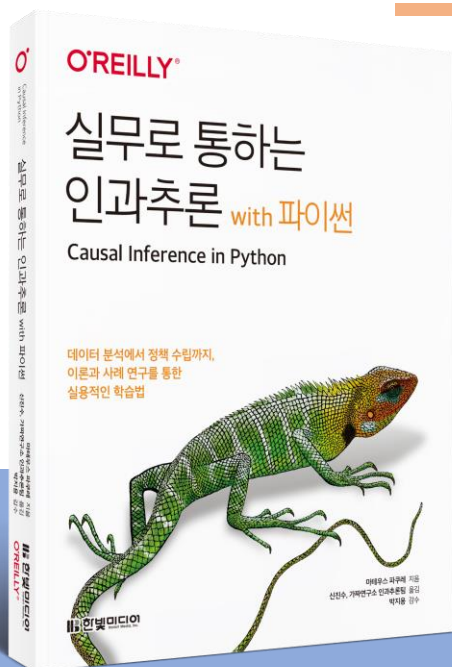
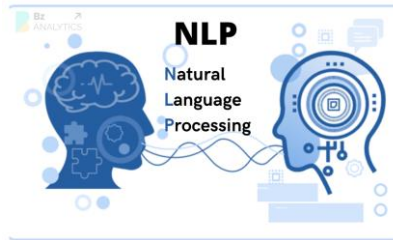


# 『실무로 통하는 인과추론 with 파이썬』 특강

## 5강: 실험 설계에서의 매칭 활용



## 유정현 (Jeonghyun Yoo)



- FC온라인(FIFA 4), 메이플스토리 등 넥슨 게임 유저 맞춤형 콘텐츠 개발
- A/B Test, 콘텐츠 및 프로덕트 분석
- 마케팅 콘텐츠 최적화(추천) 알고리즘 개발
- NLP, Data mining 전공
  - 연구 주제: Sentiment Analysis, Corpus
- 긍정적 사고방식, **경제적 자유 달성**하기

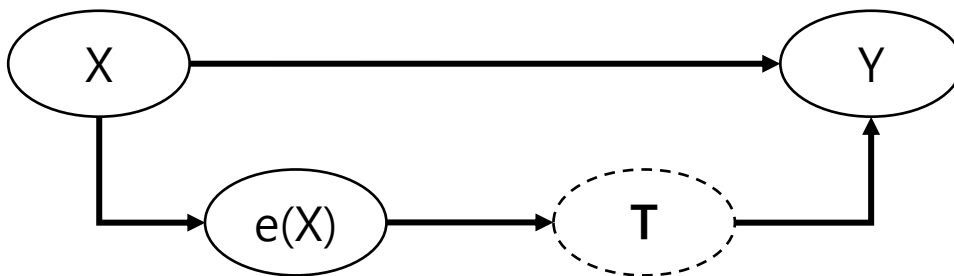


행복하게 살거

- 매칭(Matching) 을 이해하기 위한 2가지 개념
  - > 성향점수(PS), 역확률 가중치(IPW)
- 이중 강건 추정 (Doubly robust)
  - > 디자인 기반, 모델 기반
- 연속형 처치에서의 일반화 성향점수
- 요약

# 매칭(Matching)이란 무엇인가?: 성향점수

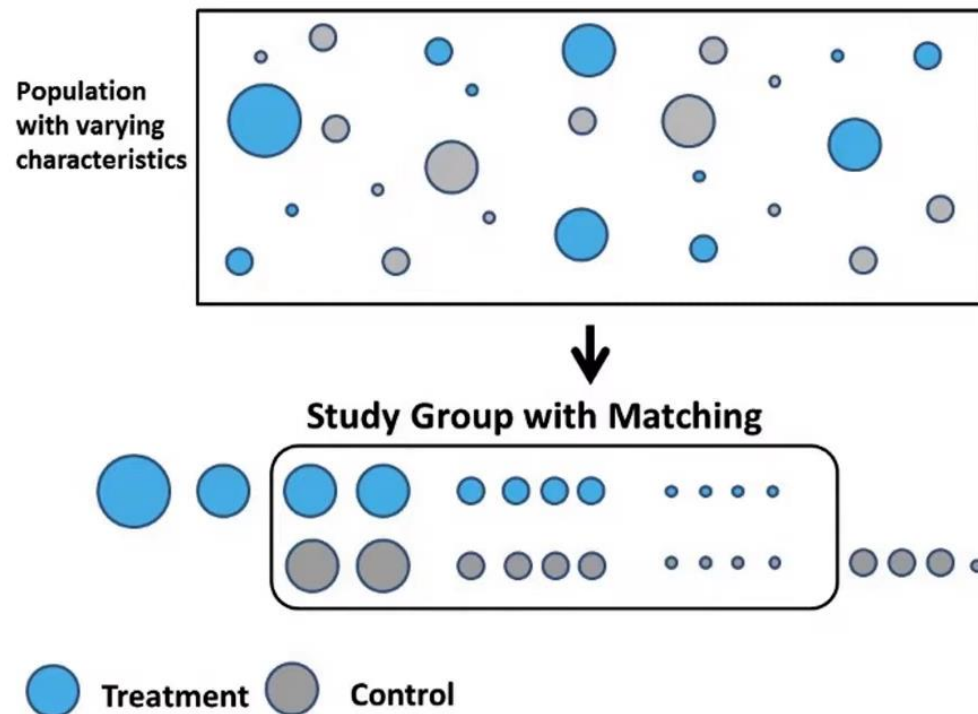
- **성향점수 (Propensity score)**
  - > 관찰 연구에서 두 그룹을 비교할 때, 각 그룹이 **여러 특성에서 비슷하게 만들어 주는 도구**
  - > 차원 축소 기법으로 볼 수 있음
- 즉,  $e(X)$ 를 통제하면, 곧  $X$ 를 직접 통제할 때와 동일한 효과가 나타난다는 가정



- **새로운 교육 프로그램, 특정 프로덕트의 이벤트 효과를 추정하고 싶을 때**
  - 단순 그룹 비교? 편향의 문제 발생
  - 이벤트 참여자들은 원래 프로덕트를 잘 이용하는 그룹일 확률이 높다

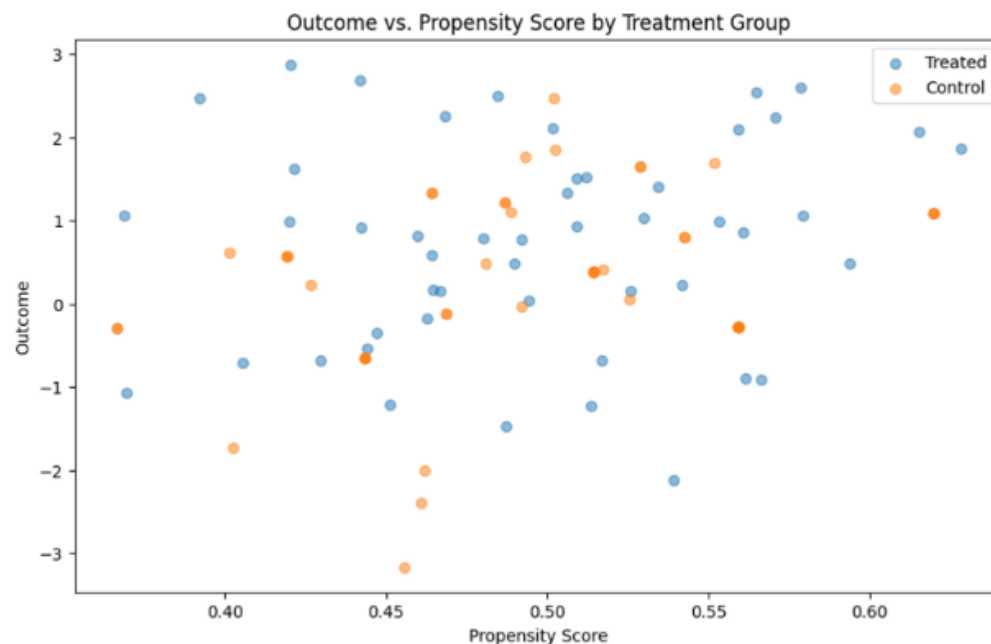
# 매칭(Matching)이란 무엇인가?: 성향점수

- 성향점수 활용 예시
  - 앱 이용 횟수, 결제 수, 주문 체결 수, 이용 시간 등을 기반으로 이벤트 참여 확률을 계산함
  - 각 유저의 참여 확률은 성향점수로 추정하며, 이벤트에 참여한 유저와 참여하지 않은 유저를 매칭함



- **성향점수 계산 및 활용 방법**

- 성향점수를 계산할 때 사용할 독립 변수를 선택함
  - > 처치에 영향을 미치는 모든 변수들 포함
- 성향점수 계산, 로지스틱 회귀 분석 모델
  - > 모델의 해석이 중요하고, 선형 가정이 적절
  - > 처치 여부를 독립 변수로 설명하는 모델을 생성, 로짓 함수 활용
- 성향점수 매칭
  - > KNN 알고리즘, K-d Tree 알고리즘
  - > 매칭 추정량의 편향

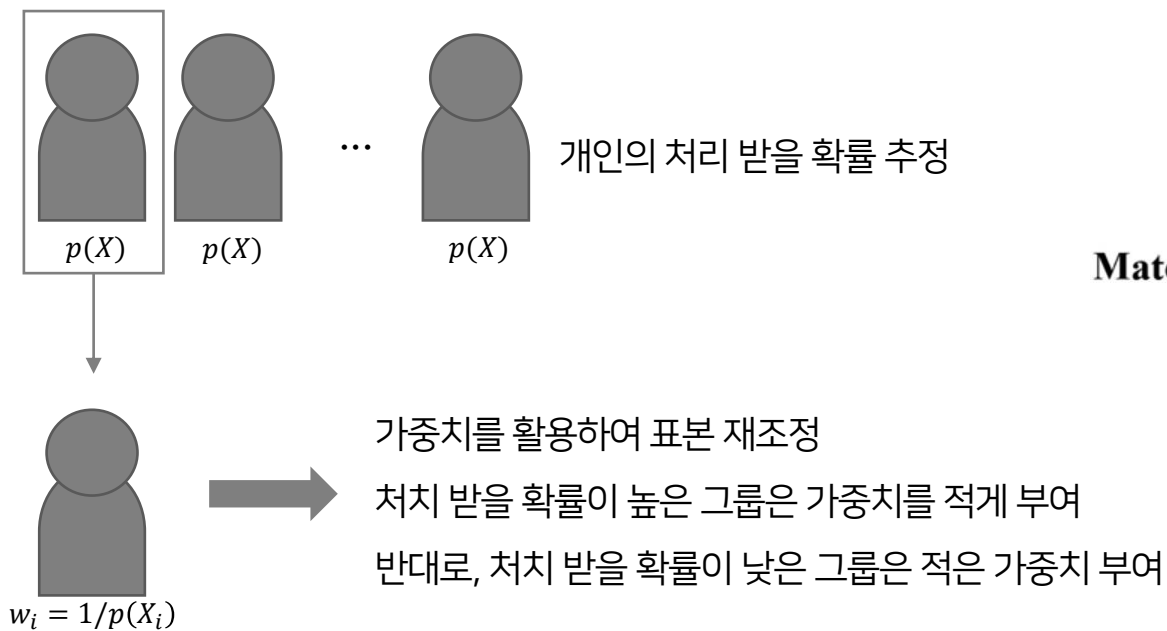


- **성향점수의 직교화**

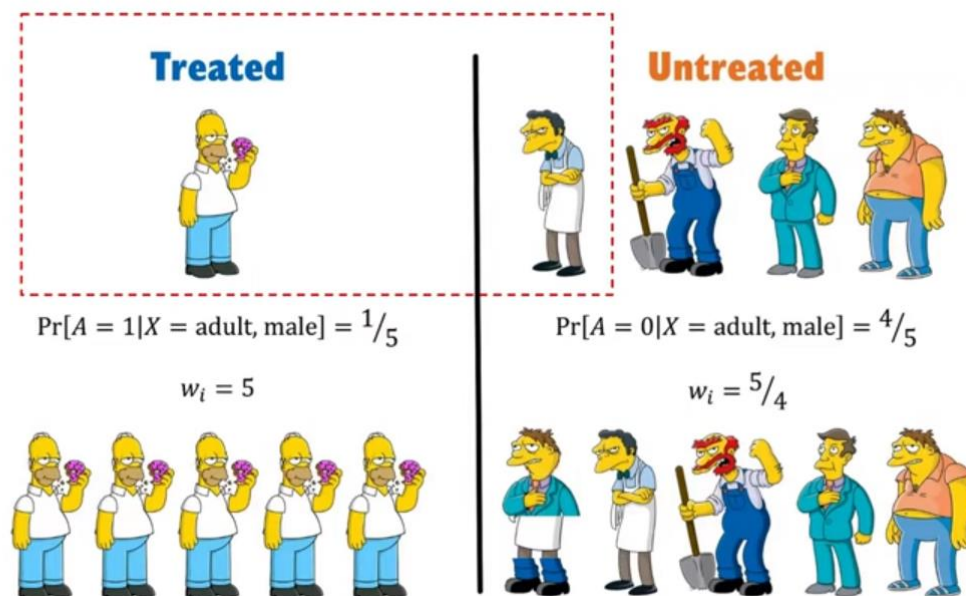
- 선형회귀도 성향점수 추정과 매우 비슷하며, OLS는 성향점수 추정과 매우 비슷하게 처치 배정 매커니즘을 모델링

# 매칭(Matching)이란 무엇인가?: 역확률 가중치

- 역확률 가중치 (IPW)
  - 개개인의 특정 처리를 받을 확률의 역수를 가중치로 부여함



## Matching



# 매칭(Matching)이란 무엇인가?: 역확률 가중치

## • 장점

- > 처리와 통제 그룹 차이를 조정할 수 있으므로, 편향을 비교적 줄일 수 있음
- > 임상 시험과 같은 실험적 연구 외, 관찰 연구에서도 인과 관계를 추정할 때 유용함

## • 단점

- > 모델 의존성
- > 매우 작은 확률로 인한 가중치의 극단값 문제 (이상치)

## • 역확률 가중치의 분산

- 처리 효과와 신뢰성을 평가하기 위한 방법
- **부트스트랩** (joblib.Parallel, delayed)

$$\omega_i = \frac{1}{p(x_i)}$$

역확률 가중치

$$\hat{\theta}_b^* = \frac{\sum_{i=1}^n \omega_i^* Y_i^*}{\sum_{i=1}^n \omega_i^*}$$

부트스트랩 샘플에 대해 가중 평균을 계산

$$\hat{\theta}_1^*, \hat{\theta}_2^*, \dots, \hat{\theta}_B^*$$

부트스트랩 추정치

$$\hat{\sigma}^2 = \frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \overline{\hat{\theta}^*})^2$$

부트스트랩 추정치의 분산



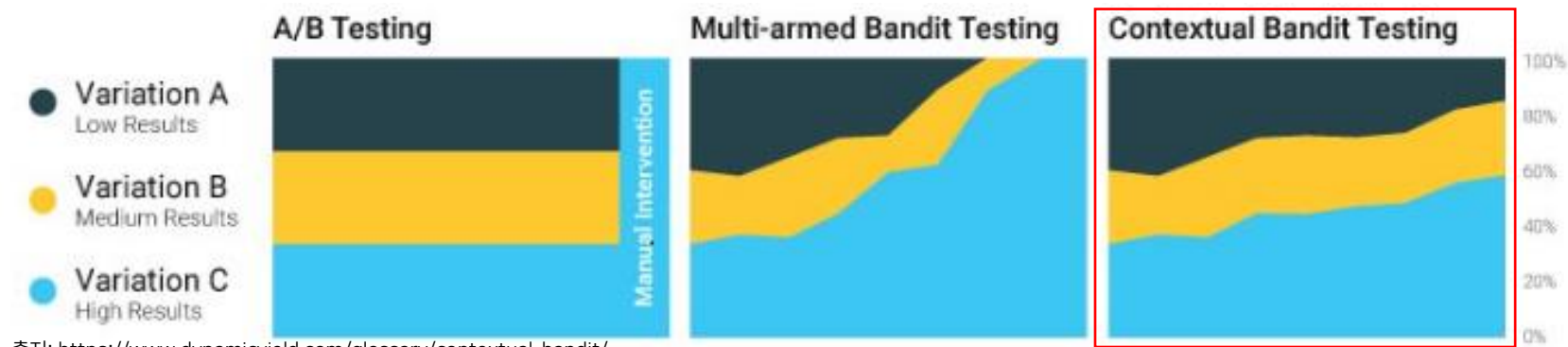
# 매칭(Matching)이란 무엇인가?

- 활용 사례: **인과적 콘텍스트 밴딧**

- > 강화학습의 한 종류(MAB), **선택적 의사결정 정책을 학습**하는 것이 목표임

- 알고리즘의 목표: **최적의 방식으로 실험(처치)군 배정**

- > 무작위가 아니기 때문에 실험 배정 확률에 가중치를 부여하여, 재조정 할 수 있음

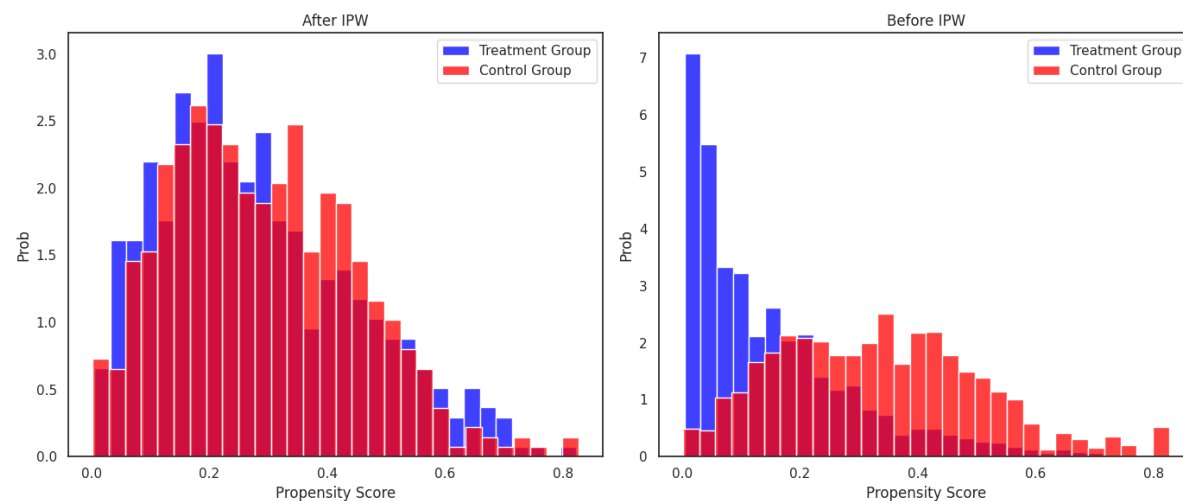
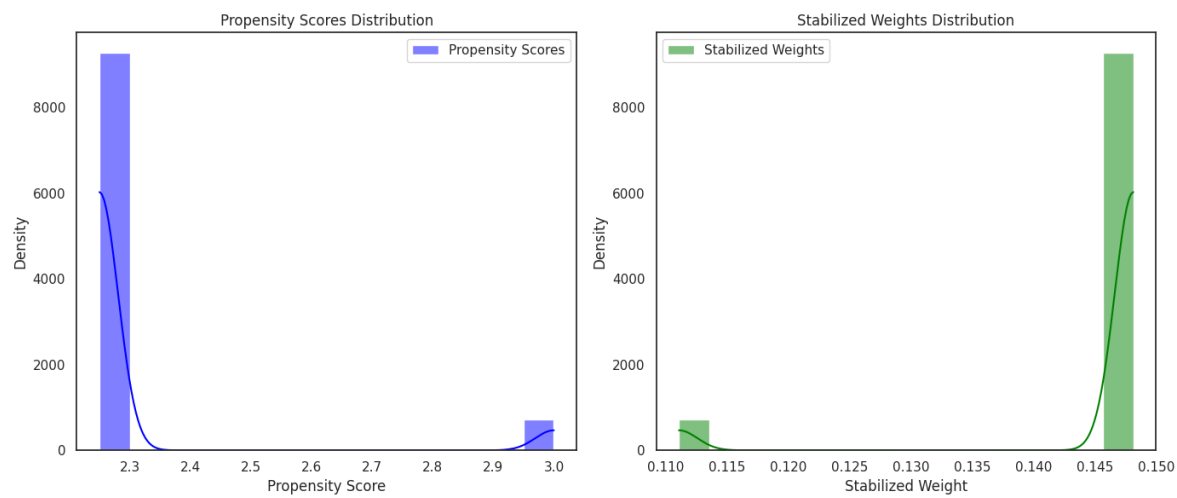


# 매칭(Matching)이란 무엇인가?

- 활용 사례: 안정된 성향점수 가중치, 유사 모집단
  - > 가중치가 극단적이지 않도록 안정성을 더하는 방식
  - > IPW가 편향을 제거하는 방식

$$\text{Stabilized Weight} = \frac{P(T = t)}{P(T = t | X)} \quad \frac{\text{행동 a가 선택될 전체 확률(사전 확률)}}{\text{주어진 컨텍스트 x에서 행동 a가 선택될 확률(성향 점수)}}$$

IPW 적용 후, 적용 전의 성향점수 분포  $Y_0 | T = 1$  를 파악하려는 목적



- 디자인 기반 식별 + 모델 기반 식별

- > 디자인 기반: 성향점수(Propensity Score)

- > 모델 기반: 회귀(Regression)

- 반사실  $Y_t$ 의 이중 강건 추정량 식

- 디자인 기반(성향점수)가 잘못되었지만, 결과 모델(회귀)가 정확하다고 가정

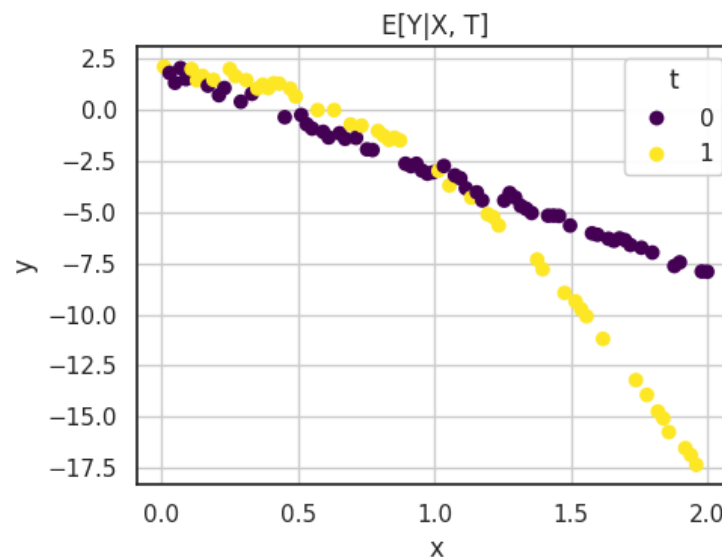
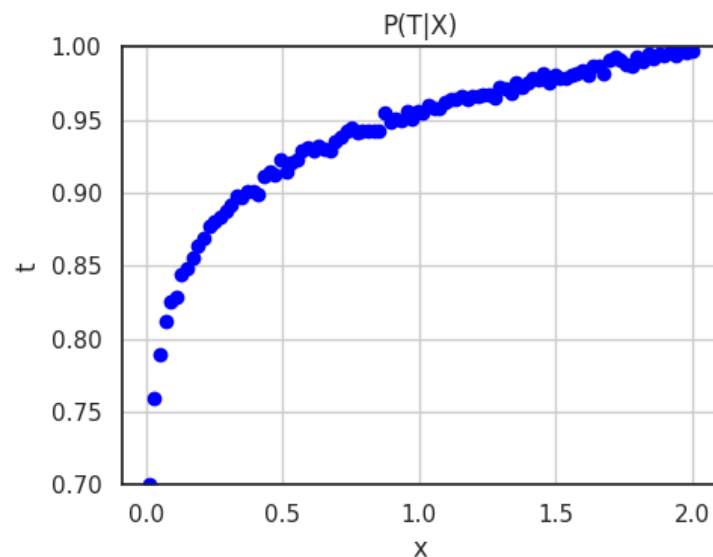
회귀 모델이 정확하다면  $E[Y - \hat{M}(X)] = 0$ 이므로 두 번째 항은 0에 수렴 
$$\hat{\mu}_t^{DR}(\hat{m}, \hat{Q}) = \frac{1}{N} \sum \hat{m}(X) + \frac{1}{N} \sum \left[ \frac{T}{\hat{e}(X)} (Y - \hat{m}(X)) \right]$$

- 결과 모델(회귀)가 잘못되었지만, 디자인 기반(성향점수)가 정확하다고 가정

성향점수가 정확하다면  $T - E^{\wedge}(X)$ 은 0으로 수렴 
$$\hat{\mu}_t^{DR}(\hat{m}, \hat{Q}) = \frac{1}{N} \sum \frac{TY}{\hat{e}(X)} + \frac{1}{N} \sum \left[ \frac{T - \hat{e}(X)}{\hat{e}(X)} \hat{m}(X) \right]$$

- 처치 모델링이 쉬운 경우

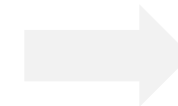
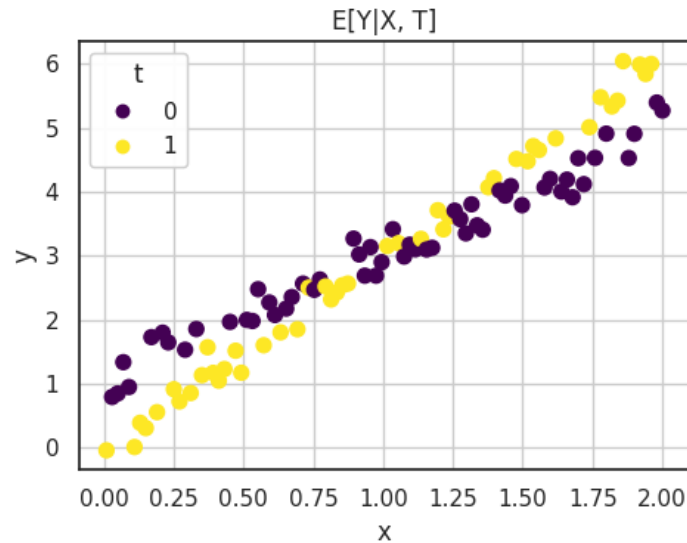
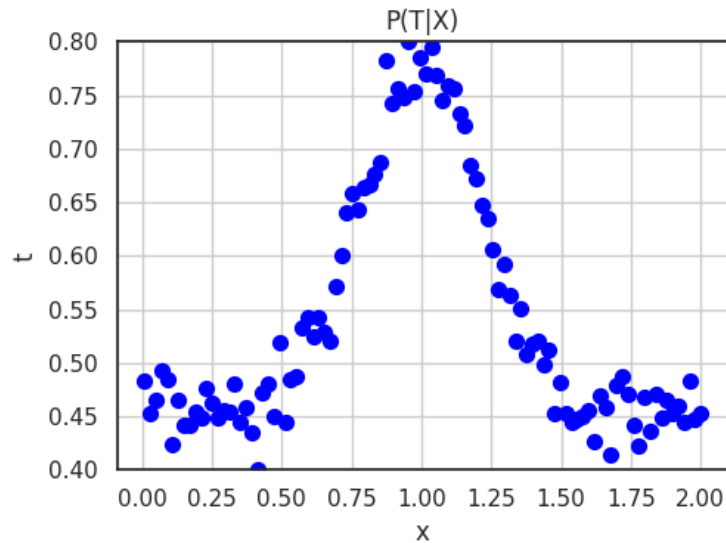
- > 처치 받을 확률을 예측하기 쉬운 상황  $e(X) = \frac{1}{1 + e^{-(1+1.5x)}}$
- > 특정 이벤트를 처치로 가정했을 때, 이벤트 참여 조건이 정의되어 있는 경우
- > 데이터의 실제 곡률(Curve)을 알면 정확하게 모델링할 수 있음
- > 그러나 현실에서는 데이터 생성 과정을 알 수 없기 때문에, **IPW를 활용**



→  $E[Y_t|X]$ 가 잘못되어도  
 $P(T|X)$ 를 모델링하기 쉬운 때

- 결과 모델링이 쉬운 경우

- > 결과를 예측하기 쉬운 상황, 결과에 미치는 요인들이 명확하고 쉽게 측정가능한 경우 (ex. Boston Housing Price)
- > 성향점수 모델링이 상대적으로 복잡하므로, IPW로는 실제 ATE를 추정하기 어려움
- > 이 때, 회귀 모델은 ATE를 정확하게 추정할 수 있음

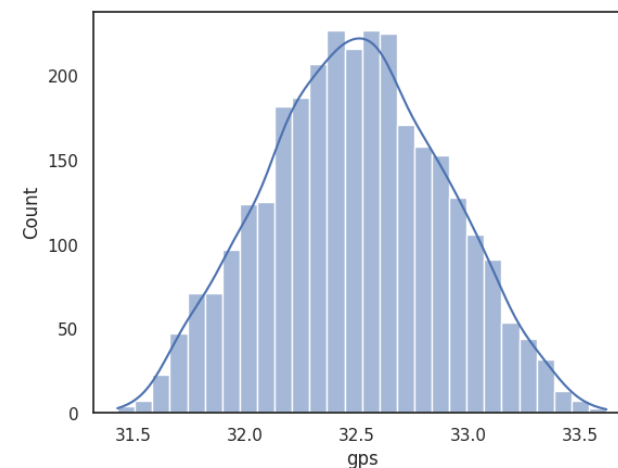
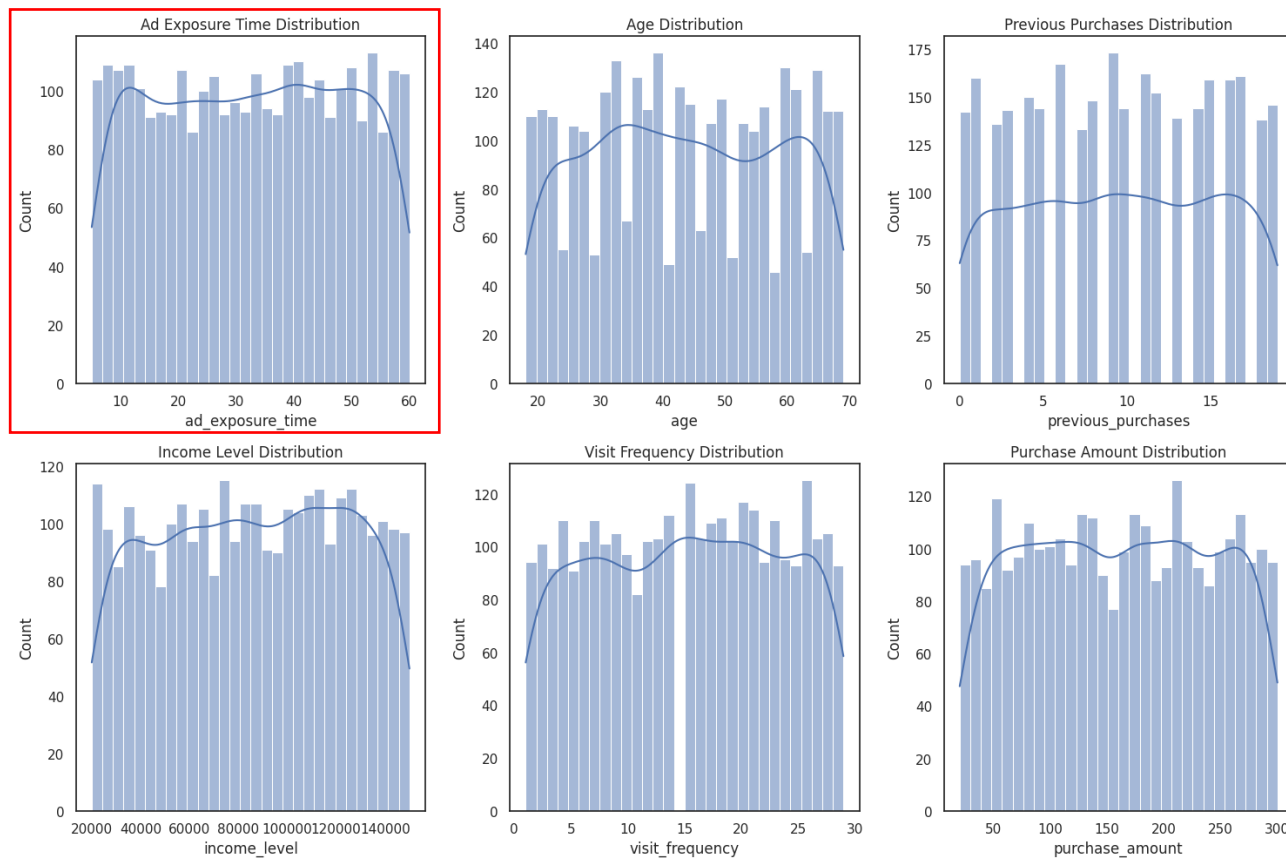


$P(T|X)$ 가 잘못되어도  
 $E[Y_t|X]$ 를 모델링하기 쉬운 때

- 성향점수 가중치에서는 비모수적 방식으로 추정되며,  $T$ 가 연속형일 때, 잠재적 결과  $Y_t$ 는 무한히 존재
- 어떤 경우일까?
  - > Binary가 아니라, 처치의 강도나 수준이 연속적인 값을 가질 때 (어떤 사람은 광고를 1.3시간 시청했고, 어떤 사람은 광고를 2시간 시청함)
  - > 이 경우, 일반화 성향점수는 각 사람이 특정 처치 수준(광고 시청 시간 시간)을 받을 확률을 모델링함
  - > 광고 노출 시간(연속형 처치 변수)에 따른 소비자의 구매 행동(잠재적 결과)을 평가, 각 소비자가 특정 시간 동안 광고에 노출될 확률 추정
- 균형 맞추기, 인과 효과 추정
  - 균형 맞추기: 처치 수준에 따라 집단 간의 차이를 보정하며, 다른 요인들이 결과에 미치는 영향 최소화
  - 인과 효과 추정: 특정 처치 수준에서의 인과 효과를 추정 (ex. 광고 시청 시간이 구매 금액에 미치는 영향 추정)
- 핵심은 각 사람마다 나이, 이전 구매 이력 등 공변량으로 광고 노출 시간을 예측하는 모델을 생성

- 광고 노출 시간과 소비자 구매 행동 간의 인과 관계를 추정하는 과정

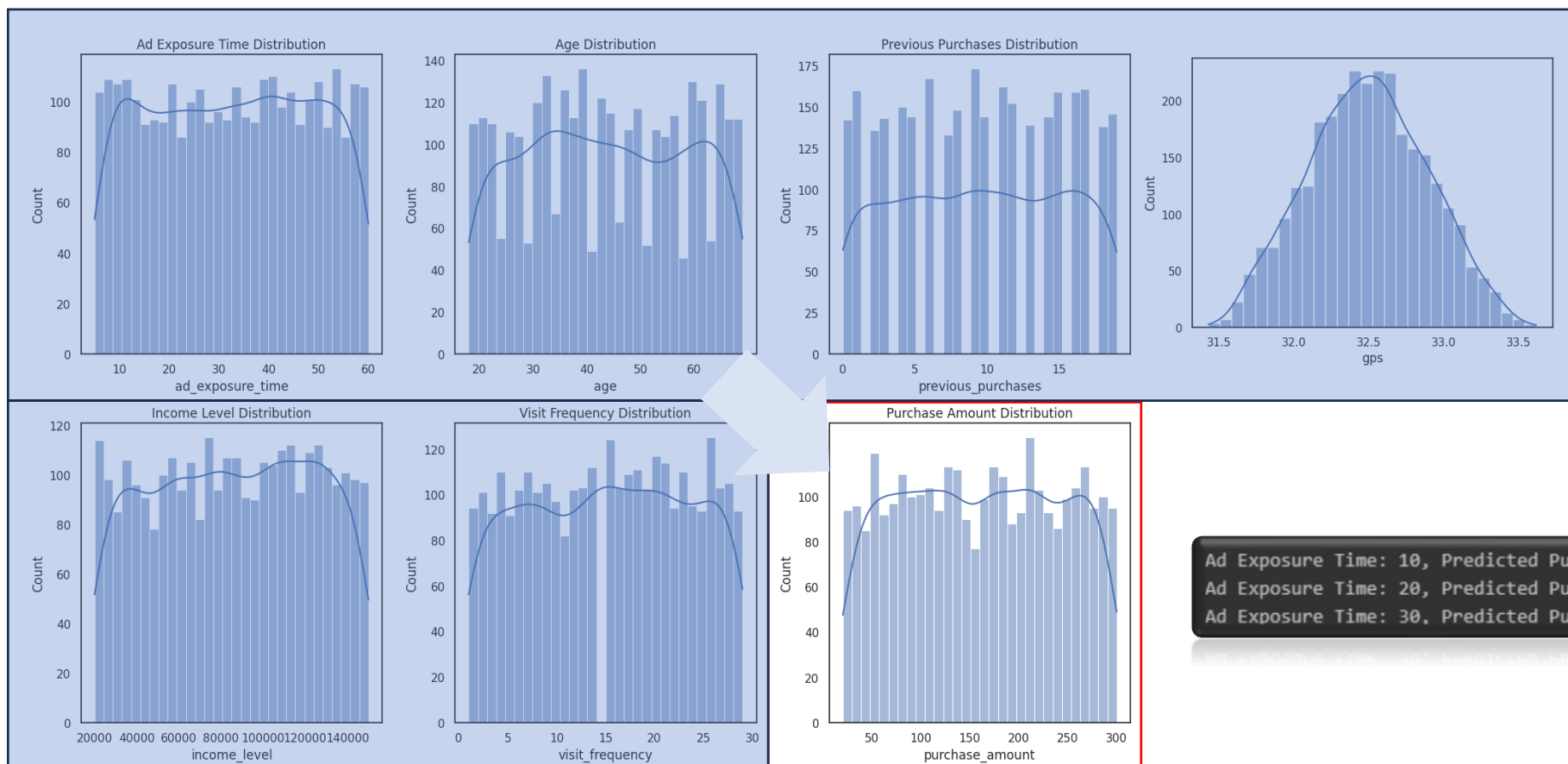
> 광고 노출 시간(연속형 변수)이 소비자의 구매 행동에 미치는 영향을 평가



# 연속형 처리에서의 일반화 성향점수 (사례)

- 광고 노출 시간과 소비자 구매 행동 간의 인과 관계를 추정하는 과정

> 광고 노출 시간(연속형 변수)이 소비자의 구매 행동에 미치는 영향을 평가



Ad Exposure Time: 10, Predicted Purchase Amount: 163.68010081914247  
Ad Exposure Time: 20, Predicted Purchase Amount: 162.1774024082255  
Ad Exposure Time: 30, Predicted Purchase Amount: 160.6747039973085



- 매칭에서 성향점수와 IPW는 중요한 개념
  - > 실험 전, 실험 후 모두 적용 가능
  - > 직교화: 처치를 잔차화, 공변량과 선형 독립인 새로운 공간으로 투영
  - > IPW: 처치의 차원을 유지, 데이터를 처치 성향점수의 역수로 재조정
- 역확률 가중치(IPW)는 편향을 보정하는 핵심 방법 중 하나이며, 처치를 모델링한다.
- 이산형 처치일 경우, IPW 사용을 선호
- 연속형 처치의 경우, IPW보다는 일반화 성향점수나, 처치 반응 등을 통해 추론하는 방법이 더 생산적