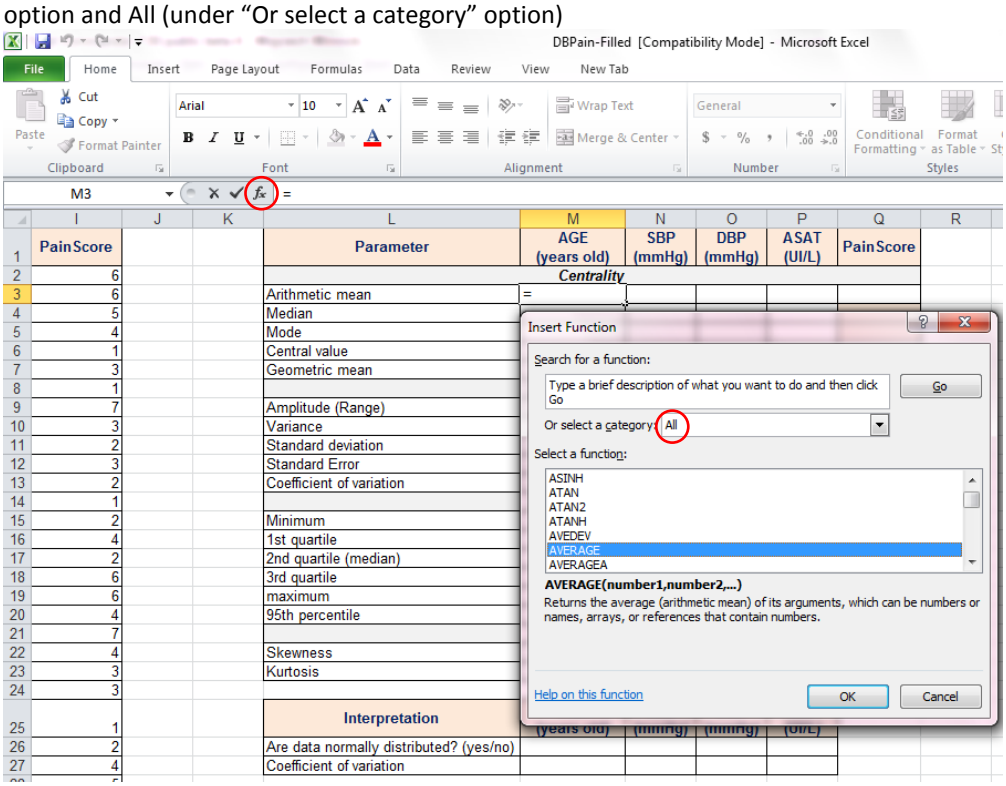
MICROSOFT EXCEL: DESCRIPTIVE STATISTICS

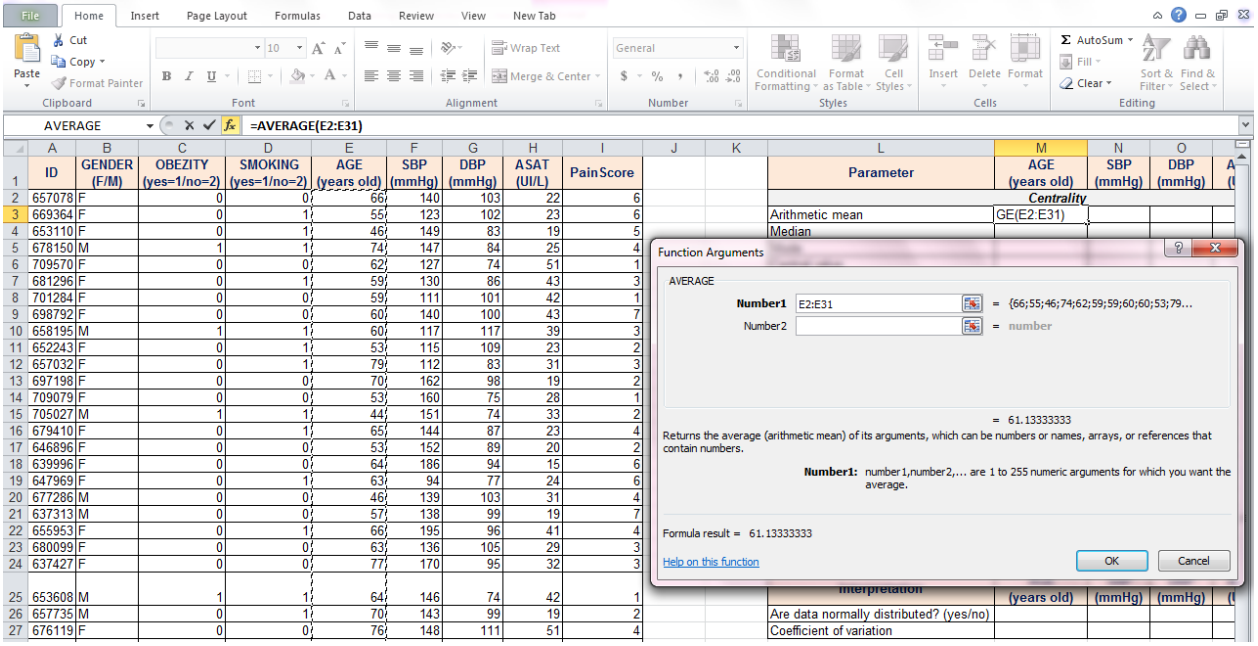
Working with predefine functions

To use predefined function:

- Place the mouse in the cell where you want to obtained the result and from **Insert** menu chouse ***fx…*** option and All (under "Or select a category" option)



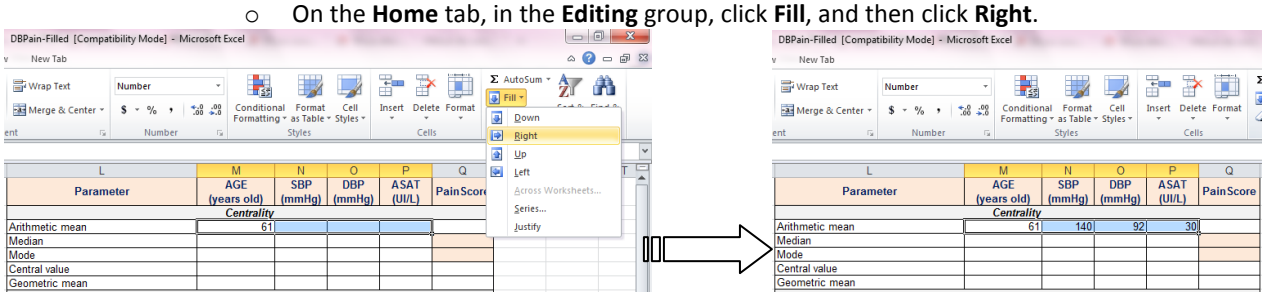
- From the "Select a function" window chouse the function according with the statistic that you want to calculate. For example, to calculate the arithmetic mean of Age the AVERAGE predefine function will be choose. In the window of the function (Function Arguments), under the Number1, please select the array were the data for age are in the data raw (e.g. E2:E31 for Age)



- Validate the selection of the data by click on OK button.

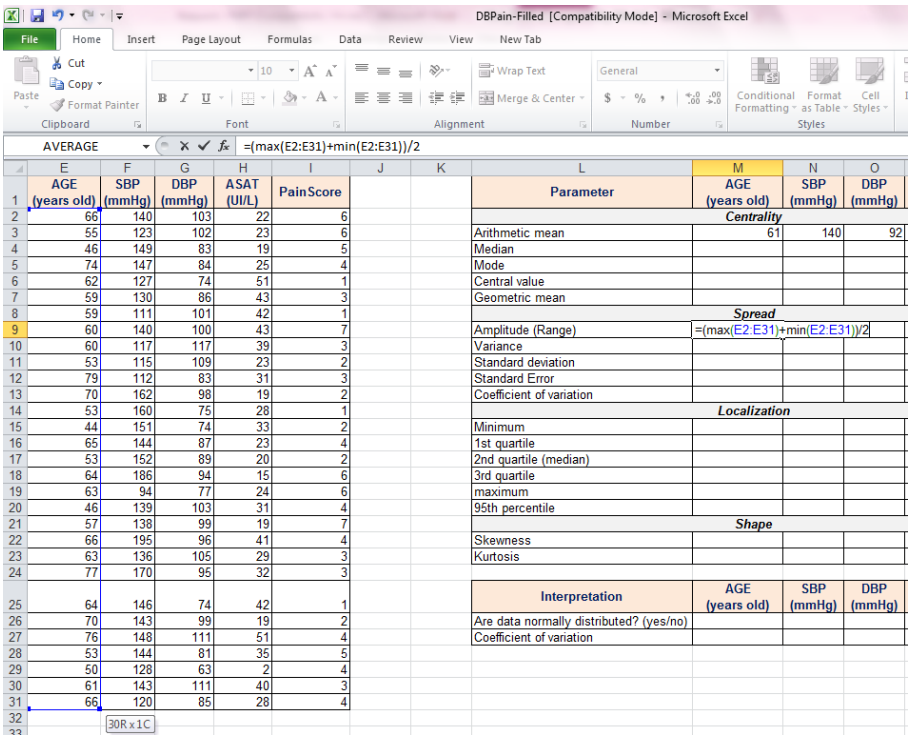
- Once you include the function in the first cell, you can use copy the inserted function for the next variables:
  - Select the cells that contain the data that you want to fill into adjacent cells.
  - Drag the fill handle across the cells that you want to fill.

  OR
  - Select the cells that contain the data that AND the cell you want to copy the function
  - On the **Home** tab, in the **Editing** group, click **Fill**, and then click **Right**.



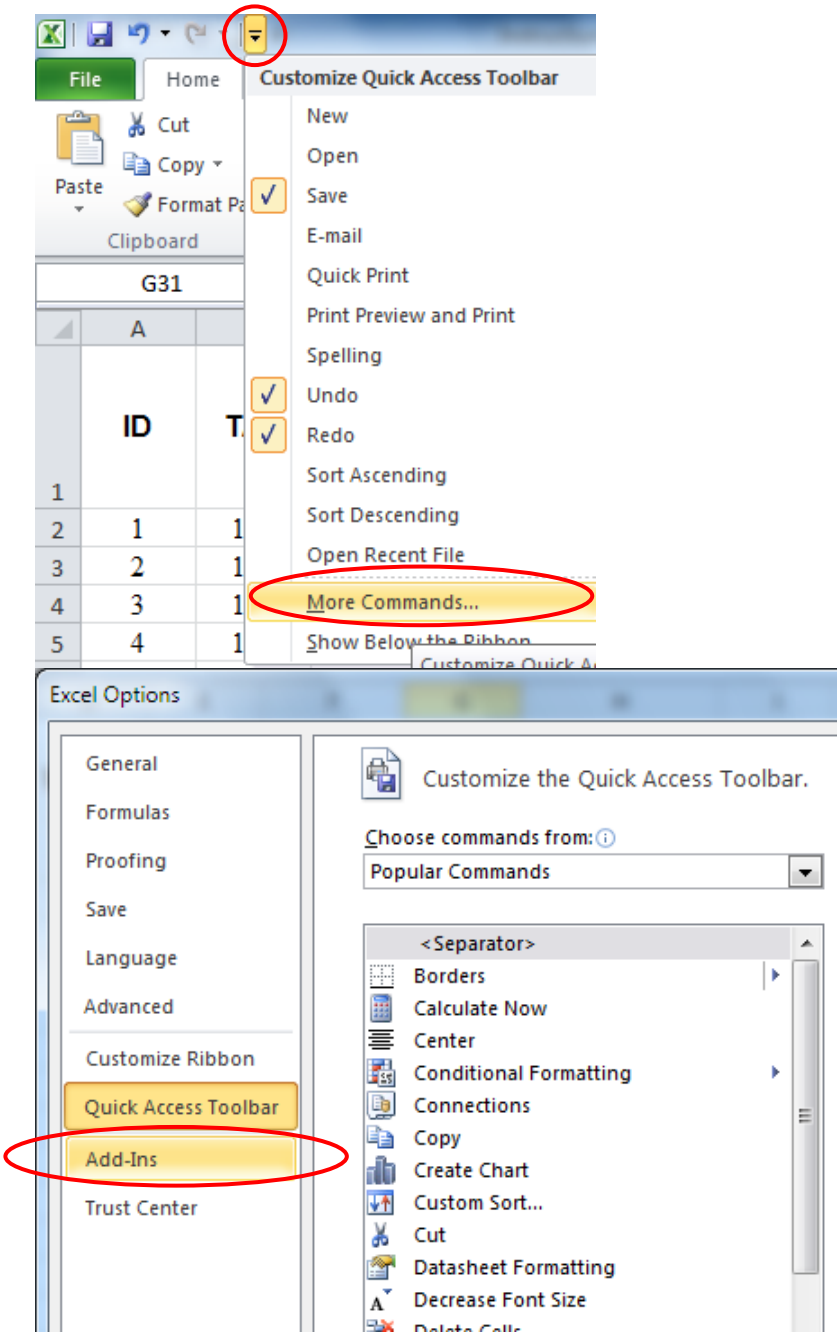
User-defined functions

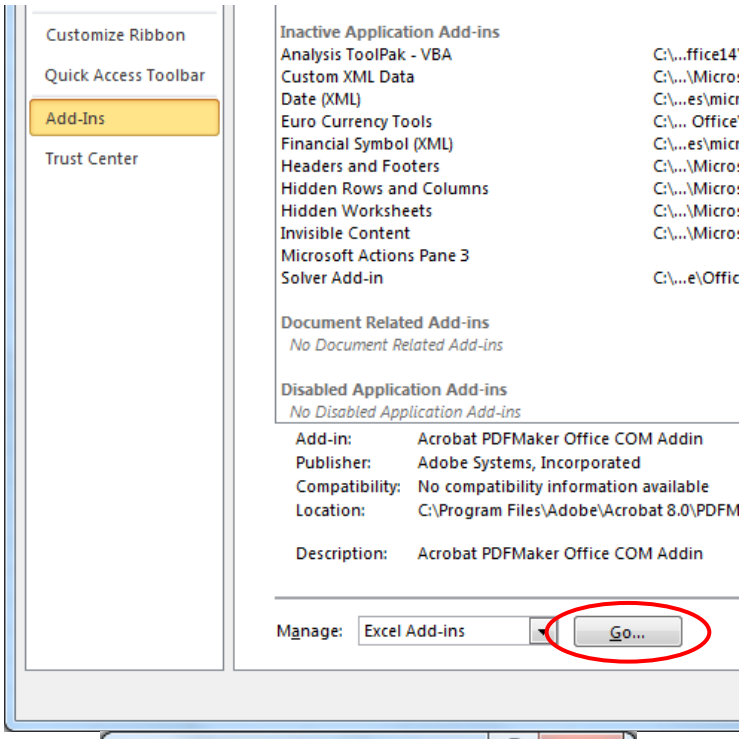
- To implement user-defined functions using predefine functions (exemplification for **central value**):
  - Select the cell were you want the results
  - = (equal sign)
  - Open the first round bracket ('(')
  - Type MAX
  - Open a second round bracket ('(')
  - Select the array of data (E2:E31) for Age
  - Close the second round bracket (')')
  - +
  - Type MIN
  - Open the third round bracket ('(')
  - Select the array of data (E2:E31) for Age
  - Close the third round bracket (')')
  - Close the first round bracket (')')
  - / (division sign)
  - 2
  - Validate the function with ENTER

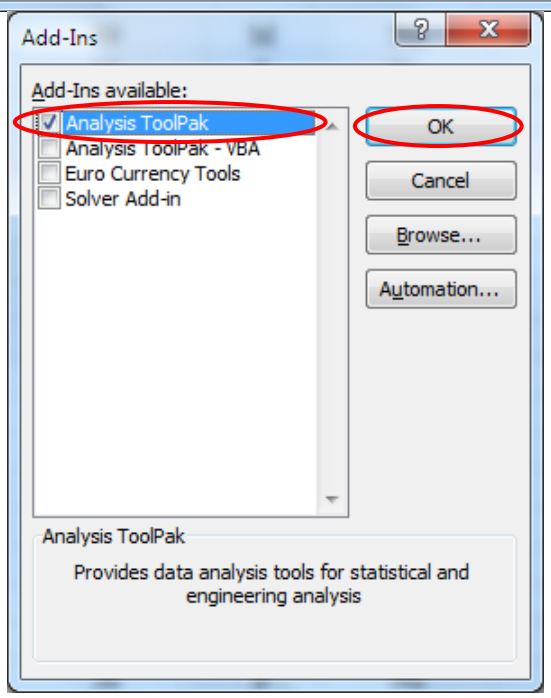


Installation of Data Analysis

[Microsoft Excel – More commands - Add-Ins- Analysis ToolPak]

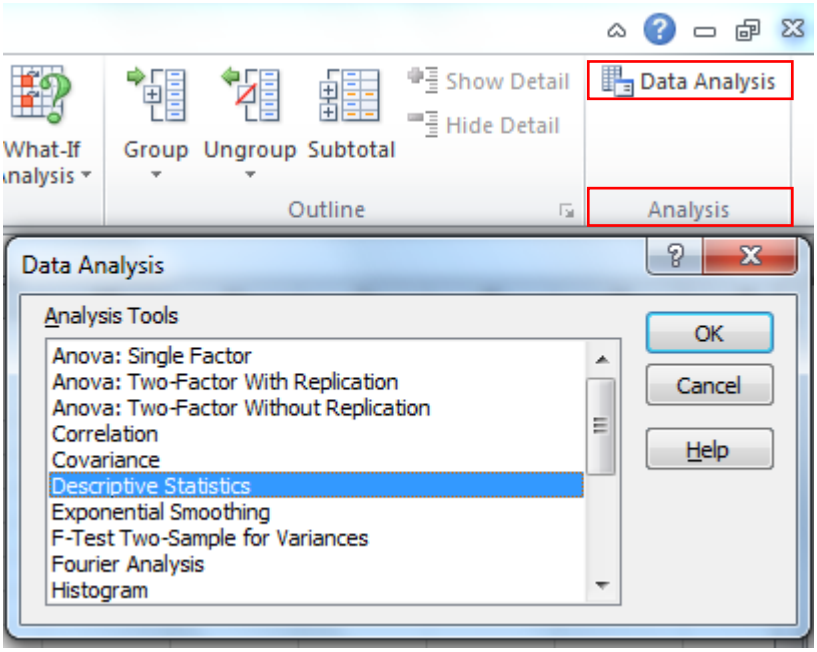




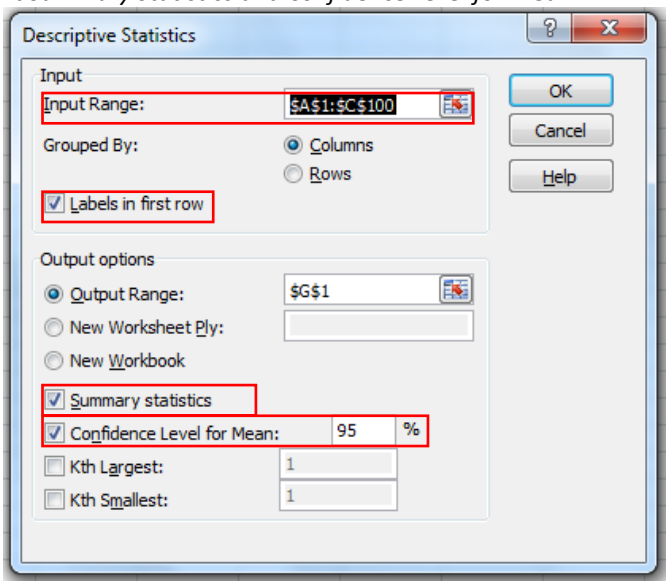


To compute descriptive statistics parameters by example

1. [Data – Analysis - Data Analysis – Descriptive statistics]



2. Descriptive Statistics dialog box:
   1. Input Range: select the range where the data (including the label of variable) are (e.g. for our request the data are $A$1:$A$100).
   2. Specify that you have Labels in first row.
   3. Output options: Output range: specify the first cell from which the output will be displayed (place the output in the same worksheet as the date).
   4. Specify that you want *Summary Statistics* and *Confidence Level for Mean*.



   5. The result will be like in the image bellow (it is just an example and did not included the data used in your exercise):

	A	B	C	D	E	F	G	H
1	*Age (Years)*		*Height (m)*		*Weight (kg)*		*PI (kg/m3)*	
2								
3	Mean	9.536585	Mean	1.434472	Mean	54.15854	Mean	17.86532
4	Standard E	0.314959	Standard E	0.017014	Standard E	1.7314	Standard E	0.235473
5	Median	9	Median	1.44	Median	50	Median	17.46244
6	Mode	7	Mode	1.48	Mode	50	Mode	17.48148
7	Standard I	3.493067	Standard I	0.188694	Standard I	19.20215	Standard I	2.611522
8	Sample Va	12.20152	Sample Va	0.035605	Sample Va	368.7226	Sample Va	6.820048
9	Kurtosis	-0.2201	Kurtosis	0.031329	Kurtosis	0.415053	Kurtosis	0.44969
10	Skewness	0.310202	Skewness	0.049093	Skewness	0.70613	Skewness	0.680054
11	Range	16	Range	0.92	Range	99	Range	13.09715
12	Minimum	2	Minimum	0.94	Minimum	18	Minimum	12.7431
13	Maximum	18	Maximum	1.86	Maximum	117	Maximum	25.84025
14	Sum	1173	Sum	176.44	Sum	6661.5	Sum	2197.434
15	Count	123	Count	123	Count	123	Count	123

6. Move the name of variables one cell to the right and delete redundant columns. Display the decimal numbers with 2 decimals. Your results will look like those in the image bellow:

K	L	M	N	O	P	Q	R	S	T
	Age (Years)	*Height (m)*	*Weight (kg)*	*PI (kg/m3)*	*Cholesterol (mg/dL)*	*HDL (mg/dL)*	*TG (mg/dL)*	*LDL1 (mg/dL)*	*LDL2 (mg/dL)*
Mean	9.54	1.43	54.16	17.87	162.60	39.61	86.63	105.67	108.23
Standard Error	0.31	0.02	1.73	0.24	3.15	0.66	3.91	3.18	3.68
Median	9	1.44	50	17	162	40	80	102.4	103.9091
Mode	7	1.48	50	17	164	40	58	110.2	#N/A
Standard Deviation	3.49	0.19	19.20	2.61	34.95	7.32	43.35	35.24	40.76
Sample Variance	12.20	0.04	368.72	6.82	1221.64	53.65	1879.46	1242.06	1661.75
Kurtosis	-0.22	0.03	0.42	0.45	15.09	-0.37	2.50	14.62	3.93
Skewness	0.31	0.05	0.71	0.68	2.47	0.01	1.33	2.45	1.22
Range	16.00	0.92	99.00	13.10	311.00	36.00	239.00	305.80	299.81
Minimum	2.00	0.94	18.00	12.74	84.00	24.00	11.00	32.80	0.91
Maximum	18.00	1.86	117.00	25.84	395.00	60.00	250.00	338.60	300.73
Sum	1173.00	176.44	6661.50	2197.43	20000.00	4872.00	10656.00	12996.80	13312.05
Count	123	123	123	123	123	123	123	123	123

7. Interpretation by example (birth weight variable):

Mean	The arithmetic average of the 99 newborn child included into the study was equal with 3143.63 gram.
Standard Error	The standard error of the mean for the birth weight was of 53.66.
Median	The observation that split the distribution of birth weight in half was equal with 3200 gram.
Mode	The observation value associated with the highest frequency is equal for our study with 3000 gram.
Standard deviation	The population standard deviation for birth weight is equal with 533.94.
Variance	The standard deviation squared for birth weight was equal with 285096.85.
Kurtosis	The distribution of the birth weight is peakedness distribution comparing with normal distribution, kurtosis being equal with 2.55.
Skewness	The negative value of the -0.85 for our sample research problem indicates that the distribution of the birth weight is negatively skewed. The negative skew indicates that the longer tail extends in the direction of low values in the distribution.
Range	The range for our distribution is found by subtracting 930 from 4400, producing a range equal to 3470.
Minimum	The lowest value of birth weight by newborn in the studies sample was of 930.
Maximum	The highest value of birth weight by newborn in the studied sample was of 4400.
Sum	The sum of the values in the distribution in the studied sample was of 311220.
Count	The number of observations in the birth weight distribution the studied sample, n = 99
Confidence level (95.0%)	The value obtained represents the amount of error subtracted from and added to the sample mean when constructing the confidence interval from the population mean. For our problem, the 95% confidence interval is: $3037.14 \leq \mu \leq 3250.13$