CD4004: Spreadsheet Task 2

This worksheet will demonstrate how you can use Microsoft Excel to perform some of the tasks that you have learnt about in class. You will learn how to use the functions that calculate the mean, mode and median of a set of data. You will learn how to sort data and to produce a frequency table and a grouped frequency table. You will also use Excel to produce a bar chart and a frequency polygon.

Step 1

Download *StatsSpreadsheet.xlsx* from the Moodle site and open it in Excel. You will see that in cells B4:C104 there is a table of exam marks (out of 20) for 100 students (figure 1).

Step 2

Enter the correct formulae, as shown in figure 1, to calculate the mean, mode and median for this data set.

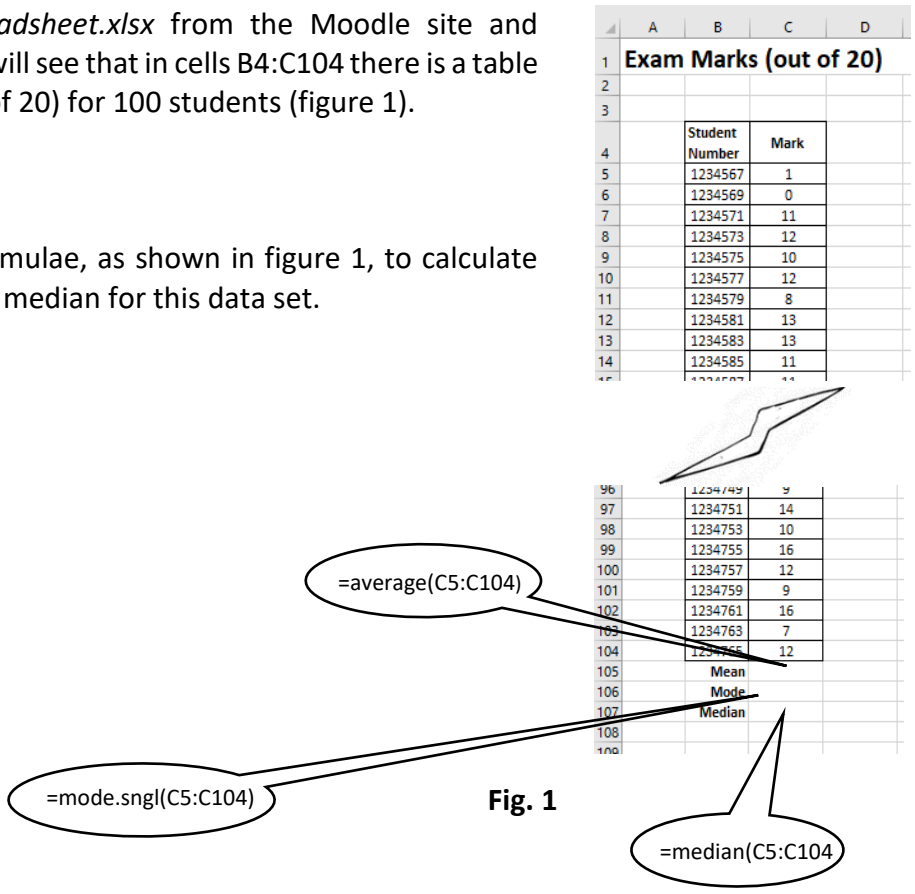

Fig. 1

Step 3

It will be easier to read the data if it is sorted from the low mark to the high mark (at the moment the table is sorted on student number).

Highlight the cells as shown in figure 2, then choose *Data*|*Sort* from the ribbon. Choose *sort by Mark* and press *OK* to sort the data by mark.

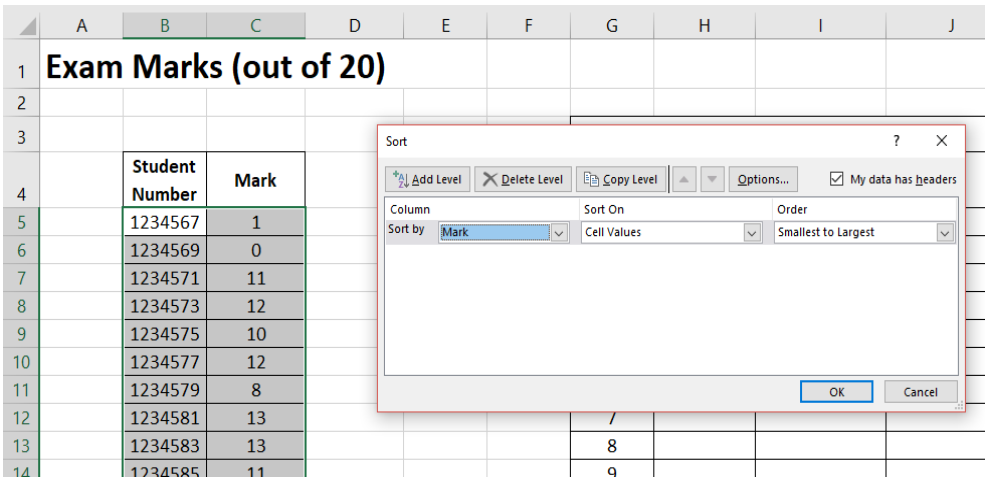

Fig. 2

Step 4

In cells G4:J26 you will see a frequency table which is not yet completed (figure 3).

To fill in the frequency column we start by adding the following formula into cell H5:

=frequency(C5:C104,G5:G25)

The first parameter is the data array that we want to analyse. The second is an array of what Excel refers to as *bins*. A bin is the range (or width) of the data we wish to be included in the frequency calculation. In this case we do not want a grouped table, so the effective width of each bin is 1. The bin is defined by its upper value, so in our case the bin array is simply the marks from 0 to 20 as represented in the range G5:G25.

Frequency Table			
Mark x	Frequency f	Cumulative Frequency	fx
0	1		
1	1		
2	1		
3	2		
4	3		
5	2		
6	4		
7	6		
8	5		
9	9		
10	10		
11	12		
12	14		
13	7		
14	7		
15	5		
16	4		
17	3		
18	1		
19	2		
20	1		
Σf		Σfx	

Fig. 4

	G	H	I	J
	Frequency Table			
	Mark x	Frequency f	Cumulative Frequency	fx
	0			
	1			
	2			
	3			
	4			
	5			
	6			
	7			
	8			
	9			
	10			
	11			
	12			
	13			
	14			
	15			
	16			
	17			
	18			
	19			
	20			
	Σf		Σfx	
			Mean	

Fig. 3

Once you have entered the formula into H5 you will see that a 1 appears in that cell, representing the number of students who obtained a mark of zero, the first of our bins. We need to reproduce this in the rest of the cells. To do this, highlight the range H5:H25 then press F2 followed be Ctrl+Shift+Enter. This produces an array of frequencies as show in figure 4.

You can now see that the mode is 12, which is what you should have got from the formula you entered step 2.

You can also check the median value by completing the cumulative frequency column. To do this enter =H5 into cell I5 so that the value in H5 is copied to I5. Then in I6 enter

=I5+H6

and then copy this into the remaining cells in column I. You will now see the cumulative frequencies as shown in figure 5.

Frequency Table			
Mark x	Frequency f	Cumulative Frequency	fx
0	1	1	
1	1	2	
2	1	3	
3	2	5	
4	3	8	
5	2	10	
6	4	14	
7	6	20	
8	5	25	
9	9	34	
10	10	44	
11	12	56	
12	14	70	
13	7	77	
14	7	84	
15	5	89	
16	4	93	
17	3	96	
18	1	97	
19	2	99	
20	1	100	
Σf		Σfx	

Fig. 5

As there are 100 items, the median value will be the average of the 50th and 51st items – you can see from the table that both of these will be 11, so the median is 11, which should tally with the result from the formula you entered in step 2.

Finally you can check your result for the mean, which you will remember is found by calculating:

$$\frac{\sum fx}{\sum f}$$

So you need to complete the final column (the product of f and x), total this column and the frequency column, then divide the first by the second. So the formula to enter into cell J28 will be:

=J26/H26

Step 5

The next step is to complete the grouped frequency table. We are going to divide the data into four groups, as shown in the table. So to complete the frequency column you follow the same procedure as you did at the beginning of step 4, but the bin array that you enter will now be N5:N8 (the upper bound of each bin). You results should be the same as in figure 6.

L	M	N	O

Grouped Frequency Table	
Mark x	Frequency f
$0 < x \leq 5$	10
$5 < x \leq 10$	34
$10 < x \leq 15$	45
$15 < x \leq 20$	11

Fig. 6

Step 6

We can now produce a bar chart from the frequency table we produced at step 4.

Highlight the data (cells G5:H25), then choose *insert* from the ribbon, then press the arrow on the charts menu as shown in figure 7. A dialogue will appear (figure 8) – choose the *All Charts* tab, then choose the second clustered column as shown.

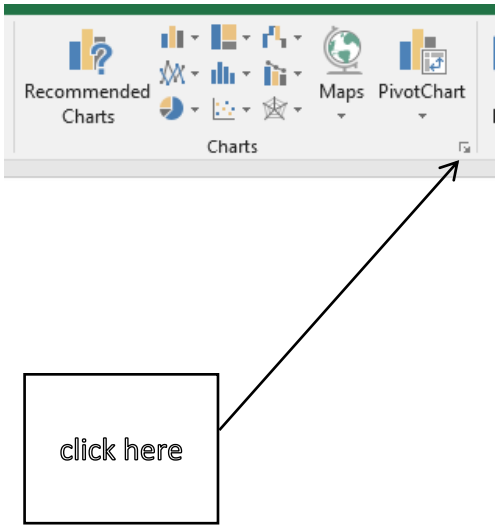

Fig. 7

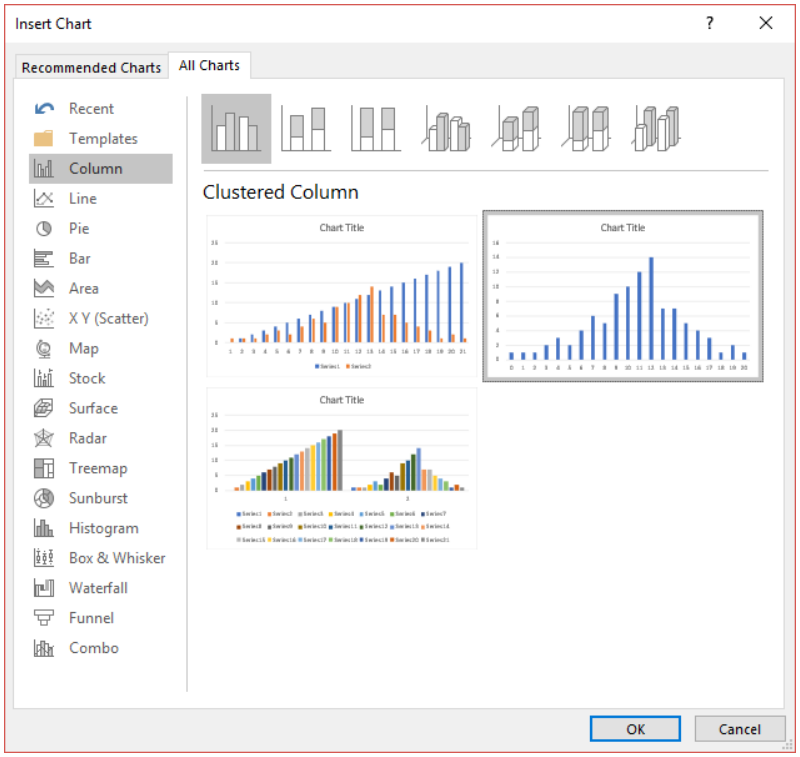

Fig. 8

The bar chart shown in figure 9 will now appear (Excel calls it a clustered column).

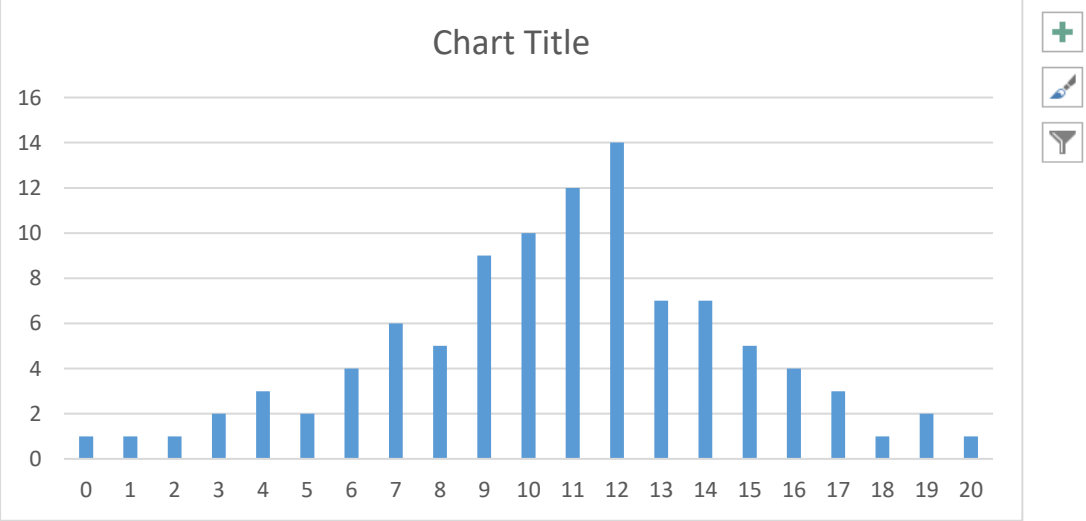

Fig. 9

You can now format the chart as in figure 10 - you can change the title, and, by clicking on the green cross that appears when the chart is selected, you can add axis titles. If you want to change the width of the bars, you can do this by double-clicking on one of them – you will then be given the option of reducing the gap between the bars, hence making them wider.

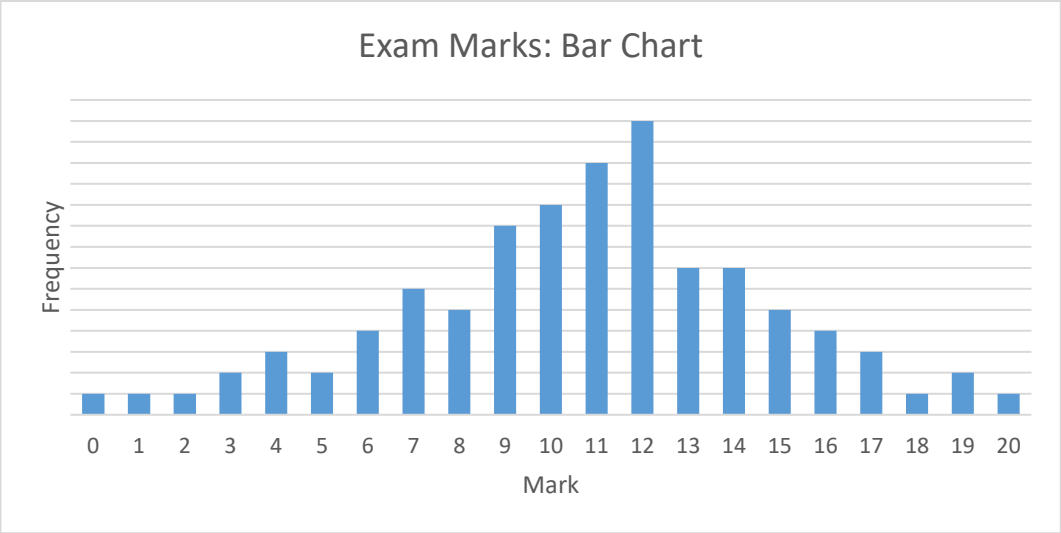

Fig. 10

Step 7 - You will need the version of Excel provided with Office 365 in order to complete this successfully (not currently available on the UEL build)

The last task will be to create a frequency polygon based on the grouped frequency table in figure 6. To do this we will produce what Excel calls a "histogram". However, this is actually a frequency polygon rather than a histogram because the frequency is represented by the height of the bars rather than the area of the bars as in a true histogram.

The easiest way to do this is to highlight the data, cells C5:C124, and this time select *histogram* when you insert the chart. The initial result is shown in figure 11.

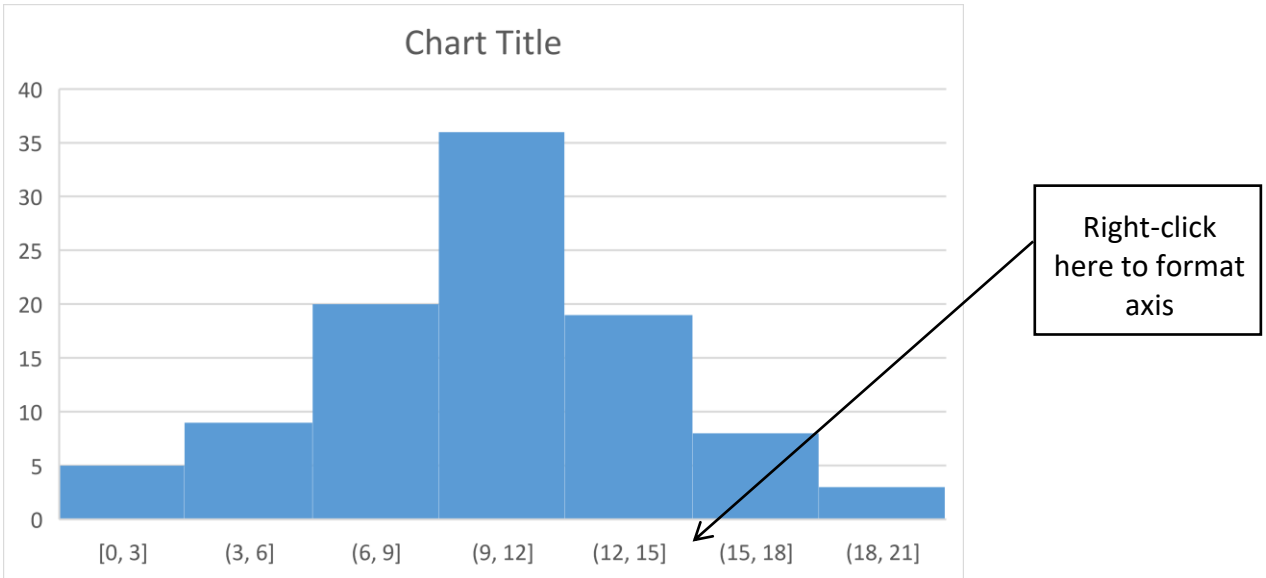

Fig. 11

As you can see, the data has been grouped into seven groups of three. To chose four groups of five (as in the grouped frequency table of figure 6), right-click on the horizontal axis as shown, and then choose *Format Axis*. In the dialogue that appears (figure 12) choose a bin width of 5, as shown.

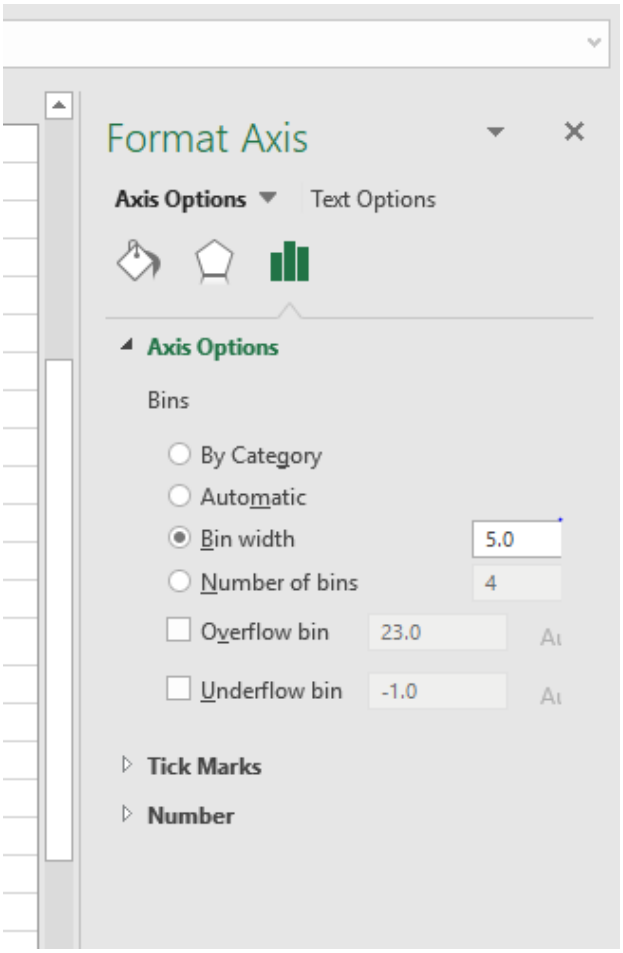

Fig. 12

The result is shown in figure 13. As before, you can change the chart title and add axis labels to get the final result that you see in figure 14.

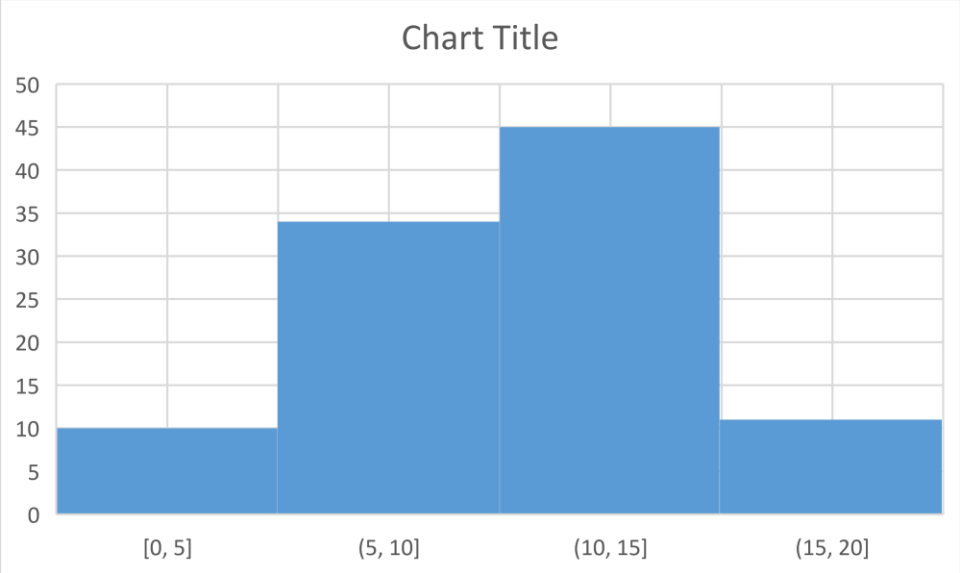

Fig. 13

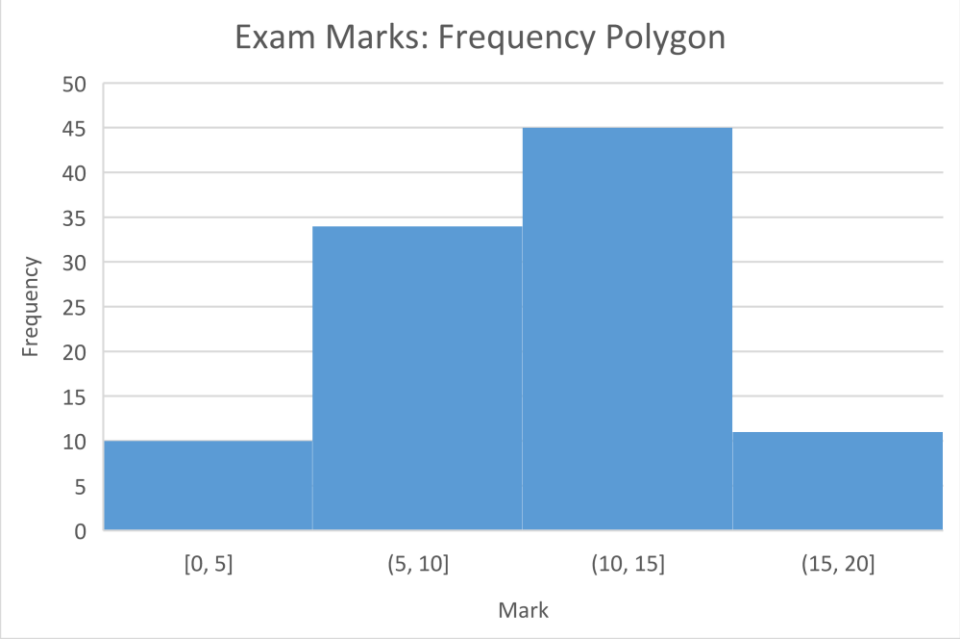

Fig. 14

Don't forget to upload your work (5 marks).