1. Introduction

Welcome to the series of E-learning modules on rank correlation. In this module, we are going to study about Spearman's rank correlation coefficient when there are no ties and ties in the ranks, derivation of the formula, limits and remarks for the coefficient, advantage, application and limitation of rank correlation coefficient.

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

- Explain the meaning of Spearman's rank correlation coefficient
- Explain the formula for calculating the correlation when there are ties and no ties in the rankings
- Explain the upper and lower limit for the coefficient
- List the remarks on rank correlation coefficient
- List the advantage, application and limitation of rank correlation coefficient

Let us consider a group of n individuals arranged in order of merit or proficiency in possession of two characteristics A and B. In general, these ranks in the two characteristics will be different.

For example, if we consider the relation between intelligence and beauty, it is not necessary that a beautiful individual is intelligent also.

Let us consider xi yi where, i is equal to 1, 2 up to n be the ranks of the i^{th} individual in two characteristics A and B respectively.

Pearsonian coefficient of correlation between ranks xi's and yi's is called rank correlation coefficient between A and B for that group of individuals.

Assuming that no two individuals are bracketed equal in either classification, each of the variables X and Y takes the values 1, 2, up to n.
Hence, x bar is equal to y bar is equal to 1 by n into 1 plus 2 plus 3 plus up to plus n is equal to n plus 1 divided by 2.
Sigma x square is equal to 1 by n into summation i ranges from 1 to n x i square minus x bar square is equal to 1 by n into 1 square plus 2 square plus up to plus n square minus n plus 1 by 2 the whole square.
Is equal to n into n plus 1 into 2 into n plus 1 divided by 6 into n minus n plus 1 divided by 2 the whole square.
Therefore, sigma x square is equal to n square minus 1 whole divided by 12.

In general, xi is not equal to yi. Let d i is equal to x i minus y i
Therefore, d i is equal to xi minus x bar minus yi minus y bar.
Squaring and summing over i from 1 to n we get,
Summation over i is equal to 1 to n d i square is equal to summation i is equal 1 to n xi minus x bar minus yi minus y bar whole square.
Is equal to summation over i, xi minus x bar whole square plus summation over i, yi minus y bar whole square minus 2 into summation over i, xi minus x bar into yi minus y bar.
Dividing both sides by n, we get,

1 by n into summation over i, d i square is equal to sigma x square plus sigma y square minus 2 into covariance of x, y
Is equal to sigma x square plus sigma y square minus 2 into row into sigma x into sigma y.

Where, row is the rank correlation coefficient between A and B.
Therefore, 1 by n into summation d i square is equal to 2 into sigma x square minus 2 into row into sigma x square, since sigma x square is same as sigma y square.
Implies, 1 minus row is equal to summation over i is equal to 1 to n d i square divided by 2 into n into sigma X square.
Implies, row is equal to 1 minus 6 summation over i is equal to 1 to n d i square divided by n into n square minus 1 by substituting for sigma x square, which is the spearman’s formula for the rank correlation coefficient.

2. Rank Correlation – Part 1

If some of the individuals receive the same rank in a ranking of merit, they are said to be tied. Let us consider m individuals say k plus 1th, k plus 2^{th} up to k plus m^{th}, are tied. Then, each of these m individuals are assigned a common rank, which is the arithmetic mean of the ranks k plus 1, k plus 2 up to k plus m.

Now, let us derive the formula for row of x y.
We have row of x y is equal to summation X minus X bar into Y minus Y bar divided by summation X minus X bar whole square into summation Y minus Y bar whole square to the whole power half is equal to summation x into y divided by summation x square into summation y square
Where, x is equal to X minus X bar and y is equal to Y minus Y bar.
If X and Y each takes values 1, 2, up to n, then X bar is equal to Y bar is equal to 1 by n into 1 plus 2 plus 3 plus up to plus n is equal to n plus 1 divided by 2
And n into sigma X square is equal to summation x square is equal to n into n square minus 1 divided by 12.
And n into sigma Y square is equal to summation y square is equal to n into n square minus 1 divided by 12.

In addition, summation d square is equal to summation X minus Y whole square is equal to summation X minus X bar minus Y minus Y bar whole square is equal to summation x minus y the whole square.
Implies summation d square is equal to summation x square plus summation y square minus 2 into summation x into y.
Implies summation x into y is equal to half into summation x square plus summation y square minus summation d square.

We shall now investigate the effect of common ranking (in case of ties), on the sum of squares of the ranks. Let S square and S 1 square denote the sum of squares of untied and tied ranks respectively. Then we have,
S square is equal to k plus 1 square plus k plus 2 square plus up to plus k plus m square.
Is equal to m into k square plus 1 square plus 2 square plus up to plus m square plus 2 into k into 1 plus 2 plus up to plus m
By substituting for sum of m natural numbers and sum of squares of m natural numbers, we get,
M into k square plus m into m plus 1 into 2m plus 1 whole divided by 6 plus km into m plus 1.

S1 square is equal to m into average rank square
Is equal to m into k plus 1 plus k plus 2 plus up to plus k plus m whole square
Is equal to m into k plus m plus 1 divided by 2 whole square
Is equal to m into k square plus m into m plus 1 square divided by 4 plus mk into m plus 1.
Therefore, S square minus S1 square is equal to m into m plus 1 whole divided by 12 into 2 into 2 m plus 1 minus 3 into m plus 1 is equal to m into m square minus 1 divided by 12.

Thus, the effect of typing m individuals (ranks) is to reduce the sum of squares by m into m square minus 1 whole divided by 12, though the mean value of the ranks remains the same namely, n plus 1 by 2.

Suppose there are s such sets of ranks to be tied in the X –series, so that the total sum of squares due to them is,
1 by 12 into summation over i is equal to 1 to s, m i into m i square minus 1 is equal to 1 by 12 into summation over i, m i cube minus m i is equal to T X.

Similarly, there are t such sets of ranks to be tied with respect to the other series Y, so that the sum of squares due to them is,
1 by 12 into summation over j from 1 to t, m j dash into m j dash square minus 1 is equal to 1 by 12 into summation over j, m j dash cube minus m j dash is equal to T Y.
Thus, in case of ties, the new sum of squares are given by,
n into variance dash of X is equal to summation x square minus T X is equal to n into n square minus 1 divided by 12 minus T X and
n into variance dash of Y is equal to summation y square minus T Y is equal to n into n square minus 1 divided by 12 minus Y Y.

And covariance dash of X Y is equal to half into summation x square minus T X plus summation y square minus T Y minus summation d square
Is equal to half into n into n square minus 1 divided by 12 minus T X plus n into n square minus one divided by 12 minus T Y minus summation d square.
Is equal to n into n square minus 1 divided by 12 minus half into T X plus T Y plus summation d square.

Hence, row X Y is equal to n into n square minus 1 by 12 minus half into T X plus T Y plus summation d square divided by
n into n square minus 1 by 12 minus T X whole power half into n into n square minus 1 by 12 minus T Y whole power half
is equal to n into n square minus 1 by 6 minus T X plus T Y plus summation d square divided by
n into n square minus 1 by 6 minus 2 into T X whole power half into n into n square minus 1 by 6 minus 2 into T Y whole power half.

Note that in the above expression, if we adjust only the covariance term, that is summation x into y and not the variances sigma x square or summation x square and sigma y square or summation y square for ties, then the formula of row of X Y reduces to
Row X Y is equal to 1 minus 6 into summation d square plus T X plus T Y divided by n into n square minus 1.

If two or more individuals have same values in the series, common ranks are given to them. This common rank is the average of the ranks, where these items would have been assumed if they were slightly different from each other. The next item will get the rank next to the ranks already assumed.

3. Limits for the Rank Correlation Coefficient

Now, let us find the limits for the rank correlation coefficient.
Spearman’s Rank correlation coefficient is given by
Row is equal to 1 minus 6 into summation over i is equal to 1 to n d i square divided by n into n square minus 1

Row is maximum if summation over i is equal to 1 to n d i square is minimum. That is if each of the deviation d i is minimum.
However, the minimum value of d i is zero in the particular case x i is equal to y i that is if the ranks of the i^{th} individual in the two characteristics are equal. Hence, the maximum value of row is plus 1, that is row is less than or equal to 1.

Row is minimum if summation over i from 1 to n d i square is maximum. That is if each of the deviation d i is maximum, if the ranks of the individuals in the two characteristics are in the opposite direction as follows.

That is X takes ranks, 1, 2, 3, up to n minus 1 and n
Y take ranks n, n minus 1, n minus 2 up to 2 and 1.
Here, we consider 2 cases when n is odd and n is even.
Consider the first case, suppose n is odd, that is n is equal to 2 into m plus 1, then the values of d are,
2 into m, 2 into m minus 2, 2 into m minus 4 up to 2, zero, minus 2, minus 4 up to minus of 2 into m minus 2, minus 2 into m.
Therefore, summation over i is equal to 1 to n d i square is equal to 2 into 2 m whole square plus 2 m minus 2 whole square plus 4 square plus 2 square.
If we take 2 square as common term, then we get,
8 into m square plus m minus 1 square plus up to plus 2 square plus 1 square
Now, by finding the sum of squares of first m natural numbers, we get
8 into m into m plus 1 into 2 m plus 1 whole divided by 6.

Hence, row is equal to 1 minus 6 into summation over i, d i square divided by n into n square minus 1.
Is equal to 1 minus 8 into m into m plus 1 into 2 m plus 1 divided by 2 m plus 1 into 2 m plus 1 square minus 1
On simplification, we get,
1 minus 8 into m into m plus 1 divided by 4 into m into m plus 1 is equal to minus 1.

Now, let us consider the second case that is n is even. That is n is equal to 2 into m. Then, the values of d are 2 into m minus 1, 2 into m minus 3 up to 1, minus 1, minus 3, up to 2 into m minus 3, 2 into m minus 1.
Therefore, summation over i from 1 to n d i square is equal to 2 into 2 m minus 1 square plus 2 m minus 3 square plus up to plus 3 square plus 1 square.
Since, we have sum of squares of odd numbers, we add and subtract sum of squares of even

numbers. Whatever we add, we write in between the odd numbers and whatever we subtract, we write separately at the end. Hence we get,
2 into 2m whole square plus 2 m minus 1 whole square plus 2 m minus 2 whole square plus up to 3 square plus 2 square plus 1 square minus,
2 m square plus 2 m minus 2 whole square plus up to plus 4 square plus 2 square.
Which is same as, 2 into 1 square plus 2 square plus 3 square plus up to plus 2 m square minus 2 square into m square plus 2 square into m minus 1 square plus up to plus 2 square.

Now, writing for sum of square of first 2m and m natural numbers, we get,
2 into 2 m minus 2 m plus 1 into 4 m plus 1 divided by 6 minus 4 into m into m plus 1 into 2 m plus 1 divided by 6
By taking 2 m by 6 outside we get,
2 into m divided by 3 into 2 m plus 1 into 4 m plus 1 minus 2 into m plus 1 into 2 m plus 1.
Taking 2 m plus 1 outside and simplifying, we get,
2 into m divided by 3 into 2 m plus 1 into 2 m minus 1
Is equal to 2 into m into 4 m square minus 1 whole divided by 3.

Therefore, the expressions of row becomes
Row is equal to 1 minus 6 into summation over i is equal to 1 to n d i square divided by n into n square minus 1 is equal to 1 minus 4m into 4 m square minus 1 divided by 2m into 4 m square minus 1 is equal to minus 1.
Thus, the limits for rank correlation coefficient are given by minus 1 less than or equal to row less than or equal to plus 1.

4. Spearman's Rank Correlation Coefficient

Now, let us see some remarks on spearman's rank correlation coefficient.

- Summation d is equal to summation x minus summation y is equal to zero, which provides a check for numerical calculations.
- Since Spearman's rank correlation coefficient row is same as Pearsonian correlation coefficient between the ranks, it can be interpreted in the same way as the Karl Pearson's correlation coefficient.
- Karl Pearson's correlation coefficient assumes that the parent population from which sample observations are drawn is normal. If this assumption is violated, then we need a measure, which is distribution free or non-parametric.

  A distribution free measure is one, which does not make any assumptions about the parameters of the population. Spearman's row is such a measure, since no assumptions are made about the form of the population from which sample observations are drawn.

Advantage:

- Spearman's formula is easy to understand and can be applied as compared with product moment formula. The value obtained by the two formulae, viz., Pearsonian r and spearman's row, are generally different.

  The differences arise due to the fact that when ranking is used instead of full set of observations, there is always some loss of information. Unless many ties exist, the coefficient of rank correlation should be slightly lower than the Pearsonian Coefficient.

Application:

- Spearman's formula is the only formula used for finding correlation coefficient if we are dealing with qualitative characteristics, which cannot be measured quantitatively but can be arranged serially. It can also be used where actual data are given. In case of extreme observations, Spearman's formula is preferred to Pearson's formula.

5. Limitations of Spearman's Formula

Limitation:

- Spearman's formula has its own limitation.

  It is not practicable in the case of bivariate frequency distribution (correlation table). For n greater than 30, this formula should not be used unless the ranks are given, since in the contrary case the calculations are quite time-consuming.

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

- The meaning of Spearman’s rank correlation coefficient
- The formula for calculating the correlation when there are ties and no ties in the rankings
- The upper and lower limit for the coefficient
- The remarks, advantage, use, limitation of the rank correlation coefficient