How to Conduct and Present Statistical Research

Kosuke Imai

Department of Politics
Princeton University

April 29, 2011

Professional Development Workshop

Conducting Statistical Analysis

1. Start with simple exploratory analysis before complex modeling
   - Get to know the data
   - Compute descriptive statistics and make graphs
2. Examine a small number of hypotheses
   - Be deductive
   - Be careful about multiple testing and avoid data snooping
3. Substance, substance, substance
   - Focus on substantive rather than statistical significance
   - Substance should motivate methods, not vice versa

Communicating Statistical Results

1. Make it easy for readers and audience
   - Exploratory analysis results (not obligatory “descriptive stats”) first
   - Clearly specify key identification assumptions
   - Report quantities of interest rather than coefficients

2. Reporting statistical results
   - Use tables and graphs but only selectively
   - Should be self-explanatory
   - Use lengthy captions if necessary

3. Tips about making graphs and tables
   - Graphs are typically better than tables
   - Small multiples and avoid legends
   - Use intuitive labels and avoid abbreviation
   - No more than 3 or 4 digits
   - Pay attention to details and “prettify” tables and graphs

Writing an Empirical Paper

1. Start with tables and graphs with detailed captions
   - They determine the "story"
   - Do not start writing before you know what to write
2. Determine title and then write abstract
   - Avoid catchy titles and be informative
   - 150 words for abstract but spend a lot of time
3. Following the abstract, write the introduction
   - The question to be answered and the problem to be solved
   - Your answers and solutions
   - Your contributions to the literature
4. Following the introduction, write the rest of the paper
   - Top-down structure
   - Whose mind are you trying to change in what way?

Presenting and Publishing Empirical Research

1. Every talk is a job talk
   - The structure of slides should follow that of the paper
   - First 5 minutes and last 5 minutes are most important
2. You will be held accountable for what you publish
   - Every submission is subject to peer review
   - Get feedback from friends and advisers first
3. Make your data and code publicly available with documentation
   - Dataverse is free and easy to use
   - Organize and comment your code from the beginning