

# Microsoft Excel – Data Analysis

#### **Statistical Functions**

Lecturer: Claudio Vallar School of Economics and Finance

• Descriptive statistics are useful because they describe and summarize a large amount of data.

 Analyse data is very easy with the Excel statistical functions. The most important is to understand the purpose of each one of these functions.



• The **arithmetic mean** is the most common indicator of central tendency of a variable. The mean is the

mathematical average of a set of two or more numbers.

=AVERAGE(range of cells with the values of interest)

• The **median** is the middle number in a sorted, ascending or descending, list of numbers.

=MEDIAN(range of cells with the values of interest)

• The **mode** refers to the most frequent, repeated or common number in the data.

=MODE(range of cells with the values of interest)

• **Minimum** (Min) is the lowest value in an array of values.

=MIN(range of cells with the values of interest)

• **Maximum** (Max) is the largest value in an array of values.

=MAX(range of cells with the values of interest)

• **Percentile**: the value below which a percentage of data falls.

=PERCENTILE.INC(array, percentile)

Percentile argument is a value between 0 and 1

• **Quartile**: the value below which a percentage of data falls. Dataset divided into four intervals.

=QUARTILE.INC(array, quart)

Quartile argument is an integer between 0 and 4

• **Countif** function returns the number of cells within a supplied range, that satisfy a given criteria.

=COUNTIF(range, criteria)

• Variance is a measurement of the spread between numbers in a data set. It measures how far each number in the set is from the mean.

It is mathematically defined as the average of the squared differences from the mean..

Function	Description	Excel Formula
VAR.P	Returns the variance of a supplied set of values (which represent an <b>entire population</b> )	=VAR.P(number1, [number2], )
VAR.S	Returns the variance of a supplied set of values (which represent a <b>sample</b> of a population)	=VAR.S(number1, [number2], )

• Population Variance: 
$$\sigma^2 = \frac{\sum_i (x_i - \mu)^2}{N}$$
; Sample Variance  $s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n-1}$ 

• The **standard deviation** is a statistic that measures the dispersion of a dataset relative to its mean and is calculated as the square root of the variance.

We have two functions: STDEV.P and STDEV.S

The covariance is a statistical measurement of the strength of the correlation between two sets of variables

Function	Description	Excel Formula
COVARIANCE.P	Calculates the population covariance of two supplied sets of values.	=COVARIANCE.P( array1, array2 )
COVARIANCE.S	Calculates the population covariance of two supplied sets of values.	=COVARIANCE.P( array1, array2 )

- The correlation coefficient is a statistical measurement of the correlation (linear association) between two sets of values.
  - Correlation coefficient lies between -1 and 1.

=CORREL(array1, array2)

Note that these arrays should be of equal length.



