WESS ECONOMETRICS

Exercise Sheet 1

The idea behind this exercise sheet is to familiarise with the statistical package STATA. The data file ***qlfs14q1.dta*** contains data on a random sample of individuals selected through the labour force survey based on data collected in 2014Q1. 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). This Stata DO file will be added to in subsequent weeks such that by the end of this term you have a whole Stata DO file, which takes you from data transformations of your series, to summary statistics and preliminary data analysis to two-variable simple regression analysis, multiple regression analysis, dummy variables, testing the stability and in general the validity of the model and IV estimation.

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 ***qlfs14q1.dta*** 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 *HOURPAY* variable (using summarize and the detailed variant of this command). Produce a suitably labelled histogram of this series – what do you conclude?
6. Produce a suitably labelled histogram of *HOURPAY* based on the 98% confidence interval limits of *HOURPAY*?

7. Reproduce the summary statistics reported in (5) separately for males and females (based on the variable *SEX*) – what do you conclude?
8. Reproduce the histogram reported in (6) separately for males and females (based on the variable *SEX*) based on the 98% confidence interval limits of HOURPAY – what do you conclude?
9. Plot the histogram of ln(*HOURPAY*), separately for males and females (based on the variable *SEX*), based on the 98% confidence interval limits of this variable - what do you conclude?
10. Calculate the mean of the variable ln(HOURPAY) for white males, white females, non-white males and non-white females (using the variables *SEX* and *ETHUKEUL*). 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 (<30, 30-37, 38-43, 44-48, 49-55 and 56+), suitably labelled. Based on this new categorical variable, produce a labelled plot of the median of ln(*HOURPAY*) for the 6 different age groups - what do you conclude?
12. From the variable EDAGE produce a new schooling variable (suitably labelled) with 7 categories as: left school at less than 16 years old, 16 years old, 17 years old, 18 years old, 19-21 years, 22-24 years old, and more than 24 years old. Produce a suitably labelled plot the median of ln(*HOURPAY*) for males and females separately according to the new categorical variable for the number of years of schooling - what do you conclude?
13. Plot the median earnings for males and females separately according to the new categorical variable for the number of years of schooling, but only for those individuals who are between 44-55 years old – what do you conclude?
14. Tabulate the variable *HIQUL11D*. Create a binary variable which is 1 if an individual has a “Degree or equivalent” and 0 otherwise. Without doing any further instructions in STATA calculate the mean and standard deviation of this variable.
15. 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.
16. Close your log file and save the DO file which you have created onto a flash drive.