

## Tema 4: Análisis conjunto de dos variables

psicologia.isipedia.com

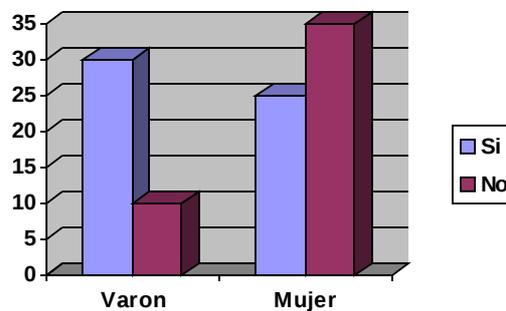
Asociación y /o relación entre dos variables: dos variables están relacionadas entre sí, cuando ciertos valores de una de las variables, se asocian con ciertos valores de la otra.

### 4.1 Asociación entre dos variables cualitativas

Tabla de contingencia:

		Y		
		Sí	No	
X	V	30	10	40
	M	25	35	60
		55	45	100

Diagrama de barras:



Para analizar la relación entre ambas variables se utiliza el estadístico Chi cuadrado  $X^2$ . Se define en función de las frecuencias empíricas ( $n_e$ ) y las frecuencias teóricas ( $n_t$ ).

$$\text{Frecuencia teórica} = n_t = \frac{\text{Total fila} \times \text{total columna}}{n}$$

		Y		
		Sí ( $n_i$ )	No ( $n_i$ )	
X	V	$40 \times 50/100$ = 22	$40 \times 45/100$ = 18	40
	M	$60 \times 55/100$ = 33	$60 \times 45/100$ = 27	60
		55	45	100

Diferencias entre las frecuencias observadas ( $n_e$ ) y las teóricas ( $n_i$ ) (la suma de filas y columnas en esta Tabla tiene que ser siempre igual a cero)

		Y	
		Sí	No
X	V	8	-8
	M	-8	8

Los valores positivos indican una relación entre ambas variables. Los negativos indican una relación negativa entre ambas variables.

Chi cuadrado: uno de los inconvenientes es su difícil interpretación puesto que desconocemos su límite superior. Sólo sabemos que tiene valor cero, cuando no hay relación entre las dos variables.

Para resolver este problema, de acuerdo al valor obtenido de  $X^2$ , se ha definido el índice o Coeficiente de Contingencia, C. Este índice toma los valores  $0 < C < 1$

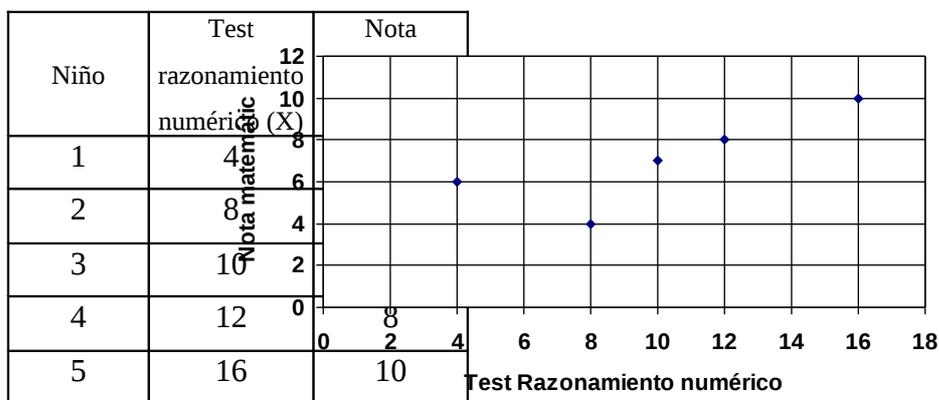
Características del coeficiente de contingencia, C :

- puede asumir valores mayores o iguales a cero y menores que 1.
- El valor 0 lo alcanza cuando  $X^2 = 0 \rightarrow$  indica que las dos variables no tienen relación y, además, las frecuencias empíricas coinciden con las teóricas
- El valor 1 sólo lo consigue si  $n = 0$ , lo que implica que no hay observaciones
- Cuanto mayor es el valor de C, mayor es la relación

- Otro aspecto más complejo de fundamentar es la causalidad en un C. Cuando existe un valor elevado en nuestro coeficiente de contingencia, no se puede afirmar que una de las variables es causa de la otra.
- Se puede estimar, en casos en que la tabla de contingencia tenga igual número de filas que de columnas, un valor máximo que puede alcanzar C.

#### 4.2 Correlación entre dos variables cuantitativas

Diagrama de dispersión:



Covarianza: hace referencia a la variación conjunta de dos variables y nos permite estudiar la posible relación entre X e Y. El signo positivo o negativo indica si la relación lineal entre ambas es directa o inversa:

- Relación lineal directa: a mayores valores de una variable, mayores de la otra
- Relación lineal inversa: a mayores valores de una variable, menores de la otra

Inconvenientes: desconocemos el rango de la covarianza. Para evitar este problema disponemos del Coeficiente de Correlación de Pearson,  $r_{xy}$ .

Propiedades de la correlación de Pearson:

- Toma valores comprendidos entre  $-1$  y  $1$ .
- Vale  $0$  cuando no existe relación lineal entre ambas variables
- $r_{xy} = \pm 1$  si una variable es la transformación lineal de la otra.

Para interpretar los resultados hay que tener en cuenta en primer lugar el valor absoluto, cuanto mayor es nos indica que la relación lineal es más fuerte. En segundo lugar, el signo: si es positivo es relación directa, si es negativo, inversa.

El coeficiente de correlación lineal sólo detecta relaciones lineales entre dos variables. Por tanto, un coeficiente de correlación lineal cercano a cero, indica que no existe relación lineal entre las variables, pero no excluye la posibilidad de que las variables tengan otras relaciones entre sí de carácter no lineal.

Cuando tenemos un valor intermedio como por ejemplo 0, 55 es difícil afirmar si la correlación es alta o baja. Un número grande de sujetos en la muestra pueden tender a bajar el valor de los coeficientes que se obtienen. El coeficiente de correlación evaluado por nosotros será bajo, si los coeficientes de correlación que obtienen otros investigadores en circunstancias similares, son mucho más altos.

Cuando existe un Coeficiente de correlación elevado entre dos variables, no se puede afirmar que una de las variables es causa de la otra. Hay que ser muy cuidadoso en este aspecto.

#### 4.3 Regresión lineal

Cuando existe una relación lineal podemos usar la denominada recta de regresión.

$$Y_i' = a + bX_i$$

Donde b es la pendiente y a es la ordenada de origen