TESTING HYPOTHESIS, CORRELATION AND REGRESSION - HINTS

This document contains general hints! You should adapt from general to particular: to your case where you work with.

A. *How to state null and alternative hypotheses for comparisons between parameters*

Twstwed parameters can be:

- One sample mean and the mean of the population,
- Two sample means,
- 3 or >3 sample means,
- One proportion from the sample and the proportion on the population
- Two or more than two proportions (percentages) from the samples
- Two distributions
- Two variances etc.

Null Hypothesis (H0):
"The differences between the tested parameters is not significant diferent than 0" (is ~ null)
Or
"There is no significant difference between the tested parameters".
Or
"The parameters are independent" (for proportions).

Alternative Hypothesis (H1) (The negation of the null hypothesis):
"The differences between the tested parameters is significant diferent than 0" (is not null)
Or
"There is significant difference between the tested parameters".
Or
"The parameters are dependent" (for proportions).

B. *Choosing the right statistical test:*

- One sample mean and the mean of the population
  - Z Test,
- Two sample means:
  - t test for equal variances (when distribution is normal, independent samples and variances are equal),
  - t test for unequal variances (when distribution is normal, independent samples and variances are not equal),

    - Mann-Whitney test (when distribution is not normal, independent samples)
    - T test for paired samples (when distribution is normal, dependent samples)
    - Wilcoxon Rank Sum test (when distribution is not normal, dependent samples)
- 3 or >3 sample means:
    - Anova test (when variances are equal)
    - Kruskal-Wallis test (when variances are not equal)
- One proportion from the sample and the proportion on the population
    - Z Test for proportions
- Two or more than two proportions (percentages) from the samples
    - Chi-square test (for independent samples, with teoretical frequency ≥5)
    - McNemar test (for dependent samples)
    - Fisher exact test (for independent samples, with teoretical frequency <5
- Two distributions
    - Kolmogorov Smirnov test
    - Lilefors test
    - Shapiro-Wilk test
- Two variances
    - Bartlet test
    - Fisher test
    - Levene test
- etc.

C. *Interpreting the result of the test: p-value.*

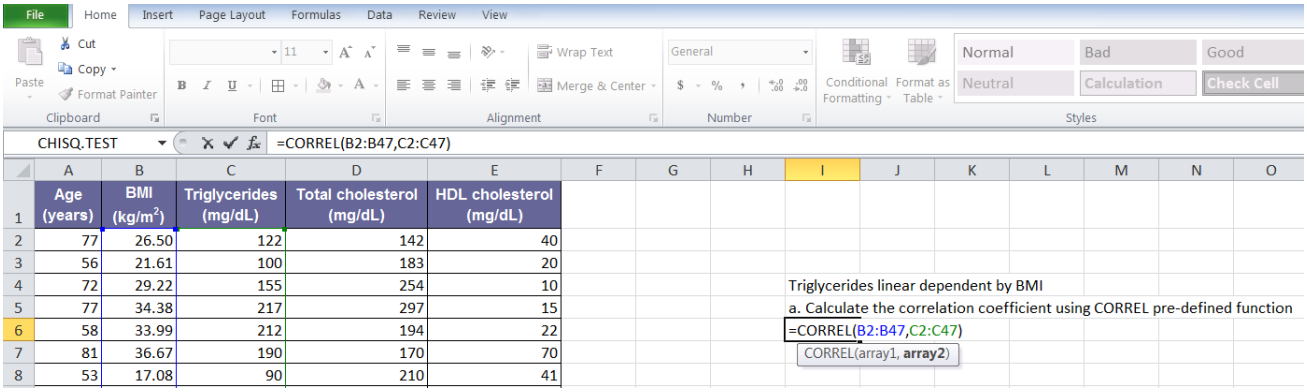
Definition: p-value is the probability that the null hypothesis is true.

We choose the significance level: $\alpha=0.05$.
If $p>0.05$ than we fail to reject null hypothesis: There is no significant difference between the tested parameters.

If $p\leq0.05$ than we reject null hypothesis and we accept alternative hypothesis: There is significant difference between the tested parameters.

D. *Calculate the correlation coefficient using CORREL pre-defined function*

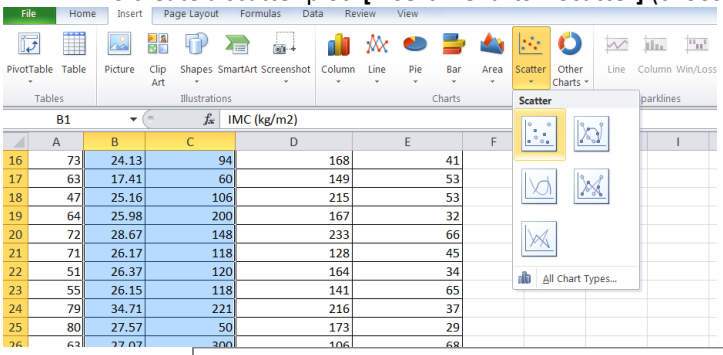
Select the cell where the results you want to be:

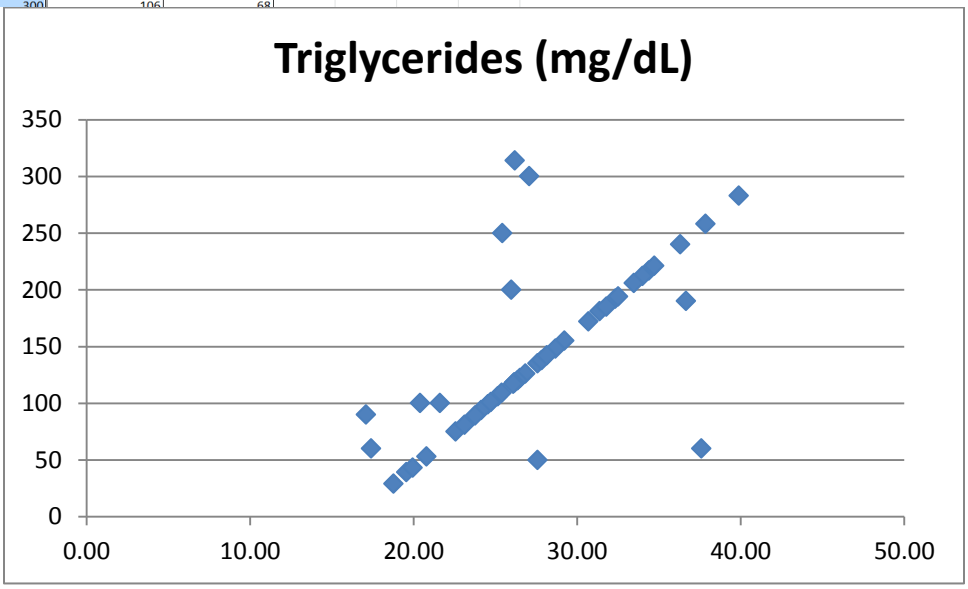


E. *Represent graphically the relation between IMC (OX) and Triglycerides (OY) (Scatter chart)*

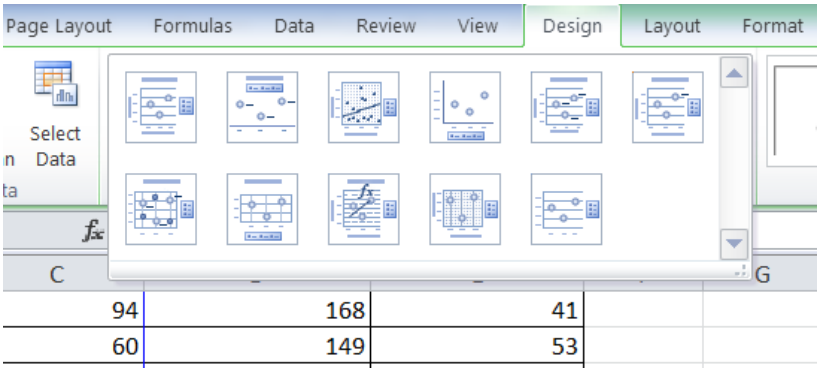
Select the columns you want to represent: First column which will be on OX, second the column that will be on the OY axes

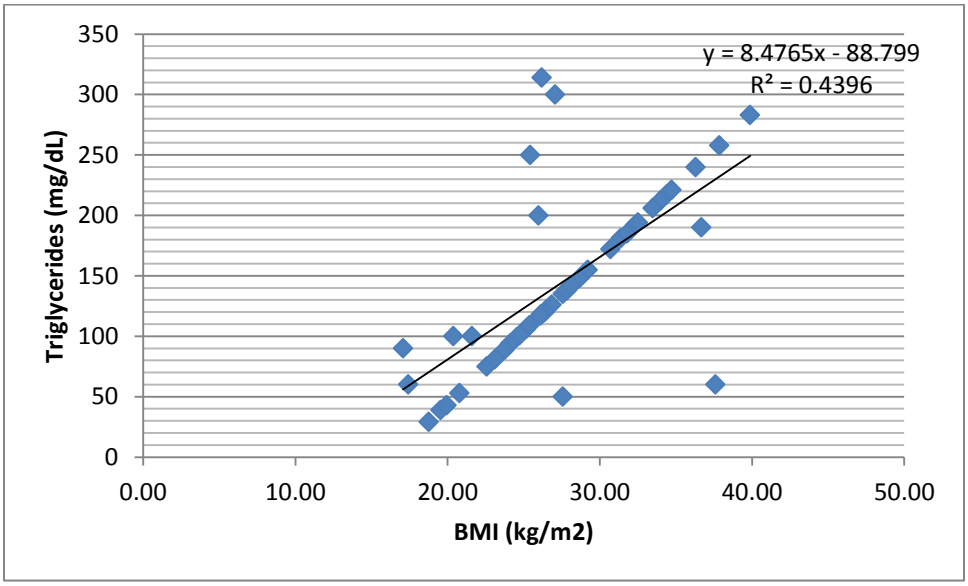
To create a scatter plot: **[Insert – Charts – Scatter]** (choose the first type of Scatter)





Work on the graphical representation to look like the one in the image bellow (choose a template):

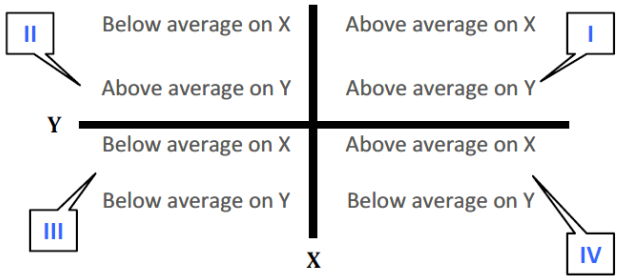




Interpretations of determination coefficient (R^2) and scatter and coefficients of regression:

- R^2 answer to the following question: how much of the percentage of variation in Y can be explained by the linear relationship between Y and X? For Request 5, $R^2 = 0.4396 \rightarrow$ 43% from variation in Tryglycerides could be explained by the linear relation between BMI and Tryglycerides.

- Interpretation of Scatter: split the scatter plot in 4 cadres using the mean of X and the mean of Y:



If a linear relationship exists between X and Y, the markers of the plot will be in cadres II and IV (negative direction – descendant trend) or I and III (positive direction – ascendant trend). If the markers are uniformly dispersed in all four cadres, the scatter indicates a null relationship between X and Y.