Using Computer Programs for Quantitative Data Analysis

Brian V. Carolan
The Graduate School's Graduate Development Conference, Montclair State University
Slides available at:

http://www.briancarolan.org/Site/Courses.html
Overview

• What this talk is not about
• Advantages & disadvantages of using software for quantitative data management and analysis
• Generic features of most popular applications: SPSS(PASW), STATA, & SAS
• Example of generic features using SPSS
• Tips for data management and analysis
• Resources to get started
A cautionary tale

• A grad student walks into my office one day holding a stack of surveys and says ...
What This Talk Not About ...

- Research design
- Measurement
- Statistics
- Applications designed for qualitative content data analysis (e.g., NVivo) or social network analysis (e.g., UCINET)
Advantages to Using Applications

• Reduce/eliminate errors in calculation
• Data management, e.g., add variables & observations, recode variables, etc.
• Graphical utilities
• Multiple users can work with the same data file
• Faster, more efficient
Disadvantages to Using Applications

- Menu, syntax, and terminology differ across applications
- Certain applications are better for different types of analyses
- Much more functionality than one typically requires
- Lag between new versions and existing documentation
- Analyses generate output that often reports results that one doesn’t need/understand
Generic Features of Popular Applications, e.g., SPSS, etc.

- “Raw” data are organized in tabular format with each observation having a row and each variable its own column (i.e., observation by attribute)
- Data, command, & output files are distinct and saved as such
- Menu or syntax can be used to create graphical displays
- Variables have to be identified in a certain format prior to analysis (most likely “numeric,” not “string”)
- Extensive “help” menus
SPSS Example: Observation by attribute
SPSS Example: Separate files
SPSS Example: Creating graphics
SPSS Example: Variable format
SPSS Example: Help menu
Tip #1: The interpretation is up to you, not the application
Tip #2: Garbage In, Garbage Out

- Only one of the below statements is correct, yet the application ran the tests that produced both results: 1) There was a non-significant relationship between students’ race and scores for planning/instruction, $F(4, 182) = .99, p = .42$ or 2) There was a significant relationship between students’ race and scores for planning instruction, $r = .20, p = .09$. Which one is right and why?

- Bad measures translate in bad analyses. For example, on a four-point scale, what does it mean when a respondent has a score of “3” on a measure of engagement? How do you know the measurement of this variable is valid and reliable?
Tip #3: Organize and Document Your Work

- Every time you alter the data file, save it as a new file. That way, if there’s a mistake in the new file, you can always go back to the previous version.
- Keep a running record of what you have done to the data, including the analyses. STATA has a nice log file that allows you to do this.
- Organize it in a way that others can follow, including labeling the variables. A codebook is a really nice tool. STATA has a nice command for this too.
Tip #4: Working with Missing Data

- In all perfect world, each observation would have a legitimate value for each variable. This, however, is rarely the case.
- Must define what the value of “missing” will be, so that the analyses do not consider these legitimate values.
- Each application handles these differently.
Tip #5: Don’t Go At It Alone

• Why do most quantitative papers have more than one author?
• Ask others, especially those who have also worked with the same data. This is especially relevant when performing a secondary analysis of existing data.
Resources

- Number of introductory statistics texts rely heavily on one of the popular applications.
- Seminars targeted to novice users
- Working with faculty