Some Useful Stata Information

John McGready
The Johns Hopkins University
Lecture Topics

- Stata’s appearance
- Entering data and holding on to it
- Re-visiting and updating Stata data sets
- Syntax diagrams and “immediate commands”
Section A

What Does Stata Look Like?
Command Icons
The Windows

1. Command
   - Window that allows user to input commands
   - Stata is a “commands driven” program—commands are entered individually as opposed to entered as a group (batch submission)
The Windows

2. Results

- Window that displays all Stata output
- When the users enter a command that produces output, this output will appear in the “results” window
The Windows

3. Review
   - This window keeps a “running tab” of all commands used in current Stata session
Review Window
The Windows

4. Variables
   - Displays all variables in data set used in the Stata session.
Variables Window
The Screen

- You, the user, can customize the screen to your liking
- You can designate which of the four windows you want to have open on the screen, and size up each window to suit your preferences
Section B

Entering and Saving Data
How can we input this into Stata?

<table>
<thead>
<tr>
<th>Age (yrs)</th>
<th>Weight (lbs)</th>
<th>Age (yrs)</th>
<th>Weight (lbs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>123</td>
<td>43</td>
<td>150</td>
</tr>
<tr>
<td>65</td>
<td>215</td>
<td>8</td>
<td>58</td>
</tr>
<tr>
<td>31</td>
<td>187</td>
<td>76</td>
<td>126</td>
</tr>
<tr>
<td>15</td>
<td>89</td>
<td>50</td>
<td>312</td>
</tr>
</tbody>
</table>
You would like to use Stata to analyze this data set collected from eight subjects

<table>
<thead>
<tr>
<th>Age (yrs)</th>
<th>Weight (lbs)</th>
<th>Age (yrs)</th>
<th>Weight (lbs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>123</td>
<td>43</td>
<td>150</td>
</tr>
<tr>
<td>65</td>
<td>215</td>
<td>8</td>
<td>58</td>
</tr>
<tr>
<td>31</td>
<td>187</td>
<td>76</td>
<td>126</td>
</tr>
<tr>
<td>15</td>
<td>89</td>
<td>50</td>
<td>312</td>
</tr>
</tbody>
</table>
Data Editor

- Stata has a very nice and easy to use data editor
Data Editor

❖ To access the data editor, click on the editor icon in the Stata toolbar
Data Editor Icon
Each Row Contains Data for a Single Observation (Person)
Each Column Contains Data for a Single Variable in the Data Set
In this data set, each of the eight subjects is an individual observation.

<table>
<thead>
<tr>
<th>Age (yrs)</th>
<th>Weight (lbs)</th>
<th>Age (yrs)</th>
<th>Weight (lbs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>123</td>
<td>43</td>
<td>150</td>
</tr>
<tr>
<td>65</td>
<td>215</td>
<td>8</td>
<td>58</td>
</tr>
<tr>
<td>31</td>
<td>187</td>
<td>76</td>
<td>126</td>
</tr>
<tr>
<td>15</td>
<td>89</td>
<td>50</td>
<td>312</td>
</tr>
</tbody>
</table>
## Data Set

- Each observation consists of two variables, **age** and **weight**

<table>
<thead>
<tr>
<th>Age (yrs)</th>
<th>Weight (lbs)</th>
<th>Age (yrs)</th>
<th>Weight (lbs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>123</td>
<td>43</td>
<td>15</td>
</tr>
<tr>
<td>65</td>
<td>215</td>
<td>8</td>
<td>58</td>
</tr>
<tr>
<td>31</td>
<td>187</td>
<td>76</td>
<td>126</td>
</tr>
<tr>
<td>15</td>
<td>89</td>
<td>50</td>
<td>312</td>
</tr>
</tbody>
</table>
Data (age, weight) on the first subject in the study has been entered in the data editor.
Notice Stata Labels the First Column “var1,” and the Second Column “var2”
Double click at the top of a column, and the variable information window pops up.
You can change the name from “var1” to “age” in the name field.
Naming Variables

- Stata is case sensitive—so far as it is concerned, “age” and “Age” are two unique variable names
All the data is now entered, and both variables (columns) are renamed.
To formally put the data into Stata’s memory, click the preserve button.
To return to the Stata windows, click on the X button.
Age and weight now appear in the Variables window
The List Command

- The `list` command allows the user to view the entire data set
- Individual variables can also be listed—for example, `list age`
Type the command in the command window and press “enter”
Results displayed in “Results” window
The List Command

- A close-up of what appears in the review window after using the `list` command
Notice, the `list` command now appears in the Review window.
Describe Command

- The `describe` command returns output giving the user a general overview of the data set.
- The `describe` command can also be used to profile individual variable (for example, `describe age`).
Describe Command

- A close up of the review window after using the describe command

```
. describe
Contains data from C:\JMCGREADY\bio611_online 2003\point\lectures\Stata review\age_weight.dta
     obs:        8
    vars:        2
     size:       56 <99.9% of memory free>

variable name storage   display value
          type   format   label   variable label
age       byte   %8.0g
weight     int     %8.0g
```

Note: dataset has changed since last saved
Describe Command

- A close up of the review window after using the `describe` command

```
describe
Contains data from C:/MCGREADY/bio611_online 2003\ппoint\lectures\Stata review\age_weight.dta
obs: 8
vars: 2 7 Aug 2000 12:32
size: 56 (99.9% of memory free)

variable name  type format label variable label
------------------ ----------- ---- --------------- ------------
age          byte  %8.0g  
weight       int   %8.0g

Sorted by:
Note: dataset has changed since last saved
```

Information on the size of the data set.

Continued 47
Describe Command

- A close up of the review window after using the `describe` command

Information on all variables in the data set.
Saving Data

* Click on the **File** option in the upper left hand corner of your screen
File option
Saving Data

- You will see a menu appear, choose the “save as” option
Save As
Saving Data

- You will get a file chooser window
- Here, you can select the proper directory path (i.e., put the file where you want it) and name the file
Naming Data Files

- With either approach, your file will be saved with extension “.dta” (the extension for Stata data sets)
- Example—if you call your file mydata, it will be saved as mydata.dta
Leaving Stata

- When you are done for the day, and have saved your data, you may want to leave Stata.
- Click on the file option in the upper left hand corner and choose exit on the drop down menu.
Exit option on drop down menu
Section C

Reusing, Updating, and Labeling Stata Data
Retrieving Saved Data

- Now that you have created your data set, and saved it, you can also revisit it at some point in the future
- The file option allows you to open a saved Stata data set
Open option on pull down menu.
Retrieving Saved Data

- As when saving, the file chooser box will appear, allowing you to designate the directory and filename of the data set you wish to retrieve
Retrieving Saved Data

- After opening the file, the names of the variables in the data set should appear in the variable window
This is a good sign!
Labeling Variables

Sure, we have given intuitive names to the two variables in this data set; however, with large data sets containing multiple observations on the same data point (example: weight measured at birth, one year, and five years) it becomes increasingly more difficult to assign variable names that tell the “whole story”
Labeling Variables

- Luckily, Stata allows us the option of labeling variables—inputting a short descriptive phrase, which will appear in the variable window alongside the variable name.
Labeling Variables

- Syntax
  - `label variable varname "label"`

- Example
  - `label variable age "Age of subject (yrs)"`
Label now appears in Variable window
Labeling Variables

- Labeling variables is optional
  - Can do it for none, some, or all of the variables in a data set
Adding Variable(s)

- Stata data sets can be updated—more variables can be added, as can more observations
- To add more data, return to the data editor screen
Adding Variable(s)

Suppose we obtained the following information about the gender of each subject:

<table>
<thead>
<tr>
<th>Age (yrs)</th>
<th>Weight (lbs)</th>
<th>Gender</th>
<th>Age (yrs)</th>
<th>Weight (lbs)</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>123</td>
<td>F</td>
<td>43</td>
<td>115</td>
<td>F</td>
</tr>
<tr>
<td>65</td>
<td>215</td>
<td>M</td>
<td>8</td>
<td>58</td>
<td>F</td>
</tr>
<tr>
<td>31</td>
<td>187</td>
<td>M</td>
<td>76</td>
<td>126</td>
<td>M</td>
</tr>
<tr>
<td>15</td>
<td>89</td>
<td>F</td>
<td>50</td>
<td>312</td>
<td>M</td>
</tr>
</tbody>
</table>
Adding Variable(s)

- Instead of entering “Male” or “Female,” let’s enter “0” for males, and “1” for females.

<table>
<thead>
<tr>
<th>Age (yrs)</th>
<th>Weight (lbs)</th>
<th>Gender</th>
<th>Age (yrs)</th>
<th>Weight (lbs)</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>123</td>
<td>F</td>
<td>43</td>
<td>115</td>
<td>F</td>
</tr>
<tr>
<td>65</td>
<td>215</td>
<td>M</td>
<td>8</td>
<td>58</td>
<td>F</td>
</tr>
<tr>
<td>31</td>
<td>187</td>
<td>M</td>
<td>76</td>
<td>126</td>
<td>M</td>
</tr>
<tr>
<td>15</td>
<td>89</td>
<td>F</td>
<td>50</td>
<td>312</td>
<td>M</td>
</tr>
</tbody>
</table>

Continued
Adding Variable(s)

• Open the data editor and add gender data

<table>
<thead>
<tr>
<th>Age (yrs)</th>
<th>Weight (lbs)</th>
<th>Gender</th>
<th>Age (yrs)</th>
<th>Weight (lbs)</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>123</td>
<td>F</td>
<td>43</td>
<td>115</td>
<td>F</td>
</tr>
<tr>
<td>65</td>
<td>215</td>
<td>M</td>
<td>8</td>
<td>58</td>
<td>F</td>
</tr>
<tr>
<td>31</td>
<td>187</td>
<td>M</td>
<td>76</td>
<td>126</td>
<td>M</td>
</tr>
<tr>
<td>15</td>
<td>89</td>
<td>F</td>
<td>50</td>
<td>312</td>
<td>M</td>
</tr>
</tbody>
</table>
Gender in Column 3
Adding Variable(s)

- When finished entering the data, change the name of the new variable from “var3” to “Gender”
You can change the name from “var3” to “Gender” in the name field.
Adding Variable(s)

- Click on the “Preserve” button to save changes and then close the data editor
- Check your update via the “List” command
## Results from List Command

```
. list, nolab

+-----------------------+
<table>
<thead>
<tr>
<th>age   weight   Gender</th>
</tr>
</thead>
</table>
1. |  12      123        1 |
2. |  65      215        0 |
3. |  31      187        0 |
4. |  15       89        1 |
5. |  43      150        1 |
6. |   8       58        1 |
7. |  76      126        0 |
8. |  50      312        0 |
+-----------------------+
```
Defining a Label

- First, we need to create a label to apply to the variable

- Syntax (in command window)
  - `label define labelname label`

- Example
  - `label define gend 1 “female” 0 “male”`
Assigning a Label

- Once we have defined a label, we need to assign to the variable(s) for which it was created.

- Syntax (in command window)
  - label values `varname labelname`

- Example
  - label values `Gender gend`
Assigning a Label

- Now, the variable “Gender” is still stored as “1” or “0,” but will appear as “Female” or “Male”
. use "C:\MCGREADY\bio611\list"

. list

<table>
<thead>
<tr>
<th>age</th>
<th>weight</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>122</td>
<td>Female</td>
</tr>
<tr>
<td>45</td>
<td>216</td>
<td>Male</td>
</tr>
<tr>
<td>31</td>
<td>187</td>
<td>Male</td>
</tr>
<tr>
<td>33</td>
<td>13</td>
<td>Male</td>
</tr>
<tr>
<td>43</td>
<td>150</td>
<td>Female</td>
</tr>
<tr>
<td>8</td>
<td>58</td>
<td>Female</td>
</tr>
<tr>
<td>76</td>
<td>176</td>
<td>Male</td>
</tr>
<tr>
<td>50</td>
<td>312</td>
<td>Male</td>
</tr>
</tbody>
</table>
## Results from List Command

```plaintext
list

+-----------------------+
<table>
<thead>
<tr>
<th>age   weight   Gender</th>
</tr>
</thead>
</table>
1. |  12      123   Female |
2. |  65      215     Male |
3. |  31      187     Male |
4. |  15       89   Female |
5. |  43      150   Female |
6. |   8       58   Female |
7. |  76      126     Male |
8. |  50      312     Male |
+-----------------------+
```
Why All this Effort?

- Why not just type in “male” or “female” into the data editor in the first place?
  - More typing
  - Data would then be a text variable instead of numeric—less flexible for using in Stata functions
General Rule

- When categorical variable is discrete (gender, race, age group, etc.)—usually better to code numerically, then label
- When it can take on many values—(favorite food, name, etc.)—enter as text
Section D

A Quick Look at Syntax Diagrams and Running Commands with Menus
Stata has a useful help menu, which allows the user to reference information about specific commands and their syntax.

The command for getting the help menu is to type `help commandname`, where `commandname` is the name of command of interest.
Stata Syntax Diagrams: The Help Menu

- For example, if you were interested in more information about the list command, you would type `help list` in the command window.
Stata Syntax Diagrams: The Help Menu

Selected results from `help list`

```
[by varlist:] list [varlist] [if exp] [in range] [, [no]display nolabel noobs ]
```
This is a syntax diagram

[by varlist:] list [varlist] [if exp] [in range]
[, [no]display nolabel noobs ]
Stata Syntax Diagrams: The Help Menu

Let’s try to demystify!

```
[by varlist:]  list [varlist] [if exp] [in range]
[, [no]display nolabel noobs ]
```
First big rule—anything in square brackets is optional

```
[by varlist:] list [varlist] [if exp] [in range] [, [no]display nolabel noobs ]
```
Stata Syntax Diagrams: The Help Menu

♦ If we take out everything in brackets, all we have left is “list”

list

♦ This is the minimum number of “words” we need to execute the list command
Stata Syntax Diagrams: The Help Menu

- “varlist”—refers to a list of variable(s)

```
[by varlist:] list [varlist] [if exp] [in range] [, [no]display nolabel noobs ]
```

- Because “varlist” is in brackets, it means listing variables is optional
Recall from before . . .

- Typing list at the command line will list the entire dataset
- Typing list age will list the age variable only for every observation in the dataset
Stata Syntax Diagrams: The Help Menu

- [by varlist]
- Optional (of course, it’s in brackets)
- Allows user to repeat the same command over various subsets of data set

by gender: list age
Stata Syntax Diagrams: The Help Menu

- [if exp]
- (exp is abbreviation for expression)
- Optional (of course, it’s in brackets)
- Allows user to run command for observations meeting certain criteria

list age if age < 35
Stata Syntax Diagrams: The Help Menu

- [in range]
- Optional (of course, it’s in brackets)
- Allows user to run command for specific observations in dataset

list age in 5/25
Stata Syntax Diagrams: The Help Menu

- [, [no]display] nolabel noobs]
- Optional (of course, it’s in brackets)
- These are all options specific to the list command—more detail in help entry
- Generally, in syntax diagram, listings after comma are command specific options
Stata Syntax Diagrams: The Help Menu

The Take-Home Message

- “help” useful for getting detailed information on a command
- Syntax diagrams cumbersome—brackets mean optional!
Using the Menus to Run Commands

- Stata 8 is the first version of Stata to include a “built-in” menus system for running data related commands
1. <m# option or -set memory-> 1.00 MB allocated to data

use "C:\\MCGEAR\\b16611_online 2003\0point\lectures\Stata review\age_weight.dta", clear

table

<table>
<thead>
<tr>
<th>age</th>
<th>weight</th>
<th>gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>Female</td>
</tr>
<tr>
<td>2</td>
<td>62</td>
<td>Male</td>
</tr>
<tr>
<td>3</td>
<td>51</td>
<td>Male</td>
</tr>
<tr>
<td>4</td>
<td>42</td>
<td>Female</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>Female</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>Male</td>
</tr>
</tbody>
</table>