Rank Correlation

Lesson 2: Spearman's Rank Correlation Coefficient

There are some characteristics, eg. beauty, intelligence etc. which cannot be measured quantitatively. In these cases, the different individuals are ranked according to their merit. The correlation calculated from two such distributions of ranks is known as the rank correlation and is measured by Spearman's rank correlation coefficient as

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

where n = number of individuals and d_i = difference in ranks of an individual

Case of Tie:

If some individuals have same rank, that is if there is a tie in merit or proficiency, then the corresponding formula of rank correlation coefficient becomes

$$R = 1 - \frac{6\left[\sum_{i=1}^{n} {d_i}^2 + \sum_{i=1}^{n} \frac{{t_i}^3 - t_i}{12}\right]}{n(n^2-1)}$$

where t_i = number of individuals in each tie in both the distributions

Problems:

Ex.1. 10 competitors in a music contest were ranked by two judges X and Y in the following manner:

Ranked by X	1	6	5	10	3	2	4	9	7	8
Ranked by Y	6	4	9	8	1	2	3	10	5	7

Calculate the rank correlation coefficient.

Solution: Let us make the table for calculation of rank correlation coefficient:

Ranked by X	1	6	5	10	3	2	4	9	7	8	-
Ranked by Y	6	4	9	8	1	2	3	10	5	7	-
$d_i = X - Y$	-5	2	-4	2	2	0	1	-1	2	1	0
${d_i}^2$	25	4	16	4	4	0	1	1	4	1	60

Then the rank correlation coefficient is given by

$$R = 1 - \frac{6\sum_{i=1}^{n} {d_i}^2}{n(n^2-1)} = 1 - \frac{6 \times 60}{10(100-1)} = 0.64$$

Ans.

Ex.2. The coefficient of rank correlation of marks obtained by 10 students in Mathematics and Statistics was found to be 0.5. Then it was detected that the difference in ranks in the two subjects for one particular student was wrongly taken to be 3 in place of 7. What should be the correct rank correlation coefficient?

Solution: We know that

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

$$\text{or,} \quad 0.5 = 1 - \frac{6\sum_{i=1}^{n} {d_i}^2}{10 \times 99}$$

$$\text{or,} \sum_{i=1}^{n} {d_i}^2 = 82.5$$

Which is the incorrect summation. Then the corrected summation will be

$$\sum_{i=1}^{n} {d_i}^2 \Big|_{\text{corrected}} = 82.5 - 3^2 + 7^2 = 122.5$$

Hence, the correct rank correlation coefficient will be

$$R = 1 - \frac{6\sum_{i=1}^{n} {d_i}^2}{n(n^2-1)} = 1 - \frac{6 \times 122.5}{990} = 0.258.$$

Ans.

Ex.3. 10 students obtained the following marks in Mathematics and statistics. Calculate the rank correlation coefficient.

Roll No.	1	2	3	4	5	6	7	8	9	10
Marks in Math	78	52	48	68	52	25	90	52	48	69
Marks in Stats.	68	42	60	58	42	30	78	42	58	61

Solution: Let us rank the students according to their marks:

Roll No.	1	2	3	4	5	6	7	8	9	10	-
Marks in Math	78	52	48	68	52	25	90	52	48	69	-
Marks in Stat	68	42	60	58	42	30	78	42	58	61	-
Rank in Math (M)	2	6	8.5	4	6	10	1	6	8.5	3	-
Rank in Stat (S)	2	8	4	5.5	8	10	1	8	5.5	3	-
$d_i = M - S$	0	-2	4.5	-1.5	-2	0	0	-2	3	0	0
${d_i}^2$	0	4	20.25	2.25	4	0	0	4	9	0	43.5

We can see that 3 students (roll nos. 2, 5 and 8) have got marks 52 in Math. As a result, their ranks should be identical. Since the tie has occurred in the 5th rank, so each one is given the average rank 6 as $\frac{5+6+7}{3} = 6$. The next rank will then be considered as 8th. Again there are 2 students (roll nos. 3 and 9) who have got the same score 48. So, instead of 8th and 9th rank, each one is given the average rank 8.5 as $\frac{8+9}{2} = 8.5$. The next rank will then be 10th.

Similarly for Stat distribution, there are 2 students (roll nos. 4 and 9) with same marks and this occurs in the 5th rank. So each one is given the average rank 5.5 and the next rank should be 7th. In this rank also, there are 3 students (roll nos. 2, 5 and 8) with the same score. So again an average rank 8th is given to each of them and the next rank becomes the last rank 10th.

Let us calculate the cases of tie for the two distributions separately.

Math Distribution: In this case, there are two ties. One involves 2 individuals and the other involves 3 individuals. Thus the correction for this distribution is:

$$\sum_{i=1}^{n} \frac{{t_i}^3 - t_i}{12} = \frac{2^3 - 2}{12} + \frac{3^3 - 3}{12} = 2.5$$

Stat Distribution: In this case also, there are two ties. Here also one tie involves 2 individuals and the other tie involves 3 individuals. Thus the correction for this distribution becomes:

$$\sum_{i=1}^{n} \frac{{t_i}^3 - t_i}{12} = \frac{2^3 - 2}{12} + \frac{3^3 - 3}{12} = 2.5$$

Hence, the total correction for both the distributions is (2.5 + 2.5) = 5

Then the rank correlation coefficient becomes

$$R = 1 - \frac{6\left[\sum_{i=1}^{n} {d_i}^2 + \sum_{i=1}^{n} \frac{{t_i}^3 - t_i}{12}\right]}{n(n^2 - 1)} = 1 - \frac{6(43.5 + 5)}{990} = 0.71$$

Ans.

Ex.4. Find the coefficient of rank correlation of the following data:

Scores of A	115	109	112	87	98	98	120	100	98	118
Scores of B	75	73	85	70	76	65	82	73	68	80

Solution: Let us rank the scores of A and B:

Scores of A	115	109	112	87	98	98	120	100	98	118	-
Scores of B	75	73	85	70	76	65	82	73	68	80	-
Rank of A (A)	3	5	4	10	8	8	1	6	8	2	-
Rank of B (B)	5	6.5	1	8	4	10	2	6.5	9	3	-
$d_i = A - B$	-2	-1.5	3	2	4	-2	-1	-0.5	-1	-1	0
${d_i}^2$	4	2.25	9	4	16	4	1	0.25	1	1	42.5

For the series A, the score 98 is identical for 3 places and this occurs at the 7th rank, so the average rank = $\frac{7+8+9}{3} = 8^{th}$ is given to each one of these 3 scores and the next score is thus 10th.

For the series B, the score 73 is same for 2 places which has occurred in 6th rank, so the average rank = $\frac{6+7}{2} = 6.5^{th}$ is given to each of these two scores. This means that the next rank will be 8th.

<u>A Series:</u> In this case, the tie involves 3 scores and thus the correction for this distribution is:

$$\sum_{i=1}^{n} \frac{{t_i}^3 - t_i}{12} = \frac{3^3 - 3}{12} = 2$$

<u>B Series:</u> In this case, the tie involves 2 scores, so the correction for this distribution is:

$$\sum_{i=1}^{n} \frac{{t_i}^3 - t_i}{12} = \frac{2^3 - 2}{12} = 0.5$$

Hence, the total correction for both the distributions is (2 + 0.5) = 2.5

Then the rank correlation coefficient becomes

$$R = 1 - \frac{6\left[\sum_{i=1}^{n} {d_i}^2 + \sum_{i=1}^{n} \frac{{t_i}^3 - t_i}{12}\right]}{n(n^2-1)} = 1 - \frac{6(42.5 + 2.5)}{990} = 0.73$$

Ans.