

Spearman's Rank Correlation Coefficient (SRCC) is a statistical test used to see if there is a significant relationship, or correlation between two data sets. The test also measures how strong any correlation is and its polarity (positively or negatively correlated).

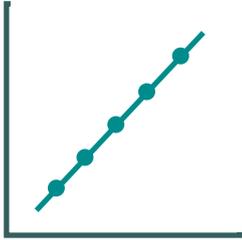


Figure 1: A perfect positive correlation

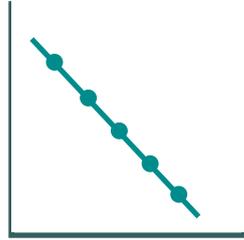


Figure 2: A perfect negative correlation

Many geographical theories relate to clear correlations between different variables, and are represented by straight line graphs (Figure 1 and 2). However, geographical data collected in the field rarely demonstrates such perfect correlations and SRCC allows geographers to see to what extent real life geography mirrors theoretical models. Before starting a SRCC, it is worth keeping in mind the theoretical correlation one might expect to see from the results.

SRCC can only be used for paired data. This means two bits of data that are relatable to each other through a factor such as time, location, or individual. It is also worth noting that SRCC only works if there are more than five pairs of data available.

Here are some examples of paired data one might come across in geographical models and the kind of correlation one might like to study:

- Rivers - the flow **velocity** of the river and how it changes with **distance** from source.
- Coasts - the size (**length**) of beach particles and how they vary with **angle** of slope.
- Weather - how **temperature** changes in relation to air **pressure**.
- Urban studies - how the **age** of individuals influences the **frequency** with which they use a space
- Economics - the **percentage** of people with more than one car and average household **income** in different areas
- Environmental studies - the average **life expectancy** and how it changes with **pollutant level**
- Ecology - average **height** of a particular tree species and the influence of soil **pH**
- Development studies - **literacy rate** and how it changes with **GDP**

How to carry out a Spearman's Rank Correlation Coefficient

For this example, we will look at how people of different ages responded to a questionnaire about a local environment. The geographical researcher wished to find out if there was a relationship between someone's age and how they viewed the environmental quality of a particular area of a local town. The area in question was a former roofed carpark and more recently the ramps and railings found there had offered an alternative recreational uses for a group of skaters. Walls and roadways had been painted by the skaters and their noise echoed around the empty area.

The researcher had a theory that the younger the person questioned, the more favourable they would view this environment. This created the following hypothesis in the mind of the researcher:

“There will be a negative correlation between the age of local people and the environmental quality score they give the Market Road car park”

The researcher collected the following data and displayed it in its paired form:

Age of respondent	34	12	25	36	55	68	18	58	40	42	17	29	34	61	70	45	39	20	14	26
Env. Quality Score (%)	54	30	51	44	33	26	69	21	48	42	74	74	68	34	18	46	52	80	70	62

The researcher put one of the variables in rank order (in this case from smallest to greatest age) and matched the second variable (EQS score) accordingly. If more than one sample share the same value, they share the rank and the next rank on, moves on one numerically. This way if there are twenty values there will be rankings that range from one to twenty.

Age of respondent	12	14	17	18	20	25	26	29	34	34	36	39	40	42	45	55	58	61	68	70
Rank (R ₁)	1	2	3	4	5	6	7	8	9	9	11	12	13	14	15	16	17	18	19	20
Env. Quality Score (%)	30	70	74	69	80	51	62	74	54	68	44	52	48	42	46	33	21	34	26	18

The EQS score was then also ranked.

Age of respondent	12	14	17	18	20	25	26	29	34	34	36	39	40	42	45	55	58	61	68	70
Rank (R ₁)	1	2	3	4	5	6	7	8	9	9	11	12	13	14	15	16	17	18	19	20
Env. Quality Score (%)	30	70	74	69	80	51	62	74	54	68	44	52	48	42	46	33	21	34	26	18
Rank (R ₂)	4	17	18	16	20	11	14	18	13	15	8	12	10	7	9	5	2	6	3	1

The difference (d) between the pairs of ranks was then calculated for each pair.

Rank (R ₁)	1	2	3	4	5	6	7	8	9	9	11	12	13	14	15	16	17	18	19	20
Rank (R ₂)	4	17	18	16	20	11	14	18	13	15	8	12	10	7	9	5	2	6	3	1
d (R ₁ - R ₂)	-3	-15	-15	-12	-15	-5	-7	-10	-4	-6	3	0	3	7	6	11	15	12	16	19

This difference was then squared to remove the negative figures

Rank (R ₁)	1	2	3	4	5	6	7	8	9	9	11	12	13	14	15	16	17	18	19	20
Rank (R ₂)	4	17	18	16	20	11	14	18	13	15	8	12	10	7	9	5	2	6	3	1
d (R ₁ - R ₂)	-3	-15	-15	-12	-15	-5	-7	-10	-4	-6	3	0	3	7	6	11	15	12	16	19
d ²	9	225	225	144	225	25	49	100	16	36	9	0	9	49	36	121	225	144	256	361

This sum of all these d² values was then calculated. In this case, $\sum d^2 = 2264$

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

Using this figure, and the equation to the left, the correlation coefficient value (r_s) was then calculated.

$$r_s = 1 - \frac{6 \times 2264}{20^3 - 20}$$

n responds to the number of paired samples.

$$r_s = 1 - \frac{13584}{8000 - 20}$$

$$r_s = 1 - \frac{13584}{7980}$$

$$r_s = 1 - 1.702$$

$$r_s = -0.702$$

This value is known as the **calculated value**.

The SRCC results indicates a negative correlation and so suggests at this stage that the researcher's hypothesis is correct. However, the calculated value is quite weak: any value of between -0.7 and +0.7 is generally considered too weak to be thought of as a significant result.

Before one can fully accept the hypothesis, a significance test should also be carried out. This tells the researcher the extent to which one can be sure that the results are meaningful and the level to which one can be sure that the results did not occur by chance. To calculate this, a significance table is required. One reads off the critical value from the table and if the **calculated value is greater than the critical value**, the result is deemed to be significant. A full significance table for the SRCC can be downloaded from [the island geographer](http://theislandgeographer.co.uk) site.

In this case, to a 95% significance level, the critical value is 0.378. Therefore, the researcher can accept the hypothesis.