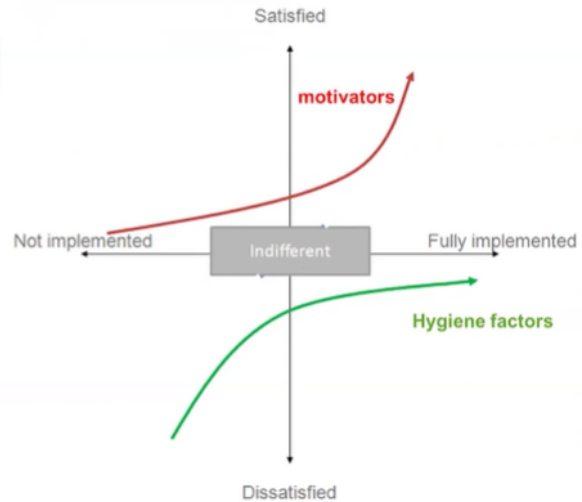
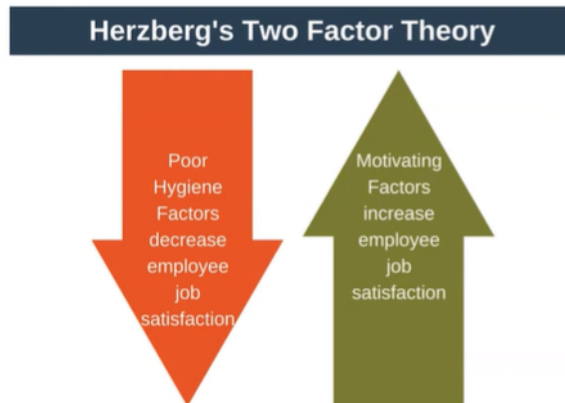


이혜람

[Session 0] 인과추론 온라인 세미나 소개

Why Does Research Methodology Matter?

What if...? 인과추론의 데이터과학



Korea Summer Session on Causal Inference 2021

Introduction to Summer Session on Causal Inference

5

The Right Tool for the Right Question

	Layer (Symbolic)	Typical Activity	Typical Question	Example
\mathcal{L}_1	Associational $P(y x)$	Seeing	What is? How would seeing X change my belief in Y ?	What does a symptom tell us about the disease?
\mathcal{L}_2	Interventional $P(y do(x), c)$	Doing	What if? What if I do X ?	What if I take aspirin, will my headache be cured?
\mathcal{L}_3	Counterfactual $P(y_x x', y')$	Imagining	Why? What if I had acted differently?	Was it the aspirin that stopped my headache?

Table 1.1: Pearl's Causal Hierarchy.

"On Pearl's Hierarchy and the Foundations of Causal Inference," ACM special volume in honor of Judea Pearl (<https://causalai.net/r60.pdf>)

Korea Summer Session on Causal Inference 2021

JUDEA PEARL
WINNER OF THE TURING AWARD
AND DANA MACKENZIE

THE
BOOK OF
WHY

THE NEW SCIENCE
OF CAUSE AND EFFECT

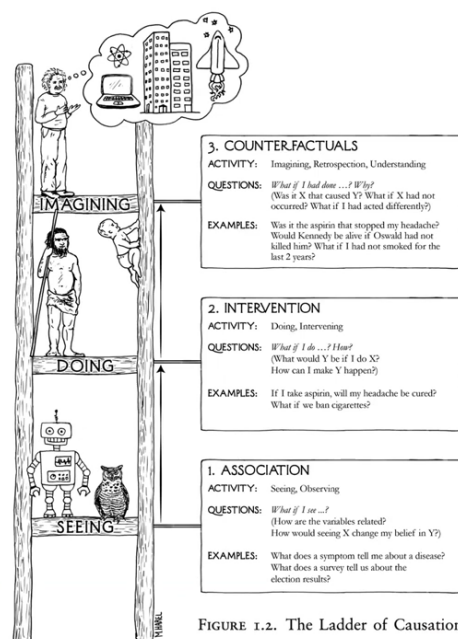


FIGURE 1.2. The Ladder of Causation



End of Document

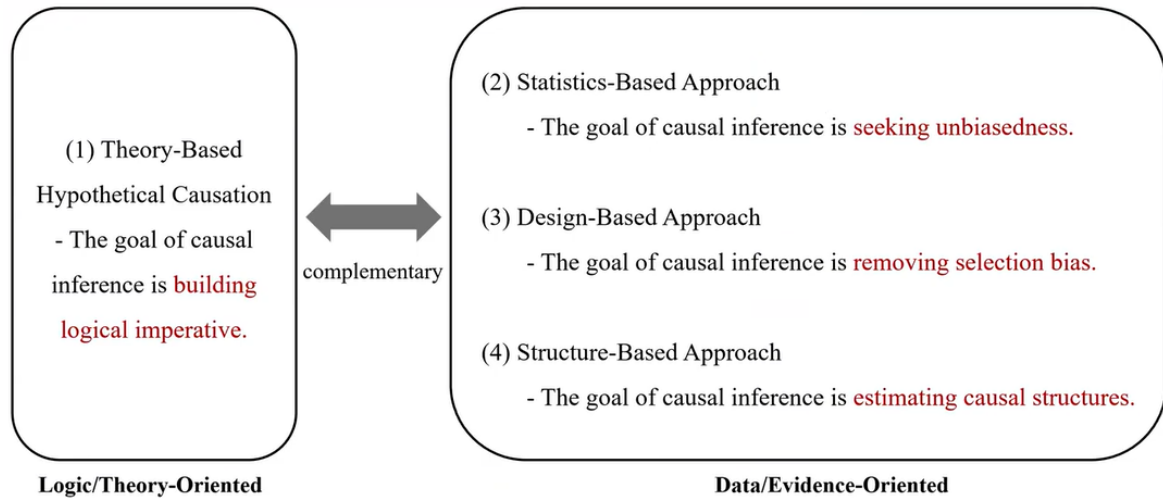
[Session 1-1] 인과추론의 다양한 접근법

데이비드 흄 - 관념적인 확신의 반복, 연결고리 (먹구름 → 비) / (무릎 쏙심 → 비?), 인과론 부정

칸트의 철학 - 흄이 자신에게 큰 영향을 미쳤다고 얘기함

실증주의, 인과관계 프레임워크: 모든 사람들이 공통적으로 인과 관계를 파악/검증할 수 있는 공통의 이해의 틀이 필요하다는 것

이 프레임워크는 공통적으로 합의된 것이 아직 없기 때문에 연사분께서 임의로 작성해 보심



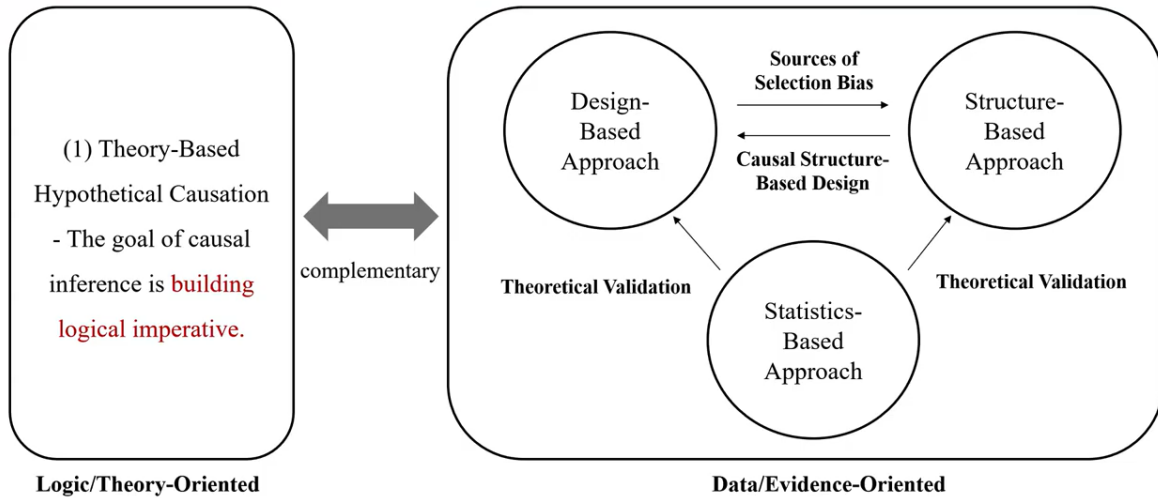
- 1) 이론 기반 : 논리적 명령
- 2) 통계 기반 : unbiasedness 비편향적 추론 / 모수, 표본
- 3) 연구 디자인 기반 : removing selection bias
- 4) 구조 기반 : 전형적인 실증주의 approach

어떤 것이 더 우월하고, 대체 가능한 것은 아님, 서로 상호 보완적으로 작용함

각 3개의 장단점

	통계 기반	디자인 기반	구조 기반
장점	내생성(Endogeneity) 를 평가하고 통계 지표를 계산하는 데에 유용하다.	인과 추론에 대한 깊은 이해 없이도 어느 정도 추론할 수 있다. ex) 아스피린	casual structures나 underlying 메커니즘에 대해 직접적으로 설명함
단점	실제로 데이터 분석 관점에서, '어떻게'에 대한 구체적인 가이드를 주지 못한다. 반면, 디자인, 구조 기반은 어느정도 데이터 분석에 대한 전략을 제시해 준다.	casual structures나 underlying 메커니즘에 대해 설명하지 않음	casual structures에 기반하고 있기 때문에 이를 잘못 산정하면 완전히 잘못된 결과를 얻을 수 있음 → casual structures에 대한 검증이 중요함

따라서 상호 보완적



Korea Summer Session on Causal Inference 2021

Session 1. Potential Outcomes Framework: Causal Mindset

9

[Session 1-2] 잠재적결과 프레임워크 (Potential Outcomes Framework)

디자인 베이스 접근 : 복잡한 통계 방법이 데이터에 적용되어있는가가 아닌, 데이터가 수집 되기 전에 얼마나 연구 디자인 설계가 잘 되어있는지에 따라 성공적 인과 추론이 가능하다는 관점

- Causal effect of the treatment = (Actual outcome for treated if treated) – (Potential outcome for treated if not treated)

- ATE on the ATT = 실제 결과 - 꺾 결과 (Counterfactual)

potential outcome은 모두 관찰할 수 없다.

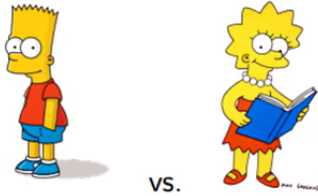
따라서 우리는 그 '꺾'을 안한 사람과 한 사람의 집단을 비교할 수 밖에 없다.

그래서 control group을 최대한 counterfactual과 가깝게 만들 수 밖에 없다.

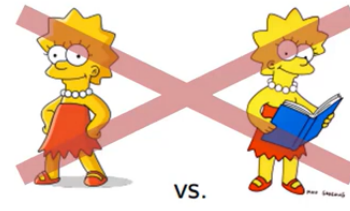
- But, in reality, we can only observe:

$$= (\text{Actual outcome for treated if treated}) - (\text{Actual outcome for untreated if not treated})$$

Estimate we can actually obtain

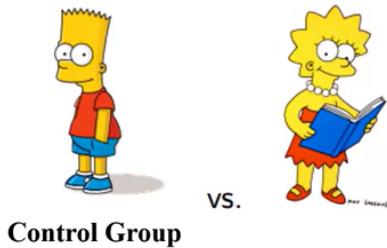


ATT we want to identify

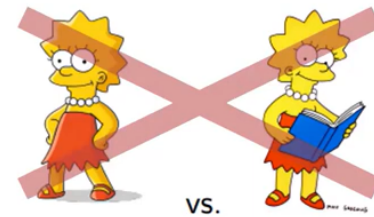


Counterfactual

Estimate we can actually obtain



ATT we want to identify



Counterfactual

책을 읽은 사람과 책을 읽지 않은 사람간에 집안,성격,학벌 등 다양하게 다른 요인이 있다면, 이 차이가 책을 읽었고/읽지 않았고의 차이라고 할 수 있을까?

- Decomposition of causal effect and selection bias

$$\begin{aligned}
 \text{Observed effect of the treatment} &= (\text{Outcome for treated if treated}) - (\text{Outcome for untreated if not treated}) \\
 &= (\text{Outcome for treated if treated}) \\
 &\quad - (\text{Outcome for treated if not treated}) + (\text{Outcome for treated if not treated}) \\
 &\quad - (\text{Outcome for untreated if not treated}) \\
 &= (\text{Outcome for treated if treated}) - (\text{Outcome for treated if not treated}) \\
 &\quad + (\text{Outcome for treated if not treated}) - (\text{Outcome for untreated if not treated})
 \end{aligned}$$

Zero sum

Causal effect

Selection bias

이 남는 것이 선택 편향 (Selection bias)

만약 이 선택 편향을 무시할 수 있을 정도로 적다면, 이 값을 구할 수 있음

이것이 causal mindset 비교 가능한 control group 을 고안하도록 연구를 설계하는 것이 중요하다

[Session 1-3] 인과적 사고방식

Casual Mindset

1) 사회제도가 경제 성장에 미치는 영향

쌍둥이 도시로 Institution and Economic Growth를 비교

2) 1995 시카고 폭염

노스론데일,사우스론데일은 서로 비교 가능했음

어떤 통계적 방법도 없지만 훌륭한 인과 추론 사례

→ potentation outcome 프레임워의 관점에서 인과 추론은 비교 가능한 대을 찾고, 비교가능한 통제 그룹이 있는 리서치 그룹을 보완하는 것이다.

3) 교육

미국 12월에 태어난 아이와 1월에 태어난 아이를 비교해 학교 교육과 미래 소득을 비교, 이처럼 연구 설계가 중요함