Problem Set: Introduction

Problem 1 (Descriptive, Predictive, or Causal?). *For each of the following research questions, classify whether it is:* descriptive, predictive, *or* causal. *Briefly justify.*

(a) *What fraction of Korean firms pay dividends in 2024?*

(b) *What will be Samsung Electronics' stock return next quarter?*

(c) *How does increasing margin requirements affect stock market volatility?*

(d) *Do firms with female CEOs have higher return on assets?*

Problem 2 (Potential Outcomes Framework). *Suppose we are studying the effect of attending college ($D_i = 1$) on annual income Y_i. For an individual i:*

$$Y_i(1) = \textit{income if attends college}, \quad Y_i(0) = \textit{income if does not attend.}$$

(a) *Define the* ***Individual Treatment Effect (ITE)****.*

(b) *Define the* ***Average Treatment Effect (ATE)****.*

(c) *Define the* ***Average Treatment Effect on the Treated (ATT)****.*

(d) *Define the* ***Average Treatment Effect on the Untreated (ATU)****.*

(e) *Explain why the Fundamental Problem of Causal Inference prevents us from observing the ITE directly.*

Problem 3 (Estimand vs. Parameter)**.** *Suppose the true population effect of college on income is*

$$\tau^* = E[Y_i(1) - Y_i(0)].$$

(a) Is τ^* a parameter or an estimand? Why?

(b) Consider the difference in conditional expectations:

$$\tau = E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0].$$

Is τ a parameter or an estimand?

(c) Suppose you compute

$$\hat{\tau}_N = \frac{1}{N_1} \sum_{i:D_i=1} Y_i - \frac{1}{N_0} \sum_{i:D_i=0} Y_i$$

from a random sample. What is $\hat{\tau}_N$ called?

(d) Explain the logical relationship between parameters, estimands, and estimators.

Problem 4 (Randomization and Identification)**.** *Suppose we could randomly assign half of the population to attend college.*

(a) Discuss how randomization implies independence between treatment assignment D_i and potential outcomes $(Y_i(0), Y_i(1))$.

(b) Under randomization and SUTVA, show that

$$ATE = E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0].$$

(c) Explain intuitively why randomization solves the selection bias problem.

Problem 5 (ATE, ATT, and ATU)**.** *Consider a population of 10 individuals. For each i, the potential outcomes $(Y_i(1), Y_i(0))$ are listed below. Suppose the observed treatment D_i equals 1 for $i = 1, \ldots, 5$ and equals 0 for $i = 6, \ldots, 10$. Define $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$ as the observed outcome.*

i	1	2	3	4	5	6	7	8	9	10
$Y_i(0)$	3	2	4	5	6	2	3	1	4	5
$Y_i(1)$	5	5	5	7	6	6	4	4	6	6
D_i	1	1	1	1	1	0	0	0	0	0

(a) Compute the ***true*** *ATE, ATT, and ATU using the potential outcomes table.*

(b) Compute the difference in observed means $\tau \equiv E[Y \mid D = 1] - E[Y \mid D = 0]$ for this population.

(c) Show the ATT decomposition*:*

$$E[Y \mid D = 1] - E[Y \mid D = 0] \;=\; ATT \;+\; \big(E[Y(0) \mid D = 1] - E[Y(0) \mid D = 0]\big).$$

Interpret the second term.