Problem set 4

Time to solve some problems!

This problem set mainly builds on this notebook: Notebook 4 - Practical sources of bias.

1 Adjusting randomization proportions during experiment

Based on "Bias case 1: Adjusting randomization proportions during experiment".

1.1 In-class foundation (25 points)

1. Let's be clear about what kind of bias this is. Calculate the NATE decomposition. What kind of bias do we have in this example?
2. Check the independence assumption for this data set. Show that it does not hold.
3. Directly calculate the ATE from the data (which we can do because we have variables we wouldn't actually have!). Show that it matches what we set in the data-generating process.
4. Modify the "perfect stratification" analysis, which uses the date of visit to stratify. If we stratify by the day of the week, do we get an unbiased estimate of the ATE? Why does this work?
5. We can also use regression to implement a stratification. First, run a simple regression of Y on D and confirm that you get the same naive estimate of the ATE. Then add 'day_of_week' as a control variable. You need to use the 'C()' dummy variable operator. Show that this gives you an unbiased estimate.

1.2 Homework extensions (30 points)

1. Note that the data set also has a field for 'hour_of_day', which ranges over 1,2,3,...,24. This is the hour that the user visited the site. Does doing a perfect stratification analysis by the hour of the day give you an unbiased estimate of the ATE? *Show and explain your answer.*
2. Modify the data-generating process so that the treatment effect also varies by the day of the week. Let's imagine that since weekend visitors are less likely to convert they are generally less interested, so they also have a lower treatment effect. To make this very stark set the treatment effect to 0.005 on the weekend and 0.10 on weekdays. Implementing this in the DGP looks very similar to the conversion rate weekly cyclicality. You can adapt the function 'daily_conversion_rates' and be sure to change the 'ATE' variable from a constant to an array. Show these results: (i) Calculate the true ATE in the data, (ii) estimate the ATE

using a regression with ‘day_of_week‘ dummy variables, and (iii) estimate the ATE using a perfect stratification on ‘day_of_week‘. Which of the two estimators gives an unbiased estimate? What did you learn from this?

3. Bonus (+5 points): Explain your answer in problem 1.2.2.

2 One-sided noncompliance in a web experiment

This problem is based on "Bias case 2: Non-compliance and "exposure" to treatment".

2.1 In-class foundations (25 points)

1. In the code, what objects correspond to the variables D and Z from the causal inference framework?
2. Consider the subjects of the experiment in terms of *compliers* and *never-takers*. What set of subjects does ‘saw_treatment_page == coin_flip‘ select? What set of subjects does ‘viewed_page == 1‘ select? What is the difference between these two sets?
3. Return to the "Identification investigation" in the example. What is the as-treated estimate of the causal effect? Why is it biased for the true ATE?
4. Use the ITT regression, but include ‘viewed_page‘ as control variable. Try also including ‘charitability‘ as a control variable. Do these let you identify the ATE? What do you learn from this?
5. For the data set in this case, show the numbers to verify that

$$\text{CACE} = \frac{\text{ITT}}{\text{proportion of compliers}}.$$

What is the intuitive explanation for this?

2.2 Heterogeneous treatment effects (20 points)

This problem used "Bias case 3: Charitability influences viewing probability *and treatment effect*" from the notebook.

1. Calculate the ATE directly from the simulated data.
2. Make a table of all four analysis approaches. Relate it to the last problem. What do you learn from this?

3. In the first problem, one of the four methods happened to identify the ATE. What happened to that method in bias case 3? Why does it work differently now?

4. Suppose that our website had a bug in the logging, so ‘viewed_page‘ is no longer reliable. The other fields in the data are still reliable. Show precisely how we can still estimate the CACE.