

## TEMA 18 : CONTRASTE DE DOS VARIABLES CUANTITATIVAS

Para estudiar la relación o dependencia entre dos variables cuantitativas se valora estadísticamente ("se contrasta") el coeficiente de correlación. En principio los datos son apareados.

Hay una prueba paramétrica, que contrasta el coeficiente de correlación de Pearson,  $r$ , y otra no paramétrica, que contrasta el coeficiente de correlación de Spearman,  $r_s$ .

### 1) PRUEBA PARAMETRICA : Contraste de r

$$Z = \frac{r\sqrt{(N-2)}}{\sqrt{1-r^2}} \quad (\text{Fórmula n}^\circ 14)$$

Condiciones de aplicación: Si la muestra es pequeña, igualdad de varianzas de x e y

$$V = \frac{S^2 \text{ mayor}}{S^2 \text{ menor}} < F(N-1, N-1, 0.05)$$

Valoración : a) si  $N > 30$ , por c de la D.N.  
b) si  $N < 30$ , por  $t(N-2, \alpha)$

Ejemplo: Ejercicio 18-1

Medimos en 5 sujetos la concentración de cafeína en sangre después de tomar cierta cantidad de café. Al mismo tiempo medimos el tiempo de reacción ante el estímulo H.

Obtenemos:

Individuo	1	2	3	4	5
Cafeína	2	4	3	6	2
Tiempo	11	9	10	7	12

Queremos contestar a la pregunta de si hay o no una relación entre la cafeína en sangre y la rapidez de reflejos.

- Se trata de un problema de contraste entre dos variables cuantitativas : cafeína en sangre y rapidez de reflejos (medida como tiempo de reacción). A resolver por la fórmula n° 14, si cumple la condición de aplicación. Ho: no hay ninguna relación entre las variables, son independientes.
- Hay que comprobar si cumple la condición de aplicación. Para ello tenemos que calcular las respectivas varianzas:

La de la cafeína es 2.8 y la del tiempo de reacción es 3.7

$V = 3.7 / 2.8 = 1.32$ , que es menor que  $F(4, 4, 0.05) = 6.39$ , por lo que sí cumple la condición de aplicación y podemos utilizar la fórmula n° 14, contraste de r

- Calculamos el coeficiente de correlación, y obtenemos  $r = -0.979$

$$Z = \frac{-0,979\sqrt{(5-2)}}{\sqrt{1-(-0,979)^2}} = -8,32, \text{ que es } > t(3, 0.01) = 5.84, \text{ por tanto se rechaza } H_0 \text{ a ese ni-}$$

vel de significación.  $p < 0,01$

Existe una relación inversa (signo negativo!) entre cafeína en sangre y rapidez de reflejos: a más cafeína, menor tiempo de reacción (es decir, más rapidez de reflejos), y a menos cafeína, más tiempo de reacción (es decir, reflejos más lentos).

El problema es experimental y por tanto puede establecerse una relación causa-efecto.

## 2) PRUEBA NO PARAMÉTRICA : contraste de $r_s$

Es el **test de correlación de rango de Spearman**. Se usa cuando no puede hacerse un contraste de  $r$  por no cumplir la condición de aplicación (igualdad de varianzas en el caso de muestras pequeñas) o los datos no proceden de una población distribuída normalmente. Hay que calcular el coeficiente de correlación de Spearman,  $r_s$ , utilizando la siguiente plantilla, que también se ofrece en el cuadernillo de fórmulas.

### Fórmula nº 15: Test de rango de Spearman ( $r_s$ )

Los datos de X e Y se ordenan por separado

Individ.	X	Y	R de X	R de Y	d	$d^2$
1						
2						
3						
4						
5						
6						
.....						
N						
					Suma	
						$\Sigma d^2$

Una vez ordenados los datos se asigna a cada uno de ellos su número de orden (Rango) y se anota en la columna R que corresponda, según se indica más abajo

**Cálculo:**

$$r_s = 1 - \frac{6 \sum d^2}{N(N^2 - 1)} \quad (\text{fórmula nº 15})$$

Z se calcula por la fórmula 14, sin condición de aplicación, dándole a  $r$  el valor de  $r_s$ . Valoración por  $t_{N-2}$  ( si  $N < 30$ ) ó DN (si  $N \geq 30$ )

Pasos:

1º- ordenar por separado los datos de ambas variables ( de mayor a menor o de menor a mayor), asignándoles números de orden correlativos. Cuando un dato se repite una o más veces (casos "ex equo") a cada uno se le asigna la media de los números de orden que les corresponderían (con un decimal).

2º- se restan los números de orden de cada individuo ( $d$ ) y esta diferencia se eleva al cuadrado ( $d^2$ ). Al final, se suma la columna de  $d^2$ , obteniendo  $\Sigma d^2$

3º- se aplica la fórmula para calcular  $r_s$

4º-  $r_s$  se valora por la fórmula nº 14 (poniendo  $r_s$  donde dice  $r$ ) y valorando por la  $t$  de Student con g.l.  $N-2$  o por la DN en función de lo que valga  $N$ .

Ejemplo. Ejercicio nº 18-2

En 10 individuos realizamos alternativamente al azar un test de memoria y otro de atención, obteniendo las siguientes puntuaciones:

Individuo	1	2	3	4	5	6	7	8	9	10
Memoria	6	4	3	5	3	2	1	5	4	1
Atención	12	6	4	12	6	2	2	14	10	1

¿Están relacionados los resultados de ambos tests?

- se trata de un problema de contraste entre dos variables CT (problema de relación).  $H_0$ : no existe ninguna relación. Hay independencia

- A resolver de entrada por la fórmula nº 14, si cumple la condición de aplicación, ya que se trata de una muestra pequeña.
- Calculamos las varianzas, obteniendo 2.93 para M. y 22.77 para A.  
 $V = 22'7 / 2'93 = 7'77$ , que es mayor que  $F(9, 9, 0'05)=3'18$ ; por tanto no cumple la condición de aplicación y no podemos aplicar la fórmula nº 14, pasando de oficio al test no paramétrico, de la  $r_s$ . Para ello, utilizaremos la plantilla, escribiendo primero los datos originales y después calculando los rangos de X (= Memoria) e Y (=Atención)

Individuos	X	Y	R de X	R de Y	d	d <sup>2</sup>
1	6	12	10	8.5	1.5	2.25
2	4	6	6.5	5.5	1	1
3	3	4	4.5	4	0.5	0.25
4	5	12	8.5	8.5	0	0
5	3	6	4.5	5.5	-1	1
6	2	2	3	2.5	0.5	0.25
7	1	2	1.5	2.5	-1	1
8	5	14	8.5	10	-1,5	2.25
9	4	10	6.5	7	-0.5	0.25
10	1	1	1.5	1	-0.5	0.25
					$\Sigma d^2 =$	8.5

- Calculamos  $r_s$ :

$$r_s = 1 - \frac{6 * 8,5}{10(10^2 - 1)} = 0,948$$

- Aplicamos la fórmula 14 con  $r = 0'948$  y obtenemos

$$Z = 0,948 \sqrt{10 - 2} / \sqrt{1 - 0,948^2} = 8,42$$

que es mayor que  $t(8, 0'001)=5'041$  rechazando a ese nivel de significación la hipótesis nula y aceptando la alternativa con  $p < 0,001$ . Por tanto, sí hay relación entre ambas puntuaciones. La relación es directa: a mayor nota en Memoria, mayor en Atención y viceversa.

### Recordatorio

Con datos independientes los problemas de relación se resuelven con las mismas fórmulas que los de comparación; en cambio si los datos son apareados, las fórmulas son distintas para cada caso.

Según el enfoque que se haga del problema, pueden surgir dudas en algunos casos entre si hay que tomar la fórmula nº 10 ó la nº 14. Ambas sirven para datos apareados. La 10 para el contraste de una variable CL y otra CT, **comparando** los datos CT obtenidos en cada modalidad de la variable CL. En la nº 14 se estudia la **relación**. Por tanto si surge la duda al tratar datos apareados de este tipo sobre la fórmula a usar, preguntarse siempre si es un problema de comparación o de relación. ¿Se pide si los valores son más o menos iguales o no, o bien se pide que se pruebe si hay o no una relación entre ellos?

Si con los datos del ejercicio 18-2 se nos preguntara: ¿Hay diferencias importantes entre los resultados de ambos tests? se trataría del contraste de una variable CL (tipo de test: M y A) y de otra CT (puntuación obtenida en el test, que medimos en ambas modalidades de la CL).

Datos apareados. A resolver por la fórmula nº 10, que no tiene condición expresa de aplicación. Ho: no hay diferencias entre las puntuaciones obtenidas en cada test.

Aplicando la fórmula 10 obtenemos :  $z = -3.416$  que es mayor en valor absoluto que  $t(9, 0.01)$  que vale 3.250. Por tanto, se rechaza  $H_0$  a ese nivel de 0.01. **Sí hay diferencias significativas** entre las puntuaciones de Memoria y Atención, en el sentido de que las de Memoria son más bajas con  $p < 0,01$