

인과추론과 실무 : 그래프 인과모델

가짜연구소 인과추론팀

발표자 : 심소현

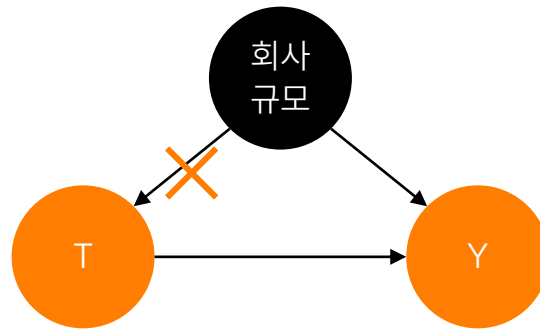
1. Structural Causal Model

RCT와 실험플랫폼 Recap

Selection Bias

RCT

- Potential Outcomes Framework : 실험 설계를 통해 selection bias를 제거하기



(회사 규모의 차이가 없다면)

T : 크리스마스 주간에 가격 할인 정책을 사용하면

Y : 제품의 판매량이 높아질까?

RCT와 실험플랫폼 Recap

무작위 실험의 난제

- Interference : 처치 그룹과 통제 그룹 간 상호작용이 발생할 가능성
- Sample Ratio Mismatch (SRM) : 무작위 처치 배정 메커니즘을 적용했으나 실제로는 selection bias가 발생

그 밖에도 실험을 할 수 없는 다양한 상황이 존재

RCT와 실험플랫폼 Recap

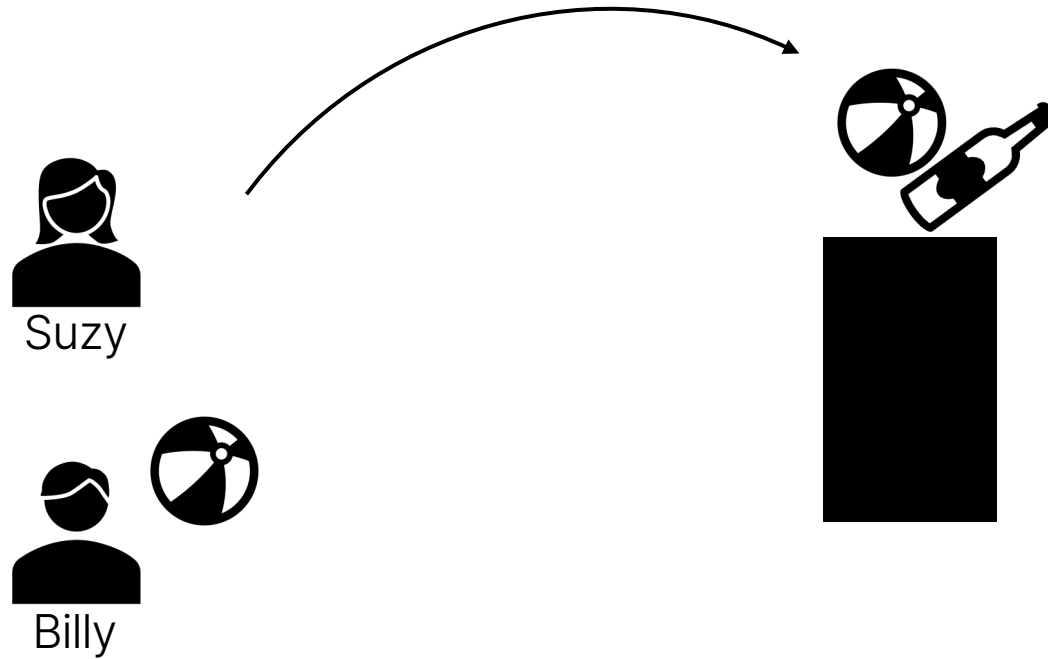
좋은 게임 아이템을 획득하면 게임 플레이를 열심히 할까?

- 똑같이 플레이를 한 유저들 중 임의로 선택한 유저에게만 아이템을 획득하게 하기 (X)

차량 소프트웨어 업데이트가 차량의 에너지 소비량을 감소시킬까?

- 특정 사용자에게 소프트웨어 업데이트를 강제하거나, 소프트웨어 업데이트를 못하게 막기 (X)

Counterfactual?

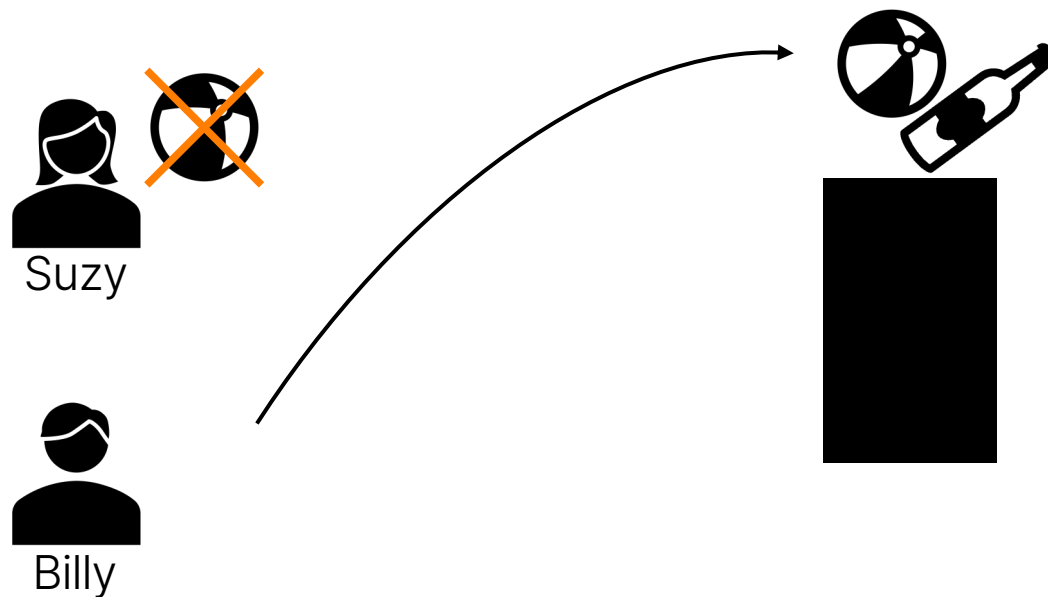


1. Suzy와 Billy가 공을 던져서 병에 맞힐 확률은 각각 100%
2. Suzy는 Billy보다 먼저 공을 던진다.

T : Suzy가 공을 던지면
Y : 병이 쓰러질까?

Counterfactual?

Counterfactual이 밝히고자 하는 관계를 둘러싼 상황을 충분히 고려하고 있을까?



1. Suzy와 Billy가 공을 던져서 병에 맞힐 확률은 각각 100%
2. Suzy는 항상 Billy보다 먼저 공을 던진다.

T : Suzy가 공을 던지면
Y : 병이 쓰러질까?

Counterfactual : Suzy가 공을 던지지 않는 것
그런데 Suzy가 공을 안 던져도 병은 쓰러진다.

$$SH \leftarrow f_{SH}(ST) = ST$$

$$BH \leftarrow F$$

$$BFO \leftarrow f_{BFO}(SH, F) = SH \vee BH$$

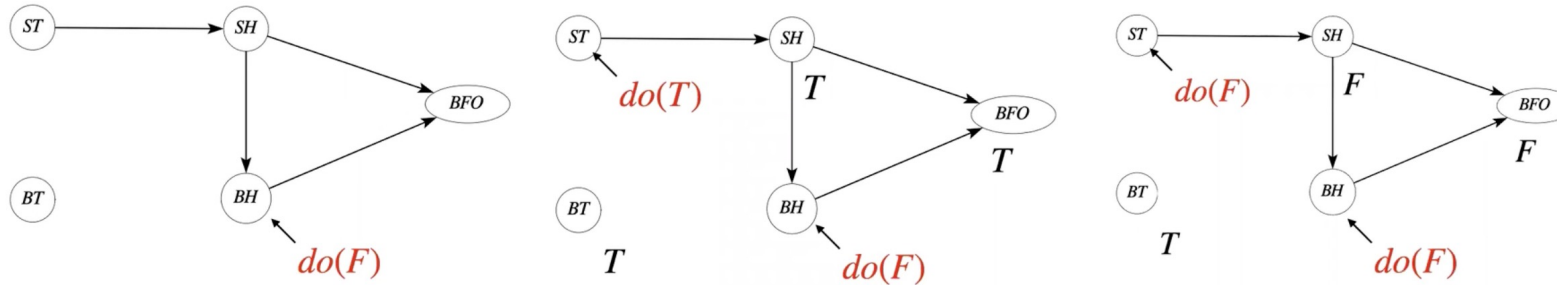
ST: Suzy Throws $\in \{T, F\}$

BT: Billy Throws $\in \{T, F\}$

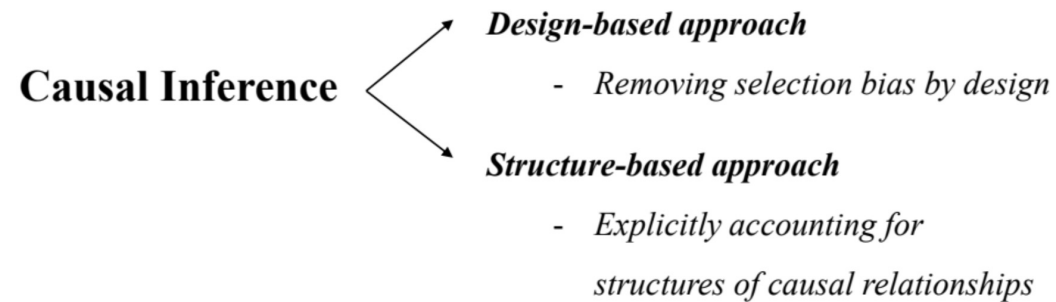
SH: Suzy's ball hit the bottle

BH: Billy's ball hit the bottle

BFO: Bottle Fall Off



2 Approaches of Causal Inference



Potential Outcomes Framework (Donald Rubin)			Structural Causal Model (Judea Pearl)
Table 1.1. Example of Potential Outcomes and Causal Effect with One Unit			
Unit	Potential Outcomes		Causal Effect
	<i>Y(Aspirin)</i>	<i>Y(No Aspirin)</i>	
You	No Headache	Headache	Improvement due to Aspirin


```

graph TD
    Drug --> BP[Blood Pressure]
    Drug --> HA[Heart Attack]
    BP --> HA
  
```

Imbens, G.W., 2020. Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics. *Journal of Economic Literature*, 58(4), pp.1129-79.

2 Approaches of Causal Inference

Potential Outcome Framework		Structural Causal Model
Gold Standard	Random Assignment	
Causal Inference Using Observational Data		
(1) Identification (Is it possible to estimate a causal effect?)	Research Design (Quasi-Experiment / Natural Experiment)	Backdoor Criterion / do-Calculus (based on Bayesian Network/DAG)
(2) Estimation (How to estimate a causal effect using data?)	Statistical Methods (DID, RD, Matching, IV, SC, etc.)	Statistical/Computational Methods (IPW, Doubly Robust Estimators, Double ML, etc.)

2. Structural Causal Model

들어가기 전에

오늘 다룰 내용 : 그래프를 활용해서 인과 관계를 식별하기

- DAG
- conditioning building block
- Intervention & do-operation
- Backdoor Criterion & Adjustment

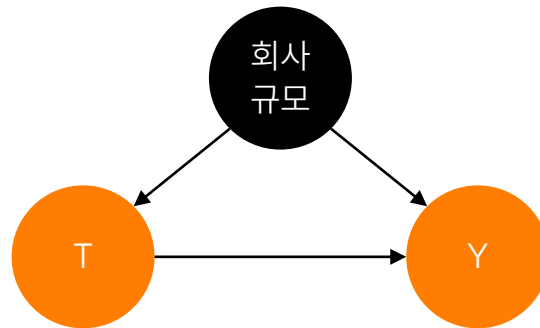
다루지 않을 내용 : 어떻게 그래프로 인과 관계를 식별하는 것이 가능할까?

- Bayesian Network Factorization
- do-calculation

DAG

그래프를 활용해서 인과관계를 표현

Directed : 방향이 있는
Acyclic : 순환하지 않는
Graph : 그래프



Node : 변수
Edge : 관계

DAG로 표현하는 관계의 종류

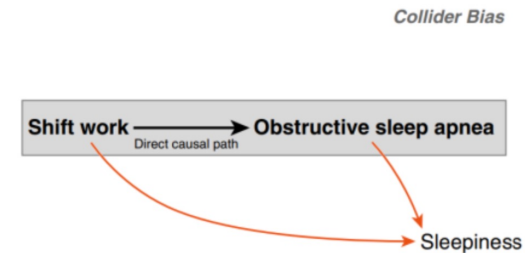
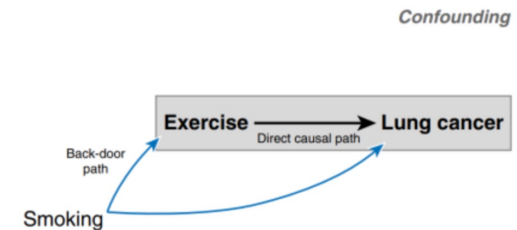
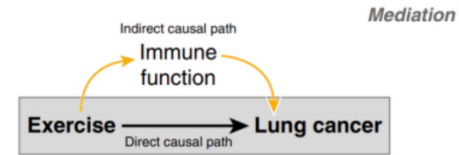
- Relationship types in DAG

- (Direct) Causal Effect $D \longrightarrow Y$

- Mediator (Chain) $D \longrightarrow X \longrightarrow Y$
(Indirect Causal Effect)

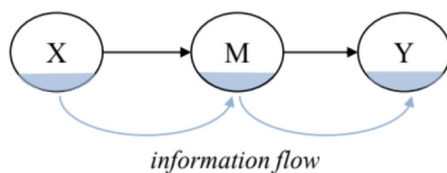
- Confounder (Fork) $D \longleftarrow X \longrightarrow Y$
common causes

- Collider (Immortality) $D \longrightarrow X \longleftarrow Y$
common effects



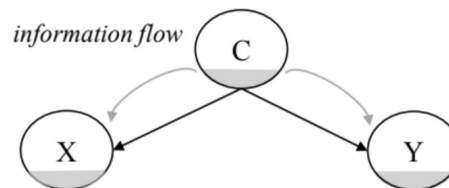
교란 변수 유형과 영향 차단하기

Mediator (Chain)



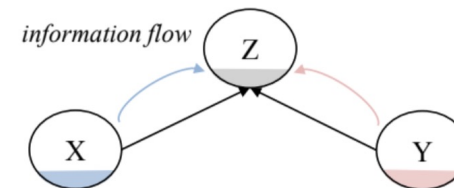
X and Y are d-connected.

Confounder (Fork)



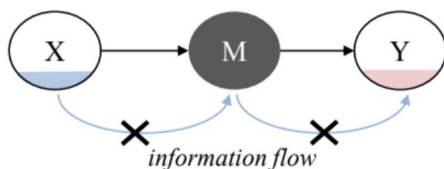
X and Y are d-connected.

Collider (Immortality)



X and Y are d-separated.

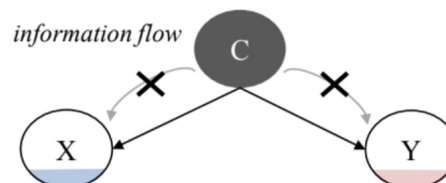
Mediator (Chain)



X and Y are d-separated.

To estimate the *direct* causal effect of X on Y, mediators should be blocked.

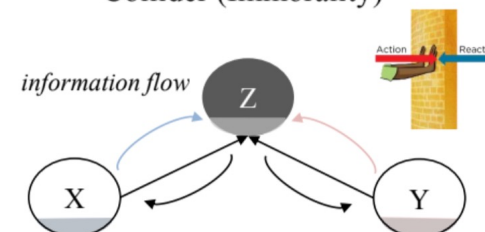
Confounder (Fork)



X and Y are d-separated.

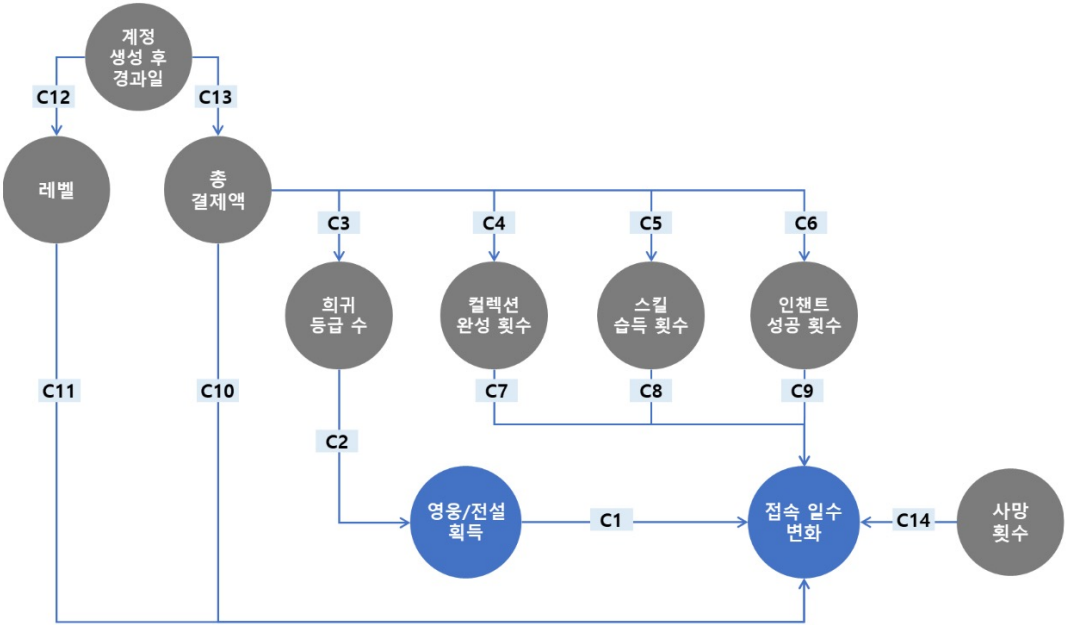
To estimate the *direct* or *indirect* causal effect of X on Y, confounders should be blocked.

Collider (Immortality)



X and Y are d-connected.

To estimate the *direct* or *indirect* causal effect of X on Y, colliders should NOT be blocked.



Intervention & do-operation

Intervention

변수의 값을 지정 (~ 처치)

> 개입을 통해 해당 변수에 미치는 영향을 제거할 수 있음

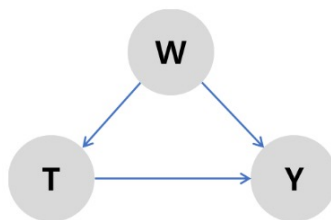
do-operation

$\text{do}(X=x)$: 변수 X 를 x 로 고정

> 변수 X 에 미치는 모든 영향을 제거함

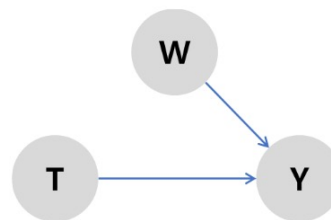
Causal effect of X on $Y = P(Y|\text{do}(X))$

Statistical distribution



T에 개입하지 않음
(관찰된 데이터)

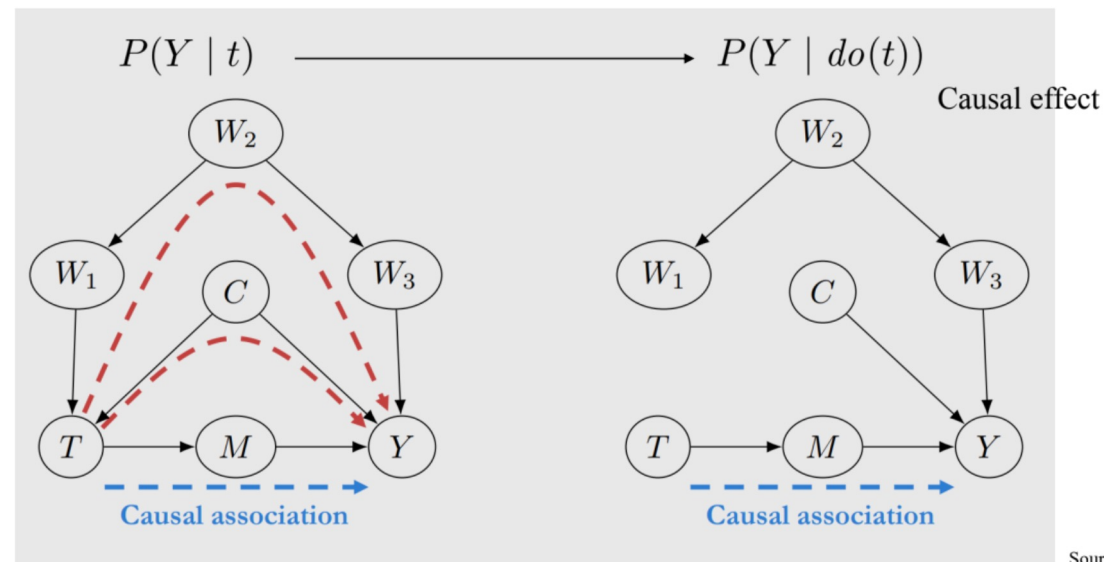
Interventional distribution



T에 개입
(실험을 통해 생성된 데이터)

Backdoor Criterion & Adjustment

- Backdoor criterion: A set of variables that block all paths, but the causal association, between X and Y
- Backdoor adjustment: Conditioning on variables that satisfy the backdoor criterion



3. 분석 패턴 – Ncsoft, Volvo

들어가기 전에

NCSoft, <게임 플레이어는 좋은 아이템을 획득하면 게임을 더 열심히 하게 될까?>

- 도메인 지식에 기반한 분석 설계
- 인과 그래프 모델링
- 회귀 모델을 활용한 인과 효과 분석

Volvo, <ON THE USE OF CAUSAL GRAPHICAL MODELS FOR DESIGNING EXPERIMENTS IN THE AUTOMOTIVE DOMAIN>

- 그래프 인과 모델을 활용한 실험 설계

Introduction

게임을 (편안하게) 플레이 하는 데 좋은 등급의 아이템이 도움을 줌


> 좋은 등급의 아이템을 획득하면 게임을 열심히 플레이할까? 그 반대는 아닐까?

실험을 위해 통제군에게는 좋은 보상을 안 줄 수는 없음



분석 과정

1. 변수 정의
 - 원인 변수와 결과 변수 정의
 - 통제 변수 후보 정의
2. 분석 기간 선정
 - 재현성 확인 : 여러 기간에 대해 동일한 분석 진행
3. 대상 선정
 - 역인과 관계 방지를 위한 조건 설정
 - 일자별 유저 매칭
4. 세부 그룹
 - 유저 특징에 따른 segment 설정
5. 인과 다이어그램을 통한 통제 변수 구성
6. 인과 효과 추정
 - 모델 선정 및 적합
 - 결과 분석



분석 설계



분석

변수 정의하기

'게임을 열심히 하게 됨'이라는 추상적인 의미를 측정 가능한 형태로 정의

결과 변수

- 게임을 열심히 하게 됨을 반영하기 위해 전후 변화량을 살펴봄
 - 1) 게임에 들이는 시간과 자본이 많다
 - 30일 전후 접속 일수 변화
 - 30일 전후 결제 금액 변화
 - 2) 게임에서 상위 콘텐츠를 즐긴다(ex. 전투)
 - 4주 후 유저 그룹 지표 상향 여부

통제 변수 후보

결과 변수에 영향을 줄 것으로 예상되는 변수를 모으기

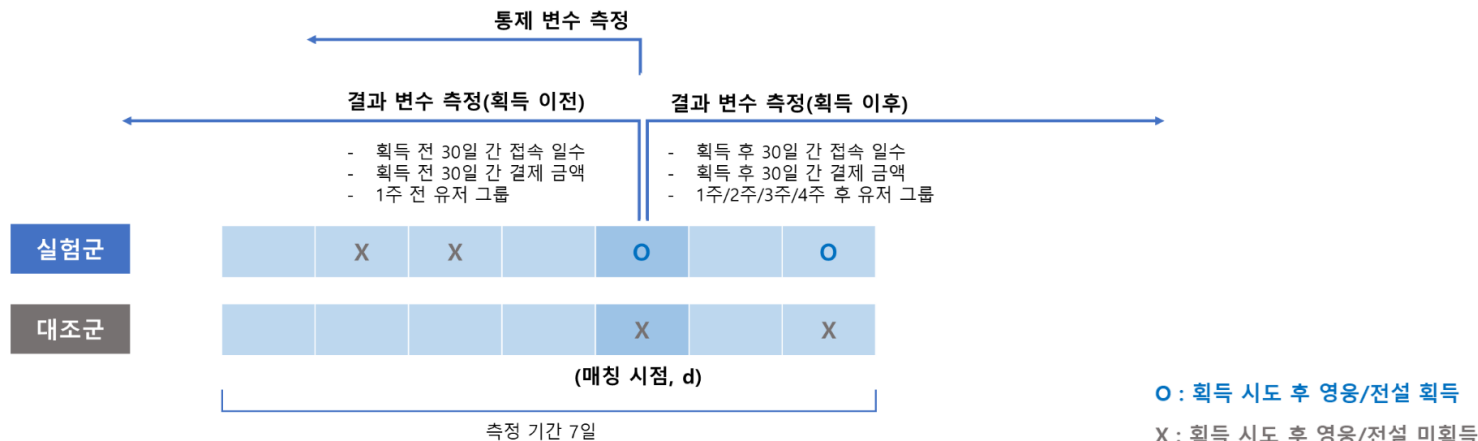
	no	변수명	영향을 받을 것으로 예상되는 결과 변수
성장	1	희귀 등급 수	접속 일수, 결제금액
	2	컬렉션 완성 횟수	접속 일수
	3	스킬 습득 횟수	접속 일수
	4	인챈트 성공 횟수	접속 일수
	5	등급 합성 성공 횟수	결제 금액
정체	6	사망 횟수	접속 일수, 결제 금액
매몰비용	7	계정 생성 후 경과일	접속 일수
	8	레벨	접속 일수
	9	총 결제액	접속 일수, 결제금액

대상 선정

역인과 관계를 방지하고 관찰 대상의 변수 집계 기간을 일치시키기 위한 일자별 유저 매칭 (not matching)

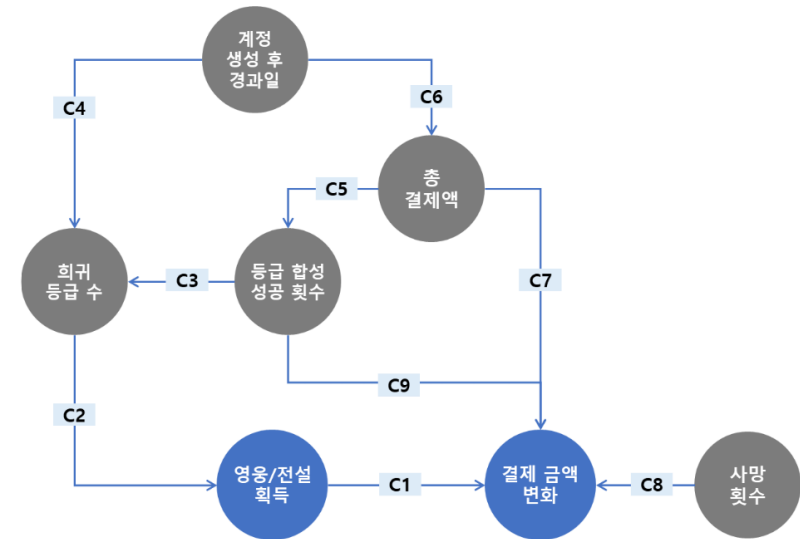
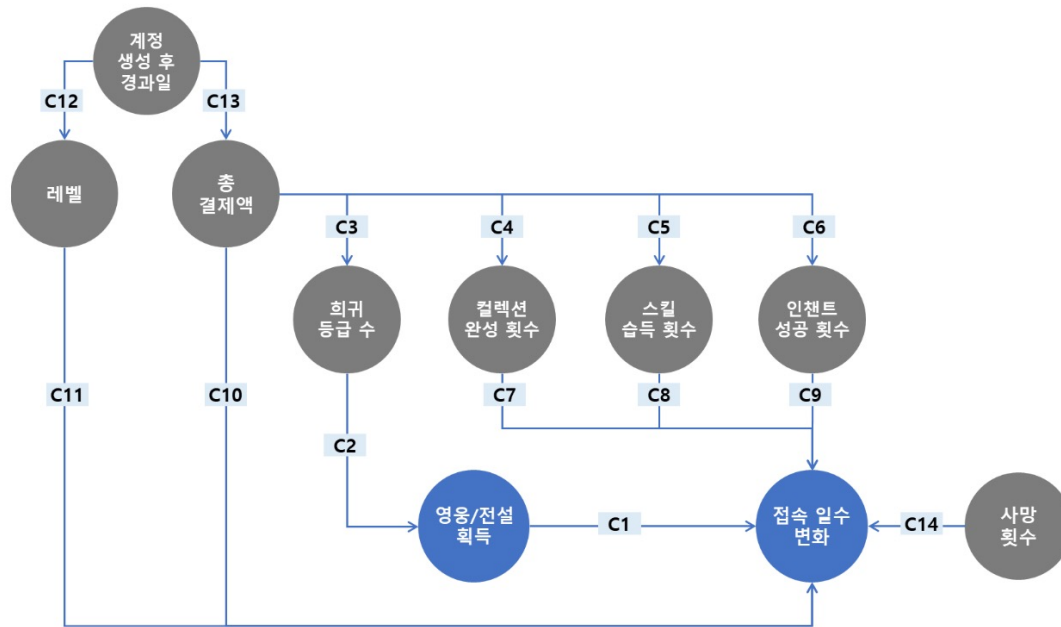
3. 대상 선정

- 측정 기간 이전에 영웅/전설 획득 이력이 없는 유저 중 측정 기간 동안 접속 이력이 있는 유저
 - 역인과 관계 방지 : 미접속으로 인한 영웅/전설 미획득(대조군)의 경우 방지
- 변수 집계 기간 설정을 위해 일자별 유저 매칭
 - 실험군(d시점) : 측정 기간 7일 간 영웅/전설 획득 유저 중 최초 획득을 d시점에 한 유저
 - 대조군(d시점) : 측정 기간 7일 간 영웅/전설 미획득 유저 중 최초 획득 시도를 d시점에 한 유저

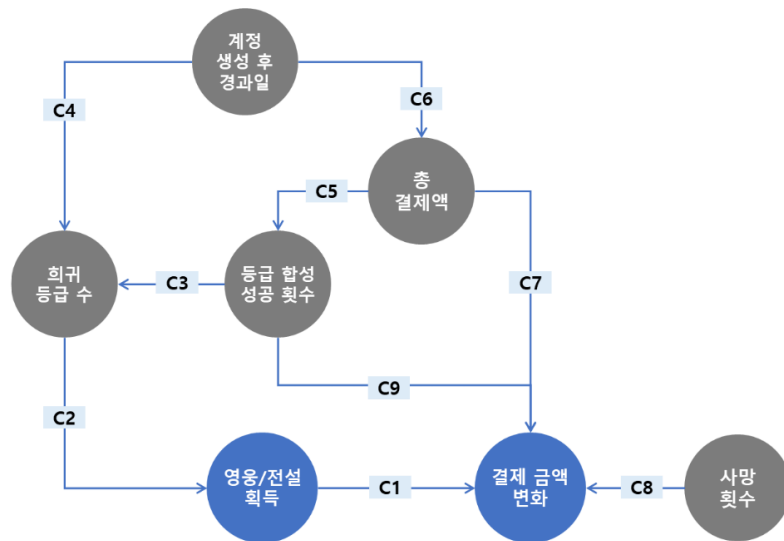


인과 그래프 모델링

원인 변수, 결과 변수, 통제 변수 후보들로 인과 그래프 모델링하기



인과 그래프 모델링과 통제 변수 구성



- step1) 원인 변수, 결과 변수에 직접적인 영향을 주지 않는 인과 관계 무시를 위한 노드 통제
- step2) 원인 변수 외에 결과 변수에 영향을 주는 노드 통제
- step3) 원인 변수, 결과 변수에 동시에 영향을 받는 노드 (collider)
- step4) 최종 통제 변수 선정
 - step1과 step2에서 선정된 전체 통제 변수들 중 step3에 포함되는 변수는 제외

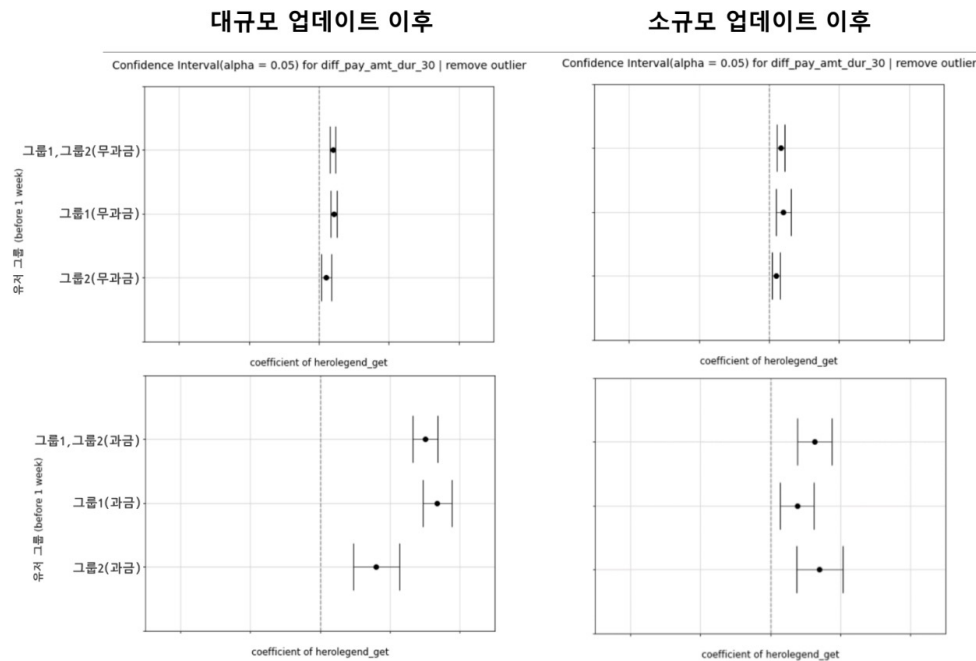
→ 결제 금액 변화 ~ 영웅/전설 획득 여부 + 희귀 등급 수 + 등급 합성 성공횟수 + 총 결제액 + 사망 횟수

인과 분석

신뢰 구간에 0인 값 포함 시 – 인과 관계 없음

신뢰 구간의 위치가 0 오른쪽 – 결과 변수 값 증가에 영향을 줌

2. 결제 금액 변화에 미치는 인과 효과 추정



- 모든 유저 그룹에서 최초 영웅/전설 획득은 30일 간 결제 금액 증가에 유의한 영향을 미침
 - 과금 여부에 따라 영향의 크기 차이가 존재
 - 과금 그룹 > 무과금 그룹
- 영웅/전설 획득은 유저의 결제 증가에 영향을 미치며, 과금 그룹에서 효과가 더 크게 나타난다

X축 범위 : 과금 그룹이 무과금 그룹에 비해 5배 넓음
결제 금액 증가 값에 대해 이상치 제거 후 인과 효과 추정

Introduction

T : 공기 조절 시스템을 업데이트하면,
Y : 차량의 에너지 소비가 감소할까?

Support / Select your car / C40 Recharge / Software updates

Software updates

Updating your car's software ensures that it remains up to date with the latest features and improvements.

Software updates are primarily done over the air (OTA) and all previous updates are included when you install one. Keep in mind that you may be charged for installing software updates at a Volvo dealer, unless installation is included in a repair or service agreement.

When a new update is available, you'll get a notification in the centre display and in the Volvo Cars app showing estimated time for installation and what's included in the update.

Functionality after updating may vary depending on market, model, model year and options.

Updates in software 3.1.2

Software release date: 13 May 2024. This update is not available over the air (OTA).

Improvements and bug fixes

- Reduced low frequency vibrations when driving at low speeds with light acceleration. This is only relevant in model year 2024 cars.
- Activating air circulation manually now also activates air conditioning to avoid condensation on the windscreen.
- Drive system and power supply adjustments that do not affect performance or other certified properties.

Backgrounds

온라인 통제 실험이 어려운 조건에서 인과 효과를 분석하기 위한 방법 찾기

- Randomised field experiments, such as A/B testing, have long been the gold standard for evaluating the value that new software brings to customers. However, running randomised field experiments is not always desired, possible or even ethical in the development of automotive embedded software. In the face of such restrictions, we propose the use of the Bayesian propensity score matching technique for causal inference of observational studies in the automotive domain. In this paper, we present a method based on the Bayesian propensity score matching framework, applied in the unique setting of automotive software engineering.

Bayesian propensity score matching in automotive embedded software engineering

Yuchu Liu
Volvo Cars
Gothenburg, Sweden
yuchu.liu@volvocars.com

David Issa Mattos
Computer Science and Engineering
Chalmers University of Technology
Gothenburg, Sweden
davidis@chalmers.se

Jan Bosch
Computer Science and Engineering
Chalmers University of Technology
Gothenburg, Sweden
jan.bosch@chalmers.se

Helena Holmström Olsson
Computer Science and Media Technology
Malmö University
Malmö, Sweden
helena.holmstrom.olsson@mau.se

Jonn Lantz
Volvo Cars
Gothenburg, Sweden
jonn.lantz@volvocars.com

분석 과정

1. Value mapping
 - 원인 변수, 결과 변수, 가드레일, 교란 요인 등이 될 수 있는 변수 후보 식별
2. Causal mapping
 - 각 변수가 가드레일, 교란 요인인지, 관측 가능한지 식별
 - 그래프 인과 모델 만들기
3. Causal mapping validation
 - 도메인 전문가의 검토
 - 그래프 인과 모델 평가
 - 조건부 교환 가능성 검증
4. Experimental design and validity
 - 통제 변수 식별하기
5. Analysis
 - 인과 효과 분석

Causal Mapping Validation

Acyclic을 충족하는지 점검

Creating a DAG

Here we model all causal relations

```
climate_dag <- dagify(  
  vehicle_variant ~ city,  
  climate_sw ~ vehicle_variant + rnd,  
  energy ~ city + vehicle_variant + climate_sw + temperature_setting,  
  temperature_setting ~ climate_performance + climate_preferences,  
  climate_performance ~ climate_sw + climate_preferences,  
  climate_preferences ~ city)
```

Double check that the DAG is acyclic

```
dagitty::isAcyclic(climate_dag)
```

```
## [1] TRUE
```


Causal Mapping Validation

조건부 독립성 검증

Checking the conditional independences

Here we check what relations should be independent. Note that we can only check with data the conditional independence of the variables that are observed.

```
dagitty::impliedConditionalIndependencies(climate_dag, max.results = 200)
```

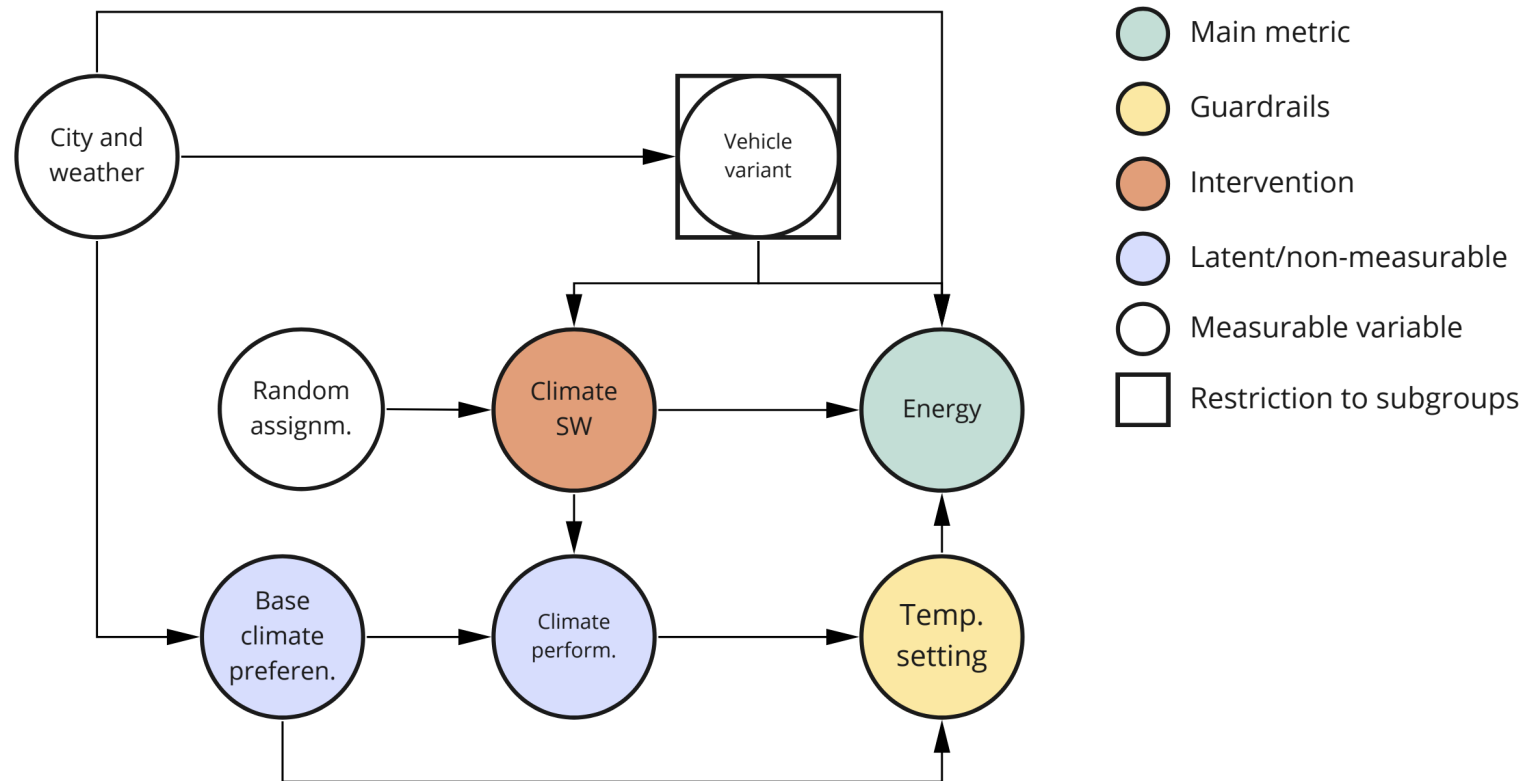
```
## city _||_ climate_performance | climate_preferences, climate_sw  
## city _||_ climate_performance | climate_preferences, vehicle_variant  
## city _||_ climate_sw | vehicle_variant
```

We can see some common conditional independence relations:

- (energy || rnd | climate_sw, vehicle_variant) is equivalent to an A/A test on the main metric
- (temperature_setting || vehicle_variant | city, climate_sw) is equivalent to the A/A test on the guardrail metric
- (rnd || vehicle_variant) is equivalent to SRM in a the restricted groups
- (city || rnd) is equivalent to SRM in one of the strata variables

The actual SRM criteria on the treatment assignment is for checking that the arrow between randomization and the treatment is actually as designed (since this is an artificially introduced variable on the DAG).

DAG



Experimental design and validity

통제 변수 식별하기

Direct and Indirect effects

Let's compute the direct effect

```
adjustmentSets(climate_dag,  
               effect = 'direct',  
               exposure = 'climate_sw',  
               outcome = 'energy')
```

```
## { city, temperature_setting, vehicle_variant }  
## { climate_preferences, temperature_setting, vehicle_variant }  
## { climate_performance, climate_preferences, vehicle_variant }
```

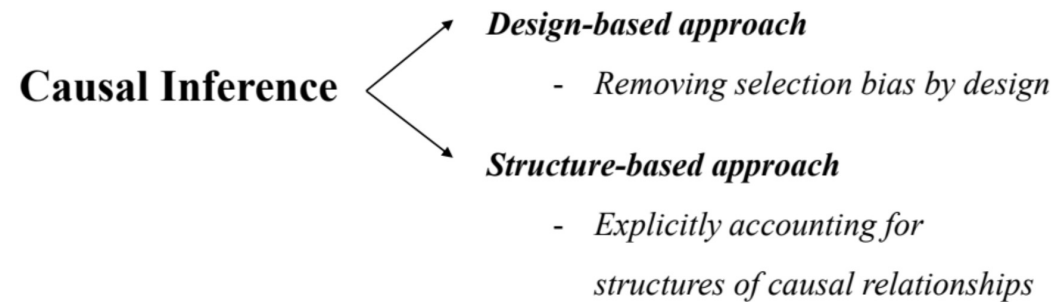
We can use only the first adjustment set, since the others contain non-observable or latent variable.

Assuming linearity in the causal effect, the indirect effect is:

$$\text{Total} = \text{Direct} + \text{Indirect}$$

4. Wrap Up

2 Approaches of Causal Inference



Potential Outcomes Framework (Donald Rubin)			Structural Causal Model (Judea Pearl)
Table 1.1. Example of Potential Outcomes and Causal Effect with One Unit			
Unit	Potential Outcomes		Causal Effect
	$Y(\text{Aspirin})$	$Y(\text{No Aspirin})$	
You	No Headache	Headache	Improvement due to Aspirin


```

graph LR
    Drug --> BP[Blood Pressure]
    Drug --> HA[Heart Attack]
    BP --> HA
  
```

Imbens, G.W., 2020. Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics. *Journal of Economic Literature*, 58(4), pp.1129-79.

그래프 인과모델을 활용한 분석 패턴

1. 문제 정의하기
2. 변수 탐색하기
 - 원인 변수, 결과 변수를 측정 가능한 형태로 정의하기
 - 통제 변수 후보 탐색하기
3. 인과 그래프 모델링
 - 통제 변수 식별하기
4. 인과 그래프 검증하기
 - 도메인 전문가의 검토
 - 조건부 교환 가능성 식별
5. 데이터 수집
6. 인과 분석
 - 인과 관계 평가
 - 인과 효과 추정

그래프 인과모델을 적용할 때의 고민

도메인 지식의 중요성

- 원인 변수와 결과 변수와 이어진 변수는 어떤 것들이 있을까?
- 변수 간 어떤 관계가 흐를까?

데이터 모으기

- 어떤 데이터를 관측할 수 있을까?
- 필요한 형태로 데이터를 얻을 수 있을까?
- 다양한 통제 변수의 조합에도 모델을 적합할 수 있는 충분한 데이터를 얻을 수 있을까?

감사합니다
Q&A