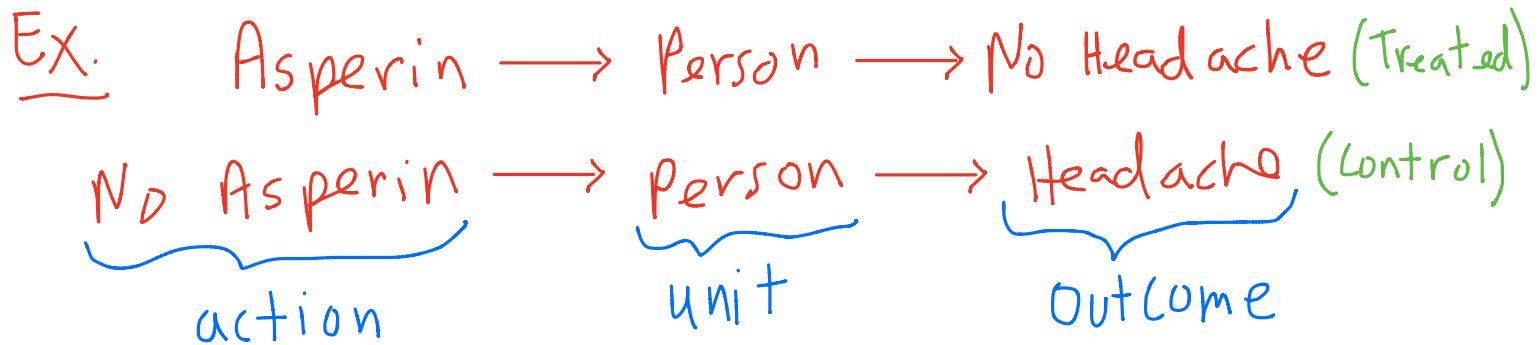


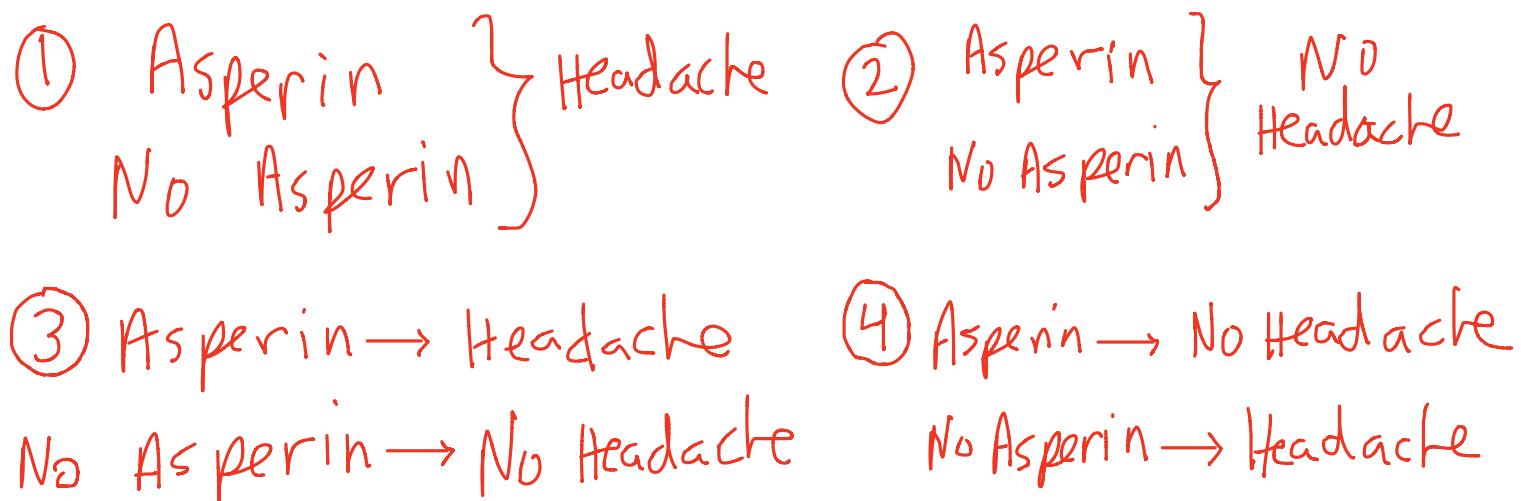
## Causal Inference Introduction

## Potential Outcomes Examples

- Framework of causal inference: action (treatment, intervention, or manipulation) is applied to a unit of interest and generates an outcome.
  - Potential outcomes are the value of the outcome under the control and treatment status



- Each (action, unit) pair is associated with an potential outcome. In the above example (aspirin, person) associated with "no headache" and (no aspirin, person) with "headache".
  - Causal inference is done by comparing potential outcomes for the same unit in the same post treatment period.
  - Definition of causal effect does not depend on observed outcome, but on potential outcome.
  - Causal statements may not be well defined if the potential outcomes cannot be realized or are unclear.
  - Let us consider all ((action, unit), potential outcome) possibilities below:



- Cases 1) and 2) above have no causal effect, but 3) and 4) do have a causal effect. For 4) there is a positive effect of aspirin, and for 4) there is a negative effect for taking aspirin.
  - "Fundamental problem of causal inference" is that only one of the potential outcomes can be realized and observed.
  - Definition of causal effect only requires one unit, but learning about causal effects from data will usually require comparing multiple units.
  - Estimation of causal effects will require comparing observed outcomes across groups.

## Stable Unit Treatment Value Assumption (SUTVA)

- Two common informal variation for causal inference are 1) compare outcomes for same unit over time under different actions, and 2) compare outcomes for different units at same time under different actions (control and treatment).

- Individuals may react differently to the same treatment over time, hence 1) above is not ideal. However variation described in 2) is better if different individuals all have same treatment choice (not vary in intensity) and are independent of each other.
- Assumption SUTVA has two components: 1) treatments applied to one unit doesn't effect the outcomes of other units and 2) treatment level is unique so that potential outcomes are well defined.
- Another interpretation of SUTVA is that 1) potential outcomes are independent of treatments status of other units, and 2) the treatment is defined the same for all units, that is the "intensity" of treatment is the same.
- Under the aspirin example, SUTVA would say 1) the effect of one unit taking aspirin does not impact the headache status of another unit, and 2) all aspirin tablets are of the same strength.

### Ex. Violates SUTVA

Unit 1 Potential Outcomes	$\begin{cases} \text{Aspirin} \\ \text{No Aspirin} \end{cases}$	Unit 2 Potential Outcomes	$\begin{cases} \text{Aspirin} \\ \text{Aspirin Old} \\ \text{No Aspirin} \end{cases}$
------------------------------	---	------------------------------	---

- In the above example unit 2 can take two types of treatment aspirin or old aspirin. But unit 1 only has access to the regular aspirin option. This violates the identical treatment condition.
- The two components of SUTVA are known as 1) "No Interference" or "No spillovers" and 2) "No hidden variation of treatments" or "Homogenous treatment".
- SUTVA is an exclusion restriction used to exclude varying possibilities and make causal inference feasible. SUTVA excludes general equilibrium effects or externalities.
- Exclusion restrictions generally cannot be verified from data but are based on prior knowledge on the subject matter of interest.
- For many economic settings SUTVA may only be feasible when units are modelled as groups such that there can be interactions within group, but independence across groups.
- SUTVA implied potential outcomes are well defined. Under SUTVA multiple units can be used to infer causality.

### Assignment mechanism (Potential Outcome Model)

- The potential outcome model can be used to study causality in a formal manner.

Let  $D_i \in \{0,1\}$  be a binary treatment indicator

$$Y_i^{\text{obs}} = D_i Y_{ii} + (1 - D_i) Y_{0i} = \begin{cases} Y_{ii}, & D_i = 1 \\ Y_{0i}, & D_i = 0 \end{cases}$$

realized Potential outcome

$$Y_i^{\text{miss}} = (1 - D_i) Y_{ii} + D_i Y_{0i} = \begin{cases} Y_{0i}, & D_i = 1 \\ Y_{ii}, & D_i = 0 \end{cases}$$

missing Potential outcome

- $Y_1$  is the potential outcome under treatment, and  $Y_0$  is the potential outcome under control.

- The model above shows that only one of the potential outcomes can be observed,  $Y_1$  for the treated units (with  $Y_0$  missing) and  $Y_0$  for the control units (with  $Y_1$  missing).
- Using the potential outcome framework we can formally state the "no interference" assumption in SUTVA as:

For all units  $i$ ,  $Y_{(D_1, D_2, \dots, D_i, \dots, D_n)i} = Y_{(D'_1, D'_2, \dots, D'_i, \dots, D'_n)i}$  if  $D_i = D'_i$

- That is the potential outcomes for a unit are independent of the treatment status of other units.

### Causal Parameters of Interest

- Unit level causal effects is comparing potential outcome under control and treatment.

Unit level Causal Effect =  $Y_{1i} - Y_{0i}$

- Causal effect definition does not depend on which treatment is observed.
- Population level causal effect is the expected outcome under treatment contrasted with expected outcome under control.
- Causal parameters are functions of only the potential outcomes. They are generally
- Cannot compute unit TE using data because of the fundamental problem of causal inference.
- Average treatment effect is mean unit level causal effect in population:

Average Treatment Effect =  $E[Y_{1i} - Y_{0i}]$

- Conditional average treatment effect is ATE conditioned on a pre-treatment covariate.

Conditional ATE =  $E[Y_{1i} - Y_{0i} | X_i = x]$

- ATT is the average treatment effect for those initially assigned the treatment

ATT =  $E[Y_{1i} - Y_{0i} | D_i = 1]$

- ATU is the average treatment effect for those assigned to the control group

ATU =  $E[Y_{1i} - Y_{0i} | D_i = 0]$

- ATE can be written as a function (weighted average) of the ATT and ATU

$$\begin{aligned} ATE &= \Pr(D_i=1) E(Y_{1i} - Y_{0i} | D_i=1) + \Pr(D_i=0) E(Y_{1i} - Y_{0i} | D_i=0) \\ &= \rho ATT + (1-\rho) ATU \end{aligned}$$

- Mean comparison of outcome across control and treatment is biased by selection and heterogeneous treatment effects.

## Identification of Causal Parameters

- The heterogeneity bias can be eliminated by assuming homogenous treatment effects.
  - Under constant treatment effect (causal effect =  $c$ ) implies 1)  $ATT = ATU$  and hence there is heterogeneity bias.

Constant effect:  $Y_{i1} = \rho + Y_{0i} \Rightarrow \text{ATT} = \text{ATU}$

- Note that under constant treatment effect the  $ATT = ATU = ATE$ .
  - Constant treatment effects allows us to reformulate the POM model into a standard regression:

$$Y_i = Y_{0i} + D_i (Y_{1i} - Y_{0i}) = Y_{0i} + \rho D_i$$

$$Y_{1i} = \rho + Y_{0i}$$

$$= E(Y_{0i}) + \rho D_i + (Y_{0i} - E(Y_{0i})) = \alpha + \rho D_i + \varepsilon_i$$

- The constant treatment effect assumption is not necessarily required for regression formulation:

$$\text{Hetero. effect: } \begin{cases} Y_{0i} = E(Y_{0i}) + u_{0i} \\ Y_{1i} = E(Y_{1i}) + u_{1i} \end{cases} \Rightarrow E(u_{0i}) = E(u_{1i}) = 0$$

$$\Rightarrow Y_i = Y_{0i} + D_i \text{ATE} + D_i (U_{1i} - U_{0i}) = E(Y_{0i}) + D_i \text{ATE} + U_{0i} + D_i (U_{1i} - Y_{0i})$$

$$\Rightarrow Y_i = \alpha + \beta D_i + \varepsilon_i$$

- Suppose a good teacher perfectly recognizes whether student's needs a tutor or not. Assignment of tutors to student's made by this teacher are not going to be independent of potential outcomes. Hence there will be selection bias.
  - If treatment is randomized to individuals, there is no selection bias and also no heterogeneity bias. Comparing across group means is causal and identifies ATE.

$$D_i \perp (Y_{1i}, Y_{0i}) \Rightarrow E[Y_{1i} - Y_{0i}] = E[Y_i | D_i = 1] - E[Y_i | D_i = 0]$$

- Rational choices made by individuals are usually dependant on potential outcomes and hence will violate the independence condition.
  - ATE is identified when treatment is randomly assigned to units. This is because now there is 1)

no selection bias and 2) no heterogeneity bias.

- Under constant treatment effect and randomly assigned treatment the ATE, ATT, and ATU are all identified and equal to each other.
- The primary assumption for identification for the causal parameters is the randomization of treatment assignment.
- Note that the ATE can be written as an integral over the potential outcome distributions.

$$ATE = \int_{Y_1} y F_{Y_1}(y) dy - \int_{Y_0} y F_{Y_0}(y) dy = \int_{Y_1} y dF_{Y_1}(y) - \int_{Y_0} y dF_{Y_0}(y)$$

- Under randomization of the treatment the marginal distributions of the potential outcomes are identified as follows.

$$D_i \perp (Y_{0i}, Y_{1i}) \Rightarrow \begin{cases} F_{Y_0}(y) = F(Y_i | D_i = 0) \\ F_{Y_1}(y) = F(Y_i | D_i = 1) \end{cases} \text{ are identified}$$

- Hence the ATE is identified under randomization of the treatment. Less well known (but clear from above) that the Quantile Treatment Effect (QTE) is also identified:

$$D_i \perp (Y_{0i}, Y_{1i}) \Rightarrow (F_{Y_1}, F_{Y_0}) \text{ identified} \Rightarrow (F_{Y_1}^{-1}, F_{Y_0}^{-1}) \text{ identified}$$

$$\Rightarrow QTE(\theta) = F_{Y_1}^{-1}(\theta) - F_{Y_0}^{-1}(\theta) = Q_{Y_1}(\theta) - Q_{Y_0}(\theta), \quad \begin{matrix} \theta \in (0,1) \\ \text{quantile} \end{matrix}$$

- Recall that the quantile function is the smallest value of  $y$  such that  $\Pr(Y < y) \geq \theta$ , where  $\theta$  is the quantile of interest.

$$Q_Y(\theta) = \inf \{y \in Y : F(y) \geq \theta\} \text{ where } \theta \in (0,1)$$

- We can also similarly show that the conditional distributions  $F(Y_0|D=0)$  and  $F(Y_1|D=1)$  are also identified under randomized treatment. This is another way of viewing the identification of the ATT and ATU in a experimental environment.

## Conclusion

- The potential outcome model can be used to study causality. Under binary treatment each unit has a potential outcome  $Y_1$  under treatment and  $Y_0$  under control.
- Note  $(Y_1, Y_0)$  do not depend on the units treatment status in the observed data. The treatment status just indicates which potential outcome is observed,  $(Y_1, Y_0)$  always exist for each unit.
- Causal parameters are functions of only potential outcomes. Common causal parameters of interest are ATE, ATT, and ATU.
- Under randomized treatment, the QTE and the common causal parameters are identified.
- The ATE can be estimated using a simple linear regression by regressing the outcome on the treatment (still need treatment to be randomized across units). This is equivalent to comparing mean outcomes across the control and treatment group.