*YOUR TA’S NAME*:

*Lecture Worksheet*

*Thursday 11/19/2020*

**MAIN POINTS OF LECTURE**

1. In establishing the causal impact of X on Y we have—to this point—only thought of third variable(s) Z as inducing spuriousness … a bad thing
2. This suggests that to estimate the true, causal impact of X on Y we should always “control” (or “adjust”) for Z (Note: “Controlling” for Z means assessing the association between X and Y within each level of Z to estimate the causal impact of X on Y.)
3. In fact there are a variety of ways in which X, Y, and Z might be related to one another
4. **Theory** and **prior evidence** should guide our decision about how Z plays into the relationship between X and Y
5. **Depending on our theoretical understanding of how Z plays into the relationship between X and Y, we might or might not want to statistically control for Z**
6. In assessing the causal effect of X on Y, another variable Z might be…

… totally orthogonal to the association between X and Y

… a confounding variable which must be controlled? (Note: A confounding variable is one that indices spuriousness in the association between X and Y)

… a mediating variable which we might or might not want to control? (Note: A mediating variable is a mechanism or causal process through which X affects Y)

… a moderating variable which we must treat with care? (Note: If the effect of X on Y differs across levels of Z, then Z is a moderating variable)

… a variable that it would be a mistake to control for?

1. Should we control for Z? How do we interpret the association between X and Y that remains after controlling for Z? It depends on the situation … so we should design analyses on the basis of **theory** and on evidence from **prior research**

**QUESTIONS**

1. [From the recorded lecture] **What are some potential confounders and some potential causal mechanisms linking X to Y in the following examples?**

Kids who frequently misbehave in high school (X) are less likely to go to college (Y)

Confounders: Gender; family socioeconomic origins; ADHD or similar issues

Mechanisms: Grades; High school graduation

Men who are married (X) live longer (Y) than men who are not married

Confounders: Income; Health; Family socioeconomic origins

Mechanisms: Stress; Loneliness; Happiness

Eating breakfast (X) improves employees’ productivity (Y) during the day

Confounders: How late people stayed out the night before? Hard to think of many.

Mechanisms: Blood sugar levels; Sleepiness; Grouchiness

1. [From the recorded lecture] **How do you interpret the following results?**

In the association between whether high school students work at paid jobs during the school year (X) and whether they drop out of high school (Y), gamma is 0.25. After statistically controlling for children’s family income and wealth (Z), gamma is reduced to 0.20

Part of the association between X and Y is confounded by Z. That is, part of the association between whether high school students work at paid jobs and whether they drop out of high school is spurious; children’s family income and wealth affect both.

The correlation between parents’ wealth (X) and their children’s wealth (Y) is 0.40. After statistically controlling for children’s education (Z), the correlation between X and Y is 0.10

Z is a causal mechanism linking X to Y … that is, it is a pathway through which X affects Y. The children of wealthier parents go farther in school; kids who go farther in school become wealthier themselves.

In a regression of final exam scores (Y) on number of hours people studied for the exam (X), the slope is estimated to be 5.0. After statistically controlling for students’ year in college (Z), the slope is estimated to be 2.0 among freshman and sophomores, 5.0 among juniors, and 8.0 among seniors.

There is a statistical interaction such that the effect of X on Y differs across levels of Z. The effect of hours spent studying on final exam scores is lower for freshman/sophomores than it is for juniors and seniors. This makes sense; juniors and seniors probably study more efficiently.

1. [From the synchronous session] According to today’s class data: How much does the number of hours you spend per week on this class (X) affect your grade (Y)?

See the recording of the synchronous session to find out!