CORRELATION PART-4

SPEARMAN RANK CORRELATION

Sometimes we are given a series of items whose no numerical measure can be made, but where best and worst or most favoured and least favoured can be identified. Rankings are often applied to put the series in order here.

For ex: Characteristics like beauty, intelligence, leadership, honesty, cannot be numerically measured, but the individuals in such group can be arranged in order thereby obtaining for each individual a number indicating its rank in the group.

Let us suppose that a group of 'n' individuals is arranged in order of merits or proficiency in possession of two characteristics A and B. These ranks in the two characteristics will, in general , be different.

Example: If we consider the relation between intelligence and beauty, is it necessary that a beautiful individual is intelligent also?

TO ANSWER THIS WE USE SPEARMAN'S COEFFICIENT OF RANK CORRELATION:

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

WHERE R_R DENOTES RANK CORRELATION COEFFICIENT AND IT LIES BETWEEN –1 AND 1 INCLUSIVE OF THESE TWO VALUES.

$D_I = X_I - Y_I$ REPRESENTS THE DIFFERENCE IN RANKS FOR THE I-TH INDIVIDUAL AND N DENOTES THE NUMBER OF INDIVIDUALS.

WHEN R = +1, THERE IS COMPLETE AGREEMENT IN THE ORDER OF RANKS AND THE RANKS ARE IN THE SAME DIRECTION.

WHEN R= –1, THERE IS COMPLETE AGREEMENT IN THE ORDER OF RANKS AND THEY ARE IN OPPOSITE DIRECTIONS

CASE 1: WHEN ACTUAL RANKS ARE GIVEN

WHERE ACTUAL RANKS ARE GIVEN, THE STEPS REQUIRED FOR COMPUTING RANK CORRELATION ARE :

1. TAKE THE DIFFERENCE OF THE TWO RANKS, IE., (R1 - R2) AND DENOTE THESE DIFFERENCES BY D.

2. SQUARE THESE DIFFERENCES AND OBTAIN THE TOTAL $\sum D_1^2$.

3. APPLY THE FORMULA

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

CASE 1: WHEN ACTUAL RANKS ARE GIVEN

QUESTION:
COMPUTE RANK CORRELATION FROM FOLLOWING DATA RELATING TO RANKS GIVEN BY 2 JUDGES IN A CONTEST

Serial No. of Candidate :	1	2	3	4	5	6	7	8	9	10
Rank by Judge A :	10	5	6	1	2	3	4	7	9	8
Rank by Judge B :	5	6	9	2	8	7	3	4	10	1

Serial No	Rank by A (x_i)	Rank by B (y_i)	$d_i = x_i - y_i$	d^2_i
1	10	5	10-5= 5	25
2	5	6	5-6= -1	1
3	6	9	6-9= -3	9
4	1	2	1-2= -1	1
5	2	8	-6	36
6	3	7	-4	16
7	4	3	1	1
8	7	4	3	9
9	8	10	-2	4
10	9	1	8	64
Total			0	166

Since N=10, rank correlation coefficient is given by:

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

$$R_r = 1 - \frac{6 \times 166}{10(10^2 - 1)}$$

$$= -0.006$$

Practise Question

Example:

Two housewires, Mrs. A and Mrs. B, were asked to express their preference for different kinds of detergents, gave the following replies.

Detergent	Mrs. A	Mrs. B
A	4	4
B	2	1
C	1	2
D	3	3
E	7	8
F	8	7
G	6	5
H	5	6
I	9	9
J	10	10

Detergent	Mrs A(X)	Mrs B(Y)	Di=(x-y)	Di^2
A	4	4	0	0
B	2	1	1	1
C	1	2	-1	1
D	3	3	0	0
E	7	8	-1	1
F	8	7	1	1
G	6	5	1	1
H	5	6	-1	1
I	9	9	0	0
J	10	10	0	0
TOTAL				6

APPLYING THE FORMULA

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

$$R = 1 - \frac{6 \times 6}{10(10^2 - 1)}$$

R =0.9636

Question: For a group of 8 students, the sum of squares of differences in ranks for Mathematics and Statistics marks was found to be 50 what is the value of rank correlation coefficient?

As given

$N = 8$ and $\sum d_1^2 = 50$.

Hence the rank correlation coefficient between marks in Mathematics and Statistics is given by

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

$$R_r = 1 - \frac{6 \times 50}{8(8^2 - 1)}$$

$$= 0.40$$

Practise Example 34

CASE 2 : WHEN RANKS ARE NOT GIVEN

When Ranks are not given, it will be necessary to assign the ranks. Ranks can be assigned by taking either the highest value as 1 or the lowest value as 1.

Question: Compute the coefficient of rank correlation between sales and advertisement expressed in thousands of dollars from the following data:

Sales :	90	85	68	75	82	80	95	70
Advertisement :	7	6	2	3	4	5	8	1

Let the rank given to sales be denoted by x and rank of advertisement be denoted by y.

We note that since the highest sales as given in the data, is 95, it is to be given rank 1, the second highest sales 90 is to be given rank 2 and finally rank 8 goes to the lowest sales, namely 68.

We have given rank to the other variable advertisement in a similar manner.

Sales (x_i)	Advertisement (y_i)	Rank for Sales (x_i)	Rank for Advertisement (y_i)	$d_i = x_i - y_i$	d_i^2
90	7	2	2	0	0
85	6	3	3	0	0
68	2	8	7	1	1
75	3	6	6	0	0
82	4	4	5	–1	1
80	5	5	4	1	1
95	8	1	1	0	0
70	1	7	8	–1	1
Total	—	—	—	0	4

Since $n = 8$ and $\sum d^2 = 4$, apply the above formula, we get

$$r = 1 - 6 \sum d^2 / n(n^2 - 1)$$

$$r = 1 - 6 \times 4 / 8(8^2 - 1)$$

$$r = 1 - 0.0476$$

$$r = 0.95$$

The high positive value of the rank correlation coefficient indicates that there is a very good amount of agreement between sales and advertisement.

PRACTISE QUESTION: The marks obtained by students in two tests are given below:

Calculate the rank correlation coefficient and comment on this.

Preliminary Test	92	89	87	86	83	77	71	63	53	50
Final Test	86	83	91	77	68	85	52	82	37	57