

[Data analysis] 인과 관계와 상관 관계 (데이터 분석의 힘 chapter.1) — 나무늘보의 개발 블로그

노트북: 첫 번째 노트북

만든 날짜: 2021-03-13 오후 6:58

URL: <https://continuous-development.tistory.com/230>

Data scientist/Data analysis

[Data analysis] 인과 관계와 상관 관계 (데이터 분석의 힘 chapter.1)

2021. 3. 11. 00:47 수정 삭제 공개

데이터의 상관 관계는 인과관계가 아니다.

ex) 2010년에 아이스크림 회사가 광고를 했더니 2009년 대비 2010년에 아이스크림 매출량이 40프로 늘었다.

광고 -> 매출 40프로의 인과 관계를 주장할 수 있을까?

다른 요소가 개입됐을수도 있다. 예를 들어 2010년의 무더위 + 2010년의 소비 회복세 등등

인과관계를 입증하기 어려운 이유

1. 다른 요인이 영향을 미쳤을 가능성이 있다.

x (광고) / y (매출 40프로)

X(광고)가 벌어진 것과 같은 시기에 여러가지 일이 일어날수 있다. v(2010년의 무더위 + 2010년의 소비 회복세 등등)

2.인과관계가 반대일 가능성이 있다.

때로는 Y가 X에 영향을 주는 역인과관계의 가능성도 부정할 수 없다.

ex) 2010년 이른 더위로 아이스크림 매출이 오르자 회사는 매출액 상승분으로 인터넷 광고를 시작했다.

인과관계와 상관관계는 다르다.

X가 Y에 영향을 주었을 가능성

Y가 X에 영향을 주었을 가능성

V가 X와 Y 양쪽에 영향을 주었을 가능성

이 세가지 가능성으로 어떤 것이 진짜인지 판단할 수 없다.

상관 관계와 인과관계를 구분하는 전통적인 방법이 있다. 잠복 변수 V 를 최대한 모은 다음 통계 분석에 의해 V의 영향을 배제하는 것이다. 하지만 이러한 방법들은 한계가 있다.

이러한 인과 관계의 문제는 **편향**이라고 한다. 분석으로 얻은 추정치의 치우침을 뜻한다.

데이터가 늘어난다고 한들 이러한 편향 문제는 해결되지 않는다. 이제 이것을 다양한 해결법으로 풀어 나가보겠다.

이 내용들은 **데이터 분석의 힘**이라는 책의 내용을 정리 및 요약한 내용입니다.

'Data scientist > Data analysis' 카테고리의 다른 글

[Data analysis] 최선의 데이터 분석법, RCT (데이터 분석의 힘 chapter.2)

[Data analysis] 인과 관계와 상관 관계 (데이터 분석의 힘 chapter.1)

데이터 분석의 힘

인과관계 상관관계

인과관계와 상관관계의 차이



나아무늘보

혼자 끄적끄적하는 블로그 입니다.