Practical Session 4

Measures of location

In this session, we illustrate how to calculate different measures of location using Excel.

1. Calculation of summary statistics given the full sample

   In the theory class we have used the following data set of times voted by a group of Madrileños.

Times	3	3	3	4	1	2	4	5	2	3	1	1	3	8	4	1	3	4	2	5	0	0	5	4	2	1	2	3	3	2
voted	1	4	3	2	3	5	0	6	3	1	3	5	4	1	4	1	2	4	4	3	3	0	7	2	2	1	3	4	2	2

   Last week we illustrated these data with a bar chart. Now we can calculate some standard measures of location.

   Firstly, for discrete, unimodal data, the function MODE (MODA) will calculate the mode of the sample. The median and mean can be calculated using the functions MEDIAN (MEDIANA) and AVERAGE (PROMEDIO) respectively.

   For the above data set, we have mode = median = 3 and mean = 2.817.

2. Calculation of the mode for multimodal data.

   Although the above sample is unimodal, we may sometimes have data with multiple modes such as the following example.

1	2	2	1	3	1	4	2

   To see if there are multiple modes, we can use the function MODE.MULT (MODA.VARIOS) in Excel. This function is an array function so, in order to use it, we first mark, for example 4 blank cells in order to output the results. Then, assuming the data are in cells **A1:A8**, we type =**MODE.MULT(A1:A8)** in the command bar and, instead of pressing the Return key, we press **Ctrl**, **Shift** and **Return** at the same time.

   We should see a result like the following:

1	2	#N/A	#N/A

   This indicates that 1 and 2 are modes and there are no further modes.

3. Calculation of quartiles and percentiles

   Excel has 3 different functions that can be used to calculate quartiles of a data set: QUARTILE (CUARTIL) which was used in older versions of Excel, QUARTILE.INC (CUARTIL.INC) and QUARTILE.EXC (CUARTIL.EXC).

If, for example we wanted to calculate the third quartile (using **QUARTILE.EXC**) of a set of data in cells **A1:A10**, then we would write =**QUARTILE.EXC(A1:A10;3)**. We could calculate other quartiles by setting the second entry to be a number in 0, …, 4. What happens when we set this to be 0, 2 or 4?

As you can see in Wikipedia, there are various, slightly different ways of calculating quartiles. In our case, the function **QUARTILE.EXC** corresponds to the approach we have used in class.

For the above data set, we should find the first quartile is 2 and the third quartile is 4.

As with the calculation of quartiles, three Excel functions: PERCENTILE, (PERCENTIL), PERCENTILE.INC (PERCENTIL.INC) and PERCENTILE.EXC (PERCENTIL.EXC) are available. The syntax for calculating the 70% percentile of data in cells **A1:A10** is =**PERCENTILE.EXC(A1:A10;0.7)**. Remember to write the fraction as 0,7 if your number settings are in Spanish though!

4. Estimating statistics from a table for continuous data.

The frequency table for town hall spending was as below.

Yearly spending (€)	Absolute frequency
≤ 30	0
(30,45]	2
(45,60]	9
(60,75]	9
(75,90]	10
(90,105]	3
(105,120]	3
> 120	0
Total	36

Last week, in order to generate a histogram, we extended the table as follows.

Lower limit	Centre	Upper limit	Absolute frequency (n)	Relative frequency (f)	Cumulative frequency (N)	Cumulative relative frequency (F)
15	22,5	30	0	0,00	0	0,00
30	37,5	45	2	0,06	2	0,06
45	52,5	60	9	0,25	11	0,31
60	67,5	75	9	0,25	20	0,56
75	82,5	90	10	0,28	30	0,83
90	97,5	105	3	0,08	33	0,92
105	112,5	120	3	0,08	36	1,00
120	127,5	135	0	0,00	36	1,00
Total			36	1,00		

Using this table, we can see immediately that the modal class is (75-90] and the median class is (60-75]. To estimate the mean, we can multiply the centre values of each interval by Relative frequency and sum, or multiply by absolute frequency, sum and divide by the total frequency (36). Remember to use the dollar trick when you do these calculations.