University of Warwick
Department of Economics
Diploma Preparation Exercise Sheet

This exercise sheet is designed to get you practice with the package Stata. Stata is a powerful statistical analysis tool, which we will be using within the module Econometrics 1.

The dataset **icfforworkbook.dta** is based on the [Living Cost and Food Survey, 2013](#) (LCF). The LCF, conducted by the Office for National Statistics collects information on spending patterns and the cost of living that reflects household budgets across the UK. This teaching dataset is a subset which has been subject to certain simplifications and additions for the purpose of learning and teaching by CMIST, Manchester.

1. Load the dataset **icfforworkbook.dta** into Stata.

2. Briefly review the dataset in the DATA EDITOR window.

3. Open a log file in which to record the output from Stata.

4. Type `describe` in the Command window to have a preliminary look at the variables in the dataset.

5. Use the `list P550tpr P344pr` command to list the values of the variables *P550tpr* (total expenditure) and *P344pr* (Gross normal weekly income).

6. Use the command `tabulate` (`tab`) to tabulate the two variables *P550tpr* and *P344pr* separately.

7. Use the command `twoway (line P55tpr casenew, sort)` to produce a line plot of the variable *P550tpr* against *casenew*.

8. Use the command `summarize P550tpr` command to produce summary statistics on the variable *P550tpr*,

9. Use the command `summarize P550tpr, detail` command to produce detailed summary statistics on the variable *P550tpr*.

10. Use the command `summarize P550tpr P344pr` command to produce summary statistics on the variables *P550tpr* and *P344pr* variables.

11. Use the command `histogram P550tpr, percent` to produce histogram a plot of the variable *P550tpr*.

12. Use the command `histogram P550tpr, width(10) percent` to produce histogram a plot of the variable *P550tpr* with a width of £10. Continue playing with the width command to see what might yiled a more informative plots. Try using the `bin` option in place of `width` to see how this impacts on your plot.

13. Use the command `twoway (scatter P550tpr P344pr)` to produce a scatter plot of *P550tpr* against *P344pr*.

14. Use the command `gen apc=P550tpr/P344pr` to generate a new variable consumption expenditure per pound on income. Use the comand `lab var apc "Average Propensity to Consume"`.

15. The issue with the variable *P550tpr* is that is contains some very strange values at the upper end of the distribution. Use the command `_pctile p550tpr, p(99)` in order to find the value above which 1% of points of *P550tpr* lie.

16. Use the information from 15 above and the command `gen exp=???` to create a new variable *exp* that replaces the top 1% of *P550tpr* values by missing values (.) – where you need to think what to use in place of ???.

17. Produce suitably labelled histogram of *exp* (try just editting the instruction issued in 11 or 12).

18. Using a command similar to that in 16 to replace the upper 1% of points in *apc* with missing values.

19. Produce suitably labelled histogram of *apc* and save this graph.

20. Tabulate some of the categorical variables *P435r, A172, SexHRP, Gorx, A049r*.

21. Report summary statistics for the variable *exp* separately for households according to their main source of income (*P425r*).

22. Produce a suitably labelled histogram of *exp* separately for households according to their main source of income (*P425r*).

23. Construct summary statistics for *exp* for households according to both the main source of income (*P425r*), and internet connection (*A172*).

24. At the 1% significance level, test the hypothesis of no difference in *exp* between households with earnings and other sources as their main source of income (*P425r*).

25. At the 1% significance level, test the hypothesis of no differences in *exp* between households with aearings and other sources as their manin source of income (*P425r*) by internet connection (*A172*).

26. From the variable *P344pr* create a suitably labelled binary variable *inc_m* which is 1 if above median income 0 otherwise. Label the values 0 and 1 within this variable.

27. At the 1% significance level, test no difference in mean for *exp* based on *inc_m*.

28. From the variable *P344pr* create a categorical variable *inc_cat* based on the quintiles of *P344pr*. Label the variable *inc_cat* and the 5 values within this variable.

29. Using *inc_cat*, produce a suitably labelled plot of the mean of *exp* for each of the 5 income categories.

30. Produce a bar graph of the mean of *exp* against *inc_cat*, separately by internet connection (*A172*).

31. 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.

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