Experiments

PP580

Harris School of Public Policy
University of Chicago

Topic 7

Experiments seem easy

- Take sample of N individuals
- Flip coin for each
  - If heads, assign health insurance (D=1)
  - If tails, assign no health insurance (D=0)

Estimation with experiments

- Observe health Y at some future date
- Estimate ATE as
- $$\hat{\Delta}^{ATE} = \overline{Y}_1 - \overline{Y}_0$$
- Provided N is reasonably large, $\hat{\Delta}^{ATE}$ is a valid estimate of ATE

The Rationale Behind Experiments

When treatment status is randomly determined, we have

$$F(X, U|D = 1) = F(X, U|D = 0) = F(X, U)$$

- distribution of observable and unobservable characteristics for treated population
- = distribution of observable and unobservable characteristics for untreated population
- = distribution of observable and unobservable characteristics for the whole population.
- Implication: no selection problem

With no selection problem,potential outcomes are independent of treatment:

$$E(Y_{1i}|D_i = 1) = E(Y_{1i}|D_i = 0) = E(Y_{1i})$$
$$E(Y_{0i}|D_i = 0) = E(Y_{0i}|D_i = 1) = E(Y_{0i})$$

This allows us to estimate the ATE by

$$\hat{\Delta}^{ATE} = \overline{Y}_1 - \overline{Y}_0$$

The estimated impact of treatment is simply the mean difference in outcomes between the treated group and the control group.

Verifying random assignment

- If randomization worked correctly,characteristics of treatment and control groups should be similar
- This can be tested, at least for observables ("balance")
- Compare covariate means between $D = 0$, $D = 1$ groups

Estimating the ATE

- ATE can be estimated via simple regression:

$$Y_i = \beta + \Delta^{ATE} D_i + u_i$$

- Easy to show that

$$\hat{\Delta}^{ATE} = \overline{Y}_1 - \overline{Y}_0$$

- and that $\hat{\Delta}^{ATE}$ is unbiased under random assignment

Generalizing the regression

- Should you add other variables to the regression?

$$Y_i = X_i\beta + \Delta^{ATE} D_i + u_i$$

- What would you expect to change if you added regressors?
- What types of change would cause you concern?

Heterogeneity

- Can allow for heterogeneity w.r.t. observables:

$$\Delta(X_1 = x_1) = E(Y_{1i}|X_1 = x_1) - E(Y_{0i}|X_1 = x_1)$$

Advantages of Experiments

- Experimental data "solves" the evaluation problem to produce internally valid estimates of average treatment effects.
- Experiments permit policymakers to observe the effects of new kinds of treatment that have not previously been observed.
- Simplicity may be a virtue when communicating results to non-specialists.

Limits of Experiments

- *ATE* may or may not be an interesting parameter
- Most experiments are a black box
- Not a panacea, nor a substitute for thought

Problems with Experiments

Problems with experiments can be characterized into three main types:

- Implementation problems
- Threats to internal validity
- Threats to external validity

Implementation Problems

- Ethical considerations
  - Experiments may raise ethical issues
    - The experimental treatment may inflict harm
    - Denial of treatment may inflict harm
  - Program officials often equate "random assignment" to "denial of services to the needy."

- Limited Duration
  - Social experiments are limited in duration. This may be problematic if participants may take time to understand the nature of the tested treatment and react to it.
  - Participants may react differently to a treatment if they are aware that it is of limited duration as opposed to a policy change that's been implemented indefinitely.

Validity: Are the estimates meaningful?

There are two types of validity: internal and external validity

- External Validity: An externally valid estimate is a treatment effect estimate that can be validly applied to other populations.
- Internal Validity: An internally valid estimate is an unbiased measure of the treatment effect in the sample actually enrolled in an experiment.

Threats to External Validity

- Randomization bias
- Equilibrium effects

Randomization bias

- Different people may participate in experiments than would participate otherwise
- In medical experiments, refusal rates are higher for experiments than non-experimental studies (34 v. 4 percent)
- In the JTPA , only 16 sites agreed to participate (90 percent refused)

Equilibrium effects

- Small experiments do not reflect the general equilibrium effect of a policy change.
- At scale, there may be unintended responses to policy

Minneapolis Domestic Violence Experiment (MDVE)

MDVE evaluated the effectiveness of various police responses to domestic violence calls.

- Before MDVE, police made few arrests for domestic violence
- MDVE experimentally tested three possible responses:
  - i) send the abuser away for eight hours,
  - ii) advice and reconciliation, and
  - iii) make an arrest.

- In MDVE, arrests reduced the rate of re-offending by half.
- The Minneapolis PD changed general policy, essentially obliging officers to make an arrest.
- Number of domestic violence calls started declining, while number of suspect hospitalizations increased.

Threats to Internal Validity

- Internal validity is experiments' strong point
- Still, noncompliance poses a threat
- Two types of noncompliance:
  - Control substitution
  - Program dropout/attrition

Control substitution

In some cases, people randomized to control can obtain similar services elsewhere. Examples:

- JTPA: job training services at nearby community colleges
- Head Start experiment: private preschools
- Oregon Medicaid experiment: other ways to get insurance

When controls find similar services, control group becomes "contaminated," since some controls are self-selecting into treatment

Program dropout/attrition

In some cases, people randomized to treatment change their minds and do nothing. Examples:

- JTPA: some people found jobs before training started
- Head Start experiment: some families chose other forms of childcare
- Oregon Medicaid experiment: lots of people weren't eligible, didn't take up offer

When treatments drop out, treatment group becomes "contaminated," since some treatments are self-selecting out of treatment

How Often Does Noncompliance Happen?

Head Start

Percent of Analysis Sample that Participated in Head Start		
	Treatment group	**Control group**
3 year olds	88.2	18.5
4 year olds	83.4	16.6

JTPA classroom training

	Adult men	Adult women	Male youth	Female youth
Treatment	48.8%	56.1%	55.7%	58.6%
Controls	27.4%	33.3%	34.5%	40.1%
Difference	21.4%	22.8%	21.2%	18.5%

What can you do when you have noncompliance?

Two possibilities:

- Estimate intent-to-treat effect (ITT)
- Estimate LATE

Intent-to-treat

In presence of non-compliance, actual treatment status (D) is non-random

However, assignment to treatment (Z) remains random

The intent-to-treat effect is given by

$$E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)$$

This can always be estimated consistently, question is whether it's interesting

Intent-to-treat

In regression-speak, the ITT

$$E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0)$$

is obtained from the reduced-form regression of the outcome on the instrument:

$$Y_i = \beta + \Delta^{ITT} Z_i + u_i$$

where you could add exogenous X's if desired to improve precision

LATE

Since Z remains random, it can be used as an instrumental variable for D. We can form the Wald estimator

$$\frac{E(Y|Z=1) - E(Y|Z=0)}{P(D=1|Z=1) - P(D=1|Z=0)}$$

This requires two things:

- The share of controls receiving treatment $P(D=1|Z=0)$ is known
- The "treatment" received by the controls is similar to the true treatment

Two ways to think about LATE

First, it is the ratio of the reduced-form to the first-stage, where the first stage

$$P(D_i = 1|Z_i = 1) - P(D_i = 1|Z_i = 0)$$

is obtained from the regression of the treatment dummy on the instrument:

$$D_i = \beta + \Delta^{FS} Z_i + u_i$$

Second, and equivalently, it is the coefficient on D_i in a 2SLS/IV regression of Y_i on D_i, using Z_i as an instrument for take-up

Moral:

IV is an important program evaluation estimator, even in experimental settings