Experimentation

Department of Government
London School of Economics and Political Science

1. What is an experiment?

2. Treatment Effects

3. Statistical Inference

Experiments

Oxford English Dictionary defines "experiment" as:

1. A scientific procedure undertaken to make a discovery, test a hypothesis, or demonstrate a known fact
2. A course of action tentatively adopted without being sure of the outcome

Experiments

"Experiments" have a very long history

Major advances in design and analysis of experiments based on agricultural and later biostatistical research in the 19th century

- R.A. Fisher
- Jerzy Neyman
- Karl Pearson
- Oscar Kempthorne

In Social Sciences

"Experiments" emerged in psychology 19th century

- Not randomized – more like "What if?" studies
- Heavily laboratory-based or clinical

In Social Sciences

"Experiments" emerged in psychology 19th century

- Not randomized – more like "What if?" studies
- Heavily laboratory-based or clinical

First randomized, controlled trial (RCT) by Peirce and Jastrow in 1884

In Social Sciences

"Experiments" emerged in psychology 19th century

- Not randomized – more like "What if?" studies
- Heavily laboratory-based or clinical

First randomized, controlled trial (RCT) by Peirce and Jastrow in 1884

RCTs came later to medicine (circa 1950)

In Social Sciences

"Experiments" emerged in psychology 19th century

- Not randomized – more like "What if?" studies
- Heavily laboratory-based or clinical

First randomized, controlled trial (RCT) by Peirce and Jastrow in 1884

RCTs came later to medicine (circa 1950)

And have been a major part of the "credibility revolution" in economics

- See, especially, LaLonde (1986)

Principles of causality

1. Correlation/Relationship
2. Nonconfounding
3. Direction ("temporal precedence")
4. Mechanism
5. Appropriate level of analysis

Principles of causality

1. **Correlation/Relationship**
2. **Nonconfounding**
3. **Direction ("temporal precedence")**
4. Mechanism
5. Appropriate level of analysis

Experiments are different

Experiments are different

1. Draw causal inferences through *design*

Experiments are different

1. Draw causal inferences through *design*
2. Randomization breaks selection bias and fixes temporal precedence

Experiments are different

1. Draw causal inferences through *design*
2. Randomization breaks selection bias and fixes temporal precedence
3. We don't need to "control" for anything

Experiments are different

1. Draw causal inferences through *design*
2. Randomization breaks selection bias and fixes temporal precedence
3. We don't need to "control" for anything
4. We see "causal effects" in the comparison of experimental groups

What kinds of questions can we answer with experiments?

What kinds of questions can we answer with experiments?

Forward causal questions

- Can X cause Y?
- What effects does X have?

What kinds of questions can we answer with experiments?

Forward causal questions

- Can X cause Y?
- What effects does X have?

Backward causal questions

- What causes Y?
- How much of Y is attributable to X?

Definitions I

The observation of units after, and possibly before, a randomly assigned intervention in a controlled setting, which tests one or more precise causal expectations

Definitions I

A randomized experiment, or randomized control trial (RCT), is:

The observation of units after, and possibly before, a randomly assigned intervention in a controlled setting, which tests one or more precise causal expectations

Definitions I

A randomized experiment, or randomized control trial (RCT), is:

The observation of units after, and possibly before, a randomly assigned intervention in a controlled setting, which tests one or more precise causal expectations

If we manipulate the thing we want to know the effect of (X), and control (i.e., hold constant) everything we do not want to know the effect of (Z), the only thing that can affect the outcome (Y) is X.

Definitions II

Definitions II

Unit: A physical object at a particular point in time

Definitions II

Treatment: An intervention, whose effect(s) we wish to assess relative to some other (non-)intervention

Definitions II

Outcome: The variable we are trying to explain

Definitions II

Potential outcomes: The outcome value for each unit that we *would observe* if that unit received each treatment

Multiple potential outcomes for each unit, but we only observe one of them

Definitions II

Causal effect: The comparisons between the unit-level potential outcomes under each intervention

This is what we want to know!

Example

Example

Unit: British Citizens Residing in England

Example

Outcome: Favorability toward EU Immigration

Example

Treatment: A stimulus showing the work of Spanish nurses working in the NHS

Example

Potential outcomes:

1. Favorability without thinking about Spanish nurses
2. Favorability while thinking about Spanish nurses

Example

Causal effect: Difference in favorability between the two conditions

Units

Units can be almost anything

Common units in experimental designs:

- Individual people
- Sites (schools, classes, surgeries)
- Areas (districts, states)

Units are period-specific

- Randomization can occur over time

Outcomes

Experiments can have many outcome concepts/measures

Quite common to think about just one at a time

Outcomes can be anything that:

- Is observable/measurable
- Can be measured at the level of randomization or lower

Treatments

Synonyms: manipulation, intervention, factor, condition, cell

Treatments are operationalizations of independent variables in a causal theory

A set of treatments generates observable variation in X

Developing Treatments

From theory, we derive testable hypotheses

- Hypotheses are expectations about differences in outcomes across levels of a putatively causal variable
- In an experiment, an hypothesis must be testable by an ATE

The experimental manipulations induce variation in the causal variable that enable tests of the hypotheses

Example: Framing and Attention[1]

Theory: Presentation of information affects politicians' attention

Hypothesis:

Information framed as a conflict draws more attention from political elites than information not framed as a conflict.

Manipulation:

Control group: Presentation of headline information
Treatment group: Same information presented as conflict

Outcome:

How likely are legislators to read full article

[1] Walgrave, Sevenans, Van Camp, Loewen (2017) – "What Draws Politicians' Attention? An Experimental Study of Issue Framing and its Effect on Individual Political Elites"

Ex.: Presence/Absence

Theory: Legislators vote in line with constituents' preferences

Hypothesis: Exposure to a poll of constituent views shifts legislative votes.

Manipulation:

> Control group receives no polling information.
> Treatment group receives a letter containing polling information.

Outcome:

> How legislators vote on relevant piece of legislation

Ex.: Levels/doses

Theory: Legislators vote in line with constituents' preferences

Hypothesis: Exposure to a poll of constituent views shifts legislative votes.

Manipulation:

- Control group receives no polling information.
- Treatment group 1 receives a letter containing polling information.
- Treatment group 2 receives two letters containing polling information.
- etc.

Outcome:

- How legislators vote on relevant piece of legislation

Ex.: Qualitative variation

Theory: Legislators vote in line with constituents' preferences

Hypothesis: Exposure to a poll of constituent views shifts legislative votes.

Manipulation:

Control group receives no polling information.
Treatment group 1 receives a letter containing polling information suggesting public support.
Treatment group 2 receives a letter containing polling information suggesting public opposition.

Outcome:

How legislators vote on relevant piece of legislation

Treatments Test Hypotheses!

Treatments Test Hypotheses!

Derive experimental design from hypotheses

Treatments Test Hypotheses!

Derive experimental design from hypotheses

Experimental “factors” are expressions of hypotheses as randomized groups

Treatments Test Hypotheses!

Derive experimental design from hypotheses

Experimental "factors" are expressions of hypotheses as randomized groups

What intervention each group receives depends on hypotheses

- presence/absence
- levels/doses
- qualitative variations

Questions?

Complexities

Experiments can have additional "moving parts"

- Control groups and placebo groups
- Pre-treatment outcome measurement
- Within-subjects design features
- Repeated measures of outcomes
- Cluster randomization
- Sampling from a population
- ...

None of these are *necessary* for causal inference

The Fundamental Problem of Causal Inference!

Units have multiple potential outcomes

We can only observe one of them!

Thus we never know the individual-level causal effect of a treatment for a given unit

Two Solutions!

1. Assume units are all "homogeneous" (i.e., identical)
2. Randomly assign units to treatments and compare *average* outcomes

"The Perfect Doctor"

Unit	Y_0	Y_1
1	?	?
2	?	?
3	?	?
4	?	?
5	?	?
6	?	?
7	?	?
8	?	?
Mean	**?**	**?**

"The Perfect Doctor"

Unit	Y_0	Y_1
1	?	14
2	6	?
3	4	?
4	5	?
5	6	?
6	6	?
7	?	10
8	?	9
Mean	**5.4**	**11**

"The Perfect Doctor"

Unit	Y_0	Y_1
1	13	14
2	6	0
3	4	1
4	5	2
5	6	3
6	6	1
7	8	10
8	8	9
Mean	**7**	**5**

Experimental Inference I

We cannot see individual-level causal effects

Experimental Inference I

We cannot see individual-level causal effects

We can see *average causal effects*

Ex.: Average difference in cancer between those who do and do not smoke

Experimental Inference I

We cannot see individual-level causal effects

We can see *average causal effects*

Ex.: Average difference in cancer between those who do and do not smoke

We want to know: $TE_i = Y_{1i} - Y_{0i}$

Experimental Inference II

We want to know: $TE_i = Y_{1i} - Y_{0i}$

Experimental Inference II

We want to know: $TE_i = Y_{1i} - Y_{0i}$

We can average:
$ATE = E[Y_{1i} - Y_{0i}] = E[Y_{1i}] - E[Y_{0i}]$

Experimental Inference II

We want to know: $TE_i = Y_{1i} - Y_{0i}$

We can average:
$ATE = E[Y_{1i} - Y_{0i}] = E[Y_{1i}] - E[Y_{0i}]$

But we still only see one potential outcome for each unit:

$$ATE_{naive} = E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0]$$

Experimental Inference II

We want to know: $TE_i = Y_{1i} - Y_{0i}$

We can average:
$ATE = E[Y_{1i} - Y_{0i}] = E[Y_{1i}] - E[Y_{0i}]$

But we still only see one potential outcome for each unit:

$$ATE_{naive} = E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0]$$

Is this what we want to know?

Experimental Inference III

What we want and what we have:

$$ATE = E[Y_{1i}] - E[Y_{0i}] \tag{1}$$

$$ATE_{naive} = E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0] \tag{2}$$

Experimental Inference III

What we want and what we have:

$$ATE = E[Y_{1i}] - E[Y_{0i}] \tag{1}$$

$$ATE_{naive} = E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0] \tag{2}$$

Are the following statements true?

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

Experimental Inference III

What we want and what we have:

$$ATE = E[Y_{1i}] - E[Y_{0i}] \tag{1}$$

$$ATE_{naive} = E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0] \tag{2}$$

Are the following statements true?

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

Not in general!

Experimental Inference IV

Only true when both of the following hold:

$$E[Y_{1i}] = E[Y_{1i}|X = 1] = E[Y_{1i}|X = 0] \tag{3}$$
$$E[Y_{0i}] = E[Y_{0i}|X = 1] = E[Y_{0i}|X = 0] \tag{4}$$

In that case, potential outcomes are *independent* of treatment assignment

If true, then:

$$\begin{aligned} ATE_{naive} &= E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0] \\ &= E[Y_{1i}] - E[Y_{0i}] \\ &= ATE \end{aligned} \tag{5}$$

Experimental Inference V

This holds in experiments because of randomization, which is a special, physical process of unpredictable sorting[2]

- Units differ only in what side of coin was up
- Experiments randomly reveal potential outcomes
- Randomization balances Z *in expectation*

[2] Not "random" in the casual, everyday sense of the word

Experimental Analysis I

The statistic of interest in an experiment is the *(sample) average treatment effect* (SATE)

This boils down to being a mean-difference between two groups:

$$\widehat{SATE} = \left(\frac{1}{n_1} \sum_{i=1}^{n_1} Y_{1i} \right) - \left(\frac{1}{n_0} \sum_{i=1}^{n_0} Y_{0i} \right) \tag{5}$$

Experimental Analysis I

The statistic of interest in an experiment is the *(sample) average treatment effect* (SATE)

This boils down to being a mean-difference between two groups:

$$\widehat{SATE} = \left(\frac{1}{n_1}\sum_{i=1}^{n_1} Y_{1i}\right) - \left(\frac{1}{n_0}\sum_{i=1}^{n_0} Y_{0i}\right) \tag{5}$$

Experiments do not require “controlling for” anything, if randomization occurred successfully

Experimental Data Structures

An experimental data structure looks like:

`unit`	`treatment`	`outcome`
A	0	5
B	0	7
C	0	9
D	0	4
E	1	9
F	1	4
G	1	13
H	1	12

Questions?

Experimental Analysis I

We don't just care about the size of the SATE. We also want to measure it precisely and know whether it is significantly different from zero (i.e., different from no effect/difference)

To know that, we need to estimate the *variance* of the SATE

The variance is influenced by:

- Total sample size
- Variance of the outcome, Y
- Relative size of each treatment group
- "Advanced" design features

Experimental Analysis II

Formula for the variance of the SATE is:

$$\widehat{Var}(SATE) = \left(\frac{\widehat{Var}(Y_0)}{n_0}\right) + \left(\frac{\widehat{Var}(Y_1)}{n_1}\right)$$

$\widehat{Var}(Y_0)$ is control group variance
$\widehat{Var}(Y_1)$ is treatment group variance

We often express this as the *standard error* of the estimate:

$$\widehat{SE}_{SATE} = \sqrt{\frac{\widehat{Var}(Y_0)}{n_0} + \frac{\widehat{Var}(Y_1)}{n_1}}$$

Intuition about Variance

Bigger sample $\rightarrow$ smaller SEs

Smaller variance $\rightarrow$ smaller SEs

Efficient use of sample size:

- When treatment group variances equal, equal sample sizes are most efficient
- When variances differ, sample units are better allocated to the group with higher variance in Y

Statistical Inference

To assess whether an effect differs from zero, we need to know the sampling distribution of the ATE

Two major ways to do this:

1. Assume a parametric distribution (e.g., t-test)
2. Randomization inference

In large samples, the latter approaches the former

Parametric Analysis Stata/R

R:

```
t.test(outcome ~ treatment, data = data)
lm(outcome ~ factor(treatment), data = data)
```

Stata:

```
ttest outcome, by(treatment)
reg outcome i.treatment
```

Questions?