**HANDOUTS**

**INTRODUCTION TO STATA**

When you open Stata, you will see a screen similar to the following:

**Example 1: View of Stata when first opened**



The top row is a menu bar with commands. Below the menu bar is a tool bar with buttons. And there are four windows labeled Review, Variable, Results, and Command. Each is described briefly below, along with other windows that can be opened.

**Menu bar**

The menu bar has lists of commands that can be opened by clicking on a word. There are too many sub-commands to list here, but below is a list of the main operations that can be done with each command. If you use Stata a lot, you probably will not use the menu bar often because the most common tasks can be done with the buttons on the tool bar and key-strokes.

****

****

**Help in Stata**

Sometimes you might need help regarding a particular Stata function or command. For every command, Stata's in-built support can be called by typing `help' followed by the command in question in

the Command window or via the Help menu icon. The information you will receive is an abbreviated version of the Stata pdf manual entry.

For a more comprehensive search or if you do not know what command to use type in `search'

or `findit' followed by specific keywords. Stata then also provides links to external (web) sources or user-written commands regarding your particular enquiry.

**Data input and saving**

One of the first steps of every statistical analysis is importing the dataset to be analysed into the software package. Depending on the format of your data there are different ways of accomplishing this task. If your dataset is already in a Stata format which is indicated by the suffix `.dta' you can click on File / Open... and simply select the dataset you would like to work with. Alternatively, you use the `use' command followed by the name and location of the dataset, e.g. use "L:\Alma Kudebayeva\Fall2017\ECN5012 Econometrics\Chapter\_1\eaef21.dta", clear'. The term `clear' tells Stata to close all work files that are currently used in memory.

If your dataset is in Excel format you click on File / Import / Excel spreadsheet (\*.xls;\*.xlsl) and select the Excel file you would like to import into Stata. A new window appears which provides a preview of the data and in which you can specify certain options as to how you would like the dataset to be imported. If you prefer to use a command to import the data you need to type in the command `import excel' into the command window followed by the file name and location as well as potential importing options.

Stata can also import text data, data in SAS format and other formats (see File / Import) or you can directly paste observations into the Data Editor. You save changes to your data using the command save or by selecting the File / Save as... option in the menu.

eaef21.dta - this is a subset of a major US data base, the National Longitudinal Survey of Youth1979- (NLSY79). NLSY79 is a panel survey with repeated interviews of a nationally representative sample of young males and females aged 14 to 21 in 1979.

**Exercise Sheet 1**

The idea behind this exercise sheet is to familiarise with the statistical package STATA. The data file eaef21.dt contains data on a random sample of individuals selected through the labour force survey. The data file contains their earnings, together with many other variables that might influence them, including personal characteristics like race, religion, education, family background and work-related variables. This dataset is going to form the basis of most of the exercise sheets, where our primary focus is going to be the determinant of wages (earnings) for this random sample of individuals. Before undertaking any form of regression analysis it is important to understand the nature of your dataset and to undertake some preliminary data analysis.

In this Exercise Sheet we are going to create a Stata DO file which instructs Stata to undertake some basic data analysis (data transformations, summary statistics and graphing).

1. Open up a new Stata DO file. We will then be copying commands from the Review window in Stata into our DO file (remembering to annotate the DO file as a reminder of what we are instructing Stata to do).

2. Load the dataset eaef21.dt into Stata.

3. Open a log file in which to record the output from Stata, this should be written to a flash drive.

4. Use describe to have a preliminary look at the variables in the dataset.

5. Calculate summary statistics on the EARNINGS and S variables (using summarize and the detailed variant of this command). Produce a suitably labelled histogram of these series – what do you conclude?

6. Produce a suitably labelled histogram of EARNINGS and S variables.

7. Reproduce the summary statistics reported in (5) separately for males and females

(based on the variable *FEMALE or MALE*) – what do you conclude?

8. Reproduce the histogram reported in (6) separately for males and females (based on the

variable *FEMALE or MALE*)– what do you conclude?

9. Plot the histogram of ln(*EARNINGS*), separately for males and females (based on the

Variable *FEMALE)*, - what do you conclude?

10. Calculate the mean of the variable ln(*EARNINGS*) for white males, white females, nonwhite

males and non-white females (using the variables *FEMALE* and ETHWHITE). For

whites, what is the gender pay gap and is it signficant? For non-whites what is the

gender pay gap and is it signficant? Is there a significant difference in the gender pay gap

between whites and non-whites?

11. From the variable *AGE* produce a categorical variable with 6 categories (37-40

and 40-45), suitably labelled. Based on this new categorical variable,

produce a labelled plot of the median of ln(*EARNINGS*) for the 2 different age groups -

what do you conclude?

12. Run a regression equation of *EARNINGS t*o *S*.

13. Copy all of your commands to a DO file, which you should then save and annotate, so

you learn to understand what each command is doing.

14. Close your log file and save the DO file which you have created