

Statistical Programming: A Brief Primer

ECOG 314 - Spring 2017

January 13, 2017

Board of Governors of the Federal Reserve System

Statistical Programming: A Brief Primer

- 1 What is data?
- 2 What is statistical software?
- 3 What is a program?
- 4 Why do economists program?

Statistical Programming: A Brief Primer

- 1 What is data?
- 2 What is statistical software?
- 3 What is a program?
- 4 Why do economists program?

da·ta

/ˈdādə, ˈdādə/

noun

facts and statistics collected together for reference or analysis.

synonyms: facts, figures, [statistics](#), details, particulars, specifics; [More](#)

- **COMPUTING**

the quantities, characters, or symbols on which operations are performed by a computer, being stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.

- **PHILOSOPHY**

things known or assumed as facts, making the basis of reasoning or calculation.

What is a data set?

What is a data set?

- A collection of related data

What is a data set?

- A collection of related data
- Used in conjunction with statistical methods to answer a question of interest

What is a data set?

- A collection of related data
- Used in conjunction with statistical methods to answer a question of interest
- Able to be manipulated by a computer

What is a data set?

- A collection of related data
- Used in conjunction with statistical methods to answer a question of interest
- Able to be manipulated by a computer
- Mathematically, a $m \times n$ matrix

Working with data sets

- “Small” data sets can be manipulated in a point-and-click manner
 - $m, n < 200, 10$
 - Excel
 - SPSS
- How should larger data sets be handled?

Statistical Programming: A Brief Primer

- 1 What is data?
- 2 What is statistical software?**
- 3 What is a program?
- 4 Why do economists program?

What is statistical software?



What is statistical software?



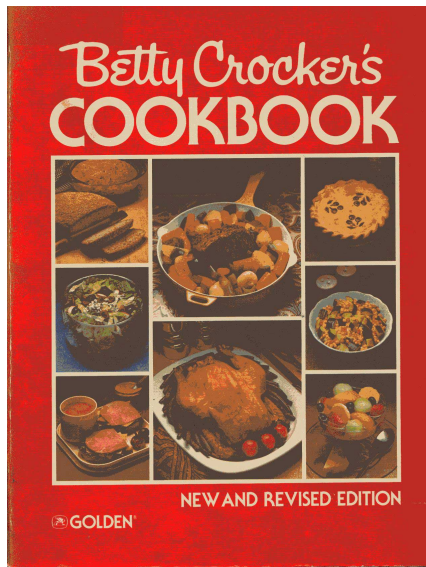
The main idea

Write every step of your research project in a **program** so that it can be done in the push of a button

Statistical Programming: A Brief Primer

- 1 What is data?
- 2 What is statistical software?
- 3 What is a program?
- 4 Why do economists program?

What is a program?



At the simplest level, a program is...

... a **set of step-by-step instructions for the computer to follow**

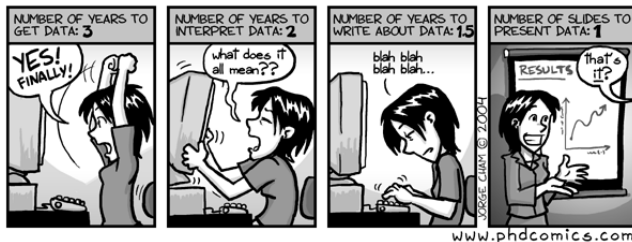
More specifically,

- Written in **syntax** the computer can recognize
- Written top-to-bottom, left-to-right
- Saved in a file with a specific extension (e.g., .do, .R)
- Written with the understanding that the computer can't do anything you don't tell it to do

Statistical Programming: A Brief Primer

- 1 What is data?
- 2 What is statistical software?
- 3 What is a program?
- 4 Why do economists program?

DATA: BY THE NUMBERS



- Data sets are usually large and messy
- Systematically transform raw data to usable data set

- Availability of well-defined statistical functions and procedures
 - Summary statistics
 - Linear and nonlinear regression
 - Monte-Carlo simulation
- No need to “reinvent the wheel”
 - $\text{lm}(\text{data} = X, \text{model} = y \sim X)$
 - $(X'X)^{-1}X'y$

- Most software can produce charts and tables
- ggplot2 in R is particularly great

Replicability of research

What your research supposedly looks like:

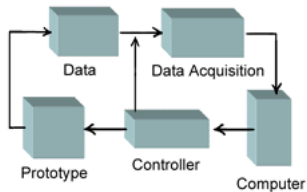


Figure 1. Experimental Diagram

What your research *actually* looks like:

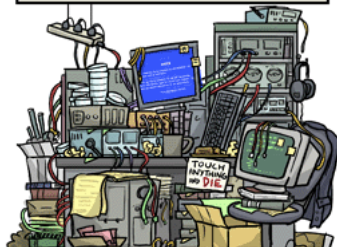


Figure 2. Experimental Mess

JORGE CHAM © 2008

WWW.PHDCOMICS.COM

The big picture

Most quantitative work involves:

- ① Creating a data set from raw data
- ② Conducting statistical analysis on the data set
- ③ Visualizing results
- ④ Presenting your work to others
 - Your boss
 - Your peers
 - Academia (publishing)

R enables you to do all of these things!

Closing thoughts

- Work thoughtfully and methodically
- Document your work
- Test copiously
- Think of the general case
- Have a peer check your code
- Don't be intimidated

Contact information

- Erik Larsson
- erik.g.larsson@frb.gov
- Piazza site