

Ocean Acidification Lesson 2

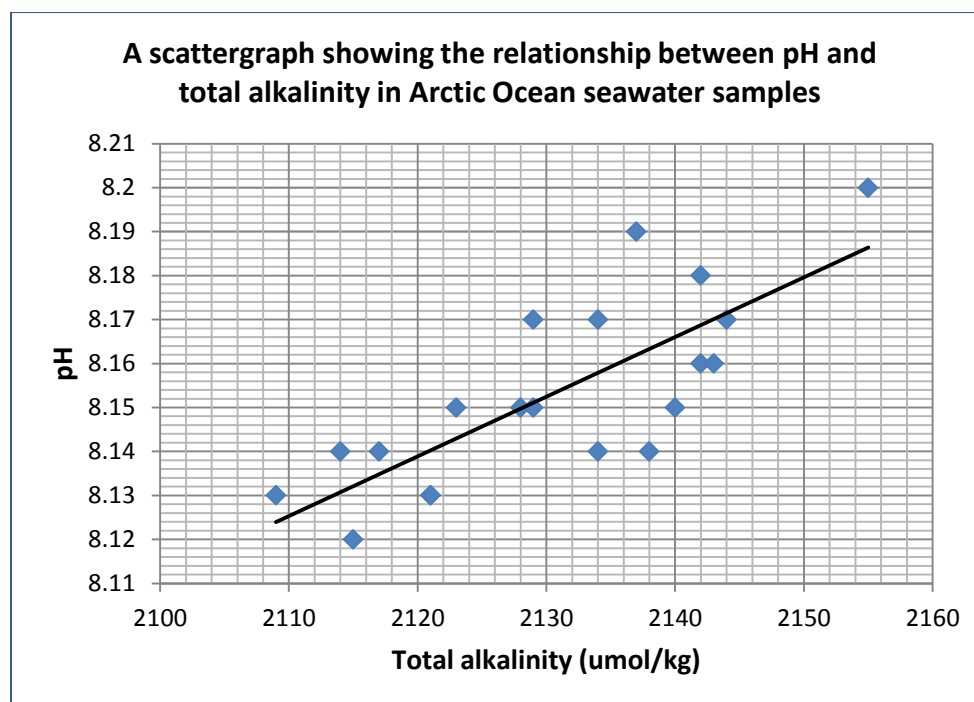
Spearman's Rank Example

Hypothesis

There is a positive correlation between the total alkalinity and pH of Arctic Ocean seawater, such that as total alkalinity increases, so does pH.

Scattergraphs

A simple, visual way to start thinking about the relationship between two sets of data is by drawing a scattergraph, either by hand or using Excel. A scattergraph plots the two datasets against each other and a best-fit line can be drawn to suggest a trend in the pattern indicated by the location of the various points.



This scattergraph suggests that there may be a positive correlation between pH and total alkalinity. However, this cannot be confirmed unless a statistical test is undertaken. Spearman's Rank correlation coefficient should now be used to establish the validity and significance of the relationship.

Spearman's Rank Correlation Coefficient

The table below lists the total alkalinity (in micromoles per kilogram) and pH for 20 samples taken during the course of Dr Findlay's research in the Arctic Ocean.

The first step in the Spearman's Rank correlation coefficient is to determine whether there is any correlation between the total alkalinity and the pH of the two samples. The instructions underneath the table take you through the process of doing this.

Sample date	Total alkalinity (umol/kg)	Rank	pH	Rank	d	d^2
17/03/2010	2137	8	8.19	2	-6	36
19/03/2010	2114	19	8.14	14.5	-4.5	20.25
21/03/2010	2134	9.5	8.17	5	-4	16
23/03/2010	2129	11.5	8.17	5	-6.5	42.25
25/03/2010	2123	14	8.15	10.5	-3.5	12.25
27/03/2010	2117	17	8.14	14.5	-2.5	6.25
29/03/2010	2142	4.5	8.18	3	-1.5	2.25
31/03/2010	2109	20	8.13	18	-2	4
02/04/2010	2115	18	8.12	20	2	4
04/04/2010	2121	15.5	8.13	18	2.5	6.25
06/04/2010	2121	15.5	8.13	18	2.5	6.25
08/04/2010	2128	13	8.15	10.5	-2.5	6.25
10/04/2010	2129	11.5	8.15	10.5	-1	1
12/04/2010	2155	1	8.20	1	0	0
14/04/2010	2142	4.5	8.16	7.5	3	9
16/04/2010	2144	2	8.17	5	3	9
18/04/2010	2138	7	8.14	14.5	7.5	56.25
20/04/2010	2134	9.5	8.14	14.5	5	25
22/04/2010	2143	3	8.16	7.5	4.5	20.25
24/04/2010	2140	6	8.15	10.5	4.5	20.25

$$\Sigma d^2 = 302.75$$

1. Rank both data sets. Ranks are shown in columns three and five of the table above. The highest value is always ranked first. If there is more than one entry with the same value, these entries should be given the same rank, for example 1, 2, 3, 4.5, 4.5, 6, 8, 8, 8, 10 (where 4.5 is the mean of ranks 4 and 5, and 8 is the mean of ranks 7, 8 and 9).
2. Calculate the difference (d) between the two rankings. The rank of the first data set should be subtracted from the rank of the second. You may obtain some negative values.
3. Calculate d^2 for each entry. All of the d^2 values should be positive.

4. Add up all of the d^2 values to obtain the value of Σd^2 . In this example, the $\Sigma d^2 = 302.75$.
5. You can now calculate the correlation coefficient using the formula below:

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

where n = the number of entries in the sample.

In our example it follows that:

$$\begin{aligned} r &= 1 - \frac{6 \times 302.75}{8000 - 20} \\ &= 1 - \frac{1816.5}{7980} \\ &= 1 - 0.228 \\ &= \underline{\underline{0.77}} \end{aligned}$$

The final answer should be given correct to two decimal places.

(You can also complete Spearman's Rank Correlation Coefficient using Excel, which will do many of the necessary calculations for you. Download the Spearman's Rank Excel Guide provided for instructions on how to do this.)

A perfect positive correlation has a value of +1, and a perfect negative correlation has a value of -1. All values of r should lie between these two extremes, but the closer the value to either +1 or -1, the stronger the likely correlation. A value of 0.77 suggests a positive correlation, but this still could have occurred by chance, so it is necessary to test the significance of the relationship.

Significance levels

The second step in your calculation is therefore to test the significance of the relationship you are investigating. You will require a table which plots the value of r against the 'degrees of freedom' ($n-2$).

Follow [this link](#) for a version of the Spearman's Rank significance graph. To use the graph, draw a horizontal line across from your value of r on the y-axis to where it meets a vertical line coming up from your value of $n-1$ on the x-axis. The three lines plotted on the graph will give you your level of significance.

If your value falls beneath the 5% significance line, you will be unable to accept your hypothesis, hence this line is referred to as the 'rejection level'.

In this example, where $r = 0.77$ and the degrees of freedom = 18, the point where the two values meet on the graph lies between the 0.1% and 1% significance levels, which means that we can say with 99.5% confidence that the correlation has not occurred by chance, hence that as total alkalinity increases, so too does the pH of sea water.

As suggested by our hypothesis, which we are now able to accept with 99.5% confidence, there is a positive correlation between total alkalinity and pH. We would expect this to be the case as pH is a measure of the acidity or alkalinity of a substance, such that the higher the pH value, the higher the level of alkalinity.

However, one thing we can't establish through this statistical test is whether the relationship between the two variables is *causal*, i.e. that a change in one results in a change in the other. Further investigation would be required in order to establish this type of relationship.

Additional link: <http://geographyfieldwork.com/SpearmanRank.htm>