

Getting Started with Stata

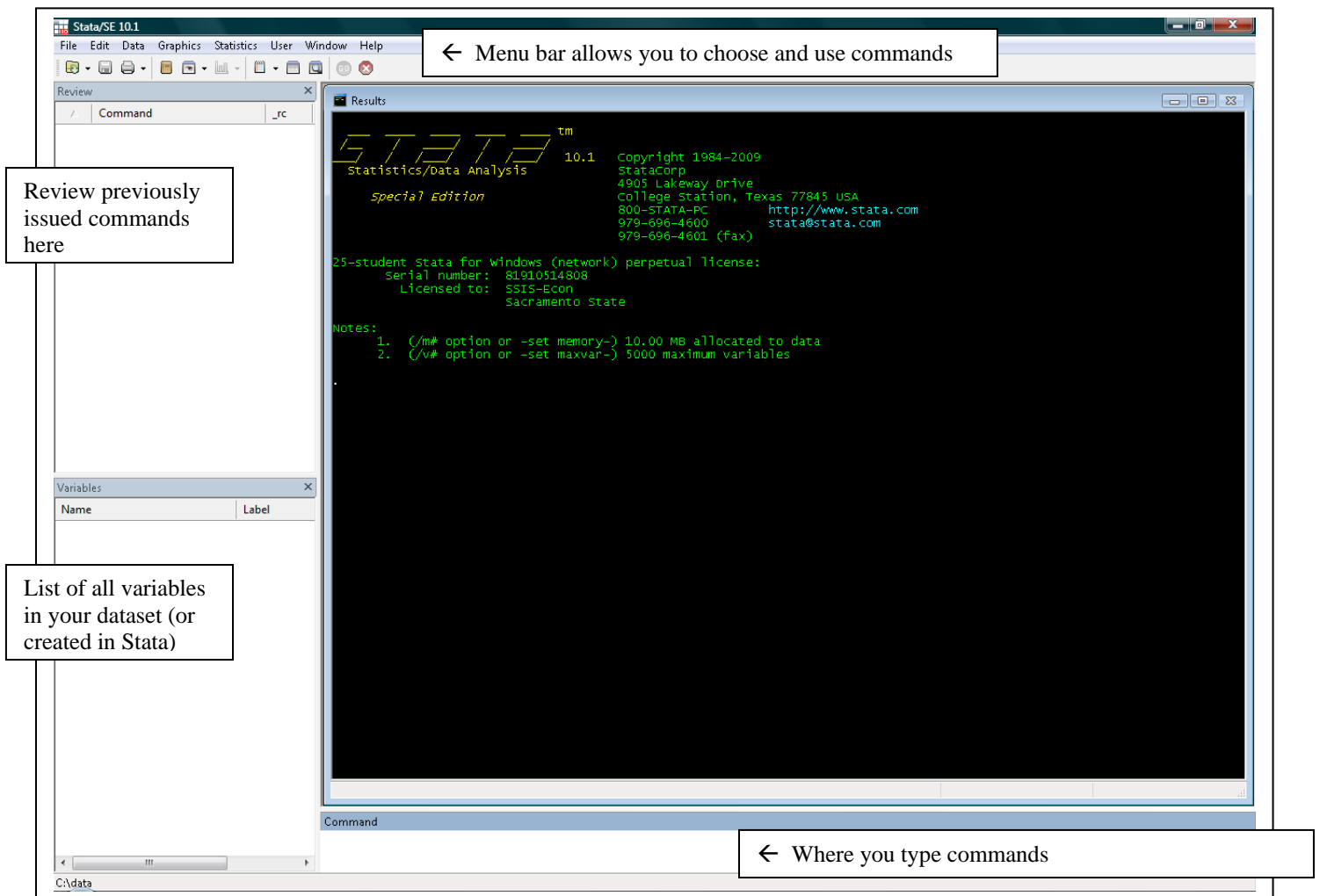
This simple introduction to Stata will allow you to open a dataset and conduct some basic analyses similar to those that we have discussed in Excel. For those who would like more information on utilizing Stata, see:

The UCLA Stata site (<http://www.ats.ucla.edu/stat/Stata/>) has a Stata starter kit that includes short videos that illustrate how to enter data, explore data, modify data, manage data, and analyze data. These videos can be useful in reviewing major commands. Please note that the videos include directions for Stata 8 and above, so they may contain some small discrepancies from Stata 9 and Stata 10.

A document at <http://fmwww.bc.edu/GStat/docs/StataIntro.pdf> provides a good overview of Stata as well as some very useful information about how to generate graphics in Stata.

A website with links to tons of additional information: <http://www-2.iies.su.se/~masa/stata.html>.

Appearance



The Stata Toolbar



Open (use)

Open a Stata dataset



Save

Save to disk the Stata dataset currently in memory



Print Results

Print a graph or the contents of the Viewer



Log Begin/Close/Suspend/Resume

Begin a new log, append to an existing log, and close or suspend the current log.



New Viewer

Open the Viewer or bring the Viewer to the front of the other Stata windows.



Bring Graph Window to Front

Bring the Graph window to the front of the other Stata windows.



New Do-File Editor

Open the Do-file Editor or bring the Do-file Editor to the front of the other Stata windows.



Data Editor

Open the Data Editor or bring the Data Editor to the front of the other Stata windows.



Data Browser

Open the Data Browser or bring the Data Browser to the front of the other Stata windows.



Clear – more – Condition

Tell Stata to continue when it has paused in the middle of long output.

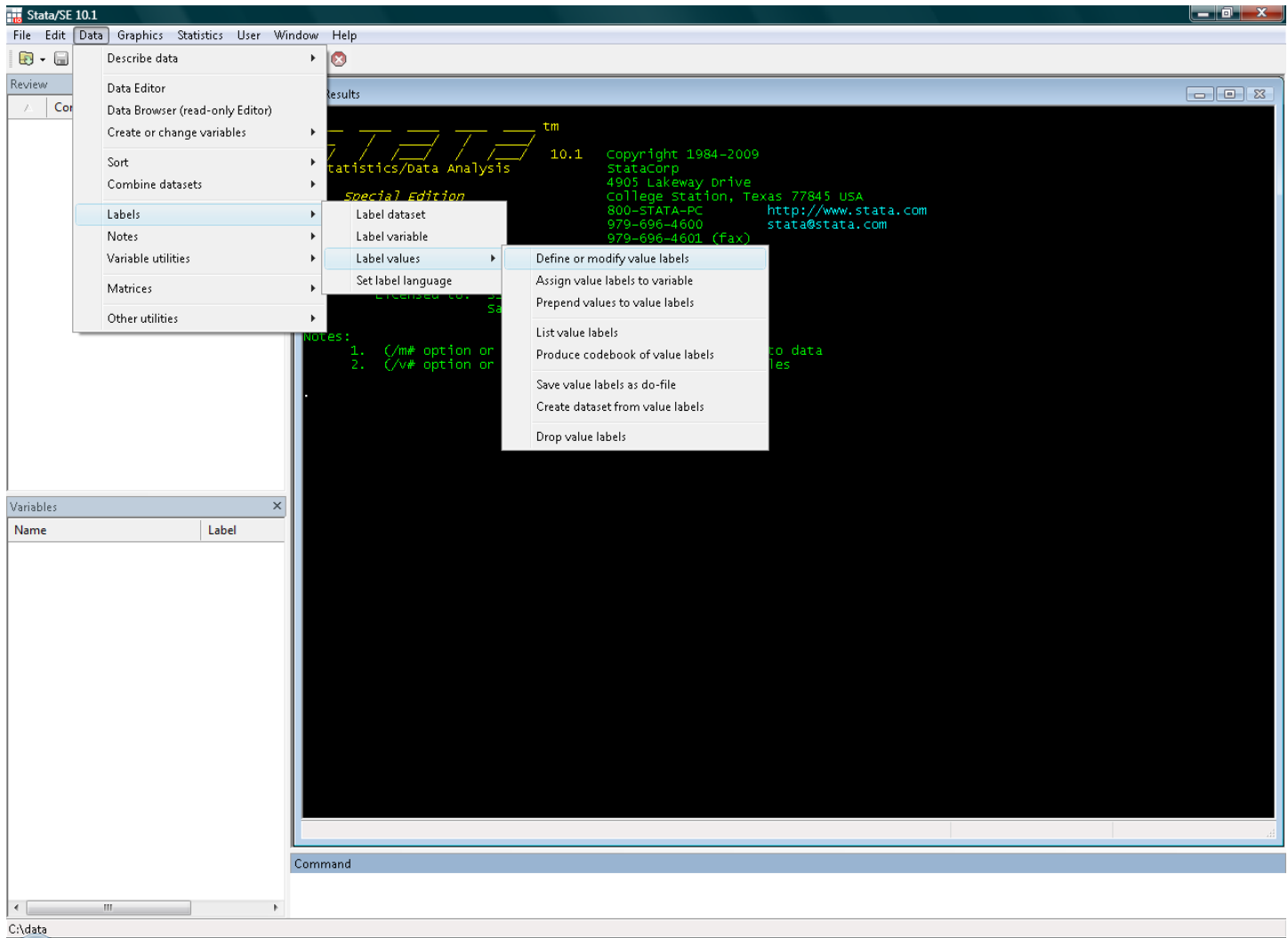


Break

Stop the current task in Stata.

The Stata Menu bar

Because it is difficult to capture screenshots of Windows menu screens and doing so would create a very long document, the following conventions will be used in this handout. Directions for the path taken from the menu bar will be given with colons in between each level of menu command. For example, to use the menu commands to define or modify value labels, you would use the following menu path:



The menu path for these commands will be expressed using this notation:

Data:Labels:Label values:Define or modify value labels

Before each command below I note its purpose and provide additional comments and instructions, which are noted with asterisks; all commands are underlined while *paths from the menu bar are italicized*. Please note that not all tasks can be performed from the Menu bar or the Tool bar, so learning the commands is very helpful.

Before you begin, you will need to download the State dataset (satscores.dta) from my website onto your computer (or the computer in the lab where you are working).

1. Open a dataset

****Open—using the Stata toolbar icon ****

use "C:\Documents and Settings\howellj\Desktop\satscores.dta"

(You will of course have to use the appropriate name and file extension for your computer)

File:Open (then navigate to the location of the data file you downloaded)

NOTE: If your dataset is sufficiently large, the default settings in Stata may not be sufficient. If you encounter an error message saying there is “no room to add more observations” immediately upon attempting to open your dataset, then type the following command in the command window: set mem 500m.

ALTERNATIVE: If you do not have a Stata dataset (e.g. a file with a .dta extension), you can still easily import your data. ****Data Editor—using the Stata toolbar icon**** and then copy/paste from Excel. Click ‘Preserve’ at the top left of the Data Editor screen and then close it. You should see your variables listed in the Variables window.

2. Open a log (record everything you do)

****Log Begin—using the Stata toolbar icon****

log using "C:\Documents and Settings\howellj\Desktop\SAT.smcl"

(You will of course have to use the appropriate name and file extension for your computer.)

File:Log:Begin

3. Commands — Managing Data and Creating Variables

****to describe the variables in terms of size, format, label**

describe

Data:Describe data:Describe variables in memory

****labeling variables**

label variable sat "aggregate SAT score"

Data:Labels:Label variable

****check to see new labels**

Describe

Data:Describe data:Describe variables in memory

****list all of the observations**

list

Data:Describe data:List data

Note- Space bar to continue; Break button or q to quit.

****list just the first 10 observations (of all variables or of just one variable)**

list in 1/10 or list sat in 1/10

(Note - This cannot be done using the Menu bar or Tool bar; it can only be done via Commands)

****sort the data by a specific variable**

sort satobs

Data:Sort:Ascending sort

****see list now sorted by satobs**

list

Data:Describe data:List data

****generate new variables**

generate male = 1-female (creates a male dummy variable)

generate white = 1-black-hisp-asian (white dummy variable)

Data:Create or change variables:Create new variable

****deleting variables (useful if you screw up)**

Drop var

4. Commands — Generating Descriptive Statistics

****Summarize all variables**

summarize

Data:Describe data:Summary statistics (leave variables list empty to summarize all variables)

****Summarize just one variable**

summarize sat

Data:Describe data:Summary statistics (choose SAT as the variable to summarize)

****Shortening command and adding some details**

summ sat, detail

Data:Describe data:Summary statistics (choose “Display additional statistics” option)

****Summarize all variables in a subsample of the data**

summarize if satobs==0 (all non-test-takers)

summarize if satobs==1 (all SAT test-takers)

Data:Describe data:Summary statistics (leave variables list empty to summarize all variables, select satobs as ‘Variables that define groups’ on “by/if/in” tab)

5. Commands — Graphing

****To save graphs, right-click on the graph and choose "save graph" in Stata; or copy by right-clicking on the graph and choosing "copy graph" (it then pastes easily into Microsoft Word).**

****Producing a Histogram**

histogram sat, freq

histogram rank, freq

Graphics:Easy graphs:Histogram

****Change the width of the bars to capture more or less variation (i.e., changing # of bins)**

histogram rank, freq bin(20)

Graphics:Easy graphs:Histogram (choose the “Options” tab to set the bins number and width)

****Producing a Histogram on a subsample of data (i.e., just SAT test-takers)**

histogram sat if satobs==1, freq

Graphics:Easy graphs:Histogram (choose the “if/in” tab to define ‘satobs==1’ subsample)

****Producing a scatterplot of Y on X (or SAT on rank)**

scatter sat rank

Graphics:Easy graphs:Scatter plot

****Producing a scatterplot of Y on X on a subsample of data (i.e., just SAT test-takers)**

Scatter sat rank if satobs==1

Graphics:Easy graphs:Scatter plot (choose the “if/in” tab to define ‘satobs==1’ subsample)

****Producing a scatterplot of Y on X and specifying a title**

scatter sat rank if satobs==1, title("SAT and Class Rank")

Graphics:Easy graphs:Scatter plot (choose the “Titles” tab)

Note - To change axes labels, change the variable labels (or choose the “Axes” tab).

6. Commands — OLS Regression

****Run a regression on your data**

regress sat rank

This command tells Stata to regress SAT score (sat) on class rank (rank). In general, regress yvar xvar1 xvar2 xvar3 tells Stata to regress yvar on xvar1, xvar2, and xvar3.

****Run a regression on a subsample with no intercept (or constant term)**

regress sat female male if satobs==1, noconstant

****Run a regression on a subsample (with an intercept, you need to have a *reference group*)**

regress sat female if satobs==1

****Run a regression and interpret *reference group***

regress sat flhs fcol if satobs==1

****Run a regression with lots of explanatory variables**

regress sat male black hisp asian rank mlhs mcol flhs fcol

****Run a regression with a binary dependent variable (Linear Probability Model)**

regress satobs rank

****Regression with a binary dependent variable (Logit Model) and generate marginal effects**

logit satobs rank

mfx compute

7. Time Series data in Stata

****Generate single time series date variable**

generate mydate=ym(year, month)

(this command assumes you have two separate variables – ‘year’ and ‘month’ – that define a monthly time series... see http://dss.princeton.edu/online_help/analysis/time_series_data.htm for other formats)

****Tell Stata you have monthly time series data**

tsset mydate, monthly

****Run regression on data within a specified date range (Jan 1987 – Dec 2007)**

Regress y x1 x2 x3 if tin(1987m1, 2007m12)

****Generate lag values of a variable x (by creating a new variable)**

generate x1lag1=x1[_n-1]

(this command uses square brackets and there is an underscore before the ‘n-1’)

****Lag values of a variable x (without creating a new variable; examine first 10 values)**

list x L.x L2.x L3.x in 1/10

(the L. operator signifies the lag of any variable, L2 lags 2 periods, L3 lags 3 periods, etc.)

****Run regression with lag values of x (without actually creating the lagged variables)**

regress y x1 L.x1 L2.x1 L3.x1

this command regresses the variable y on the contemporaneous value of x1 and three lags of x1)

****Generate Durbin-Watson statistic to test for serial correlation**

dwstat

(this command generates the Durbin-Watson based on the previous ‘regress’ command issued)

****First difference data (examine first 10 values)**

list D.y D.x1 in 1/10

(the D. operator signifies the change or first difference in any variable)

****Run regression with data in first difference form (to eliminate serial correlation)**

regress D.y D.x1

(the D. operator generates the change in any variable)

****Run regression with data in first difference form and lagged values of x**

regress D.y D.x1 D.L.x1 D.L2.x1

8. Help in Stata (Stata’s internal help or try Google)

help regress help scatter help logit