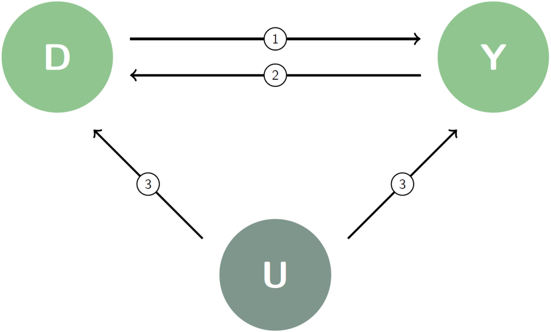
Financial Econometrics

Randomized Controlled Trials and Matching

Tim C.C. Hung 洪志清

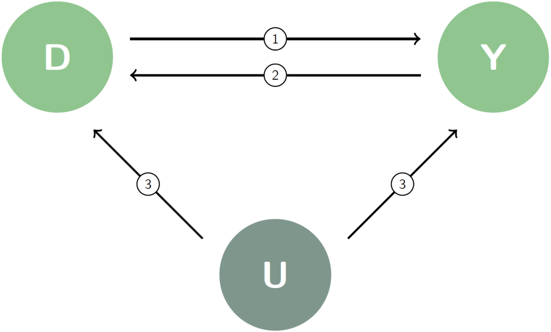
May 9^{th}, 2022

Identify Causal Effect



- There are three possible reasons why we observe treatment D is correlated with outcome Y.
  1. D causes Y.
  2. Y causes D.
  3. Other confounding factor U affects D and Y.
     (The famous example: number of drowned and ice cream sales.)

Identify Causal Effect



- To identify causal effect of treatment, we need to make sure the observed relationship between treatment D and outcome Y is due to 1.

- Identification strategies help us establish causal relationship from the **observed data** by imposing some assumptions.

Randomized Controlled Trial

- The most credible identification strategy!
- Randomized controlled trial (RCT):
  - Each observation (e.g. individual, household, school, state, or country) is randomly assigned to treatment and control group.
- RCT has two features that can help us hold **other things equal** and eliminates selection bias:
  1. Randomly assign treatment
  2. Sufficiently large sample size
- Randomly assign treatment (such as a coin flip) ensures that the probability of receiving treatment is unrelated to any other confounding factors (individual characteristics).
- RCT is called the **gold standard** for causal inference.

Randomly Assign Treatment

- Randomly assign treatment implies:
  - The values of potential outcomes are independent of treatment assigned.
  - (Y_i^1, Y_i^0) are independent of D_i

$$(Y_i^1, Y_i^0) \perp D_i$$

- **Intuition**: Treatment assignment D is random and not based on an individual's value of potential outcome (Y_i^1, Y_i^0).

- Randomly assign treatment makes treatment and control group to be similar along all characteristics (including Y_i^1, Y_i^0).
  - $E[Y_i^0|D_i = 1] = E[Y_i^0|D_i = 0]$
  - $E[Y_i^1|D_i = 1] = E[Y_i^1|D_i = 0]$

- Therefore, we can eliminate selection bias:
  - $\underbrace{E[Y_i^0|D_i = 1] - E[Y_i^0|D_i = 0]}_{\text{Selection Bias}} = 0$

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.

- RCT Identifies ATT and ATE.

$$
\begin{aligned}
&\underbrace{E[Y_i|D_i=1]-E[Y_i|D_i=0]}_{\text{Observed Difference in Average Outcome}} \\
&= \underbrace{E[Y_i^1-Y_i^0|D_i=1]}_{\text{Causal Effect (ATT)}} + \underbrace{E[Y_i^0|D_i=1]-E[Y_i^0|D_i=0]}_{\text{Selection Bias}} \\
&= \underbrace{E[Y_i^1-Y_i^0|D_i=1]}_{\text{Causal Effect (ATT)}} + \underbrace{0}_{\text{Selection Bias}} \\
&= \underbrace{E[Y_i^1-Y_i^0]}_{\text{Causal Effect (ATE)}}
\end{aligned}
$$

RCT: Estimation

- Now we know ATE and ATT are identified, how will we estimate it?
- Example: 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.
$\rightarrow$Randomly assign master degree to every Taiwanese.
- We can obtain the ATE of master degree on earning:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \underbrace{=}_{\text{Due to RCT}} E[Y_i^1 - Y_i^0] = \alpha_{ATE}$$

- But, we do not have population data.
$\rightarrow$We will come back to this point at the end.

RCT: Estimation

- 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 the treatment group and the control group:

$$\widehat{\alpha_{ATE}} = \bar{Y_1} - \bar{Y_0}$$

- where $\bar{Y_1} = \frac{1}{N_1} \sum_{D_i=1} Y_i$, $\bar{Y_0} = \frac{1}{N_0} \sum_{D_i=0} Y_i$

- Now we want to use **sample estimator** 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.

RCT: Hypothesis Testing

National Taiwan University

1 Choose a nul hypothesis:
  - $H_0 : \alpha_{ATE} = 0$ or $H_0 : \alpha_{ATE} = \mu$
  - The goal is to see if we can reject the null.

2 Choose a test statistic:

$$t = \frac{\widehat{\alpha_{ATE}} - \alpha_{ATE}}{\widehat{SE}(\alpha_{\hat{ATE}})}$$

3 Estimate standard error of the estimator:

$$\widehat{SE}(\alpha_{\hat{ATE}}) = \hat{\sigma_Y}\sqrt{[\frac{1}{N_1} + \frac{1}{N_0}]}$$

RCT: Hypothesis Testing

4 Determine the distribution of the test statistic under the null.
   - If sample size is sufficient large, using CLT, t-statistic will have standard normal distribution.

5 Calculate the probability of wrongly reject null hypothesis given null hypothesis is true (p-value).
   - We reject the null hypothesis $H_0 : \alpha_{ATE} = 0$ against the alternative $H_1 : \alpha_{ATE} \neq 0$ at the 5% significance level if $|t| > 1.96$.

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.
- Threats to Internal Validity:
  - Failure of randomization
  - Non-compliance with experimental protocol
  - Attrition

- **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.
- Threats to External Validity:
  - Non-representative sample
  - Non-representative program
    $\rightarrow$Actual implementations are not randomized (nor full scale).

Observational Studies

National Taiwan University

- RCT eliminates selection bias by randomly assigning treatment. Thus, treatment group and control group are comparable.
- But implementing a randomized experiment in social science is very expensive and sometimes has ethical issues.
- In social science, many empirical studies use **non-experimental data**. We call this type of empirical researches as **observational studies**.
- In contrast to RCT, in observational studies, **researchers can NOT control the assignment of treatment**.
- We need to directly control for the observed variables and use indirect methods to adjust for unobserved variables.
  $\rightarrow$ Force other thing equal in observed and unobserved variables.
- We want to design observational studies that approximate experiments.

Main Idea of Matching

- Assume all confounding factors are observable to researchers.
- Matching is a way to eliminate selection bias by controlling observable covariates.
- By constructing a comparison sample of untreated units with the same characteristics as the sample of treated units.
- After controlling observed covariates X_i, we could identify causal effect of treatment.
- This can be accomplished by **matching** treated and untreated units with the same characteristics.
- **Example**: We want to estimate the causal effect of job training program on worker's earnings. Suppose **age** is the only confounding factors that affect both earnings and job training decision.

Matching: an Example

Trainees			Non-Trainees			Matched Sample		
unit	age	earnings	unit	age	earnings	unit	age	earnings
1	28	17700	1	43	20900			
2	34	10200	2	50	31000			
3	29	14400	3	30	21000			
4	25	20800	4	27	9300			
5	29	6100	5	54	41100			
6	23	28600	6	48	29800			
7	33	21900	7	39	42000			
8	27	28800	8	28	8800			
9	31	20300	9	24	25500			
10	26	28100	10	33	15500			
11	25	9400	11	26	400			
12	27	14300	12	31	26600			
13	29	12500	13	26	16500			
14	24	19700	14	34	24200			
15	25	10100	15	25	23300			
16	43	10700	16	24	9700			
17	28	11500	17	29	6200			
18	27	10700	18	35	30200			
19	28	16300	19	32	17800			
			20	23	9500			
			21	32	25900			
Avg:	28.5	16426	Avg:	33	20724			

Matching: an Example

National Taiwan University

Trainees			Non-Trainees			Matched Sample		
unit	age	earnings	unit	age	earnings	unit	age	earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000			
3	29	14400	3	30	21000			
4	25	20800	4	27	9300			
5	29	6100	5	54	41100			
6	23	28600	6	48	29800			
7	33	21900	7	39	42000			
8	27	28800	8	28	8800			
9	31	20300	9	24	25500			
10	26	28100	10	33	15500			
11	25	9400	11	26	400			
12	27	14300	12	31	26600			
13	29	12500	13	26	16500			
14	24	19700	14	34	24200			
15	25	10100	15	25	23300			
16	43	10700	16	24	9700			
17	28	11500	17	29	6200			
18	27	10700	18	35	30200			
19	28	16300	19	32	17800			
			20	23	9500			
			21	32	25900			
Avg:	28.5	16426	Avg:	33	20724			

Matching: an Example

National Taiwan University

Trainees			Non-Trainees			Matched Sample		
unit	age	earnings	unit	age	earnings	unit	age	earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000	14	34	24200
3	29	14400	3	30	21000			
4	25	20800	4	27	9300			
5	29	6100	5	54	41100			
6	23	28600	6	48	29800			
7	33	21900	7	39	42000			
8	27	28800	8	28	8800			
9	31	20300	9	24	25500			
10	26	28100	10	33	15500			
11	25	9400	11	26	400			
12	27	14300	12	31	26600			
13	29	12500	13	26	16500			
14	24	19700	14	34	24200			
15	25	10100	15	25	23300			
16	43	10700	16	24	9700			
17	28	11500	17	29	6200			
18	27	10700	18	35	30200			
19	28	16300	19	32	17800			
			20	23	9500			
			21	32	25900			
Avg:	28.5	16426	Avg:	33	20724			

Matching: an Example

Trainees			Non-Trainees			Matched Sample		
unit	age	earnings	unit	age	earnings	unit	age	earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000	14	34	24200
3	29	14400	3	30	21000	17	29	6200
4	25	20800	4	27	9300			
5	29	6100	5	54	41100			
6	23	28600	6	48	29800			
7	33	21900	7	39	42000			
8	27	28800	8	28	8800			
9	31	20300	9	24	25500			
10	26	28100	10	33	15500			
11	25	9400	11	26	400			
12	27	14300	12	31	26600			
13	29	12500	13	26	16500			
14	24	19700	14	34	24200			
15	25	10100	15	25	23300			
16	43	10700	16	24	9700			
17	28	11500	17	29	6200			
18	27	10700	18	35	30200			
19	28	16300	19	32	17800			
			20	23	9500			
			21	32	25900			
Avg:	28.5	16426	Avg:	33	20724			

Matching: an Example

National Taiwan University

Trainees

unit	age	earnings
1	28	17700
2	34	10200
3	29	14400
4	25	20800
5	29	6100
6	23	28600
7	33	21900
8	27	28800
9	31	20300
10	26	28100
11	25	9400
12	27	14300
13	29	12500
14	24	19700
15	25	10100
16	43	10700
17	28	11500
18	27	10700
19	28	16300
Avg:	28.5	16426

Non-Trainees

unit	age	earnings
1	43	20900
2	50	31000
3	30	21000
4	27	9300
5	54	41100
6	48	29800
7	39	42000
8	28	8800
9	24	25500
10	33	15500
11	26	400
12	31	26600
13	26	16500
14	34	24200
15	25	23300
16	24	9700
17	29	6200
18	35	30200
19	32	17800
20	23	9500
21	32	25900
Avg:	33	20724

Matched Sample

unit	age	earnings
8	28	8800
14	34	24200
17	29	6200
15	25	23300

Matching: an Example

Trainees

unit	age	earnings
1	28	17700
2	34	10200
3	29	14400
4	25	20800
5	29	6100
6	23	28600
7	33	21900
8	27	28800
9	31	20300
10	26	28100
11	25	9400
12	27	14300
13	29	12500
14	24	19700
15	25	10100
16	43	10700
17	28	11500
18	27	10700
19	28	16300
Avg:	28.5	16426

Non-Trainees

unit	age	earnings
1	43	20900
2	50	31000
3	30	21000
4	27	9300
5	54	41100
6	48	29800
7	39	42000
8	28	8800
9	24	25500
10	33	15500
11	26	400
12	31	26600
13	26	16500
14	34	24200
15	25	23300
16	24	9700
17	29	6200
18	35	30200
19	32	17800
20	23	9500
21	32	25900
Avg:	33	20724

Matched Sample

unit	age	earnings
8	28	8800
14	34	24200
17	29	6200
15	25	23300
17	29	6200

Matching: an Example

National Taiwan University

Trainees		
unit	age	earnings
1	28	17700
2	34	10200
3	29	14400
4	25	20800
5	29	6100
6	23	28600
7	33	21900
8	27	28800
9	31	20300
10	26	28100
11	25	9400
12	27	14300
13	29	12500
14	24	19700
15	25	10100
16	43	10700
17	28	11500
18	27	10700
19	28	16300
Avg:	28.5	16426

Non-Trainees		
unit	age	earnings
1	43	20900
2	50	31000
3	30	21000
4	27	9300
5	54	41100
6	48	29800
7	39	42000
8	28	8800
9	24	25500
10	33	15500
11	26	400
12	31	26600
13	26	16500
14	34	24200
15	25	23300
16	24	9700
17	29	6200
18	35	30200
19	32	17800
20	23	9500
21	32	25900
Avg:	33	20724

Matched Sample		
unit	age	earnings
8	28	8800
14	34	24200
17	29	6200
15	25	23300
17	29	6200
20	23	9500

Matching: an Example

Trainees

unit	age	earnings
1	28	17700
2	34	10200
3	29	14400
4	25	20800
5	29	6100
6	23	28600
7	33	21900
8	27	28800
9	31	20300
10	26	28100
11	25	9400
12	27	14300
13	29	12500
14	24	19700
15	25	10100
16	43	10700
17	28	11500
18	27	10700
19	28	16300
Avg:	28.5	16426

Non-Trainees

unit	age	earnings
1	43	20900
2	50	31000
3	30	21000
4	27	9300
5	54	41100
6	48	29800
7	39	42000
8	28	8800
9	24	25500
10	33	15500
11	26	400
12	31	26600
13	26	16500
14	34	24200
15	25	23300
16	24	9700
17	29	6200
18	35	30200
19	32	17800
20	23	9500
21	32	25900
Avg:	33	20724

Matched Sample

unit	age	earnings
8	28	8800
14	34	24200
17	29	6200
15	25	23300
17	29	6200
20	23	9500
10	33	15500

Matching: an Example

國立臺灣大學 National Taiwan University

Trainees			Non-Trainees			Matched Sample		
unit	age	earnings	unit	age	earnings	unit	age	earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000	14	34	24200
3	29	14400	3	30	21000	17	29	6200
4	25	20800	4	27	9300	15	25	23300
5	29	6100	5	54	41100	17	29	6200
6	23	28600	6	48	29800	20	23	9500
7	33	21900	7	39	42000	10	33	15500
8	27	28800	8	28	8800	4	27	9300
9	31	20300	9	24	25500			
10	26	28100	10	33	15500			
11	25	9400	11	26	400			
12	27	14300	12	31	26600			
13	29	12500	13	26	16500			
14	24	19700	14	34	24200			
15	25	10100	15	25	23300			
16	43	10700	16	24	9700			
17	28	11500	17	29	6200			
18	27	10700	18	35	30200			
19	28	16300	19	32	17800			
			20	23	9500			
			21	32	25900			
Avg:	28.5	16426	Avg:	33	20724			

Matching: an Example

Trainees

unit	age	earnings
1	28	17700
2	34	10200
3	29	14400
4	25	20800
5	29	6100
6	23	28600
7	33	21900
8	27	28800
9	31	20300
10	26	28100
11	25	9400
12	27	14300
13	29	12500
14	24	19700
15	25	10100
16	43	10700
17	28	11500
18	27	10700
19	28	16300
Avg:	28.5	16426

Non-Trainees

unit	age	earnings
1	43	20900
2	50	31000
3	30	21000
4	27	9300
5	54	41100
6	48	29800
7	39	42000
8	28	8800
9	24	25500
10	33	15500
11	26	400
12	31	26600
13	26	16500
14	34	24200
15	25	23300
16	24	9700
17	29	6200
18	35	30200
19	32	17800
20	23	9500
21	32	25900
Avg:	33	20724

Matched Sample

unit	age	earnings
8	28	8800
14	34	24200
17	29	6200
15	25	23300
17	29	6200
20	23	9500
10	33	15500
4	27	9300
12	31	26600

Matching: an Example

Trainees			Non-Trainees			Matched Sample		
unit	age	earnings	unit	age	earnings	unit	age	earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000	14	34	24200
3	29	14400	3	30	21000	17	29	6200
4	25	20800	4	27	9300	15	25	23300
5	29	6100	5	54	41100	17	29	6200
6	23	28600	6	48	29800	20	23	9500
7	33	21900	7	39	42000	10	33	15500
8	27	28800	8	28	8800	4	27	9300
9	31	20300	9	24	25500	12	31	26600
10	26	28100	10	33	15500	11,13	26	8450
11	25	9400	11	26	400			
12	27	14300	12	31	26600			
13	29	12500	13	26	16500			
14	24	19700	14	34	24200			
15	25	10100	15	25	23300			
16	43	10700	16	24	9700			
17	28	11500	17	29	6200			
18	27	10700	18	35	30200			
19	28	16300	19	32	17800			
			20	23	9500			
			21	32	25900			
Avg:	28.5	16426	Avg:	33	20724			

Matching: an Example

National Taiwan University

Trainees			Non-Trainees			Matched Sample		
unit	age	earnings	unit	age	earnings	unit	age	earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000	14	34	24200
3	29	14400	3	30	21000	17	29	6200
4	25	20800	4	27	9300	15	25	23300
5	29	6100	5	54	41100	17	29	6200
6	23	28600	6	48	29800	20	23	9500
7	33	21900	7	39	42000	10	33	15500
8	27	28800	8	28	8800	4	27	9300
9	31	20300	9	24	25500	12	31	26600
10	26	28100	10	33	15500	11,13	26	8450
11	25	9400	11	26	400	15	25	23300
12	27	14300	12	31	26600			
13	29	12500	13	26	16500			
14	24	19700	14	34	24200			
15	25	10100	15	25	23300			
16	43	10700	16	24	9700			
17	28	11500	17	29	6200			
18	27	10700	18	35	30200			
19	28	16300	19	32	17800			
			20	23	9500			
			21	32	25900			
Avg:	28.5	16426	Avg:	33	20724			

Matching: an Example

Trainees

unit	age	earnings
1	28	17700
2	34	10200
3	29	14400
4	25	20800
5	29	6100
6	23	28600
7	33	21900
8	27	28800
9	31	20300
10	26	28100
11	25	9400
12	27	14300
13	29	12500
14	24	19700
15	25	10100
16	43	10700
17	28	11500
18	27	10700
19	28	16300
Avg:	28.5	16426

Non-Trainees

unit	age	earnings
1	43	20900
2	50	31000
3	30	21000
4	27	9300
5	54	41100
6	48	29800
7	39	42000
8	28	8800
9	24	25500
10	33	15500
11	26	400
12	31	26600
13	26	16500
14	34	24200
15	25	23300
16	24	9700
17	29	6200
18	35	30200
19	32	17800
20	23	9500
21	32	25900
Avg:	33	20724

Matched Sample

unit	age	earnings
8	28	8800
14	34	24200
17	29	6200
15	25	23300
17	29	6200
20	23	9500
10	33	15500
4	27	9300
12	31	26600
11,13	26	8450
15	25	23300
4	27	9300

Matching: an Example

Trainees			Non-Trainees			Matched Sample		
unit	age	earnings	unit	age	earnings	unit	age	earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000	14	34	24200
3	29	14400	3	30	21000	17	29	6200
4	25	20800	4	27	9300	15	25	23300
5	29	6100	5	54	41100	17	29	6200
6	23	28600	6	48	29800	20	23	9500
7	33	21900	7	39	42000	10	33	15500
8	27	28800	8	28	8800	4	27	9300
9	31	20300	9	24	25500	12	31	26600
10	26	28100	10	33	15500	11,13	26	8450
11	25	9400	11	26	400	15	25	23300
12	27	14300	12	31	26600	4	27	9300
13	29	12500	13	26	16500	17	29	6200
14	24	19700	14	34	24200			
15	25	10100	15	25	23300			
16	43	10700	16	24	9700			
17	28	11500	17	29	6200			
18	27	10700	18	35	30200			
19	28	16300	19	32	17800			
			20	23	9500			
			21	32	25900			
Avg:	28.5	16426	Avg:	33	20724			

Matching: an Example

Trainees			Non-Trainees			Matched Sample		
unit	age	earnings	unit	age	earnings	unit	age	earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000	14	34	24200
3	29	14400	3	30	21000	17	29	6200
4	25	20800	4	27	9300	15	25	23300
5	29	6100	5	54	41100	17	29	6200
6	23	28600	6	48	29800	20	23	9500
7	33	21900	7	39	42000	10	33	15500
8	27	28800	8	28	8800	4	27	9300
9	31	20300	9	24	25500	12	31	26600
10	26	28100	10	33	15500	11,13	26	8450
11	25	9400	11	26	400	15	25	23300
12	27	14300	12	31	26600	4	27	9300
13	29	12500	13	26	16500	17	29	6200
14	24	19700	14	34	24200	9,16	24	17700
15	25	10100	15	25	23300			
16	43	10700	16	24	9700			
17	28	11500	17	29	6200			
18	27	10700	18	35	30200			
19	28	16300	19	32	17800			
			20	23	9500			
			21	32	25900			
Avg:	28.5	16426	Avg:	33	20724			

Matching: an Example

National Taiwan University

Trainees

unit	age	earnings
1	28	17700
2	34	10200
3	29	14400
4	25	20800
5	29	6100
6	23	28600
7	33	21900
8	27	28800
9	31	20300
10	26	28100
11	25	9400
12	27	14300
13	29	12500
14	24	19700
15	25	10100
16	43	10700
17	28	11500
18	27	10700
19	28	16300
Avg:	28.5	16426

Non-Trainees

unit	age	earnings
1	43	20900
2	50	31000
3	30	21000
4	27	9300
5	54	41100
6	48	29800
7	39	42000
8	28	8800
9	24	25500
10	33	15500
11	26	400
12	31	26600
13	26	16500
14	34	24200
15	25	23300
16	24	9700
17	29	6200
18	35	30200
19	32	17800
20	23	9500
21	32	25900
Avg:	33	20724

Matched Sample

unit	age	earnings
8	28	8800
14	34	24200
17	29	6200
15	25	23300
17	29	6200
20	23	9500
10	33	15500
4	27	9300
12	31	26600
11,13	26	8450
15	25	23300
4	27	9300
17	29	6200
9,16	24	17700
15	25	23300

Matching: an Example

國立臺灣大學 National Taiwan University

Trainees

unit	age	earnings
1	28	17700
2	34	10200
3	29	14400
4	25	20800
5	29	6100
6	23	28600
7	33	21900
8	27	28800
9	31	20300
10	26	28100
11	25	9400
12	27	14300
13	29	12500
14	24	19700
15	25	10100
16	43	10700
17	28	11500
18	27	10700
19	28	16300
Avg:	28.5	16426

Non-Trainees

unit	age	earnings
1	43	20900
2	50	31000
3	30	21000
4	27	9300
5	54	41100
6	48	29800
7	39	42000
8	28	8800
9	24	25500
10	33	15500
11	26	400
12	31	26600
13	26	16500
14	34	24200
15	25	23300
16	24	9700
17	29	6200
18	35	30200
19	32	17800
20	23	9500
21	32	25900
Avg:	33	20724

Matched Sample

unit	age	earnings
8	28	8800
14	34	24200
17	29	6200
15	25	23300
17	29	6200
20	23	9500
10	33	15500
4	27	9300
12	31	26600
11,13	26	8450
15	25	23300
4	27	9300
17	29	6200
9,16	24	17700
15	25	23300
1	43	20900

Matching: an Example

National Taiwan University

Trainees			Non-Trainees			Matched Sample		
unit	age	earnings	unit	age	earnings	unit	age	earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000	14	34	24200
3	29	14400	3	30	21000	17	29	6200
4	25	20800	4	27	9300	15	25	23300
5	29	6100	5	54	41100	17	29	6200
6	23	28600	6	48	29800	20	23	9500
7	33	21900	7	39	42000	10	33	15500
8	27	28800	8	28	8800	4	27	9300
9	31	20300	9	24	25500	12	31	26600
10	26	28100	10	33	15500	11,13	26	8450
11	25	9400	11	26	400	15	25	23300
12	27	14300	12	31	26600	4	27	9300
13	29	12500	13	26	16500	17	29	6200
14	24	19700	14	34	24200	9,16	24	17700
15	25	10100	15	25	23300	15	25	23300
16	43	10700	16	24	9700	1	43	20900
17	28	11500	17	29	6200	8	28	8800
18	27	10700	18	35	30200			
19	28	16300	19	32	17800			
			20	23	9500			
			21	32	25900			
Avg:	28.5	16426	Avg:	33	20724			

Matching: an Example

Trainees			Non-Trainees			Matched Sample		
unit	age	earnings	unit	age	earnings	unit	age	earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000	14	34	24200
3	29	14400	3	30	21000	17	29	6200
4	25	20800	4	27	9300	15	25	23300
5	29	6100	5	54	41100	17	29	6200
6	23	28600	6	48	29800	20	23	9500
7	33	21900	7	39	42000	10	33	15500
8	27	28800	8	28	8800	4	27	9300
9	31	20300	9	24	25500	12	31	26600
10	26	28100	10	33	15500	11,13	26	8450
11	25	9400	11	26	400	15	25	23300
12	27	14300	12	31	26600	4	27	9300
13	29	12500	13	26	16500	17	29	6200
14	24	19700	14	34	24200	9,16	24	17700
15	25	10100	15	25	23300	15	25	23300
16	43	10700	16	24	9700	1	43	20900
17	28	11500	17	29	6200	8	28	8800
18	27	10700	18	35	30200	4	27	9300
19	28	16300	19	32	17800			
			20	23	9500			
			21	32	25900			
Avg:	28.5	16426	Avg:	33	20724			

Matching: an Example

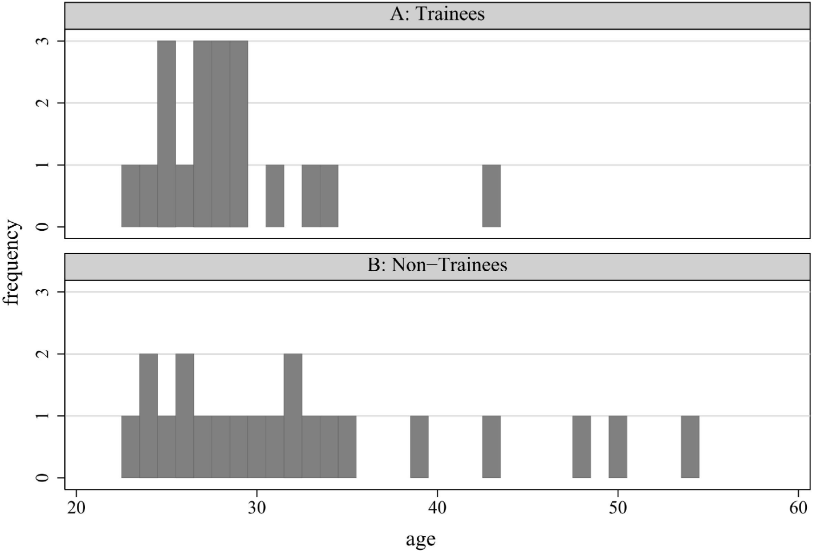
National Taiwan University

Trainees			Non-Trainees			Matched Sample		
unit	age	earnings	unit	age	earnings	unit	age	earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000	14	34	24200
3	29	14400	3	30	21000	17	29	6200
4	25	20800	4	27	9300	15	25	23300
5	29	6100	5	54	41100	17	29	6200
6	23	28600	6	48	29800	20	23	9500
7	33	21900	7	39	42000	10	33	15500
8	27	28800	8	28	8800	4	27	9300
9	31	20300	9	24	25500	12	31	26600
10	26	28100	10	33	15500	11,13	26	8450
11	25	9400	11	26	400	15	25	23300
12	27	14300	12	31	26600	4	27	9300
13	29	12500	13	26	16500	17	29	6200
14	24	19700	14	34	24200	9,16	24	17700
15	25	10100	15	25	23300	15	25	23300
16	43	10700	16	24	9700	1	43	20900
17	28	11500	17	29	6200	8	28	8800
18	27	10700	18	35	30200	4	27	9300
19	28	16300	19	32	17800	8	28	8800
			20	23	9500			
			21	32	25900			
Avg:	28.5	16426	Avg:	33	20724			

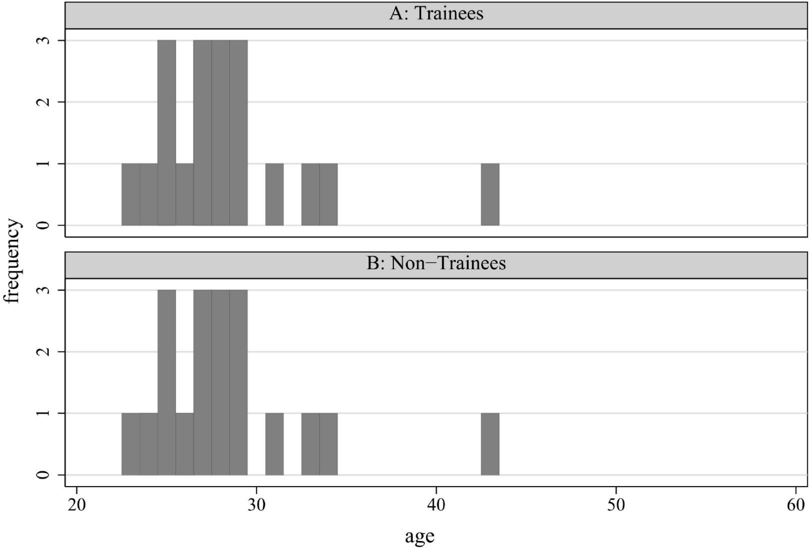
Matching: an Example

Trainees			Non-Trainees			Matched Sample		
unit	age	earnings	unit	age	earnings	unit	age	earnings
1	28	17700	1	43	20900	8	28	8800
2	34	10200	2	50	31000	14	34	24200
3	29	14400	3	30	21000	17	29	6200
4	25	20800	4	27	9300	15	25	23300
5	29	6100	5	54	41100	17	29	6200
6	23	28600	6	48	29800	20	23	9500
7	33	21900	7	39	42000	10	33	15500
8	27	28800	8	28	8800	4	27	9300
9	31	20300	9	24	25500	12	31	26600
10	26	28100	10	33	15500	11,13	26	8450
11	25	9400	11	26	400	15	25	23300
12	27	14300	12	31	26600	4	27	9300
13	29	12500	13	26	16500	17	29	6200
14	24	19700	14	34	24200	9,16	24	17700
15	25	10100	15	25	23300	15	25	23300
16	43	10700	16	24	9700	1	43	20900
17	28	11500	17	29	6200	8	28	8800
18	27	10700	18	35	30200	4	27	9300
19	28	16300	19	32	17800	8	28	8800
			20	23	9500			
			21	32	25900			
Avg:	28.5	16426	Avg:	33	20724	Avg:	28.5	13982

Age Distribution: Before Matching



Age Distribution: After Matching



Treatment Effect Estimates

Difference in average earnings between trainees and non-trainees:

- Before matching:

$$16426 - 20724 = -4298$$

- After matching:

$$16426 - 13982 = 2444$$

Conditional Independence Assumption

国立臺灣大學 National Taiwan University

Assumption (Conditional Independence Assumption)

$$(Y_i^1, Y_i^0) \perp D_i | X_i$$

- This assumption is also called **selection on observable**.
- CIA asserts that conditional on observable characteristics X_i, potential outcomes are independent of treatment assigned.
- Thus, after controlling for value of covariates X_i, both groups should have similar potential outcomes:
  - $E[Y_i^0 | X_i, D_i = 1] = E[Y_i^0 | X_i, D_i = 0]$
  - $E[Y_i^1 | X_i, D_i = 1] = E[Y_i^1 | X_i, D_i = 0]$

Common Support Assumption

National Taiwan University

Assumption (Common Support Assumption)

$$0 < Pr(D_i = 1|X_i) < 1$$

- For each value of covariates X_i, there is a positive probability of being both treated and untreated.

- In other words, it is NOT possible to perfectly predict one's treatment status by using specific value of X_i.
  $\rightarrow$ Exclude: $Pr(D_i = 1|X_i = x) = 1$ or $Pr(D_i = 1|X_i = x) = 0$.

- It ensures that there is sufficient overlap in the characteristics of treated and untreated units to find adequate matched sample.

Identification Results for Matching

国立臺灣大學 National Taiwan University

- Remember CIA ensures $E[Y_i^0|X_i, D_i = 1] = E[Y_i^0|X_i, D_i = 0]$

$$
\begin{aligned}
\alpha_{match}(X) = \quad & \underbrace{E[Y_i|X_i, D_i = 1] - E[Y_i|X_i, D_i = 0]}_{\text{Observed Difference in Average Outcome at given } X_i} \\
& = \underbrace{E[Y_i^1 - Y_i^0|X_i, D_i = 1]}_{\text{Causal Effect (CATT) at } X_i} \\
& + \underbrace{E[Y_i^0|X_i, D_i = 1] - E[Y_i^0|X_i, D_i = 0]}_{\text{Selection Bias}} \\
& = \underbrace{E[Y_i^1 - Y_i^0|X_i, D_i = 1]}_{\text{Causal Effect (CATT) at } X_i} + \underbrace{0}_{\text{Selection Bias}} = \underbrace{E[Y_i^1 - Y_i^0|X_i]}_{\text{Causal Effect (CATE)}}
\end{aligned}
$$

Identification Results for Matching

- Under CIA, matching estimator represent **conditional average treatment effect (CATE)**

- How to obtain ATE?
  $\rightarrow$Take expectation of CATE over all subgroups (all possible X-values).

- Applying LIE, we can identify ATE by averaging CATEs

$$E[\ \underbrace{E[Y_i^1 - Y_i^0 | X_i]}_{\text{Causal Effect (CATE)}}\ ] = \underbrace{E[Y_i^1 - Y_i^0]}_{\text{Causal Effect (ATE)}}$$

Matching Estimation

- Suppose we want to estimate ATT.

$$\widehat{\alpha_{ATT}} = \frac{1}{N_1} \sum_{D_i=1} (Y_i - Y_{j(i)})$$

- $Y_{j(i)}$ is the outcome of an untreated observation j such that $X_{j(i)}$ is the closest value to X_i among the untreated observations.

- We can also estimate ATC and ATE similarly.

$$\widehat{\alpha_{ATC}} = \frac{1}{N_0} \sum_{D_i=0} (Y_{j(i)} - Y_i)$$

$$\widehat{\alpha_{ATE}} = \frac{1}{N} [\sum_{D_i=1} (Y_i - Y_{j(i)}) + \sum_{D_i=0} (Y_{j(i)} - Y_i)]$$

Matching

- We need to define a **distance metric** to measure closeness to construct a matched sample.

- Euclidean Distance: $||X_i - X_j|| = \sqrt{(X_i - X_j)'\widehat{V}^{-1}(X_i - X_j)}$
  $\rightarrow$ Curse of dimensionality!

- Propensity Score Matching:

$$\text{Propensity Score:} p(X_i) = E[D_i|X_i] = Pr(D_i = 1|X_i)$$

$\rightarrow$ Match based on the propensity score!

Propensity Score Matching

- There are two ways to estimate causal effect of treatment using PSM.

  1. Nearest Neighbor:
     By matching each treated observation to the untreated observation with the same or similar values of the propensity score.

  2. Weighting Approach:
     Skip the cumbersome matching procedure and re-weight sample.

- Drawbacks:
  - Selection on observables (CIA) is usually an unconvincing assumption.
    $\rightarrow$ In other words, selection bias might still exist due to **unobservable omitted variables**.
  - The choice of set of covariates is usually arbitrary and may encourage over-fitting or allow rooms for p-hacking.