PhD Math Camp: Linear Regression

Shangze Rudy Dai

Food and Resource Economics Department
University of Florida

Aug 13, 2025

Contents

Section 1

Introduction

What Do Economists Care

Economists are often interested in how one variable can be **linearly represented** or **predicted** by another.

- **Wages and Education:** Years of schooling $\rightarrow$ predict log(wage).
- **Consumption and Income:** Household income $\rightarrow$ consumption expenditure.
- **Demand and Price:** Product price $\rightarrow$ quantity demanded.
- **CO_2 Emissions and GDP:** Economic output $\rightarrow$ environmental impact.
- **Investment and Interest Rates:** Interest rate $\rightarrow$ investment level.

What Do Economists Care

Economists are often interested in how one variable can be **linearly represented** or **predicted** by another.

- **Wages and Education:** Years of schooling $\rightarrow$ predict log(wage).
- **Consumption and Income:** Household income $\rightarrow$ consumption expenditure.
- **Demand and Price:** Product price $\rightarrow$ quantity demanded.
- **CO_2 Emissions and GDP:** Economic output $\rightarrow$ environmental impact.
- **Investment and Interest Rates:** Interest rate $\rightarrow$ investment level.

These relationships are often studied using **simple linear regression**, where:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Our Goal: Estimating β

In simple linear regression, we assume:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- β_0 — the **intercept**: expected value of Y when $X = 0$.
- β_1 — the **slope**: change in Y for a one-unit change in X.

Our Goal: Estimating β

In simple linear regression, we assume:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- β_0 — the **intercept**: expected value of Y when $X = 0$.
- β_1 — the **slope**: change in Y for a one-unit change in X.
- ε — the **error term**: all other factors affecting Y that are **not captured** by X.
  - Examples: measurement error, omitted variables, random shocks.
  - In economics: captures unobserved heterogeneity.

Our Goal: Estimating β

In simple linear regression, we assume:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

- β_0 — the **intercept**: expected value of Y when $X = 0$.
- β_1 — the **slope**: change in Y for a one-unit change in X.
- ε — the **error term**: all other factors affecting Y that are **not captured** by X.
  - Examples: measurement error, omitted variables, random shocks.
  - In economics: captures unobserved heterogeneity.

Our task: Use sample data to **estimate** β_0 and β_1 so that the model best fits the observed relationship between X and Y.

Remember? You've learnt estimation yesterday!

How Do We Estimate β?

We know the model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

How Do We Estimate β?

We know the model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Question for you: How can we use the sample data to find good estimates of β_0 and β_1?

How Do We Estimate β?

We know the model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Question for you: How can we use the sample data to find good estimates of β_0 and β_1?
Possible methodological principles:

- **Minimize variance of errors** — find β so that the sum of squared residuals

  $$\sum_{i=1}^{n} \varepsilon_i^2$$

  is as small as possible. $\Rightarrow$ This is the idea behind **Ordinary Least Squares (OLS)**.

How Do We Estimate β?

We know the model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Question for you: How can we use the sample data to find good estimates of β_0 and β_1?
Possible methodological principles:

- **Minimize variance of errors** — find β so that the sum of squared residuals

$$\sum_{i=1}^{n} \varepsilon_i^2$$

  is as small as possible. $\Rightarrow$ This is the idea behind **Ordinary Least Squares (OLS)**.

- **Maximum Likelihood Estimation (MLE)** — choose β to maximize the probability of observing our data, given the model.

How Do We Estimate β?

We know the model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Question for you: How can we use the sample data to find good estimates of β_0 and β_1?

Possible methodological principles:

- **Minimize variance of errors** — find β so that the sum of squared residuals

  $$\sum_{i=1}^{n} \varepsilon_i^2$$

  is as small as possible. $\Rightarrow$ This is the idea behind **Ordinary Least Squares (OLS)**.

- **Maximum Likelihood Estimation (MLE)** — choose β to maximize the probability of observing our data, given the model.

These two approaches often lead to the same estimates (e.g., when $\varepsilon \sim N(0, \sigma^2)$), but their interpretations differ.

Section 2

Methodology: OLS

OLS Objective Function

Given the simple regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

OLS Objective Function

Given the simple regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

Our goal in OLS: Find β_0 and β_1 that minimize the sum of squared residuals:

$$\min_{\beta_0, \beta_1} \; S(\beta_0, \beta_1) = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1 X_i\right)^2$$

Is this constrained or unconstrained?

OLS Objective Function

Given the simple regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

Our goal in OLS: Find β_0 and β_1 that minimize the sum of squared residuals:

$$\min_{\beta_0, \beta_1} \; S(\beta_0, \beta_1) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

Is this constrained or unconstrained?

- No explicit restrictions on β_0, β_1.
- $\Rightarrow$ This is an **unconstrained optimization problem**.

How do we solve unconstrained optimization?

OLS Objective Function

Given the simple regression model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

Our goal in OLS: Find β_0 and β_1 that minimize the sum of squared residuals:

$$\min_{\beta_0, \beta_1} S(\beta_0, \beta_1) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

Is this constrained or unconstrained?

- No explicit restrictions on β_0, β_1.
- $\Rightarrow$ This is an **unconstrained optimization problem**.

How do we solve unconstrained optimization?

- Take partial derivatives of $S(\beta_0, \beta_1)$ w.r.t. β_0 and β_1.
- Set them equal to zero (**first-order conditions**).
- Solve the resulting system of equations.

OLS: First-Order Conditions

Model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \ldots, n$$

Objective (sum of squared residuals):

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} \big(Y_i - \beta_0 - \beta_1 X_i\big)^2$$

OLS: First-Order Conditions

Model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \ldots, n$$

Objective (sum of squared residuals):

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} \big(Y_i - \beta_0 - \beta_1 X_i\big)^2$$

Compute partial derivatives:

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^{n} \big(Y_i - \beta_0 - \beta_1 X_i\big) \stackrel{!}{=} 0 \quad \Rightarrow \quad \sum_{i=1}^{n} Y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} X_i$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^{n} X_i \big(Y_i - \beta_0 - \beta_1 X_i\big) \stackrel{!}{=} 0 \quad \Rightarrow \quad \sum_{i=1}^{n} X_i Y_i = \beta_0 \sum_{i=1}^{n} X_i + \beta_1 \sum_{i=1}^{n} X_i^2$$

OLS: First-Order Conditions

Model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, \dots, n$$

Objective (sum of squared residuals):

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} \big(Y_i - \beta_0 - \beta_1 X_i\big)^2$$

Compute partial derivatives:

$$\frac{\partial S}{\partial \beta_0} = -2 \sum_{i=1}^{n} \big(Y_i - \beta_0 - \beta_1 X_i\big) \stackrel{!}{=} 0 \quad \Rightarrow \quad \sum_{i=1}^{n} Y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} X_i$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum_{i=1}^{n} X_i \big(Y_i - \beta_0 - \beta_1 X_i\big) \stackrel{!}{=} 0 \quad \Rightarrow \quad \sum_{i=1}^{n} X_i Y_i = \beta_0 \sum_{i=1}^{n} X_i + \beta_1 \sum_{i=1}^{n} X_i^2$$

These are the **normal equations**. Next: solve for β_0, β_1.

OLS: Closed-Form Estimators

Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

From the normal equations, solve for (β_0, β_1):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad = \quad \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

OLS: Closed-Form Estimators

Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

From the normal equations, solve for (β_0, β_1):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad = \quad \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Equivalent “raw-sum” form:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - n\,\bar{X}\,\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\,\bar{X}^2}, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

OLS: Closed-Form Estimators

Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

From the normal equations, solve for (β_0, β_1):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \quad = \quad \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Equivalent "raw-sum" form:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - n\,\bar{X}\,\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\,\bar{X}^2}, \qquad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Interpretation: Slope $\hat{\beta}_1$ scales how much Y co-moves with X (Cov) per unit variability in X (Var).

Practice

Let's compute $\hat{\beta}_0$ and $\hat{\beta}_1$ by hand.

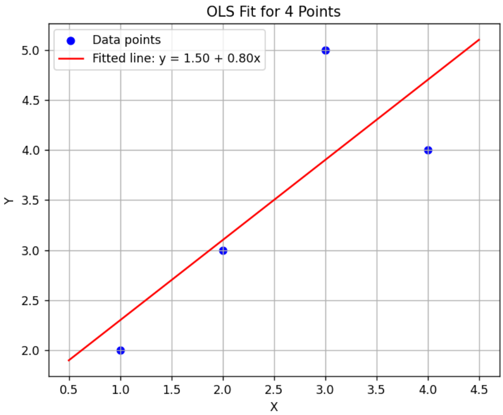
X_i	Y_i
1	2
2	3
3	5
4	4

Tasks:

1. Compute $\bar{X}$ and $\bar{Y}$.
2. Use the formula:
$$\hat{\beta}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$
3. Compute $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$.

OLS: Visualizing the Fit

We can visualize the data points and the fitted regression line:



This confirms our hand calculation: the red line is the **best linear fit** minimizing the sum of squared residuals.

Section 3

Methodology: ML

Review: Maximum Likelihood Estimation (MLE)

Idea: Choose parameter values that make the observed sample **most likely**.

Review: Maximum Likelihood Estimation (MLE)

Idea: Choose parameter values that make the observed sample **most likely**.
Step-by-step:

1. **Specify** the probability distribution of the data, with parameters θ.
2. **Write** the likelihood function:

$$L(\theta) = \prod_{i=1}^{n} f(Y_i \mid \theta)$$

   where $f(\cdot)$ is the PDF of the model.
3. **Maximize** $L(\theta)$ (or $\ell(\theta) = \log L(\theta)$) with respect to θ.

Review: Maximum Likelihood Estimation (MLE)

Idea: Choose parameter values that make the observed sample **most likely**.
Step-by-step:

1. **Specify** the probability distribution of the data, with parameters θ.
2. **Write** the likelihood function:

$$L(\theta) = \prod_{i=1}^{n} f(Y_i \mid \theta)$$

   where $f(\cdot)$ is the PDF of the model.

3. **Maximize** $L(\theta)$ (or $\ell(\theta) = \log L(\theta)$) with respect to θ.

Next: Apply MLE to the linear regression model and compare with OLS.

MLE for Linear Regression

At first glance, there seems to be no PDF for us to use here. So, what can we do?

MLE for Linear Regression

At first glance, there seems to be no PDF for us to use here. So, what can we do?

Idea: We can not directly assume a distribution for Y_i itself. Instead, we make an assumption about the distribution of the **error term** ε_i.

MLE for Linear Regression

At first glance, there seems to be no PDF for us to use here. So, what can we do?

Idea: We can not directly assume a distribution for Y_i itself. Instead, we make an assumption about the distribution of the **error term** ε_i.

Assumption:

$$\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

- Mean $\mu_\varepsilon = 0$ — ensures **unbiasedness**:

$$E[Y_i \mid X_i] = \beta_0 + \beta_1 X_i$$

- Variance σ^2 — measures the spread around the regression line.

MLE for Linear Regression

At first glance, there seems to be no PDF for us to use here. So, what can we do?

Idea: We can not directly assume a distribution for Y_i itself. Instead, we make an assumption about the distribution of the **error term** ε_i.

Assumption:

$$\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

- Mean $\mu_\varepsilon = 0$ — ensures **unbiasedness**:

$$E[Y_i \mid X_i] = \beta_0 + \beta_1 X_i$$

- Variance σ^2 — measures the spread around the regression line.

From the model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

we get the conditional distribution:

$$Y_i \mid X_i \sim \mathcal{N}(\beta_0 + \beta_1 X_i, \sigma^2)$$

Now we have a **PDF** to build the likelihood function for MLE.

MLE for Linear Regression (Normal Errors)

Model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

Then the conditional density:

$$f(Y_i \mid X_i; \beta_0, \beta_1, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2} \right)$$

MLE for Linear Regression (Normal Errors)

Model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

Then the conditional density:

$$f(Y_i \mid X_i; \beta_0, \beta_1, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2}\right)$$

Likelihood & Log-likelihood:

$$L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} f(Y_i \mid X_i)$$

$$\ell(\beta_0, \beta_1, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 X_i)^2$$

MLE for Linear Regression (Normal Errors)

Model:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$$

Then the conditional density:

$$f(Y_i \mid X_i; \beta_0, \beta_1, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{2\sigma^2} \right)$$

Likelihood & Log-likelihood:

$$L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} f(Y_i \mid X_i)$$

$$\ell(\beta_0, \beta_1, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

Key result: For fixed σ^2, maximizing $\ell \iff$ minimizing

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_i)^2$$

$\Rightarrow \hat{\beta}$ by MLE $= \hat{\beta}$ by OLS.

MLE for Linear Regression (Normal Errors)

Estimators:

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\varepsilon}_i^2 \quad \big(\text{note: unbiased estimator uses } \tfrac{1}{n-1-1}\big)$$

MLE/OLS in Matrix Form (Simple Regression)

Model (matrix form):

$$\underbrace{Y}_{n\times 1} = \underbrace{X}_{n\times 2}\underbrace{\beta}_{2\times 1} + \underbrace{\varepsilon}_{n\times 1}, \qquad \varepsilon \sim \mathcal{N}(0,\ \sigma^2 I_n).$$

What do Y, X, and β look like?

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} \mathbf{1} & X \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

MLE/OLS in Matrix Form (Simple Regression)

Model (matrix form):

$$\underbrace{Y}_{n\times 1} = \underbrace{X}_{n\times 2}\underbrace{\beta}_{2\times 1} + \underbrace{\varepsilon}_{n\times 1}, \qquad \varepsilon \sim \mathcal{N}(0,\ \sigma^2 I_n).$$

What do Y, X, and β look like?

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} \mathbf{1} & X \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

Log-likelihood (Normal errors):

$$\ell(\beta, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y - X\beta)^\top (Y - X\beta).$$

MLE/OLS in Matrix Form (Simple Regression)

Model (matrix form):

$$\underbrace{Y}_{n\times 1} = \underbrace{X}_{n\times 2}\underbrace{\beta}_{2\times 1} + \underbrace{\varepsilon}_{n\times 1}, \qquad \varepsilon \sim \mathcal{N}(0,\ \sigma^2 I_n).$$

What do Y, X, and β look like?

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} \mathbf{1} & X \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}.$$

Log-likelihood (Normal errors):

$$\ell(\beta, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(Y - X\beta)^\top (Y - X\beta).$$

Estimators (MLE = OLS for β):

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}, \qquad \hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}(Y - X\hat{\beta})^\top (Y - X\hat{\beta}).$$

$$\text{(Unbiased variance: } \hat{\sigma}^2_{\text{unb}} = \frac{1}{n-2}(Y - X\hat{\beta})^\top (Y - X\hat{\beta})\text{)}.$$

Interpreting and Estimating σ^2

What does σ^2 represent?

- The **variance of the error term** ε_i.
- Measures how far Y_i typically deviates from the regression line.
- Economically: captures unobserved factors affecting Y that are not explained by X.

Interpreting and Estimating σ^2

What does σ^2 represent?

- The **variance of the error term** ε_i.
- Measures how far Y_i typically deviates from the regression line.
- Economically: captures unobserved factors affecting Y that are not explained by X.

Why estimate it this way?

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\varepsilon}_i^2$$

- Comes directly from maximizing the normal log-likelihood.
- Dividing by n treats parameters β_0, β_1 as **known** (as in MLE theory).

Why the Unbiased Estimator Uses $n-2$*

Model: $Y = \beta_0 + \beta_1 X + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$. Let $\widehat{\varepsilon} = Y - \widehat{\beta}_0 \mathbf{1} - \widehat{\beta}_1 X$ be the OLS residuals.

Why the Unbiased Estimator Uses $n-2$*

Model: $Y = \beta_0 + \beta_1 X + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$. Let $\widehat{\varepsilon} = Y - \widehat{\beta}_0 \mathbf{1} - \widehat{\beta}_1 X$ be the OLS residuals.

Matrix/probabilistic view (formal)

- Write $X = [\mathbf{1}\ X]$ $(n \times 2)$, hat matrix $H = X(X'X)^{-1}X'$, residual maker $M = I_n - H$.
- $\widehat{\varepsilon} = MY$, and M is symmetric idempotent with $\operatorname{rank}(M) = n-2$.
- Since $Y \sim \mathcal{N}(X\beta, \sigma^2 I_n)$, we have

$$\frac{1}{\sigma^2}\widehat{\varepsilon}'\widehat{\varepsilon} = \frac{1}{\sigma^2} Y' M Y \ \sim\ \chi^2_{\operatorname{rank}(M)} = \chi^2_{n-2}.$$

- Therefore $E[\widehat{\varepsilon}'\widehat{\varepsilon}] = (n-2)\sigma^2 \ \Rightarrow\ \hat{\sigma}^2_{\text{unbiased}} = \dfrac{1}{n-2}\widehat{\varepsilon}'\widehat{\varepsilon}.$

Section 4

Hypothesis Testing

What Do Economists Care About in Hypothesis Testing?

After we estimate the regression model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

and obtain $\hat{\beta}_1$, the next big question is:

Is $\beta_1 \neq 0$?

What Do Economists Care About in Hypothesis Testing?

After we estimate the regression model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

and obtain $\hat{\beta}_1$, the next big question is:

$$\textbf{Is } \beta_1 \neq 0\,?$$

Why is this important?

- If $\beta_1 = 0$: X has no linear effect on Y (in our model).
- If $\beta_1 \neq 0$: X has a statistically detectable linear impact on Y.

What Do Economists Care About in Hypothesis Testing?

After we estimate the regression model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

and obtain $\hat{\beta}_1$, the next big question is:

$$\textbf{Is } \beta_1 \neq 0\,?$$

Why is this important?

- If $\beta_1 = 0$: X has no linear effect on Y (in our model).
- If $\beta_1 \neq 0$: X has a statistically detectable linear impact on Y.

Formal statement:

$$H_0 : \beta_1 = 0 \quad \text{vs.} \quad H_1 : \beta_1 \neq 0$$

What Do Economists Care About in Hypothesis Testing?

After we estimate the regression model:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

and obtain $\hat{\beta}_1$, the next big question is:

$$\textbf{Is } \beta_1 \neq 0 \text{ ?}$$

Why is this important?

- If $\beta_1 = 0$: X has no linear effect on Y (in our model).
- If $\beta_1 \neq 0$: X has a statistically detectable linear impact on Y.

Formal statement:

$$H_0 : \beta_1 = 0 \quad \text{vs.} \quad H_1 : \beta_1 \neq 0$$

Next: How do we use our sample data to decide whether to reject H_0? We need a **test statistic** and a **decision rule**.

Basic Idea of Hypothesis Testing

Step 0: Formulate hypotheses.

$$H_0 : \text{“Our hypothesis” (no effect, equal 1, etc.)}$$

$$H_1 : \text{“Alternative hypothesis” (effect exists, not equal to 1, etc.)}$$

Basic Idea of Hypothesis Testing

Step 0: Formulate hypotheses.

$$H_0 : \text{“Our hypothesis” (no effect, equal 1, etc.)}$$

$$H_1 : \text{“Alternative hypothesis” (effect exists, not equal to 1, etc.)}$$

Step 1: Assume H_0 is true.

- Under H_0, the test statistic has a known sampling distribution (e.g., normal, t, χ^2, F).

Basic Idea of Hypothesis Testing

Step 0: Formulate hypotheses.

$$H_0 : \text{“Our hypothesis” (no effect, equal 1, etc.)}$$

$$H_1 : \text{“Alternative hypothesis” (effect exists, not equal to 1, etc.)}$$

Step 1: Assume H_0 is true.

- Under H_0, the test statistic has a known sampling distribution (e.g., normal, t, χ^2, F).

Step 2: Calculate the test statistic from the data.

Basic Idea of Hypothesis Testing

Step 0: Formulate hypotheses.

$$H_0 : \text{“Our hypothesis” (no effect, equal 1, etc.)}$$

$$H_1 : \text{“Alternative hypothesis” (effect exists, not equal to 1, etc.)}$$

Step 1: Assume H_0 is true.

- Under H_0, the test statistic has a known sampling distribution (e.g., normal, t, χ^2, F).

Step 2: Calculate the test statistic from the data.
Step 3: Locate this statistic in the reference distribution.

- If it falls near the center, the data are consistent with H_0.
- If it falls far into the tails, the data are unlikely under H_0.

Basic Idea of Hypothesis Testing

Step 0: Formulate hypotheses.

$$H_0 : \text{“Our hypothesis” (no effect, equal 1, etc.)}$$

$$H_1 : \text{“Alternative hypothesis” (effect exists, not equal to 1, etc.)}$$

Step 1: Assume H_0 is true.

- Under H_0, the test statistic has a known sampling distribution (e.g., normal, t, χ^2, F).

Step 2: Calculate the test statistic from the data.
Step 3: Locate this statistic in the reference distribution.

- If it falls near the center, the data are consistent with H_0.
- If it falls far into the tails, the data are unlikely under H_0.

Step 4: Make a decision.

- If the statistic is too far from the center (beyond a critical value), **reject** H_0.
- Otherwise, **fail to reject** H_0.

Basic Idea of Hypothesis Testing

Step 1: Assume the null hypothesis is true.

$$H_0 : \beta_1 = 0$$

If H_0 holds and errors are normal, the sampling distribution of our test statistic will follow a known distribution (e.g., t-distribution).

Basic Idea of Hypothesis Testing

Step 1: Assume the null hypothesis is true.

$$H_0 : \beta_1 = 0$$

If H_0 holds and errors are normal, the sampling distribution of our test statistic will follow a known distribution (e.g., t-distribution).

Step 2: Compute the test statistic from the data.

$$t = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)}$$

Under H_0: $t \sim t_{n-2}$.

Basic Idea of Hypothesis Testing

Step 1: Assume the null hypothesis is true.

$$H_0 : \beta_1 = 0$$

If H_0 holds and errors are normal, the sampling distribution of our test statistic will follow a known distribution (e.g., t-distribution).

Step 2: Compute the test statistic from the data.

$$t = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)}$$

Under H_0: $t \sim t_{n-2}$.

Step 3: Compare where our statistic falls in this distribution.

- If t is close to the center (around 0), data is consistent with H_0.
- If t is far from 0 (in the tails), data is unlikely under H_0.

Basic Idea of Hypothesis Testing

Step 1: Assume the null hypothesis is true.

$$H_0 : \beta_1 = 0$$

If H_0 holds and errors are normal, the sampling distribution of our test statistic will follow a known distribution (e.g., t-distribution).

Step 2: Compute the test statistic from the data.

$$t = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)}$$

Under H_0: $t \sim t_{n-2}$.

Step 3: Compare where our statistic falls in this distribution.

- If t is close to the center (around 0), data is consistent with H_0.
- If t is far from 0 (in the tails), data is unlikely under H_0.

Step 4: Decision rule.

- If $|t| > t_{\alpha/2,\, n-2}$, **reject** H_0 at significance level α.
- Otherwise, **fail to reject** H_0.

Standard Error of $\hat{\beta}_1$ in Simple Regression

Model: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,\ i = 1, \ldots, n$, with $\mathrm{Var}(\varepsilon_i) = \sigma^2$.

Definition and closed form

$$\mathrm{Var}(\hat{\beta}_1 \mid X) = \frac{\sigma^2}{S_{xx}} \quad \Rightarrow \quad \mathrm{SE}(\hat{\beta}_1) = \sqrt{\frac{\sigma^2}{S_{xx}}}, \qquad S_{xx} = \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Since σ^2 is unknown, estimate it by the residual variance:

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2, \qquad \hat{\varepsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

$$\boxed{\mathrm{SE}(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}}$$

Standard Error of $\hat{\beta}_1$ in Simple Regression

Model: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,\ i = 1, \ldots, n$, with $\operatorname{Var}(\varepsilon_i) = \sigma^2$.

Definition and closed form

$$\operatorname{Var}(\hat{\beta}_1 \mid X) = \frac{\sigma^2}{S_{xx}} \quad \Rightarrow \quad \operatorname{SE}(\hat{\beta}_1) = \sqrt{\frac{\sigma^2}{S_{xx}}}, \qquad S_{xx} = \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Since σ^2 is unknown, estimate it by the residual variance:

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2, \qquad \hat{\varepsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

$$\boxed{\operatorname{SE}(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}}$$

Note: With homoskedastic normal errors, $t = \dfrac{\hat{\beta}_1 - \beta_1}{\operatorname{SE}(\hat{\beta}_1)} \sim t_{n-2}$.

Why the t-statistic has a t-distribution*

Setup (normal linear model).

$$Y \sim \mathcal{N}(X\beta,\ \sigma^2 I_n), \qquad \hat{\beta} = (X'X)^{-1}X'Y, \qquad p = \text{number of parameters.}$$

Why the t-statistic has a t-distribution*

Setup (normal linear model).

$$Y \sim \mathcal{N}(X\beta,\ \sigma^2 I_n), \qquad \hat{\beta} = (X'X)^{-1}X'Y, \qquad p = \text{number of parameters.}$$

1) Numerator is normal.

$$\hat{\beta} \sim \mathcal{N}\big(\beta,\ \sigma^2 (X'X)^{-1}\big) \ \Rightarrow\ \frac{\hat{\beta}_j - \beta_j}{\sigma\sqrt{(X'X)^{-1}_{jj}}} \ \sim\ \mathcal{N}(0,1).$$

Why the t-statistic has a t-distribution*

Setup (normal linear model).

$$Y \sim \mathcal{N}(X\beta,\ \sigma^2 I_n), \qquad \hat{\beta} = (X'X)^{-1}X'Y, \qquad p = \text{number of parameters.}$$

1) Numerator is normal.

$$\hat{\beta} \sim \mathcal{N}\big(\beta,\ \sigma^2(X'X)^{-1}\big) \ \Rightarrow\ \frac{\hat{\beta}_j - \beta_j}{\sigma\sqrt{(X'X)^{-1}_{jj}}} \ \sim\ \mathcal{N}(0,1).$$

2) Denominator involves a chi-square. Let $\hat{\varepsilon} = Y - X\hat{\beta}$, $M = I - H$, $H = X(X'X)^{-1}X'$. Then M is symmetric idempotent with $\text{rank}(M) = n - p$, and

$$\frac{\hat{\varepsilon}'\hat{\varepsilon}}{\sigma^2} = \frac{Y'MY}{\sigma^2} \ \sim\ \chi^2_{n-p}.$$

Why the t-statistic has a t-distribution*

Setup (normal linear model).

$$Y \sim \mathcal{N}(X\beta,\ \sigma^2 I_n), \qquad \hat\beta = (X'X)^{-1}X'Y, \qquad p = \text{number of parameters.}$$

1) Numerator is normal.

$$\hat\beta \sim \mathcal{N}\big(\beta,\ \sigma^2(X'X)^{-1}\big) \ \Rightarrow\ \frac{\hat\beta_j - \beta_j}{\sigma\sqrt{(X'X)^{-1}_{jj}}} \ \sim\ \mathcal{N}(0,1).$$

2) Denominator involves a chi-square. Let $\hat\varepsilon = Y - X\hat\beta$, $M = I - H$, $H = X(X'X)^{-1}X'$. Then M is symmetric idempotent with $\operatorname{rank}(M) = n - p$, and

$$\frac{\hat\varepsilon'\hat\varepsilon}{\sigma^2} = \frac{Y'MY}{\sigma^2}\ \sim\ \chi^2_{n-p}.$$

3) Independence. Because HY and MY are orthogonal projections and Y is multivariate normal,

$$\hat\beta\ \perp\ \hat\varepsilon \quad\Rightarrow\quad \frac{\hat\beta_j - \beta_j}{\sigma\sqrt{(X'X)^{-1}_{jj}}} \text{ is independent of } \frac{(n-p)\hat\sigma^2}{\sigma^2}.$$

Practice: Test $H_0: \beta_1 = 0$ (Simple Regression)

Data: $(X_i, Y_i) = (1, 2),\ (2, 3),\ (3, 5),\ (4, 4) \quad (n = 4)$

Practice: Test $H_0: \ \beta_1 = 0$ (Simple Regression)

Data: $(X_i, Y_i) = (1,2),\ (2,3),\ (3,5),\ (4,4) \quad (n = 4)$
Formulas to use:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$S_{xx} = \sum_{i=1}^{n}(X_i - \bar{X})^2, \quad S_{xy} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\varepsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i, \quad \hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} \hat{\varepsilon}_i^2$$

$$\mathrm{SE}(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$$

$$t\text{-statistic} \quad t = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)} \ \sim \ t_{n-2} = t_2 \ \text{ under } H_0$$

Practice: Test $H_0: \ \beta_1 = 0$ (Simple Regression)

Data: $(X_i, Y_i) = (1,2),\ (2,3),\ (3,5),\ (4,4) \quad (n=4)$
Formulas to use:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$S_{xx} = \sum_{i=1}^{n}(X_i - \bar{X})^2, \quad S_{xy} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\varepsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i, \quad \hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

$$\mathrm{SE}(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}$$

$$t\text{-statistic} \quad t = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)} \ \sim \ t_{n-2} = t_2 \ \text{ under } H_0$$

Let's first compute the t-statistics

Practice: Test $H_0: \ \beta_1 = 0$ (Simple Regression)

Decision (two-sided at level α):

Reject H_0 iff $|t| > t_{\alpha/2,\, n-2}$. Equivalently: p-value $= 2\big(1 - F_{t_{n-2}}(|t|)\big)$.

Practice: Test $H_0: \beta_1 = 0$ (Simple Regression)

Decision (two-sided at level α):

$$\text{Reject } H_0 \text{ iff } |t| > t_{\alpha/2,\,n-2}. \quad \text{Equivalently: } p\text{-value } = 2\big(1 - F_{t_{n-2}}(|t|)\big).$$

Common convention: set $\alpha = 5\%$ (two-sided). Then the rejection rule is

$$|t| > t_{0.025,\,n-2}.$$

Practice: Test $H_0: \ \beta_1 = 0$ (Simple Regression)

Decision (two-sided at level α):

Reject H_0 iff $|t| > t_{\alpha/2,\, n-2}$. Equivalently: p-value $= 2\big(1 - F_{t_{n-2}}(|t|)\big)$.

Common convention: set $\alpha = 5\%$ (two-sided). Then the rejection rule is

$$|t| > t_{0.025,\, n-2}.$$

Critical values at 5% (two-sided):

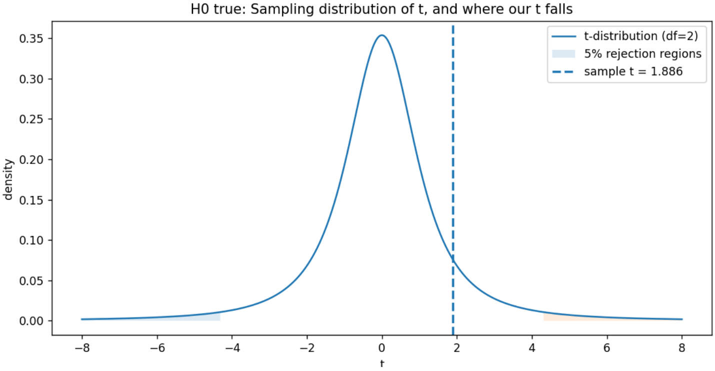
$$t_{0.025,\,2} = 4.303, \quad t_{0.025,\,10} = 2.228, \quad t_{0.025,\,30} = 2.042, \quad t_{0.025,\,\infty} \approx 1.96.$$

*For our example (*n=4 $\Rightarrow$ d.f. $= 2$*): reject if* $|t| > 4.303$.

Assumptions: i.i.d. errors; $E[\varepsilon_i | X_i] = 0$; $\mathrm{Var}(\varepsilon_i | X_i) = \sigma^2$; normality for exact t.

Visualizing the t-test under H_0

Under H_0 (two-sided, $\alpha = 5\%$), we compare our sample t to the t-distribution with $n-2$ d.f. The shaded tails mark the rejection regions at 5%.



Note: For our example, n=4 $\Rightarrow$ d.f. $= 2$, critical values ± 4.303.

Section 5

BLUE Property

What is the BLUE Property?

BLUE = **B**est **L**inear **U**nbiased **E**stimator.

Gauss–Markov Theorem: In the linear model $Y = X\beta + \varepsilon$, if

1. (Linearity & full rank) X has full column rank,
2. (Unbiasedness) $E[\varepsilon \mid X] = 0$,
3. (Spherical errors) $\text{Var}(\varepsilon \mid X) = \sigma^2 I$ (homoskedastic, no autocorrelation),

then the OLS estimator

$$\hat{\beta}_{\text{OLS}} = (X'X)^{-1}X'Y$$

is the **best** (minimum variance) among all **linear unbiased** estimators of β.

What is the BLUE Property?

BLUE = **B**est **L**inear **U**nbiased **E**stimator.

Gauss–Markov Theorem: In the linear model $Y = X\beta + \varepsilon$, if

1. (Linearity & full rank) X has full column rank,
2. (Unbiasedness) $E[\varepsilon \mid X] = 0$,
3. (Spherical errors) $\mathrm{Var}(\varepsilon \mid X) = \sigma^2 I$ (homoskedastic, no autocorrelation),

then the OLS estimator

$$\hat{\beta}_{\text{OLS}} = (X'X)^{-1}X'Y$$

is the **best** (minimum variance) among all **linear unbiased** estimators of β.
Meaning of "best": For any linear unbiased $\tilde{\beta} = AY$ with $AX = I$,

$$\mathrm{Var}(\hat{\beta}_{\text{OLS}} \mid X) \ \preceq \ \mathrm{Var}(\tilde{\beta} \mid X) \quad \text{(Loewner order)},$$

and

$$\mathrm{Var}(\hat{\beta}_{\text{OLS}} \mid X) = \sigma^2 (X'X)^{-1}.$$

What is the BLUE Property?

BLUE = **B**est **L**inear **U**nbiased **E**stimator.

Gauss–Markov Theorem: In the linear model $Y = X\beta + \varepsilon$, if

1. (Linearity & full rank) X has full column rank,
2. (Unbiasedness) $E[\varepsilon \mid X] = 0$,
3. (Spherical errors) $\mathrm{Var}(\varepsilon \mid X) = \sigma^2 I$ (homoskedastic, no autocorrelation),

then the OLS estimator

$$\hat{\beta}_{\mathrm{OLS}} = (X'X)^{-1}X'Y$$

is the **best** (minimum variance) among all **linear unbiased** estimators of β.
Meaning of "best": For any linear unbiased $\tilde{\beta} = AY$ with $AX = I$,

$$\mathrm{Var}(\hat{\beta}_{\mathrm{OLS}} \mid X) \preceq \mathrm{Var}(\tilde{\beta} \mid X) \quad \text{(Loewner order)},$$

and

$$\mathrm{Var}(\hat{\beta}_{\mathrm{OLS}} \mid X) = \sigma^2(X'X)^{-1}.$$

Notes:

- Normality of errors is *not* required for BLUE; it is needed for exact t/F inference and for OLS=MLE.

Sketch Proof of BLUE (Gauss–Markov)*

$Y = X\beta + \varepsilon$, with $E[\varepsilon \mid X] = 0$, $\mathrm{Var}(\varepsilon \mid X) = \sigma^2 I_n$, and X full column rank.

1) OLS is linear and unbiased.

$$\hat{\beta}_{\mathrm{OLS}} = (X'X)^{-1}X'Y \quad \Rightarrow \quad E[\hat{\beta}_{\mathrm{OLS}} \mid X] = (X'X)^{-1}X'X\beta = \beta.$$

Sketch Proof of BLUE (Gauss–Markov)*

$Y = X\beta + \varepsilon$, with $E[\varepsilon \mid X] = 0$, $\mathrm{Var}(\varepsilon \mid X) = \sigma^2 I_n$, and X full column rank.

1) OLS is linear and unbiased.

$$\hat{\beta}_{\mathrm{OLS}} = (X'X)^{-1}X'Y \quad \Rightarrow \quad E[\hat{\beta}_{\mathrm{OLS}} \mid X] = (X'X)^{-1}X'X\beta = \beta.$$

2) Any linear unbiased estimator has the form $\tilde{\beta} = AY$ with $AX = I_p$. Write

$$A = (X'X)^{-1}X' + C \quad \text{where } CX = 0$$

(since $AX - (X'X)^{-1}X'X = I_p - I_p = 0$).

Sketch Proof of BLUE (Gauss–Markov)*

$Y = X\beta + \varepsilon$, with $E[\varepsilon \mid X] = 0$, $\mathrm{Var}(\varepsilon \mid X) = \sigma^2 I_n$, and X full column rank.

1) OLS is linear and unbiased.

$$\hat{\beta}_{\mathrm{OLS}} = (X'X)^{-1}X'Y \quad \Rightarrow \quad E[\hat{\beta}_{\mathrm{OLS}} \mid X] = (X'X)^{-1}X'X\beta = \beta.$$

2) Any linear unbiased estimator has the form $\tilde{\beta} = AY$ with $AX = I_p$. Write

$$A = (X'X)^{-1}X' + C \quad \text{where } CX = 0$$

(since $AX - (X'X)^{-1}X'X = I_p - I_p = 0$).

3) Compare variances.

$$\mathrm{Var}(\tilde{\beta} \mid X) = \sigma^2 AA' = \sigma^2\left[(X'X)^{-1} + CC'\right]$$

because $X'C' = (CX)' = 0$ makes cross terms vanish, and

$$\mathrm{Var}(\hat{\beta}_{\mathrm{OLS}} \mid X) = \sigma^2(X'X)^{-1}.$$

Sketch Proof of BLUE (Gauss–Markov)*

$Y = X\beta + \varepsilon$, with $E[\varepsilon \mid X] = 0$, $\mathrm{Var}(\varepsilon \mid X) = \sigma^2 I_n$, and X full column rank.

1) OLS is linear and unbiased.

$$\hat{\beta}_{\mathrm{OLS}} = (X'X)^{-1}X'Y \quad \Rightarrow \quad E[\hat{\beta}_{\mathrm{OLS}} \mid X] = (X'X)^{-1}X'X\beta = \beta.$$

2) Any linear unbiased estimator has the form $\tilde{\beta} = AY$ with $AX = I_p$. Write

$$A = (X'X)^{-1}X' + C \quad \text{where } CX = 0$$

(since $AX - (X'X)^{-1}X'X = I_p - I_p = 0$).

3) Compare variances.

$$\mathrm{Var}(\tilde{\beta} \mid X) = \sigma^2 AA' = \sigma^2 \left[(X'X)^{-1} + CC'\right]$$

because $X'C' = (CX)' = 0$ makes cross terms vanish, and

$$\mathrm{Var}(\hat{\beta}_{\mathrm{OLS}} \mid X) = \sigma^2 (X'X)^{-1}.$$

4) Loewner ordering. Since $CC' \succeq 0$,

$$\mathrm{Var}(\tilde{\beta} \mid X) \; - \; \mathrm{Var}(\hat{\beta}_{\mathrm{OLS}} \mid X) = \sigma^2 \, CC' \; \succeq \; 0.$$

Equality iff $C = 0 \Rightarrow A = (X'X)^{-1}X'$ (the OLS estimator).

Summary & Course Note

What we covered today

- **Modeling:** $Y = \beta_0 + \beta_1 X + \varepsilon$, with $E[\varepsilon|X] = 0$, $\mathrm{Var}(\varepsilon|X) = \sigma^2$.
- **Estimation (OLS):** minimize $S(\beta) = \sum(Y_i - \beta_0 - \beta_1 X_i)^2$. $\hat{\beta}_1 = \dfrac{S_{xy}}{S_{xx}}$, $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$.
- **Estimation (MLE under normal errors):** same $\hat{\beta}$ as OLS; $\hat{\sigma}^2_{\mathrm{MLE}} = \frac{1}{n}\sum \hat{\varepsilon}_i^2$, unbiased $\hat{\sigma}^2 = \frac{1}{n-2}\sum \hat{\varepsilon}_i^2$.
- **Inference:** $\mathrm{SE}(\hat{\beta}_1) = \sqrt{\hat{\sigma}^2 / S_{xx}}$, $t = \dfrac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)} \sim t_{n-2}$ (under H_0). 5% two-sided: reject if $|t| > t_{0.025,\, n-2}$.
- **BLUE (Gauss–Markov):** with homoskedastic, uncorrelated errors and full-rank X, OLS is the **Best Linear Unbiased Estimator**. Normality not required for BLUE.

Course note: In econometrics class we will consistently use matrix notation. Please be comfortable with matrix algebra (transpose, inverse, rank), projections, and moving between summation and vector forms.

References

- Moss, C. B. (2014). *Mathematical Statistics for Applied Econometrics*. CRC Press.
- Wikipedia contributors. *Various entries on statistical distributions and estimation*. Retrieved from `https://en.wikipedia.org`.
- Figures generated with Python.

Thank You!

Five days (20 hours) have flown by in the blink of an eye. We have explored, learned, questioned, and grown together.

Thank you all for your participation and support! Your feedback is invaluable for improving our future sessions.

Wishing you all the best in your future studies and research!