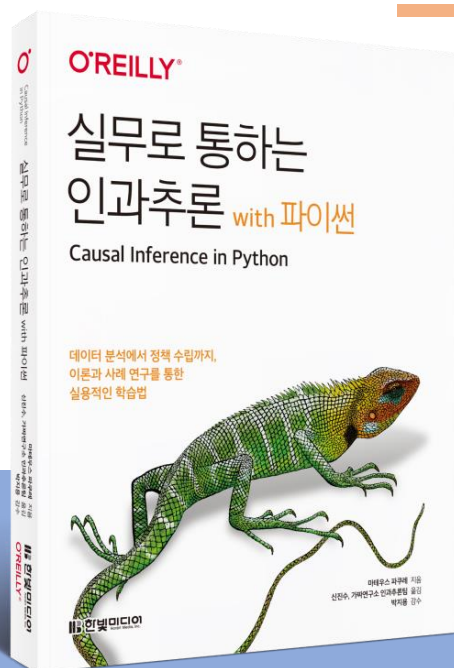


# 『실무로 통하는 인과추론 with 파이썬』 특강

## Orientation





신진수 (Jinsoo Shin)

## 본캐

- 데이터 분석가 (4년)

**KRAFTON** NEW STATE Mobile, PUBG Mobile



던전앤파이터 (PC/Mobile)

- 관심분야 : 인과추론, 온라인 통제실험, 조직문화

## 부캐

- 가짜연구소 위원회 / 인과추론팀 운영 (22.03~)



- 번역가

번역 신진수

인물정보

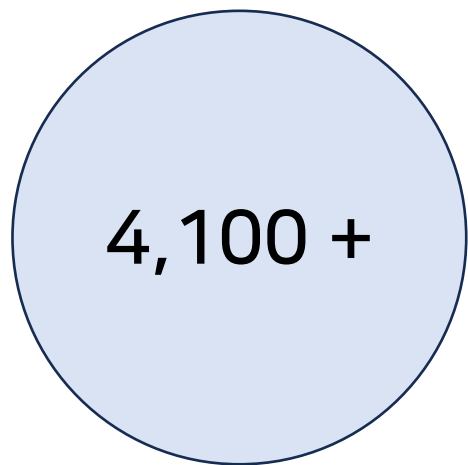
데이터베이스전문가

네오플을 거쳐 크래프톤(KRAFTON)의 데이터 분석가로 일하고 있습니다. 게임 업계에서 쌓은 커리어를 기반으로 <던전앤파이터>, <뉴스테이트 모바일>, <배틀그라운드 모바일> 등 다양한 장르의 게임에서 데이터 분석과 실험을 통해 유저 경험을 개선하는 데 기여했습니다. 비영리 데이터 사이언스 커뮤니티인 가짜연구소에서 인과추론팀을 운영 중입니다. 마테우스 파쿠레의 웹북 「Causal Inference for The Brave and True」를 한국어로 번역하는 작업을 주도했습니다(<https://github.com/CausalInferenceLab/Causal-Inference-with-Python>).

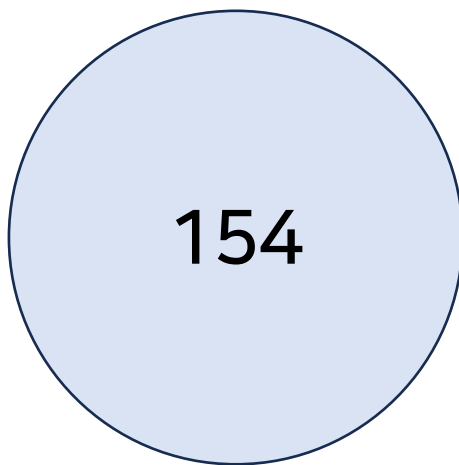
- 가짜연구소 & 인과추론팀 소개
- 특강 소개 & 구성
- 번역서 소개

머신러닝/데이터사이언스 연구를 중심으로 모인 비영리 커뮤니티 (Since 20.09)

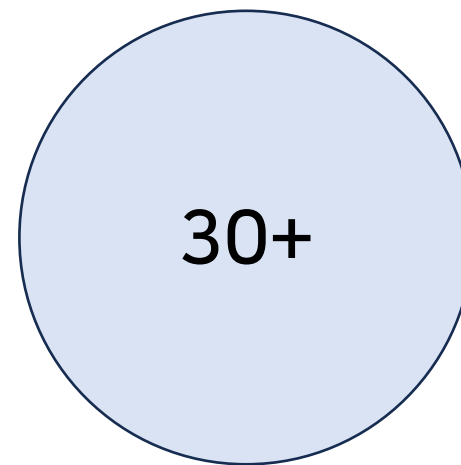
DS/ML 분야의  
구성원 수



23년까지 완료된  
프로젝트 수



24년 인과추론팀  
구성원 수



## 24년 진행중인 프로젝트

1. “실무로 통하는 인과추론 with 파이썬” 특강 – 신진수 빌더
2. 온라인 통제 실험 연구자로 거듭나기 – 방태모 빌더
3. Smart Causal Inference – 박상호 빌더
4. 인과추론과 실무 – 김성수 빌더



한국어 자료가 많지 않은 인과추론과 온라인 통제실험에 대한

접근성을 높일 수 있도록 기여

## 한국어 자료 만들기 for 인과추론 허들 낮추기

[Python으로 하는 인과추론:  
개념부터 실습까지 \(Github\)](#)

[“실무로 통하는 인과추론 with 파이썬” \(한빛미디어\)](#)



 README  MIT license

---

### Python으로 하는 인과추론 : 개념부터 실습까지

---

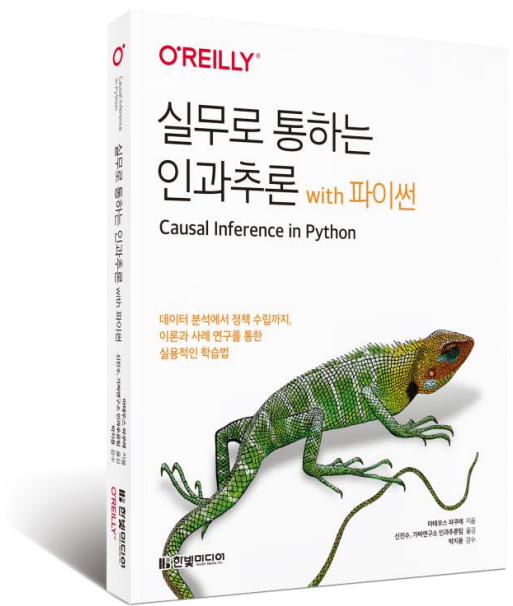
#### Causal Inference for The Brave and True 책 번역

 DOI [10.5281/zenodo.4445778](https://doi.org/10.5281/zenodo.4445778)  hits [63 / 21837](#)

안녕하세요. [가짜연구소](#) Causal Inference 팀입니다.

이 책은 Matheus Facure (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. 해당 한국어 번역 자료는 [Jupyter Book](#)으로도 배포되어 있으니 많은 관심 부탁드립니다.

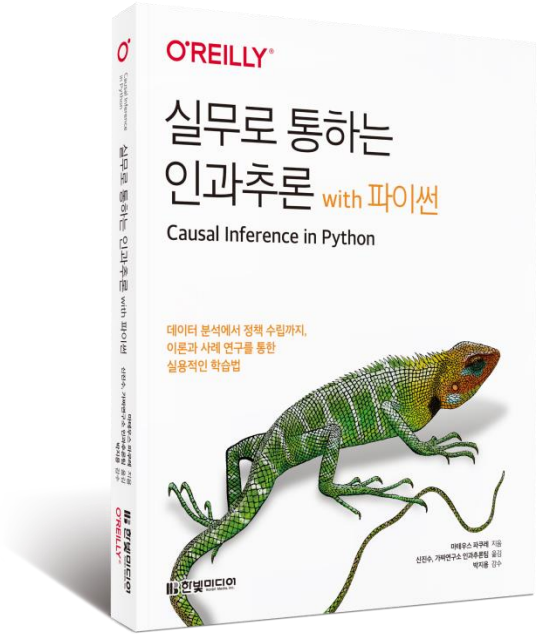
가짜연구소 Causal Inference 팀은 전문 번역가가 아닌, 데이터분석가/과학자/연구자로 구성 되어 있습니다. 따라서, 한국어 번역 과정에서 이상하거나 어색한 부분이 발생할 수 있어요. 해당 부분에 대한 여러분의 피드백을 부탁드립니다. 또한, 한국어 표현이 익숙하지 않은 부분에 대해서는 영어로 그대로 표현했습니다. 해당 부분 이해부탁드려요.



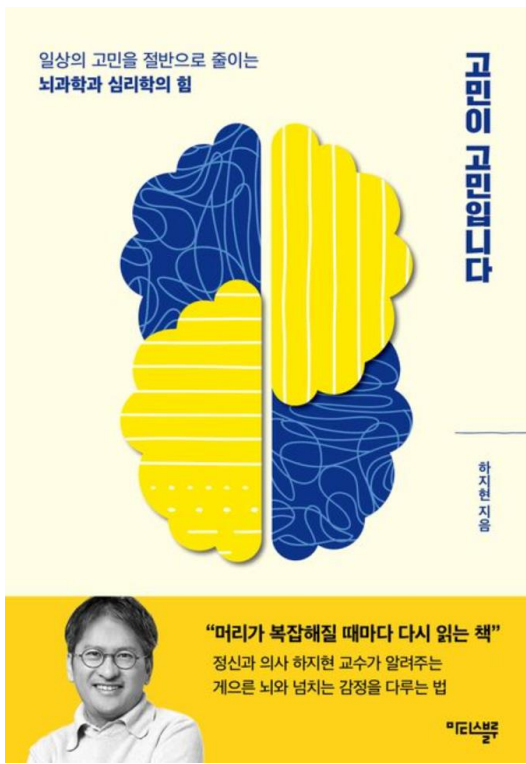


배경

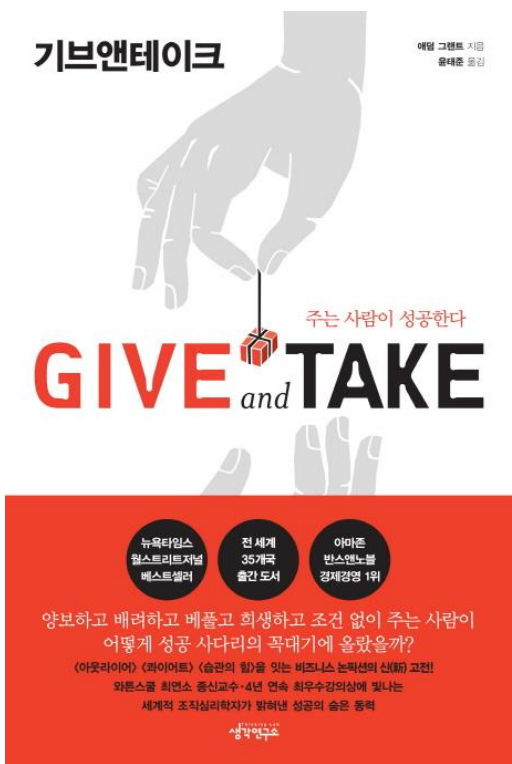
(1) 인과추론 기본서 출간



(2) 인과추론 실무 경험 및 고민 공유 채널 필요

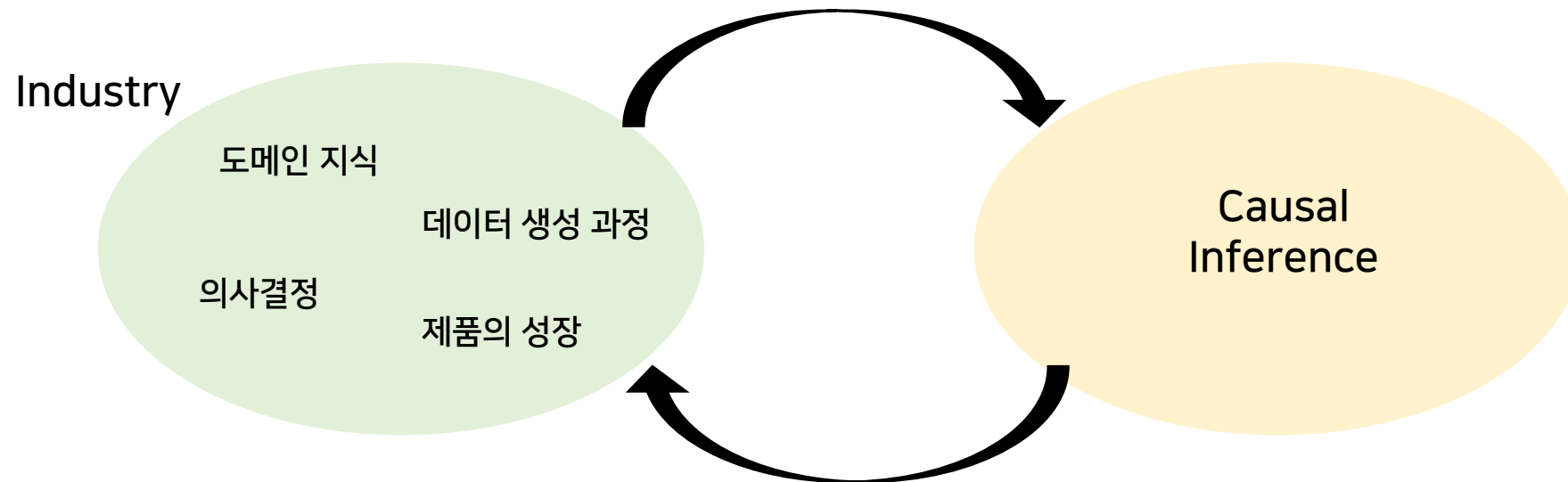


(3) 받은 만큼 환원하기



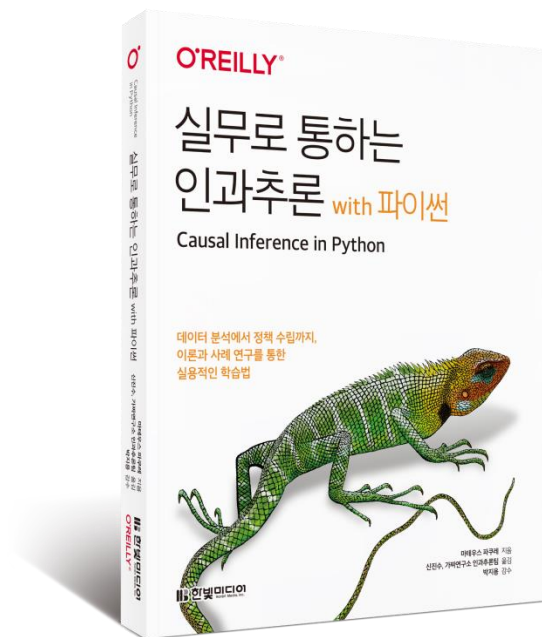
## 목표

- 해당 도서 기반으로, 실무/연구자 관점에서 내용을 재해석 및 경험 사례를 공유
- 특강 및 번역서를 기반으로, 인과추론에 대한 실무/연구자들의 다양한 의견 활성화



## 역자와 함께하는 “실무로 통하는 인과추론 with 파이썬” 온라인 강의

### 1. 범위 : 번역서에 담긴 인과추론 개념



- Part1 : 인과추론 기초
  - 1장. 인과추론 소개
  - 2장. 무작위 통제실험 및 기초 통계 리뷰
  - 3장. 그래프 인과모델
- Part2: 편향 보정
  - 4장. 유용한 선형회귀
  - 5장. 성향점수
- Part3 : 이질적 효과와 개인화
  - 6장. 이질적 처치효과
  - 7장. 메타러너
- Part4 : 패널데이터
  - 8장. 이중차분법
  - 9장. 통제집단합성법
- Part5: 대안적 실험 설계
  - 10장. 지역 실험과 스위치백 실험
  - 11장. 도구변수
  - 12장. 더 배울 내용

## 역자와 함께하는 “실무로 통하는 인과추론 with 파이썬” 온라인 강의

### 2. 내용

- 책 기반의 내용을 실무/연구자 관점에서 재해석
- 인과추론 실무 적용 사례 위주로 구성

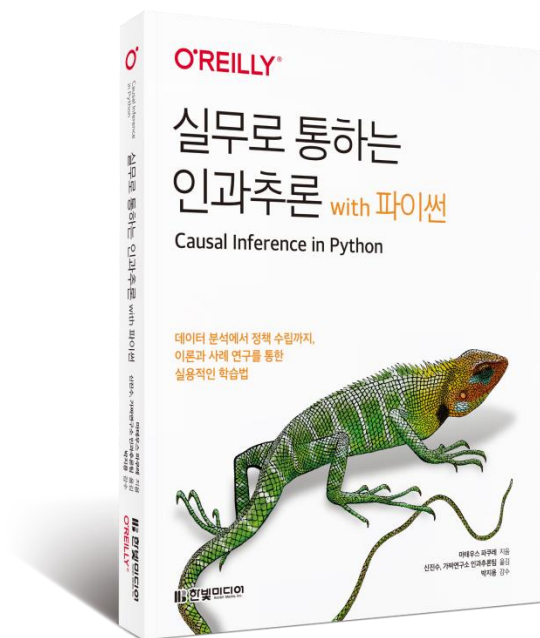
### 3. 진행 기간 : 24.05.01 ~ 24.10.20 (5개월)

### 4. 업로드 방식

- 특강 : [“인과추론의 데이터과학 유튜브 채널”](#)에 1~2주 주기로 업로드
- 자료 : [가짜연구소 인과추론팀 Github](#)

### 5. 강의 대상

- “실무로 통하는 인과추론 with 파이썬” 독자
- 인과추론 & 데이터 기반 의사결정에 관심이 있는 실무자 및 대학(원)생



## 역자와 함께하는 “실무로 통하는 인과추론 with 파이썬” 온라인 강의

### 6. 참여 구성원 : 가짜연구소 인과추론팀 (17명)

✓ 총괄 : 신진수 (크래프톤)

✓ 엔터테인먼트

- 김소희 (TVING)
- 김지연 (EA)
- 박병수 (넥슨코리아)
- 박시온 (넥슨코리아)
- 박이삭 (하이브IM)
- 임종언 (넥슨코리아)
- 유정현 (넥슨코리아)
- 조슬지 (넷마블에프앤씨)
- 최은희 (넥슨코리아)

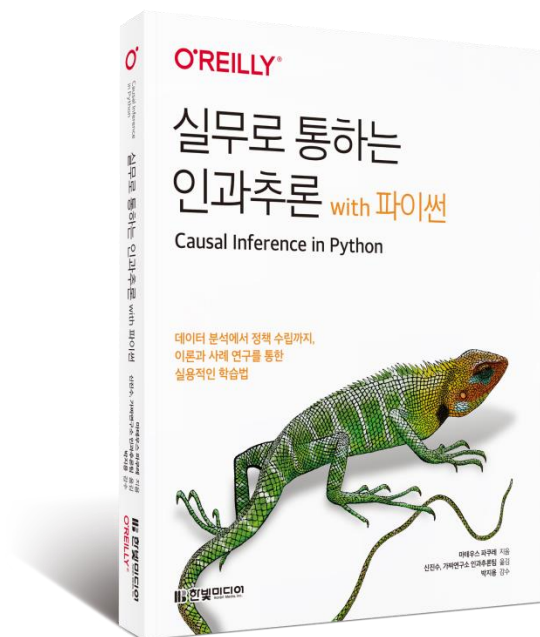
✓ 금융 및 정부기관

- 정호재 (롯데캐피탈)
- 박상우 (한국은행)
- 조인서 (KOICA)

✓ 커머스 : 방태모 (Gmarket)

✓ 리서치

- 김성수 (LG CNS)
- 김준영 (컬럼비아대학교)
- 박상호 (성균관대학교)



## 계획

- 1부 : 인과추론 기초 (프레임워크 / 식별 가정 / 그래프 모델) & 무작위 통제 실험
- 2부 : 편향 제거 방법 (회귀분석 / 매칭) & 디자인/모델 기반 인과추론

챕터	컨텐츠	담당자 (소속)	공유 일정
OT	인과추론 번역서 및 강의 계획 소개	신진수 (크래프톤)	5.1
1장 : 인과추론 기초	프로모션 효과분석으로 인과추론 기초 배우기	신진수 (크래프톤)	5.12
2장 : 무작위 통제 실험	온라인 통제 실험 소개 및 분석 기초	방태모 (Gmarket)	5.19
	경제성 평가를 위한 무작위 통제 실험	조인서 (KOICA)	5.26
3장 : 그래프 인과모델	게임 유저의 생존분석 관점에서의 그래프 인과모델 소개	박시온 (넥슨코리아)	6.9
4장 : 회귀분석	실무자가 바라본 고객 신용한도와 채무 불이행률의 인과적 관계	정호재 (롯데캐피탈)	6.16
	마케팅 믹스 모델링과 다중회귀분석	최은희 (넥슨코리아)	6.23
5장 : 매칭	실험 설계에서의 매칭 활용	유정현 (넥슨코리아)	6.30

## 계획

- 3부 : 개인화 효과분석 (CATE 추정 / 모델 평가 방법) & 인과추론과 머신러닝 (메타러너)
- 4부 : 패널데이터로 실험 디자인 설계 (이중차분법 / 통제집단합성법 / 합성 이중차분법)

챕터	컨텐츠	담당자 (소속)	공유 일정
6장 : 이질적 처치효과	유저 세그멘테이션과 개인화 관점에서의 인과추론	박병수 (넥슨코리아)	7.7
	게임 업데이트 효과에 따른 이질적 처치효과 추정	박이삭 (하이브IM)	7.14
7장 : 메타러너	머신러닝 모델로 인과추론하기	김준영 (컬럼비아대학교)	7.21
8장 : 이중차분법	은행 정책관점에서의 패널데이터와 인과추론	박상우 (한국은행)	7.28
	외부 요인 (월드컵 16강 진출 / 손흥민 득점왕 / 챔스 결승)이 축구 게임 유입에 영향을 준 요인 찾기	임종언 (넥슨코리아)	8.4
	게임에서 이중차분법 적용하고 비교하기	김지연 (EA Korea)	8.11
9장 : 통제집단합성법	통제집단합성법을 활용한 업데이트 효과분석	박시온 (넥슨코리아)	8.18

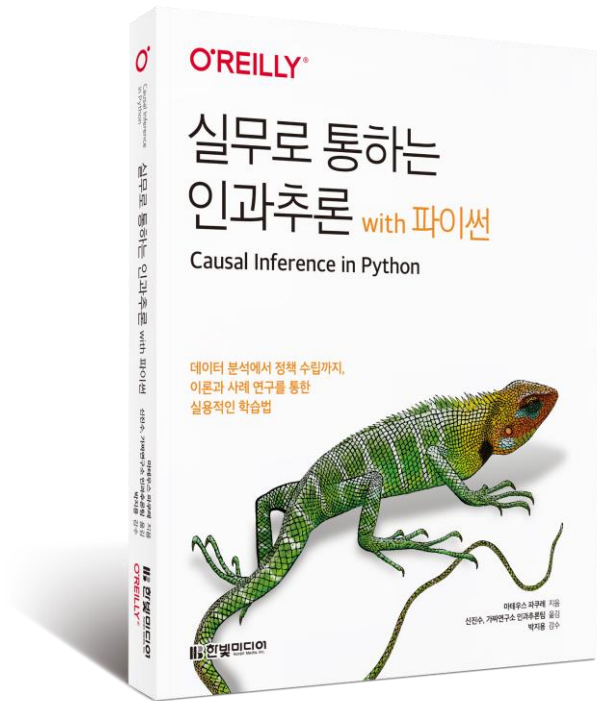
## 계획

- 5부 : 스위치백 실험 & 불응 문제를 다루기 위한 방법 (도구변수 및 불연속 설계)
- Last but not least : 심화 내용 + 인과추론 실무 사례

챕터	컨텐츠	담당자 (소속)	공유 일정
10장 : 스위치백 실험	생소하지만 중요한 스위치백 실험	신진수 (크래프톤)	9.1
11장 : 도구변수	도구변수를 통한 프로덕트의 인앱 푸시 효과 측정	신진수 (크래프톤)	9.8
12장 : 더 배울 내용	인과관계 발견 및 모델 개발/운영 (Causal Ops) 소개	김성수 (LG CNS)	9.15
13장 : 인과추론과 실무	인과추론으로 바라보는 게임 내 길드 영향력 분석	조슬지 (넷마블에프앤씨)	9.22
	구독 상품에 따른 retention의 인과효과 추정하기	김소희 (TVING)	9.29
	임상 관찰 연구에서 RCT 모사를 위한 Target Trial Emulation	박상호 (성균관대학교)	10.6
	실무에 적용하는 인과추론 템플릿	김성수 (LG CNS)	10.13
종강	에필로그 소개 및 특강 갈무리	신진수 (크래프톤)	10.20



실무로 통하는 인과추론 with 파이썬



2024년 03월 3주

2024.03.13 ~ 2024.03.19

20개씩 보기

전체선택

장바구니

Excel다운로드

1

교보문고 Best 1

MD의 선택 이벤트 사은품

소독공제

진짜 쓰는 실무 엑셀

오빠두(전진권) · 제이펍

10% 18,900원 (1,050p)

9.62 / 66 도움돼요

2

MD의 선택 이벤트 사은품

소독공제

혼자 공부하는 파이썬

윤인성 · 한빛미디어

10% 19,800원 (1,100p)

9.79 / 66 집중돼요

3

MD의 선택 사은품

소독공제

Do it! 점프 투 파이썬

박용훈 · 이지스퍼블리싱

10% 19,800원 (1,100p)

9.76 / 66 도움돼요

4

MD의 선택 이벤트

소독공제

회사는 엑셀을 가르쳐주지 않아요

쏘피(박성희) · 메가스터디박스

10% 16,200원 (900p)

9.69 / 66 도움돼요

5

소독공제

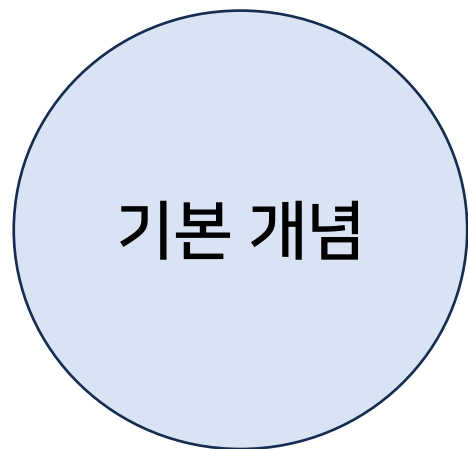
실무로 통하는 인과추론 with 파이썬

마테우스 파쿠레 · 한빛미디어

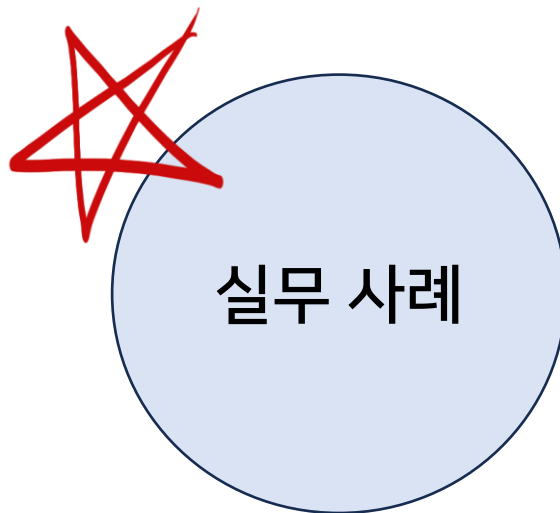
10% 34,200원 (1,900p)

10 / 66 도움돼요

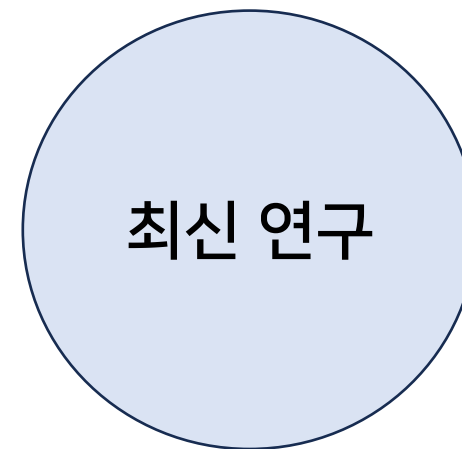
## 특징 1 : 인과추론 기본 개념 / 실무 사례 / 최신 연구 및 심화 방법론을 다루는 책



- ✓ 인과추론 식별 가정  
(교환 가능성, 양수성, SUTVA, 패널데이터/IV 식별)
- ✓ 편향 제거 방법  
(회귀분석, 매칭, 도구변수)
- ✓ 인과추론 프레임워크  
(잠재적 결과, 구조적 인과 모델)



- ✓ 신용 한도와 채무불이행률
- ✓ 어트리뷰션과 증분  
(마케팅 믹스 모델)
- ✓ 인과추론 라이브러리  
(CausalML, CausalImpact, fkearn)



- ✓ 코로나19 백신의 효과
- ✓ 패널데이터 연구  
(시차 도입, 합성 이중차분법)
- ✓ 이질적 효과의 모델 평가  
(분위수 곡선, 누적효과 곡선, 목표 변환)

특징 2 : 실무자가 고민하는 방법론 활용에 대한 실무적 조언을 포함 with 파이썬



+  python<sup>TM</sup>

## 특징 2 : 실무자가 고민하는 방법론 활용에 대한 실무적 조언을 포함 with 파이썬

### (5장 성향점수)

개인적으로 필자는 매칭 추정량을 크게 선호하지 않습니다. 그 이유는 첫째, 편향될 가능성이 있고, 둘째, 분산을 추정하기 어렵고, 셋째, 데이터 과학 활용 경험에 비추어보았을 때 KNN에 다소 회의적이기 때문입니다. 특히 KNN은  $X$ 가 고차원인 경우, 대체로 효율이 크게 떨어집니다. 이 마지막 문제는 성향점수만으로 매칭한다면 문제가 되지 않지만, 처음 두 가지 문제는 여전히 남습니다. 그럼에도 불구하고 매칭 추정량을 다룬 이유는 해당 방법이 많이 알려져 있기 때문입니다.

#### NOTE 더 알아보기

게리 킹 Gary King과 리처드 닐슨 Richard Nielsen의 연구<sup>3</sup>는 성향점수 매칭 문제에 관한 더 전문적인 내용을 다룹니다.

### (11장 도구변수) 표준오차를 줄이기 위한 조언

이 식을 보면, 표본 크기를 늘리는 방법 외에도 표준오차를 줄이는 세 가지 방법이 있음을 알 수 있습니다.

- 1. 1단계의  $R^2$  증가시키기. 이는 순응을 잘 예측하는 강력한 도구변수를 찾아서 해결할 수 있습니다. 해당 변수들은 결과에 영향을 미치지 않으면서 배제 제약을 만족해야 합니다.
- 2.  $T$ 에 대한 예측력이 높은 변수를 제거하여  $\sigma(\tilde{T})$  증가시키기
- 3. 결과의 예측력이 높은 변수를 찾아서 2단계 잔차 줄이기

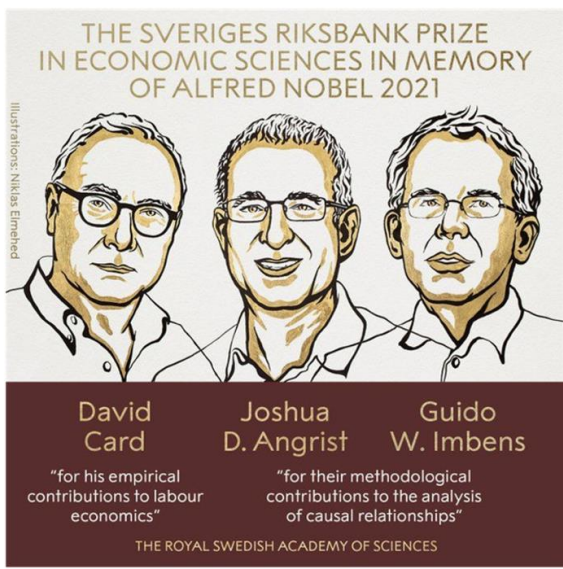
필자는 이 중 마지막 방법을 선호합니다. 앞서 말했듯이, 실전에서 도구변수를 찾기란 매우 어렵고  $\sigma(\tilde{T})$ 를 줄이기 위해 변수를 제거하는 부분도 한계가 있습니다. 따라서 신뢰할 수 있는 유일한 분산 감소 방법은 결과를 잘 예측하는 변수를 찾는 것이 현실적이라고 생각합니다.

## 특징 3 : 잠재적 결과 프레임워크 중심으로 인과추론을 설명

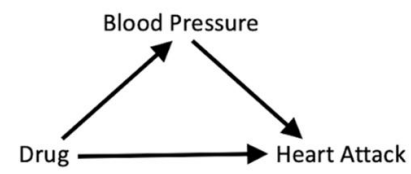
### Potential Outcomes Framework (Rubin Causal Model)

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



### Structural Causal Model (Pearl Causal Model)



ACM A. M. Turing Award

Judea Pearl  
United States – 2011  
**Citation:**  
*For fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning.*



### 잠재적 결과 (1.4.6절)

- ✓ 사실적(factual) 결과
- ✓ 반사실적(counterfactual) 결과

$$Y_i = T_i Y_{1i} + (1 - T_i) Y_{0i}$$

$$\tau_i = Y_{1i} - Y_{0i}$$

( $T$  : 이진 처치,  $\tau$  : 인과효과)

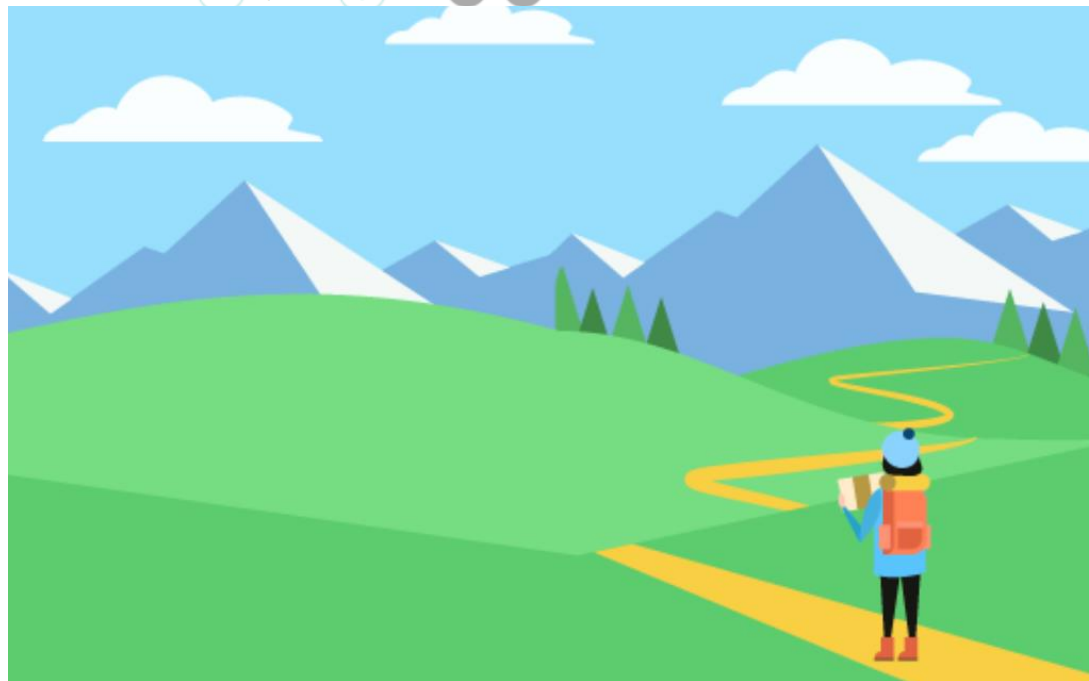
## 특징 4 : 실무에서 실험 / 인과추론 적용을 통해 배운점 및 역자의 생각이 담긴 "에필로그" 반영

### 실무에 인과추론 적용하기 위한 험난한 5단계 여정

M 한빛미디어 | 2024-03-12 | by 신진수

3,542 4

- ① 도메인 지식 습득
- ② 신뢰 쌓기
- ③ 실험 설계
- ④ 실험 진행 및 분석
- ⑤ 실험을 바탕으로 의사결정 및 피드백



## Summary

- 인과추론 기본 개념 / 실무 사례 / 최신 연구 및 심화 방법론을 다루는 책
- 파이썬 코드와 함께 실무자가 고민하는 방법론 활용에 대한 실무적 조언을 포함
- 잠재적 결과 프레임워크 중심으로 인과추론을 설명
- 실무에서 실험 / 인과추론 적용을 통해 배운점 및 역자의 생각이 담긴 "에필로그" 반영

## 목차

- Part1 : 인과추론 기초
  - 1장. 인과추론 소개
  - 2장. 무작위 통제실험 및 기초 통계 리뷰
  - 3장. 그래프 인과모델
- Part2: 편향 보정
  - 4장. 유용한 선형회귀
  - 5장. 성향점수
- Part3 : 이질적 효과와 개인화
  - 6장. 이질적 처치효과
  - 7장. 메타러너
- Part4 : 패널데이터
  - 8장. 이중차분법
  - 9장. 통제집단합성법
- Part5: 대안적 실험 설계
  - 10장. 지역 실험과 스위치백 실험
  - 11장. 도구변수
  - 12장. 더 배울 내용



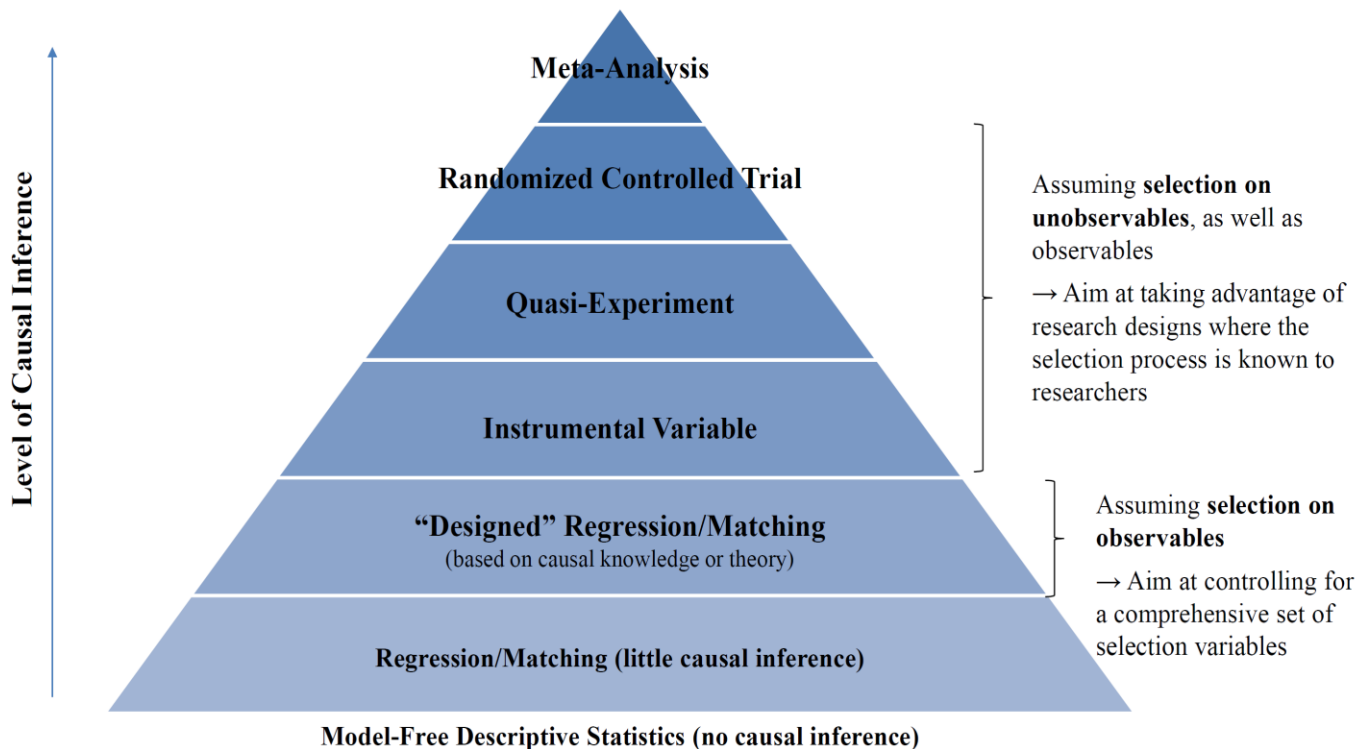
## 목차

- Part1 : 인과추론 기초
  - 1장. 인과추론 소개
  - 2장. 무작위 통제실험 및 기초 통계 리뷰
  - 3장. 그래프 인과모델
- Part2: 편향 보정
  - 4장. 유용한 선형회귀
  - 5장. 성향점수
- Part3 : 이질적 효과와 개인화
  - 6장. 이질적 처치효과
  - 7장. 메타러너
- Part4 : 패널데이터
  - 8장. 이중차분법
  - 9장. 통제집단합성법
- Part5: 대안적 실험 설계
  - 10장. 지역 실험과 스위치백 실험
  - 11장. 도구변수
  - 12장. 더 배울 내용

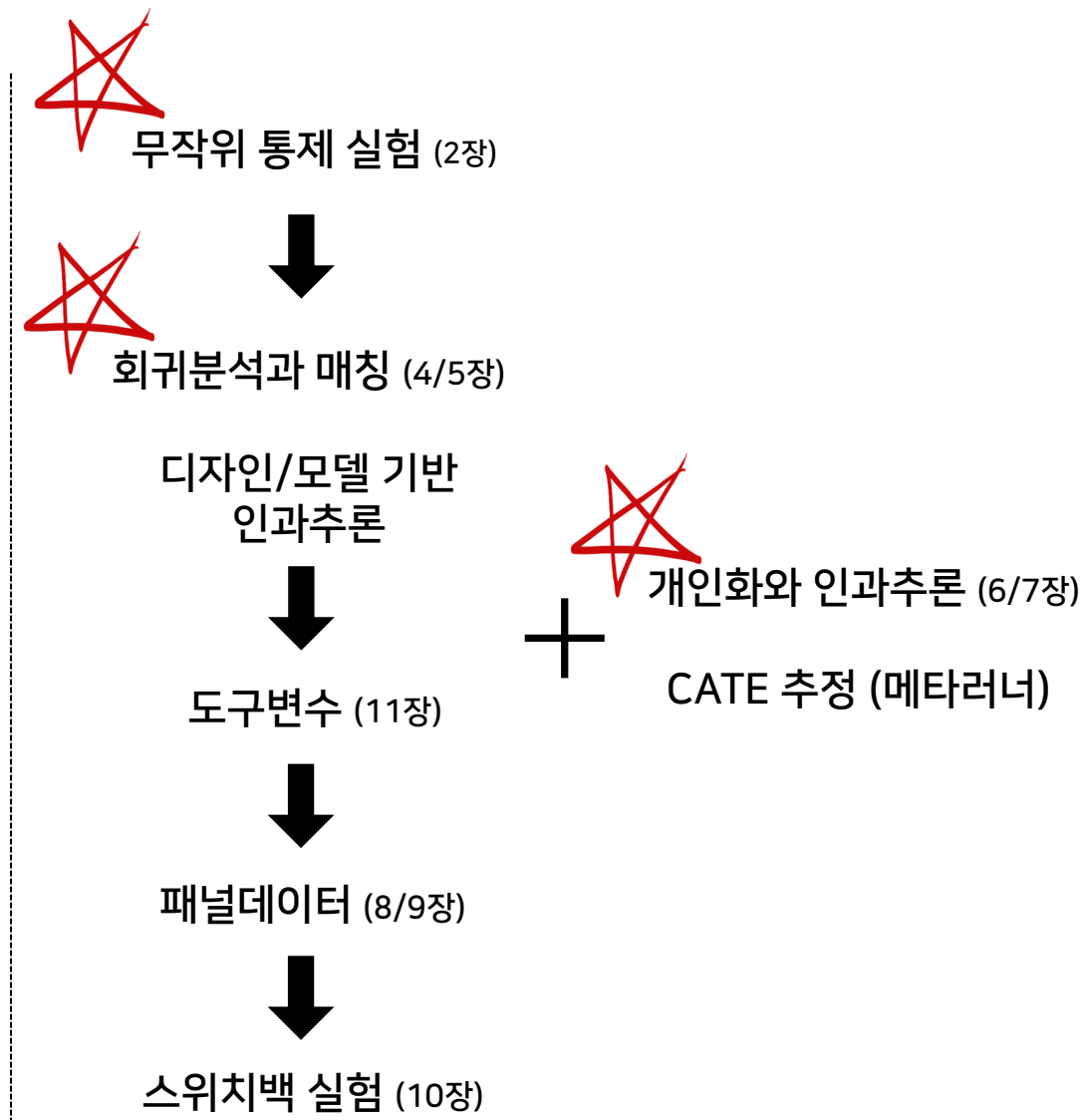
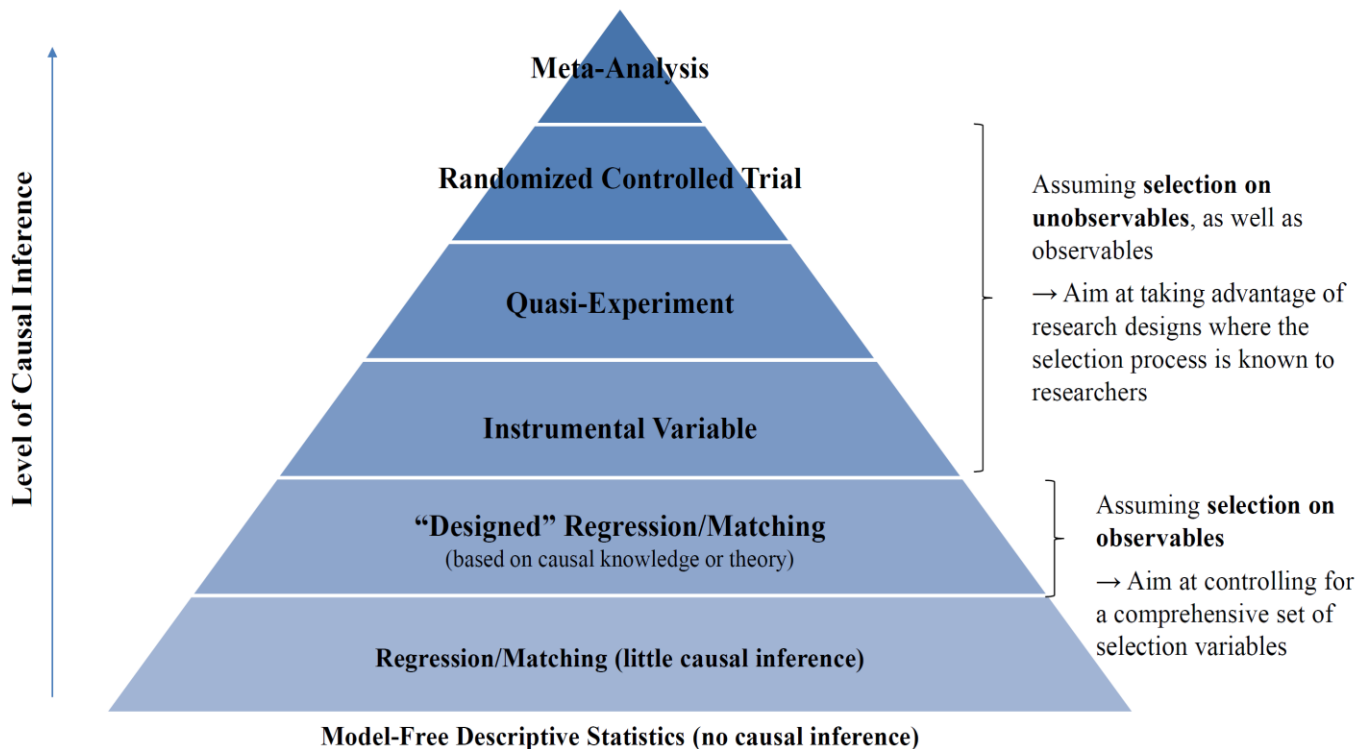
## 책 구조화

- ➡ 인과추론의 근본적인 문제 & 편향
- ➡ 인과추론 프레임워크
- ➡ 인과추론 가정
- ➡ 식별 과정 (Causal Estimand)
- ➡ 추정 방법 = 편향(endogeneity) 제거
- ➡ 결과 신뢰성 검증

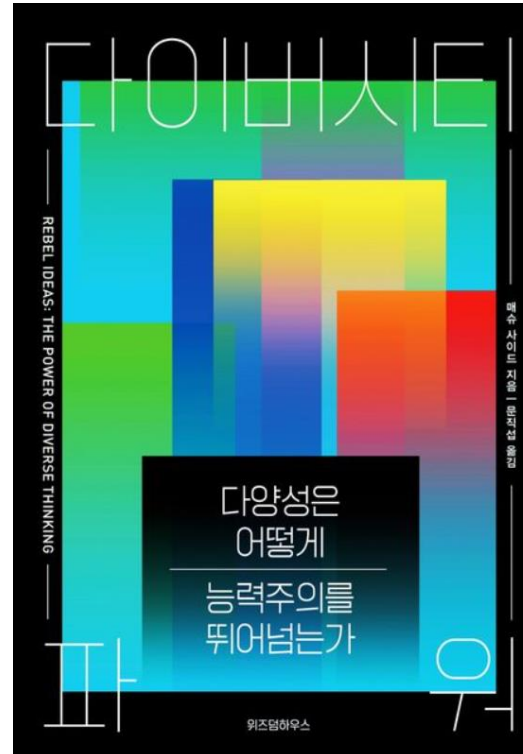
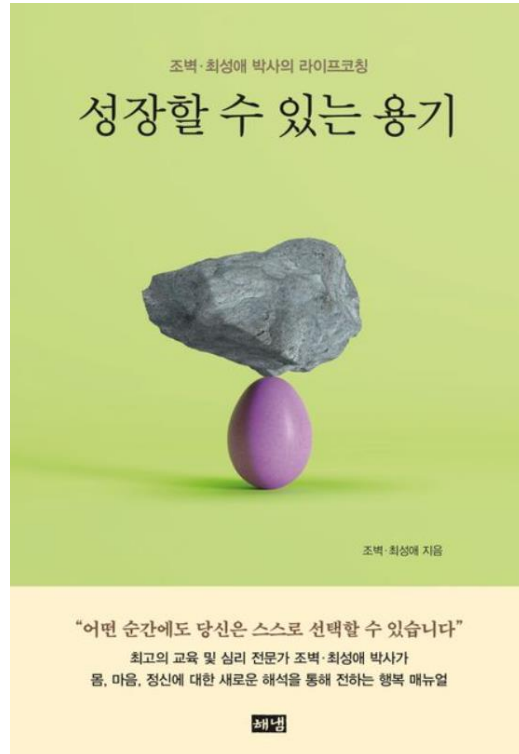
## 인과추론의 위계



## 역자가 추천하는 권장 커리큘럼



## 부탁드리고 싶은 부분



감사합니다

---