Practical 0
Introduction to Stata version 12

A Demonstration of STATA

1) Load Stata.

2) There are a number of windows.

(a)	**Command**	for typing and running commands.
(b)	**Results**	for displaying output (except high resolution graphics).
(c)	**Review**	for displaying past commands.
(d)	**Variables**	for listing the variables in your data.
(e)	**Properties**	for information about your data.

A type of window which does not open automatically is a **Graph** window. This window is used to display high resolution graphical output.

3) Open the **Data Editor** window using **Window → Data Editor**.

Enter the following small dataset on 4 subjects, where gender is coded as 1 = female and 2 = male.

bmi	age	gender
22.4	20	1
24.9	29	1
26.0	35	2
24.2	25	2

4) Type the variable names in the **Properties** window.

Assign **bmi** a variable label, say, **Body Mass Index**.

Create a value label called **genderfmt** (**Female** to **1** and **Male** to **2**) and assign it to the **gender** variable.

5) Note Stata is case sensitive. Hence **BMI** is different from **bmi**.

6) Save the data to a Stata data file called **health.dta** using **File → Save As**.

7) Obtain some basic summary statistics for the **bmi** variable.

Use **Statistics → Summaries, tables, and tests → Summary and descriptive statistics → Summary statistics**. Select **bmi** into the **Variables** box.

To obtain summary statistics by **gender** use the **by/if/in** tab.

8) To produce a scatter-plot of **bmi** against **age** use:

- **Graphics → Twoway graph (scatterplot, line etc.)**.
- **Click** on **Create**, and ensure **Basic plots** is selected from under **Choose a plot category and type**.
- From the **Basic Plots** list select **Scatter**. Select **bmi** and **age** into the **Y variable** and **X variable** box respectively. **Click, Accept, OK** when finished.

Note the Stata command language is more efficient.

9) When first opening a Stata data file (using **File → Open**) it is useful to examine the content.

Use **Data → Describe data → Describe data in memory** to list the variables and attached labels.

For a more detailed summary use **Data → Describe data → Describe data contents (codebook)**

10) The Stata **help** system is most useful for the command language, e.g. help summarize. Note options of the main command need to come after a comma.