Simple Linear Regression: Estimation

EC 320: Introduction to Econometrics

SM Shihab Siddiqui

Prologue

Housekeeping

Short Assignments:

- **Short Assignments** are due by 5 PM **Monday** every week other than Week 1 and the week after Midterm week.
- Tentatively, **Problem Sets are due 5 PM Fridays of Week 2, 4, 7 and 9**.
- I will drop the three lowest scores of short assignments, and lowest score of problem set.
- No late submission is accepted.
- Last reminder.

Where Are We?

Where we've been

- Reviewed core ideas from statistics

Where we're going

Consider how to think about how two variables related to each other.

- We will learn the mechanics of Ordinary Least Squares (OLS) regressions
- Interpret regression results (mechanically and critically)
- Lays a foundation for more-sophisticated regression techniques.

Simple Linear Regression

Addressing Questions

Example: Effect of education on wages

Policy Question: Does more education increase wages ?

- **Empirical Question:** Does the years of education increase wages ? If so, by how much ?

How can we answer these questions?

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

Addressing Questions with Data

These are data from the 1976 Current Population Survey.

- Taken from R package `wooldridge`, the dataset is called *wage1*.
- 526 observations on 24 variables:
- The variables are measurements of wages, years of education, level of work experience, demographics (sex, race, marital status, number of dependents), location in US, type of industry.

Take 1: Let's "*Look*" at Data

Example: Effect of education on wages

Search:

	Wages	Education
1	3.1	11
2	3.24	12
3	3	11
4	6	8
5	5.3	12
6	8.75	16

Showing 1 to 6 of 526 entries

Previous Next

Take 2

Example: Effect of education on wages

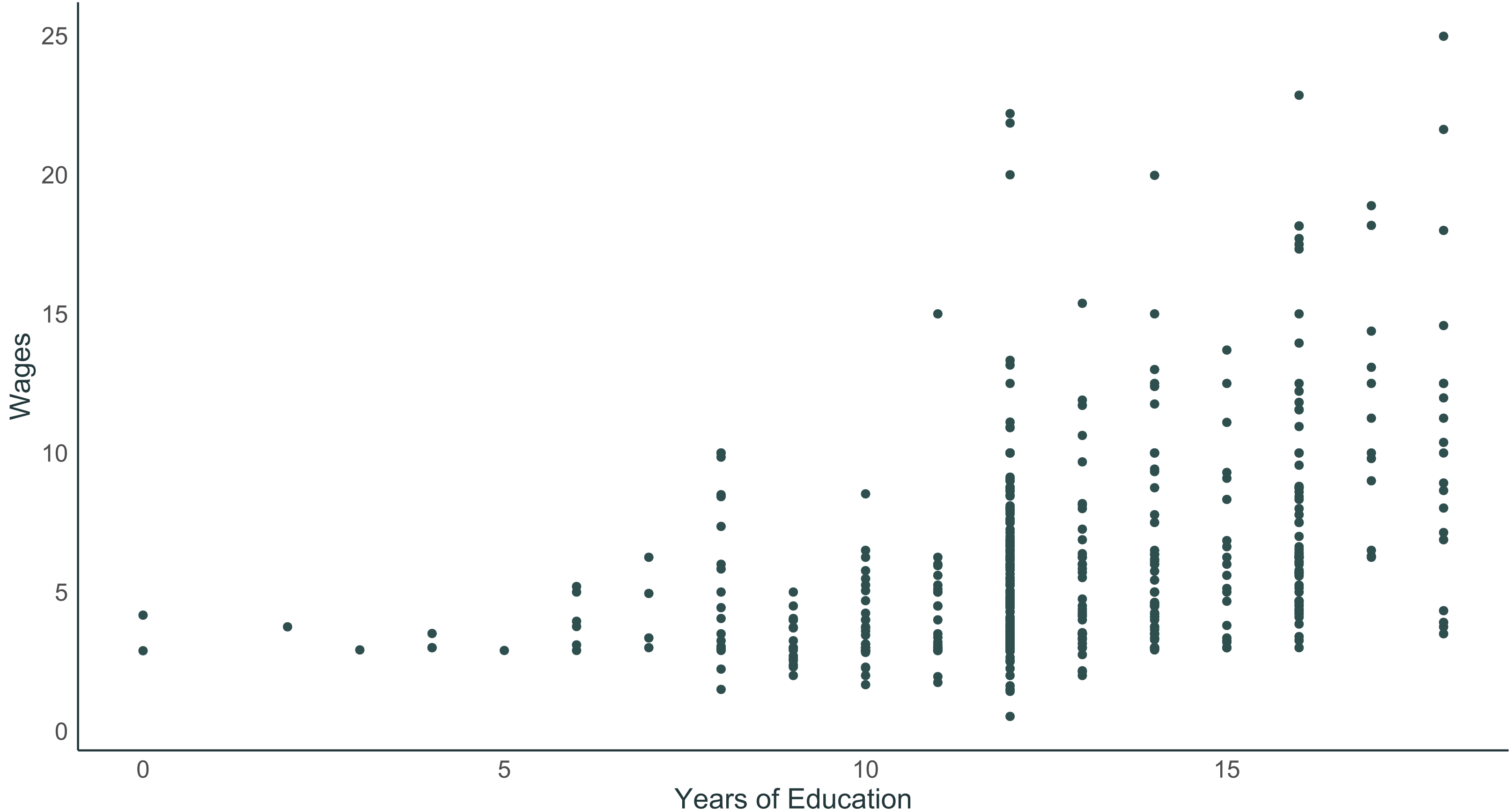
"Looking" at data wasn't especially helpful.

Let's try using a scatter plot.

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

Take 2

Example: Effect of education on wages



- Sample correlation coefficient of 0.41 confirms a positive correlation.

Take 3

Example: Effect of education on wages

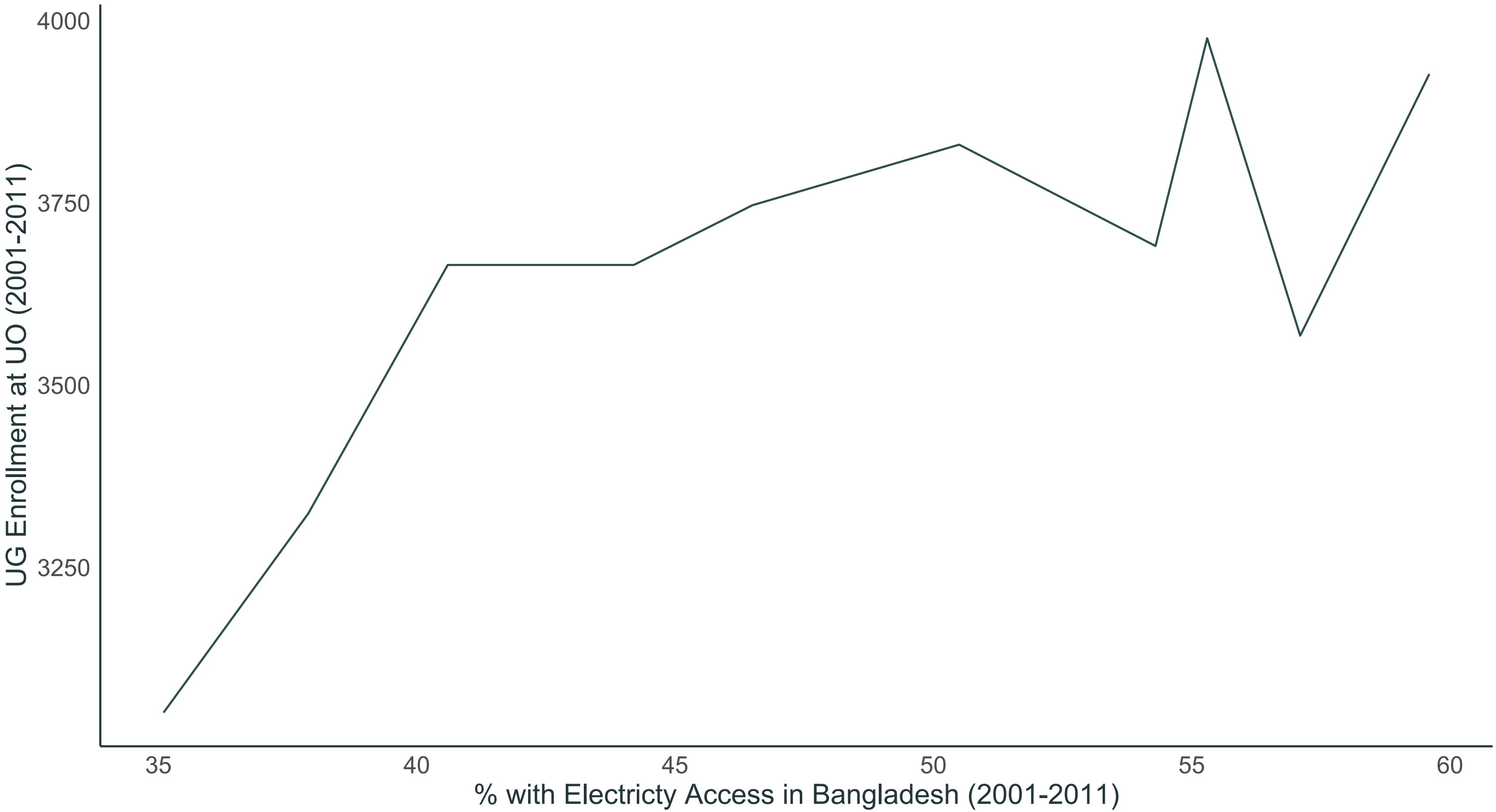
But our question was

> Does the years of education increase wages ? If so, by how much ?

The scatter plot and correlation coefficient provide only a partial answer, if even.

It **ONLY** tells us that we usually observe higher years of education and higher wages together.

Correlation is not causation



Correlation coefficient is 0.76 even larger that before!

Take 3

Example: Effect of education on wages

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

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

"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 education on wages

How might we apply the simple linear regression model to our question about the effect of on education on wages ?

- Which variable is X? Which is Y?

$$\mathbf{Wage}_i = \beta_1 + \beta_2 \mathbf{Education}_i + u_i.$$

Take 3, continued

Example: Effect of education on wages

How might we apply the simple linear regression model to our question about the effect of on education on wages ?

- Which variable is X? Which is Y?

$$\mathbf{Wage}_i = \beta_1 + \beta_2 \mathbf{Education}_i + u_i.$$

- β_1 is the wage rate without college.
- β_2 is the increase in wages when years of education increase by one.

Take 3, continued

Example: Effect of education on wages

$$\mathbf{Wage}_i = \beta_1 + \beta_2 \mathbf{Education}_i + u_i.$$

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

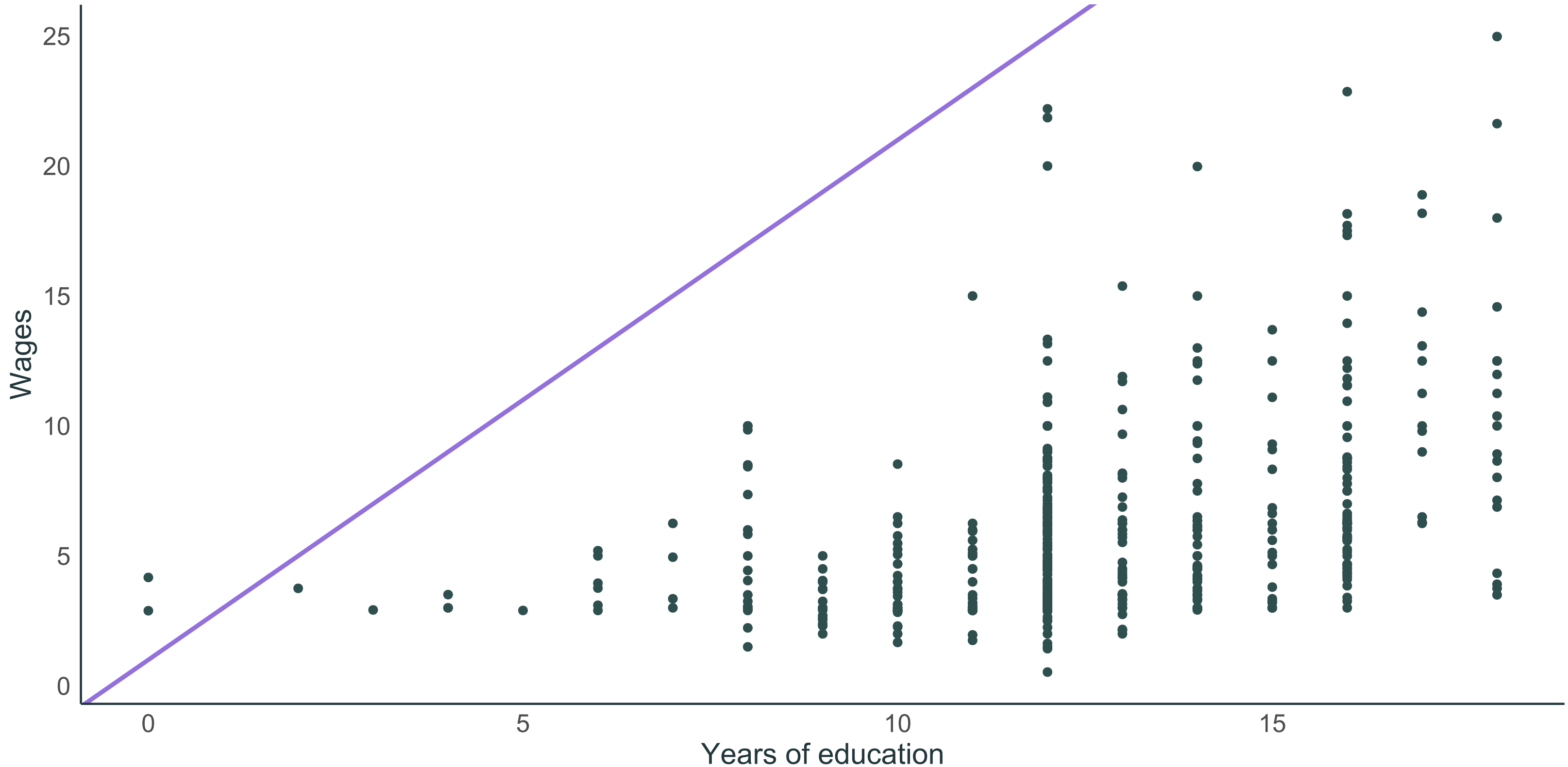
Instead, we must estimate the population parameters.

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

Take 3, attempted

Example: Effect of education on wages

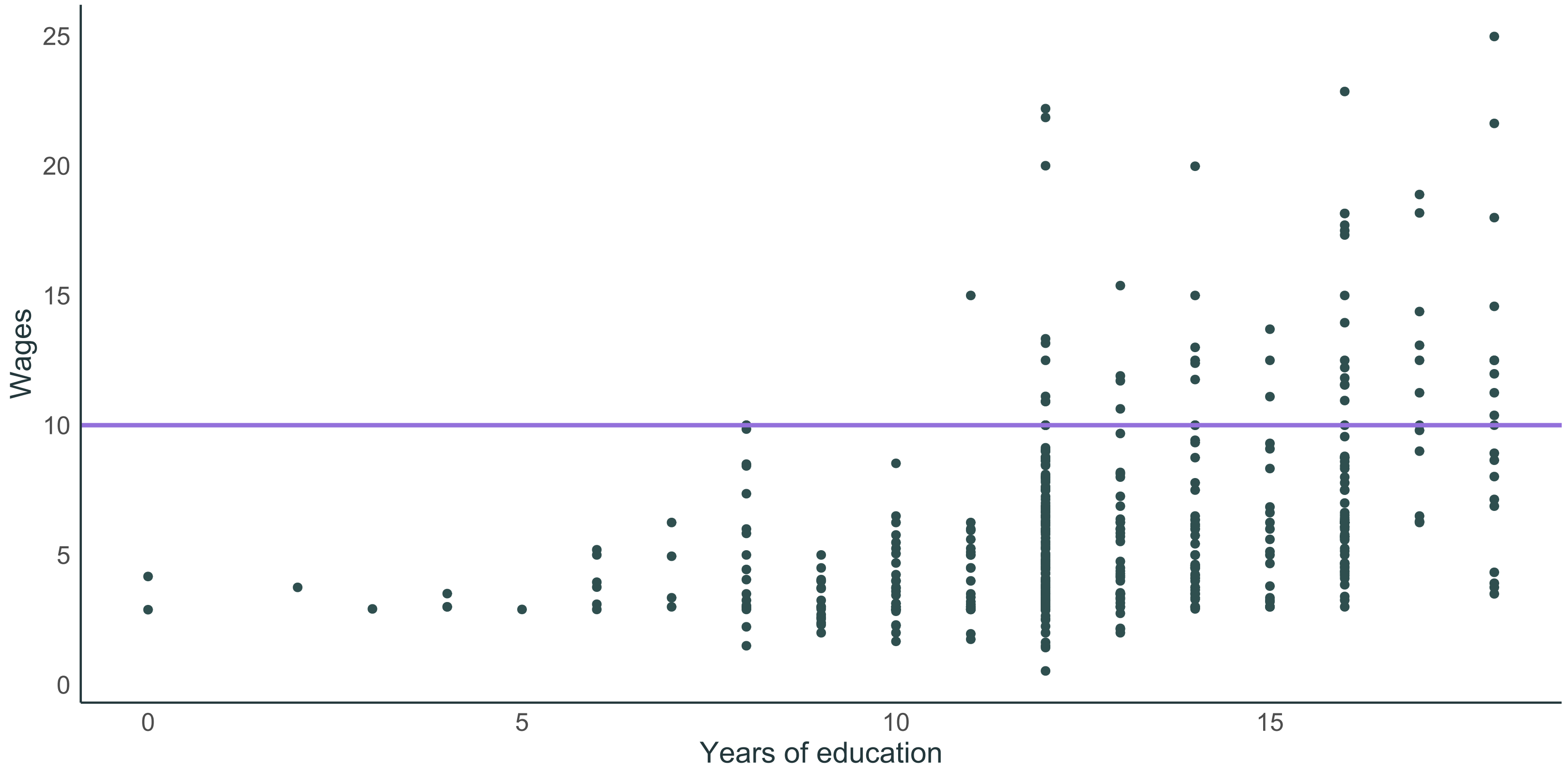
Guess: $\hat{\beta}_1 = 1$ and $\hat{\beta}_2 = 2$.



Take 4

Example: Effect of education on wages

Guess: $\hat{\beta}_1 = 10$ and $\hat{\beta}_2 = 0$.



Residuals

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

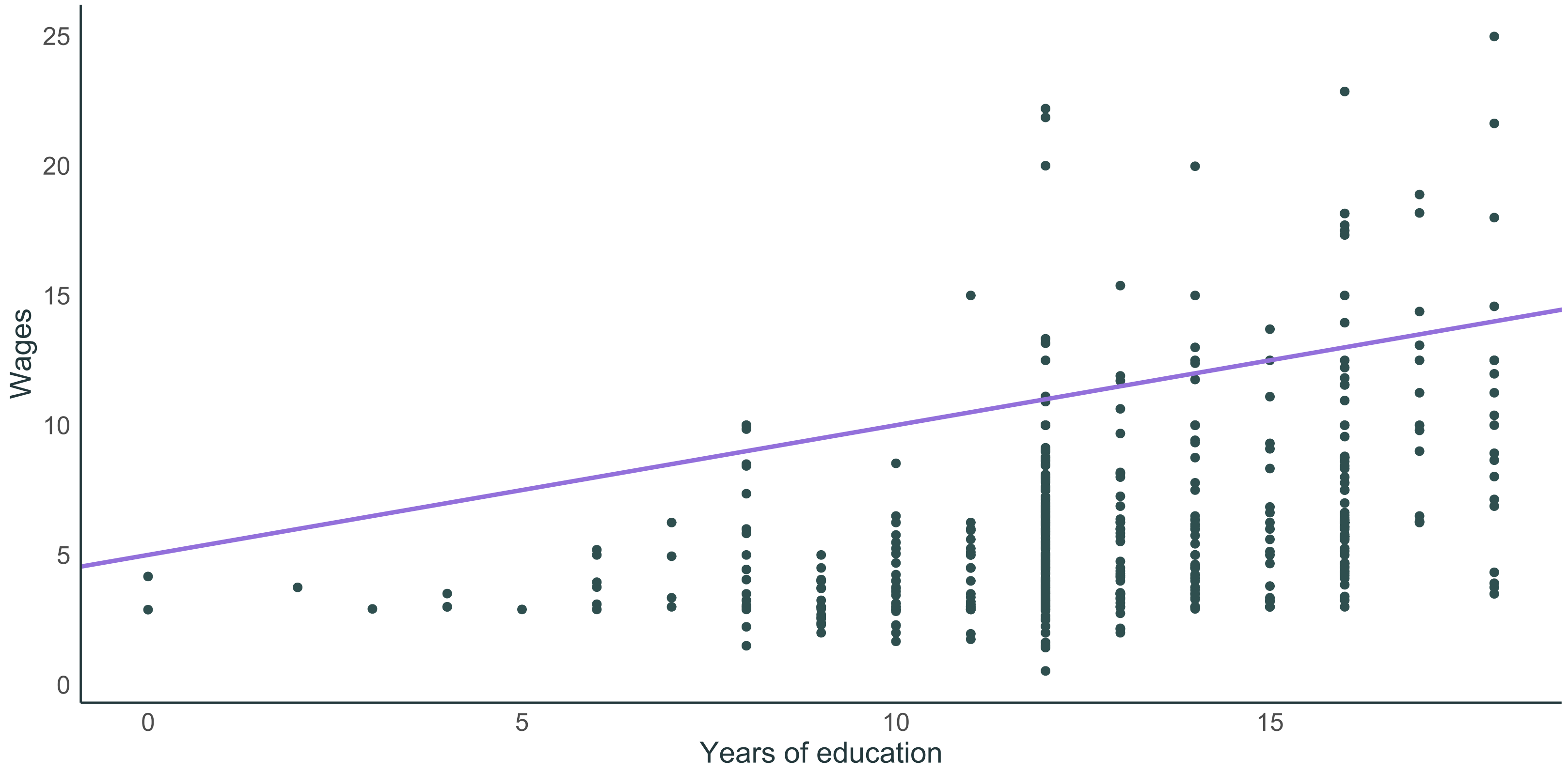
$$\hat{u}_i = Y_i - \hat{Y}_i.$$

- Sometimes denoted e_i.

Residuals

Example: Effect of education on wages

Using $\hat{\beta}_1 = \mathbf{5}$ and $\hat{\beta}_2 = \mathbf{0.5}$ to make $\hat{\text{wages}}_i$ generates **residuals**.



Residuals

We want an estimator that makes less mistakes in our prediction.

$\hat{u}_i$ is a measure of mistake for observation i.

- So, one measure of total mistake for all observations is $\sum_{i=1}^{n} \hat{u}_i$

But mistakes can be positive and negative. When we add them, it cancels each other. So, $\sum_{i=1}^{n} \hat{u}_i$ is a bad measure of total mistakes.

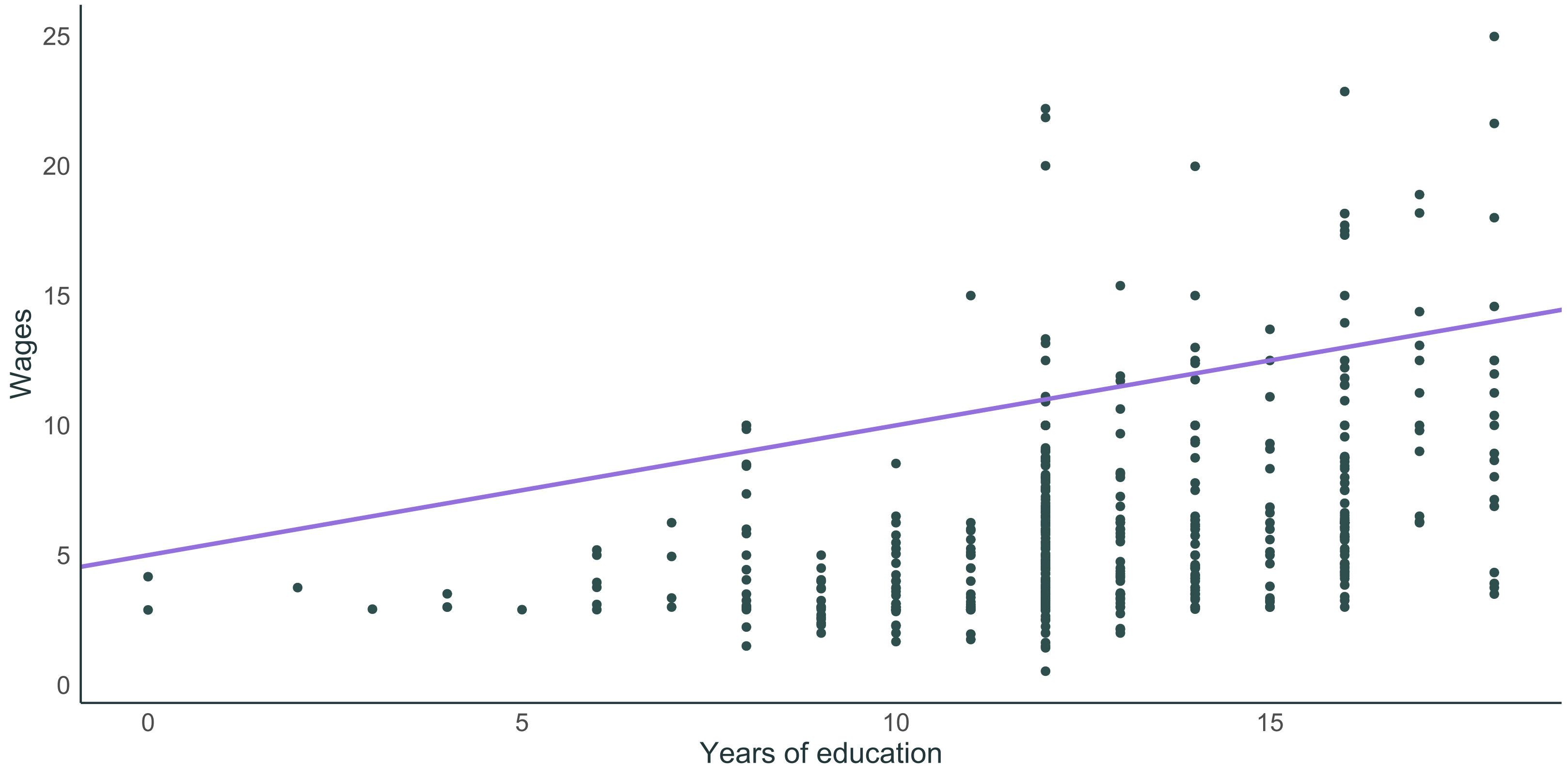
Solution: Minimize the sum of squared residuals a.k.a. the **residual sum of squares (RSS)**.

- Squared numbers are never negative.

Residuals

Example: Effect of education on wages

RSS Bigger penalties are given for bigger mistakes in prediction.



Residuals

Minimizing RSS

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

- In fact, some estimation process involves doing just that. But that is not what we are interested in here.

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.**

Deriving the OLS Estimator

Outline

For details, see textbook. In summary:

- **Step 1.** Replace $\sum_{i=1}^{n} \hat{u}_i^2$ with an equivalent expression involving $\hat{\beta}_1$ and $\hat{\beta}_2$.
- **Step 2.** Take partial derivatives of our RSS expression with respect to $\hat{\beta}_1$ and $\hat{\beta}_2$ and set each one equal to zero (first-order conditions).
- **Step 3.** Use the first-order conditions to solve for $\hat{\beta}_1$ and $\hat{\beta}_2$ in terms of data on Y_i and X_i.
- **Step 4.** Check second-order conditions to make sure we found the $\hat{\beta}_1$ and $\hat{\beta}_2$ that minimize RSS.

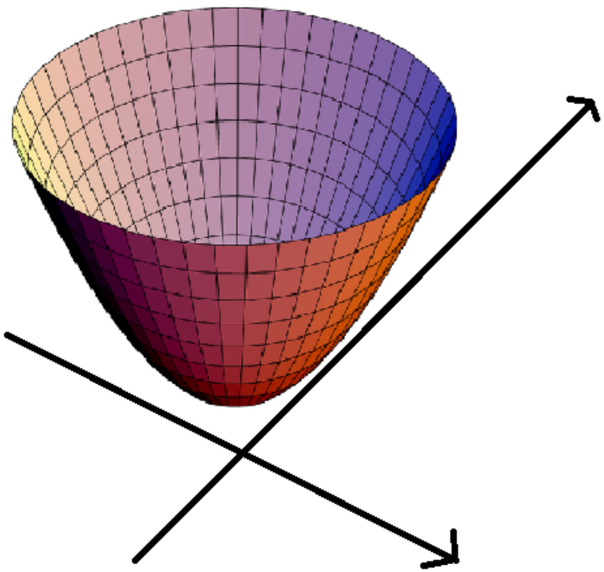
Deriving OLS estimator, step 1

$$\begin{aligned}
\mathrm{RSS}(\hat\beta_1, \hat\beta_2) &= \sum_{i=1}^{n} \hat u_i^2 \ \ \dots\ \ \textit{We substitute expression for } \hat u_i \\
&= \sum_{i=1}^{n} (Y_i - \hat\beta_1 - \hat\beta_2 X_i)^2 \ \ \dots\ \ \textit{We expand square term} \\
&= \sum_{i=1}^{n} (Y_i^2 + \hat\beta_1^2 + \hat\beta_2^2 X_i^2 - 2\hat\beta_1 Y_i - 2\hat\beta_2 X_i Y_i + 2\hat\beta_1 \hat\beta_2 X_i)
\end{aligned}$$

Deriving OLS estimator, step 2

Minimization review

- $\frac{\partial RSS}{\partial \hat{\beta}_1} = 0$
- $\frac{\partial RSS}{\partial \hat{\beta}_2} = 0$



Deriving OLS estimator, step 2 & 3

We take the partial derivatives and set them to zero:

- $\frac{\partial RSS}{\partial \hat{\beta}_1} = 0 \implies 2n\hat{\beta}_1 - 2\sum_{i=1}^{n} Y_i + 2\hat{\beta}_2 \sum_{i=1}^{n} X_i = 0 \ \ldots \ (1)$
- $\frac{\partial RSS}{\partial \hat{\beta}_2} = 0 \implies 2\hat{\beta}_2 \sum_{i=1}^{n} X_i^2 - 2\sum_{i=1}^{n} X_i Y_i + 2\hat{\beta}_1 \sum_{i=1}^{n} X_i = 0 \ \ldots \ (2)$
- These are called **normal equations.**

This is a 2x2 simultaneous equation system where we are solving for $\hat{\beta}_1$ & $\hat{\beta}_2$. We know how to solve this !

Step 4 is beyond our scope. Trust me.

OLS Formulas

After solving the simultaneous equation system above, we get:

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

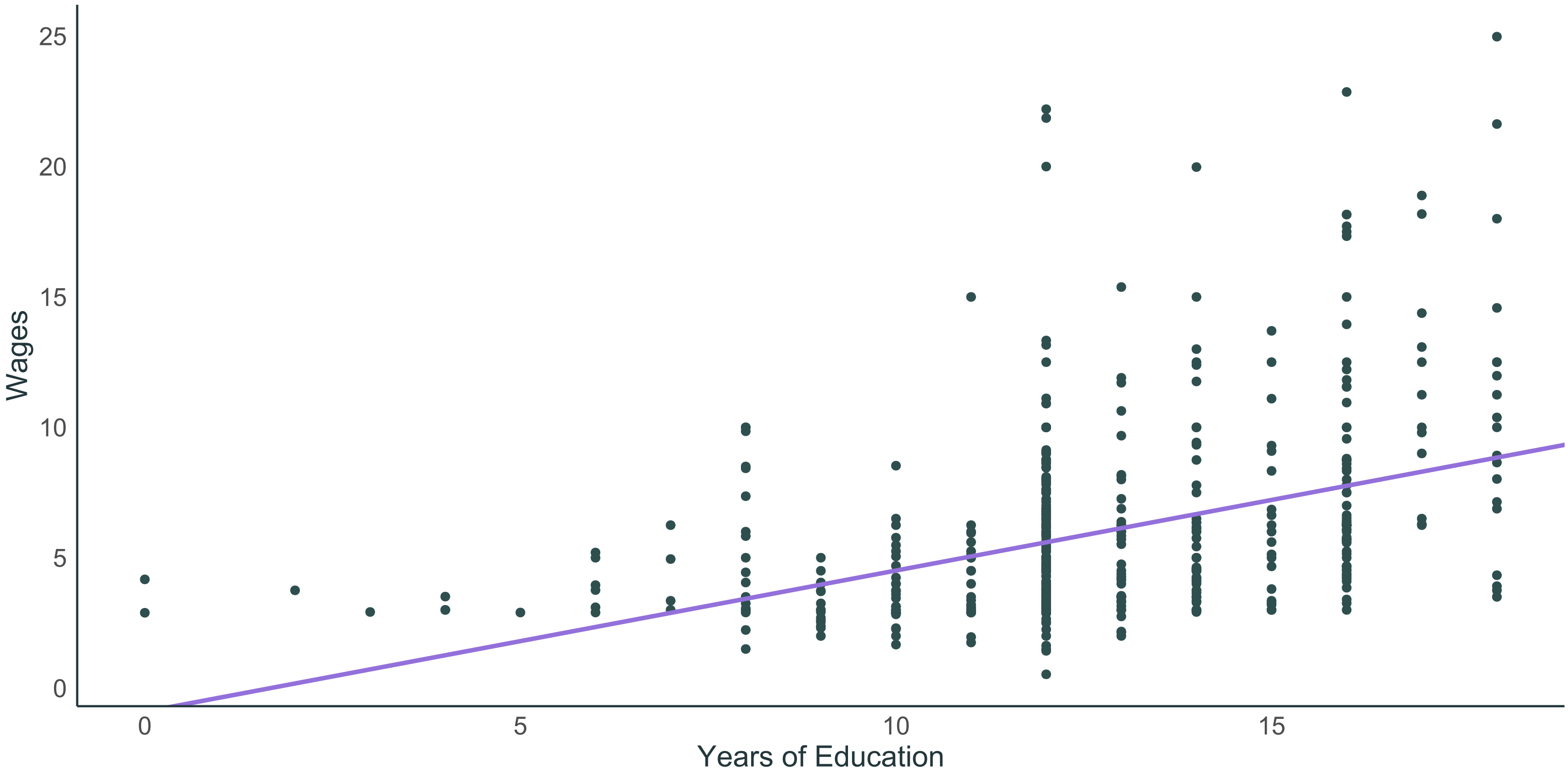
The slope estimator is equal to the sample covariance divided by the sample variance of X:

$$
\begin{aligned}
\hat{\beta}_2 &= \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \\
&= \frac{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} \\
&= \frac{S_{XY}}{S_X^2}.
\end{aligned}
$$

Coefficients

Example: Effect of education on wages, take 4

Using the OLS formulas, we get $\hat{\beta}_1$ = -0.9 and $\hat{\beta}_2$ = 0.54.



Coefficient Interpretation

Example: Effect of education on wages

Using OLS gives us the fitted line

$$\textbf{Wage}_i = \hat{\beta}_1 + \hat{\beta}_2 \textbf{Education}_i \ \textbf{Wage}_i = -\mathbf{0.9} + \mathbf{0.54}\ \textbf{Education}_i$$

What does $\hat{\beta}_1$ = -0.9 tell us?

What does $\hat{\beta}_2$ = 0.54 tell us?

Gut check: Does this mean that people without any education *pay* to work ?
Gut check: Does this mean that one extra year of education *cause* wages to go up by $0.54 ?

- Probably not. **Why?**

Coefficient Interpretation

Correlation is not causation

These points would be discussed in future. I just want to contain your excitement!

There are many issues with this analysis. Let us discuss a few.

We must think through the **data generating process** before we interpret the coefficients.

> In statistics and in empirical sciences, a data generating process is a process in the real world that "generates" the data one is interested in. *(Prof. Wiki)*

Coefficient Interpretation

Correlation is not causation

- Government regulation leads to a situation where most people undergo 10 years of education at the least.
- Loosely speaking, we are extrapolating to say things like wage = -0.9 if years of education = 0.
- People with higher educational ability goes to college. They may have fewer *behavioral* problems. They may comes from richer families. Wages are also determined by many other factors - experience, field of study, so many other things. *We will tackle that in Multiple Linear Regression.*
- Our econometric procedure simply captures the association between lower(**higher**) levels of education and lower(**higher**) wages.

Coefficient Interpretation

- we cannot say that each unit increase in years of education **causes** wages to go up by \$0.54.
- Do we think an additional year of education will have the same impact regardless of the level of education ?
  - What about grade 1 vs 2 ? completing 3 years of college vs 4 (and getting the degree ?)

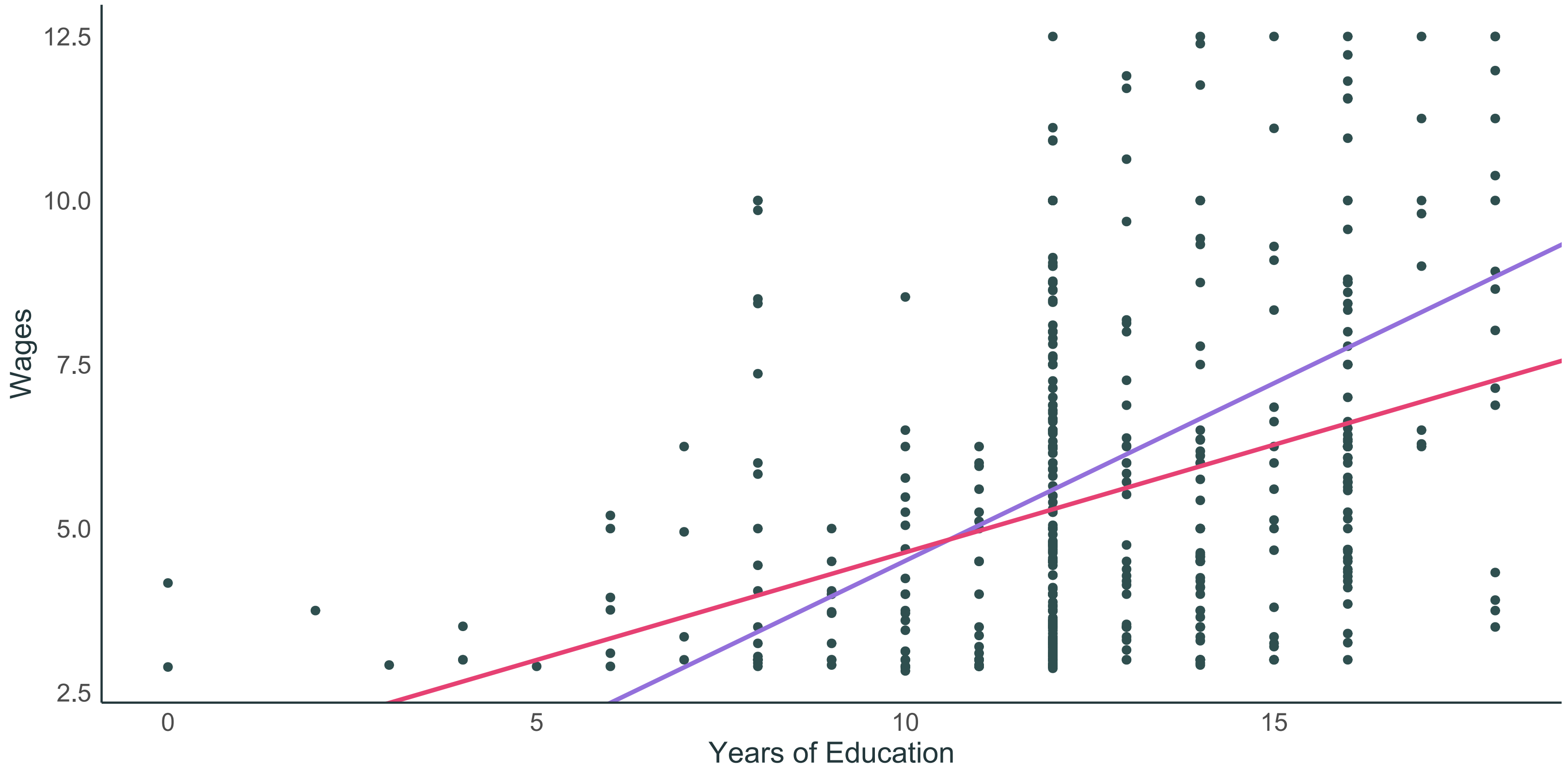
The correct interpretation is a humble one:

$\hat{\beta}_2 = \mathbf{0.54}$ means that one more year of education is **associated** with a **0.54** increase in wage rate on **average, given everything else remains constant.**

Outliers

Example: Effect of education on wages

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



OLS Properties

The way we selected OLS estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ gives us three important properties:

- Residuals sum to zero: $\sum_{i=1}^{n} \hat{u}_i = 0$.
  - By extension, the sample mean of the residuals are zero.
- The sample covariance between the independent variable and the residuals is zero: $\sum_{i=1}^{n} X_i \hat{u}_i = 0$.
- The point $(\bar{X}, \bar{Y})$ is always on the regression line.
- You will have a chance to prove some of these later.

Where are we at

We considered a simple linear regression of Y_i on X_i:

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

- β_1 and β_2 are **population parameters** that describe the *"true"* relationship between X_i and Y_i.

- **Problem:** We don't know the population parameters. The best we can do is to estimate them.

Where are we, continued

We derived the OLS estimators for parameters β_1 and β_2 given a dataset (X,Y) by picking estimates that minimize $\sum_{i=1}^{n} \hat{u}_i^2$.

- **Intercept:**

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

- **Slope:**

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

Where are we

With the OLS estimates of the population parameters, we constructed a regression line:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i.$$

- $\hat{Y}_i$ are predicted or **fitted** values of Y_i.
- You can think of $\hat{Y}_i$ as an estimate of the average value of Y_i given a particular of X_i.

OLS still produces prediction errors: $\hat{u}_i = Y_i - \hat{Y}_i$.

- Put differently, there is a part of Y_i we can explain and a part we cannot: $Y_i = \hat{Y}_i + \hat{u}_i$.

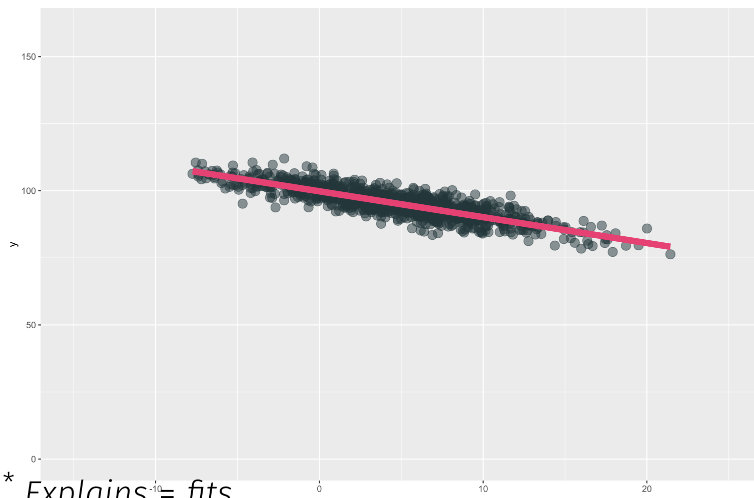
Goodness of Fit

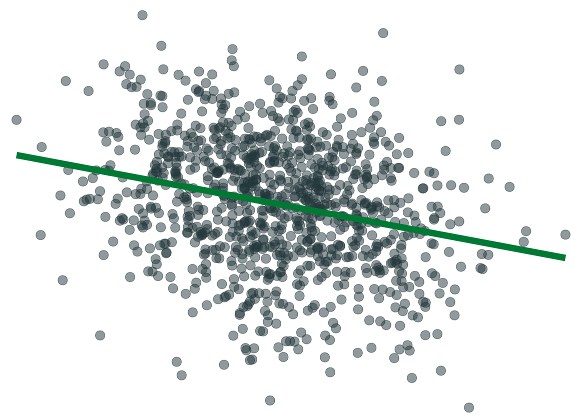
Regression 1 *vs.* Regression 2

- Same slope.

- Same intercept.

Q: Which fitted regression line *"explains"*[*] the data better?





[*] *Explains* = *fits*.

Regression 1 *vs.* Regression 2

The **coefficient of determination** R^2 is the fraction of the variation in Y_i *"explained"* by X_i in a linear regression.

- $R^2 = 1 \implies X_i$ explains *all* of the variation in Y_i.
- $R^2 = 0 \implies X_i$ explains *none* of the variation in Y_i.

R^2 = 0.72

R^2 = 0.07

Explained and Unexplained Variation

Residuals remind us that there are parts of Y_i we can't explain.

$$Y_i = \hat{Y}_i + \hat{u}_i$$

- Sum the above, divide by n, and use the fact that OLS residuals sum to zero to get $\bar{\hat{u}} = 0 \implies \bar{Y} = \bar{\hat{Y}}$.

Total Sum of Squares (TSS) measures variation in Y_i:

$$\text{TSS} \equiv \sum_{i=1}^{n} (Y_i - \bar{Y})^2.$$

- We will decompose this variation into explained and unexplained parts.

Explained and Unexplained Variation

Explained Sum of Squares (ESS) measures the variation in $\hat{Y}_i$:

$$\text{ESS} \equiv \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2.$$

Residual Sum of Squares (RSS) measures the variation in $\hat{u}_i$:

$$\text{RSS} \equiv \sum_{i=1}^{n} \hat{u}_i^2.$$

Goal: Show that $\text{TSS} = \text{ESS} + \text{RSS}$.

Step 1: Plug $Y_i = \hat{Y}_i + \hat{u}_i$ into TSS.

$$
\begin{aligned}
&\text{TSS} \\
&\quad = \textstyle\sum_{i=1}^n (Y_i - \bar{Y})^2 \\
&\quad = \textstyle\sum_{i=1}^n ([\hat{Y}_i + \hat{u}_i] - [\bar{\hat{Y}} + \bar{\hat{u}}])^2
\end{aligned}
$$

Step 2: Recall that $\bar{\hat{u}} = 0$ and $\bar{Y} = \bar{\hat{Y}}$.

$$
\begin{aligned}
&\text{TSS} \\
&\quad = \textstyle\sum_{i=1}^n \left([\hat{Y}_i - \bar{Y}] + \hat{u}_i \right)^2 \\
&\quad = \textstyle\sum_{i=1}^n \left([\hat{Y}_i - \bar{Y}] + \hat{u}_i \right) \left([\hat{Y}_i - \bar{Y}] + \hat{u}_i \right) \\
&\quad = \textstyle\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n \hat{u}_i^2 + 2 \sum_{i=1}^n \left((\hat{Y}_i - \bar{Y}) \hat{u}_i \right)
\end{aligned}
$$

Step 3: Notice **ESS** and **RSS**.

$$
\begin{aligned}
&\text{TSS} \\
&\quad = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n} \hat{u}_i^2 + 2 \sum_{i=1}^{n} \left((\hat{Y}_i - \bar{Y}) \hat{u}_i \right) \\
&\quad = \text{ESS} + \text{RSS} + 2 \sum_{i=1}^{n} \left((\hat{Y}_i - \bar{Y}) \hat{u}_i \right)
\end{aligned}
$$

Step 4: Simplify.

$$
\begin{aligned}
&\mathrm{TSS} \\
&\quad = \mathrm{ESS} + \mathrm{RSS} + 2 \sum_{i=1}^n \left((\hat{Y}_i - \bar{Y}) \hat{u}_i \right) \\
&\quad = \mathrm{ESS} + \mathrm{RSS} + 2 \sum_{i=1}^n \hat{Y}_i \hat{u}_i - 2\bar{Y} \sum_{i=1}^n \hat{u}_i
\end{aligned}
$$

Step 5: Shut down the last two terms. Notice that

$$
\begin{aligned}
&\sum_{i=1}^n \hat{Y}_i \hat{u}_i \\
&\quad = \sum_{i=1}^n (\hat{\beta}_1 + \hat{\beta}_2 X_i) \hat{u}_i \\
&\quad = \hat{\beta}_1 \sum_{i=1}^n \hat{u}_i + \hat{\beta}_2 \sum_{i=1}^n X_i \hat{u}_i \\
&\quad = 0
\end{aligned}
$$

Goodness of Fit

Calculating R^2

- $R^2 = \frac{\text{ESS}}{\text{TSS}}$.
- $R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}$.

R^2 is related to the correlation between the actual values of Y and the fitted values of Y.

- Can show that $R^2 = (r_{Y,\hat{Y}})^2$.

Goodness of Fit

So what?

In the social sciences, low R^2 values are common.

Low R^2 doesn't mean that an estimated regression is useless.

- In a randomized control trial, R^2 is usually less than 0.1.

High R^2 doesn't necessarily mean you have a *"good"* regression.

- Worries about selection bias and omitted variables still apply.