Rank Correlation Coefficient or Spearman's Correlation Coefficient

Let $\{(x_i, y_i);\ i = 1, 2, 3, \ldots, n\}$ be the given set of paired data. Then the rank correlation coefficient is defined by

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2-1)}$$

where d_i is the difference of ranks assigned to x_i and y_i.

Note:

(i) We assign the rank 1 to highest value of x_i or y_i and remaining x_i's or y_i's are ~~assi~~ assigned the rank in descending order.

(ii) When there are ties in rank, then we assign the common rank to these observations and this common rank is equal to the average of ranks they would have assumed in case of marginal differences.

(iii) When there is no ties in the rank, then $r_s = r$ (Correlation Coefficient)

Q. Show that Spearman's rank correlation coefficient r_s also lies between -1 and 1.

Sol^n

Spearman's rank correlation coefficient is given by $r_s = 1 - \dfrac{6 \sum d_i^2}{n(n^2-1)}$ —— (1)

It is maximum when $\sum_{i=1}^{n} d_i^2$ is minimum

and we know that min. $\left(\sum_{i=1}^{n} d_i^2\right) = 0$

$\Rightarrow d_i^2 = 0 \quad \forall i$

$\Rightarrow d_i = 0 \quad \forall i$

$\Rightarrow$ rank x_i − rank $y_i = 0 \ \forall i$ $\Rightarrow$ ~~$\text{rank}(x_i - y_i) = 0 \quad \forall i$~~

$\Rightarrow \text{rank } x_i = \text{rank } y_i \quad \forall i = 1, 2, 3, \ldots n$

Hence maximum value of r_s is

$$\text{max. } r_s = 1 - \frac{6(0)}{n(n^2-1)}$$

$\Rightarrow$ max. $r_s = 1$.

Now, r_s is minimum when x_i and y_i have opposite rank

that is, rank x_i + rank $y_i = n+1, \ \forall i = 1, 2, \ldots n$.

$\therefore$ In this case, $\sum_{i=1}^{n} d_i^2 = \cancel{\sum_{i=1}^{n}(x_i - y_i)^2}$

$$= \sum_{i=1}^{n} \left(\text{rank}(x_i) - \text{rank}(y_i)\right)^2$$

$$= \sum_{i=1}^{n} \left[2\,\text{rank}(x_i) - (n+1)\right]^2$$

$$= 4\sum_{i=1}^{n} \left(\text{rank}(x_i)\right)^2 + n(n+1)^2 - 4(n+1)\sum_{i=1}^{n}\text{rank}(x_i)$$

$$= 4\,\frac{n(n+1)(2n+1)}{6} + n(n+1)^2 - 4(n+1)\cdot\frac{n(n+1)}{2}$$

$\left[\because \text{rank}(x_i) \text{ varies from } 1 \text{ to } n\right]$

$$= n(n+1)\left[\frac{2(2n+1)}{3} - (n+1)\right] = \frac{n(n^2-1)}{3}$$

Hence, $\min. r_s = 1 - \dfrac{6\,\frac{(n)(n^2-1)}{3}}{n(n^2-1)} = 1 - 2$

$\Rightarrow \min. r_s = -1.$

Q. The ranking of 10 students in two subject A and B are as follows

marks in A	68	64	75	50	64	80	75	40	55	64
marks in B	62	58	68	45	81	60	68	48	50	70

Calculate the rank correlation coefficient

Soln

Marks in A (x_i)	Marks in B (y_i)	Rank of x_i (R_1)	Rank of y_i (R_2	$d_i = R_1 - R_2$	d_i^2
68	62	4	5	-1	1
64	58	6	7	-1	1
75	68	2.5	3.5	-1	1
50	45	9	10	-1	1
64	81	6	1	5	25
80	60	1	6	-5	25
75	68	2.5	3.5	-1	1
40	48	10	9	1	1
55	50	8	8	0	0
64	70	6	2.	4	16

$\Sigma d_i^2 = 72$

Here we can see that highest value of $x_i = 80$

$\therefore$ rank 1 is assigned to 80.

After 80, maximum value of x_i is 75. But 75 is repeated twice against rank 2 and 3

$\therefore$ ~~rank of 75~~ $\therefore$ rank assigned againsed 75 is

$$R_i = \frac{2+3}{2} = 2.5$$

After that next maximum value of x_i is 68

therefore, it has assigned rank 4.

Similarily, we can assign the rank to all values of x_i and y_i.

average rank $\left(6 = \frac{5+6+7}{3}\right)$ is assigned to the value 64, that is repeated thrice.

Now, $r_s = 1 - \frac{6 \sum d_i^2}{n(n^2-1)} = 1 - \frac{6(72)}{10 \times 99}$

$= 1 - 0.4364 = 0.5636$

Q. Calculate the rank correlation coefficient for the following data.

No. of hours (x)	8	5	11	13	10	5	18	15	2	8
Score (y)	56	44	79	72	54	94	85	33	70	65

Here in x_i rank 1 is assigned against 18.

repeatation on 6th and 7th rank of entry 8

∴ rank $\frac{6+7}{2} = 6.5$ has assigned against 6th 8.

Similarily, we can assign rank to all x_i's and y_i's.

No. of hours x	Score (y)	Rank(x) R_1	Rank(y) R_2	$d_i = R_1 - R_2$	d_i^2
8	56	6.5	7	-0.5	0.25
5	44	8.5	9	-0.5	0.25
11	79	4	3	1	1
13	72	3	4	-1	1
10	54	5	8	-3	9
5	94	8.5	1	7.5	56.25
18	85	1	2	-1	1
15	33	2	10	-8	64
2	70	10	5	5	25
8	65	6.5	6	0.5	0.25

$\boxed{\sum d_i^2 = 158}$

∴ The rank correlation coefficient is

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2-1)} = 1 - \frac{6(158)}{(10)(99)}$$

$$\Rightarrow r_s = 1 - \frac{158}{165} = 1 - 0.9576$$

$$\Rightarrow r_s = 0.0424$$

Hence, there is almost zero correlation between no. of hours and score.