

## Lab Worksheet II

### Stata's interface

This is probably the first time you use Stata. As described in their own website, “Stata is a complete, integrated statistics package that provides everything you need for data analysis, data management, and graphics.”

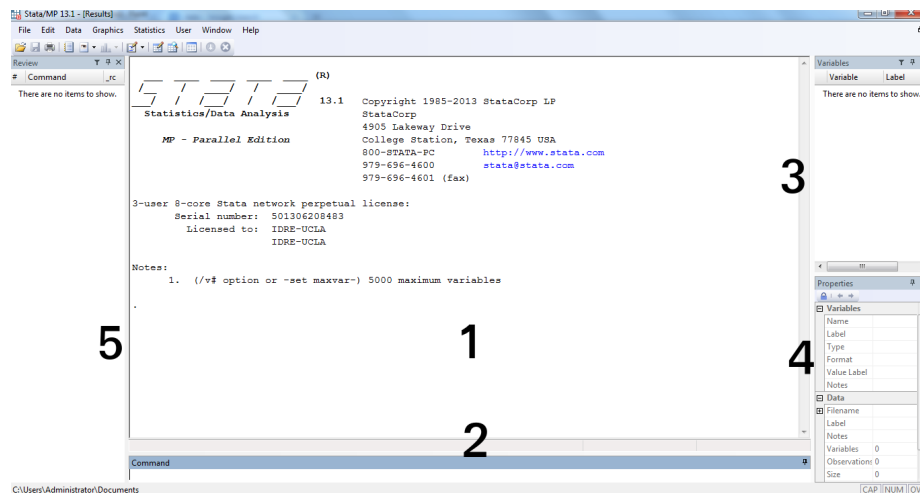


Figure 1: Stata's interface.

Upon starting Stata you will see the following sections:

1. The results and outputs of your commands will be displayed in the **results window**.
2. The **command window** is where you want to type your commands and send them to Stata by pressing the Enter key.
3. If you have a dataset loaded its variables will appear in the **variables pane**.
4. You can browse the properties of variables in the **properties pane**.
5. The **review pane** lists all the actions you have taken in Stata.

At the moment you have no data loaded, that is the reason why your variables window is empty. Today you are going to input manually some data on UK universities. You can find these data in the table below.

Table 1: Some universities in the UK

Name	Students	Ranking	Location	Nobel laureate	Vice chancellor	Year
Edinburgh University	35258	20	Scotland	20	Male	1582
Napier University	17264	64	Scotland	0	Female	1992
University of Essex	11939	47	England	2	Male	1965
University of Cambridge	18271	1	England	91	Male	1209
Cardiff University	30180	27	Wales	2	Male	1997
University of Liverpool	21276	59	England	9	Female	1903
Queen's University Belfast	24955	45	Northern Ireland	3	Male	1908
Glasgow Caledonian	18410	89	Scotland	1	Female	1993
Ulster University	26200	82	Northern Ireland	0	Male	1982
University of Warwick	23420	6	England	1	Male	1965

As you can see there are seven variables and ten observations. You need the following variables:

- Name of university.
- Total number of students (numeric).
- Position in the The Guardian Ranking.
- Geographical location (categorical, see below).
- Number of affiliated Nobel laureates (numeric).
- Gender of the Vice-chancellor (binary, see below).
- Year in which it was established (numeric).

The variable “Location” has the following four categories: Scotland = 1, Northern Ireland = 2, Wales = 3, and England = 4. The variable “Vice-chancellor” should be coded as Male = 0 and Female = 1.

You need to think of suitable **variable names**, **variable labels**, and **value labels**. Keep in mind that is customary, but not mandatory, to name variables with small letters.

Stata accepts both upper cases and lower cases, but it is case sensitive! Also, variable names cannot contain spaces and the use of underscores is customary when signalling spaces. After thinking of some variable names, open up the data editor by clicking on the icon with a pencil over a spreadsheet (see Figure 2). Stata's data editor resembles very much Excel. Your "spreadsheet" should be empty, but if it was not you could type `clear all` in the command window to get rid of previous data.

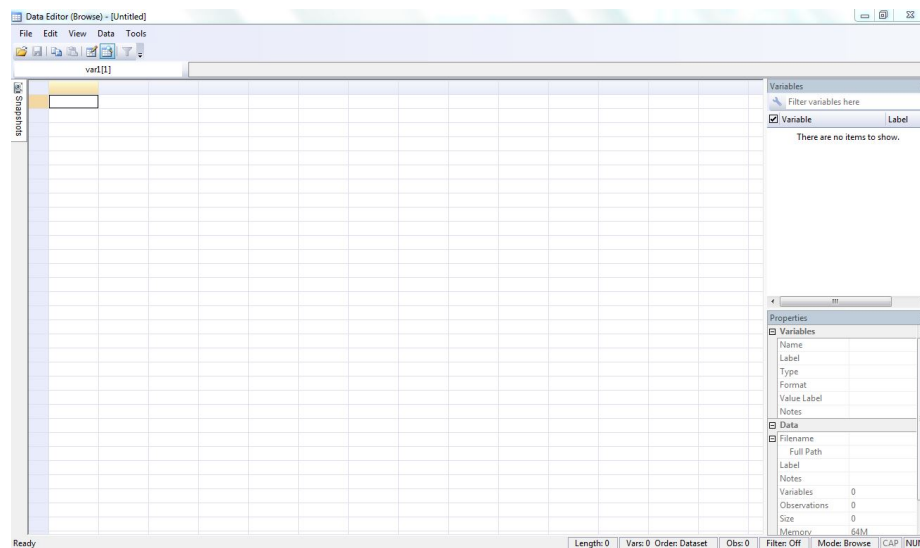


Figure 2: Stata's data editor.

Now it is time to enter the data; start by typing in the first observation (University of Edinburgh). If you press the Tab key you will move onto the next cell. Keep entering the values for all variables for the University of Edinburgh (your first observation). Here it is important to remember that rows represent observations whereas columns represent variables. You might notice that Stata is naming your variables as *var1*, *var2*, and so on. To change **variable names** simply select a variable in the variables pane (top right corner) and change its name in the properties pane (bottom right corner). Repeat the process for the rest of variables. Let me remind you again that we need a number for geographical location and gender (the categories are specified above).

From the properties pane you can also define **variable labels**. The process is identical as before: select a variable in the variables pane, switch to the properties pane and give the variable a label (but this time using the box named "Label"). Variable labels allow spaces. Remember to be precise and accurate with your labelling, but try to keep your labels parsimonious (not too short as to make them ambiguous, but not too long either). Once you have named and labelled your variables you should see the new names and labels in the variables window.

Now that you have assigned names and labels to all your variables, you are going to assign **value labels** to two variables: geographical location and vice-chancellor (remember the above-mentioned specifications). But first try typing in the command window:

```
tabulate gender
```

In the command above, `tabulate` creates a tabulation of a variable named “gender.” If you did not name your variable “gender”, you will have to substitute in your variable’s name instead of “gender.” In the results window you can see now a table with two categories: 0 and 1. You know that 0 = Male and 1 = Female, but when you report these results you want your readers to know exactly what is going on. Changing 0 to “Male” and 1 to “Female” is what you do when you assign **value labels** to your vice-chancellor’s gender variable.

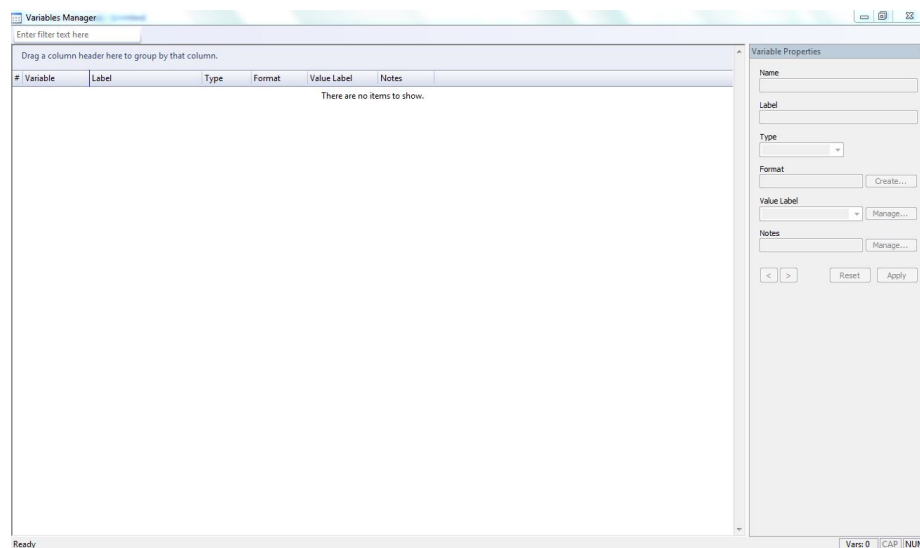


Figure 3: Stata’s variables manager.

To assign value labels you need to exit the data editor and open the *variables manager*. This can be accessed by clicking on the icon with a spreadsheet. Once the variables manager shows up (see Figure 3 for the generic, blank window) you need to create first a set of value labels which you will be attaching to a specific variable later on. Click on **Manage...** (situated on the right side of the window). In the new window click on **Create Label**. Yet a new window named **Create Label** will appear on screen. Give to your new set of value labels a name (in the **Label name** box). It is best if you name it with something that resembles the name of the variable you are modifying. For example, if you created a variable named “location” for the geographical location of universities you can name your set of value labels as “locationlab” or “locationl.” Keep in mind that sets of value labels cannot use variable names. Now type

in the box named **Value** the original value you gave to your variable (i.e., 0 = Male, so you would type “0” here) and in the **Label** box type its new label (i.e., “Male”). Then click on **Add** and repeat the process for 1 = Female. Once this is done click on **OK**.

Now that you have a set of value labels, the last thing you need to do is attach that set to the original variable. For this, go back to the variables manager window (Figure 3), select the variable at stake, and click on the down arrow located to the right of **Value Label**. A dropdown menu in which you can select the appropriate set of value labels will appear. Finally, click on **Apply**. If you use the `tabulate` command again you will see that your tabulation, this time, reports “Female” and “Male” instead of 1 and 0. Well done!