MICROSOFT EXCEL: DESCRIPTIVE STATISTICS

Learning Objectives:

- Working with user-define functions (relative and absolute cell references). (Exercise 1 & 2)
- Working with predefined functions / formulas– descriptive statistics functions [Insert – Functions ...]. (Exercise 1)
- Interpretation of descriptive statistics parameters (coefficient of variation, skewness and kurtosis) (Exercise 1).

EXERCISE 1

A study has been conducted on a sample of 30 patients who presented in Emergency room for chest pain. The following data has been collected for each patient: gender, age (years), obesity (yes/no), smoking statius (yes/no), systolic blood pressure (mmHg), diastolic blood pressure (mmHg), alanine transaminase (mg/dL) - AST, and the pain at 30 minutes after admission in emergency room (no medication was applied). The pain was measured using Likert Scale and the significance of the scale are as follows: 1 = The pain has become intolerable; 2 = The pain is substantially worse; 3 = The pain is a bit worse; 4 = The pain is the same; 5 = The pain has improved a bit; 6 = The pain has improved substantially; and 7 = The pain has disappeared. Data were collected in **DBPain.xls** file.

1. Copy the data from **DBPain.xls** file into a new Microsoft Excel file named **Data_PA05.xlsx** and save it in the in **Lab05** folder. Name the sheet with the data as Data.

2. Calculate for each patient the systemic pulse pressure (SPS column) of using the following formula:

   SPS = SBP - DBP

3. Calculate for each patient the mean arterial pressure (MAP column) of using the following formula:

   MAP = DBP+1/3*(SBP-DBP)

4. Display the ASTstatus of each patient using the following criterion:

   A patient is considering to be with positive ASTStatus (*yes* in the column) IF AST is >=40

5. Insert a new sheet and name it Descriptive Statistics. Copy in *Descriptive Statistics* sheet all data from Data sheet using Paste Special option **[Home – Paste – Paste Special – Values and Numbers Format]**.

6. Move the column as shown in the figure bellow:

	A	B	C	D	E	F	G	H	I	J	K	L
1	ID	GENDER (F/M)	OBESITY (yes/no)	SMOKING (yes/no)	ASTStatus (pos/neg)	AGE (years)	SBP (mmHg)	DBP (mmHg)	SPS (mmHg)	MAP (mmHg)	AST (IU/L)	PainLikert Scale
2	657078	F	no	no	negative	66	140	103	37	115	22	6
3	669364	F	no	yes	negative	55	123	102	21	109	23	6
4	653110	F	no	yes	negative	46	149	83	66	105	19	5

7. Create in the Descriptive Statistics sheet a table with the structure presented in the image bellow and compute for each variable the measures of centrality, spread, symmetry and localization:

Parameter	AGE (years)	SBP (mmHg)	DBP (mmHg)	SPS (mmHg)	MAP (mmHg)	AST (IU/L)	PainLikert Scale
Measures of centrality							
Aritmetic mean							
Median							
Mode							
Central value							
Geometric mean							
Harmonic mean							
Measures of spread							
Range							
Standard deviation							
Standard error							
Coefficient of variatio (CV)							
Measures of symmetry							
Skewness							
Kurtosis							
Measures of localization							
1st Quartile							
2nd Quartile (median)							
3rd Quartile							
95th Percentile							

8. Compute for each variable all descriptive statistics which have sense to be computed according with the type of variable. For example you will compute just median, mode, and measures of localization for *Pain Likert Scale*.

9. Under the table with descriptive statistics parameter, create a new table with interpretation of folowing statistics:

Interpretation	AGE (years)	SBP (mmHg)	DBP (mmHg)	SPS (mmHg)	MAP (mmHg)	AST (IU/L)
Coefficient of variation (CV)						
Skewness						
Kurtosis						

For interpretation use the following information:

▪ Coefficient of variation:

Coefficient of Variation (CV)	Interpretation: The population could be considered
CV < 10%	Homogenous
10% ≤ CV < 20%	Relative homogenous
20% ≤ CV < 30%	Relative heterogeneous
> 30%	Heterogeneous

▪ Skewness:

a. If skewness is less than −1 or greater than +1, the distribution is highly skewed.

b. If skewness is between −1 and −½ or between +½ and +1, the distribution is moderately skewed.

c. If skewness is between −½ and +½, the distribution is approximately symmetric.

▪ Kurtosis:

d. A normal distribution has kurtosis exactly 3 (excess kurtosis exactly 0). Any distribution with kurtosis $\cong$ 3 (excess $\cong$ 0) is called mesokurtic.

e. A distribution with kurtosis <3 (excess kurtosis <0) is called platykurtic. Compared to a normal distribution, its central peak is lower and broader, and its tails are shorter and thinner.

f. A distribution with kurtosis >3 (excess kurtosis >0) is called leptokurtic. Compared to a normal distribution, its central peak is higher and sharper, and its tails are longer and fatter.

Based on the interpretation of skewness and kurtosis what can you say about distribution of your data? Write your answer under the table with interpretation of CV, skewness and kurtosis.

EXERCISE 2

The next table contains marks (not real data) of 15 students at University of Medicine and Pharmacy Cluj-Napoca, Faculty of Medicine, first year of study obtained in the first session (winter session). Data could be fourn on **WMean.xls** file.

1. Save the **Wmean.xls** file on your partition of the server in the folder Lab05.
2. Compute for each student the arithmetic mean and display its value in the column named **_Mean_**.
3. Compute for each student the weighed mean and display its value in the column named **_Weighted Mean (WMean)_**. Use the following formula for weighted mean (where W_i = is the weight according to the number of credits and X_i = the mark):

$$m_x = \frac{\sum_{i=1}^{n} W_i X_i}{\sum_{i=1}^{n} W_i}$$

4. Compute for each student the general mean and display its values in the column **_General Mean (GMean)_**. For computing the general mean use the following formula:

General mean = (arithmetic mean + weighted mean)/2