

# Practical Session 5

## Measures of dispersion and other measures of form

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

1. Calculation of the range and IQR.

Yearly spending of different town halls (€)

114579	73896	59003	86165	53428	93844	61536	90628	49501
56767	78063	87750	82409	107664	60479	88872	66325	78268
38360	82436	83531	81364	63210	112842	56206	59052	52660
45000	91562	66308	50397	79964	65369	71803	60108	49264

Two weeks ago, we generated a histogram for these data. Now we shall calculate some measures of spread.

Firstly, the range of the data can be calculated by using the functions **MAX** and **MIN** and then calculating the difference.

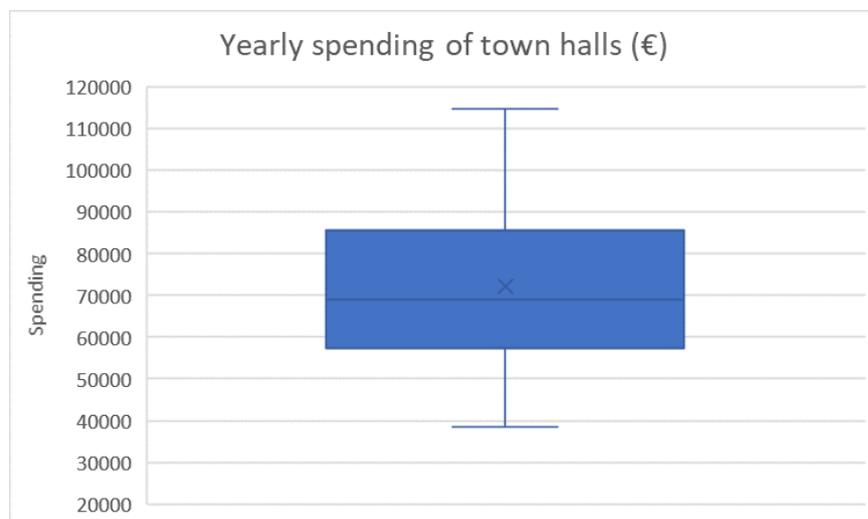
Similarly, the interquartile range can be calculated by using the function **QUARTILE.EXC** to calculate the first and third quartiles and again evaluating the difference.

For these data, the range is €76219 and the IQR is €28180.5.

2. Using the box and whisker plot to illustrate the sample.

In Excel 2016 and Excel 365, the box and whisker plot is available as one of the standard graphics options. To use this, we should first put the data into a single column and then, mark the data and select the box and whisker plot option (caja y bigotes) option from the graphics options in Excel.

The graph for the town hall data should look something like the following.



The box and whisker plot indicates no outliers for this data. (Note that mild outliers will be marked but Excel does not distinguish between mild and extreme outliers.)

In Excel 2013, there is no direct method of constructing a box and whisker plot but, for example, you can follow the instructions [here](#) in order to construct the plot. Using this method, outliers are not identified on the plot however.

3. Calculation of variance, standard deviation, coefficient of variation, skewness and kurtosis.

In Excel, variance can be calculated using the functions **VAR** (for compatibility with old versions and equivalent to **VAR.S**), **VAR.P** and **VAR.S**. The difference between **VAR.P** and **VAR.S** is that **VAR.P** uses a divisor of N and corresponds to  $(\hat{\sigma}^2)$  as we have done in class and **VAR.S** uses a divisor of N-1 corresponding to  $s^2$ .

The standard deviation can be calculated using the square root of the variance with the function **SQRT (RAIZ)** or directly using the function **STDEV.P (DESVEST.P)** or **STDEV.S (DESVEST.S)**.

Although Excel does not have a direct function to calculate the coefficient of variation, you can just divide the standard deviation by (the absolute value of) the mean to generate this.

Skewness can be calculated using the functions **SKEW.P (COEFICIENTE.ASIMETRIA.P)**, which corresponds to the formula used in class or **SKEW (COEFICIENTE.ASIMETRIA)**.

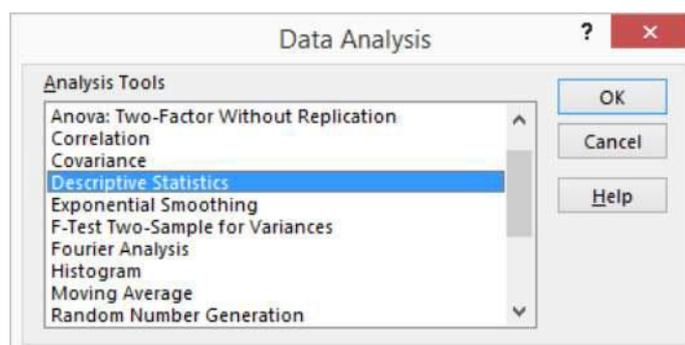
Finally, kurtosis can be calculated using the function **KURT (CURTOSIS)**. This corresponds to the sample or unbiased kurtosis estimate which is slightly different to the formula provided in class.

Try to calculate some of these statistics for the town hall data. How can you interpret them?

4. A fast way of calculating various summary statistics.

It is slightly tedious to calculate all these statistics individually. However, the **Descriptive Statistics (Estadística Descriptiva)** option in the **Data Analysis Add On (Complemento de Análisis de Datos)** can be used to allow us to perform many of these calculations.

First, we need to install the add on by going to **File, (Archivo) Options, (Opciones) Add Ons (Complementos)** and **Go (Ir)** and then marking the **Data Analysis (Análisis de Datos)** option. If you are doing this in the Computer Labs, you will need to repeat this procedure every time you want to use the Data Analysis Add On but on your home pc, you should only need to do this once.



When the Data Analysis Add On has been installed, you can go to the option Data on the menu bar and then the Data Analysis option. In the Data Analysis dialogue box, you then use the **Descriptive Statistics (Estadística Descriptiva)** option as shown above.

You need to then input the data range, (if the data have a title in the first row, mark this option) and the **Summary Statistics (Resumen de estadísticas)** options. Note that you should have all the data in the same row or column to use this option. In this case, the results should appear something like the following:

<i>Gastos de ayuntamientos</i>	
Media	72831,7714
Error típico	3201,98757
Mediana	71803
Moda	#N/A
Desviación estándar	18943,214
Varianza de la muestra	358845355
Curtosis	-0,32571676
Coefficiente de asimetría	0,41751189
Rango	76219
Mínimo	38360
Máximo	114579
Suma	2549112
Cuenta	35

Note that standard deviation, variance, skewness and kurtosis are calculated using the alternative formulas (VAR.S, ...) to those used in class.

One extra statistic produced is the standard error, which is not relevant for us as a summary statistic at the moment.