

인과추론과 실무 : 인과추론 Overview

가짜연구소 인과추론팀

발표자 : LG CNS 김성수

1. 인과추론과 예측 모델

비즈니스에서 인과추론의 목적



고객의 넷플릭스 구독 갱신을 예측하는 모델을 만들어주세요.

의사결정권자



DB에서 다음 데이터를 수집했습니다! 해당 변수들이 구독 갱신을 예측하는 데 도움이 될 것으로 생각됩니다.

데이터 과학자

$X = \{\text{할인률}, \text{앱 업데이트 이력}, \text{앱 월별 사용량}, \text{광고}, \text{고객이 신고한 버그}, \text{상호 작용(views, Clicks)}, \text{전화 판매}, \text{거시경제 활동}\}$

작년 데이터를 사용하여, SOTA 머신러닝 모델에 적합시켰습니다. 이 모델은 X 변수를 통해 $y = \{\text{갱신}\}$ 을 예측합니다.

```
X, y = user_retention_dataset()  
model = fit_xgboost(X, y)
```

비즈니스에서 인과추론의 목적



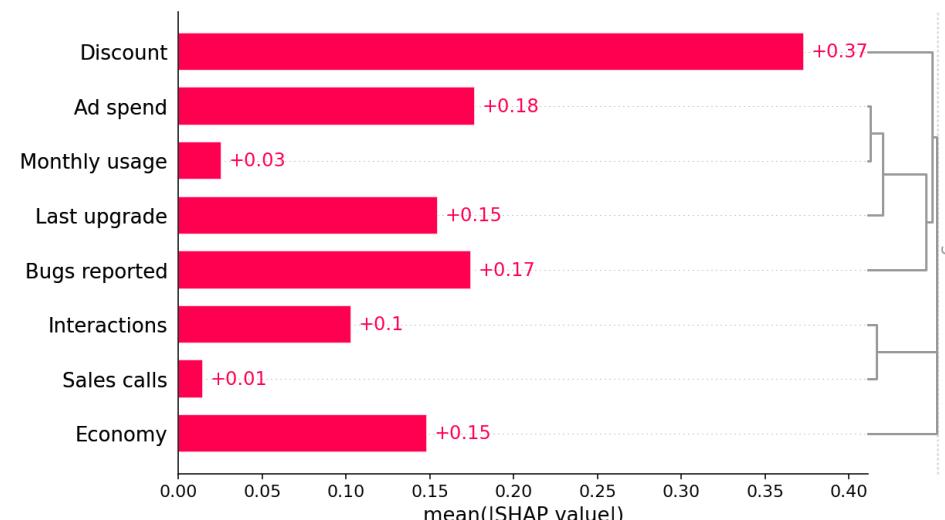
새로운 데이터에 대해서 얼마나 정확하게 예측하나요?

의사결정권자



테스트 데이터에 대해서 약 99%의 정확도로 예측을 제공합니다.
또한 머신러닝 모델의 Feature Importance를 활용해 고객 이탈을 방지하기 위한 예방 정책을 세울 수 있겠네요.
Feature Importance를 보니 "할인과 광고 비용이 중요하다는 것 같습니다. 그리고 버그 리포트도요"

데이터 과학자



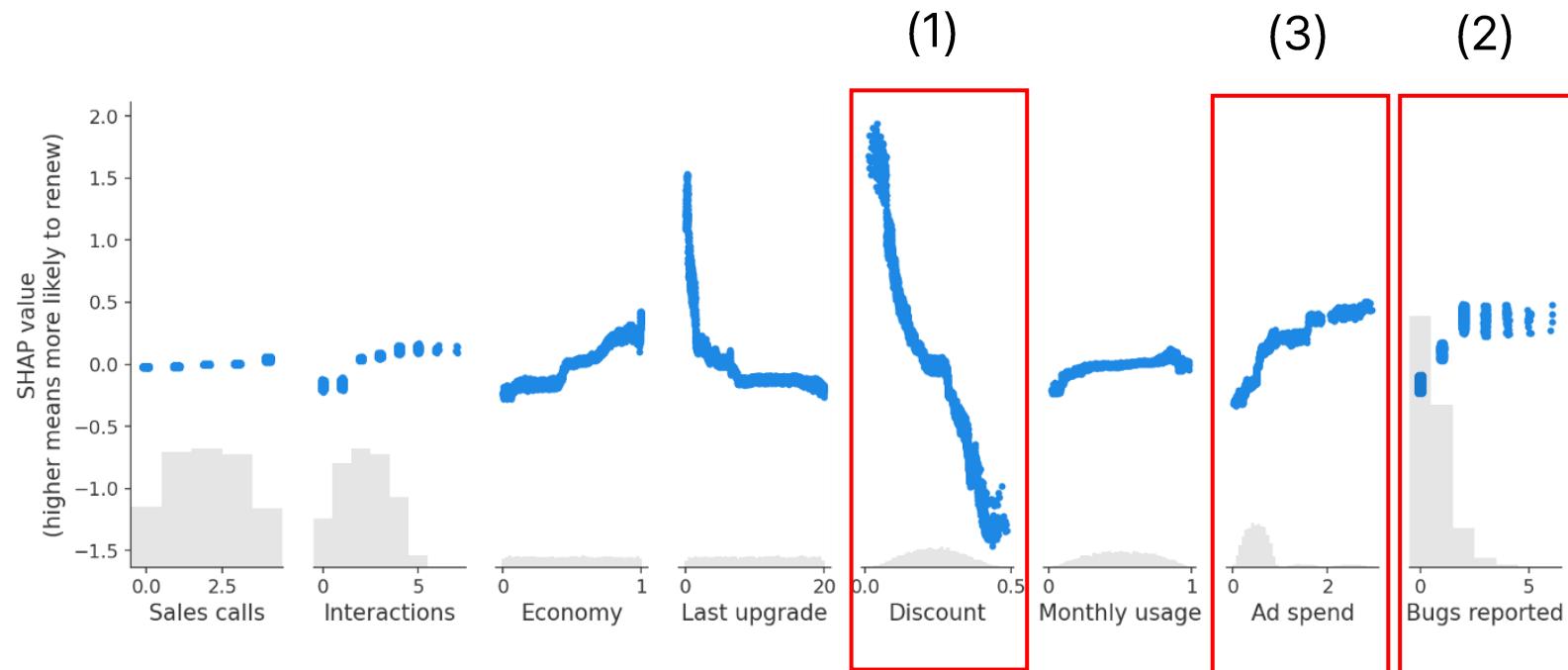
비즈니스에서 인과추론의 목적



데이터 과학자

각각 변수가 결과에 어떤 영향을 미치는지 더욱 자세히 살펴보겠습니다.

- (1) 더 큰 할인은 구독 갱신을 줄이는군요.
- (2) 더 많은 버그 리포트 보고는 구독 갱신으로 이어지는 것 같아요.
- (3) 광고도 구독 갱신에 매우 중요하네요!



비즈니스에서 인과추론의 목적



가격을 올리고 더 많은 버그를 서비스에 추가하고 모두에게 광고를 보내는 의사결정을 하면 고객 이탈을 방지할 수 있겠네요.

의사결정권자

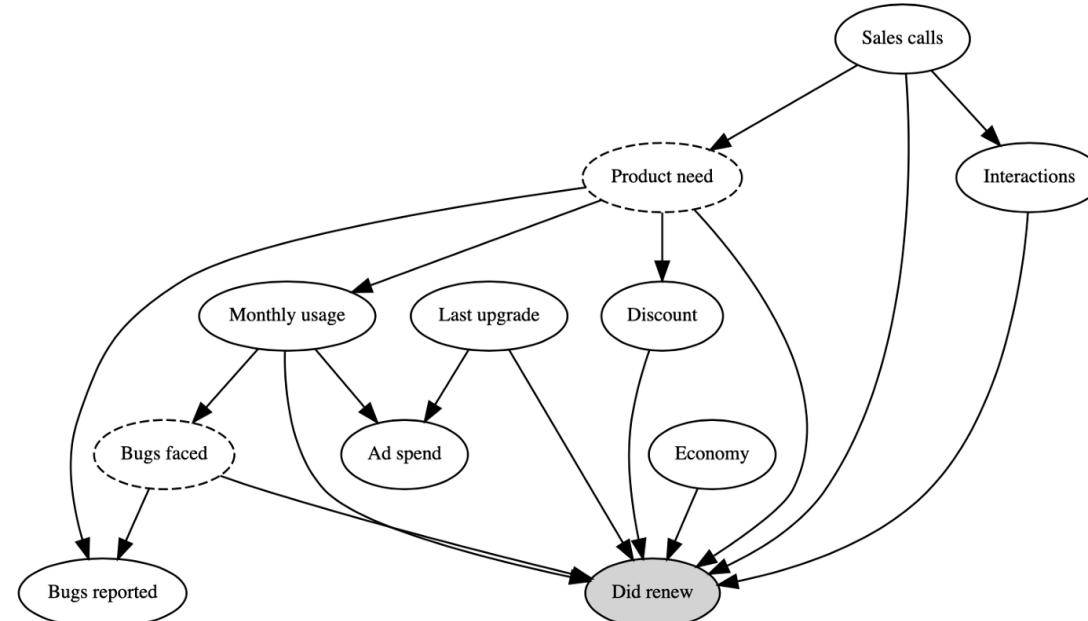
비즈니스에서 인과추론의 목적



해당 분석 결과를 받은
도메인 전문가 혹은 협업

(데이터 분석은 사기다. 내가 다시 믿나 봐라)

- (1) 제품을 오래 사용하는 구독자는 시스템 상의 버그를 발견하고, 구독을 갱신할 가능성이 당연히 더 높음
- (2) 제품 구독에 관심이 없는 사람에게 더 높은 할인율을 부여하니, 이러한 고객들의 이탈율은 당연히 더 높음

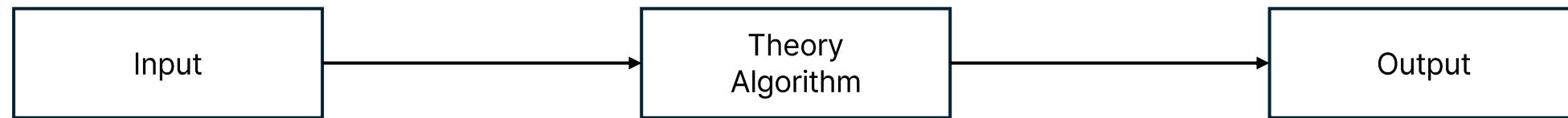


왜 이러한 차이가 나는 걸까요?

이러한 차이는 우리가 데이터를 통해 풀고자 하는 비즈니스 문제와 목표가 다르기 때문에 나타남

왜 이러한 차이가 나는 걸까요?

이러한 차이는 우리가 데이터를 통해 풀고자 하는 비즈니스 문제와 목표가 다르기 때문에 나타남



Intervening and
Manipulating inputs



Predicting
Outputs

특정 변수에 개입하여 Y 에 대한
변화를 관측하고 싶다면? Then

-> **인과추론 Tool을 선택**

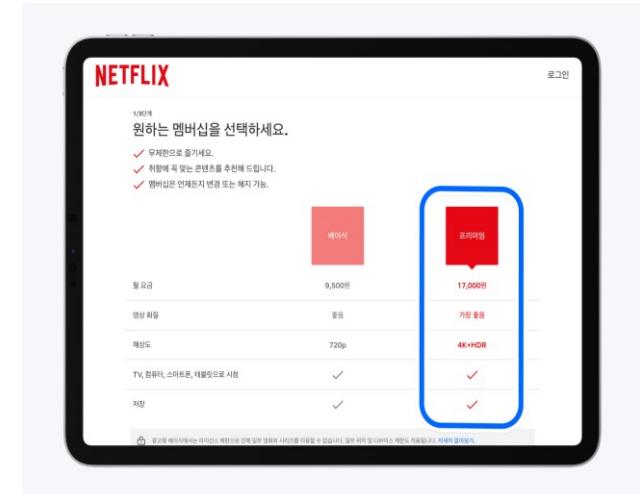
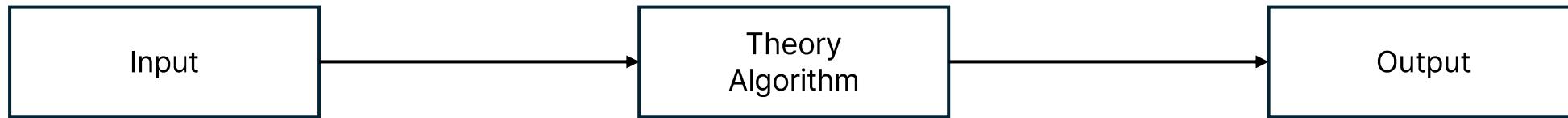
변수를 활용해 정확하게 Y 를
예측하고 싶다면? Then

-> **예측 모델링 Tool을 선택**

왜 이러한 차이가 나는 걸까요?

다시 사례로 돌아와서, 사례의 목표는 무엇이었죠?

- 요구사항: 고객의 다음 넷플릭스 구독 갱신을 예측하는 모델을 만들어주세요.



X=

{할인률, 앱 업데이트 이력, 앱 월별 사용량, 광고, 고객이 신고한 버그, 상호 작용(views, Clicks), 전화 판매, 거시경제 활동}

사용자가 구독을 갱신할까?

왜 이러한 차이가 나는 걸까요?

비즈니스 현장에서도 인과추론을 예측 모델과 다른 목적과 목표를 가진 방법론으로 생각해야 함

왜 이러한 차이가 나는 걸까요?

비즈니스 현장에서도 인과추론을 예측 모델과 다른 목적과 목표를 가진 방법론으로 생각해야 함

- 데이터를 활용해 해결하고자 하는 목표와 활용 방법론이 다름
- 스타틴(Statin) - X을 줄이면(Manipulating) 평균적으로 여성의 뇌졸중 위험(Y – Change?)을 줄일 수 있나요?

Table Examples of Tasks Conducted by Data Scientists Working with Electronic Health Records

	Description	Data Science Task Prediction	Causal inference
Example of scientific question	How can women aged 60–80 years with stroke history be partitioned in classes defined by their characteristics?	What is the probability of having a stroke next year for women with certain characteristics?	Will starting a statin reduce, on average, the risk of stroke in women with certain characteristics?
Data	<ul style="list-style-type: none"> Eligibility criteria Features (symptoms, clinical parameters ...) 	<ul style="list-style-type: none"> Eligibility criteria Output (diagnosis of stroke over the next year) Inputs (age, blood pressure, history of stroke, diabetes at baseline) 	<ul style="list-style-type: none"> Eligibility criteria Outcome (diagnosis of stroke over the next year) Treatment (initiation of statins at baseline) Confounders Effect modifiers (optional)
Examples of analytics	Cluster analysis ...	Regression Decision trees Random forests Support vector machines Neural networks ...	Regression Matching Inverse probability weighting G-formula G-estimation Instrumental variable estimation ...

먼저 해야 할 것

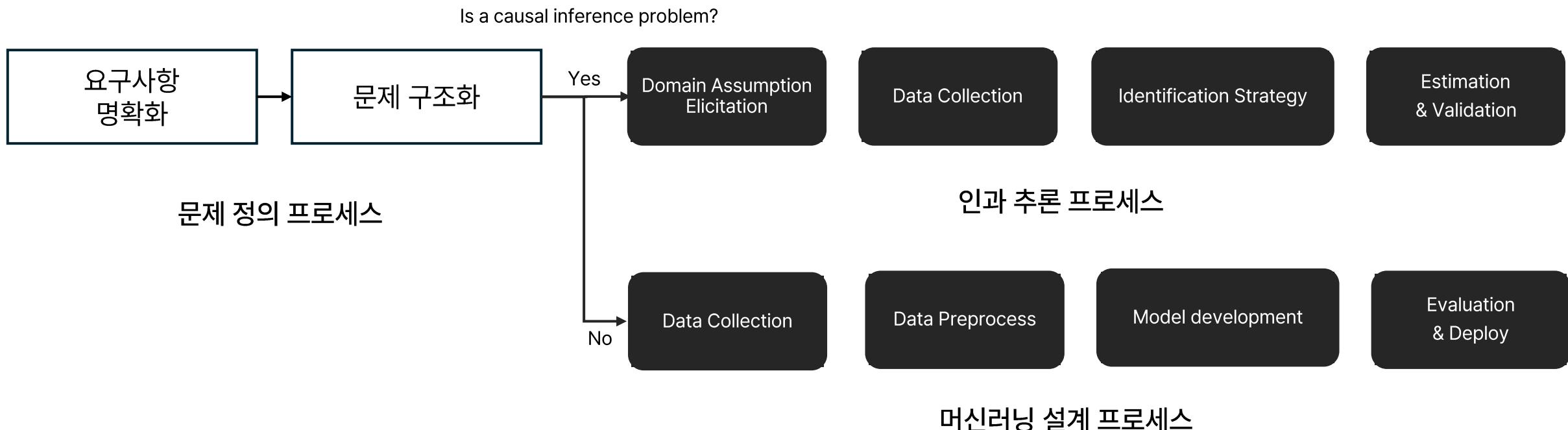
문제 정의: 우리는 데이터를 통해서 어떠한 문제를 풀고자 하는가 or 궁금해 하는가?

- X 변수를 통해 Y 예측을 잘 했으면 좋겠어요 vs (다른 변수는 고정하고) 특정 X 변수 변화에 따른 Y의 변화를 확인하고 싶어요

먼저 해야 할 것

문제 정의: 우리는 데이터를 통해서 어떠한 문제를 풀고자 하는가 or 궁금해 하는가?

- X 변수를 통해 Y 예측을 잘 했으면 좋겠어요 vs (다른 변수는 고정하고) 특정 X 변수 변화에 따른 Y의 변화를 확인하고 싶어요
- 실무에서 주어진 Task가 Causal Problem인가 Predictive Problem인가를 명확히 구분해야 하며, 이에 맞는 방법론과 분석 프로세스를 활용해야 함



2. 인과추론 소개

Causality: Potential Outcome Framework Perspective

2주차부터 인과추론 ToolBox를 열기 전에, 교재 Chapter 1을 활용해 인과추론 소개

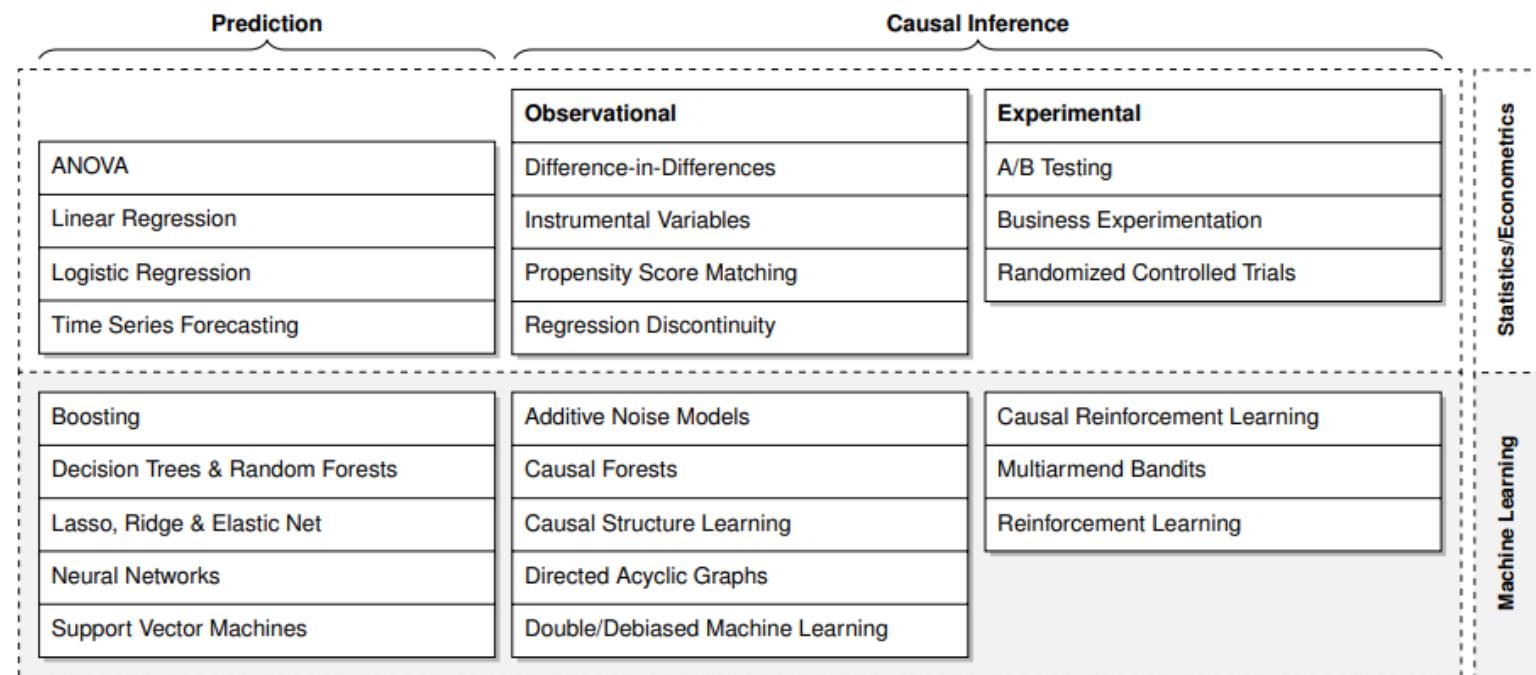
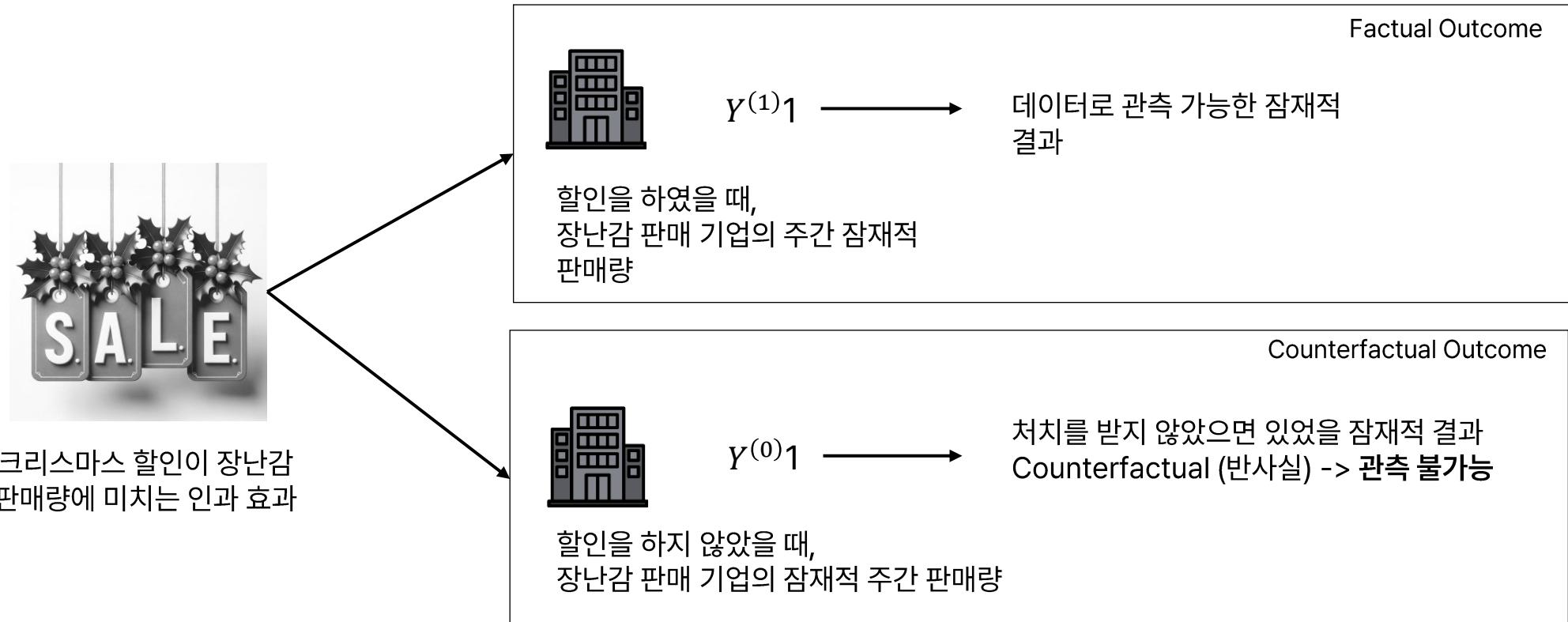


Figure 1: Examples of popular data analysis algorithms in statistics and econometrics, as well as machine learning and artificial intelligence, classified according to prediction and causal inference methods. Causal inference methods are further differentiated according to observational (based on ex-post observed data) and experimental approaches.

Causality: Potential Outcome Framework Perspective

- 인과 문제: 어린이 장난감을 판매하는 기업의 12월 크리스마스 상품 가격할인이 주간 판매량에 미치는 영향 (Unit of analysis : 기업)
- 인과추론의 근본적인 문제 : *Causal Effect* : $Y^{(1)} - Y^{(0)}$



Causality: Potential Outcome Framework Perspective

- 인과 문제: 어린이 장난감을 판매하는 기업의 12월 크리스마스 상품 가격할인이 주간 판매량에 미치는 영향 (Unit of analysis : 기업)
- 인과추론의 근본적인 문제 : *Causal Effect* : $Y^{(1)} - Y^{(0)}$

우리가 신(?)이라서 Counterfactual Outcome을 볼 수 있다고 해보자

	i	y0	y1	t	x	y
0	1	200	220	0	0	200
1	2	120	140	0	0	120
2	3	300	400	0	1	300
3	4	450	500	1	0	500
4	5	600	600	1	0	600
5	6	600	800	1	1	800

(i = 장난감 기업)

Causality: Potential Outcome Framework Perspective

- 인과 문제: 어린이 장난감을 판매하는 기업의 12월 크리스마스 상품 가격할인이 주간 판매량에 미치는 영향 (Unit of analysis : 기업)
- ATE : 처치 효과의 평균, ATT: 처치받은 대상에 대한 처치 효과, CATE: 변수 X로 정의된 그룹에서의 처치효과**
- 크리스마스 주간(22.5증가)에 가격을 할인했을 때보다, 크리스마스 1주일 전에 할인했을 때(150 증가) 가 훨씬 많은 이득

$$ATE = (20 + 20 + 100 + 50 + 0 + 200) / 6 = 65$$

$$ATT = (50 + 0 + 200) / 3 = 83.33$$

$$CATE(x=1) = (100 + 200) / 2 = 150$$

$$CATE(x=0) = (20 + 20 + 50 + 0) / 4 = 22.5$$

우리가 신(?)이라서 Counterfactual Outcome을 볼 수 있다고 해보자

	i	y0	y1	t	x	y	te
0	1	200	220	0	0	200	20
1	2	120	140	0	0	120	20
2	3	300	400	0	1	300	100
3	4	450	500	1	0	500	50
4	5	600	600	1	0	600	0
5	6	600	800	1	1	800	200

X = 1 -> 크리스마스 1주일 전

Causality: Potential Outcome Framework Perspective

- 연관관계를 상관관계로?



(연관관계를 인과관계로 생각하는 분석가)

크리스마스 할인의 효과는?

$$\text{ATE} = (500 + 600 + 800) / 3 - (200 + 120 + 300) / 3 = 426.67 \text{ (weekly sales)}$$

실제 계산한 값(65)과 약 300이상 차이,
처치 효과를 과대 추정

	i	y0	y1	t	x	y	te
0	1	200	NaN	0	0	200	NaN
1	2	120	NaN	0	0	120	NaN
2	3	300	NaN	0	1	300	NaN
3	4	NaN	500	1	0	500	NaN
4	5	NaN	600	1	0	600	NaN
5	6	NaN	800	1	1	800	NaN

(i = 장난감 기업)

Causality: Potential Outcome Framework Perspective

- 연관관계를 상관관계로?

우리는 신이 아니라서, 빨간색 테두리 Missing data를 채우지 못 함

그럼 어떻게 인과를 추론할 수 있을까?

⇒ 이상적인 Counterfactual과 가장 가까운 집단을 찾아서 비교
(Exchangeability)

Causal effect:

Potential Outcome with treatment – Potential outcome without treatment

-> Consistency

Observed outcome with treatment – **Potential outcome without treatment**

	i	y0	y1	t	x	y	te
0	1	200	NaN	0	0	200	NaN
1	2	120	NaN	0	0	120	NaN
2	3	300	NaN	0	1	300	NaN
3	4	NaN	500	1	0	500	NaN
4	5	NaN	600	1	0	600	NaN
5	6	NaN	800	1	1	800	NaN

(i = 장난감 기업)

Causality: Potential Outcome Framework Perspective

- 두 방식으로 구한 효과는 왜 다를 수 밖에 없는가? – 편향
- 426.67 vs 65의 차이는 왜 발생하는가?

$$\text{Association } (E[Y | T = 1] - E[Y | T = 0]) = E[Y_1 | T = 1] - E[Y_0 | T = 0]$$

$$E[Y | T = 1] - E[Y | T = 0] = (E[Y_1 | T = 1] - E[Y_0 | T = 0]) + (E[Y_0 | T = 1] - E[Y_0 | T = 1])$$

$$E[Y | T = 1] - E[Y | T = 0] = (E[Y_1 - Y_0 | T = 1]) + (E[Y_0 | T = 1] - E[Y_0 | T = 0])$$

ATT BIAS(처치군과 대조군 사이에 존재하는 Y_0 의 차이)

$$BIAS \quad E[Y_0 | T = 1] > E[Y_0 | T = 0]$$

- ✓ 할인할 여력이 있는 기업이 할인 여부와 관계없이 더 많이 판매하는 경향이 있을 수 있음
- ✓ 처치 전에는 둘 집단간의 차이가 없어야 한다는 것 즉, 할인한 회사와 하지 않은 회사가 처치와 관계없이 비슷해야함
- ✓ 이는 인과추론을 위해서는 비교가능한 통제군을 구성해야 한다는 의미
- ✓ 처치받는 그룹과 받지 않은 그룹이 처치에 유사하게 반응하면($E[Y_1 - Y_0 | T = 1] = E[Y_1 - Y_0 | T = 0]$) 평균의 차이가 ATE

Causality: Potential Outcome Framework Perspective

- 관측 가능한 데이터에서 인과 추정량을 찾아내는 방법: 식별 과정 -> 편향을 없애는 과정 ($E[Y_0 | T = 1] = E[Y_0 | T = 0]$)
- 독립성 가정 : 처치 T가 잠재적 결과와 독립인가? 처치 T가 기업의 특징(기업 규모)에 따라서 발생하지 않았나?
- 모든 회사가 할인을 하지 않았더라면, 할인하기로 한 회사와 그렇지 않은 회사를 구분할 수 없다는 의미
- 처치 T를 랜덤하게 배정한다면, 처치 T가 잠재적 결과와 독립

두 그룹의 기업(1,2,3 그리고 4,5,6)은 비교 가능 한가 ? 서로 비슷한가?

	i	y0	y1	t	x	y	te
0	1	200	NaN	0	0	200	NaN
1	2	120	NaN	0	0	120	NaN
2	3	300	NaN	0	1	300	NaN
3	4	NaN	500	1	0	500	NaN
4	5	NaN	600	1	0	600	NaN
5	6	NaN	800	1	1	800	NaN

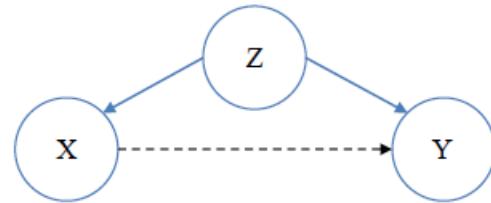
(i = 장난감 기업)

Causality: Potential Outcome Framework Perspective

- 두 처치 효과는 왜 다를 수 밖에 없는가? – 인과 그래프 관점

Omitted variable bias (Selection bias) -> Endogeneity

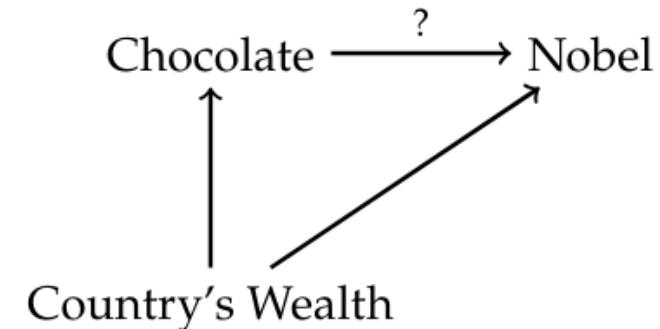
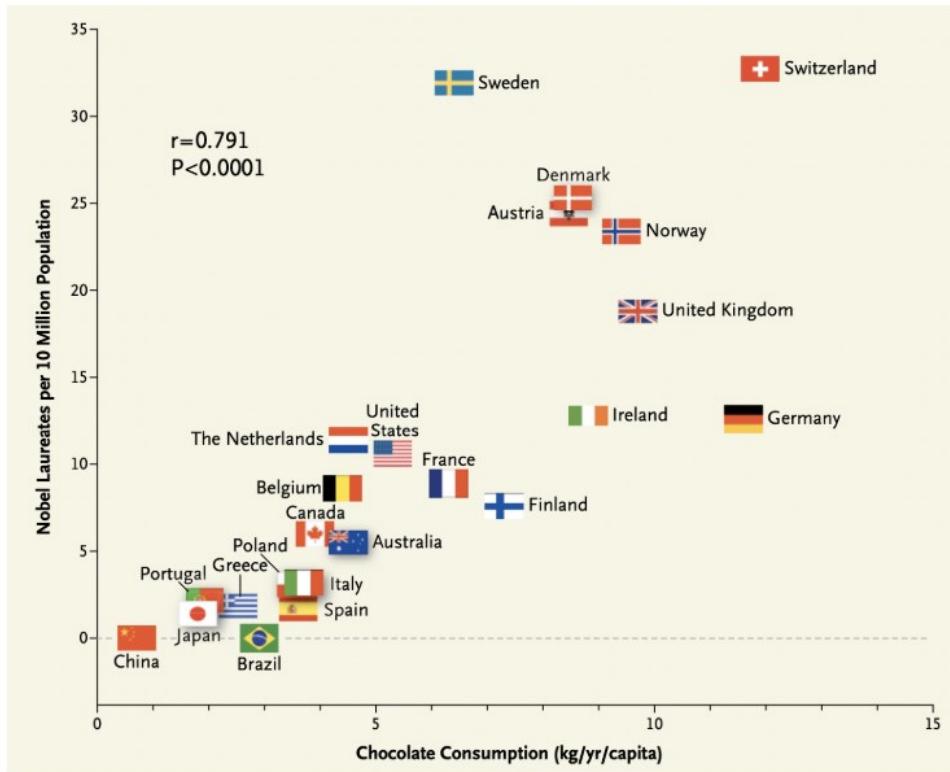
- Unobserved or omitted variable is confounding both independent and dependent variables



Selection Bias: There might be systematic differences between treated and untreated sample

Causality: Potential Outcome Framework Perspective

- 두 처치 효과는 왜 다를 수 밖에 없는가? – 인과 그래프 관점



Causality: Potential Outcome Framework Perspective

- 두 처치 효과는 왜 다를 수 밖에 없는가? – 회귀 관점의 이해

Potential Outcomes (counterfactual regression function)

$$\begin{aligned} Y_{0i} &= \alpha + \varepsilon_{0i} \\ Y_{1i} &= \alpha + \boxed{\beta} + \varepsilon_{1i} \end{aligned}$$

개인 i의 처치 효과
X=0 이면 오른쪽 항이 사라지고 X=1이면 왼쪽 항이 사라짐

Regression Model

$$\begin{aligned} Y_i &= (1 - X_i)Y_{0i} + X_iY_{1i} \\ &= \alpha + \beta X_i + [\varepsilon_{0i} + (\varepsilon_{1i} - \varepsilon_{0i})X_i] \\ &= \alpha + [\beta + (\varepsilon_{1i} - \varepsilon_{0i})]X_i + \varepsilon_{0i} \end{aligned}$$

Selection bias

Error term
 ε_{1i} if $X_i = 1$
 ε_{0i} if $X_i = 0$

Estimated Treatment Effect

$$E[Y_{1i}] - E[Y_{0i}] = \beta + E[\varepsilon_{1i}] - E[\varepsilon_{0i}] \rightarrow E[\varepsilon_i | X_i=1] = E[\varepsilon_i | X_i=0]$$

가 언제 가능한가?
 ->처치(X)와 error term(ε)이 독립인 경우

In the presence of **selection bias**,
 the independent variable (X_i) is correlated
 to the error term ($[\varepsilon_{0i} + (\varepsilon_{1i} - \varepsilon_{0i})X_i]$),
 which is called **endogeneity in regression**.

Causality: Potential Outcome Framework Perspective

- 두 처치 효과는 왜 다를 수 밖에 없는가? – 사례
- 다각화 결정이 기업의 재무성과에 미치는 영향? – 모델에 따라 Premium 또는 Discount
- 편향(Bias)은 인과 추정치를 반대로 만들기도 하여, 잘못된 의사결정의 상황으로 이어짐

	Sales Multipliers				
	OLS (BO)	OLS	Fixed Effects	IV	Self- selection
Constant	-0.36 (33)	-0.75 (26)	-0.09 (1.67)	-0.72 (30)	-0.68 (30)
D	-0.13 (10)	-0.11 (9.13)	-0.06 (2.88)	0.30 (5.03)	0.18 (4.03)
Log of total assets	0.04 (19)	0.61 (36)	0.33 (16)	0.55 (40)	0.54 (39)
EBIT/SALES	1.15 (42)	0.69 (19)	0.39 (13)	0.44 (13)	0.44 (13)

D: diversification

"We use three alternative econometric techniques to control for the **endogeneity** of the diversification decision... the diversification discount always drops, and sometimes turns into a premium." (Campa & Kedia 2002)

Summary

- 인과추론은 연관관계로부터 인과관계를 추론하고 언제, 그리고 왜 서로 다른지 이해하는 과학
- 인과추론의 목적은 예측 모델링과 다르며, 처치의 효과를 추론하는 것
- Counterfactual을 관측할 수 없다는 인과추론의 근본적인 문제가 존재
- Bias의 존재로 인해 우리는 연관관계를 인과관계로 해석할 수 없음
- 관측 가능한 데이터에서 인과 추정량을 찾아내는 방법은 Bias를 제거하는 과정이며, 비교가능한 그룹을 설정하는 것이 중요
- Bias를 제거하기 위해 처치(Treatment)가 어떻게 배정되었는지가 중요 (처치 배정 매커니즘) -> 독립성 가정
- Bias로 인하여 때로는 완전히 잘못된 처치 효과를 추론함 – 인과추론의 필요성

3. 인과추론 Toolbox 소개와 주의점

3장 시작에 앞서

- 실무에서 우리가 미쳐 고려하지 못한 변수(Confounders)가 처치의 효과 추론에 영향을 미칠 수 있음
- 실제 현상이 어떻게 발현되었는지 모두 알지 못하고, 모든 상황을 데이터로 통해 관측하기 어려움

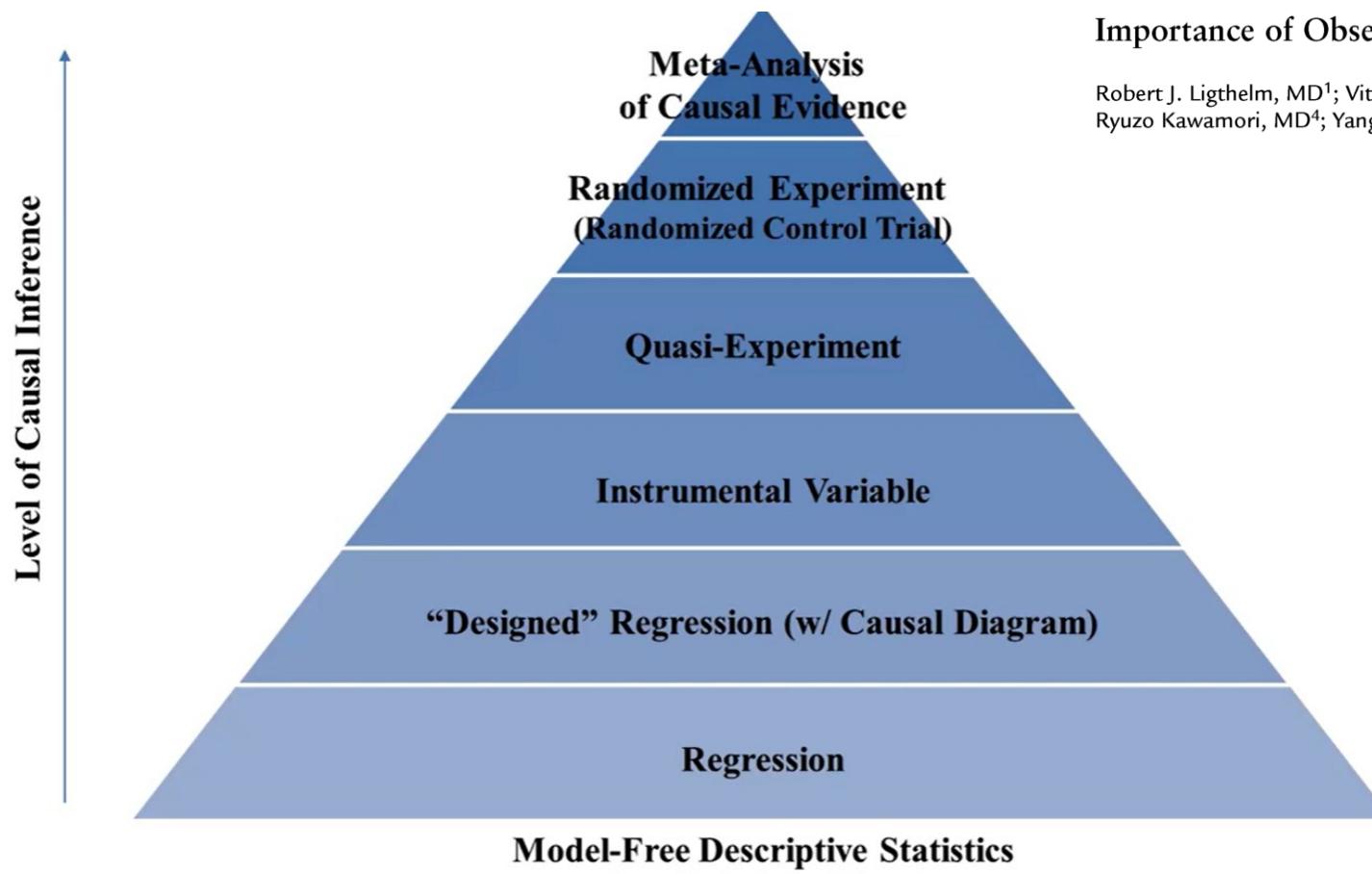
Q. 기계의 Quality에 영향을 미치는 공정 Process가 모두 알려졌으며, 관측 가능하다면?

- Causal Diagram & Regression
- 처치 배정을 실무자가 통제하는 실험 (RCT) – 독립성 가정을 강제 충족 -> Unobserved Confounder 통제
- 실험이 불가능한 상황에는 RCT를 모방할 수 있는 **상황과 디자인 그리고 식별 가정 필요**

Design-based Approach

- ✓ RCT(Random Controlled Trial)
- ✓ Quasi-experiments (DID, RDD, IV etc)
- 신뢰성 있는 결과를 위해 (1) 인과추론의 표준적인 방법과 식별 가정 (2) 분석이 가능한 상황과 디자인인가를 설명하는 것이 중요

Causal inference Toolbox



Importance of Observational Studies in Clinical Practice

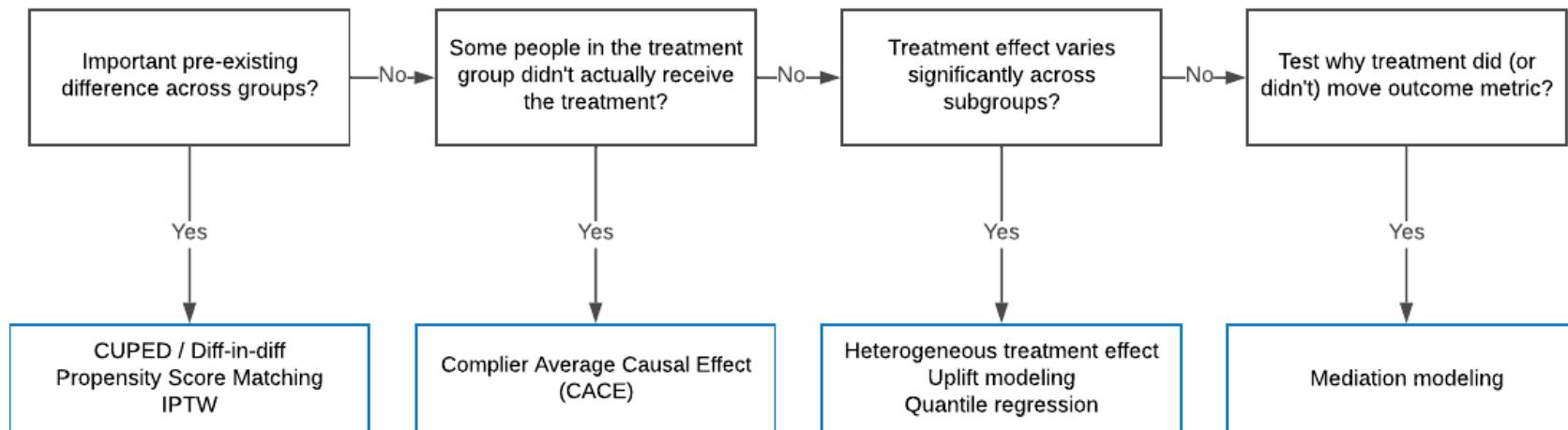
Robert J. Lighelmin, MD¹; Vito Borzi, PhD²; Janusz Gumprecht, MD, PhD³; Ryuzo Kawamori, MD⁴; Yang Wenying, MD⁵; and Paul Valensi, MD⁶

Causal inference Toolbox - 어려운 점

- Since causal inference is a family of loosely connected methods, it can feel overwhelming for a beginner to form a structural understanding of the various methods (Uber, 2019)

Causal inference Toolbox - 어려운 점

- Since causal inference is a family of loosely connected methods, it can feel overwhelming for a beginner to form a structural understanding of the various methods (Uber, 2019)
- 인과 추론의 방법론은 느슨하게 연결되어 있기 때문에 각 방법론에만 집중하면, 실제 인과추론 문제를 만났을 때 이를 적용하기 어려움
- 인과추론은 근본적으로 관측하지 못하는 값을 추론하는 과정
- 인과추론의 표준적인 접근법과 실제 산업 적용 케이스를 바탕으로 End-to-End 예제를 학습하고, 실무 예제를 템플릿화 필요



Causal inference Toolbox

Overview of data generation, modeling and interpretation in statistical perspectives

Randomized Experiments

(With randomization over treatment and control groups)

Classic Experiments (Non-recurring)

Univariate Tests

A/B Tests

Continuous Experiments (Recurring)

Statistical Techniques

Allocation %
Eg. Thompson Sampling

Rollout %
Eg. Power Based via Sequential Tests
Eg. Risk Based

Model-Based Techniques

Contextual MAB

Bayesian Optimization

Observational Studies

(Pure observation with no randomization)

Different Methods To Estimate Associated Lifts

Synthetic Control

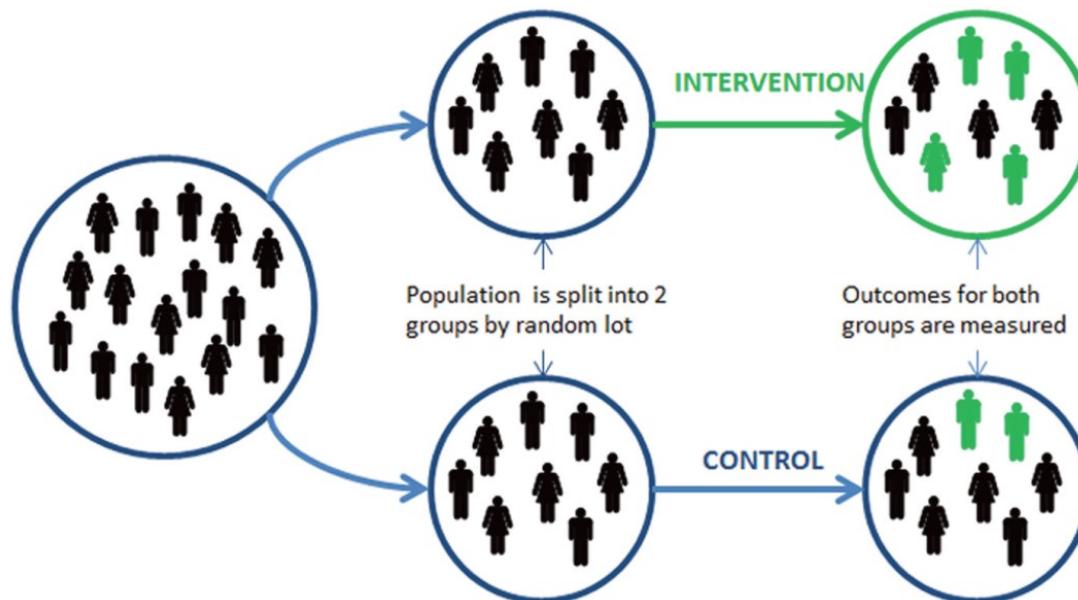
A/B-like:
How to construct a weighted "control" group

Regression

Causal inference Toolbox - RCT

RCT : Gold standard

- 처치(Treatment)를 무작위로 할당 - 처치 배정을 실무자가 통제하는 실험
- 처치에 대한 유일한 차이를 제외하고는 통계적으로 구별하기 어려운 두 개의 집단을 얻을 수 있음
 - 두 집단의 개개인은 서로 다를 수 있어도, 평균적으로 유사
- 처치를 받지 않았다면, 두 집단은 평균적으로 비슷하며 비교 가능함
- 두 집단 간의 결과에서 통계적 차이가 발생하면 이는 오직 처치에서 기인하는 인과 효과



Relying on the law of large number,
RCT guarantees orthogonality asymptotically
treatment and treatment group are exchangeable

Causal inference Toolbox - RCT

실무에서 온라인을 통한 RCT를 수행할 때, 다양한 문제로 인해 신뢰성 있는 실험을 하는 것은 어려움

- Stable Unit Treatment Value Assumption (SUTVA) – Spillover Effect, Network Effect
- 불응 문제 (Imperfect Compliance)
- 표본 크기 추정 (minimal sample size requirements Power analysis)
- 실험의 장기적 효과에 대한 추정 (Estimating Long-Term Effects)
- 실험의 평가 기준 (Overall evaluation criterion)
- 중단 문제 (Optional Stopping)
- Multiple testing, metrics coupling, incorrectly chosen statistics, and uneven test groups

Causal inference Toolbox - RCT

실무에서 온라인을 통한 RCT를 수행할 때, 다양한 문제로 인해 신뢰성 있는 실험을 하는 것은 어려움

- Stable Unit Treatment Value Assumption (SUTVA) – Spillover Effect, Network Effect
- 불응 문제 (Imperfect Compliance)
- 표본 크기 추정 (minimal sample size requirements Power analysis)
- 실험의 장기적 효과에 대한 추정 (Estimating Long-Term Effects)
- 실험의 평가 기준 (Overall evaluation criterion)
- 중단 문제 (Optional Stopping)
- Multiple testing, metrics coupling, incorrectly chosen statistics, and uneven test groups

대규모 RCT를 수행하기 위한 엔지니어링 시스템과 도구의 부재, 비용적인 문제, 윤리적인 문제 등 RCT를 수행할 수 없는 경우 존재

The Experimentation Gap – Davis Treybig

Causal inference Toolbox – Quasi Experiment

준실험(Quasi-experiment)은 무작위 실험을 모방하지만, 실험자가 임의로 그룹을 무작위로 배정하지 않음

- 특정 연구자의 의도나 개입 없이 존재하는(자연적인) 기준을 활용하여 그룹을 나누는 것
- Treatment Assignment Machinesim -> *Exogenous shock/events or Endogenous events*
- 자연적인 기준이나 사건으로 인해 처치그룹과 대조 그룹이 구분되지만, 완벽하게 무작위 할당처럼 작동한다 보장할 수 없음
- 준실험적으로 볼 수 있는지에 대한 디자인과 그에 맞는 식별 가정이 필요

예: Airbnb의 SuperHost이라는 기능이 호스트의 사전 공지없이 도입 -> 숙소 예약에 효과?

- 특정 에어비앤비 숙소에게만 영향을 주게 되는 자연적인 상황이 발생
- SuperHost와 SuperHost를 받지 않는 숙소에는 여러가지 특성의 차이가 존재
- 이중차분법(Difference in Difference)에서는 그룹간의 특성이 다를 수도 있으며, 숙소 예약 패턴의 **Time Trend**가 비교가능하면 식별 가능



Causal inference Toolbox – Quasi Experiment

- 처치 할당 매커니즘(Treatment Assignment mechanism)과 식별 가정
- 실험이 불가능한 상황에는 RCT를 모방할 수 있는 **상황과 디자인 그리고 식별 가정 필요**

Two questions you must ask (디자인)

1. What influences the assignment to treatment?
(Is it **as good as random**? Is it independent to outcomes?)
2. **How similar is the counterfactual** to the treatment group in the absence of treatment?
→ [Prove parallel trends \(Key Assumption of DID\)](#)

Causal inference Toolbox – Quasi Experiment

준실험으로 볼 수 있는 디자인과 방법론에 따른 Identification Assumption을 충족

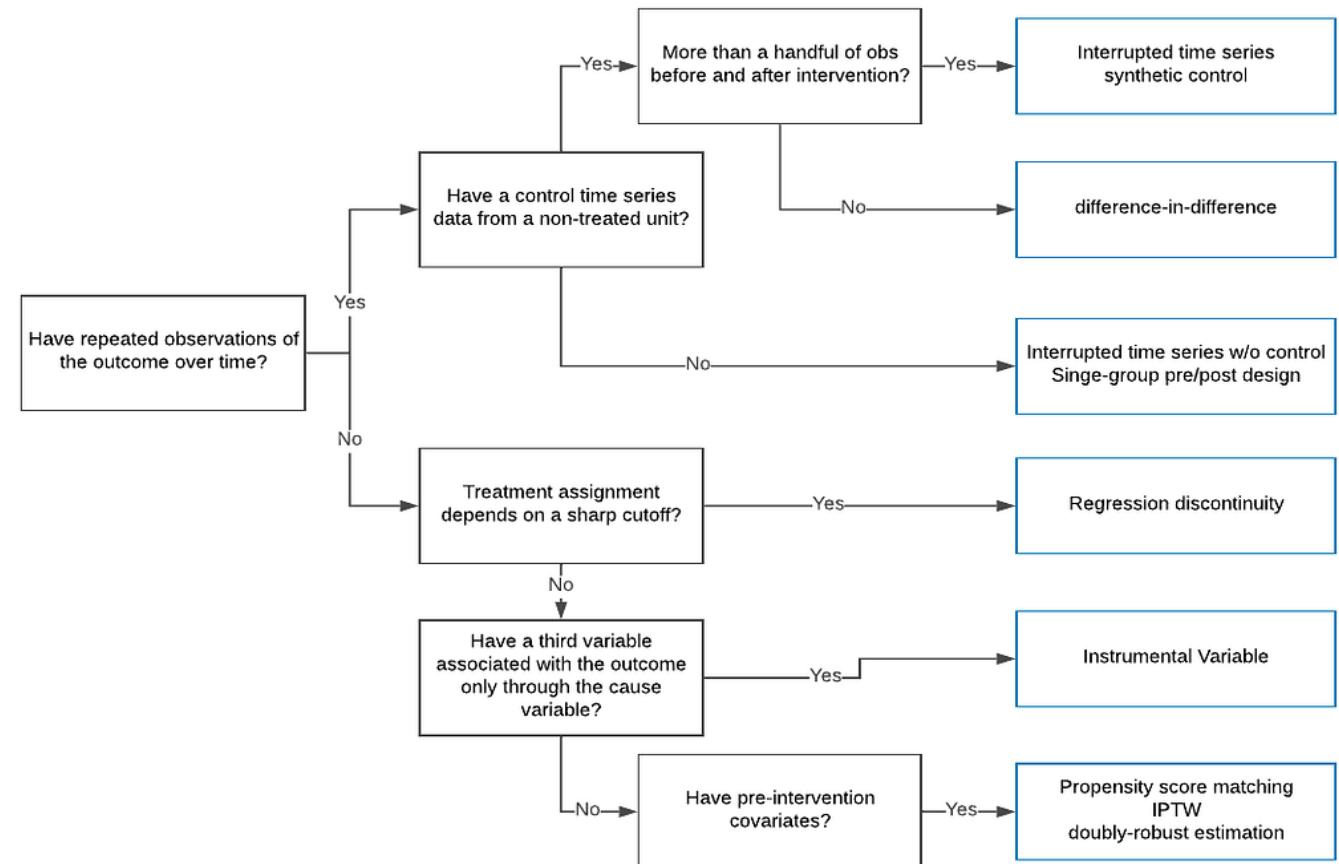
Two questions you must ask

1. What influences the assignment to treatment?

(Is it **as good as random**? Is it independent to outcomes?)

2. How similar is the counterfactual to the treatment group in the absence of treatment?

→ Prove parallel trends (Key Assumption of DID)

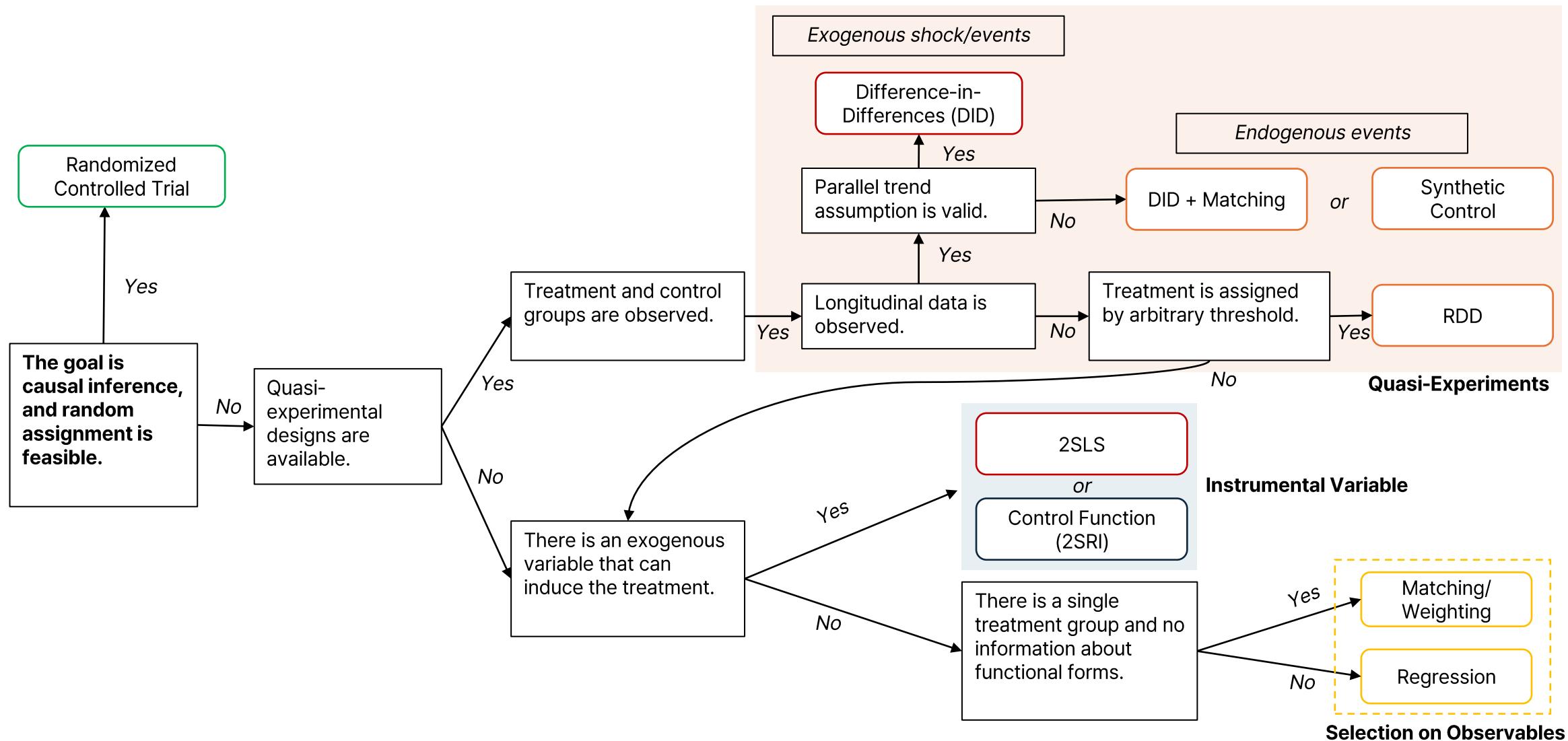


Causal inference Toolbox – Quasi Experiment

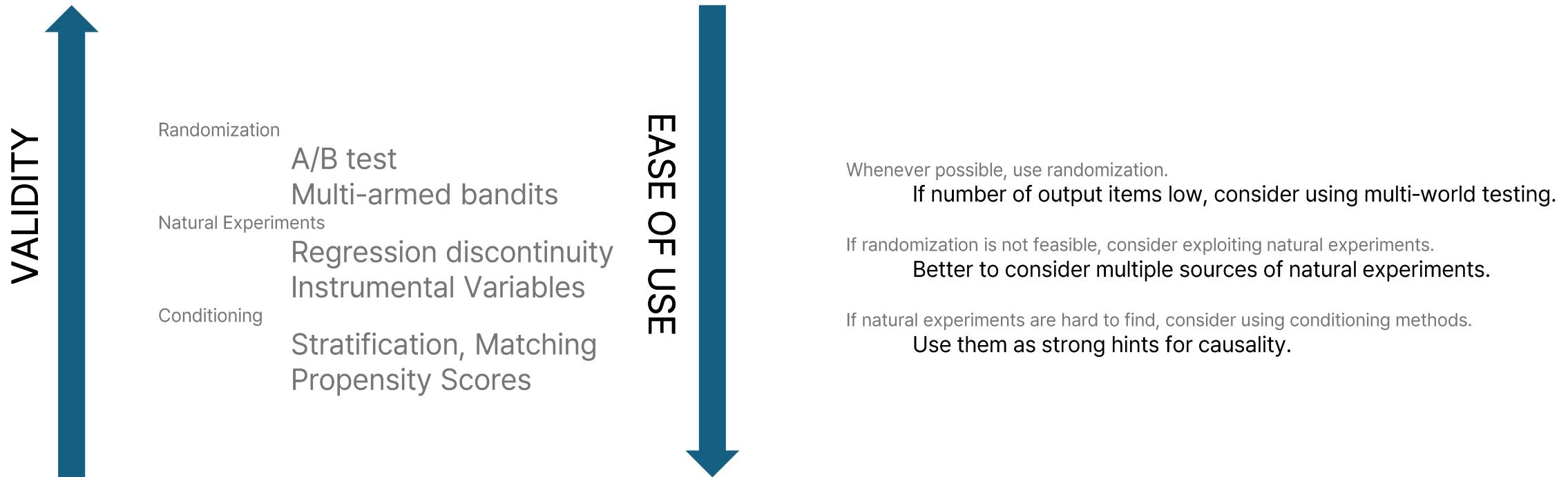
Table 2. Quasi-Experimental Regression Analysis Frameworks.

	Difference-in-Differences	Regression Discontinuity	Instrumental Variables
Identification	<p>Clarify the source of the shock, provide evidence why the shock can be seen as quasi-experimental, be clear on the identifying assumptions, and be transparent on the potential confoundedness.</p> <p>Shock의 근원을 제시 준실험적으로 볼 수 있는 증거를 제시 식별 가정 명확하게 제시 잠재적인 혼동 가능성 제시</p>	<p>Justify the source of the fixed threshold, and whether the assignment to the treatment is determined, either completely or partly, by the value of the predictor on either side of a fixed threshold.</p> <p>고정 임계값의 이유 및 근거 제시 처치에 대한 배정이 고정된 임계값에 의해 완전히 혹은 부분적으로 결정되는지 정당화</p>	<p>Justify why the IV moves the endogenous covariate as if they are an experiment; explain the exclusion restriction.</p> <p>Relevance Condition 배제 제한(Exclusion restriction)을 설명</p>
Raw data	<p>Test whether those who receive the treatment are similar to those who do not; whether the parallel assumptions are satisfied; illustrate the trajectory.</p> <p>처치 그룹과 실험 그룹이 유사한지 테스트 평행 가정이 충족되는지 테스트</p>	<p>Provide evidence that the threshold is arbitrarily determined and not linked to underlying discontinuities in effects.</p> <p>Threshold가 임의로 결정됨을 증명</p>	<p>Regress the outcome directly on the instrument and show that the instrument has the expected direct effect.</p> <p>2SLS</p>
Standard errors	<p>Cluster at the level of treatment to account for within-unit correlation of the error term over time.</p>	<p>Use robust standard errors, do not cluster on a discrete variable</p>	<p>Cluster at the level of treatment to account for within-unit correlation of the error term over time.</p>

Causal inference Toolbox



Causal inference Toolbox

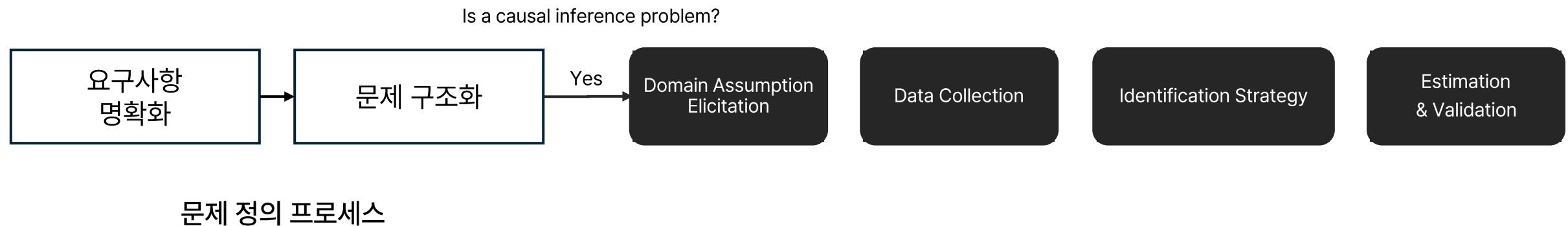


Causal inference Toolbox – 정리

(1) 분석 상황을 밝히고 (2) 상황에 맞는 방법론을 선택하고 (3) 방법론에 해당하는 가정을 밝히기. 마지막으로 (4) 추가 분석 제공

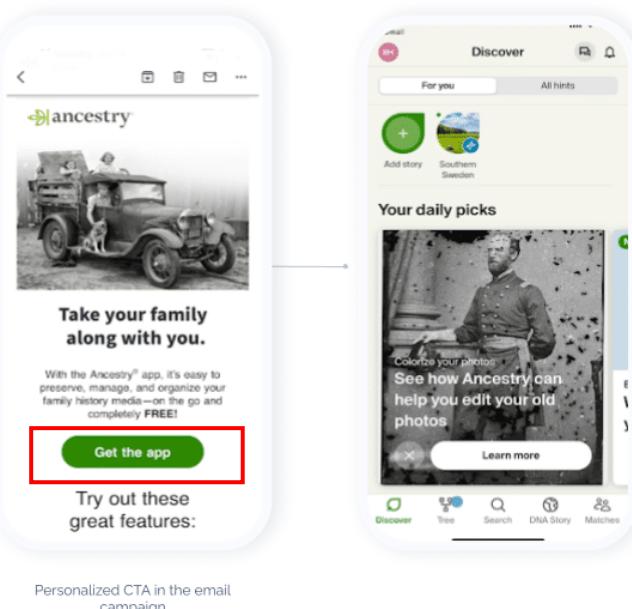
- 분석하려는 대상 및 상황이 인과추론을 할 수 있는지에 대한 증거 혹은 설명 제시
준실험이라면 준실험적으로 볼 수 있는 증거를 제시
- 상황에 맞는 Research Design 선택 -> Treatment assignment mechanism
 - ✓ Shock가 시스템 외부에서 발생하는 이벤트나 변화인가? (e.g., 코로나)
 - ✓ 해당 Shock가 모든 그룹에 동일하게 적용되었는가?
 - ✓ Shock가 시스템 내부에서 알 수 없는 특정 조건에 의해 발생하는가? (e.g., SuperHost) -> 처치가 무작위 X
 - > Matching + DID
- Research Design의 식별 가정 (Exchangeability)
 - ✓ 특정 모델을 사용하고 인과 효과를 추정하는 것만으로는 인과 효과를 보장할 수 없음 (실무자가 주의해야 할 케이스)
 - ✓ Prove parallel trends (Key Assumption of DID)
- 추가 분석 제공
 - Propensity Score Matching + CEM, Look ahead Matching 등 다양한 방식 활용하여 Robustness 제공
 - ✓ PSM의 상황에서 Unobserved confounders에 대한 Sensitivity 검증 방식 활용
 - Rosenbaum Sensitivity Bounds

4. 인과추론의 라이프 사이클과 예제



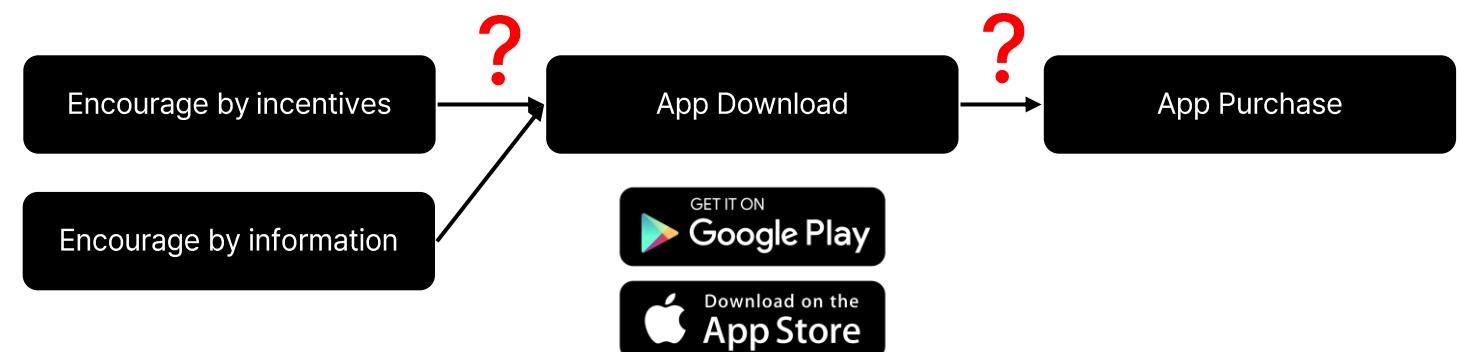
Domain Assumption Elicitation

- 플랫폼에서 보유한 채널(데스크톱, 모바일 웹, 모바일 앱) 중 극히 일부 유저만 활용하는 모바일 앱으로 다운로드 및 구매 유도
- 마케팅을 통해 사용자가 모바일 앱을 다운로드하도록 유도, 마케팅 방식은 앱에 대한 정보제공과 다운로드 인센티브 제공 이메일 전송
 - (1) Organic 다운로드 뿐만 아니라, 마케팅 활동을 통한 앱 다운 유도에 대한 인과적 효과를 이해
 - (2) 정보제공으로 앱을 다운로드 받은 유저와 인센티브로 다운받은 유저 중에 누가 더 많은 In-App 구매로 이어질까?



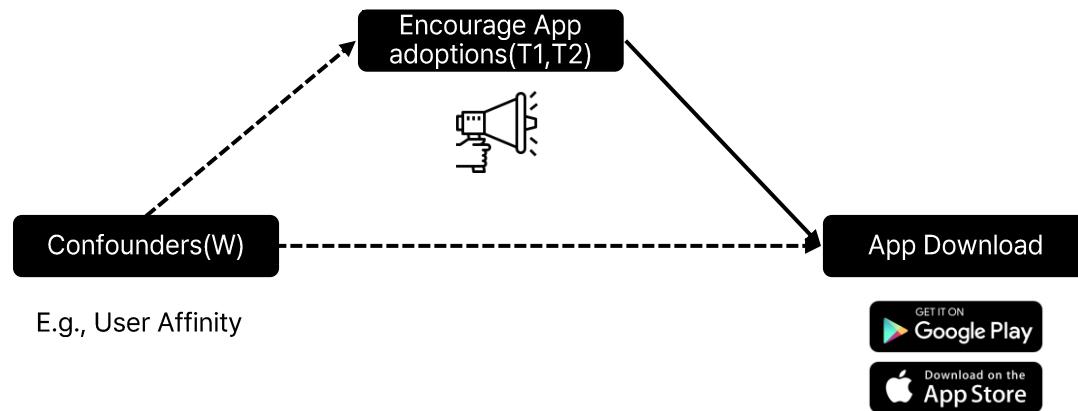
Q1. App 다운 인센티브 제공 vs App의 정보를 담은 마케팅이 고객의 App 다운로드에 미치는 영향

Q2. App 다운 인센티브 제공 vs App의 정보를 담은 마케팅이 고객의 향후 In-app 구매에 미치는 영향

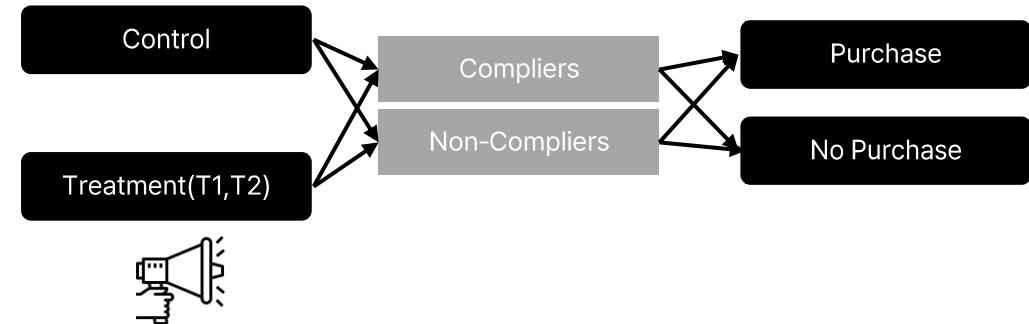


Identification Strategy

Confoundedness -> RCT



Imperfect Compliance -> LATE



- 관찰 데이터에서는 유저의 서비스 선호도를 통한 Organic 다운로드와 App 다운로드 유도 마케팅을 통한 다운로드에 대한 효과를 밸라내기 어려움 (Selection bias)
- 단순 관찰 데이터에서는 App 광고를 받은 사람 vs App 광고를 받지 않는 사람 두 집단을 동일하게 비교를 할 수 없음

- A/B 테스트 수행 중에, App 다운로드를 사용자에게 강제할 수 없음
- 유저가 참여 여부를 결정할 수 있는 경우 Imperfect Compliance 문제 발생
- LATE 개념을 활용해 Complier의 app 다운로드 효과를 분리

Identification Strategy

- Any Pre-existing Difference? : No

Table 1. Randomization Check

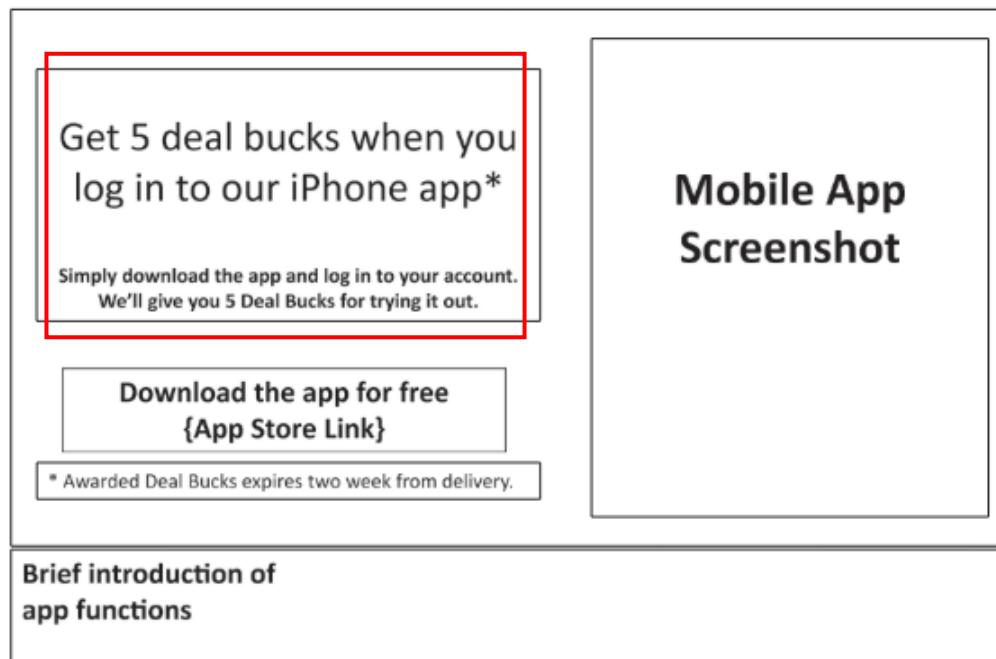
	Control, n = 137,195	T1, n = 48,027	T2, n = 48,070	p-Value (Control = T1 = T2)
Customer tenure, days				
Mean	0	-1.32	0.01	0.407
SD	193.4	192.9	193.97	
Total no. of purchases				
Mean	0	0.01	0.01	0.909
SD	3.82	3.69	3.74	
Total desktop purchases				
Mean	0	0.02	0.01	0.726
SD	3.6	3.48	3.52	
Total mobile web purchases				
Mean	0	-0.01	0	0.384
SD	0.84	0.81	0.83	
Total revenue				
Mean	0	-0.02	0.21	0.983
SD	237.37	228.19	236.89	
Average price of deal purchases				
Mean	0	-0.39	-0.53	0.481
SD	89.88	82.10	80.83	
No. of unique deal categories purchased				
Mean	2.08	2.08	2.09	0.460
SD	1.44	1.46	1.45	

Notes. The figures provided are demeaned values obtained by subtracting the mean value of treatment groups from that of control group. Demeaning preserves the difference in mean value between test groups as well as the t test (i.e., randomization check). SD, standard deviation.

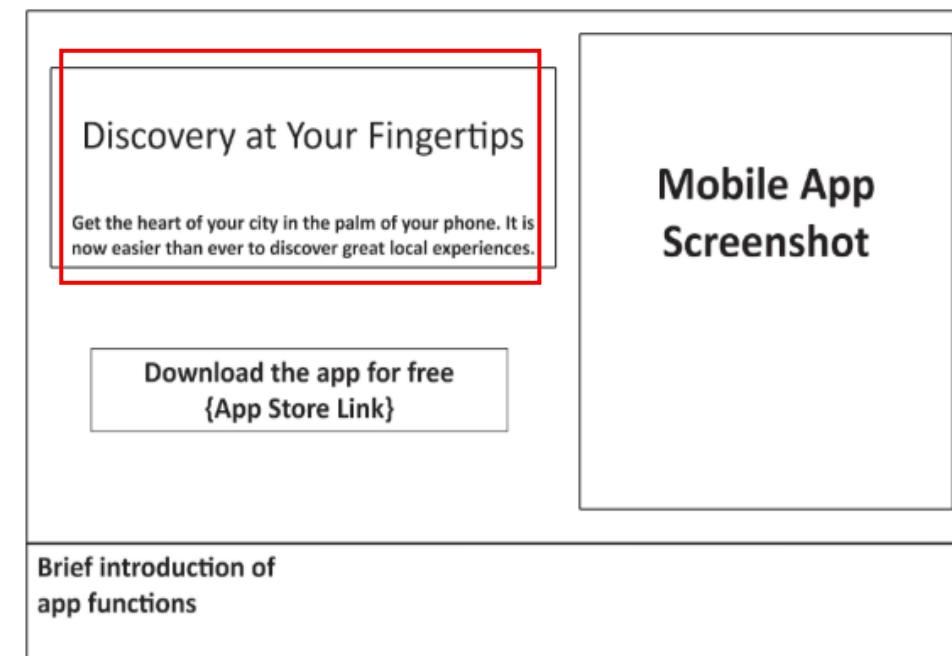
Identification Strategy

- 첫째, 인센티브/정보는 오직 이메일 채널을 통해서만 제공되기 때문에, 고객들은 다른 채널을 통해 캠페인에 참여할 수 없음 (**Spillover effect**)
- 둘째, 서로 다른 처치 그룹에 있는 고객들이 앱을 다운로드 한 후(컨트롤, T1, 그리고 T2), 모바일 앱에서 정확히 동일한 경험과 정보를 받음

(T1) Email Template for Treatment 1: Highlight incentive for app adoption



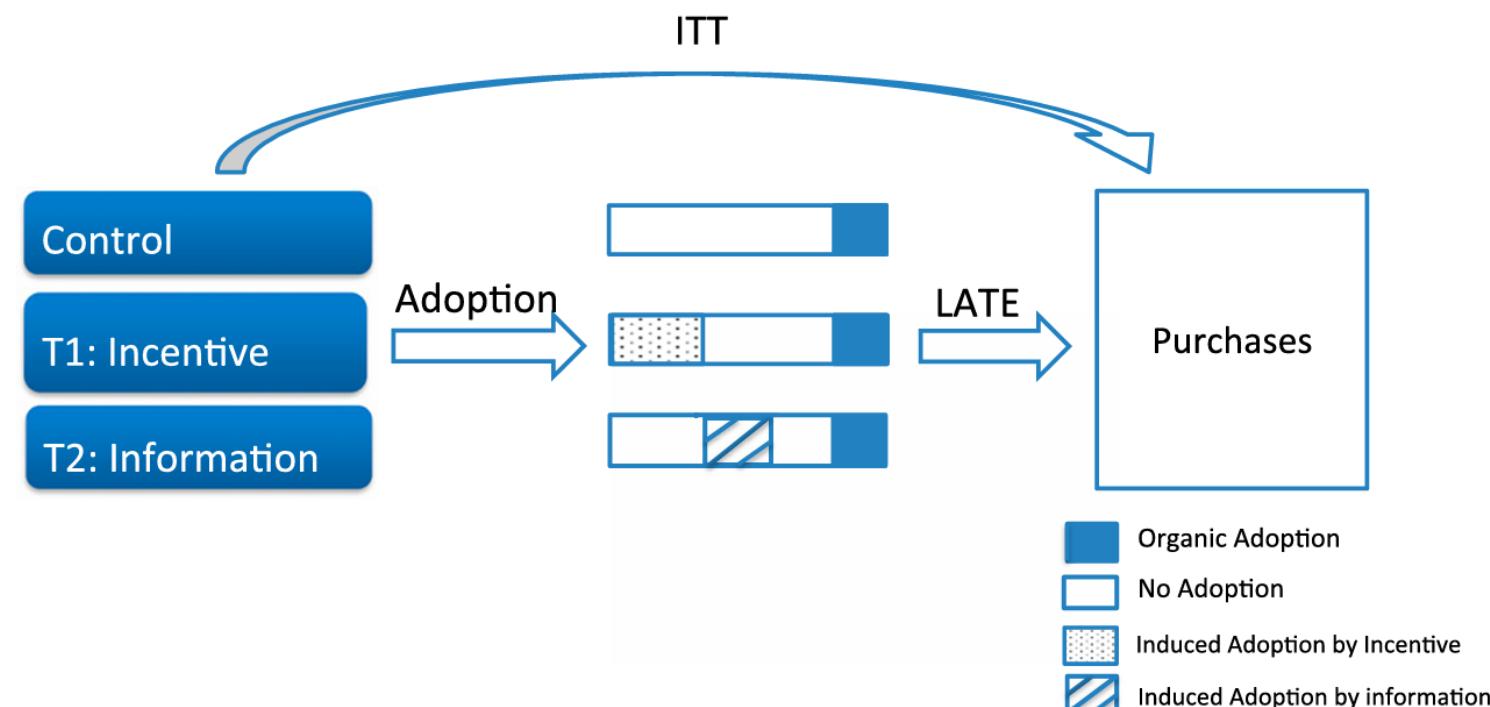
(T2) Email Template for Treatment 2: Highlight information about the benefits of the app



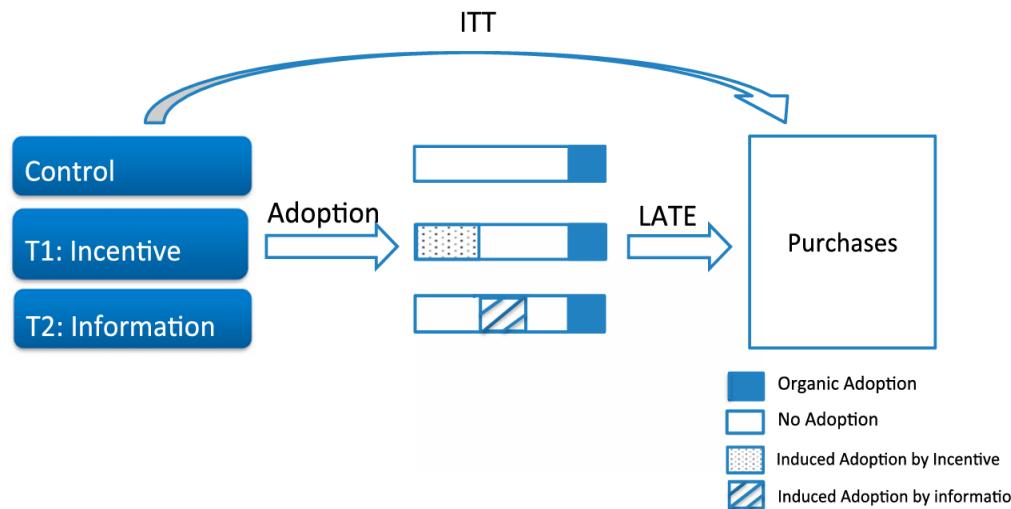
- (1) 정보나 인센티브에 대한 정보를 받지 못한 통제 그룹 (대상자 137,195명), (2) 앱 채택에 대한 인센티브(5달러)를 강조한 이메일을 받은 처치그룹 T1 (대상자 48,027명),
 (3) 앱을 유용성에 대한 정보 이메일을 받은 처치그룹 T2 (대상자 48,070명)

Identification Strategy

- Organic 다운로드는 LATE 프레임워크 하에서 'always -takers' : 어떠한 외부 개입 없이도 앱을 자연스럽게 다운로드할 고객
- 처치 받은 모든 대상(ITT)의 효과를 보고 싶은 것이 아닌, organic을 제외한 인센티브 혹은 정보 제공을 받은 사람들이 인앱 구매 여부가 관심



Identification Strategy



도구변수 선정 및 추정

도구변수 선정 : 처치 할당(T)

1) First Stage

$$Download_i = \sum \alpha_k \times T_{ik} + \varepsilon_i,$$

2) Second Stage

$$Purchase_i = \beta \times \widehat{Download}_i + \sigma_i. \quad (1)$$

도구변수 식별 가정

- (1) 단조성(monotonicity) : 사용자가 처리를 받았을 때, 앱 채택의 확률이 통제집단에 비해 증가
 - Treatment와 반대로 움직이는 Definer 제거
 - 이메일 캠페인은 앱 채택을 위한 더 많은 정보나 인센티브를 제공하기 때문에 가정을 충족
- (2) 배제 제한(exclusion restriction) : 정보 제공이나 인센티브 제공의 처치가 인앱 구매에 영향 X
 - 정보 제공: 인앱 구매를 장려하는 것이 아닌, 앱의 혜택만을 언급 (?)
 - 인센티브 제공: 인센티브는 발송 후에 2주 이내에 완료되기 때문에, 단기적으로 구매가 증가
 - LATE 추정에 실험 후 3주 동안의 구매를 테스트 그룹에서 배제

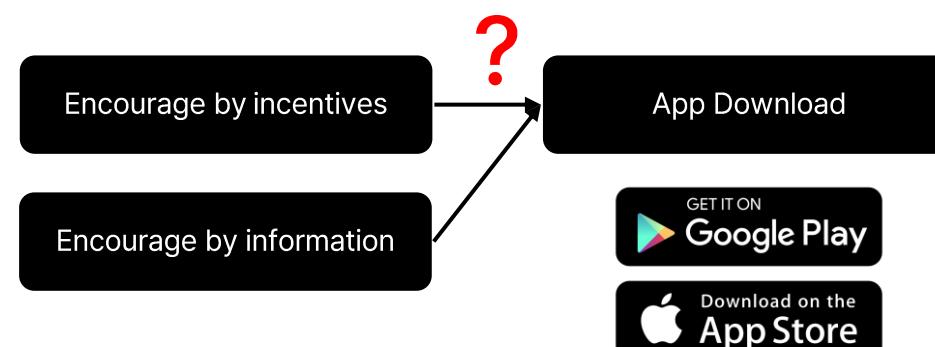
Estimation/Inference

Table 2. Effect of Treatments on Mobile App Adoptions (Q1a)

	Download_1day	Download_3day	Download_1week	Download_2week
T1	0.00533*** (0.000347)	0.00931*** (0.000497)	0.00995*** (0.000543)	0.0101*** (0.000626)
T2	0.00124*** (0.000189)	0.00287*** (0.000340)	0.00328*** (0.000398)	0.00322*** (0.000502)
Constant	0.000357*** (5.10e-05)	0.00200*** (0.000121)	0.00325*** (0.000154)	0.00666*** (0.000220)
Observations	233,292	233,292	233,292	233,292
p(T1 = T2)	8.08e-26	7.44e-29	1.30e-25	1.35e-20

Notes. The coefficients of T1 and T2 are significantly different based on the p -value($T1 = T2$). Robust standard errors in parentheses.

*** $p < 0.01$.



Estimation/Inference

Table 4. Effect of Induced App Adoptions on Purchase Behaviors in the Long Run:
LATE (Q1b)

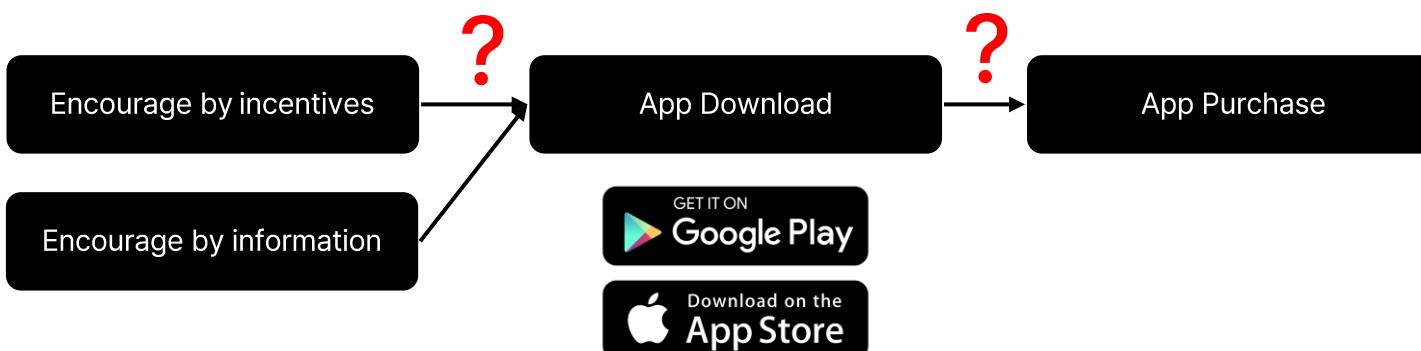
	T1		T2	
	Purch_6month	Purch_1year	Purch_6month	Purch_1year
Induced adoption	-0.274 (0.830)	-0.500 (1.476)	5.081** (2.589)	10.25** (4.650)
Constant	0.877*** (0.00613)	1.685*** (0.0109)	0.859*** (0.0113)	1.650*** (0.0201)
Observations	185,222	185,222	185,265	185,265

Notes. We exclude customers' purchases within the first 3 weeks to maintain exclusion restriction and focus on customers' purchase behavior in the long run. The results are robust when excluding purchases within the first three months and six months (Online Appendix, Table A.1). Robust standard errors in parentheses.

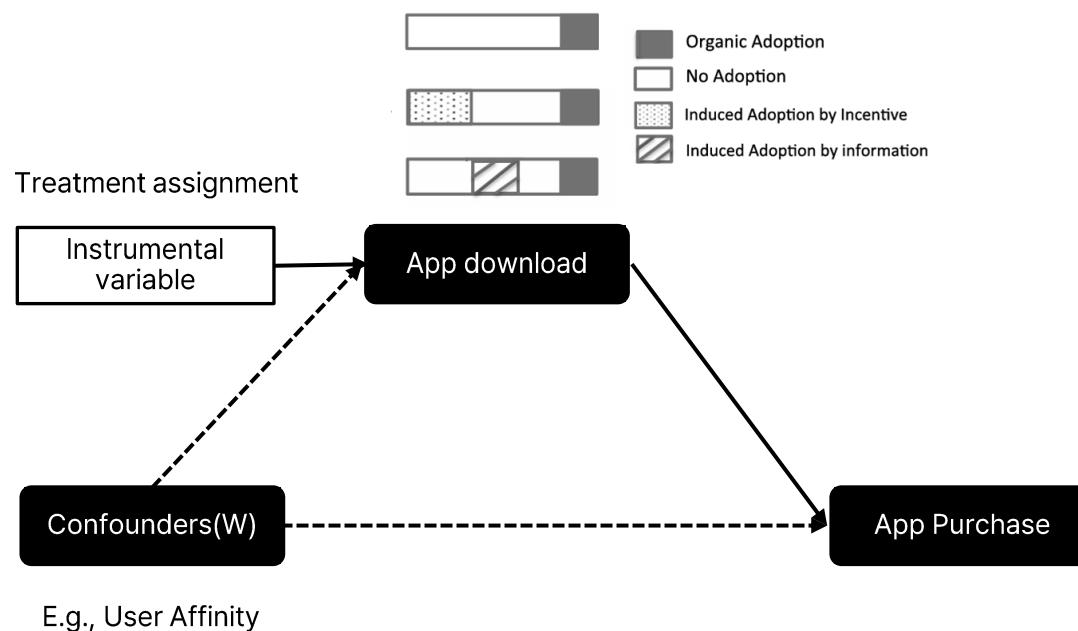
** $p < 0.05$; *** $p < 0.01$.

- 정보(T2)에 의해 유도된 앱 채택만이 장기적으로 고객의 구매 증가에 유의미한 영향
- 인센티브(T1)에 의해 유도된 앱 채택은 고객의 구매 행태에 인과적 영향을 미치지 않음
- 금전적 인센티브(T1)가 사람들이 앱을 채택하도록 유도하는 데 효과적이지만, 이러한 모집 접근 방식은 그들이 앱을 다운로드 한 후 더 많은 구매로 이어지지 않음

반대로, 정보를 제공하는 것은 앱 채택에서 더 작은 증가를 가져오지만, 그러한 채택은 장기적으로 고객의 구매에서 지속 가능한 증가



Wrap Up



- Q1. 인센티브와 정보가 고객의 앱 채택에 미치는 영향은 무엇인가요?
- Q2. 앱 채택 유도가 장기적으로 고객의 행동에 미치는 인과관계는 무엇인가요? 인센티브에 의한 채택과 정보에 의한 채택의 효과는 다른가요?
- 앱 다운로드는 사용자 참여가 필요한 행동
Always takers - organic 채택
(non-compliance시 편향된 추정치로 이어질 수 있음)

사용자를 기준으로 랜덤으로 인센티브 or 정보 제공 이메일 할당

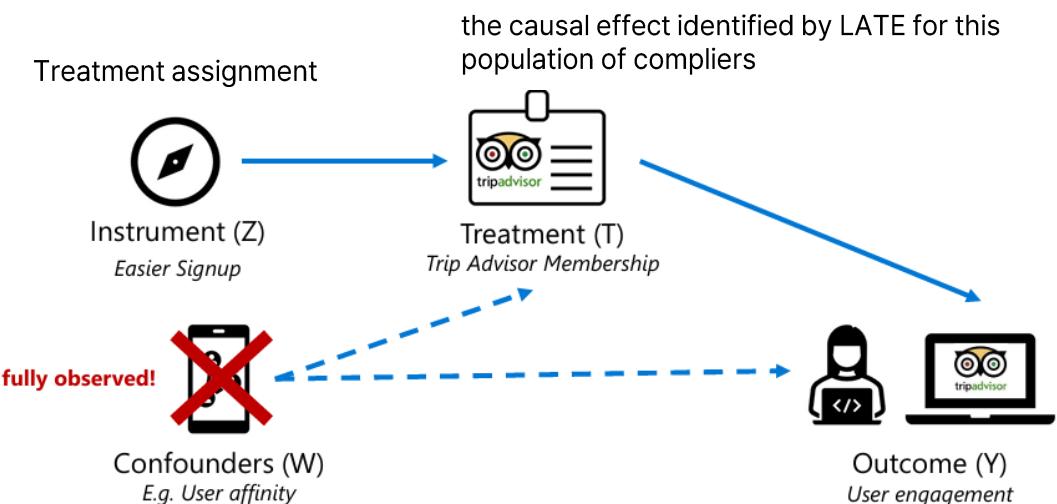
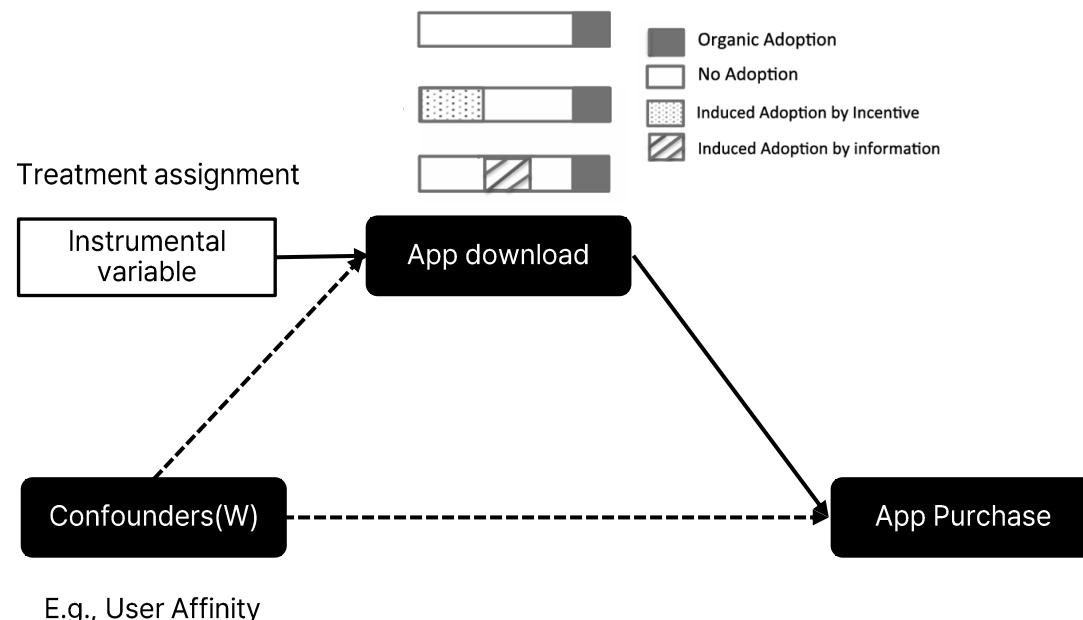
도구 변수: 처치 할당 T에 영향을 미치지만 처치 이외의 결과 Y에는 영향을 미치지 않는 임의의 변수 Z 설정 (단조성, 배제제한)

Assumptions

- Stable Unit Treatment Value Assumption (SUTVA) : Non-interference
- Excludability Assumption
- Monotonicity Assumption

Causal Design Pattern

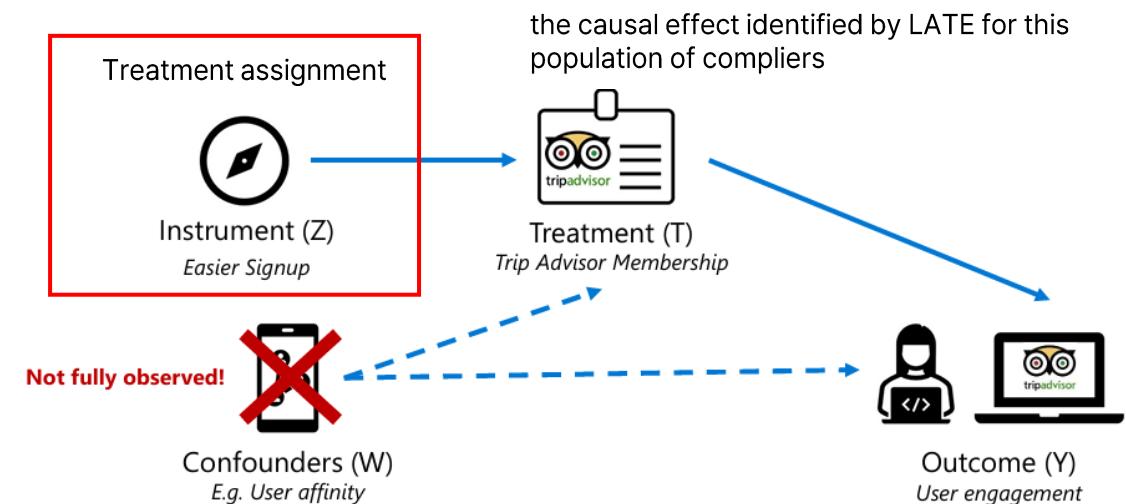
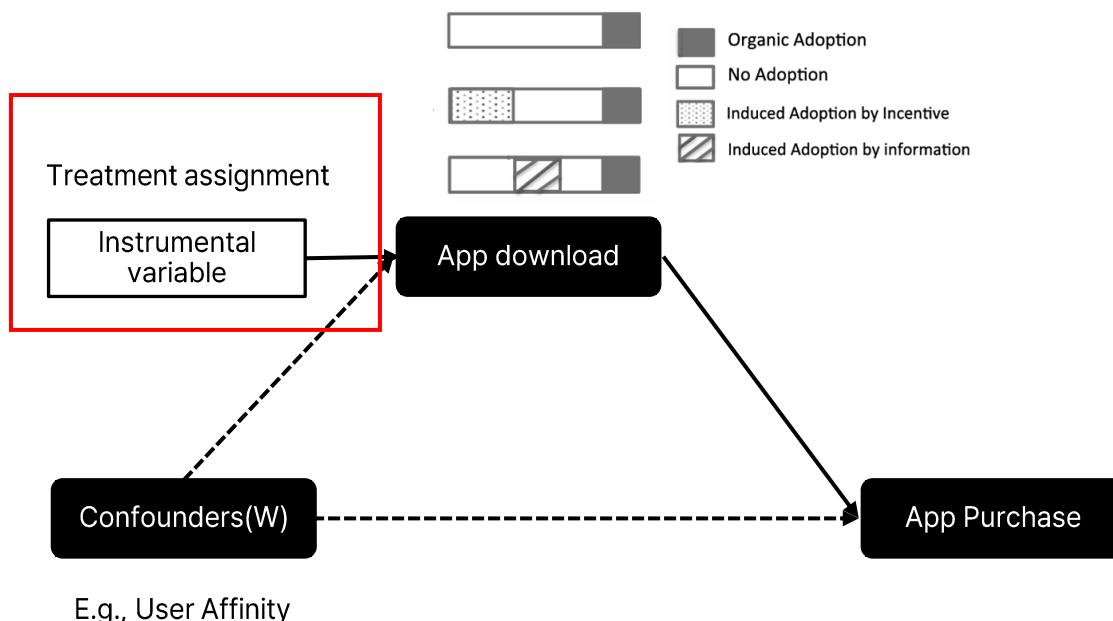
- 아래 사례에서 관찰할 수 있는 패턴이 있나요?



Customer Segmentation at TripAdvisor with Recommendation AB Tests
KDD tutorial 2021

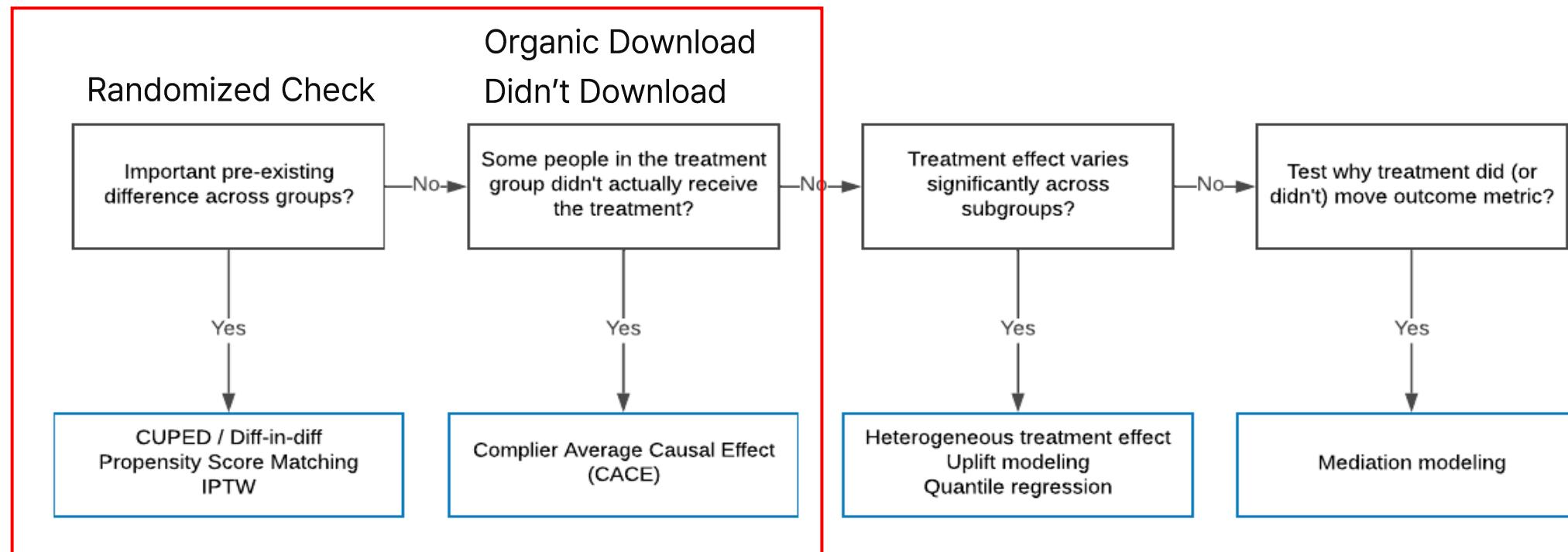
Causal Design Pattern

- 아래 사례에서 관찰할 수 있는 패턴이 있나요?
- 도구변수 정당화에 대한 가장 좋은 방법은 도구변수를 **Random assignment**
- 도구변수가 Random하게 Assign되면, **observable covariates**와 상관이 없음

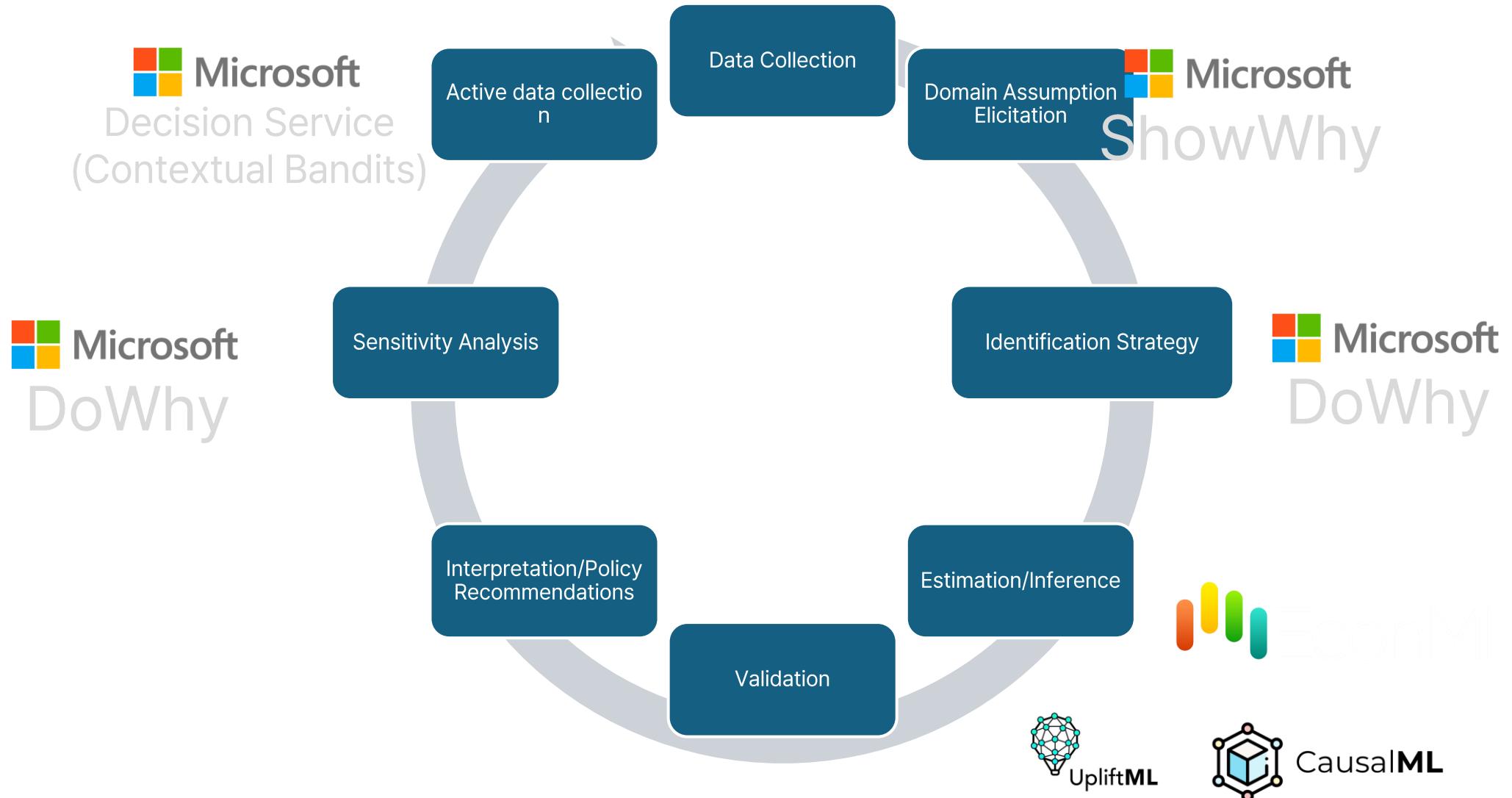


Customer Segmentation at TripAdvisor with Recommendation AB Tests
KDD tutorial 2021

Causal Design Pattern



Effect of Incentive & Information app
download on In-app Purchase



5. 앞으로 하고자 하는 것 : 사례 템플릿화

Causal inference with experimental data

No	Question	Method	사례
1	Important pre-existing difference across groups?	CUPED / Diff-in-diff Propensity Score Matching IPTW	
2	Some people in the treatment group didn't actually receive the treatment?	Complier Average Causal Effect (CACE)	Week 2
3	Treatment effect varies significantly across subgroups?	Heterogeneous treatment effect Uplift modeling Quantile regression	
4	Test why treatment did (or didn't) move outcome metric?	Mediation modeling	

Causal inference with observational data

No	Criteria	Method	사례
1	Time-series data before and after treatment are available, but no control group	Interrupted Time-Series Analysis	
2	Exogenous shock/events + Valid Parallel trend assumption	Difference-in-Differences (DID)	
3	Exogenous events + Invalid Parallel trend assumption	DID + Matching or Synthetic Control	
4	Endogenous events + Arbitrary threshold	Regression Discontinuity	
5	There is an instrumental variable that induces treatment	Two-Stage Least Squares or Selection Bias Correction Method	
6	Selection on observables is possible	Matching	
7	Matching is not possible	Regression	

감사합니다

Q&A