**P8122 - STATISTICAL METHODS FOR CAUSAL INFERENCE**
**MIDTERM**
**DUE OCTOBER 30TH 11:59PM**

**INSTRUCTIONS**

- Upload a pdf of your MIDTERM EXAM on Canvas along with your Script on the due date.
- The script should be self-contained, so someone else should be able to run it as is and get your results.
- The script should also be well-commented, so it is clear which code goes with which question.
- You ARE NOT ALLOWED to discuss these problems with each other nor share code with each other. This must be your own work.
- Please be concise!
- Please type your answers in a Word document, R markdown, or LateX. We will not accept scanned or pictures of handwritten answers.

**QUESTION 1**
**(70 points)**

Consider the following table of 20 individuals and suppose you know all the potential outcomes. In this table:

A=1: assigned new treatment
A=0: assigned standard treatment
Y=1: disease prevented
Y=0: disease not prevented

| "Truth" | | |
|---|---|---|
| **Individual** | **$Y_1$** | **$Y_0$** |
| 1 | 1 | 0 |
| 2 | 1 | 1 |
| 3 | 1 | 0 |
| 4 | 0 | 0 |
| 5 | 1 | 1 |
| 6 | 1 | 1 |
| 7 | 1 | 0 |
| 8 | 0 | 0 |
| 9 | 1 | 0 |
| 10 | 0 | 0 |
| 11 | 0 | 1 |
| 12 | 0 | 0 |
| 13 | 0 | 0 |
| 14 | 1 | 1 |
| 15 | 1 | 0 |
| 16 | 0 | 1 |
| 17 | 1 | 0 |
| 18 | 1 | 0 |
| 19 | 0 | 0 |
| 20 | 1 | 0 |

1) (10 points) Define and calculate the appropriate measure for the effect of the treatment and interpret it.

Suppose you conduct a study in the real world on the same 20 individuals to estimate the effect of the new treatment vs. the standard treatment on preventing disease. Here are the data from **Study 1**.

| **Study 1** | | | |
|---|---|---|---|
| **Individual** | **A** | **Y\|A=1** | **Y\|A=0** |
| 1 | 1 | 1 | . |
| 2 | 1 | 1 | . |
| 3 | 0 | . | 0 |
| 4 | 0 | . | 0 |
| 5 | 0 | . | 1 |
| 6 | 0 | . | 1 |
| 7 | 1 | 1 | . |
| 8 | 0 | . | 0 |
| 9 | 0 | . | 0 |
| 10 | 0 | . | 0 |
| 11 | 1 | 0 | . |
| 12 | 1 | 0 | . |
| 13 | 1 | 0 | . |
| 14 | 0 | . | 1 |
| 15 | 0 | . | 0 |
| 16 | 1 | 0 | . |
| 17 | 1 | 1 | . |
| 18 | 1 | 1 | . |
| 19 | 1 | 0 | . |
| 20 | 0 | . | 0 |

2) (5 points) Write the formula for the appropriate measure for this study and calculate it.

3) (10 points) Compare this estimate with what you obtained from Question 1. If you obtain the same number, explain why. If you obtain a different number, explain why (hint: make sure your explanation discusses the assignment mechanism).

4) (5 points) Explain how this type of data might arise in (a) an observational study and (b) a randomized controlled trial.

5) (5 points) Can you rule out a particular study design given what you observe?

Now let's suppose you want to perform a study (**Study 2**) that is a randomized controlled trial (RCT) looking at the effect of the new treatment vs. the standard treatment on preventing disease. You perform this study on the 20 individuals from Question 1. Assuming you have knowledge of every individuals' potential outcomes under both treatments (i.e., you know the truth), how would you create an RCT in which the assignment mechanism was randomized using a block randomization method, ensuring it is probabilistic and unconfounded.

6) (10 points) Please describe your process in words and provide a table showing your study (i.e., your observed data) under this assignment mechanism.

7) (10 points) Write and test the sharp null hypothesis of no causal effect in your study (describe your process and plot the randomization distribution). Interpret your results.

8) (5 points) What is your scientific question of interest for this randomized study under Neyman's approach to inference?

9) (10 points) Provide the point estimate and confidence interval of the marginal average causal effect in your study using Neyman's approach. Interpret your results. Compare your estimate with what you computed in **Question 1.1**.

**Question 2**
**(20 points)**

Suppose that the federal Department of Education surveys school districts to find out how many male and female teachers are employed in each district.
The Department learns that some school districts have been hiring only female teachers. Rather than intervene directly in hiring, the Department creates a workshop for school administrators that focuses on the benefits of diversity.
The workshop is given at all school districts that currently have only female teachers. At other school districts across the country, the workshop is given if requested by school administrators. Two years later, the Department again asks all school districts to indicate the number of male and female teachers employed.

1. Identify the units, potential outcomes, treatment, and any observed covariates. Explain each in one sentence.

2. Is the assignment observational or randomized? Explain your reasoning.

3. Is the assignment mechanism probabilistic? Explain your reasoning.

4. Is the assignment mechanism unconfounded? Explain your reasoning.

**Question 3**
**(10 points)**

Answer **one** of the two questions that we had selected for small group discussion after the group presentations.
If you were one of the presenters, you need to answer the question pertaining to the article that you have not presented on.

1. **ARTICLE:** Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology. 2008;19(6):766-779.
   **QUESTION:** After getting ITT estimates for the NHS study, why do the authors compare these estimates with the previous NHS study? What do you think was the intention of the researchers?

2. **ARTICLE:** Jackson LA, Jackson ML, Nelson JC, Neuzil KM, Weiss NS. Evidence of bias in estimates of influenza vaccine effectiveness in seniors. Int J Epidemiol. 2006;35(2):337-344.
   **QUESTION:** What are the confounders in this paper? What did the authors do to address them?