STAT 3005 – Nonparametric Statistics

ASSIGNMENT 5

Due: 22 November (Friday) at 1800

Fall 2024

Exercise 5.1 (★★☆ — Causal inference on cloud seeding (100%)). "Cloud seeding is a weather modification technique that improves a cloud's ability to produce rain or snow by artificially adding condensation nuclei to the atmosphere, providing a base for snowflakes or raindrops to form." (*Adopted from Desert Research Institute*)

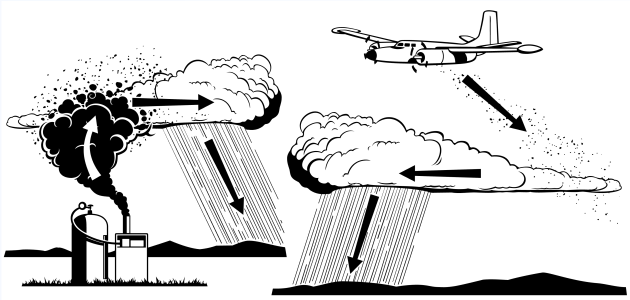

Photo source: https://en.wikipedia.org/wiki/Cloud_seeding.

The aim of this exercise is to analyze the effectiveness of cloud seeding for producing rainfall. The modified dataset (`clouds.csv`) is obtained in a series of weather modification experiments conducted in south Florida from 1968 to 1972; see Simpson, Alsen and Eden (1975). In the experiments, $n = 40$ independent clouds were considered. They were randomly selected to seed with silver nitrate with probability 30%. Let

$$A_i = \mathbb{1}\,(\text{the } i\text{th cloud was seeded with silver nitrate}),$$
$$X_i = \text{the observed amount of rainfall (in acre-feet) of the } i\text{th cloud},$$

for $i = 1, \ldots, n$. We also denote the **potential outcomes** by the following RVs:

$$X_i(0) = \text{the amount of rainfall (in acre-feet) of the } i\text{th cloud if it was not seeded},$$
$$X_i(1) = \text{the amount of rainfall (in acre-feet) of the } i\text{th cloud if it was seeded},$$

for each i. We assume that $X_i(a) = \mu_a + \varepsilon_{ia}$ for $a = 0, 1$ and $i = 1, \ldots, n$, where ε_{i0} and ε_{i1} are of mean zero and identically distributed. Note, however, that the n pairs of RVs $(\varepsilon_{i0}, \varepsilon_{i1})$'s may not be identically distributed. The treatment effect is defined as $\theta = \mu_1 - \mu_0$. The major goal is to test whether cloud seeding leads to more rainfall on average.

1. Estimation. Consider two estimators:
$$\bar{\theta} = \frac{1}{n}\sum_{i=1}^{n}\{X_i(1) - X_i(0)\} \qquad \text{and} \qquad \widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{A_i X_i}{0.3} - \frac{(1-A_i)X_i}{0.7}\right\}.$$
   (a) (10%) Express X_i in terms of A_i, $X_i(0)$ and $X_i(1)$.
   (b) (10%) Prove that $\mathsf{E}(A_i X_i) = 0.3\mu_1$ and $\mathsf{E}\{(1 - A_i)X_i\} = 0.7\mu_0$.
   (c) (10%) Explain in layman's terms why $\bar{\theta}$ is not a practical estimator. (Use $\lesssim 50$ words.)
   (d) (10%) Prove that $\widehat{\theta}$ is an unbiased estimator for θ.
2. Hypothesis testing.
   (a) (10%) Write down the null hypothesis H_0 and the alternative hypothesis H_1.
   (b) (10%) Explain why rank sum test is not the best choice for this problem. (Use $\lesssim 50$ words.)
   (c) (10%) Describe a permutation procedure under H_0. (Use $\lesssim 50$ words.)
   (d) (10%) Propose an appropriate test statistic $T = T(X_{1:n}, A_{1:n})$ for performing a permutation test.
   (e) (10%) Compute the permutation p-value. What is your conclusion? (Use $\lesssim 50$ words.)
   (f) (10%) Plot the permutation distribution of T and draw a vertical line to indicate the observed T.
3. (✄ bonus) In this part, we instead suspect that the amount of rainfall with seeding is at least 5 times more than that without seeding. Perform an analysis similar to parts 2(c)–(f).

Hints: See Remark 5.1. Don't read the hints unless you have no ideas and have tried for more than 15 mins.

Remark 5.1 (Hints for Exercise 5.1).

1. (a) $X_i = A_i X_i(1) + \cdots$.
   (b) Note that (i) $X_i A_i = X_i(1) A_i$, (ii) $A_i \overset{\text{IID}}{\sim} \text{Bern}(\cdots)$, and (iii) $A_i \perp\!\!\!\perp (X_i(0), X_i(1))$.
   (c) What are observable?
   (d) Use part (1b). Recall that $\widehat{\theta}$ is said to be unbiased for θ if $\mathsf{E}(\widehat{\theta}) = \theta$.
2. (a) H_1 is a statement that you suspect true.
   (b) What are the assumptions of rank sum test?
   (c) An appropriate permutation W is chosen so that $W(\text{data})$ and data are identically distributed under H_0.
   (d) T can be un-standardized (by standard derivation). It is the beauty of permutation test.
   (e) Is H_1 one-sided or two-sided?
   (f) See Example 6.9 of the lecture note.

Instructions: Please follow the instructions stated on the last page of assignment 1.