Randomized Control Trial 1: Framework Trial

Instructor: Yuta TOYAMA (遠山 祐太)

Last updated: 2021-05-20

Introduction

Introduction

- Program Evaluation, or Causal Inference
  - Estimation of "treatment effect" of some intervention (typically binary)
  - Hereafter, I use "treatment effect" and "causal effect" interchangeably (acknowledging abuse of language).

- Example:
  - effects of job training on wage
  - effects of advertisement on purchase behavior
  - effects of distributing mosquito net on children's school attendance

- Difficulty: treatment is **endogenous decision**
  - selection bias, omitted variable bias.
  - especially in **observational data** (in comparison with experimental data)

Overview

- Introduce **Rubin's causal model**
  - also known as **potential outcome framework) (潜在アウトカムモデル)**

- Introduce **randomized controlled trial (ランダム化比較試験)**
  - Framework
  - Inference: Estimation and hypothesis testing
  - (next week) Application: Field Experiment on Energy Demand in Japan (Ito et al 2018)

Reference

- Angrist and Pischke "Mostly Harmless Econometrics"
- Cunningham

Rubin's Potential Outcome Framework

Framework

- Y_i: **observed outcome** for person i
- D_i: binary **treatment (処置)** status

$$D_i = \begin{cases} 1 & treated\ (treatment\ group) \\ 0 & not\ treated\ (control\ group) \end{cases}$$

- Define **potential outcomes**
  - Y_{1i}: outcome for i when she is treated
  - Y_{0i}: outcome for i when she is not treated

- With this, we can write

$$\begin{aligned} Y_i &= D_i Y_{1i} + (1 - D_i) Y_{0i} \\ &= \begin{cases} Y_{1i} & if\ D_i = 1 \\ Y_{0i} & if\ D_i = 0 \end{cases} \end{aligned}$$

Example: College Choice

- Let D_i be whether go to a college.
- Y_{1i}: potential income if i goes to college, Y_{0i} potential income if not
- Y_i: actual observed outcome

	Y_{1i}	Y_{0i}	D_i	Y_i
Adam	80000 USD	70000 USD	1	80000 USD
Bob	70000 USD	60000 USD	0	60000 USD
Cindy	90000 USD	70000 USD	1	90000 USD
Debora	80000 USD	60000 USD	0	70000 USD

Parameters of Interest

- **Individual treatment effect** $Y_{1i} - Y_{0i}$
  - Key: allowing for heterogenous effects across people
  - Individual treatment effect cannot be obtained due to the fundamental problem (see next).
- Instead, we focus on the average effects
- **Average treatment effect (平均処置効果)**: $ATE = E[Y_{1i} - Y_{0i}]$
- Average treatment effect on treated: $ATT = E[Y_{1i} - Y_{0i}|D_i = 1]$
- Average treatment effect on untreated: $ATT = E[Y_{1i} - Y_{0i}|D_i = 0]$
- Average treatment effect conditional on covariate (共変量): $ATE(x) = E[Y_{1i} - Y_{0i}|D_i = 1, X_i = x]$

Relation to Regression Analysis

- Assume that
  1. linear (parametric) structure in Y_{0i}, and
  2. constant (homogeneous) treatment effect,

$$\begin{aligned} Y_{0i} &= \beta_0 + \epsilon_i \\ Y_{1i} - Y_{0i} &= \beta_1 \end{aligned}$$

- You will have

$$Y_i = \beta_0 + \beta_1 D_i + \epsilon_i$$

- Program evaluation framework is nonparametric in nature.
  - Though, in practice, estimation of treatment effect relies on a parametric specification.

Key Point 1 (/2) Counterfactual outcome is never observed.

- We can observe (Y_i, D_i) for each person i
- However, can never observe Y_{0i} and Y_{1i} **simultaneously**.
  - Ex: Cannot know Adam's income if he had not attended a college.
- Once person i took a particular treatment, observed outcome is potential outcome for that treatment.
- Known as **fundamental problem of program evaluation**

Key Point 2 (/2): No spillover of treatment effect

- **Stable Unit Treatment Value Assumption (SUTVA)**: Treatment effect for a person does **not depend on the treatment status of other people.**
- It rules out **externality (外部性)** and **general equilibrium effects (一般均衡効果)**.
- Ex: If **everyone** takes a job training, equilibrium wage would change, which affects the individual outcome.
- Question: Any example of treatment effect that violates the SUTVA?

Selection Bias (セレクションバイアス)

- Goal: Estimate treatment effect parameters (AET, ATT, conditional ATE, etc)
- The simplest way is to compare average outcomes between treatment and control group
- Does this tell you average treatment effect? No in general!
- To see this, first, for $d = \{0, 1\}$,

$$E[Y_i|D_i = d] = E[Y_{di}|D_i = d]$$

  - LHS: Average of observed outcome for group d
  - RHS: Average of **potential outcome** for group d

- Then,

$$\underbrace{E[Y_i|D_i=1] - E[Y_i|D_i=0]}_{simple\ comparison} = E[Y_{1i}|D_i=1] - E[Y_{0i}|D_i=0]$$

$$= \underbrace{E[Y_{1i} - Y_{0i}|D_i=1]}_{ATT}$$

$$+ \underbrace{E[Y_{0i}|D_i=1] - E[Y_{0i}|D_i=0]}_{selection\ bias}$$

Example: College Choice

	Y_{1i}	Y_{0i}	D_i	Y_i
Adam	80000 USD	70000 USD	1	80000 USD
Bob	70000 USD	60000 USD	0	60000 USD
Cindy	90000 USD	70000 USD	1	90000 USD
Debora	80000 USD	60000 USD	0	70000 USD

- Simple Difference: $E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = 85000 - 60000 = 25000$
- ATT: $E[Y_{1i} - Y_{0i}|D_i = 1] = ((80000 - 70000) + (90000 - 70000))/2 = 15000$
- Bias: $E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] = 70000 - 60000 = 10000$

Simple difference = ATT + Bias

$$\underbrace{E[Y_i|D_i=1] - E[Y_i|D_i=0]}_{simple\ comparison} = \underbrace{E[Y_{1i} - Y_{0i}|D_i=1]}_{ATT} + \underbrace{E[Y_{0i}|D_i=1] - E[Y_{0i}|D_i=0]}_{selection\ bias}$$

- The bias is not zero in general:
  - Those who go to a college **would earn a lot even without a college degree**
- We cannot observe $E[Y_{0i}|D_i=1]$:
  - the outcome of people in treatment group if they **WERE NOT treated (counterfactual)**.

Solutions

- Randomized Control Trial
  - Assign treatment D_i randomly
- Matching (regression):
  - Using observed characteristics of individuals to control for selection bias
- Instrumental variable
  - Use the variable that affects treatment status but is not correlated to the outcome
- Panel data (difference-in-differences)
- Regression discontinuity

Randomized Control Trial: Overview

What is Ranomized controlled trial (RCT, ランダム化比較試験) ?

- Measure treatment effect by
  1. randomly assigning treatment to subjects (people)
  2. measure outcomes of subjects in both treatment and control group.
  3. the difference of outcomes between these two groups is treatment effect.
- Since treatment is randomly assigined, no worry for selection bias (see later).
- It began in a clinical trial (治験), but now is widely used in social science.

RCTs in Social Science and Business

- Development economics: Esther Duflo "Social experiments to fight poverty"
- Health economics: Amy Finkelstein "Randomized evaluations & the power of evidence | Amy Finkelstein"
- Buisiness: Ron Kohavi et al "Trustworthy Online Controlled Experiments" (和訳「A/Bテスト実践ガイド」
- Andrew Lee "Randomistas" (和訳：「RCT大全」)

Example: A/B Test in Fund Raising by President Obama

- How Obama Raised $60 Million by Running a Simple Experiment
- あの大統領も140%の成果改善。アメリカ大統領とA/Bテストの意外な関係

Framework

- Key assumption: Treatment D_i is independent with potential outcomes (Y_{0i}, Y_{1i})

$$D_i \perp (Y_{0i}, Y_{1i})$$

- Under this assumption,

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

- The sample selection does not exist! Thus,

$$\underbrace{E[Y_i|D_i=1]-E[Y_i|D_i=0]}_{simple\ comparison} = \underbrace{E[Y_{1i}-Y_{0i}|D_i=1]}_{ATT}$$

- ATT can be estimated (identified) by a simple comparison of outcomes between treatment and control groups.

(A bit technical) What is identification (識別)?

- Roughly speaking, a parameter of the model is **identified** if that parameter can be written by **observable objects**.
- In the previous slide, the parameter of interest is ATT $E[Y_{1i} - Y_{0i}|D_i = 1]$.
- This is written as $E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$, the difference of the conditional expectations of observed outcome Y_i for each group.
- Conditional expectation $E[Y_i|D_i = d]$ is an observable object if you have the knowledge on the joint distribution of (Y_i, D_i).

Limitations of RCTs

- Some people say "RCT is a gold standard for causal inference".
- There are limitations that we should acknowledge.

1. SUTVA assumption
    - not specific to RCT though).
2. Ethical criticism
    - Is this fair for everyone?
3. Cannot do RCTs in many settings.
    - Topics that are not suitable to randomized experiment.
    - It requires a lot of money and effort.
4. External Validity (外的妥当性)

Internal and External Validity

- **Internal validity (内的妥当性)**
  - Can the analysis establish a credible result about causal effect?
  - RCT is strong in this aspect.
- **External validity (外的妥当性)**:
  - Can you extrapolate your results from an experiment to a general population?
  - A population in an experiment may differ from the population of interest.

Inference 1: Estimation

Overview of Inference

- So far, we show identification of treatment effect parameter.
- In practice, we have a sample of people (data) and use it to infer the unknown parameter.
- I explain statistical inference in the context of RCT.
  - (Point) Estimation (点推定)
  - Hypothesis testing (仮説検定)

Estimation of ATT parameter

- Remember that ATT is written as

$$E[Y_{1i} - Y_{0i}|D_i = 1] = E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$$

- Estimate the conditional expectation by the **conditional sample mean**

$$\hat{E}[Y_i|D_i = 1] = \frac{1}{N_1}\sum_{i=1}^{N} Y_i \cdot \mathbf{1}\{D_i = 1\} = \frac{\frac{1}{N}\sum_{i=1}^{N} Y_i \cdot \mathbf{1}\{D_i = 1\}}{\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{D_i = 1\}}$$

- Difference of the sample average is an estimator for the ATT

$$\hat{ATT} = \frac{\frac{1}{N}\sum_{i=1}^{N} Y_i \cdot \mathbf{1}\{D_i = 1\}}{\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{D_i = 1\}} - \frac{\frac{1}{N}\sum_{i=1}^{N} Y_i \cdot \mathbf{1}\{D_i = 0\}}{\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{D_i = 0\}}$$

- Question: Is this a good way to estimate ATT good? See this next.

Alternative: Linear Regression

- You can run a linear regression of Y on D along with other covariates X_i

$$Y_i = \beta_0 + \beta_1 D_i + \beta' X_i + \epsilon_i$$

Properties of Estimators

Consider the estimator $\hat{\mu}_N$ for the unknown parameter μ.

1. **Unbiasdeness (不偏性)**: The expectation of the estimator is the same as the true parameter in the population.

$$E[\hat{\mu}_N] = \mu$$

2. **Consistency (一致性)**: The estimator converges to the true parameter in probability.

$$\forall \epsilon > 0, \lim_{N \to \infty} \ Prob(|\hat{\mu}_N - \mu| < \epsilon) = 1$$

   - Intuition: As the sample size gets larger, the estimator and the true parameter is close with probability one.
   - Note: a bit different from the usual convergence of the sequence.

The estimator above is consistent

- **Law of large numbers (大数の法則)** Sample mean converges to population mean in probability.

$$\frac{1}{N}\sum_{i=1}^{N} X_i \xrightarrow{p} E[X]$$

- Can be applied to the above (using continuous mapping theorem)

$$\frac{\frac{1}{N}\sum_{i=1}^{N} Y_i \cdot \mathbf{1}\{D_i = 1\}}{\frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{D_i = 1\}} \xrightarrow{p} \frac{E[Y_i D_i]}{E[D_i]} = E[Y_i | D_i = 1]$$

- Exercise: Show the last equality (Hint: law of iterated expectation).

Inference 2: Hypothesis Testing

Hypothesis Testing

- Testing (検定): use the sample to decide whether the hypothesis (仮説) about the population parameter is true
- Example 1: Is the average age 45 in population?
- Example 2: Are test scores of male and female students are different in population?
- Issue: Sample statistic is random! How to distinguish between
  - just random phenomenon, or
  - true effects (difference) in the population

Example in Population Mean

1. Calculate sample mean $\bar{Y}$
2. Define **null hypothesis (帰無仮説)** and **alternative hypothesis (対立仮説)**: For a chosen value of μ.
   - Null: $H_0 : E[Y] = \mu$
   - Alternative: $H_1 : E[Y] \neq \mu$
3. If the null hypothesis H_0 is true, then $\bar{Y}$ should be close to μ
4. If $\bar{Y}$ is “vary far” from μ, then we should **reject (棄却)** H_0.

- Question: How to determine whether it is "very far"?

Preliminary: Standard Errors

- Let $V(\bar{Y})$ denote (population) variance of the sample mean.
- If Y_i is **independently and identifally distributed (i.i.d.)**

$$V(\bar{Y}) = \frac{1}{N^2} \sum_{i=1}^{N} V(Y_i) = \frac{V(Y)}{N}$$

- **Standard errors (標準誤差)**: standard deviation of the sample mean

$$SE(\bar{Y}) = \sqrt{V(Y)/N}$$

- We usually use **estimated** standard errors by replacing $V(Y)$ with sample variance $S(Y)$

$$\hat{SE}(\bar{Y}) = \sqrt{\hat{V}(Y)/N}$$

where $\hat{V}(Y) = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$

t-statistics

- Consider the null hypothesis $H_0 : E[Y] = \mu$.
- Define **t-statistics (t 統計量)**

$$t(\mu) = \frac{\bar{Y} - \mu}{\hat{SE}(\bar{Y})}$$

- When the null hypothesis is true, $t(\mu)$ follows some distribution.
- If the realized value of $t(\mu)$ is unlikely under the distribution, we reject the hypothesis.
- Question: What is the distribution?

Central Limit Theorem (CLT, 中心極限定理)

- Consider the i.i.d. sample of $Y_1, \cdots, Y_N$ drawn from the random variable Y with mean μ and variance σ^2. The following Z converges in distribution to the normal distribution.

$$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{Y_i - \mu}{\sigma} \xrightarrow{d} N(0,1)$$

- In this context

$$t(\mu) = \frac{\bar{Y} - \mu}{\hat{SE}(\bar{Y})} = \frac{1}{N} \sum_{i=1}^{N} \frac{Y_i - \mu}{\sqrt{\hat{V}(Y)/N}} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{Y_i - \mu}{\sqrt{\hat{V}(Y)}} \overset{approx}{\sim} N(0,1)$$

Simulation of CLT using R

- Consider the random variable Y_i that follows *binomial distribution (*二項分布*)* with probability 0.4.
- Here, $E[Y] = 0.4$ and $V[Y] = 0.4 \times (1 - 0.4)$.
- Define

$$Z = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{Y_i - E(Y)}{\sqrt{(V(Y)}}$$

- We demonstrate that as N gets larger, the distrubution of Z gets closer to the standard normal distribution.

Define a function

- This function draws `samplesize` observations from binomial distribution, calculate Z for each sample, and repeat this `Nreps` times.

```
f_simu_CLT <- function(Nreps, samplesize, distp ){

  output = numeric(Nreps)
  for (i in 1:Nreps ){

    test <- rbinom(n = samplesize, size = 1, prob = distp)

    EY <- distp
    VY <- (1 - distp)*distp

    output[i] <- ( mean(test) - EY ) / sqrt( VY / samplesize )
  }

  return(output)

}
```

```r
# Set the seed for the random number
set.seed(12345)

# Run simulation
Nreps = 500
result_CLT1 <- f_simu_CLT(Nreps, samplesize = 10 , distp = 0.4 )
result_CLT2 <- f_simu_CLT(Nreps, samplesize = 1000, distp = 0.4 )

# Random draw from standard normal distribution as comparison
result_stdnorm = rnorm(Nreps)

# Create dataframe
result_CLT_data <- data.frame(  Ybar_standardized_10 = result_CLT1,
                              Ybar_standardized_1000 = result_CLT2,
                              StandardNormal = result_stdnorm )
```

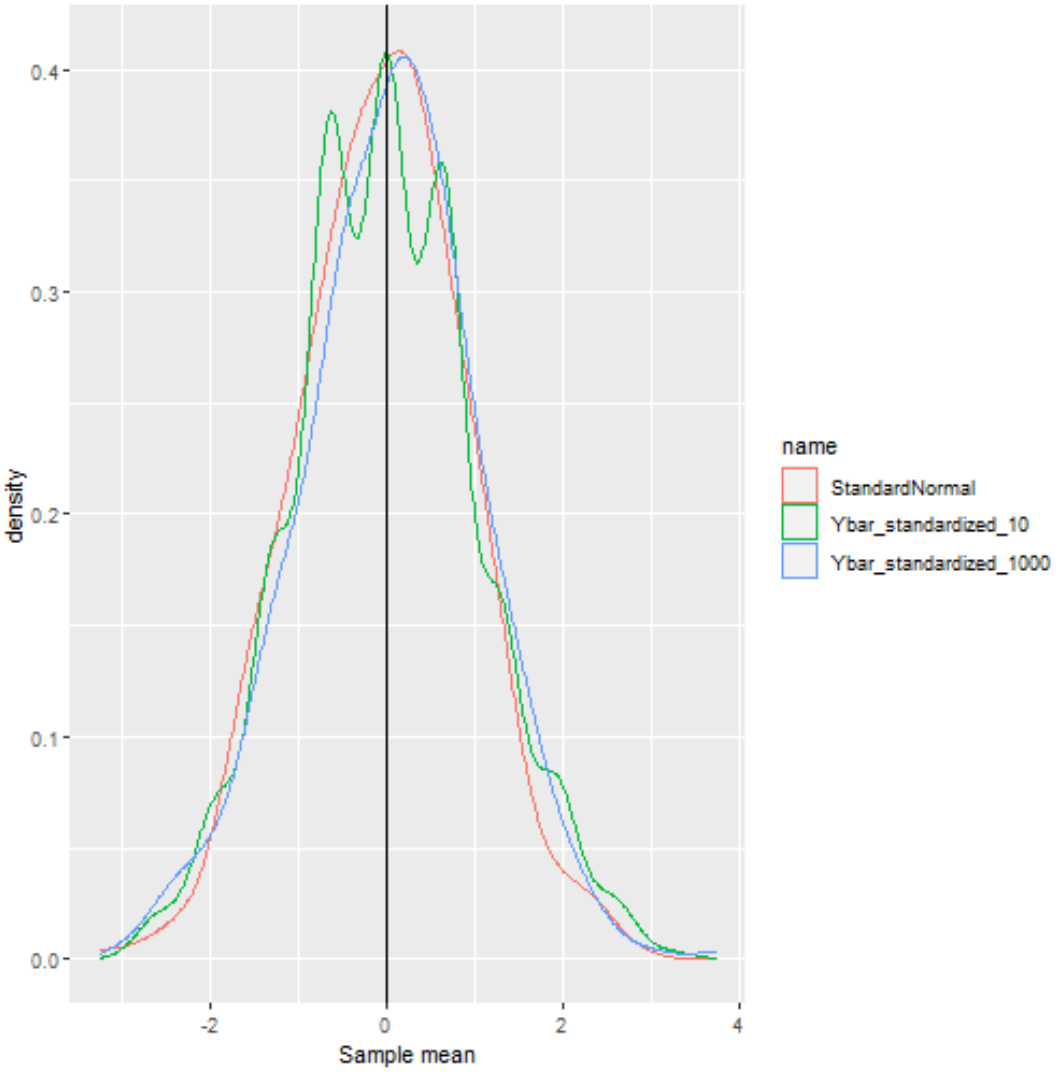
- Now take a look at the distribution.

```
# load tidyverse
library("tidyverse")

# Use "melt" to change the format of result_data
data_for_plot <- tidyr::pivot_longer(data = result_CLT_data, cols = everything())

# Use "ggplot2" to create the figure.
fig <-
  ggplot(data = data_for_plot) +
  xlab("Sample mean") +
  geom_density(aes(x = value, colour = name ),  ) +
  geom_vline(xintercept=0 ,colour="black")
```

```
plot(fig)
```


0.4
0.3
0.2
0.1
0.0
density
-2
0
2
4
Sample mean
name
StandardNormal
Ybar_standardized_10
Ybar_standardized_1000

Hypothesis Testing based on CLT

- Standard normal dist has mean 0 and standard deviation 1.
- Under this distribution, values larger than ± 2 appeared only about 5%!!!
- We say if $t(\mu)$ is larger than 2 in absolute value, we judge the hypothesis is unlikley to be true at 5%.
- We often say the sample mean is “significantly” different from 0.

Testing the difference of average between two groups

- Suppose that you want to test whether treatment effect is zero or not.
- The null hypothesis

$$H_0 : E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = 0$$

- t-statistics in this case is

$$t = \frac{\bar{Y}_1 - \bar{Y}_0}{\hat{SE}(\bar{Y}_1 - \bar{Y}_0)}$$

- Here, $\bar{Y}_d$ is conditional sample mean of each group d.

- The standard error is

$$SE(\bar{Y}_1 - \bar{Y}_0) = \sqrt{\frac{V^1(Y)}{N_1} + \frac{V^0(Y)}{N_0}}$$

  where $V^d(Y)$ is the population variance of observations in group d.