Getting Started with Quantitative Empirical Methods

Dr. Su Li
University of California, Berkeley, School of Law
09/18/2014
Outline

• What are empirical methods? What are quantitative empirical methods?

• Quantitative empirical methods and law
  ---lawyering and statistics
  ---legal scholars and statistics

• Overview procedures of quantitative empirical research
  ---design your project
  ---primary data/secondary data
  ---Analyzing data

• Stata Demonstration of data management and analysis (if time allows)
FAQs

• How can I determine whether my research questions are appropriate for quantitative methods?

• Where can I find data that might be useful for the research questions being asked and how can it be set up for statistical analysis?

• How can one collect original data and set it up to conduct statistical analysis?

• What are the tradeoffs between analyzing original versus secondary data?

• What does it mean by sampling and statistical inference? (another workshop)
Overview of quantitative empirical methods

• What are empirical methods?
  Research methods that are based on observations of the world, data, which is a term for facts about the world.

• Qualitative Data vs. Quantitative Data
  --Qualitative data: Historical archives, ethnography observations, in-depth interviews, case studies
  --Quantitative Data: large sample survey, coded law cases, survey experiment

• What are quantitative empirical methods?
  Research methods that utilize quantitative data when attempting to make a general statement about a certain issue in the world on the basis of a sample of observations of the world. Usually the sample needs to be large enough and representative of the world.
Quantitative Empirical Methods and Law

- Lawyering and statistics: (certainty vs. uncertainty; inductive thinking vs. deductive thinking) (Luna, 2007)
  - e.g. in employment discrimination cases; discrimination in primary strike


- Legal scholarship and statistics

  ---judicial behavior (judges; juror): vote

  ---court opinions: strict scrutiny; punitive damages


  ---criminal behavior and punishment

  ---law and society

  impact /effectiveness of law; enforcement (discretion); factors affect the formation of law
Goal of this workshop

• Overview of how to conduct quantitative empirical legal research (peer review or law review articles)

• Form a study group (support group) for legal scholars who intend to write a quantitative research paper on law
Example research question:

Do judges’ ideology impact their voting outcomes?

What would you do to answer this question?

• Dependent vs. independent variable
  Ideology → Voting outcome
• Measurement
  ---Is Ideology easily measurable?
• Explanatory and control variables
  ---Is ideology the only factor that affect judge’s behavior?
• Unit of Analysis: judge? Case?
Overview of quantitative empirical research (stepwise procedures)

1. Research Question
2. Hypothesis construction
3. Research design: concepts quantification (variables)
4. Data collection/Data finding (data coding and cleaning)
5. Data analysis (model construction)
6. Results interpretation

Operationalization & Generalization

Interpretation
Design your project

• Decide if quantitative method fits your need NOW
  • Is your question explorative or confirmatory?
  • Is experimental design viable?
  • Are you interested in looking for “trends”, “likelihood”, “distribution”?
  • Is quantification of the concepts you want to measure possible/viable?

• Design your project: abstract → concrete: clarification is a major issue in project design, clarifying abstract concepts, making them concrete.
  • Dependent variable vs. independent variable
  • Variable types (nominal, ordinal, interval-ratio)
Collect/find and code/recode the data

• Collect and code the data: coding--taking information and translate it into ways that can be analyzed.
  -- from where? Secondary data sources (next page)
  -- help yourself? original data collection and coding
  -- combine secondary and original data
• Trade-offs secondary data vs. original data
• how much data do you need?
  • as much as possible
  • at least enough for statistical inference if data is a sample (sample size workshop)
Secondary Data Resources

Some public/restricted quantitative data resources

• The Interuniversity Consortium for Political and Social Research (ICPSR): http://www.icpsr.umich.edu/icpsrweb/ICPSR/access/index.jsp
• The Roper Center Public Opinion Archive: http://www.ropercenter.uconn.edu/data_access/data/search_for_datasets.html
• The Social Science Electronic Data Library (SSEDL data archives) http://www.socio.com/members/memonly.htm
• Minnesota Population Center Integrated Public Use Microdata Series (IMPUS): http://www.ipums.org/
• Wharton Research Data Services (WRDS): http://wrds.wharton.upenn.edu/
• The Census Bureau: http://www.census.gov/
• American Bar Foundation After the J.D. (AJD )data http://www.americanbarfoundation.org/publications/AftertheJD/AJD_Data_Access.html
• The U.S. Supreme Court Database http://supremecourtdatabase.org/

UC Berkeley resources:

• CSLS BELS data link http://www.law.berkeley.edu/4501.htm
• Immigration topics data (Irene Bloemraad): http://www.popcenter.berkeley.edu/resources/migration_data_sets/intro.shtml.
• U.C. Data: http://ucdata.berkeley.edu/; California Census Research Data Center (public and restricted census data)
• Berkeley Library Data lab: http://sunsite.berkeley.edu/wikis/datalab/Main/HomePage
Tips for collecting and coding your own data

• Decide your population and sample
• Try get random samples (simple random sample, stratified random sample, clustered sample)
• Take extreme caution in conducting your OWN survey
  SurveyMonkey.com
  Qualtrics.com
• Construct your own codebook (see example white-collar project)
• Mutually exclusive and exhaustive categories/values for each variable (e.g. wc #6)
• Missing values (e.g. wc #5)
• Reduce human error (clear instruction, less ambiguity in variable definitions) (e.g. wc #8)
• Control inter-coder reliability (another workshop)
Recoding Secondary Data

- Construct new variables from existing variables (e.g. ideology majority from ideology composition)
- Recode existing variables (e.g. ideology (1,2) to ideology (0,1)
- Scales and indexes (e.g. psychological wellbeing scale from 12 variables)
- Missing values

see example (supreme court data; Sunstein data)
Analyze the data

- Analyze the data: putting the data results together with the hypothesis.
  - Summarize the data (summary statistics, cross-tabulation, graphs)
  - Inference: generalize the findings from a sample of cases to all cases, want to assess the uncertainty about your inference.
FAQs

• How does one determine whether research questions are appropriate for quantitative methods?
• Where can one find data that might be useful for the research questions being asked and how can it be set up for statistical analysis?
• How can one collect original data and set it up to conduct statistical analysis?
• What are the tradeoffs between analyzing original versus secondary data?
Stata Demo (IP data)

- Input data
- Recode data
- Tabulation and cross-tabulation
• Thank you!