ECON 256: Poverty, Growth & Inequality

Jack Rossbach

Regression Analysis

Suppose we have a dataset with values for two series, X and Y

Correlation tells us the direction of a relationship and the strength in terms of linearity

- If X and Y are positively correlated, then if X increases, on average, so does Y

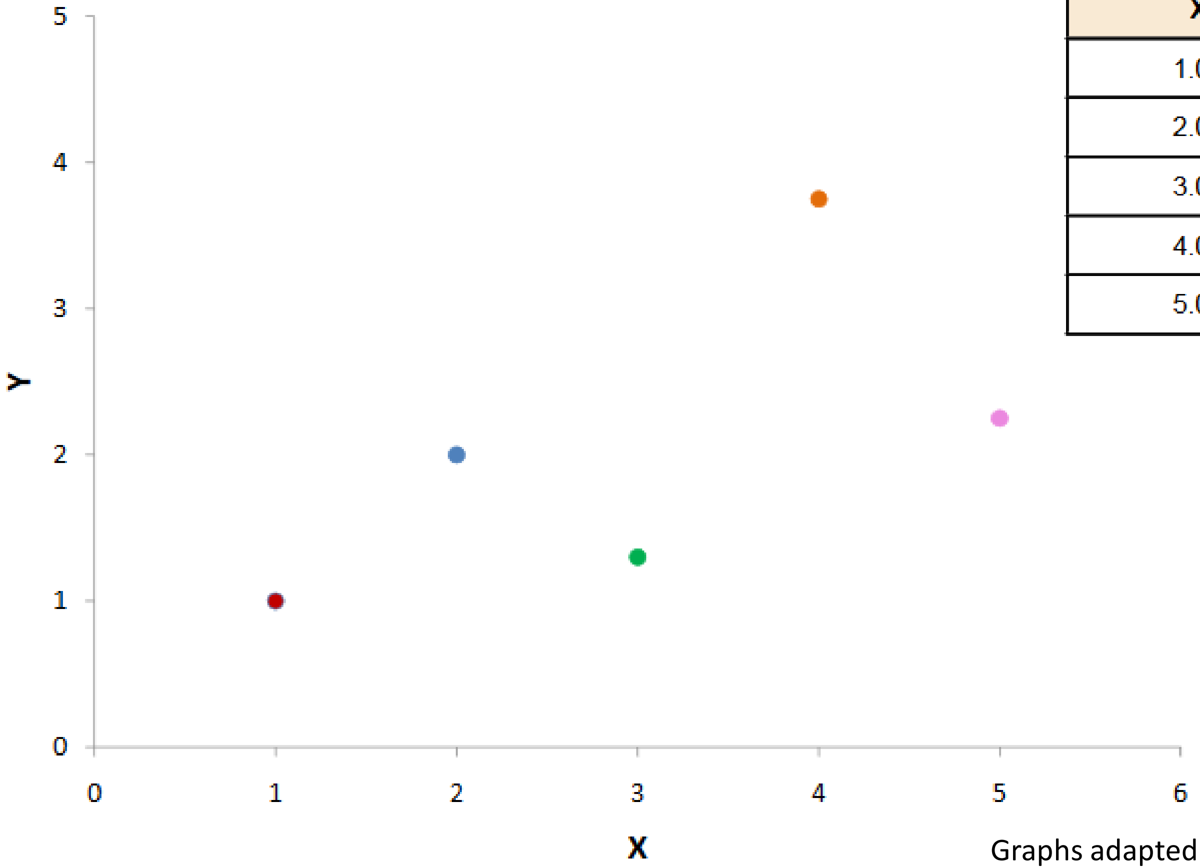
Regression Analysis tells us how much Y increases when X increases by a given amount

- Can fit the data with a linear equation of the form

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

- Where α and β are constants indicating best fit line, and ϵ_i is an error term for observation i
- Use estimated β to make predictions: $\hat{Y} = \alpha + \beta$X (Expected Y for a given value of X)

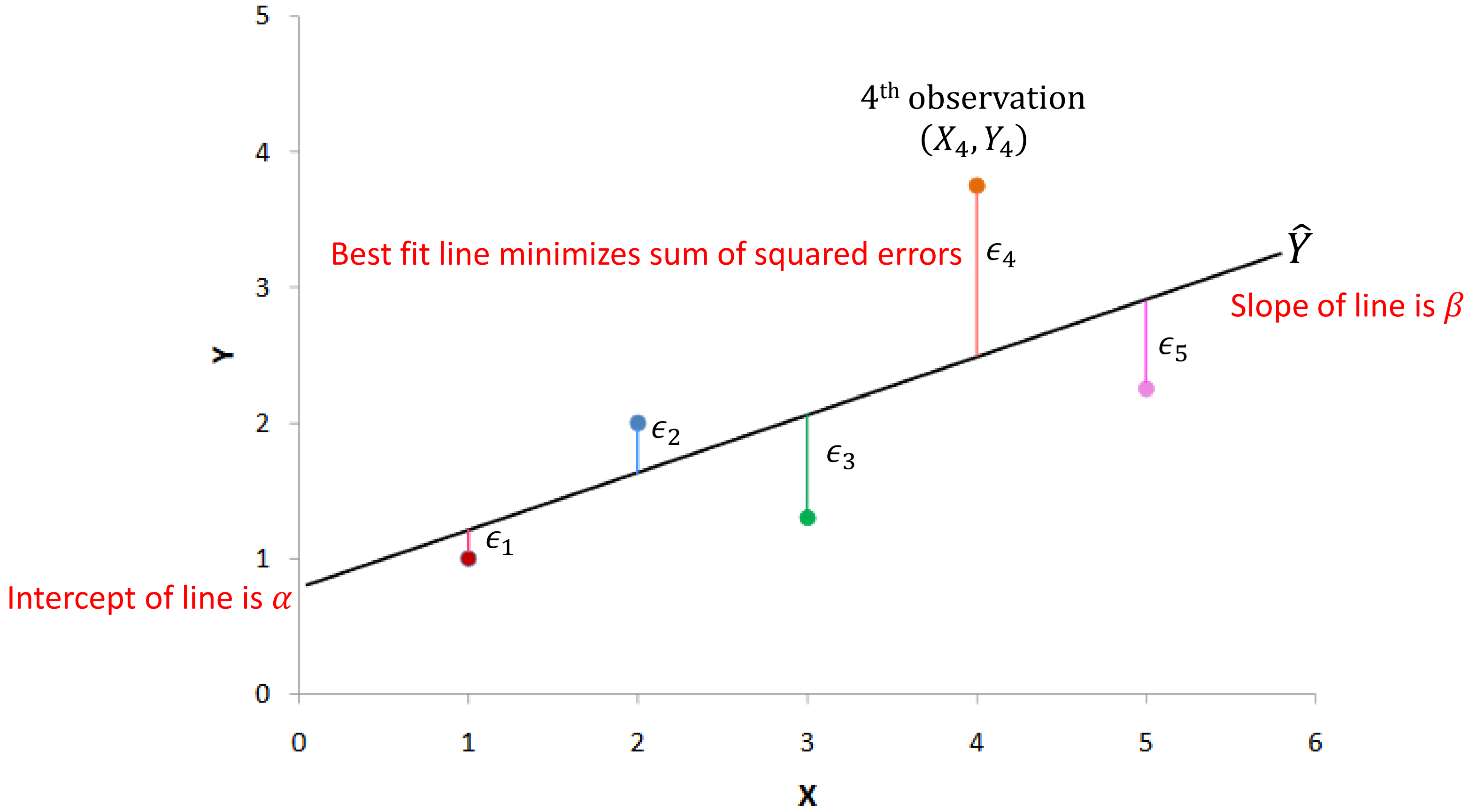
Regression Analysis: Start with Some Data



X	Y
1.00	1.00
2.00	2.00
3.00	1.30
4.00	3.75
5.00	2.25

Graphs adapted from [Introduction to Statistics]

Regression Analysis: Construct Best Fit Line



Interpreting Regression Lines

End up with regression line of form

$$\hat{Y} = \alpha + \beta X$$

- α is intercept of best fit line, β is slope of best fit line

Interpreting Regression Lines

End up with regression line of form

$$\hat{Y} = \alpha + \beta X$$

- α is intercept of best fit line, β is slope of best fit line

Our focus will be on how changing one variable affects the other

- Suppose X increases to X', then we predict Y will increase by

$$\Delta\hat{Y} \equiv \hat{Y}' - \hat{Y} = \beta(X' - X)$$

- Example, suppose β is 0.5, then increasing X from 15.5 to 17.5 will increase Y by 1 as:

$$\Delta\hat{Y} = 0.5 \times (17.5 - 15.5) = 0.5 \times (2) = 1$$

Standard Errors and Significance

When we run regressions we end up with estimated coefficients and a standard errors

- Standard Errors tell us how precise estimated coefficients are

- Can do confidence intervals for estimated coefficient

$$\text{95\% Confidence Interval for } \beta \approx [\beta_{est} - 2 \times \text{SE}, \beta_{est} + 2 \times \text{SE}]$$

Standard Errors and Significance

When we run regressions we end up with estimated coefficients and a standard errors

- Standard Errors tell us how precise estimated coefficients are
- Can do confidence intervals for estimated coefficient

$$95\% \text{ Confidence Interval for } \beta \approx [\beta_{est} - 2 \times \text{SE}, \beta_{est} + 2 \times \text{SE}]$$

- Suppose we have $\beta_{est} = 5$ and $SE = 1$. Then

$$95\% \text{ Confidence Interval for } \beta \approx [5 - 2 \times 1, 5 + 2 \times 1] = [3,7]$$

- This means we are 95% sure, that β falls somewhere between 3 and 7 (if our regression specification is correct! It usually isn't)
- Our estimate is **significant** at a 0.05 significance level if 0 is **NOT** in that confidence interval

Interpreting Regression Lines

Can also do regressions of the form

$$\log \hat{Y} = \alpha + \beta \log X$$

- Interpretation is how much 1% increase in X increases Y in % (for small changes)
- Example, $\alpha = 10$, $\beta = 0.5$

$$\log \hat{Y} = 10 + 0.5 \times \log X$$

- Then 1% increase in X will be a 0.5% increase in Y
- Note that the form we run regressions in matters. Won't get same results with linear and log regressions. When you take Econometrics you will talk about when to take transformations.