



# Tecnológico nacional de México

## Instituto tecnológico de Tijuana

Subdirección Académica  
Departamento de sistemas y computación

SEMESTRE:  
ENERO - JULIO 2020

CARRERA:  
Ing. Informática

MATERIA:  
DATOS MASIVOS

Nombre del trabajo:  
Correlación de pearson

UNIDAD A EVALUAR:  
Unidad 1

NOMBRE Y NÚMERO DE CONTROL DEL ALUMNO:  
Castro Arenas Flavio Daniel #16210783

NOMBRE DEL MAESTRO (A):  
Jose Cristian Romero Hernandez

El coeficiente de correlación de Pearson es una prueba que mide la relación estadística entre dos variables continuas. Si la asociación entre los elementos no es lineal, entonces el coeficiente no se encuentra representado adecuadamente.

El coeficiente de correlación puede tomar un rango de valores de +1 a -1. Un valor de 0 indica que no hay asociación entre las dos variables. Un valor mayor que 0 indica una asociación positiva. Es decir, a medida que aumenta el valor de una variable, también lo hace el valor de la otra. Un valor menor que 0 indica una asociación negativa; es decir, a medida que aumenta el valor de una variable, el valor de la otra disminuye.

**Para llevar a cabo la correlación de Pearson es necesario cumplir lo siguiente:**

- La escala de medida debe ser una escala de intervalo o relación.
- Las variables deben estar distribuida de forma aproximada.
- La asociación debe ser lineal.
- No debe haber valores atípicos en los datos.
- Cómo se calcula el coeficiente de correlación de Pearson
- La fórmula del coeficiente de correlación de Pearson es la siguiente:

**Donde:**

“x” es igual a la variable número uno, “y” pertenece a la variable número dos, “zx” es la desviación estándar de la variable uno, “zy” es la desviación estándar de la variable dos y “N” es es número de datos.

Hemos especificado los términos "valores absolutos" ya que en realidad si se contempla el signo el coeficiente de correlación de Pearson oscila entre  $-1$  y  $+1$ . No obstante ha de indicarse que la magnitud de la relación viene especificada por el valor numérico del coeficiente, reflejando el signo la dirección de tal valor. En este sentido, tan fuerte es una relación de  $+1$  como de  $-1$ . En el primer caso la relación es perfecta positiva y en el segundo perfecta negativa. Pasamos a continuación a desarrollar algo más estos conceptos.

En los fenómenos humanos, fuertemente cargados de componentes aleatorios, no suelen ser posible establecer relaciones funcionales exactas. Dado un cierto valor en la variable X no encontraremos uno y solo un único valor en la variable Y. Por ejemplo, si relacionamos horas de estudio con el rendimiento académico obtendremos mayor rendimiento a mayor inteligencia, pero será prácticamente imposible saber con exactitud la puntuación que obtendrá un sujeto para unas horas determinadas. Dado un cierto número de personas con un mismo número de horas, por ejemplo 10, no

todos obtendrán exactamente la misma puntuación en rendimiento. Unos obtendrán más o menos en función de otras variables, tales como motivación o personalidad.

