

Stata 13

HELP

Getting help

Type **help command** (e.g., help regress). If you don't know the command name, type **lookup topic** (e.g., lookup regression).

Email: tech-support@stata.com. Put your Stata serial # in the subject line of the email.

STARTING AND QUITTING STATA

To start

Double-click on the Stata icon or on a Stata dataset.

To quit

Stata menu → Quit Stata (or C-Q).

SETTING THE "SCHEME"

After opening Stata for the first time: **set scheme s1mono, permanently**

COMMANDS

Use commands rather than point-and-click

Nearly everything in Stata can be done via the menus. But you're better off typing commands into a word processing file and saving them, then cutting-and-pasting them into the Stata "Command" window.

How to enter commands in the "Command" window

Type the command or paste it in; then Enter.

To repeat an earlier command without having to retype it: Look at the list of your previous commands in the "Review" window and click on the one you want; then Enter.

DATA FILES

Creating and opening data files

To create a new data file: Click on the "Data Editor" toolbar button.

To get into an existing data file: Click on the "Open" toolbar button (or File menu → Open) and select the file.

Saving a data file

To save a data file: Exit the data editor by clicking on the X in the upper-left corner of the Stata screen, then File menu → Save (or C-S).

The extension for Stata data files is .dta.

Entering data

Type it in manually: In a Stata data file, type the datum in the appropriate cell, then Enter (or one of the arrow keys).

Copy and paste from an Excel file: In Excel, highlight the cells you want; then Edit menu → Copy (or C-C); then in the Stata Data Editor, put the cursor in the upper-left cell and Edit menu → Paste (or C-V).

To change a string variable into a numeric variable

destring *variablename*, replace

Listing data

list *variablename1 variablename2 variablename3* (option **clean** leaves out table lines)

Variable names and labels

Variable names: Click on the variable name. Variable names must begin with a letter; they can't begin with a number. Stata is case sensitive: **race** is different from **Race** or **RACE**. Variable names can't include a dash; use an underscore instead.

Variable labels: Click on the variable name.

Variable format: Click on the variable name. %8.2g indicates that the variable can stretch for 8 characters, 2 of which follow the decimal point. %8.0gc indicates that a comma will display.

Value labels: Type **label define race 1 "white" 2 "black" 3 "other"** and then Enter. Then type **label values race race** and Enter.

Creating a new variable

To create a new variable: **generate *variablename* = *operation*** (e.g., **generate unionization_sqr = unionization^2**). Once you execute this command, Stata creates the new variable and adds it to the data set. For any recomputations, use the replace command instead of generate.

To create a new variable that corresponds to observation numbers (i.e., 1, 2, 3, etc.): **generate *variablename* = _n**.

Recoding a variable

To recode a variable: First make a copy of the variable: **generate race2 = race**. Then **recode race2 3=0 2=1 1=2**.

To recode a value into a missing value: **recode race 3=.** (3=dot).

An alternative to the recode command is the replace command: First **generate income2=income**. Then **replace income2=1 if income<=10000**. Then **replace income2=2 if income>10000**.

Ordering variables and cases

To reorder variables in the data file: **order city stateabbr year population** will put those variables, in the order listed, at the beginning of the data set. Options for the order command include **alphabetic, before(), after(), first, last**.

To reorder cases in the data file: **sort city year**.

OUTPUT

Pasting output into a word processing document

In the "Results" window, highlight the output you want to copy; then Edit menu → Copy (or C-C). In the word processing document, put the cursor where you want the results to appear; then Edit menu → Paste (or C-V).

Printing output

To print from the Results window: Highlight the output you want to print; then click on the "Print" toolbar button (or C-P).

CONDITIONAL EXPRESSIONS, OPERATORS

Conditional expressions

IF command: **command if expression**. For example: **regress income educ age agesq if race==2**.

BY command: **by variablename: command**. For instance: **by race: regress income educ age agesq**.

Operators and functions

and: **&**

or: **|**

equals: **==**

does not equal: **~=**

greater than or equal to: **>=**

less than or equal to: **<=**

addition: **+**

subtraction: **-**

multiplication: *****

division: **/**

to the power of: **^**

square root: **sqrt(variablename)**

GRAPHS

Histogram

histogram *variablename*, percent bin(*numberofdesiredbars*). The **percent** option requests that relative frequencies, rather than counts, be displayed on the vertical axis. The **bin** option tells Stata the number of bars you want. For example: **graph income, percent bin(8)**.

Line graph

To add a loess curve: **scatter *variablename* year, connect(direct)**.

Scatterplot

scatter *yvariablename* *xvariablename*

To add a regression line: **scatter *yvariablename* *xvariablename* || lfit *yvariablename* *xvariablename*, connect(direct)**.

To add a loess curve: **scatter *yvariablename* *xvariablename* || lowess *yvariablename* *xvariablename*, connect(direct)**.

To spread out data points that otherwise would lie on top of each other and thus be undecipherable: **scatter *yvariablename* *xvariablename*, jitter(*number*)**. A good jitter number to start with is 7.

Scheme

When first opening Stata, type **set scheme s1mono, permanently**. For other schemes, type **help scheme**.

MISSING VALUES

How to enter missing values

Stata's recognized code for missing values is a period (.). Note, however, that Stata treats missing values as larger than nonmissing values, so beware when using the generate or if commands. It's probably best to leave missing values blank.

STATISTICAL ANALYSIS

Coefficient of variation

**summarize *variablename*
display r(sd) / r(mean)**

Correlation

correlate *variablename1* *variablename2* *variablename3*
For pairwise deletion: **pwcorr *variablename1* *variablename2*
*variablename3***

Crosstabs

tabulate *rowvariablename columnvariablename, column*. The column option requests column percentages. Usually with crosstabs we put the (presumed) y variable as the row variable and the x variable as the column variable, and then examine the column percentages.

Descriptive statistics

summarize *variablename*. Shows mean, standard deviation, smallest value, largest value.

summarize *variablename, detail*. In addition to mean etc., this shows the median (50th percentile), some other percentiles (1st, 5th, 10th, 25th, 75th, 90th, 95th, 99th), variance, skewness statistic, kurtosis statistic.

Frequency distribution

tab1 *variablename1 variablename2 variablename3*

Regression (OLS)

regress *yvariablename x1variablename x2variablename*

The option **beta** shows standardized coefficients.

The option **robust** adds heteroskedasticity-consistent standard errors.

The option **level(80)** shows 80% confidence intervals for coefficients (default is 95).

Skewness statistic

summarize *variablename, detail*

Z-score (standardized score)

summarize *variablename*

generate *variablename_zscore = (variablename - r(mean)) / r(sd)*

UPDATING STATA VIA THE WEB

In the "Command" window, type **update all**