Econometrics

Practical Session 1

Introduction to OLS

Ricardo Gouveia-Mendes
rgouveiamendes@ucp.pt

Spring 2025-26

Católica-Lisbon School of Business and Economics

Theoretical Wrap-up

Simple Linear Regression

- **Theory** states that a variable X has **some effect** on a variable Y

$$Y = f(X) \tag{1}$$

Simple Linear Regression

- **Theory** states that a variable X has **some effect** on a variable Y

$$Y = f(X) \tag{1}$$

- The exact form of $f(X)$ is usually unknown → simplest form is linear:

$$Y = \underbrace{\beta_0 + \beta_1 X}_{\mathbb{E}[Y|X]} + u \tag{2}$$

Simple Linear Regression

- **Theory** states that a variable X has **some effect** on a variable Y

$$Y = f(X) \tag{1}$$

- The exact form of $f(X)$ is usually unknown → simplest form is linear:

$$Y = \underbrace{\beta_0 + \beta_1 X}_{\mathbb{E}[Y|X]} + u \tag{2}$$

- β_1 qualifies the **sign and magnitude of the effect** of X on Y → **How to estimate β_0 and β_1?**

Simple Linear Regression

- **Theory** states that a variable X has **some effect** on a variable Y

$$Y = f(X) \tag{1}$$

- The exact form of $f(X)$ is usually unknown → simplest form is linear:

$$Y = \underbrace{\beta_0 + \beta_1 X}_{\mathbb{E}[Y|X]} + u \tag{2}$$

- β_1 qualifies the **sign and magnitude of the effect** of X on Y → **How to estimate β_0 and β_1?**

- Collecting a **sample** → N pairs of observations (X_i, Y_i) assuming:

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad \forall i = 1, ..., N$$

Simple Linear Regression

- **Theory** states that a variable X has **some effect** on a variable Y

$$Y = f(X) \tag{1}$$

- The exact form of $f(X)$ is usually unknown → simplest form is linear:

$$Y = \underbrace{\beta_0 + \beta_1 X}_{\mathbb{E}[Y|X]} + u \tag{2}$$

- β_1 qualifies the **sign and magnitude of the effect** of X on Y → **How to estimate β_0 and β_1?**

- Collecting a **sample** → N pairs of observations (X_i, Y_i) assuming:

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad \forall i = 1, \dots, N$$

- The estimates for the coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ would allow to predict a value for $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

Simple Linear Regression

- **Theory** states that a variable X has **some effect** on a variable Y

$$Y = f(X) \tag{1}$$

- The exact form of $f(X)$ is usually unknown → simplest form is linear:

$$Y = \underbrace{\beta_0 + \beta_1 X}_{\mathbb{E}[Y|X]} + u \tag{2}$$

- β_1 qualifies the **sign and magnitude of the effect** of X on Y → **How to estimate β_0 and β_1?**

- Collecting a **sample** → N pairs of observations (X_i, Y_i) assuming:

$$Y_i = \beta_0 + \beta_1 X_i + u_i, \quad \forall i = 1, \dots, N$$

- The estimates for the coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ would allow to predict a value for $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$

- Recipe → **Ordinary Least Squares**:

$$\min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i\right)^2 = \min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{N} \hat{u}_i^2$$

Terminology

Econometric Model

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad (3)$$

- **Dependent** Variable: Y
- **Independent** Variable or **Regressor**: X
- Regression **Coefficients**: β_0 and β_1
- **Error** Term: u

Estimated Model

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i \qquad (4)$$

- **Estimated** Regression **Coefficients**: $\hat{\beta}_0$ and $\hat{\beta}_1$
- **Predicted Value**: $\hat{Y}_i$
- **Residual**: $\hat{u}_i = Y_i - \hat{Y}_i$
- Notice that:

$$\hat{\beta}_1 = \frac{\Delta \hat{Y}}{\Delta X}$$

The OLS Estimators

$$\min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{i=1}^{N} \hat{u}_i^2 = \min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{i=1}^{N} \left(Y_i - \hat{Y}_i\right)^2 = \min_{\hat{\beta}_0,\hat{\beta}_1} \sum_{i=1}^{N} \left(Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_i\right)^2$$

- Solving for $\hat{\beta}_0$:

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X} \tag{5}$$

- Solving for $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)\left(X_i - \overline{X}\right)}{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2} = \frac{s_{Y,X}}{s_X^2} \tag{6}$$

Exercise

Wage and Education

Consider the following simple linear regression model:

$$\text{wage}_i = \beta_0 + \beta_1 \text{schooling}_i + u_i$$

where:

- wage_i represents the wage of an individual,
- schooling_i is the number of years of education,
- u is the error term.

Wage and Education

You are given the following dataset:

TABLE 1. Wage and Schooling Data

id	wage	schooling
1	6.00	8
2	5.30	12
3	8.75	16
4	11.25	18
5	5.00	12
6	3.60	12
7	18.18	17
8	6.25	16
9	8.13	13
10	8.77	12

Using the least squares method, find the estimated coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$:

1. By hand
2. Using Excel
3. Using R

Wage and Education

By Hand

- **Step 1:** Compute Means

$$\overline{X} = \frac{8 + 12 + 16 + 18 + 12 + 12 + 17 + 16 + 13 + 12}{10} = 13.6$$

$$\overline{Y} = \frac{6 + 5.3 + 8.75 + 11.25 + 5 + 3.6 + 18.18 + 6.25 + 8.13 + 8.77}{10} = 8.123$$

Wage and Education

- **Step 2:** Compute $\hat{\beta}_1$

$$\sum_{i=1}^{10}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right) = 72.562, \quad \sum_{i=1}^{10}\left(X_i - \overline{X}\right)^2 = 84.4$$

$$\hat{\beta}_1 = \frac{72.562}{84.4} = 0.8597$$

- **Step 3:** Compute $\hat{\beta}_0$

$$\hat{\beta}_0 = 8.123 - (0.8597 \times 13.6) = -3.569$$

- **Conclusion:** The estimated regression is given by:

$$\widehat{\text{wage}}_i = -3.569 + 0.8597\text{schooling}_i$$

Wage and Education

Using R

```r
# Load necessary package
library(readxl)

# Load dataset
df <- read_excel("path/to/wage_educ_data.xlsx")

# Run linear regression model
m <- lm(wage ~ schooling, data = df)

# Display summary of the model
summary(m)
```