Identification: Randomized Experiments

Bakhrom Mirkasimov

June 5, 2015

WIUT
Tashkent, Uzbekistan

Randomized Experiments: **Literature**

- Angrist and Pischke, Chapter 2, Mostly Harmless Econometrics (2009)
- Angrist and Pischke, Chapter 1, Mastering 'Metrics (2015)
- Heckman, J. and J. Smith (1995), "Assessing the Case for Social Experiments". Journal of Economic Perspectives, 9(2): 85-110
- Robert J. Lalonde (1986), "Evaluating the Econometric Evaluations of Training Programs with Experimental Data". American Economic Review, 76(4): 604-620

Introduction

- We are considering "Social Experiments" as opposed to laboratory experiments. That is, field experiments in real world settings.
- A lot of the language in the experimental literature is borrowed from the medical literature.
- Question: Does class size have a causal effect on student achievement?

Correlation vs. Causation

- Let's refresh: not all correlations are causal relationships.
- Many reasons why, for example, education or health & income may be correlated:
  - Causality
  - Reverse causality
  - Simultaneity
  - Omitted variables / confounding
  - Spurious correlation

Let’s practice skepticism

Read the following two correlations:

- Job applicants with names that are common among African-Americans are less likely to get an interview.
- More educated people are likely to earn more than less educated people.

A field experiment on Labor Market Discrimination (2003)

- For causality, we need to know that if the name was the *only* thing that changed, then they would still be less likely to be called for an interview.
- See paper: “Are Emily and Greg are more Employable than Lakisha and Jamal?” by Bertrand and Mullainathan
- In this experiment, fake resumes were sent; each resume is *randomly* assigned either a very African-American sounding name or a very White sounding name.
- Main finding: those with “white-sounding” names were 50% more likely to get an interview.

More educated people have higher income

- Key problem: omitted variables
- Higher wages may reflect ability, not schooling
- Actual cause of outcome is unobserved
- Intervention is non-random (ex: more schools, more teachers, etc).

Challenge: the counterfactual

- We can’t observe the same person both with and without the change in the treatment – the counterfactual is unobservable.
- Finding the right counterfactual is challenging; need to estimate counterfactual.
- Only reason for the difference in outcomes is due to the intervention; no other reason for differences in outcomes of treated and counterfactual.

Identification

- Key to identifying the causal effect of treatment: removing and dealing with potential sources of bias.
- One way to remove the bias terms: randomization.
- Key Idea – Randomization: people are randomly assigned treatment and control groups.
- With a large sample, we can make the two groups more or less identical.

Randomization

- Selection into treatment is purely due to chance, so that everyone has the same probability of receiving the treatment.
- By construction, selection into treatment is independent of potential outcomes (particularly, from the idiosyncratic gains from treatment).

The Selection Problem: Example

- Simple example introducing notation and the selection problem: "Do hospitals make people healthier?"
- Data from National Health Interview Survey (NHIS) 2005
- Health status measured from 1 (poor health) to 5 (excellent health)

Group	Sample Size	Mean Health Status	Std. Error
Hospital	7,774	3.21	0.014
No hospital	90,049	3.93	0.003

Basic Notation (1)

- Think of hospital treatment as a binary random variable: $D_i = [0,1]$
- Outcome of interest (here health status) is denoted: Y_i
- Is Y_i affected by hospital care?
- For each individual there are two potential outcomes (health variables):

$$\textit{Potential outcome} = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

Basic Notation (2)

- Y_{0i} is the health status of an individual had she not gone to hospital (irrespective of whether she actually went) and Y_{1i} is the health status of an individual if she goes to hospital.
- The causal effect of hospital treatment is:

  $\mathbf{Y_{1i} - Y_{0i}}$

The observed outcome Y_i can be written terms of potential outcomes as: $Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i$

This is useful because $Y_{1i} - Y_{0i}$ is the causal effect of hospitalization.

A fundamental problem is that we cannot observe both Y_{1i} and Y_{0i} for each individual. We can therefore not directly observe:

$E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$

We therefore have to estimate the average effect of hospitalization by comparing average health of those who were and those who were not hospitalized.

What are we actually measuring if we compare these averages?

$$\underbrace{E[Y_i|D_i = 1] - E[Y_i|D_i = 0]}_{\textit{Observed difference in average health}} = \underbrace{E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]}_{\textit{average treatment effect on the treated}}$$
$$+ \underbrace{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]}_{\textit{Selection bias}}$$

Random assignment solves the Selection Problem

- Random assignment of D_i (treatment) makes treatment D_i independent of potential outcomes Y_i.
- If D_i is independent of Y_i, we can swap selection bias terms.
- Therefore, with random assignment, the selection bias term vanishes!

- Question: but who is being randomized?
- Depends on your research goal (ATE or ATOT)
  - Average causal effect on a randomly selected person?
  - Average causal effect on the treated, when treatment is randomly assigned?
  - Average causal effect on the treated of a treatment that is usually not randomly assigned?

Average Treatment Effect (ATE)

- ATE = $E[\Delta] = E[Y_1] - E[Y_0]$
- ATE is the expected treatment effect for a person randomly selected from the entire population
- ATE is consistently estimated by the difference between sample mean outcomes of participants and non-participants
- Pure randomization is the theoretical ideal for ATE and the benchmark for non-experimental methods

$E[Y_i]$

"The mathematical expectation of a variable, Y_i, written $E[Y_i]$, is the population average of this variable. If Y_i is a variable generated by a random process, such as throwing a die, $E[Y_i]$ is the average in infinitely many repetitions of this process. If Y_i is a variable that comes from a sample survey, $E[Y_i]$ is the average obtained if everyone in the population from which the sample is drawn were to be enumerated."

Average Treatment Effect on the Treated (ATOT)

- We are aiming at ATOT – randomization conditional on *D*
- ATOT = $E[Y_1| D=1] – E[Y_0| D=1]$
- **D=1 means**: receiving the treatment under normal conditions
- In practice, most treatments are not randomized. For example: publicly sponsored training programs operate normally without randomization and are administered to (self)-selected volunteers.

ATOT

- Pure randomization is rare
- If we pursue ATOT, we can apply randomization at different stages if the program (treatment) is run under normal conditions:
  - eligibility *for* …
  - application *for* …
  - acceptance *into* …

Confounders?

- Question: suppose there are other observable factors (X) that influence outcomes besides the treatment. Shouldn’t we “control for” these other factors (by regression techniques?)

- Answer: **NO!** If we have enough data, randomization will balance the distributions of these factors (and also the unobservable factors) in the treatment group and control group

Tennessee Project STAR

- Question: Causal effect of class size on student achievement?
- Non-experimental studies find <u>little or no link</u>
- Experimental study in Tennessee (known as Tennessee Student/Teacher Achievement Ratio – STAR)
- Cost: $12 million
- 1985-1986 cohort of kindergartners (11,600 children) followed into 3rd grade for four years

Tennessee STAR experiment

- Krueger (1999) econometrically re-analyzed a randomized experiment of the effect of class size on student achievement.
- 3 treatment groups: small classes (13-17) students, regular classes (22-25) and regular classes (22-25) students with a full-time teacher’s aide.
- Randomization occurred within schools.

TABLE 2.2.1

Comparison of treatment and control characteristics in the Tennessee STAR experiment

	Class Size			P-value for equality
Variable	Small	Regular	Regular/Aide	across groups
Free lunch	.47	.48	.50	.09
White/Asian	.68	.67	.66	.26
Age in 1985	5.44	5.43	5.42	.32
Attrition rate	.49	.52	.53	.02
Class size in kindergarten	15.10	22.40	22.80	.00
Percentile score in kindergarten	54.70	48.90	50.00	.00

Notes: Adapted from Krueger (1999), table I. The table shows means of variables by treatment status for the sample of students who entered STAR in kindergarten. The P-value in the last column is for the F-test of equality of variable means across all three groups. The free lunch variable is the fraction receiving a free lunch. The percentile score is the average percentile score on three Stanford Achievement Tests. The attrition rate is the proportion lost to follow-up before completing third grade.

TABLE 2.2.2

Experimental estimates of the effect of class size on test scores

Explanatory Variable	(1)	(2)	(3)	(4)
Small class	4.82	5.37	5.36	5.37
	(2.19)	(1.26)	(1.21)	(1.19)
Regular/aide class	.12	.29	.53	.31
	(2.23)	(1.13)	(1.09)	(1.07)
White/Asian	—	—	8.35	8.44
			(1.35)	(1.36)
Girl	—	—	4.48	4.39
			(.63)	(.63)
Free lunch	—	—	−13.15	−13.07
			(.77)	(.77)
White teacher	—	—	—	−.57
				(2.10)
Teacher experience	—	—	—	.26
				(.10)
Teacher Master's degree	—	—	—	−0.51
				(1.06)
School fixed effects	No	Yes	Yes	Yes
R^2	.01	.25	.31	.31

Notes: Adapted from Krueger (1999), table V. The dependent variable is the Stanford Achievement Test percentile score. Robust standard errors allowing for correlated residuals within classes are shown in parentheses. The sample size is 5,681.

More experimental examples

- See Angrist and Pischke’s book (2015), Chapter 1
- The RAND Health Insurance Experiment (HIE) – ran from 1974 to 1982
- Oregon’s health insurance lottery – the state-run Oregon Health Plan (OHP)
- The RAND and Oregon findings are remarkably similar
- DellaVigna, List and Malmendier (2010) – charitable giving

Maybe Parents Don’t Like Boys Better: A follow-up to the recent column about whether daughters cause divorce (by Steven Landsburg)

“Other readers accepted the reality of the 5 percent difference but questioned the conclusion that daughters cause divorce. After all, **marriages differ in all sorts of ways that might be relevant—financial stresses, infidelity, emotional distance. The phrase "correlation does not imply causation" popped up a lot. But** in this case, correlation does imply causation, and here's why: If you take 3 million people, have them all flip coins, and divide them into two groups according to whether their coins came up heads or tails, **then the two groups are going to look statistically identical in every way—same average income, same average intelligence, same average height. That's called the law of large numbers,** and it works for two reasons—first, the sample size is huge, and second, coin flips are random. Now do the same thing, dividing your 3 million people according to the gender of their last-born child. The same thing happens—parents of boys are going to be statistically identical in every way to parents of girls, because you've still got a huge sample size and because the sex of a child is as random as a coin flip. Since everything else is equal, the only thing that can be causing the difference in divorce rates is the gender of the children”

http://www.slate.com/articles/arts/everyday_economics/2003/10/maybe_parents_dont_like_boys_better.html

Angrist and Pischke (2009) conclude:

- “The STAR study, an exemplary randomized trial in the annals of social science, also highlights the logistical difficulty, long duration, and potentially high cost of randomized trials. In many cases, such trials are impractical”
- “Still, hypothetical (ideal) experiments are worth contemplating because they help us pick fruitful research topics. … The mechanics of an ideal experiment highlight the forces you’d like to manipulate and the factors you’d like to hold constant. Research questions that cannot be answered by an experiment are FUQs: fundamentally unidentified questions”

BREAK!

Roy Model

- Literature: A.D. Roy (1951), “Some Thoughts on the Distribution of Earnings”. Oxford Economic Papers, 3(2): 135-146
- Quick Review of Rubin’s Potential Outcome Model of Causality
- Remember? Difference in group means = Average causal effect + Selection bias

What is the causal effect of *D* on *Y*?

- Y_1 = Potential Outcome as a Hunter
- Y_0 = Potential Outcome as a Fisherman

- D = 1 Hunting
- D = 0 Fishing (not hunting)
- $Y = Y_1 \times D + Y_0 \times (1-D)$ observed outcome
- Question: Who becomes a hunter? Who becomes a fisherman?

- Roy, 1951: “to begin with, let a very simple community be considered in which a member of the working population has the choice of only two occupations: hunting and fishing, for example”.
- Distribution of incomes in the population:

  potential income from hunting: $Y_1 = E[Y_1] + U_1$

  potential income from fishing: $Y_0 = E[Y_0] + U_0$
- Here, U_1 and U_0 are a person’s talents for hunting and fishing, respectively

Self-selection

- Who goes into hunting? (who selects the treatment?)

  Those for which $Y_1 > Y_0$
- **Self-Selection (rule)** into D=1 on the basis $Y_1 > Y_0$
- A person chooses the occupation where he or she earns more.

- Therefore, selection is based on (perceived) rewards! Treatment dummy explicitly depends on outcomes.

Two Papers to Discuss

- On Education (2007): “Remedying Education: Evidence from two randomized experiments in India” by Banerjee, Cole, Duflo and Linden. The Quarterly Journal of Economics, 122(3): 1235-1264
- On Deworming (2004): “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities” by Miguel and Kremer. Econometrica, 72(1): 159-217
- We will replicate ‘Deworming’ paper (see files)

Health: Worms

- 1 in 4 people are infected with intestinal worms around the globe.
- Diarrhea is frequent among children.
- Common: anemia, fatigue, lethargy, stunting
- Treatment care is very cheap; only requires one or two pills per year (less than 0.50 USD per child).
- Question: What is the effect of health interventions on economic outcomes?

Deworming Program

- Seventy-five of 89 rural primary schools in rural western Kenya (broadly representative of rural Kenya in education, health and worms).
- Randomization took place at *school level*
- Randomized into three treatment groups, 25 schools each:
  - Group 1: deworming drugs and health education in 1998-2003
  - Group 2: 1999-2003
  - Group 3: 2001-2003

Aim

- This study tracks down children who were in primary school during the deworming period (group 1 and 2 = treatment group, group 3 = control) in 2007-2009.
- Evaluate the effect of a deworming program implemented in Western Kenya (1998-2001) on health, school absenteeism and test scores.

Findings

- Rates of serious worm infection fell from 52% to 25%.
- Absenteeism fell by ¼, or 7.5 percentage points.
- Deworming also reduced re-infection among other community members: positive externalities.

Long-run Findings

- New paper (2011) by Baird, Hicks, Kremer and Miguel
- Data: Kenya Life Panel Survey (KLPS-1, 2003-2005 and KLPS-2, 2007-2009)
- Followed the same children who participated in the deworming project ten years later; most participants are 19-26 years old.
- Earnings are over 20% higher and hours worked increased by 17% in the treatment group.

Education: Balsakhi

- Education quality is low (the problem!) in developing countries:
  - High teacher absence
  - High student absence
  - Low achievement
- Question: How to improve school quality (pupil performance) in a cost-effective way?
- The Balsakhi program – one of Pratham's first program.
- Pratham was established in 1994 – the largest NGO that provides support to education in India.

Balsakhi

- The child’s friend!
- A young woman, from the children’s community, with 10-12 grade education who is working in the classroom with the students who are lagging behind in class (20 students), for about 2 hours per day, and focus on basic skills.
- She receives basic 2 weeks training and on the job support.

Issues/Solutions

- In two cities, Vadodara and Mumbai, half of schools randomly selected to get a Balsakhi teacher for either 3rd or 4th grade.
- “Intention-to-Treat” (ITT)
- Check for partial compliance (program participation) and attrition (tracking subjects).
- External validity: two separate locations/cities

Evaluation design: Vadodara, Gujarat (year 1)

	GROUP A	GROUP B
GRADE 3	TREATMENT	CONTROL
GRADE 4	CONTROL	TREATMENT

Evaluation design: Vadodara, Gujarat (year 2)

	GROUP A	**GROUP B**
GRADE 3	CONTROL	TREATMENT
GRADE 4	TREATMENT	CONTROL

Findings

- 0.14 standard deviation increase in average test scores in year 1.
- 0.26 standard deviation increase in year 2.
- One year after the program, only gains for bottom third remain, fading to 0.10 standard deviation increase.
- Balsakhi has very large direct effects; the CAL program had a strong effect on math scores.

Some concerns about RCTs

- “Hawthorne” / “John Henry” effects
- Generalizability / external validity
- Externalities / supply side changes
- General equilibrium effects
- Can’t apply randomization to all research questions
- Randomization bias
- Attrition bias
- Substitution bias

What is a good experimental paper?

Combine experimental data with economic theory

ANY QUESTIONS / COMMENTS?

THANK YOU VERY MUCH!

Additional Video

- NBER Summer Institute Econometric Lecture (2009) by John List and Michael Kremer

  www.streamingmeeting.com/webmeeting/matrixvideo/nber/index.html

  www.nber.org/SI_econometrics_lectures.html
- J-PAL: www.povertyactionlab.org
- Harvard Dataverse: https://dataverse.harvard.edu