

인과추론과 실무 : Week1 OT

가짜연구소 인과추론팀

발표자 : 김성수

빌더 소개

안녕하세요,
이번 가짜연구소
인과추론과 실무의
빌더 **김성수** 입니다.



LG CNS Data Scientist
Enterprise Data 분석 Team 소속
Causal inference Lab 3-8기

From Lab To Life
배운 것을 어떻게 실무에 적용할지 고민하고 있음

관심 분야
MLOps & Machine Learning System &
Experiment Platform & Causal inference

프로젝트 개요

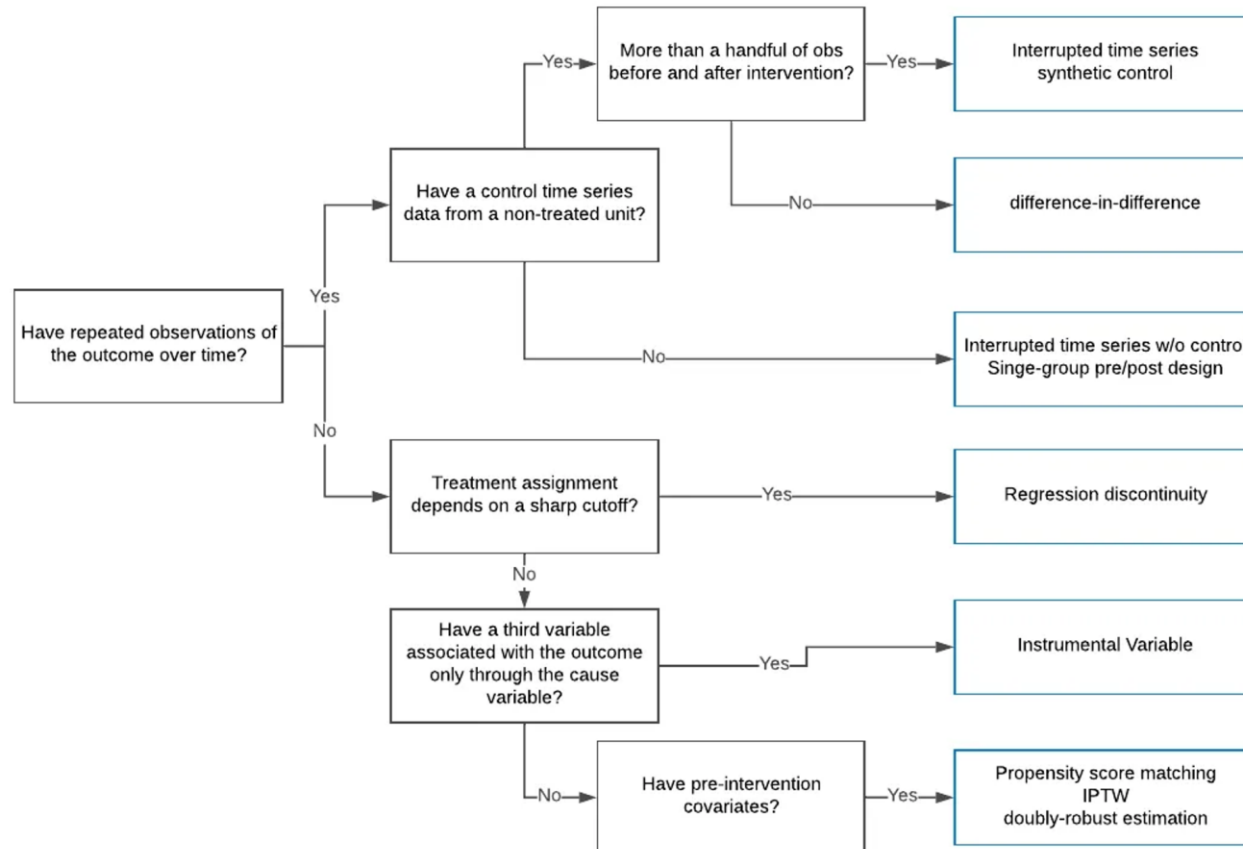
- 인과추론은 경제, 정치, 통계, 수학, 의학 등 다양한 도메인에서 활용되고 발전해옴
- 인과추론 방법론 및 프레임워크가 다양한 도메인에서 활용되고 있지만, 도메인별로 언어가 혼동되는 경우도 있음
Exchangeability = Ignobility = Exogeneity = Independence = Unconfoundedness
- 변수의 인과관계를 이해하자는 목표 자체는 동일하지만, 인과추론의 방법론은 다양한 도메인에 분산되어 있음
따라서 (1) 실무자는 현재 직면한 문제를 인과추론 문제로 변환하지 못하고, (2) 적절한 인과추론 방법을 선택하기 어려움

프로젝트 목표

- 인과추론의 실무 사례 템플릿化
- 실무를 위한 인과추론의 디자인 패턴 제시
 - 실제 실무 문제를 해결하는 방법을 패턴화
 - 반복적으로 발생하는 문제를 어떻게 해결할지에 대한 솔루션
- 기존 빅테크 기업의 인과추론 실무 적용 패턴 한글화 및 본인의 상황에 맞도록 변경

프로젝트 목표

- 문제 상황에 맞는 인과추론 방법론을 선택하여 적용할 수 있는 실무자
- 실무에서 문제 상황을 만났을 때 아래와 같은 인과추론 Flowchart가 머리속에서 바로 나올 수 있는 실무자



프로젝트 진행방식

- 기본 교재 : 실무로 통하는 인과추론 with 파이썬
- 학습 방식:
 1. 교재를 활용해 특정 방법론 및 실무 사례 학습
 2. 해당 방법론을 적용한 실무 사례 및 연구 검색 및 정리 (모든 관련 Docs 허용, 빅테크 기업 아티클, Paper)
 3. (중요) 사례를 정리하여 실무에서 어떠한 문제를 풀기 위해, 인과추론을 어떻게 적용했는지 정리하여 템플릿화
- 발표 방식: PPT & PDF

드래곤볼처럼 흩어져 있는, 실무 인과추론 사례를 모아서 나만의 인과추론 프레임워크를 만들자

O'REILLY*

실무로 통하는
인과추론 with 파이썬

Causal Inference in Python

데이터 분석에서 정책 수립까지,
이론과 사례 연구를 통한
실용적인 학습법



프로젝트 진행방식

추가 자료:

인과 추론 실무에 관한 모든 자료 (실무에 적용할 수 있는 Research Paper 격하게 환영)

1. <https://research.netflix.com/research-area/experimentation-and-causal-inference>
2. [KDD 2023 Workshop - Causal Inference and Machine Learning in Practice](#)
3. <https://www.nabe.com/tec2022>
4. <https://www.uber.com/en-KR/blog/causal-inference-at-uber/>
5. [PDAMS 23](#)

프로젝트 진행방식

- 시간: 수요일 오후 8시 – 합의 필요
- 오프라인 모임 : 월 1회 or 2개월에 한 번 – 합의 필요
- 결과물 : 가짜연구소 Causal Inference Team · GitHub 공유 – 양식 PDF (필수)

프로젝트 진행방식

Git Convention

- Causal-inference-practice GitHub 저장소를 만들 예정
- 스터디 중에 진행한 모든 토의 및 질문 사항 저장소에 정리
- 발표 자료와 레퍼런스 주소는 정리해서 본인이 올려주세요 (For your Contribution)
- 만들어지면 GitHub Star 눌러주세요!

예상 프로젝트 Output 참고

- 실무에서 발생하는 인과추론 적용 상황을 MECE하게 분류하여 그에 대한 실무 사례를 템플릿화

현업에서 마주치는 문제 상황의 분류

1. A/B 테스트가 불가능할 때

1. TripAdvisor 사례 (IV)
2. Roblox 사례 (IV)

2. A/B 테스트 기반 추가적인 인사이트 도출할 때

1. QANDA 광고 사례 (IV)
2. Uber Eats 광고 타겟팅 사례 (Uplift)

3. 이미 만들어진 결정 및 제품에 대한 효과를 검증할 때

1. QANDA 퀴즈 사례 (DiD)
2. QANDA 퀴즈 사례 (CausalImpact)

+ 4. 진행된 A/B 테스트의 Bias를 보정할 때

예상 프로젝트 Output 참고

- 실무에서 발생하는 인과추론 적용 상황을 MECE하게 분류하여 그에 대한 실무 사례를 템플릿화

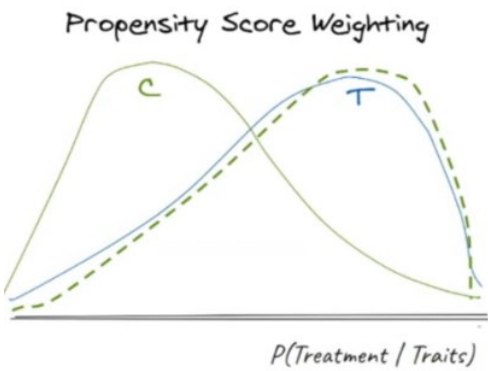
No	Question	Method	사례
1	Important pre-existing difference across groups?	CUPED / Diff-in-diff Propensity Score Matching IPTW	
2	Some people in the treatment group didn't actually receive the treatment?	Complier Average Causal Effect (CACE)	
3	Treatment effect varies significantly across subgroups?	Heterogeneous treatment effect Uplift modeling Quantile regression	
4	Test why treatment did (or didn't) move outcome metric?	Mediation modeling	
5	Time-series data before and after treatment are available, but no control group	Interrupted Time-Series Analysis	
6	Exogenous shock/events + Valid Parallel trend assumption	Difference-in-Differences (DID)	
7	Exogenous events + Invalid Parallel trend assumption	DID + Matching or Synthetic Control	
8	Endogenous events + Arbitrary threshold	Regression Discontinuity	
9	There is an instrumental variable that induces treatment	Two-Stage Least Squares or Selection Bias Correction Method	
10	Selection on observables is possible	Matching	
11	Matching is not possible	Regression	

예상 프로젝트 Output 참고

- 실무 인과추론 적용 사례 템플릿화

상황: Airbnb 슈퍼호스트 배지가 숙소 예약에 미치는 인과 효과추론 실무 사례

A/B 테스트를 진행하지 않았지만, 어떻게 슈퍼호스트가 숙소 예약에 미치는 효과를 알 수 있을까?



Situation	Hypothesis	Defining Estimand	Identification Strategy	
There are Airbnb accommodations that have implemented a superhost badge and those that have not	How does superhost affect consumer demand?	Situation Change to Causal inference problem : ATE	Random assignment is not feasible ▼ Pre-intervention and post-intervention outcome data available ▼ Is there homogeneity between the control group and the treatment group? ▼ YES	Use Propensity Score Matching to discard poor counterfactuals, DID to control for unobserved variables ▼ Estimation



TODO

TODO: 아이스브레이킹

TODO: 학습 방식

TODO: 난이도 조절

TODO: 스터디 순서

아이스 브레이킹