IR 514: Multivariate Analysis

Benjamin Graham

STAND BACK
I'M GOING TO TRY SCIENCE

Saturday, January 26, 13
Plan for Today

• Introductions
  – Your motivations/objectives for this course
  – A bit about POIR methods training

• Context and Course objectives
  – Pedagogy

• Business and logistics

• Establishing the research design context
Quantitative vs. Qualitative Research

- This distinction is WAY overhyped
  - The stereotype of quantitative research is that it lacks nuance and context
    - The stereotype of qualitative research is that it lacks rigor
  - Either can be (and often are) done badly
  - Either can be (and sometimes are) done well
  - Really we’re looking at a broad spectrum of methods

- Your selection of method should always be question driven
  - OK, but I can’t become an expert in every method...
  - That’s what co-authors are for
Training for Your Research Agenda

• Step 1: Enough training in all the core methods of the field so that:
  • You can read intelligently
  • You can look at a question and know the right method to use
• Step 2: Pick your research area
• Step 3: Get more specialized training in the most appropriate method
The Efficient Research Portfolio

- Your book or your core set of articles that use the methods you know to answer questions on your core topic

- Some articles in your core area where a co-author does the heavy lifting on the methods side and you write the theory

- Some articles outside your core area where you do the heavy lifting on the methods side and your co-author writes the theory
Why quantitative training is so awesome

• Math is scary. Stats is hard.
• Fewer people get the training, so your skill set is more valuable
  • Applies both inside and outside academia
• Because it’s hard, you want to learn from a class (to start).
• Stats is new
  • So senior people don’t know it
  • And there are still a lot of questions no one has gotten to yet
• Data is rapidly becoming cheaper to collect and store
  • The scarce complement is people who can make meaning from that data
POIR Methods Training

• Currently this is the only quant course offered
• We’re working on that
  – Diana O’Brien -- maximum likelihood and related topics
• Also working on interdisciplinary social science methods training
• Does this course work for MA students?
Course Objectives

• 1. Enable you to read quantitative research intelligently
• 2. Teach you several basic techniques for quantitative hypothesis testing
• 3. Enable you to manage and manipulate datasets for large-n research projects
Course Objectives(2)

• 4. Provide you skills for self-teaching or working with co-authors using more sophisticated techniques in the future

• 5. Provide you with a working knowledge of Stata to facilitate items 2-4
Notes on Pedagogy

• Where I can, I will try to flip the classroom

• This means web videos at home

• In class, it means simulations, demos, examples, Q&A, and labs

• Videos for next week will go up by Saturday night
Business and Logistics

• Any problems acquiring Stata?
  – Stata lab in class next week
• Any problems enrolling?
• I run this course off my USC website
  • Dropbox for homework assignments
• The syllabus is a work in progress
  – Many readings will change
The Paper

• Half your grade is a final paper
• Replication paper or original research
  – This isn’t about theory, its about testing
The Scientific Method

- **The scientific method has four steps**
  - 1. Observation and description of a phenomenon or group of phenomena.
  - 2. Formulation of a theory to explain the phenomena.
    - Key concept: Induction
  - 3. Derivation of hypotheses that make falsifiable predications about new observations.
    - Key concepts: falsifiability and “out of sample tests”
  - 4. Performance of multiple tests of these hypotheses by multiple research teams.
    - Key Concept: Replicability

- If the experiments bear out the hypothesis it may come to be
Types of Studies (1)

- **Descriptive:** To observe and describe the world around us. E.g., *The Origin of Species*
  - World War I killed a whole lot of people

- **Relational:** Designed to examine the relationship between two or more variables.
  - e.g. the democratic peace

- **Causal:** Designed to determine whether one or more variables causes or affects one or more outcomes.
Induction and Deduction

- Induction

  Observation

  Descriptive Study

  Pattern

  Relational Study

  Theory

This is your puzzle!
Induction and Deduction

- Deduction

  Theory → Hypothesis → Observation → Confirmation or Refutation
Social science theories are usually causal

- The pattern we observe can usually be described as a relationship between two things
  - Democracy and interstate war
- The dependent variable is the outcome we are trying to explain
- The independent variable is another item that we think causes the outcome in question
Why are we fixated on causation?

• Because we want to change the world
• A causal theory tells you what lever to pull

• We want to improve normatively bad outcomes
  • Save the world
• We want to alter the behavior of others or ourselves
  • Make someone buy something
  • Stop someone from harming us
The big picture
Deductive theory testing is a Growth Industry

- Program evaluation
  - Public Sector
  - Private sector
    - for profit
    - not for profit
- Ideally, the evaluation is built into the implementation in the first place.
Purely Academic Research

- Academics may go back and forth without moving to implementation:
  - Induction creates theory
  - Deduction tests it. Then we revise the theory and test it again. And we do it again.
Research Design

• When quantitative research is done badly, research design is usually at fault
• But this class is focused on what comes after the research design is in place
Is a project/topic feasible?

- Does your dependent variable vary?
- Can we define and measure everything that needs to be defined and measured?
- If the study involves some condition, can we define it? Can we be sure we'll recognize it when we see it?
  - What is improved human welfare?
  - What is political competition? How do we observe it?
  - What is crop yield--what gets harvested or what makes it to market, or both? What are sales--what gets manufactured or what gets into the hands of the consumer?
  - What's the difference between civil war and other wars?
  - How do you measure norms, beliefs, values?
Measurement

- Are we measuring what we think we're measuring?
- How should we choose among different measurement techniques?
- Can measurements be made consistently?
- How accurate are the measurements?
- If there are errors in the data, are they random?
- Do we have missing data? If so, are those missing completely at random?
Collecting Your Own Data

• All of the relevant data must be collected. If a critical piece of data cannot be obtained, perhaps the study should not be undertaken.
• It is equally important to guard against collecting data unrelated to the research question.
Using Other People’s Data

• User beware
  – Read the codebook.
  – Talk to the authors of the data if you have questions.
  – Take this as seriously as if you were writing the codebook yourself
  – Spend some time getting to know the data, one variable at a time
Causal Variables/Treatments

- What is a causal variable/treatment?
- Treatments must be clearly identified
- It is possible that multiple things cause the observed effects
  - What makes someone vote for Obama?
  - How would we know that age increases probability of voting for Obama?
Types of Studies (2)

• The typical study can be classified into one of two types:
  – observational studies (non-experimental)
  – Intervention studies (experimental)

• The distinction is based on whether an intervention is involved, that is, whether the investigator changes some aspect of subjects' behavior.
Experiments = Easy Analysis

• If we have randomization, the stats are simple
  – Difference in means test
• The farther we are from an experiment, the more complicated the statistics
The Order of Things

• We’re going to start with describing a single population
• Then we’ll move to comparing two (or more) populations on a single dimension
• Then we’ll move to comparing multiple populations on multiple dimensions
Ways to Be A Good Student

• Watch the videos with intention
  – Write down your questions and bring them in
• In class, two useful statements are always
  – I don’t get it
  – I don’t get how I would use this
• Do the homework yourself
• Pick a paper that is actually useful to you