Module-4
Part – 5
Rank Correlation

Rank Correlation

- **A group of individuals may be arranged in order to merit with respect to some characteristic. The same group would give different orders for different characteristics.**
- **Considering the orders corresponding to two characteristics A and B, the correlation between these n pairs of ranks is called the rank correlation in the characteristics A and B for that group of individuals.**

Let x_i and y_i be the ranks of the i^{th} individuals in A and B respectively. Assuming that no two individuals are bracketed equal in either case, each of the variables taking the values 1, 2, ... , n,

we have, $$\bar{x} = \bar{y} = \frac{1+2+3+...+n}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

If X and Y be the deviations of x and y from their means, then

$$\Sigma X_i^2 = \Sigma\left(x_i - \bar{x}\right)^2 = \Sigma x_i^2 + n\left(\bar{x}\right)^2 - 2\bar{x}\Sigma x_i$$

$$= \Sigma n^2 + \frac{n(n+1)^2}{4} - 2\left(\frac{n+1}{2}\right)\Sigma n$$

$$= \frac{n(n+1)(2n+1)}{6} + \frac{n(n+1)^2}{4} - \frac{n(n+1)^2}{2}$$

$$= \frac{1}{12}\left(n^3 - n\right)$$

Similarly, $\Sigma Y_i^2 = \frac{1}{12}\left(n^3 - n\right)$

Now, let $d_i = x_i - y_i$ so that $d_i = \left(x_i - \overline{x}\right) - \left(y_i - \overline{y}\right) = X_i - Y_i$

$\therefore$ $$\Sigma d_i^2 = \Sigma X_i^2 + \Sigma Y_i^2 - 2\Sigma X_i Y_i$$

$\Rightarrow$ $$\Sigma X_i Y_i = \frac{1}{2}\left[\Sigma X_i^2 + \Sigma Y_i^2 - \Sigma d_i^2\right]$$

$\Rightarrow$ $$\Sigma X_i Y_i = \frac{1}{12}\left(n^3 - n\right) - \frac{1}{2}\Sigma d_i^2$$

Hence the correlation coefficient between these variables is

$$r = \frac{\Sigma X_i Y_i}{\sqrt{\Sigma X_i^2 \Sigma Y_i^2}} = \frac{\frac{1}{12}\left(n^3 - n\right) - \frac{1}{2}\Sigma d_i^2}{\frac{1}{12}\left(n^3 - n\right)}$$

$$\Rightarrow \qquad r = 1 - \frac{6\Sigma d_i^2}{\left(n^3 - n\right)}$$

This formula is called the rank correlation coefficient or Spearman's Rank Correlation Coefficient and is denoted by ρ.

$$\text{i.e.,} \quad \rho = 1 - \frac{6\Sigma d_i^2}{\left(n^3 - n\right)}$$

- Spearman's rank is probably one of the most useful statistical tests that we can do in Geography to prove a relationship between two different sets of data.
- The Spearman rank correlation coefficient, *is the non-parametric* version of the Pearson correlation coefficient. Data must be ordinal, interval or ratio. Spearman's returns a value from –1 to 1, where:

+1 = a perfect positive correlation between ranks

–1 = a perfect negative correlation between ranks

0 = no correlation between ranks.

Note: To assign the rank for the given set of values, order the scores from greatest to smallest; assign the rank 1 to the highest score, 2 to the next highest and so on. If ranks are tied, i.e., Tied ranks are where two items in a column have the same rank then each tied data point assigned a mean rank.

Example 4.4.1: Calculate the rank correlation coefficient and comment on the following data on sunflower:

Height of a sunflower (in cm)	**183**	**134**	**234**	**256**	**190**	**89**	**112**
Width of the stem (in mm)	**21**	**14**	**24**	**32**	**29**	**18**	**20**

We have, the rank correlation coefficient is

$$\rho = 1 - \frac{6\Sigma d_i^2}{\left(n^3 - n\right)}$$

We prepare a relevant table as follows

Height of a sunflower (in cm)	Rank x_i	Width of the stem (in mm)	Rank y_i	$d_i = x_i - y_i$	d_i^2
183	4	21	4	0	0
134	5	14	7	-2	4
234	2	24	3	-1	1
256	1	32	1	0	0
190	3	29	2	1	1
89	7	18	6	1	1
112	6	20	5	1	1
				Σ	8

Here, $n = 7$, $\Sigma d_i^2 = 8$

$$\rho = 1 - \frac{6\Sigma d_i^2}{(n^3 - n)} = 1 - \frac{6(8)}{(7^3 - 7)} = 1 - 0.143 = 0.857$$

As ρ is close to one, we can conclude that the wider the stem to higher the sunflower grows.

Example 4.4.2: The scores for 9 students in Physics and Maths are as follows:

Physics:	**35**	**23**	**47**	**17**	**10**	**43**	**9**	**6**	**28**
Mathematics:	**30**	**33**	**45**	**23**	**8**	**49**	**12**	**4**	**31**

Compute the ranks of students in the two subjects and compute the Spearman's rank correlation.

We have, the rank correlation coefficient is

$$\rho = 1 - \frac{6\Sigma d_i^2}{\left(n^3 - n\right)}$$

We prepare a relevant table as follows

Physics	Rank x_i	Mathematics	Rank y_i	$d_i = x_i - y_i$	d_i^2
35	3	30	5	-2	4
23	5	33	3	2	4
47	1	45	2	-1	1
17	6	23	6	0	0
10	7	8	8	-1	1
43	2	49	1	1	1
9	8	12	7	1	1
6	9	4	9	0	0
28	4	31	4	0	0
				Σ	12

Here, $n = 9$, $\Sigma d_i^2 = 12$

$$\rho = 1 - \frac{6\Sigma d_i^2}{(n^3 - n)} = 1 - \frac{6(12)}{(9^3 - 9)} = 1 - 0.1 = 0.9$$

Example 4.4.3: Ten participants in a contest are ranked by two judges as follows:

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

Calculate the rank correlation coefficient.

We have, the rank correlation coefficient is

$$\rho = 1 - \frac{6\Sigma d_i^2}{\left(n^3 - n\right)}$$

We prepare a relevant table as follows

Rank x_i	Rank y_i	$d_i = x_i - y_i$	d_i^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
		Σ	60

Here, $n = 10$, $\Sigma d_i^2 = 60$

$$\rho = 1 - \frac{6\Sigma d_i^2}{(n^3 - n)} = 1 - \frac{6(60)}{(10^3 - 10)} = 1 - 0.36 = 0.64$$

Example 4.4.4: Three judges A, B and C give the following ranks. Find which pair of judges has common approach

A	1	6	5	10	3	2	4	9	7	8
B	3	5	8	4	7	10	2	1	6	9
C	6	4	9	8	1	2	3	10	5	7

We have, the rank correlation coefficient is

$$\rho = 1 - \frac{6\Sigma d_i^2}{\left(n^3 - n\right)}$$

We prepare a relevant table as follows

Ranks by A x_i	Ranks by B y_i	Ranks by C z_i	d_1 $x_i - y_i$	d_2 $y_i - z_i$	d_3 $z_i - x_i$	d_1^2	d_2^2	d_3^2
1	3	6	-2	-3	5	4	9	25
6	5	4	1	1	-2	1	1	4
5	8	9	-3	-1	4	9	1	16
10	4	8	6	-4	-2	36	16	4
3	7	1	-4	6	-2	16	36	4
2	10	2	-8	8	0	64	64	0
4	2	3	2	-1	-1	4	1	1
9	1	10	8	-9	1	64	81	1
7	6	5	1	1	-2	1	1	4
8	9	7	-1	2	1	1	4	1
					Σ	200	214	60

Here, $n = 10$

$$\rho(x,y) = 1 - \frac{6\Sigma d_1^2}{(n^3 - n)} = 1 - \frac{6(200)}{(10^3 - 10)} = -0.2$$

$$\rho(y,z)=1-\frac{6\Sigma d_2{}^2}{\left(n^3-n\right)}=1-\frac{6(214)}{\left(10^3-10\right)}=-0.3$$

$$\rho(z,x)=1-\frac{6\Sigma d_3{}^2}{\left(n^3-n\right)}=1-\frac{6(60)}{\left(10^3-10\right)}=0.6$$

Since $\rho(z,x)$ is maximum, the pair of judges A and C have the nearest common approach.

EXERCISE

1. Find the rank correlation for the following data:

x	56	42	72	36	63	47	55	49	38	42	68	60
y	147	125	160	118	149	128	150	145	115	140	152	155

2. Calculate the rank correlation coefficient from the following data showing ranks of 10 students in two subjects:

Maths:	3	8	9	2	7	10	4	6	1	5
Physics:	5	9	10	1	8	7	3	4	2	6

3. Find the rank correlation coefficient for the following data

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

(VTU 2018)