

Coeficiente de Correlación

El *coeficiente de correlación* es una herramienta de cálculo que sirve para asegurarse de que la relación entre dos variables que se muestran en un diagrama no sea resultado de una construcción errónea del diagrama de dispersión (por ejemplo, el tamaño y las escalas). Además, nos sirve para cuantificar la magnitud de la correlación lineal en términos numéricos.

Para un conjunto de n valores del tipo (x_i, y_i) , obtenidos a partir de n unidades o productos, este coeficiente se calcula con:

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$$

Donde:

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{\left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n}$$

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n}$$

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{\left(\sum_{i=1}^n y_i \right)^2}{n}$$

Coeficiente de Correlación

Para obtener el cálculo de r es recomendable apoyarse en un programa computacional. Por ejemplo, en Excel se utiliza la siguiente función:

COEF.DE.CORREL(matriz1;matriz2)

En la que *matriz1* es el rango de celdas donde están los valores de X , y *matriz2* es el correspondiente rango de celdas donde se encuentran los valores de Y . Los valores que toma el coeficiente de correlación, r , están entre -1 y 1 , incluyendo $(-1 \leq r \leq 1)$. Los valores de r cercanos o iguales a cero implican poca o nula relación lineal entre X y Y . En contraste, los valores de r cercanos a 1 indican una relación lineal muy fuerte, y los valores de r próximos a -1 muestran una fuerte correlación negativa. Los valores de r cercanos a -0.85 o 0.85 indican una correlación fuerte; mientras que los valores de r cercanos a -0.50 o 0.50 se refieren a una correlación de moderada a débil. Por último, los valores de r iguales o menores que -0.30 o 0.30 indican una correlación lineal prácticamente inexistente. En los diagramas de dispersión de la figura 6.10 se tienen los siguientes valores para el coeficiente de correlación: en *a*) $r=0.97$; en *b*) $r=-0.99$; en *c*) $r=0.06$, y en *d*) $r=-0.33$. No olvidemos que cuando r tiene valores cercanos a cero, no existe una relación *lineal* y de ninguna manera dice que no hay ningún tipo de relación; esto se puede observar en el diagrama de dispersión de la figura 6.10*d*, donde, a pesar de que existe una clara e importante relación entre X y Y , el coeficiente de correlación es pequeño ($r=-0.30$) debido a que los puntos no se agrupan en torno a una línea recta. De esta forma, la función del coeficiente de correlación es comprobar la magnitud de la relación lineal que muestra el diagrama de dispersión.

REFERENCIA:

Gutiérrez, H. y De la Vara, R. (2009). Control estadístico de calidad y seis sigma. Recuperado de:
<https://www.uv.mx/personal/ermeneses/files/2018/05/6-control-estadistico-de-la-calidad-y-seis-sigma-gutierrez-2da.pdf>