Causal Inference - Assignment 1
UCLA - 2025

Due: Monday, April 14th (before class)

Email your report to falco@ucla.edu amd kmishimoto@g.ucla.edu before the deadline. Reports can be in doc or pdf format. Please use the following format for files' names: ASSIGNMENT1_lastname.xxx

Problem 1: Referee Report

Choose one of the following papers:

- **Deaton, A., Cartwright, N. (2018)** "Understanding and misunderstanding randomized controlled trials," *Social Science & Medicine*, 210: 2-21.
  Available at: https://drive.google.com/file/d/1GuFOBBRIs5V_CtHIwQb0Zz0iRqEOzlWo/view?usp=sharing
- **Young, A. (2019)** "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results" *The Quarterly Journal of Economics*,
  Available at: https://drive.google.com/file/d/1avf3AH5hZZ5yH8uO1dWl8e6dW5zxsby3/view?usp=sharing

Write a short referee report no longer than two pages. The report must include: 1. A short summary of the paper; 2. a section with main comments; 3. a section with minor comments.

Problem 2: Computational Problem

If you want, you can use a statistical software or programming language of your choice to complete the following problem. You should submit a clear and concise report of your results.
The table below contains 12 observations from a (hypothetical) randomized experiment. S is sex at birth (F=female, M=male), X is a measurement of cholesterol level at the beginning of the experiment, and Y is a measurement of cholesterol level at the end of the experiment, after a new drug (treatment) or a placebo (control).

1. **Fisher**
   Assume that the population of interest includes only these 12 units. For parts 1b, 1c assume a completely randomized experiment.

Table 1: Cholesterol: Observed Data for 12 Units

Control Group			Treatment Group		
S	X	Y^{obs}	S	X	Y^{obs}
F	288	293	F	285	282
F	302	306	F	299	256
F	358	367	F	366	307
M	315	320	M	334	284
M	273	279	M	266	233
M	257	254	M	251	237

(a) Provide some evidence that the assignment was indeed random (e.g., how do the pre-treatment covariate distributions compare?)

(b) Using only observed outcome data, calculate exact Fisher p-values for the absolute value of the difference in sample averages of Y(1) and Y(0) and for the absolute value of the difference in Y(1) and Y(0) sample medians. Use also a third alternative test statistics of your choice. Use a sharp null hypothesis of zero treatment effects.

(c) How would you change the three test statistics if the design was a block randomized experiment, with two blocks of females and males?

(d) Repeat parts 1b-1c using "gain scores" rather than only Y data. Briefly discuss the differences. Is this what you would have expected to see? Explain in one sentence.

2. **Neyman**
Assume data are from a completely randomized experiment.

(a) Calculate an unbiased estimate of the average treatment effect for the 12 units.

(b) Apply Neyman's method to construct a 90% large sample confidence interval for the average treatment effect (ATE_{FP} and ATE_{SP}).

(c) Obtain an estimate of the average treatment effect for the 12 units specifying and estimating a linear regression model with W as a covariate and then one with W and X as covariate using OLS estimator. How do point estimates and variances compare to Neyman's? Comment in a few sentences.

(d) Use "gain scores" to construct a 90% large sample confidence interval for the average treatment effect using Neyman's method. How does this compare to the interval calculated in part 2b? Explain in one sentence.

3. In a few sentences, discuss the pros and cons of Neyman v. Fisher for causal inference in a classical randomized experiment.

Problem 3: Conceptual Problem

We want to study how effective a new drug is at reducing blood pressure in patients that have hypertension. A group of 100 people are recruited for this study from a hospital that treats a diverse range of people.

(a) Choose what classical randomized experiment you would use to randomize people into the treatment and control groups.

(b) Discuss the pros and cons of your choice.

(c) Choose what type of inference you would use on the collected data.

(d) Discuss the pros and cons of your choice.