Notes 2: Observational studies, confounding

ECO 231W - Undergraduate Econometrics

Prof. Carolina Caetano

1 Experiments, continuation from last class

1.1 Review

We were talking about the two scientific questions:

- How much will attending one extra class impact the final grade?
- How much smoking one more cigarette per day during pregnancy impacts the baby's weight?

and we thought about two theoretically possible experiment designs:

1. Take two random groups of students. Force the first group to go to 10 classes, and the other to go to 11. The average difference in the grades across the groups has to be caused by the differences in classes, because the two groups are comparable.
2. Take two random groups of pregnant women. Force one to smoke 4 cigarettes, and the other to smoke 5 cigarettes daily throughout pregnancy. The difference in the average weight of the babies across the groups has to be caused by the differences in cigarettes, because the two groups are comparable.

These experiment designs suffer from two problems:

1. They are unethical.
2. There is no placebo.

The lack of a placebo turns out to be an important issue that we cannot really circumvent. However, we could try to modify such experiments so that they would be ethical by modern standards.

1.2 Real world experiments

In the best case scenario, this is what you might be authorized to do:

1. Take two random groups of students. Leave the first group alone. Then give the second group some incentive to go to one more class. For example, sing the praises of going to class, or give them a prize for going.

2. Take two random groups of pregnant women. Leave the first group alone. Then give the second group some incentive to smoke less (an Ethics Board would never approve an experiment where you incentivize women to smoke more.) Either tell them that smoking is bad, or give them a prize if they smoke less.

These seem to be acceptable solutions. Suppose that the students in the first group went to 10 classes in average. We could compare their grades to the students in the second group that did go to an extra class. Suppose the mothers in the first group smoked 4 cigarettes per day. We could compare them to the mothers in the second group that smoked one cigarette less. However, these designs are actually quite bad. The restrictions of reality completely ruined our ability to arrive at the causal effects we wanted to discover. Here are some reasons:

1. _____

   Who took the incentives, and who ignored them? The students that decided to go to one more class may be the best of the second group. Hence the difference between the groups can be due not to the extra class, but to the fact that we are comparing the first group against the very best of the second. The same way, the mothers that decided to smoke one less cigarette in the second group may be the ones that care the most about the pregnancy. Hence, the differences in birth weights may not be due to smoking amounts, but to the fact that we are comparing the first group to the mothers that care the most in the second (and therefore they may be more careful with their pregnancy in other ways).

2. _____

   The incentives themselves can have an effect. If you sang the praises of going to classes, students may have become more conscious of the importance of studying. If you told mothers they should not smoke, you may have made them guilty. The differences across groups may not be due to the class, or to the amount smoked. Instead, it could just be that the students in the second group began studying more, and that the mothers in the second group began to take better care of their pregnancy.

The lack of placebo, as well as the two reasons above are very common occurrences in human behavior problems.

1.3 Conclusion (about experiments)

Experiments are the ideal way in which to study causal questions. However, when studying behavioral questions, it is often impossible to design a feasible and ethical experiment which does not suffer from issues of _____

_____

2 Observational Studies

Experiments: _____

_____

Observational studies: _____

_____

- **Advantages of observational studies:**

  1. _____

     There are still concerns about privacy breach, embarrassment, and coercion to participate.

  2. _____

     It is often cheaper to acquire information on many people than to run large scale experiments.

- **Disadvantages of observational studies:** they have to be done extremely carefully to be useful. This entire course is about this.

Observational studies in the two problems we are discussing would take the form:

1. A group of students is observed. Data is collected on how many classes each student attends, and the final grade each student receives.
2. A group of mothers is observed. Data is collected on how many cigarettes the mother smokes per day, and also the baby's weight at birth.

We are tempted to compare those subjects that took the treatment with those that didn't. For example, compare the students that went to 11 classes with those that went to 10 classes. Analogously, we could compare the women that smoke 5 cigarettes daily with those that smoked 4.

- Can you find the causal effect of class attendance, or smoking in pregnancy, using this approach?

  _____

- Why?

  _____

  _____

  _____

- So what?

  _____

  If students that took 11 classes do better than students that took 10 classes, we cannot know if it is due to the classes, or to the inherent differences between the students that took different decisions. Analogously, if babies born to mothers that smoke 4 cigarettes are heavier than babies born to mothers that smoked 5 cigarettes, we cannot know if the difference was caused by cigarettes or by the inherent differences between the mothers who smoked 4 and the mothers who smoked 5.

- What differences are you talking about?

  For example, students that <u>choose</u> to go to more classes could be more responsible, or like the course material more, or find the material more understandable, or are more patient.

  _____

  _____

Confounders: _____

_____

A confounder must be: _____

1. _____

2. _____

Let's check why the following variables are possible confounders

1. Natural ability on econometrics is a confounder in the problem of finding the causal effect of class attendance on the final grade.

   (a) Students with more natural ability may _____

   _____

   (b) Students with more natural ability may _____

   _____

2. How much a mother wants the pregnancy is a confounder in the problem of finding the causal effect of smoking during pregnancy on the baby's birth weight.

   (a) Mothers that want the pregnancy more may _____

   _____

   (b) Mothers that want the pregnancy more may _____

   _____

   _____