Introduction to Quantitative Research

Senior Thesis Tutorial
October 18, 2013

Ekédi Mpondo-Dika
empondo@fas.harvard.edu

What is research?

- *Controlled* collection and analysis of information in order to understand a phenomenon
- Originates with a question, a problem, a puzzling fact
- Requires both theory and data.
  - Previous theory helps us form an understanding of the data we see (no blank slate).
  - Data lets us tests our hypotheses.

Quantitative research

- Quantitative methods allow us to learn about the world by quantifying some variation(s) in it.
  - Example: how do suicide rates vary across demographic categories (Durkheim)?
- In order to learn about the world, we use *inference*:
  - General definition: "Using facts you know to learn about facts you don't know" (Gary King)
  - Example: Using a sample to learn about the world
- It is important to *design* our research so that our inference will be correct.

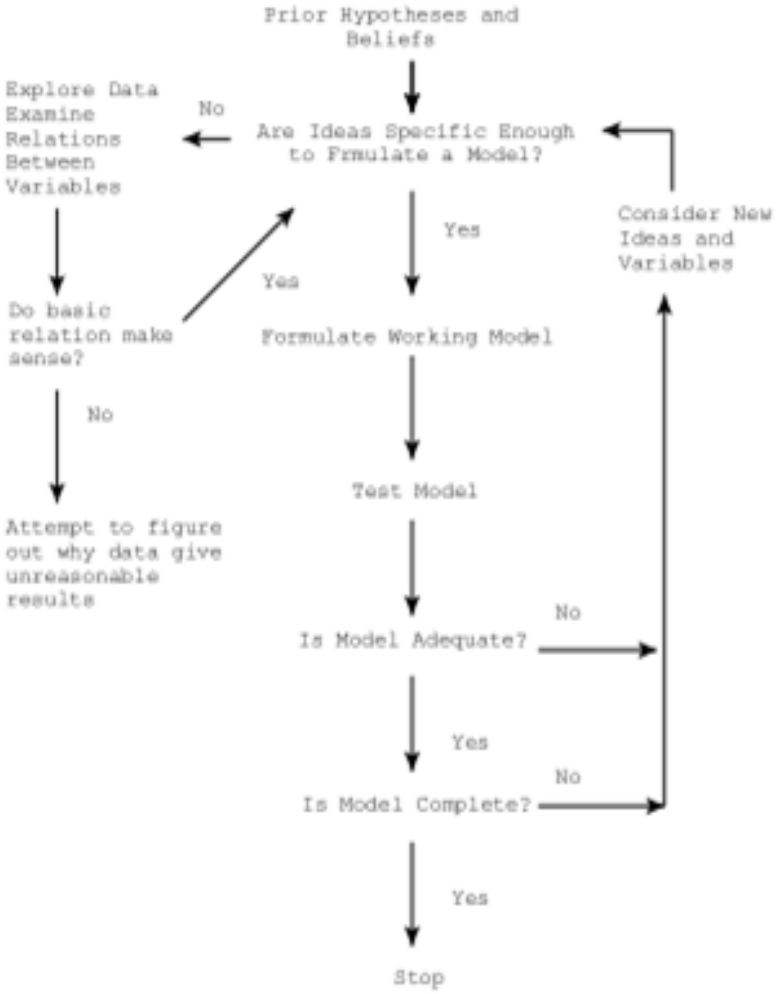
Research process
(confirmatory or deductive model)

- Reviewing literature and identifying a question
  - Sometimes question helps you identify relevant literature, sometimes literature helps you identify unsolved puzzle
  - What makes the question interesting? Real world implications and theoretical contribution
- Form hypotheses
- Think about data you need to test them
- Think about methods to analyze your data

Research process (actually)

- Back and forth between theory and data
  - Each is going to highlight relevant features of the other
  - That's how a contribution takes shape
- But: you can't use the same data to generate and test hypotheses. It would be tautological
- Importance of cross-validation

Exploratory and confirmatory analysis



Research Questions

- Have to be specific enough and operationalizable
- Examples:
  - Macro: How does a society's level of economic inequality affect its population's health? (Wilkinson and Pickett 2009)
  - Micro: How does the distribution of control over financial resources between father and mother affect children's food security? (Kenney 2008)
  - Macro and micro: How do cultural norms regarding gender roles affect fertility outcomes? (Brinton et al, in progress)

Research questions

- Let's pause and consider your questions

Data Collection

- Think hard about the population you want to study
- Think hard about selection
  - Interviewing method: face to face, phone
  - Sampling: random? Purposive?
    - Time of day
    - Place
    - Who agrees to respond?

Models

- A model is a strategic simplification
  - Never true or false
  - Only useful or not (does it capture the features you are interested in?)
- Examples:
  - Obesity rate=f(absolute wealth, inequality)
  - Food security=f(total hh income, distribution of control over income)
  - Fertility=f(individual factors, structural constraints*cultural norms)

Models

- Let's think about models for your questions

Hypotheses

- Inequality and health
  - H0: life expectancy only affected by level of GDP
  - Ha: autonomous effect of economic inequality on life expectancy
  - To test this, I need data on countries which have similar levels of GDP but different levels of inequality (eg, Sweden versus US)

Statistical model

- Depends on the process you want to study:
- Continuous outcome (rate, income…): linear or log-linear regression
- Binary outcome (marriage, incarceration): logistic regression
- Count data (civil wars): Poisson regression (advanced)
- Timing of events: Survival analysis (advanced)

STATA resources

- UCLA website (v. popular): http://www.ats.ucla.edu/stat/stata/
- Princeton web page (the one I learnt from, good to start): http://data.princeton.edu/stata/
- List of various books: http://www.stata.com/links/resources-for-learning-stata/
  - My personal bible: Cameron & Trivedi *Microeconometrics using Stata,* Stata Press.
- RTC team in CGIS building (awesome!): http://projects.iq.harvard.edu/rtc
  - Consulting
  - Stata workshops

Excel to Stata

If your data are currently in Excel, you need to convert and import them:

- http://www.stata.com/support/faqs/data-management/converting-excel-files/

- Also: “help insheet” in Stata

STATA gold rules

- Always use a **do-file**
- Always **comment** on everything you do within the do-file
  - You don't want to be lost months from now
  - Also in some cases you might want to show me or your adviser your do-file
- (Almost) never save the *data* at the end of a session
  - That will replace your original data set
  - Better to "**save as**"

Exploring your data

- Always useful: **help** command
- What do my variables mean?
  - Command **describe [name/list of var]**
  - Command **codebook [name/list of var]**
- Basic statistics
  - **Summarize [list of variables], detail**
  - **Tabulate [one or two variables], options**

Example

- Let's explore the data set used by Matthew Hunt in his article "African American, Hispanic, and White Beliefs about Black/White Inequality, 1977–2004" (2007)
- GSS Data, but we'll reduce that huge data set to something we can use more easily

- See do-file "Intro to Stata"