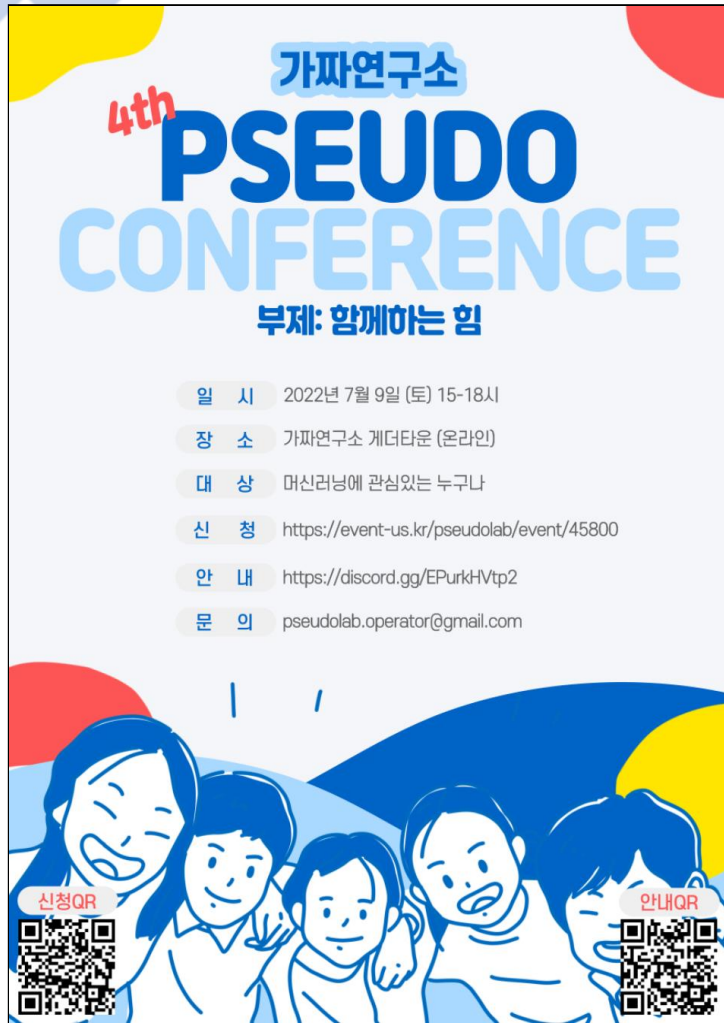




4th PseudoCon : 함께하는 힘

Casual하게 Causality 이해하기





◦ 스터디 이름 : Casual하게 Causality 이해하기

◦ 발표자 : 신진수 (빌더)

◦ 스터디 활동 내용 :

인과추론 한국어 자료 - [Causal하게 Causality 이해하기 노션](#)

팀 블로그 - [가짜연구소 Causal Inference Lab 블로그](#)

◦ 스터디원 (9명) :

김상돈, 김성수, 김소희, 김준영, 남궁민상,
방효석, 정호재, 최은희, 홍성철



스터디 소개

- 스터디 기간 : 22.03.12(토) ~ 22.06.25(토)
- 스터디 개설 이유



$$\text{인생} = 0 \uparrow \times \infty \uparrow$$

[나의 역량] [기회]



Casual하게 Causality 이해하기

스터디/크루 소개

- 스터디 내용
 - 인과관계/추론에 대한 기본적인 개념학습
 - Reading Material 및 Guest Lecture를 통한 다양한 인과관계 사례 및 논문 리뷰
 - 공부 내용을 블로그 콘텐츠로 정리
- 스터디 목적
 - 분석/실험 환경에서 자주 쓰이는 인과관계에 대한 기본적인 내용을 학습
 - 한국어로 된, Causality 자료가 많지 않습니다, 우리 함께 같이 만들어 봐요!

스터디 소개

- 스터디 학습 : 이론 공부 + 사례 리뷰 (온라인 Gather Town)

(1) 이론 : Introduction to Causal Inference (Brady Neal)

(2) 사례 : 논문, Uber Tech Blog, Python Causality Handbook, 인과추론 Summer Seminar 등

(3) 정리 : 스터디 강의 (아카이브) + 참고한 링크 + Causal Inference 라이브러리

Causal Inference 학습에 도움이 되는 자료 & 링크

[Causal Inference 자료]

Week	자료 설명 및 링크
OT	<p>같이 학습하면 좋은 Lecture Material 정리!!!</p> <ul style="list-style-type: none"> - 박지용 교수님 세미나 : https://sites.google.com/view/causal-inference2021 (교수님 깃헙 : https://github.com/jiyong-park) - Microsoft Research Scientist Amit Sharma의 블로그 : http://www.amitsharma.in/ - Uber Engineering 블로그 : https://eng.uber.com/ - 인과추론 톨아보기 한국어 자료 : https://assaeunji.github.io/bayesian/2020-04-10-causal/ <p>사례 발표시 참고하기 좋은 블로그</p> <ul style="list-style-type: none"> - https://matheusfacure.github.io/python-causality-handbook/10-Matching.html <p>Causality 주요 커뮤니티</p> <ul style="list-style-type: none"> - (1) Facebook : Causal Inference KR : https://ko-kr.facebook.com/groups/causalinferencekr/ - (2) Product Analytics PlayGroup : https://playinpap.github.io/ - (3) Awesome Causality (인과추론 아카이브 블로그) : https://shubhanshu.com/awesome-causality/

[학습 링크]

Archives

 갤러리 보기
  Table
  Gallery
  List
  Board
  Timeline

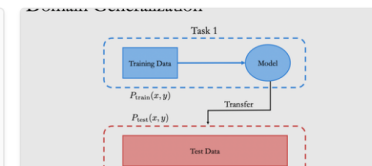
Causal Inference 아카이빙


Chapter 14. Counterfactual Mediation

Content 구성

- Counterfactuals Basics
- Important Application : Mediation

 Chapter 14. Counterfactual Mediation



 Chapter 13. Transfer learning and Transportability

Chapter 12. Causal Discovery from Interventions

Content 구성

- Structural Interventions (1) Single-Node Interventions

 Chapter 12. Causal Discovery from Interventions

[스터디 강의 + 사례 (아카이브)]

숫자로 보는 Causal하게 Causality 이해하기 스터디



총 스터디 원
10 명

빌더와 러너 전원 포함한 총 스터디
참여 인원 수입니다



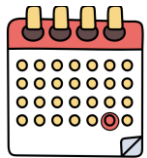
누적 학습시간
1440 분

총 스터디 시간인 24시간을 분으로
환산한 결과입니다



발표 횟수
3 회

스터디에서 스터디원의
평균 발표 횟수입니다.



함께한 토요일
14 번

중간 리뷰 제외, 3/12 ~ 6 /25 동안
매 주 학습을 진행했습니다



정리한 자료 개수
28 개

Notion에 학습한 Causal Inference
내용을 정리한 개수입니다

스터디 목표

- 3개월 간, **스터디의 목표** & **각자 스터디에서 이루고 싶은 목표**

(1) Causality 강의 & 책을 끝까지 스터디원들과 함께 마무리 해보는 것

(2) 공부한 내용을 정리해, 인과추론에 대한 한국어 자료 만들어 보기

(3) 본인이 가지고 있는 문제를 해결할 수 있는 실마리 찾기 Or 적용

(4) 기타 : 방향성, 시간관리, 블로그, 좋은 스터디원 등

Casual하게 Causality 이해하기

🎯 Study 목표

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

Causal Inference?

- Causal Inference란 무엇일까요?

(0) 현상(문제)에 대한 원인을 찾고 해당 원인에 대한 효과를 추론하는 것

- (1) "Causal Inference is concerned with a **very specific kind of prediction problem**
: Predicting the results of an action, manipulation, or Intervention"
〈 Making Things Happen 〉 (2003, James Woodward) 〉



[왜 그런거예요??]

1. 사례 - 사회 제도와 경제 성장

논문 제목 : *The Colonial Origins of Comparative Development*

✓ 논문 요약

- 질문 :
 - 국가간 1인당 국민 소득(GDP)에 큰 차이가 발생하는 근본적인 원인은 무엇인가?
 - 제도적 차이의 영향이라면 어떻게 **사회 제도 차이**에 대한 **경제적 성장**의 인과 효과를 추정할 수 있을까?
- Contribution :
 - 유럽인에 의해 식민화된 국가들 사이의 제도적 차이에 대한 이론을 제안
 - 해당 이론을 이용하여 외생적 변동성의 원인을 도출
 - 사회 제도적인 부분을 도구변수를 활용해, Causal Effect **정량적으로 추정**

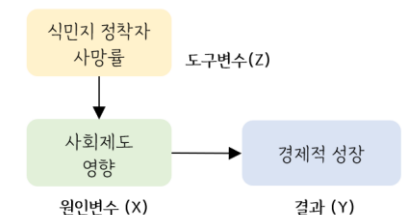
[스터디 맛보기]

[Endogeneity 원인]

- 생략된 변수 (Omitted variables)
- 역인과관계 (Reverse Causality)
- 측정 오차 (Measurement Errors)



[Endogeneity 다루기 - 도구변수]



Causal Inference?

- Causal Inference Framework

(1) 인과추론의 근본적인 문제 : Counterfactuals



◦ 유 퀴즈 온 더 블럭 : [구준엽 편 Youtube](#)

스터디를 통해 얻은 것s

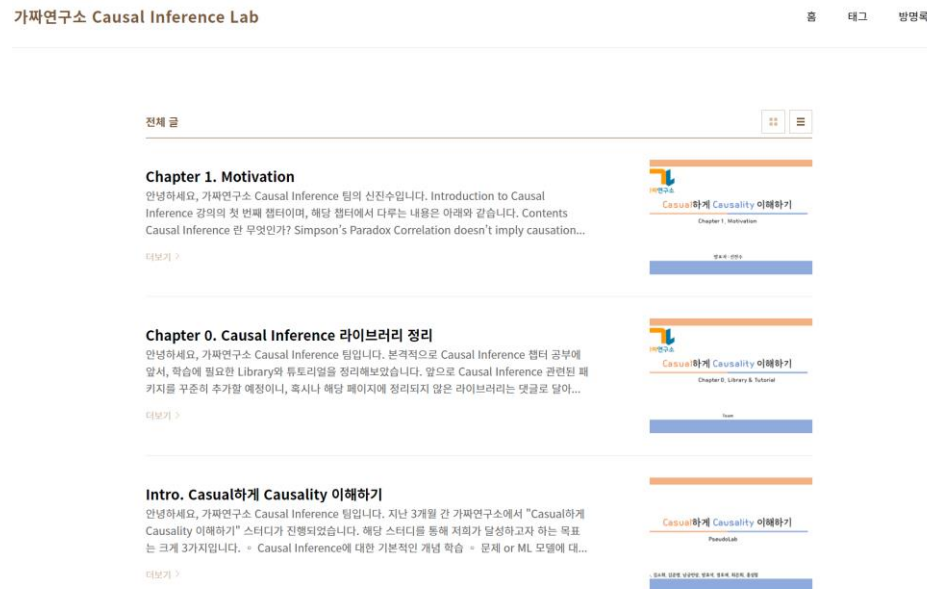
- 비록 짧은 시간이지만, 우리가 얻은 것!

(1) Team Blog

(2) 가짜연구소 Notion

(3) 좋은 스터디원 !

(4) 팀 로고!



[팀 블로그]

Chapter 14. Counterfactual Mediation

Content 구성

- Counterfactuals Basics
- Important Application : Mediation

(1) Causal Inference의 근본적인 문제 : Counterfactuals

💡 Counterfactual : a probabilistic answer to a "what would have happened if" question

- 근본적인 문제 : 실험 대상이 Treatment를 받았다면, Treatment를 받지 않은 경우에 대해 관측 X
 - 결과
 - 통계적 관점 : Unit-level Causal Effect 추정 X
 - 의사결정 관점 : 유저에 대한 Treatment의 개인화된 효과를 알 수 없음
 - 1) Potential Outcomes에 대해서 동시에 관측이 불가능
 - Causal Effect에 대한 계산을 위해, Missing Value에 대한 해결이 필요
 - 2) 첫 번째 Outcome을 관찰하고, 그 이후에 다른 Outcome을 관찰하는 경우
 - 첫 번째 Outcome 관측 이후, 해당 결과에 의해 다른 Outcome이 영향을 받을 수 있음

[가짜연구소 Notion]

Feedback

• 스터디 회고

(1) 이론 위주의 학습

(2) 스터디 발표 내용에 대해 토의와 공유할 시간을 만들어 드리지 못한 점..

(3) 스터디원과 인과추론 이외 부분에 대해 이야기하는 시간의 부재

학습 활동

1. 스터디 (시작전)

- 모두가 해당 주차의 강의를 듣고, 공부합니다.
 - 해당 챕터를 공부하면서 느낀점 or 질문 내용을 준비합니다
 - 해당 챕터를 공부하는데, 참고하거나 좋은 Material이 있다면, 계획표에 링크를 남겨주세요

2. 스터디 (모임)

- 발표자는 해당 주차에 학습한 내용을 노선에 정리합니다.
 - 이해를 못해도, 완벽하지 않아도 됩니다! → 스터디를 통해, 어떻게 이해했는지가 더 중요해요!
- 발표자의 발표가 끝난 후에는,
 - 해당 주차에서 이해가 되지 않았던 부분을 질문 or 같이 알아두면 좋은 내용을 토의, 공유합니다.
 - 스터디원이 선정한 인과관계 사례에 대한 부분을 토론합니다.

3. 스터디 (모임 이후)

- 발표자가 해당 주차의 나온 내용을 노선으로 정리합니다 (Q&A, 자료 링크 포함)
- 권장) 문제 상황 및 다음주 토의 내용에 대해 제안해보는 걸 추천드립니다.
- 권장) 인과관계 내용이 담긴 도서 읽어보는 걸 추천드려요 (저에게 하는 말이에요.....)
 - ex) 데이터 분석의 힘, The Book of Why 등등

Future Work

- 우리 Causal Inference 스터디원분들과 다음에도 **함께** 나아갈 예정입니다!

(1) 문제해결 위주의 학습 / 진행

- 문제해결) 카드 금액 연체된 사람들에게 **이메일을 보내는 것은 자금 회수에 긍정적인 효과**를 가져오는가?

(1) 목표 : 이메일 발신 여부에 따른 연체 금액 회수 효과

- 대상 : 카드 금액이 연체된 고객
- 상황 : Control / Treatment Group에 **실험** 대상을 **Random**하게 할당
- **Treatment** : 이메일 보내기 (T = 1 : 이메일 발송 / T = 0 : 이메일 발송 X)
- **Outcome** : 연체 금액 회수
- 그 외 변수 : 신용 한도 (credit_limit), 리스크 등급 (risk_score),
이메일을 받은 후 채무 협상을 위한 계약 여부(Agreement)

	payments	email	opened	agreement	credit_limit	risk_score
0	740	1	1.0	0.0	2348.495260	0.666752
1	580	1	1.0	1.0	334.111969	0.207395
2	600	1	1.0	1.0	1360.660722	0.550479

[Data Set]

(2) 스터디 발표 내용에 대해 **토의와 공유할 시간** ↑

(2) ATE 추정

- 선형 회귀 : $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ (Y: 연체금, X: 이메일 발송)
- 추정된 ATE : $E[Y | T=1] - E[Y | T=0] = -0.62$
- 결과해석 : 이메일 발송한 경우, 연체 금액 효과가 -0.62이다.

```
vanila_model = smf.ols('payments ~ email', data = df).fit()
vanila_model.summary()
```

OLS Regression Results					
Dep. Variable:	payments	R-squared:	0.000		
Model:	OLS	Adj. R-squared:	-0.000		
Method:	Least Squares	F-statistic:	0.04447		
	coef	std err	t	P> t	[0.025 0.975]
Intercept	669.9764	2.061	325.116	0.000	665.937 674.016
email	-0.6203	2.941	-0.211	0.833	-6.387 5.146

[선형회귀 결과]

(3) 스터디원과 **인과추론 이외** 부분에 대해 이야기하는 시간 ↑

→ 실험대상에 대해 **Randomisation**도 했고...시간과 비용을 투자해
이메일 발송을 했는데, **금액 회수 효과가 마이너스다?**

→ 이걸 하지 말았어야 하나?? 괜히 했나... **No No!**

[What's Next?]



감사합니다

