



*Nombre del Alumno: Juan Antonio Espinosa Hernández*

*Nombre del tema: super nota*

*Parcial: 3*

*Nombre de la Materia: estadística descriptiva*

*Nombre del profesor: **Andrés Alejandro reyes molina***

*Nombre de la Licenciatura: administración y estrategias de negocios*

*Cuatrimestre: **3***

# Relaciones entre variables

**Introducción:** En el análisis de los estudios clínico-epidemiológicos surge muy frecuentemente la necesidad de determinar la relación entre dos variables cuantitativas en un grupo de sujetos. Los objetivos de dicho análisis suelen ser

- a. Determinar si las dos variables están correlacionadas, es decir si los valores de una variable tienden a ser más altos o más bajos para valores más altos o más bajos de la otra variable.
- b. Poder predecir el valor de una variable dado un valor determinado de la otra variable.
- c. Valorar el nivel de concordancia entre los valores de las dos variables

**Correlación:** En este artículo trataremos de valorar la asociación entre dos variables cuantitativas estudiando el método conocido como correlación. Dicho cálculo es el primer paso para determinar la relación entre las variables. La predicción de una variable dada un valor determinado de la otra precisa de la regresión lineal.

a. El valor del coeficiente de correlación es independiente de cualquier unidad usada para medir las variables.

b. El valor del coeficiente de correlación se altera de forma importante ante la presencia de un valor extremo, como sucede con la desviación típica.

El coeficiente de correlación mide solo la relación con una línea recta.

d. El coeficiente de correlación no se debe extrapolar más allá del rango de valores observado de las variables a estudio ya que la relación existente entre X e Y puede cambiar fuera de dicho rango.

e. La correlación no implica causalidad. La causalidad es un juicio de valor que requiere más información que un simple valor cuantitativo de un coeficiente de correlación.

**El coeficiente de correlación de Pearson ( $r$ )** puede calcularse en cualquier grupo de datos, sin embargo la validez del test de hipótesis sobre la correlación entre las variables requiere los sentidos estrictos

a) que las dos variables procedan de una muestra aleatoria de individuos.

b) que al menos una de las variables tenga una distribución normal en la población de la cual la muestra procede.

El cálculo del coeficiente de correlación ( $r$ ) entre peso y talla de 20 niños varones se muestra en la tabla 1. La covarianza, que en este ejemplo es el producto de peso (kg) por talla (cm), para que no tenga dimensión y sea un coeficiente, se divide por la desviación típica de X (talla) y por la desviación típica de Y (peso) con lo que obtenemos el coeficiente de correlación de Pearson que en este caso es de 0.885 e indica una importante correlación entre las dos variables.

## Coeficiente de correlación de Pearson

### Covarianza:

$$\text{Covarianza} = \frac{\sum(\bar{X} - X) * (\bar{Y} - Y)}{n - 1} = \frac{290.8}{19} =$$

$$r = \frac{\text{covarianza}}{S_x * S_y} = \frac{15.30}{8.087 * 2.137} = 0.885$$

$$S_x = \text{Desviación típica } x = 8.087$$

$$S_y = \text{Desviación típica } y = 2.137$$

Y	X	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X}) * (Y - \bar{Y})$
Peso (Kg)	Talla (cm)			
9	72	5.65	1.4	7.91
10	76	9.65	2.4	23.16
6	59	-7.35	-1.6	11.76
8	68	1.65	0.4	0.66
10	60	-6.35	2.4	-15.24
5	58	-8.35	-2.6	21.71
8	70	3.65	0.4	1.46
7	65	-1.35	-0.6	0.81
4	54	-12.35	-3.6	44.46
11	83	16.65	3.4	56.61
7	64	-2.35	-0.6	1.41
7	66	-0.35	-0.6	0.21
6	61	-5.35	-1.6	8.56
8	66	-0.35	0.4	-0.14
5	57	-9.35	-2.6	24.31
11	81	14.65	3.4	49.81
5	59	-7.35	-2.6	19.11
9	71	4.65	1.4	6.51
6	62	-4.35	-1.6	6.96
10	75	8.65	2.4	20.76

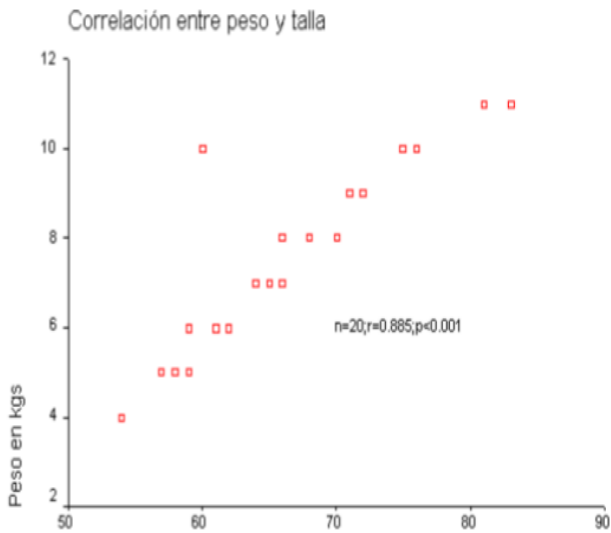
$\Sigma 290.8$

$$X(\text{Media de } \bar{X} = 66.35)$$

$$Y(\text{Media de } \bar{Y} = 7.6)$$

**Test de hipótesis de r** Tras realizar el cálculo del coeficiente de correlación de Pearson (r) debemos determinar si dicho coeficiente es estadísticamente diferente de cero. Para dicho calculo se aplica un test basado en la distribución de la t de student

$$\text{Error estandard de } r = \sqrt{\frac{1 - r^2}{n - 2}}$$



Si el valor del r calculado (en el ejemplo previo  $r = 0.885$ ) supera al valor del error estándar multiplicado por la t de Student con  $n-2$  grados de libertad, diremos que el coeficiente de correlación es significativo

El nivel de significación viene dado por la decisión que adoptemos al buscar el valor en la tabla de la t de Student.

En el ejemplo previo con 20 niños, los grados de libertad son 18 y el valor de la tabla de la t de student para una seguridad del 95% es de 2.10 y para un 99% de seguridad el valor es 2.88.

Como quiera que  $r = 0.885 > a \cdot 0.109 = 2.30$  podemos asegurar que el coeficiente de

correlación es significativo ( $p < 0.05$ ). Si aplicamos el valor obtenido en la tabla de la t de Student

para una seguridad del 99% ( $t = 2.88$ ) observamos que como  $r = 0.885$  sigue siendo  $> 2.88 \cdot$

$0.109 = 0.313$  podemos a su vez asegurar que el coeficiente es significativo ( $p < 0.001$ ). Este

proceso de razonamiento es válido tanto para muestras pequeñas como para muestras

grandes. En esta última situación podemos comprobar en la tabla de la t de student que para

una seguridad del 95% el valor es 1.96 y para una seguridad del 99% el valor es 2.58

Tras calcular la confianza los intervalos de z debemos proceder a hacer el cálculo inverso para obtener los intervalos de confianza de coeficiente de correlación r que era lo que buscábamos en un principio antes de la transformación logarítmica

$$\frac{e^{2 \cdot 0.922} - 1}{e^{2 \cdot 0.922} + 1} \quad \alpha \quad \frac{e^{2 \cdot 1.873} - 1}{e^{2 \cdot 1.873} + 1}$$

0.726 a 0.953 son los intervalos de confianza (95%) de r.

Presentación de la correlación

Se debe mostrar siempre que sea posible la gráfica que correlaciona las dos variables de estudio

(gráfica). El valor de r se debe mostrar con dos decimales junto con el valor de la p si el test de

hipótesis se realizó para demostrar que r es estadísticamente diferente de cero. El número de

observaciones debe a su vez estar indicado

**Interpretación de la correlación:** El coeficiente de correlación como previamente se indicó oscila entre  $-1$  y  $+1$  encontrándose en medio el valor  $0$  que indica que no existe asociación lineal entre las dos variables a estudio. Un coeficiente de valor reducido no indica necesariamente que no exista correlación ya que las variables pueden presentar una relación no lineal como puede ser el peso del recién nacido y el tiempo de gestación.

El coeficiente de correlación no debe utilizarse para comparar dos métodos que intentan medir el mismo evento, como por ejemplo dos instrumentos que miden la tensión arterial. El coeficiente de correlación mide el grado de asociación entre dos cantidades, pero no mira el nivel de acuerdo o concordancia.