

# Assignment Project Exam Help

ECON3350/7350: Applied Econometrics for  
Macroeconomics and Finance

<https://powcoder.com>

Eric Eisenstat

The University of Queensland

Add WeChat powcoder

Tutorial 1: Introduction to Stata

# What Makes Stata Popular?

Stata is a powerful statistical package for applied economics.

- straightforward commands and simple syntax
- easy to code programs and record results
- convenient to control execution process
- powerful data management and analysis
- update frequently: `update query`
- many functions tailored for applied economics
  - robust and clustered SE, marginal effects, prediction, etc.
- infinite extensibility and large user community
  - code your own commands if necessary
  - download user-written commands
  - statistical software component (SSC): `ssc install pkgname`

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Useful Stata References

## Getting Started with Stata - for first-time users

- <http://www.stata.com/bookstore/getting-started>

The Stata Users Guide and Forum

- <http://www.stata.com/bookstore/users-guide>
- <http://www.statalist.org>

## Textbooks

- A Gentle Introduction to Stata (2014)
- Microeconometrics Using Stata (2010)
- Statistics with Stata v12 (2012)

## Online Tutorials

- Princeton: <http://data.princeton.edu/stata>
- UCLA IDRE: <https://stats.idre.ucla.edu/stata/>

## Basic Stata command syntax

```
[prefix:] cmd [varlist][=exp][if][in][weight][using filename][, options]
```

# Assignment Project Exam Help

- Stata is case sensitive
- Stata (default) views each line as a separate command

R-class commands: data analyzing commands

- **summarize**, **describe**, **tabulate**, **tabstat**, **codebook**, **return list**

E-class commands: estimation commands

- **regress**, **nl**, **ivregress**, **logit**, **probit**, **gmm**, **ereturn list**

Shortcuts

- wildcard \*: e.g., **drop v\* z\***
- abbreviation: e.g., **reg = regress**, **var1-var7**, **\_all**

# <https://powcoder.com>

# Add WeChat powcoder

# Template Do-file

Do-file is a text file collecting Stata commands, which can be executed by Stata when you type `do filename`.

## Head of Do-file

```
clear all // clear the memory
capture log close // just in case!
set more off // pause for the "more"
set matsize 5000 // set max number of variables
cd "your_path" // change the working path
log using logfilename.log, replace // start log-file
use your_raw_data, clear // read data into Stata
```

## End of Do-file

```
save your_modified_data, replace // save the modified data
log close // close log-file, stop recording
```

**capture:** e.g., capture log close

**preserve and restore**

- preserve: preserve current data when executing commands
- restore: restore the preserved data

**quietly:** prefix execute command but suppress output

Loops: **forvalues**, **foreach**, **while**

Add comments to your do-file

- single line comment: \*
- multiple line comment: /\* ... \*/
- single line after command: //
- use /\* ... \*/ to span one-line command to several lines

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Arithmetic operators:  $+$ ,  $-$ ,  $*$ ,  $/$ ,  $^$

- $+$  can be also applied to strings,  $"A" + "B" \rightarrow "AB"$

Logic operators:

- $\&$  (and),  $\|$  (or),  $!$  (not),  $==$  (equal),  $!=$  (not equal),  $>=$ ,  $<=$ ,  $>$ ,  $<$

Range: in 1/10, in 0.1(0.1)1, in -10/-1

Double quote: string and path contains space(s)

- drop if state `== "QLD"`
- `cd "c:/Users/My Documents/Data"`

- Always use do-file and keep log-file

- Keep a copy of your raw data in a safe place

- Create a separate folder for each project and put all the stuff (data, do-file) in it

- Divide and rule: divide a complex task to several simple tasks, create do-file for each of them, and collect all do-files in a master.do file

- Add comment to help your group members and yourself understand the code

- Watch missing data carefully and make sure you know where missing values come from

- Use loops as much as possible, but not copy and paste codes

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



## Assignment Project Exam Help

- `help cmdname`: useful if user knows the command name
- `search keyword`: keyword search, can be local or net or both
- `findit keyword = search keyword, all`: broadest keyword search, can be used for searching user-written commands
- `hsearch`: whole word local search, useful if you want to know what Stata can do for a particular task
- Ask for help online: <http://www.statalist.org>

<https://powcoder.com>  
Add WeChat powcoder

# Input and Output Data

Import data saved in Stata format (.dta file)

- **use** filename, **clear**

Import data saved by Excel

- save the data as a .csv (ASCII, comma delimited) file and
- **insheet** [varlist] **using** filename.csv, **clear** comma nonames/names

Import data saved by other softwares (SPSS, SAS, Matlab, etc)

- Stat/Transfer (not free), then read data with **use**
- or export the data to ASCII format and use **insheet**

Save data in Stata format

- **save** filename, **replace** - save in Stata format
- **saveold** filename, **replace** - save for Stata of older versions

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Familiarize Yourself with the Data

## Browse the data

- `describe`

- `set [varlist] if fin, sepb()`

- missing values: be very careful! typically coded as ".", larger than any real number for Stata

- `codebook`: generate a codebook for all variables and get a quick overview

## Summary statistics

- `summarize [varlist], detail`

- `tabstat [varlist], stat() col(stat) by()`

- `tabulate varlist, row col chi2` (two-way)

- `table varlist` (two-way and three-way tables)

- `table rowvar, contents(N colvar mean colvar ...)`

# Manipulate the Data

## Rename and label variables

- **rename** oldname newname
- **label** variable varname
- **label define** lbnam #1 label1 #2 label2 ...
- **label values** varlist lbnam

## Generate new variables

- **gen** newvar = exp [if] [in]
- **egen** newvar = egenfcn(arguments) [if] [in]
- **egen** functions: **total()**, **max()**, **count()**, **mean()**, etc.
- install user-written egen functions: **ssc inst egenmore**

## Replace values of existing variables

- **replace** oldvar = exp [if] [in]
- **recode** varlist (rule), **gen**(newvarlist)

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

# Manipulate the Data (cont'd)

Drop observations and variables

- `keep` varlist / `[if]` `[in]`

- `drop` varlist / `[if]` `[in]`

System variables

- `_n` refer to the current observation number.
- `_N` refer to the highest observation number (= sample size).
- `_cons`/`_b`(varname) refer to estimated intercept and slope.

By-prefix: repeat Stata command on each subsets of the data

- `by`sort varlist: stata\_command

Sort data

- `sort` varlist (ascending order)
- `gsort` (+/-) var1 (+/-) var2, `gen`(newvar)

# Manipulate the Data (cont'd)

## Forward and lag variables

- `gen fvar = f.var`
- `gen lvar = l.var`
- `f2, l2, f3, l3, ...`
- must use after `xtset` panelvar timevar or `tsset` timevar
- any difference from `gen lvar = var[_n-1]`?

## Dummy variables and interaction terms

- `xi i.var, xi i.var1*i.var2, xi i.var1*i.var2`
- generate dummy variables using `if`
- generate dummy variables using logic operators
- watch missing data! use `if` is preferred

# Commands Handling Datasets

Merge datasets: create wider dataset

- `merge 1:1 / 1:m / m:1 / m:m varlist using filename [, nogen]`

Append datasets: create longer dataset

- `append using filename`
- two datasets must have the same variables

`collapse`: make dataset of summary statistics

- sometimes it is better to use `egen`

Reshape datasets: transform between long and wide data datasets

- wide data vs. long data
- `reshape wide varlist, i(id) j(period)`
- `reshape long varlist, i(id) j(period)`

## Histogram

- `histogram varname [options]`
- describe probability density (or mass) functions
- can be used with `kdensity` option

## Kernel density plot

- `kdensity varname[, options]`
- a better alternative to histogram for continuous variable

## Scatter plot

- `scatter var1 var2 [, options]`
- provide quick look at the relationship between 2 variables
- `graph matrix`: bivariate scatterplots between several variables

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



# Edit, Save and Export Your Graphs

## Edit graphs

- add more elements to the graph
  - e.g., `graph twoway (scatter v1 v2) (lfit v1 v2)`
- combine multiple graphs into a single figure

- `graph combine` graph1 graph2
- graph1 and graph2 must be saved first in `.gph` format

- edit graphs using options: `title`, `yaxis`, `xaxis`, `text`, `legend`, etc
- edit graphs using Stata's point-and-click graph editor

## Save and export graphs

- save graphs in `.gph` format (Stata default/internal format)
  - `graph save` graph
- save graphs in other formats (recommend `.png` or `.eps`)
  - `graph export` graph.suffix

# Command List for this Class

The following commands will be used in your lectures and/or PS. We will revisit them for more details in future.

## Estimation commands

- OLS and 2SLS: `regress`, `ivregress`
- MLE: `logit`, `probit`, `poisson`
- NLS: `nl`
- GMM: `gmm`

## Post-estimation commands

- prediction: `predict`, `predictnl`, `lincom`, `nlcom`
- marginal effects: `margins`
- test: `test`, `testnl`, `lrtest`
- estimation results: `estimates store`/`restore`, `estimates save`/`use`
- post-estimation statistics: `estat summarize`, `estat vce`, `estat ic`

# Assignment Project Exam Help

# <https://powcoder.com>

# Add WeChat powcoder

## Assignment Project Exam Help

Stata functions

- quick reference for Stata functions

Generate random samples

- `set obs 500` // setup sample size
- `gen id = _n` // generate id indicator
- `set seed 10001` // set seed for a random draw
- `gen X = rnormal(0,1)` // generate 500 samples of  $N(0,1)$

<https://powcoder.com>

Add WeChat powcoder