**Directions:** Email your TAs (see syllabus for their emails) by the beginning of class on Tuesday. You will be graded simply on the basis of the email. You want to tell them how well your progress is going. The only assignment you need to send is Assignment 1, which is your background assignment.

**Discord**

Use the discord channel to get to know your classmates and to get help on the material. The Discord channel is:

<https://discord.com/invite/yWjZRumpAg>

**Gentzkow and Shapiro**

Read Gentzkow and Shapiro “Code and Data for the Social Sciences” in the “Helpful stuff” Github directory and answer the following

1. Why do Genztkow and Shapiro think these elements of modern empirical work are so important? What problems does each element solve?
2. Give an example of the sort of problem that could arise in the course of an empirical project if someone were to fail to adopt these principles.
3. How do you plan to incorporate these solutions into your own work?

**Git and R**

These next questions concern the software “git” and “github”.

Read the “Telling Stories with Data” sections on R and Git. Install R Studio and Github Desktop and clone the Causal Inference github repository.

<https://www.tellingstorieswithdata.com>

<https://github.com/scunning1975/causal-inference-class>

**Potential outcomes, selection bias and the perfect doctor**

1. Consider the simple hypothetical example in Table 1. This example involves eleven patients each of whom is infected with coronavirus. There are two treatments: ventilators and bedrest. Table 1 displays each patient’s potential outcomes in terms of years of post-treatment survival under each treatment. Larger outcome values correspond to better health outcomes.

Table 1: Perfect doctor example

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Patient** | **Y1** | **Y0** | **Age** | **TE** | **D** | **Y** |
| 1 | 1 | 10 | 29 |  |  |  |
| 2 | 1 | 5 | 35 |  |  |  |
| 3 | 1 | 4 | 19 |  |  |  |
| 4 | 5 | 6 | 45 |  |  |  |
| 5 | 5 | 1 | 65 |  |  |  |
| 6 | 6 | 7 | 50 |  |  |  |
| 7 | 7 | 8 | 77 |  |  |  |
| 8 | 7 | 10 | 18 |  |  |  |
| 9 | 8 | 2 | 85 |  |  |  |
| 10 | 9 | 6 | 96 |  |  |  |
| 11 | 10 | 7 | 77 |  |  |  |

* 1. Provide an example of how SUTVA might be violated for treatments of covid-19.
  2. Calculate each unit’s treatment effect (TE).
  3. What is the average treatment effect for ventilators compared to bedrest? Which type of intervention is more effective on average?
  4. Suppose the “perfect doctor” knows each patient’s potential outcomes and as a result chooses the best treatment for each patient. If she assigns each patient to the treatment more beneficial for that patient, which patients will receive ventilators and which will receive bedrest? Fill in the remaining missing columns based on what the perfect doctor chooses.
  5. Calculate the simple difference in outcomes. How similar is it to the ATE?
  6. Calculate the ATT and the ATU. How similar are each of these to the SDO? How similar are each of these to the ATE?
  7. Show that the SDO is numerically equal to the sum of ATE, selection bias and heterogeneous treatment effects bias. You will need to calculate the ATE, selection bias and heterogenous treatment effects bias, combine them in the appropriate way, and show that their sum is equivalent to the SDO.

1. Programming exercise version of #1.
   1. Create a dataset with 10,000 observations where Y1 and Y0 differ across the population. Make the ATE equal to 10.
   2. Use the perfect doctor example to assign treatment to each unit. Decompose the SDO, as we did before, into selection bias and weighted heterogenous treatment effect bias.
      1. Hint: use randomization for Y1 and Y0 such that you have a treatment effet of 10 + Y0 for the Y1 variable.
   3. Show that E[Y1|D=1] = E[Y1|D=0].
   4. Show that E[Y0|D=1] = E[Y0|D=0].
   5. So if that is the case in c and d, then why isn’t the case that E[Y1|D=1] = E[Y0|D=0]? What’s the different reasons for (a) vs (b) and (c)?
   6. Put in your own words what independence means.
      1. Recall: (Y1,Y0) \_||\_ D is the math formula for independence.