Class 12 A/B Testing

Dr. Wei Miao

UCL School of Management

November 6, 2024

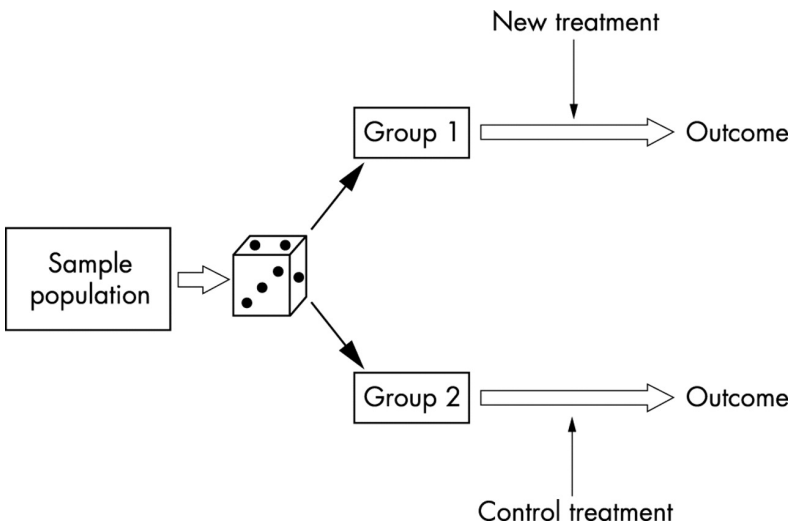
Section 1

Randomized Controlled Trials

Randomized Controlled Trials

Randomized Controlled Trials

A **randomized controlled trial** (RCT) is an experimental form of impact evaluation in which the population receiving the intervention is chosen *at random* from the eligible population, and a control group is also chosen *at random* from the same eligible population.



Types of RCTs: Based on Location

	Lab Experiment	Field Experiment
Location	In a controlled, laboratory environment	In the field
Internal validity	High	Low
External validity	Low (Hawthorne effect)	High

- Internal validity refers to the extent to which the experiment is free from selection bias.
- External validity refers to the extent to which the results can be generalized to the real world or other contexts.

Types of RCTs: Based on Treatment Design

- A/B testing (treatment group + control group)
  1. Loyalty program
  2. No loyalty program
- A/B/N testing (multiple treatment groups + control group)
  1. Point-based loyalty program; points can be redeemed for price vouchers
  2. Point-based loyalty program; points can be redeemed for gifts
  3. Point-based loyalty program; points can be redeemed for free top ups
  4. No loyalty program
- Factorial design
  - more than 2 dimensions of treatments, used if we care about the interaction effects

Section 2

Procedures of A/B Testings

Motivating Example of Tom's Loyalty Program

- Should we introduce a loyalty program for our customers?
  - Cons: increased costs due to free drinks
  - Pros: it may increase spending and retention rate, and hence future CLV
- How to estimate the causal effect of introducing a loyalty program?

Step 1: Decide on the Unit of Randomization

We decide **the granularity of randomization unit** based on the research question at hand.

- **individual**
- household
- store
- other levels more granular (e.g., device level) or even less granular (e.g., city level)

Step 1: Decide on the Unit of Randomization

Proposal 1: Randomize the treatment based on West London and East London.

- Do you expect the "randomize" to be true randomization?[1]

Proposal 2: Randomize the treatment among individual customers.

- Is this true randomization?
- What problems can we still have?

[1] All answers to questions in the slides are on the webpage version of lecture notes.

Step 1: Pros and Cons of Granularity

Disadvantages of granularity:

- Costs and logistics
- Spillovers and crossovers

Advantages of granularity:

- Increase the chance of successful randomization, thereby mitigating any systematic unbalance of covariates before the experiment.

Exercise:

- If we would like to randomize prices, how can we randomize individualized price discounts to customers?

Step 2: Mitigate Spillover and Crossover Effects

- **Crossover Effects**: A crossover occurs when an individual who was supposed to be assigned to one treatment is accidentally exposed to another or more treatments.
  - e.g., For online A/B testing, a notorious crossover effect is that when browsers reset the cookies, the same individual customer may be treated as a different new customer.
- **Spillover effects**: The behavior of the treatment group can affect control group as well
  - e.g., customers may share the promotions with family members and friends.

Question: How should we mitigate spillover and crossover effects?

Step 3: Decide on Randomization Allocation Scheme

- Individuals (or the relevant unit of randomization) are allocated at random into a treatment condition based on some decision rules.
- Due to the high costs and potential risks of A/B testing, we often select a small percentage of customers into the treatment condition, while the remaining customer should do "business-as-usual".

Step 4: Collect Data

- Any field experiment should be aware of the need for a sufficiently large sample size, or sufficient statistical power.
  - The larger sample size, the higher statistical power for the experiment; meanwhile, larger sample size brings higher costs and risks.
  - Run a power calculation in R
- Collect both data on the outcome variables of interest and consumer characteristics data

Proposal: We need to collect customers' spending and retention data and link the data with their treatment assignment.

Step 5: Interpreting Results from a Field Experiment

Step 5.1: Randomization check

- We need to check if the treatment group and control group are well-balanced in terms of their **pre-treatment** characteristics, especially the outcome variables.

Step 5.2: Analyze the data and estimate the ATE

- **t-test** to examine the difference in the average outcome between the treatment group and control group. In R, we can use `t.test()`
- **Regression analysis** when analyzing A/B/N testing or multivariate experiments.

After-Class Readings

- (optional) Varian, Hal R. 'Causal Inference in Economics and Marketing'. Proceedings of the National Academy of Sciences 113, no. 27 (5 July 2016): 7310–15.