

## Using SPSS 20: Handout 1, Getting Started:

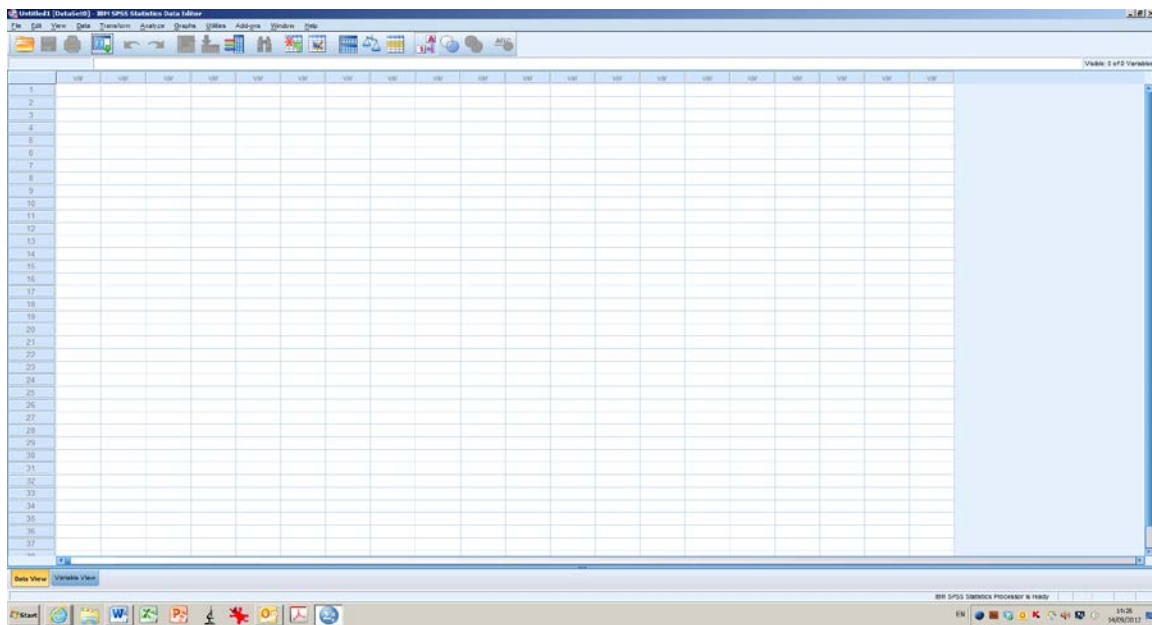
SPSS stands for "Statistical Package for the Social Sciences". It's a very powerful program that can do all of the statistics that you are ever likely to want to use. It's actually fairly easy to use, but because it can do statistics for "grown ups" as well as novices, it may seem quite daunting at first. It presents you with a bewildering array of options that you will probably never need to use. When it comes to giving you statistical results, it will give you what you want - as well as a lot of extra stuff that you may not need! The secret to using SPSS is to take it one small step at a time. This series of handouts is aimed at showing you how to use SPSS to do the statistics referred to in the lectures. There are now many helpful books which explain using SPSS well: the only catch with these is that SPSS exists in various versions, and so, depending on which book you get, it may not correspond *exactly* to what happens with the version we will be using (version 20).

The easiest way to learn SPSS is by example, so this handout will show you how to

- (a) start up SPSS;
- (b) enter data (numbers that you want to do some statistics on);
- (c) save the data to a file;
- (d) retrieve the data;
- (e) obtain some simple descriptive statistics (mean and standard deviation for a couple of variables);
- (f) print out the statistics;
- (g) shut down SPSS.

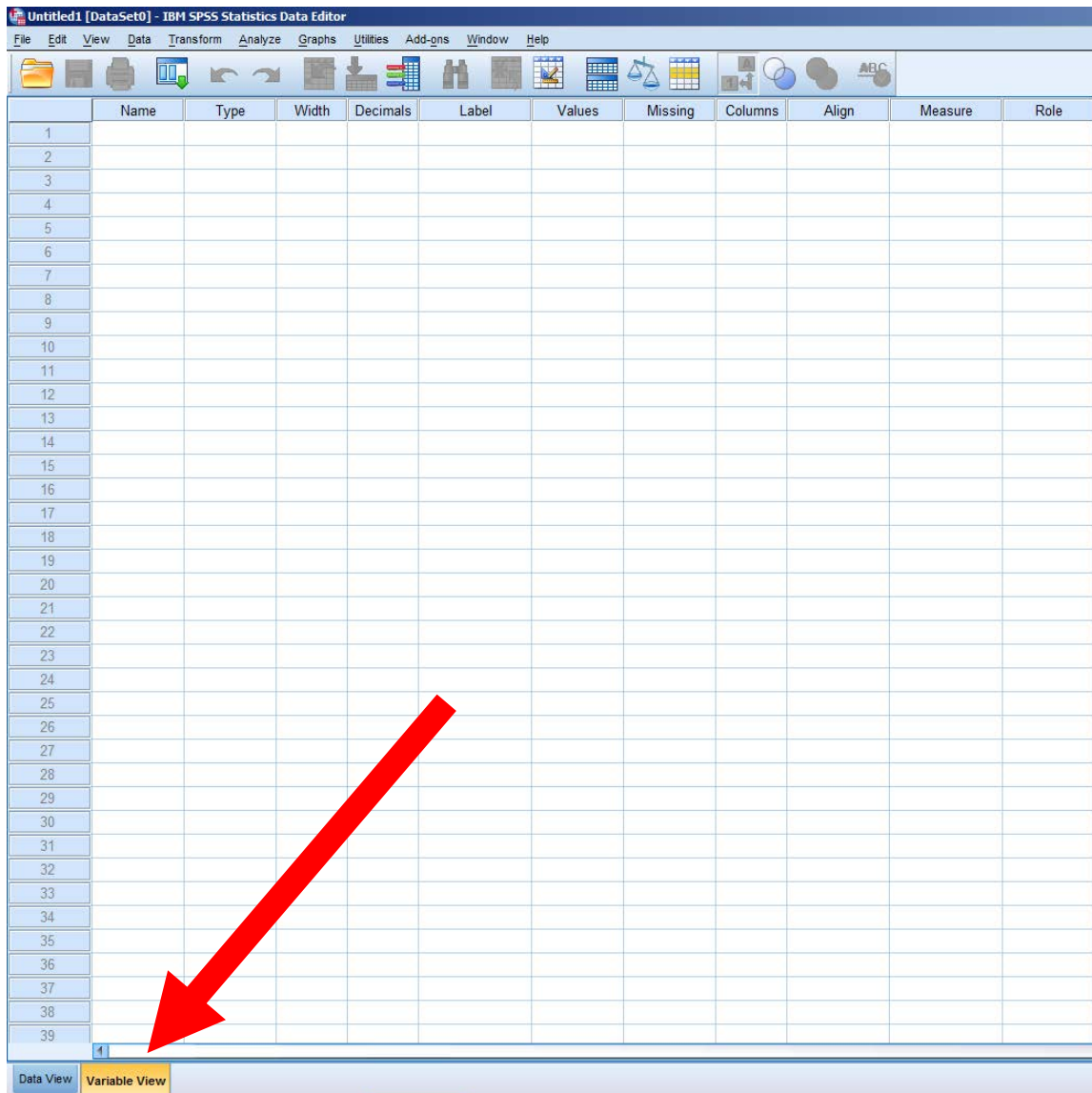
### 1. Starting up SPSS:

Double-click on the SPSS icon. After a few seconds, a window like the following should appear on your screen. This is the "Data View" window. It has the default name "Untitled1".



At the top of this window, there is a row of commands, (File, Edit, Help, etc.) Clicking on any of these will produce a drop-down menu, much as in other programs that you might be familiar with (such as Word or Excel). At this stage, many of the options on the menus will appear quite meaningless to you. "File", "Edit", "Analyze" and "Window" are the options that we will use the most. As you might expect, "File" enables you to open and save files, and "Edit" enables you to cut, copy and paste things. We will use "Analyze" to perform various statistical tests. "Help" provides you with information about SPSS, but isn't particularly user-friendly.

If you click on the tab at the bottom left of the window, it switches to a new window, the "Variable View". (You can also do this by using the "Window" or "View" commands at the top of the screen). You can toggle between these two windows at any time. You use the Data View window when you want to input data, and you use the Variable View window to change various properties of the data. (More details on this in a moment).



There is a third window, the "Output Window": this will contain the results of any statistical analyses that you perform. (The Output Window will only be available once you have some output to see, so it won't actually be accessible just yet). Each window has the controls for SPSS at the top of the screen (the words "File", "Edit" and so on, and the row of icons beneath them). Most of the SPSS controls will remain visible all the time, but you can switch between the three windows whenever you like. You switch between the "Data View" and "Variable View" windows by using the tabs at the bottom left of either of these two windows. You switch between

these two windows and the "Output Window" by clicking on "window" at the top left of the screen and selecting the one that you want from the menu that appears.

## 2. Entering data:

In SPSS, each row of the grid is a "case" and each column is a "variable". To make this clear, imagine we have the heights and weights of six people. Each person is a separate case. We have three variables: height, weight and sex. We could therefore enter the data in such a way that each row represents a different individual: one column has the height data, another column has the weight data, and a third column tells SPSS whether the person was male or female.

To enter values, make sure you have the "Data view" window selected. Move the cursor to the square into which you wish to make an entry, and click on it. Enter the value, followed by a press on the "enter" key. To move around the grid, you can use the arrow keys or the mouse.

A "variable": each column contains data of a particular kind (in this instance, "gender")

	VAR00001	VAR00002	VAR00003	VAR00004	var
1	Matilda	2150.00	120.00	2.00	
2	Sheila	1994.00	100.00	2.00	
3	Florrie	1803.00	130.00	2.00	
4	Tom	2000.00	180.00	1.00	
5	Dick	2200.00	240.00	1.00	
6	Harry	1800.00	200.00	1.00	
7					

A "case": each row contains a single person's set of data

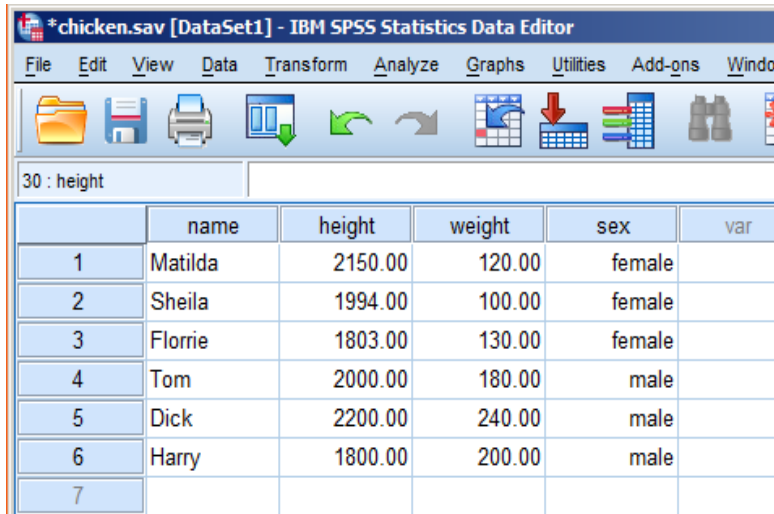
The screenshot above shows data entered for six people: three women and three men. The first column gives their names; the second column their heights; the third their weights; and the fourth, their sex. I've used "1" as a label to represent male and "2" to represent "female". Thus, the first row gives the values of these variables (height, weight and sex) for the case of "Matilda". Copy these data into the appropriate places in your SPSS grid. (You can type the numbers in as "2000", "2200", "1", etc.: SPSS will automatically add the decimal point and the zeros that follow it).

You can label each column with a more meaningful title than "VAR0001" and so on. There are two ways to do this. You can either double-click on the grey box at the top of a column, or go down to the tab that selects "Variable View". Either method will take you to the "Variable View" window.



You can associate a label with each number - thus, in this case, you can tell SPSS that "1" means "male" and "2" means "female". (Don't forget to click on "add" each time you enter a label).

When you return to the "Data view" window, click on "view" on the top menu and then click on "value labels". SPSS will then show the labels instead of the numbers (i.e., instead of the "sex" variable reading "2, 2, 2, 1, 1, 1", it will read "female, female, female, male, male, male").



The screenshot shows the IBM SPSS Statistics Data Editor window for a file named 'chicken.sav [DataSet1]'. The window title bar includes the file name and the application name. Below the title bar is a menu bar with options: File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, and Window. A toolbar with various icons is located below the menu bar. The main area of the window displays a data table with the following content:

	name	height	weight	sex	var
1	Matilda	2150.00	120.00	female	
2	Sheila	1994.00	100.00	female	
3	Florrie	1803.00	130.00	female	
4	Tom	2000.00	180.00	male	
5	Dick	2200.00	240.00	male	
6	Harry	1800.00	200.00	male	
7					

(ii) *Width:*

*Width* merely specifies the width of each column in the "Data View" window. By default, a column that contains *numbers* is eight characters wide. However, an annoying complication of this version of SPSS is that, if the first entry in a column is a *string* of a certain length, SPSS assumes that all of the subsequent cells will contain strings of the same length. This is why the names start with "Matilda": had I started the column with "Tom", SPSS would assume that all of the strings in this column are going to be 3 characters long. Consequently "Dick" would have been truncated to "Dic", "Harry" to "Har" and so on. You can stop SPSS doing this by changing the column width to suit the length of your strings - or, as I did here, by making sure that the longest string goes in the first cell of a column!

(iii) *Decimals:*

If the variable is a number, this column specifies how many decimal places each case will be displayed to. So, if you select 0, Tom's height will appear as "2000"; if you select 2, it will appear as "2000.00"; if you select 3, his height will be "2000.000"; and so on. If the variable is a string, this is irrelevant, and so SPSS will show a 0 in the relevant cell of the "Variable View" window.

(iv) *Missing:*

Sometimes a data-set is incomplete - perhaps someone forgot to tell you their height, for example. If a numerical entry is blank, SPSS assumes it's missing data anyway. However, sometimes you might want to enter a code for missing values - perhaps "999" to show that the value is missing because the person forgot to enter it, and "99" to show that it's missing because the person refused to give it. *Missing* enables you to do this. At the moment, it's simplest to show missing data by leaving blank the relevant cell in the "Data View" window.

(v) *Measure:*

This will be explained more fully in the statistics lectures. Essentially this column shows what kind of data SPSS thinks is in the column: nominal (a name, i.e. a string), ordinal (rating data) or scale (interval or ratio data). At the moment, it will suffice to keep clear the distinction between using numbers as numbers, and using them as labels (strings).

### 3. Saving a file:

Once you have entered your data, *always save the data before doing anything else*. If you are using a campus computer, the easiest way is to save the file onto a USB stick. Alternatively you can write it to a CD, or send yourself an email with the file as an attachment. Saving the file will save you a lot of heartache in the future. It is really demoralising to spend hours typing numbers in, only to lose the lot by some accident or mistake. Click on "File" on the SPSS controls (it's top left). A menu will appear. Now click on "Save As", and enter a filename in the top left-hand box - where it says " \*.sav". Any combination of letters and numbers will do, but let's call the file "chicken". SPSS will automatically add the suffix ".sav". The ".sav" bit is important, as it tells SPSS that this file is a data-file of a kind that it likes. (SPSS can read other types of data-file as well, but it's simplest to stick with .sav types for the moment). So, all you have to do, is type "chicken" (or, in future, any filename you choose) in the box to the right of "filename".

Press "enter", and SPSS will save the data into a file. You can now carry on, secure in the knowledge that whatever happens, your data will be safe on the computer; or you can quit SPSS and come back another time. To quit SPSS, click on "exit", which is at the bottom of the "File" menu.

When you save your data, an "Output Window" will open automatically, containing information that you've successfully saved the file. If this is all it contains, just close it without saving it. However if you have run some statistical analyses and hence have some output in the "Output Window", you can save this in a separate file. To do so, make the "Output Window" the active window (as described earlier), and then click on "File" and then "Save As" in the same way as for saving the data. This time, SPSS will prompt you to supply a filename ending in ".spo", to show that it's an output file rather than a data file. So, you could call it "chicken.spo", and SPSS will then save the contents of the "Output" window as a file. This file can be read into a word-processor such as Microsoft Word, and then treated like any other text document.

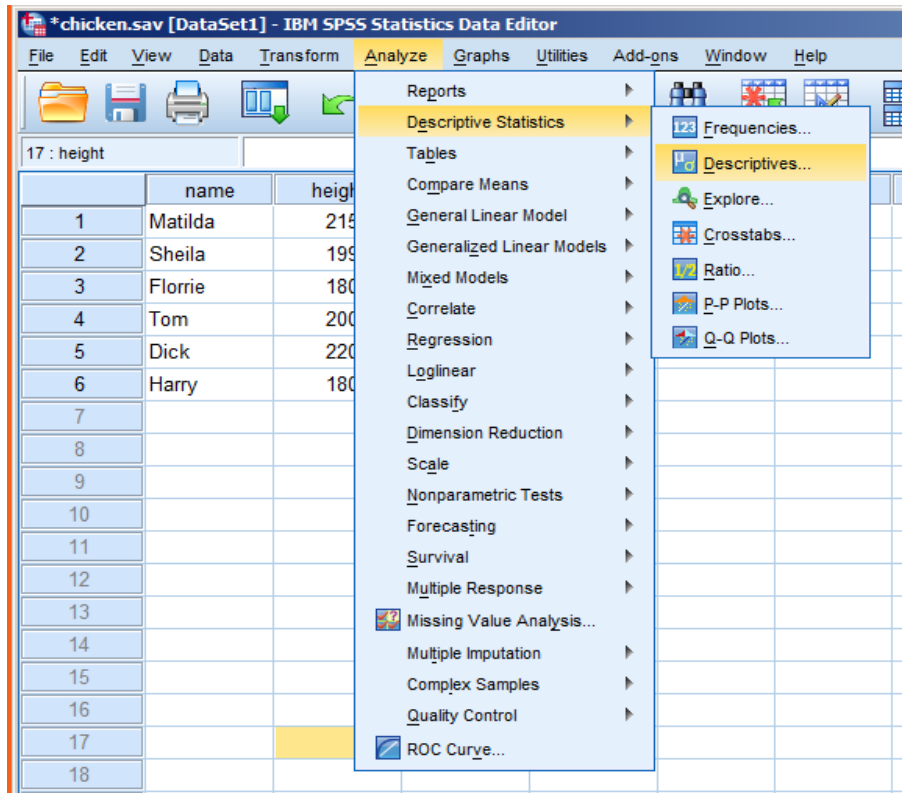
### 4. Retrieving a file:

Let's assume you have just opened SPSS and you want to retrieve a file you saved earlier. I saved my data as "chicken.sav". Go to the "File" menu, and click on "Open". Off to the right of "Open", a smaller menu will appear: select "Data" on that one, if you want to retrieve a data-file, or "Output" if you want to retrieve an output file that you saved earlier. Enter the name of your file in the box at the top. For the moment, assume we want to re-open the file containing our heights and weights. Select "Data" from the "Open" option on the "File" menu, and then type into the box at the top, "chicken.sav". Then click the "OK" button. Your data should appear! Notice that the word "Untitled1" at the top of the window is now replaced by the full name of the file that you have just opened.

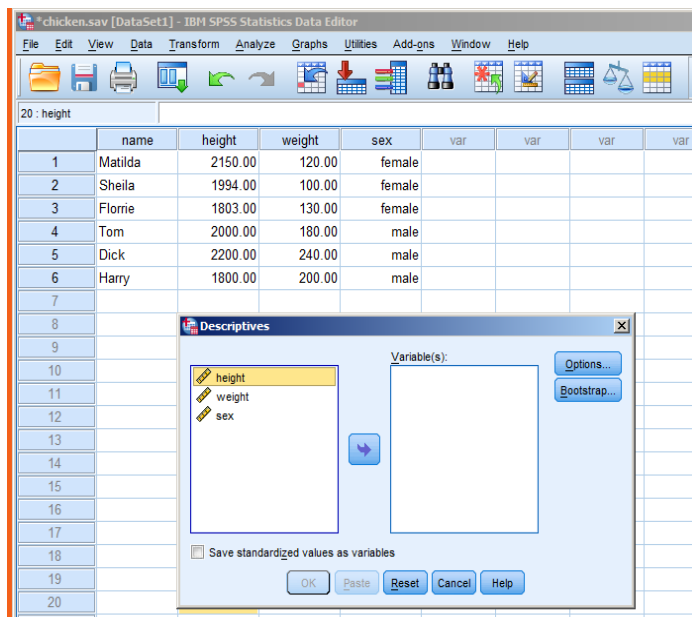
### 5. Calculating some basic descriptive statistics:

This section tells you how to work out the mean and standard deviation of a set of numbers, using SPSS.

- (i) Click on "Analyze" on the SPSS controls at the top of the screen.
- (ii) Click on "Descriptive Statistics".
- (iii) Click on "Descriptives...".

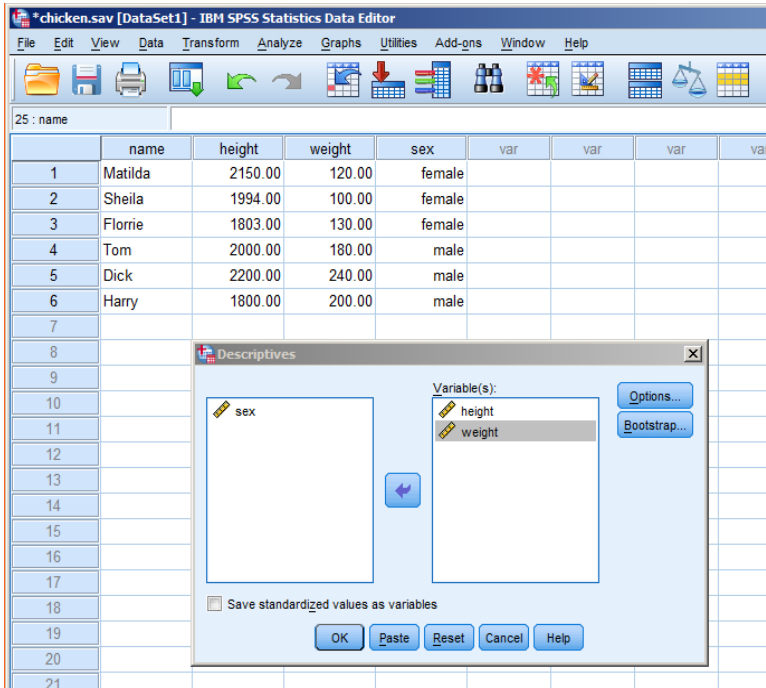


The screen should now look like this:

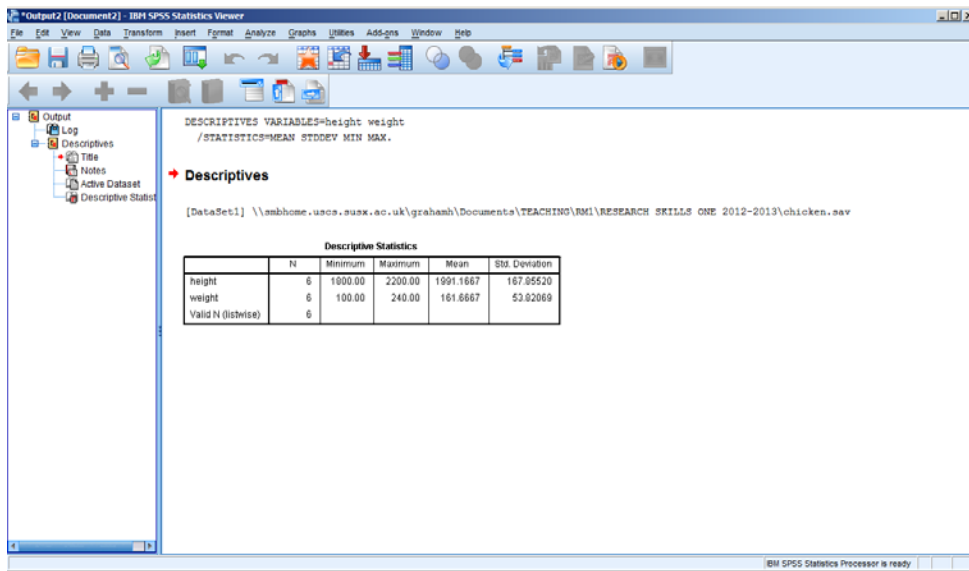


The box on the left shows variables for which you could produce descriptive statistics such as means, etc. (Note that SPSS doesn't show you the first column, as that contains data in the form of words, and you can't calculate means on this kind of data! However, we have fooled it with the variable "sex". SPSS would allow us to work out descriptive statistics for "sex", even though they would be meaningless - remember we used "1" and "2" merely as labels for "male" and "female" respectively, so they are not really "numbers" in an arithmetical sense at all).

You can move any or all of the variable names on the left, into the box on the right. Highlight a variable name by clicking on it, and then click on the little arrow between the boxes. If variables are moved to the box on the right, SPSS will calculate basic descriptive statistics on them. As a default option, SPSS will work out the mean, standard deviation, and minimum and maximum values for each variable placed in the right-most box. If you want other descriptive statistics, try clicking on the "Options" button. For this example, we will content ourselves with the default statistics, so move "height" and "weight" into the right-hand box, and then press "OK".



The screen will abruptly switch from the data window, to the "Output Window", and the statistics will be displayed.





Here's the table enlarged, for ease of reading:

	N	Minimum	Maximum	Mean	Std. Deviation
height	6	1800.00	2200.00	1991.1667	167.85520
weight	6	100.00	240.00	161.6667	53.82069
Valid N (listwise)	6				

The first column in the table, "N", tells you how many valid observations there were, basically a reassurance that SPSS has used as many participants' data in its calculations as we thought it was going to, and that it hasn't dropped participants from the analysis because they had missing data. (With a small data-set like ours, this isn't too useful, but if you had zillions of entries, it's always possible you made a mistake in entering the data and failed to notice).

There follow means, standard deviations and minimum and maximum values for our two variables of height and weight.

You have now done your first statistical analysis using SPSS - much quicker than working things out by hand!

#### **6. Printing the output:**

Easy - go to the "File" menu, click on "Print", and then click on "OK" in the little dialog box that pops up. You might want to use Print Preview beforehand, to see how many pages the output will come to; SPSS tends to use up an awful lot of paper if you are not careful. To delete output that you don't need, first click on "Close" on the top menu, in order to return to the Output Window. Then click on an individual piece of output in the righthand window to select it, and press the delete key to remove it. Conversely, you can copy and paste tables and graphs from SPSS straight into Word. That's handy if you only want to keep certain parts of the output, for example. However you should never just cut and paste SPSS output into lab reports - always redo tables in Word and graphs in Excel, as they will then look much nicer.

#### **7. Adding and removing cases:**

##### **(a) To insert more data:**

Recall that SPSS calls each person's data a "case", and that each row of the datasheet represents a single case.

In our example, we have data for six cases: three men and three women. The "sex" column contains 1's and 2's to tell us each person's sex. So far, we have obtained means and standard deviations of height and weight for all the subjects together, irrespective of their sex. However, later on we are going to consider males and females separately.

If you simply wanted to add more participants, you could move the cursor to the first empty row underneath the data you already have, and type the numbers in as before. As long as the computer has the 1's and 2's in the same row as the rest of the new peoples' data, it will be able to separate them out by sex at analysis time. However, for our own sakes, it would be nice to add the new data in such a way that all the males are in one block, followed by all the females: it just makes it easier to see what we are doing. If we want to do this, we need to insert the new rows of data, in between the ones we have already. Let's suppose we want to insert the details of two new cases, "Fred" and "Mary", into the data. Inserting Fred's data is straightforward: move to the first unused row underneath the data we already have, and enter the data. In the case of Mary, however, we might want to squeeze her in after all the other females, and before the males. To do this, we need to use the *insert new case* command:

(i) click the cursor anywhere on the row immediately *below* where you want to insert a new row. We want to insert a new female, so we click on the row that contains Tom's data.

(ii) click on "Edit", on the SPSS controls at the top of the screen. Then click on "Insert Cases". An empty row (with a dot in each empty cell) will be inserted *above* the row on which you originally clicked in (i). To insert more cases, simply repeat these steps.

**(b) To remove data:**

(i) Click on the shaded box at the extreme left of the row you want to delete. This will highlight the entire row in a different colour (yellow on my computer).

(ii) Go to the "Edit" menu at the top, and click on "Clear": the row you highlighted will disappear, and all of the data below it will move up to close the gap.

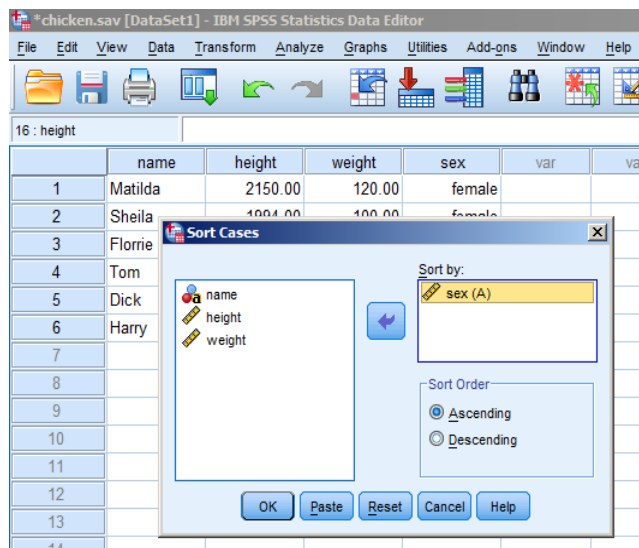
You can remove the contents of any single cell by clicking in the cell, and then pressing the "delete" key on the keyboard. Whatever was in the cell will be replaced by a dot, to show that the cell is empty.

**(c) The "Sort Cases" command:**

Suppose you had entered the males' and females' rows of data haphazardly, so that they were all mixed up: you could use the "sort cases" command on the "Data" menu to sort them out, so that all the male rows were together in one block, followed by all the female rows.

(i) Click on "Data" on the SPSS controls at the top of the screen.

(ii) Click on "Sort Cases". A dialog box will open, called "Sort Cases". It has two boxes in it. The left-hand one includes all the variables (columns) in your file. To sort the cases by a particular variable, click on that variable and drag it to the right-hand box. So, in this case, I drag "sex" over to the right.



(iii) Click on "OK", and SPSS will sort all the cases, so that (in this case), all of the males are together, followed by all of the females. (Have a play: try sorting by the other variables, and see what happens. Try sorting by more than one variable at a time).

**8. Shutting down SPSS:**

First, save your files if you want to, as described earlier in the handout. Anything not saved will be lost forever - although if you have simply opened an existing file, and worked with it, but not changed the data in it, you won't need to save it again. (If you forget to save files of data or output before trying to close SPSS, SPSS will remind you before shutting down, but it's best to get into the good habit of saving the files yourself as a matter of routine). Then, go to the "File" menu, and click on "Exit" at the bottom. Don't forget that you can save both the data and the output, in separate files. Saving the output file is useful if you want to print it out in some other place, at some other time.