# ECON 351* -- NOTE 2

# Ordinary Least Squares (OLS) Estimation of the Simple CLRM

## 1. The Nature of the Estimation Problem

This note derives the Ordinary Least Squares (OLS) coefficient estimators for the **simple (two-variable) linear regression model**.

**1.1** The **population regression equation**, or **PRE**, for the simple (two-variable) linear regression model takes the form:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \qquad (i = 1, ..., N) \qquad (1)$$

where $u_i$ is an iid random error term.

**1.2** The **OLS sample regression equation** (or **OLS-SRE**) corresponding to equation (1) can be written as

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i = \hat{Y}_i + \hat{u}_i \qquad (i = 1, ..., N). \qquad (2)$$

where:

$\hat{\beta}_0 =$ the OLS estimator of the ***intercept*** **coefficient** $\boldsymbol{\beta_0}$;

$\hat{\beta}_1 =$ the OLS estimator of the ***slope*** **coefficient** $\boldsymbol{\beta_1}$;

$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ = the OLS estimated (or predicted) values of $E(Y_i \mid X_i) = \beta_0 + \beta_1 X_i$ for sample observation i, and is called the **OLS sample regression function** (or **OLS-SRF**);

$\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$ = the **OLS residual** for sample observation i.

**1.3 The Estimation Problem:** The estimation problem consists of constructing or deriving the OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ for any given sample of N observations $(Y_i, X_i)$, i = 1, ..., N on the observable variables Y and X.

**Definition of an Estimator**

An **estimator** of a population parameter is a **rule, formula, or procedure** for computing a numerical estimate of an unknown population parameter from the sample values of the observable variables.

*Example:* The estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ of the population parameters $\beta_0$ and $\beta_1$ in the simple PRE (1) are therefore formulas that tell us how to compute numerical estimates of $\beta_0$ and $\beta_1$ from the given sample data on the observable variables Y and X.

**Properties of an Estimator**

An estimator possesses *two* critical properties.

1. An estimator is a **function only of the given sample data**; this function **does not contain any *unknown* parameters**.

   *Examples:* In the context of the simple linear regression model represented by PRE (1), the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ of the regression coefficients $\beta_0$ and $\beta_1$ are functions of the sample data on Y and X, which consist of a random sample of N observed values $(Y_i, X_i)$, i = 1, ..., N. Thus, the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ take the general form

$$\hat{\beta}_0 = f_0(Y_1, Y_2, \ldots, Y_N; X_1, X_2, \ldots, X_N) = f_0(Y_i, X_i) \qquad i = 1, ..., N$$
$$\hat{\beta}_1 = f_1(Y_1, Y_2, \ldots, Y_N; X_1, X_2, \ldots, X_N) = f_1(Y_i, X_i) \qquad i = 1, ..., N$$

   where the functions $f_0(\cdots)$ and $f_1(\cdots)$ contain no unknown parameters.

2. An estimator is a **random variable** (or **statistic**) because it is a function of the sample data $(Y_i, X_i)$, i = 1, ..., N. An estimator therefore has a probability distribution, which is called the **sampling distribution of the estimator**.

**Distinction Between “Estimators” and “Estimates”**

- The term ***estimator*** refers to the **formula or rule or procedure** by which a numerical estimate of an unknown population parameter is computed from any given sample data.

  *Examples:*

  - The estimator $\hat{\beta}_0$ of the intercept coefficient $\beta_0$ is the function $f_0(Y_1, Y_2, \ldots, Y_N; X_1, X_2, \ldots, X_N) = f_0(Y_i, X_i)$ which tells us how to compute a numerical estimate of the parameter $\beta_0$ for any given sample values ($Y_i$, $X_i$), i = 1, ..., N of the observable variables Y and X.

  - Similarly, the estimator $\hat{\beta}_1$ of the slope coefficient $\beta_1$ is the function $f_1(Y_1, Y_2, \ldots, Y_N; X_1, X_2, \ldots, X_N) = f_1(Y_i, X_i)$ which tells us how to compute a numerical estimate of the parameter $\beta_1$ for any given sample values ($Y_i$, $X_i$), i = 1, ..., N of the observable variables Y and X.

- The term ***estimate*** refers to the **specific numerical value given by the formula** for a specific set of sample values ($Y_i$, $X_i$), i = 1, ..., N of the observable variables Y and X. That is, an estimate is the **value of the estimator** obtained when the formula is evaluated for a particular set of sample values of the observable variables.

  *Example:* Suppose that for a particular sample of 50 observed values of $Y_i$ and $X_i$ that the formula or function $f_1(\cdots)$ yields the value $\hat{\beta}_1 = 0.83$. Then the number 0.83 is an *estimate* of the slope coefficient $\beta_1$.

# 2. The OLS Estimation Criterion

The OLS coefficient estimators are those formulas or expressions for $\hat{\beta}_0$ and $\hat{\beta}_1$ that **minimize the sum of squared residuals RSS** for any given sample of size N.

The **OLS estimation criterion** is therefore:

$$\underset{\{\hat{\beta}_j\}}{\text{Minimize}}\ \text{RSS}\left(\hat{\beta}_0, \hat{\beta}_1\right) = \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2 \qquad \textbf{(3)}$$

## Rationale for the OLS Estimation Criterion

- ***Theoretical Rationale:*** The OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ implied by the OLS estimation criterion have several **desirable statistical properties** when the assumptions of the CLRM (Classical Linear Regression Model) are satisfied.

- ***Intuitive Rationale:*** The OLS estimation criterion corresponds to the **idea of "best fit"** of the estimated sample regression function (SRF) to the given sample data ($Y_i$, $X_i$), i = 1, ..., N.

  Note that the OLS criterion minimizes the **sum of *squared* residuals** $\sum_i \hat{u}_i^2$, *not* the sum of the residuals $\sum_i \hat{u}_i$.

  Squaring the residuals $\hat{u}_i$ does two things:

  **(1)** It avoids the possibility that large *positive* residuals and large *negative* residuals could offset each other and still lead to a small (or even zero) value of $\sum_i \hat{u}_i$.

  **(2)** It implicitly assigns a larger weight to numerically large residuals (regardless of whether they are positive or negative). The larger the numerical or absolute value of a particular $\hat{u}_i$, the larger the corresponding value of $\hat{u}_i^2$.

## 3. Derivation of the OLS Coefficient Estimators

Derivation of the OLS formulas for the regression coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ is performed in ***two* stages**:

<u>**STAGE 1**</u> consists of ***deriving* the first-order conditions (or FOCs)** for minimizing the residual sum of squares function $\text{RSS}(\hat{\beta}_0, \hat{\beta}_1)$ given by equation (3). These FOCs are called the **OLS normal equations**.

<u>**STAGE 2**</u> consists of ***solving* the first-order conditions (FOCs) -- i.e., the OLS normal equations --** to obtain explicit formulas or expressions for the OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$.

## **STAGE 1**: Derivation of the OLS Normal Equations, or FOCs

**Step 1.1**: **Partially differentiate** the $RSS(\hat{\beta}_0, \hat{\beta}_1)$ function in (3) with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$.

- First, **rewrite the $RSS(\hat{\beta}_0, \hat{\beta}_1)$ function** in (3) as follows:

$$RSS(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} f(\hat{u}_i) \quad \text{where} \qquad f(\hat{u}_i) = \hat{u}_i^2$$

$$\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i \qquad \textbf{(3)}$$

*Note:* The function $f(\hat{u}_i) = \hat{u}_i^2$ is a function of $\hat{u}_i$, and $\hat{u}_i$ is in turn a function of $\hat{\beta}_0$ and $\hat{\beta}_1$.

- Using the **chain rule of differentiation**, each partial derivative of the $RSS(\hat{\beta}_0, \hat{\beta}_1)$ function takes the general form

$$\frac{\partial RSS}{\partial \hat{\beta}_j} = \sum_{i=1}^{N} \frac{df}{d\hat{u}_i} \frac{\partial \hat{u}_i}{\partial \hat{\beta}_j}. \qquad (4)$$

- Using the **power rule of differentiation**, the derivative $df / d\hat{u}_i$ is

$$\frac{df}{d\hat{u}_i} = \frac{d(\hat{u}_i^2)}{d\hat{u}_i} = 2\hat{u}_i.$$

- The partial derivatives $\partial RSS / \partial \hat{\beta}_j$ for j = 0, 1 are therefore

$$\frac{\partial RSS}{\partial \hat{\beta}_j} = \sum_{i=1}^{N} \frac{df}{d\hat{u}_i} \frac{\partial \hat{u}_i}{\partial \hat{\beta}_j} = \sum_{i=1}^{N} 2\hat{u}_i \frac{\partial \hat{u}_i}{\partial \hat{\beta}_j} = 2\sum_{i=1}^{N} \hat{u}_i \frac{\partial \hat{u}_i}{\partial \hat{\beta}_j} \qquad j = 0, 1. \qquad (5)$$

- Since the i-th residual is $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$, the partial derivatives $\partial \hat{u}_i / \partial \hat{\beta}_j$ for j = 0, 1 are:

for j = 0: $\dfrac{\partial \hat{u}_i}{\partial \hat{\beta}_0} = -1;$

for j = 1: $\dfrac{\partial \hat{u}_i}{\partial \hat{\beta}_1} = -X_i.$

- Substitute the partial derivatives $\partial \hat{u}_i / \partial \hat{\beta}_j$ for j = 0, 1 into equation (5):

$$\frac{\partial RSS}{\partial \hat{\beta}_j} = 2 \sum_{i=1}^{N} \hat{u}_i \frac{\partial \hat{u}_i}{\partial \hat{\beta}_j} \qquad j = 0, 1. \tag{5}$$

The partial derivatives $\partial RSS / \partial \hat{\beta}_j$ for j = 0, 1 thus take the form:

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = 2 \sum_{i=1}^{N} \hat{u}_i \frac{\partial \hat{u}_i}{\partial \hat{\beta}_0} = 2 \sum_{i=1}^{N} \hat{u}_i (-1) = -2 \sum_{i=1}^{N} \hat{u}_i \tag{6.1}$$

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = 2 \sum_{i=1}^{N} \hat{u}_i \frac{\partial \hat{u}_i}{\partial \hat{\beta}_1} = 2 \sum_{i=1}^{N} \hat{u}_i (-X_i) = -2 \sum_{i=1}^{N} X_i \hat{u}_i \tag{6.2}$$

**<u>Step 1.2</u>: Obtain the first-order conditions (FOCs)** for a minimum of the RSS function by setting the partial derivatives (6.1)-(6.2) equal to zero, then dividing each equation by –2, and finally setting $\hat{u}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$:

- $$\frac{\partial RSS}{\partial \hat{\beta}_0} = -2\sum_{i=1}^{N}\hat{u}_i \qquad (6.1)$$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = 0 \quad \Rightarrow \quad -2\sum_{i=1}^{N}\hat{u}_i = 0 \quad \Rightarrow \quad \sum_{i=1}^{N}\hat{u}_i = 0 \qquad (7.1)$$

$$\Rightarrow \quad \sum_{i=1}^{N}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0 \qquad (8.1)$$

- $$\frac{\partial RSS}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{N}X_i\hat{u}_i \qquad (6.2)$$

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = 0 \quad \Rightarrow \quad -2\sum_{i=1}^{N}X_i\hat{u}_i = 0 \quad \Rightarrow \quad \sum_{i=1}^{N}X_i\hat{u}_i = 0 \qquad (7.2)$$

$$\Rightarrow \quad \sum_{i=1}^{N}X_i\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0 \qquad (8.2)$$

- Equations (7.1) and (7.2) are the most compact way of writing the FOCs for the OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$:

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{N}\hat{u}_i = 0 \qquad \textbf{(7.1)}$$

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = 0 \quad \Leftrightarrow \quad \sum_{i=1}^{N}X_i\hat{u}_i = 0 \qquad \textbf{(7.2)}$$

- But equations (8.1) and (8.2) are used to obtain the formulas for the OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$:

$$\sum_{i=1}^{N}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0\,; \qquad \textbf{(8.1)}$$

$$\sum_{i=1}^{N} X_i\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0\,. \qquad \textbf{(8.2)}$$

**<u>Step 1.3</u>:** Rearrange each of the equations (8.1) and (8.2) to put them in the conventional form of the OLS normal equations. Thus, taking summations and rearranging terms, we obtain the **OLS normal equations**:

$$\sum_{i=1}^{N}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0 \qquad \textbf{(8.1)}$$

$$\begin{aligned} \textstyle\sum_i Y_i - N\hat{\beta}_0 - \hat{\beta}_1 \textstyle\sum_i X_i &= 0 \\ - N\hat{\beta}_0 - \hat{\beta}_1 \textstyle\sum_i X_i &= -\textstyle\sum_i Y_i \\ N\hat{\beta}_0 + \hat{\beta}_1 \textstyle\sum_i X_i &= \textstyle\sum_i Y_i \end{aligned} \qquad (9.1)$$

$$\sum_{i=1}^{N} X_i\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right) = 0 \qquad \textbf{(8.2)}$$

$$\begin{aligned} \textstyle\sum_i (X_i Y_i - \hat{\beta}_0 X_i - \hat{\beta}_1 X_i^2) &= 0 \\ \textstyle\sum_i X_i Y_i - \hat{\beta}_0 \textstyle\sum_i X_i - \hat{\beta}_1 \textstyle\sum_i X_i^2 &= 0 \\ -\hat{\beta}_0 \textstyle\sum_i X_i - \hat{\beta}_1 \textstyle\sum_i X_i^2 &= -\textstyle\sum_i X_i Y_i \\ \hat{\beta}_0 \textstyle\sum_i X_i + \hat{\beta}_1 \textstyle\sum_i X_i^2 &= \textstyle\sum_i X_i Y_i \end{aligned} \qquad (9.2)$$

- **<u>RESULT</u>:** The last equations in (9.1) and (9.2) are the **OLS normal equations**:

$$N\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{N} X_i = \sum_{i=1}^{N} Y_i \qquad \textbf{(N1)}$$

$$\hat{\beta}_0 \sum_{i=1}^{N} X_i + \hat{\beta}_1 \sum_{i=1}^{N} X_i^2 = \sum_{i=1}^{N} X_i Y_i \qquad \textbf{(N2)}$$

The **OLS normal equations (N1) and (N2)** constitute ***two linear equations*** **in the *two unknowns*** $\hat{\beta}_0$ and $\hat{\beta}_1$. Their solution yields explicit expressions for $\hat{\beta}_0$ and $\hat{\beta}_1$; these expressions are the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ of the regression coefficients $\beta_0$ and $\beta_1$.

## **STAGE 2: Solution of the OLS Normal Equations**

There is more than one way to solve the OLS normal equations (N1) and (N2) for the two unknowns $\hat{\beta}_0$ and $\hat{\beta}_1$. The following steps constitute one such solution method.

**Step 2.1:** Divide normal equation (N1) by N and then solve for $\hat{\beta}_0$:

$$N\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{N} X_i = \sum_{i=1}^{N} Y_i \qquad \textbf{(N1)}$$

$$\hat{\beta}_0 + \hat{\beta}_1 \left( \frac{\sum_i X_i}{N} \right) = \left( \frac{\sum_i Y_i}{N} \right)$$

$$\hat{\beta}_0 + \hat{\beta}_1 \overline{X} = \overline{Y} \qquad \text{since } \overline{X} \equiv \sum_i X_i / N \text{ and } \overline{Y} \equiv \sum_i Y_i / N.$$

Therefore, solving for $\hat{\beta}_0$:

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}. \qquad \textbf{(10)}$$

**Step 2.2:** Substitute the expression for $\hat{\beta}_0$ given by equation (10) into normal equation (N2):

$$\hat{\beta}_0 \sum_{i=1}^{N} X_i + \hat{\beta}_1 \sum_{i=1}^{N} X_i^2 = \sum_{i=1}^{N} X_i Y_i \qquad \textbf{(N2)}$$

$$\left( \overline{Y} - \hat{\beta}_1 \overline{X} \right) \Sigma_i X_i + \hat{\beta}_1 \Sigma_i X_i^2 = \Sigma_i X_i Y_i \qquad (11)$$

**Step 2.3:** Since $\overline{X} \equiv \sum_i X_i / N$ by definition, $\sum_i X_i = N\overline{X}$. Set $\sum_i X_i = N\overline{X}$ in equation (11):

$$\begin{aligned} \left( \overline{Y} - \hat{\beta}_1 \overline{X} \right) N\overline{X} + \hat{\beta}_1 \Sigma_i X_i^2 &= \Sigma_i X_i Y_i \\ N\overline{X}\,\overline{Y} - \hat{\beta}_1 N\overline{X}^2 + \hat{\beta}_1 \Sigma_i X_i^2 &= \Sigma_i X_i Y_i. \end{aligned} \qquad (12)$$

**Step 2.4:** Solve the last equation in (12) for $\hat{\beta}_1$:

$$N\bar{X}\bar{Y} - \hat{\beta}_1 N\bar{X}^2 + \hat{\beta}_1 \Sigma_i X_i^2 = \Sigma_i X_i Y_i$$

$$- \hat{\beta}_1 N\bar{X}^2 + \hat{\beta}_1 \Sigma_i X_i^2 = \Sigma_i X_i Y - N\bar{X}\bar{Y}$$

$$\hat{\beta}_1 \left(\Sigma_i X_i^2 - N\bar{X}^2\right) = \Sigma_i X_i Y - N\bar{X}\bar{Y}.$$

Therefore, solving for $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum_i X_i Y_i - N\bar{X}\bar{Y}}{\sum_i X_i^2 - N\bar{X}^2} \qquad \textbf{(13)}$$

- **RESULT:** Equations (10) and (13) are the OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$.

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}. \qquad \textbf{(10)}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} X_i Y_i - N\bar{X}\bar{Y}}{\sum_{i=1}^{N} X_i^2 - N\bar{X}^2}. \qquad \textbf{(13)}$$

They represent the solution of the OLS normal equations; that is, they represent the solution of the FOCs for minimizing the residual sum-of-squares function $RSS\left(\hat{\beta}_0, \hat{\beta}_1\right)$ given by equation (3).

- **Computational Note:** If we were actually using formulas (10) and (13) to compute estimates of $\beta_0$ and $\beta_1$ for a given sample of N observations ($Y_i$, $X_i$), i = 1, ..., N, we would employ the following **two-step computational procedure**.

  **1.** First, use equation (13) to compute the estimate of $\beta_1$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{i=N} X_i Y_i - N\bar{X}\bar{Y}}{\sum_{i=1}^{i=N} X_i^2 - N\bar{X}^2} \qquad \textbf{(14.1)}$$

  **2.** Second, substitute $\hat{\beta}_1$, the estimate of $\beta_1$ computed by equation (14.1), into the formula for $\hat{\beta}_0$ to obtain the estimate of $\beta_0$:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} \qquad \textbf{(14.2)}$$

- **Characteristics of the OLS Coefficient Estimators $\hat{\beta}_0$ and $\hat{\beta}_1$:**

  **1.** The OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are **functions *only* of the observed sample values ($Y_i$, $X_i$): i = 1, ..., N** of the observable variables Y and X. They can therefore be computed for any given set of sample data.

  **2.** The OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are ***point* estimators**. That is, each estimator provides only a single (point) value of the corresponding population parameter.

# 4. Alternative Expressions for the OLS Slope Coefficient Estimator

## 4.1 Deviation-From-Means Formula for $\hat{\beta}_1$: Derivation

Formula (14.1) for the OLS slope coefficient estimator $\hat{\beta}_1$ can conveniently be re-written in ***deviation-from-means form***.

- **Deviation-from-means notation** uses ***lower case letters*** to denote the deviations of each observed sample value from its corresponding sample mean.

  **1.** Define the deviations of the $Y_i$ sample values from their sample mean $\overline{Y}$ as

  $$y_i \equiv Y_i - \overline{Y} \qquad (i = 1, ..., N)$$

  where $\overline{Y} = \Sigma_i Y_i / N = \dfrac{\Sigma_i Y_i}{N}$ = the sample mean of the $Y_i$ values.

  **2.** Define the deviations of the $X_i$ sample values from their sample mean $\overline{X}$ as

  $$x_i \equiv X_i - \overline{X} \qquad (i = 1, ..., N)$$

  where $\overline{X} = \Sigma_i X_i / N = \dfrac{\Sigma_i X_i}{N}$ = the sample mean of the $X_i$ values.

- **Formula (14.1)** for the OLS slope coefficient estimator $\hat{\beta}_1$ is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{i=N} X_i Y_i - N\bar{X}\bar{Y}}{\sum_{i=1}^{i=N} X_i^2 - N\bar{X}^2} \qquad (14.1)$$

- The **numerator of formula (14.1)** for OLS slope coefficient estimator $\hat{\beta}_1$ can be shown to equal

$$\sum_i x_i y_i = \sum_i \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right). \qquad (15.1)$$

*Proof:* Expand the expression on the right-hand side of (15.1):

$$\begin{aligned}
\sum_i x_i y_i &= \sum_i \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) \\
&= \sum_i \left(X_i Y_i - \bar{X} Y_i - \bar{Y} X_i + \bar{X}\bar{Y}\right) \\
&= \sum_i X_i Y_i - \bar{X}\sum_i Y_i - \bar{Y}\sum_i X_i + N\bar{X}\bar{Y} \\
&= \sum_i X_i Y_i - N\bar{X}\bar{Y} - N\bar{X}\bar{Y} + N\bar{X}\bar{Y} \quad \text{since } \sum_i Y_i = N\bar{Y}, \sum_i X_i = N\bar{X} \\
&= \sum_i X_i Y_i - N\bar{X}\bar{Y} = \text{ the numerator of formula (14.1) for } \hat{\beta}_1.
\end{aligned}$$

- The **denominator of formula (14.1)** for OLS slope coefficient estimator $\hat{\beta}_1$ can be shown to equal

$$\sum_i x_i^2 = \sum_i \left(X_i - \bar{X}\right)^2. \qquad (15.2)$$

*Proof:* Expand the expression on the right-hand side of (15.2):

$$\begin{aligned}
\sum_i x_i^2 &= \sum_i \left(X_i - \bar{X}\right)^2 \\
&= \sum_i \left(X_i^2 - 2\bar{X}X_i + \bar{X}^2\right) \\
&= \sum_i X_i^2 - 2\bar{X}\sum_i X_i + N\bar{X}^2 \\
&= \sum_i X_i^2 - 2N\bar{X}^2 + N\bar{X}^2 \quad \text{since } \sum_i X_i = N\bar{X} \\
&= \sum_i X_i^2 - N\bar{X}^2 = \text{ the denominator of formula (14.1) for } \hat{\beta}_1.
\end{aligned}$$

- **<u>RESULT</u>:** The **OLS slope coefficient estimator $\hat{\beta}_1$** can be written in (at least) ***three* equivalent ways**:

$$\hat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

$$= \frac{\sum_i \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_i \left(X - \bar{X}\right)^2} \qquad (16)$$

$$= \frac{\sum_i X_i Y_i - N\overline{XY}}{\sum_i X_i^2 - N\bar{X}^2}$$

But the easiest and most useful formula for $\hat{\beta}_1$ is the deviations-from-means formula

$$\hat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}.$$

## 4.2 Deviation-From-Means Formula for $\hat{\beta}_1$: Interpretation

The deviation-from-means form of the OLS slope coefficient estimator $\hat{\beta}_1$ can be given a convenient and easy-to-remember interpretation. This interpretation makes use of two sample statistics: (1) the sample covariance of $X_i$ and $Y_i$; and (2) the sample variance of $X_i$.

- Define the **sample *covariance* of the observed $X_i$ and $Y_i$ values** as:

$$\text{Cov}(X_i, Y_i) = \frac{\sum_i x_i y_i}{N} = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{N}.$$

- Define the **sample *variance* of the observed $X_i$ values** as:

$$\text{Var}(X_i) = \frac{\sum_i x_i^2}{N} = \frac{\sum_i (X_i - \bar{X})^2}{N}.$$

Note: Var($X_i$) as defined above is a biased but consistent estimator of the population variance of X, conventionally denoted as $\sigma_X^2$. The unbiased (and consistent) estimator of $\sigma_X^2$ is given by $s_X^2 = \sum_i x_i^2 / N - 1$.

- Now divide both the numerator and denominator of the deviation-from-means formula for the OLS slope coefficient estimator $\hat{\beta}_1$ by N:

$$\hat{\beta}_1 = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \frac{\sum_i x_i y_i / N}{\sum_i x_i^2 / N} = \frac{\text{Cov}(X_i, Y_i)}{\text{Var}(X_i)}. \qquad (17)$$

❑ **RESULT:** The **OLS slope coefficient estimator $\hat{\beta}_1$** can be interpreted as **the ratio of (1) the sample *covariance* of $X_i$ and $Y_i$ to (2) the sample *variance* of $X_i$.**