CORE CONCEPTS OF EXPERIMENTAL DESIGN

PROF. SEBASTIÁN VALLEJO VERA

HAPPY BIRTHDAY GRAMSCI!

- Tomorrow is the 134th birthday of Antonio Gramsci. According to Davidson's (1977) biography of Gramsci, Gramsci believed that "[…] the problem of education is the most important class problem".
- Last year, the editor of The Lancet wrote that "we must engage in a *war of position,*" citing Gramsci's idea that the dominant group uses culture to exert its controlling influence.
- We all should read more Gramsci.

- From our last session, we know that:
  - The fundamental problem of causal inference is the **counterfactual problem**
- We also showed (somewhat superficially) that *experiments* allow us to provide a solution to our lack of observable counterfactual.

- Experiments are a type of quantitative method in which researchers randomly assign individuals to experimental conditions (i.e., a treatment condition where the explanatory factor is present and a control condition where it is not) in order to test causal arguments.

A BRIEF HISTORY OF RANDOMIZATION

(AND GREAT WAY TO MAKE YOU FEEL UNACCOMPLISHED)

GOOD OL' HUME:

- "We may define a cause to be an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second. Or in other words where, if the first object had not been, the second never had existed."

STATISTICAL INFERENCE

- At the beginning of the 19th century, the astronomer Giuseppe Piazzi discovered the dwarf planet Ceres, between Jupiter and Mars. Piazzi observed the planet 24 times before it disappeared. Carl Friedrich Gauss (yes, that Gauss) proposed a method that would be able to predict the location of Ceres based on its previous locations. His method minimized the sum of the squares of the errors…
- Gauss invented OLS…
- When he was 18… year… old.

- In 1899, an economist named Udny Yule used regressions to determine the cause (*cause* according to him) of poverty in England. The *poor* relied on public assistance. Yule wanted to know if public assistance increased the number of beggars.
- Yule used the following equation to find this *causal* relation:
  - $Beggars = \alpha + \delta\, Assistance + \beta_1\, Age + \beta_2\, Population + \mu$

- Bad in different ways, but the important thing is that he tried and was the first to do it in the social sciences.

RANDOMIZATION

- The first time that randomization was used for research was in psychology to trick patients (rather than to determine causality).
- During the first half of the previous century, randomization was mainly used in agriculture.
- The first formal use of formal randomization was used to estimate the effect of the Salk vaccine (aka the polio vaccine).

- Experiments are considered the gold standard in quantitative methods for causal inference and represent a benchmark by which the results from other methods are judged.
- The random assignment of individuals to experimental conditions, and the variation of only the explanatory variable across these conditions, isolates the effect of the explanatory variable and helps to eliminate alternative explanations for the outcome of interest.

- Experiments are frequently evaluated according to their level of internal and external validity.
- Internal validity refers to the degree to which the relationship between the explanatory variable and the outcome variable is causal.
  - In other words, internal validity refers to the ability to attribute changes in the outcome variable exclusively to changes in the explanatory variable.
- External validity refers to the extent to which inferences from an experiment are generalizable.

- There are three basic types of experiments:
  - laboratory experiments;
  - field experiments;
  - survey experiments.
- Each type of experiment varies in terms of the internal and external validity they can afford.

- (There are also natural experiments, for when you cannot send postcards to police officers making traffics stops or put presidential candidates in a lab)

- Still… why does randomization work?

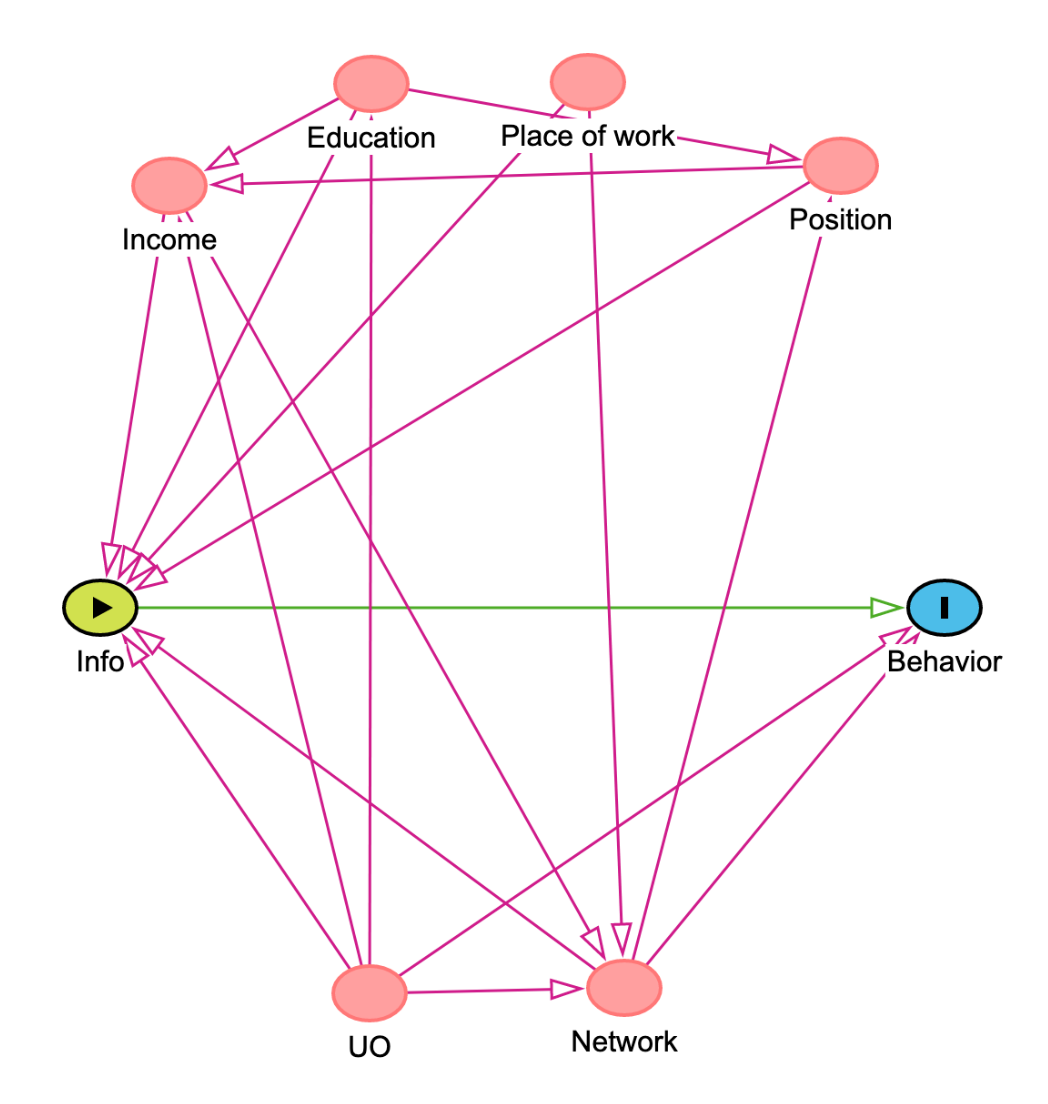
- Let's look at two explanations:
  - The selection bias explanation (why does not randomizing does not work?)
  - The potential outcomes approach (why does randomization work?)

SELECTION BIAS

WHY THINGS DON'T WORK

EXAMPLE

- Liebman and Luttmer (2015) wanted to know the effect of information on the decisions taken by individuals regarding social security benefits in the U.S.
- Specifically, they wanted to know if obtaining more information about social security benefits helped people make more optimal decisions regarding their pension plan. For example, understand the impact to their pension of extending a year of employment



Education
Place of work
Income
Position
Info
Behavior
UO
Network

- If we were to compare the behavior of those with information and those without information we would run into some problems.
- Individuals with information and individuals without information are different in many other respects. For example, the individuals with information are more educated, more likely to have higher incomes, and more likely to have a high-information network than individuals without information.
- The problem is that these characteristics are correlated with both information and behavior.

- In other words, workers with information behave more optimally for all sorts of reasons, including, perhaps, the causal effect of information. But people with information also behave more optimally because they have a high-information network, among other things. This is a problem of selection bias.

- The comparison of informed and uninformed, therefore, is a comparison of *apples* to *oranges* not *apples* to *apples*, preventing us from using the difference in means between the informed and uninformewd as an estimate of the causal effect of information on behavior.
- If the only source of selection bias is a set differences in characteristics that we can observe and measure, selection bias is (relatively) easy to fix using regression analysis, for example.
- The problem, however, is that not all relevant differences are measurable (e.g., networks), or worst, observable.

FORMAL DEMONSTRATION OF SELECTION BIAS

- Consider Y_{1i} the behavior of individual i with information and Y_{0i} the behavior of the same individual i without information.
- Thus, the individual causal effect of information on behavior is $Y_{1i} - Y_{0i}$
- In a group of n people, average causal effects are written as:
  - $Avg_n[Y_{1i} - Y_{0i}] = \frac{1}{n}\sum_{i=1}^{n}[Y_{1i} - Y_{0i}]$
- But, as we already know, Y_{1i} and Y_{0i} cannot be both observed.

- Consider D_i a dichotomous variable indicating the information status of individual i, where $D_i = 1$ if i has information and $D_i = 0$ otherwise.
- We have that the difference in group means is:
  - $Avg_n[Y_i|D_i = 1] - Avg_n[Y_i|D_i = 0]$
- But the average Y_i for the informed ($D_i = 1$) is necessarily an average of the outcome Y_{1i} and contains no information about Y_{0i} . Likewise, the average Y_i for the uninformed ($D_i = 0$) is an average of outcome Y_{0i} and this average is devoid of information about the corresponding Y_{1i}.
- Thus, we can rewrite the difference in group means as:
  - $Avg_n[Y_{1i}|D_i = 1] - Avg_n[Y_{0i}|D_i = 0]$

- Now, let's suppose that the individual (and average) causal effect of information on behavior is κ. Such that:
  - $\kappa = Y_{1i} - Y_{0i}$
- Which we can rewrite as:
  - $Y_{1i} = \kappa + Y_{0i}$
- By using this expression, we can rewrite the difference in group means as:
  - $Avg_n[Y_{1i}|D_i = 1] - Avg_n[Y_{0i}|D_i = 0] = Avg_n[(\kappa + Y_{0i})|D_i = 1] - Avg_n[Y_{0i}|D_i = 0]$
  - $= \kappa + (Avg_n[Y_{0i}|D_i = 1] - Avg_n[Y_{0i}|D_i = 0])$

- The equation:
  - $\kappa + (Avg_n[Y_{0i}|D_i = 1] - Avg_n[Y_{0i}|D_i = 0])$
- reveals that behavior comparisons between the informed and uninformed equal the causal effect of interest (κ) plus the difference in average Y_{0i} (behavior) between the informed and uninformed. This second term describes selection bias.
- In other words, the difference in average behavior by information status can be written as:
  - $Difference\ in\ group\ means = Average\ causal\ effect + Selection\ bias$

*We saw this in effect in a simulation last week

- The question is how to eliminate selection bias?
- One way is to know EVERYTHING that affects Y_{0i} and also affects getting the treatment. That is hard (impossible?).
- The other way to eliminate Selection bias is by random assignment.
- The idea is simple. Select a sample of people that are currently uninformed and provide information to a randomly chosen subset of this sample (by a coin toss, for example). Later, the behavior of the informed and uninformed groups can be compared to estimate the causal effect of information on behavior.
- Random assignment makes this comparison *ceteris paribus*: informed groups and uninformed groups created by random assignment differ only in their information status and any consequences that follow from it.
  - The risk of following in either group is the same for any individual in the sample (going back to the previous week).

- Two randomly chosen groups from a same population, when large enough, are comparable (remember last week). In fact, randomly assigned groups should be similar in every way, including ways that we cannot easily measure or observe (due to the powerful statistical property of the Law of Large Numbers). This is the root of random assignment's awesome power to eliminate selection bias.

- Why? Well, one way to explain it is through the idea of potential outcomes…

POTENTIAL OUTCOMES

A *LESS MATHY* VERSION OF THE ORIGINAL

- Suppose we want to know the causal effect of the treatment X_i on the *outcome* Y_i.
- For example, let's assume that Y_i is behavior and that the treatment is information such that:
  - $X_i = 1$ when i gets information
  - $X_i = 0$ when i does not get information
- For each individual i there are two potential *outcomes*.
  - $Y_i(1)$ is the outcome for individual i if they get information
  - $Y_i(0)$ is the outcome for individual i if they do not get information
- The causal effect of the treatment on the *outcome* of individual i is:
  - $CausalEffect_i = Y_i(1) - Y_i(0)$

- The observed outcome for Y_i can be written as a potential outcome:
  - $Y_i = Y_i(1) \cdot X_i + Y_i(0) \cdot (1 - X_i)$
- If the individual i received the treatment ($X_i = 1$):
  - $Y_i = Y_i(1) \cdot 1 + Y_i(0) \cdot 0 = Y_i(1)$
- If the individual i did not received the treatment ($X_i = 0$):
  - $Y_i = Y_i(1) \cdot 0 + Y_i(0) \cdot 1 = Y_i(0)$

- But… here we have an identification problem. Why?
- That's right, we cannot identify the causal effect for the individual i because we either observe $Y_i(1)$ or we observe $Y_i(0)$. But never both (once again, the counterfactual problem).

- Even though we will never be able to observe the causal effect of the treatment on individual i, maybe we can estimate the average causal effect of the treatment on the population.
- The average treatment effect (ATE):
  - $E[Y_i(1) - Y_i(0)] = E[Y_i(1)] - E[Y_i(0)]$
- In the ATE equation we still need to know the potential outcomes of individual i. Since we can only know one of these outcomes, the ATE is inherently unknown. Therefore, the ATE can only be *estimated*.

- Ok… here is where the magic happens:
- The potential outcome can be different for two individuals:
  - $Y_i(1) \neq Y_j(1) \quad and \quad Y_i(0) \neq Y_j(0)$
- However, if treatment X_i is randomly assigned, the distribution of potential outcomes will be the same for the treated group $(X_i = 1)$ and the control group $(X_i = 0)$.
- With randomization, the potential outcomes are independent of the treatment. Therefore,
  - $E[Y_i(1)|X_i = 1] = E[Y_i(1)|X_i = 0]$
  - $E[Y_i(0)|X_i = 1] = E[Y_i(0)|X_i = 0]$
- In other words, the groups are exchangeable.

- In an experimental setting, individuals are randomly assigned to a treatment group and a control group, thus:
  - $E[Y_i(1)] = E[Y_i(1)|X_i = 1] = E[Y_i|X_i = 1]$
  - $E[Y_i(0)] = E[Y_i(0)|X_i = 0] = E[Y_i|X_i = 0]$
- This would suggest that:
  - $E[Y_i(1) - Y_i(0)] = E[Y_i(1)] - E[Y_i(0)] = E[Y_i|X_i = 1] - E[Y_i|X_i = 0]$
- This translates as ”the potential outcomes are the same on average” since (from the previous slide):
  - $E[Y_i(1)|X_i = 1] = E[Y_i(1)|X_i = 0]$
  - $E[Y_i(0)|X_i = 1] = E[Y_i(0)|X_i = 0]$

- And the selection bias?
- No selection bias!! YEY!

EXAMPLE (CONT.)

- What Liebman and Luttmer did was randomly send an informational pamphlet and video about social security and pensions to a group of workers >50 years old (N=2,483).
- We now have two groups:
  - A treatment group ($X_i = 1$) who gets pamphlet and video
  - A control group ($X_i = 0$) who does not get pamphlet and video
- They were interested in the effect of information on the behavior of workers (in this case, if they worked more or not).

- For each individual surveyed we have two potential outcomes:
  - $Y_i(1)$ is the outcome for individual i if they get pamphlet and video
  - $Y_i(0)$ is the outcome for individual i if they do not get pamphlet and video
- The causal effect of the pamphlet and video in the behavior of i is $Y_i(1) - Y_i(0)$.
  - But this we cannot observe.
- Since each individual surveyed was assigned randomly to the treatment group and the control group, we can estimate $E[Y_i(1) - Y_i(0)]$ if we compare
  - The average behavior of those who got the pamphlet and video $E[Y_i|X_i = 1]$
  - The average behavior of those who did not get the pamphlet and video $E[Y_i|X_i = 0]$

- Let’s see this in action (go to code for 4 Core Concepts of Experimental Design.R / Vignette 3.1)

- Why can we use an OLS model to estimate the effects?
- Remember that we said:
  - $Y_i = Y_i(1) \cdot X_i + Y_i(0) \cdot (1 - X_i)$
  - $Y_i = Y_i(0) + [Y_i(1) - Y_i(0)] \cdot X_i$
  - $Y_i = E[Y_i(0)] + [Y_i(1) - Y_i(0)] \cdot X_i + ([Y_i(0)] - E[Y_i(0)])$
  - Reminds you of something?

  - $Y_i = \beta_0 + \beta_1 \cdot X_i + \mu_1$

IN-CLASS EXERCISE

- Using the ’fertil2’ dataset from ’wooldridge’ on women living in the Republic of Botswana in 1988,
  - calculate the means and mean differences between those with and without electricity for the following two characteristics: education (educ) and age when first child was born (agefbrth);
  - evaluate if the mean differences are statistically significant at the 0.01 and 0.05 levels;
  - interpret the results;
  - from this analysis, would you say that the comparisons between women with and without electricity are apples to apples or apples to oranges?
  - How does that affect our ability to conclude anything about the relationship between access to electricity and number of children?