

인과추론을 위한 매칭 방법 (결과 분석)

통계학과

김지현
한서현
명세훈

Table of Contents

- | | | |
|--|-------------------------------------|-----------------------------------|
| <p>1
매칭 방법 이후 결과 분석이 어떻게 진행되어야 하는가</p> | <p>4
세분화(Subclassification)</p> | <p>7
불확실성을 고려해야 하는가</p> |
| <p>2
K:1 매칭 이후</p> | <p>5
ATT vs ATE</p> | <p>8
추정된 성향 점수를 사용할 때의 이점</p> |
| <p>3
교체 매칭</p> | <p>6
여러 하위집단 통합</p> | |

매칭 방법 이후 결과 분석이 어떻게 진행되어야 하는가

매칭 방법 자체는 인과추정 방법이 아니다.
매칭은 공변량이 적절하게 균형을 이루도록 만드는 것이다.

무작위 실험에서의 회귀 보정과 유사하다.
(= 공변량이 통제된 상태에서 처치 효과를 비교)

특정 결과 모형의 세부 사양에 대한 처리효과 추정치의
민감도를 낮춘다.
(= 공변량을 기준으로 처치군과 대조군이 비슷하기 때문)

매칭의 목적은 모형 의존성을 줄이는 것이다.

매칭 방법 이후 결과 분석이 어떻게 진행되어야 하는가

일반최소자승

- 모든 관측값의 오차 분산이 동일하다
- 잔차 제곱의 단순 합을 최소화
- 관측값의 신뢰도가 동등하다고 간주하지만, 이분산성이 존재하면 초정치가 비효율적

가중최소자승

- 관측값에 따라 오차 분산이 다르다
- 잔차 제곱의 합을 최소화할 때 가중치를 반영
- 가중치를 회귀 모형에 적용하면 처리 그룹과 통제 그룹의 공변량 분포가 인위적으로 균형을 이룸

비모수적 순열 기반 검정

-> 관측값을 무작위로 재배열하여 가설 하에서의 분포를 생성하고, 이를 통해 관찰된 결과의 통계적 유의성을 평가

주변처리효과

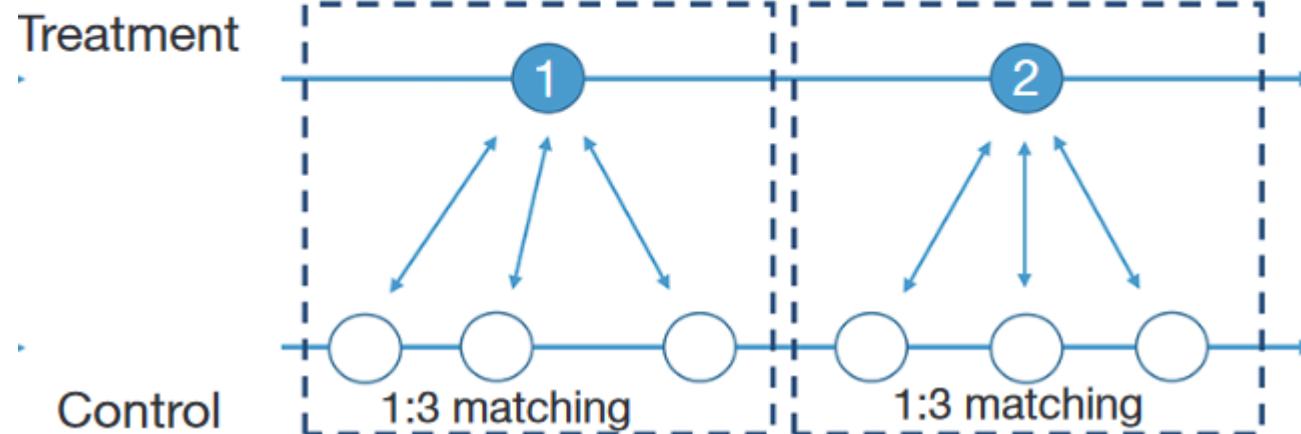
평균 차이: 처리 그룹의 평균 결과와 통제 그룹의 평균 결과 간의 차이

비율 차이: 처리 그룹의 성공 비율과 통제 그룹의 성공 비율 간의 차이

모형적 접근 뿐 아니라 비모수적 순열 기반 검정의 사용이나 주변처리효과와 같은 방법들도 전체 집단의 효과를 추정하는데 매우 강력하다

K:1 매칭 이후

F



<결과 분석>

1. 매칭된 표본을 사용하여 진행한다.
2. 표본이 무작위 배정에 의해 생성된 것처럼 분석한다.
(= 매칭 후에는 무작위 실험처럼 분석해야 한다.)
3. 매칭쌍을 고려해야 하는지에 대해 생각해봐야 한다.

매칭쌍: 매칭을 통해 서로 대응되는 치치군과 대조군의 묶음

기본 매칭쌍은 1대1
- K:1 매칭인 경우, k개의 대조군을 하나의 묶음으로 생각한다.

매칭쌍을 고려할 필요가 없는 두 가지 이유

1. 공변량을 회귀모형에서 조건부로 포함하면 충분하다.

매칭 과정에서 사용한 공변량들을 회귀식에 그대로 넣고 보정하면 매칭쌍 단위로 분석할 필요가 없다.

2. 성향점수는 쌍 단위로 완벽한 매칭을 보장하지 않는다.

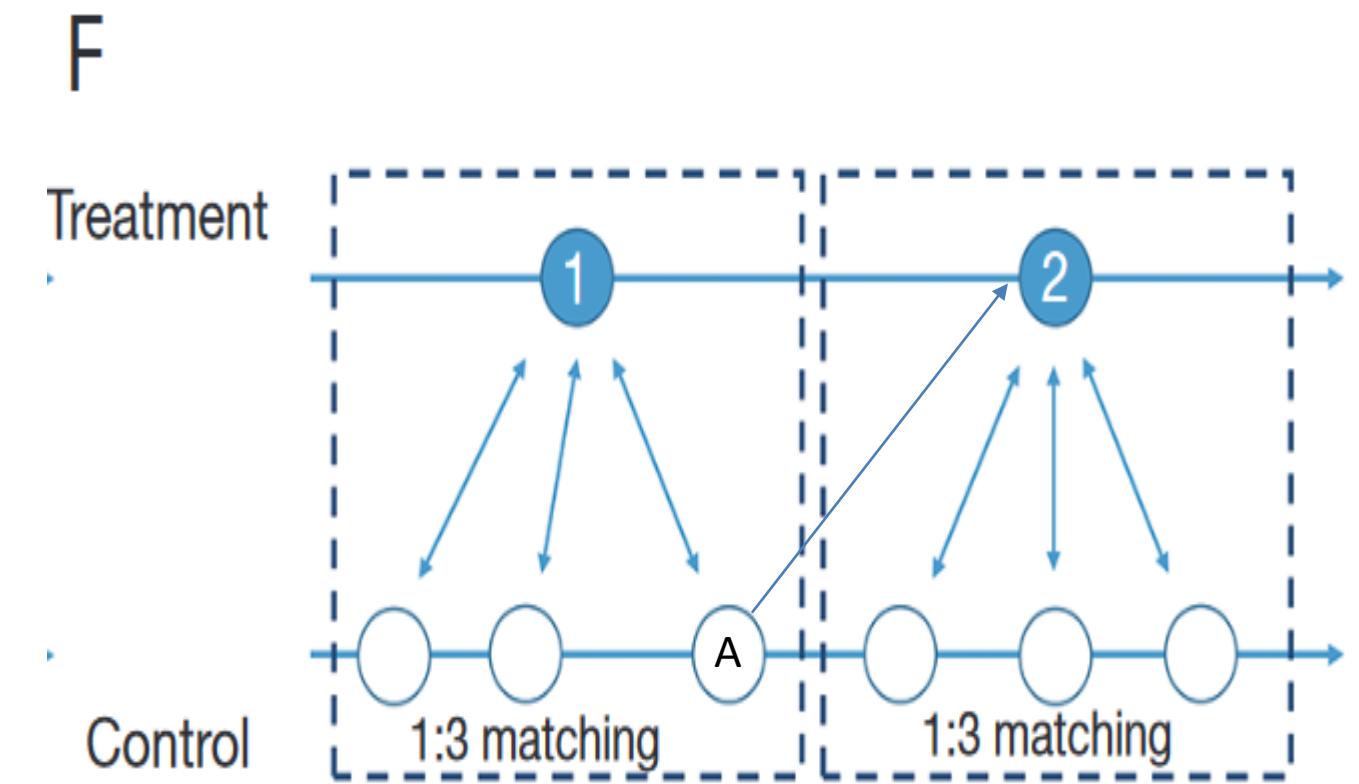
각 쌍이 공변량 전체에서 동일한지는 보장하지 않지만 공변량 분포가 비슷하다는 것은 보장한다. 따라서 공변량이 적절히 균형을 이루었는지를 판단할 수는 있지만 개인의 각각의 특성이 동일하다는 뜻은 아니다.

매칭된 처리군과 대조군 전체를 합쳐서 분석하는 것이 흔한 방법이다.
즉, 매칭된 사람들 전부를 하나의 데이터셋으로 합쳐서 분석하라는 뜻이다.

교체 매칭

같은 대조군 개체를 여러 번 사용할 수 있는 매칭 방식

하나의 대조군이 여러 처리군에게 반복 매칭되면 그 대조군은 매칭된 횟수만큼
빈도를 부여한다.



위의 사진처럼 대조군 A가 처치군 1과 2 모두에게 매칭이 되었다면,
A의 가중치는 1/2이 된다.

가중치를 분석에 넣어야 불편(unbiased) 추정이 된다. = 대조군이 각 처치군에게 공정하게 기여함을 보여주고 분석의 왜곡을 막아준다.

세분화(Subclassification)

정의

- 전체 데이터를 여러 하위 집단(성향점수 기준)으로 분할
- 각 하위집단에서 효과 추정
- 가중평균으로 전체 효과 계산

예시

- 성향점수 기준 5개 그룹 분할:
- 1집단: 성향점수 0~0.2 (처치 확률 낮음)
 - 2집단: 성향점수 0.2~0.4
 - 3집단: 성향점수 0.4~0.6
 - 4집단: 성향점수 0.6~0.8
 - 5집단: 성향점수 0.8~1.0 (처치 확률 높음)

ATT vs ATE

ATT

- 하위집단별 '처리군' 인원수로
가중치를 줄 때 이용
- 실제 처치를 받은 집단에서의 평균
효과

ATE

- 전체 하위집단 인원수로 가중치를 줄
때 이용
- 전체 집단(처치군과 비처치군 모두)에
대한 평균 효과

예시: 학생 프로그램

가정

- 하위집단1: 처리군 20명, 대조군 10명, 효과 +5점
- 하위집단2: 처리군 30명, 대조군 40명, 효과 +2점

ATT(처치군 중심)

- 프로그램 처치를 받은 학생들의 평균 효과
- $(20/50) \times 5 + (30/50) \times 2 = 3.2$ 점
- > 프로그램 참가자(처치군)의 평균 효과

ATE(전체 집단 중심)

- 전체 학생(참가 + 비참가 모두)의 평균 효과
- $(30/100) \times 5 + (70/100) \times 2 = 2.9$ 점
- > 전체 학생의 평균 프로그램 효과

- ATE는 참가자 효과 ATE는 전체 집단 효과임을 알 수 있음.
- 또한 연구 목적에 따라 누구에게 대표성을 부여하느냐에 따라 평균 효과 계산과 해석이 달라질 수 있음.

여러 하위집단 통합

공동 모델(Joint Model)구성

$$Y_{ij} = \beta_{0j} + \beta_{1j}T_{ij} + \gamma X_{ij} + e_{ij}$$

i: 개별학생, j: 하위집단

T_{ij} : 처치 여부(참/비참)

X_{ij} : 공변량(학생 특성 등)

β_{1j} : 각 집단 내 처치 효과

γ : 공변량 효과가 모든 집단에서 동일하다고 가정

결과 결합 방법

각 하위 집단의 처치 효과(β_{1j})를 집단 인원 비율로
가중평균하여 전체 효과 산출

$$\text{전체 효과} = \sum_{j=1}^J \left(\frac{N_j}{N} \right) \beta_{1j}$$

N_j =집단 인원수, N=전체 인원수

- 하위 집단 수가 많을 때 공동 모델은 계산적 통계적으로 효율적인 해법
- 집단별 추정치를 비중(가중치)대로 합쳐 전체 효과 구할 수 있음
- ATT/ATE 구분은 가중치 기준(참가자/전체인원 수 차이)

불확실성을 고려해야 하는가

불확실성을 무시해야 한다

- 성향 점수를 추정하는데 사용된 공변량이 실제 무작위 배경의 기초가 된 것처럼 가정
- 성향 점수 자체를 추정하는 과정에서 발생하는 불확실성을 무시
- 추정된 성향 점수를 참 성향 점수인 것처럼 매칭, 가중치 부여 등에 사용

불확실성을 고려해야 한다

- 성향 점수가 자료로부터 추정된 것이기 때문에 필연적으로 오류와 불확실성 포함
- 불확실성을 처리 효과 추정 및 표준 오차 계산에 모두 통합해야 한다고 주장

추정된 성향 점수를 사용하는 것(불확실성을 무시)은 성향 점수 추정의 이득을 반영하지 못해 분산이 과대추정됨
필요 이상의 넓은 신뢰구간, 보수적 결과를 제공

추정된 성향 점수를 사용할 때의 이점

분산이 줄어드는 효과

- 매칭은 기본적으로 공변량에 대해 조건부로 처리 효과를 추정
- 추정된 성향 점수를 통해 매칭하면 해당 표본에 존재하는 우연한 공변량 불균형까지 줄이는 효과
- 공변량의 평균 차이가 줄어들어 처리군의 변동성과 처리 효과 추정치의 분산이 함께 감소
- 추정된 성향 점수가 참 성향 점수보다 표본의 불균형을 더 잘 잡아내기 때문에 최종 결과 추정치의 분산이 더 작아지게 됨

정밀도의 증가

- 가정: 실제, 추정된 성향 점수 둘 다 편향 제거에 성공
- 추정된 성향 점수를 사용하는 것이 분산을 감소
- 추가적인 분산 감소 이득으로 인해 정밀도가 높아짐(통계적 검정력의 증가)

감사합니다