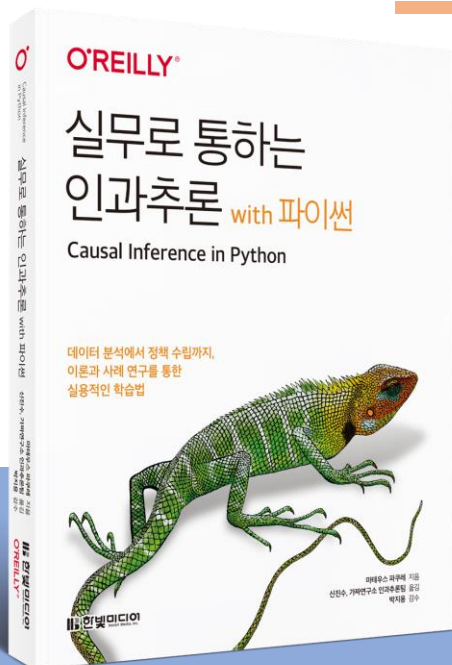


# 『실무로 통하는 인과추론 with 파이썬』 특강

## 유저 세그먼트와 이질적 처치 효과



1. HTE 그게 뭔데, 그거 어떻게 하는건데.

2. 이질적 처치 효과로 개인화하기 (사례)



3. 유저 세그먼트 관점에서의 HTE에 대한 생각

1. H T E 그제 뉼데, 그거 어떻게 하는건데.

# 1. HTE 그게 뭔데, 그거 어떻게 하는건데.

인과추론의 근본적인 도전 과제 : 반사실

어떤 히어로를 없애야 우주를 효과적으로 정복할 수 있을까?



타노스의 숨겨진 아들 타노스 Jr.



핑거스냅으로 무작위 추출



처치했을 때 가장 효과가 좋은 히어로는?

# 1. HTE 그게 뭔데, 그거 어떻게 하는건데.

인과추론의 근본적인 도전 과제 : 반사실

만약 스파이더맨이 사망한다면? 악당들의 범죄율은 얼마나 상승할까?



스파이더맨

생존



사망..



Real World:  $\text{do}(T=1)$



Counterfactual World:  $\text{do}(T=0)$



범죄율?



# 1. HTE 그게 뭔데, 그거 어떻게 하는건데.

## 이질적 처치 효과의 아이디어

만약 스파이더맨이 사망한다면? 악당들의 범죄율은 얼마나 상승할까?

히어로	세계 구분	사망(처치) 여부	범죄율
	지구-1610B	0	9%
	지구-50101B	1	68%
	지구-14512B	0	11%
	지구-90124	1	73%
	지구-616B	0	8%

지구-1610B 스파이더맨의 반사실은 알 수 없으니  
다른 스파이더맨들의 반사실을 모아본다면?

같은 히어로라는 공변량  $X$ 를 고려해서  
조건부 평균 처치효과(CATE)를 구해보자

처치했을 때,  
가장 효과가 뛰어난 히어로를 우선적으로 없애볼까?

# 1. HTE 그게 뭔데, 그거 어떻게 하는건데.

CATE를 추정하려면 어떻게 해야할까?

$$y_i = \beta_0 + \beta_1 t_i + \beta_2 X_i + e_i$$

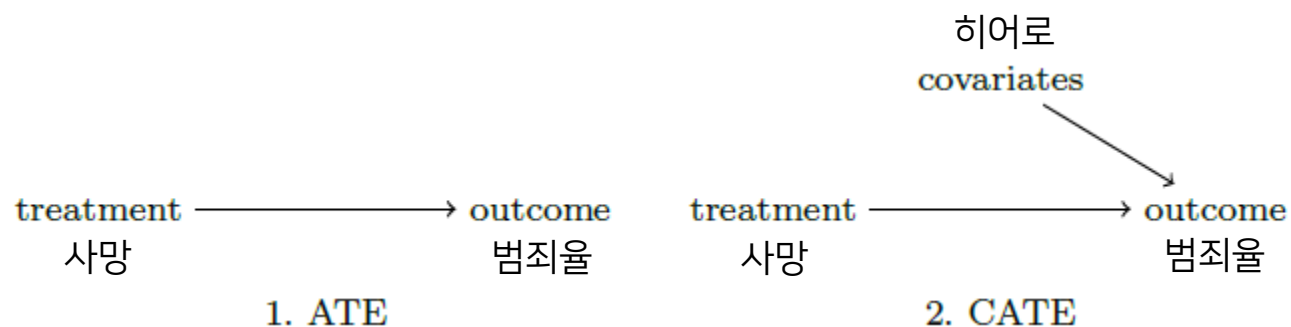


$$\text{ATE} \quad \frac{\delta y_i}{\delta t_i} = \beta_1$$

처치(t)에 대해 미분하면 기울기를 예측할 수 있지만,  
전체 대상에 대한 기울기이기 때문에 ATE를 구하는 것

# 1. HTE 그게 뭔데, 그거 어떻게 하는건데.

CATE를 추정하려면 어떻게 해야할까?





# 1. HTE 그게 뭔데, 그거 어떻게 하는건데.

CATE를 추정하려면 어떻게 해야할까?

처치와 공변량 간의 상호작용 항을 추가!

$$y_i = \beta_0 + \beta_1 t_i + \beta_2 X_i + e_i$$



$$y_i = \beta_0 + \beta_1 t_i + \beta_2 X_i + \boxed{\beta_3 t_i X_i} + e_i$$

상호작용항



$$\frac{\delta y}{\delta t} = \frac{y(t + \epsilon) - y(t)}{(t + \epsilon) - t}$$



$$\frac{\delta y}{\delta t} \approx \hat{y}(t + 1) - y(t)$$

도함수를 정의해서 미분을 풀면,  
좀 더 쉬운 예측값(근사값)을 얻을 수 있음!

CATE

$$\frac{\widehat{\delta y_i}}{\delta t_i} = \widehat{\beta}_1 + \widehat{\beta}_3 X_i$$

상호작용항을 추가해서  
공변량에 따라 처치 효과가 어떻게 변하는 지 학습

# 1. HTE 그게 뭔데, 그거 어떻게 하는건데.

## CATE 예측 모델의 성능 평가하기

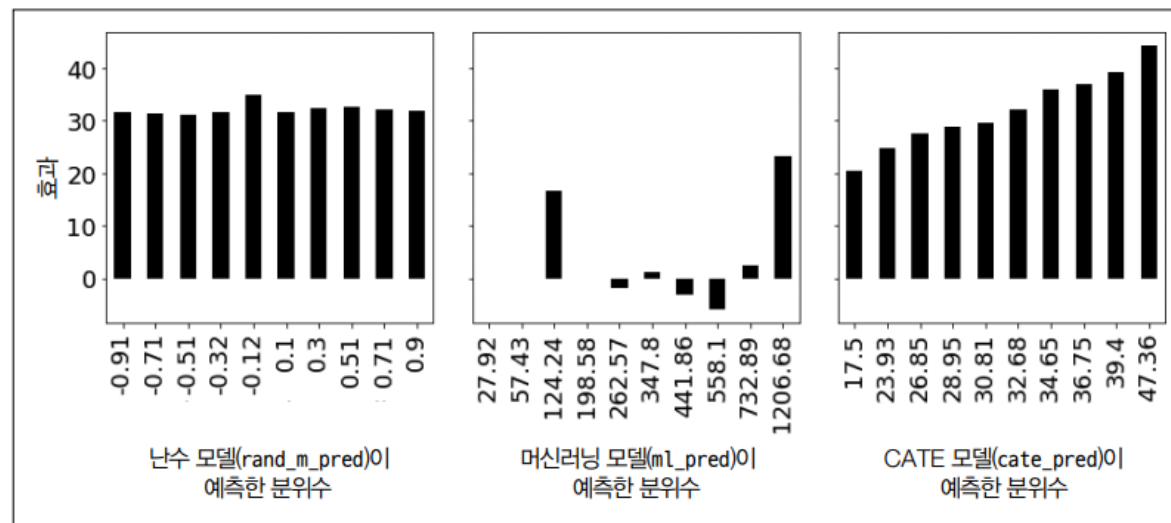
개인마다 처치 / 미처치를 모두 관측할 수 없기 때문에 그룹 수준의 지표로 개별 대상의 처치효과를 판단!

### 모델 평가 순서

1. 데이터를 Train Set / Test Set으로 나눈다.
2. CATE를 예측하는 모델을 생성한다.
3. 모델의 예측값의 분위수를 기준으로 데이터를 나누고 분위수 그룹별로 효과를 추정한다.
4. 분위수 순으로 정렬된 그룹별로 처치효과(기울기)를 추정

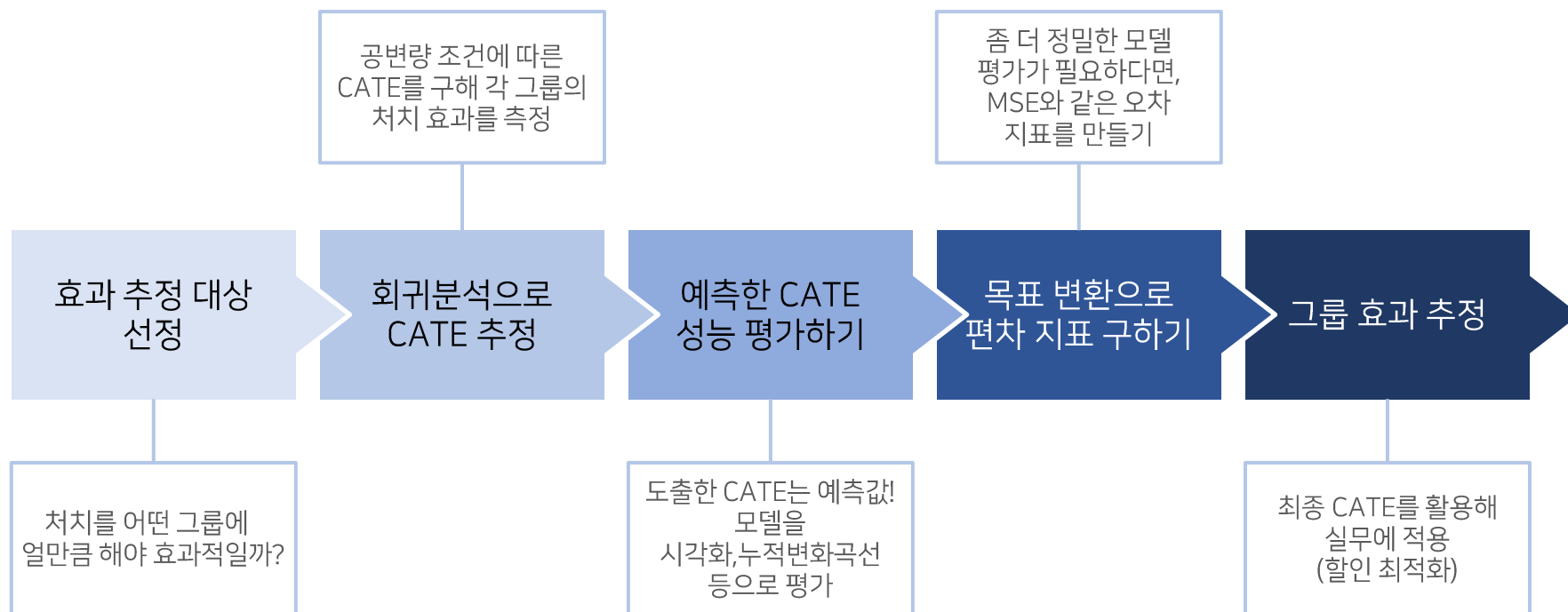
분위 순서대로 점차 상승하는 형태일 수록 좋은 모델이라고 평가할 수 있음!

분위 수 그룹에 대한 대표값이 추정 효과와 일치 할 수록 CATE를 잘 예측한다고 볼 수 있음!



# 1. HTE 그게 뭔데, 그거 어떻게 하는건데.

## 이질적 처치 효과 전체 프로세스 개요



## 2. 이질적 처치 효과로 개인화하기 (사례)

## 2. 이질적 처치 효과로 개인화하기 (사례)

### 청년내일채움공제 정책의 이질적 처치 효과

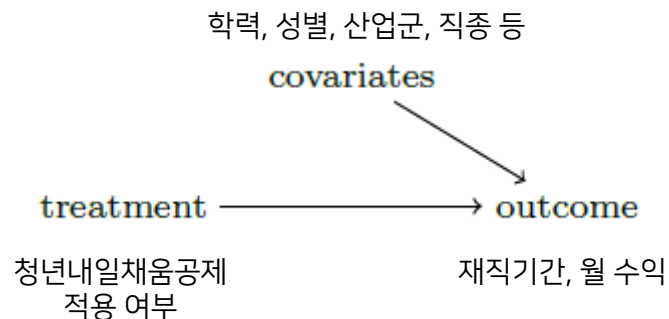


#### 청년내일채움공제?

중소기업에 취업한 청년이 2년간 400만원을 적립하면 정부가 400만원, 기업이 400만원을 공동 적립하여 만기 시 1,200만원의 목돈 마련을 할 수 있는 제도.

#### 논문 요약

- 청년내일채움공제 정책이 중소기업에서 청년 근로자의 근속 기간을 증가시키는 긍정적 효과가 있음
- 그러나 소득 감소와 같은 부작용도 동반되므로, 정책 설계 시 이를 고려한 세밀한 타겟팅 전략이 필요
- 이질적 효과 분석은 정책 수혜자를 식별하고 예산을 효율적으로 배분하는 데 유용한 방법임을 제안



## 2. 이질적 처치 효과로 개인화하기 (사례)

### 청년내일채움공제 정책의 이질적 처치 효과

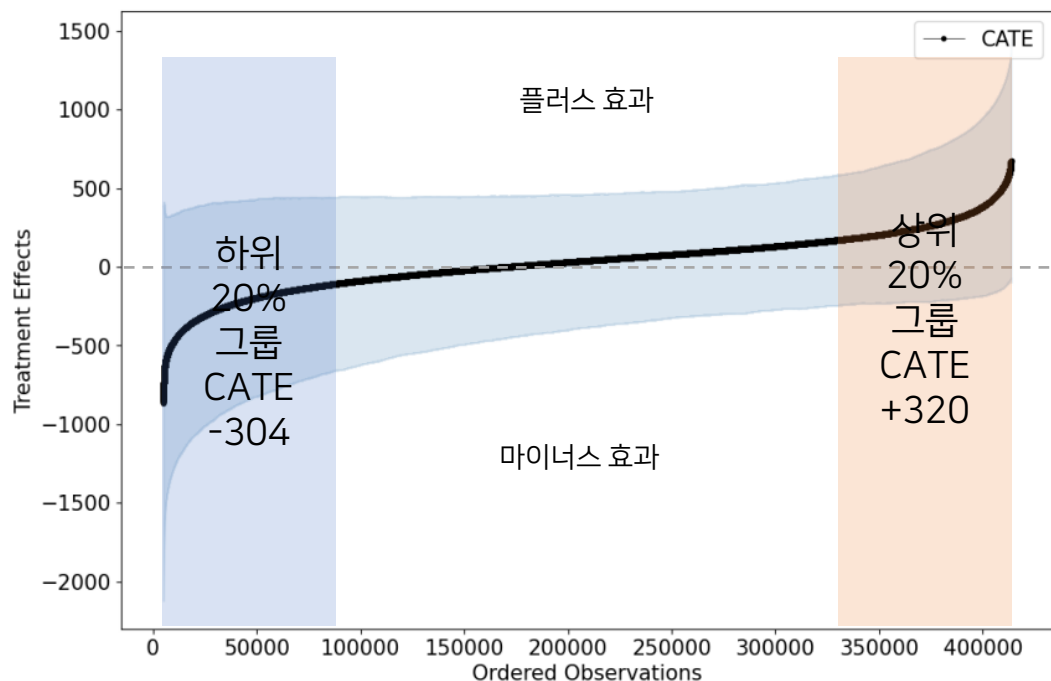


Figure 4.2: Observations sorted by treatment effect level

공변량(covariates)을 기반으로 CATE의 결과를 도출하고 정책 시행에 초점을 맞춰야 할 하위 그룹을 제안

CATE 예측 모델에서 생성된 추정된 처치 효과의 분포를 보여줍니다. 관찰치는 처치 효과 값에 따라 정렬 후 처치 효과를 누적 평균으로 시각화

Table 4.3: CATE results

	20% Least	ATE	20% Most
CATE	-304.02	25.25	320.82

상위 20% 그룹은 +320 만큼의 처치 효과  
하위 20% 그룹은 -304만큼의 처치 효과  
ATE와 두 그룹을 비교했을 때도 차이가 큰 것을 볼 수 있음

가장 영향을 많이 받은 관찰치와 가장 적게 받은 관찰치 간의 차이를 예상함으로써, 이러한 그룹의 평균 특성을 비교할 수 있습니다. 이 비교는 관찰된 이질성에 기여하는 잠재적 요인을 식별하는 데 도움을 줄 수 있습니다.

## 2. 이질적 처치 효과로 개인화하기 (사례)

### 청년내일채움공제 정책의 이질적 처치 효과

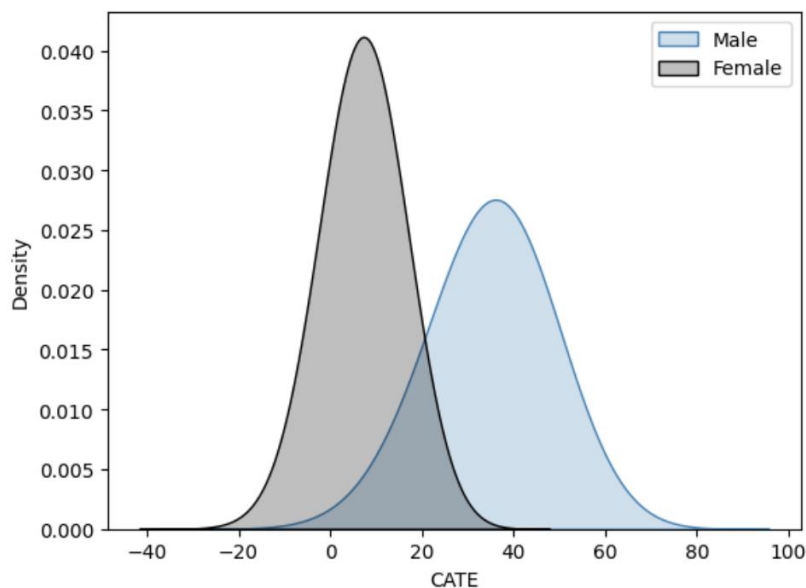


Figure 4.3: CATE distribution by gender

남성이 여성보다 더 큰 정책 효과를 경험

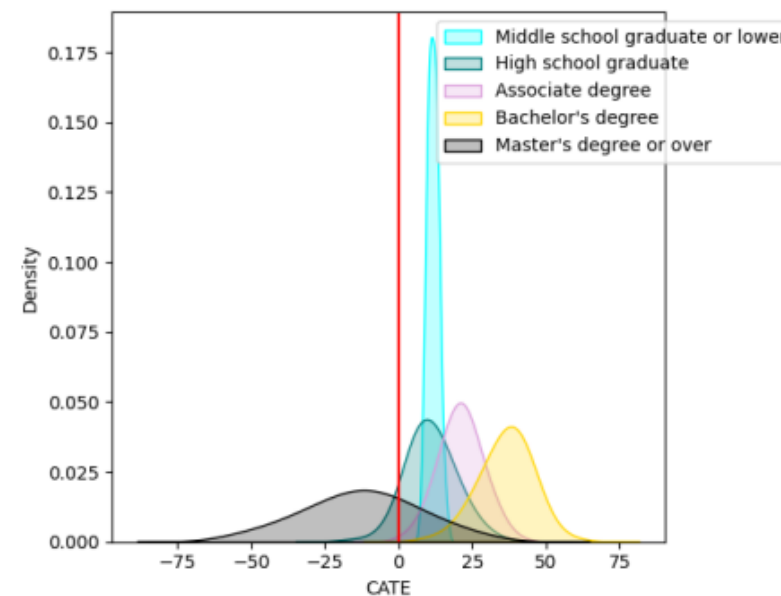
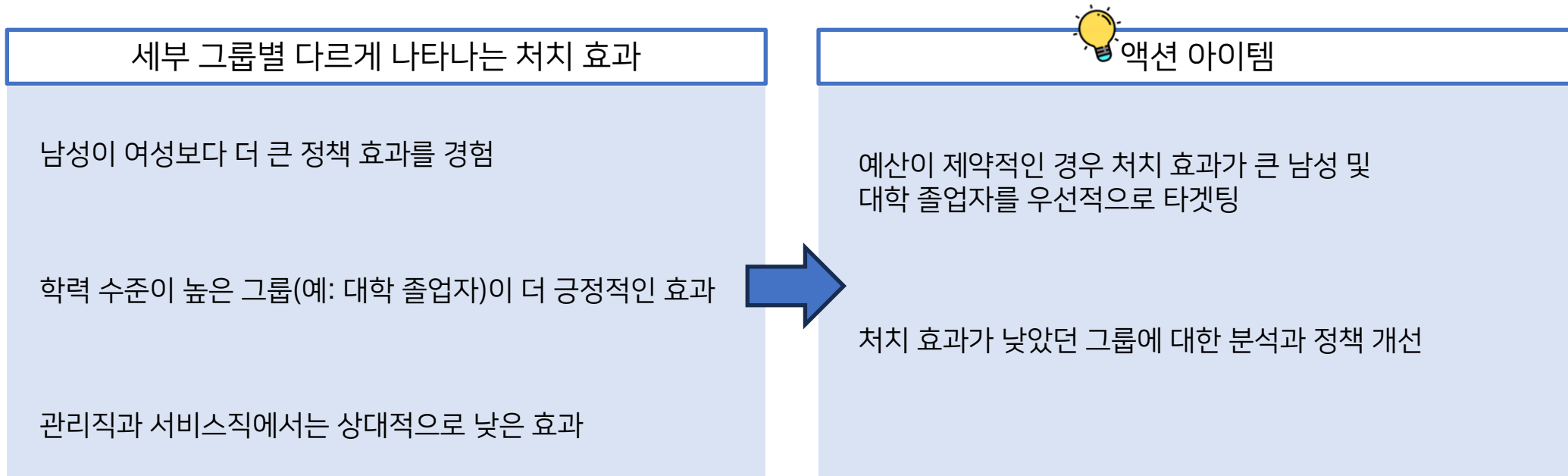


Figure 4.4: CATE distribution by education level

학력 수준이 높은 그룹(예: 대학 졸업자)이 더 긍정적인 효과

## 2. 이질적 처치 효과로 개인화하기 (사례)

### 청년내일채움공제 정책의 이질적 처치 효과





### 3. 유저 세그먼트 관점에서의 H T E에 대한 생각

### 3. 유저 세그먼트 관점에서의 H T E에 대한 생각

유저 세그먼트는 왜 중요할까요?

특정 그룹에 대한 맞춤 전략을 수립 할 수 있다.

모든 유저가 동일한 특성을 가질 가능성은 굉장히 낮음

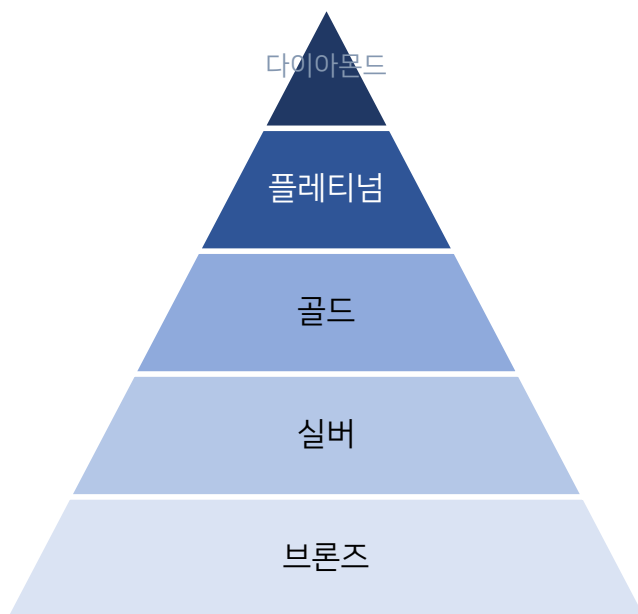
같은 나이와 성별을 가진 게임 유저라고 하더라도 플레이 스타일, 성향 등은 완전히 다를 수 있음

이런 세분화 그룹들을 구분하여 실무에서의 목표를 달성하기 위해서는 어떤 유저 그룹에 집중해야할 지 포커싱을 할 수 있음

### 3. 유저 세그먼트 관점에서의 H T E에 대한 생각

유저 세그먼트는 어떻게 활용할 수 있을까요?

도메인 기반의 세그먼트



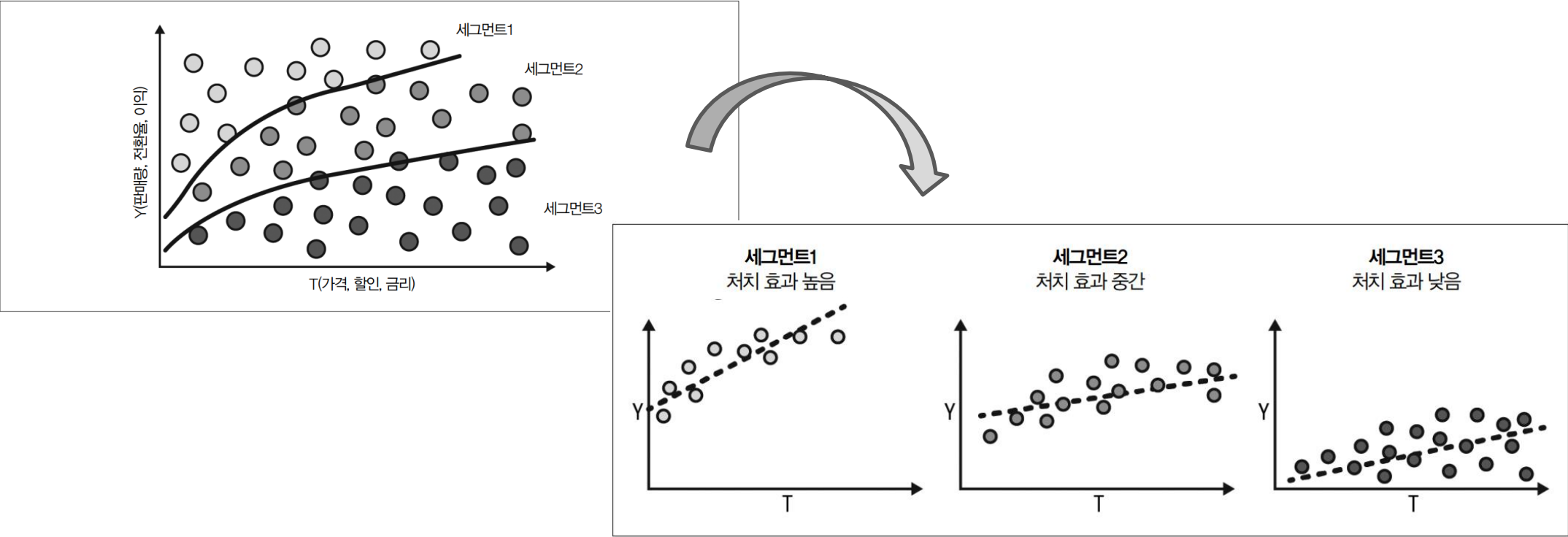
모델 기반의 세그먼트



인구통계학, 행동 기반, 성향 기반, 감정 기반..

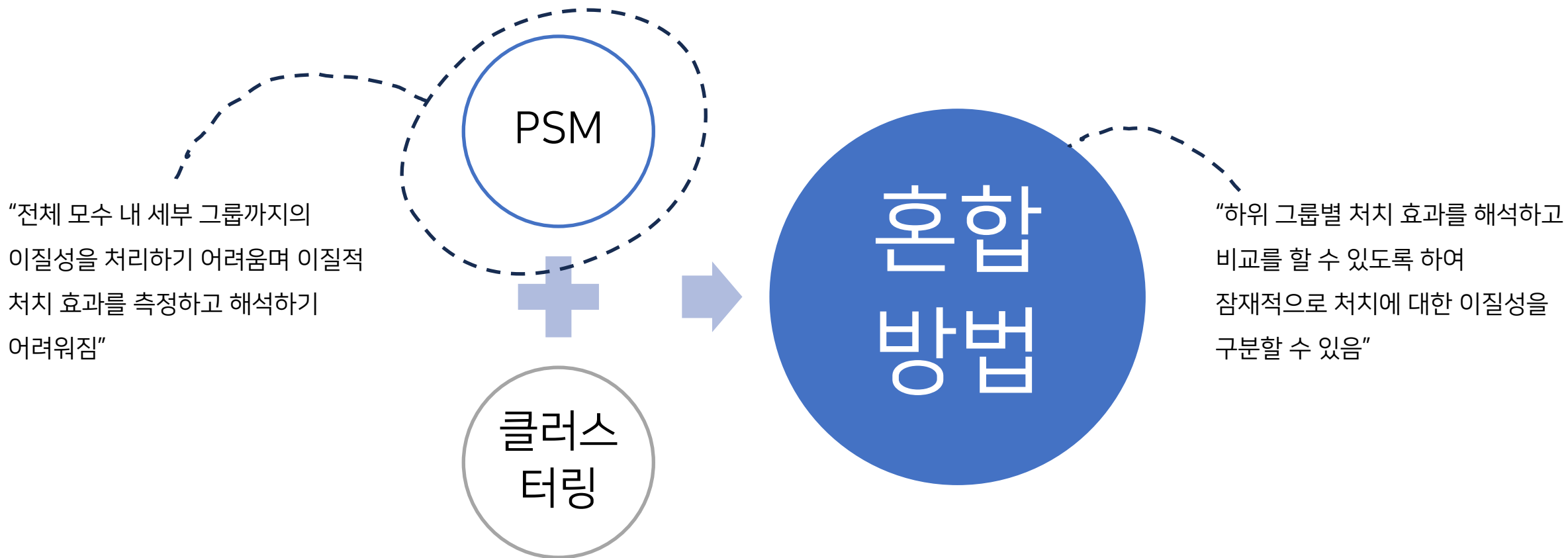
# 3. 유저 세그먼트 관점에서의 H T E에 대한 생각

유저 세그먼트는 어떻게 활용할 수 있을까요?



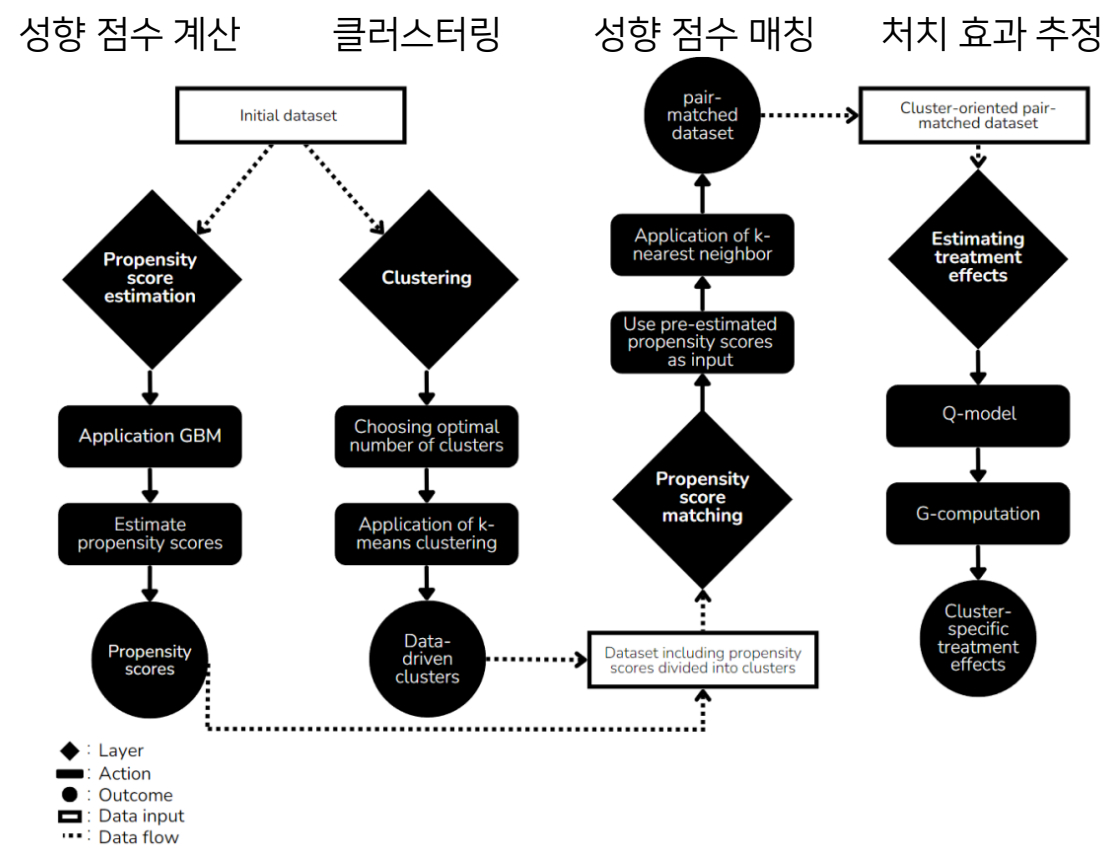
### 3. 유저 세그먼트 관점에서의 H T E에 대한 생각

유저 세그먼트를 활용한 이질적 처치 효과 분석 사례



# 3. 유저 세그먼트 관점에서의 H T E에 대한 생각

유저 세그먼트를 활용한 이질적 처치 효과 분석 사례

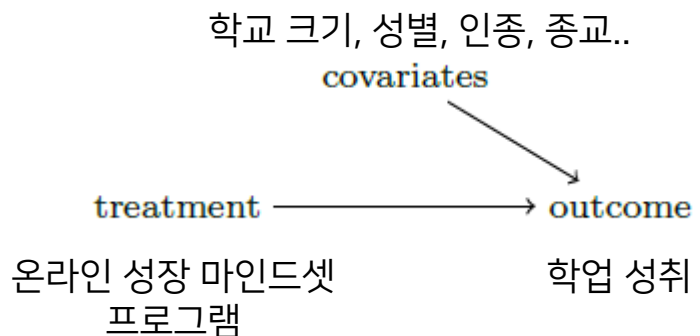


Lips, Mike, Nuno Almeida Camacho, and Michel van de Velden. "The Application of Cluster-Based Heterogeneity Analysis on an Observational Study." (2023).

### 3. 유저 세그먼트 관점에서의 H T E에 대한 생각

#### 유저 세그먼트를 활용한 이질적 처치 효과 분석 사례

어떤 조건의 9학년 학생들에게 성장 마인드셋(처치) 프로그램을 했을 때 효과적일까?

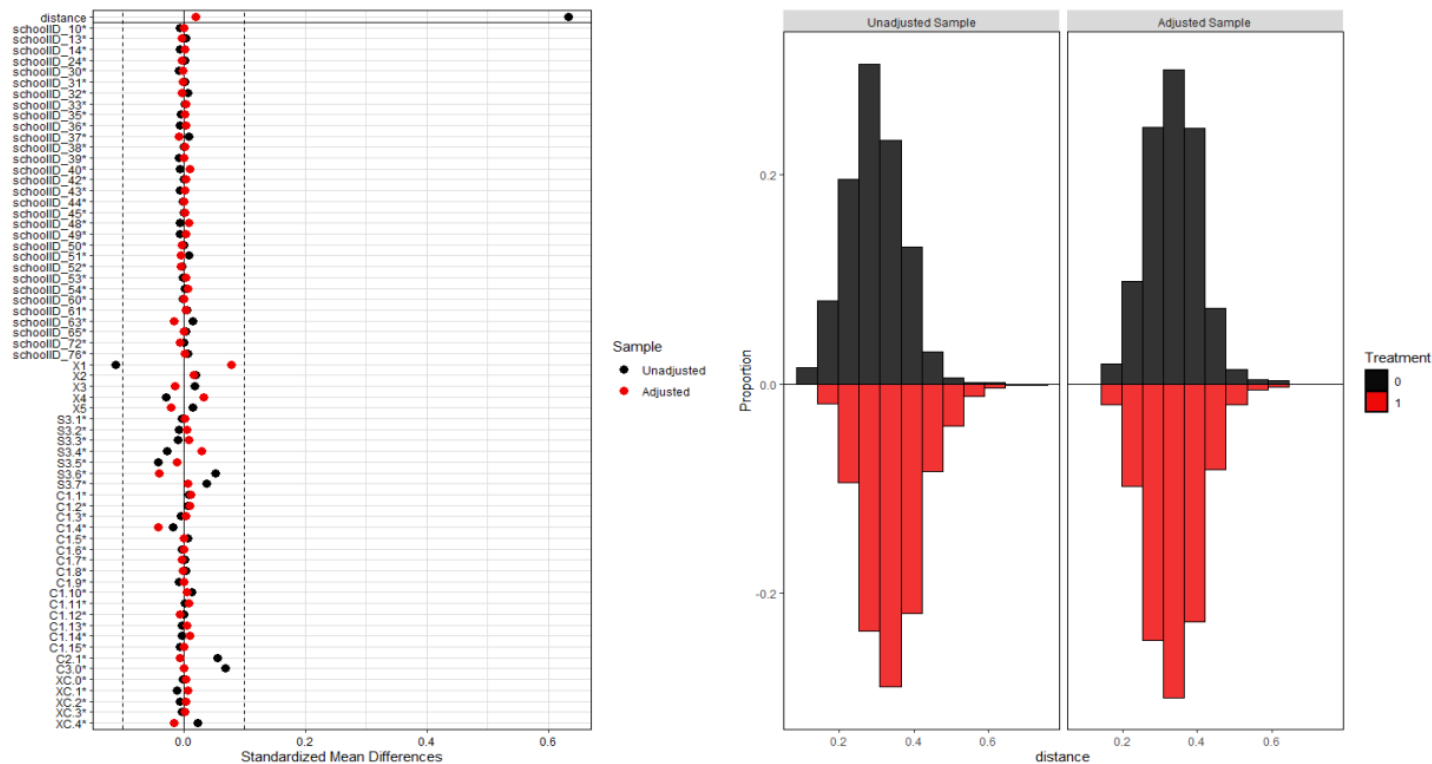


무작위로 선정된 76개 미국 고등학교 내에서  
9학년 학생들을 처치 그룹 또는 통제 그룹에 무작위로 배정.

### 3. 유저 세그먼트 관점에서의 H T E에 대한 생각

#### 유저 세그먼트를 활용한 이질적 처치 효과 분석 사례

동일 클러스터 내에서 성향점수매칭을 활용하여 매칭의 정밀도를 올림





### 3. 유저 세그먼트 관점에서의 H T E에 대한 생각

#### 유저 세그먼트를 활용한 이질적 처치 효과 분석 사례

클러스터 6번 그룹이 가장 처치 효과가 좋게 나타남

6번 그룹은 자체적인 마인드셋 교육 경험이 없고 집중력이 낮은 도시 학생들

Table 7: Average treatment effect (ATT) estimates for each cluster

Cluster	Estimate	Std. Error	2.5%	97.5%
1	0.234***	0.032	0.171	0.297
2	0.234***	0.035	0.166	0.302
3	0.277***	0.026	0.225	0.328
4	0.221***	0.028	0.167	0.275
5	0.266***	0.030	0.207	0.324
6	0.317***	0.036	0.247	0.387

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$

Table 4: Cluster-specific content containing the mean of covariates

Cluster	1	2	3	4	5	6
School-level fixed mindset (X1)	-0.883	1.250	-0.421	0.107	0.434	-1.160
School achievement level (X2)	1.170	-1.220	0.395	-0.072	-0.383	0.749
School ethnic minority composition (X3)	-0.227	1.320	-0.915	-0.767	0.649	0.032
School poverty concentration (X4)	-0.989	1.140	-0.839	0.189	-0.387	1.070
School size (X5)	1.060	-0.673	0.140	-1.050	-0.101	1.600
Rural (XC.0)	0.199	0.100	0	0.083	0.114	0
Countryside (XC.1)	0.307	0.095	0.437	0.301	0.086	0
Town (XC.2)	0.494	0	0.438	0.063	0.092	0
Suburban (XC.3)	0	0.287	0	0.433	0	0
City (XC.4)	0	0.518	0.124	0.120	0.708	1

## 4. 정리

1. 동일한 처치라도 특정 조건이나 그룹에 따라 다르다 (CATE)
2. 사람은 모두 다르기에 비슷한 특징을 가진 유저 세그먼트를 정의하고 활용하는게 중요하다
3. 유저 세그먼트 기반의 이질적 처치 효과 분석으로 세부 그룹을 타겟팅하고 비즈니스 전략 세울 수 있다!

감사합니다

KIM, Dohui. Estimating heterogeneous impacts using causal forest. Diss. KDI School, 2023.

Lips, Mike, Nuno Almeida Camacho, and Michel van de Velden. "The Application of Cluster-Based Heterogeneity Analysis on an Observational Study." (2023).

실무로 통하는 인과추론 with 파이썬 -마테우스 파쿠레 저자(글) · 신진수 , 가짜연구소 인과추론팀 번역 · 박지용 감수

게임 플레이어는 좋은 아이템을 획득하면 게임을 더 열심히 하게 될까? [https://danbi-ncsoft.github.io/works/2021/05/13/class\\_get\\_causal\\_analysis.html](https://danbi-ncsoft.github.io/works/2021/05/13/class_get_causal_analysis.html)

길드는 유저를 응집시킬 수 있을까? 인과추론으로 바라보는 길드 영향력 분석 <https://blessedby-clt.tistory.com/60> [데이터 탐험 노트:티스토리]

Causal Machine Learning for Econometrics: Causal Forests <https://towardsdatascience.com/causal-machine-learning-for-econometrics-causal-forests-5ab3aec825a7>

우리 서비스의 유저 세그멘테이션 하기 <https://brunch.co.kr/@lee880728/10>