

# Problem Set #1

## *Experiments and Causality*

*September 13, 2017*

### 1. Potential Outcomes Notation

- Explain the notation  $Y_i(1)$ .

Answer:  $Y_i(1)$  would be the outcome for an individual subject if the subject was exposed to the treatment of the experiment. Breaking it down, the  $Y_i$  part stands for the response of the  $i$ th entity (from subjects 1 through  $N$ ; where  $N$  is total number of subjects inside the treatment cases) of an experiment and the  $(1)$  signifies that it is the treatment cases (not control cases which is:  $Y_i(0)$ )

- Explain the notation  $E[Y_i(1)|d_i = 0]$ .

Answer:  $E[Y_i(1)|d_i = 0]$  stands for the the expectation of the treatment outcome ( $Y_i(1)$  or outcome while receiving treatment) when one subject is selected at random from the subjects that did not receive the treatment. These outcomes are potential outcomes that can be imagined but not observed during the experiment.

- Explain the difference between the notation  $E[Y_i(1)]$  and the notation  $E[Y_i(1)|d_i = 1]$ .

Answer: The difference is that the notation for:  $E[Y_i(1)]$  represents the expected value of the outcomes in the treatment group for a random selected subject vs.  $E[Y_i(1)|d_i = 1]$  represents the expected value of the outcome when subjects are chosen at random from subjects that were actually received the treatment.

(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.

Answer: The difference in notation is for:  $E[Y_i(1)|d_i = 1]$  means expected value of the outcome when subjects are chosen at random from subjects that actually were treated vs. the notation:  $E[Y_i(1)|D_i = 1]$  stands for the expected value of the potential outcome of the treated group for a subject who would be treated under some hypothetical allocation of treatments. This hypothetical expectation can range in value and the range of the values is determined by how biased the estimator potentially can be. If the estimator is unbiased then the expectations should be very similar in expectation, but if it is biased then the range can be very large based on what hypothetical situation is being imagined.

### 2. FE 2.2

Use the values depicted in Table 2.1 to illustrate that  $E[Y_i(0)] - E[Y_i(1)] = E[Y_i(0) - Y_i(1)]$ .

Answer: \* Note testing the theory of expectations. We will test average of columns first then subtract, then subtract first and then average to show they are same result

```
table_2.1 = data.frame(Yi0 = c(10, 15, 20, 20, 10, 15, 15), Yi1 = c(15, 15, 30, 15, 20, 15, 30))

# E[Y_{i}(0)] - E[Y_{i}(1)]:
ind_exp_value = mean(table_2.1$Yi0) - mean(table_2.1$Yi1)
# E[Y_{i}(0) - Y_{i}(1)]
com_exp_value = mean(table_2.1$Yi0 - table_2.1$Yi1)
ind_exp_value == com_exp_value
```

```
## [1] TRUE
```

The above statement shows that the statement:  $E[Y_i(0)] - E[Y_i(1)] = E[Y_i(0) - Y_i(1)]$  is TRUE.

### 3. FE 2.3

Use the values depicted in Table 2.1 to complete the table below.

$Y_i(0)$	15	20	30	Marginal $Y_i(0)$
10	n: %:	n: %:	n: %:	
15	n: %:	n: %:	n: %:	
20	n: %:	n: %:	n: %:	
Marginal $Y_i(1)$				1.0

- Fill in the number of observations in each of the nine cells.
- Indicate the percentage of all subjects that fall into each of the nine cells.
- At the bottom of the table, indicate the proportion of subjects falling into each category of  $Y_i(1)$ .
- At the right of the table, indicate the proportion of subjects falling into each category of  $Y_i(0)$ .

Answer: \* NOTE: Rounded all values to thousandth position for visual ease and all tables entries are: n %

$Y_i(0)$	15	20	30	Marginal $Y_i(0)$
10	1 0.143	1 0.143	0 0.000	2 0.286
15	2 0.286	0 0.000	1 0.143	3 0.429
20	1 0.143	0 0.000	1 0.143	2 0.286
Marginal $Y_i(1)$	4 0.572	1 0.143	2 0.286	7 1.0

- Use the table to calculate the conditional expectation that  $E[Y_i(0)|Y_i(1) > 15]$ .

Answer: \*Note using same logic as Box 2.4 in FE,  $E[Y_i(0)|Y_i(1) > 15] = \sum Y_i(0) * \frac{Pr[(Y_i(0)=Y, Y_i(1)>15]}{Pr[Y_i(1)>15]}$  Since the statement  $Y_i(1) > 15$  is always the case,  $E[Y_i(0)|Y_i(1) > 15] = E[Y_i(0)]$  (shown in table 2.1 in book as equal to expected value of 15) shown below in formula form:

```
# $E[Y_{i}(0)|Y_{i}(1) > 15]$
exp_Yi0_condYi1_gt15 = ((2/7) / (7/7)) * 10 + ((3/7) / (7/7)) * 15 + ((2/7) / (7/7)) * 20
exp_Yi0_condYi1_gt15
```

```
## [1] 15
```

- Use the table to calculate the conditional expectation that  $E[Y_i(1)|Y_i(0) > 15]$ .

Answer: \*Note using same logic as Box 2.4 in FE,  $E[Y_i(1)|Y_i(0) > 15] = \sum Y_i(1) * \frac{Pr[Y_i(1)=Y, Y_i(0)>15]}{Pr[Y_i(0)>15]}$ . Because of the  $Y_i(0) > 15$  being violated in some situations, we will take total percentage of each  $Y_i(1)$  and subtract out the percentage when  $Y_i(0) > 15$  is not true in the column and divide by the total probability of  $Y_i(0) > 15$

```
# $E[Y_{i}(1)|Y_{i}(0) > 15]$
exp_Yi1_condYi0_gt15 = ((4/7 - 3/7) / (2/7)) * 15 + ((1/7 - 1/7) / (2/7)) * 20 +
  ((2/7 - 1/7) / (2/7)) * 30
exp_Yi1_condYi0_gt15
```

```
## [1] 22.5
```

## 4. 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.)

```
d <- data.frame(child = 1:10,
                y0 = c(1.1, 0.1, 0.5, 0.9, 1.6, 2.0, 1.2, 0.7, 1.0, 1.1),
                y1 = c(1.1, 0.6, 0.5, 0.9, 0.7, 2.0, 1.2, 0.7, 1.0, 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.

- 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.)

Answer: \*Note using equation 2.1 from FE to calculate treatment effect for each group, shown in dataframe format Treatment Effect:  $\tau_i = Y_i(1) - Y_i(0)$

```
answer.P0a <- d$y1 - d$y0
data.frame(child = d$child, treatment_effect = answer.P0a)
```

##	child	treatment_effect
## 1	1	0.0
## 2	2	0.5
## 3	3	0.0
## 4	4	0.0
## 5	5	-0.9
## 6	6	0.0
## 7	7	0.0
## 8	8	0.0
## 9	9	0.0
## 10	10	0.0

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

Answer: The first thing that stands out for the reason why the distribution looks the way it does is the sample size is small ( $N=10$ ). We have no idea how the samples were chosen and for the majority of the subjects there is no treatment effect. We have no idea if the differences is a localized effect or realitive to the study mechanisms. In addition to the sample size, the distribution can also represent an experiment that has no effect or small effect on outcome variables because of some omitted variable bias not properly taken into account or controlled for. Selection bias (or convenience bias) could also be a reason for the distribution shape. If random selection was not used, there is smoothing effect for intrinsic quality of participants and heterogeneity of the participants can be very relevant to the current group of participants.

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

Answer: A lot of instances could account for children observing different treatment effects. One could be that some children could live more into the wild than others and have easier access to deadly animals or insects and poisonous plants that could hinder eyesight quality. Another is if some kids play more contact sports than other kids and of those kids some of them obtained head trauma that lead to blurred vision. Some other reasons could be if the children had more access to modern conveniences (based on wealthier parents) they could have more access to electronics and screens that could either hinder or improve vision. Hinder:

could come from ruined eyesight from more tv watching. Improvement: could come from improvement of eyesight (from those suffering from lazy eye) with access to very expensive training software that improves eyesights for those who have lazy eye condition. In summary, humans have a lot of heterogeneity and you need to account for this in order to conduct a proper experiment.

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

Answer: \*Note from equation 2.3 in FE,  $ATE = \frac{1}{N} \sum_{i=1}^N \tau_i$

```
answer.P0d <- mean(answer.P0a)
answer.P0d
```

```
## [1] -0.04
```

The average treatment effect of playing outside is -0.04. This is due to majority of the population (8/10) have a treatment effect of 0.00, with two non-zero cases (child 2 with ATE = 0.5 and child 5 with ATE = -0.9). Since the majority of kids were 0, the negative treatment effect from child 5 is larger in magnitude than the positive treatment effect from child 2, so the ATE gets dragged to a lower value (negative) after being averaged over all participants

e. 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.)

Answer: \*Note split groups up into treatment (odd integers) and control (even integers) and then found treatment effect (like in 4a) for each group and then calculated ATE difference between two groups

```
treatment_POe = d[which(d$child %in% c(1, 3, 5, 7, 9)),]
control_POe = d[which(d$child %in% c(2, 4, 6, 8, 10)),]
```

```
answer_POe <- mean(treatment_POe$y1 - control_POe$y0)
answer_POe
```

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

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

Answer: The estimate for ATE has the same sign, but is slightly larger in magnitude than “the truth” (but not practically significantly different). Intuitively, there is a difference because we are averaging over a smaller amount of samples in the two groups. So outliers will have a bigger effect averaged on five samples vs averaged on 10 samples hence why the ATE of playing outside with controlling for hours has a larger magnitude than the answer from question 4d.

g. 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)?

Answer: \*Note made a loop starting from 10 choose 1 all the way up to 10 choose 9 in order to calculate the total number in each instance (to sum later) of different ways to split the groups up into control and treatment with always one needing at least one in control group. N choose K -> from 10 samples (N) choose (1, 2, ..., 9) for the control (K).

```
sum_group_create_ways = function(total, minValue){
  temp_list = c()
  while(minValue <= 9) {

    temp_number = (factorial(total) / (factorial(minValue) * factorial(total - minValue)))
    # print(temp_number)
    temp_list = c(temp_list, temp_number)
  }
}
```

```

    minValue = minValue + 1
  }
  return(temp_list)
}

answer.P0g <- sum(sum_group_create_ways(10, 1))
answer.P0g

## [1] 1022

```

- h. 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.

Answer:

```

treatment_P0h = d[which(d$child %in% c(1, 2, 3, 4, 5)),]$y1
control_P0h = d[which(d$child %in% c(6, 7, 8, 9, 10)),]$y0

answer.P0h <- mean(treatment_P0h) - mean(control_P0h)
answer.P0h

## [1] -0.44

```

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

Answer: The value from (h) is much larger in magnitude than the true ATE from (a), but with same sign (negative). The difference is that we can see the delta from the same subject if they were in treatment or in control in question 4a. so in 4(h) we are averaging the differences between the treatment and control (true apples to apples comparison), but the true ATE experiment does not work in reality because in reality we can not be in two different groups at the same time for this experiment. The answer from question 4h takes the average of a group of kids who played over 10 hours per week and subtracted the average of a group of kids who played under 10 hours per week and computed the difference. We have no idea what would have happened if the kids were in the other group. This situation has selection bias in it (convenience picking 1-5 and 6-10) with no mention on how the kids where numbered (was it random selection or not?) and also averaged over 5 people instead of 10 like in the perfect experiment hypothetical case. Therefore we are not comparing apples to apples and have no experimental mechanism to help out with approaching an apples to apples comparison.

## 5. FE, exercise 2.5

*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”)*

- (a) Discuss the strengths and weaknesses of each approach

Answer:

Method 1: Coin flipping

Strengths:

- This is truly a independent random experiment, with no circumstances from the past effecting the current assignment
- You can get the highest number of total volunteering :  $6 * 60$  or 360 minutes

Weaknesses:

- You can get the lowest number of total volunteering:  $6 * 30$  or 180 minutes
- You can get unequal amount subjects in 30 minute bin and 60 minute bin
- Highest variability in results if repeated (expected value isnt reached every time the study is ran)

Method 2: Card Shuffling

Strengths:

- You will have a consistant amount of total volunteering time:  $3 * 30 + 3 * 30 = 270$  minutes everytime
- expected value = total time of volunteering everytime
- Will never get the minimum value you could obtain from method 1

Weaknesses:

- Order of operation makes a difference on volunteering time given to participant
- Will never get a higher amount of total volunteering time

Method 3 is same results as method 2, method 3 just has more necessary steps in order to perform the task then method 2 (i.e put paper with number on it into an envelope and seal it vs just writting down the number)

- (b) In what ways would your answer to (a) change if the number of subjects were 600 and not 6.

Answer: The strengths and weaknesses do not vary based on the size (scale) all that much. The expected value of assigned number of minutes per participant is still the same. The only thing to consider with these different methods is: can you take on the risk of method 1 in order to get total volunteering amount which is above the expected value for assignmend number of minutes at the risk of getting an assigned volunteering minutes lower than expected value

- (c) What is the expected value of  $D_i$  (the assigned number of minutes) if the coin toss method is used? What is the expected value of  $D_i$  if the envelope method is used

```
# coin toss
coin_toss_expected_value = (1/2) * 30 + (1/2) * 60
coin_toss_expected_value
```

```
## [1] 45
```

```
# Envelope
envelope_expected_value = (1/2) * 30 + (1/2) * 60
envelope_expected_value
```

```
## [1] 45
```

## 6. FE, exercise 2.6

Many programs strive to help students prepare for college entrance exams, such as the SAT. In an effort to study the effectiveness of these preparatory programs, a researcher draws a random sample of students attending public high school in the US, and compares the SAT scores of those who took a preparatory class to those who did not. Is this an experiment or an observational study? Why?

Answer: This is an observational study. Even though a researcher randomly selected a sample of students, there is no intervention inside of the study and merely just binned two populations and sampled randomly from the two groups. In order for this to be a experiment, the researcher would have had to randomly select from the entire population of students who attend public school and then assign the random individual to either a control (no SAT prep) or the treatment (SAT prep). The researcher instead compared SAT scores

from participants who had the drive to study additionally to normal school studies with participants who had no desire to take additional practice or SAT prep, leading to a self selection bias.

## 7: Skip in 2017

## 8. FE, exercise 2.9

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.

- a. Critically evaluate this assumption.

Answer: Just because the lottery chooses winners at random doesn't mean that the amount that people have reported winning is random. There are all sorts of values attached to a various different kind of lottery. Many times people won't play the lottery until it hits a certain amount (like for instance me, I won't buy a ticket until the jack pot is above 500 million). In addition there are a number of lottery prizes that have max values, like scratch off lottery. Some have a max payoff of \$1000, \$10,000, \$50,000, etc, and very specific non jackpot winning amounts if you got anything that lined up with a winning ticket (i.e. 2/10 letters on a scratch off then you get \$2 back). Participants of the lottery could also just play one type of lottery game, and therefore the winning will not be random, and the outcome of the lottery win is inside the range of possible winnings from that specific lottery game. There are a lot of predefined fixed values with a predefined fixed amount in distribution which makes this variable less random than the researcher above believes it is. In addition scratch of tickets are given out in series, you can't get to the 10th card until you go through 1-9.

- b. 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?

Answer: There is a lot more people who lose the lottery or win an amount under \$10,000 than win \$10,000 or more, so the potential outcomes of views about the estate tax will probably be different amounts the group. Outliers will have more effects in the smaller group, which also won more money so are intrinsically different in attitude than the group who won a little or nothing. For example If I give you 100 dollars for free and someone comes by and takes half, you are upset but still probably can go out for dinner or fill your tank of gas up and continue with life as nothing ever happened. However, if I give you \$10 million dollars and someone comes around and takes \$5 million of it, you still have \$5 million dollars and can still buy probably everything you ever dreamed about with and still have some money in the bank for retirement, however you will complain and fixate on how you lost \$5 million dollars and what could have been. The same can be said about the attitude of the estate tax (government taking ~ 50 percent of winnings) and therefore this task has intrinsic bias built into the interview responses There is too much heterogeneity of humans characteristics and beliefs and too much bias inside the observation in order to assume the two groups view on estate tax is similar.

### *Clarifications*

1. 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?"
2. 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  $E[Y_i0|D = 1] = E[Y_i0|D = 0]$ , comparing what would have happened to the actual winners, the  $|D = 1$  part, if they had not won, the  $Y_i(0)$  part, and what actually happened to those who did not win, the  $Y_i(0)|D = 0$  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?"

3. Assume lottery winnings are always observed accurately and there are no concerns about under- or over-reporting.

## 9. FE, exercise 2.12(a)

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.

- a. 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.

Answer: Looking at equation 2.15 in the FE book, it states that  $E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]$  is equal to selection bias. Under random assignment,  $E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]$  is equal to zero. However, in this situation, there was no random assignment or experiment at all involved in making all prisoners read over three hours. In fact, it was an observation from a group (with some self selection bias inside it) that was applied to the entire population. Because of no experiment with intervention and no random selection, the statement that  $E[Y_i(0)|D_i = 0] = E[Y_i(0)|D_i = 1]$  and  $E[Y_i(1)|D_i = 0] = E[Y_i(1)|D_i = 1]$  is not an accurate statement.