Problem Set #1

W241: Field Experiments

Due the day before Week 3 live session.

Some guidelines for submitting problem sets in this course:

- Please submit a PDF document rather than a Word document or a Google document.
- Please put your name at the top of your problem set.
- Please bold or highlight your numerical answers to make them easier to find.
- If you'll be using R or Python code to calculate your answers, please put the code and its output directly into your Problem Set PDF document.
- It is highly recommended, although not required, that you use the RMarkdown feature in RStudio to compose your problem set answers. RMarkdown allows you to easily intermingle analysis code and answers in one document.
- You do not need to show work for trivial calculations, but showing work is always allowed.
- For answers that involve a narrative response, please feel free to describe the key concept directly and briefly, if you can do so, and do not feel pressure to go on at length.
- Please ask us questions about the problem set if you get stuck. Don't spend more than 20 minutes puzzling over what a problem means - it may just not be written clearly enough.

1. On the notation of potential outcomes:
    a. Explain the notation $Y_i(1)$.
    b. Explain the notation $E[Y_i(1)|d_i=0]$.
    c. Explain the difference between the notation $E[Y_i(1)]$ and the notation $E[Y_i(1)|d_i=1]$.
    d. (Extra credit) Explain the difference between the notation $E[Y_i(1)|d_i=1]$ and the notation $E[Y_i(1)|D_i=1]$. Use exercise 2.7 from FE to give a concrete example of the difference.
2. FE, exercise 2.2.
3. FE, exercise 2.3.
4. More practice with potential outcomes. 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.)

	$Y_i(0)$: Visual acuity after not playing outside	$Y_i(1)$: Visual acuity after playing outside.
Child 1	1.1	1.1
Child 2	0.1	0.6
Child 3	0.5	0.5
Child 4	0.9	0.9
Child 5	1.6	0.7
Child 6	2.0	2.0
Child 7	1.2	1.2
Child 8	0.7	0.7
Child 9	1.0	1.0
Child 10	1.1	1.1

In the table, state (1) means "playing outside an average of *at least* 10 hours per week from age 3 to age 6," and state (0) means "playing outside an average of *less than* 10 hours per week from age 3 to age 6." Y_i represents visual acuity measured at age 6.

a. Compute the individual treatment effect for each of the ten children. Note that this is only possible because we are working with hypothetical potential outcomes; we could never have this much information with real-world data. (We encourage the use of computing tools on all problems, but please describe your work so that we can determine whether you are using the correct values.)
b. In a single paragraph, tell a story that could explain this distribution of treatment effects. What might cause some children to have different treatment effects than others?
c. For this population, what is the true average treatment effect (ATE) of playing outside.
d. 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? (Again, please describe your work.)
e. How different is the estimate from the truth? Intuitively, why is there a difference?
f. 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)?

   g. 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.
   h. Compare your answer in (g) to the true ATE. Intuitively, what causes the difference?

5. FE, exercise 2.5. By way of clarification, note that the book typically defines D to be 0 for control and 1 for treatment. However, it doesn't have to be 0/1. In particular, one can have more than two treatments, or a continuous treatment variable. Here, the authors want to define D to be the number of minutes the subject is asked to donate. (This is because “D” stands for “dosage.”)
6. FE, exercise 2.6.
7. FE, exercise 2.8.
8. FE, exercise 2.9.
   a. Please think of the outcome variable as an individual’s answer to the survey question “Are you in favor of raising the estate tax rate in the United States?”
   b. The hint about potential outcomes could be rewritten as follows: Do you think those who won the lottery would have had the same views about the estate tax if they had actually not won it as those who actually did not win it? (That is, is $E[Yi0|D=1] = E[Yi0|D=0]$, comparing what would have happened to the actual winners, the $|D=1$ part, if they had not won, the $Yi0$ part, and what actually happened to those who did not win, the $Yi0|D=0$ part.) In general, it is just another way of asking, "are those who win the lottery and those who have not won the lottery comparable?"
   c. Assume lottery winnings are always observed accurately and there are no concerns about under- or over-reporting.
9. FE, exercise 2.12(a). In your answer, give some intuitive explanation in English for what the mathematical expressions mean.