Lecture 2B: Estimation and Inference in RCT

Introduction to Econometrics,Fall 2018

Zhaopeng Qu

Nanjing University

9/20/2018

1. **Estimation and Inference in Randmoized Controlled Trial(RCT)**
2. **Reivew: Estimation and Inference**
3. **Hypothesis Testing: Concerning the Population Mean**
4. **Estimation and Inference in Randmoized Controlled Trial(RCT)**
5. **Examples of Randomized Controlled Trials**

Estimation and Inference in Randmoized Controlled Trial(RCT)

Randomly Assignment Eliminate Selection Bias

- RCT identify ATT and ATE

$$\begin{aligned}
\alpha_{assc} &= E[Y_{1i}|D_i = 1] - E[Y_{1i}|D_i = 0] \\
&= E[Y_{1i} - Y_{0i}|D_i = 1] + E[Y_{1i}|D_i = 0] - E[Y_{0i}|D_i = 0] \\
&= \alpha_{ATT} + Selection\ Bias \\
&= E[Y_{1i} - Y_{0i}] \\
&= \alpha_{ATE}
\end{aligned}$$

- Now that we know that the ATE and ATT are identified when we run a randomly experiment but how do we estimate it?

Introduction

- Question: whether attending college can increase your wage?
- Research Design: a randomized trial for attending college
    - Our sample is N individuals
    - Randomly assign treatment
        - Treatment group: N_1 individuals attend college
        - Control group: N_0 individuals do NOT attend college
- Compare a difference in wages between these two groups.

Reivew: Estimation and Inference

Reivew: Estimator and Estimate

- Given a random sample $\{Y_1, Y_2, ..., Y_n\}$ drawn from a population distribution Y that depends on an unknown parameter θ, thus $Y(\theta)$ and an **estimator** $\hat{\theta}$ is a function of the sample: thus $\hat{\theta}_n = h(Y_1, Y_2, ..., Y_n)$
- An estimator is a r.v. because it is a function of r.v.
- $\{\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_n\}$ is a sequence of r.v.s, so it has convergence in probability/distribution.
- If we would like to know the characteristics of a distribution, then we just need to know the properties of the r.v.
- eg. $Y \sim N(\mu, \sigma)$

Definition:Estimate

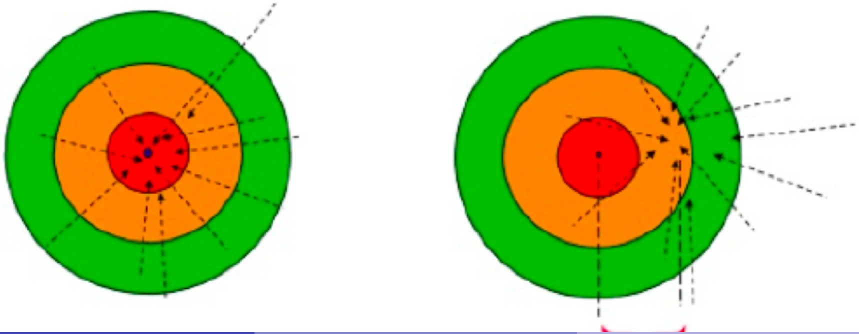
- An **estimate** is the *numerical value* of the **estimator** when it is actually computed using data from a specific sample. Thus if we have the actual data $\{y_1, y_2, ..., y_n\}$,then $\hat{\theta} = h(y_1, y_2, ..., y_n)$
- True or False and Why?
  - "My estimate was the sample mean and my estimator was 0.5"?

Review: Three Characteristics of an Estimator

- Let $\hat{\mu}_Y$ denote some estimator of μ_Y and $E(\hat{\mu}_Y)$ is the mean of the sampling distribution of $\hat{\mu}_Y$,

1. **Unbiasedness:** the estimator of μ_Y is unbiased if

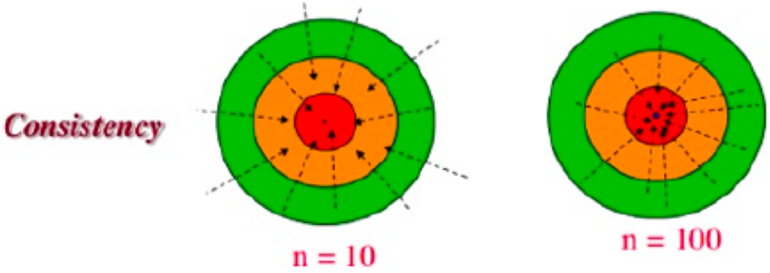
$$E(\hat{\mu}_Y) = \hat{\mu}_Y$$



Review: Three Characteristics of an Estimator

2. **Consistency:** the estimator of μ_Y is consistent if

$$E(\hat{\mu}_Y) \to \hat{\mu}_Y$$



Review: Three Characteristics of an Estimator

3. **Efficiency:** Let $\tilde{\mu}_Y$ be another estimator of μ_Y and suppose that both $\tilde{\mu}_Y$ and $\hat{\mu}_Y$ are unbiased.Then $\hat{\mu}_Y$ is said to be more efficient than $\tilde{\mu}_Y$

$$Var(\hat{\mu}_Y) < Var(\tilde{\mu}_Y)$$

- Comparing variances sometime is difficult if we do not restrict our attention to unbiased estimators.

Review: Properties of the sample mean

- Let μ_Y and σ_Y^2 denote the mean and variance of Y, then

$$E(\bar{Y}) = \frac{1}{n}\sum_{i=1}^{N} E(Y_i) = \mu_Y$$

- so $\bar{Y}$ is an **unbiased** estimator of μ_Y
- Based on the L.L.N.,$\bar{Y} \xrightarrow{p} \mu_Y$ so $\bar{Y}$ is also **consistent**.

Review: Properties of the sample mean

- The variance of the sample mean

$$\sigma_{\bar{Y}}^2 = Var(\bar{Y}) = Var(\frac{1}{n}\sum_{i=1}^{N} Y_i) = \frac{1}{n^2}\sum_{i=1}^{N} Var(Y_i) = \frac{\sigma_Y^2}{n}$$

- The standard deviation of the sample mean is

$$\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}}$$

Review: Properties of the sample mean

- Because efficiency entails a comparison of estimators, we need to specify the estimator or estimators to which Y is to be compared.
- An weighted average:

$$\widetilde{Y} = \frac{1}{n}\left(\frac{1}{2}Y_1 + \frac{3}{2}Y_2 + \frac{1}{2}Y_3 + \frac{3}{2}Y_4 + ... + \frac{1}{2}Y_{n-1} + \frac{3}{2}Y_n\right)$$

(Exercise 3.11)

- Then

$$Var(\widetilde{Y}) = 1.25\frac{\sigma_Y^2}{n} > \frac{\sigma_Y^2}{n} = Var(\overline{Y})$$

Exercise 3.11)

- Thus $\overline{Y}$ is more **efficient** than $\widetilde{Y}$
- In fact,$\overline{Y}$ is **BLUE**, the Best Linear Unbiased Estimator,thus the most efficient estimator among all estimators that are unbiased and are linear functions of $Y_1, Y_2, ..., Y_n$

the Sample Variance, Sample Standard Deviation and Standard Error

- Let μ_Y and σ_Y^2 denote the mean and variance of Y_i , then
- the **sample variance**, which is an estimator of the population variance σ_Y^2,

$$s_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

- The **sample standard deviation** is an estimator of the population standard deviation σ_Y,
- the **standard error** of the sample average $\bar{Y}$ is an estimator of the standard deviation of the sampling distribution of $\bar{Y}$.

Properties of the Sample Variance

1. Unbiasedness: $E(s_Y^2) = \sigma_Y^2$, thus S^2 is an unbiased estimator of σ_Y^2. It is also the reason why the average uses the divisor $n-1$ instead of n. [Exercise:3.18]
2. Consistency: $s_Y^2 \xrightarrow{p} \sigma_Y^2$, thus the sample variance is a consistent estimator of the population variance.[Appendix:3.3]

- **Prove it by yourself**

the Standard Error of $\bar{Y}$

- Because $\sigma_{\overline{Y}} = \frac{\sigma_Y}{\sqrt{n}}$ and $S_Y^2 \xrightarrow{p} \sigma_Y^2$,
- so the statement above justifies using $\frac{s_Y}{\sqrt{n}}$ as an estimator of the standard deviation of the sample mean, $\sigma_{\overline{Y}}$.
- It is called the **standard error** of the sample mean,$\bar{Y}$ and it denoted as $SE[\overline{Y}]$ or $\hat{\sigma}_{\overline{Y}}$.

Hypothesis Testing: Concerning the Population Mean

Review: Hypothesis Testing

Definition. A hypothesis is a *statement* about a **population** parameter, thus θ. Formally, we want to test whether is significantly different from a certain value μ_0

- **null hypothesis**:

$$H_0 : \theta = \mu_0$$

- The **alternative hypothesis**:

$$H_1 : \theta \neq \mu_0$$

(two-sided alternative)

Review: Hypothesis Testing

- If the value μ_0 does not lie within the calculated confidence interval, then we **reject** the null hypothesis.
- If the value μ_0 lie within the calculated confidence interval, then we **fail to reject** the null hypothesis.
- eg: college graduates earn 5000RMB monthly
  - the null hypothesis: $E(Y) = \theta = \mu_0 = 5000$
  - the alternative hypothesis: $E(Y) = \theta \neq \mu_0 = 5000$

Two Type Errors

- **Type I error**: A Type I error is when we reject the null hypothesis H_0 when it is in fact true.("left-wing").
- The probability of **Type I** error is denoted by α and called significance level or size of a test.

$$P(Type\ I\ error) = P(reject\ H_0 \mid H_0\ is\ true) = \alpha$$

  - By convention, α is chosen to be a small number, for example, a = 0.01, 0.05, or 0.10.

- **Type II Error**: A Type II error is when we fail to reject the null hypothesis when it is false.("right-wing")

$$P(Type\ II\ error) = P(accept\ H_0 \mid H_0 is\ false)$$

the significance of level

- Unfortunately, the probabilities of Type I and II errors are inversely related.
- By decreasing the probability of Type I error α, one makes the critical region smaller, which increases the probability of the Type II error. Thus it is impossible to make both errors arbitrary small.
- And we will never know with certainty whether we committed and error.
- So we have to set a standard of evaluation in the test, thus the significance of level α.

Decision Rule

- Usually, we has to carry the "burden of proof," thus the statement as H_1.
- We would like to prove that the assertion H_1 is true by showing that the data rejects H_0.
- The decision rule that leads us to reject or not to reject H_0 is based on a **test statistic**, which is a function of the data $T_n = T(Y_1, ..., Y_n)$
- Usually, one rejects H_0 if the test statistic falls into a **critical region**. A critical region is constructed by taking into account the probability of making a wrong **decision**.

t-statistic

- The **t-statistic** or **t-ratio**

$$\frac{\bar{Y}-\mu}{SE(\bar{Y})} \sim t_{n-1}$$

- let c denote the 97.5^{th} percentile in the t_{n-1} distribution.

$$P(-c < t \leq c) = 0.95$$

- where $c_{\frac{\alpha}{2}}$ is the critical value of the t distribution.

Testing procedure

- The following are the steps of the hypothesis testing:
  1. Specify H_0 and H_1.
  2. Choose the significance level α.
  3. Define a decision rule (critical region).
  4. Given the data compute the test statistic and see if it falls into the critical region.

Hypothesis Test of $\bar{Y}$

- Make $H_0 : E[Y] = \mu_{Y,0}$ and $H_1 : E[Y] \neq \mu_{Y,0}$
  - Step1: Compute the sample average $\bar{Y}$
  - Step2: Compute the **standard error** of $\bar{Y}$

$$SE(\overline{Y}) = \frac{s_Y}{\sqrt{n}}$$

  - Step3: Compute the **t-statistic**

$$t^{act} = \frac{\bar{Y} - \mu_{Y,0}}{SE\left(\bar{Y}\right)}$$

  - Step4: Reject the null hypothesis if
    - $\mid t^{act} \mid > critical\ value$
    - or if $p - value < significance\ level$

Estimation and Inference in Randmoized Controlled Trial(RCT)

Estimation in RCT

- We know the **ATE** of attending college

$$\alpha_{ATE} = E[Y_{1i} - Y_{0i}] = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$$

- α_{ATE} is a **population** parameter, while we only have a part of data of population, the sample.
- So we have to use the sample to estimate the population parameter, which is the **estimation**.

Estimation in RCT

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

$$\hat{\alpha}_{ATE} = \overline{Y_1} - \overline{Y_0}$$

$$\overline{Y}_1 = \frac{1}{N_1} \sum_{D_{i=1}} Y_i$$

$$\overline{Y}_0 = \frac{1}{N_0} \sum_{D_{i=0}} Y_i$$

Estimation of Sample Average Treatment Effect

- Estimator = Observed difference in means in the sample

$$\hat{\alpha}_{ATE} = \overline{Y_1} - \overline{Y_0}$$

- is this the estimator **unbiased** and **consistent** for ATE?
- **Prove it by yourself**!

Estimation in RCT

- Variance of $\hat{\alpha}_{ATE}$

$$
\begin{aligned}
Var(\hat{\alpha}_{ATE}) &= Var(\overline{Y}_1 - \overline{Y}_0) \\
&= Var(\overline{Y}_1) - Var(\overline{Y}_0) \\
&= Var(\overline{Y}_1) - Var(\overline{Y}_0) \\
&= \frac{\sigma^2_{Y1}}{N_1} + \frac{\sigma^2_{Y2}}{N_0}
\end{aligned}
$$

- The standard deviation of $\hat{\alpha}_{ATE}$

$$
SD(\hat{\alpha}_{ATE}) = \sqrt{(\frac{\sigma^2_{Y1}}{N_1} + \frac{\sigma^2_{Y2}}{N_0})}
$$

Estimation in RCT

- Since σ_Y is unknown,in practice, we use the value of sample,$\hat{\sigma}_Y$ to replace population parameter σ_Y
- Thus the Standard error of $\hat{\alpha}_{ATE}$

$$SE(\hat{\alpha}_{ATE}) = \sqrt{(\frac{s_{Y1}^2}{N_1} + \frac{s_{Y2}^2}{N_0})}$$

Hypothesis Testing

- To illustrate a test for the difference between two means, let μ_m be the mean hourly earning in the population of women recently graduated from college and let μ_w be the population mean for recently graduated men.
- Then the null hypothesis and the two-sided alternative hypothesis are

$$H_0 : \mu_m = \mu_w$$

$$H_1 : \mu_m \neq \mu_w$$

- Consider the null hypothesis that mean earnings for these two populations differ by a certain amount, say d_0. The null hypothesis that men and women in these populations have the same mean earnings corresponds to $H_0 : d_0 = \mu_m - \mu_w = 0$

ATE in t-statistic

- The t-statistic for testing the null hypothesis is constructed analogously to the t-statistic for testing a hypothesis about a single population mean,thus t-statistic for comparing two means is

$$t = \frac{\bar{Y_m} - \bar{Y_w} - d_0}{SE(\bar{Y_m} - \bar{Y_w})}$$

- If both n_m and n_w are large, then this t-statistic has a standard normal distribution when the null hypothesis is true.

Confidence Intervals for the Difference Between Two Population Means

- the 95% two-sided confidence interval for d consists of those values of d within ± 1.96 standard errors of $\bar{Y_m} - \bar{Y_w}$ thus $d = \mu_m - \mu_w$ is:

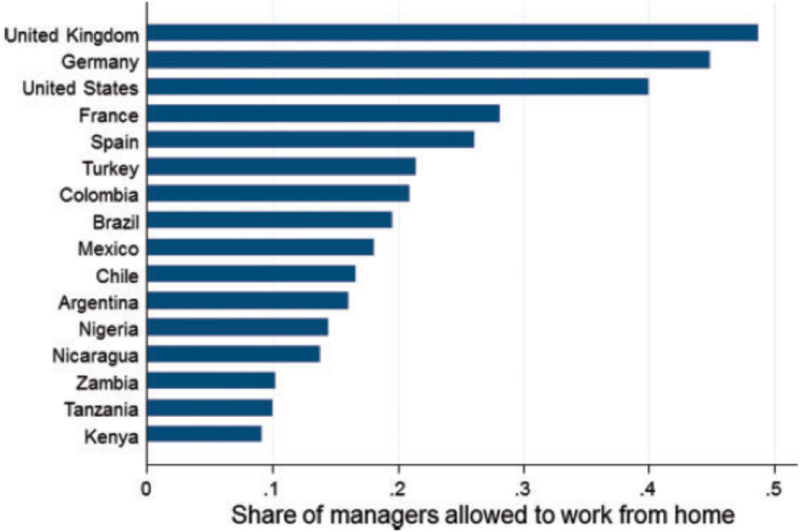
$$(\bar{Y_m} - \bar{Y_w}) \pm 1.96 SE(\bar{Y_m} - \bar{Y_w})$$

Examples of Randomized Controlled Trials

working from home (WFH)

- "Does Working from Home Work? Evidence from a Chinese Experiment",by Nicholas A. Bloom, *James Liang*, John Roberts, Zhichun Jenny Ying The **Quarterly Journal of Economics**,February 2015, Vol. 130, Issue 1, Pages 165-218.
- Basic Question: *WFH=SFH(Shirking from Home)*?

working from home (WFH)

- 

Motivations

- Working from home is a modern management practice which appears to be stochastically spreading in the US and Europe
- 20 million people in US report working from home at least once per week
- Little evidence on the effect of workplace flexibility
    - productivity
    - employee satisfaction
    - shirking

Ctrip Experiment

- Ctrip, China's largest travel-agent, with16,000 employees, $6bn NASDAQ
- the CEO of Ctrip, James Liang, was an Econ PhD at Stanford and decided to run a experiment to test WFH.

Ctrip Experiment

- 

Headquarters in Shanghai

Main Lobby

The Experimental Design

- The experiment runs on airfare & hotel departments in Shanghai.
- Main Work: Employees take calls and make bookings.
- Treatment: work 4 shifts (days) a week at home and to work the 5th shift in the office on a fixed day.
- Control: work in the office on all 5 days.

The Experimental Design

- In early November 2010, employees in the airfare and hotel booking departments were informed of the WFH program.
- Of the 994 employees in the airfare and hotel booking departments, 503 (51%) volunteered for the experiment.
- Among the volunteers, 249 (50%) of the employees met the eligibility requirements and were recruited into the experiment.
- The treatment and control groups were then determined from this group of 249 employees through a public lottery.

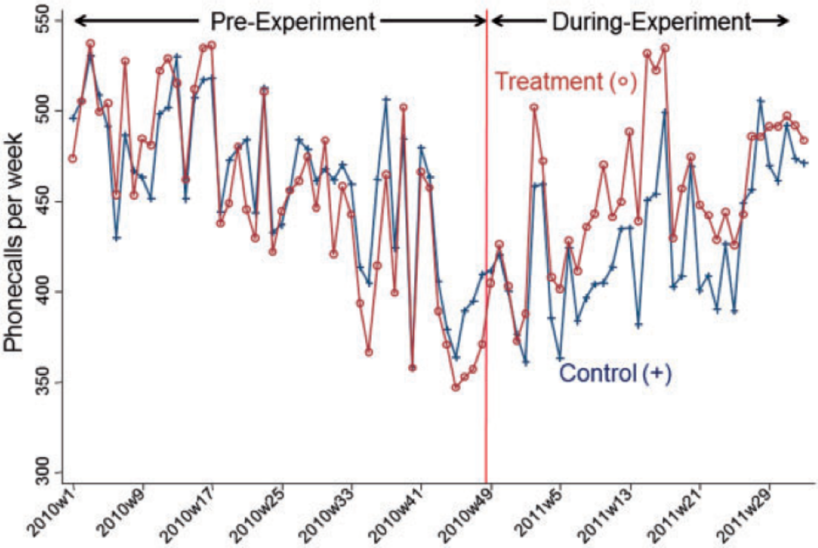
Experimental Design

- 

Treatment groups were determined by a lottery

Working at home

Result

- 
Pre-Experiment
During-Experiment
Treatment (∘)
Control (+)
Phonecalls per week
550
500
450
400
350
300
2010w1
2010w9
2010w17
2010w25
2010w33
2010w41
2010w49
2011w5
2011w13
2011w21
2011w29

Result

-

Variables	(1) Minutes on the phone	(2) Minutes on the phone/days worked	(3) Days worked	(4) Minutes on the phone	(5) Minutes on the phone/days worked	(6) Days worked
$Experiment_t$*$Treatment_i$	0.088***	0.063***	0.025**	0.069**	0.049*	0.021
	(0.027)	(0.024)	(0.012)	(0.030)	(0.027)	(0.013)
$Experiment_t$*$Treatment_i$*				0.069*	0.055*	0.014
$[\text{total commute} > 120\,\text{min}]_i$				(0.036)	(0.031)	(0.017)
Number of employees	134	134	134	134	134	134
Number of weeks	85	85	85	85	85	85
Observations	9,426	9,426	9,426	9,426	9,426	9,426

Notes. The regressions are run at the individual by week level, with a full set of individual and week fixed effects. *Experiment*treatment* is the interaction of the period of the experimentation (December 6, 2010, until August 14, 2011) by an individual having an even birthdate (2nd, 4th, 6th, etc. day of the month). The pre-experiment period refers to January 1, 2010, until November 28, 2010. During the experiment period refers to December 6, 2010, to August 14, 2011. In columns (4)–(6), *Experiment* × *Treatment* is further interacted with a dummy variable indicating whether an employee's total daily commute (to and from work) is longer than 120 minutes (21.3% of employees have a commute longer than 120 minutes). Standard errors are clustered at the individual level. Once employees quit they are dropped from the data. *** denotes 1% significance, ** 5% significance, and * 10% significance. Minutes on the phone are recorded from the call logs.

Conclusion

- They found a highly significant 13% increase in employee performance from WFH,
- of which about 9% was from employees working more minutes of their shift period (fewer breaks and sick days)
- and about 4% from higher performance per minute.
- Home workers also reported substantially higher work satisfaction and psychological attitude scores, and their job attrition rates fell by over 50%.