Rank Correlation

In calculating the Pearson's correlation coefficient, it is necessary that the data should be definitely measured. Sometimes exact magnitude of data cannot be determined, but they can be ranked according to some criterion.

For example, a teacher may want to know how far a boy's ability in sports can determine his popularity. But it is extremely difficult to give a numerical value either of this qualities. However, it is easy to place the boys in order for each of this qualities and to measure the association between the two rankings.

Sometimes one variable may be in ordinal measure and the other is in either ratio or interval data.

When the data are not available in the form of numerical measures but it is possible to assign ranks to the data, then nonparametric procedures can be applied in correlation analysis.

When two variables can be ranked separately in ordered series, it is possible to compute a rank correlation coefficient. This rank correlation coefficient is a measure of the degree of association between two sets of ranks.

Different types of rank coefficients can be applied to the data. Spearman rank order correlation coefficient is the most widely used measure of rank correlation.

The situations when rank correlation method is applicable:

- Rank correlation is applicable to find the strength or degree of association of data that are expressed in ranks or scores, i.e. the data expressed in ordered series.
- Allows us to correlate between two sets of qualitative observations, which can be ranked or ordered.
- Can be used as an alternative in finding the degree of association for two sets of data originally given in measurements.
- Can be useful where there are extreme observations in the original numerical valued data.

Spearman's Rank Correlation

Spearman's rank correlation is the nonparametric version of the Pearson product-moment correlation. When the assumptions of the Pearson correlation is violated, then Spearman's correlation can be used.

Spearman's correlation coefficient measures the strength and direction of association between two ranked variables.

Spearman's rank correlation does not carry any assumptions about the distribution of the data and is the appropriate correlation analysis when the variables are measured on an ordinal scale of measurement. This was developed by Charles Edward Spearman in 1904.

Assumptions of Spearman's Rank Correlation

- The measurement scale is at least ordinal.
- There are two variables.
- The scores in one variable must be monotonically related with the other variable.
- Each criterion under study is ranked separately on each variable.

Formula and Interpretation of Rank Correlation

Spearman rank correlation coefficient is equivalent to Pearson correlation coefficient on ranks.

Spearman rank correlation coefficient can also be computed by:

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

where,

r_s = Spearman rank correlation coefficient

d = Difference between the ranks for each pair

n = the number of paired observations

Interpretation: The rank correlation coefficient can be interpreted in the same way as Karl Pearson's correlation coefficient.

Computation of Spearman Rank Correlation Coefficient

In the rank correlation, we may have three types of situations

- When ranks are given
- When ranks are not given
- When repeated ranks occur

Computation of Rank Correlation Coefficient

When actual ranks are given to us, then the steps required for computing Spearman's rank correlation are:

- Take the difference of the two ranks and denote these differences by *d.*
- Square these differences and obtain the total.
- Apply the formula.

When we are given the actual data and not the ranks, it will be necessary to assign the ranks. Ranks can be assigned by taking either highest value as 1 or the lowest value as 1. But whether we start with the lowest value or the highest value we must follow the same method in case of both variables.

Dealing with Tied Ranks

Sometimes more than one item is given the same rank. In this case new ranks are allocated to the items that have tied ranks. In such cases the items are given the average of the ranks they would have received.

For example, if two observations are placed in the 5^{th} place , they are given the rank $[5 + 6]/2 = 5.5$ each, which is common rank to be assigned and the next rank will be 7; and if three observations are placed in the 5^{th} place, they are given the rank $[5 + 6 + 7]/3 = 6$ each, which is common rank to be assigned and the next assigned rank will be 8.

Example 1

A survey company evaluated the top internet companies and their reputations. The following table shows how 10 internet companies ranked in terms of reputation and desirable purchase.

Internet Company	Reputation, x	Probable purchase, y	$d = x - y$	d^2
Microsoft	1	3	-2	4
Intel	2	4	-2	4
Dell	3	1	2	4
Lucent	4	2	2	4
Texas Instruments	5	9	-4	16
Cisco Systems	6	5	1	1
Hewlett-Packard	7	10	-3	9
IBM	8	6	2	4
Motorola	9	7	2	4
Yahoo	10	8	2	4

$r_s = 1 - \frac{6\sum_{i=1}^{n} {d_i}^2}{n(n^2-1)}$, $\sum d^2 = 54$

$= 1 - \frac{6*54}{10(10^2-1)}$ $\quad n = 10$

$= 0.673$

Comment: There is moderate positive correlation between company's reputation and desirable purchase.

Example 2

The relationship between the age of online shoppers and the number of minutes spent in browsing on the internet can be measured from the following sample of 9 internet shoppers.

Age, x	Browsing time, y	R_x	R_y	$d = R_x - R_y$	d^2
32	257	7	5	2	4
38	185	9	2	7	49
33	241	8	3	5	25
28	342	6	6.5	-0.5	0.25
21	342	3.5	6.5	-3	9
22	141	5	1	4	16
19	583	2	9	-7	49
17	394	1	8	-7	49
21	249	3.5	4	-0.5	0.25
					$\sum d^2 = 201.5$

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

$= 1 - \frac{6*201.5}{9(9^2-1)}$ $\quad n = 9$

$= -0.679$

Comment: There is moderate negative correlation between age of the internet shopper and the minutes spent in browsing.

Merits and Demerits of Spearman Rank Correlation:

Merits

- Can be used as a measure of degree of association between qualitative data.
- Simple and easily understandable.
- Less sensitive to bias due to the effect of outliers.
- Does not require assumption of normality.

Demerits

- It is only an approximate measure as the actual values are not used for calculations.
- Calculations may become tedious when a large number of observation are given.
- Cannot be used for finding out correlation in a grouped frequency distribution.

Practice Problems:

Textbook: Statistical Techniques in Business & Economics (LIND MARCHAL WATHEN)

Chapter 16: NONPARAMETRIC METHODS: ANALYSIS OF ORDINAL DATA

Page 611: 26, 27(a)