

An overview of Stata

Interface commands

- Working directory
 - Set working directory: `cd "c:\path\to\file"`. You must make the slashes appropriate to your OS
 - Get the current working directory: `pwd`
- Read a dataset
 - Stata-formatted: use `filename.dta` or use `"c:\path\to\file\filename.dta"` if the file is not in your working directory
 - Comma-delimited (CSV): `insheet using filename.csv, comma`
 - Tab-delimited: `insheet using filename.txt, tab`
- Export a dataset
 - Stata format: `save using filename.dta, replace`. The `replace` option overwrites the current file
 - Delimited: `outsheet using filename.txt, comma`. Use the `, tab` option if you want a tab-delimited file. The extension doesn't really matter
- Get help on a command: `help commandname`

Basic statistical commands

- Summary statistics
 - Get summary statistics for all variables: `sum`
 - Get summary statistics for a list of variables: `sum var1 var2 var3, detail`. The `detail` option gives more information about the distribution of the variables
- Examine a two-way table: `tab x1 x2`
- Correlation matrix: `correlate x1 x2 x3`
- Basic graphs
 - Histogram of `var1`: `hist var1`

– Plot of `var1` against `var2`: `scatter var1 var2`

- Subsetting commands

– Look at the results of a command according to levels of a grouping variable

```
bysort groupvar: commandname
```

– Look at the results of a command for observations that meet a particular criterion:

```
commandname, if (var1 <= 5 & var2==3)
```

A useful application is `commandname, if !missing(var1)`, which only performs `commandname` on observations without missing values for `var1`

Data manipulation

- Logical operators

a and b	a & b
a or b	a b
a = b	a == b
a ≠ b	a != b
a ≥ b	a >= b

- Create a new variable

```
gen newvar = var1/100 + var3^2 + ...
```

- Recode an existing variable

```
replace x = 0 if x <= 5  
replace x = . if y!=3
```

The “.” stands for missing value

- Recode many variables in the same way:

```
foreach x in var1 var2 var3 {  
  replace 'x'=1 if 'x'==5  
  replace 'x'=2 if 'x'==4  
  replace 'x'=0 if (!missing(x) & ('x'!=5 & 'x'!=4))  
}
```

This changes all 5s and 4s to 1s and 2s, and maps everything else that isn't missing to zero (for the variables `var1` through `var3`)

Basic modeling commands

- Regressions
 - OLS: `reg y x1 x2 x3, robust`. The robust option gives Eicker-Huber-White standard errors. You can also use `, cluster(groupvar)` to cluster the standard errors at the level of `groupvar`
 - 2SLS: `ivregress 2sls y (x1 x2 = z1 z2) x3 x4`. This treats `x1` and `x2` as endogenous, and uses `z1` and `z2` as instruments
 - Probit: `probit y x1 x2 x3`
 - Logit: `logit y x1 x2 x3`
- Post-estimation objects and commands
 - Test a hypothesis: `test x1 x2 x3` tests the hypothesis that $\beta_1 = \beta_2 = \beta_3$. `test x1=x2=0` tests $\beta_1 = \beta_2 = 0$
 - Generate predicted values after running a regression: `predict yhat`. If you have run a Probit or Logit, you can use `predict yhat, pr` to get predicted probabilities instead of predictions of $x_i\beta$
 - Generate residuals: `predict newvar, resid`
 - See all post-estimation objects available: `ereturn list`

I will introduce new commands as we learn new estimation procedures

A do file template

```
set more off
clear
capture log close
cd "c:\path\to\files\"
//log using XX.log , replace
//set mem XXm
/*-----*/

use filename.dta

// commands go here

/*-----*/
capture log close
```

`set more off` tells Stata to keep running even if the screen fills up. `clear` drops any data currently loaded into memory. `capture log close` turns off any old log that might be running, and `log using...` tells Stata to save a record of all of your commands (and the results). `//` and `/* comment here */` are comments (the latter can span multiple lines).

Other (free) Stata resources

- “Stata tutorial.” German Rodriguez. <http://data.princeton.edu/stata/>
- “An introduction to Stata 8.” Svend Juul. <http://www.folkesundhed.au.dk/uddannelse/software/besked.pdf>.
- “A little bit of Stata programming goes a long way.” Christopher Baum. <http://fmwww.bc.edu/EC-P/WP612.pdf>.