4e – A Guide to
Spearman's Rank

Royal Geographical Society
with IBG

Advancing geography
and geographical learning

The **Spearman's Rank Correlation Coefficient** is a statistical test that examines the degree to which two data sets are correlated, if at all.

Why would we use Spearman's Rank?

While a scatter graph of the two data sets may give the researcher a hint about whether the two have a correlation, Spearman's Rank gives the researcher a numerical value on the degree of correlation, or indeed, the degree of non-correlation. It is a relatively straightforward analysis for those researchers who are not wholly confident in their mathematical skills, it is an appropriate technique to use.

In order to use Spearman's Rank the researcher must have paired sets of data that are in some way related (such as the geographical site where they were collected in the field). It is a good idea to have at least ten pairs of data to use: any fewer and the result will be highly insignificant and more likely be the result of chance than of true correlation.

Worked Example:

In the following example, the researcher is looking at whether for River X, the channel width increases as the distance from the source increases. Theoretically, this should be true, but the Spearman's Rank analysis will tell the researcher whether it is true in this case that there is a correlation and the strength of any such correlation. The following provides a step-by-step guide to the application of the Spearman's Rank test to this data.

1. The researcher should arrange the paired data in a table to allow for ease of analysis. This can be done in a spreadsheet or through handwritten methods.

Site	Distance from source (m)	Rank (R_1)	Width (m)	Rank (R_2)	d $(R_1 - R_2)$	d^2
1	150		0.40			
2	300		0.80			
3	450		1.00			
4	600		0.95			
5	750		1.20			
6	900		1.10			
7	1050		1.30			
8	1200		1.40			
9	1350		1.85			
10	1500		2.40			
11	1650		2.55			
12	1800		3.20			
13	1950		3.80			
14	2100		3.60			
15	2250		3.20			
					Total	

2. Then researcher should rank each data, starting with 1 as the smallest figure and (in this case) 15 as the largest. Where there might be two values that are equal, (known as tied ranks) the researcher should average the ranks and omit the ranking values they cover. For example, a set of rankings may read: 1, 2, 3.5, 3.5, 5, 6, 7.25, 7.25, 7.25, 7.25, 11, 12, etc.

Site	Distance from source (m)	Rank (R_1)	Width (m)	Rank (R_2)	d $(R_1 - R_2)$	d^2
1	150	1	0.40	1		
2	300	2	0.80	2		
3	450	3	1.00	4		
4	600	4	0.95	3		
5	750	5	1.20	6		
6	900	6	1.10	5		
7	1050	7	1.30	7		
8	1200	8	1.40	8		
9	1350	9	1.85	9		
10	1500	10	2.40	10		
11	1650	11	2.55	11		
12	1800	12	3.20	12.5		
13	1950	13	3.80	15		
14	2100	14	3.60	14		
15	2250	15	3.20	12.5		
					Total	

3. The difference (d) between the two ranks should then be calculated by subtracting R_1 from R_2:

Site	Distance from source (m)	Rank (R_1)	Width (m)	Rank (R_2)	d ($R_1 - R_2$)	d^2
1	150	1	0.40	1	0	
2	300	2	0.80	2	0	
3	450	3	1.00	4	-1	
4	600	4	0.95	3	1	
5	750	5	1.20	6	-1	
6	900	6	1.10	5	1	
7	1050	7	1.30	7	0	
8	1200	8	1.40	8	0	
9	1350	9	1.85	9	0	
10	1500	10	2.40	10	0	
11	1650	11	2.55	11	0	
12	1800	12	3.20	12.5	-0.5	
13	1950	13	3.80	15	-2	
14	2100	14	3.60	14	0	
15	2250	15	3.20	12.5	2.5	
					Total	

4. d should then be squared to remove any negative values. The total value of all the d^2 can also be calculated at this stage.

Site	Distance from source (m)	Rank (R_1)	Width (m)	Rank (R_2)	d ($R_1 - R_2$)	d^2
1	150	1	0.40	1	0	0
2	300	2	0.80	2	0	0
3	450	3	1.00	4	-1	0
4	600	4	0.95	3	1	1
5	750	5	1.20	6	-1	0
6	900	6	1.10	5	1	1
7	1050	7	1.30	7	0	0
8	1200	8	1.40	8	0	0
9	1350	9	1.85	9	0	0
10	1500	10	2.40	10	0	0
11	1650	11	2.55	11	0	0
12	1800	12	3.20	12.5	-0.5	0.25
13	1950	13	3.80	15	-2	4
14	2100	14	3.60	14	0	0
15	2250	15	3.20	12.5	2.5	6.25
					Total	12.5

5. One should then apply the Spearman's Rank equation to calculate the coefficient value (R) (the value that tells the researcher the strength of the correlation).

$$R \ = \ 1 \ - \ \frac{6\Sigma d^2}{n\,(n^2-n)}$$

where n is the number of pairs of data collected and used (in this case 15). The sum of the d^2 values (Σd^2) in this example is 12.5.

Therefore, the equation can be calculated as follows:

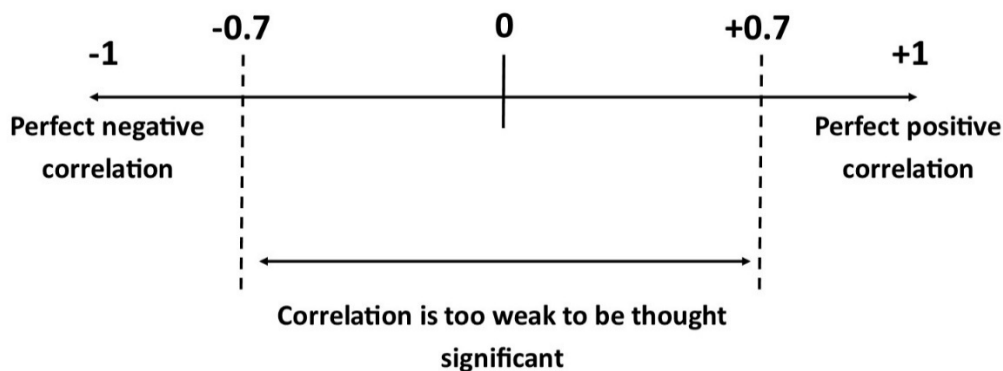$$R = 1 - \frac{6 \times 12.5}{15\,(15^2 - 15)} \qquad R = 1 - \frac{75}{15\,(225 - 15)} \qquad R = 1 - \frac{75}{15 \times 210}$$

$$R = 1 - \frac{75}{3150} \qquad\qquad R = 1 - 0.024 \qquad\qquad R = 0.976$$

The coefficient (R) will be between a value of -1 and +1 where -1 indicates a perfect negative correlation and +1 indicates a perfect positive correlation. A value of between -0.7 to +0.7 is generally seen as being too weak to be thought of as a significant result.

Therefore, the data in this example - with a Spearman's Rank result of 0.976 - shows a strong positive correlation between channel width and distance from the source.

6. To check whether the result is meaningful or just down to chance, the value for R can be compared with the critical value for n in the Spearman's Rank significance table.

Below is the significance table for some values of n, but for analysis of larger sets of data, extended significance tables can be found online.

n	Significance Level		
	0.1	0.05	0.01
4	1.000	1.000	1.000
5	0.700	0.900	1.000
6	0.657	0.771	0.943
7	0.571	0.679	0.857
8	0.548	0.643	0.810
9	0.483	0.600	0.767
10	0.442	0.564	0.733
11	0.418	0.527	0.700
12	0.399	0.504	0.671
13	0.379	0.478	0.648
14	0.367	0.459	0.622
15	0.350	0.443	0.600
16	0.338	0.427	0.582
17	0.327	0.412	0.558
18	0.317	0.400	0.543

The critical value for this example, where there are 15 pairs of data (= 15), is 0.443. As the value of R is greater than the critical value, we can say with 95% certainty that the results we have observed have not occurred by chance. This means the results are highly significant and sound conclusions can be drawn from them.

Note: There are ready-made spreadsheets available online to do this.