
Causal Inference: In Statistical Aspects

Yitae Kwon

SFERS

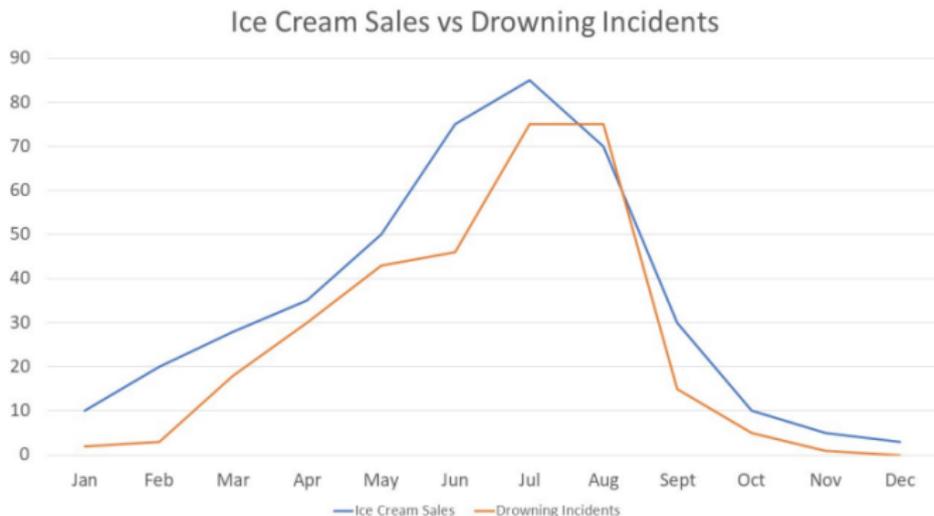
2023-2

Contents

- 1 Causality vs. Associativity
 - Confounding Effect
 - Causal Effect Parameters
- 2 Causal Inference in Randomized Experiments
- 3 Causal Inference in Observational Studies
 - Outcome Regression
 - Balancing the Covariates: Matching
- 4 Weighting
 - Inverse Probability Weighting
 - Doubly Robust Estimator

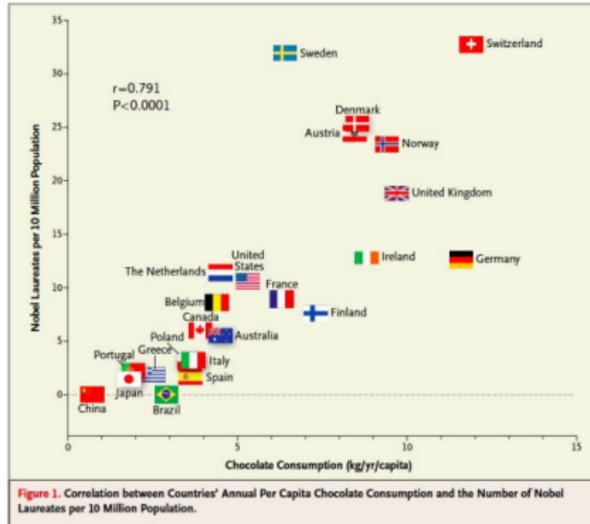
Causality vs. Associativity

The Danger of Ice Cream...



- 아이스크림의 판매량이 증가할수록 익사사고가 증가하는 경향성이 나타난다. 너무 맛있는 나머지 정신을 잃어버리는 게 아닐까?

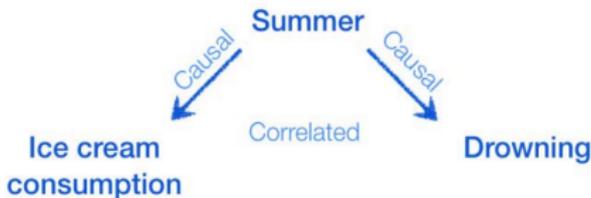
Sweet Chocolate Makes Nobel Prize



- 초콜릿 소비량이 많은 국가일수록 인구당 노벨상 수가 증가한다고 한다. 초콜릿이 연구 능력에 도움을 주는 게 아닐까? (Messerli, 2012)

Counfounding Effect

- 둘 사이에는 **인과적인**(causal) 관계가 존재하는지 알 수 없다.
- 단지 둘 사이에 **연관성**(association)이 존재할 뿐이다.



- 여름, 혹은 뜨거운 기온은 모두 인과적으로 아이스크림 소비량과 익사사고 수를 증가시킨다.
- 이처럼 보려고 하는 두 요인에 모두 인과적으로 영향을 미쳐 두 요인 사이의 상관관계를 만드는 요인을 **교락요인**(confounder)라고 부른다.
- 마찬가지로 초콜릿 소비량과 노벨상 수상자 비율 간에는 '유럽'이라는 지역적 원인과 '소득' 등이 교락요인이 될 수 있다.

Counfounding Effect

- 인과추론 연구에서는 이러한 교락요인의 통제가 필요하다.
- 관찰연구 상에서 어떤 약물의 콜레스테롤 수치 하락 효과를 보려고 한다.
- 임상시험에서는 그냥 동전 던지기로 무작위로 피험자들에게 약물 혹은 위약을 제공하고 그 효과 차이를 **Wilcoxon test** 등으로 검정하면 된다.
- 하지만 사후적으로 그 효과를 분석하고자 하는 관찰연구 상황에서는 이러한 추론이 불가능하다.
- 만약 원래 콜레스테롤 수치가 높은 환자들만 약물을 제공받았다면, 약물 처리군의 콜레스테롤 수치가 대조군보다 높게 관측될 수도 있다.
- 이는 '처리 전 콜레스테롤 수치'가 교락요인으로 작용해 '처리 후 콜레스테롤 수치'와 '약물 처리' 사이의 상관관계를 비틀기 때문이다.

Counfounding Effect

- SFERS의 모 팀장(들)은 '담배가 폐암에 미치는 영향이 없다. 사실 담배를 마렵게 하는 유전자가 있어 그 유전자가 폐암 발병 확률도 높인다'라고 주장하려 한다.
- 이 경우 특정 유전자가 교락요인으로 작용해, 흡연 가능성과 폐암 발병 가능성을 모두 높이기 때문에 흡연자가 폐암에 걸릴 확률이 높다고 주장할 수 있다.
- 한 연구에 따르면, 이 유전자가 폐암에 미치는 영향이 담배가 폐암에 미치는 영향의 9배 이상은 되어야 담배의 영향이 유의하지 않다고 말할 수 있는 수준이다. (Cornfield, 1959)
- 보건 분야의 경우 9배가 아니라 1.5~2배만 되어도 교락요인이 없다고 본다. 즉 흡연이 폐암 발병 확률을 높임은 거의 확실하다고 주장할 수 있다.

Setting

- 관측값은 Y_1, \dots, Y_n 으로 주어진다.
- 각 개체의 관측된 **공변량**(covariates)는 X_1, \dots, X_n 으로 주어진다.
- 각 개체의 관측되지 않은 공변량은 u_1, \dots, u_n 으로 주어진다.
- 각 개체의 처리 여부는 Z_1, \dots, Z_n 으로 주어진다.
- 각 개체가 처리를 받았을 때와 처리를 받지 않았을 때의 **잠재적 결과**(potential outcome)은 $Y_i(\cdot)$ 처럼 주어진다.
- 따라서, group이 2개인 모형에서 관측값 Y_i 는

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

처럼 쓸 수 있다.

Setting

- 처리의 인과적 효과는

$$\tau_i = Y_i(1) - Y_i(0)$$

으로 주어진다.

- 이는 각 개인 i 마다 다르기 때문에, 평균적인 처리 효과를 **average treatment effect**(ATE)로 정의한다.

$$ATE = E[\tau_i] = E[Y_i(1)] - E[Y_i(0)]$$

- 한편 처리를 받은 대상에 대하여 처리 효과를 보기 위하여서는 **average treatment effect on treated**(ATT)를 사용한다.

$$ATT = E[Y_i(1) - Y_i(0) | Z_i = 1]$$

- 혹은 처리를 받지 않았던 대상이 처리를 받았더라면 얻는 처리 효과를 **average treatment effect on controls**(ATC)로 분석하기도 한다.

$$ATC = E[Y_i(1) - Y_i(0) | Z_i = 0]$$

Setting

- 언뜻 보기엔 ATE, ATT, ATC가 모두 같아 보인다.
- 포인트는 어떠한 교락요인의 존재로 처리에 따른 효과의 크기와 처리 여부 사이에 상관성이 존재할 수 있다는 것이다.
- 관측된 공변량 X_i 이 $U(0, 1)$ 을 따르고,

$$Y_i(1) = 5X_i, \quad Y_i(0) = 3X_i, \quad Z_i = \begin{cases} 1 & X_i > 0.5 \\ 0 & X_i \leq 0.5 \end{cases}$$

처럼 결정된다고 해보자.

-

$$ATE = E[5X_i - 3X_i] = 1$$

$$ATT = E[5X_i - 3X_i | Z_i = 1] = 1.5$$

$$ATC = E[5X_i - 3X_i | Z_i = 0] = 0.5$$

로 상이함을 알 수 있다.

Setting

- 위의 예시를 조금 더 극단적으로 만들면, 어떠한 parameter를 사용하냐에 따라 처리효과의 부호가 완전히 바뀌어 버릴 수도 있다.
- 우리는 어떤 처리, 어떤 치료, 혹은 어떤 정책이 응당 받아야 할 사람에게 주어졌을 때 그 사람이 받는 효과를 보고자 한다. (특히, 사회과학 혹은 보건 분야에서)
- 이에 따라 인과추론 연구에서는 ATT를 가장 많이 사용하며, 우리의 가이던스에서도 마찬가지이다.

Fundamental Problem of Causal Study

개체 i	처리여부 Z_i	소득		경제학부 X_{i1}	GPA X_{i2}	부모소득 u_i
		$Y_i(1)$	$Y_i(0)$			
1	1	2000	1500	1	4.2	2000
2	0	2500	2100	0	3.8	2600
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	
n	1	6000	3500	0	2.2	9000

- 우리가 원하는 테이블은 위처럼 각 개체의 처리여부부터 이에 따른 potential outcomes, 심지어는 공변량들까지 모두 주어진 테이블이다.
- 하지만 실제 관찰연구에서는...

Fundamental Problem of Causal Study

개체 i	처리여부 Z_i	소득		경제학부 X_{i1}	GPA X_{i2}	부모소득 u_i
		$Y_i(1)$	$Y_i(0)$			
1	1	2000	?	1	4.2	???
2	0	?	2100	0	3.8	???
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	
n	1	6000	?	0	2.2	???

- 관측되지 않은 공변량 u_i 는 알아보지 않았으니 모르는 게 당연하다.
- potential outcome도 $Y_i(Z_i)$ 부분만 알 뿐, $Y_i(1 - Z_i)$ 는 모른다.
- 다기간, 다그룹 연구에서는 이러한 문제가 더 심해진다. 2개 중 1개를 모르는 것에서 넘어서서, n 개 중 $n - 1$ 개를 모르기 때문이다.

SUTVA

- 마지막으로 인과추론 연구에서 가장 많이 사용되는 가정 하나를 추가한다.
- **SUTVA**(Stable Unit Treatment Values Assumption):
 - ① **No inference**: 각 개체의 potential outcome은 다른 개체의 potential outcome, 혹은 그들의 처리 여부나 공변량 등에 의해 변하지 않고 완전히 독립적이다.
 - ② **No hidden variations of treatment**: 각 개체는 동일한 처리를 받는다.

Causal Inference in Randomized Experiments

Causal Inference in Randomized Experiments

- **Randomized Experiments**(랜덤화 실험)에서는 처리 여부를 무작위로 할당한다.
- 대조군과 처리군이 어느 정도의 동질성을 가지게 된다. 즉

$$(Y(1), Y(0)) \perp\!\!\!\perp Z$$

이다. 이를 **causal assumption** 혹은 **identification assumption**이라고 부른다.

- 처리 여부에 관계없이 개인 특성에 의하여 potential outcome이 결정된다.
- 따라서 $ATE = ATT = ATC$ 이다.

Causal Inference in Randomized Experiments

- ATE(ATT 혹은 ATC)의 추정량도 쉽게 구할 수 있다.

-

$$\begin{aligned}\widehat{ATE} &= \hat{E}[Y(1) - Y(0)] \\ &= \hat{E}[Y(1)] - \hat{E}[Y(0)] \\ &= \hat{E}[Y(1)|Z = 1] - \hat{E}[Y(0)|Z = 0] \\ &= \bar{Y}_t - \bar{Y}_c\end{aligned}$$

- \bar{Y}_t 와 \bar{Y}_c 는 각각 처리군과 대조군의 표본평균을 의미한다.
- 이는 우리가 일반적으로 생각할 수 있는 처리효과의 추정량과 동일하다.

Balancing the Groups

- 하지만 대부분의 관찰연구에서는 처리가 랜덤하게 주어지지 않는다.
- 관측한, 혹은 관측하지 못한 공변량들이 교락요인으로 작용하여 potential outcome과 처리 여부 사이에 상관성을 만든다.
- 이는 처리군과 대조군의 공변량 분포가 서로 다른 결과를 낳는다.
- 이를 두고 두 그룹 사이의 **균형**(balance)가 깨졌다고 말한다.
- 두 그룹 사이에 균형이 깨지게 되면, $\hat{E}[Y(1)]$ 와 $\hat{E}[Y(1)|Z = 1]$ 가 다르기에 추정량에 bias가 생긴다.

Balancing the Groups

- 관측할 수 있는 공변량으로부터 발생하는 bias를 **overt bias**라고 부른다.
- 이는 공변량을 고려한 **outcome regression**, IPW를 위시한 **weighting**, 간단한 **matching** 등으로 극복할 수 있다.
- 한편 관측할 수 없는 공변량으로부터 발생하는 bias를 **hidden bias**라고 한다.
- 이는 관측되지 않은 공변량의 영향을 설명할 수 있도록 **instrumental variable**을 도입하거나, **difference-in-differences**를 통해 극복된다.
- 경제학에서는 공변량을 모두 파악하기가 힘들기 때문에 hidden bias를 통제하는 방법을 선호한다.
- 물론 overt bias를 통제하는 것 역시 중요하다. 오늘은 이를 다루볼 것이다.

Causal Inference in Observational Studies

Identification Assumption

- 우리는 공변량들이 같다면 처리 여부는 랜덤하게 정해진다고 가정을 완화할 수 있다. 앞을 예시로 두면, 경제학부에 GPA가 4.0인 사람들을 100명 모아두면 그들의 처리 여부는 동전 던지기(앞면과 뒷면의 등장 확률이 다를 수 있는)로 결정하자는 것이다.

- 즉

$$(Y(1), Y(0)) \perp\!\!\!\perp Z | X$$

임을 가정해줄 수 있다. 이를 **unconfoundness** 혹은 **ignorability** 조건이라 한다.

- 추가적으로 이후의 논의를 위하여 모든 X 에 대하여 동전은 앞면과 뒷면 모두 나올 수 있는 것으로 한다. 즉 **positivity(overlap)** 가정

$$0 < P(Z = 1 | X) < 1$$

을 적용해 **strong ignorability**를 가정한다.

Identification Assumption

- 이제 ATE를 추정하여 보자.
- 먼저 $E[Y(1)]$ 은

$$E[Y(1)] = E[E[Y(1)|X]] = E[E[Y(1)|X, Z = 1]] = E[E[Y|X, Z = 1]]$$

으로 추정해줄 수 있다.

- 따라서

$$\hat{E}[Y(1)] = \frac{1}{n} \sum_{i=1}^n \hat{E}[Y(1)|X_i]$$

처럼 추정해줄 수 있다.

- 주의할 것은 $\hat{E}[Y(1)]$ 를 구할 때 전체 n 개의 자료를 모두 사용한다는 것이다.

Outcome Regression

- 이제 관심사는 $\hat{E}[Y(1)|X_i]$ 를 어떻게 추정하냐는 것이다.
- 가장 편하면서 자주 사용되는 것이 **outcome regression**으로,

$$E[Y|X] = \beta_0 + \beta_1 Z + \beta_2 X$$

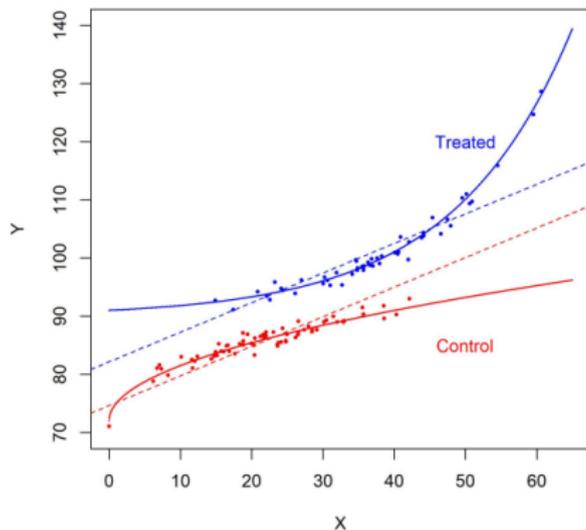
처럼 모형화하는 것이다. 회귀식의 형태는 해당 분야에서 이미 사용되고 있는 이론을 바탕으로 한다.

- 그렇다면 $\hat{E}[Y(1)|X]$ 과 $\hat{E}[Y(0)|X]$ 의 차이는 $\hat{\beta}_1$ 으로 주어진다.
- 따라서 ATE의 추정량은 $\hat{\beta}_1$ 처럼 쉽게 쓸 수 있다.

Outcome Regression

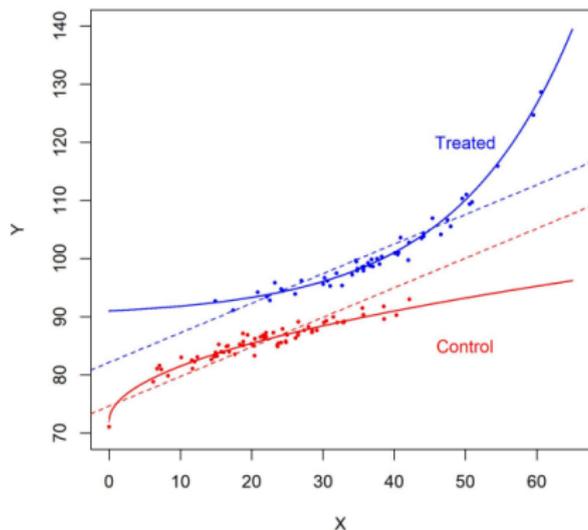
- outcome regression의 문제는 네 가지 정도이다.
 - ① Model misspecification 문제.
회귀모형이 잘못 추정되는 경우 이는 큰 오차를 불러오게 된다. 특히 선형모형을 적합할 수 없는데 선형모형을 적합하거나, interaction term을 누락하거나 하면 **effect modification**을 파악해줄 수 없다.
 - ② Outlier의 문제.
회귀모형의 적합에서 outlier의 존재는 회귀직선의 기울기를 변화시킴으로써 제대로 된 추론이 불가능하게 만든다.
 - ③ Extrapolation의 문제.
거듭제곱항 등이 들어가는 경우, 회귀직선의 X 범위 바깥에서 회귀모형을 이용한 추정량에 큰 편향이 발생할 수 있다.
 - ④ Cherry picking의 문제.
outcome regression의 과정에서 관측값들을 모두 보게 된다. 이때 가설에 맞지 않는 데이터들을 이상점으로 취급해 없애버리고 원하는 결론을 얻어버릴 수도 있다.

Outcome Regression



- 점선은 관측값들을 모두 모아 outcome regression을 수행한 결과다.
- 실선은 실제로 관측값이 비롯되는 곡선이다.

Outcome Regression

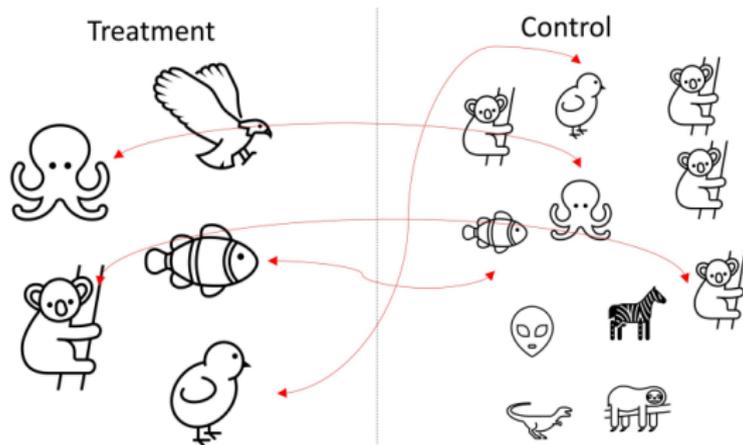


- 실제로는 극단적인 X 로 가면 처리효과가 극대화된다.
- 하지만 outcome regression 상에서는 그 효과가 항상 상수인 것으로 보인다. 이는 overlap condition이 위배되었기 때문이다.

Idea of Matching

- X_i 의 potential outcome에의 영향을 통제한다면, overt bias를 효과적으로 통제할 수 있을 것이다.
- 우리가 랜덤화 실험에서는 랜덤하게 처리 여부를 결정했었으므로, 처리군과 대조군의 공변량 X_i 분포 역시 동일하였다. 따라서 각 X_i 가 potential outcome에 미치는 영향이 처리군과 대조군에서 동일하기에, ATE의 계산에 무리가 없었다.
- outcome regression에서는 회귀분석으로써 X_i 의 영향을 통제한다. 따라서 그 영향을 고려하여 ATE를 추정해줄 수 있다.
- 여기에서는 randomized experiment를 모방하기 위하여, 처리군과 대조군의 X_i 분포 사이에 balance를 맞추는 방법을 알아보려 한다.
- 가장 대표적이고 간단한 것이 **matching**이다.

Idea of Matching



- matching에서는 대조군과 처리군에서 유사한 기질을 가진 것들끼리 매칭한다.
- '유사한 기질'을 가진 것들은 유사한 공변량들을 가지지만 오직 처리 여부에서만 차이가 있다.
- 따라서 두 그룹 사이 balance가 맞춰질 수 있다.

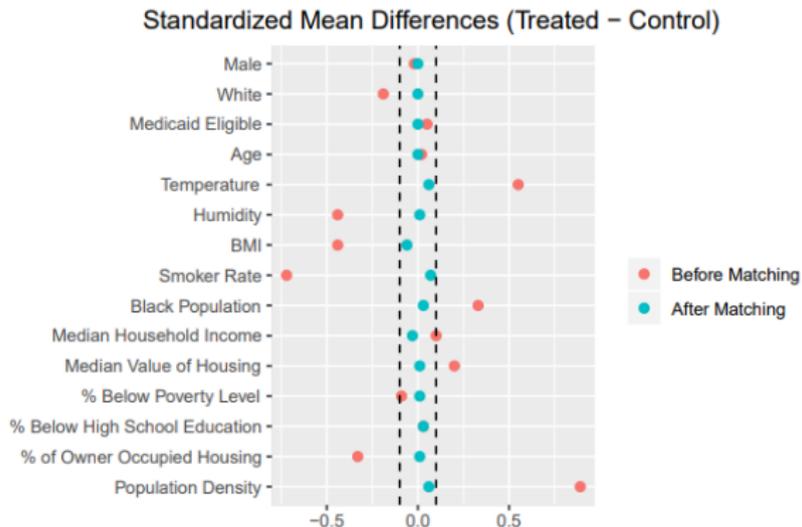
Standardized Difference

- **standardized difference**는 balance가 잘 이루어졌는지를 확인하는 지표다.
- 처리군과 대조군의 공변량 사이의 표준화된 차이를 제공한다.

$$d = \frac{\bar{x}_{treated} - \bar{x}_{control}}{\sqrt{\frac{s^2_{treated} + s^2_{control}}{2}}}$$

- 일반적으로 $d < 0.1$ 수준이도록 매칭하는 것이 좋다.

Standardized Difference



- 매칭 전(붉은색)에 비해 매칭 후(푸른색)에서 balance가 잘 맞는 것을 볼 수 있다.

Matching

- 매칭은 공변량끼리 유사한 것들끼리 매칭시킨 뒤 처리 여부에 따른 관측값 차이를 관찰하기에, 직관적이다.
- 관측값을 사용하지 않기 때문에 cherry-picking의 위험이 적다.
- '분석하기 좋은' 개체들만 선택해서 분석하기 때문에, extrapolation의 문제가 없다.
- 반면 매칭되지 않는 개체들은 버려버리기 때문에, 표본 수가 적은 경우 효율성이 많이 떨어진다.
- 특정 범위의 공변량을 갖는 경우에 대해서만 처리효과를 분석할 수 있다.

Propensity Score

- **propensity score**는 관측된 공변량에 대한 처리의 조건부 확률이다.

$$e(\mathbf{x}) = P(Z = 1|\mathbf{x})$$

- 만약 $e(\mathbf{x})$ 가 같다면, 두 group 사이에 balance가 맞춰짐이 알려져 있다. 즉

$$X \perp\!\!\!\perp Z | e(\mathbf{x})$$

이다.

- 또한, propensity score에 조건부로 potential outcome과 처리 여부가 독립적이다. 즉

$$(Y(1), Y(0)) \perp\!\!\!\perp Z | e(\mathbf{x})$$

이다.

Propensity Score

- propensity score는 곧 모든 covariate들을 1차원으로 축소시켜 둔 것이나 마찬가지다.
- p 개의 covariates를 이용해 매칭하는 것보다, $e(\mathbf{x})$ 를 통해 매칭하는 것이 훨씬 간단하다.
- 단, 문제는 propensity score 자체는 우리가 모르는 값이기에 그 추정량 $\hat{e}(\mathbf{x})$ 를 사용해야 한다는 것이다.
- 여기에는 회귀분석 등이 사용될 수 있다.
- outcome regression에서와 마찬가지로, $\hat{e}(\mathbf{x})$ 를 잘못 추정하면 matching과 balancing이 제대로 이루어지지 않으므로, bias가 커지게 된다.

Propensity Score

- **propensity score matching (PSM)**은 propensity score에 기반하여 매칭을 수행하는 방법이다.
- 처리군 개체의 index를 \mathcal{T} , 대조군 개체의 index를 \mathcal{C} 라고 하자.
 - ① $e(\mathbf{x})$ 의 추정값 $\hat{e}(\mathbf{x})$ 를 구한다.
 - ② $\delta_{t,c} = |\hat{e}_t - \hat{e}_c|$ 를 처리군의 t 와 대조군의 c 에 대해 구한다. 이를 바탕으로 $|\mathcal{T}| \times |\mathcal{C}|$ 의 거리행렬을 구한다.

③

$$\min_{m(\cdot)} \sum_{t \in \mathcal{T}} \sum_{m(t) \in \mathcal{C}} \delta_{t,m(t)}$$

인 mapping $m : \mathcal{T} \rightarrow \mathcal{C}$ 을 찾는다.

Paired Randomized Experiments

- 매칭의 결과로 처리군과 대조군의 쌍 n 개를 얻었다고 하자.
- propensity score의 성질에 의하여, 쌍을 이루는 두 개체의 처리 여부는 무작위로 결정된 것이다. 즉 두 개체 중 하나만 처리를 받아야 한다면, 어떤 개체가 처리를 받을지는 반반 코인토스로 결정된다.(근사적으로)
- $i = 1, 2, \dots, n$ 과 $j = 0, 1$ 는 각각 쌍의 번호와 처리 여부를 의미한다.
- 이를 수식적으로 표현하면

$$Z_{i0} + Z_{i1} = 1$$

$$P(Z_{i0} = 1 | Z_{i0} + Z_{i1} = 1) = P(Z_{i1} = 1 | Z_{i0} + Z_{i1} = 1) = 0.5$$

$$Z_{i0} \perp\!\!\!\perp Z_{j0}$$

Paired Randomized Experiments

- 이제 모든 데이터의 쌍을

$$\mathcal{F} = \{(Y_{ij}(0), Y_{ij}(1), \mathbf{x}_{ij}, u_{ij}), i = 1, 2, \dots, n, j = 0, 1\}$$

으로 표현하자.

- $\mathbf{Z} = (Z_{10}, Z_{11}, \dots, Z_{n0}, Z_{n1})$ 은 n 개 쌍의 처리 상황을 늘어놓은 것이다.
- \mathcal{Z} 는 \mathbf{Z} 의 값으로 가능한 2^n 개의 결과를 모아둔 집합이다.
- $P(\mathbf{Z} = \mathbf{z} | \mathcal{F}, \mathcal{Z}) = \frac{1}{2^n}$ 이다.

Paired Randomized Experiments

- 이제 처리 효과가 있는지는 귀무가설

$$H_0 : Y_{ij}(1) = Y_{ij}(0) \quad \text{for all } i, j$$

를 검정해주면 된다. (정확히는 두 값의 분포의 위치모수가 같은지)

- treated-minus-control difference $D_i = (Z_{i0} - Z_{i1})(Y_{i0} - Y_{i1})$ 를 고려하자.
- H_0 하에서

$$D_i = \begin{cases} Y_{i0} - Y_{i1} & \text{if } Z_{i0} = 1 \\ Y_{i1} - Y_{i0} & \text{if } Z_{i1} = 1 \end{cases}$$

Paired Randomized Experiments

- 이제 D_i 들에 대하여 Wilcoxon test를 적용해줄 수 있다.
- 혹은 M-estimator를 이용해 추정하는 방법도 있다.
- 여기에서는 Wilcoxon Test를 이용할 때의 상황을 조금 더 심도깊게 들여다보고자 한다.
- 지난 시간에 Wilcoxon Test를 간단히 소개하기만 하고 실제 검정 과정은 R의 결과로 대체했었다.
- 좋은 조건 하에서, $n \rightarrow \infty$ 라면 Wilcoxon의 검정통계량 W 는 동점이 없을 때

$$\frac{W - n(n+1)/4}{n(n+1)(2n+1)/24} \xrightarrow{d} N(0, 1)$$

로 정규근사가 가능함이 알려져 있다.

Paired Randomized Experiments

- 따라서 W 의 영분포는 n 이 클 때

$$W \sim N\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right)$$

이다.

- 유의수준 α 에서 양측검정의 기각역은

$$W_{obs} \geq \frac{n(n+1)}{4} + z_{\alpha/2} \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

과

$$W_{obs} \leq \frac{n(n+1)}{4} - z_{\alpha/2} \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

이다.

Paired Randomized Experiments

- 단측검정의 경우에는

$$W_{obs} \geq \frac{n(n+1)}{4} + z_{\alpha} \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

또는

$$W_{obs} \leq \frac{n(n+1)}{4} - z_{\alpha} \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

이다.

- p 값은 가설의 종류에 따라 W_{obs} 의 W 의 영분포상 분위수를 파악함으로써 쉽게 계산해줄 수 있다.
- 귀무가설이 $H_0 : Y_{ij}(1) = Y_{ij}(0) + \tau$ 와 같이 둘 간에 상수 차이가 있도록 주어진다면, $D_i - \tau_0$ 에 대하여 wilcoxon의 검정통계량을 만들면 똑같이 검정해줄 수 있다.

Paired Randomized Experiments

- 한편 검정만이 아니라 추정 역시 수행해줄 수 있다.
- D_1, \dots, D_n 에 대하여 **왁시 평균**을 정의하자.

$$W_{st} = \frac{D_s + D_t}{2}, \quad 1 \leq s \leq t \leq n$$

- 실제 $Y_{ij}(1) - Y_{ij}(0)$ 이 어떠한 상수라고 할 때, 해당 상수 τ 의 $1 - \alpha$ 신뢰구간은

$$W_{(Q_W(\alpha/2))} \leq \tau < W_{(n(n+1)/2 - Q_W(\alpha/2) + 1)}$$

처럼 주어진다. Q_W 는 W 의 영분포상 분위수를 의미한다.

Paired Randomized Experiments

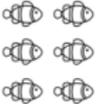
- 부호와 순위를 이용한 전통적인 비모수적 상황에서 자주 사용하는 점추정량은 **Hodges-Lehmann point estimate**(HL estimate)이다.
- 이는 구간추정량을 점점 줄여나감에 따라 최종적으로 남게 되는 값을 의미한다.
- 이 경우에는 $\hat{\tau}^{HL} = \text{median}(W_{st})$ 이다.
- 일반적으로 사용할 수 있는 \bar{D} 에 비해, 평균이 아닌 median을 사용하기 때문에 더 robust한 모습을 보인다.
- matching이 이루어진 경우, 우리는 근사적인 paired randomized experiment를 수행하는 것이나 마찬가지다.
- 따라서 매칭된 쌍에 대해 이러한 방법을 적용하면 로버스트한 처리효과의 추정량을 얻어줄 수 있다.

Weighting

Weighting

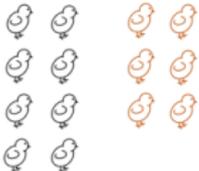
- matching은 잘 들여다보면 추정 과정에서 몇몇 개체는 버려버리고 (가중치를 0으로 두고), 몇몇 개체만 택함으로써(가중치를 1로 두고) 처리효과를 추정하였다.
- weighting은 이를 확장하여 각 개체가 추정량에 주는 weight를 변경하여 weighted average를 구함으로써 인과적인 효과를 추정한다.
- matching에 비해 직관적이지는 못하지만, weight를 주는 다양한 방식이 개발되어 있으며 그 효율이 matching보다 좋은 경우가 많다.
- 또한 몇몇 개체를 버려버리는 matching과 비교할 때, weighting은 모든 개체를 사용하기에 효율성이 좋다.

Weighting

	처리군	대조군
경제학부		
경제X		

- 공변량이 오직 하나만 존재하며, 그것이 binary인 경우를 고려해 보자.
- $P(Z = 1|\text{경제학부}) = 4/60$ 이 $P(Z = 1|\text{경제X}) = 6/14$ 와 다르다.
- 즉 propensity score가 X마다 다르다. 만약 X에 무관하게 propensity score를 동일하게 맞춰줄 수 있다면...? 자동으로 밸런싱이 되지 않을까?

Inverse Probability Weighting

	처리군	대조군
경제학부		
경제X		

- treatment 여부가 measured covariates와 independency를 유지할 수 있도록 주황색의 **pseudo-population**을 추가한다.
- $P(Z = 1 | \text{경제학부}) = 6/12$ 이 $P(Z = 1 | \text{경제X}) = 14/28$ 과 같다.

Inverse Probability Weighting

- pseudo-population을 추가한다는 것은 곧 해당 집단에 가중치를 추가함을 의미한다.
- 위에서는 binary case여서 단순히 반대 집단의 원래 개체 수를 더해 주면 되었다.
- 그 가중치는 특정 covariate 조건 하에서의 처리, 혹은 비처리 확률의 역수와 같다. 즉 처리군에 대해서는 $1/e(\mathbf{x})$ 배, 대조군에 대해서는 $1/(1 - e(\mathbf{x}))$ 배를 해준 것이나 마찬가지다.
- 이 아이디어를 적용하면 **inverse probability weighting**을 수행할 수 있다.

Inverse Probability Weighting

- SUTVA, Strong Ignorability, Binary Treatment를 가정한다.

$$\begin{aligned} E[Y(1)] &= E[E[Y(1)|X]] \\ &= E \left[\frac{E[Z|X]}{e(X)} E[Y(1)|X] \right] \\ &= E \left[E \left[\frac{ZY(1)}{e(X)} \mid X \right] \right] \\ &= E \left[\frac{ZY}{e(X)} \right] \end{aligned}$$

- 같은 이유로 $E[Y(0)] = E \left[\frac{(1-Z)Y}{1-e(X)} \right]$ 이다.

Inverse Probability Weighting

- 따라서 ATE의 **HT estimator**(Horvitz Thompson estimator)는

$$\hat{\tau}_{IPW} = \frac{1}{N} \sum_{i=1}^N \frac{Z_i Y_i}{\hat{e}_i} - \frac{1}{N} \sum_{i=1}^N \frac{(1 - Z_i) Y_i}{1 - \hat{e}_i}$$

으로 주어진다.

- 다만 만약 propensity score가 overlap을 만족스럽게 충족하지 못한다면, \hat{e}_i 혹은 $1 - \hat{e}_i$ 가 0에 가까워지며 과도한 inflating이 발생하고 variance가 커질 수 있다.
- 이를 해결하기 위해 SIPW/truncation 등의 방법이 개발되어 있으나, 생략한다.

Inverse Probability Weighting

- ATT를 추정하려면 weight만 조금 바꾸어 주면 된다.

-

$$\widehat{ATT} = \frac{1}{n_1} \sum_{i=1}^n \left(Z_i Y_i - \frac{\hat{e}_i (1 - Z_i) Y_i}{1 - \hat{e}_i} \right)$$

- weight의 비율은 항상 $(1 - e) : e$ 로 유지되어야 한다. 어떤 상수를 곱해서 어떤 표본 크기로 나누어 줄지를 다르게 만들면 추정하는 estimand가 달라질 뿐이다.

Doubly Robust Estimator

- IPW는 $\hat{\tau}$ 의 정확한 추정이 필요하다.
- 이는 propensity score에 대한 회귀모형을 통해 추정된다.
- outcome regression과 유사하게, IPW는 propensity score를 추정하는 회귀모형이 망가지면 큰 타격을 입는다.
- **doubly robust estimator**는 outcome에 대한 regression과 propensity score에 대한 regression 둘 중 하나만 맞아도 robust한 추정량을 낼 수 있도록 설계되었다.

Doubly Robust Estimator



$$\hat{\tau}_{DR} = \frac{1}{n} \sum_{i=1}^n \left[\hat{\mu}_1(X_i) + \frac{Z_i(Y_i - \hat{\mu}_1(X_i))}{\hat{e}(X_i)} \right] - \frac{1}{n} \sum_{i=1}^n \left[\hat{\mu}_0(X_i) + \frac{(1 - Z_i)(Y_i - \hat{\mu}_0(X_i))}{1 - \hat{e}(X_i)} \right]$$

- outcome regression이 맞다면 $\hat{\mu}_1$ 과 $\hat{\mu}_0$ 의 기댓값으로써 $E[Y(1)]$ 과 $E[Y(0)]$ 을 잘 추정해줄 수 있다. 뒷 항은 $Y_i - \hat{\mu}$ 를 하면서 기댓값이 0으로 간다.
- propensity score의 추정이 맞다면, 앞에 나와 있는 항을 안으로 그냥 넣어버릴 수 있다. \hat{e} 가 잘 추정되었기 때문이다. 그러면 그냥 IPW랑 똑같은 식이기 때문에 추정이 잘 된다.

Doubly Robust Estimator

- 문제 상황과 estimand에 따라
 - ① outcome regression model
 - ② propensity score regression model
 - ③ number of groups
 - ④ number and types of treatment
 - ⑤ estimand
 - ⑥ weight
 - ⑦ which estimator?(OR, IPW, or DR?)

가 다 다르다.

- 문제 상황에 맞춰 이걸 다 조절하는 건 (우리에겐) 사실상 불가능에 가깝다.
- 따라서 우리는 그냥 IPW와 DR의 개념을 이 정도에서만 알아보고자 한다. 원래는 예시도 상세히 들어서 보여드리려 했는데, 쉽지 않고 시간도 없어 포기했다.
- 자세한 것은 논문의 식을 그냥 받아들이도록 하자...