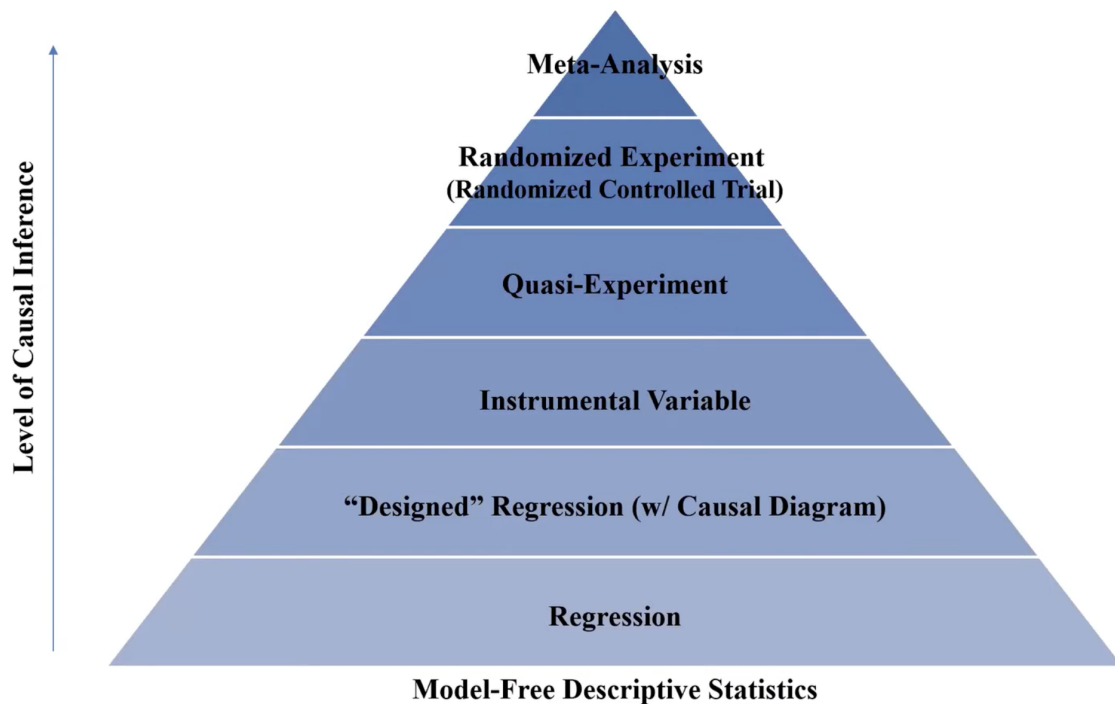


Session 2

Session 2-1 인과추론을 위한 연구 디자인

Causal Hierarchy of Research for Causal Inference



Korea Summer Session on Causal Inference 2021

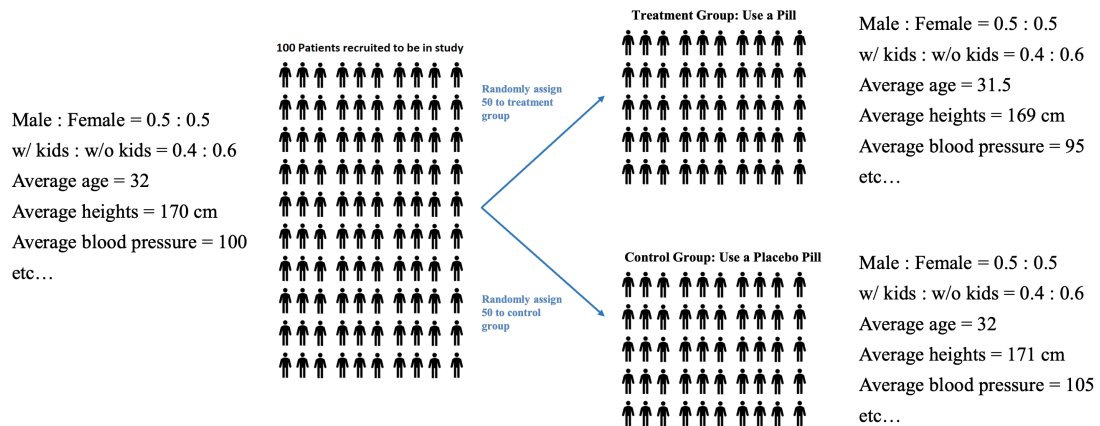
Session 2. Overview of Research Design for Causal Inference

- 피라미드 위로 올라갈수록 인과추론의 수준이 가장 높다.
- 즉, 아래로 갈수록 인과관계의 신뢰도가 낮고 인과관계를 입증하기 어렵다.
- Meta-Analysis : 기존의 여러 인과추론의 결과들을 종합적으로 분석하는 방법론.
- RCT : 단일 인과추론 방법론 중 가장 수준이 높다.
- Quasi-Experiment : 현실상황에서 randomized experiment를 하기는 쉽지 않다. 따라서 실험과 매우 유사한 상황을 찾아서 분석하는 준 실험은 특정 가정하에서는 randomized experiment에 가까울 것.
- Instrumental Variable : 준 실험상황도 어려울 때 인위적 도구를 사용한다. 이는 인과추론을 방해하는 요인(내생성)을 인위적으로 제거하기 위함.
- 'Designed' Regression : 적절한 도구변수도 없을 때 적절한 통제변수의 디자인을 통한 회귀분석도 적절한 인과관계를 추론할 수 있을 것. 이런 상황에서 Causal Diagram이 매우 유용할 것.

Session 2-2 인과추론의 정석: 무작위 통제실험

Random Assignment

- treatment 그룹과 control그룹을 동전던지기로 나눈다고 하면, 랜덤 배정된 샘플 수가 충분하다면, 실험참가자들의 특성이 평균적으로 두 그룹에 균일하게 분포될 것.
- Relying on the law of large numbers, random assignment ensures that the subjects with various characteristics are distributed evenly across the treatment and control groups, so that they are comparable.



- 여기서 중요한 점은 Random Assignment를 할 때 sample size가 너무 적지 않게 주의해야한다.
- 또, 다른 요인에 의해서 random성을 유지할 수 없을 때는 두 그룹이 균등하게 나뉘었는지 확인해야한다.

Randomized Experiments

- Impact of computer use in the classroom on academic performance of college students (Carter et al. 2017)

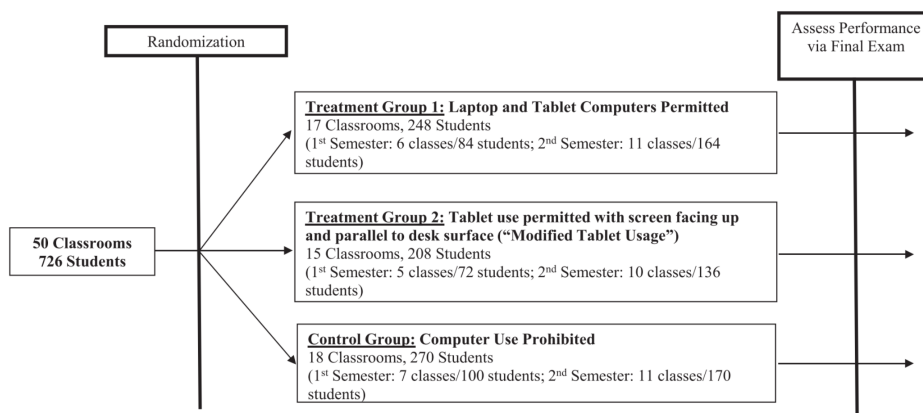


Fig. 1. Experimental design.

- 이 실험에서 태블릿을 제외한 나머지요인들(성별, 인종, 이전성적 등)은 세 그룹에서 큰 차이가 없어야한다.

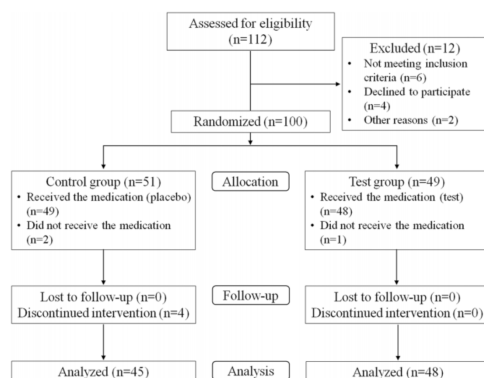
- Impact of computer use in the classroom on academic performance of college students (Carter et al. 2017)

	Control (1)	Treatment 1 (laptops/tablets) (2)	Treatment 2 (tablets, face up) (3)	Both treatments vs. control (4)	Treatment 1 vs. control (5)	Treatment 2 vs. control (6)	
A. Baseline characteristics							Other confounders
Female	0.17	0.20	0.19	0.03 (0.03)	0.06 (0.04)	0.00 (0.04)	
White	0.64	0.67	0.66	0.02 (0.04)	0.02 (0.04)	0.02 (0.05)	
Black	0.11	0.10	0.11	-0.02 (0.03)	-0.02 (0.03)	-0.03 (0.04)	
Hispanic	0.13	0.13	0.09	0.00 (0.03)	0.02 (0.03)	-0.03 (0.03)	
Age	20.12 [1.06]	20.15 [1.00]	20.15 [0.96]	0.03 (0.08)	0.05 (0.09)	0.06 (0.10)	
Prior military service	0.19	0.19	0.16	-0.02 (0.03)	0.00 (0.04)	-0.01 (0.04)	
Division I athlete	0.29	0.40	0.35	0.05 (0.04)	0.07* (0.04)	0.04 (0.05)	
GPA at baseline	2.87 [0.52]	2.82 [0.54]	2.89 [0.51]	-0.01 (0.04)	-0.05 (0.05)	0.03 (0.05)	
Composite ACT	28.78 [3.21]	28.30 [3.46]	28.30 [3.27]	-0.34 (0.26)	-0.37 (0.31)	-0.54 (0.33)	
P-Val (Joint χ^2 Test)				0.610	0.532	0.361	
B. Observed computer (laptop or tablet) use							Main treatment
any computer use	0.00	0.81	0.39	0.62*** (0.02)	0.79*** (0.03)	0.40*** (0.04)	
Average computer use	0.00	0.57	0.22	0.42*** (0.02)	0.56*** (0.02)	0.24*** (0.03)	
Observations	270	248	208	726	518	478	

Carter, S.P., Greenberg, K. and Walker, M.S., 2017. The impact of computer usage on academic performance: Evidence from a randomized trial at the United States Military Academy. *Economics of Education Review*, 56, pp.118-132.

(Wrong) Example of Randomized Experiments

- Effect of Medicine F on gingival inflammation (Hong et al. 2019)



Hong, J.Y., Lee, J.S., Choi, S.H., Shin, H.S., Park, J.C., Shin, S.I. and Chung, J.H., 2019. A randomized, double-blind, placebo-controlled multicenter study for evaluating the effects of fixed-dose combinations of vitamin C, vitamin E, lysozyme, and carbazochrome on gingival inflammation in chronic periodontitis patients. *BMC Oral Health*, 19(1), pp.1-8.

The lower, the healthier gum is

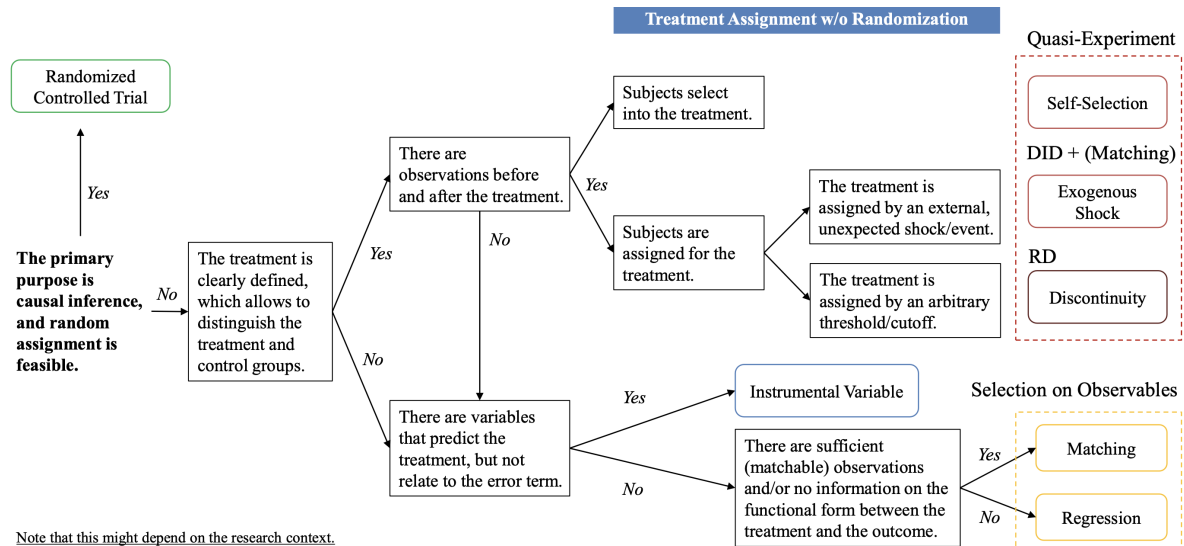
Table 2 Clinical parameters of the control and test groups at baseline, 4 weeks, and 8 weeks (mean \pm SD)

Clinical parameters	Control (N = 45)	Test (N = 48)	p-value
GI			0.042
Baseline	1.00 ± 0.46	1.19 ± 0.51	
4 weeks	1.01 ± 0.46	1.02 ± 0.44 [†]	
8 weeks	0.90 ± 0.50	0.95 ± 0.49 [†]	
Δ baseline – 4 weeks	0.01 ± 0.38	−0.18 ± 0.33 [‡]	
Δ baseline – 8 weeks	−0.10 ± 0.40	−0.24 ± 0.38	
PI			0.138
Baseline	1.50 ± 0.68	1.61 ± 0.67	
4 weeks	1.45 ± 0.68	1.55 ± 0.58	
8 weeks	1.48 ± 0.61	1.42 ± 0.52 [†]	
Δ baseline – 4 weeks	−0.05 ± 0.40	−0.06 ± 0.48	
Δ baseline – 8 weeks	−0.02 ± 0.39	−0.18 ± 0.52	
PD			0.381
Baseline	2.49 ± 0.39	2.63 ± 0.47	
4 weeks	2.47 ± 0.37	2.52 ± 0.49 [†]	
8 weeks	2.39 ± 0.36	2.51 ± 0.51 [†]	
Δ baseline – 4 weeks	−0.02 ± 0.25	−0.11 ± 0.29	
Δ baseline – 8 weeks	−0.10 ± 0.36	−0.11 ± 0.34	

- 몇명의 sample이 필요한지의 기준은 없지만 적어도 random assignment를 한 후 각 그룹이 균일하게 잘 배정되었는지 확인해야한다.
- 하지만 실험 시작 전부터 애초에 Control 그룹과 treatment 그룹의 염증지수 차이가 크다.
- 이 예시는 비교가능하지 않은 대상의 변화를 비교분석하는 것은 인과적 효과를 충분히 추론하지 못할 것이라는 것을 보여준다.

Session 2-3 실험 아닌, 실험 같은 준실험

What's Your Research Design?



- Randomized Experiment와 Quasi-Experiment의 유일한 차이점은 treatment에 대한 assignment방법.
- **Self-Selection** : 연구 대상들이 스스로 treatment를 받을지 결정하는 방법.
- **Exogenous Shock** : 외부요인에 의해서 treatment와 control 그룹이 나뉘는 방법. 보통 자연실험이라고 부른다.
- **Discontinuity** : 임의의 경계값을 기준으로 treatment와 control 그룹을 나누는 방법.
- 많은 경우에는 treatment와 control 그룹을 명확히 나누기 쉽지않다. 또한 나눌 수 있다 하더라도 전, 후의 데이터를 관찰할 수 없다면 위에서 언급한 방법을 활용하기 힘들다.
- 이럴 경우, treatment를 예측할 수 있고 동시에 오차항(결과에 영향을 줄 수 있지만 관찰할 수 없는 모든 변수)과 연관성이 없는 변수를 찾을 수 있는지의 여부를 판단해야 한다. 만약 찾을 수 있다면 Instrumental Variable을 사용할 수 있을 것.
- Instrumental Variable를 활용하기 어려운 경우, 관찰가능한 변수들만 가지고 선택편향을 통제하자는 접근이 가능. 이러한 방법으로는 Matching과 Regression이 있다.
- 가지고 있는 변수와 결과변수에 대한 관계가 있다고 가정의 접근이 가능하다면 Regression을 사용. 하지만 가정의 접근이 불가능하다면 Matching방법론을 사용할 수 있다. Matching은 가지고 있는 변수를 균일한 양 집단으로 나누는 것.

Random Assignment is not Always Feasible

- Randomized experiment와 Quasi-experiment 모두 비교 가능한 control 그룹을 구성하는 것이 목표이다.
- 하지만, Quasi-experiment는 counterfactual에 대한 명시적으로 증명을 해야할 책임이 있다.

Examples of Exogenous Shock for Quasi-Experiments

- Impact of the access to local finance on firm formation
 - 이런 상황은 실험이 불가능하다. 그래서 이 연구에서는 실험 대신 Shale boom(Natural Shock)을 통해 treatment 그룹을 정의했다.
 - shale boom에 의해 정의된 treatment 그룹과 control 그룹이 treatment인 local finance의 접근성을 제외하고 다른 요인은 얼마나 유사한지 확인해야한다.
 - 이런 경우 Exogenous Shock은 아무도 예상하지 못하는 경우이기 때문에 어느정도 random하게 배정한 것과 유사하다고 인정할 수 있다.

Examples of Self-Selection for Quasi-Experiments

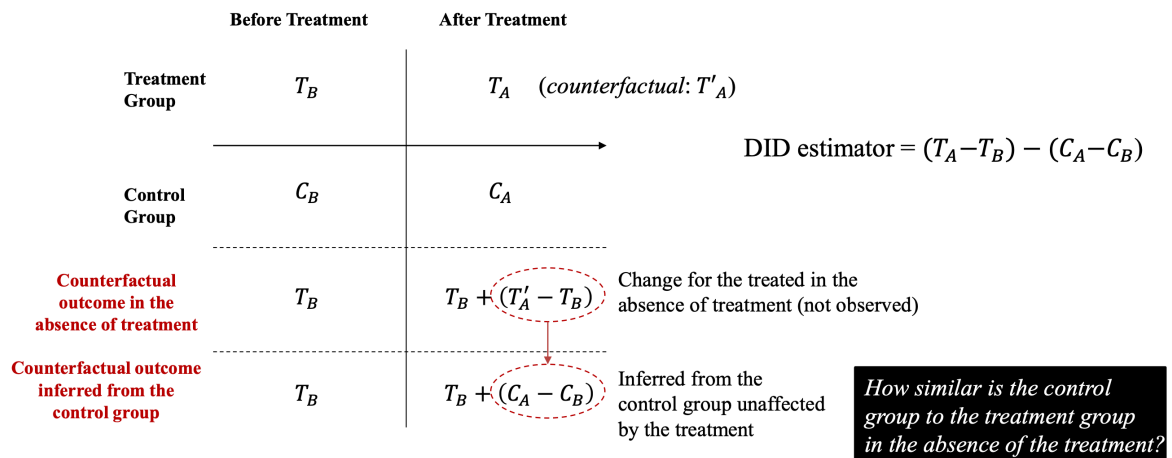
- Effect of customers' social media participation on their visit frequency
 - 이 연구의 treatment는 기업 소셜미디어 참여
 - 그렇기 때문에 treatment와 control그룹을 쉽게 나눌 수 있다.
 - 하지만, 소비자들이 어떤 의도를 가지고 참여를 했는지 알수 없기 때문에 두 집단이 비교 가능한지 의구심이 들 수밖에 없다.
- Example : Effect of electronic word of mouth(sns상에서 친구의 입소문이 어떤 역할을 하는지)
 - treatment : 팔로워 중 해당 기업을 팔로우한 그룹, control : 팔로워 중 해당 기업을 팔로우하지 않은 그룹
 - 이런 경우는 위의 실험보다는 좀 더 안전한 경우. 왜냐하면, 두 그룹 모두 나의 팔로워이기 때문에 어느정도 특성이 유사할 것이라고 주장하기 용이하기 때문.

Examples of Discontinuity for Quasi-Experiments

- Drinking and public health/death
 - 음주 가능 연령은 임의로 정한 숫자. 하지만 이 숫자를 기준으로 사망률의 차이가 생긴다면 음주가 미치는 인과적 효과라고 인정할 수 있다.

Session 2-4 준실험 분석도구: 이중차분법 & 회귀불연속

Difference-in-Difference (DID)

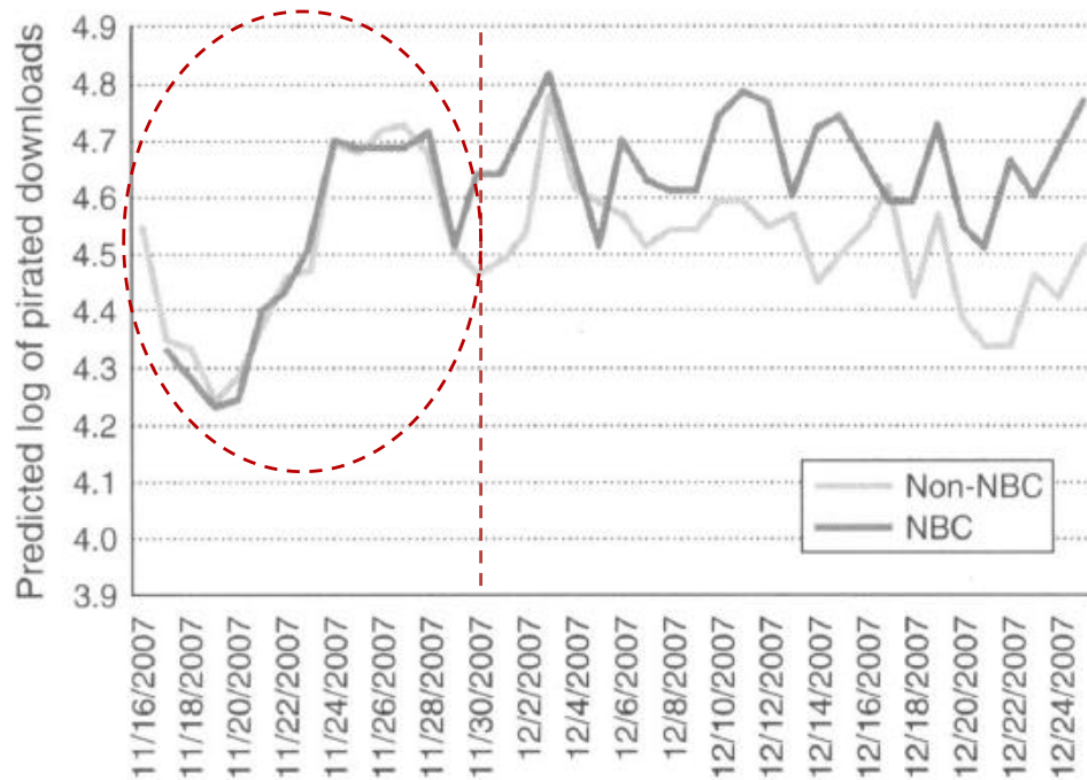


- counterfactual은 treatment가 없었다면 있었을 잠재적 결과. 그럼 T'_A 는 $T_B + (T'_A - T_B)$ 로 표현가능하고 이는 시간에 따라 변하지 않는 값과 시간에 따라 변하는 값을 구분할 수 있다.
- 하지만 $(T'_A - T_B)$ 는 관찰할 수 없으니 counterfactual과 가까운 control그룹을 찾아서 treatment를 받지 않는 control 그룹의 시간에 따른 변화량과 비슷하다고 가정할 수 있다면 대체하여 counterfactual을 추정할 수 있지 않을까? ⇒ DID의 핵심 아이디어.
- 즉, $T_B + (T'_A - T_B) \Rightarrow T_B + (C_A - C_B)$,
- DID estimator = $(T_A - T_B) - (C_A - C_B) = T_A - [T_B + (C_A - C_B)]$
- $[T_B + (C_A - C_B)]$: Inferred Counterfactual
- 따라서 control그룹이 treatment가 없을때의 treatment 그룹과 얼마나 비교가능한지 여부가 DID분석의 신빙성을 결정하게 될 것.
- 그럼 왜 Before Period가 필요할까?
 - 목표는 $T_A - T'_A$ 이고, 이는 $T_A - C_A$ 로 대체할 수도 있을 것. 하지만 DID에서는 T'_A 전체를 control그룹으로 치환하는 것이 아니라 시간에 따른 변화량만 control그룹으로 치환한다. 즉, DID는 시간에 따라 변화하는 부분만 비슷하면 된다는 이점이 있다.
 - 또한 시간에 따라 변화하는 양은 데이터를 통해 어느정도 검증을 할 수 있다.

Identification Assumption for DID

- parallel trends assumption : 시간에 따른 변화량이 같아야한다.

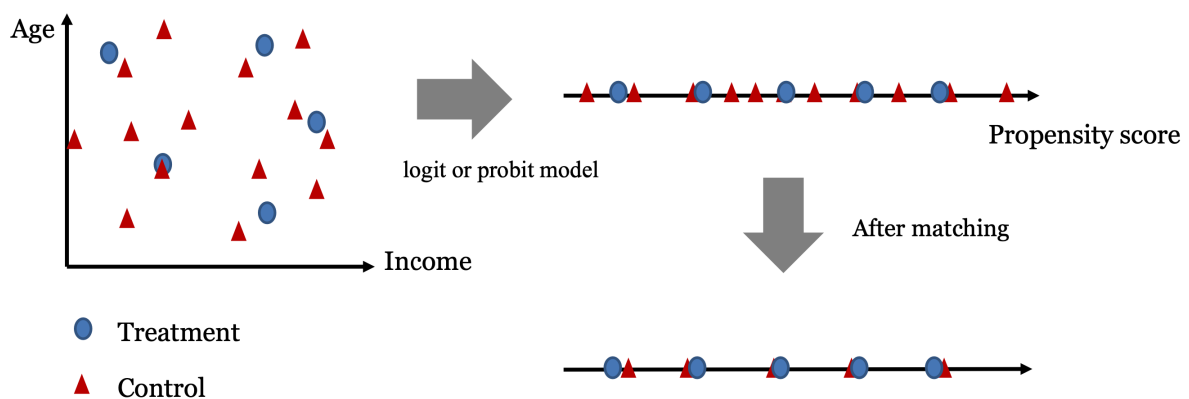
Figure 1 NBC vs. Non-NBC Piracy Surrounding December 1, 2007



Matching Techniques

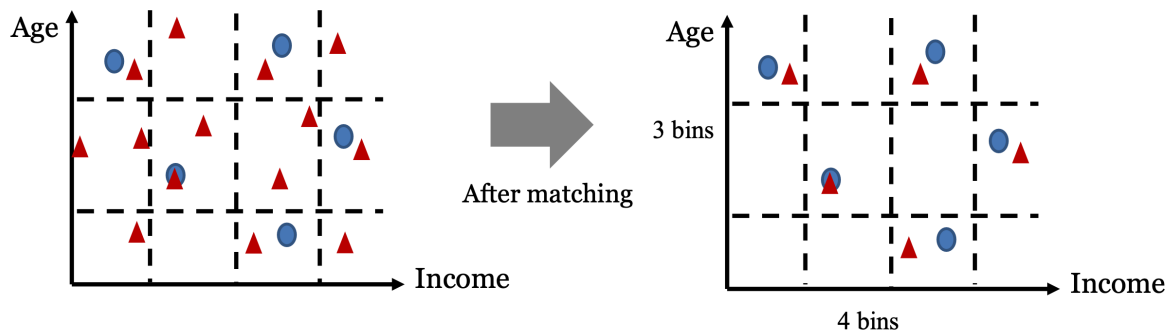
- 비교가능한 control 그룹을 찾을 수 없을 때 사용할 수 있는 방법.
- 이는 우리가 가질 수 있는 변수들 중 평균적으로 유사한 샘플들만 서로 matching하는 방법.
- Matching에는 PSM, CEM이라는 두가지 방법이 있다.

Propensity Score Matching(PSM)



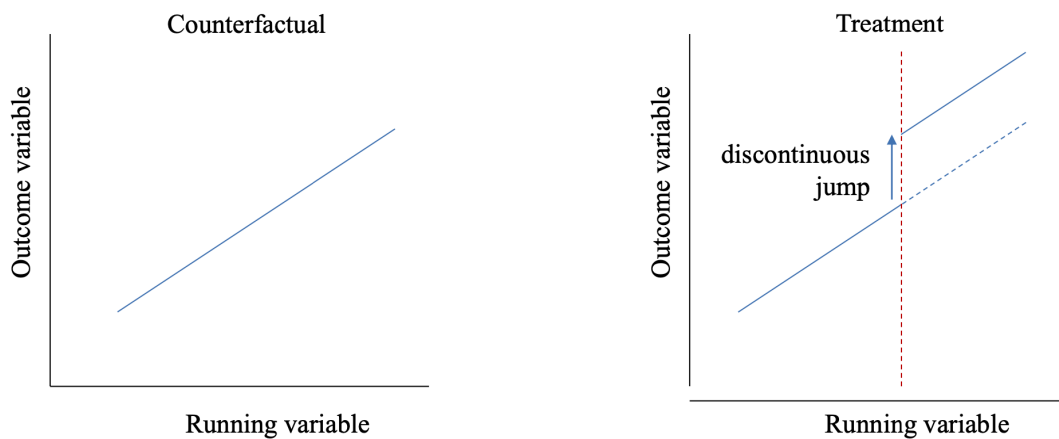
- propensity score를 계산하고 이 score가 유사한 샘플들끼리 서로 matching을 하자.
- propensity score : treatment 그룹에 속할 확률
- 모든 샘플을 score 하나로만 matching을 하기때문에 만약에 변수가 굉장히 많다면 경우에 따라 어떤 변수들은 차이가 많이 날 수 있다. 이를 해결하기 위한 방법이 CEM.

Coarsened Exact Matching(CEM)



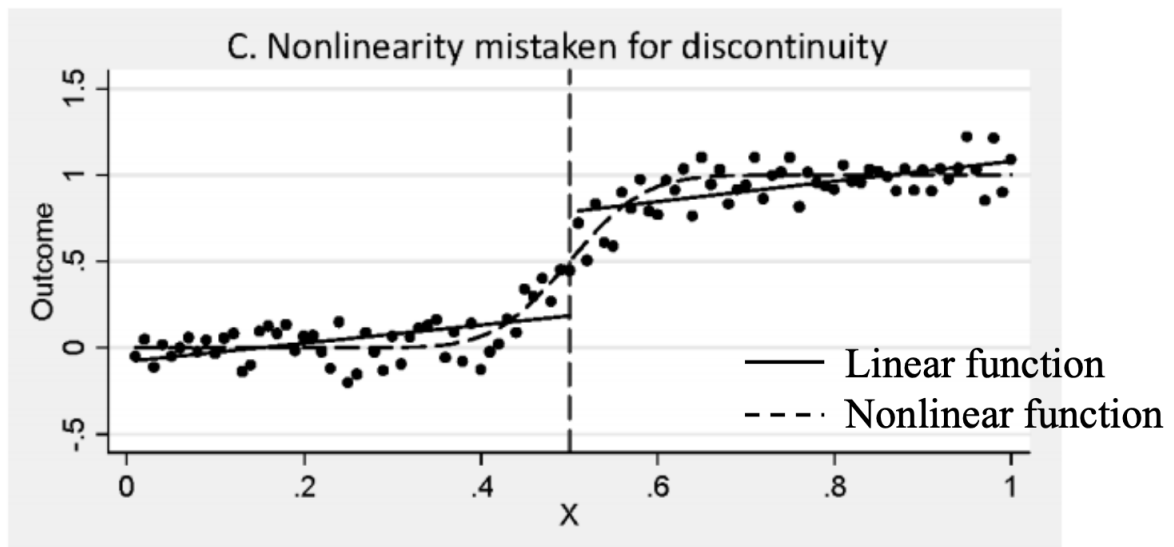
- 각 변수별로 구간을 나누고 구간에 따라 일치하는 샘플끼리 매칭하자.

Regression Discontinuity (RD)



- RD는 running variable의 modeling이 핵심.
- 즉, discontinuous jump가 있을 때와 없을 때의 차이를 바탕으로 treatment effect를 구하고자 함.

Identification Assumption for RD



- 모델링을 선형으로 할때와 비선형으로 할 때, jump가 일어날수도 안일어날수도 있다. 즉, 모델링에 굉장히 민감하다는 한계점이 있다.