



Causal-Lab
가계연구소

5th PseudoCon : 함께할 의지

Casual하지 않게 Causality 이해하기

CONTENTS

1. 인과추론 스터디 소개

2. 인과추론 소개 & 사례

3. 인과추론을 통해 느낀점

[Prequel]



4th PseudoCon : 함께하는 힘

Casual하게 Causality 이해하기



Casual하게 Causality 이해하기

₩ 1개의 비용크

Goal

1. Causal Inference (인과추론, 분석)에 대한 기본적인 개념 학습
2. 문제 or ML 모델에 대한 원인 분석 능력 키우기
3. 풍부한 내용을 정리해, 인과추론에 대한 한국어 자료 만들어 보기

Greeting

- 가짜연구소 소개
- 팀원 소개
- 출석 체크



1. 인과추론 스터디 소개

- Causal Inference 팀 소개
 - 스터디 기간 : 22.03.12 ~ 22.11.26 (8개월)
 - 빌더 : 신진수 (KRAFTON)
 - 러너 : 김상돈 (서울시립대)
김성수 (연세대)
김소희 (TVING)
김준영 (Columbia Uni.)
남궁민상 (KAIST)
정호재 (롯데캐피탈)
최은희 (NEXON KOREA)
홍성철 (서울시립대)

[스터디 Notion 페이지]



Casual하지 않게 Causality 이해하기

태그 Causal Inference Data Science Decision making ML
가수 571
담당 빌더 Jinsoo Shin
날짜 2022년 8월 6일 ~ 2022년 11월 26일
속성 비어 있음
발대 연혁 비어 있음
대표 링크 <https://pseudolabcausal.inference.tistory.com/>
팀 인원 9

[스터디 목표]



1. Causal Inference 실습에 대한 한국어 자료 만들기

2. Causal Inference에 대한 개인 TASK를 설정 후, TASK 결과를 팀 블로그에 정리

3. 스터디에서의 의견 공유를 통해, 주 1회 1시간씩 깊은 고민을 해보는 습관 만들기

1. 인과추론 스터디 소개

- 스터디 방향성

[4기 : **Casual하게** Causality 이해하기]



마법책을 통해서 **주문을 배웠**다면



[5기 : **Casual하지 않게** Causality 이해하기]

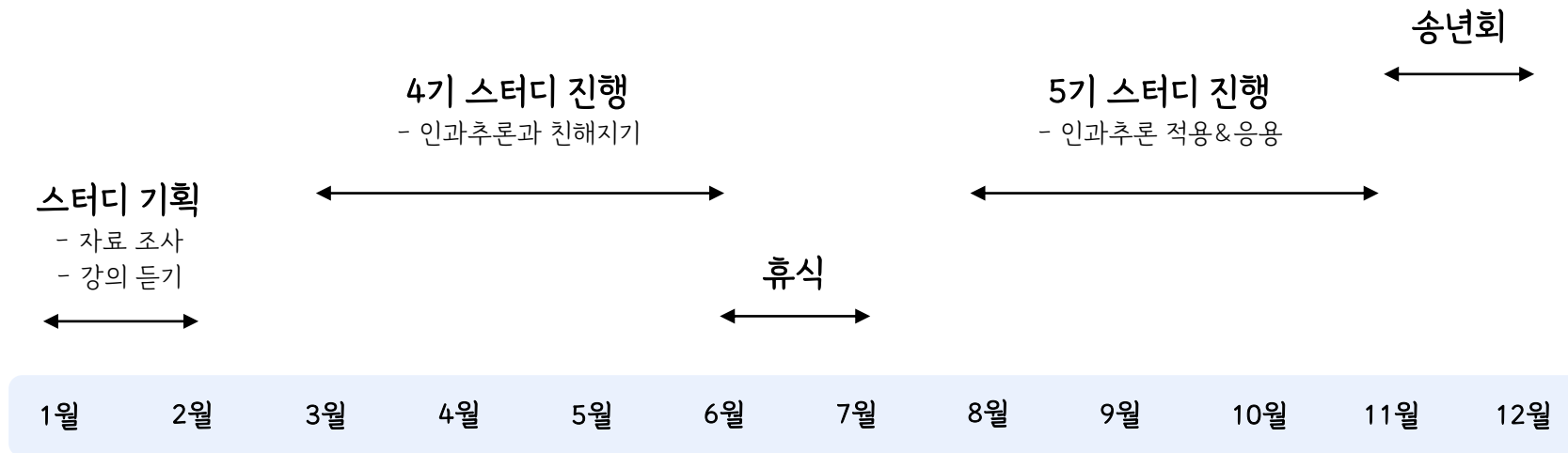


이제는 직접 **주문을 써봐야** 합니다!

1. 인과추론 스터디 소개



- 2022 Study Roadmap



1. 인과추론 스테디 소개

- Achievement - 한국어 자료

[인과추론팀 Github]



☰ README.md

Python으로 하는 인과추론 : 개념부터 실습까지

Causal Inference for The Brave and True 책 번역

안녕하세요. 가짜연구소 Causal Inference 팀입니다.

이 책은 Matheus Fature (Nubank Data Scientist)의 [Causal Inference for The Brave and True](#)을 한국어로 번역한 자료입니다. 이 책은 인과추론에 대한 기본 개념과 Python 실습, 나아가 최신 사례까지 모두 다루고 있습니다. 인과추론에 대한 이론과 실습자료가 많지 않은 상황에서 이 책은 인과추론 처음 접하거나 인과추론에 대한 이해를 바탕으로 사례를 접하고 싶은 분 모두를 만족시킬 수 있는 책이라고 생각합니다. 한국에서 인과추론을 공부하시는 분들께 많은 도움이 되었으면 좋겠습니다.

가짜연구소 Causal Inference 팀은 전문 번역가가 아닌, 데이터분석가/과학자/연구자로 구성 되어 있습니다. 따라서, 한국어 번역 과정에서 이상하거나 어색한 부분이 발생할 수 있어요. 해당 부분에 대한 여러분의 피드백을 부탁드립니다. 또한, 한국어 표현이 익숙하지 않은 부분에 대해서는 영어로 그대로 표현했습니다. 해당 부분 이해부탁드려요.

마지막으로, 이 책의 한국어 번역에 동의해 주신 Matheus Fature에게 감사의 말씀을 전합니다.

그리고, 번역과 실습자료를 만들기 위해 같이 고생해준 Causal Inference 팀 멤버들에게도 진심으로 감사드립니다!

[실습 및 번역에 참고한 책]

Causal Inference for The Brave and True



DOI: 10.5281/zenodo.4445778

A light-hearted yet rigorous approach to learning impact estimation and sensitivity analysis. Everything in Python and with as many memes as I could find.

[Check out the book here!](#)

If you want to read the book in Chinese, @xieliang was very kind to make a translation:

[因果推断：从概念到实践](#)

If you want to read the book in Spanish, @doneliang was very kind to make a translation:

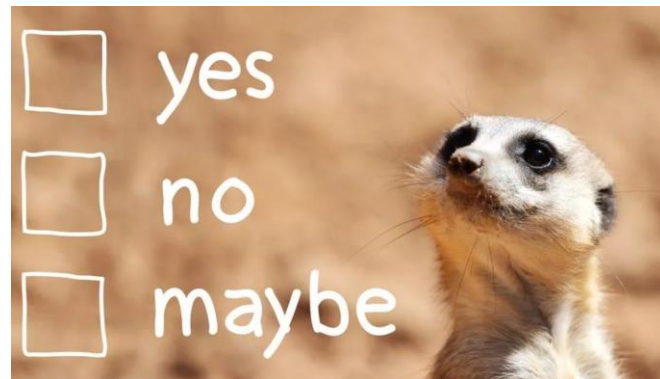
[Inferencia Causal para los Valientes y Verdaderos](#)

If you want to read it in Korean, @jsshin2019 has put up a team to make the that translation possible:
[Python으로 하는 인과추론 : 개념부터 실습까지](#)

Github Star 부탁드립니다!

2. 인과추론 소개 & 사례

- 여러분은 혹시 인과추론에 대해 알고 계신가요?
 - 알고보면, 우리가 일상 생활에서 마주하는 문제를 다루는 방법입니다.
 - 축구를 보면, 항상 우리팀이 지는 것 같아... ☹



2. 인과추론 소개 & 사례

- 인과추론 이란?
 - 문제에 대한 원인을 찾고 해당 원인에 대한 효과를 추론하는 것



2. 인과추론 소개 & 사례

- 인과추론이란?
 - 문제에 대한 원인을 찾고 해당 원인에 대한 효과를 추론하는 것
 - 즉, **Treatment**를 주었을 때, 이에 따른 **Outcome**이 어떻게 변하는지를 추정
 - 게임에서 이벤트를 하면, 유저들의 접속이 늘어날까?
(Treatment) (Outcome)



2. 인과추론 소개 & 사례

- 인과추론 이란?
 - 문제에 대한 원인을 찾고 해당 원인에 대한 효과를 추론하는 것
 - 즉, **Treatment**를 주었을 때, 이에 따른 **Outcome**이 어떻게 변하는지를 추정
 - 게임에서 이벤트를 하면, 유저들의 접속이 늘어날까?
(Treatment) (Outcome)
- ➡ 과연 이벤트 때문일까? 원래 잘 접속하는 유저가 아닐까?
- ➡ 단순히, 이벤트와 유저 접속간의 **연관성**이 있는건 아닐까?



2. 인과추론 소개 & 사례

- 상관관계는 인과관계를 의미하지 않는다!
 - 연간 니콜라스 케이지의 영화 출현 횟수와 연간 익사 사망건수는 높은 상관관계를 보임
 - 서로 연관성이 없는 변수가 높은 상관관계를 보이는 경우가 존재 (Spurious Correlation)

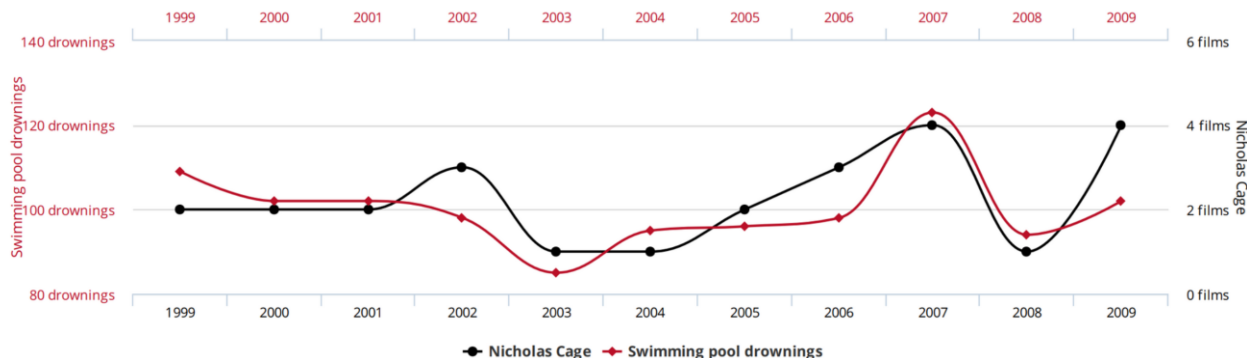


Figure 1.3: The yearly number of movies Nicolas Cage appears in correlates with the yearly number of pool drownings [1].

2. 인과추론 소개 & 사례

- 상관관계는 인과관계를 의미하지 않는다!
 - 연간 니콜라스 케이지의 영화 출현 횟수와 연간 익사 사망건수는 높은 상관관계를 보임
 - 서로 연관성이 없는 변수가 높은 상관관계를 보이는 경우가 존재 (Spurious Correlation)

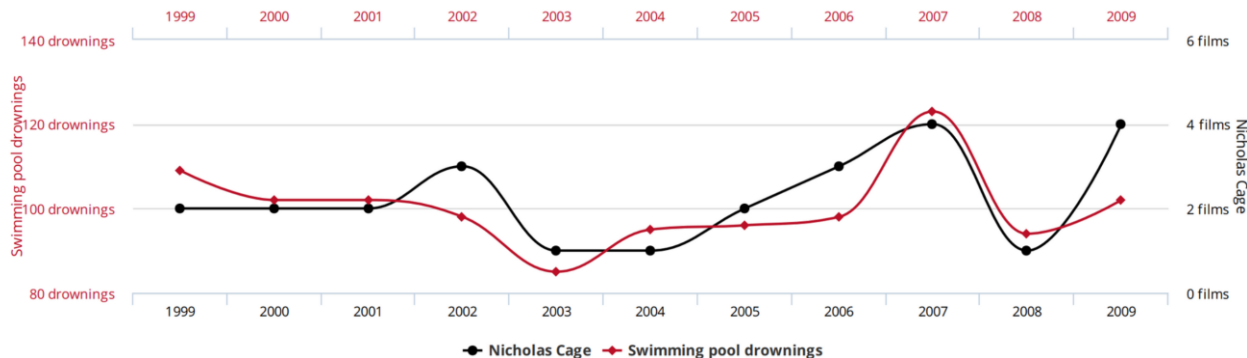


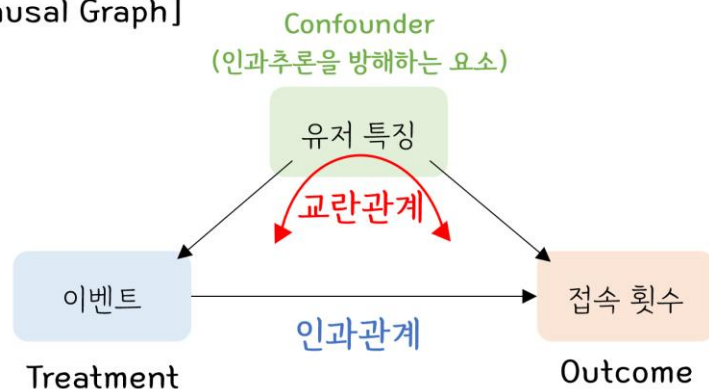
Figure 1.3: The yearly number of movies Nicolas Cage appears in correlates with the yearly number of pool drownings [1].

➡ 잘못된 의사결정을 유도!

2. 인과추론 소개 & 사례

- 다시 사례로 돌아와서 ..
 - 게임에서 이벤트를 하면, 유저들의 접속이 늘어날까?
(Treatment) (Outcome)
 - 과연 이벤트 때문일까? 원래 잘 접속하는 유저 일수도...
(Confounder)
 - 단순히, 이벤트와 유저 접속간의 연관성이 있는건 아닐까?

[Causal Graph]



2. 인과추론 소개 & 사례

- 가설설정 & 문제 정의

- 게임에서 이벤트를 하면, 유저들의 접속이 늘어날까?
(Treatment) (Outcome)

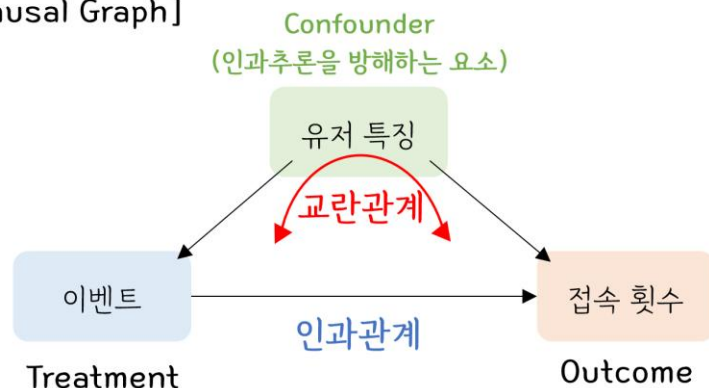
- 과연 이벤트 때문일까? 원래 잘 접속하는 유저 일수도...
(Confounder)

- **Confounder**로 인해 제대로된 인과효과 추정이 어려워짐

➡ 연관성 (상관관계) = **교란관계** + **인과관계**

➡ 인과추론에 방해되는 bias를 최소화! 실험설계를 잘해보자!

[Causal Graph]



2. 인과추론 소개 & 사례

- 실험설계

- Apple to apple : 비교 대상 (실험군, 대조군)은 최대한 유사하게 설계!

T = 1 (이벤트 참여 유저)



T = 0 (이벤트 미참여 유저)



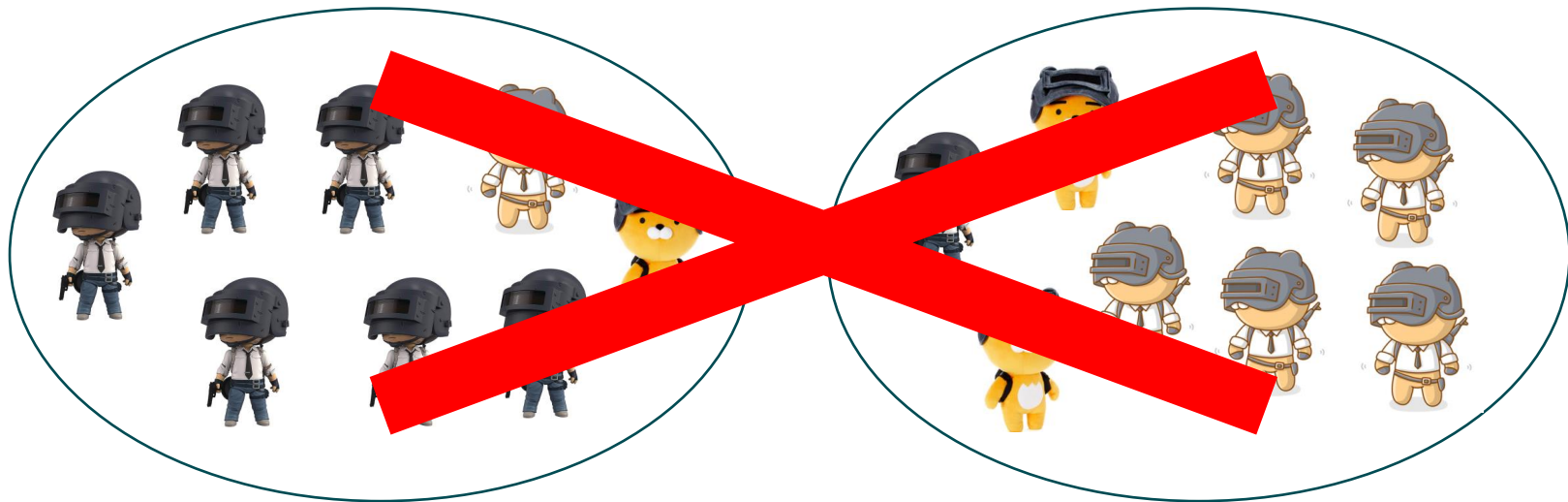
2. 인과추론 소개 & 사례

- 실험설계

- Apple to apple : 비교 대상 (실험군, 대조군)은 최대한 유사하게 설계!

T = 1 (이벤트 참여 유저)

T = 0 (이벤트 미참여 유저)



2. 인과추론 소개 & 사례

- 실험설계

- Apple to apple : 비교 대상 (실험군, 대조군)은 최대한 유사하게 설계!

T = 1 (이벤트 참여 유저)



T = 0 (이벤트 미참여 유저)



2. 인과추론 소개 & 사례

- 실험설계

- Counterfactuals : 타임머신이 있다면, 이벤트 시점으로 돌아가서 동일한 유저에게 실험을 해볼 수 있을텐데...



현재



타임머신타고
돌아온 현재

➡ 비교가능함!



2. 인과추론 소개 & 사례

- 실험설계
 - Apple to apple : 비교 대상 (실험군, 대조군)은 최대한 유사하게 설계!
 - Counterfactuals : 타임머신이 있다면, 이벤트 시점으로 돌아가서 동일한 유저에게 실험을 해볼 수 있을텐데...



인과추론의 근본적인 문제!

2. 인과추론 소개 & 사례

- 실험설계

- Apple to apple : 비교 대상 (실험군, 대조군)은 최대한 유사하게 설계!
- Counterfactuals : 타임머신이 있다면, 이벤트 시점으로 돌아가서 동일한 유저에게 실험을 해볼 수 있을텐데...



대조군이 최대한 Counterfactuals과 비슷해야 한다!
(Selection Bias 해결)



대조군



타임머신타고
돌아온 현재



비슷하게 구성

2. 인과추론 소개 & 사례

- 인과효과 추정

- Random Assignment (RCT) : 실험군과 대조군에 랜덤하게 할당 (동전 던지기로 팀 나누기)
- Research Design : RCT 처럼 실험 환경을 구성할 수는 없지만, 실험 디자인을 통해 간접적으로 구성

Potential Outcome Framework vs. Structural Causal Model

What if...? 인과추론의 데이터과학

	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. 인과추론 소개 & 사례

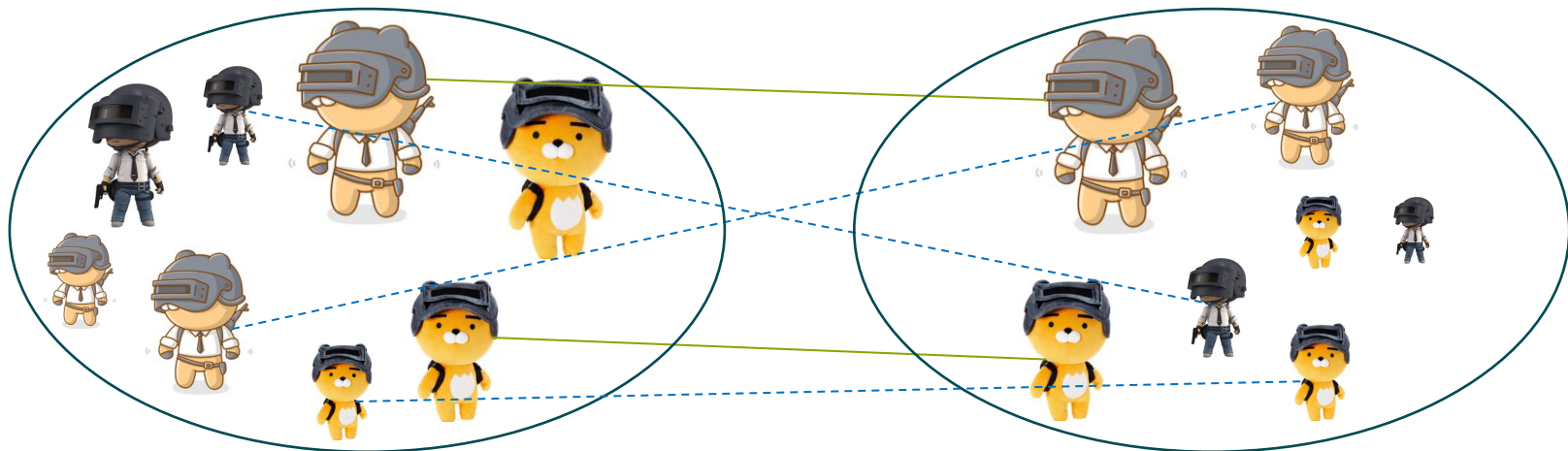
- 인과효과 추정

- Matching : 실험군과 대조군에서 유사한 대상들만 비교하는 방법

'비슷함'이 무엇을 의미하는지는 실험설계에 따라 다양한 방법으로 결정!

T = 1 (이벤트 참여 유저)

T = 0 (이벤트 미참여 유저)



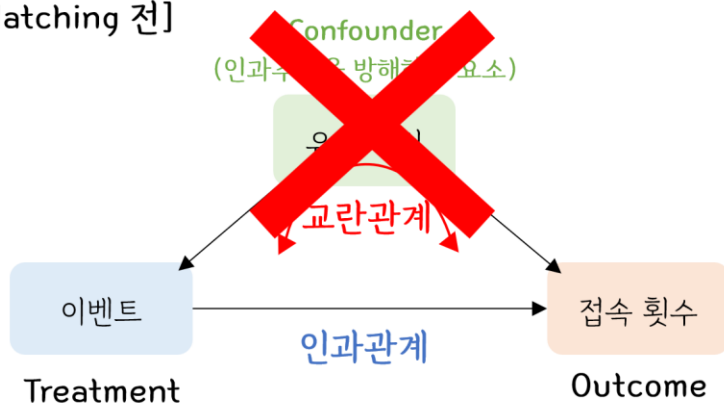
2. 인과추론 소개 & 사례

- 인과효과 추정

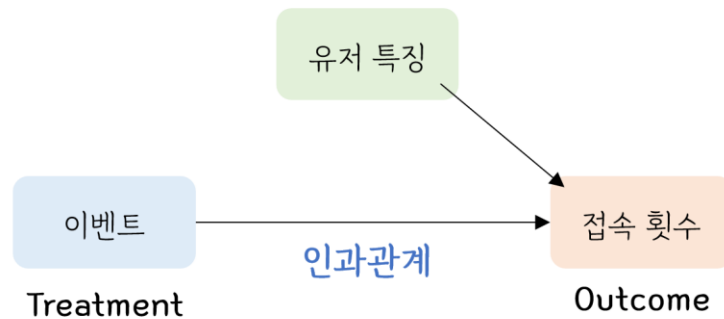
- Matching : 실험군과 대조군에서 유사한 대상들만 비교하는 방법

'비슷함'이 무엇을 의미하는지는 실험설계에 따라 다양한 방법으로 결정!

[Matching 전]



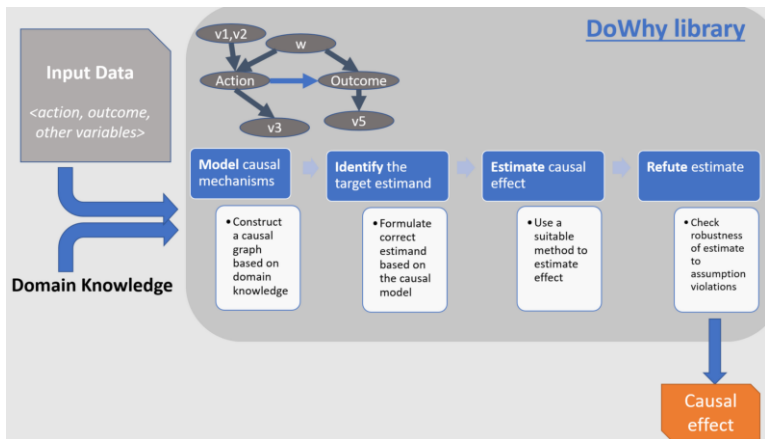
[Matching 후]



2. 인과추론 소개 & 사례

- 인과추론 적용을 위한 Flow
 - 문제 Scope 정의
 - 문제 상황 파악 & 문제를 해결을 위한 가설설정
 - 어떤 지표를 바라볼 것인가?
 - 데이터 수집 (Data generating process)
 - 로그를 통해 필요한 변수를 확인할 수 있는가?
 - 실험설계 및 인과효과 추정
 - A/B 테스트가 가능한 실험환경인가?
 - 불가능하다면 어떤 방법을 적용해야 하는가?
 - 어떤 변수를 활용/통제해서, Causal Model을 만들것인가?
 - 결과에 대한 비판적인 해석 & 합리적인 의심

[Microsoft 인과추론 패키지 : DoWhy]



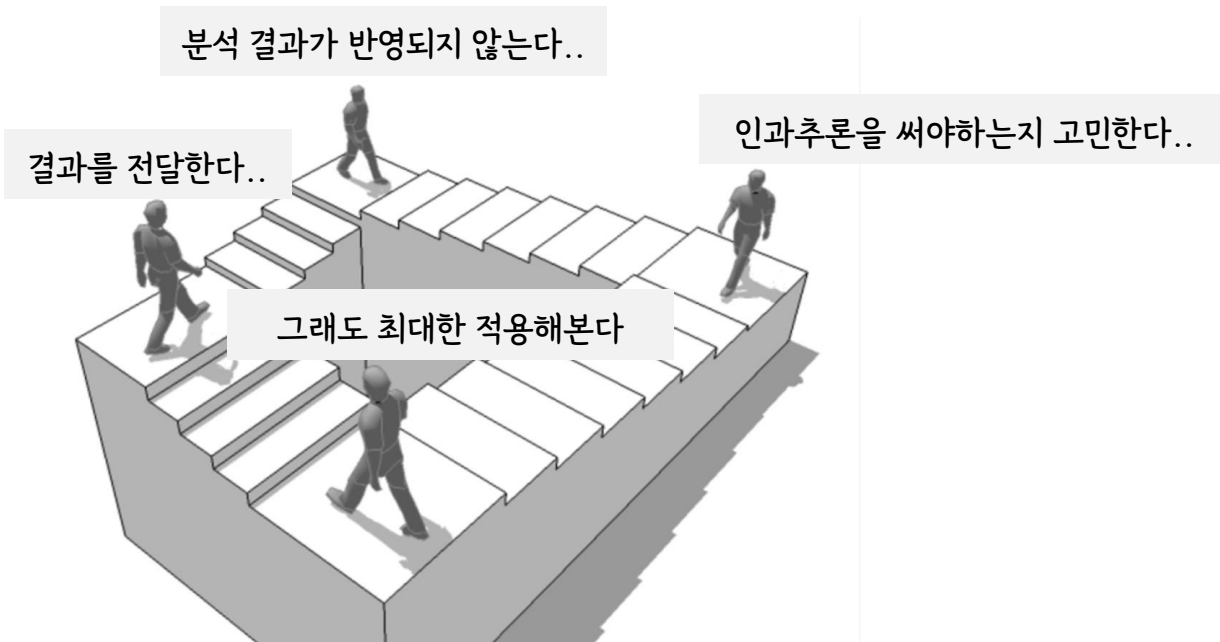
2. 인과추론 소개 & 사례

- 인과추론 어떠신가요?
 - 어 려 워 요



3. 인과추론을 통해 느끼는 점

- 인과추론은 왜 어려울까? (빙빙 돌아가는 회전목마처럼..)



3. 인과추론을 통해 느낀점

- 인과추론은 왜 어려울까? (빙빙 돌아가는 회전목마처럼..)



3. 인과추론을 통해 느낀점

- 인과추론은 왜 어려울까? (Feat. Teammates)



배움의 필요성

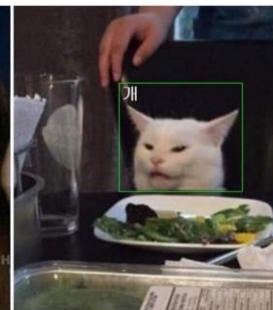


커뮤니케이션

AI가 세계를 지배할 거라는
AI 알못들:



내가 만든 AI:



괴리감

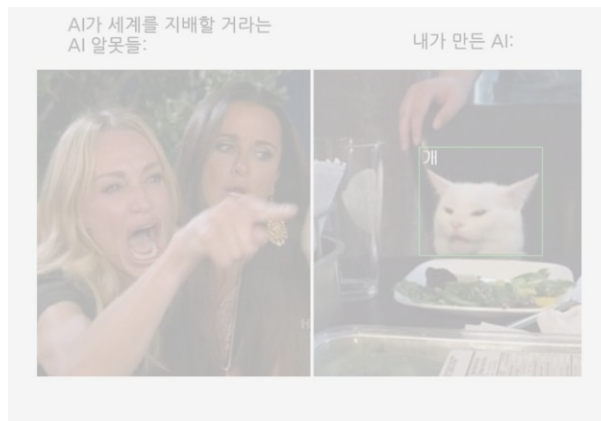
3. 인과추론을 통해 느낀점

- 인과추론은 왜 어려울까? (Feat. Teammates)
 - 당위성 : 알아두면 좋지만 항상 적용할 수 있는 상황 or 환경이 아님
 - 접근성 : 한국어 자료가 많지 않고 접근하기 쉬운 분야는 아님



3. 인과추론을 통해 느낀점

- 인과추론은 왜 어려울까? (Feat. Teammates)
 - 커뮤니케이션 : 결국 결과를 전달하는 입장에서 쉽게 전달해야 하는데...
 - Domain 지식 : 게임을 잘 이해하고 있어야, 의사결정 지원이 가능



3. 인과추론을 통해 느낀점

- 인과추론은 왜 어려울까? (Feat. Teammates)
 - Discrepancy : 이론과 실제의 큰 괴리
 - Demotivation : 의사결정에 바로 반영이 될 수 있을까?



AI가 세계를 지배할 거라는
AI 알못들:



내가 만든 AI:



3. 인과추론을 통해 느낀점

- 그런데, 왜 인과추론은 필요할까?

◦ 데이터 분석가/과학자 역할 : 실제 비즈니스 환경에서, 가장 근본적인 원인에 대한 질문을 많이 받음

[가장 많이 받은 질문]



[질문을 받은 나의 모습]



3. 인과추론을 통해 느낀점

- 그런데, 왜 인과추론은 필요할까?



◦ 유저 개인화 : 결국, 원인을 찾아 유저의 경험 개선 & 유저 행동 변화를 만드는 것이 목적



자료사진/사진=쿠팡



Summary

◦ 인과추론팀 Github 페이지 Star를 꼭 누르자!

◦ 상관관계와 인과관계

◦ Treatment / Outcome / Confounders

◦ 인과추론의 근본적인 문제 / Random Assignment / Research Design

◦ Selection Bias

☰ README.md

Python으로 하는 인과추론 : 개념부터 실습까지

Causal Inference for The Brave and True 책 번역

안녕하세요. 가짜연구소 Causal Inference 팀입니다.

이 책은 Matheus Fature (Nubank Data Scientist)의 [Causal Inference for The Brave and True](#)을 한국어로 번역한 자료입니다. 이 책은 인과추론에 대한 기본 개념과 Python 실습, 나아가 최신 사례까지 모두 다루고 있습니다. 인과추론에 대한 이론과 실습자료가 많지 않은 상황에서 이 책은 인과추론 처음 접하거나 인과추론에 대한 이해를 바탕으로 사례를 접하고 싶은 분 모두를 만족시킬 수 있는 책이라고 생각합니다. 한국에서 인과추론을 공부하시는 분들께 많은 도움이 되었으면 좋겠습니다.

가짜연구소 Causal Inference 팀은 전문 번역가가 아닌, 데이터분석가/과학자/연구자로 구성 되어 있습니다. 따라서, 한국어 번역 과정에서 이상하거나 어색한 부분이 발생할 수 있어요. 해당 부분에 대한 여러분의 피드백을 부탁드립니다. 또한, 한국어 표현이 익숙하지 않은 부분에 대해서는 영어로 그대로 표현했습니다. 해당 부분 이해부탁드려요.

마지막으로, 이 책의 한국어 번역에 동의해 주신 Matheus Fature에게 감사의 말씀을 전합니다.
그리고, 번역과 실습자료를 만들기 위해 같이 고생해준 Causal Inference 팀 멤버들에게도 진심으로 감사드립니다!

Q & A





Causal-Lab
가계연구실

THANK YOU