

## MICROSOFT EXCEL: CORRELATIONS & REGRESSIONS

### Learning Objectives:

- Correlation:
  - Pearson Correlation Coefficient (CORREL function, Correlation Matrix, Interpretation of Results)
  - Scatterplot
- Regression analysis. Interpreting the results:
  - Simple linear regression
  - Multiple linear regression

### Problem

Maintaining glycemic control is essential to avoid long-term consequences of diabetes in children (e.g. neurological, vision, and kidney problems). What is less clear is whether maintaining diabetes control affects growth and development of childhood. For this purpose a group of adolescent boys aged 9-15 were investigated periodically (3 times). The degree of diabetes control was assessed by measuring the Hemoglobin A1c (abbreviated as HbA1c = a form of hemoglobin used primarily to identify the average plasma glucose concentration over prolonged periods of time (reference range: 4%–7%). The value of HbA1c increased, both in diabetic patients and in patients with renal failure. The higher the HbA1c is, the poorer the diabetic control. The age, height, and weight of each child were determined at each visit. The data were collected and stored in **Diabetes.xlsx** file.

The main question: Is any overall relationship between glycemic control and growth?.

### Requests

1. Download the **Diabetes.xlsx** file and save it in **Lab08** folder.
2. Display in **eAG columns** the estimated average glucose level, expressed in mg/dl using the following formula [Nathan DM, Kuenen J, Borg R, Zheng H, Schoenfeld D, Heine RJ. Translating the A1C assay into estimated average glucose values. Diabetes Care 2008;31(8):1473-8]:

$$eAG(\text{mg/dl}) = 28.7 \times \text{HbA1c} - 46.7$$

3. Display in **BMI columns** the body mass index, expressed in kg/m<sup>2</sup> using the following formula:

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

4. Compute the mean of HbA1c for each patient.

$$m\text{-HbA1c} = ((\text{HbA1c (\%)-1}^{\text{st}} \text{ visit}) + (\text{HbA1c (\%)-2}^{\text{nd}} \text{ visit}) + (\text{HbA1c (\%)-3}^{\text{rd}} \text{ visit}))/3$$

5. Compute the overall median of HbA1c (use predefined MEDIAN function and display the result in cell W80).
6. Display for each patient if the control status was good or not (use IF function) considering the following:

Control status is **yes** if the mean of HbA1c for a patient is below the median of all investigations;  
otherwise is **no**.

7. Calculate the measure of growth for each patient using the formula:

$$\text{Measure of growth} = [(\text{weight at 3}^{\text{rd}} \text{ visit}) - (\text{weight at 1}^{\text{st}} \text{ visit})]/[(\text{age at 3}^{\text{rd}} \text{ visit}) - (\text{age at 1}^{\text{st}} \text{ visit})]$$

8. Create a new sheet named Correlation. Copy the following variables in this sheet (used Paste Special – Values and number formats option): m-HbA1c, GoodControl, Measure of growth.

9. Under the assumption of normal distribution of experimental data, compute correlation coefficient (CORREL predefined function) between m-HbA1c and Measure of growth:
  - a. For all patients
  - b. For patients with GoodControl
  - c. For patients without GoodControlInterpret the results.
  
10. Create a new sheet named Simple Regression. Copy in this file the data corresponding to the 3<sup>rd</sup> visit. We assume that all experimental data follow a normal distribution. Based on these data please answer to the following questions:
  - a. Is height (dependent variable) linear dependent by age (independent variable)? Interpret the following parameters (in the same sheet beside the value of the parameter): Multiple R; R Square; F.
  - b. Represent graphically the relation between height (OY) and age (OX) (use SCATTER chart). Include on the graphical representation both the R square and the regression equation.
  
11. Insert a new sheet named Multiple Regression. Copy in this file the data corresponding to the 3<sup>rd</sup> visit. We assume that all experimental data follow a normal distribution. Based on these data please answer to the following questions:
  - a. Is there any linear relationship between HbA1c as dependent variable and age, height and weight as independent variables?
  - b. Interpret the following parameters (in the same sheet beside the value of the parameter): Multiple R; R Square; F.
  - c. Is any of these variables linearly related to HbA1c?