P8122 - STATISTICAL METHODS FOR CAUSAL INFERENCE
FINAL
DUE DECEMBER 18TH 11:59PM

INSTRUCTIONS

- Upload a pdf of your FINAL 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!

PART I (20 points)

Consider the following table of 40 individuals sampled from the population. In this table:

A=1: assigned new treatment
A=0: assigned standard treatment
Y=1: disease prevented
Y=0: disease not prevented
L is a baseline covariate

Study				
Individual	**Y|A=1**	**Y|A=0**	**A=a**	**L=l**
1	1	.	1	1
2	.	0	0	1
3	.	0	0	0
4	.	0	0	0
5	1	.	1	1
6	1	.	1	0
7	.	0	0	1
8	.	0	0	0
9	.	0	0	0
10	0	.	1	1
11	.	1	0	0
12	1	.	1	1
13	1	.	1	1
14	.	1	0	0
15	.	1	0	0
16	1	.	1	1
17	1	.	1	0
18	1	.	1	0

19	1	.	1	0
20	1	.	1	1
21	.	1	0	1
22	1	.	1	0
23	0	.	1	0
24	0	.	1	0
25	.	1	0	0
26	.	0	0	0
27	.	0	0	0
28	1	.	1	0
29	.	0	0	0
30	0	.	1	0
31	.	1	0	1
32	1	.	1	1
33	.	0	0	0
34	.	1	0	0
35	1	.	1	1
36	1	.	1	1
37	.	1	0	1
38	1	.	1	0
39	.	0	0	0
40	.	1	0	0

1) Based on the information given in the Table, can you evaluate whether L is a confounder of the exposure-outcome relationship? Explain your reasoning.

2) Compute the association between A and Y.

3) Compute the ACE using the g-formula for observational studies assuming $Y_a \perp A|L$. Interpret the result.

4) Compare your result with what you obtained in **Question 2**. If you get the same result, explain why. If you get something different, explain why.

PART II (Points 40)
Let's say you want to parametrically estimate the average causal effect in your sample, using the propensity score subclassification method to adjust for L.

5) Estimate the propensity score using a model.

6) Evaluate covariate balance and overlap.

7) Provide the estimate and 95% confidence interval for the marginal average causal effect using propensity score stratification.

8) Provide the estimate and 95% confidence interval for the marginal average causal effect using the outcome regression to adjust for confounders. For this question, run a linear probability model. Compare your results with what you obtained in **Question 3**.

Note: the linear probability model might lead to outcome predictions outside the (0,1) allowed bounds for the probability. For now, ignore this issue.

9) List two advantages of using propensity score methods to adjust for confounding versus a traditional outcome regression approach that includes all confounders in the model.

10) Under which assumptions does the outcome regression that you ran in **Question 8** coincides with a parametric version of the g-formula for observational studies.

PART III (20 points)

Let's say you want to estimate the average causal effect in the treated in your sample using 1:1 exact matching to adjust for L.

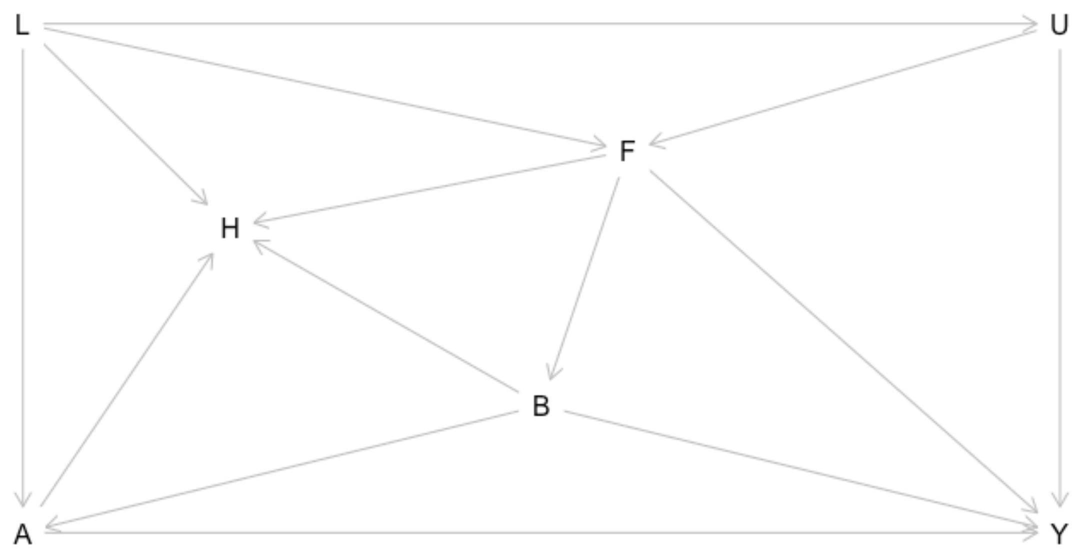
11) Find the matches and check covariate balance for the ATT. Are you satisfied with the covariate balance?

12) Compute the estimate, the standard error and confidence interval for the ATT.

EXTRA CREDIT: Conduct the analysis using and inexact matching procedure based on the propensity score. Describe your work and interpret the results. Compare the results with what you obtained with the exact matching procedure.

PART IV (20 points)

In truth, the relationship between A and Y is much more complex. See below the true DAG for the observational study where A is the treatment, Y is the outcome, L, H, F and B are measured variables and U is unmeasured.



13) Identify a variable or set of variables in the DAG that when conditioned on would close a back-door path between A and Y.

14) What is the relationship between NUCA assumption of potential outcomes and a DAG?

15) Identify a variable in the DAG that when conditioned on would open a closed path from A to Y.

16) Conceptually, what is a collider and why is it problematic to adjust for a collider? Can you provide an example of a collider in the DAG?