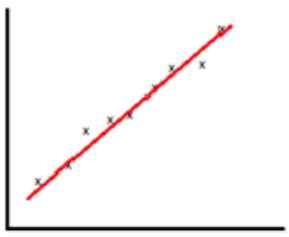
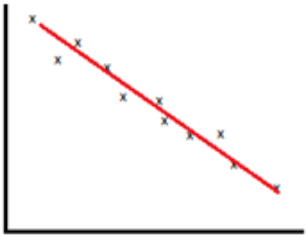
Statistical skills example sheet: Spearman's Rank

Spearman's rank correlation is a statistical test that is carried out in order to assess the degree of **association** between different **measurements** from the same sample. That is, if you are looking for a positive or negative correlation between two variables.



Positive correlation

No correlation



Negative correlation

When the points on a graph clearly fit onto a line of best fit it is easy to determine whether a correlation exists. However, as the points become further placed from each other it is hard to make an accurate judgement. This is where statistics is used; to clarify how confident we are that a correlation exists.

Example:

During monitoring for the MoorLIFE project, an ecologist collected data from an area of moorland that had been restored in 2003. In order to assess whether a correlation existed between the two plant species she studied or whether they were growing independently of one another, she carried out the following statistical test.

The MoorLIFE project is co-funded by the European Union's Life+ programme and delivered by Moors for the Future Partnership.

Partners include: Environment Agency, Natural England, National Trust, Peak District National Park Authority, United Utilities and Yorkshire Water.

Quadrat	% Cover of Bilberry	% Cover Common Heather
1	5	0
2	40	0
3	50	5
4	5	0
5	10	0
6	25	0
7	0	1
8	4	0
9	0	0
10	0	1
11	10	6
12	2	0.5

First of all, form a null hypothesis. This hypothesis assumes there is no relationship between the two variables *e.g. There is no correlation between percentage cover of Bilberry and percentage cover of Common Heather.*

Step one: Rank the data

For each set of data assign ranks from lowest to highest. The lowest value in a column will be given the rank of 1, the next smallest number will be given a 2 etc. If there are tied scores each of those will share the ranks and be given the average (mean) rank.

For example, there are 7 quadrats with a percentage cover of Heather value of 0. As these are the lowest values in that data set, they share the ranks of 1, 2, 3, 4, 5, 6 and 7 (a total of 28) giving each value an average rank of 4 (28÷7). The next rank that can be given is therefore an 8 as ranks 1 to 7 have already been assigned. Similarly in the same data set there are two values for Heather cover of 1%, these two then share the ranks of 8 and 9 (8 +9 = 17, 17÷2=9.5). Each place is assigned the average rank of 9.5 and the next rank to be given out is that of 10.

The MoorLIFE project is co-funded by the European Union's Life+ programme and delivered by Moors for the Future Partnership.

Partners include: Environment Agency, Natural England, National Trust, Peak District National Park Authority, United Utilities and Yorkshire Water.

Quadrat	% Cover of Bilberry	**Rank variable 1**	% Cover Heather	**Rank variable 2**
1	5	**6.5**	0	**4**
2	40	**11**	0	**4**
3	50	**12**	5	**11**
4	5	**6.5**	0	**4**
5	10	**8.5**	0	**4**
6	25	**10**	0	**4**
7	0	**2**	1	**9.5**
8	4	**5**	0	**4**
9	0	**2**	0	**4**
10	0	**2**	1	**9.5**
11	10	**8.5**	6	**12**
12	2	**4**	0.5	**8**

Step two: Find the differences between rank 1 and rank 2 for each value in the table.

Rank variable 1	Rank variable 2	**Difference d**
6.5	4	**2.5**
11	4	**7**
12	11	**1**
6.5	4	**2.5**
8.5	4	**4.5**
10	4	**6**
2	9.5	**-7.5**
5	4	**1**
2	4	**-2**
2	9.5	**-7.5**
8.5	12	**-3.5**
4	8	**-4**

The MoorLIFE project is co-funded by the European Union's Life+ programme and delivered by Moors for the Future Partnership.

Partners include: Environment Agency, Natural England, National Trust, Peak District National Park Authority, United Utilities and Yorkshire Water.

Step three: Square the differences to get rid of negative numbers then find the sum of d^2

Rank variable 1	Rank variable 2	Difference d	d^2
6.5	4	**2.5**	**6.25**
11	4	**7**	**49**
12	11	**1**	**1**
6.5	4	**2.5**	**6.25**
8.5	4	**4.5**	**20.25**
10	4	**6**	**36**
2	9.5	**-7.5**	**56.25**
5	4	**1**	**1**
2	4	**-2**	**4**
2	9.5	**-7.5**	**56.25**
8.5	12	**-3.5**	**12.25**
4	8	**-4**	**16**
		$\sum d^2 =$	**264.5**

The MoorLIFE project is co-funded by the European Union's Life+ programme and delivered by Moors for the Future Partnership.

Partners include: Environment Agency, Natural England, National Trust, Peak District National Park Authority, United Utilities and Yorkshire Water.

Step four: Calculate Spearman's Rank using the formula below, where n is the number of samples in each category.

$$\text{Rank (R)} = 1 - \frac{6\sum d^2}{n(n^2-1)}$$

$$R = 1 - \frac{6 \times 264.5}{12\,(144-1)}$$

$$R = 1 - 0.925$$

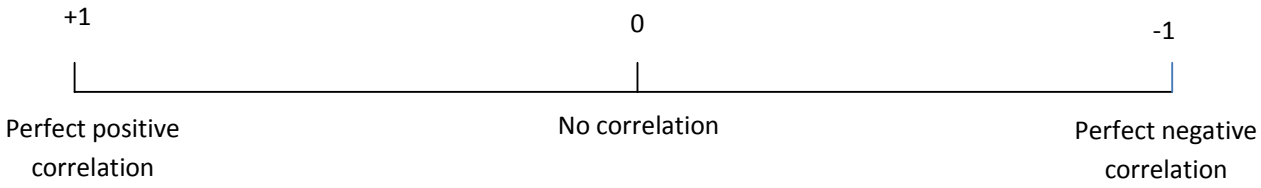
$$\mathbf{R = 0.075}$$

The MoorLIFE project is co-funded by the European Union's Life+ programme and delivered by Moors for the Future Partnership.

Partners include: Environment Agency, Natural England, National Trust, Peak District National Park Authority, United Utilities and Yorkshire Water.

Step five: work out the meaning of the R value

The closer R is to 1 or -1 the more likely the correlation. A perfect positive correlation has an R value of 1, a perfect negative correlation has a value of -1.



If the value lies between -1 and 1 we need to carry out a test for **significance.**

The significance level is calculated by finding the degrees of freedom for your test. This is the number of pairs in your sample minus 2.

In our example this would be:

12-2 = **10 degrees of freedom**

The MoorLIFE project is co-funded by the European Union's Life+ programme and delivered by Moors for the Future Partnership.

Partners include: Environment Agency, Natural England, National Trust, Peak District National Park Authority, United Utilities and Yorkshire Water.

In Biology we work to a significance level of 0.05, that is, at this level we are 95% confident that the results are due to a correlation and not due to chance. Looking at the table below we can see that at a significant level of 0.05, for 10 degrees of freedom the R value must be 0.564 or above for it to be considered a statistically significant correlation. Our result does not reach this value, therefore *the probability of our result being due to chance alone is greater than 5%* and we must accept our null hypothesis, the two variables are not correlated. We can conclude that the percentage cover of Bilberry and Common Heather are not correlated (positively nor negatively) and the two plants therefore grow independently of one another.

Degrees of Freedom	Significance level	
	0.05	**0.01**
4	1.000	
5	0.900	1.000
6	0.829	0.943
7	0.714	0.893
8	0.643	0.833
9	0.600	0.783
10	0.564	0.745
11	0.523	0.736
12	0.497	0.703
13	0.475	0.673
14	0.457	0.646
15	0.441	0.623
16	0.425	0.601
17	0.301	0.582
18	0.399	0.564
19	0.388	0.549
20	0.377	0.534

The MoorLIFE project is co-funded by the European Union's Life+ programme and delivered by Moors for the Future Partnership.

Partners include: Environment Agency, Natural England, National Trust, Peak District National Park Authority, United Utilities and Yorkshire Water.

Exercise: Results analysis

Following restoration of some Blanket Bog sites on Bleaklow by the MoorLIFE team, several areas of Blanket Bog were monitored in order to measure the success of the process. The table shows an extract from the results table for an area of Blanket Bog, (Joseph's Patch), which was restored in 2003.

Quadrat_ID	Percentage cover of D. flexuosa	Rank 1	Percentage cover of Agrostis sp	Rank 2
JP001	30		15	
JP003	35		1	
JP004	15		5	
JP005	15		10	
JP006	15		2	
JP007	10		10	
JP007b	3		3	
JP008	5		1	
JP009	10		5	
JP010	25		0.25	
JP011	35		30	
JP012	50		20	
JP013	50		10	
JP014	30		15	
JP015	40		15	
JP016	50		5	
JP017	30		4	
JP018	15		10	
JP019	40		7	
JP020b	35		15	
JP021	30		30	
JP022	50		5	
JP023	15		10	
JP024	40		10	
JP025	40		12	
JP026	10		4	

www.moorsforthefuture.org.uk/moorlife

The MoorLIFE project is co-funded by the European Union's Life+ programme and delivered by Moors for the Future Partnership.

Partners include: Environment Agency, Natural England, National Trust, Peak District National Park Authority, United Utilities and Yorkshire Water.

1. The team ecologist wanted to find out if there was any relationship between *D.flexuosa* and the *Agrostis species*. The null hypothesis was that *there is no correlation between the percentage cover of D.flexuosa and Agrostis species*. Complete the table by ranking the two sets of data.
2. The data above has an R value of **0.41**. Work out how many degrees of freedom there are for this data set and then use the significance table to comment upon the significance of the results. Remember to include the terms *chance* and *probability* in your answer.

The MoorLIFE project is co-funded by the European Union's Life+ programme and delivered by Moors for the Future Partnership.

Partners include: Environment Agency, Natural England, National Trust, Peak District National Park Authority, United Utilities and Yorkshire Water.