

[Data analysis] 급격한 변화의 경계선을 찾는 RD 디자인 (데이터 분석의 힘 chapter.3) — 나무늘보의 개발 블로그

노트북: 첫 번째 노트북

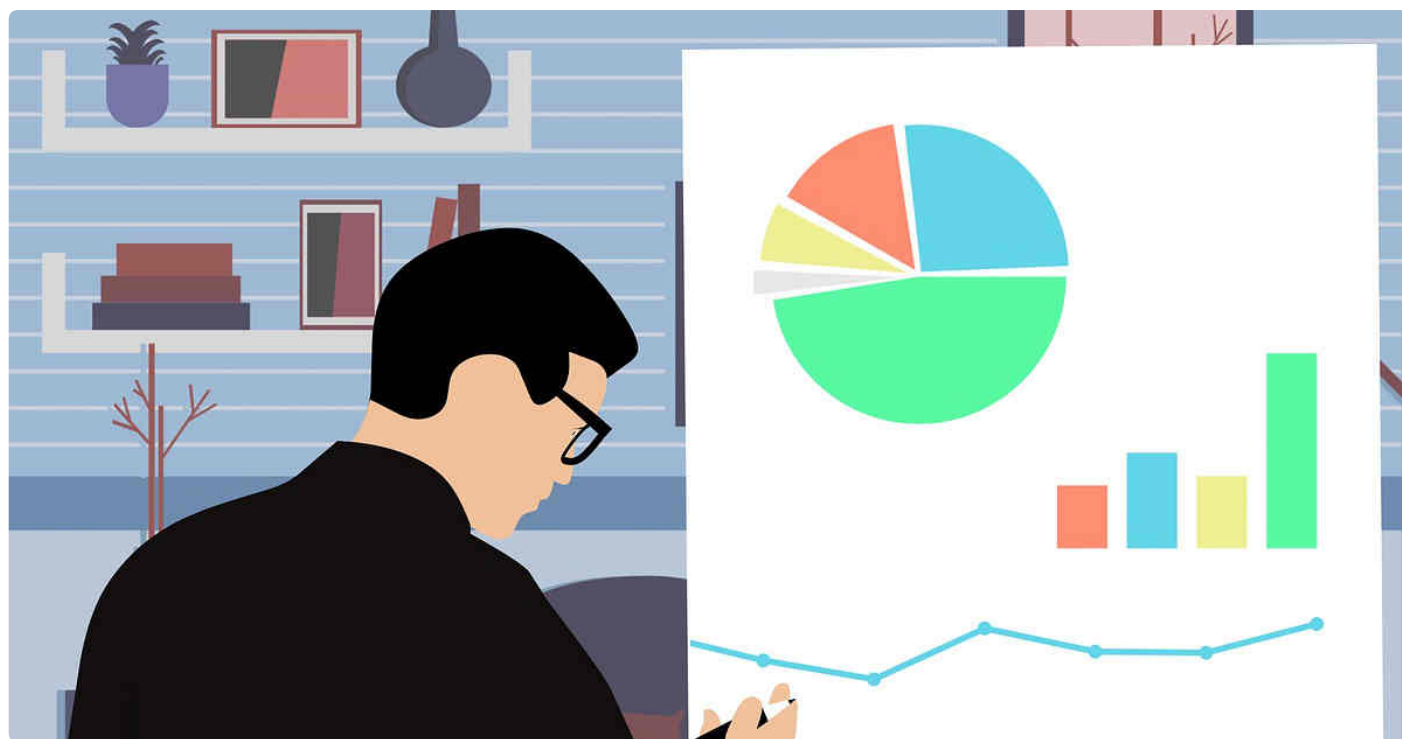
만든 날짜: 2021-03-15 오후 10:52

URL: <https://continuous-development.tistory.com/232>

Data scientist/Data analysis

[Data analysis] 급격한 변화의 경계선을 찾는 RD 디자인 (데이터 분석의 힘 chapter.3)

2021. 3. 15. 09:08 수정 삭제 공개



급격한 변화의 경계선을 찾는 RD 디자인

RCT가 불가능 할 때 인과관계를 밝혀내기 위해 자연 실험 기법을 사용한다.

이것은 실험에 가까운 상황을 이용한다는 발상에서 비롯 되었다.

대표적인 자연실험 기법은 RD 디자인이다.

이것은 RCT와는 보완 관계이므로, 상황이나 문제에 따라 결정된다.

RD디자인은 회귀불연속설계법을 줄인 말이다.

일본의 의료비 부담 사례

일본에서는 70세 이상의 환자는 의료비의 본인 부담금이 30프로에서 10프로로 줄어 든다.

이것을 통해 본인부담금이 의료 서비스 이용에 영향을 미친다는 가설을 세웠다.

일본의 월 연령별 외래환자의 수를 보게 되면 70세를 기점으로 10퍼센트 이상이 증가 된다.

원래의 그래프에서는 완만한 상승곡선이다가 70세를 기점으로 큰 점프가 보인다는 점이다.

여기서 변화는 본인부담금의 변화 외에는 다른 변수는 설명 되지 않는다. 즉 경계선에서의 점프를 통해 의료비 본인부담금(x)이 의료서비스 이용(Y)에 미치는 인과관계를 측정할 수 있다.

이것이 경계선을 활용한 RD디자인의 기본적인 사고 방식이다.

RD 디자인에 필요한 가정

RD 디자인의 가정 - 만약 경계선에서 본인부담금이 변화하지 않는다면
의료 서비스 이용자 수도 점프 하지 않는다

이 가정이 성립하는지를 검증하는 방법은 존재하지 않다. 분석자는 이것
을 가정이 성립 할 것이라고 주장만 할 수 있다.

이것이 RCT와 자연실험의 차이점 이다. RCT에서는 무작위로 집단을 나누
기만 하면 인과관계를 분석하기 위한 가정이 수학적으로 증명되지만 자연
실험에서는 수학적으로 증명하지 못하고

가능성만 제기 할 수 있다.

가정에 대한 확인 과정

나이가 들면 병에 걸릴 가능성이 높아지기 때문에 의료 서비스 이용률
이 높아지는게 당연하다.

=> 확실히 맞는 말이긴 하지만 그전 데이터를 봤을때 65세부터 70세 생
일까지 환자 수가 완만하게 증가 되고 있다. 나이와 건강의 관계가 연속적
이라면 70세를 기점으로 급격히 올라가는 것을 설명 할 수 는 업산.

나이가 들어감에 따라 취업률, 노동시간, 수입 등에 변화가 있을 것

=> 이것도 맞는 말이긴 하지만 그전까지 연속적이던 것이 70세를 기점으로 오른다는 것은 설명할 수 없다

RD 디자인의 가정이 무너지는 때

70세를 기준으로 본인부담금 이외의 무언가가 비연속적으로 변화하여 의료 서비스 이용에 미칠 때

분석의 대상이 그래프 가로축의 변수를 조작할 수 있을 때

경계선 부근에서는 RCT와 비슷해 진다.

RCT와는 달리 데이터 분석자가 두 집단을 만들지는 않았다. 그러나 70 세 생일을 기준으로 본인 부담금이 30퍼센트에서 10퍼센트로 크게 바뀌는 일본의 제도가 마치 실험 같은 상황을 만들어 낸 것이다. 그래서 이것을 자연 실험이라고 불리운다.

RD디자인의 강점과 약점

약점

- RD디자인의 약점은 경계선 부근에 있는 사람에 대한 인과관계만 측정할 수 있다는 점이다.

- 가정이 성립할 것이라는 근거를 제시할 수는 있지만 실제로 성립한다는 것을 입증할 수는 없다.

강점

- RCT를 실시하지 않고도 RCT에 가까운 상황을 만들어낼 수 있다는 점이다. 실험에 비용과 노력이 많이 들어가는 RCT에 비해 이미 확보된 데이터만으로 분석이 가능한 RD 디자인의 장점이다.
- 그래프를 이용해 분석 결과나 가정을 시각적으로 설명해준다.
- RD 디자인을 이용할 수 있는 상황이 많다.

'Data scientist > Data analysis' 카테고리의 다른 글

[Data analysis] 급격한 변화의 경계선을 찾는 RD 디자인 (데이터 분석의 힘 c...

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

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

RD 디자인

RD디자인

자연실험 기법

회귀불연속설계법



나아무늘보

혼자 끄적끄적하는 블로그 입니다.