

Quantitative Techniques – Term 2

Laboratory 1

19 January 2006

Overview

The objectives of this lab are:

- Familiarize with Stata's basic functions;
- Calculate summary statistics;
- Estimate a linear regression model with OLS.

Start Stata 9 from the Windows Start menu (Programs, F. Departmental Software, E. Social Science, Stata).

Do not update the version that is installed on your machine.

Before starting your tasks, make sure that the following windows are visible on your screen and choose their size so that they do not overlap: Review, Variables, Results, Command.

Task 1 – Import a data file¹

- Go to www.staff.city.ac.uk/a.banal-estanol/teaching.htm and download Lab1.csv to your directory.

Note that the names of the variables are in the first row, immediately followed by your data. If you do not specify names, then Stata will use var1, var2, etc. You can change the names later, if you need to.

The source of the file is Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data*, Chapter 4, Example 4.1.

- Type in the Command window:

insheet using “*your directory*\Lab1.csv”

This command loads the data in Stata. Note that Stata recognizes commands in lower case letters. Use “” when you specify a filename.

Alternatively, you can use the Menu bar: File, Import, ASCII data created by a spreadsheet.

- Save the file as Lab1.dta. This is the extension used by Stata for datasets.
- To view your data, on the Menu bar click on Window, Data Editor. Alternatively, look for the Data Editor icon on the bar (it is on the right-hand side).

¹ Reference: <http://www.stata.com/help.cgi?infilng>

In the Variables window, you can see the names of the variables in the dataset:

- wage: wage of married women
- exper: years of experience
- exper2: years of experience squared
- educ: years of schooling
- age: woman's age
- kidslt6: number of children less than six
- kidsge6: number of children at least six years of age
- inlf: in labour force

If you already have a Stata dataset (e.g. download Lab1.dta directly from www.staff.city.ac.uk/a.banal-estanol/teaching.htm), then you can open it by either:

- Open – using the Menu bar or the icon on the toolbar; or
- Type in the command window:

use "your directory\Lab1.dta"

Again, note the use of “ ” when you specify a filename.

Task 2 – Descriptive statistics

- Use the command **summarize** to obtain the mean and the standard deviation of the variables in your dataset. You can also use this command with a subset of variables;²
- You may want to obtain summary statistics for separate groups, e.g. summarize the wages of married women depending on the number of children. Type:³

tabulate kidslt6, summarize(lwage)

- Calculate the pair-wise correlations between all your variables. Type:⁴

correlate

You can also use this command with some specific variables, e.g. **correlate lwage kidslt6**.

Task 3 – OLS regression

By using OLS, estimate the parameters of the equation⁵

$$lwage = a + b * exper + c * expersq + d * educ + e * age + f * kidslt6 + g * kidsge6 + u$$

- First you need to generate lwage, by typing **gen lwage=ln(wage)**
- Then estimate the above model by typing reg, followed by the dependent variable and all the regressors:

reg lwage exper expersq educ age kidslt6 kidsge6

² <http://www.stata.com/help.cgi?summarize>

³ http://www.stata.com/help.cgi?tabulate_summarize

⁴ <http://www.stata.com/help.cgi?correlate>

⁵ <http://www.stata.com/help.cgi?regress>

You can also click on the variables in the left-hand side Variables window. The output is the same as in Excel, but it is organized differently. Can you spot which coefficients are significant?

Note: you don't need to specify the constant, as Stata includes it automatically.

Source	SS	df	MS			
Model	35.3398089	6	5.88996815	Number of obs =	428	
Residual	187.987632	421	.446526442	F(6, 421) =	13.19	
Total	223.327441	427	.523015084	Prob > F =	0.0000	
				R-squared =	0.1582	
				Adj R-squared =	0.1462	
				Root MSE =	.66823	

lwage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
exper	.039819	.013393	2.97	0.003	.0134936	.0661444
expersq	-.0007812	.0004022	-1.94	0.053	-.0015718	9.37e-06
educ	.1078319	.0144021	7.49	0.000	.079523	.1361409
age	-.0014653	.0052925	-0.28	0.782	-.0118682	.0089377
kidslt6	-.0607106	.0887626	-0.68	0.494	-.2351836	.1137625
kidsge6	-.014591	.0278981	-0.52	0.601	-.069428	.0402459
_cons	-.4209078	.316905	-1.33	0.185	-1.043821	.2020053

- Copy your output into Word⁶

This is an easy way to save your output, but you also have the option to create a so-called log file in which the output in the 'Stata results' window is copied.

Select the table with your results and right click. Choose 'copy, text' from the menu. Then open a Word document and Paste. Select the text in Word, and then change the font to 'Courier' or 'Courier New' in order to preserve the right column width.

- Calculate fitted values⁷

In order to obtain the estimated values for the log of wage, type the command:

predict name of variable

e.g. predict fitted

The estimated values will be created as a new variable and you can see it in the Data Editor. The default of the 'predict' command is given by fitted values.

- Calculate fitted residuals

You can use the same command as above, but this time you need to specify what 'prediction' you need:

predict name of variable, residuals

e.g. predict resid, residuals

⁶ <http://www.ats.ucla.edu/stat/stata/faq/outgraph.htm>

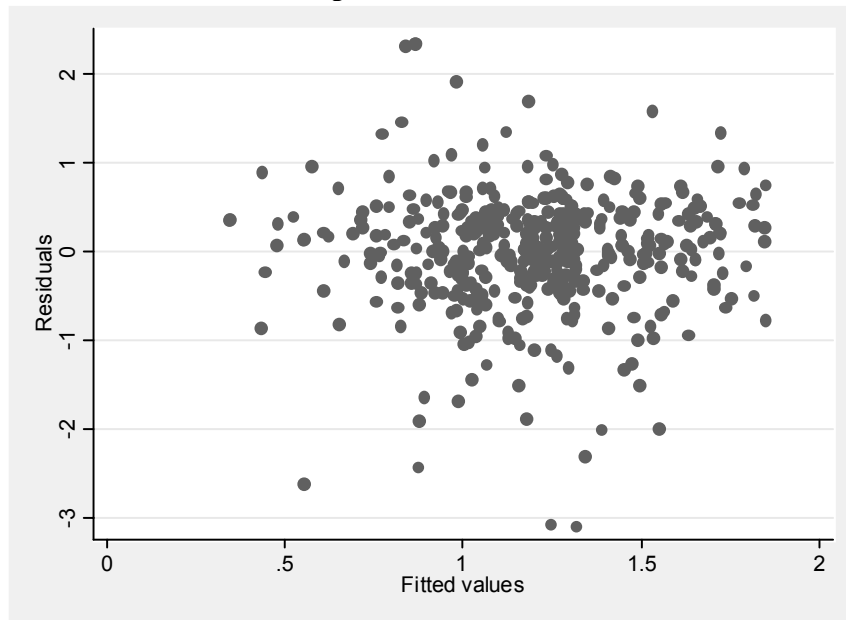
⁷ <http://www.stata.com/help.cgi?regress+postestimation#predict>

Task 4 – Graphs⁸

Plot the residuals from the above regression in order to detect the presence of heteroscedasticity.

You have two options:

- use the predicted values from above: **scatter resid fitted**
- use the in-built command: **rvfplot**



Note: you can copy the graph into Word by right-clicking on it and choosing ‘Copy, Picture’.

Task 5 – F test

Test the following hypothesis:

$H_0: e = 0, f = 0, g = 0$

```
test age kidslt6 kidsge6
```

```
( 1)  age = 0
( 2)  kidslt6 = 0
( 3)  kidsge6 = 0
```

```
F( 3, 421) = 0.24
Prob > F = 0.8705
```

As an exercise, go through the commands in:

<http://www.stata.com/capabilities/session.html>

⁸ http://www.stata.com/help.cgi?twoway_scatter