

인과추론과 실무 : 성향점수

가짜연구소 인과추론팀

발표자 : 손지우

1. 복습

핵심개념 복습

(1) 잠재적 결과 (Counterfactual)

$Y_i(1)$ = 만일 i 번째 환자가 신약을 투여받았다면 나타날 잠재적 결과

$Y_i(0)$ = 만일 i 번째 환자가 위약을 투여받았다면 나타날 잠재적 결과

Y_i = i 번째 환자의 실제 관측값

$$Y_i = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$$

where $Z_i = 1$ if 신약 투여, $Z_i = 0$ if 위약 투여

핵심개념 복습

(2) 무작위 배정 (Random Assignment)

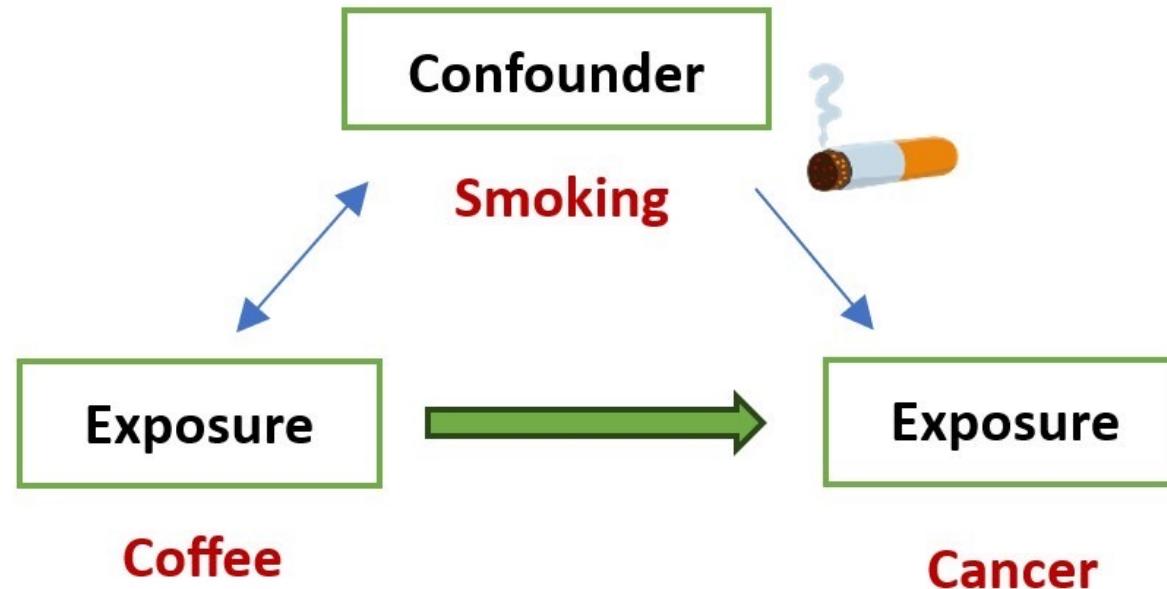
$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp Z_i$$

시험군으로 배정되는 경우($Z_i = 1$)와 대조군으로 배정되는 경우($Z_i = 0$)는 랜덤하게 결정되므로,
 Z_i 의 값은 각 환자들의 $Y_i(1)$ 및 $Y_i(0)$ 과 아무런 연관이 없다.

* 유의사항: $Y_i \perp\!\!\!\perp Z_i$ 과는 다른 의미이다.

핵심개념 복습

(3) 교란 변수 (confounder)



공변량(Covariates)과 구분하기!

cf) $X_i = (X_1, \dots, X_p)$: i 번째 환자의 p 개의 공변량(나이, 성별, 흡연 여부 등)

2. 성향점수 Overview

성향점수란?

1. What : 성향점수의 정의

2. Why : 왜 성향점수를 활용하는가

3. How : 어떻게 성향점수를 계산하는가

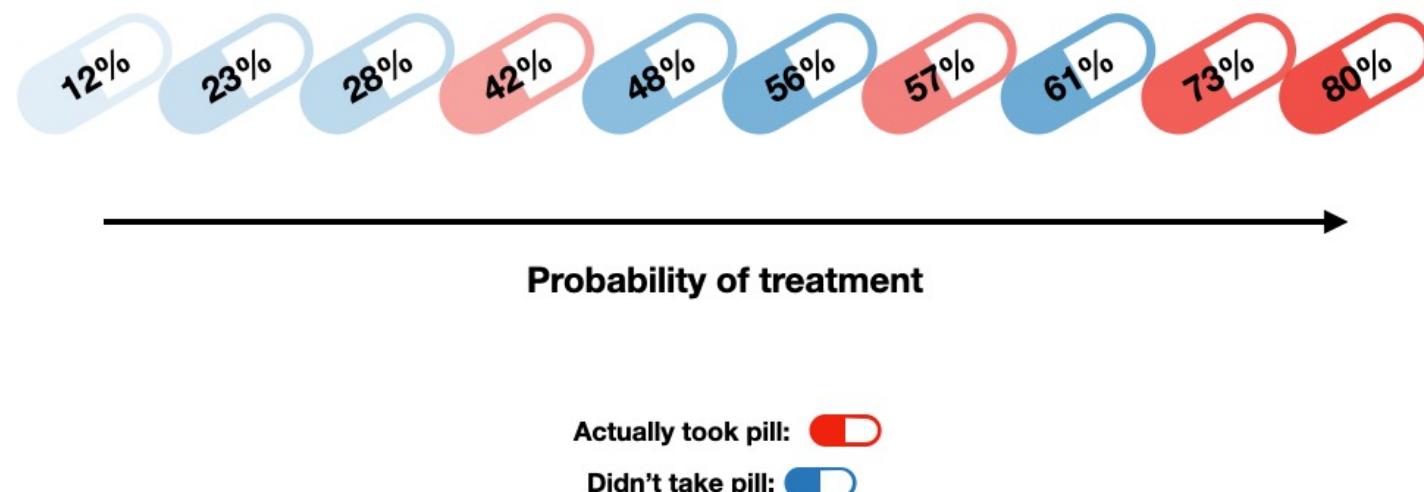
성향점수 정의

$$e(X_i) = P[Z_i = 1 | X_i]$$

공변량 X_i 가 주어진 경우, 각 환자들이 시험군에 속할 조건부 확률

성향점수 정의

Propensity Score



성향점수 정의

공변량 X_i 가 주어진 경우, 각 환자들이 시험군에 속할 조건부 확률

Q. 왜 이런 정의가 필요할까?

성향점수가 같은 사람들끼리 모아준다.

= 모인 사람들끼리는 시험군에 배정될 확률이 같다.

= 모인 사람들의 '각' 그룹은 무작위배정 실험을 수행한 것과 같다.

성향점수는 흔히 '공변량의 균형을 보존해주는 방법'이라고 소개됩니다.

그런데 공변량의 균형을 보존하는 것이 왜 중요할까?

왜 성향점수를 활용하는가?



**OBSERVATIONAL
STUDY**

QuestionPro

실험이 아닌 '관찰' 상황에서도 인과추론을 필요로 하기 때문

왜 성향점수를 활용하는가?



관찰연구의 가장 큰 특징

: 무작위 배정이 이루어지지 않았다는 점 (=편향이 존재할 수 있다.)

(Point)

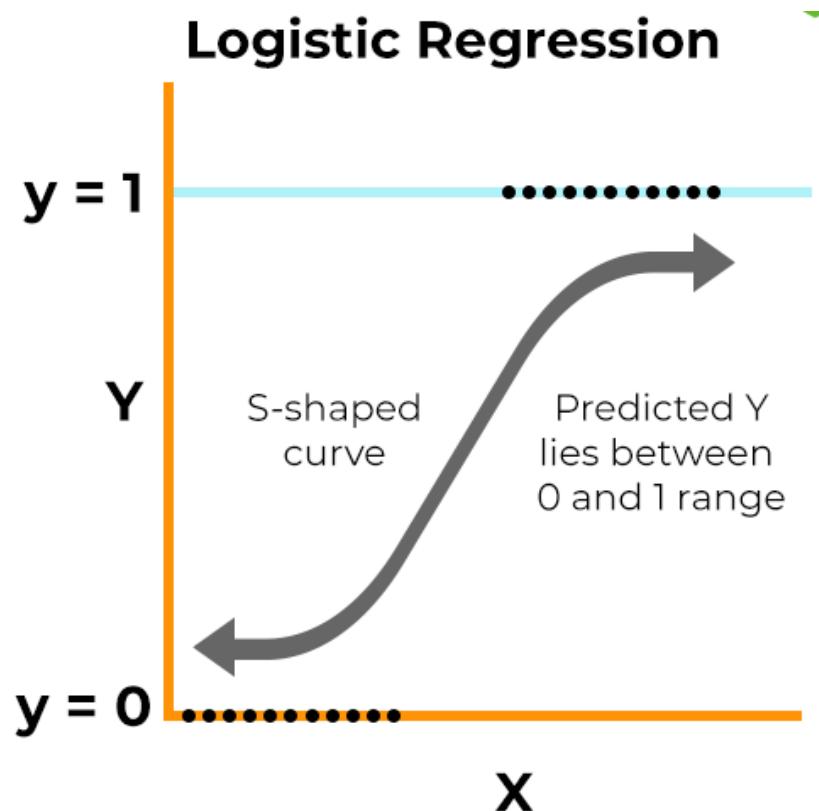
무작위배정이 이루어지지 않은 시험군과 대조군으로 이루어진 연구에 모두 적용 가능하다.

즉, 성향점수 분석에 사용되는 데이터는 실제로 '관찰'을 한 데이터일 필요는 없다.

어떻게 성향점수를 계산하는가?

성향점수는 알 수 없는 미지의 값이므로, 관측된 자료 $(Z_i, X_i), i = 1, 2, \dots, n$ 를 이용하여 추정하게 된다.

가장 흔한 방법 : Logistic Regression



$$e(\mathbf{X}_i) = \frac{1}{1 + \exp(-\mathbf{X}_i^t \boldsymbol{\beta})}$$

$$\ln\left(\frac{e(\mathbf{X}_i)}{1 - e(\mathbf{X}_i)}\right) = \mathbf{X}_i^t \boldsymbol{\beta}$$

3. 성향점수에 필요한 가정

1) Consistency Assumption

$$Y_i = Z_i Y_i(1) + (1 - Z_i) Y_i(0)$$

너무나 당연한 가정 아니야?

2) No unmeasured Confounders

$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp Z_i | \mathbf{X}_i$$

where $\mathbf{X}_i = (X_1, \dots, X_p)$: i 번째 환자의 p개의 공변량(나이, 성별, 흡연 여부 등)

3) Positivity

$$P(Z = z | X = x) > 0 \text{ for all } (z, x)$$

$P(Z = z | X = x) = 0$ 이면 성향점수 분석을 하는 데에 있어서 논리적 결함이 생긴다.

4) SUTVA

Stable Unit Treatment Value Assumption

한 개인의 결과가 다른 사람의 결과에 영향을 미치지 않는다.

- 1) Consistency Assumption**
- 2) No unmeasured Confounders**
- 3) Positivity**
- 4) SUTVA**

이중에서 가장 중요한 가정은?

4. 시뮬레이션 예시

시뮬레이션 데이터 설명

#1. 노인 남성				
	사망	생존	총합	사망률
시험군	32,001	48,121	80,122	39.9%
대조군	7,980	11,898	19,878	40.1%
#2. 노인 여성				
	사망	생존	총합	사망률
시험군	24,012	55,712	79,724	30.1%
대조군	6,130	14,146	20,276	30.2%
#3. 젊은 남성				
	사망	생존	총합	사망률
시험군	3,969	16,070	20,039	19.8%
대조군	16,088	63,873	79,961	20.1%
#4. 젊은 여성				
	사망	생존	총합	사망률
시험군	1,993	17,638	19,631	10.2%
대조군	8,051	72,318	80,369	10.0%

- 네 개의 층은 10만명의 환자들이 모집되었다.
- '각 층별로' 별도로 무작위 배정을 실시하였다. (관찰연구와 유사한 세팅)
 - (2-1) 노인남성과 노인여성은 시험군에 배정될 확률이 80%
 - (2-2) 젊은남성과 젊은여성은 시험군에 배정될 확률이 20%
- 시험약의 효과는 없다고 가정
= 시험군과 대조군의 사망률은 같다고 가정
- 두 개의 공변량 중 하나만이 교란 변수이다.
 - (4-1) 공변량: 나이(교란 변수), 성별
 - (4-2) 시험군에 배정될 확률이 오직 나이와만 연관됨
 - (4-3) 사망률은 나이와 성별 모두와 연관됨

*주의사항

나이와 성별을 무시하고 시험군과 대조군에 무작위 배정한 결과가 아니다.

시뮬레이션 데이터 설명

#1. 노인 남성				
	사망	생존	총합	사망률
시험군	32,001	48,121	80,122	39.9%
대조군	7,980	11,898	19,878	40.1%
#2. 노인 여성				
	사망	생존	총합	사망률
시험군	24,012	55,712	79,724	30.1%
대조군	6,130	14,146	20,276	30.2%
#3. 젊은 남성				
	사망	생존	총합	사망률
시험군	3,969	16,070	20,039	19.8%
대조군	16,088	63,873	79,961	20.1%
#4. 젊은 여성				
	사망	생존	총합	사망률
시험군	1,993	17,638	19,631	10.2%
대조군	8,051	72,318	80,369	10.0%

- 네 개의 층은 10만명의 환자들이 모집되었다.
- '각 층별로' 별도로 무작위 배정을 실시하였다. (관찰연구와 유사한 세팅)
 - (2-1) 노인남성과 노인여성은 시험군에 배정될 확률이 80%
 - (2-2) 젊은남성과 젊은여성은 시험군에 배정될 확률이 20%
- 시험약의 효과는 없다고 가정
= 시험군과 대조군의 사망률은 같다고 가정
- 두 개의 공변량 중 하나만이 교란 변수이다.
 - (4-1) 공변량: 나이(교란 변수), 성별
 - (4-2) 시험군에 배정될 확률이 오직 나이와만 연관됨
 - (4-3) 사망률은 나이와 성별 모두와 연관됨

*주의사항

나이와 성별을 무시하고 시험군과 대조군에 무작위 배정한 결과가 아니다.

시뮬레이션 데이터 분석

#1. 노인 남성

	사망	생존	총합	사망률
시험군	32,001	48,121	80,122	39.9%
대조군	7,980	11,898	19,878	40.1%

#2. 노인 여성

	사망	생존	총합	사망률
시험군	24,012	55,712	79,724	30.1%
대조군	6,130	14,146	20,276	30.2%

#3. 젊은 남성

	사망	생존	총합	사망률
시험군	3,969	16,070	20,039	19.8%
대조군	16,088	63,873	79,961	20.1%

#4. 젊은 여성

	사망	생존	총합	사망률
시험군	1,993	17,638	19,631	10.2%
대조군	8,051	72,318	80,369	10.0%

성향점수의 직관을 이해하기 위하여

옆의 표가 '관찰연구'에서 수집된 데이터라고 생각해보자.

그리고 우리는 이미 정답을 알고 있습니다.

"시험약은 효과가 없다."

시뮬레이션 데이터 분석

#1. 노인 남성				
	사망	생존	총합	사망률
시험군	32,001	48,121	80,122	39.9%
대조군	7,980	11,898	19,878	40.1%
#2. 노인 여성				
	사망	생존	총합	사망률
시험군	24,012	55,712	79,724	30.1%
대조군	6,130	14,146	20,276	30.2%
#3. 젊은 남성				
	사망	생존	총합	사망률
시험군	3,969	16,070	20,039	19.8%
대조군	16,088	63,873	79,961	20.1%
#4. 젊은 여성				
	사망	생존	총합	사망률
시험군	1,993	17,638	19,631	10.2%
대조군	8,051	72,318	80,369	10.0%

나이가 유일한 교란변수라고 가정한다면

해당 데이터는 관찰연구에서 얻어진 자료라고 하더라도

'각 층은' 무작위배정 실험으로부터 얻어진 것으로 간주할 수 있게 된다.

시뮬레이션 데이터 분석

#1. 노인 남성				
	사망	생존	총합	사망률
시험군	32,001	48,121	80,122	39.9%
대조군	7,980	11,898	19,878	40.1%
#2. 노인 여성				
	사망	생존	총합	사망률
시험군	24,012	55,712	79,724	30.1%
대조군	6,130	14,146	20,276	30.2%
#3. 젊은 남성				
	사망	생존	총합	사망률
시험군	3,969	16,070	20,039	19.8%
대조군	16,088	63,873	79,961	20.1%
#4. 젊은 여성				
	사망	생존	총합	사망률
시험군	1,993	17,638	19,631	10.2%
대조군	8,051	72,318	80,369	10.0%



Case #1. 나이와 성별을 무시하고 합친 경우

ALL IN ONE				
	사망	생존	총합	사망률
시험군	61,975	137,541	199,516	31.1%
대조군	38,249	162,235	200,484	19.1%

시험약의 효과가 있다고 결론 내려도 될까?

안된다면, 왜 안 될까?

시뮬레이션 데이터 분석

#1. 노인 남성				
	사망	생존	총합	사망률
시험군	32,001	48,121	80,122	39.9%
대조군	7,980	11,898	19,878	40.1%
#2. 노인 여성				
	사망	생존	총합	사망률
시험군	24,012	55,712	79,724	30.1%
대조군	6,130	14,146	20,276	30.2%
#3. 젊은 남성				
	사망	생존	총합	사망률
시험군	3,969	16,070	20,039	19.8%
대조군	16,088	63,873	79,961	20.1%
#4. 젊은 여성				
	사망	생존	총합	사망률
시험군	1,993	17,638	19,631	10.2%
대조군	8,051	72,318	80,369	10.0%



Case #2. 성별은 무시하고 나이로만 합친 경우

#1. 노인				
	사망	생존	총합	사망률
시험군	56,013	103,833	159,846	35.0%
대조군	14,110	26,044	40,154	35.1%
#2. 젊은이				
	사망	생존	총합	사망률
시험군	5,962	33,708	39,670	15.0%
대조군	24,139	136,191	160,330	15.1%

시험약의 효과가 없다는 올바른 결과로 이어진다.

왜 그럴까?

시뮬레이션 데이터 분석

1. 네 개의 층은 10만명의 환자들이 모집되었다.

2. '각 층별로' 별도로 무작위 배정을 실시하였다. (관찰연구와 유사한 세팅)

(2-1) 노인남성과 노인여성은 시험군에 배정될 확률이 80%

(2-2) 젊은남성과 젊은여성은 시험군에 배정될 확률이 20%

3. 시험약의 효과는 없다고 가정

= 시험군과 대조군의 사망률은 같다고 가정

4. 두 개의 공변량 중 하나만이 교란 변수이다.

(4-1) 공변량: 나이(교란 변수), 성별

(4-2) 시험군에 배정될 확률이 오직 나이와만 연관됨

(4-3) 사망률은 나이와 성별 모두와 연관됨

Case #2. 성별은 무시하고 나이로만 합친 경우

#1. 노인				
	사망	생존	총합	사망률
시험군	56,013	103,833	159,846	35.0%
대조군	14,110	26,044	40,154	35.1%
#2. 젊은이				
	사망	생존	총합	사망률
시험군	5,962	33,708	39,670	15.0%
대조군	24,139	136,191	160,330	15.1%

시험약의 효과가 없다는 올바른 결과로 이어진다.

왜 그럴까?

시뮬레이션 데이터 분석

2. '각 층별로' 별도로 무작위 배정을 실시하였다. (관찰연구와 유사한 세팅)

(2-1) 노인남성과 노인여성은 시험군에 배정될 확률이 80%

(2-2) 젊은남성과 젊은여성은 시험군에 배정될 확률이 20%

"시험군에 배정될 확률이 같은 층끼리만 합침"



$$\{Y_i(1), Y_i(0)\} \perp\!\!\!\perp Z_i | X_i$$

where $X_i = (X_1, \dots, X_p)$: '교란변수'를 포함한 공변량

Case #2. 성별은 무시하고 나이로만 합친 경우

#1. 노인				
	사망	생존	총합	사망률
시험군	56,013	103,833	159,846	35.0%
대조군	14,110	26,044	40,154	35.1%
#2. 젊은이				
	사망	생존	총합	사망률
시험군	5,962	33,708	39,670	15.0%
대조군	24,139	136,191	160,330	15.1%

시험약의 효과가 없다는 올바른 결과로 이어진다.

왜 그럴까?

시뮬레이션 데이터 분석

#1. 노인 남성				
	사망	생존	총합	사망률
시험군	32,001	48,121	80,122	39.9%
대조군	7,980	11,898	19,878	40.1%
#2. 노인 여성				
	사망	생존	총합	사망률
시험군	24,012	55,712	79,724	30.1%
대조군	6,130	14,146	20,276	30.2%
#3. 젊은 남성				
	사망	생존	총합	사망률
시험군	3,969	16,070	20,039	19.8%
대조군	16,088	63,873	79,961	20.1%
#4. 젊은 여성				
	사망	생존	총합	사망률
시험군	1,993	17,638	19,631	10.2%
대조군	8,051	72,318	80,369	10.0%



Case #3. 나이는 무시하고 성별로만 합친 경우

#1. 남성				
	사망	생존	총합	사망률
시험군	35,970	64,191	100,161	35.9%
대조군	24,068	75,771	99,839	24.1%
#2. 여성				
	사망	생존	총합	사망률
시험군	26,005	73,350	99,355	26.2%
대조군	14,181	86,464	100,645	14.1%

시험약의 효과가 있다고 결론 내려도 될까?

안된다면, 왜 안 될까?

시뮬레이션 요약

Case #2. 성별은 무시하고 나이로만 합친 경우

#1. 노인				
	사망	생존	총합	사망률
시험군	56,013	103,833	159,846	35.0%
대조군	14,110	26,044	40,154	35.1%
#2. 젊은이				
	사망	생존	총합	사망률
시험군	5,962	33,708	39,670	15.0%
대조군	24,139	136,191	160,330	15.1%

Case #3. 나이는 무시하고 성별로만 합친 경우

#1. 남성				
	사망	생존	총합	사망률
시험군	35,970	64,191	100,161	35.9%
대조군	24,068	75,771	99,839	24.1%
#2. 여성				
	사망	생존	총합	사망률
시험군	26,005	73,350	99,355	26.2%
대조군	14,181	86,464	100,645	14.1%

(Case #2) 시험군에 배정될 확률이 같은 층을 합치는 방식 -> 공변량의 균형을 보존

(Case #3) 시험군에 배정될 확률이 다른 층을 합치는 방식 -> 공변량의 균형을 파괴

시뮬레이션 요약

(Case #2) 시험군에 배정될 확률이 같은 층을 합치는 방식 -> 공변량의 균형을 보존

(Case #3) 시험군에 배정될 확률이 다른 층을 합치는 방식 -> 공변량의 균형을 파괴

공변량의 균형을 보존해줄 수 있다면

인과추론을 할 수 있지 않을까?!

⇒ 시험군에 배정될 확률이 같은 층끼리 합쳐진 것처럼 만들어준다면

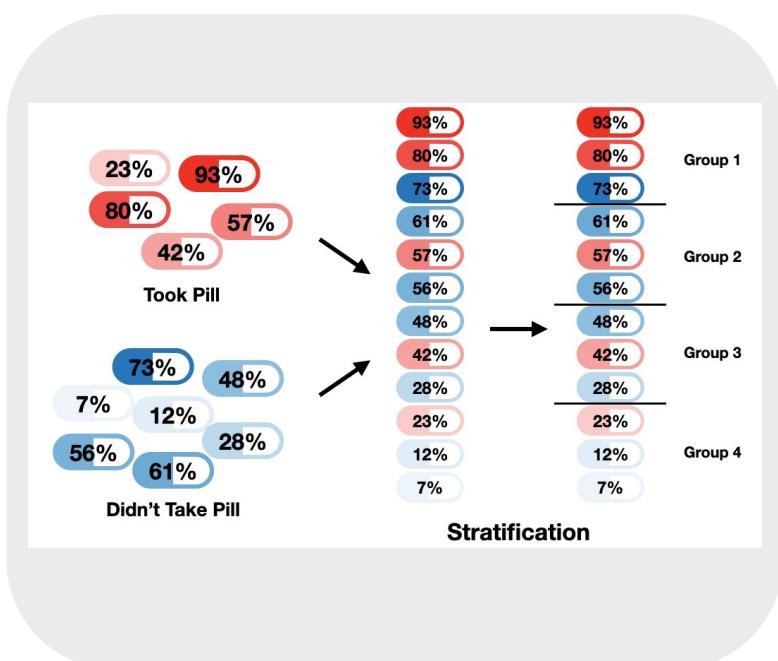
인과추론을 할 수 있지 않을까?

⇒ 성향점수의 아이디어

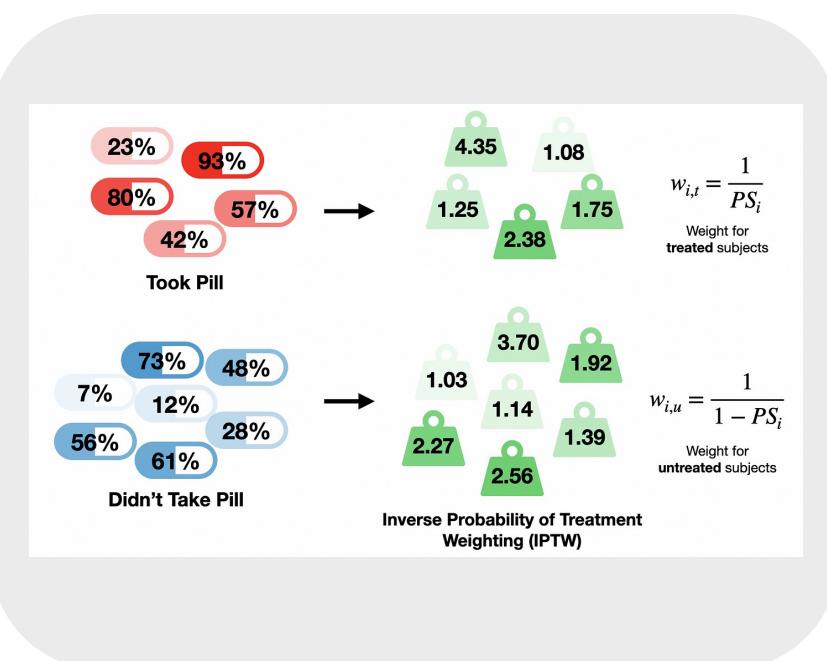
5. 성향점수를 활용한 ATE 추정

ATE(Δ_E)를 추정하는 세 가지 방법

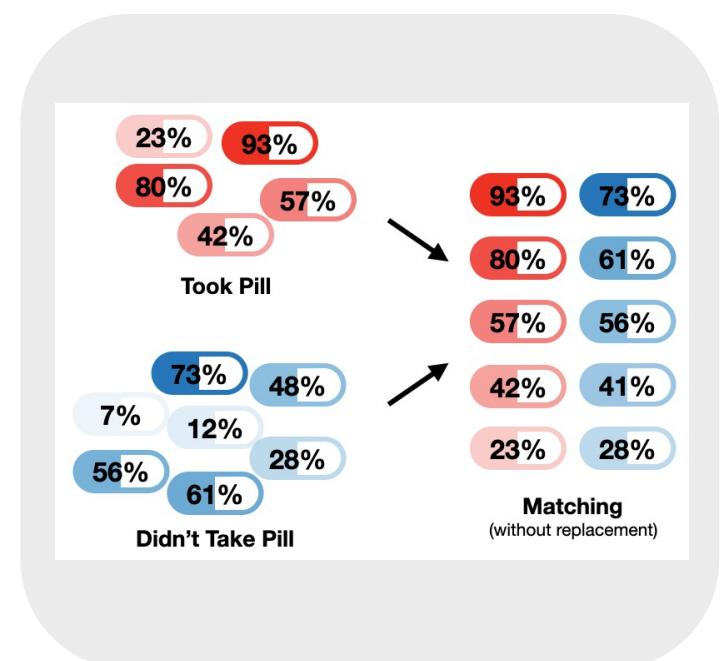
총화(stratification)



가중치(weighting)

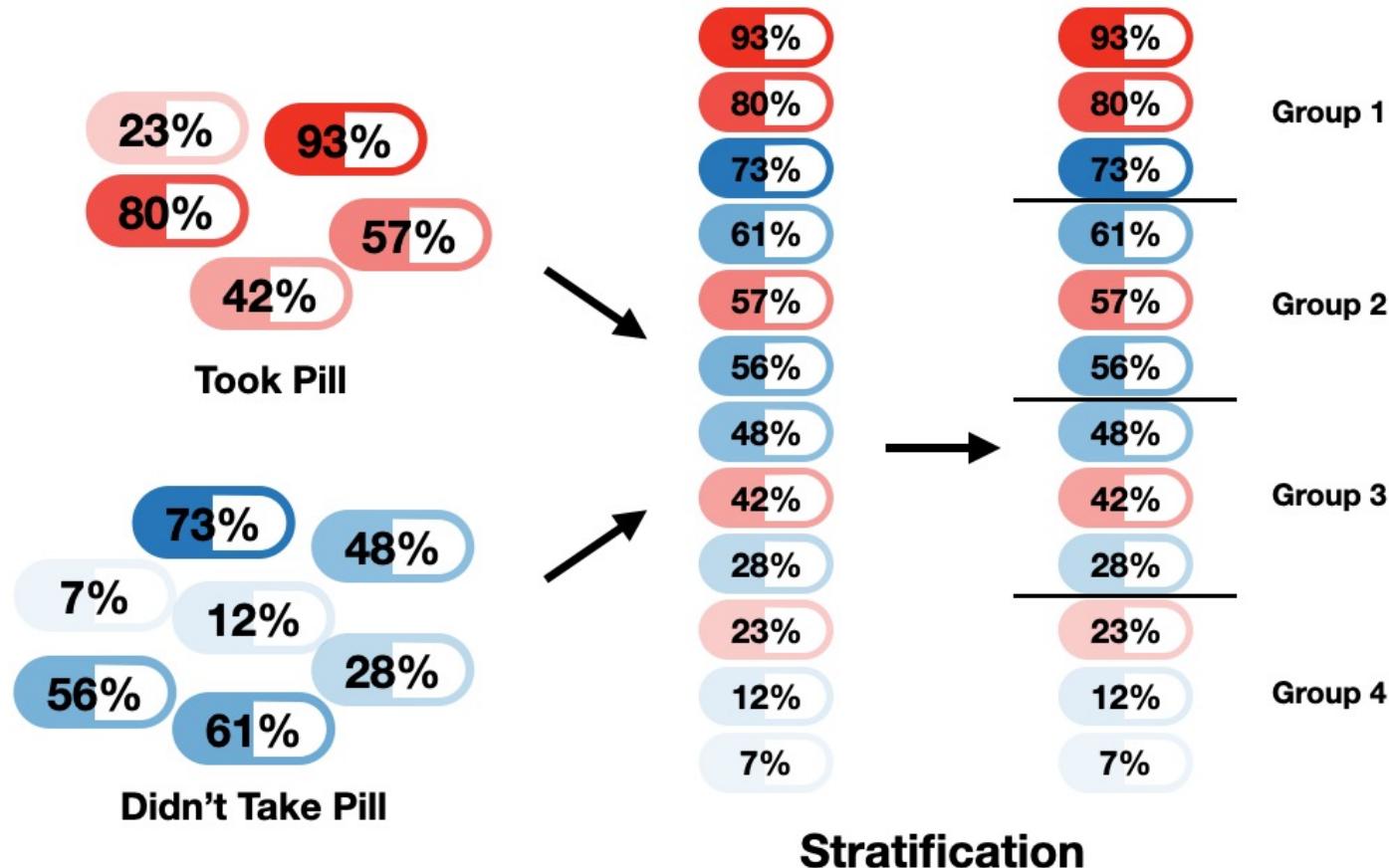


짝맞기(matching)



1) Stratification

층화(stratification)를 활용한 ATE(Δ_E) 추정



1) Stratification

층화(stratification)를 활용한 ATE(Δ_E) 추정

$$\widehat{\Delta}_E = \sum_{j=1}^k \frac{n_j}{n} \left[\frac{1}{n_{1j}} \sum_{i=1}^n Z_i Y_i I(\widehat{e}_i \in \widehat{Q}_j) - \frac{1}{n_j - n_{1j}} \sum_{i=1}^n (1 - Z_i) Y_i I(\widehat{e}_i \in \widehat{Q}_j) \right]$$

(용어 설명)

$\widehat{Q}_j : j$ 번째 층 ($j = 1, \dots, K$)

$n_j = \sum_{i=1}^n I(\widehat{e}_i \in \widehat{Q}_j) : j$ 번째 층에 속하는 관측치의 개수

$n_{1j} = \sum_{i=1}^n Z_i I(\widehat{e}_i \in \widehat{Q}_j) : j$ 번째 층에 속하고, 시험군에 속한 관측치의 개수

1) Stratification

층화(stratification)를 활용한 ATE(Δ_E) 추정

$\widehat{Q}_j : j$ 번째 층 ($j = 1, \dots, K$)

$n_j = \sum_{i=1}^n I(\widehat{e}_i \in \widehat{Q}_j) : j$ 번째 층에 속하는 관측치의 개수

$n_{1j} = \sum_{i=1}^n Z_i I(\widehat{e}_i \in \widehat{Q}_j) : j$ 번째 층에 속하고, 시험군에 속한 관측치의 개수

$$\widehat{\Delta}_E = \sum_{j=1}^k \frac{n_j}{n} \left[\frac{1}{n_{1j}} \sum_{i=1}^n Z_i Y_i I(\widehat{e}_i \in \widehat{Q}_j) - \frac{1}{n_j - n_{1j}} \sum_{i=1}^n (1 - Z_i) Y_i I(\widehat{e}_i \in \widehat{Q}_j) \right]$$

Step1. 성향 점수 추정 (feat. β 추정)

$$\widehat{e}_i = e(\mathbf{X}_i, \widehat{\beta})$$

Step2. \widehat{e}_i 를 기준으로 K 개의 층을 만든다.

Step3. 각 층 안에서 Y_i 에 대한 시험군과 대조군 간의 표본평균 차이를 계산한다.

Step4. Step3에서 계산한 차이의 가중평균으로 $\widehat{\Delta}_E$ 를 추정한다. (가중치: 각 층의 관측치 비율)

1) Stratification

층화(stratification)를 활용한 ATE(Δ_E) 추정

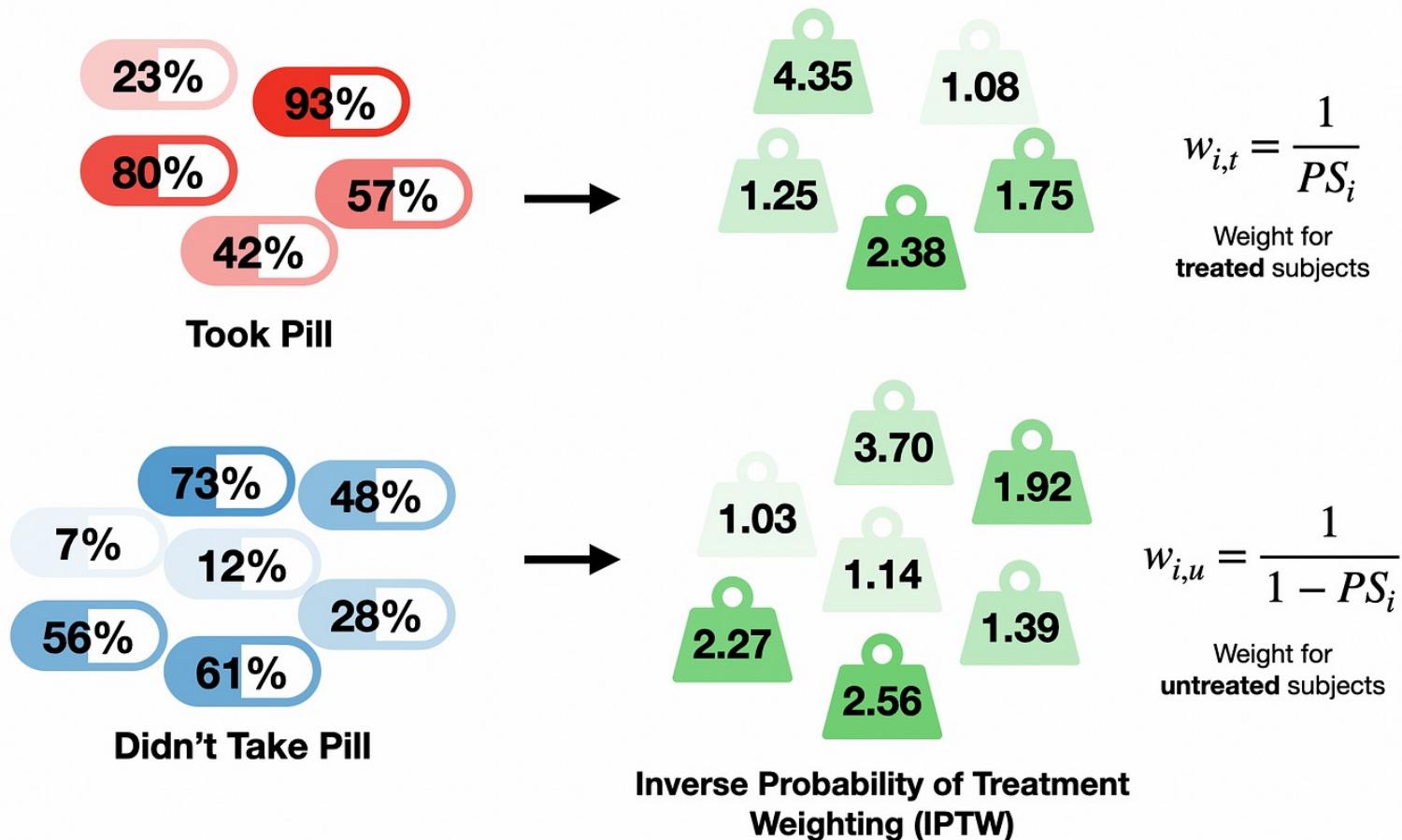
#1. 노인 남성				
	사망	생존	총합	사망률
시험군	32,001	48,121	80,122	39.9%
대조군	7,980	11,898	19,878	40.1%
#2. 노인 여성				
	사망	생존	총합	사망률
시험군	24,012	55,712	79,724	30.1%
대조군	6,130	14,146	20,276	30.2%
#3. 젊은 남성				
	사망	생존	총합	사망률
시험군	3,969	16,070	20,039	19.8%
대조군	16,088	63,873	79,961	20.1%
#4. 젊은 여성				
	사망	생존	총합	사망률
시험군	1,993	17,638	19,631	10.2%
대조군	8,051	72,318	80,369	10.0%

$$\widehat{\Delta}_E = \frac{1}{4} (39.9 - 40.1) + \frac{1}{4} (30.1 - 30.2) + \frac{1}{4} (19.8 - 20.1) + \frac{1}{4} (10.2 - 10.0)$$

$$= -0.1$$

2) Weighting

가중치(weighting)를 활용한 ATE(Δ_E) 추정



2) Weighting

가중치(weighting)를 활용한 ATE(Δ_E) 추정

$$\widehat{\Delta}_E = \frac{1}{n} \sum_{i=1}^n \frac{Z_i Y_i}{\widehat{e}_i} - \frac{1}{n} \sum_{i=1}^n \frac{(1 - Z_i) Y_i}{1 - \widehat{e}_i}$$

특징(1) 각 관측치들을, 그 관측치가 그 군에 속할 확률의 역수를 가중치로 주고 있다.

특징(2) 모든 대상이 실험군에 배정받았을 경우와 같이 유사 모집단을 생성

개인적으로 신기했던 부분(직관과 달랐던 부분)은

각 항을 n_t 또는 n_c 으로 나눠주는 것이 아니라

공통적으로 n 으로 나눠주고 있다는 부분

2) Weighting

가중치(weighting)를 활용한 ATE(Δ_E) 추정

$$\widehat{\Delta}_E = \frac{1}{n} \sum_{i=1}^n \frac{Z_i Y_i}{\hat{e}_i} - \frac{1}{n} \sum_{i=1}^n \frac{(1 - Z_i) Y_i}{1 - \hat{e}_i}$$

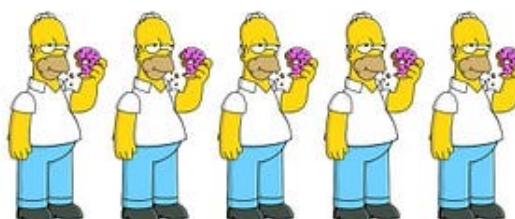
- ii) 모든 대상이 실험군에 배정받았을 경우와 같이 **유사 모집단**을 생성

생성 전

Treated



생성 후



$$\Pr[A = 1 | X = \text{adult, male}] = 1/5$$

$$w_i = 5$$

Untreated



$$\Pr[A = 0 | X = \text{adult, male}] = 4/5$$

$$w_i = 5/4$$



2) Weighting

가중치(weighting)를 활용한 ATE(Δ_E) 추정

$$\widehat{\Delta}_E = \frac{1}{n} \sum_{i=1}^n \frac{Z_i Y_i}{\widehat{e}_i} - \frac{1}{n} \sum_{i=1}^n \frac{(1 - Z_i) Y_i}{1 - \widehat{e}_i}$$

증명

$$\begin{aligned}\Delta_E &= E[Y_i(1)] - E[Y_i(0)] \\ &= E[E[Y_i(1)|\mathbf{X}_i] - E[E[Y_i(0)|\mathbf{X}_i]] \\ &= E\left[\frac{E[Z_i|\mathbf{X}_i]E[Y_i(1)|\mathbf{X}_i]}{e(\mathbf{X}_i)} - \frac{E[1 - Z_i|\mathbf{X}_i]E[Y_i(0)|\mathbf{X}_i]}{1 - e(\mathbf{X}_i)}\right] \\ &\quad (\because e(\mathbf{X}_i) = P[Z_i = 1|\mathbf{X}_i] = E[Z_i|\mathbf{X}_i]) \\ &= E\left[\frac{E[Z_i Y_i(1)|\mathbf{X}_i]}{e(\mathbf{X}_i)} - \frac{E[(1 - Z_i) Y_i(0)|\mathbf{X}_i]}{1 - e(\mathbf{X}_i)}\right] \\ &\quad (\because e(\mathbf{X}_i) = P[Z_i = 1|\mathbf{X}_i] = E[Z_i|\mathbf{X}_i]) \\ &= E\left[\frac{Z_i Y_i}{e(\mathbf{X}_i)} - \frac{(1 - Z_i) Y_i}{1 - e(\mathbf{X}_i)}\right]\end{aligned}$$

...

2) Weighting

가중치(weighting)를 활용한 ATE(Δ_E) 추정

$$\Delta_E = E[Y_i(1)] - E[Y_i(0)] \\ = \dots$$

$$= E \left[\frac{Z_i Y_i}{e(X_i)} - \frac{(1-Z_i)Y_i}{1-e(X_i)} \right]$$



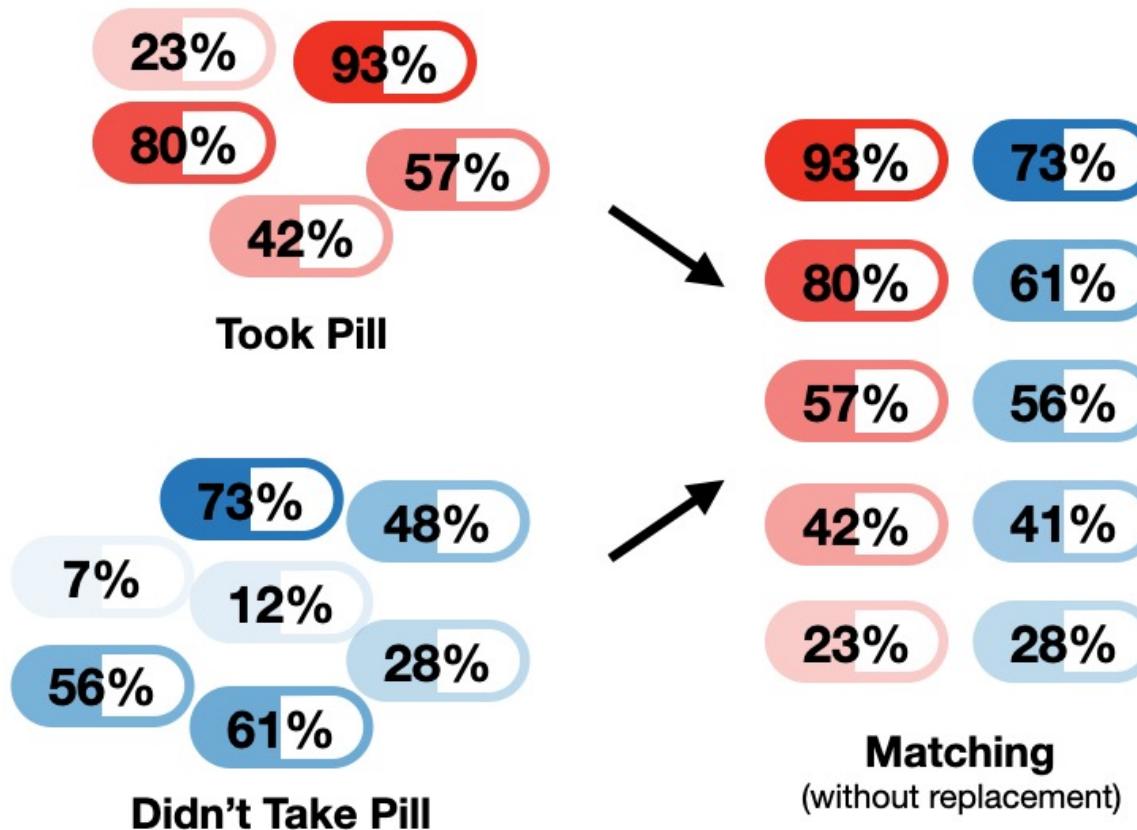
$$\widehat{\Delta}_E = \frac{1}{n} \sum_{i=1}^n \frac{Z_i Y_i}{\widehat{e}_i} - \frac{1}{n} \sum_{i=1}^n \frac{(1-Z_i)Y_i}{1-\widehat{e}_i}$$

OR

$$\widehat{\Delta}_E = \left(\sum_{i=1}^n \frac{Z_i}{\widehat{e}_i} \right)^{-1} \sum_{i=1}^n \frac{Z_i Y_i}{\widehat{e}_i} - \left(\sum_{i=1}^n \frac{1-Z_i}{1-\widehat{e}_i} \right)^{-1} \sum_{i=1}^n \frac{(1-Z_i)Y_i}{1-\widehat{e}_i}$$

3) Matching

짝맞기(Matching)를 활용한 ATE(Δ_E) 추정



3) Matching

짝맺기(Matching)를 활용한 ATE(Δ_E) 추정 (1) Exact Matching

$$\widehat{\Delta}_E = \frac{1}{n_t} \sum_{i:Z_i=1} \left(Y_i - Y_{m_i^c} \right)$$

용어 설명

$I_t = \{1, \dots, n_t\}$: set of indices for the n_t treated units

$I_c = \{n_t + 1, \dots, n_t + n_c\}$: set of indices for the n_c control units

$M_i^c = \{m_i^c\} \subset I_c$: set of control indices containing the matched for treated unit i

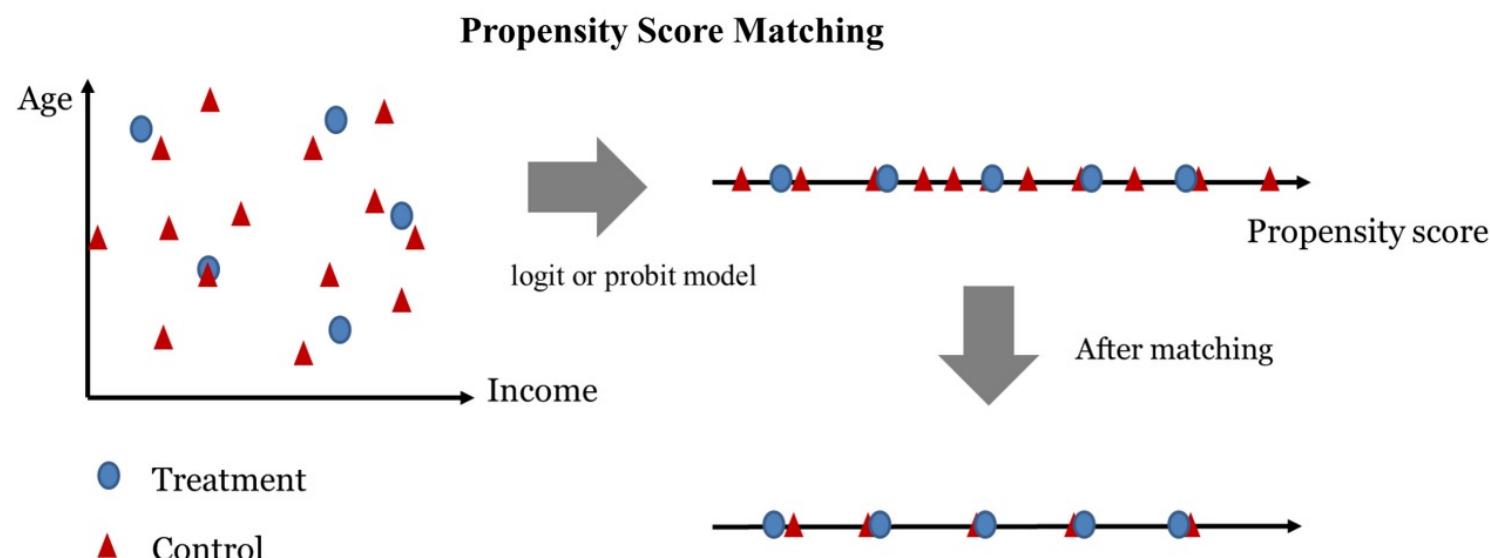
가정사항

1. 실험군 환자들은 각자 다른 대조군 환자와 매칭되어야 한다.
2. 실험군 환자에 매칭되는 대조군은 서로 공변량의 값이 정확히 똑같아야 한다.
3. 실험군 환자 수보다 대조군 환자 수가 더 많아야 한다.

3) Matching

짝맞기(Matching)를 활용한 ATE(Δ_E) 추정 (2) Close Matching

$$\widehat{\Delta}_E = \frac{1}{n_t} \sum_{i:Z_i=1} (Y_i - Y_{m_i^c})$$



3) Matching

짝맺기(Matching)를 활용한 ATE(Δ_E) 추정 (2) Close Matching

$$\widehat{\Delta}_E = \frac{1}{n_t} \sum_{\substack{i: Z_i=1}} \left(Y_i - Y_{m_i^c} \right)$$

$$\text{where } m_i^c = \arg \min_{i' \in I_c} |e(\mathbf{X}_i) - e(\mathbf{X}_{i'})|$$

$$\text{or } m_i^c = \arg \min_{i' \in I_c} \left| \log \left(\frac{e(\mathbf{X}_i)}{1 - e(\mathbf{X}_i)} \right) - \log \left(\frac{e(\mathbf{X}_{i'})}{1 - e(\mathbf{X}_{i'})} \right) \right|$$

$$\text{or } \arg \min_{i' \in I_c} \sum_{i=1}^{n_t} |e(\mathbf{X}_i) - e(\mathbf{X}_{m_i^c})| \text{ subject to } m_i \neq m_j \text{ for } i \neq j (i, j \in I_c)$$

가정사항

1. ~~실험군 환자들은 각자 다른 대조군 환자와 매칭되어야 한다.~~
2. ~~실험군 환자에 매칭되는 대조군은 서로 공변량의 값이 정확히 똑같아야 한다.~~
3. ~~실험군 환자 수보다 대조군 환자 수가 더 많아야 한다.~~

ATE(Δ_E)를 추정하는 세 가지 방법 요약

총화(stratification)
$$\widehat{\Delta}_E = \sum_{j=1}^k \frac{n_j}{n} \left[\frac{1}{n_{1j}} \sum_{i=1}^n Z_i Y_i I(\widehat{e}_i \in \widehat{Q}_j) - \frac{1}{n_j - n_{1j}} \sum_{i=1}^n (1 - Z_i) Y_i I(\widehat{e}_i \in \widehat{Q}_j) \right]$$

가중치(weighting)
$$\widehat{\Delta}_E = \frac{1}{n} \sum_{i=1}^n \frac{Z_i Y_i}{\widehat{e}_i} - \frac{1}{n} \sum_{i=1}^n \frac{(1 - Z_i) Y_i}{1 - \widehat{e}_i}$$

짝맺기(matching)
$$\widehat{\Delta}_E = \frac{1}{n_t} \sum_{i:Z_i=1} \left(Y_i - Y_{m_i^c} \right)$$

6. 성향점수 활용 심화

1) 이중 강건 추정 (Doubly Robust Estimation)

성향점수 모델 $\hat{e}(\mathbf{x})$ 이나 결과 모델 $\hat{m}(X)$ 둘 중 하나만 정확하면 되는 추정치

$$\hat{\mu}_t^{DR}(\hat{m}, \hat{e}) = \frac{1}{N} \sum \hat{m}(X) + \frac{1}{N} \sum \left[\frac{T}{\hat{e}(\mathbf{x})} (Y - \hat{m}(X)) \right] \quad \dots \text{Case #2}$$

$$= \frac{1}{N} \sum \frac{TY}{\hat{e}(\mathbf{x})} \hat{m}(X) - \frac{1}{N} \sum \left[\frac{T - \hat{e}(\mathbf{x})}{\hat{e}(\mathbf{x})} \hat{m}(X) \right] \quad \dots \text{Case #1}$$

Case #1. 성향점수 모델이 정확한 경우

$$\frac{1}{N} \sum \left[\frac{T - \hat{e}(\mathbf{x})}{\hat{e}(\mathbf{x})} \hat{m}(X) \right] \rightarrow 0$$

$(\because \hat{e}(\mathbf{x}) \rightarrow T)$

Case #2. 결과모델이 정확한 경우

$$\frac{1}{N} \sum \left[\frac{T}{\hat{e}(\mathbf{x})} (Y - \hat{m}(X)) \right] \rightarrow 0$$

$(\because \hat{m}(X) \rightarrow Y)$

1) 이중 강건 추정 (Doubly Robust Estimation)

성향점수 모델 $\hat{e}(\mathbf{x})$ 이나 결과 모델 $\hat{m}(\mathbf{X})$ 둘 중 하나만 정확하면 되는 추정치

$$\hat{\mu}_t^{DR}(\hat{m}, \hat{e}) = \frac{1}{N} \sum \hat{m}(\mathbf{X}) + \frac{1}{N} \sum \left[\frac{T}{\hat{e}(\mathbf{x})} (Y - \hat{m}(\mathbf{X})) \right]$$

$$= \frac{1}{N} \sum \frac{TY}{\hat{e}(\mathbf{x})} \hat{m}(\mathbf{X}) - \frac{1}{N} \sum \left[\frac{T - \hat{e}(\mathbf{x})}{\hat{e}(\mathbf{x})} \hat{m}(\mathbf{X}) \right]$$

$$\therefore \widehat{\Delta_E} = \hat{\mu}_1^{DR}(\hat{m}, \hat{e}) - \hat{\mu}_0^{DR}(\hat{m}, \hat{e})$$

2) 연속형 처치

확실한 해결책은 없다. But...

Case #1. 연속형 처치에 대한 처치 반응의 함수 형태를 가정하기

(예시) $y = a + bt$ 또는 $y = a + b\sqrt{t}$

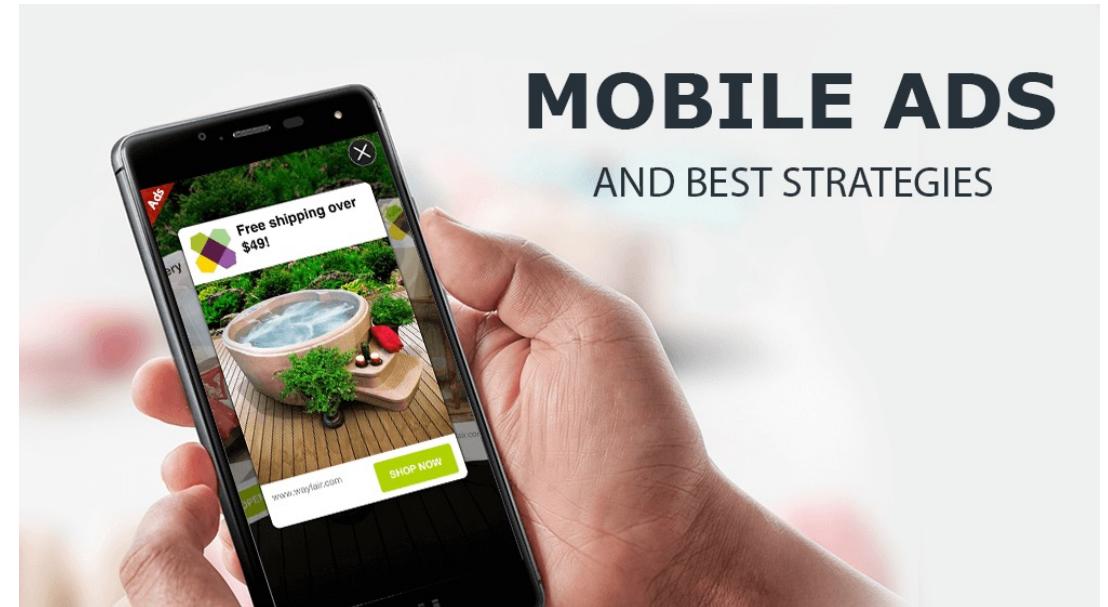
Case #2. 연속형 처치를 이산화(discretize)

Case #3. 일반화 성향점수 (generalized propensity score, GPS)

- 정규분포를 가정하면 OLS 사용
- 아니면, glm 적합

7. 성향점수 활용 사례

1) 군중밀집정도가 모바일 광고에 미치는 영향



1) 군중밀집정도가 모바일 광고에 미치는 영향



1. 데이터 설명

1) 수집 데이터

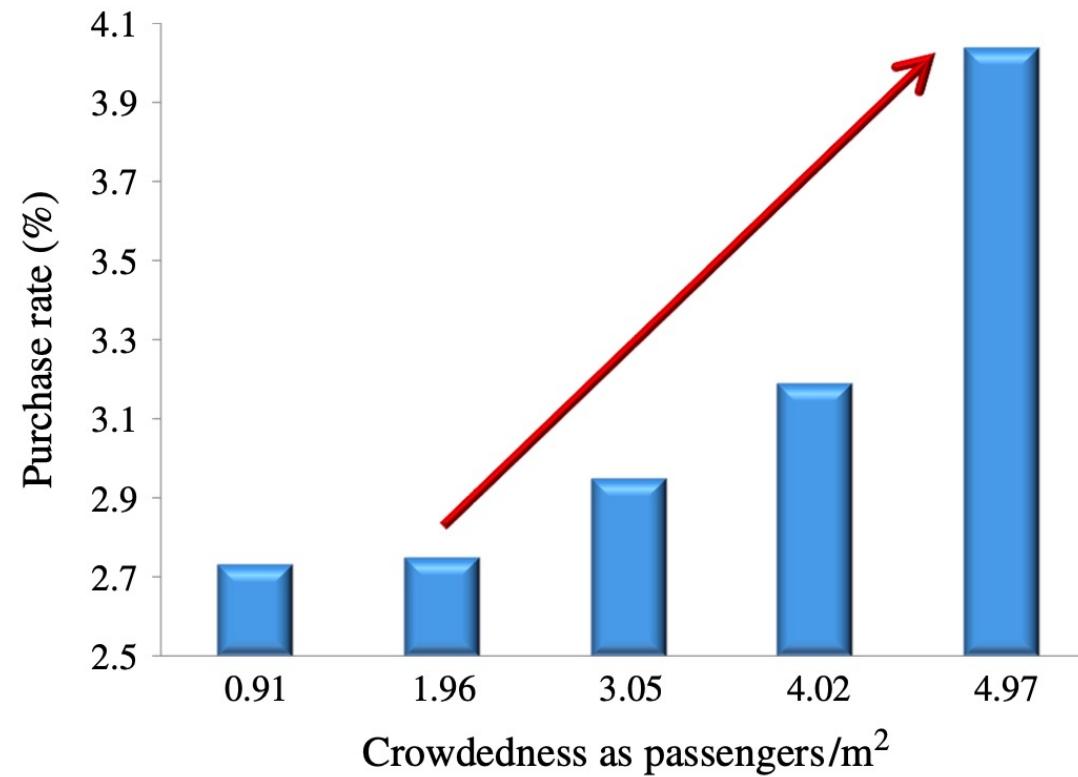
- Peak hour
- Weekend
- ARPU (average revenue per user)
- MOU (minutes of use)
- SMS (number of texts sent and received per user)
- GRPS (data use with the wireless provider)

2) 준실험 2가지 상황

- 교통 통제 (Street Closure)
- 열차 지연 (Train Delay)

1) 군중밀집정도가 모바일 광고에 미치는 영향

Figure 1 (Color online) Crowdedness and Purchase Rates



1) 군중밀집정도가 모바일 광고에 미치는 영향

1. 데이터 설명

- 1) 공통 공변량: Peak hour, Weekend, ARPU, MOU, SMS, GPRS
- 2) 준실험 2가지 상황: 교통 통제(Street Closure), 열차 지연(Train Delay)

2. 성향점수 활용

- 목적: Robustness Check

the field data mimics randomized field experiments and becomes a quasi-experiment design with
pseudo-treatment and pseudo-control groups

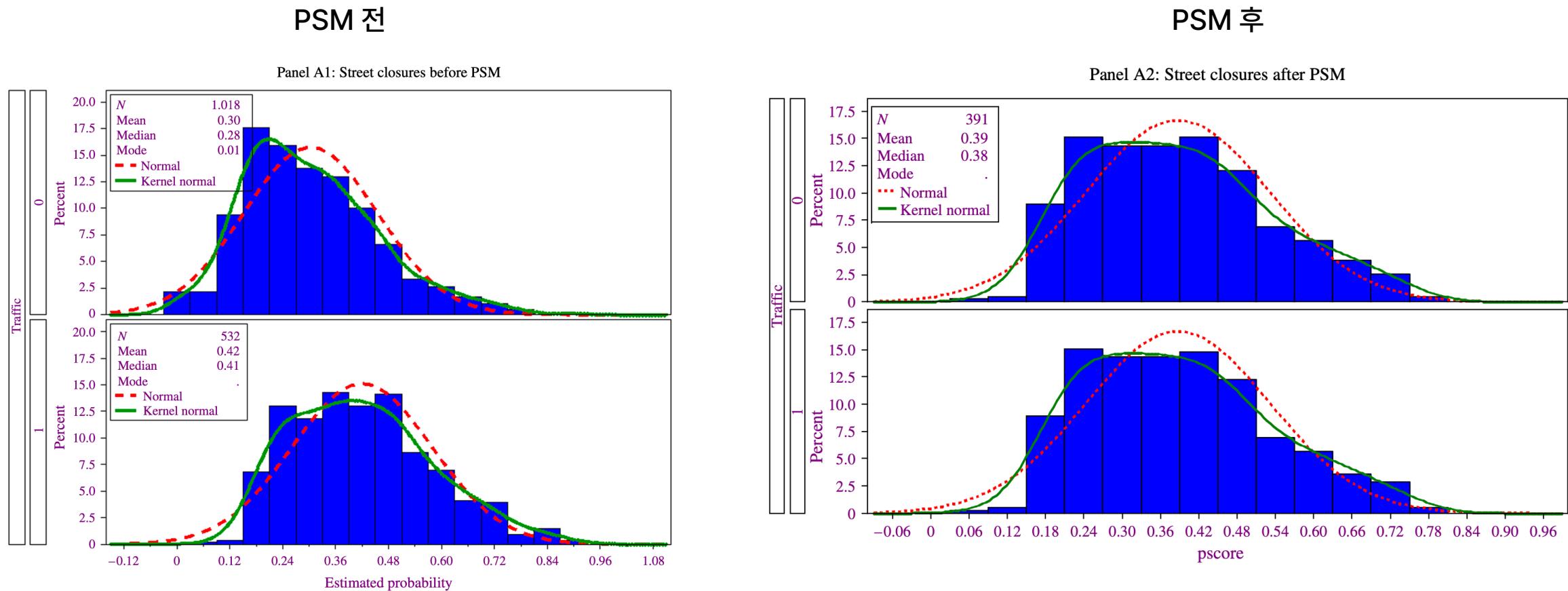
- 방법론: Propensity Score Matching (PSM)

- 성향점수 추정: Logistic Regression

- 타겟: $Y=1$ if a commuter enter trains with sudden variations in crowding, and $Y=0$ if otherwise

1) 군중밀집정도가 모바일 광고에 미치는 영향

1-1) 교통 통제 (Street Closure)



1) 군중밀집정도가 모바일 광고에 미치는 영향

1-1) 교통 통제 (Street Closure)

Table 3 Evidence with Sudden Variations in Crowdedness

	Panel A: Sudden street closures effects			
	model 1	model 2	model 3	model 4
Crowdedness × Sudden street closures				0.492** (0.187)
Crowdedness			0.126** (0.041)	0.114** (0.042)
Sudden street closures		-0.120 (0.117)	-0.142 (0.177)	-1.887 (1.057)
Ln(ARPU)	0.301** (0.118)	0.308** (0.119)	0.308** (0.119)	0.306** (0.119)
Ln(MOU)	-0.043 (0.065)	-0.043 (0.065)	-0.044 (0.065)	-0.044 (0.065)
Ln(SMS)	0.014 (0.069)	0.014 (0.069)	0.015 (0.069)	0.013 (0.069)
Ln(GPRS)	-0.001 (0.024)	-0.001 (0.023)	-0.001 (0.023)	-0.001 (0.023)
Day effects (weekend dummy)	Yes	Yes	Yes	Yes
Time effects (peak hour dummy)	Yes	Yes	Yes	Yes
Observations	11,960	11,960	11,960	11,960
	model 1	model 2	model 3	model 4

$$PurchaseLikelihood_i = \frac{\exp(U_i)}{1 + \exp(U_i)}$$

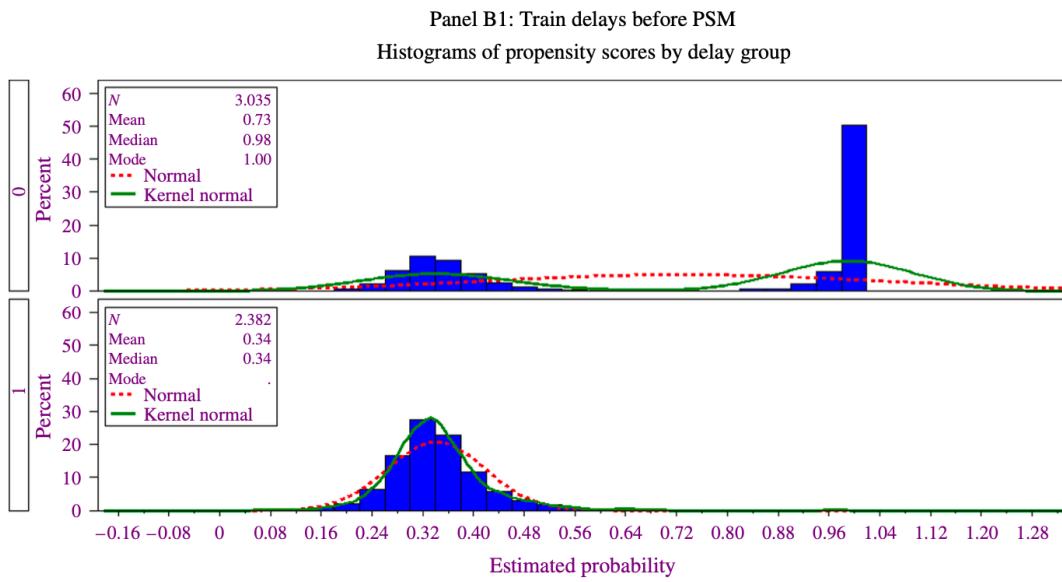
$$U_i = \alpha + \beta \times crowdedness_i + \phi \times street\ closure_i \\ + \varphi \times crowdedness_i \times street\ closure_i + \gamma \times weekday_i \\ + \delta \times peak\ hour_i + \xi \times location_i + \tau \times X_i + \varepsilon_{i2}, \quad (2)$$

교통 통제 자체로는 구매 확률 증가로 이어지지 않는다.
하지만 군중밀집정도를 고려한다면,
구매 확률 증가로 이어지는 효과를 확인할 수 있다.

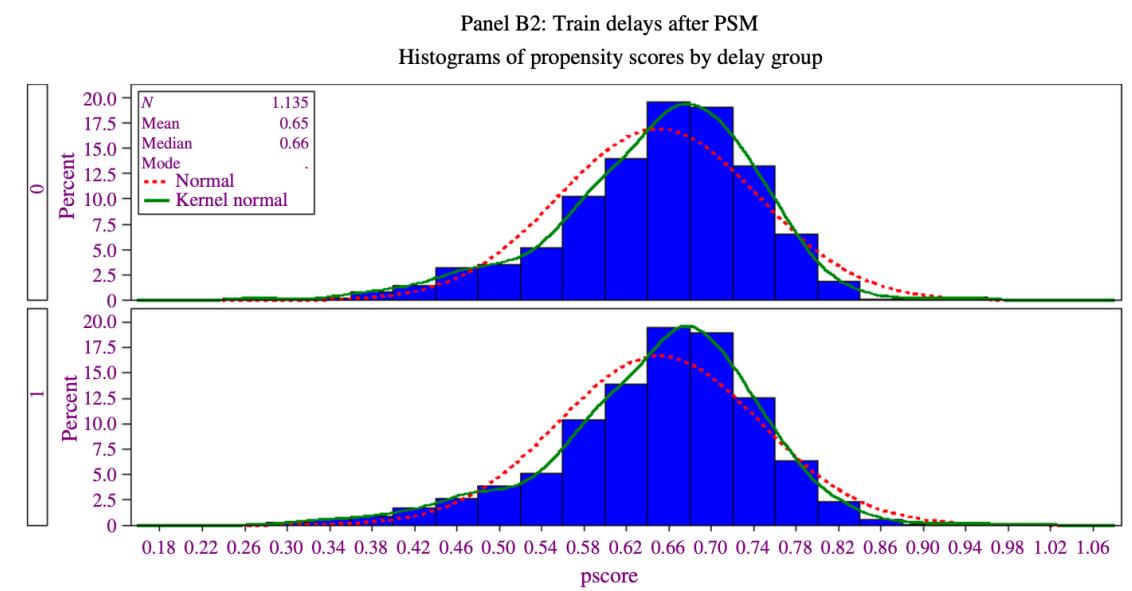
1) 군중밀집정도가 모바일 광고에 미치는 영향

1-2) 열차 지연 (Train Delay)

PSM 전



PSM 후



1) 군중밀집정도가 모바일 광고에 미치는 영향

1-2) 열차 지연 (Train Delay)

Panel B: Sudden train delays effects			
Crowdedness × Sudden train delays		0.669*** (0.148)	
Crowdedness		0.306*** (0.108)	0.282*** (0.119)
Sudden train delays	0.142 (0.161)	0.130 (0.184)	0.162 (0.174)
Ln(ARPU)	-0.0003 (0.0016)	-0.0003 (0.0016)	-0.0003 (0.0016)
Ln(MOU)	0.0001 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)
Ln(SMS)	0.0001 (0.0003)	0.0001 (0.0003)	0.0001 (0.0003)
Ln(GPRS)	-0.0000 (0.0000)	-0.0000 (0.0000)	-0.0000 (0.0000)
Day effects (weekend dummy)	Yes	Yes	Yes
Time effects (peak hour dummy)	Yes	Yes	Yes
Observations	13,702	13,702	13,702
	model 1	model 2	model 3
			model 4

$$PurchaseLikelihood_i = \frac{\exp(U_i)}{1 + \exp(U_i)}$$

$$\begin{aligned} U_i = & \alpha + \beta \times crowdedness_i + \phi \times train\ delay_i \\ & + \varphi \times crowdedness_i \times train\ delay_i + \gamma \times weekday_i \\ & + \delta \times peak\ hour_i + \tau \times X_i + \varepsilon_{i3}, \end{aligned} \quad (3)$$

열차 지연 자체로는 구매 확률 증가로 이어지지 않는다.
 하지만 군중밀집정도를 고려한다면,
 구매 확률 증가로 이어지는 효과를 확인할 수 있다.

1) 군중밀집정도가 모바일 광고에 미치는 영향

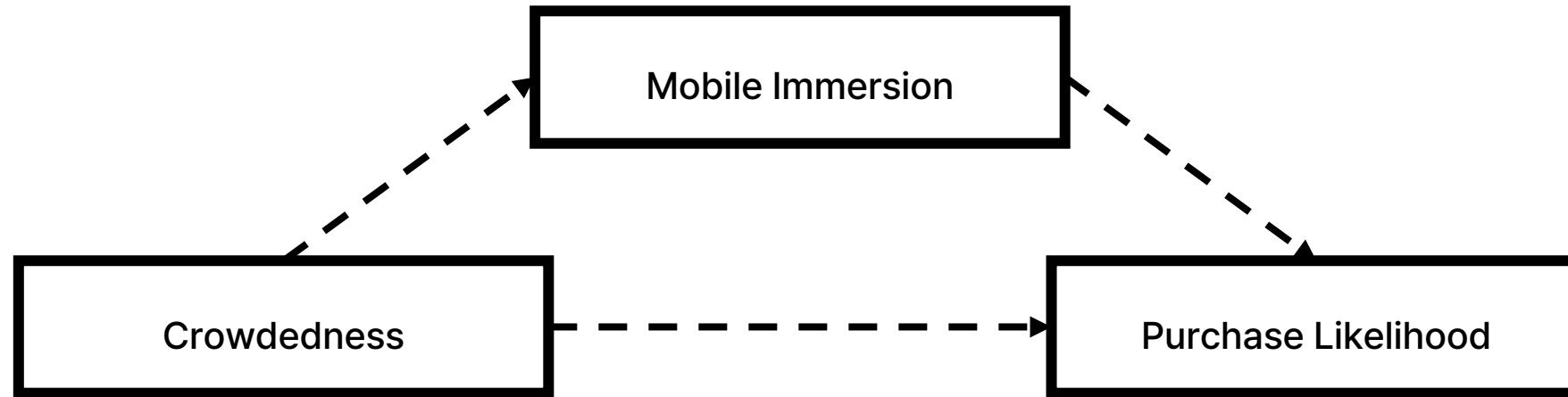
개인적으로 흥미로웠던 점



1. Crowdedness를 측정하지 않으면 Street Closure 또는 Train Delay의 상황의 영향력이 유의미하지 않았음
2. Crowdedness 측정하고 이것을 그대로 변수화하지 않고
두 가지 상황(교통 통제, 열차 지연)에 따라 별도로 분석을 진행했다는 점

1) 군중밀집정도가 모바일 광고에 미치는 영향

추가 설문조사) Mobile Immersion



1) 군중밀집정도가 모바일 광고에 미치는 영향

참고) 각 상황별 공변량 분포 비교

Table 4 Summary Statistics of Mobile Users' Wireless Usage Behavior

	Regular passengers (N = 10,690)		Street closures passengers (N = 1,270)		Train delays passengers (N = 3,012)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
ARPU	59.424	41.144	59.551	41.488	59.118	40.147
MOU	567.024	731.090	561.063	736.778	550.804	719.550
SMS	311.913	222.189	312.346	224.979	315.522	228.972
GPRS	27,856.6	8,867.1	28,445.8	8,944.3	27,336.8	8,506.7

그룹간 공변량 차이 없음을 확인

2) 블린사이토 (Blincyto)

블린사이토: 치료제 (성인 및 소아 필라델피아 염색체 양성인 재발 또는 불응성 전구 B세포 급성림프모구성백혈병)



Package and vials are not shown to scale

2) 블린사이토 (Blincyto)

블린사이토: 치료제 (성인 및 소아 필라델피아 염색체 양성인 재발 또는 불응성 전구 B세포 급성림프모구성백혈병)

1. 상황 요약

- 1상 임상시험 : 치료 효과 확인
- 2상 임상시험 : 치료 효과 확인 (**But, 단일군으로 진행**)
⇒ FDA와 논의하여, real-world data로 과거대조군을 만들어서 2상 임상시험결과와 비교하기로 협의

2. 데이터 설명

- European National Study Group 및 미국과 유럽의 여러 곳에서 얻어진 1,112명으로 구성
- 주요 공변량 : 나이, 성별, 초기진단으로부터 salvage therapy까지의 기간, 지역, prior HSCT 여부, 사전에 투여받은 salvage therapy 횟수, primary refractory in the first salvage 여부, refractory to last salvage therapy 여부

3. 성향점수 활용 방안

- **IPTW** (Inverse Probability of Treatment Weighting)
 - 성향점수 추론 방법론 : Logistic Regression

4. 분석 대상

- (1) 완전관해 비율 : Logistic Regression
- (2) 전체 생존 기간 : Cox Proportional Hazards Model

*완전관해: 치료후 골수검사에서 아세포가 5%이하이며 골밀도가 20%이상이면서, 피검사 수치가 정상으로 회복하여 4주 이상 유지되는 것

2) 블린사이토 (Blincyto)

블린사이토: 치료제 (성인 및 소아 필라델피아 염색체 양성인 재발 또는 불응성 전구 B세포 급성림프모구성백혈병)

Table 1			
Demographics and patient characteristics—historical comparator and clinical trial data			
	Historical data set	Clinical trial data	
	Patients with CR data	Patients with OS data	Blinatumomab trial (MT103-211)
	N=694	N=1112	N=189 ^a
Sex, n (%)			
Male	421 (60.7)	644 (57.9)	119 (63%)
Female	273 (39.3)	468 (42.1)	70 (37%)
Mean (s.d.) age, years	38.8 (14.8)	37.4 (14.2)	41.1 (17.3)
Age group (years)			
18–34	317 (45.6)	527 (47.4)	90 (48)
35–54	256 (36.9)	428 (38.5)	46 (24)
≥55	80 (11.5)	115 (10.3)	53 (28)
≥65	41 (5.9)	42 (3.8)	25 (13)
Disease status, n (%)			
Primary refractory			16 (8.5)
In 1st salvage	61 (8.8)	56 (5.0)	4 (2.1)
In 2nd or greater salvage	50 (7.2)	61 (5.5)	12 (6.3)
Relapsed, with prior alloHSCT			64 (33.9)
In 1st salvage	44 (6.3)	130 (11.7)	9 (4.8)
In 2nd or greater salvage	45 (6.5)	57 (5.1)	55 (29.1)
Relapsed, without prior alloHSCT			109 (57.7)
In 1st salvage	245 (35.3)	543 (48.8)	25 (13.2)
In 2nd or greater salvage	249 (35.9)	265 (23.8)	84 (44.4)
Year of initial diagnosis, n (%)			
1990 to 1999	246 (35.4)	364 (32.7)	0 (0)
2000 or later	448 (64.6)	748 (67.3)	189 (100)
Region, n (%)			
Europe	393 (56.6)	811 (72.9)	95 (50)
USA	301 (43.4)	301 (27.1)	94 (50)

[Open in a separate window](#)

^aNumbers by salvage treatment in disease status categories not totalling due to missing data for salvage history (need to verify).

1. 상황 요약

- 1상 임상시험 : 치료 효과 확인
- 2상 임상시험 : 치료 효과 확인 (**But, 단일군으로 진행**)
⇒ FDA와 논의하여, real-world data로 과거대조군을 만들어서 2상 임상시험결과와 비교하기로 협의

2. 데이터 설명

- European National Study Group 및 미국과 유럽의 여러 곳에서 얻어진 1,112명으로 구성
- 주요 공변량 : (1) 나이, (2) 성별, (3) 초기진단부터 salvage therapy까지의 기간, (4) 지역, (5) prior HSCT 여부, (6) 사전에 투여받은 salvage therapy 횟수, (7) primary refractory in the first salvage 여부, (8) refractory to last salvage therapy 여부

2) 블린사이토 (Blincyto)

블린사이토: 치료제 (성인 및 소아 필라델피아 염색체 양성인 재발 또는 불응성 전구 B세포 급성림프모구성백혈병)

3. 성향점수 활용 방안

- IPTW (Inverse Probability of Treatment Weighting)

Supplemental Table 5. Covariate balance before and after propensity score adjustments, based on survival data analysis set

Factor	Before Adjustments				After Adjustments ¹			
	Historical Dataset (N=1131)	Blinatumomab Trial (MT103-211) (N=189)	Standardized Difference	p-value	Historical Dataset (N=1131)	Blinatumomab Trial (MT103-211) (N=189)	Standardized Difference	p-Value ²
Age Mean (SD)	37.4 (14.2)	41.1 (17.3)	0.233	0.0014	38.1 (14.5)	36.9 (15.7)	-0.078	0.4694
Female n (%)	477 (42%)	70 (37%)	-0.105	0.1850	475 (42%)	68 (36%)	-0.122	0.2913
Duration since initial diagnosis (months) Mean (SD)	12.2 (12.3)	28.1 (36.5)	0.585	<0.0001	13.8 (15.1)	15.6 (18.0)	0.106	0.1740
Region-- Europe n (%)	822 (73%)	95 (50%)	-0.473	<0.0001	780 (69%)	89 (47%)	-0.452	0.0001
Prior alloHSCT n (%)	209 (18%)	64 (34%)	0.355	<0.0001	238 (21%)	38 (20%)	-0.019	0.8475
Number of prior salvage therapies ³ Mean (SD)	1.52 (0.82)	2.36 (0.99)	0.924	<0.0001	1.64 (0.89)	1.69 (0.87)	0.061	0.5334
Primary refractory and in first salvage n (%)	62 (5%)	4 (2%)	-0.177	0.0587	57 (5%)	19 (10%)	0.194	0.1882
Refractory to preceding salvage n (%)	259 (23%)	98 (52%)	0.627	<0.0001	305 (27%)	51 (27%)	-0.002	0.9833

¹ p-value is from a logistic regression model for the binary variables and a linear regression for the continuous variables

² Includes the last line of treatment, which is blinatumomab for blinatumomab subjects

(1) 성향점수 보정 전 : 8개 공변량 중 6개에서 유의한 차이 관측

Ex) 블린사이토 2상 임상시험 환자들이 과거대조군에 비해

이전에 다른 치료제로 치료를 받은 경험이 더 많았다.

- Average line of salvage therapy: 2.36 vs. 1.52, p<0.0001

- Refractory to the last line of salvage therapy: 52% vs. 23%, p<0.0001

(2) 성향점수 보정 후: 8개 공변량 중 1개에서만 유의한 차이 관측

2) 블린사이토 (Blincyto)

결과 1. 완전 관해 (Complete Remission)

Table 4

Propensity score analysis of historical data set and blinatumomab clinical trial data: CR and overall survival^a

<i>End point</i>	<i>Statistic</i>	<i>Historical data set</i>	<i>Blinatumomab trial (MT103-211)</i>
CR/CRh rate	Predicted rate (95% CI)	26.7% (23.4–30.3%)	49.3% (33.4–65.3%)
Overall survival	6-month survival rate (95% CI)	33.4% (31.0–36.1%)	57.6% (54.9–60.4%)
Overall survival	12-month survival rate (95% CI)	17.2% (15.3–19.4%)	39.0% (36.0–42.2%)

95% 신뢰구간 비교

시험군: 49.3% (33.4, 65.3)
 대조군: 26.7% (23.4, 30.3)

결론: 완전관해에 도달할 odds는 시험군이 대조군에 비해 2배 이상 높았다.

[통계량: OR=2.68, 95% CI=(1.67, 4.31)]

2) 블린사이토 (Blincyto)

결과 2. 전체생존기간 (Overall Survival Time)

Table 4

Propensity score analysis of historical data set and blinatumomab clinical trial data: CR and overall survival^a

<i>End point</i>	<i>Statistic</i>	<i>Historical data set</i>	<i>Blinatumomab trial (MT103-211)</i>
CR/CRh rate	Predicted rate (95% CI)	26.7% (23.4–30.3%)	49.3% (33.4–65.3%)
Overall survival	6-month survival rate (95% CI)	33.4% (31.0–36.1%)	57.6% (54.9–60.4%)
Overall survival	12-month survival rate (95% CI)	17.2% (15.3–19.4%)	39.0% (36.0–42.2%)

95% 신뢰구간 비교

(1) 6개월

시험군: 57.6% (54.9, 60.4)
 대조군: 33.4% (31.0, 36.1)

(1) 12개월

시험군: 39.0% (36.0, 42.2)
 대조군: 17.2% (15.3, 19.4)

결론: 시험군이 대조군보다 생존률이 유의미하게 높았다.

[통계량: HR(Hazard Ratio)=0.54, 95% CI=(0.40, 0.73)]

*HR이 1보다 작으면, 상대적으로 안전하다는 의미이다.

감사합니다

Q&A

1. 책

- [1] 실무로 통하는 인과추론 with 파이썬
- [2] Real world health care data analysis: causal methods and implementation using SAS

2. 수업자료

- [1] 빅데이터활용사례연구 (2022-2)

3. 블로그

- [1] <https://yeong-jin-data-blog.tistory.com/entry/Matching-Weighting>

4. 논문

- [1] Mobile Ad Effectiveness: Hyper-Contextual Targeting with Crowdedness
(<https://pubsonline.informs.org/doi/abs/10.1287/mksc.2015.0905>)
- [2] Blinatumomab vs historical standard therapy of adult relapsed/refractory acute lymphoblastic leukemia
(<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5056974/>)