Causal Inference HW4

October 2026

Problem 1

Suppose that you are interested in the effect of a certain drug on the lifetime of a given population, and now you have a random sample of size n from the population of interest. Answer the following questions that are designed to help you get familiar with potential outcomes. Try to keep your answers brief and your language precise. Throughout the problem, assume that the Stable Unit Treatment Value Assumption (SUTVA) holds.

1. Contrast the meaning of $Y_i(0)$ with the meaning of Y_i^{obs}.

2. Explain the notation $E\{Y_i(0) \mid W_i = 1\}$, where W_i is a binary variable that gives the treatment status for subject i, 1 if treated (taking a certain amount of the drug), 0 if control (taking a certain amount of placebo).

3. Contrast the meaning of $E\{Y_i(0)\}$ with the meaning of $E\{Y_i^{\text{obs}} \mid W_i = 0\}$.

4. Contrast the meaning of $E\{Y_i(0) \mid W_i = 1\}$ with the meaning of $E\{Y_i(0) \mid W_i = 0\}$.

5. Which of the following quantities (that you explained in parts (d) through (f)) can be identified from observed information? Do not make any additional assumptions about the distributions of Y or W, except SUTVA, and that there is at least one observation with $W_i = 1$ and at least one with $W_i = 0$ in the observed data: $E\{Y_i(0) \mid W_i = 1\}$, $E\{Y_i(0)\}$, $E\{Y_i^{\text{obs}} \mid W_i = 0\}$, and $E\{Y_i(0) \mid W_i = 0\}$.

6. Now, assume that n is even, and we randomly assign $\frac{n}{2}$ units in this sample to the treatment group. Please justify whether $\vec{W}$ satisfies individualistic assumption and unconfounded assumption in this case. Furthermore, which of the above quantities in (5) can be identified from the observed data?

Problem 2

This is a classic kidney stone example. The treatment and outcome data can be summarized in the following two-by-two table:

	success	failure
an open surgical procedure	273	77
a small puncture procedure	289	61

Table 1: Summary Data

1. If you got a kidney stone, which procedure would you prefer?

2. Suppose then you get more information. For patients with smaller stones, the treatment and outcome data can be summarized in the following two-by-two table:

	success	failure
an open surgical procedure	81	6
a small puncture procedure	234	36

Table 2: Summary data for smaller stones

For patients with larger stones, the treatment and outcome data can be summarized in the following two-by-two table:

	success	failure
an open surgical procedure	192	71
a small puncture procedure	55	25

Table 3: Summary data for larger stones

Again, if you got a kidney stone, which procedure would you prefer? Why? Further, please explain why the paradox exists.

3. Denote $W_i = 1$ if the patient uses an open surgical procedure and $W_i = 0$ if the patient uses a small puncture procedure. Denote Y_i as success or not (1 or 0). Denote X_i as the stone being large or not (1 or 0).

   Is it possible for us to observe the three data tables above under the assumption that W_i is independent

of $(Y_i(0), Y_i(1))$? If not, please give your reason. If so, please provide a reasonable estimation using the Neyman approach. Further, is the result conflict with your result in (2)? Please comment on your results.

Problem 3

Consider the causal question that you chose in problem 5 of homework 1 again.

- For the data that you will collect or the existing data that you can find, is it from a random experiment or an observational study?
- Is it reasonable to assume the 3 assumptions for the assignment mechanism?
- Why does it satisfy or not satisfy the unconfoundedness assumption? Please give your reason.