Notes for Cors 236 test 5

Correlations and causation

At the start of this test period, we (finally) discussed the survey about an internal email to UI advisors recommending that all students take at least 15hr/semester. That example was discussed in the context of correlation does not imply causation, third variables and Simpson's paradox. The topics for this example overlap test 4, but because we discussed them during this period, they are fair game.

3rd variable models: can you identify them?
The method to use for this is that you take the correlation given between X and Y. Then ask if, under the model, changing X (and its downstream consequences) but leaving everything else as is, will change the outcome. If the outcome does change, there is no 3rd variable. If the outcome does not change when X is changed, there is a 3rd variable.

Controls and controlled variables
To control a variable is to make it the same (on average) between treatment and control groups. This can be done either by randomizing assignment to treatment and control groups (before the data are gathered), or by knowingly making sure control and treatment groups have the same composition of the 3rd variable. With the randomization approach, you don't need to identify the 3rd variables, but with the 'knowing' approach you do.

A treatment variable is one that is arranged to differ between treatment and control groups.

A table of presence/absence is an easy way to allow one to determine which variables are controlled

Experiments:
They are manipulations of nature to gather data in a new setting to test a model.
The manipulation of nature is what makes it an experiment (in contrast to correlational data).
Experimental data may have replication, blind, randomization, standards and controls, but it is the manipulation that makes it experimental.

Quizzes are fair game if you have completed them and can access them (up to the exam).