

[Data analysis] 그림에도 데이터 분석은 불완전하다 (데이터 분석의 힘 chapter.7) — 나무늘보의 개발 블로그

노트북: 첫 번째 노트북

만든 날짜: 2021-05-25 오후 11:42

URL: <https://continuous-development.tistory.com/242>

Data scientist/Data analysis

## [Data analysis] 그림에도 데이터 분석은 불완전하다 (데이터 분석의 힘 chapter.7)

2021. 5. 25. 00:58 수정 삭제 공개



**그림에도 데이터 분석은 불완전하다 : 불량 분석을 피하기 위한 방법**

### 데이터 자체에 문제가 있다면 모든게 무용지물

데이터 자체에 문제가 있으면 아무리 뛰어난 분석 기법을 써도 신뢰성 있는 결과를 내기 어렵다. 다음과 같은 경우 데이터에 문제가 있는것이다..

1. 데이터 측정에 문제가 있고 수치도 바르게 기재되어 있지않다

2. 관측치에 대량의 결측치가 있다.
3. 데이터를 모든 세대가 아닌, 편중된 표본에서만 확보했다.

## 분석 결과의 외적 타당성이라는 문제

RCT나 자연실험으로 발견한 인과관계가 분석 표본 이외의 다른 대상에게도 적용되는지에 대해서는 신중한 논의가 필요하다.

내적 타당성이 확보되었다고 해도 다른 실험 결과를 다른 세대에도 적용할 수 있는지는 다른 문제이다.

분석방법	외적 타당성의 범위 : 개입효과를 분석할 수 있는 대상	내적 타당성의 강도
RCT(강제 참가형)	실험 대상자	매우 높다
RCT(자발적 참가형)	실험 대상자 중 자발적 참가자	매우 높다
RD 디자인	경계선 부근의 대상	높다
집군분석	집적한대상	높다
패널 데이터 분석	개입집단 전체	약간 떨어진다.

이 부분에 대해서는 외적 타당성이 중요하다는 점을 강조하고 싶었다.

## 출력 편향과 협력 관계 편향이라는 문제

외적 타당성과 연관해서 출판 편향과 협력관계 편향이라는 것이 있다.

**출판 편향**은 다음과 같은 상황에서 나타난다. 인과관계를 검증하는 상황에서 분석 결과 영향이 0이었을때 이 결과 자체는 유용한 과학적 발견이다.

하지만 이 결과는 논문으로서 좋은 평가를 받기 어렵다.

또한 실험을 설계하는 단계에서 연구자는 영향을 미쳤을 상황에서 실험 하려는 경향이 있다.

### 협력관계편향

파트너를 고르는 과정 자체가 연구자의 의도나 분석의 실현 가능성에 영향을 받는다면 더 나아가 분석의 외적 타당성까지 흔들리게 된다.

## 개입에 파급효과가 존재할 때의 주의점

RCT나 자연실험은 개입집단과 비교집단을 나누는 것으로 부터 시작한다. 여기서 가정은 개입 집단에 대한 개입이 비교집단에는 영향을 주지 않는다는 가정이 필요하다.  
이것에 대한 영향을 주는 현상을 개입의 파급효과라고 부른다.

실험자는 이것에 대해 어떤 대책을 세울수 있을까?

- 1.개입집단을 어느 수준으로 설정할 지 깊이 생각한다.
- 2.개입효과 외에 개입의 파급효과도 분석할 수 있도록 실험을 설계하는 것이다.

## 일반 균형적 효과가 존재할 때의 주의점

개입이 일반 균형적인 효과를 가져올 경우 소규모 실험의 결과와 대규모 정책의 결과가 다를 수 있으므로 RCT나 자연실험의 성과를 대규모 개입에 활용할 때는 주의해야 한다.

'Data scientist > Data analysis' 카테고리의 다른 글

[Data analysis] 그럼에도 데이터 분석은 불완전하다 (데이터 분석의 힘 chapter.7)

[Data analysis] Google Analytics 란? (기본적인 설명)

[Data analysis] 데이터는 어떻게 전략이 되는가 (데이터 분석의 힘 chapter.6)

[Data analysis] 기본적인 시계열 분석(Time Series Analysis)의 종류와 개념 설명

[Data analysis] 시간의 흐름에 따라 패널 데이터 분석 (데이터 분석의 힘 chapter.5)

[Data analysis] 계단식 변화가 있는 곳엔 집군분석 (데이터 분석의 힘 chapter.4)

데이터 분석의 힘



나아무늘보

혼자 끄적끄적하는 블로그 입니다.