Rank Correlation

The correlation coefficient is a measure of relationship between the variables X and Y. There is another measure that also gives a measure of relationship using the position or 'ranks' of the items in the X and Y series. In Rank correlation, we find how the ranks in each series is correlated.

One advantage of finding the rank correlation we can avoid cumbersome calculation involving x_i, y_i and need to deal only with numbers 1,2, …, n.

The Spearman's rank correlation coefficient is given by

1. $\rho = 1 - \dfrac{6\sum d_i^2}{n(n^2-1)}$ ...(*) **if there are no repeating ranks**

2. $\rho = 1 - \dfrac{6\left[\sum d_i^2 + \dfrac{1}{12}\left[m_i(m_i^2-1) + m_k(m_k^2-1) + \ldots\right]\right]}{n(n^2-1)}$ ...(**)

if the ranks repeat, where

m_i is the number of times rank i repeats, m_k is the number of times rank k repeats etc...

1. Calculate the Spearman's rank correlation coefficient for the following data of marks in two subjects Maths and Physics:

Maths	80	75	78	93	98	100
Physics	45	65	68	72	71	69

Maths (x)	Physics (y)	Rank x	Rank y	d_i	d_i^2
80	45	4	6	-2	4
75	65	6	5	1	1
78	68	5	4	1	1
93	72	3	1	2	4
98	71	2	2	0	0

100	69	1	3	-2	4

$\sum d_i^2 = 14$, n=6

By Spearman's rank correlation,

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

$$= 1 - \frac{6\times 14}{6(6^2-1)} = 0.6$$

2. Obtain the rank correlation coefficient for the following data:

Marks in X	10	12	18	18	15	40
Marks in Y	12	18	25	25	50	25

Note: Some of the students in the above data have the same marks. How do you decide their ranks? For the marks in X, 2 students have 18 marks. You can see that they both should have the second rank. So considering they will together occupy the 2nd and 3rd place, their ranks will be the average of the places, i.e. (2+3)/2=2.5. So, both the students who got 18 marks will have the rank 2.5. Similarly, for the marks in Y, 3 students have 25 marks. They are in the 2nd place. But collectively, they will occupy places 2,3,4. Hence, their rank is (2+3+4)/3=3.

Marks in X	Marks in Y	Rank for X	Rank for Y	d_i	d_i^2
10	12	6	6	0	0
12	18	5	5	0	0
18	25	2.5	3	-0.5	0.25
18	25	2.5	3	-0.5	0.25
15	50	4	1	3	9
40	25	1	3	-2	4

$\sum d_i^2 = 13.5$, n=6

By Spearman's rank correlation,

$$\rho = 1 - \frac{6\left\{\sum d_i^2 + \frac{1}{12}\left[m_i\left(m_i^2 - 1\right) + m_k(m_k^2 - 1) + \cdots\right]\right\}}{n(n^2 - 1)}$$

$$= 1 - \frac{6\left\{13.5 + \frac{1}{12}[2(2^2 - 1) + 3(3^2 - 1) + \cdots]\right\}}{6(6^2 - 1)}$$

$$= 1 - \frac{6\times 16}{6(6^2-1)} = 0.5428$$

Prof. Nancy Sinollin