Randomized Controlled Trial

Prof. Tzu-Ting Yang
楊子霆

Institute of Economics, Academia Sinica
中央研究院經濟研究所

September 10, 2024

Randomized Controlled Trial

隨機試驗

- The most credible identification strategy is **randomized controlled trial/experiment**
- Randomized controlled trial (RCT):
    - Each observation (e.g. individual, household, school, state, or country) is **randomly assigned** to treatment and control group

Randomized Controlled Trial

- There are two types of RCT

  1. Lab experiment

     - An experiment conducted under highly controlled conditions (laboratory)
     - Usually let individuals play some games in the computers as a proxy for the real-world treatment

  2. Field experiment

     - A experiment used to test a hypothesis in a natural setting, such as a school or a workplace
     - This type of experiment allows researchers to observe the effects of treatment in a real-world setting

Randomized Controlled Trial

- RCT has two features that can help us hold "other things equal" and then eliminates selection bias
    1. Randomly assign treatment
    2. Sufficiently "large" sample size

Randomly Assign Treatment

- Randomly assign treatment (such as a coin flip) ensures that:
  - The probability of receiving treatment is unrelated to any other confounding factors
  - Every observation has the same probability of being assigned to the treatment group
- **Example:**
  - Randomly assign master degree (D)
  - An individual's probability of receiving master degree (D) is unrelated to his/her family wealth or ability

Randomly Assign Treatment

- **Randomly assign treatment implies:**
  - The values of potential outcomes are **independent** of treatment assigned
  - $(\mathrm{Y}_i^{\mathbf{1}}, \mathrm{Y}_i^{\mathbf{0}})$ are independent of D_i

$$(\mathrm{Y}_i^{\mathbf{1}}, \mathrm{Y}_i^{\mathbf{0}}) \perp\!\!\!\perp D_i$$

  - Two variables are said to be **independent** means:
    - The occurrence of one variable (D) does not affect the probability of occurrence of the other $(\mathrm{Y}_i^{\mathbf{1}}, \mathrm{Y}_i^{\mathbf{0}})$
- **Intuition:** treatment assignment D is random so that it is not based on an individual's value of potential outcome $(\mathrm{Y}_i^{\mathbf{1}}, \mathrm{Y}_i^{\mathbf{0}})$

Randomly Assign Treatment

- Randomly assign treatment makes treatment and control group to be similar along all characteristics, including $\mathrm{Y}_i^0, \mathrm{Y}_i^1$
  - $\mathrm{E}[\mathrm{Y}_i^{\mathbf{0}}|D_i = 1] = \mathrm{E}[\mathrm{Y}_i^{\mathbf{0}}|D_i = 0]$
  - $\mathrm{E}[\mathrm{Y}_i^{\mathbf{1}}|D_i = 1] = \mathrm{E}[\mathrm{Y}_i^{\mathbf{1}}|D_i = 0]$

(隨機分配)

- So that we can eliminate selection bias
  - $\underbrace{\mathrm{E}[\mathrm{Y}_i^0|D_i = 1] - \mathrm{E}[\mathrm{Y}_i^0|D_i = 0]}_{\text{Selection Bias}} = 0$

Sufficiently Large Sample Size

- Randomly assign treatment can ensure the **average** characteristics of two groups are similar
  - How about each group only has one individual?
- We also need **large sample size** to ensure that the group differences in individual characteristics wash out

Randomized Controlled Trial: Potential Outcome Framework

Randomized Controlled Trial

Identify ATT and ATE

- RCT can identify ATT, ATU and ATE

研究者隨機給定

$$\underbrace{\mathrm{E}[\mathrm{Y}_i|D_i = 1] - \mathrm{E}[\mathrm{Y}_i|D_i = 0]}_{\text{ODO}}$$

$$= \underbrace{\mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0|D_i = 1]}_{\text{ATT}} + \underbrace{\mathrm{E}[\mathrm{Y}_i^0|D_i = 1] - \mathrm{E}[\mathrm{Y}_i^0|D_i = 0]}_{\text{Selection Bias}}$$

$$= \underbrace{\mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0|D_i = 1]}_{\text{ATT}} + \underbrace{0}_{\text{Selection Bias}}$$

$$= \underbrace{\mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0|D_i = 0]}_{\text{ATU}}$$

$$= \underbrace{\mathrm{E}[\mathrm{Y}_i^1 - \mathrm{Y}_i^0]}_{\text{ATE}}$$

Randomized Controlled Trial: Estimation

Estimation and Inference in RCT

- Up until now, we've talked about identification.
- Now that we know that the ATE and ATT are identified, how will we estimate them?
- Remember: identification first, then estimation.

- We want to know whether getting master degree can increase monthly salary for Taiwanese people
- Suppose we can implement a RCT for whole population in Taiwan
  - Randomly assign master degree to every Taiwanese
- We can obtain the ATE of master degree on earning:
- But we do NOT have population data
  - Use **sample estimates** to guess population value

- Suppose we get a nationally representative sample: N individuals
- Randomly assign treatment (master degree)
  - N_1 individuals obtain master degree: treatment group
  - N_0 individuals do not have it ($N_0 = N - N_1$): control group
- Compare difference in monthly salary between treatment group and control group

- Using the analogy principle, we construct the following sample estimator:

$$\hat{\alpha}_{\mathsf{ATE}} = \bar{\mathrm{Y}}_1 - \bar{\mathrm{Y}}_0$$

- where $N_1 = \sum_i D_i$ and $N_0 = N - N_1$

$$\bar{\mathrm{Y}}_1 = \frac{1}{N_1} \sum_{D_i=1} \mathrm{Y}_i$$

$$\bar{\mathrm{Y}}_0 = \frac{1}{N_0} \sum_{D_i=0} \mathrm{Y}_i$$

- Now we want to use **sample estimator** $\hat{\alpha}_{\text{ATE}}$ to infer whether outcomes (e.g. monthly salary) are different in treatment and control group at **population level** α_{ATE}
- Statistical inference (Hypothesis Testing) helps us answer this question

Review: Hypothesis Testing

Sample Estimator

- Suppose that the sample estimator $\hat{\alpha}_{\text{ATE}} = \bar{\mathrm{Y}}_1 - \bar{\mathrm{Y}}_0$ is $NT\$5000$
- Does it mean those who get master degree have more monthly salary than those do not hold master degree ?
  - Note that sample estimator can be somewhat different when drawing another sample from the same population
- It's possible that this is a **chance finding**
  - Need to show this is not just a chance finding
  - How?
  - Show that the probability of obtaining this estimate is small

Summary of Hypothesis Testing

1 Choose a null hypothesis:

- Since we do NOT know exact population value of α_{ATE}
- What we can do is assume the true α_{ATE} is exactly μ
- This is what we call the null hypothesis (H_0)
- Suppose we want to test there is **no average treatment effect** of master degree:
    - $H_0 : \alpha_{\text{ATE}} = 0$

Summary of Hypothesis Testing

2. Choose a test statistic to examine a null hypothesis:
   - We usually use t-statistics to measure the difference between sample estimator and null hypothesis
   - Need to consider variation of sample estimator $\hat{\mathrm{SE}}(\hat{\alpha}_{ATE})$
   - $t = \dfrac{(\hat{\alpha}_{ATE} - \alpha_{ATE})}{\hat{\mathrm{SE}}(\alpha_{\hat{ATE}})}$
   - Estimate standard error of the sample estimator
     - $\hat{\mathrm{SE}}(\hat{\alpha}_{ATE}) = \hat{\sigma_Y}\sqrt{[\frac{1}{N_1} + \frac{1}{N_0}]}$
     - $\hat{\sigma_Y}$ is sample variance of Y

Summary of Hypothesis Testing

3. Evaluate whether the sample estimator is against null hypothesis or not
   - **Goal:** Calculate **p-value**
     - **p-value:** Given null hypothesis is true, the probability of obtaining the sample estimates or more extreme ones
     - If this probability is high, it means the sample estimator might support for null hypothesis
     - If this probability is low, it means the sample estimator might be against null hypothesis

Summary of Hypothesis Testing

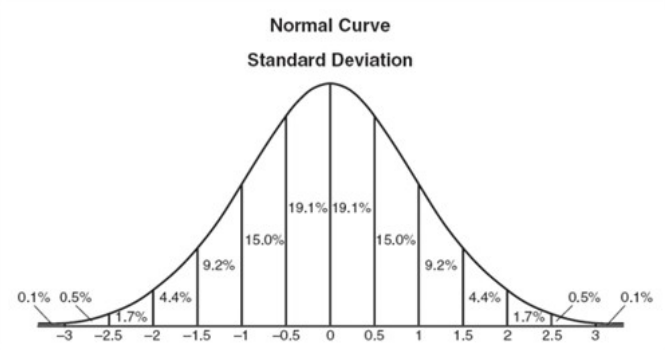
3. Evaluate whether the sample estimator is against null hypothesis or not

   - In order to calculate this probability (p-value), we need to know the distribution of the t-statistic under the null hypothesis

     - If sample size is sufficient large, using **Central Limit Theorem (CLT)**, t-statistic will have standard normal distribution
     - **Standard normal distribution** has a mean of 0 and standard deviation of 1 : $N(0,1)$

$$\boxed{t = \frac{\hat{\alpha}_{\text{ATE}} - 0}{\hat{\text{SE}}(\hat{\alpha}_{ATE})} \xrightarrow{d} N(0,1)}$$

Summary of Hypothesis Testing



Summary of Hypothesis Testing

3. Evaluate whether the sample estimator is against null hypothesis or not

   - Based on standard normal distribution and sample estimator, we can get p-value
   - We usually select an arbitrarily pre-defined threshold value α, which is referred to as the **level of significance**
     - The value of α is instead set by the researcher before examining the data
     - By convention, α is commonly set to 0.05 or 0.01
   - If p-value is smaller than α, we would say the sample estimate is **significantly different from the null hypothesis**

Empirical Example: Leonardo Bursztyn, Thomas Fujiwara, and Amanda Pallais (2017)

Empirical Example: Leonardo Bursztyn, Thomas Fujiwara, and Amanda Pallais (2017)

Leonardo Bursztyn, Thomas Fujiwara, and Amanda Pallais (2017) “**'Acting Wife': Marriage Market Incentives and Labor Market Investments**”, AER

- General question: Why is there a gender gap in labor market outcomes?
- Specifically, they want to examine whether marriage market incentives lead to gender differences in labor market investments.

- Do single women **avoid career-enhancing actions** because these actions could signal personality traits, like ambition?
- Since these features are undesirable in the marriage market
  - Bertrand et al. (2015): It is relatively unlikely that a woman will earn more than her husband, and when she does, marital satisfaction is lower and divorce is more likely
  - Folke and Rickne (2016): Promotions increase the chance of divorce for women, but not for men.

- Single women may thus face a trade-off:
    - Actions that lead to **professional success** might be sanctioned in the marriage market because they **signal ambition** and assertiveness.
- Single women might try to improve their marriage options by "acting wife."

Empirical Example: Leonardo Bursztyn, Thomas Fujiwara, and Amanda Pallais (2017)

- The authors test for the existence and the implications of this trade-off by studying students in an elite U.S. MBA program.
  - Graduate school is a natural place to study this trade-off.
  - Many students are both investing in their professional career and looking for a long-term partner.
- They conducted a field experiment to test this trade-off

- On the first day of the MBA program, a career counselor asked students to complete a questionnaire about their job preferences for their summer internships.
    - This questionnaire asked about students desired compensation, hours of work, and days per month of travel.
    - It also asked students to rate their leadership abilities and professional ambition.

- The answer to this questionnaire can impact their placement of summer internships
- Summer internships would affect their post-graduation job offer
    - They will answer these questions seriously

- They created two versions of the instructions of questionnaire
- Which version a student received was **randomized**

隨機分配

  - 公開 **Public version**: students were also told that “**your**” answers will be discussed in the career class
  - 私人 **Private version**: students were told that “**anonymized**” answers will be discussed in the career class

- Two versions only differed by one word (“your” v.s. “anonymized”)
- Such type of experiment is called **Survey Experiment**

Empirical Example: Leonardo Bursztyn, Thomas Fujiwara, and Amanda Pallais (2017)

Experimental Design

Appendix Figure 3. Primary Experiment Questionnaire

The information on this survey will help the career office get to know you and help it find the right fit for your first-year internship. This information will not be shared with employers, so please express your true preferences, not just what you think employers want to hear. This information will be shared with your career advisor and [your/anonymized] answers will be discussed during the [name of career class].

UID Number: ______________________ **Name:** ____

Gender Identity (Optional): Male Female Other ________________________ **Age:** ___________

Marital Status: Single In a serious relationship Cohabiting Engaged Married

Do you have children, either biological or adopted? Yes No

What industries are you interested in working in? List these below.

Tell us about any geographic preferences.

For the questions below, please circle only one answer.

What is your desired compensation level in your first year after graduation? Include base pay, performance pay, and equity, but not the signing bonus.

Under $75,000 $75,000-$100,000 $100,000-$125,000 $125,000-$150,000 $150,000-$175,000

$175,000-$200,000 $200,000-$225,000 $225,000-$250,000 Above $250,000

How often are you willing to travel for work?

Rather not travel A few days a month 1-2 days a week

4-5 days a week As much as necessary

Empirical Example: Leonardo Bursztyn, Thomas Fujiwara, and Amanda Pallais (2017)

Results

Table 3. Randomization Assessment by Subgroup

Primary Experiment

	Private Treatment	Public Treatment	p-Value of Difference	Private Treatment	Public Treatment	p-Value of Difference
		A. Single Women			B. Non-Single Women	
Age	27.4	27.2	0.715	27.7	27.3	0.483
Has Children	3.3%	0.0%	0.338	0.0%	0.0%	-
GMAT Score	703	712	0.205	701	700	0.974
Years of Work Experience	5.0	4.8	0.644	5.0	4.9	0.743
U.S. Citizen	61.3%	55.2%	0.638	77.8%	64.0%	0.282
Observations	31	29	60	27	25	52
		C. Single Men			D. Non-Single Men	
Age	27.5	27.7	0.471	28.4	28.9	0.350
Has Children	0.0%	0.0%	-	12.1%	17.2%	0.418
GMAT Score	719	719	0.924	720	720	0.929
Years of Work Experience	5.3	5.2	0.876	5.4	5.5	0.824
U.S. Citizen	58.3%	71.4%	0.165	65.7%	51.6%	0.103
Observations	48	56	104	67	64	131

Notes: The first and second columns of each panel contain the means of each demographic variable for the sample indicated by the panel among those in the private and public treatments, respectively. The third column shows the p-value of the difference in the means from a two-tailed t-test. *Non-Single* students are those who are in serious relationships, cohabiting, engaged, or married.

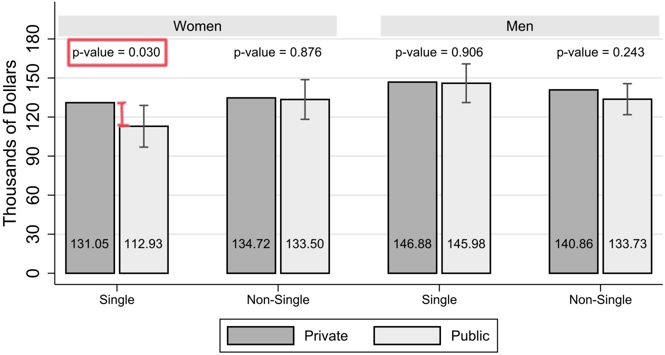
Empirical Example: Leonardo Bursztyn, Thomas Fujiwara, and Amanda Pallais (2017)

Results

- Q: their desired compensation in their first year after graduation

Figure 4. Desired Compensation

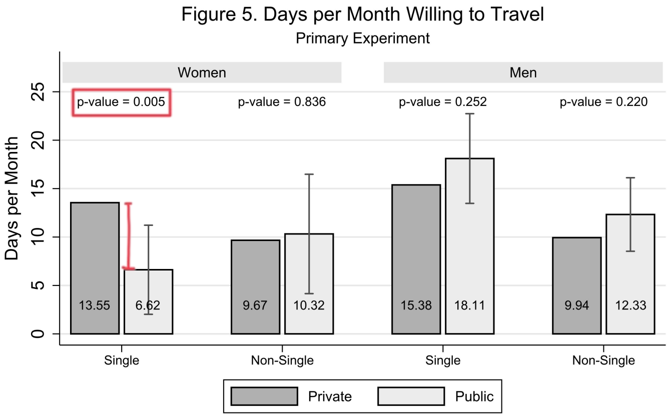
Primary Experiment



Empirical Example: Leonardo Bursztyn, Thomas Fujiwara, and Amanda Pallais (2017)

Results

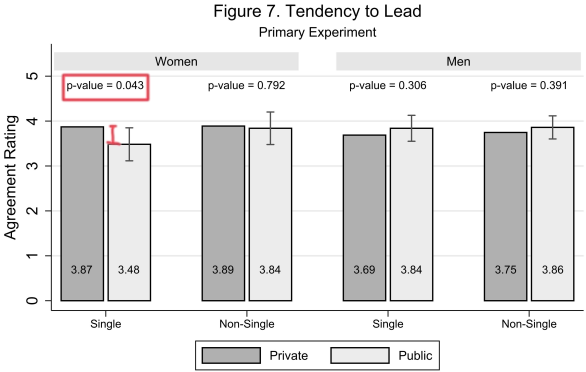
- Q: how often they are willing to travel for work.



Empirical Example: Leonardo Bursztyn, Thomas Fujiwara, and Amanda Pallais (2017)

Results

- Q: Students rated agreement with the statement "You tend to lead in your day-to-day interactions" on a 1-5 scale, where 1 is Strongly Disagree and 5 is Strongly Agree.

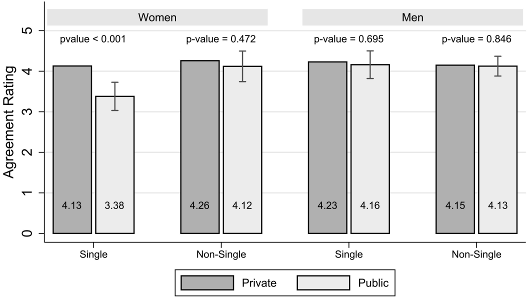


Empirical Example: Leonardo Bursztyn, Thomas Fujiwara, and Amanda Pallais (2017)

- Q: Students rated agreement with the statement "You are more professionally ambitious than your most recent work colleagues" on a 1-to-5 scale, where 1 is Strongly Disagree and 5 is Strongly Agree.

Figure 8. Professional Ambition

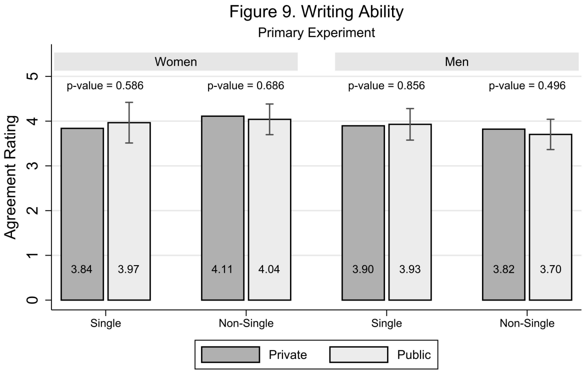
Primary Experiment



Empirical Example: Leonardo Bursztyn, Thomas Fujiwara, and Amanda Pallais (2017)

Placebo Test

- Q: Students rated agreement with the statement "You have above-average writing skills" on a 1-to-5 scale, where 1 is Strongly Disagree and 5 is Strongly Agree.



Empirical Example: Leonardo Bursztyn, Thomas Fujiwara, and Amanda Pallais (2017)

Experimental Design

- Their results have implications for understanding gender gaps in labor market outcomes
- Their results highlight the importance of social norms - particularly **what is differentially expected from a husband and a wife** - in explaining gender gaps

Threats to Validity of Randomized Experiments

Threats to Validity of Randomized Experiments

- **Internal validity:**
  - Can we estimate treatment effect for this particular sample?
  - We fail to do so when there are differences between treated and untreated sample

Most Common Threats to Internal Validity

- Failure of randomization
- Non-compliance with experimental protocol
- Attrition

Threats to Validity of Randomized Experiments

- **External validity:**
  - Can we extrapolate our estimates to other populations?
  - We fail to do so when the treatment effect is different outside the evaluation environment

Most Common Threats to External Validity

- Non-representative sample
- Non-representative program
  - The treatment differs in actual implementations
  - Actual implementations are not randomized (nor full scale)

Suggested Readings

- Chapter 1, Mastering Metrics: The Path from Cause to Effect
- Chapter 2, Mostly Harmless Econometrics