

AN EFFICIENT INTRODUCTION TO STATA

(Pierre DUBOIS, dubois@toulouse.inra.fr)

- To start STATA, run the file `c:\stata\wstata.exe`
- Program file "profile.do" in directory of wstata.exe automatically executed when STATA starts: example

```
Set mem 50m                                sets memory at 50 megabytes (larger than data file size)
Set matsize 800                             sets matrix size to 800
Sysdir set PERSONAL "c:\stata\ado\personal"  sets directory system PERSONAL as "c:\stata\ado\personal"
Sysdir set STBPLUS "c:\stata\ado\stbplus"    sets directory system STBPLUS as "c:\stata\ado\stbplus"
cd c:\Data                                  Define the current working directory c:\Data
```
- Regularly update your STATA software by typing the command `update query`
- Several windows can be opened through the window list of the menu: command, results, review, variables, graph, viewer, data editor and .do file editor.
- Data files have extension `.dta`
- Program files can have extension `.do` or `.ado`
- .do files must be in working directory and `prog.do` is run with the command: `do prog`
- .ado files are searched in sub-directories of `c:\stata\ado\` and are run directly with the name of the program as a command.
- STATA works with observations, variables, scalars, labels, macro variables, temporary variables and programs.
- STATA commands are separated by paragraph return or another sign
- Help for STATA: commands `help command name` and `search keyword`. Use the STATA help to learn STATA. STATA allows to search programs and commands available on the net with `webseek keyword`. `webseek` displays links for installation.
- Input data with data editor or use `infile` or `insheet` commands to read non STATA files.
- Read STATA files with command `use file name` (transform formats with `Stattransfer` or `Dbmscopy`)
- Save a file with command `save file name` (the default extension is `.dta`)
- Record results in a `.log` or `.smcl` file (see also `help translatetext`): open log file with `log using resultsname`, close recording log file with `log close`
- Description of variables in memory: `describe` or `d`. Also: `codebook`
- Compression of data in memory space: `compress`
- Data are recorded as observations (rows) for which variables (columns) are defined.
- Missing observation is coded with `.`
- Observations are numbered with system variables `_n` and total number is `_N`
- Commands syntax: `prefix : command name suffix , options`
- Use of prefix: `by varname : for data sorted by command: sort varname`
- Suffix: `if condition` or `in 20/30` (range of observations)
- Useful commands (see STATA help corresponding):

- Descriptive statistics commands: codebook, list, tabulate **or** tab, table, summarize **or** sum, corr, reshape, pctlile, compare,
- Variable management commands generate **or** gen, egen, replace, recode, rename, drop, keep, label, creating interactions with xi, doing loops with for
- Data management commands sort, merge, append, compress, save, collapse, reshape
- Econometric commands : regress **or** reg, ivreg, mvreg, sureg, mvreg, qreg, probit, logit, mlogit, tobit, predict, test, lrtest, testnl, streg, stcox, _robust,
- Panel data commands (most of them begin with xt): tsset, xt, xtdata, xtdes, xtglsl, xtivreg, xtlogit, xtpois, xtprobit, xtreg, xtsum, xttab, xttobit. Time series commands : ac, arima, arch, .
- Survey commands for data using stratification, clustering and sampling weights (beginning with svy): svyset, svydes, svyreg, svyivreg, svylogit, svypois, svyprobit, svysum, svytab, ..
- Graph commands: graph **or** gr, grhist, kdensity, cumul,
- STATA creates graphs in format .gph, .ps **or** .eps (see help translategph). Can also copy and paste directly a graph in a word processor. A .wmf file is created automatically if pasting in Scientific Word.

- STATA programs:

- .do files: set of successive commands saved in prog.do and run with command do prog
 - Defining a global variable: global globname azerty, then \$globname is equal to azerty
 - Defining a local variable: local localname azerty, then `localname' is equal to azerty
 - Useful macro commands: tokenize, macro shift, ..

- Program definition as commands: example

#delimit;	<i>changes the command delimiter to ;</i>
program define myprog ;	<i>beginning of program definition</i>
version 7.0;	<i>Stipulates version 7.0 for the syntax of the program</i>
syntax [varlist] [if] [in] [, Beta(real 1.0)];	<i>program uses macros `varlist', `if', `in', `dof', and `beta'. `beta' defaults to 1.0</i>
if condition1 { ; commands ;};	<i>conditions</i>
else if condition2 { ; commands ; }; else { ; commands ;};	
end;	<i>end of the program definition</i>

- If recorded in an .ado file, automatically searched in any other STATA session.
 - Program commands: program dir, program drop myprog, program list
- Maximum likelihood estimation: see help ml and STATA documentation
- Non linear least squares estimation: see help nl and STATA documentation
- Simulation command: simul. Bootstrap commands: sample, bsample, bs, bstrap,
- Debugger: set trace on, set trace off
- DOS commands can be used inside STATA with : ! DOS command
- Very useful web sites:
 - STATA official web site: <http://www.stata.com/>
 - Resources for learning STATA: <http://www.stata.com/links/resources1.html>
 - UCLA resources to help you learn and use STATA: <http://www.ats.ucla.edu/stat/stata/>