Math 455 Class 15 February 24

Reminder: Midterm 1: Monday March 3, in class.

Covers roughly Ch. 1-5 of textbook.

Note that the topics on the sample midterm are not "complete": there are testable topics not on the sample midterm.

One example: the "Hill Criteria". (Sec 5.7).

We want to make causal conclusions from regression.

Regression, by itself, does not imply a causal connection.

Ideally we would do some kind of controlled study: we could do "double-blind" experiments:

- Patients don't know if drug / placebo.
- Doctors _____.

We could match patients to control group.

All of this is expensive!

What can we do with observational data?

(Not from a controlled experiment.)

Hill Criteria: common-sense rules of thumb for when we can take regression as evidence of causation.

- Most commonly used criterion:
  does the result hold after controlling for covariates?

  This means "put extra variables" (covariates) in your regression.

  Show that the association still holds.

  (This may depend on choice of extra variables!)

Controlling for Covariates is covered under "Strength" and "Consistency".

Strength: various measures, like p-values and size of coefficient, We can ask if these measures remain strong after controlling for covariates.

Consistency: Various data sets / times / places / experiments.

Specificity: The supposed causal factor is associated with a specific response and not a lot of other possible responses.

Example: we want to demonstrate that asbestos exposure $\longrightarrow$ lung cancer.

The evidence is stronger if workers exposed to asbestos don't have elevated rates of other diseases.

Temporality: If X ~~cc~~ causes Y, X must happen before Y.

Plausibility: Is there some sensible theory under which X causes Y?

Gradient: Does more of X associate with more of Y?

Natural Experiments: e.g. Card - Krueger, Newmark on min wage. NJ/PA.

Section 5.5: Matching

Another way of "controlling" for covariates:
divide data into "matched pairs".

R packages for doing this automatically.

Less frequently used in practice.

Know the difference between "explanation" (Ch 5)
and prediction. (Ch 4)

A question might describe a problem and
ask "is it a prediction problem or an
explanation problem."

In prediction, know the difference between
"confidence" and "prediction" intervals.
(see explanation in 4.1)