TECNOLÓGICO NACIONAL DE MÉXICO
INSTITUTO TECNOLÓGICO DE TIJUANA

CARRERA:
INGENIERÍA EN SISTEMAS COMPUTACIONALES

PERIODO:
AGOSTO - DICIEMBRE 2020

ASIGNATURA:
BDD-1704TI9A DATOS MASIVOS

TRABAJO:
CORRELACION DE PEARSON

NOMBRE DEL ALUMNO
16210504 ALCANTAR BALCON ADRIAN GIOVANNY

ASESOR:
JOSE CHRISTIAN ROMERO HERNANDEZ

LUGAR Y FECHA:
TIJUANA, BAJA CALIFORNIA, 16 DE OCTUBRE DEL 2020

Pearson's correlation

Pearson's correlation coefficient is a test that measures the statistical relationship between two continuous variables. If the association between the elements is not linear, then the coefficient is not adequately represented.

The correlation coefficient can take a range of values from +1 to -1. A value of 0 indicates that there is no association between the two variables. A value greater than 0 indicates a positive association. That is, as the value of one variable increases, so does the value of the other. A value less than 0 indicates a negative association; that is, as the value of one variable increases, the value of the other decreases.

Pearson's correlation coefficient is used to study the relationship (or correlation) between two quantitative random variables (minimum interval scale); for example, the relationship between weight and height.
It is a measure that gives us information about the intensity and direction of the relationship. In other words, it is an index that measures the degree of covariation between different linearly related variables.

$$r = \frac{N\Sigma xy - \Sigma x \Sigma y}{\sqrt{\{N\Sigma x^2 - (\Sigma x)^2\} \times \{N\Sigma y^2 - (\Sigma y)^2\}}}$$

To carry out the Pearson correlation it is necessary to fulfill the following:
- The measurement scale must be an interval or ratio scale.
- Variables must be roughly distributed.
- The association must be linear.
- There should be no outliers in the data.

Advantages and disadvantages of the Pearson correlation coefficient
Among the main advantages of the Karl Pearson correlation coefficient are:
- The value is independent of whatever unit is used to measure the variables.
- If the sample is large, the accuracy of the estimate is more likely. Some of the disadvantages of the correlation coefficient are:
- The two variables need to be measured at a continuous quantitative level.
- The distribution of the variables must be similar to the normal curve.

Reference sources:

[1] https://www.questionpro.com/blog/es/coeficiente-de-correlacion-de-pearson/

[2] https://psicologiaymente.com/miscelanea/coeficiente-correlacion-pearson

[3] https://bookdown.org/dietrichson/metodos-cuantitativos/coeficientes-de-correlacion.html