

Data Processing

with Stata 14.1 Cheat Sheet

For more info see Stata's reference manual (stata.com)

Useful Shortcuts

F2 — keyboard buttons

describe data

Ctrl + **9**

open a new .do file

Ctrl + **8**

open the data editor

Ctrl + **D**

highlight text in .do file,
then ctrl + d executes it
in the command line

clear

delete data in memory

AT COMMAND PROMPT

PgUp **PgDn** scroll through previous commands

Tab autocompletes variable name after typing part

cls clear the console (where results are displayed)

Set up

pwd

print current (working) directory

cd "C:\Program Files (x86)\Stata13"

change working directory

dir

display filenames in working directory

fs *.dta

List all Stata data in working directory

capture log close

close the log on any existing do files

log using "myDoFile.txt", replace

create a new log file to record your work and results

search mdesc

find the package mdesc to install

ssc install mdesc

install the package mdesc; needs to be done once

Import Data

sysuse auto, clear

load system data (Auto data)

for many examples, we
use the auto dataset.

use "yourStataFile.dta", clear

load a dataset from the current directory

import excel "yourSpreadsheet.xlsx", /*

import an Excel spreadsheet

import delimited "yourFile.csv", /*

import a .csv file

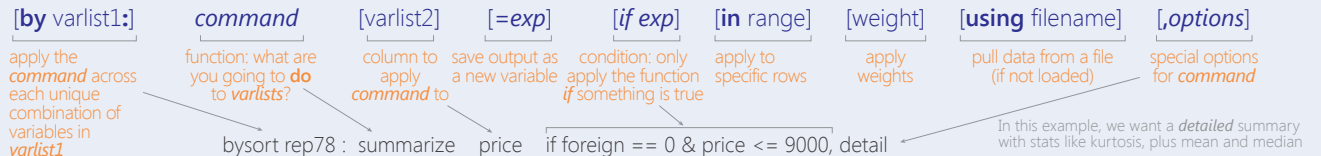
webuse set "https://github.com/GeoCenter/StataTraining/raw/master/Day2/Data"

webuse "wb_indicators_long"

set web-based directory and load data from the web

Basic Syntax

All Stata functions have the same format (syntax):



To find out more about any command – like what options it takes – type **help command**

Basic Data Operations

Arithmetic

+ add (numbers)
combine (strings)

- subtract

***** multiply

/ divide

^ raise to a power

Logic

& and

! or **~** not

| or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

or

== tests if something is equal

< less than

<= less than or equal to

> greater than

>= greater or equal to

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

if foreign != 1 | price >= 10000

if foreign != 1 & price >= 10000

count

count if price > 5000

number of rows (observations)

Can be combined with logic

ds, has(type string)

lookfor "in."

search for variable types,

variable name, or variable label

isid mpg

check if mpg uniquely

identifies the data

histogram mpg, **frequency**

plot a histogram of the

distribution of a variable

codebook make price

overview of variable type, stats,

number of missing/unique values

summarize make price mpg

print summary statistics

(mean, stdev, min, max)

for variables

inspect mpg

show histogram of data,

number of missing or zero

observations

histogram mpg, **frequency**

plot a histogram of the

distribution of a variable

codebook make price

overview of variable type, stats,

number of missing/unique values

summarize make price mpg

print summary statistics

(mean, stdev, min, max)

for variables

inspect mpg

show histogram of data,

number of missing or zero

observations

histogram mpg, **frequency**

plot a histogram of the

distribution of a variable

codebook make price

overview of variable type, stats,

count

count if price > 5000

number of rows (observations)

Can be combined with logic

ds, has(type string)

lookfor "in."

search for variable types,

variable name, or variable label

isid mpg

check if mpg uniquely

identifies the data

histogram mpg, **frequency**

plot a histogram of the

distribution of a variable

codebook make price

overview of variable type, stats,

number of missing/unique values

summarize make price mpg

print summary statistics

(mean, stdev, min, max)

for variables

inspect mpg

show histogram of data,

number of missing or zero

observations

histogram mpg, **frequency**

plot a histogram of the

distribution of a variable

codebook make price

overview of variable type, stats,

number of missing/unique values

summarize make price mpg

print summary statistics

(mean, stdev, min, max)

for variables

inspect mpg

show histogram of data,

number of missing or zero

observations

histogram mpg, **frequency**

plot a histogram of the

distribution of a variable

codebook make price

overview of variable type, stats,

count

count if price > 5000

number of rows (observations)

Can be combined with logic

ds, has(type string)

lookfor "in."

search for variable types,

variable name, or variable label

isid mpg

check if mpg uniquely

identifies the data

histogram mpg, **frequency**

plot a histogram of the

distribution of a variable

codebook make price

overview of variable type, stats,

number of missing/unique values

summarize make price mpg

print summary statistics

(mean, stdev, min, max)

for variables

inspect mpg

show histogram of data,

number of missing or zero

observations

histogram mpg, **frequency**

plot a histogram of the

distribution of a variable

codebook make price

overview of variable type, stats,

number of missing/unique values

summarize make price mpg

print summary statistics

(mean, stdev, min, max)

for variables

inspect mpg

show histogram of data,

number of missing or zero

observations

histogram mpg, **frequency**

plot a histogram of the

distribution of a variable

codebook make price

overview of variable type, stats,

count

count if price > 5000

number of rows (observations)

Can be combined with logic

ds, has(type string)

lookfor "in."

search for variable types,

variable name, or variable label

isid mpg