Exercise: Comparative Case Studies & Process Tracing

Let's simulate some fake data and see whether we are able to recover the correct treatment effect using comparative case study methods. Our research question is whether being a high-income country causes democracy.

1. First, let's generate some confounder variables for 100 countries (we have only cross-section data).
   (a) The variable 'Inequality' should be drawn randomly from the normal distribution with mean 0.5 and standard deviation 0.2.
   (b) The variable 'Continent' should be randomly drawn with equal probability from four possibilities: Europe, Americas, Asia, and Africa.
2. Our outcome is going to be a binary variable for whether the country is a democracy or not. The relationships we will assume are based on existing theory and evidence:
   - European countries are more likely to be high-income and democracies
   - Inequality increases the chances of democracy (Meltzer-Richards theory), but reduces the level of development
   - Our treatment variable, high income, increases the log-odds of democracy by 0.5 (so our treatment effect in a logit model should be 0.5).

Use the expressions below to simulate potential outcomes

$$y_0 = \begin{cases} 1 \text{ if } (2 * inequality + 0.5 * Europe) > 1 \\ 0 \text{ else} \end{cases}$$

$$y_1 = \begin{cases} 1 \text{ if } y_0 = 1 \\ 1 \text{ if } y_0 = 0 \ \& \ Binom(0.5) = 1 \\ 0 \text{ else} \end{cases}$$

3. Treatment D is the level of income of the country. It depends on the confounding variables and we will simplify to High versus Low income based on whether the country has an income over USD10,000. Imagine we know the treatment assignment mechanism so that binary (1/0) treatment is determined by the following expression:

$$D = \begin{cases} 1 \text{ if } (N(10000, 2000) + 2000 * Europe - 1000 * inequality) > 10000 \\ 0 \text{ else} \end{cases}$$

4. Calculate observed outcomes based on potential outcomes and treatment.
5. Draw the causal diagram that represents the relationships in the simulated data. (Make sure you indicate the direction of the effect).
6. Now we will implement a comparative case study methodology for just two units. Select one treated (**high income**) unit that is a **democracy**. What are the values of the two confounding variables, inequality and continent?
7. From all of the control (low income) units that are NOT democracies, pick one at random.
   (a) Is it *possible* that being European explains why your treated unit is a democracy and the control unit is not?

(b) Is it *possible* that inequality explains why your treated unit is a democracy and the control unit is not?
(c) Can we conclude anything about the role of income level (treatment) on democracy from this comparison?

8. What would be the ideal characteristics of a control unit that allow you to answer "No" to questions (a)-(b) in Question 7 and will help demonstrate the effect of income level (treatment) on democracy? I.e. what continent should the control unit be from, and what level of inequality should it have (compared to the treated unit)?

9. Identify one control unit that is a non-democracy and has values of the confounders you identified in Question 8. It may not be possible - if not, try picking a different treated unit and then search for an appropriate control unit.

10. Once you succeed in finding an appropriate pairing of treated and control units, what can you conclude about the role of development in explaining democracy?

11. How **representative** are your treated and control cases? Use a table or graph or numerical comparison to compare them to the overall distribution of countries on the confounding variables.

12. Let's take a closer look at Europe. We know democracies are more common in Europe, and so if we had a treated unit in Europe we would need a control unit also in Europe to avoid the risk of bias. How easy is it to find these units? Specifically, how many units do you have in Europe and how many of them are control (low-income) units? How many are also non-democracies?

13. Return to your selected treated case and ignore the control cases. If we were to conduct process tracing on our single treated case to generate additional evidence that it was development that caused democracy in this case, what observable implications of this theory/mechanism would you look for?

14. A competing theory of democratization (think Acemoglu and Robinson) is the role of inequality in generating a stronger demand for redistribution, which is expressed through protests by the poor/left and is ultimately only possible to guarantee through democracy. What observable implications can you think of that might help us *exclude* the mechanism of inequality as an explanation of our treated case using process tracing?

15. Comparing the observable implications you have suggested in Q13 and Q14, which are most likely to help us distinguish the influence of development versus inequality?