.

Assume $u = \begin{bmatrix} -0.5 \\ 0 \\ 0.5 \end{bmatrix}$ for simplicity

$$\implies x_2 = \begin{bmatrix} 5.5 \\ 5 \\ 4.5 \end{bmatrix}$$

Then we have

$y = 1 + 2x_1 + 1.5x_2 + \varepsilon$, say $\varepsilon = \begin{bmatrix} 0.2 \\ -0.1 \\ 0.3 \end{bmatrix}$

$$\implies y = \begin{bmatrix} 17.45 \\ 18.4 \\ 20.05 \end{bmatrix}$$

OLS ...

$\beta = \begin{bmatrix} 1 \\ 2 \\ 1.5 \end{bmatrix}$, $\hat\beta = (X^TX)^{-1}X^Ty$

$X = \begin{bmatrix} 1 & 4 & 5.5 \\ 1 & 5 & 5 \\ 1 & 6 & 4.5 \end{bmatrix}$, $X^TX = \begin{bmatrix} 3 & 15 & 15 \\ 15 & 77 & 74 \\ 15 & 74 & 75.5 \end{bmatrix}$

$\det(X^TX) = 0 \implies X^TX$ is singular and not invertible

$\implies$ This is a consequence of multicollinearity

Because the inverse is unstable (we can approx. computationally) we will get very high estimates for $\text{Var}(\hat\beta)$

Solve for this inverse with R and see what you get

Example 4: OVB

There's always some latent variable we would like to measure but we can't. Can you think of any relevant ones for our example?

IQ or intelligence is a good one.

Higher IQ probably means you have to study less and therefore you can sleep more.

To make this effect obvious in the math, suppose we have the true model

$$y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \varepsilon$$

$\beta_0 = 1$, $\beta_1 = 2$, $\beta_3 = \frac{1}{2}$, $\varepsilon \sim \mathcal{N}(0,1)$

Suppose $x_3 = \begin{bmatrix} 90 \\ 100 \\ 110 \end{bmatrix}$ and $x_1 = 10 - 0.1x_3 + u$

$u = \begin{bmatrix} 0.1 \\ 0 \\ -0.1 \end{bmatrix}$

$\implies x_1 = \begin{bmatrix} 1.1 \\ 0 \\ -1.1 \end{bmatrix}$, and $\varepsilon = \begin{bmatrix} 0.2 \\ -0.1 \\ 0.3 \end{bmatrix}$

$$\implies y = \begin{bmatrix} 48.4 \\ 50.9 \\ 54.1 \end{bmatrix}$$

Now suppose we omit IQ, so we estimate

$$y = \beta_0 + \beta_1 x_1 + \varepsilon$$

$X = \begin{bmatrix} 1 & 1.1 \\ 1 & 0 \\ 1 & -1.1 \end{bmatrix}$, $X^TX = \begin{bmatrix} 3 & 0 \\ 0 & 2.42 \end{bmatrix}$

$$(X^TX)^{-1} = \begin{bmatrix} 1/3 & 0 \\ 0 & 1/2.42 \end{bmatrix}$$

$$X^Ty = \begin{bmatrix} 153.4 \\ -6.27 \end{bmatrix} \implies \hat\beta = \begin{bmatrix} 51.133 \\ -2.59 \end{bmatrix}$$

Completely off! Even the sign flipped.

This is because x_1 and x_3 are correlated and by omitting x_3 from our regression then we are missing key information to explain the true/observed values of x_1.