Econ 174, Section 101/103 Week 5

Joshua Blumenstock

jblumenstock@berkeley.edu

Please take out a piece of paper, and write the following on the paper:
1. Stuff you would like to do more of in section
2. Stuff you would like to do less of in section
3. Any other feedback (optional)

Do not write your name on the paper (unless you want to)
Put in in the folder by the door on your way out today

Today

- Questions?
  - Email me if you need a group!
- Difference-in-Difference
- Clustering
- Stata (woohoo!)
- Duflo

Diff-in-Diff

- Example: Tennesee STAR experiment on class size
  - The Tennessee STAR experiment randomly assigned 11,600 elementary school students and their teachers to a small class, regular-size class or regular-size class with a teacher-aide. The experiment began with the wave of students who entered kindergarten in 1985, and lasted for four years. After the third grade, all students returned to regular-size classes.
  - Say we have a dependent variable, test scores, Y_{ig} for each student i in group g, and the independent variable X_g which only varies at the group level.
- What is the "simple difference"?
- What is the "double difference"?
  - What's the "identifying assumption"?
- Why might you not need diff-in-diff?
- How measure in a regression?
- How to measure in stata?

- Need help? read this: http://bit.ly/eF2mqc

Clustering

$$Y_{it} = \alpha + \beta X_{it} + u_i + e_{it}$$

- What is clustering?
- Why cluster?
  - OLS assumes errors are uncorrelated (e_{it} are i.i.d.)
  - i.e. each observation is independent of all other observations in the dataset
  - Clustering allows for correlation across observations
- How to cluster?
  - Rule of thumb: cluster at level of randomization
- Where to cluster in STAR? In PROGRESA?
- More info? Try http://bit.ly/e6JyzP

STATA Exercise

1. Which cars have the lowest and highest values of mpg?

2. Do foreign cars have the same mpg as non-foreign cars? What is the p-value associated with this difference? Use `regress` and/or `ttest`

3. Rerun the regression in q2, but include a number of control variables (you choose). How do the controls affect the point estimate? The standard error? The R-squared? What does all this mean?

4. Test to see if there is an interaction between being foreign cars and the gear ratio, on mpg. Is there a differential effect of the gear ratio on mpg for foreign and domestic cars?

STATA Exercise

- Create a new variable rank that gives the relative price of the car: Use `sort, gen rank = _n`
- Create the variable `cost_quartile`:

  =1 if cost in first quartile (0-25%), 2 if (26-50%), etc…

  Hint: `xtile`
- Compute average mpg for each cost quartile, and save this in a new variable `quartile_mpg`
  - Hint: `bysort, egen`
- Plot average mpg by cost quartiles
  - Hint: `scatter`

STATA Exercises

- Plot the relationship between mpg and price
  - Draw the regression line and add a title to the plot. Use + signs for point markers, use a dashed yellow regression line.
  - Is the relationship statistically significant?
  - Bonus: label the points with the car model
  - Bonus: add confidence intervals

Duflo

- Why not compare 2-6 to 12-17
  - Age effects
  - Overestimate vs. underestimate
- Counterfactual assumption / identifying assumption

Duflo

TABLE 3—MEANS OF EDUCATION AND LOG(WAGE) BY COHORT AND LEVEL OF PROGRAM CELLS

	Years of education			Log(wages)		
	Level of program in region of birth			Level of program in region of birth		
	High (1)	Low (2)	Difference (3)	High (4)	Low (5)	Difference (6)
Panel A: Experiment of Interest						
Aged 2 to 6 in 1974	8.49	9.76	−1.27	6.61	6.73	−0.12
	(0.043)	(0.037)	(0.057)	(0.0078)	(0.0064)	(0.010)
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Difference	0.47	0.36	0.12	−0.26	−0.29	0.026
	(0.070)	(0.038)	(0.089)	(0.011)	(0.0096)	(0.015)
Panel B: Control Experiment						
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Aged 18 to 24 in 1974	7.70	9.12	−1.42	6.92	7.08	−0.16
	(0.059)	(0.044)	(0.072)	(0.0097)	(0.0076)	(0.012)
Difference	0.32	0.28	0.034	0.056	0.063	0.0070
	(0.080)	(0.061)	(0.098)	(0.013)	(0.010)	(0.016)

Notes: The sample is made of the individuals who earn a wage. Standard errors are in parentheses.

- "Placebo test"
- Standard Errors vs. T-stats vs. p-values

Duflo

$$(1) \quad S_{ijk} = c_1 + \alpha_{1j} + \beta_{1k} + (P_j T_i)\gamma_1$$

$$+ (\mathbf{C_j} T_i)\delta_1 + \varepsilon_{ijk}$$

where S_{ijk} is the education of individual i born in region j in year k, T_i is a dummy indicating whether the individual belongs to the "young" cohort in the subsample, c_1 is a constant, β_{1k} is a cohort of birth fixed effect, α_{1j} is a district of birth fixed effect, P_j denotes the intensity of the program in the region of birth, and $\mathbf{C_j}$ is a vector of region-specific variables.

- Fixed effects
- Interactions