

Instituto Tecnológico de Tijuana
Ingeniería en Sistemas Computacionales



Investigación II:
Correlación de Pearson

Materia: Datos Masivos

Unidad: Unidad I

Facilitador:
José Christian Romero Sánchez

Alumno: Hernández Negrete Juan Carlos

Fecha:
Tijuana Baja California a 16 de Octubre del 2020.

CORRELACION DE PEARSON

El Coeficiente de Correlación de Pearson es una medida de la correspondencia o relación lineal entre dos variables cuantitativas aleatorias. En palabras más simples se puede definir como un índice utilizado para medir el grado de relación que tienen dos variables, ambas cuantitativas.

Teniendo dos variables, la correlación facilita que se hagan estimaciones del valor de una de ellas, con conocimiento del valor de la otra variable.

Este coeficiente es una medida que indica la situación relativa de los sucesos respecto a las dos variables, es decir, representa la expresión numérica que indica el grado de correspondencia o relación que existe entre las 2 variables. Estos números varían entre límites de +1 y -1.

¿Cómo se calcula?

Para contar con una guía que permita:

- Establecer la variación contigua de las dos variables
- Comparar los distintos casos entre sí

Para ello se hace uso del coeficiente de correlación de Pearson, definido como la covarianza que se da entre dos variables tipificadas y se calcula con la siguiente expresión:

$$r_{xy} = \frac{\sum Z_x Z_y}{N}$$

¿Cómo se interpreta el coeficiente de correlación de Pearson?

Su dimensión indica el nivel de asociación entre las variables.

- Cuando es menor a cero ($r < 0$) Se dice que hay correlación negativa: Las variables se correlacionan en un sentido inverso.

A valores altos en una de las variables, le suelen corresponder valores bajos en la otra variable y viceversa. Cuánto el valor esté más próximo a -1 dicho coeficiente de correlación más evidente será la covariación extrema.

Datos Masivos

Si $r = -1$ se habla de correlación negativa perfecta, la cual supone una determinación absoluta entre ambas variables, en sentido directo coexiste una relación lineal perfecta de pendiente negativa.

- Cuando es mayor a cero ($r > 0$) Se dice que hay correlación positiva: Ambas variables se correlacionan en un sentido directo.

A valores altos en una de las variables, le corresponden valores altos en la otra variable e igualmente en una situación inversa sucede con los valores bajos. Cuánto más próximo a $+1$ se encuentre el coeficiente de correlación más evidente será la covariación.

- Si $r = 1$ Se habla de correlación positiva perfecta, la cual supone una determinación absoluta entre las variables, en sentido directo coexiste una relación lineal perfecta de pendiente positiva).

Cuando es igual a cero ($r = 0$) Se dice que las variables están incorrectamente relacionadas, no puede es posible establecer algún sentido de covariación.

No existe relación lineal, pero esto no implica necesariamente que las variables sean independientes, pudiendo existir relaciones no lineales entre las variables.

Cuando las dos variables son independientes se dice que no están correlacionadas, aunque el resultado de reciprocidad no es necesariamente cierto.

Examinar la relación lineal entre las variables (Pearson)

Se utiliza el coeficiente de correlación de Pearson para examinar la fuerza y la dirección de la relación lineal entre dos variables continuas.

Fuerza

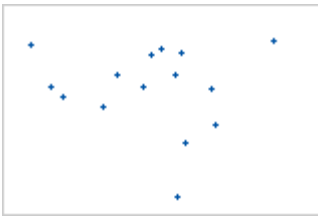
- El valor del coeficiente de correlación puede variar de -1 a $+1$. Mientras mayor sea el valor absoluto del coeficiente, más fuerte será la relación entre las variables.
- Para la correlación de Pearson, un valor absoluto de 1 indica una relación lineal perfecta. Una correlación cercana a 0 indica que no existe relación lineal entre las variables.

Dirección

- El signo del coeficiente indica la dirección de la relación. Si ambas variables tienden a aumentar o disminuir a la vez, el coeficiente es positivo y la línea que representa la correlación forma una pendiente hacia arriba. Si una variable tiende a incrementarse mientras la otra disminuye, el coeficiente es negativo y la línea que representa la correlación forma una pendiente hacia abajo.

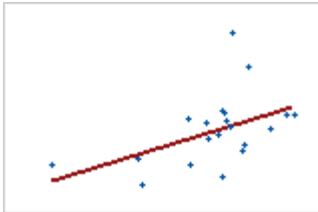
Datos Masivos

Las siguientes gráficas muestran datos con valores específicos del coeficiente de correlación para ilustrar diferentes patrones en la fuerza y la dirección de las relaciones entre las variables.



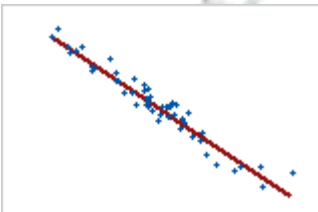
Ninguna relación: Pearson $r = 0$

Los puntos se ubican de forma aleatoria en la gráfica, lo que significa que no existe relación lineal entre las variables.



Relación positiva moderada: Pearson $r = 0.476$

Algunos puntos están cerca de la línea, pero otros puntos están lejos de ella, lo que indica que solo existe una relación lineal moderada entre las variables.



Relación positiva grande: Pearson $r = 0.93$

Los puntos se ubican cerca de la línea, lo que indica que existe una relación lineal fuerte entre las variables. La relación es positiva porque a medida que una variable aumenta, la otra variable también aumenta.

Determinar si el coeficiente de correlación es significativo

Para determinar si la correlación entre las variables es significativa, compare el valor p con su nivel de significancia. Por lo general, un nivel de significancia (denotado como α o alfa) de 0.05 funciona adecuadamente. Un α de 0.05 indica que el riesgo de concluir que existe una correlación, cuando en realidad no es así, es 5%. El valor p indica si el coeficiente de correlación es significativamente diferente de 0. (Un coeficiente de 0 indica que no existe una relación lineal).

Valor $p \leq \alpha$: La correlación es estadísticamente significativa

- Si el valor p es menor que o igual al nivel de significancia, entonces usted puede concluir que la correlación es diferente de 0.

Valor $p > \alpha$: La correlación no es estadísticamente significativa

- Si el valor p es mayor que el nivel de significancia, entonces usted no puede concluir que la correlación es diferente de 0.

Ventajas y desventajas del coeficiente de correlación de Pearson

Entre las principales ventajas del coeficiente de correlación de Pearson se encuentran:

- El valor es independiente de cualquier unidad que se utiliza para medir las variables.
- Si la muestra es grande, es más probable la exactitud de la estimación.

Alguna de las desventajas del coeficiente de correlación es:

- Es necesario las dos variables sean medidas a un nivel cuantitativo continuo.
- La distribución de las variables debe ser semejantes a la curva normal.

Referencias

- Matias Riquelme. (11 de Mayo del 2019). ¿Qué es y cómo se interpreta el coeficiente de correlación de Pearson?. 15 de Octubre del 2020, de Web y Empresas Sitio web: <https://www.webyempresas.com/coeficiente-de-correlacion-de-pearson/>
- Minitab. (2019). Interpretar los resultados clave para Correlación. 2020, de Minitab Sitio web: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/statistics/basic-statistics/how-to/correlation/interpret-the-results/key-results/>
- QuestionPro. (2020). ¿Qué es el coeficiente de correlación de Pearson?. 2020, de QuestionPro Sitio web: <https://www.questionpro.com/blog/es/coeficiente-de-correlacion-de-pearson/>