

# Midterm

Petra Ivanovic - pi2018 - Section 005

Due Oct 25, 2023

This midterm must be turned in on Brightspace by Oct 25, 2023. It must be your own work, and your own work only – you must not copy anyone’s work, or allow anyone to copy yours. This extends to writing code. You **may not** consult with others. All work must be independent.

Your homework submission must be written and submitted using Rmarkdown. No handwritten solutions will be accepted. You should submit:

1. A compiled PDF file named yourNetID\_solutions.pdf containing your solutions to the problems.
2. A .Rmd file containing the code and text used to produce your compiled pdf named your-NetID\_solutions.Rmd.

Note that math can be typeset in Rmarkdown in the same way as Latex. Please make sure your answers are clearly structured in the Rmarkdown file:

1. Label each question part
2. Do not include written answers as code comments.
3. The code used to obtain the answer for each question part should accompany the written answer. Comment your code!

## Problem 1 (25 points)

A cafe is testing out a promotion set to determine which pastry goes well with their new espresso blend. Customers are told that the promotion set is \$5 for a cup of espresso and a random pastry item. After receiving the promotional set, they are asked to rate the product. There are two types of pastries: a sweet scone and a savory bagel, customers are randomly assigned to receive either type. Let  $D_i = 1$  if the customer receives the bagel (the “treatment”) and  $D_i = 0$  if they receive the scone. Let  $Y_i$  denote the observed rating from the  $i$ th customer.

### Part a (12 points)

In your own words, explain what the following quantities represent in this setting and indicate whether this quantity is observable without making assumptions: (4 points each)

1.  $Y_i(1)$  represents the bagel rating of the customer  $i$ . We can observe this only for those customer that have actually received the bagel but cannot observe it without making assumptions for those that have not received a bagel. Thus it is not observable for all customers without making assumptions.
2.  $E(Y_i(1)|D_i = 0)$  represents the expected rating of a bagel by customer  $i$  given that they received a scone (and not a bagel). We cannot directly observe this without assumptions as the customer did not receive the treatment whos outcome we are trying to observe. We can never (unfortunately) observe what the outcome would be if the customer received the alternate treatment (fundamental problem of causal inference.)
3.  $E(Y_i|D_i = 0)$  represents the expected rating from customer  $i$  given that they received a scone. This can be observed for all the customers that received a scone (instead of a bagel).

### Part b (4 points)

Suppose we have 6 customers who bought the set this morning, the observed randomization and potential outcomes are:

Customer	$D_i$	$Y_i(1)$	$Y_i(0)$
1	1	5	5
2	1	9	5
3	0	8	6
4	0	4	1
5	1	8	5
6	0	7	5

Write down the individual treatment effects (ITE) and observed outcome for each customer.

- Individual Treatment Effect (ITE) can be defined as a difference between individuals (customers) outcome if they received the treatment and if they had not. Thus for our problem it would be the difference between ratings if same customer received a scone vs a bagel. To put it mathematically,  $ITE = Y_i(1) - Y_i(0)$ .
- Lets calculate this for each customer:
  - $ITE_1 = Y_1(1) - Y_1(0) = 5 - 5 = 0 \rightarrow ITE_1 = 0$
  - $ITE_2 = Y_2(1) - Y_2(0) = 9 - 5 = 4 \rightarrow ITE_2 = 4$
  - $ITE_3 = Y_3(1) - Y_3(0) = 8 - 6 = 2 \rightarrow ITE_3 = 2$
  - $ITE_4 = Y_4(1) - Y_4(0) = 4 - 1 = 3 \rightarrow ITE_4 = 3$
  - $ITE_5 = Y_5(1) - Y_5(0) = 8 - 5 = 3 \rightarrow ITE_5 = 3$

$$- ITE_6 = Y_6(1) - Y_6(0) = 7 - 5 = 2 \rightarrow ITE_6 = 2$$

- Observed Outcome for each customer depends on what their treatment was, for instance if  $D_i = 1$  observable outcome is  $Y_i(1)$  and if  $D_i = 0$  observable outcome is  $Y_i(0)$ . More specifically, observable outcome is the rating of the treatment (bagel or scone) that the customer receives (only one, not both!).
- Lets calculate this for each customer:
  - Observed Outcome 1: 5
  - Observed Outcome 2: 9
  - Observed Outcome 3: 6
  - Observed Outcome 4: 1
  - Observed Outcome 5: 8
  - Observed Outcome 6: 5

### Part c (4 points)

Estimate the difference in means (treatment - control) in this case using the table in part b, assuming consistency holds. Is this quantity equal to a causal effect in this case? Why or why not?

- To estimate the difference of means we first need to calculate the mean for both groups, treatment and control:
- Treatment( $D_i = 1$ ) group mean =  $\frac{5+9+8}{3} = \frac{22}{3} = 7.33$
- Control( $D_i = 0$ ) group mean =  $\frac{6+1+5}{3} = \frac{12}{3} = 4$
- Difference in Means = Treatment group mean - Control group mean
- Difference in Means =  $7.33 - 4 = 3.33$

We have calculated that our difference in means is 3.33 and we can see that bagels have received higher ratings.

There are multiple variables that we have to take into account to make a conclusion if this difference in means is equal to our causal effect: - 1. Positivity: Positivity means that there is a non-zero probability that each subject/individual will receive either treatment or control. In our case, this means that there is a non-zero probability that each customer will receive the bagel or the scone. As we learned from the question set-up this is true as every person who purchases the promotion set will be randomly assigned a pastry (with non-zero probability) and thus this is satisfied. - 2. Ignorability: Ignorability means that “Treatment assignment is independent of potential outcomes” in other words the way we distribute our treatment or control - bagel or scone is independent of ratings. This is true and it holds as we are randomly assigning the treatment without knowing customers preferences for pastries (unlike in part D). - 3. SUTVA: SUTVA means that there is no spillover in our experiment, one unit's treatment does not influence another unit's outcome, as well as that there is a single version of treatment. In our case, single version of treatment means that all bagels assigned will be the same (no different flavours/toppings) and all scones will be the same. We can assume this is true as it was mentioned that we are only trying out 2 pastries (sweet scone and a savory bagel) so we know that this holds as well. As always, spillover can get a bit tricky, but assuming that customers are coming on their own (in USA coffee shops are places where people get food to go more than to stay and share) and not sharing their pastries with other participating customers and thus not biasing their rating we can say that spillover is not happening and thus SUTVA is satisfied.

Additionally, it is worth noting that there might be other external factors that could bias our study such as if some customers were more hungry when they received the treatment so that is why they rated higher than others. However, this kind of bias should be eliminated since we are using randomization especially if we have large number of subjects (customers).

Since all of these assumptions hold we can say that difference in means  $E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$  is equal to ATE  $\tau = E[Y_i(1)] - E[Y_i(0)]$ .

#### Part d (5 points)

The cafe hired a new barista who is very considerate. She asks each customer whether they prefer sweet or savory things, and then gives them their preferred pastry item with their espresso. Is it possible to estimate the average treatment effect of getting the bagel on ratings with data collected after this new barista was hired? Why or why not?

- No, it is not possible to estimate the average treatment effect of getting the bagel on ratings with data collected after this new barista was hired. While this barista might be good for the business since customers will most likely get the pastry they like more and will come back again, she is very bad for our experiment. This idea of people getting the pastry they prefer based on their sweet vs savory preference is strongly biasing (selection bias) our experiment as the treatment is no longer being assigned randomly. In other words, the difference in ratings is no longer based only on the pastry quality but also the differences in preferences of the customers in treatment and control groups. Thus, we do not know if the ratings are due to the pastry quality or the fact that the customers are getting the pastry they like more. If we wanted to estimate the causal effect correctly we need to make sure that the treatment is **randomly** assigned to customers regardless of their preferences for sweet or savory. Doing this makes it so that we can assume both groups are identical in all manner appart the treatment they are receiving and thus we can compare their ratings and estimate the causal effect of the treatment (bagel).

## Problem 2 (25 points)

The STAR (Student–Teacher Achievement Ratio) Project is a four-year longitudinal study examining the effect of class size in early grade levels on educational performance and personal development (whether they finish high school). A longitudinal study is one in which the same participants are followed over time. This particular study lasted from 1985 to 1989 and involved 11,601 students. During the four years of the study, students were randomly assigned to small classes, regular-sized classes, or regular-sized classes with an aid. In all, the experiment cost around \$12 million. Even though the program stopped in 1989 after the first kindergarten class in the program finished third grade, the collection of various measurements (e.g., performance on tests in eighth grade, overall high-school GPA) continued through to the end of participants' high-school attendance.

The variables of interest are:

1. classsize - Treatment variable - size of class before the fourth grade.
2. sex
3. race
4. g4math - total scaled score for the math portion of the fourth-grade standardized test
5. g4reading - total scaled score for the reading portion of the fourth-grade - standardized test
6. gpa - high school gpa
7. grad - finish high school, 1 yes, 0 no

### Part a (8 points)

Consider the variables *sex*, *classsize*, *gpa*, and *grad*. Draw a DAG representing the causal relationship between them in this experiment.

```
library(dagitty)
# ?dagitty(): See the help file

DAG <- dagitty('dag{ ClassSize -> GPA ClassSize -> Grad GPA -> Grad Sex -> GPA Sex -> Grad}')
coordinates(DAG) <- list(x=c(ClassSize=0, GPA=2, Grad=3, Sex = 0),
                        y=c(ClassSize=0, GPA=1, Grad=0, Sex=1))

plot(DAG)
```



- We are using class size as treatment and we want to see its impact on students graduation status and GPA. We are thus creating a relationship between ClassSize and Graduation Status and ClassSize and GPA. Further, we can assume that students GPA has something to do with their Graduation Status, for instance higher GPA should mean a person is going to graduate and low GPA means they might not be able to graduate. Thus, we are drawing a relationship between GPA and Grad.
- Finally, we are considering Sex a variable that influences both of the outcomes we would like to observe, thus Sex is a covariate and we are drawing a relationship between sex and graduation status and sex and GPA.

### Part b (10 points)

Suppose in the experiment, the researcher found out the CATE for female students is different from CATE for male students. We want to know whether these two CATEs are statistically different from each other. Can we conclude anything about this from the fact that one of them is statistically different from zero and the other is not? Why or why not?

- CATE (Conditional Average Treatment Effect) refers to average effect of our treatment (class size) on the outcome (GPA and Graduation status) for a specific subgroup of our 'subjects' (students).
- If we want to know if our CATE for female students is different from CATE for male it is not enough to know that one of them is statistically different from zero and the other is not. If, for instance, CATE for female students is statistically different from zero this means that there is enough evidence to say that the effect of class size is statistically significant for female group. Further, if then CATE for male students is not statistically different from zero this does not mean the treatment is (definitely) not effective but just that we have not found enough evidence to conclude that (there might be other reasons for this such as too small sample size.) Further, these don't tell us anything about the relationship

among CATE for female students and CATE for male student, for instance they could be very similar but one of them just being over the limit so it counts are statistically different from zero and other being just below the limit and thus is not statistically different from zero.

- If we wanted to know if CATE for female students and CATE for male students are statistically different from each other we would have to conduct a hypothesis testing using statistical test (using t-statistic or z-statistic) and p-value and then comparing p-value to our significance level (e.g. 0.05) and then if  $p\text{-value} < \text{significance level}$  we can conclude that CATE for female students and CATE for male students are statistically different.

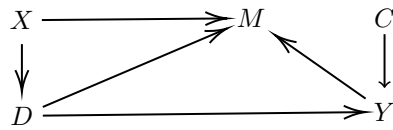
### Part c (7 points)

Imagine we wanted to estimate the effect of class size on finishing high school in this experiment. What would be necessary for you to control to estimate an unbiased treatment effect? How would you estimate the treatment effect? Explain your answer.

- We define confounders as variables that influence both our treatment and our outcome and we define colliders as variables that are influenced by both our treatment and outcome. As there is no variable that influences our treatment, class size as it is randomized, there is no confounder and thus there is nothing for us to control for to get more unbiased effect. Further, we can notice that there is also no colliders as there is nothing that is being influenced by our outcome - graduation status (we don't control for colliders but just mentioning this for the sake of analysis).
- Since there are no confounders to control for (at least among the mentioned variables) we would expect that we can get an unbiased effect of our treatment, class size on our outcome, graduation status / rate, without controlling for any variables. However, in order to increase the precision (and with that reduce the standard error) of our causal estimate we can control for race and sex since those might/do have influence on our outcome variable. In our problem set 1 we saw that there is a disparity in graduation rates when controlling for race and we can suspect same might be the case for sex. Thus taking into consideration these variables can help us potentially decrease the standard error.
- We should not control for GPA, g4math and g4reading score as they can be considered mediators (can be descendants / caused by class size) and we should not control for those.

### Problem 3 (25 points)

Consider the following Directed Acyclic Graph:



#### Part a (15 points)

List all of the paths from D to Y. On each path, identify confounders and colliders.

PATH 1:  $D \rightarrow Y$

- The only direct path from D to Y in this diagram is the direct path  $D \rightarrow Y$  (direct effect). There are no indirect (directed) paths / effects from D to Y. We can notice that apart from this path X has an influence on D and C has an influence on Y and both D and Y (and X) have an influence on M.
- We define confounders as variables that influence both our treatment and our outcome and as such, in DAGs we would expect to see paths from a confounding variable to both D and Y. Since there are no such variables we can say that according to this DAG there is no confounder on our path from D to Y.
- We define colliders as variables that are influenced by both our treatment and outcome and as such we would expect to see arrows from both pointing to the same variable (collider). We do see this in our graph since both D and Y point to M, and thus M is the collider.
- It is worth to note that while C is not a collider or confounder it can have significant influence on our outcome as C is a parent of Y (C is a covariate) which can make our point that D caused Y hard to prove as C has only influence on Y if we don't do something about that (if we do not control for it).

PATH 2:  $D \leftarrow M \rightarrow Y$

- $D \leftarrow M \rightarrow Y$  is an indirect (backdoor) path from D to Y over mediator M.
- We can see that there is no variable that influences both D and Y and thus there is no confounder on D and Y. However, there is a confounder X on the path from D to M (on the way to Y) as it points (influences) both D and M.
- There are no variables that are influenced by D and Y or D and M and thus there are no colliders on this entire path.

PATH 3:  $D \leftarrow X \rightarrow M \leftarrow Y$

- $D \leftarrow X \rightarrow M \leftarrow Y$  is also an indirect path from D to Y, this way with 2 mediators, X and M.
- We can see that from D to X there is no variable that influences both D and X, same for from X to M and M to Y and thus there is no confounders on this path. It is worth noting that, as shown in path 2, if we are considering the path from D to M, X is a confounder on it as it influences both D and M.
- Finally, we can notice that there are no variables that are influenced by both D and M so on that part of the path there are no colliders. However, we can also notice that D and X are both influencing M as well as D and Y, so M could be considered a collider on both of those parts of the path.



**Part b (10 points)**

Are there any variables that we should condition on in order to identify the causal effect of D on Y? Explain.

- Generally, we need to condition on confounders (back-door paths) because they are the variables that influence both our treatment and output. We do this because conditioning on the confounder “blocks” the back-door path and then we can understand the effect of our treatment on the outcome. As there are no confounders in this DAG (no variables that have influence (arrows towards) on both D and Y). Further, we know that we should never condition on colliders (M) as that could have an opposite effect on our outcome by introducing bias rather than minimizing it.
- However, if we wanted to increase the precision (and with that reduce the standard error) of our causal estimate we can control for C as it has influence on Y but it is not a confounder so we should not condition on it!
- Finally, as we saw in the example in class about being paid to exercise, our pre-treatment variables can also influence our post-treatment outcome, for instance if someone exercised before, regardless of our treatment they will most likely keep exercising. Further, we saw that stratified standard error (when accounting for pre-treatment variable) was lower than unstratified standard error, this means that (post)-stratification here reduces the estimated standard error, narrowing the confidence interval and improving precision. Thus we can conclude that depending on our experiment/study just like with controlling for C we might consider controlling for X as this can reduce SE and with that increase precision.

## 4 Design Your Study (25 points)

Design your own study from start to finish. Choose an *interesting* question that we have not mentioned in class. Answer the following questions: (1) Explain the effect you wish to estimate in words and why you think it's interesting. Carefully explain both your treatment, outcome, and the research question you wish to answer. (2) What is the "ideal experiment" for your question? (3) Draw the ideal experiment in a DAG. Can you estimate the effect of your treatment on your outcome? Is it identifiable and how do we know? (4) If you were to collect observational data on this topic, what potential confounders and mediators would exist? Please explain them in words. (5) Draw out a DAG that corresponds to this observational study. Please include at least one confounder and one mediator. (6) Using the DAG you drew in question 5, can you estimate the impact of your treatment on your outcome? Is the effect identifiable? Explain why or why not.

\*Note: You cannot reuse an example we went over in class nor an example you used in a previous problem set.

1.

As a computer and data science senior, I have been trying to find time to practice LeetCode questions for technical interviews but it is very hard to regularly find time to do them. Due to this, I want to design a study that estimates how regular LeetCode practice (let's say a problem per day) influences the hiring result (e.g. 1 = hired, 0 = not hired). As we all know the current economical climate is making it very hard to find a job in technical fields such as software engineering or data science. I think it would be interesting if we were able to pinpoint how this one variable, regular LeetCode practice, influences job outcomes. The results of this study can be used to help people understand the importance of regular practice (or lack thereof) while they are looking for a job.

- Treatment: Regular LeetCode practice (e.g. 1 specified (same for all participants) question per day or weekly equivalent during the period of the study)
- Outcome: Securing a job - binary outcome, 1 = hired, 0 = not hired
  - $Y_1$  Recruitment outcome (hired/not hired) for an individual who regularly practices LeetCode questions
  - $Y_0$  Recruitment outcome (hired/not hired) for an individual who does not regularly practice LeetCode questions
- Research Question: Does regular practice of LeetCode questions increase the chances of securing a job in a technical CS/DS field?

2. Ideal Experiment:

Goal:

Determine if regular LeetCode practice increases the chances of finding a job in a technical field. (such as Software Engineering or Data Science) In an experiment setting we would want to focus on one specific category as the test might differ, so let's focus on the outcomes of finding a job in Software Engineering specifically. We would also want to focus on candidates with similar amount of experience so let's only use students in their junior year of 4 year college.

Treatment and control assignment:

In an ideal experiment we would **randomly** assign a group (as large as possible but preferably 100 or more) of job seekers in Software Engineering field that are in their junior year of 4 year college in 2 groups:

- Treatment group: individuals in treatment group would be given a certain number of LeetCode questions they must answer/complete to best of their ability on regular basis during the entire duration of the experiment, for instance 1 specified problem per day or 7 specified problems per week.

- Control group: individuals in the control group would not be given any LeetCode questions and would not be encouraged (discouraging them or making sure they are not practicing would be even better but it might be an ethical issue) to do them on regular basis at all during the experiment period.

Throughout the experiment period we would remind our treatment group to make sure they are doing the practices as ‘prescribed’.

We would put both of these groups of individuals through the job application and general interview process to see how they perform by keeping track how many job offers members of each group receive.

Since our assignment would be randomized in perfect conditions we would not have to worry about confounders such as some candidates being more prepared/experience to begin with than others. Randomization would take care of this by assigning people in both groups regardless of their experience.

At the end we would have (hopefully) unbiased data of how big effect regular LeetCode practice has on recruitment results.

Potential Issues:

If some of the participants in the treatment group were not doing the LeetCode practice regularly as ‘prescribed’ this could cause the experiment to be much less valuable and correct/precise. Same would happen if members of the control group were practicing LeetCode questions even though they were discouraged of it. As mentioned some confounders could also be that some candidates have more experience/knowledge or even connections but those should be taken care of by using randomization. Finally, we would have to consider if the experiment is ethical as discouraging students for practicing LeetCode questions might have detrimental result on their careers.

To sum up, an ideal experiment would make sure only the randomly assigned treatment group is practicing LeetCode questions regularly and control group is not with a goal to isolate the effect of LeetCode practice from other factors that might influence the hiring process. However this can be very hard to achieve in real life.

3.

```
library(dagitty)
# ?dagitty(): See the help file

DAG3 <- dagitty('dag{ LeetCode_Practice -> Recruitment_Outcome}')
coordinates(DAG3) <- list(x=c(LeetCode_Practice=0, Recruitment_Outcome=2),
                          y=c(LeetCode_Practice=0, Recruitment_Outcome=0 ))

plot(DAG3)
```



In an ideal experiment (very hard/unrealistic to achieve in real world) there would be no confounders or other variables influencing our treatment and outcome as they would have been isolated due to randomization and previously mentioned experiment rules. Thus, our effect of regular LeetCode Practice on the Recruitment Outcome would be identifiable as we could compare how many of our candidates that regularly practiced LeetCode questions secured jobs and how many of those that did not regularly practice secured jobs.

4.

As mentioned in the part 2, there are many variables that could cause ones Recruitment Outcome regardless of their LeetCode practice habits. There are many variables that could influence our outcome such as pre-existing knowledge, pre-existing experience, professional network and connections, educational background, motivation and most likely a many of other factors however not many of those are confounders but rather variables that only influence our outcome and not our treatment as well. Lets look at some potential confounders and mediators:

Potential Confounder(s):

- Time availability / number of hours of classes taken -> amount of free time student has will influence how many LeetCode questions they do and if they do them regularly, in similar manner amount of free time also influences whether students apply for jobs as they do not have time to write cover letters and put applications in if they do not have any free time
- Motivation -> some students might be more motivated than others, for instance international students might be more motivated to practice LeetCode questions as they know they can lead to better results on technical exams which can help them with recruiting. In similar manner, they will be applying for more job opportunities (and potentially putting in more detail focused applications) since they know that if they do not find a job by the time they graduate they will not be able to stay in US.

Potential Mediator(s):

- GPA -> for a Software Engineering / CS student, more and regular LeetCode practice could cause higher GPA which could in turn help them / cause them to get hired.
- Improvement in Coding skills -> improved coding skills (and generally problem solving) are kind of an implied result of practicing LeetCode questions (assuming people are not memorizing them or not paying attention to how to do them efficiently). Further, since technical (coding and conceptual) exams are a big part of recruitment for Software Engineers improved coding skills can help a lot.

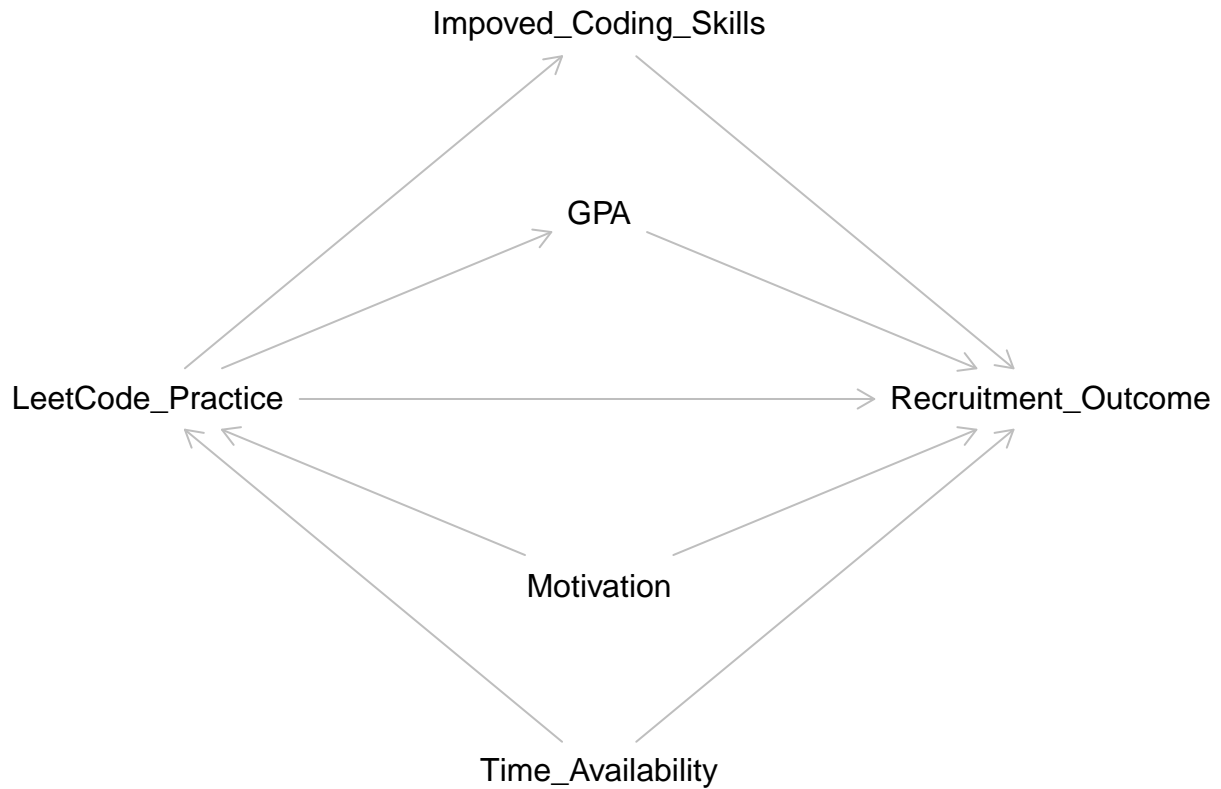
5.

```
library(dagitty)
# ?dagitty(): See the help file

DAG3 <- dagitty('dag{ LeetCode_Practice -> Recruitment_Outcome LeetCode_Practice -> GPA
                    GPA -> Recruitment_Outcome Motivation -> LeetCode_Practice
                    Motivation -> Recruitment_Outcome Time_Availability -> LeetCode_Practice
                    Time_Availability -> Recruitment_Outcome
                    LeetCode_Practice -> Improved_Coding_Skills
                    Improved_Coding_Skills -> Recruitment_Outcome}')

coordinates(DAG3) <- list(x=c(LeetCode_Practice=0, GPA = 1, Recruitment_Outcome=2,
                             Time_Availability = 1, Motivation = 1, Improved_Coding_Skills = 1),
                        y=c(LeetCode_Practice=1, GPA=0, Recruitment_Outcome=1,
                             Motivation=2, Time_Availability= 3, Improved_Coding_Skills= -1))

plot(DAG3)
```



6. Using the DAG you drew in question 5, can you estimate the impact of your treatment on your outcome? Is the effect identifiable? Explain why or why not.

If we want to estimate the impact of our treatment of the outcome in an observational study shown above, we need to make sure we are controlling for confounders and adjusting for all other variables that might influence our outcome. Further, to make sure the difference in means of our observed outcomes is equal to ATE (and with that identifiable) we need to make sure that the positivity, ignorability, and SUTVA are also satisfied.

- To learn more about causal effect in this study we want to learn more about Average Treatment Effect or 'ATE'. To compute ATE,  $ATE = E[Y_1 - Y_0]$ , we need to observe  $E[Y_1]$ , average outcome for those who regularly do LeetCode practice (treatment) and  $E[Y_0]$ , average outcome for those who do not regularly do LeetCode practices (control). However, since unlike our ideal experiment, observational study has confounders and mediators, thus these observed outcomes may not give us an unbiased estimate of ATE. To achieve this unbiased estimate, we need to control for all the confounders. In our case this means we would need to control for Time Availability and Motivation and in that way block backdoor paths between our treatment and control. Controlling for these confounders would enable us to estimate the impact of our treatment on the outcome. This can be done as we have identified our confounders so we know what to control on when conducting the study, thus, we can estimate the impact on the outcome. One thing to note is that we need to make sure that we are controlling for all confounders we can think off (2 were mentioned but we should try to think if there are any others that have not been mentioned).
- However, if we want to make sure our effect is identifiable we need to identify our assumptions of Positivity, Ignorability and SUTVA and prove they can be established / valid.

- **Positivity:** Positivity means that there is a non-zero probability that each subject/individual will receive either treatment or control. In our case this means that there is a non-zero probability for each student that they will either be assigned to practicing LeetCode regularly or they will not. This condition is satisfied as we will be randomly choosing students that are and are not regularly practicing to both groups regardless of any preexisting conditions or opinions and thus there is a 50 - 50 chance for each student to be assigned to either group.
- **Ignorability:** Ignorability means that “Treatment assignment is independent of potential outcomes” in other words the way we distribute our treatment or control is independent of our outcome. Ignorability can be a bit harder to achieve as we cannot place students in random groups ourselves (in the matter we would in an experiment) but we can make sure to get a very large and diverse set of students. I believe ignorability can be achieved with large enough sample since we can group subjects solely based on the treatment and no other variable. I believe if we choose participants randomly many of their similarities will cancel out (or we will control for them as cofounders) as for instance there are many very smart students that practice LeetCode a lot and do not get hired regardless but there are also very smart students that do not practice LeetCode a lot and do get hired. Thus, with large enough sample we could make sure to have people with different backgrounds in all the groups regardless of their outcomes. We should remain aware of these characteristics as they may be confounders and thus we should for them.
- **SUTVA:** SUTVA means that there is no spillover in our experiment, one unit's treatment does not influence another unit's outcome, as well as that there is a single version of treatment. In our case there is only one unit of treatment as all the participants will be solving specified LeetCode questions on a specified (daily/weekly) basis thus the single unit of treatment is satisfied. If we cannot ensure that they are practicing exactly the same questions we can make sure that the average level of difficulty (since LeetCode questions are grouped in easy, medium, and hard) of the questions participants completed is approximately equal for those that we admit in the study, and with this ensure we still have a single unit of treatment. Further, we can choose all of our participants randomly (those that fit out criteria) from different schools and assuming there are not a lot of students that are in the same class (shouldn't happen as we can get students from all over the country) we can assume that there is no (significant) spillover. Further, this is especially true since the outcome of the experiment is not going to be immediate as recruiting is a long process and regular LeetCode practice is just a small part of the process so even if students discussed the fact they were doing regular LeetCode practice they will not know the direct outcome and with that cannot influence others. Thus, SUTVA is also satisfied.

Since we have shown we can conduct the experiments so that the Positivity, Ignorability and SUTVA hold and we can condition on all our confounders (we do not condition on mediators), we can say that we can estimate the impact of our treatment of the outcome and that the effect is identifiable.