AREC 345: Global Poverty & Economic Development

Lecture 2:

The Problem of Causal Inference

Professor: Pamela Jakiela

Department of Agricultural and Resource Economics
University of Maryland, College Park

Research Questions

Three questions:

1. Did the slave trade cause under-development in Africa?
2. Does foreign aid lead to higher economic growth in poor countries?
3. Do entrepreneurship programs increase young people's incomes?

Each is about the relationship between a **treatment** and an **outcome**

1. **Outcome**: the variable or measure that we are interested in studying (e.g. under-development, growth, or individual income)
2. **Treatment**: the event or policy that might have impacted the outcome (e.g. the slave trade, aid, or an entrepreneurship program)

You Can't Always Get [the Answers] You Want

To answer these questions, we need to know what would have happened to "treated" countries or individuals in the absence of treatment

1. Did the slave trade cause under-development in Africa?
   - Angola and Benin were (very) exposed to the Atlantic slave trade.
   - How high would per capita income be in those countries today if the Atlantic slave trade had never happened?
2. Does foreign aid lead to higher economic growth in poor countries?
   - Countries (should) receive aid because they are poor/struggling; would they be better or worse off in the absence of aid?

Unfortunately, we can never know the answers to these questions

You Can't Always Get [the Answers] You Want

This is not because the world only contains so many countries, or because economic history only happened in one particular way

Our micro-level research question is equally unanswerable:

3. Do entrepreneurship programs increase young people's incomes?
   - What would participants' incomes be in the absence of the program?

We can never know the answers to this question

- Why not? Because each person either participated in the program or did not participate in the program (you only live once).
  - We cannot observe the **counterfactual**
- What if we compared program participants to non-participants?

Comparing Participant and Non-Participants

Example: young women from Nairobi slums who were invited to participate in a **microfranchising** program that provided vocational and life skills training and helped them to start small businesses (e.g. salons)

Comparing Participant and Non-Participants

Only 60 percent of eligible applicants participate in the program

How do their incomes compare to those who didn't participate?

	Participation in Program	
	Did Program	**No Program**
Weekly income (in USD)	9.98	8.38
Living with a parent	0.49	0.43
Any work experience	0.57	0.51
Completed secondary school	0.39	0.44
Any vocational training	0.29	0.36

Incomes higher among participants, but there are other differences

⇒ Did the program **cause** the difference in income?

Terminology: Potential Outcomes

We are interested in the relationship between program participation and some outcome that may be impacted by the program (eg. income)

- Borrowing a term from medicine, we call the program the **treatment**

Outcome of interest:

- Y = outcome we are interested in studying (e.g. income)
- Y_i = value of outcome of interest *for individual i*

For each individual, there are two **potential outcomes**:

- $Y_{0,i}$ = i's outcome she **doesn't** receive treatment
- $Y_{1,i}$ = i's outcome she **does** receive treatment

For any individual, we can only observe one potential outcome

Potential Outcomes: Example

Two women are invited to participate in the microfranchising program:

- **Alice** comes from a very poor family. She had to leave school after 9^{th} grade. She is extremely excited to start her own business.
  - $Y_{0,Alice} = \$8$
  - $Y_{1,Alice} = \$10$

$\Rightarrow$ **Impact of treatment on Alice:**

  - $Y_{1,Alice} - Y_{0,Alice} = \$10 - \$8 = \$2$

Potential Outcomes: Example

Two women are invited to participate in the microfranchising program:

- **Betty** comes from a better off family. She finished secondary school. She'd prefer an office job and doesn't see herself as an entrepreneur.
  - $Y_{0,Betty} = \$12$
  - $Y_{1,Betty} = \$11$

⇒ **Impact of treatment on Betty:**

- $Y_{1,Betty} - Y_{0,Betty} =$

Potential Outcomes: Example

Outcomes and Treatments

	Alice	**Betty**
Potential outcome without program: $Y_{0,i}$	8	12
Potential outcome with program: $Y_{1,i}$	10	11
Treatment effect: $Y_{1,i} - Y_{0,i}$	2	-1
Participates in the program?		
Actual income: Y_i		

Potential Outcomes: Example

Alice chooses to participate, while Betty chooses not to.

Question: Can a comparison of Alice's and Betty's incomes teach us anything about the impact of the microfranchising program?

- Answer: No.

Difference between Alice's actual income and Betty's actual income:

$$Y_{Alice} - Y_{Betty} = 10 - 12 = -2$$

Comparison suggests that microfranchising lowers income:

Alice's income is lower than Betty's income even though
Alice participated in the microfranchising program and Betty did not

Potential Outcomes: Example

What are we comparing when we compare Alice's income to Betty's?

$$Y_{Alice} - Y_{Betty} = Y_{1,Alice} - Y_{0,Betty}$$

$$= \underbrace{Y_{1,Alice} - Y_{0,Alice}}_{=2} + \underbrace{Y_{0,Alice} - Y_{0,Betty}}_{=-4}$$

Comparing those who choose treatment to those who do not conflates:

- $Y_{1,Alice} - Y_{0,Alice}$ = treatment effect on Alice (who is treated)
- $Y_{0,Alice} - Y_{0,Betty}$ = **selection bias**
  - Those who choose treatment and those who do not choose treatment would have different outcomes *even in the absence of treatment*

To Summarize

The **causal effect** of the program is not the same for Alice and Betty:

- Alice: $Y_{1,Alice} - Y_{0,Alice} = 2$
- Betty: $Y_{1,Betty} - Y_{0,Betty} = -1$

$\Rightarrow$ **Alice chooses to participate in the program because it makes her better off than she would be otherwise**

Comparing outcomes post-treatment does not identify the causal effect

- Alice and Betty have different incomes in the absence of the program

At this point, you might be wondering:

Wouldn't this issue disappear if we looked at a larger cross-section of the population? This is, after all, just a made up example...

Terminology: Average Causal Effects

What we actually want to know is the **average causal effect**, but that is not what we get from a difference in means comparison

Difference in group means

= average causal effect of program on participants + selection bias

Even in a large sample:

- People will choose to participate in a program when they expect the program to make them better off (i.e. when $Y_{1,i} - Y_{0,i} > 0$)
- The people who choose to participate are likely yo be different than those who choose not to... *even in the absence of the program*

Terminology: Average Causal Effects

P_i is a **dummy variable** equal to 1 if individual i participated

- $P_i = 1$ if i individual i participated (i.e. received treatment)
- $P_i = 0$ if i individual i did not participate

Y_i is the observed outcome for individual i

- $Y_i = Y_{1,i}$ if individual i if $P_i = 1$
- $Y_i = Y_{0,i}$ if individual i if $P_i = 0$

Terminology: Average Causal Effects

If n is the number of people, then the **unconditional mean** of Y_i is:

- $AVG_n[Y_i] = \frac{1}{n}\sum_{i=1}^{n} Y_i$

We'd like to estimate the **average causal effect** of the program, either on those who chose to participate or on the population at large

- $AVG_n[Y_{1,i} - Y_{0,i}] = \frac{1}{n}\sum_{i=1}^{n} Y_{1,i} - \frac{1}{n}\sum_{i=1}^{n} Y_{0,i}$

Our problem: we can't observe $Y_{0,i}$ for participants

- If we want to know the average causal effect on the *population*, we'd also need to observe $Y_{1,i}$ for non-participants — but we can't

Terminology: Average Causal Effects

The average Y_i among program participants is the **conditional mean**:

$$AVG_n[Y_i|P_i = 1] = \frac{1}{n_{P_i=1}} \sum_{i=1}^{n_{P_i=1}} Y_i$$

$$= \frac{1}{n_{P_i=1}} \sum_{i=1}^{n_{P_i=1}} Y_{1,i}$$

The conditional mean among non-participants is:

$$AVG_n[Y_i|P_i = 0] = \frac{1}{n_{P_i=0}} \sum_{i=1}^{n_{P_i=0}} Y_{0,i}$$

Terminology: Average Causal Effects

When we compare means for participants and non-participants:

$$\textbf{Difference in group means} = AVG_n[Y_i|P_i = 1] - AVG_n[Y_i|P_i = 0]$$

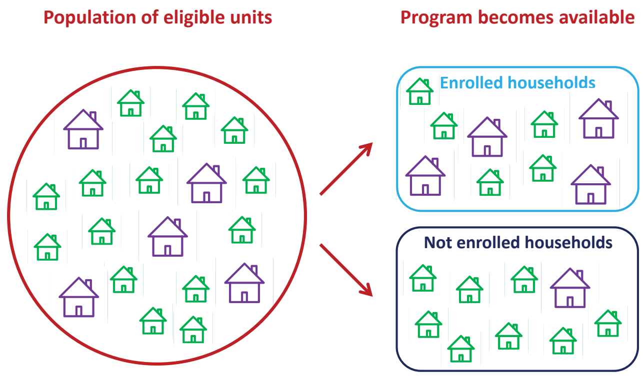
$$= AVG_n[Y_{1,i}|P_i = 1] - AVG_n[Y_{0,1}|P_i = 0]$$

Adding in $\underbrace{-AVG_n[Y_{0,i}|P_i = 1] + AVG_n[Y_{0,i}|P_i = 1]}_{=0}$, we get:

Difference in group means

$$= \underbrace{AVG_n[Y_{1,i}|P_i = 1] - AVG_n[Y_{0,i}|P_i = 1]}_{\text{average causal effect on participants}} + \underbrace{AVG_n[Y_{0,i}|P_i = 1] - AVG_n[Y_{0,i}|P_i = 0]}_{\text{selection bias}}$$

Selection Bias



How Do We Estimate Causal Impacts?

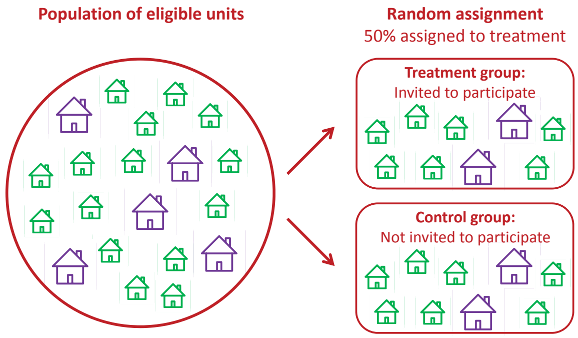
Experimental approach:

- **Random assignment to treatment:** eligibility for program is literally determined at random, eg. through pulling names out of hat

When this isn't feasible, we're left with **quasi-experimental approaches:**

- Natural experiments: historical accidents that create seemingly as-good-as-random variation in exposure to some treatment
  - Classic example: rainfall as a(n essentially random) shock
    conditional on where you grew up (and the weather pattern there), experiencing more sunny Fourth of July holidays as a child makes you more conservative and nationalistic as an adult
- Other approaches, e.g. controlling for time trends in both treatment and comparison groups ("difference-in-difference estimation")

Random Assignment Solves the Selection Problem



"Randomization works not by eliminating individual difference but rather by ensuring that the mix of individuals being compared is the same. Think of this as comparing barrels that include equal proportions of apples and oranges."

Random Assignment & the Law of Large Numbers

The **law of large numbers** tells us that a sample average can be brought as close as we like to the population average just by enlarging the sample

Example: imagine that I want to evaluate the impact (on test scores) of my brand new "get an A or your money back guarantee" program

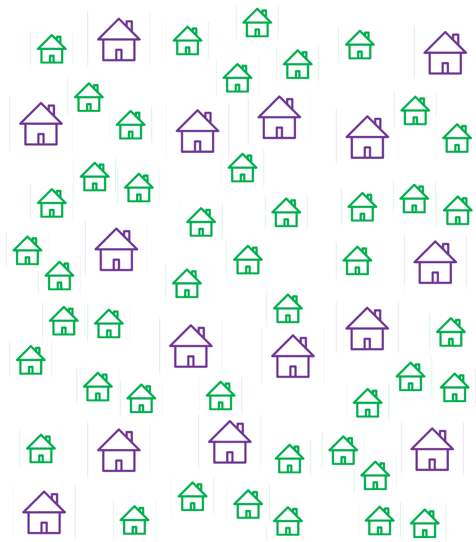
- I choose two students at random, one treatment and one control
- I flip a coin to determine which one gets an automatic A grade
- The other student is "the counterfactual" — no guaranteed A

How does the **treated** student's grade compare to the **control** student's?

Are these students comparable in the absence of the program?

Random Assignment & the Law of Large Numbers

Population of eligible households



25% purple households

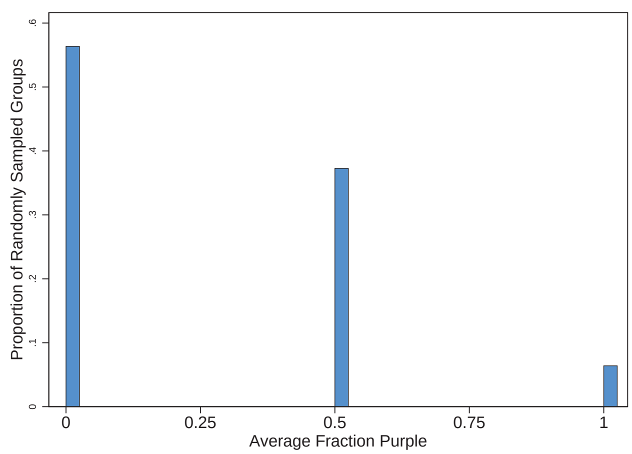
If you chose one at random, probability it is purple:
0.25

However,
any one house
(chosen at random)
is either purple or green.

What if you chose 2 HHs?

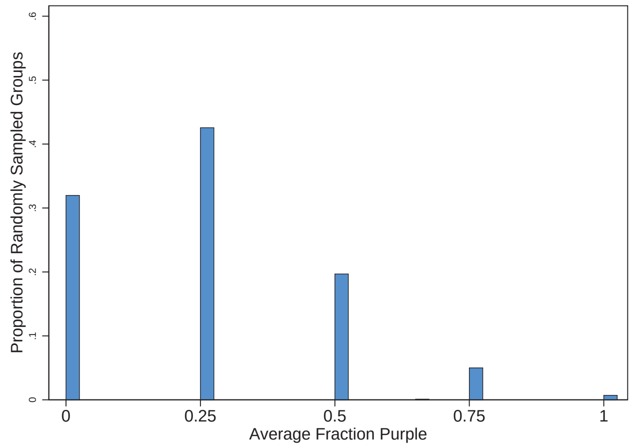
Random Assignment & the Law of Large Numbers

When you randomly sample groups of 2:



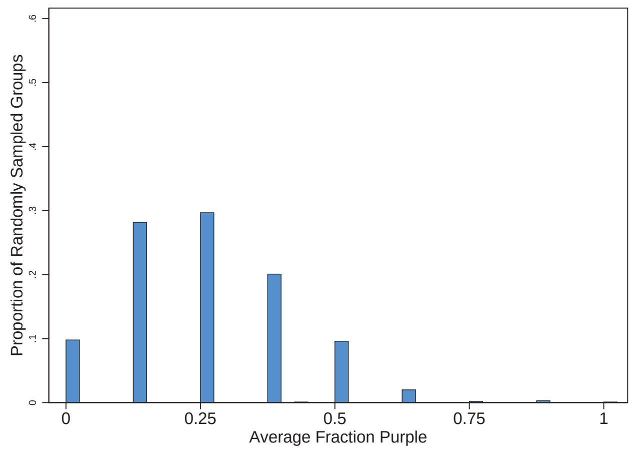
Random Assignment & the Law of Large Numbers

When you randomly sample groups of 4:



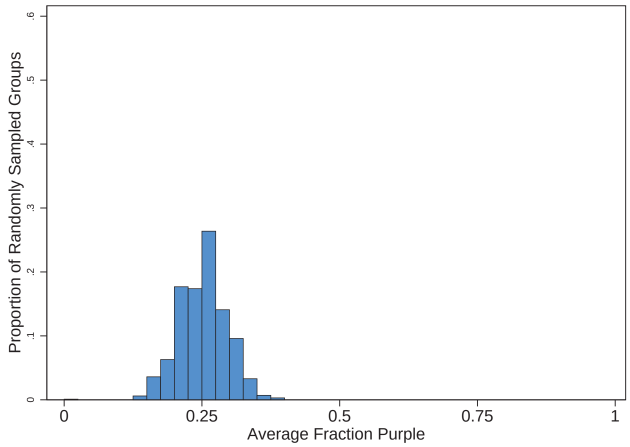
Random Assignment & the Law of Large Numbers

When you randomly sample groups of 8:



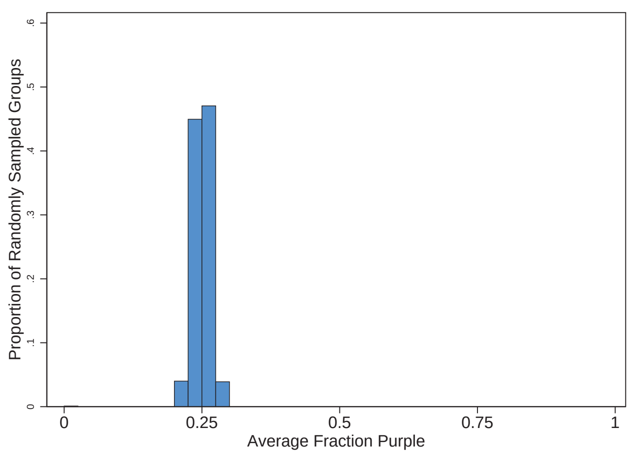
Random Assignment & the Law of Large Numbers

When you randomly sample groups of 100:



Random Assignment & the Law of Large Numbers

When you randomly sample groups of 1000:



Mathematical Expectation

The **mathematical expectation**, $E[Y_i]$, is the population average of Y_i

- Equivalent to the sample average in an infinitely large population
- For example, the mathematical expectation tells us the probability that a coin lands with heads facing up when you flip it; it is equivalent to the fraction heads after a (very) large number of flips

In a small sample, the average of Y_i might be anything; but the law of large numbers tells us that the average of Y_i gets very close to $E[Y_i]$ as the number of observations (that we are averaging over) gets large

Random Assignment Eliminates Selection Bias

Conditional expectation:

$E[Y_i|P_i = 1]$

The conditional expectation of Y_i given a dummy variable $P_i = 1$, is the average of Y_i in the population that has $P_i = 1$.

$E[Y_i|P_i = 0]$

The conditional expectation of Y_i given a dummy variable $P_i = 0$, is the average of Y_i in the population that has $P_i = 0$.

When treatment is randomly assigned,
the treatment, control groups are random samples of a single population (e.g. the population of all eligible applicants for the program)

$$\Rightarrow E[Y_{0,i}|P_i = 1] = E[Y_{0,i}|P_i = 0] = E[Y_{0,i}]$$

Expected outcomes are the same in the absence of the program

Random Assignment Eliminates Selection Bias

If treatment is random and $E[Y_{0,i}|P_i = 1] = E[Y_{0,i}|P_i = 0] = E[Y_{0,i}]$, the difference in means estimator gives us the average causal effect:

$$\textbf{Difference in group means} = E[Y_i|P_i = 1] - E[Y_i|P_i = 0]$$
$$= E[Y_{1,i}|P_i = 1] - E[Y_{0,1}|P_i = 0]$$

Adding in $\underbrace{-E[Y_{0,i}|P_i = 1] + E[Y_{0,i}|P_i = 1]}_{=0}$, we get:

Difference in group means

$$= E[Y_{1,i}|P_i = 1] - E[Y_{0,i}|P_i = 1] + E[Y_{0,i}|P_i = 1] - E[Y_{0,i}|P_i = 0]$$

$$= \underbrace{E[Y_{1,i}|P_i = 1] - E[Y_{0,i}|P_i = 1]}_{\text{average causal effect on participants}} + \underbrace{E[Y_{0,i}] - E[Y_{0,i}]}_{=0}$$

Summary: the Problem of Causal Inference

We want to know:

- What historical factors caused poverty (or slowed development)?
- What policies contribute to the elimination of poverty?

Knowing whether A causes B is really hard

(Because of selection bias)

Questions of causality can only be answered by analyzing data

In order to the answer them, you need a full set of empirical (i.e. statistical, econometric) tools in your toolkit

Study Guide: Key Terms

- average causal effect
- causality
- counterfactual
- difference in means
- law of large numbers
- mathematical expectation
- outcome variable
- potential outcomes
- selection bias
- treatment
- (un)conditional expectation
- (un)conditional mean