**Stat 490, Spring 2025** Due: Tuesday February 18, 11:59 PM

Homework 2

*Tirthankar Dasgupta*

**Collaboration policy:** you are free to discuss the problems with others, though it is strongly recommended that you try the problems on your own first. Copying is not allowed, and write-ups must be your own explanations in your own words.

**Problem 1**

Three $10 \times 2$ tables of potential outcomes for ten units and two treatments are shown in Tables 1 - 3.

Table 1:

| Unit | Treatment | |
|---|---|---|
| | 0 | 1 |
| 1 | 10.04 | 14.20 |
| 2 | 9.63 | 13.51 |
| 3 | 7.26 | 11.52 |
| 4 | 8.80 | 13.97 |
| 5 | 10.59 | 13.48 |
| 6 | 10.78 | 12.18 |
| 7 | 7.58 | 10.09 |
| 8 | 9.27 | 11.61 |
| 9 | 6.75 | 13.85 |
| 10 | 9.49 | 12.97 |
| Avg | | |
| Med | | |
| Var | | |
| Ran | | |

Table 2:

| Unit | Treatment | |
|---|---|---|
| | 0 | 1 |
| 1 | 2.77 | 4.77 |
| 2 | 2.26 | 4.26 |
| 3 | 0.69 | 2.69 |
| 4 | 1.49 | 3.49 |
| 5 | 3.65 | 5.65 |
| 6 | 4.01 | 6.01 |
| 7 | 0.81 | 2.81 |
| 8 | 1.89 | 3.89 |
| 9 | 0.53 | 2.53 |
| 10 | 2.10 | 4.10 |
| Avg | | |
| Med | | |
| Var | | |
| Ran | | |

Table 3:

| Unit | Treatment | |
|---|---|---|
| | 0 | 1 |
| 1 | 2.77 | 8.18 |
| 2 | 2.26 | 5.79 |
| 3 | 0.69 | 2.14 |
| 4 | 1.49 | 7.30 |
| 5 | 3.65 | 5.71 |
| 6 | 4.01 | 2.97 |
| 7 | 0.81 | 1.05 |
| 8 | 1.89 | 2.24 |
| 9 | 0.53 | 6.86 |
| 10 | 2.10 | 4.41 |
| Avg | | |
| Med | | |
| Var | | |
| Ran | | |

(a) From each table, calculate

(i) The missing summaries (average, median, variance and range) of each column.

(ii) The unit-level causal effects $\tau_1, \ldots, \tau_{10}$.

(iii) The average treatment effect $\overline{\tau} = \overline{Y}(1) - \overline{Y}(0)$.

(iv) Two typical causal effects: median of unit-level causal effects: $\underset{i}{\text{Med}}(\tau_i)$ and the difference of the medians of the potential outcomes under treatments 0 and 1: $\underset{i}{\text{Med}}[Y_i(1)] - \underset{i}{\text{Med}}[Y_i(0)]$. Are they equal in any of the three cases? If so, why or why not?

(b) Does additivity hold in any of the three tables? If so, in which?

(c) For Table 1, using a Bernoulli treatment assignment mechanism (i.e., treatment assignment for each unit is determined by independent tosses of a fair coin), obtain a realization of the assignment vector $\mathbf{W} = (W_, \ldots, W_{10})^{\mathrm{T}}$ and thus show the observed and missing outcomes for each unit. Explain how you obtained the assignment (showing appropriate R codes).

(d) For Table 2, using a completely randomized treatment assignment mechanism to assign treatments 0 and 1 to five units each, obtain a realization of the assignment vector $\mathbf{W} = (W_, \ldots, W_{10})^{\mathrm{T}}$ and thus show the observed and missing outcomes for each unit. Explain how you obtained the assignment (showing appropriate R codes).

**Problem 2**

If a completely randomized treatment assignment mechanism is used to assign treatments 0 and 1 to four and six units respectively, and the observed assignment vector $\mathbf{w}$ is $(1, 1, 0, 0, 1, 1, 1, 0, 1, 0)^{\mathrm{T}}$, create a table of observed and missing outcomes from the potential outcomes in Table 3 in Problem 1. Estimate the average causal effect $\overline{\tau}$ from the observed data. What is the TRUE sampling variance of this estimator? What will be the standard estimate of this sampling variance from the observed outcomes?

**Problem 3**

Lyle et al. (1987) reported the results of a randomized experiment on twenty one men conducted to investigate the effect of calcium on blood pressure in African American men. The treatment (a calcium supplement) was assigned to ten men selected at random from the twenty one, and the remaining eleven men instead received a placebo. All the men had their systolic blood pressures tested before and after the 12-week period. The data from the experiment are given in Table 4, and is also available at

`http://lib.stat.cmu.edu/DASL/Datafiles/Calcium.html.`

(a) Which of the following causal estimands do you care about more? Justify your answer.

  (i) The difference between average post-treatment BP for the treatment and the control group.

  (ii) The difference between average BP reduction for the treatment and the control group.

(b) Propose two unbiased estimators for estimand (i) defined in part (a) and explain why they are unbiased. Which of these two estimators is more efficient (i.e., should have smaller sampling variance) and why?

Table 4: Effect of Calcium on BP

| Treatment | BP before ($X$) | BP after ($y$) | Observed (negative) reduction ($y - X$) |
|---|---|---|---|
| Calcium | 107 | 100 | -7 |
| Calcium | 110 | 114 | 4 |
| Calcium | 123 | 105 | -18 |
| Calcium | 129 | 112 | -17 |
| Calcium | 112 | 115 | 3 |
| Calcium | 111 | 116 | 5 |
| Calcium | 107 | 106 | -1 |
| Calcium | 112 | 102 | -10 |
| Calcium | 136 | 125 | -11 |
| Calcium | 102 | 104 | 2 |
| Placebo | 123 | 124 | 1 |
| Placebo | 109 | 97 | -12 |
| Placebo | 112 | 113 | 1 |
| Placebo | 102 | 105 | 3 |
| Placebo | 98 | 95 | -3 |
| Placebo | 114 | 119 | 5 |
| Placebo | 119 | 114 | -5 |
| Placebo | 112 | 114 | 2 |
| Placebo | 110 | 121 | 11 |
| Placebo | 117 | 118 | 1 |
| Placebo | 130 | 133 | 3 |

(c) Conduct a Fisher randomization test to compare the effects of the treatment on BP. Explain your choice of test statistic. What is the smallest p-value that can be obtained from such a test in this experiment? Show your R codes and outputs.

(d) Obtain approximate 95% confidence intervals for estimands (i) and (ii).