Notes 1: Scientific questions, experiments

ECO 231W - Undergraduate Econometrics

Prof. Carolina Caetano

1 General Information

2 General Information

- Instructor: Carolina Caetano

- Class website: www.carolinacaetano.net/teaching/eco231w

  Everything you need to know will be on the website: assignments, weights, curriculum, sections, book, software, etc.

- In the first week of this course, you must take the following **mandatory** actions:

  1. Go to the class website and look for a tab in the top left-hand side where it is written "Check-in." This is the course syllabus. You must read it. A student that remains in this course after the first week acknowledges that they read that page and are therefore aware of all the course assignments, dates and resources available to them.

  2. Make a note of all the dates regarding this course. Everything is already scheduled. In particular, pay attention to the midterm dates, as you will need to come to class on those days. Here are the dates.

     - 02/16: Homework 1 due
     - 02/21: **Midterm 1**
     - 03/09: Homework 2 due
     - 03/23: Replication due
     - 04/04: **Midterm 2**
     - 05/04: Project due

  3. The official source of news and assignment rules in this course is the Announcements page: www.carolinacaetano.net/teaching/eco231w/announcements. It supersedes everything else you may think you have heard in class, that I or any

TA or student tell you. The rules that are on the Announcements page are the official rules. You must make sure to check this page frequently, or arrange to be notified when any change is made to it.

4. Go to the Downloads page: www.carolinacaetano.net/teaching/eco231w/download and download the document "Notes 2: Observational studies, confounding" to bring to the next class. Every class, you must print and bring the corresponding notes in order to fill out the blank spaces. If you are confused about which notes to print, check the Class Plan page (www.carolinacaetano.net/teaching/eco231w/classplan) in the course website.

3 What will we be studying in this course?

This is a course about applied research. The objective is that you gain a good measure of independence conducting a research project of your own. When I say research, I don't necessarily mean academic research, I also mean research for companies, banks, and governmental institutions.

You will not learn everything there is to know about the subject, of course, but what you will learn will be very valuable. The techniques we will study are still the most widely used in all the profession. The research skills taught in this course are real college-level skills, and are also quite marketable.

What is applied research? It is the practice of answering scientific questions using data. We won't be concerning ourselves with all forms of scientific questions. We will focus on behavioral questions, those involving human decisions.

Topics involving human behavior are special in many ways. To name a few ways in which studying humans is different, remember that humans forget, lie, exaggerate, round, omit, refuse to participate, disobey, change opinions, and take revenge. We will see how these and other actions impact our ability to answer questions. We will see how to conceive models so as to prevent several of these problems, and also when to admit that our models are unsatisfactory.

4 Scientific Questions

Mainly in this course we are interested in causal questions. Moreover, we are concerned with questions that can be studied using statistical data sets. A causal question must have two variables, named

- _____: the variable which we influence directly. The "causal variable."

- _____: the variable which we will only influence through the treatment variable, the "caused variable."

These variables must be well defined, in the sense that they must be quantifiable. We want precise units of measurement. Consider the following scientific questions:

1. How much will attending one extra class impact the final grade?

2. How much will smoking one more cigarette per day during pregnancy affect the baby weight at birth?

What are the treatment variables in the examples above?

1. _____

2. _____

What are the outcome variables in the examples above?

1. _____

2. _____

Consider the following attempts at writing a scientific question. Why are they unsuccessful?

- I want to study class attendance.

- Is class attendance important?

- Is class attendance important for grades?

Another example

- I want to study about pregnant women that smoke.

- Is smoking during pregnancy bad?

- Is smoking during pregnancy bad for the baby?

5 Experiments

Experiments (when well designed) are the gold standard of scientific research in causal questions. Why?

5.1 Counterfactuals

Take the 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?

What is the ideal way in which to answer them?

1. Take a person that went to 10 classes. Suppose that this person got a grade of 76. Then suppose I go back in time and make that person go to 11 classes instead. Then the person ends up getting a grade of 81. Then I can say that one extra class caused the grade to go up by 5.

2. Take a woman that smoked an average of 4 cigarettes per day during pregnancy, and had a baby weighting 3100 grams. Now I go back in time and force her to smoke 5 cigarettes per day instead. The baby is born with 3000 grams. We can then say that one extra daily cigarette caused a weight loss of 100 grams.

- Why can we make such causal statements?
  Because the only thing that changed between the two situations is the ⎯⎯⎯⎯⎯⎯⎯

  ⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯

- Why is finding the causal effects this way impossible?
  Fundamental problem of causality: we can't time travel. Each person will take one action, so we can never observe the counterfactual if the person had taken a different action instead.

Counterfactual: ⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯⎯

⎯⎯

5.2 Ceteris Paribus

One idea is to get two different people equal in every possible aspect. Then force them to take different actions.

1. Take two students exactly alike, and force one to go to 10 classes, another to go to 11. The difference in their grades has to be caused by the differences in classes, because nothing else changed.

2. Take two women exactly alike, pregnant at exactly the same time from the same man (kinky...) Force one to smoke 4 cigarettes, and another to smoke 5 cigarettes daily throughout pregnancy. The difference in the weight of their babies was caused by the differences in cigarettes, because everything else is the same.

- Why can we make such causal statements?

  Because the only thing that changed between the two situations is the _____

  _____

  You see, the essence of finding the causal effect is *ceteris paribus*, that is, leaving everything else constant. We changed the classes and the amount smoked, but kept everything else the same.

  Ceteris paribus: _____

- Why is finding the causal effects this way impossible?

  _____

  _____

5.3 Experiments

However, as it turns out, even though we can never find two people exactly alike, if we distribute people in two groups in a random manner, the two groups will turn out to be comparable. The experiments below are ideal, but not necessarily feasible.

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.

- Why can we make such causal statements?

  Because the two groups are comparable, since they were selected randomly. Since we didn't intervene in any way other than to force one group to take one more unit of treatment (go to one further class, or to smoke one further cigarette), the difference in the outcomes across the groups has to be caused by the extra unit of treatment.

- Why is finding the causal effects this way hard? (observe that it is not, in principle, impossible)

  1. _____

     Can you force anybody to go to more classes? Can you force a mother to smoke?

  2. _____

     Even if you force them, they know whether they are going to one extra class, or smoking one extra cigarette. Hence, they may decide to change other actions because of that. For example, a student forced to go to one more class may decide to not go to an office hour he was planning to attend. A mother that was forced to smoke one more cigarette may decide to go to the doctor for prenatal visits more often. The problem with the inexistence of a placebo is that the treatment and control groups may not longer be comparable. Students in the group forced to go to more classed may go to less office hours. Mothers in the group forced to smoke may go to more prenatal visits. "*Ceteris paribus*" fails.