Simple Linear Regression: Estimation

EC 320: Introduction to Econometrics

Philip Economides
Winter 2022

Prologue

Housekeeping

Problem Set 3 to be released in a week! (Jan 31st)

- Will be focused on simple linear regressions and underlying assumptions
- Will not be due until **after** your midterm exam (Feb 7th)
- Computational component will be using recent lab knowledge

Navigating Metrics

Where are we?

- Reviewed core ideas from statistics
- Developed a framework for thinking about causality
- Dabbled in regression analysis.

Also, **R**.

Navigating Metrics

Where we're going

- Learn the mechanics of *how* OLS works
- Interpret regression results (mechanically and critically)
- Extend ideas about causality to a regression context
- Think more deeply about statistical inference
- Lay a foundation for more-sophisticated regression techniques.

Also, **more R**.

Simple Linear Regression

Addressing Questions

Example: Effect of police on crime

Policy Question: Do on-campus police reduce crime on campus?

- **Empirical Question:** Does the number of on-campus police officers affect campus crime rates? If so, by how much?

How can we answer these questions?

- Prior beliefs.
- Theory.
- **Data!**

Let's "*Look*" at Data

Example: Effect of police on crime

Search:

	Police per 1000 Students	Crimes per 1000 students
1	20.42	1.1
2	0.15	2
3	0.47	1.41
4	14.68	2.06
5	23.75	1.52
6	7.68	2.76

Showing 1 to 6 of 96 entries

Previous Next

Take 2

Example: Effect of police on crime

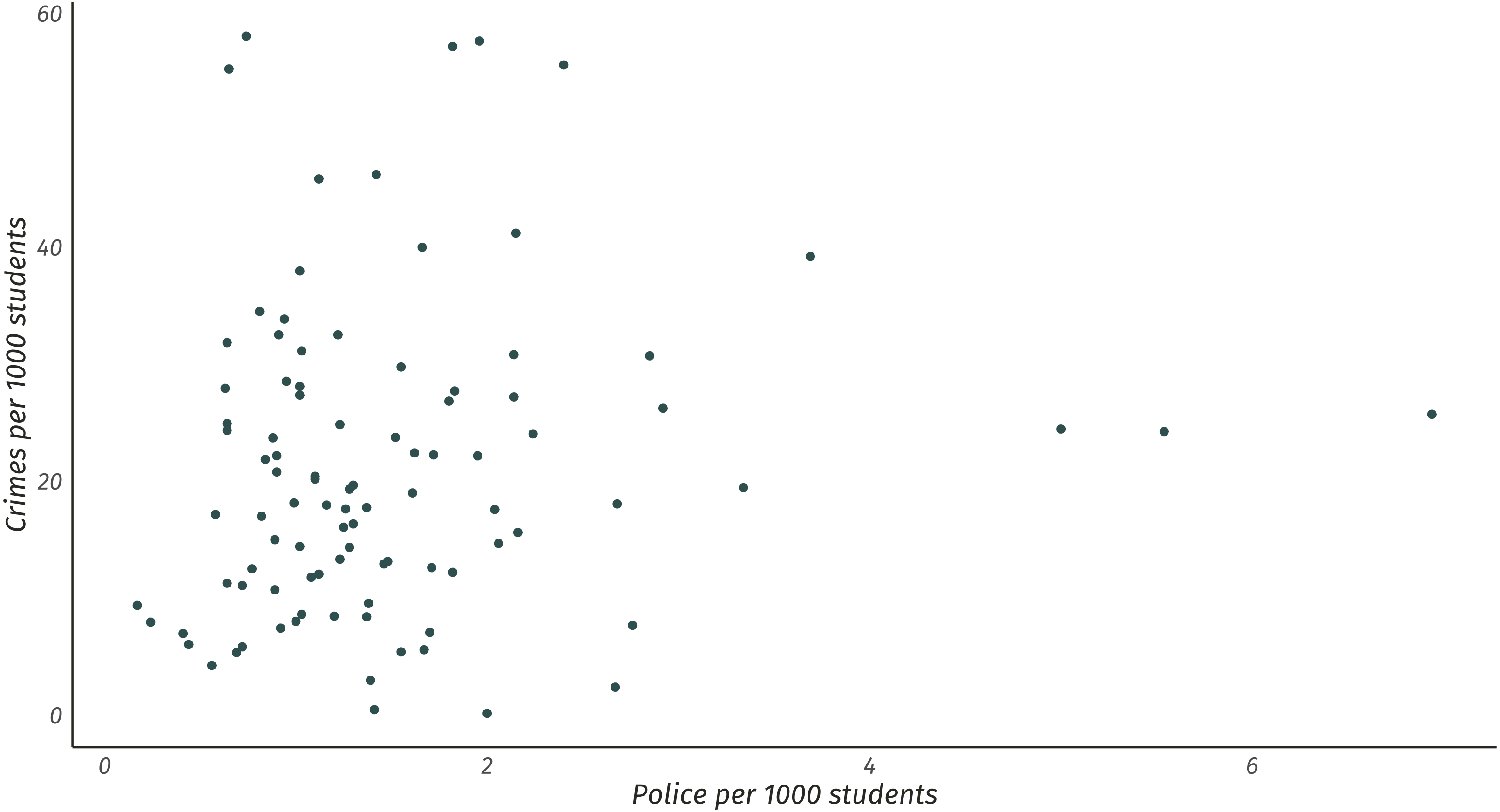
"Looking" at data wasn't especially helpful.

Let's try using a scatter plot.

- Plot each data point in (X, Y)-space.
- `Police` on the X-axis.
- `Crime` on the Y-axis.

Take 2

Example: Effect of police on crime



Take 2

Example: Effect of police on crime

The scatter plot tells us more than the spreadsheet.

- Somewhat weak *positive* relationship.
- Sample correlation coefficient of 0.14 confirms this.

But our question was

> Does the number of on-campus police officers affect campus crime rates? If so, by how much?

- The scatter plot and correlation coefficient provide only a partial answer.

Take 3

Example: Effect of police on crime

Our next step is to estimate a **statistical model.**

To keep it simple, we will relate an **explained variable** Y to an **explanatory variable** X in a linear model.

Simple Linear Regression Model

We express the relationship between a **explained variable** and an **explanatory variable** as linear:

$$Y_i = \beta_1 + \beta_2 X_i + u_i.$$

- β_1 is the **intercept** or constant.
- β_2 is the **slope coefficient**.
- u_i is an **error term** or disturbance term.

Simple = Only one explanatory variable.

Simple Linear Regression Model

The **intercept** tells us the expected value of Y_i when $X_i = 0$.

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Usually not the focus of an analysis.

Simple Linear Regression Model

The **slope coefficient** tells us the expected change in Y_i when X_i increases by one.

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

"A one-unit increase in X_i is associated with a β_2-unit increase in Y_i."

Under certain (strong) assumptions about the error term, β_2 is the *effect of X_i on Y_i*.

- Otherwise, it's the *association of X_i with Y_i*.

Simple Linear Regression Model

The **error term** reminds us that X_i does not perfectly explain Y_i.

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Represents all other factors that explain Y_i.

- Useful mnemonic: pretend that u stands for "*unobserved*" or "*unexplained*."

Take 3, continued

Example: Effect of police on crime

How might we apply the simple linear regression model to our question about the effect of on-campus police on campus crime?

- Which variable is X? Which is Y?

$$\text{Crime}_i = \beta_1 + \beta_2 \text{Police}_i + u_i.$$

- β_1 is the crime rate for colleges without police.
- β_2 is the increase in the crime rate for an additional police officer per 1000 students.

Take 3, continued

Example: Effect of police on crime

How might we apply the simple linear regression model to our question?

$$\mathbf{Crime}_i = \beta_1 + \beta_2 \mathbf{Police}_i + u_i$$

β_1 and β_2 are the population parameters we want, but we cannot observe them.

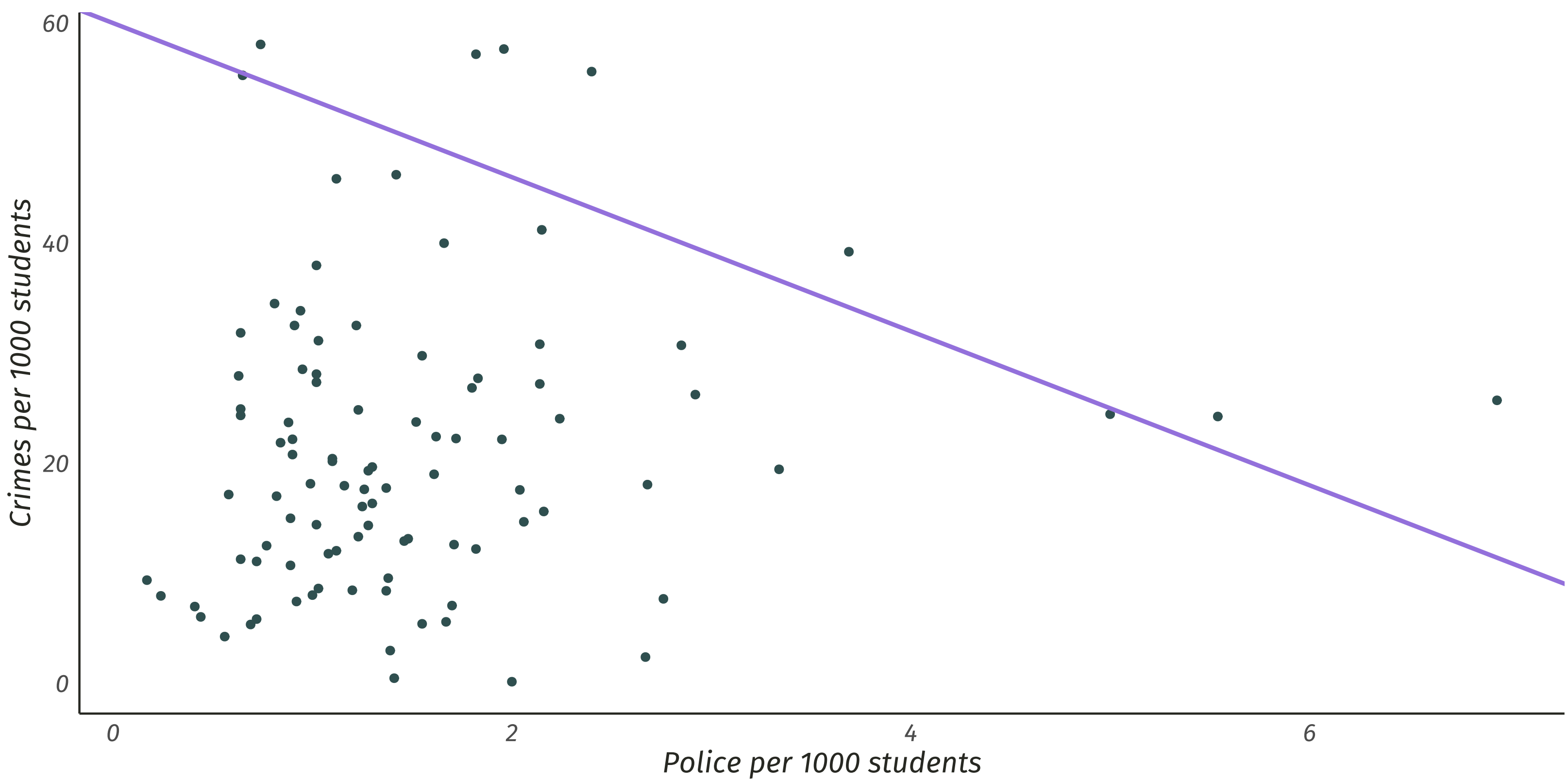
Instead, we must estimate the population parameters.

- $\hat{\beta}_1$ and $\hat{\beta}_2$ generate predictions of $\mathbf{Crime}_i$ called $\widehat{\mathbf{Crime}}_i$.
- We call the predictions of the dependent variable **fitted values.**
- Together, these trace a line: $\widehat{\mathbf{Crime}}_i = \hat{\beta}_1 + \hat{\beta}_2 \mathbf{Police}_i$.

Take 3, attempted

Example: Effect of police on crime

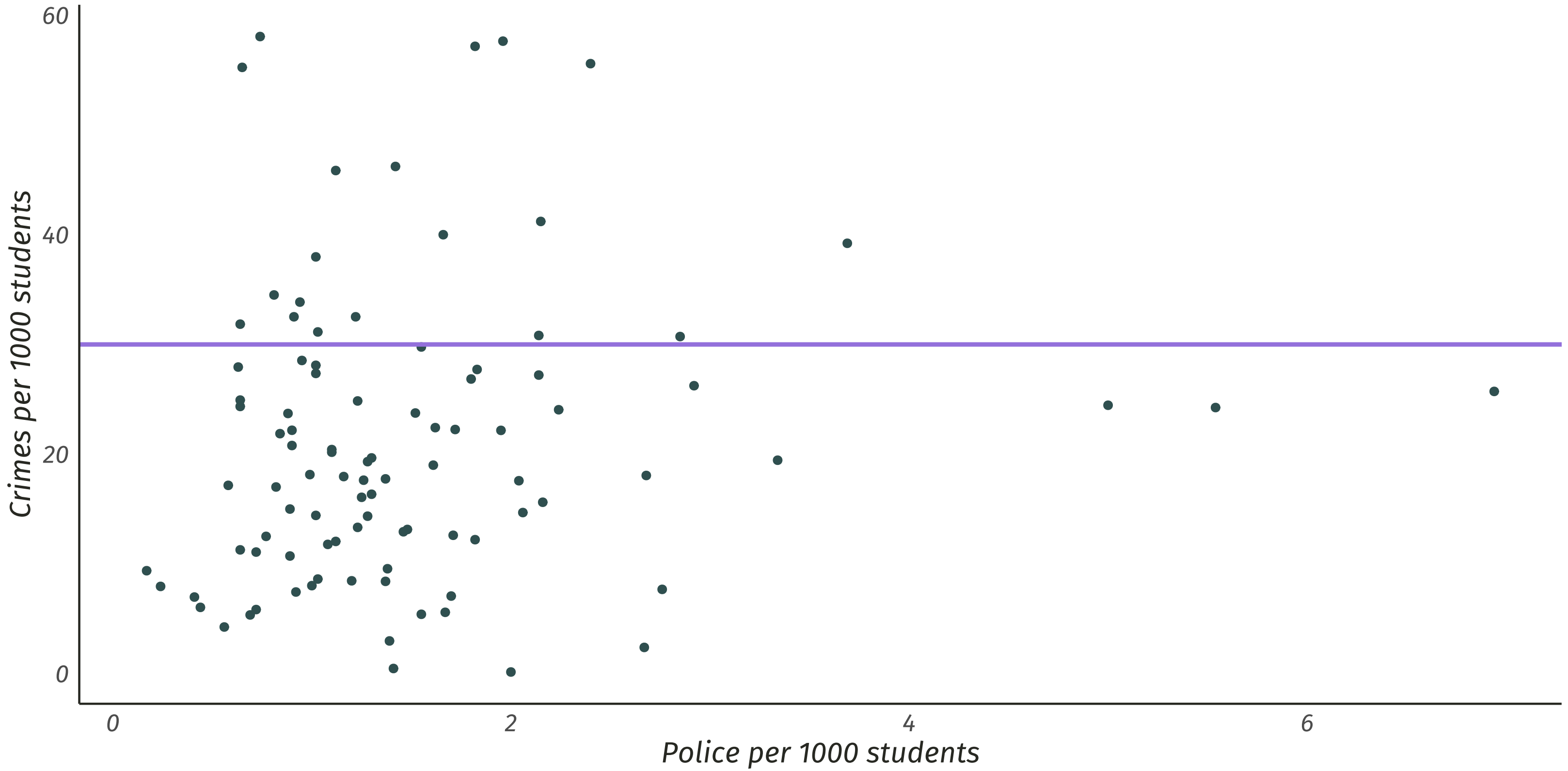
Guess: $\hat{\beta}_1 = \mathbf{60}$ and $\hat{\beta}_2 = -\mathbf{7}$.



Take 4

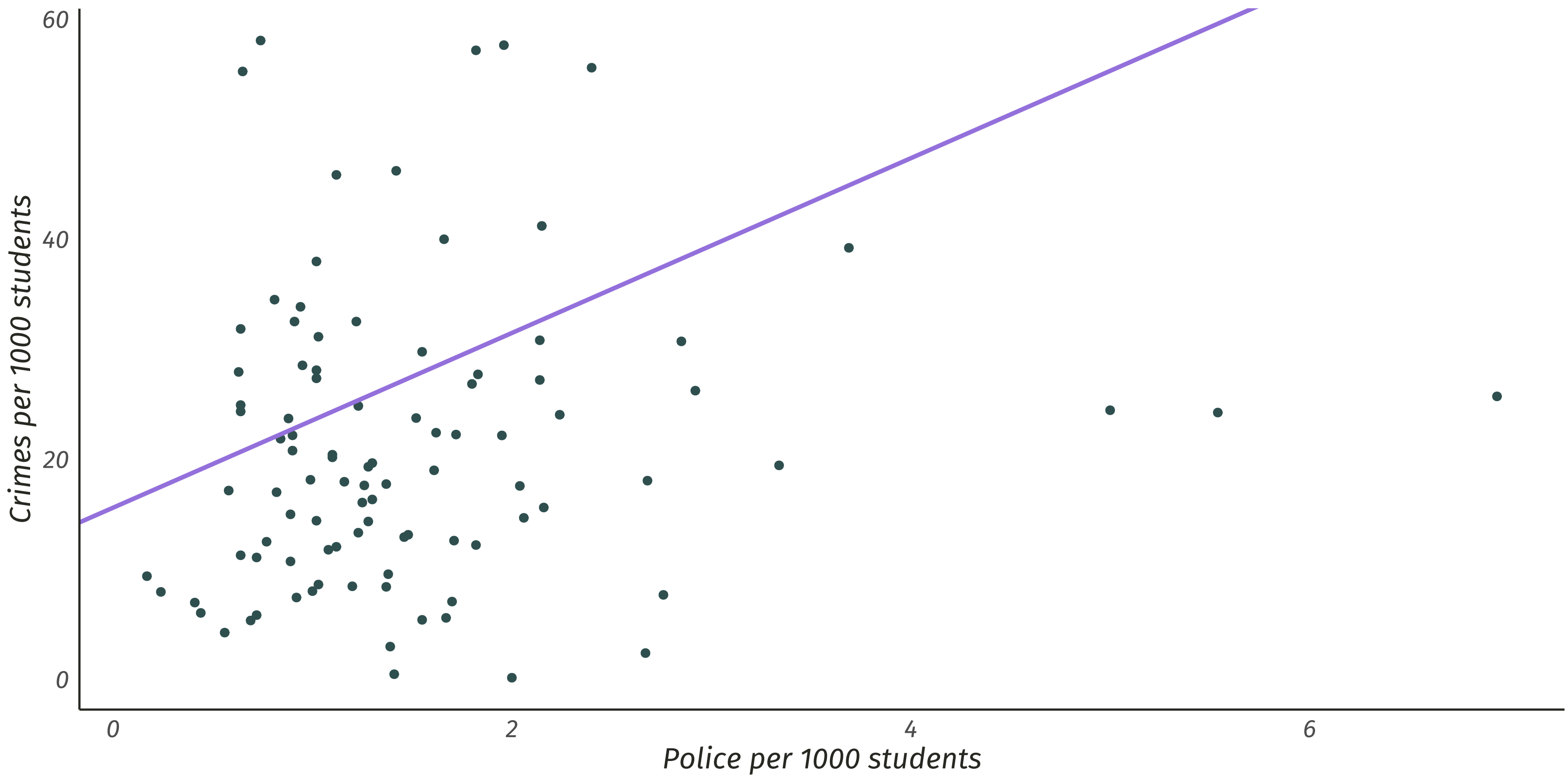
Example: Effect of police on crime

Guess: $\hat{\beta}_1 = \mathbf{30}$ and $\hat{\beta}_2 = \mathbf{0}$.



Example: Effect of police on crime

Guess: $\hat{\beta}_1 = \mathbf{15.6}$ and $\hat{\beta}_2 = \mathbf{7.94}$.



Residuals

Using $\hat{\beta}_1$ and $\hat{\beta}_2$ to make $\hat{Y}_i$ generates misses called **residuals**:

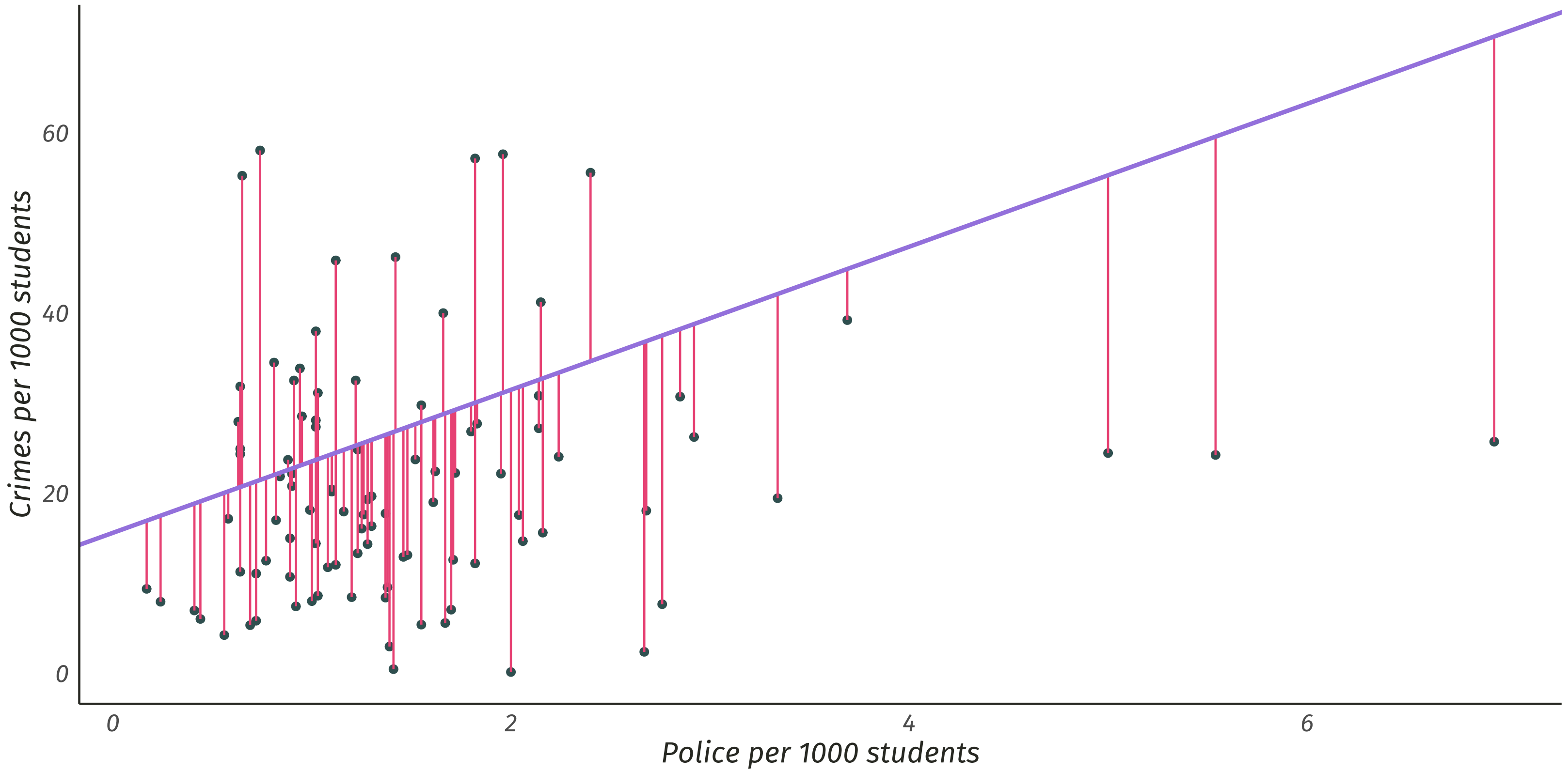
$$\hat{u}_i = Y_i - \hat{Y}_i.$$

- Sometimes called e_i.

Residuals

Example: Effect of police on crime

Using $\hat{\beta}_1 = \mathbf{15.6}$ and $\hat{\beta}_2 = \mathbf{7.94}$ to make $\hat{\text{Crime}}_i$ generates **residuals**.



Residuals

We want an estimator that makes fewer big misses.

Why not minimize $\sum_{i=1}^{n} \hat{u}_i$?

- There are positive *and* negative residuals $\implies$ no solution (can always find a line with more negative residuals).

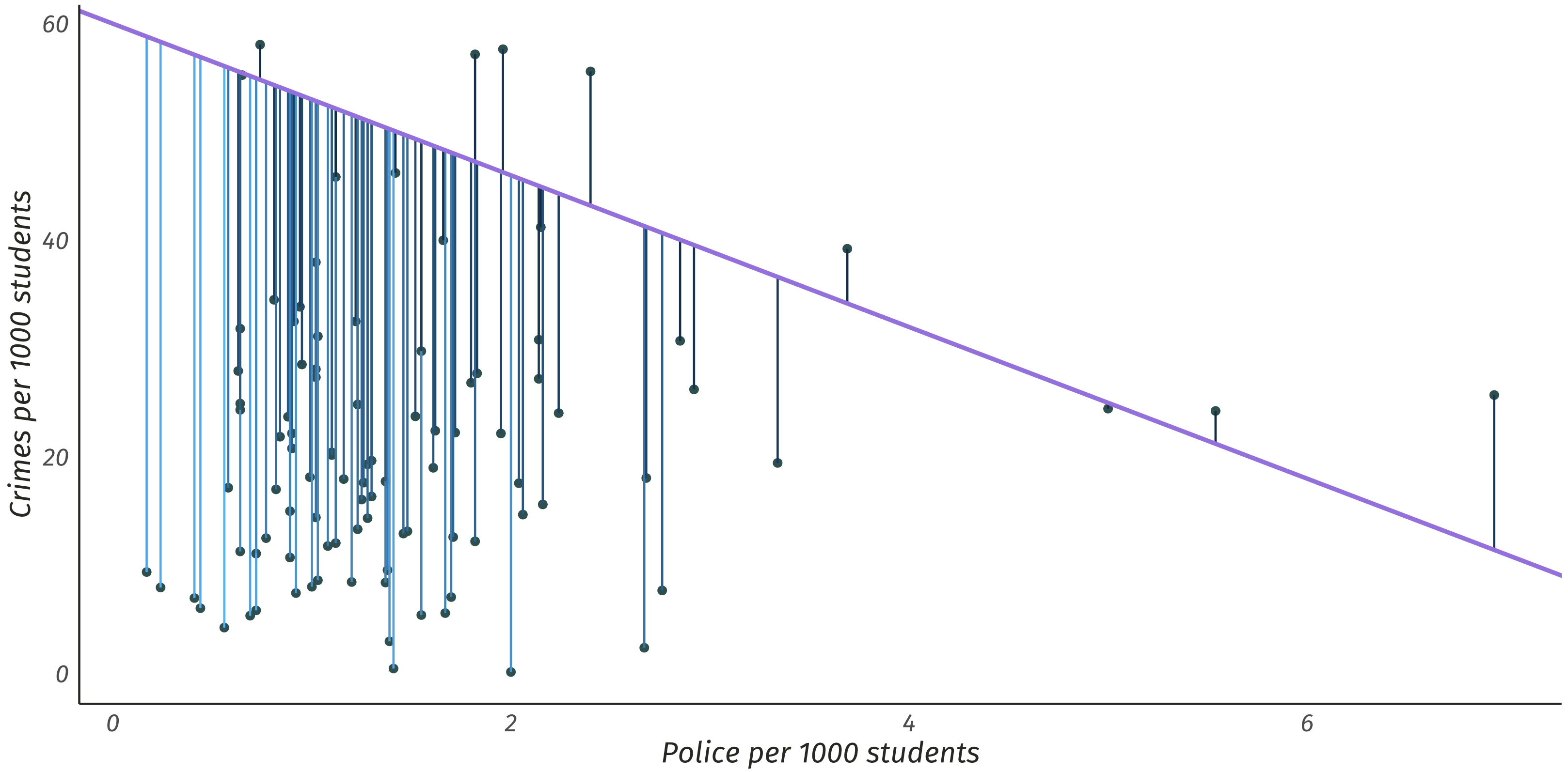
Alternative: Minimize the sum of squared residuals a.k.a. the **residual sum of squares (RSS)**.

- Squared numbers are never negative.

Residuals

Example: Effect of police on crime

RSS gives bigger penalties to bigger residuals.



Residuals

Minimizing RSS

We could test thousands of guesses of $\hat{\beta}_1$ and $\hat{\beta}_2$ and pick the pair that minimizes RSS.

- Or we just do a little math and derive some useful formulas that give us RSS-minimizing coefficients without the guesswork.

Ordinary Least Squares (OLS)

OLS

The **OLS estimator** chooses the parameters $\hat{\beta}_1$ and $\hat{\beta}_2$ that minimize the **residual sum of squares (RSS)**:

$$\min_{\hat{\beta}_1,\, \hat{\beta}_2} \quad \sum_{i=1}^{n} \hat{u}_i^2$$

This is why we call the estimator ordinary **least squares.**

OLS Formulas

For details, see the handout posted on Canvas.

Slope coefficient

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Intercept

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

Slope coefficient

The slope estimator is equal to the sample covariance divided by the sample variance of X:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

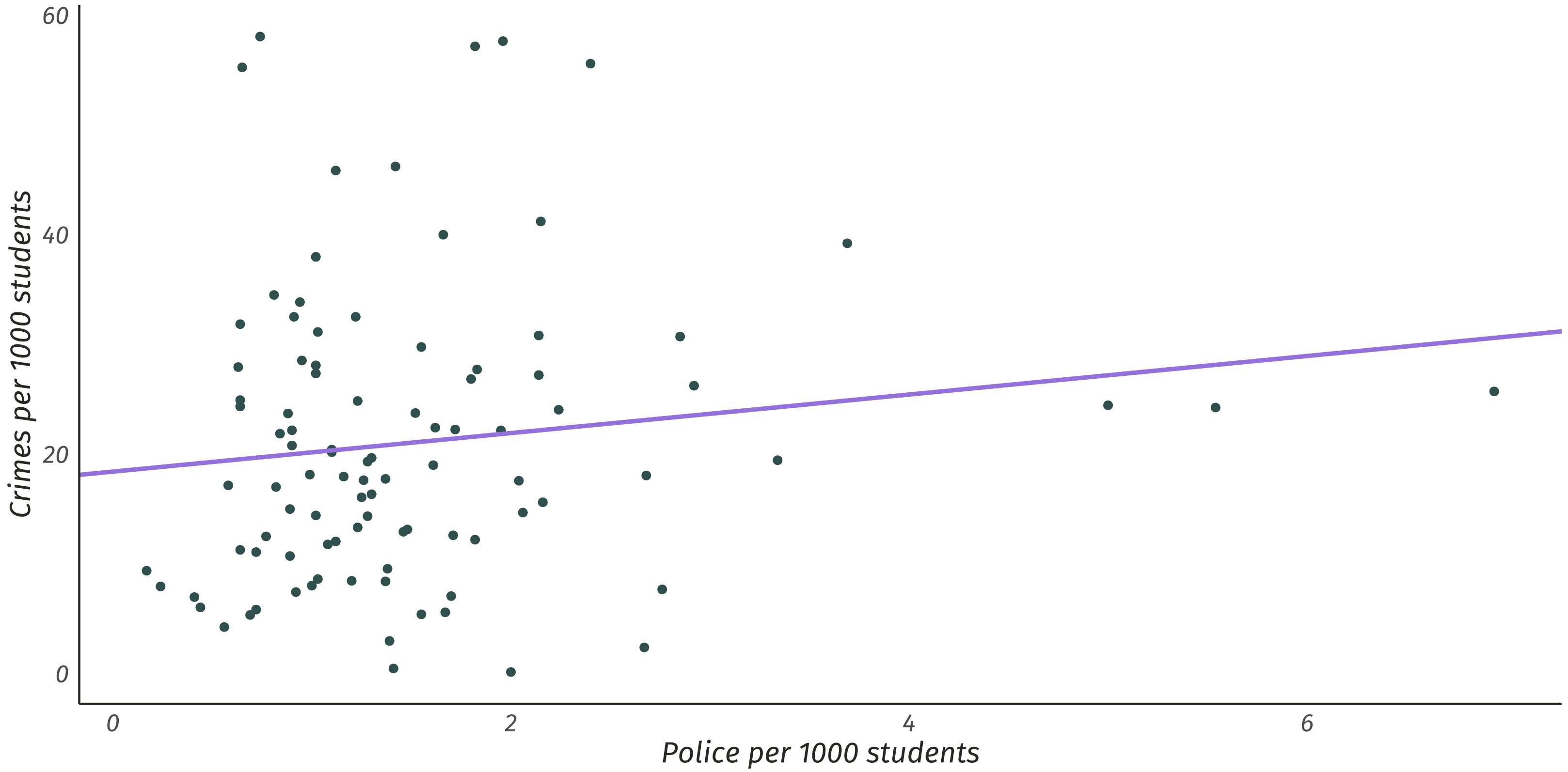
$$= \frac{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$= \frac{S_{XY}}{S_X^2}.$$

Take 6

Example: Effect of police on crime

Using the OLS formulas, we get $\hat{\beta}_1$ = 18.41 and $\hat{\beta}_2$ = 1.76.



Coefficient Interpretation

Example: Effect of police on crime

Using OLS gives us the fitted line

$$\hat{\text{Crime}}_i = \hat{\beta}_1 + \hat{\beta}_2 \text{Police}_i.$$

What does $\hat{\beta}_1$ = 18.41 tell us?

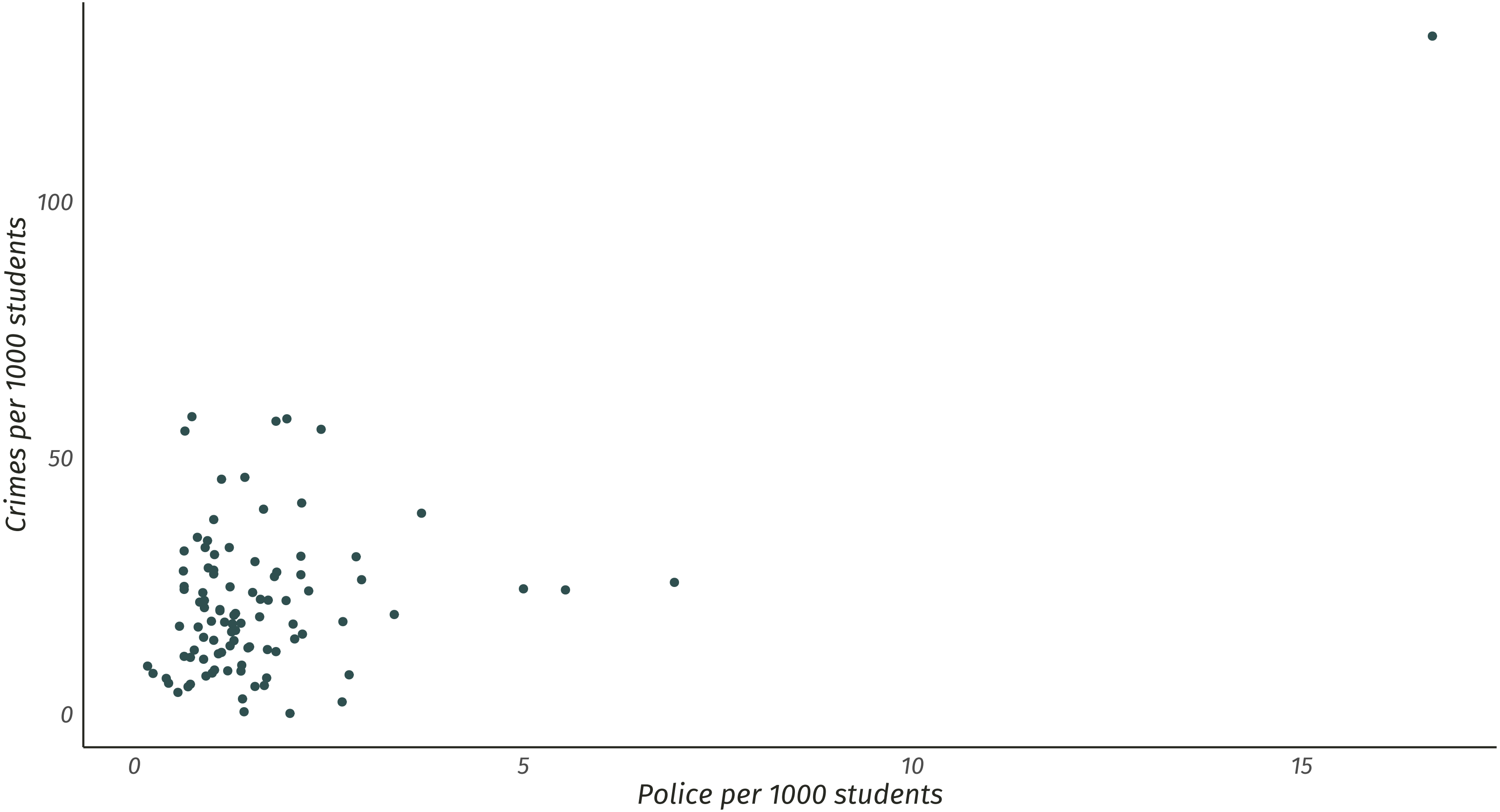
What does $\hat{\beta}_2$ = 1.76 tell us?

Gut check: Does this mean that police *cause* crime?

- Probably not. **Why?**

Outliers

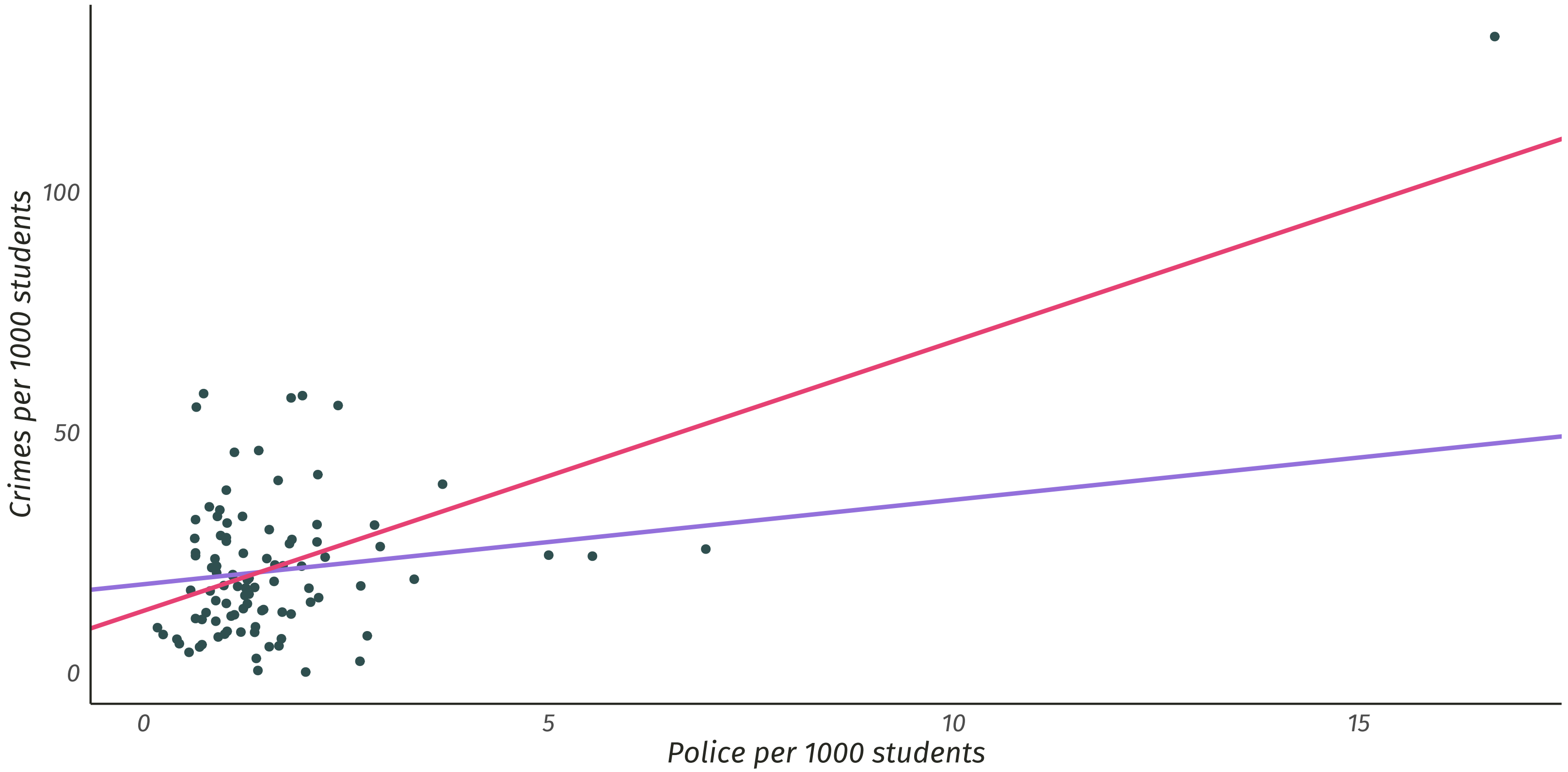
Example: Association of police with crime



Outliers

Example: Association of police with crime

Fitted line without outlier. **Fitted line** with outlier.



OLS Application

Suppose we do not yet have an empirical question, but wish to observe the mechanics involved in generating parameter estimates.

Consider the following **mini sample** $\{X, Y\}$ data points:

Example: $n = 4$

i	x_i	y_i
1	1	4
2	2	3
3	3	5
4	4	8

Regression Model:
$\boldsymbol{Y_i = \beta_1 + \beta_2 X_i + u_i}$
Fitted Line: $\boldsymbol{\hat{Y}_i = b_1 + b_2 X_i}$

Lets calculate the estimated parameters $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$ using the OLS estimator

OLS Application

Recall that OLS focuses on minimizing the RSS. We will take four steps.

1. Calculate the residuals, $\hat{u}_i = Y_i - \hat{Y}_i$
2. Summate the squared residuals, $RSS = \sum_{i=1}^{n} \hat{u}_i$
3. Differentiate for $\frac{\partial RSS}{\partial b_j}$ such that our number of unknown parameters is equal to the number of partial differentiation equations
4. Solve for the unknown parameters

We'll use the **mini sample** to get an idea of the mechanics involved. Given larger datasets and more covariates, **R** comes to the rescue.

Warning: Check the second derivatives to ensure minimization of the functions, where all the second-order partial derivatives are greater than zero.

OLS Application

Step 1: Calculate the residuals

$$\hat{u}_1 = Y_1 - \hat{Y}_1 = Y_1 - b_1 - b_2 X_1$$
$$\hat{u}_2 = Y_2 - \hat{Y}_2 = Y_2 - b_1 - b_2 X_2$$
$$\hat{u}_3 = Y_3 - \hat{Y}_3 = Y_3 - b_1 - b_2 X_3$$
$$\hat{u}_4 = Y_4 - \hat{Y}_4 = Y_4 - b_1 - b_2 X_4$$

Plug in values from our given data for $\{X, Y\}$

$$\hat{u}_1 = 4 - b_1 - 1 * b_2$$
$$\hat{u}_2 = 3 - b_1 - 2 * b_2$$
$$\hat{u}_3 = 5 - b_1 - 3 * b_2$$
$$\hat{u}_4 = 8 - b_1 - 4 * b_2$$

Next we'll square each of these terms and summate for RSS

OLS Application

Step 2: Calculate the RSS

$$
\begin{aligned}
RSS &= \sum_{i=1}^{n} \hat{u_i}^2 = \hat{u_1}^2 + \hat{u_2}^2 + \hat{u_3}^2 + \hat{u_4}^2 \\
&= (4 - b_1 - b_2)^2 + (3 - b_1 - 2b_2)^2 + (5 - b_1 - 3b_2)^2 + (8 - b_1 - 4b_2)^2 \\
&= 114 + 4b_1^2 + 30b_2^2 - 40b_1 - 114b_2 + 20b_1b_2
\end{aligned}
$$

Recall that OLS minimizes the RSS expression with respect to the specific parameters involved.

To find the values that minimize a particular expression, we need to apply differentiation.

OLS Application

Step 3: Differentiate RSS by parameters

To differentiate by a particular variable, multiply each term by its power value and subtract 1 from the power of each of its terms.

e.g. for $y = 2x^3, \partial y/\partial x = 2 * 3x^{3-1} = 6x^2$

$$\frac{\partial RSS}{\partial b_1} = 0 \implies (4*2)b_1^{2-1} - (40*1)b_1^{1-1} + (20*1)b_1^{1-1}b_2 = 0$$
$$\implies 8b_1 - 40 + 20b_2 = 0 \qquad Eq(1)$$

$$\frac{\partial RSS}{\partial b_2} = 0 \implies (30*2)b_2^{2-1} - (114*1)b_2^{1-1} + (20*1)b_1 b_2^{1-1} = 0$$
$$\implies 60b_2 - 114 + 20b_1 = 0 \qquad Eq(2)$$

OLS Application

Step 4: Solve for parameters

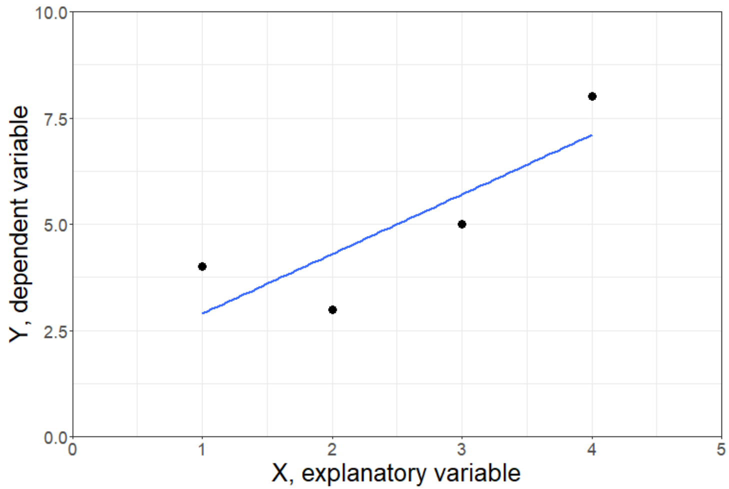
With two unknowns $\{b_1, b_2\}$ and two equations in which these unknowns satisfied the first order conditions $\left\{\frac{\partial RSS}{\partial b_1}, \frac{\partial RSS}{\partial b_2}\right\}$, we can solve for our parameters.

How? Substitute one expression into the other.

$$\begin{aligned} 20b_2 = 40 - 8b_1 &\implies 60b_2 = 120 - 24b_1 \\ &\textbf{substitute into second equation} \\ Eq(2): \ &120 - 24b_1 - 114 + 20b_1 = 0 \\ &6 = 4b_1 \implies b_1 = 1.5 \\ Eq(1): \ &20b_2 = 40 - 8 \times 1.5 = 28 \implies b_2 = 1.4 \end{aligned}$$

OLS would prescribe $\{1.5, 1.4\}$ for our set of parameter estimates.

OLS Application



Fitting a line through the data points, with the aim of minimizing the RSS, results in the same implied parameters

- Such parameters will always be estimated computationally
- We will perform an exercise by hand in **PBS3** to understand the mechanics underlying the values we hang our hats on