

The “Debugging” Workshop

Hsueh-Sheng Wu
CFDR Workshop Series
April 13, 2015

BGSU



Center for Family and
Demographic Research

Outline

- Introduction to syntax and logical errors
- Using PFE and Excel to help your programming
- Rules of good programming
- Three examples of Stata command files
 - 1) A bad command file
 - 2) A good command file to create the racial background for respondents
 - 3) A good command file to identify who the reference person's spouse or partner was and whether the respondent was in a different- or same-sex union
- Conclusions

Introduction to Syntax and Logical Errors

- Syntax error: Software does not understand what you want it to do and will issue an error message.
Examples:
 - Cannot find the data because a wrong file name or location is provided
 - Cannot find the variables because types in variable name
 - Cannot find the necessary variables (e.g., using one variable for a cross-tabulation)
 - Wrong variables are provided for a certain analysis (e.g., adding value label to a string variable)
- Logical error: Software understands and does what you ask and does not issue an error message.
Examples:
 - Wrong analytic sample
 - Variables are inaccurately created
 - Wrong model are chosen for the analysis (e.g., using OLS for binary dependent variables)
- Logical errors are harder to detect because software does not issue an error message.
- When creating a command file, you need to know what your data look like and how to detect possible logical errors in your code.
- We will use Programmer's File Editor (PFE) and Excel to help debug the programs.

Using PFE and Excel to Help Your Programming

Good things about Programmer's File Editor (PFE):

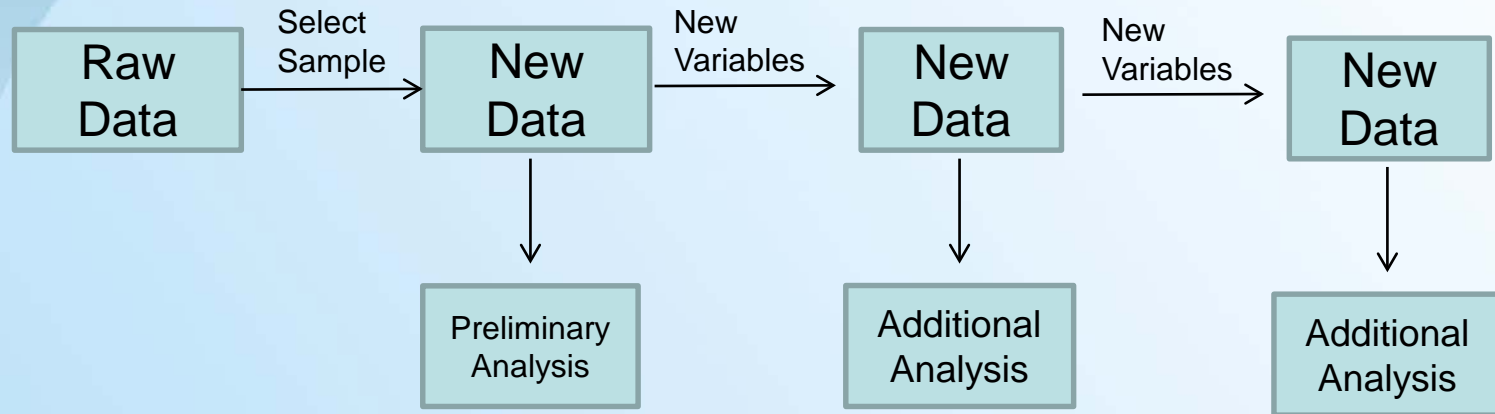
- Allows you to search and replace special characters
- Has a intuitive user interface
- It does not lock the log file and you will be asked if you want to load the newest log file.
- Allows you to open multiple files and examine them simultaneously
- Portable; it can be run from a thumb drive
- File can be downloaded here (<http://www.lancaster.ac.uk/staff/steveb/cpaap/pfe/>)

Good things about Excel:

- Has built-in function to change the letter case
- Has function to compare observations
- Allows you to better look your data and mark problematic observations
- Allows you to quickly write repeated command code

Rules of Good Programming

1. Always write command files and create log files
2. Create separate command files for constructing data and for analyzing data. Don't mix these two command files together.



3. Provide a summary statement at the beginning of your command file
4. Include notes and comments in your command file
5. Keep your command clean and easy to spot any errors
6. Always consider what can go wrong with your code and how to check for the possible errors.
7. Use the same name for the command, log, and data files

Conclusions

- Command files should be used to generate variables, construct data files, and run analyses.
- Separate command files for constructing data and for analyzing data. Don't mix these two types of command files together.
- Keep your command files short and neat. You can always merge it with other command files.
- Include detailed notes and comments in your command files.
- Always think about what can go wrong with your code and always check if your code achieves what you intend them to do.
- If you have problem debugging your command code, please contact me at wuh@bgsu.edu.