Problem Set 1

Experiments and Causality

1. Potential Outcomes Notation

- Explain the notation (Y_{i}(1)). The notation is the outcome if you were in treatment
- Explain the notation (E[Y_{i}(1)|d_{i}=0]). The notation is the expectation of the outcome if you were in treatment and the treatment dosage is 0
- Explain the difference between the notation (E[Y_{i}(1)]) and the notation (E[Y_{i}(1)|d_{i}=1]). (Extra credit) One is the outcome if you were in treatment for any treatment dosage and the second is the outcome if you were in treatment for a treatment dosage of one.
- 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. The notation (E[Y_{i}(1)|D_{i}=1]) refers to the expectation of the outcome if the individuals is in the treatment group and the dosage is 1. The notation (E[Y_{i}(1)|d_{i}=1]) refers to the expectation of the outcome if the individual is in the treatment group and the dosage is one based on a sample of the population data. # 2. Potential Outcomes Practice Use the values in the following table to illustrate that (E[Y_{i}(1)] - E[Y_{i}(0)] = E[Y_{i}(1) - Y_{i}(0)]).

	(Y_{i}(0))	(Y_{i}(1))	(\tau_i)
Individual 1	5	6	1
Individual 2	3	8	5
Individual 3	10	12	2
Individual 4	5	5	0
Individual 5	10	8	-2

(E[Y_{i}(0)])=(5+3+10+5+10)/5=6.6 (E[Y_{i}(1)])=(6+8+12+5+8)/5=7.8 (E[Y_{i}(1) - Y_{i}(0)])=(1+5+2+0-2)/5=1.2 7.8-6.6=1.2 # 3. More Practice with Potential Outcomes 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.)

child	y0	y1
1	1.1	1.1
2	0.1	0.6
3	0.5	0.5
4	0.9	0.9
5	1.6	0.7
6	2.0	2.0
7	1.2	1.2
8	0.7	0.7
9	1.0	1.0
10	1.1	1.1

In the table, state (Y_{i}(1)) means “playing outside an average of at least 10 hours per week from age 3 to age 6,” and state (Y_{i}(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.

1. 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.)

Trteff
0.0
-0.5
0.0
0.0
0.9
0.0
0.0

0.0
0.0
0.0

One of the neat things about Rmarkdown documents is that you can bring results from your computation into the reading space. Do so, by opening an inline math chunk by calling for a single “back-tick” followed by the letter R.

So, an inline call would look like, 0.04. Please report all of your answers from an object using this form. And also, while you’re at it, please **embolden** your answers. So, **0.04** is the right form.

2. In a single paragraph, tell a story that could explain this distribution of treatment effects.

The mean treatment effect is **0.04**. This means that children playing outside had on average had worse eyesight. This could be because children who play outside may hurt their eyes due to being exposed to sunlight.

3. What might cause some children to have different treatment effects than others?

Some children might be more resistant to sunlight or some children’s eyes may respond to sunlight in a different manner. For instance, some children’s eyes may actually be nourished by sunlight, however, other children’s eyes may be hurt. Also, some children’s eyes may be hurt by sunlight but helped by other conditions outside.

4. For this population, what is the true average treatment effect (ATE) of playing outside.

The average treament effect of playing outside for this population is **0.04**.

5. 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.)

```
## [1] -0.06
```

6. How different is the estimate from the truth? Intuitively, why is there a difference?

The estimate is **-0.1** off from the truth. The reason for the difference is that we only considered a subset of samples for each group, and we did not control by individual, as we put different individuals in each group, although randomization helped in making other factors in the two groups the same.

7. We just considered one way (odd-even) an experiment might split the children. How many different ways (every possible way) 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)?

```
## [1] 1022
```

There are **1022** different ways to split children into control and treatment groups.

8. 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.

```
## [1] -0.44
```

1. Compare your answer in (h) to the true ATE. Intuitively, what causes the difference? The answer:**-0.44** is much lower than the true ATE. This is because we did not control for other variables which may affect eyesight, such as genetics, environment(factors like pollution), and nourishment of eyes.

5. Randomization and Experiments

Suppose that a reasearcher wants to investigate whether after-school math programs improve grades. The researcher randomly samples a group of students from an elementary school and then compare the grades between the group of students who are enrolled in an after-school math program to those who do not attend any such program. Is this an experiment or an observational study? Why?

This is an observational study, as the researcher does not control for variables other than enrollment in after-school programs that might affect the results. For instance, individuals who enroll in after-school programs might be motivated to study additional hours. Thus, there could be factors other than enrollment which may affect the results of the experiment.

6. Lotteries

A researcher wants to know how winning large sums of money in a national lottery affect people’s views about the estate tax. The research interviews a random sample of adults and compares the attitudes of those who report winning more than $10,000 in the lottery to those who claim to have won little or nothing. The researcher reasons that the lottery choose winners at random, and therefore the amount that people report having won is random.

1. Critically emvaluate this assumption.

The lottery choose winners at random, however, not everyone plays in the lottery, so the amount that people report having won is not random, as some groups of individuals may be more likely to play to lottery than others. .

2. Suppose the researcher were to restrict the sample to people who had played the lottery at least once during the past year. Is it safe to assume that the potential outcomes of those who report winning more than $10,000 are identical, in expectation, to those who report winning little or nothing?

No, it is not safe to assume this as an individual who plays the lottery will be more in favor of raising the estate tax if he or she had not won than a random person as people who play the lottery in hopes of getting rich are usually not rich to begin with. Thus, people who won the lottery, had they not won, would be less likely to support raising the estate tax, as they would be on average not as rich as the general population who did not play in or win the lottery. This means that (E[Y_{i}0|D=1]) is not equal to (E[Y_{i}0|D=0])

7. Inmates and Reading

A researcher studying 1,000 prison inmates noticed that prisoners who spend at least 3 hours per day reading are less likely to have violent encounters with prison staff. The researcher recommends that all prisoners be required to spend at least three hours reading each day. Let (d_{i}) be 0 when prisoners read less than three hours each day and 1 when they read more than three hours each day. Let (Y_{i}(0)) be each prisoner’s PO of violent encounters with prison staff when reading less than three hours per day, and let (Y_{i}(1)) be their PO of violent encounters when reading more than three hours per day.

In this study, nature has assigned a particular realization of (d_{i}) to each subject. When assessing this study, why might one be hesitant to assume that ({E[Y_{i}(0)|D_{i}=0] = E[Y_{i}(0)|D_{i}=1]}) and (E{[Y_{i}(1)|D_{i}=0] = E[Y_{i}(1)|D_{i}=1]})? In your answer, give some intuitive explanation in English for what the mathematical expressions mean.

({E[Y_{i}(0)|D_{i}=0]) is the expectation that a person will have violent encounters with staff given that he or she did not read three hours per day. (E[Y_{i}(0)|D_{i}=1]}) is the expectation that a person will not have violent encounters with staff given that he or she read three hours per day. (E{[Y_{i}(1)|D_{i}=0]) is the expectation that a person will have violent encounters with staff given that he or she did not read 3 hours a day. (E[Y_{i}(1)|D_{i}=1]}) is the expectation that a person will have violent encounters with staff given that he or she read 3 hours a day. A person is less likely to be violent with staff if he or she likes to read as reading usually indicates a calmer individual. This is why we might be hesitant to assume ({E[Y_{i}(0)|D_{i}=0] = E[Y_{i}(0)|D_{i}=1]}) and (E{[Y_{i}(1)|D_{i}=0] = E[Y_{i}(1)|D_{i}=1]}).