Business Analytics

Session 6b. Randomized Experiments

Renyu (Philip) Zhang

New York University Shanghai

Spring 2019

Why Randomized Experiments?

- Randomization solves selection bias: Assignment of treatment and control is independent of outcome.

$$\mathbb{E}[Y(1)|W=0]=\mathbb{E}[Y(1)|W=1],\ \mathbb{E}[Y(0)|W=0]=\mathbb{E}[Y(0)|W=1]$$

$$\mathbb{E}[Y(1)|W=1]-\mathbb{E}[Y(0)|W=0]=\mathbb{E}[Y(1)-Y(0)|W=1]=\mathbb{E}[Y(1)-Y(0)|W=0]$$

- Completely randomized assignment of treatment and control eliminates selection bias.
- Thanks to the Internet technology, unprecedentedly large scale randomized experiments are taking place each day.
  - Google, Facebook, Tencent, Alibaba, Kuaishou, etc.

Essential Role of Identification in Causal Inference

- Causal inference requires good identification strategy (i.e., the treatment assignment mechanism).
- Treatment is randomized by the researcher $\star\star\star$
  - Lab experiments (Mendel's experiments on inheritance).
  - Field experiments (Oregon health insurance experiment).
- Treatment is haphazard (natural experiment) $\star\star$
  - Weather, birthdays, child gender, arbitrary administrative rules, etc.
- Treatment is "as-if" random after statistical control $\star$
  - Regression, matching.
- Treatment is self-selected and no plausible control is available :(

Example: Social Pressure Experiment

Background

- Rational self-interested behavior generally fails to predict significant turnout.

- Two aspects of the utility associated with performing civic duty: Intrinsic satisfaction from behaving in accordance with a norm and extrinsic incentives to comply.

- A large scale randomized experiment by applying varying degrees of extrinsic pressure on voters using to series of mailings to 180,002 households before the August 2006 primary election in Michigan.
  - $Y_i =$ whether voted in primary
  - $W_i =$ type of mailing sent (randomized across different households)

Treatment Assignment

- Civic Duty
  - Encouraged to vote.
- Hawthorne
  - Encouraged to vote.
  - Told that researchers would be checking on whether they voted.
- Self
  - Encouraged to vote.
  - Told that researchers would be checking on whether they voted.
  - Shown whether members of their own household voted in the last two elections.
- Neighbors
  - Like Self but in addition recipients are shown whether the neighbors on the block voted in the last two elections.

“Neighbors” Letter Mailed to Each Household

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We’ve been talking about the problem for years, but it only seems to get worse. This year, we’re taking a new approach. We’re sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

MAPLE DR	Aug 04	Nov 04	Aug 06
9995 JOSEPH JAMES SMITH	Voted	Voted	______
9995 JENNIFER KAY SMITH		Voted	______
9997 RICHARD B JACKSON		Voted	______
9999 KATHY MARIE JACKSON		Voted	______
9999 BRIAN JOSEPH JACKSON		Voted	______
9991 JENNIFER KAY THOMPSON		Voted	______
9991 BOB R THOMPSON		Voted	

Social Pressure Experiment Results

Group	Number of Individuals	Voting Percentage
Control (not mailed)	191,243	29.7%
Civic Duty	38,218	31.5%
Hawthorne	38,204	32.2%
Self	38,218	34.5%
Neighbors	38,201	37.8%

- Question: Does extrinsic pressure really motivate individuals to vote?
- We address this question with the causal inference analysis of the social pressure experiment.

Causal Inference with Randomized Experiments

Identification vs. Estimation

- Essence of Causal Inference:
  - Learn about a counterfactual quantity of interest (QoI) using *finite*, *observed* data.

- Two inferential challenges:
  - Identification: If you can observe data from an entire population, can you learn about your QoI?
  - Estimation: Given your finite amount of data on a sample, how well can you learn about your QoI?

- Gold rule of causal inference: Identification precedes estimation.

Setup of Randomized Experiment

- Subjects: $i = 1, 2, , \cdots, N$ (N is the sample size)
- Treatment: $W_i \in \{0, 1\}$, randomly assigned
- Potential outcomes: $Y_i(1)$, $Y_i(0)$
- # of treated/untreated subjects: $N_1 = \sum_{i=1}^{N} W_i$ and $N_0 = N - N_1$
- Random assignment of treatment:
  - Complete randomization: Randomly select N_1 subjects into the treatment group.
  - Simple (Bernoulli) randomization: Each unit independently assigned to treatment with probability p.
  - Randomization implies that W_i is independent of $Y_i(1)$ and $Y_i(0)$.

Identifaction and Estimation of ATE

- Thanks to the randomized assignment, selection bias is eliminated.

$$\widehat{ATE} = \frac{1}{N_1}\sum_{W_i=1} Y_i(1) - \frac{1}{N_0}\sum_{W_i=0} Y_i(0) \rightarrow \mathbb{E}[Y(1)|W=1] - \mathbb{E}[Y(0)|W=0]$$
$$= \mathbb{E}[Y(1) - Y(0)] = ATE \text{ (random assignment)}$$

$$\widehat{SE} = \sqrt{\frac{\hat{\sigma}_1^2}{N_1} + \frac{\hat{\sigma}_0^2}{N_0}}$$

- If we have a large sample, $\widehat{ATE}$ is a good estimate of ATE and $\widehat{SE}$ is a good estimate of the standard deviation of ATE.
    - If the sample size is small, $\widehat{SE}$ overestimates (i.e., greater than the true value) for the standard deviation of ATE.

- Alternatively, we can directly use OLS to estimate ATE, its standard deviation, and $p-$value.

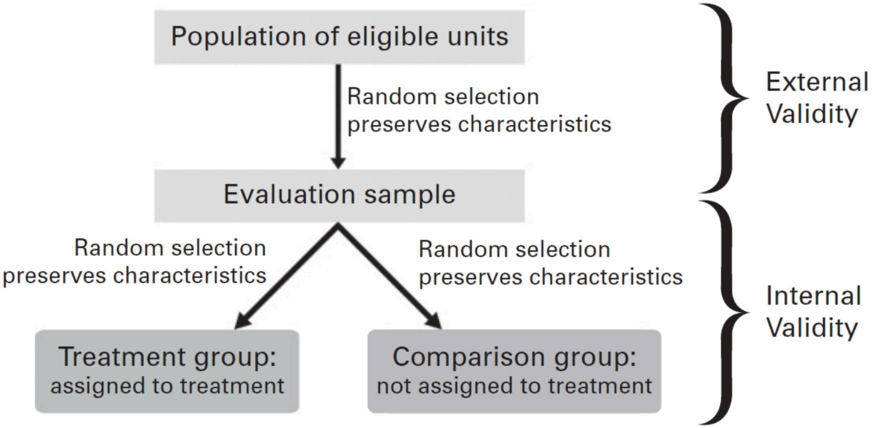
Experiment Design

External and Internal Validity

- Internal Validity: Can we estimate the treatment effect for our particular sample?
  - Valid control group used, so no confounding factors in the estimated factor.
  - Random assignment such that control group is statistically equivalent to the baseline.
  - Fails when there are differences between treated and controls (other than the treatment itself) that affect the outcome and that we cannot control for.

- External validity: Evaluation sample accurately represents population.
  - Random sampling of population.
  - Fails when outside the experimental environment the treatment has a different effect.

External and Internal Validity



Common Threats to Internal Validity

- Failure of randomization
  - Implementing partners assign their favorites to treatment group; imbalance due to small sample size, etc.
- Noncompliance with experimental protocol
  - e.g., failure to treat or "crossover": Some members of the control group receive the treatment and some members of the treatment group go untreated.
- Differential attrition
  - e.g., control group subjects are more likely to drop out of a study than treatment group subjects.

Common Threats to External Validity

- Non-representative sample
  - e.g., laboratory experiment using a convenience sample.
  - Subjects are randomly sampled, but not from the population of interest.
- Non-representative treatment
  - Treatment differs from actual implementations.
  - Actual implementations are not randomized (nor full scale).

Internal vs. External Validity

- Internal validity comes first:
  "If you do not know the effects of the treatment on the units in your study, you are not well-positioned to infer the effects on units you did not study who live in circumstances you did not study." –Rosenbaum (2010)
- Well-excuted randomization ensures internal validity.
- External validity may be partially addressed by comparing the results of several internally valid studies conducted in different circumstances and at different times.
  - Fully addressing external validity is impossible as long as your sample is the full population.

Randomized Experiments with Covariates

- Randomization balances both observed and unobserved pre-treatment covariates between the treated and untreated in large samples.
  - No need to take into account covariates under well-excuted randomization.
- In small samples, you may get unlucky and suffer from imbalance.
- Balance check with respect to observed pre-treatment covariates in different groups.
  - $t-$test and bias check (estimated mean difference/average standard deviation of treatment and control groups).
  - Regress treatment on covariates.
- Value of balance check:
  - Ensure randomization is well-executed (This is important for large scale online platforms like Google where thousands of experiments are conducted simultaneously).
  - With modest data imbalance, proper analysis can be applied to address it (regression, matching, weighting, etc.).
  - Help strengthen external validity when sample and population are not identical.

Stable Unit Treatment Value Assumption

- Potential outcomes for one individual is independent with the assignment of treatment/control for another individual.
  - This is called the Stable Unit Treatment Value Assumption (SUTVA).

- An example:
  - In AirBnb, we provide a new feature that dramatically streamlines the booking process to a randomized group of customers.
  - We find that customers given the new feature book much more frequently than customers in the control group.
  - This is an over-estimate of ATE, why?

- Two aspects of SUTVA
  1. No interference.
  2. One version of treatment level exists.

Homework

- Finish Homework 6 (NO need to submit it).

- Think about your final project.
  - A list of project topis will be distributed later this week.
  - Approval is needed if you choose a topic outside of the list.
  - Choice of final project due: Sunday, April 14, 10:00PM