

MICROSOFT EXCEL: DESCRIPTIVE STATISTICS II

Learning Objectives:

- Working with user defines functions. (Exercise 1 & 2)
- Creating r×n contingency table [Data - Pivot Table and PivotChart Reports]. (Exercise 1 & 2)
- Creating graphical representations of qualitative variables (one and two variables). (Exercise 1 & 2)
- Creating Histograms [Data – Data Analysis - Histogram]. (Exercise 1 and 2)
- Working with Descriptive Statistics option. (Exercise 1 and 2)

EXERCISE 1

1. Copy the data from **ChoIDB.xls** file into a new Microsoft Excel file named **Data_PA06.xlsx** and save it in the in **Lab06** folder. Name the sheet with the data as Data.

2. Calculate for each patient the value of ponderal index (PI column) of using the following formula:

$$PI = \text{weight(kg)}/\text{height(m}^3)$$

3. Calculate for each patient the value for LDL1 (LDL1 column) of using the following formula:

$$LDL1 = \text{Cholesterol} - \text{HDL} - \text{TG}/5.0$$

4. Calculate for each patient the value for LDL2 (LDL2 column) of using the following formula:

$$LDL2 = \text{Cholesterol}/1.19 + \text{TG}/1.9 - \text{HDL}/1.1 - 38$$

5. Display the IF-TG status as normal / high value of each patient using the following criterion:

A patient has a normal value IF TG < 150 mg/dL

6. Insert a new sheet and name it Graphs. Copy in Graphs sheet using Paste Special option [**Home – Paste – Paste Special – Values and Numbers Format**] all data from Data sheet. Create in this file the following graphical representations:

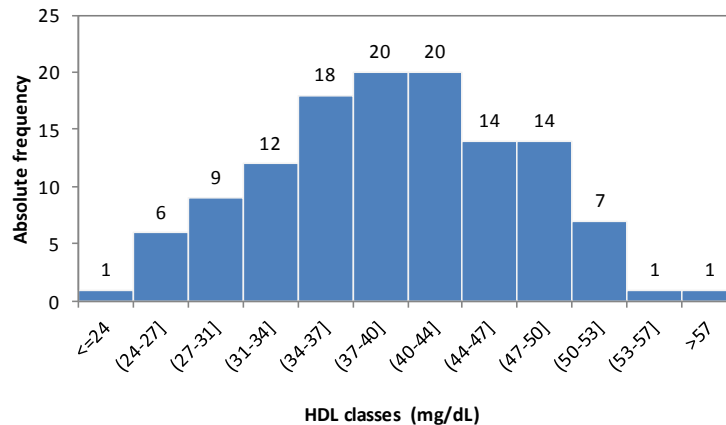
- Pie chart:** Use Pivot Table and PivotChart Reports option to create the frequency table gender. Based on the obtained frequency table create in Graphs sheet a Pie chart.
- 100% Stacked Column chart:** Use Pivot Table and PivotChart Reports option to create a contingency table with gender on rows and genetic susceptibility on columns. Copy the values of the contingency table in a new table as:

	Genetic susceptibility=yes	Genetic susceptibility=no
Female		
Male		

Based in the above table, create a 100% Stacked Column chart.

- Histogram (no constrains regarding the classes of frequencies):** Create the distribution of HLD. Your graphical representation is expected to be like the one in the image bellow:

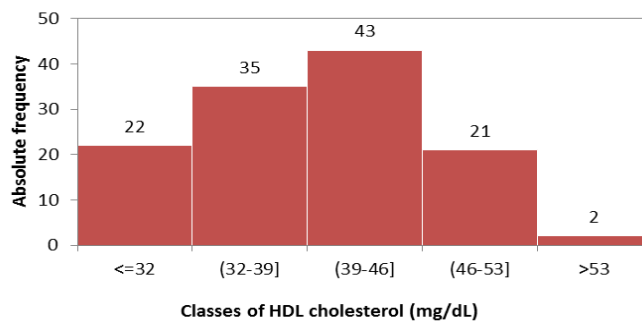
Histogram of HDL (mg/dL)



Is data presented in the Histogram graph normal distributed? Write your answer under the graphical representation and provide also the reason able to sustain your answer.

- d. **Histogram (classes of frequencies provided):** Create the distribution of HLD using the following frequency classes <=32, (32-39], (39-46], (46-53], and >53. Your graphical representation will look like the one in the image bellow:

Histogram of HDL cholesterol



7. 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**].
- a. Compute descriptive statistics parameters for all quantitative variables. Formating the obtained results to look like the one in the following image:

K	L	M	N	O	P	Q	R	S	T
	Age (Years)	Height (m)	Weight (kg)	PI (kg/m ³)	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

- b. Under the descriptive statistic table, calculate for each quantitative variable the value of coefficient of variation ($CV=(\text{standard deviation}/\text{mean})\cdot 100$):

K	L	M	N	O	P	Q	R	S	T
	Age (Years)	Height (m)	Weight (kg)	PI (kg/m ³)	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
Coefficient of variation									
Interpretation of CV									

- c. Interpret the value of coefficient of variation.
d. Sort the data ascending according to group and compute descriptive statistics parameters for each group (separately for case group and for control group).

EXERCISE 2

The **CholDB.xls** will be also use for the requests of the second exercise.

- Copy the data from **CholDB.xls** file into a new Microsoft Excel file named **Data2_PA06.xlsx** and save it in the in **Lab06** folder. Name the sheet with the data as **Data**.
- Insert a new sheet and name it **Histograms**. Copy in the **Histogram** sheet Gender and Cholesterol columns. Sort data according to gender and create 2 histograms of cholesterol: one for male and the other one for female.
- Insert a new sheet and name it **Column Chart**. Copy in this sheet the Group and Genetic susceptibility columns. Create based on these two columns a 100% Stacked Column chart.
- Insert a new sheet and name it **Descriptive statistics**. Copy here the following variables: Gender, Age (Years), Height (cm), Weight (kg), Cholesterol (mg/dL), HDL (mg/dL), and TG (mg/dL). Sort data according to gender and compute descriptive statistics parameters separately for each gender (male and female).