***Exercise 1 - Introduction to the Stata software, and to Stata commands and results***

In the following exercises you will

(i) Familiarise yourself with the Stata windows interface
(ii) submit commands using the Graphical User Interface (GUI)
(iii) Save the commands created by the GUI in a do-file; you will edit and resubmit the commands
(iv) Save a log-file containing the commands and the output

**This exercise covers the materials in Chapters 1 and 2 in the Introduction to Stata module notes**

Before starting this exercise:

- On your computer, create a folder called *Stats Computing* and a sub-folder called *Stata*. You can add a sub-folders for R later.

- Copy all the contents of *U:/Download/Teach/Med_St_Intro_to_Stata/* to the *Stata* folder created above.

***Exercise 1.1: Introduction to Stata***

Start by launching Stata. The aim of this set of exercises is to begin to familiarise yourself with the Stata working environment. Each time you submit a command look at the output in the Results Window and check the other windows.

- Use the ***File*** drop down menu to:
  - Change the current working directory to the *C:/Stats Computing/Stata/Data/Exercise 1* folder.
  - Open the *bl_demog*.dta dataset.
  - What do the *History Window* and *Variables Window* now contain?

- Select the ***Data Editor (Browse)*** from the Shortcut toolbar:
  - Scroll across and down the variables - which variables are string and which are numeric?
  - How many variables and observations are there?
  - Close the Data Editor.

- In the ***History Window*** click on the `use` command so that it appears in the *Command Window*.
  - Edit the command and resubmit so that you now load the *vitals_long* dataset.
  - View the data in the *Data Browser*.

- Within the ***Command Window***:
  - Use the page-up button to retrieve the command to load the *bl_demog* dataset.
  - Resubmit the command.
  - Use the `dir` command to (i) get a list of all the files in the current directory and (ii) get a list of all the Stata datasets in the current directory.

- Right click the mouse within the ***Results Window***:
  - Try changing the Window Preferences and Font Size.
  - Find a set up that is comfortable for you.

- Selecting variables from the ***Variables Window***:
  - Within the *Command Window* type "`tabulate`"
  - Using the mouse select the variable *agegroup* from the *Variables Window*. You can either double click on the variable or click on the arrow that appears when you hover the mouse over the variable name. The variable *agegroup* should appear in the *Command Window* after `tabulate`. Press the return (enter) key to submit the command.
  - Type "`summarize`" in the *Command Window* and then from the *Variables Window* select *age*, *sbp* and *hrate* using the mouse. Now submit the command.
  - From the *History Window* select the command *tabulate agegroup*. Now select the variable *diab* from the *Variables Window* (the command should then read `tabulate agegroup diab`). Resubmit the command.
  - Within the *Command Window* use the page-up key to bring the last command back into the *Command Window*. Now type ", row" (type the comma but not the quotes) and then submit the command. Look at the output in the *Results Window*.
  - From the *History Window* select the summarize *sbp* command.
  - Delete *sbp* and replace with *wt*. Press enter.

***Exercise 1.2: Drop down menus and Graphical User Interface (GUI)***

NB: As you run through the GUI exercises look at the Stata commands that are echoed to the Results Window or History Window and try to identify the various parts of the command syntax. Also look at any output in the Results Window.

- Continue with the *bl_demog.dta* dataset, which should still be loaded

- Using the ***Data*** *drop down menu*:
  - Obtain a description of the dataset in memory
  - Produce a description of *vitals_long.dta*
  - Obtain a codebook for the variables *ptid*, *birthdt*, *age*, *smkstat*, *sbp* and *diab.*
  - Add the label "Systolic Blood Pressure (mmHg)" to the variable *sbp*
  - Add the label "Heart Rate (bpm)" to the variable *hrate*.
  - List *ptid*, *birthdt*, *age* and *sex* in the first 10 rows of the dataset.
  - Sort the dataset by *age*.
  - List *ptid*, *birthdt*, *age* and *sex* in the first 10 rows of the dataset.

- Use the ***Graphics*** *drop down menu* to produce:
  - A histogram of *wt*.
  - On the Main tab:
    - Compare the "continuous" and "discrete" options.
    - Try out the frequency and bin options
  - A histogram of *smkstat*.
    - Is the data continuous or discrete?

  - A bar graph displaying the mean of *egfr* over categories of *agegroup*.
    - Use the y-axis tab to add a y-axis title.
    - Add *sex* as a second over categories variable
  - A scatterplot matrix of *age*, *wt*, *ht*, *wc*, *sbp* and *hrate.*
    - Use a small hollow-circle for the marker symbols.
    - Plot just a half-matrix.
  - A vertical box plot of *egfr.*
    - Reproduce the plot over categories of *agegroup.*
    - Add *sex* as a second over category.
    - Swap the order of *agegroup* and *sex* in the over categories.
- Use the ***Statistics>Summaries, Tables and Tests*** *drop down menu* to obtain:
  - Summary statistics for *age*, *wt* and *sbp*.
    - Select the option for additional statistics.
  - A pair-wise correlation matrix of *age*, *wt*, *ht*, *wc*, *sbp* and *hrate*.
    - Add the number of observations for each pair-wise correlation.
  - One-way frequency tables of (i) *agegroup* and (ii) *diab.*
  - A two-way frequency table of *agegroup* and *diab.*
    - Add row percentages.
    - Repeat the table by categories of *sex.*

Don't exit Stata. We will now create and save a do-file containing the commands you submitted in exercise 1.2.

***Exercise 1.3: Creating a Do-file***

- Remove any commands with errors (those in red text) from the History Window.
- Send all the remaining commands in the History Window to a new do-file.
- Save the do-file as *Stata_Exercise1.do* in the Exercise 1 folder.
- Add appropriate comments throughout the do-file.
- Remove any duplicated commands and make sure each command works in turn i.e. submit one command at a time. Correct any errors. Remember to keep saving the file as you make changes.
- Try running the whole do-file.
- Try editing any of the longer commands so that they run over more than one line.

***Exercise 1.4: Creating a Log-File***

- Once your do-file is working correctly add a command immediately after the change directory command (`cd`) to open a log-file called *Log_Exercise1.log*.
- Add a command at the end of the do-file to close the log-file.
- Re-run the whole do-file.
- Try re-running again. What will you need to add to the `log using` command to stop the do-file breaking at this point?
- What happens to the graphs as you run through the do-file? Try adding the option `name(`*`nameofgraph, replace`*`)` to each of the graph box commands – where *`nameofgraph`* is whatever you want to call the graph. Re-run the graph box section of the do-file.

- Try opening and viewing the log-file in Word or a text-file editor e.g. Notepad.
- Save the updated do-file.

***Exercise 1.5: Help***

- Try using `help` to find out what options there are for the `summarize`, `codebook` and `pwcorr` commands.