Causal Inference HW1

September 2026

Note: There is some computation required for this assignment. You can use any software or programming language that you like. Codes should be submitted along with your results.

Problem 1 (Fisher's test)

The table below gives data from a randomized experiment carried out around 1970 in Southern Florida to evaluate cloud seeding (a technique that involves injecting clouds with a chemical to increase rainfall). On each of 52 days that were suitable for cloud seeding, a random mechanism was used to decide whether to seed or not. A plane flew through the clouds each suitable day (both seeding and non-seeding days). The experimenter set up the plane to inject or not depending on the random outcome. The plane's pilot did not know whether the seeding mechanism was loaded to inject or not. Precipitation was measured as total rain volume (in acre-feet) resulting from the clouds in the 24 hours following the plane's run. The data:

Seeded days ($n = 26$): 2745.6, 274.7, 115.3, 1697.1, 274.7, 92.4, 1656.4, 255.0, 40.6, 978.0, 242.5, 32.7, 703.4, 200.7, 31.4, 489.1, 198.6, 17.5, 430.0, 129.6, 7.7, 334.1, 119.0, 4.1, 302.8, 118.3;

Unseeded days ($n = 26$): 1202.6, 87.0, 26.0, 830.1, 81.2, 24.4, 372.4, 68.5, 21.4, 345.5, 47.3, 17.3, 321.2, 41.1, 11.5, 244.3, 36.6, 4.9, 163.0, 29.0, 4.9, 147.8, 28.6, 1.0, 95.0, 26.3.

a. Why did the plane fly through the cloud even on non-seeding days? Why were the pilots not informed as to whether the cloud seeding mechanism was set to inject or not?

b. Suppose we want to test Fisher's sharp null hypothesis of no treatment effect. Use the difference in sample means as a test statistic and use simulation to approximate the p-value. (Generate 1000 simulated random assignments under the null hypothesis, compute the test statistic for each, and determine the proportion of trials with a larger outcome than the observed statistic.)

c. How accurate is the simulation approximation? Explain by referring to the expected variablity in a sample proportion of this type.

d. Report the p-value you would obtain using a traditional t-test to compare the two group means. Is this related to the p-value in (b)? Explain.

e. The test statistic used in (b) is sensitive to a change in the mean amount of rainfall. Suppose that you expected the treatment to increase variability of rainfall but not the mean. Develop a test statistic that would be sensitive to this alternative hypothesis and obtain an estimate of the p-value using this test statistic.

f. The wide range of data values suggest that the impact of cloud seeding may be multiplicative rather than additive. Repeat the test once more using the difference in the mean of the log rainfall as a test statistic. What p-value do you obtain?

g. What conclusion do you draw about these data based on the analysis above?

Problem 2 (Another test statistic for Fisher's test)

Another test statistic that can be used in Fisher's test is based on a statistic called Wilcoxon statistic. The concrete form of the Wilcoxon statistic is:

$$T_W = \sum_{i=1}^{n} W_i R_i,$$

where R_i is the number of elements in the set $\{j : Y_j \leq Y_i\}$ and is called the rank of Y_i in the pooled sample $\mathbf{Y}$. Please answer the following questions:

a. Intuitively, why T_W can serve as a statistic for testing the hypothesis of no treatment effect?

b. Prove that $E(T_W) = n_1(n+1)/2$ and $\text{var}(T_W) = n_1 n_0 (n+1)/12$.

c. Use $T_W - n_1(n+1)/2$ to conduct Fisher's test using data in Problem 1.

d. Under Fisher's sharp null hypothesis, we have the finite population central limit theorem:

$$\frac{T_W - E(T_W)}{\sqrt{\text{var}(T_W)}} \xrightarrow{d} N(0,1).$$

Can you use this result to test Fisher's sharp null hypothesis in Problem 1?

e. Compare the Wilcoxon statistic with the difference in sample means statistic, what advantages do they

have respectively? Hint: consider extreme observations.

Problem 3

Suppose that you are conducting an experiment. Your goal is to access whether the languages used in questions would influence the correctness of answers in math quiz. Suppose you randomly choose 8 volunteers for participation. They sit at a table, and each person wears a single earbud on his/her right ear. $D_i, i = 1, \cdots, 8$ gives the treatment status for each person, where $D_i \in \{0, 1\}$, 1 represents this person's single earbud provides questions in English, and 0 represents this person's single earbud provides questions in Chinese. Denote Y as the the error rate of responses. Without single earbud on left ears, each person might be influenced by the language received by his/her left side person, as indicated by the arrows in Figure 1. So you can define the potential outcome as $Y_i(D_i, D_{i-1})$ for the ith person.

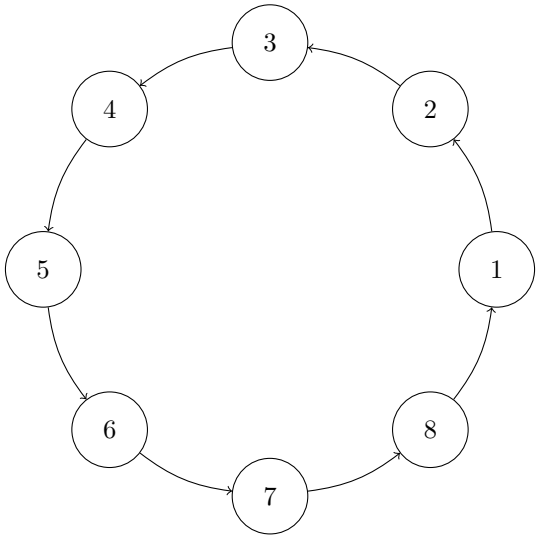


Figure 1: An experiment to access whether the languages used in questions would influence the correctness of answers in math quiz.

- Using Figure 1, translate Table 1 into a table formatted like Table 2, where the latter columns give $Y_i(D_i, D_{i-1})$ for the combinations of possible values for D_i and D_{i-1}. (And for $i = 1$, $Y_1(D_1, D_8)$, per Figure 1). Where you can, fill the cells of your table with values from Table 1; leave blank the cells representing unobserved potential outcomes.

- Formally express each of the estimands described below using potential outcomes; propose appropriate

Table 1: Observed Data: Treatment and Error Rate

Unit	D_i	Y_i
1	1	4%
2	1	2%
3	0	8%
4	1	9%
5	0	12%
6	0	13%
7	0	4%
8	1	1%

Table 2: Observed and Unobserved Potential Outcomes

Unit	D_i	D_{i-1}	$Y_i(1,1)$	$Y_i(1,0)$	$Y_i(0,1)$	$Y_i(0,0)$
1						
⋮						
8						

estimators for each; and finally estimate them with the help of your new table.

(i) The average causal effect (average treatment effect), conditional on a left neighbor taking treatment.

(ii) The average causal effect (average treatment effect), conditional on a left neighbor taking control.

(iii) The magnitude of effect modification due to assignment of treatment to a left neighboring unit.

Problem 4

With potential outcomes $\{Y_i(1), Y_i(0)\}_{i=1}^n$ for n units under the treatment and control, the difference in means equals the mean of the individual tratment effects: $\bar{Y}(1) - \bar{Y}(0) = n^{-1} \sum_{i=1}^n \{Y_i(1) - Y_i(0)\}$. Therefore, the average treatment effect is a *linear* causal estimand in the sense that the difference in average potential outcomes equals the average of the differences in individual potential outcomes.

Other estimands may not be linear. For instance, we can define the median treatment effect as

$$\delta_1 = \text{median}\{Y_i(1)\}_{i=1}^n - \text{median}\{Y_i(0)\}_{i=1}^n,$$

which is, in general, different from the median of the individual treatment effect

$$\delta_2 = \text{median}\{Y_i(1) - Y_i(0)\}_{i=1}^n.$$

1. Give numerical examples that have $\delta_1 = \delta_2$, $\delta_1 > \delta_2$, and $\delta_1 < \delta_2$, respectively.

2. Which estimand makes more sense, δ_1 or δ_2? Why? Use examples to justify your conclusion. If you feel that both δ_1 and δ_2 can make sense in different applications, you can also give examples to justify both estimands.

Problem 5

Have you ever encountered some questions that are related to causal inference in your academic experience or daily life? Or, what causal question would you like to investigate? Please think it over and answer the following questions. Try to keep your answers brief and your language precise.

- Would you please briefly describe the question and point out the corresponding Unit, Actions (e.g., treatment, control), Potential Outcomes and Covariates (if applicable)?
- Is it reasonable to assume SUTVA here?
- To solve the question, how could you collect the data? Please briefly describe your plan.
- 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?