Econ 5/Poli 5D
Data Analytics/ Social Sciences

Shubhro Bhattacharya
UC San Diego

Lab-4
January 29, 2026

How to access course materials?

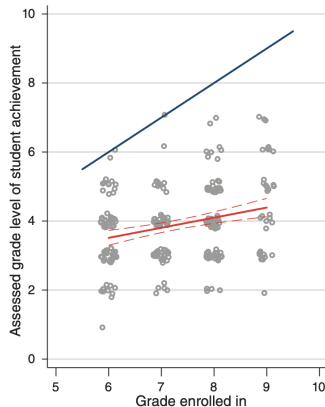Course materials on my website

TODAY: MINDSPARK RCT (LAB 4)

- **Motivation:** can personalized, computer-assisted learning raise test scores?
- **Data:** Mindspark student-level dataset (Muralidharan, Singh & Ganimian 2019)
- **Agenda:**
  1. Workflow & reproducibility (paths, clean start)
  2. Descriptives & graphs: treated vs control
  3. Regression + interpretation
  4. (Bonus) heterogeneity by gender and age
- Then: we run the do-file top-to-bottom and fill in code/interpretation together

WHY THIS STUDY? (RCT LOGIC)

- **Treatment** (Mindspark): adaptive practice + instant feedback
- **Control**: business-as-usual schooling
- Because assignment is randomized, a simple treated-vs-control comparison has a causal interpretation
- Today we will focus on:
  - Visualizing mean differences
  - Estimating linear relationships with regression
  - Checking whether patterns differ across subgroups

Panel A. Math

Panel B. Hindi
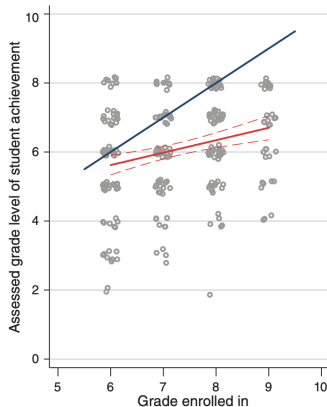
Many students are behind their enrolled grade level: **teaching at the right level** is crucial

KEY CONCEPT 1: DESCRIPTIVES + CLEAR FIGURES

- We will plot **average math scores** by treatment status
- Baseline vs endline outcomes:
    - `per_math1` = baseline math score (
    - `per_math2` = endline math score (
- Stata tools used:
    - Set a readable theme: `set scheme plotplainblind`
    - Bar means: `graph bar (mean) var, over(treat)`
    - Always label axes/titles so figures are self-contained

KEY CONCEPT 2: REGRESSION + INTERPRETATION

- Regression in the do-file: `reg per_math1 ses_index`
- Interpreting the slope β:
  - A 1-unit increase in `ses_index` changes predicted math score by β
  - Because scores are in 0–1 units, multiply by 100 for % points
- Prediction and comparisons:
  - Predicted score at `ses_index = 0.5`: $\hat{y} = \hat{\alpha} + 0.5\hat{\beta}$
  - Difference between 0.5 and -0.5: $\hat{\beta}(0.5 - (-0.5))$

KEY CONCEPT 3: VISUALIZING + SUBGROUP ANALYSIS

- Scatter + line of best fit:
  - `twoway (scatter per_math1 ses_index) (lfit per_math1 ses_index)`
  - This shows both **raw variation** and the **linear trend**
- Optional: **conditional regressions** (heterogeneity)
  - By gender: `if st_female1==1` vs `==0`
  - By age: create an indicator `above_med_age` using the median
- When comparing groups, keep the outcome and covariate the same; only change the sample