Introduction to Stata: Practical Session 1

David Clark

12th April 2018

Exercise One

1. Open Stata (either IC or SE will do - if after you choose one, it displays a prompt citing insufficient licences, choose the other version)
2. Find out what Stata's present working directory is using the **command method**.
3. Change directory to your personal M drive (M:/), if it isn't set there already.
4. Create a folder using the command prompt in Stata called **stata_course**.
5. Within the file you downloaded at the beginning of the course, move the **data folder** to **stata_course** folder you've just created (You'll have to do this using the Windows file explorer outside of Stata).
6. Using the appropriate command, change Stata's present working directory to the **data folder that you moved in the previous step**.
7. Confirm the folder contains the data files but using the **list directory command** (Be careful: this is not done by using the command **list**).
8. Open the datafile called **PWT.dta** (hint: using the drop down menu is probably easiest here)
9. How many variables are in the dataset?
10. Rename each of the variables so the each start with a capital letter.
11. What is the variable type of the following variables?
    - Country
    - Kg
    - Year
12. Give these labels to the following variables:

- Country - Name of the country
- Year - Year that data was collected
- Pop - Population
- Cgdp - GDP per capita

13. Browse the dataset - what do you notice about the number of observations? Is all the data present in the dataset?

14. Attain the summary statistics for the following variables - **year pop cgdp**

15. Attain the summary statistics for the same variables, **excluding missing values** (Note, you'll need to use the conditional and relational operators here).

16. What do you notice about the summary statistics calculated with and without missing values? Is there a difference? Explain.

Exercise Two

1. Clear what is in Stata's working memory before we proceed any further.

2. Import the Excel datafile called **auto.xls**, remembering to format the variable names correctly.

3. Save the datafile in Stata's native file format (again, using the drop down menus with probabily be easiest here).

4. Which command can be used to find the number of observations, number of variables and each of the variable types, simultaneously? Use this command to find the number of observations in this dataset and the variable type of the following:

   - make
   - foreign
   - price

5. Browse the first 10 observations in the dataset.

6. Find the correlation between the variables in the dataset (If this results in an error, think about the variable types).

7. What is the relationship between the following:

   - mpg and price

- headroom and price

8. Find the mean price of cars in the dataset.

9. Using conditional operators, find the mean price of cars for the domestic group and foreign group (keep in mind the **type** of variable that foreign is when writing this command).

10. Using the **tabulate** command, find the frequency of cars that are **foreign-made** and have a repair record of **3** (the variable for repair record is rep78).

11. Attain the distribution of observations by percentiles for the variable **mpg** using commands that we have covered in the slides.

12. Generate a variable called **mpg50** which is equal to one if **mpg** is greater than the value at the 50^{th} percentile and equal to zero otherwise.

13. Using the variable you've just generated, tabulate the variable **rep78** for cars with an mpg above and below the 50^{th} percentile by both two-way tabulation or using conditional operators.