A brief introduction to Stata

First steps in Stata

– Opening Stata
– Getting to know the different windows
output, command, review, variables, toolbox
– How to get help ?
yet another window : the viewer
– The general command syntax
[prefix :] `command` *[varlist] [=exp] [if] [in] [weight] [using filename] [, options]*
– Making oneself comfortable and opening a dataset
changing the directory, setting the memory (optional in most situations), `use` *for Stata datasets,* `insheet` *for csv files,* `infile` *for other formats, parenthesis about StatTransfer*
– Saving a file
`save[, replace]`
– The best way to feed Stata with commands : do-files
how to manage them and why it is usually better than just the command window

The stuff that data are made of

– What's in a dataset ?
variables, formats, `describe`
– Looking into the data
`browse`, `edit`, `list` *and its (most useful) options, parenthesis about logical operators and missing data*
– Some very useful commands to describe data
`summarize` *for basic univariate statistics,* `codebook` *for a general description,* `tabulate` *for tables of frequencies (discrete variables),* `corr` *for a matrix of correlations, parenthesis about abbreviations :* `su` *for* `summarize` *etc.*
– Your first graph
`graph` *for the general command, the simplest (and most useful) type of graph :* `scatter`
– Your first regression
`reg` *depvar [indepvars] [if] [in] [weight] [, options]*

Managing data

– Creating / modifying a variable
`generate` *and* `egen`, `replace`, *example : dummy variables (the tedious way / the easy way),* `rename`, `recode`, *accessing lagged values (e.g. with [_n-1])*
– Keeping / dropping observations or variables
`keep`, `drop`
– Sorting your data
`sort`, *why sort the data ? The* `by` *option, example :* `by` *varname :* `summarize` *varname*
– The dangerous `collapse` command
– Merging or appending datasets
`merge`, `joinby`, `append`

– Labels
– Dealing with string format variables
  Some string functions and string operators : `string`, `destring`, *substr to take a substring, trim, + to concatenate strings..*

A quick glance at more advanced stuff

– Loops and macros
  `foreach`, `forvalues`, *accessing the results of a command*
– Keeping a log file
– Reshaping a dataset
  `reshape`
– Dealing with time-series format variables
– Post-estimation commands
  hypothesis testing, predictions and much more
– Programming your own commands
  amazingly useful and not as complicated as it sounds

References

– `http://rlab.lse.ac.uk/data/stata_info/`
  for tutorials (one by Arnaud Chevalier, the other by Martin Stewart : both are pretty good and go quite far), exercices and other resources
– `http://www.ats.ucla.edu/STAT/stata/`
  for a very comprehensive website with online lectures, tutorials etc.
– `http://www.stata.com/statalist/index.html`
  Statalist is a mailing-list run by Stata-addicts, who are ready to help whoever has a problem with Stata