Spearman's Rank Correlation Coefficient

Spearman's Rank Correlation Coefficient is used to measure the nature and strength of any relationship between two sets of data. The test uses rankings of pairs of data to see how closely the data 'fits' with a perfect correlation line on a graph.

Note: You can use the **Spearman's Rank** for

- data that is paired together
- data sets of five pairs of samples

Worked example:

Bedload was collected from different sites along the length of a river. At each site, the distance from the source was noted along with the average size of the bedload's longest length at that point. The researcher wanted to see if there was a correlation between bedload size and distance from the source of the river. They devised the following hypotheses to be tested through the Spearman's Rank Correlation Coefficient:

H_1 There is a correlation between the size of the bedload and the distance from the source of the river.

H_0 There is no correlation between the size of the bedload found and the distance from the source of the river.

Distance (m)	Rank (R_1)	Size of bedload (mm)	Rank (R_2)	Difference between ranks (d)	d^2
10	1	112	8	-7	49
250	2	86	6	-4	16
840	3	91	7	-4	16
1570	4	45	5	-1	1
2110	5	34	4	1	1
2750	6	10	2.5	3.5	12.25
3540	7	10	2.5	4.5	20.25
4890	8	3	1	7	49

All the data is tabulated in its paired form. Each variable is then given a rank. Both sets of variables should be ranked from lowest value to highest.

If more than one sample has the same value, they share a rank.

The correlation coefficient value (r_s) is then calculated:

$$r_s = 1 - \frac{6 \Sigma d^2}{n^3 - n}$$

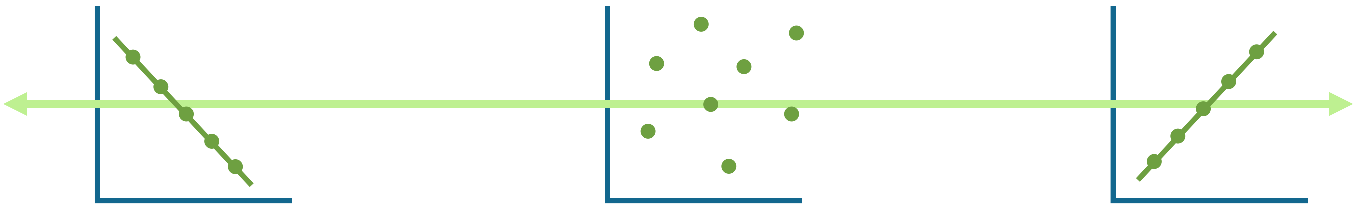
n = number of paired samples

Σd^2 = total of squared differences

$$r_s = 1 - \frac{6 \times 164.5}{8^3 - 8} \qquad r_s = 1 - \frac{987}{504} \qquad r_s = 1 - 1.96 \qquad r_s = -0.97$$

The negative calculated value of the correlation coefficient indicates a negative correlation between the two variables. Therefore it appears that as the distance from the source increases, the size of the bedload decreases.

The calculated value also tells the researcher about the strength of the correlation. In this case, the calculated value of - 0.97 indicates a strong negative correlation.



A calculated value of between -0.7 and +0.7 is considered too weak to be thought a significant result.

It is important to also check the significance of the result by using a significance table (below). The calculated value needs to be compared with a critical value for the appropriate number of samples taken. This tells the researcher whether the correlation simply occurred by chance or is likely to be the same if repeated.

	Level of significance			
n	0.05	0.025	0.01	0.005
4	1.000	1.000	1.000	1.000
5	0.900	0.900	1.000	1.000
6	0.771	0.829	0.943	0.943
7	0.679	0.786	0.857	0.893
8	0.643	0.738	0.890	0.857
9	0.600	0.683	0.767	0.817
10	0.564	0.649	0.733	0.782
11	0.528	0.609	0.700	0.755
12	0.504	0.587	0.671	0.727
13	0.478	0.560	0.648	0.698
14	0.459	0.539	0.622	0.675
15	0.443	0.518	0.600	0.654
16	0.427	0.503	0.582	0.632
17	0.412	0.482	0.558	0.606
18	0.400	0.468	0.543	0.590
19	0.389	0.456	0.529	0.575
20	0.378	0.444	0.516	0.561

The level of significance means the level to which one can be sure that the results are meaningful and did not occur just by chance. In this test, a 95% (or 0.05) level is appropriate.

From the significance table we can read off a critical value of **0.643**. As the **calculated value** (regardless of correlation direction) **is greater than the critical value, the null hypothesis can be rejected**.

Therefore, the is a significant and strong negative correlation between distance from source and bedload size for the river in question.