Problem Set 1
EC140

Identify a causal questions in the real world

Due Friday July 1st, 5pm on bCourses

Find a news article, video clip, or personal story of a potential causal relationship that you are interested in, and use it to answer the following questions:

1. What is one dimension or characteristic over which you are interested to investigate the effect over? Later in the course, we will call this an outcome, or a dependent variable.
2. What is the event, variable, or policy that is causing something? Later in the course we will call this the treatment or independent variable of interest. (Example: ban on abortion at the state level)
3. Could you think of a data set, real or fictitious, where you could observe these variables? Describe what each variable (column) means. What are the observations (rows)? In addition to these two variables, could you think of other variables that might help us better understand this relationship? Later in the course we will call these control variables.
4. What could be the population of interest? And what a sample?
5. Please present fictional statistics for what might be a plausible mean and standard deviation of each. Discuss when standard deviations might not be informative.
6. Provide an example of how selection bias could be threatening the analysis
7. Describe how an RCT could be used to estimate the true causal effect of interest.

Max 2 pages and 700 words. See example below.
Each question will be graded on a 4 point scale: (1) no response; (2) response does not demonstrate understanding of the question; (3) demonstrates understanding; (4) demonstrates clear understanding. The difference between understanding and clear understanding will be assessed on the basis of clarity, brevity and ability to connect with concepts discussed in class.

Example:
Article from NYT today (June 20th, 2022) “With Roe Set to End, Many Women Worry About High-Risk Pregnancies”. Link here.

1. Dimension of interest: Health of women with high risk pregnancies.
2. Causing variable: ban on abortion at the state level.
3. Data:
    a. Each observation (row) could represent a pregnant woman, across all US states in some past period, for example 2010 - 2020.
    b. Dimension of interest: Health information for women with high risk pregnancy. Defined as mortality 1 year after birth or abortion. This variable would take the value of 1 if the woman died, and 0 if she survived after 30 days.
    c. Causing variable: Information if there is an abortion ban in the state where each women resides. Operationalized as 1 if state has strong regulations against abortion at time of diagnosis (of high risk) and zero otherwise.
    d. Other variables: Women's age, and Education could be of interest when assessing differences between those who were subject to the policy and those who did not.

4. The data above could be the population of interest. It could also be all pregnant women at some point in the future, or in a different country. A sample could be data from one health care provider (e.g. Kaiser) in two states, for example California and Arizona, for a narrower time frame, say 2016 to 2020. In this case our causing variable (treatment) could simply be residing in Arizona.
5. Fictional statistics:

Variable	Mean	SD
1-year maternal mortality rate	0.00019	-*
Strong anti-abortion regulation	0.3	-
Age	27.5	5
Education	11.3	3

* SD of binary variables are not so meaningful. If you have the mean and the sample size, you can calculate the standard deviation

6. If we compare Mortality Rates in California vs Arizona we might be violating the “all else equal” assumption (ceteris paribus). For example, let's assume that women get pregnant at a later age in CA relative to AZ. Given that high risk pregnancies are more likely to happen later in life, it could be the case that we observe higher mortality in CA *regardless* of its abortion laws. If we do not account for this, we could incorrectly claim that lower abortion laws cause and *increase* in mortality of pregnant women.

7. [WARNING: just a thought exercise!] Randomly assign anti-abortion regulation across states, and compare mortality rates between states that receive the treatment of anti-abortion, and those that did not (status quo). In this case treatment and control are equally likely to differ in other characteristics (like age, and others).