

**TECNOLÓGICO NACIONAL DE MÉXICO**

**INSTITUTO TECNOLÓGICO DE TIJUANA**

**SUBDIRECCIÓN ACADÉMICA**

**DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN**

**SEMESTRE SEP-ENE 2020-2021**

**INGENIERÍA EN SISTEMAS COMPUTACIONALES**

|  |  |
| --- | --- |
| **DATOS MASIVOS** | **BDD-1704TI9A** |

# **Tarea 1- Correlación de Pearson**

|  |  |
| --- | --- |
| **Angeles Valadez Jonathan** | **15211883** |

**Profesor:**

**JOSE CHRISTIAN ROMERO HERNANDEZ**

**Introducción**

La correlación es una medida de la relación (covariación) lineal entre dos variables cuantitativas continuas (x, y). La manera más sencilla de saber si dos variables están correlacionadas es determinar si co-varían (varían conjuntamente). Es importante hacer notar que esta covariación no implica necesariamente causalidad, la correlación puede ser fortuita, como en el caso clásico de la correlación entre entre el número de venta de helados e incendios, debido al efecto de una tercera variable, la temperatura ambiental.

* La correlación cuantifica como de relacionadas están dos variables, mientras que la regresión lineal consiste en generar una ecuación (modelo) que, basándose en la relación existente entre ambas variables, permita predecir el valor de una a partir de la otra.
* El cálculo de la correlación entre dos variables es independiente del orden o asignación de cada variable a X e Y, mide únicamente la relación entre ambas sin considerar dependencias. En el caso de la regresión lineal, el modelo varía según qué variable se considere dependiente de la otra (lo cual no implica causa-efecto).
* A nivel experimental, la correlación se suele emplear cuando ninguna de las variables se ha controlado, simplemente se han medido ambas y se desea saber si están relacionadas. En el caso de estudios de regresión lineal, es más común que una de las variables se controle (tiempo, concentración de reactivo, temperatura…) y se mida la otra.
* Por norma general, los estudios de correlación lineal preceden a la generación de modelos de regresión lineal. Primero se analiza si ambas variables están correlacionadas y, en caso de estarlo, se procede a generar el modelo de regresión.

**Desarrollo**

La correlación es en esencia una medida normalizada de asociación o covariación lineal entre dos variables. Esta medida o índice de correlación r puede variar entre -1 y +1, ambos extremos indicando correlaciones perfectas, negativa y positiva respectivamente. Un valor de r = 0 indica que no existe relación lineal entre las dos variables. Una correlación positiva indica que ambas variables varían en el mismo sentido. Una correlación negativa significa que ambas variables varían en sentidos opuestos. Lo interesante del índice de correlación es que r es en sí mismo una medida del tamaño del efecto, que suele interpretarse de la siguiente manera:

* correlación **despreciable**: r < |0.1|
* correlación **baja**: |0.1| < r <= |0.3|
* correlación **mediana** : |0.3| < r<= |0.5|
* correlación **fuerte o alta**: r > |0.5|

La correlación se define en términos de la varianza ()de las variables x e y, así como de la covarianza cov de x,y. Es por tanto una medida de la variación conjunta de ambas variables (cov(x,y)).

### **varianza** ()

La varianza de una muestra representa el promedio de la desviación de los datos con respecto a la media



### **covarianza** cov(x,y)

La covarianza entre dos variables x e y es una medida de la relación “promedio” éstas. Es la desviación promedio del producto cruzado entre ellas:



**Conclusión**

A diferencia de la covarianza, la correlación de Pearson es independiente de la escala de medida de las variables, se utiliza para estudiar la relación (o correlación) entre dos variables aleatorias cuantitativas (escala mínima de intervalo); por ejemplo, la relación entre el peso y la altura. Y en nuestro caso de la práctica, la relación entre las columnas que tenemos en el archivo csv, High, Low, Close.

**Bibliografía**  
**[1]En Línea:** Tema 8 - Correlación: teoría y práctica, Pablo Vinuesa, CCG-UNAM., 14 de Octubre, 2016,

<https://www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema8_correlacion.html>