Simple Linear Regression: Estimation

EC 320: Introduction to Econometrics

Winter 2022

Prologue

Housekeeping

Grading: Midterm 1 grade out.

Problem Set 2: Due Monday, Feb 7th by 11:59pm *on Canvas*.

Lab & Exercise: Wednesday, Feb 2nd by 11:59pm.

Where Are We?

Where we've been

High Concepts

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

Also, **R**.

Where Are We?

Where we're going

The Weeds!

- 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

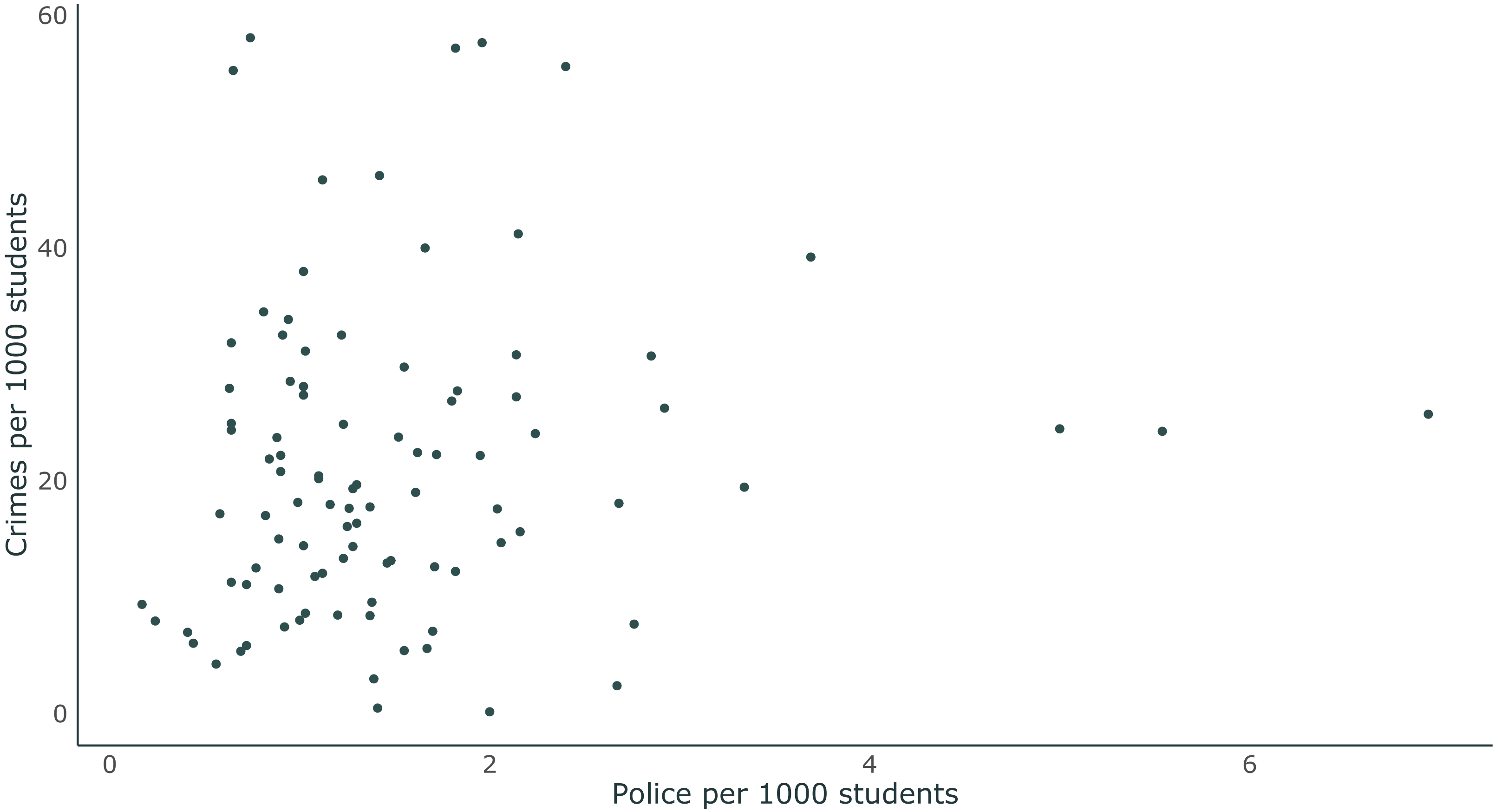
"Looking" at data wasn't especially helpful.

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:

$$Y_i = \beta_0 + \beta_1 X_i + u_i.$$

- β_0 is the **intercept** or constant.
- β_1 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 $X_i = 0$.

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

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.

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

"A one-unit increase in X_i is associated with a β_1-unit increase in Y_i."

Under certain (strong) assumptions about the error term, β_1 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.

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

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?

$$\text{Crime}_i = \beta_0 + \beta_1 \text{Police}_i + u_i.$$

- β_0 is the crime rate for colleges without police.
- β_1 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?

$$\mathbf{Crime}_i = \beta_0 + \beta_1 \mathbf{Police}_i + u_i$$

β_0 and β_1 are the population parameters we want, but we cannot observe them.

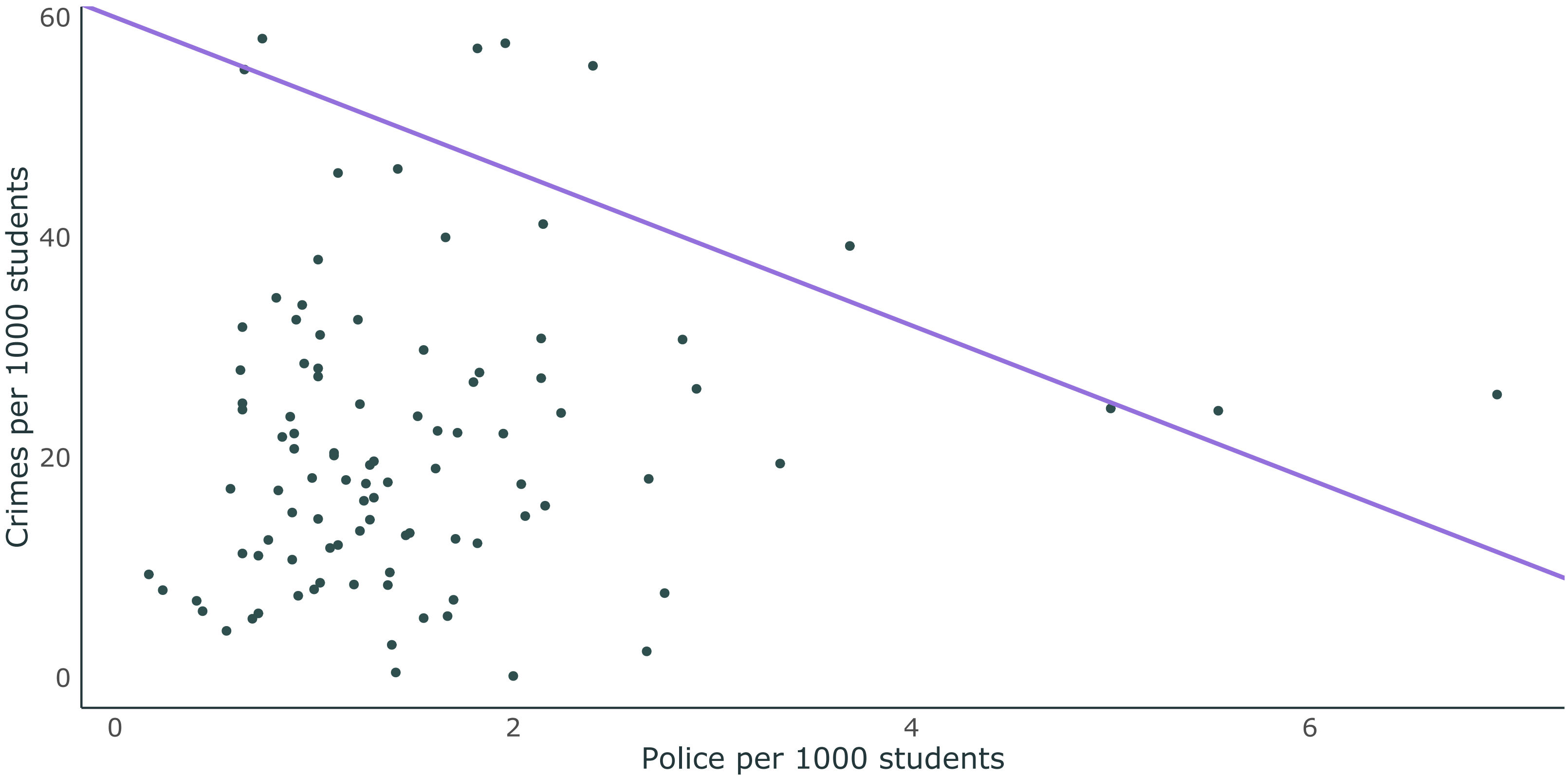
Instead, we must estimate the population parameters.

- $\hat{\beta}_0$ and $\hat{\beta}_1$ generate predictions of $\mathbf{Crime}_i$ called $\hat{\mathbf{Crime}}_i$.
- We call the predictions of the dependent variable **fitted values.**
- Together, these trace a line: $\hat{\mathbf{Crime}}_i = \hat{\beta}_0 + \hat{\beta}_1 \mathbf{Police}_i$.

Take 3, attempted

Example: Effect of police on crime

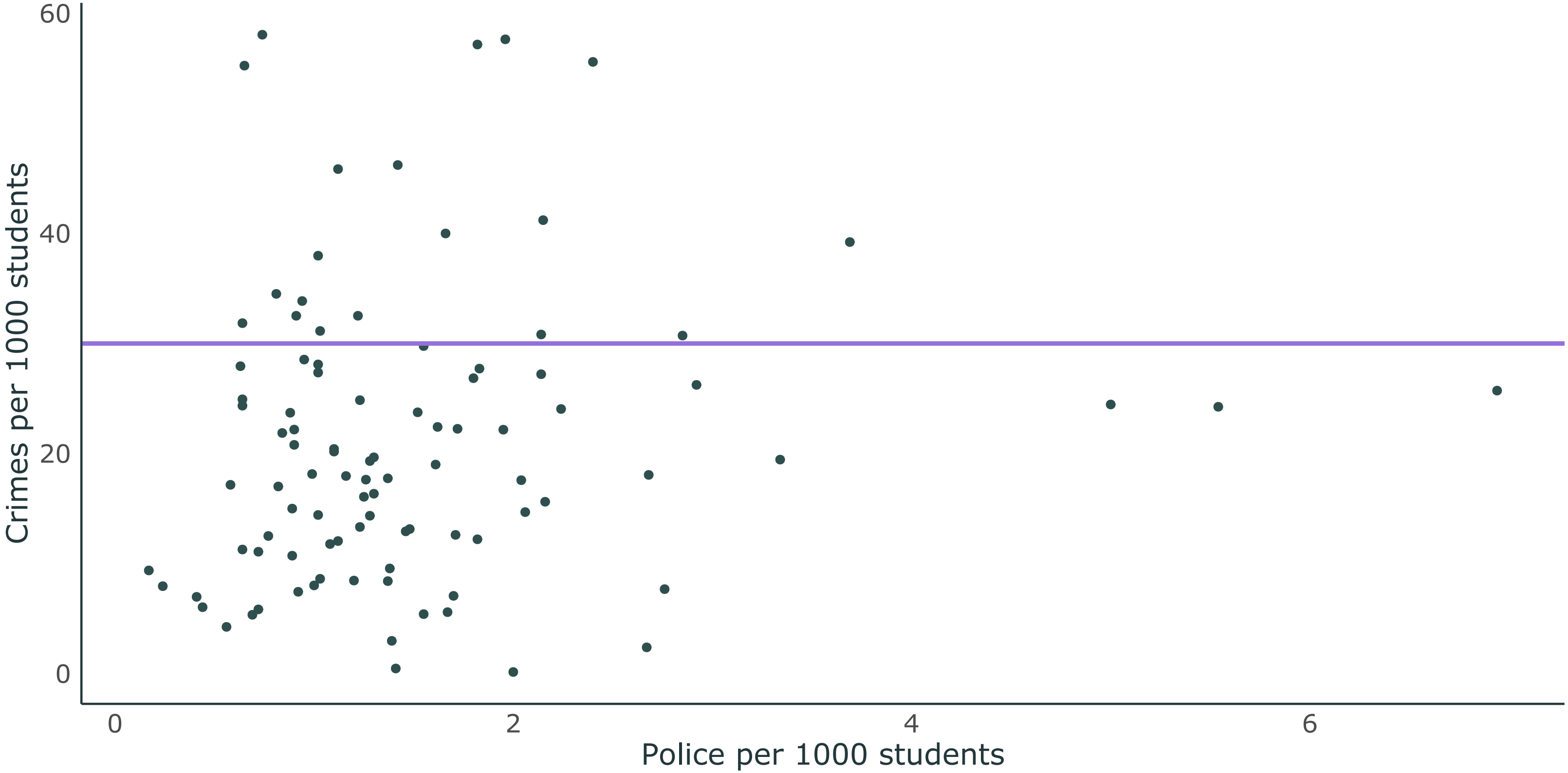
Guess: $\hat{\beta}_0 = \mathbf{60}$ and $\hat{\beta}_1 = -\mathbf{7}$.



Take 4

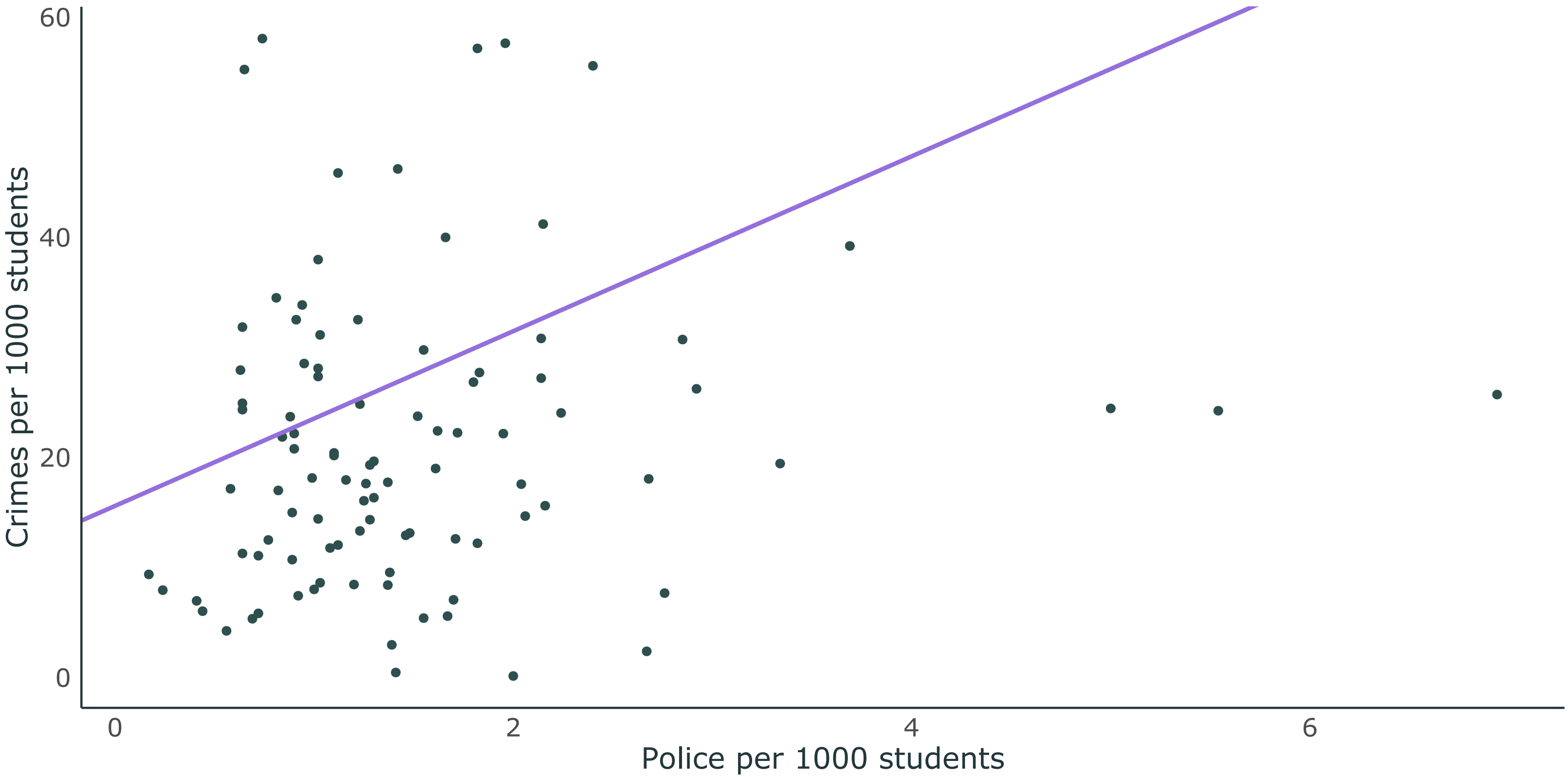
Example: Effect of police on crime

Guess: $\hat{\beta}_0 = \mathbf{30}$ and $\hat{\beta}_1 = \mathbf{0}$.



Example: Effect of police on crime

Guess: $\hat{\beta}_0 = \mathbf{15.6}$ and $\hat{\beta}_1 = \mathbf{7.94}$.



Residuals

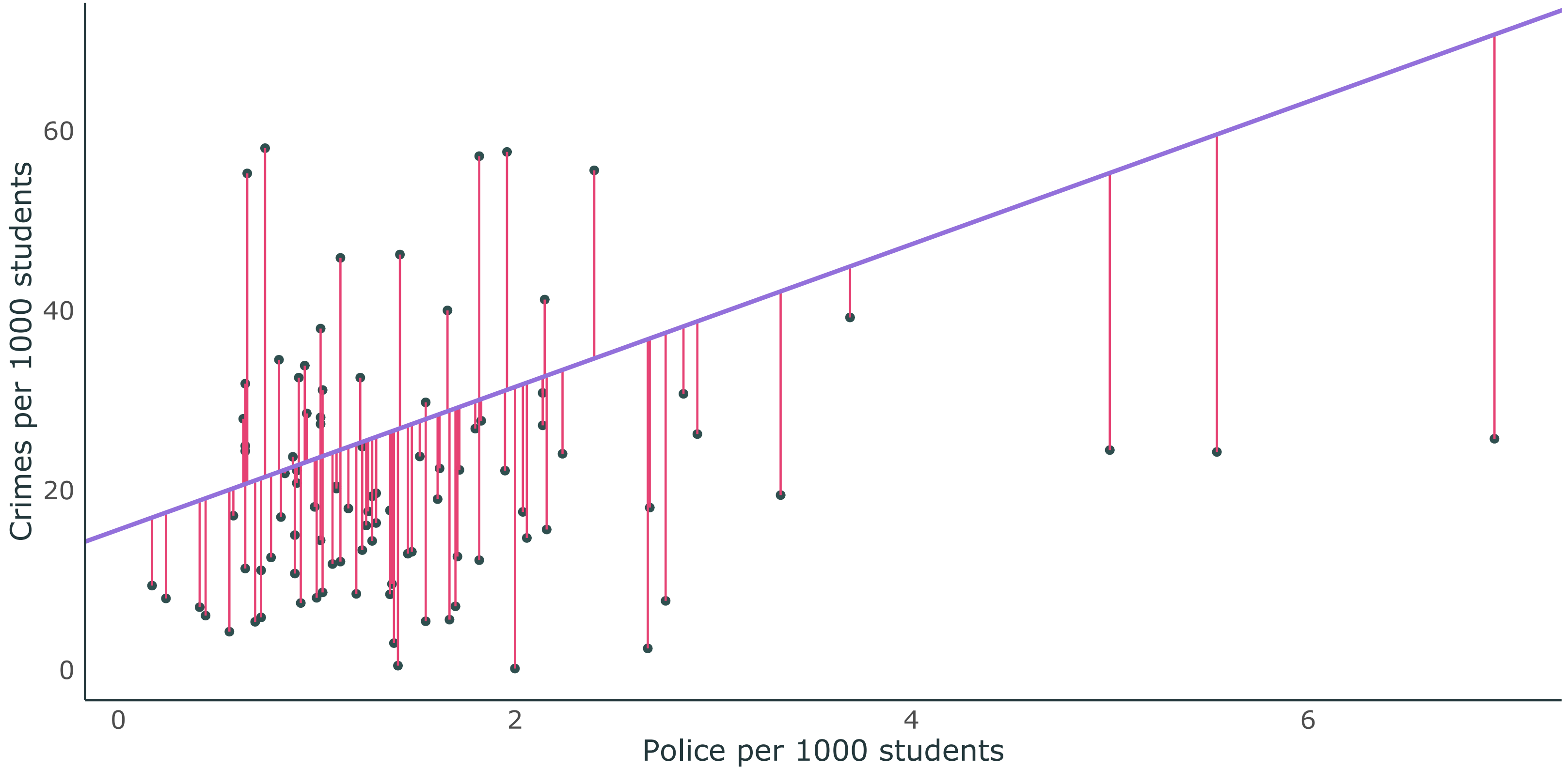
Using $\hat{\beta}_0$ and $\hat{\beta}_1$ to make $\hat{Y}_i$ generates misses called **residuals**:

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

- Sometimes called e_i.

Example: Effect of police on crime

Using $\hat{\beta}_0 = \mathbf{15.6}$ and $\hat{\beta}_1 = \mathbf{7.94}$ 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).

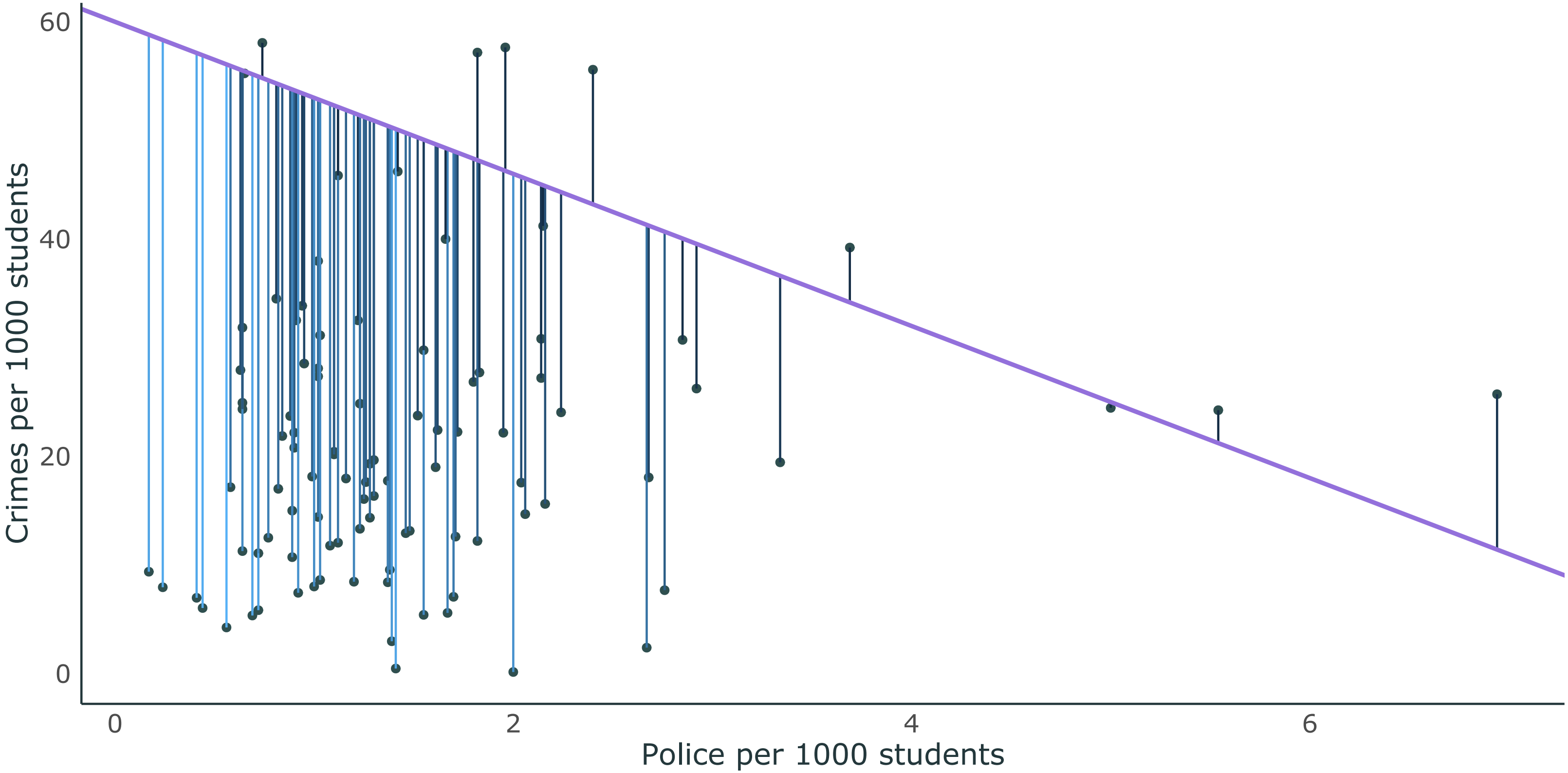
Alternative: Minimize the sum of squared residuals a.k.a. the **residual sum of squares (RSS)**.

- 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 $\hat{\beta}_0$ and $\hat{\beta}_1$ 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 $\hat{\beta}_0$ and $\hat{\beta}_1$ 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.**

Deriving the OLS Estimator

Outline

1. Replace $\sum_{i=1}^{n} \hat{u}_i^2$ with an equivalent expression involving $\hat{\beta}_0$ and $\hat{\beta}_1$.
2. Take partial derivatives of our RSS expression with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$ and set each one equal to zero (first-order conditions).
3. Use the first-order conditions to solve for $\hat{\beta}_0$ and $\hat{\beta}_1$ in terms of data on Y_i and X_i.
4. Check second-order conditions to make sure we found the $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize RSS.

OLS Formulas

For details, see the handout posted on Canvas.

Slope coefficient

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

Intercept

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_2\bar{X}$$

Slope coefficient

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

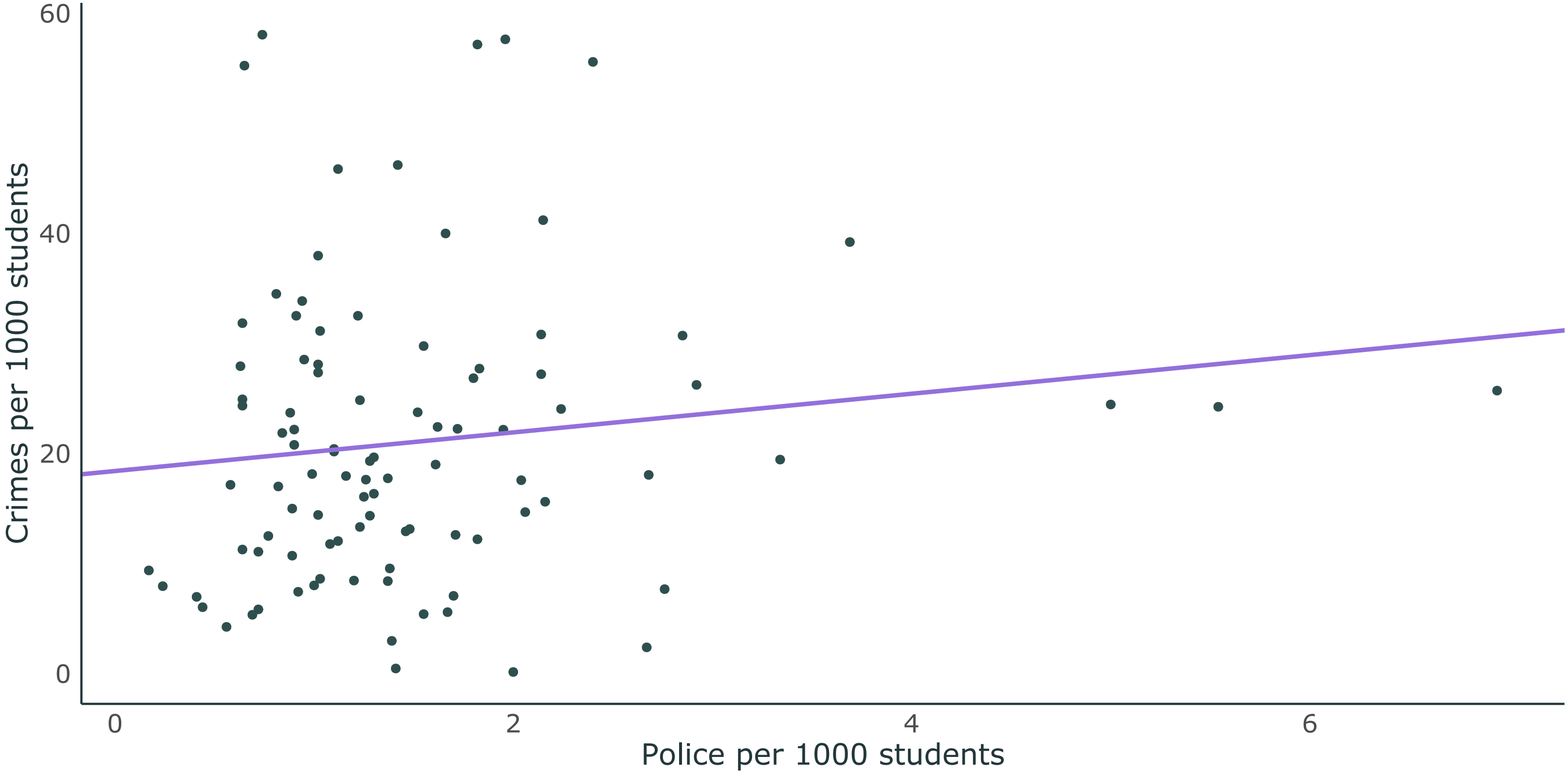
$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

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

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

Example: Effect of police on crime

Using the OLS formulas, we get $\hat{\beta}_0$ = 18.41 and $\hat{\beta}_1$ = 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 $\hat{\beta}_0$ = 18.41 tell us?

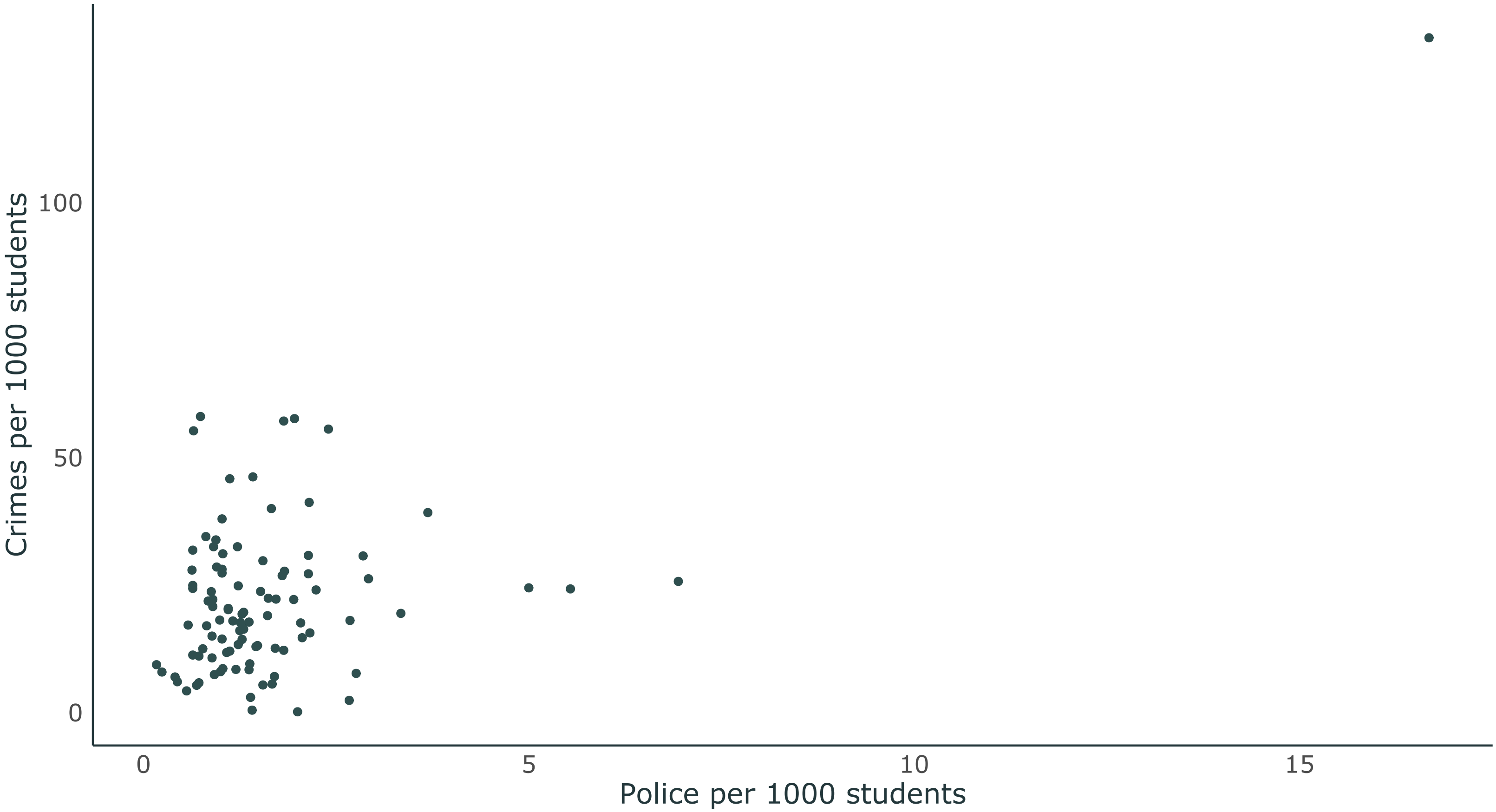
What does $\hat{\beta}_1$ = 1.76 tell us?

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.