

Tópicos en investigación clínica

PRUEBA NO PARAMÉTRICA DE CORRELACIÓN DE SPEARMAN

Fredy Mendivelso¹, Milena Rodríguez²

1. MD. MPH. MSc. FETP. Clínica Reina Sofía. Departamento de Cirugía
2. MV. Epidemióloga. MSc. FETP. Fundación Universitaria Sanitas. Facultad de Medicina

RESUMEN

La correlación es una medida estadística que permite conocer el grado de asociación lineal entre dos variables cuantitativas u ordinales (X, Y). También determina la fuerza de asociación y dirección que toma esta relación mediante el cálculo del coeficiente de correlación, cuyo resultado puede variar en el intervalo [-1, +1]. Entre más cercano a 1 sea el coeficiente de correlación, mayor la fuerza de asociación. Cuando una relación es aleatoria o no existe, el coeficiente tiende a cero. Los coeficientes de correlación solo miden relaciones lineales de variables continuas con distribución normal (Pearson) o monótonas con variables ordinales organizadas en rangos o jerarquías (Spearman), las cuales tienden a cambiar al mismo tiempo, pero no necesariamente a un ritmo constante. Las pruebas de correlación son muy usadas en la investigación biomédica para determinar la tendencia de dos variables a ir juntas, a lo que también se denomina covarianza; lo que no necesariamente significa que cuando se reporta una correlación (+/-) esta represente una relación de causa y efecto.

Palabras clave: Estadísticas no paramétricas; Correlación de datos; Medicina Clínica; Investigación

DOI: <https://doi.org/10.26852/01234250.578>

Recibido: 23/02/2022

Aceptado: 25/04/2022

Correspondencia: fmendivelso@colsanitas.com

SPEARMAN'S CORRELATION NONPARAMETRIC TEST

ABSTRACT

Correlation is a statistical measure that allows knowing the degree of linear association between two quantitative or ordinal variables (x, y). It also determines the strength of the association and the direction that this relationship takes by calculating the correlation coefficient, the result of which can vary in the interval [-1, +1]. The closer the correlation coefficient is to 1, the stronger the association. When a relationship is random or does not exist, the coefficient tends to be zero. Correlation coefficients only measure linear relationships of continuous variables with normal (Pearson) or monotone distribution with ordinal variables organized in ranks or hierarchies (Spearman), which tend to change at the same time, but not necessarily at a constant rate. Correlation tests are widely used in biomedical research to determine the tendency of two variables to go together, which is also called covariance; This does not necessarily mean that when a correlation (+/-) is reported, it represents a cause and effect relationship.

Keywords: Statistics, Nonparametric; Correlation of Data; Clinical Medicine; Research

PRUEBA DE CORRELACIÓN DE RANGOS DE SPEARMAN

Propósito

Determinar el grado de correlación entre dos series de observaciones pareadas (1)

Supuestos

Variables cuantitativas de libre distribución o datos ordinales

Las observaciones x_i, y_i se han obtenido por parejas
Tendencia monótona en el análisis de dispersión (2,3)

Prueba de hipótesis

H_0 : Los datos son independientes

H_1 : Los datos están relacionados

Cálculo

A todas las observaciones de x_i se les asignan los números de rango en orden descendente: 1, 2, ..., n.

Un procedimiento similar se realiza para todas las observaciones de y_i . Para cada pareja de observaciones se calcula su rango d_i . Finalmente, la diferencia entre los rangos [R] se calcula en la fórmula 1 (4,5):

$$R = \sum_{i=1}^n d_i^2$$

Donde:

R= diferencia entre los rangos

n= número de observaciones

d= rango calculado para cada pareja de observaciones

Luego se rempazan los valores en la fórmula 2:

$$r_s = 1 - \frac{6R}{n(n^2 - 1)}$$

Ejemplo

Se desea evaluar si existe algún grado de correlación entre las calificaciones en el curso de estadística y física en nueve estudiantes universitarios. Los datos se mencionan en el siguiente cuadro:

Paso 1. Crear una tabla con los datos reportados

ESTUDIANTE (N)	ESTADÍSTICA (X)	FÍSICA (Y)
1	35	30
2	23	33
3	47	45
4	17	23
5	10	8
6	43	49
7	9	12
8	8	4
9	28	31

Fuente: elaboración propia a partir de datos simulados

Paso 2. Frente a cada columna de calificaciones se ordenan los valores de manera descendente

ESTUDIANTE (N)	ESTADÍSTICA (X)	RANGO _X	FÍSICA (Y)	RANGO _Y
1	35	3	30	5
2	23	5	33	3
3	47	1	45	2
4	17	6	23	6
5	10	7	8	8
6	43	2	49	1
7	9	8	12	7
8	8	9	4	9
9	28	4	31	4

Paso 3. Se calcula el valor absoluto del rango |d| y la diferencia entre los rangos será la suma total de los cuadrados de la diferencia (R) descrita en la fórmula 1.

ESTUDIANTE (N)	ESTADÍSTICA (X)	RANGO _X	FÍSICA (Y)	RANGO _Y	D= X-Y	D ²
1	35	3	30	5	3-5= -2	(2) ² = 4
2	23	5	33	3	5-3= 2	(2) ² = 4
3	47	1	45	2	1-2= -1	(1) ² = 1

4	17	6	23	6	6-6= 0	(0) ² = 0
5	10	7	8	8	7-8= -1	(1) ² = 1
6	43	2	49	1	2-1= 1	(1) ² = 1
7	9	8	12	7	8-7= 1	(1) ² = 1
8	8	9	4	9	9-9= 0	(0) ² = 0
9	28	4	31	4	4-4= 0	(0) ² = 0
Σ						12

|d| = valor absoluto de la diferencia

Paso 4. Se reemplazan los valores en la fórmula 2

$$r_s = 1 - \frac{6R}{n((n^2 - 1))}$$

$$r_s = 1 - \frac{6(12)}{9(9^2 - 1)}$$

$$r_s = 1 - \frac{72}{9(81 - 1)}$$

$$r_s = 1 - \frac{72}{720}$$

$$r_s = 1 - 0,1$$

$$r_s = 0,90$$

Paso 5. Interpretación. El coeficiente de correlación de Spearman para estos datos es de 0,9, lo que significa que entre los rangos de las calificaciones de los nueve alumnos, existe una relación directa (función monótona creciente) entre los puntajes obtenidos por los estudiantes en los cursos de estadística y física.

CONCLUSIÓN

A diferencia de los estudios de regresión, en los de correlación simple no se ejerce control sobre ninguna de las variables (x_i, y_i) y las dos cumplen papeles simétricos. Se trata de estudiar la tendencia en el comportamiento que tienen las dos variables como pareja, dejando la función de pronóstico en un segundo plano. El coeficiente de correlación de Spearman se basa en los rangos $r(x_i), r(y_i)$ de los valores x_i, y_i al ordenar cada muestra por separado. El estudio de los residuos permite identificar posibles anomalías por tendencias,

por debilidad de la relación, por la presencia de valores extremos y eventualmente de valores influyentes que puedan distorsionar la forma como se ven las variables relacionadas.

REFERENCIAS

1. Arndt S, Turvey C, Andreasen NC. Correlating and predicting psychiatric symptom ratings: Spearman's r versus Kendall's tau correlation. *J Psychiatr Res.* 1999;33(2):97-104. [https://doi.org/10.1016/S0022-3956\(98\)90046-2](https://doi.org/10.1016/S0022-3956(98)90046-2)
2. Pinilla JO, Rico AFO. ¿ Pearson y Spearman, coeficientes intercambiables? *Comun en Estad.* 2021;14(1):53. <https://doi.org/10.15332/23393076.6769>
3. Talavera JO, Rivas-Ruiz R. Clinical research IV. Relevancy of the statistical test chosen. *Rev Med Inst Mex Seguro Soc.* 2011;49(4):401-5.
4. Hazra A, Gogtay N. Biostatistics Series Module 6: Correlation and Linear Regression. *Indian J Dermatol.* 2016;61(6):593-601. <https://doi.org/10.4103/0019-5154.193662>.
5. Fredricks GA, Nelsen RB. On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables. *J Stat Plan Inference.* 2007;137(7):2143-2150. <https://doi.org/10.1016/j.jspi.2006.06.045>



Fundación
Universitaria Sanitas
Sanitas Internacional