TECNOLÓGICO NACIONAL DE MÉXICO

INSTITUTO TECNOLÓGICO DE TIJUANA

SUBDIRECCIÓN ACADÉMICA
DEPARTAMENTO DE SISTEMAS COMPUTACIONALES
Semestre: AGOSTO - DICIEMBRE 2020

CARRERA

Ingeniería en Sistemas Computacionales

MATERIA
Datos Masivos

TÍTULO
¿Qué es la correlación de Pearson?

UNIDAD
1

ALUMNO(S) Y NO. DE CONTROL
Encarnacion Ocampo Gustavo 16211993

DOCENTE
JOSE CHRISTIAN ROMERO HERNANDEZ

What is Pearson's correlation?

Pearson's Correlation Coefficient is a measure of the correspondence or linear relationship between two random quantitative variables. In simpler words, it can be defined as an index used to measure the degree of relationship between two variables, both quantitative.

Having two variables, the correlation facilitates estimates of the value of one of them, with knowledge of the value of the other variable.

This coefficient is a measure that indicates the relative situation of the events with respect to the two variables, that is, it represents the numerical expression that indicates the degree of correspondence or relationship that exists between the 2 variables. These numbers vary between limits of +1 and -1.

Covariation is the degree of agreement of the relative positions of the two-variable data. Consequently, the Pearson correlation coefficient operates with standardized scores (which measure relative positions) and is defined:

$$r_{xy} = \frac{\sum z_x z_y}{N}$$

Example.

X	Y	z_x	z_y
2	4	-1.41	-1.41
3	5	-0.71	-0.71
4	6	-0.00	-0.00
5	7	0.71	0.71
6	8	1.41	1.41

Notice that the standardized data (expressed as z scores) in the two columns on the right have the same values in both variables, since the relative positions are the same in variables X and Y.

If we obtain the products of the typified values for each case, the result is:

X	Y	z_x	z_y	$z_x z_y$
2	4	-1.41	-1.41	2
3	5	-0.71	-0.71	0.5
4	6	0.00	0.00	0
5	7	0.71	0.71	0.5
6	8	1.41	1.41	2
				$\Sigma = 5$

The quotient of dividing the sum of products (5) by N (it must be taken into account that N is the number of cases, not the number of data) is equal to 1:

$$= \frac{\sum z_x z_y}{N} = \frac{5}{5} = 1$$

Sources.

https://www.webyempresas.com/coeficiente-de-correlacion-de-pearson/
https://www.uv.es/webgid/Descriptiva/31_coeficiente_de_pearson.html