Problem Set 1

Alex, Micah, and Scott

8/22/2020

Potential Outcomes Notation

1. Explain the notation $Y_i(1)$.
2. Explain the notation $Y_1(1)$.
3. Explain the notation $E[Y_i(1)|d_i = 0]$.
4. Explain the difference between the notation $E[Y_i(1)]$ and $E[Y_i(1)|d_i = 1]$.

Potential Outcomes and Treatment Effects

1. Use the values in the table below to illustrate that $E[Y_i(1)] - E[Y_i(0)] = E[Y_i(1) - [Y_i(0)]$.
2. Is it possible to collect all necessary values and construct a table like the one below in real life? Explain why or why not.

```
kable(table)
```

subject	y_0	y_1	tau
1	10	12	2
2	12	12	0
3	15	18	3
4	11	14	3
5	10	15	5
6	17	18	1
7	16	16	0

Visual Acuity

Suppose we are interested in the hypothesis that children playing outside leads them to have better eyesight.

Consider the following population of ten representative children whose visual acuity we can measure. (Visual acuity is the decimal version of the fraction given as output in standard eye exams. Someone with 20/20 vision has acuity 1.0, while someone with 20/40 vision has acuity 0.5. Numbers greater than 1.0 are possible for people with better than "normal" visual acuity.)

```
kable(d)
```

child	y_0	y_1
1	1.2	1.2
2	0.1	0.7
3	0.5	0.5
4	0.8	0.8
5	1.5	0.6
6	2.0	2.0
7	1.3	1.3
8	0.7	0.7
9	1.1	1.1
10	1.4	1.4

In this table, `y_1` means means the measured *visual acuity* if the child were to play outside at least 10 hours per week from ages 3 to 6. `y_0` means the measured *visual acuity* if the child were to play outside fewer than 10 hours per week from age 3 to age 6. Both of these potential outcomes *at the child level* would be measured at the same time, when the child is 6.

1. Compute the individual treatment effect for each of the ten children.
2. Tell a "story" that could explain this distribution of treatment effects. In particular, discuss what might cause some children to have different treatment effects than others.
3. For this population, what is the true average treatment effect (ATE) of playing outside.
4. Suppose we are able to do an experiment in which we can control the amount of time that these children play outside for three years. We happen to randomly assign the odd-numbered children to treatment and the even-numbered children to control. What is the estimate of the ATE you would reach under this assignment? (Please describe your work.)
5. How different is the estimate from the truth? Intuitively, why is there a difference?
6. We just considered one way (odd-even) an experiment might split the children. How many different ways (every possible ways) are there to split the children into a treatment versus a control group (assuming at least one person is always in the treatment group and at least one person is always in the control group)?
7. Suppose that we decide it is too hard to control the behavior of the children, so we do an observational study instead. Children 1-5 choose to play an average of more than 10 hours per week from age 3 to age 6, while Children 6-10 play less than 10 hours per week. Compute the difference in means from the resulting observational data.
8. Compare your answer in (7) to the true ATE. Intuitively, what causes the difference?

Randomization and Experiments

1. Assume that researcher takes a random sample of elementary school children and compare the grades of those who were previously enrolled in an early childhood education program with the grades of those who were not enrolled in such a program. Is this an experiment or an observational study? Explain!
2. Assume that the researcher works together with an organization that provides early childhood education and offer free programs to certain children. However, which children that received this offer was not randomly selected by the researcher but rather chosen by the local government. (Assume that the government did not use random assignment but instead gives the offer to students who are deemed to need it the most) The research follows up a couple of years later by comparing the elementary school grades of students offered free early childhood education to those who were not. Is this an experiment or an observational study? Explain!
3. Does your answer to part (2) change if we instead assume that the government assigned students to treatment and control by "coin toss" for each student?

Moral Panic

Suppose that a researcher finds that high school students who listen to death metal music at least once per week are more likely to perform badly on standardized test. As a consequence, the researcher writes an opinion piece in which she recommends parents to keep their kids away from "dangerous, satanic music". Let $Y_i(0)$ be each student's test score when listening to death metal at least one time per week. Let $Y_i(1)$ be the test score when listening to death metal less than one time per week.

1. Explain the statement $E[Y_i(0)|D_i = 0] = E[Y_i(0)|D_i = 1]$ in words. First, state the rote english language translation; but then, second, tell us the *meaning* of this statement.
2. Do you expect the above condition to hold in this case? Explain why or why not.

MIDS Admission

Suppose a researcher at UC Berkeley wants to test the effect of taking the MIDS program on future wages. The researcher convinces the School of Information to make admission into the MIDS program random among those who apply. The idea is that since admission is random, it is now possible to later obtain an unbiased estimate of the effect by comparing wages of those who where admitted to a random sample of people who did not take the MIDS program. Do you believe this experimental design would give you an unbiased estimate? Explain why or why not. Assume that everybody who gets offer takes it and that prospective students do not know admission is random.