STAT3005 Assignment 5
Lam Hoi Chun (1155192755)

Exercise 5.1(a)

The observed amount of rainfall X_i of the ith cloud can be expressed in:

$$X_i = A_i X_i(1) + (1 - A_i)X_i(0)$$

Exercise 5.1(b)

$$\begin{aligned}
\mathrm{E}(A_i X_i) &= \mathrm{E}\big\{A_i\big[A_i X_i(1) + (1 - A_i)X_i(0)\big]\big\} \\
&= \mathrm{E}\big[A_i^2 X_i(1) + A_i(1 - A_i)X_i(0)\big] \\
&= \mathrm{E}\big[A_i X_i(1) + 0\big] \\
&= \mathrm{E}\big[A_i X_i(1)\big] \\
&= \mathrm{E}\big[A_i\big] \cdot \mathrm{E}\big[X_i(1)\big] \\
&= 0.3\mu_1
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\big[(1 - A_i)X_i\big] &= \mathrm{E}\big[X_i - A_i X_i\big] \\
&= \mathrm{E}\big[X_i\big] - \mathrm{E}\big[A_i X_i\big] \\
&= \mathrm{E}\big[A_i X_i(1) + (1 - A_i)X_i(0)\big] - 0.3\mu_1 \\
&= \mathrm{E}\big[A_i X_i(1)\big] + \mathrm{E}\big[(1 - A_i)X_i(0)\big] - 0.3\mu_1 \\
&= \mathrm{E}\big[A_i\big] \cdot \mathrm{E}\big[X_i(1)\big] + \mathrm{E}\big[1 - A_i\big] \cdot \mathrm{E}\big[X_i(0)\big] - 0.3\mu_1 \\
&= 0.3\mu_1 + (1 - 0.3)\mu_0 - 0.3\mu_1 \\
&= 0.7\mu_1
\end{aligned}$$

Exercise 5.1(c)

$\overline{\theta}$ is not practical since it needs to observe both potential outcomes, which is for each cloud, how much rain it would produce in both situations: seeded or not seeded. They cannot be measured simultaneously in reality.

Exercise 5.1(d)

$$\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{A_i X_i}{0.3} - \frac{(1 - A_i)X_i}{0.7}\right\}$$

$$\begin{aligned}
\mathrm{E}(\widehat{\theta}) &= \mathrm{E}\left[\frac{A_i X_i}{0.3}\right] - \mathrm{E}\left[\frac{(1 - A_i)X_i}{0.7}\right] \\
&= \frac{0.3\mu_1}{0.3} - \frac{0.7\mu_0}{0.7} \\
&= \mu_1 - \mu_0 \\
&= \theta \\
&\Rightarrow \widehat{\theta} \text{ is unbiased.}
\end{aligned}$$

Exercise 5.2(a)

The null hypothesis and the alternative hypothesis are:

$$H_0 : \theta = 0 \quad \text{and} \quad H_1 : \theta > 0$$

Exercise 5.2(b)

Rank sum test assumes that the samples are identically distributed under the H_0 null hypothesis and detects differences in location. However, our data: X_0, X_1 have unequal noises, thus violating the assumptions.

Exercise 5.2(c)

Under H_0, cloud seeding has no effect on rainfall, i.e. treatment $A_i \perp\!\!\!\perp$ the observed outcomes X_i. We apply order permutation test by randomly shuffling the treatment labels A_i among the observed X_i. A null distribution of the test statistic can be generated, reflecting the randomness under H_0. Reject $H_0 \iff$ permutation p-value < 0.05.

Exercise 5.2(d)

The proposed test statistic is

$$\sum_{i=1}^{n} \left[\frac{A_i X_i}{p_1} - \frac{(1 - A_i) X_i}{p_0} \right], \text{where}$$

- $p_1 = \mathrm{P}(\text{a seeded cloud is randomly chosen}) = 0.3$;
- $p_0 = \mathrm{P}(\text{a non-seeded cloud is randomly chosen}) = 0.7$

Exercise 5.2(e)

The permutation test can be applied using the following R code:

```r
# Data
A = data$A
X = data$X

# OLS estimator
FUN = function(x,y){
  n = length(x)
  sum((x[1:n]*y[1:n])/0.3 + ((1-x[1:n])*y[1:n])/0.7)
}

# Compute the Permutation p-value
# Using the Order Permutation Test Function for Paired Sample From Tut09
set.seed(1)
order.ptest(A,X, FUN, alternative="greater", plot=TRUE) # p-value = 7e-04
```

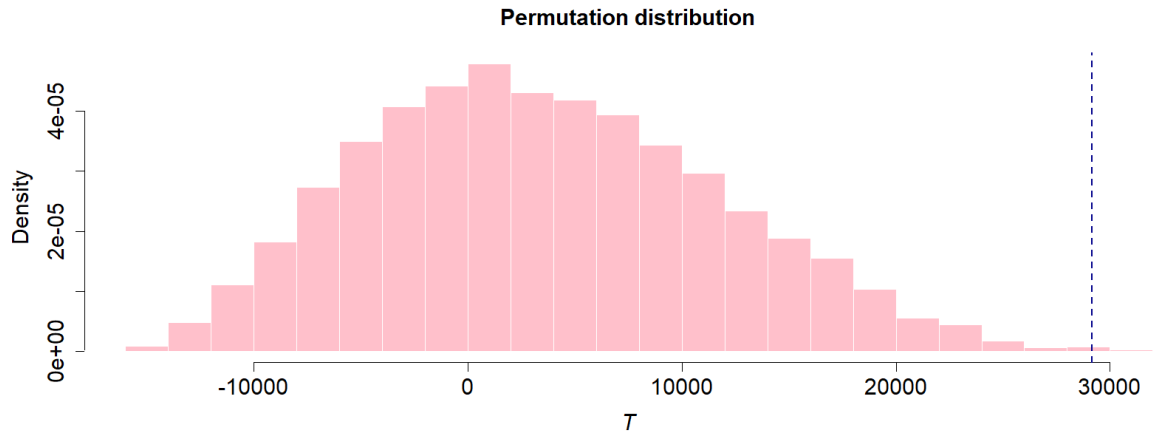
Since the computed p-value $< \alpha = 0.05$, we reject H_0, i.e. there is significant evidence to suggest that cloud seeding increases rainfall.

Exercise 5.2(f)

The plot of the permutation distribution of T can be produced using the following R code:

```r
set.seed(1)
order.ptest(A,X, FUN, alternative="greater", plot=TRUE)
```

And the result is:



Exercise 5.3(c)

Under H_0, mean rainfall for unseeded clouds ≥ 5 times mean rainfall for seeded clouds: $\mu_1 \leq 5\mu_0$. We apply order permutation test by randomly shuffling the treatment labels A_i among the observed X_i, A null distribution of the test statistic can be generated, reflecting the randomness under H_0. Reject $H_0 \iff$ permutation p-value < 0.05.

Exercise 5.3(d)

The proposed test statistic is

$$\sum_{i=1}^{n} \left[\frac{A_i X_i}{p_1} - 5 \cdot \frac{(1 - A_i) X_i}{p_0} \right], \text{where}$$

- $p_1 = \mathrm{P}(\text{a seeded cloud is randomly chosen}) = 0.3$;
- $p_0 = \mathrm{P}(\text{a non-seeded cloud is randomly chosen}) = 0.7$

Exercise 5.3(e)

The permutation test can be applied using the following R code:

```r
# OLS estimator
FUN2 = function(x,y){
  n = length(x)
  sum((x[1:n]*y[1:n])/0.3 - 5 * ((1-x[1:n])*y[1:n])/0.7)
}

# Permutation p-value
set.seed(1)
order.ptest(A,X, FUN2, alternative="greater", plot=TRUE) # p-value = 7e-04
```

Exercise 5.3(f)

The plot of the permutation distribution of T can be produced using the following R code:

```
set.seed(1)
order.ptest(A,X, FUN2, alternative="greater", plot=TRUE)
```

And the result is: