

Session 0~1

Session 0 인과추론 세미나 소개

Motivation

- 한글자료의 제공
- 최근의 연구 소개

Herzberg's Two Factor Theory : 불만족의 반대는 만족이 아니라 불만족 아님

1. Hygiene Factor
 - 적절한 데이터분석 방법론과 연구 방법론을 적용하는 것.
 - 즉, 연구 자체의 가치를 제대로 인정받기 위한 기본 요건.
2. Motivating Factor

The Right Tool for the Right Question

Two Paradigms of Data Science / Analysis

- 인과관계에 기반한 Causal inference
 - Design-based approach : 연구 설계를 적절하게 고안하여 인과추론에 방해가 되는 요인을 제거하고자 하는 접근방법
 - Structure-based approach : 인과관계의 구조를 명시적으로 나타내고, 이 구조를 기반해서 인과관계를 추론하는 접근방법
- 상관관계에 기반한 Prediction
 - model fitting approach
 - information filtering approach

Two Frameworks of Causal Inference

- Potential Outcomes Framework
 - 약을 먹었을 때 결과와 먹지 않았을 때의 결과의 차이를 바탕으로 약이 병의 완화에 미치는 인과적 효과를 추론하고자 하는 접근이고, 이러한 추론을 위한 적절한 연구 디자인을 강조하는 framework

- Structural Causal Model
 - 결과에 영향을 미칠 수 있는 여러 요인들의 복합인 그래프를 그리고 그러한 구조를 직접적으로 통제하여 인과관계를 추론하고자 하는 접근방법

Session 1-1 인과추론의 다양한 접근법

Frameworks for Causation

- 모든 사람들이 동일하게 이해하고 동일한 방식으로 검증할 수 있는 수단이 필요하다.
- 그런 역할을 하는 것이 인과관계에 대한 framework(공통의 이해에 대한 틀)

Various Approaches to Causation

1. 이론에 기반한 가설적 인과관계

- 수학적 법칙이나 논리적 추론에 따라서 인과관계가 형성될 수 밖에 없는 당위성 (logical imperative)을 구축하여 인과관계를 밝히고자 하는 접근방법.

2. Statistics-Based Approach

- 통계적 비편향성으로 인과관계를 정의하는 접근방법.
- 모집단에서 여러번 샘플링을 하고 각 샘플에서 원인과 결과에 대한 관계를 추론했을 때 평균적으로 모집단에서 가지고 있는 실제 관계에 가깝게 추론할 수 있다는 방법.
- 인과추론을 방해하는 요인으로써 Endogeneity를 강조하고 있음.
- 장점 : Endogeneity에 대해 수학적으로 계산할 수 있고, 이는 Endogeneity에 대해 얼마나 효과적으로 줄일 수 있는지에 대해 평가하거나 평가를 위한 통계지표를 개발하는 데 있어 유용하다.
- 단점 : 이론적으로는 Endogeneity를 측정할 수 있지만, 실제로 데이터분석 관점에서서는 구체적 가이드를 주지 못한다.

3. Design-Based Approach

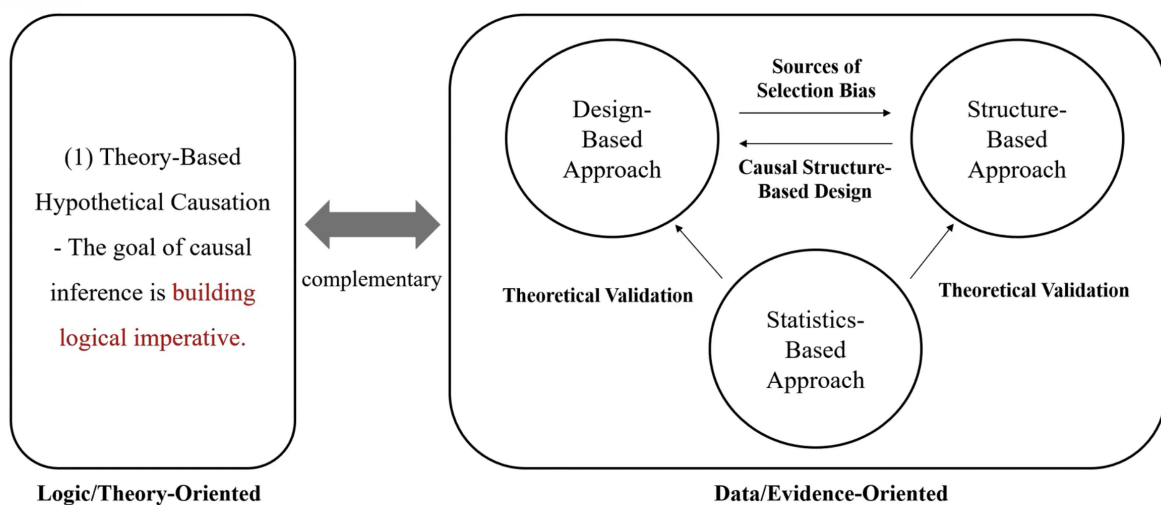
- 인과관계를 연구 디자인적 방법으로 인과관계를 정의.
- 장점 : 적절한 연구 디자인만 설정할 수 있다면, 인과추론에 대한 깊이있는 이해 없이도 충분히 인과적 효과를 추정할 수 있다.
- 단점 : 구체적으로 어떤 인과구조에 대해 직접적으로 알 수 없다.

4. Structure-Based Approach

- 원인과 결과가 서로 얽혀있는 인과적 구조를 추정하고자 하는 접근방법.
- 장점 : 인과구조에 대한 관계를 직접적으로 추정하여 원인이 결과에 어떤 영향을 주는지에 대한 path에 대해 보여줄 수 있다.
- 단점 : 추정 결과는 인과구조에 기반하고 있기 때문에 구조를 잘못산정하게 된다면 완전히 잘못된 결과를 도출할 수 있다.

Various Approaches to Causation

What if...? 인과추론의 데이터과학



Korea Summer Session on Causal Inference 2021

Session 1. Potential Outcomes Framework: Causal Mindset

9

- Statistics-Based Approach는 인과추론에 대한 이론적 기반을 제공할 수 있고, Design-Based Approach와 Structure-Based Approach는 데이터분석의 전략을 고안해줄 수 있다.
- Design-Based Approach는 어떤 요인들이 인과구조를 그리는 데 있어 포함이 되어야 하는지에 대한 인사이트를 줄 수 있고, 반대로 인과구조에 대한 이해는 연구디자인을 고안하는데 있어 중요한 인사이트를 줄 수 있다.
- 즉, 인과추론에 있어 완벽한 framework는 없고, 각각이 서로 상호보완적으로 작동할 수밖에 없다.
- 때문에 서로 다른 접근 방법에 대해 충분히 이해해야 하고, 연구 목적에 맞는 적절한 framework와 상응하는 적절한 data를 수집하고 그에 맞는 적절한 분석방법론을 채택하는 것이 가장 중요하다.

Session 1-2 Potential Outcomes Framework

Design-Based Approach to Causation

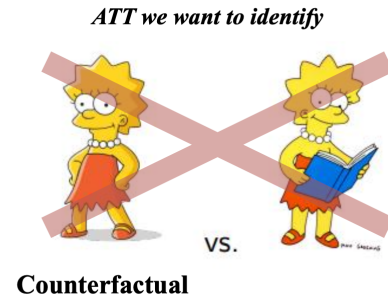
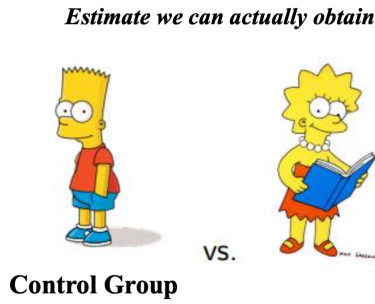
- 연구대상에 행해질 수 있는 구체적인 treatment를 정의할 수 없다면 그것에 대한 인과적 효과도 정의할 수 없다.
- 이런 측면에서는 research design이 중요하다. 즉, 인과추론을 하기위해 중요한 것은 빅데이터나 복잡한 통계모형이 아닌 데이터를 모으기 전에 연구자가 얼마나 적절한 연구 디자인을 고안했는지의 여부이다.

Potential Outcomes Framework

- 특정 treatment의 인과적 효과에 대한 잠재적 결과의 차이로 정의하는 관점이 Potential Outcomes Framework(그때 알았으면 좋았을 걸..)
 - 예를 들어, 독서가 성적에 미치는 인과적인 효과를 알아보기 위해 실제 책을 읽었을 때 성적과 만약 그때 책을 읽지 않았을 때의 잠재적 성적과의 차이를 통해 인과적 효과를 정량화할 수 있다.
 - treatment의 인과적 효과 = treatment를 받았을 때의 결과 - 받지 않았을 때의 잠재적 결과
 - “What if the treatment was not applied?”
 - Causal effect of the treatment = (Actual outcome for treated if treated) - Counterfactual (Potential outcome for treated if not treated)
- Average Treatment Effect (ATE)
on the Treated (ATT)
- Potential Outcomes Framework하에서는 같은 대상에 대한 잠재적 결과를 비교해야 한다.

Fundamental Problem of Causal Inference

- Potential Outcomes Framework의 근본적 문제점은 Potential Outcome을 모두 관찰할 수 없다는 것.
- 결국 가능한 것은 treatment를 받은 outcome과 treatment를 받지 않은 outcome을 비교하는 것.



Selection Bias

- Counterfactual과 Control Group의 차이를 selection bias라고 한다.

- Decomposition of causal effect and selection bias

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

Causal Mindset

- Selection Bias를 줄이는 것이 가장 중요하다.
- 즉, Counterfactual에 최대한 가까운 Control Group를 찾을 수 있는 적절한 연구디자인을 고안하는 것이 핵심 목표.
- 이러한 목표를 Ceteris Paribus라고 한다.

CETERIS PARIBUS:
all other things
being equal



**Treatment Group
w/o Treatment
(Counterfactual)**

Control Group

Session 1-3 Examples of Causal Mindset

(1) Institution and Economic Growth

- 남한과 북한은 사회제도를 제외하고 모두 비교가능한가? 이는 불가능하다. 그러면 비교 가능한 쌍둥이 도시를 찾아서 서로 비교해보면 인과적 효과를 추론할 수 있지 않을까?
- 대표적 사례가 노갈레스라고 하는 두 도시. 사회제도를 제외한 모든 부분이 비슷한 이 두 도시를 통해 서로다른 사회제도가 경제성장에 미치는 영향에 대해 어느정도 추론할 수 있다.

(2) 1995 Chicago Heat Wave

- 원인변수에 대한 디자인을 통해 결과를 바꾸고자 하는 극적인 사례.
- 비교 가능한 두 도시의 폭염피해 차이에 대한 인과관계를 합리적으로 추론할 수 있다.

(3) Returns to Education

- 일란성 쌍둥이는 여러 요인에 대해 유사하기 때문에 서로 비교가능하다. 이 쌍둥이의 교육수준에 따라 장래 수입이 영향을 미치는지 추론할 수 있다.

- 미국 법에 의해 4분기에 태어난 아이들이 1분기에 태어난 아이들보다 학교를 다닌 연수가 더 높다. 이들의 장래 수입을 비교해볼 수 있다.
- 만약 충분히 합리적인 통제변수를 디자인할 수 있다면 회귀분석 만으로도 충분히 훌륭한 인과추론을 할 수 있다.

(4) Difference in Same-Donor Kidneys

- 같은 사람에게서 기증된 두개의 신장은 서로 충분히 비교가능하다.
- 만약 신체적 요인만 영향을 준다면 이 두개의 신장은 서로 유사한 패턴을 보여야 하지만 그렇지 않다.