1. Introduction & Steps in Finding Rho (ρ)

Welcome to the series of E-learning modules on practical's on rank correlation coefficient.

By the end of this session, you will be able to:

- Explain how to find rank correlation coefficient when there are ties and no ties
- Explain the importance in finding the correlation

Charles Edward Spearman developed this method in 1904.

Spearman's rank correlation coefficient is usually denoted by row and is given by,
Row is equal to 1 minus 6 into summation d square divided by n cube minus n,

Where, d is the difference between the pair of ranks of the same individual in the two characteristics and n is the number of pairs.

Above formula can be used only when the ranks are different. If there are ties or repeated ranks, we use following formula.

Row is equal to 1 minus 6 into summation d square plus 1 by 12 into m cube minus m plus 1 by 12 into m cube minus m etc.

Where, m is number of values involved in a tie.

Let's look at the Steps in finding row.

- Determine the ranks of X and Y.
  - If ranks are given in the question, the above step is not needed
  - If rank is not given, the highest value is given rank 1, 2nd highest value is given rank 2 and we proceed like this for remaining value
  - If there are two or more equal values then we get an obstruction in giving ranks. We have a practice that if two students score equal marks then we give them equal ranks
- Take the difference of two ranks that is R one minus R two and denote them as d
- Square these differences and total them to get summation d square
- Apply required formula

2. Exercise on Coefficient of Rank Correlation – (Part-1)

Exercise 1:
The ranks of students in Hindi and Economics are given as follows.

Figure 1

Hindi	Economics
6	3
1	1
5	4
2	2
4	5
3	6

Calculate the coefficient of rank correlation.

Solution:
To find the rank correlation coefficient, we find the following table.

Figure 2

Hindi	Economics	d	d^2
6	3	3	9
1	1	0	0
5	4	1	1
2	2	0	0
4	5	-1	1
3	6	-3	9
			$\Sigma d^2=20$

First two columns of the table are written as it is in the question.

The third column is obtained by finding the difference between the ranks in the first and second columns.

The last column is obtained by squaring the numbers in the 3rd column.
Sum of the last column gives the value of summation d square.

Since, the ranks are not repeated, we use the following formula to find the rank correlation coefficient.
Row is equal to 1 minus 6 into summation d square divided by n cube minus n.

By substituting, we get,
Row is equal to 1 minus 6 into 20 divided by 6 cube minus 6 is equal to zero point 4, 2, 9.

Exercise 2:
10 students got the following percentage of marks in Economics and Statistics.

Figure 3

Economics	78	36	98	25	75	82	92	62	65	39
Statistics	84	51	91	60	68	62	86	58	35	49

Calculate rank correlation coefficient.

Solution:

Figure 4

Economics	Statistics	R_1	R_2	d	d^2
78	84	4	3	1	1
36	51	9	8	1	1
98	91	1	1	0	0
25	60	10	6	4	16
75	68	5	4	1	1
82	62	3	5	-2	4
92	86	2	2	0	0
62	58	7	7	0	0
65	35	6	10	-4	16
39	49	8	9	-1	1
					$\Sigma d^2 = 40$

First two columns are written as it is in the question.

The third column, R 1 denotes the ranks of marks scored in economics. The highest score is 98 and it is given rank 1. The second highest score is 92, which is given 2 and we proceed like this to give the rank for all the scores.

Similarly, we obtain the fourth column, where the numbers written are the ranks given to the scores of Statistics.

The fifth column gives the value of d which is obtained by subtracting R2 from R 1 and the last column is obtained by squaring the numbers in the 5^{th} column.

We find the sum of the numbers in the last column which gives summation d square.

Since, there are no ties, we use the following formula.

Row is equal to 1 minus 6 into summation d square divided by n cube minus n.
Is equal to 1 minus 6 into 40 divided by 10 cube minus 10.
Is equal to 1 minus zero point 2, 6, which is same as zero point 7, 4.

It shows a high degree of positive correlation between the percentage of marks in Economics and Statistics.
That is, those who score more in Economics are expected to score more in Statistics also.

3. Exercise on Coefficient of Rank Correlation – (Part-2)

Exercise 3:
Ten competitors in a beauty contest are ranked by three judges in the following order. Use the rank correlation coefficient to discuss which pair of judges has the nearest approach to common taste in beauty.

Figure 5

First Judge	1	6	5	10	3	2	4	9	7	8
Second Judge	3	5	8	4	7	10	2	1	6	9
Third Judge	6	4	9	8	1	2	3	10	5	7

Solution:
In order to determine which pair of judges has the nearest approach to common taste in beauty we shall have to calculate the rank coefficient of correlation between the ranking of

1. First and Second judge
2. Second and Third Judge
3. First and Third Judge

Now let us consider one by one and find the coefficient of rank correlation.

Figure 6

First Judge	Second Judge	d	d^2
1	3	-2	4
6	5	1	1
5	8	-3	9
10	4	6	36
3	7	-4	16
2	10	-8	64
4	2	2	4
9	1	8	64
7	6	1	1
8	9	-1	1
			$\Sigma d^2 = 200$

First column indicate the rankings given by first judge, which is considered as R1 and the second column indicate the rankings given by the second judge, which is considered as R2.

The third column denotes d is equal to R1 minus R2.
The last column is obtained by finding square of the numbers written in the third column.

Since, there are no ties, we use the following formula.
Row is equal to 1 minus 6 into summation d square divided by n cube minus n.

Where, n is equal to 10 and summation d square is equal to 200.

By substituting the values in the above formula, we get,
Row is equal to 1 minus 6 into 200 divided by 10 cube minus 10.
Is equal to 1 minus 1 point 2, 1.
Is equal to minus zero point 2, 1.

Now, let us find the rank correlation coefficient between second and third judges.

Consider the following table.

Figure 7

Second Judge	Third Judge	d	d^2
3	6	-3	9
5	4	1	1
8	9	-1	1
4	8	-4	16
7	1	6	36
10	2	8	64
2	3	-1	1
1	10	-9	81
6	5	1	1
9	7	2	4
			$\Sigma d^2 = 214$

In this table, the first column gives the ranks given by the second judge, considered as R1 and the second column is the ranks given by the third judge, considered as R2.
Third column is obtained by finding the difference, that is, R1 minus R2.
The last column is obtained by squaring the values in the third column.

Sum of last column gives summation d square.

Since, there are no ties, we use the following formula.

Row is equal to 1 minus 6 into summation d square divided by n cube minus n. where, n is equal to 10 and summation d square is equal to 214.

By substituting the values in the above formula, we get,
Row is equal to 1 minus 6 into 214 divided by 10 cube minus 10.
Is equal to 1 minus 1 point 2, 9, 7.
Is equal to minus zero point 2, 9, 7.

Now, let us find the rank correlation coefficient between first and third judges.
Consider the following table.

Figure 8

First Judge	Third Judge	d	d^2
1	6	-5	25
6	4	2	4
5	9	-4	16
10	8	2	4
3	1	2	4
2	2	0	0
4	3	1	1
9	10	-1	1
7	5	2	4
8	7	1	1
			$\Sigma d^2 = 60$

The first column denotes the ranks given by the first judge, considered as R 1 and the second column denotes the ranks given by the third judge, considered as R2.
We obtain the third column by finding the difference between the first and second column that is R1 minus R2.
The last column is obtained by squaring the numbers in the third column.
The sum of the last column gives summation d square.

Since, there are no ties, we use the following formula.

Row is equal to 1 minus 6 into summation d square divided by n cube minus n. where, n is equal to 10 and summation d square is equal to 60.

By substituting the values in the above formula, we get,
Row is equal to 1 minus 6 into 60 divided by 10 cube minus 10.
Is equal to 1 minus zero point 3, 6.
Is equal to zero point 6, 4.

If we compare the rank correlation coefficient in all the above cases, we get highest value in case of first and third judges.
Hence, first and third judge pair has the nearest approach to common taste in beauty.

4. Exercise on Coefficient of Rank Correlation – (Part-3)

Exercise 4:
The rank correlation coefficient between marks obtained by some students in Statistics and Accountancy is zero point 8.
If the total of squares of rank differences is 33, find the number of students.

Solution:
We know that,

By substituting the given information in the above, we get,
row is equal to 1 minus 6 into summation d square divided by n cube minus n

Zero point 8 is equal to 1 minus 6 into 33 divided by n cube minus n.
Is equal to 1 minus 198 divided by n into n square minus 1.

Implies, 198 divided by n into n square minus 1 is equal to 1 minus zero point 8.

Implies, 198 divided by n into n square minus 1 is equal to zero point 2.

Implies zero point 2 into n into n square minus 1 is equal to 198.

Or n into n square minus 1 is equal to 198 divided by zero point 2 is equal to 990.
By observation we find that, 10 into 10 square minus 1 is equal to 990.
Hence, the number of students is 10.

Exercise 5:
The coefficient of rank correlation of the marks obtained by 10 students of English and Economics was found to be 0.5.
It was later discovered that the difference in ranks in the two subjects obtained by one of the students was wrongly taken as 3 instead of 7.
Find the correct coefficient of rank correlation.

Solution:
We know that, row is equal to 1 minus 6 into summation d square divided by n cube minus n.
By substituting the given values in the above, we get,
Zero point 5 is equal to 6 into summation d square divided by 10 cube minus 10.
Is equal to 1 minus 6 into summation d square divided by 990.

Implies, 6 into d square dived by 990 is equal to 1 minus zero point 5 is equal to zero point 5.

That is, 6 into summation d square is equal to zero point 5 into 990.
Implies, summation d square is equal to 495 divided by 6 is equal to 82 point 5.

But this is based on the incorrect d. Hence after correction,
Summation d square is equal to 82 point 5minus 3 square plus7 square is equal to 122 point 5 and the corrected rank correlation is,
Row is equal to 1 minus 6 into summation d square dived by n cube minus n.
Is equal to 1 minus 6 into 122 point 5 divided by 10 cube minus 10.
On simplification we get, 1 minus zero point 7, 4, which is equal to zero point 2, 6.

Exercise 6:
Obtain the rank correlation coefficient for the following data with regard to two series x and y.

Figure 9

x	68	64	75	50	64	80	75	40	55	64
y	62	58	68	45	81	60	68	48	50	70

Solution:
To find rank correlation coefficient, we need to find the ranks of X and Y series.

Figure 10

x	y	R_1	R_2	d	d^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
					Σd^2 =72

Hence, the first two column in the table gives the values of x and y series, which are given in the question.

The third column is the ranks given to the x series. In the X series we see that the value 75 occurs 2 times. The common rank given to these values is 2 point 5 which is the average of 2 and 3, the ranks which these values would have taken if they were different. The next value 68, then gets the next rank which is 4.
Again we see that value 64 occur thrice. The common rank given to it is 6 which is the average of 5, 6 and 7.

Fourth column gives ranks given to y series. In the Y series the value 63 occurs twice and its common rank is 3 point 5 which is average of 3 and 4.

Fifth column gives the difference d of ranks of series x and y.
The last column gives the squared values of the numbers in the fifth column. Sum of this

column, which is written in bold numbers gives summation d square.

As we have repeated ranks in the above problem, we use the following formula.
Row is equal 1 minus 6 into summation d square plus 1 by 12 into m cube minus m plus 1 by 12 into m cube minus m plus etc. divided by n cube minus n.

Observe that in the above table, 3 ranks are repeated.
Therefore, we find 1 by 12 into m cube minus m for each value repeated.

In x series 75 is repeated twice. Hence, m is equal to 2.
Therefore, 1 by 12 into m cube minus m is equal to 1 by 12 into 2 cube minus 2 is equal to half.

64 is repeated thrice. Hence m is equal to 3. Therefore, 1 by 12 into m cube minus m is equal to 1 by 12 into 3 cube minus 3 is equal to 2.
In y series, 63 occur twice. Hence m=2. Therefore,
1 by 12 into m cube minus m is equal to 1 by 12 into 2 cube minus 2 is equal to
Half.
Therefore, coefficient of rank correlation is given by,
Row is equal to 1 minus 6 into 72 plus half plus 2 plus half divided by 10 cube minus 10 is equal to zero point 5, 4, 5.

5. Exercise on Coefficient of Rank Correlation – (Part-4)

Exercise 7:
Find rank correlation coefficient between the marks of the subjects Mathematics and Hindi

Figure 11

Mathematics	29	32	53	47	45	32	70	45	70	53
Hindi	56	60	72	48	72	35	67	67	75	31

Solution:

Figure 12

Mathematics	**Hindi**	R_1	R_2	**d**	d^2
29	56	10	7	3	9
32	60	8.5	6	2.5	6.25
53	72	3.5	2.5	1	1
47	48	5	8	-3	9
45	72	6.5	2.5	4	16
32	35	8.5	9	-0.5	0.25
70	67	1.5	4.5	-3	9
45	67	6.5	4.5	2	4
70	75	1.5	1	0.5	0.25
53	31	3.5	10	-6.5	42.25
					$\Sigma d^2 = 97$

First two columns are written as it is in the question. That is the first column indicates marks scored in mathematics and the second column indicates the marks scored in Hindi.

Now, let us obtain third and fourth column by giving the ranks to the marks scored in Mathematics and Hindi.

Observe that there are repeated marks in both mathematics and Hindi.

In mathematics, 70 is repeated twice. Hence rank for 70 is given by taking average of 1 and 2, that is 1 point 5.

Second highest 53 is also repeated twice. Hence rank for 53 is given by taking the average of 3 and 4, that is 3 point 5. The next rank 5 is given to 47. 45 is repeated twice. Hence the rank is the average of 6 and 7 that is 6 point 5. Again the next highest number 32 is also repeated twice. Hence the rank for 32 is the average of 8 and 9 that is 8 point 5. Finally the least number 29 has rank 10.

In the same way we fill the fourth column also.
The second highest mark in Hindi 72 is repeated twice. Hence rank of 60 is the average of 2 and 3, that is 2 point 5 and the next highest mark 67 is repeated twice. Hence the rank of 67 is the average of 4 and 5 that is 4 point 5.
Other marks are occurred only once and hence the ranks are given accordingly.

The fifth column gives the difference between the ranks of Mathematics and Hindi and the last column gives the squares of the numbers in the fifth column.
The total of this column gives the value of summation d square, which is equal to 97.

As we have repeated ranks in the above problem, we use the following formula.
Row is equal to 1 minus 6 into summation d square plus 1 by 12 into m cube minus m plus 1 by 12 into m cube minus m plus etc. divided by n cube minus n.

Observe that in the above table, altogether 6 ranks are repeated, 4 in Mathematics and 2 in Hindi. Further each rank is repeated twice.

Hence, for each repeated ranks, the value of 1 by 12 into m cube minus m is equal to 1 by 12 into 2 cube minus 2 is equal to half.

Therefore, coefficient of rank correlation is given by,
Row is equal to 1 minus 6 into 97 plus half plus half plus half plus half plus half plus half divided by 10 cube minus 10.
Is equal to 1 minus 6 into 100 divided by 990 is equal to zero point 3, 9, 4.

From the following data calculate Spearman's Rank Coefficient of Correlation, where we have given the rank differences and one difference is unknown.

Figure 13

Sl. No.	Rank Differences
1	-2
2	-4
3	-1
4	3
5	2
6	0
7	?
8	3
9	3
10	-2

We know that, the sum of the rank differences is always equal to zero. So assuming the

unknown rank difference as x,
Minus 2 minus 4 minus 1 plus 3 plus 2 plus zero plus x plus 3 plus 3 minus 2 is equal to zero.

Implies, x plus 2 is equal to zero or x is equal to minus 2.

Therefore, the unknown rank difference is minus 2.
Now let us write the complete table and find d square values.

Figure 14

Sl. No.	Rank Differences	d^2
1	-2	4
2	-4	16
3	-1	1
4	3	9
5	2	4
6	0	0
7	-2	4
8	3	9
9	3	9
10	-2	4
		$\Sigma d^2 = 60$

The table is written by substituting minus 2 in the missing place.
The last column is obtained by squaring the numbers in the second column. Sum of this column gives summation d square, which is equal to 60.

The formula is given by, row is equal to 1 minus 6 into summation d square divided by n cube minus n.
Is equal to 1 minus 6 into 60 divided by 10 cube minus 10 is equal to 1 minus zero point 3, 6, 4.
Is equal to zero point 6, 3, 6.

Here's a summary of our learning in this session, where we have understood:

- How to find rank correlation coefficient when there are ties and no ties
- Its importance in finding the correlation