POLC 4325 Assignment #2

Posted: September 30, 2024
Due: October 14, 2024 (11:59pm)

1 Potential Outcomes Model (10%)

In the standard setup we discussed, which of the following is NOT true? Explain why.

A. An individual treatment effect $\tau_i = Y_i(1) - Y_i(0)$ is always not identifiable

B. An average treatment effect $E[\tau_i] = E[Y_i(1)] - E[Y_i(0)]$ is always identifiable

C. The value of τ_i depends on unit i

D. If all units from a subgroup are not treated, we cannot estimate the average treatment effect for the group

2 SUTVA (10%)

The stable-unit-treatment-value assumption (SUTVA) is a critical assumption in causal identification. Which of the following does NOT violate the SUTVA? Explain why.

A. Some control subjects received the treatment

B. Some control and treated units switched their treatment status by chance

C. Some treated subjects contacted their fellow treated units

D. Some treated units contacted their friends in the control group

3 Natural Experiments (10%)

Read McCauley and Posner (2017) "The Political Sources of Religion Identification: Evidence from the Burkina Faso-Côte d'Ivoire Border" published in the *British Journal of Political Science*.

1. What is **endogenous sorting**? (5%)
2. How does endogenous sorting threaten the identification strategy taken in the study? Explain while using the word "potential outcomes." (5%)

4 The Difference-in-Means Estimator (10%)

Read McCauley and Posner (2017) "The Political Sources of Religion Identification: Evidence from the Burkina Faso-Côte d'Ivoire Border" published in the *British Journal of Political Science*.

1. Find and download the replication dataset of McCauley and Posner (2017). Do not change the name of the original file. As a proof, **report the DOI** of the dataset. DOI is an URL starting from " https://doi.org/..." (5%)

2. Reproduce the first two rows of Table 1 (see below, you can ignore the standard errors). This means that you need to guess what the authors did to produce these reported values. You do not have to recreate the same table format. Please also note that Columns 3-4 in Row 2 cannot be exactly reproduced due to an error in the original paper.

When reproducing the ATE (Column 4), use two approaches:

- The simple difference-in-means estimator (i.e., compute two averages and compute the difference between them).
- The OLS. Compare the result to the one in 1.

Submit your code and make sure that your code is reproducible. It means that your code should generate the same result when the grader executes your code. If the grader encounters any errors while running your code, you will receive 0 point. (5%)

TABLE 1 *The Salience of Religious Identity*

	Full sample (%)	Burkina Faso (%)	Côte d'Ivoire (%)	CI – BF difference
	(1)	(2)	(3)	(4)
Lists religion as most important identity	18.8	10.0	27.8	17.8* (0.04)
Lists religion among top two identities	36.0	23.0	49.7	26.7 (0.14)
Willing to marry across religious lines	66.3	75.0	57.1	−17.9 (0.08)
Feels closer to co-nationals than co-religionists	59.7	77.3	40.3	−37.0* (0.02)

5 Stratification or Subclassification (35%)

As everyone knows, the Titanic ocean cruiser hit an iceberg and sank on its maiden voyage. Slightly more than 700 passengers and crew survived out of the 2,200 people on board. It was a horrible disaster. One of the things about it that was notable, though, was the role that wealth and norms played in passengers' survival.

Imagine that we wanted to know whether or not being seated in first class made someone more likely to survive. Given that the cruiser contained a variety of levels for seating and that wealth was highly concentrated in the upper decks, it's easy to see why wealth might have a leg up for survival. But the problem was that women and children were explicitly given priority for boarding the scarce lifeboats. If women and children were more likely to be seated in first class, then maybe differences in survival by first class is simply picking up the effect of that social norm.

Using the data set on the Titanic, we calculate a simple difference in mean outcomes (SDO), which finds that being seated in first class raised the probability of survival by 35.4%. But note, since this does not adjust for observable confounders age and gender, it is a biased estimate of the ATE. So next we use subclassification weighting to control for these confounders. Here are the steps that will entail:

1. Assess the covariate balance across the treatment and control groups with respect to the two conditioning variables (age and gender). In this dataset, Age == 0 if young and Age == 1 if old, whereas

Sex == 0 if woman and Sex == 1 if man. Discuss the result and what is the main problem with respect to causal identification (5%).

2. Stratify the data into four groups: young males, young females, old males, old females (5%).
3. Calculate the difference in survival probabilities for each group (5%).
4. Calculate the proportion of each of the four groups. These are our strata-specific weights (5%).
5. Calculate the weighted average survival rate using the strata weights.
6. Compare the above to the naive difference in mean survival rates in the data (5%).
7. In order to claim that the effect in 4. is a *causal* effect, what assumption do you have to make? Discuss also when such an assumption may break (5%).

6 Matching (25%)

Continue working on the Titanic dataset.

1. What is the causal estimand that one can identify with matching? (5%)
2. Estimate the estimand via the Coarsened Exact Matching. Report its estimated value (5%).
3. Estimate the estimand via the Nearest-Neighbor Matching. Report its estimated value (5%).
4. Discuss when the matching-based identification strategy (overall) breaks (5%).
5. Discuss the plausibility of the assumption required for matching to recover the causal estimand with the running example (hint: discuss whether there are possible unobserved confounders and what they may be) (5%).