STATISTICAL ANALYSIS OF MEDICAL DATA: CORRELATION AND REGRESSION ANALYSIS

Learning Objectives:

- Correlation:
  - Pearson Correlation Coefficient (CORREL function, Correlation Matrix, Interpretation of Results)
  - Scatter graph.
  - Determination coefficient
- Regression analysis. Simple linear regression

A study has been conducted on a sample of 30 adult subjects. The following medical parameters were collected from each subject: age (years), weight (kg), systolic blood pressure (SBP (mmHg)), diastolic blood pressure (DBP (mmHg)), glycemia (mg/dL), baseline and follow-up cholesterol (mg/dL). Collected data are in **CorReg.xls** file. Save the Excel file on your partition, in **Lab12** folder.

Requests:

1. Calculate for each subject the value of BMI (body mass index) by applying the following formula:

   $$BMI = \text{weight (kg)}/\text{height (m}^2\text{)}$$

   Display the results of BMI as number without decimals.
2. Under assumption of normal distribution of data, compute using CORREL predefine function the Pearson correlation coefficient (R) between Age and Weight. In the cell next to the result, specify the direction of the association (positive/negative) as well as the degree of association. For the degree of association use the Colton empirical rules [Colton T. Statistics in Medicine. Little Brown and Company, New York, NY 1974]:
   - $R \subset [-0.25 \text{ to } +0.25] \rightarrow$ No relation
   - $R \subset (0.25 \text{ to } +0.50] \cup (-0.25 \text{ to } -0.50] \rightarrow$ weak relation
   - $R \subset (0.50 \text{ to } +0.75] \cup (-0.50 \text{ to } -0.75] \rightarrow$ moderate relation
   - $R \subset (0.75 \text{ to } +1) \cup (-0.75 \text{ to } -1) \rightarrow$ strong relation
3. Under the assumption of normal distribution, compute and interpret the correlation coefficient between BMI and SBP.
4. The following table contains the Pearson correlation coefficient (R) for pairs of variables used in our study. Copy this table in a new sheet named **Correl** and fill in the last two columns (Degree of association & Direction).

Variable 1	Variable 2	R (p-value)	Degree of association	Direction
Age	Weight	-0.1378 (0.2339)	No relation	Negative
Age	BMI	0.0518 (0.3929)	No relation	Positive
Age	SBP	0.7158 (<0.0001)	Moderate relation	positive
Age	DBP	0.4624 (0.0050)		
Age	Cholesterol baseline	0.4050 (0.0132)		
Weight	Height	0.4826 (0.0034)		
Weight	SBP	0.0665 (0.3635)		
Weight	DBP	0.4948 (0.0027)		
Weight	Cholesterol baseline	-0.0526 (0.3912)		
BMI	DBP	0.5077 (0.0021)		
BMI	Glycemia	0.4748 (0.0040)		
BMI	Cholesterol baseline	0.1313 (0.2446)		
Cholesterol baseline	Cholesterol follow-up	0.9007 (<0.001)		

5. Is body mass index (dependent variable) linear dependent by age (independent variable)? To answer this question, represent graphically the relation between age (OX) and body mass index (OY) (Scatter chart). Display on the graphical representation both the R square (R^2) and the regression equation. Beside graphical representation interpret:

a. The plot.
b. The R^2.
c. The coefficients of the regression equation.

6. Is systolic blood pressure (dependent variable) linear dependent by age (independent variable)? To answer this question, represent graphically the relation between age (OX) and body mass index (OY) (Scatter chart). Display on the graphical representation both the R square (R^2) and the regression equation. Beside graphical representation interpret:
    a. The plot.
    b. The R^2.
    c. The coefficients of the regression equation.