

PL 9239: Assignment 4

Deadline: Tuesday May 10th

Please write up all your answers in either a R script or a Word document if you prefer.

- Irrespective of submitting R code or a word document, this document has to contain all code. The R code should be running from top to bottom in one go once the working directory is adapted to the respective machine.
- Any answers that you are adding in your own words should also go into the document.
- Please make sure that you structure your script well so that it is easy to read.

Experimental Data

Foundations

- What is the Fundamental Problem of Causal Inference? (4-6 sentences, 1 point)
- Do experiments have internal validity, external validity or both? Explain your answer. (4-6 sentences, 1 point)

Analysing an Experiment

There is a persistent gender gap in motivations to run for political office. For this exercise, you are analysing data from an experiment to study what could increase women's political ambition. (NB: The design of the experiment is from Foos and Gilardi (2019), but I am generating artificial data for the sake of simplicity).

For the experiment female students were randomly assigned to receive emails inviting them to an event that included career workshops with female politicians, or no email. One day after the treatment email went out, all participants were asked on a scale from 0–10 whether they would have any ambitions to pursue a political career. For the analysis it is OK to assume that the answers between 0-10 are continuous data.

Load the data `a4_experiment_pol_career.csv`. Analyse the treatment effect from the invitation email on the students' political ambition.

- Interpret the substantive causal effect from the experiment. What do you conclude? (2-3 sentences, 1 point)
- Interpret the statistical uncertainty of the causal effect from the experiment. What do you conclude? (2-3 sentences, 1 point)

Treatment Heterogeneity in an Experiment

Let us imagine that some scientists criticize the design of the first study. They claim that results are different if one distinguishes between the young 'Fridays-for-Future' cohort and older students. Repeating the original experiment, they include two further variables: whether the person receiving the email is younger than 25 and whether she is self identifying with a politically left political ideology.

Load the data `a4_experiment_pol_career_2.csv`. Analyse the treatment effect from the invitation email on the students' political ambition. This time, also analyse the scientists' hunch that the treatment effect might be different for those younger than 25. (NB: all data is of course again artificial and not 'real' data.)

- Specify a regression model that correctly captures the causal effect (1 point)
- Interpret the results from the regression model:

- What is the combined substantive causal effect of the treatment for those younger than 25, and what is the combined substantive causal effect for those 25 years and older? What can you say about the uncertainty of the two effects? (1 point)
- What is the average political ambition that you would expect from someone who did not receive the email, is younger than 25 and does not identify as left? (1 point)

Difference-in-Difference

Foundations

- What is the core assumption that you have to make when choosing the control group in a difference-in-difference research design? (4-6 sentence, 1 point)

Analysis

You are analysing the impact of reducing the speed in a road from 40mph to 20mph on the air quality in that region. Air quality will be measured using dust pollution (or particulate matter, PM10). This dust pollution occurs in a metropolitan area when tiny particles of dust or matter—e.g. from car emissions—becomes airborne.

You have a measurement of the PM10 air quality on a daily basis for 4 roads over the course of two months. You know that reducing the speed limit from 40mph to 20mph was imposed for the first road on the 20th day. All other roads did not receive any reduction in speed and continue at 40mph. You want to estimate the causal effect of the speed limit reduction on the air quality measured with PM10. (NB: All data are artificial data.)

- What road should you pick as the correct control group? Explain why you chose this road (3-5 sentences, 1 point)
- Calculate the substantive causal effect of the speed limit reduction using the difference-in-difference estimator. (1 point)