Potential Outcomes Causal Model

John Stuart Mill (A System of Logic, Ratiocinative and Inductive)

↳ 5 methods for inferring causation. (Canons of Empirical Science)

1. Method of agreement → Factor necessary for effect
2. ** Method of difference → Comparison of counterfactuals
3. Joint Method. → agreement + difference.
4. Method of concomitant variation (variation in one causes variation in the other)
5. Method of residues. (Through process of elimination)

Correlation ≠ Causation

Regression alone is insufficient.

RANDOMIZATION

Splawa-Neyman

Potential outcomes notation.

↓ (Social sciences)

D. Rubin

Roland A. Fisher

Randomization for causal inference

Comparison of 2 states of the world

Actual

Counterfactual.

Universe 1 → takes aspirin

Universe 2 → doesn't take aspirin.

Effect of aspirin = Difference in reported outcomes of Universe 1 & 2.

Counterfactual data → missing data.

"Potential outcomes exist ex-ante as a set of possibilities, but once a decision is made, all but one outcome disappears"

$$D_i = \begin{cases} 1, & \text{unit } i \text{ receives treatment} \\ 0, & \text{unit } i \text{ doesn't receive treatment} \end{cases}$$

Potential outcomes → Y_i^1 → receives treatment

Y_i^0 → doesn't

↳ control state

Y_i } actual outcome.

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$$

causal effect $\delta_i = Y_i^1 - Y_i^0$

Average Treatment Effect

$$ATE = E[\delta_i]$$
$$= E[Y_i^1 - Y_i^0]$$
$$= E[Y_i^1] - E[Y_i^0]$$

ATE is inherently unknowable. It cannot be calculated, only estimated.
$\because$ Only 1 potential outcome can be known, the other is imagined and hence missing.

Average Treatment Effect for Treated Group

Population mean treatment effect for group of units that had been assigned to treatment.

ATT is unknowable. (but estimable)

$$ATT = E[\delta_i | D_i = 1]$$
$$= E[Y_i^1 - Y_i^0 | D_i = 1]$$
$$= E[Y_i^1 | D_i = 1] - E[Y_i^0 | D_i = 1]$$

Average treatment for untreated group.

$$ATU = E[\delta_i \mid D_i = 0]$$
$$= E[Y_i^1 - Y_i^0 \mid D_i = 0]$$
$$= E[Y_i^1 \mid D_i = 0] - E[Y_i^0 \mid D_i = 0]$$

In observational settings, given heterogenous treatment effects,

$$ATT \neq ATU \text{ (Probably)}$$

[Heterogenous treatment effect
– Non random variability in the direction or magnitude of treatment effect]

Concrete Example — 4.1.3. Scott Cunningham's Book.

Simple difference in outcomes = ATE + Selection bias + heterogeneous treatment effect bias

If we assume no heterogeneity, then $ATU = ATT$

$\Rightarrow SDO = ATE +$ selection bias

Independence Assumption

$(Y^1, Y^0) \perp D$

treatment assignment has nothing to do with gains of treatment

$\left.\begin{array}{l}\text{Selection bias} \\ \quad \& \\ \text{heterogenous} \\ \quad\text{treatment effect}\end{array}\right\}$ eliminated by randomization of treatment assignment.

If treatment is independent of potential outcomes, $SDO = ATE$

But, this is not realistic. – People make choices based on perceived increase in gain.

$\Rightarrow$ selection bias cannot be eliminated.

SUTVA - Stable Unit Treatment Value Assumption.