

Ocean Acidification

Lesson 2: Data Analysis

Key question: How can ocean acidification data be analysed and presented?

Aim: To handle data collected during Dr Findlay's ocean acidification research visit to the Arctic Ocean, including:

- the setting of a hypothesis for the data set used
- the presentation of data in graphical form
- statistical analysis of data
- written analysis and explanation of findings

Starter

Watch the video clip "[In the Field](#)" which highlights some of the challenges of collecting data in the Arctic region.

Discuss the following questions with your group:

- What other challenges do you think researchers face living on the frozen ice of the Arctic Ocean for two months?
- What characteristics would it be important for a researcher collecting this type of data to have?
- What are the positives of an experience like this? Why do you think researchers dedicate so much time to collecting and analysing scientific and geographical data?

You will be analysing data collected in these conditions for the remainder of the lesson.

Main

The main part of the lesson follows the structure of an AQA Geographical Skills paper (Unit 2) question. However, it is also relevant to those studying other specifications:

Edexcel: Unit 1 Global Challenges calls for a review of recent scientific research into climate change.

OCR: As practice for the Geographical Skills (Unit F764) paper, which calls for the analysis and interpretation of geographical data.

WJEC: To practise the skills required for the Contemporary Themes and Research in Geography unit of the course (Unit G3), particularly Section B.

For all of the boards, you are required to have practical knowledge and experience of a range of statistical processes. For example, the OCR syllabus states that: "Statistical analysis is likely to be useful in most cases. Statistical analysis includes the simple grouping of data, through descriptive statistics (such as measures of central tendency and dispersion) to tests for correlation (such as Spearman's Rank) and for difference (such as Chi-squared, Mann-Whitney U test)."

In this lesson, you'll be using the Spearman's Rank Correlation Coefficient statistical test. This is used to show the strength and significance of the relationship between two variables, for example, the relationship between the GNP of a country and its birth rate. In this case, you'll be investigating the

relationship between different variables such as temperature, pH and dissolved organic carbon in the Arctic seawater samples collected by Dr Findlay during her research investigation.

The equation for Spearman's Rank Correlation Coefficient is

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

You need a data set (sample size) of at least 15 for the test to work.

If you're new to Spearman's Rank, download the [Spearman's Rank example document](#) which shows you how it works. The example investigates the relationship between total alkalinity and pH in Arctic sea water samples. It also tells you how you can complete Spearman's Rank in Excel, although you'll need to know how to do the calculations yourself for your exam.

Now it's your turn!

The [exam question](#) provided is based on the AQA Geographical Skills Paper (Unit 2), but the techniques are useful practice for students of all boards.

The question forms half of an actual paper, which means that you should be allowed 30 minutes to complete it. You are allowed to use a calculator.

Plenary

When you've completed your exam question, swap with your partner and use the [mark scheme](#) provided to assess each other's work. This is really useful practice for you as it makes you aware of the way papers are marked, and what the examiners are looking out for.

The total number of marks that can be awarded for the question is 25.

How did you do? If you think you could do with some more practice with Spearman's Rank, [two additional data sets](#) are provided for you to work on. The [answers](#) are given in a separate document so you can check how you've done.

Don't forget to spend some time analysing the results of your Spearman's Rank calculations. You will need to determine the significance level and see if you can explain the relationship you have identified. The information on [this website](#) will help you to do this.

You can also read Dr Findlay's [correlation explanations](#) in the document provided.

Ocean Acidification Lesson 2

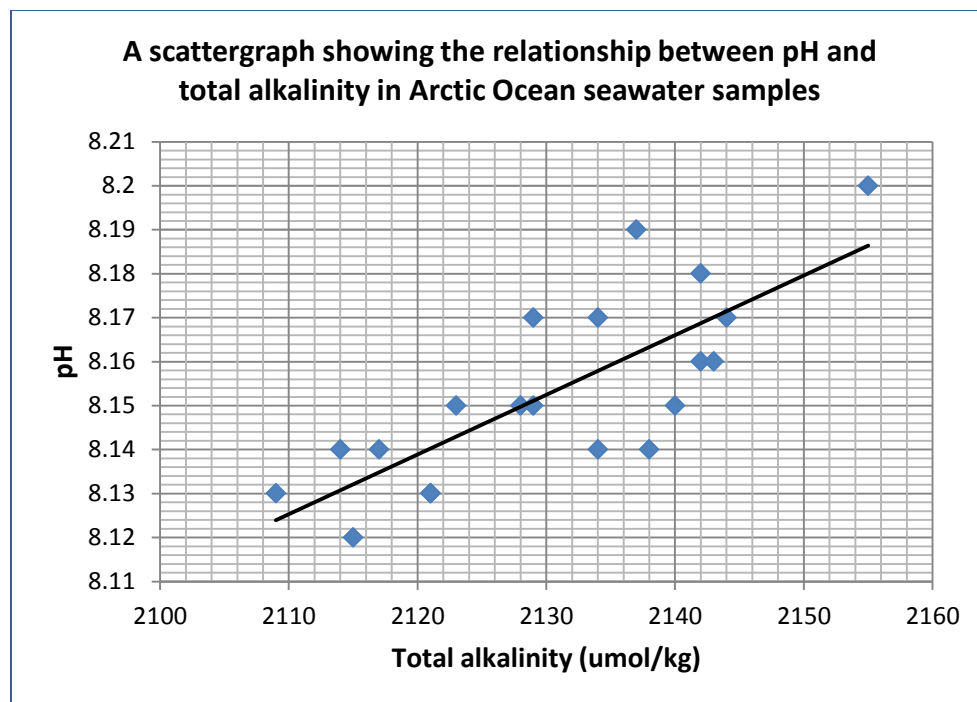
Spearman's Rank Example

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, 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:

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

where n = the number of entries in the sample.

In our example it follows that:

$$r = 1 - \frac{6 \times 302.75}{8000 - 20}$$

$$= 1 - \frac{1816.5}{7980}$$

$$= 1 - 0.228$$

$$= \underline{\underline{0.77}}$$

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.

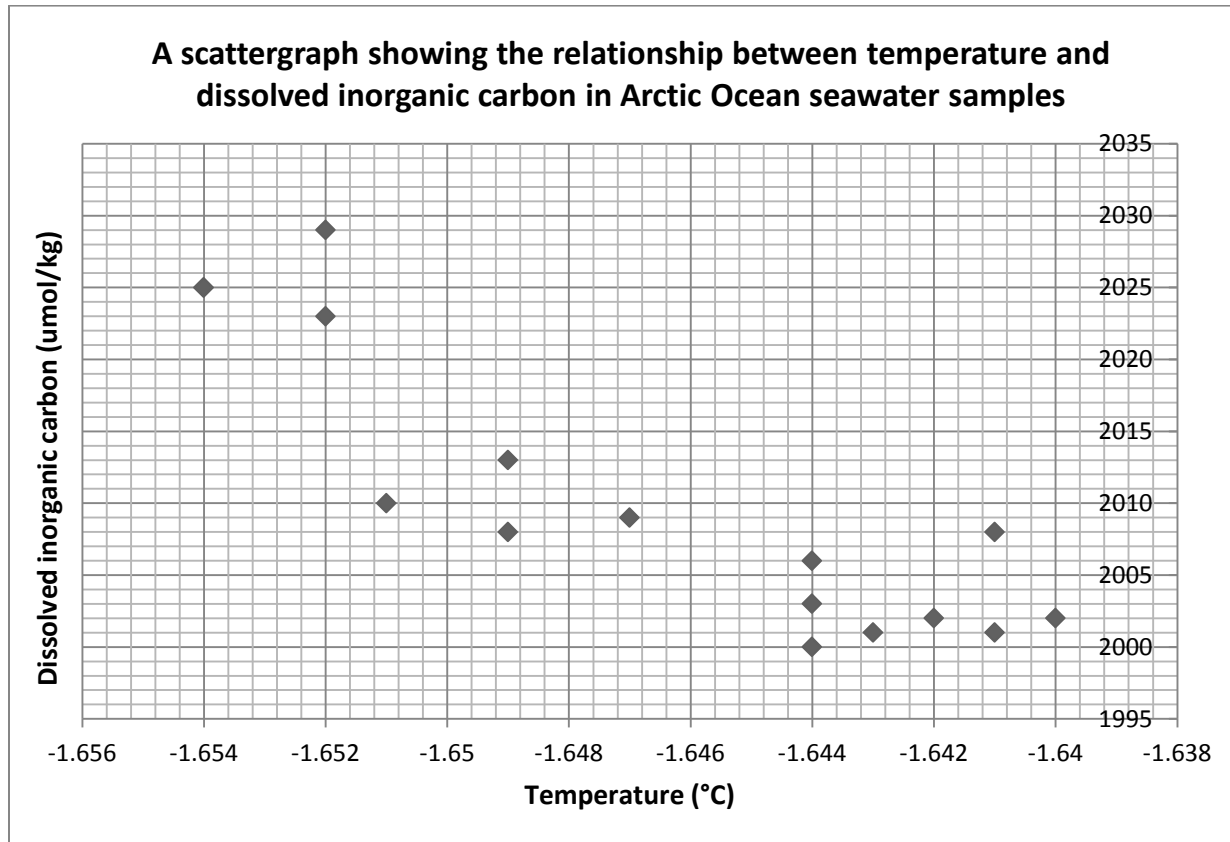
Additional link: <http://geographyfieldwork.com/SpearmanRank.htm>

Ocean Acidification sample exam question

Answer **all** questions

- 1 (a) Study **Figure 1**, a scattergraph which shows the relationship between the temperature and dissolved carbon content of seawater samples collected in the Arctic Ocean.

Figure 1



- 1 (a) (i) Complete the scattergraph by adding the following data

Sample date	Temperature (°C)	Dissolved inorganic carbon (umol/kg)
04/04/10	-1.646	2007
10/04/10	-1.645	2010
16/04/10	-1.652	2018
22/04/10	-1.650	2024

(4 marks)

- 1 (a) (ii) Draw a best-fit line on the graph to show the trend between the two sets of data.

(1 mark)

1 (a) (iii) Describe the relationship between the two sets of variables, as indicated by the scattergraph in **Figure 1**.

.....

.....

.....

.....

.....

.....

(3 marks)

The next step when investigating the correlation between two variables is to use statistical testing to determine the validity and strength of the relationship.

1 (a) (iv) Write a hypothesis for this dataset that you could investigate through statistical testing.

.....

.....

.....

.....

(2 marks)

Question 1 continues on the next page

1 (b) Study **Figure 2** below. This table sets out the preliminary workings for a Spearman's Rank Correlation Coefficient calculation.

Figure 2

Sample date	Temperature (°C)	Rank	Dissolved inorganic carbon (umol/kg)	Rank	<i>d</i>	<i>d</i> ²
17/03/10	-1.644	7	2003	15	8	64
19/03/10	-1.644	7	2000	19	12	144
21/03/10	-1.644	7	2006	14	7	49
23/03/10	-1.643	5	2001	18.5	13.5	182.25
25/03/10	-1.641	2.5	2001	18.5	16	256
27/03/10	-1.640	1	2002	16.5	15.5	240.25
29/03/10	-1.641	2.5	2008	11.5	9	81
31/03/10	-1.642	4	2002	16.5	12.5	156.25
02/04/10	-1.647	11	2009	10	-1	1
04/04/10	-1.646	10	2007	13	3	9
06/04/10	-1.649	12.5	2013	7	-5.5	30.25
08/04/10	-1.649	12.5	2008	11.5	-1	1
10/04/10	-1.645	9	2010	8.5	-0.5	0.25
12/04/10	-1.651	15	2010	8.5	-6.5	42.25
14/04/10	-1.652	17	2023	5	-12	144
16/04/10	-1.652	17	2018	6	-11	121
18/04/10	-1.652	17	2029	1	-16	256
20/04/10	-1.654	19.5	2025	2.5	-17	289
22/04/10	-1.650	14	2024	4	-10	100
24/04/10	-1.654	19.5	2025	2.5	-17	289

1 (b) (i) Using the formula $r = 1 - \frac{6\sum d^2}{n^3 - n}$ (where *n* = the number of samples) calculate the Spearman's Rank Correlation Coefficient for this data set. Show your workings in the box below and give your answer to two decimal places. (5 marks)

r =

1 (b) (ii) Explain how you would establish the significance of the correlation between water temperature and dissolved carbon content.

.....

.....

.....

.....

.....

.....

.....

.....

(4 marks)

Question 1 continues on the next page

1 (b) (iii) Use your knowledge of the process of ocean acidification to explain the significance of a correlation between water temperature and dissolved carbon content in the Arctic Ocean.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

(6 marks)

(Extra space)

.....

.....

.....

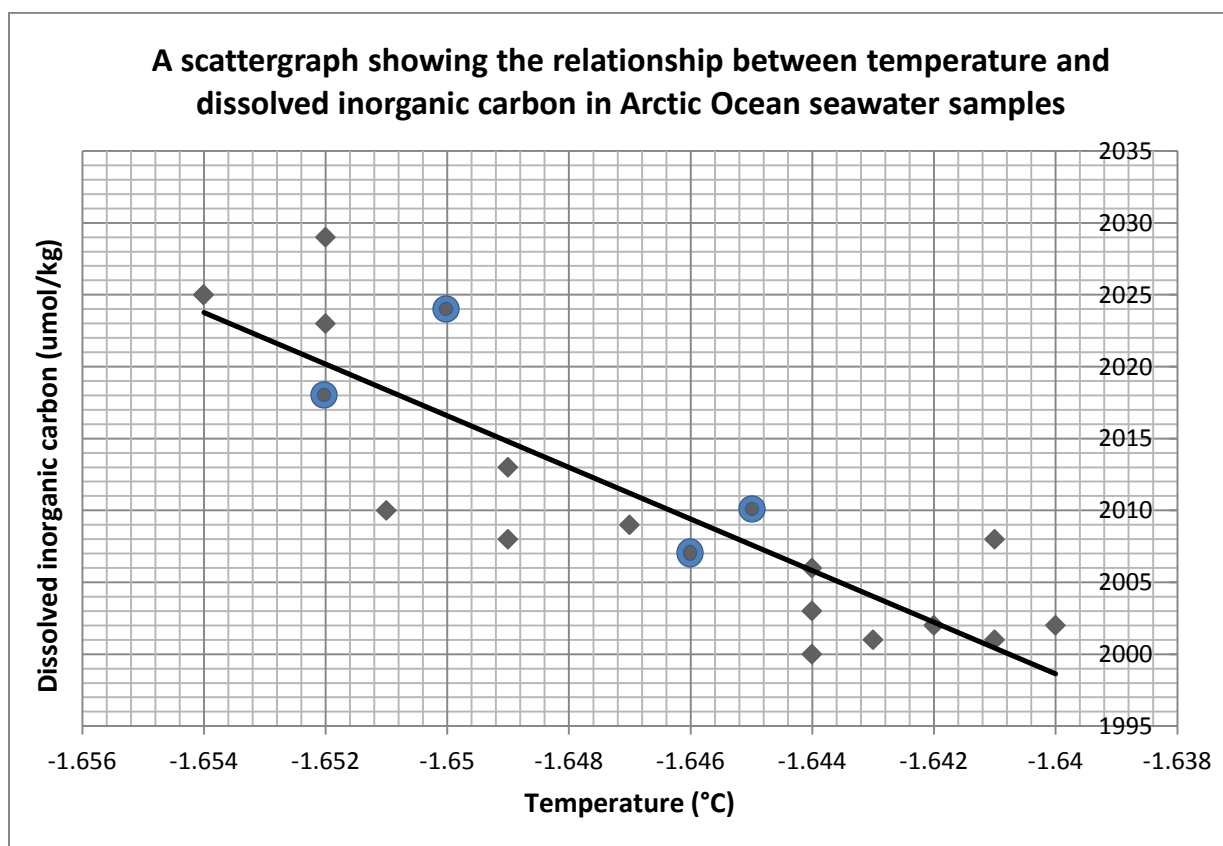
.....

END OF QUESTION

Total marks: 25

Mark scheme for Ocean Acidification Lesson 2 main activity

1 (a) (i)	1 mark for each accurate plot (see highlighted points on graph below)	(4 marks)
1 (a) (ii)	1 mark for straight line through the middle of the plots, approximating that shown below. There should be an even number of plots above and below the line.	(1 mark)
1 (a) (iii)	1 mark per valid point with additional credit for correct use of data from the graph (maximum 1 mark for this), e.g. As temperature decreases, dissolved inorganic carbon increases. Suggestion of a negative correlation between the variables. Distribution of points suggests this is a relatively strong but not perfect correlation. Examples given, e.g. at -1.640°C dissolved inorganic carbon at a level of 2002 $\mu\text{mol/kg}$ whereas at -1.654°C the level is 2025 $\mu\text{mol/kg}$.	(3 marks)
1 (a) (iv)	1 mark for mention of both sets of data, 1 mark for Maximum 1 mark for incorrectly worded hypothesis. e.g. There is a significant negative correlation between the temperature and dissolved inorganic carbon content of Arctic Ocean seawater samples, meaning that as temperature decreases, the dissolved carbon content decreases.	(2 marks)



1 (b) (i)	1 mark for each correct stage (shown by workings – see overleaf) in the calculation (maximum 4 marks). Alternative approaches to workings permitted, as long as they are accurate and logical. 1 mark for correct answer given to 2 decimal places.	(5 marks)
1 (b) (ii)	1 mark per valid stage in the process of establishing the strength and validity of a correlation, e.g. The process requires a graph which plots r against degrees of freedom. Degrees of freedom are the number of pairs in the sample minus 2 (i.e. $n - 2$). Graph indicates significance levels of 0.1%, 1% and 5%. Plotting value of r against degrees of freedom determines whether hypothesis can be accepted at 99.9% / 99% / 95% level, or must be rejected. Levels mean there is a 0.1% / 1% / 5% likelihood that the correlation occurred by chance. If value of r falls below the 5% level, hypothesis is rejected. No result can prove a causal relationship between variables.	(4 marks)
1 (b) (iii)	<p>Notes for answers</p> <p>Responses should link the significance of the temperature/dissolved carbon relationship to an awareness and understanding of the process of ocean acidification, and of the implications of this for the Arctic Ocean.</p> <p>Ocean acidification is the process by which carbon dioxide from the atmosphere is absorbed by the oceans, where it reacts with seawater to form carbonic acid. As the concentration of carbon dioxide in the atmosphere increases, so to does the amount absorbed by the oceans. As a result, the acidity of oceans globally is increasing (the pH is decreasing). The results of the correlation suggest that colder water contains more dissolved carbon, hence has more potential to become acidic. This has serious implications for the Arctic Ocean because it is in the cold Polar regions that the effects and impacts of ocean acidification are being observed, and are likely to increase the most.</p> <p>Ocean acidification has the potential to impact on the whole Arctic ecosystem. The increased acidity of the water can start to dissolve the calcium carbonate shells of microscopic plankton, reducing their resilience and reducing population numbers. This affects the food supply for consumers and impacts on the entire foodweb.</p>	(6 marks)

	<p>Level 1 (1 – 4 marks) Ranging from no real understanding of the science behind ocean acidification (1-2 marks) to some inaccuracies (3-4 marks). No clear explanation of the importance of the relationship to the Arctic Region. Simplistic comments. Predominantly descriptive. Lacking explanation.</p> <p>Level 2 (5 – 6 marks) Accurate description of the process of ocean acidification. Clear explanation of the relationship indicated by the data. Links made between the data and the implications for the Arctic region. Mention of specific impacts of ocean acidification on the ecosystem for full marks.</p>	
--	---	--

Calculations for Question 1 (b) (i):

$r = 1 - \frac{6 \sum d^2}{n^3 - n}$		
Step 1:	$\sum d^2 = 2455.5$	
Step 2:	$1 - \frac{6 \times 2455.5}{8000 - 20}$	Answer correct to 2 decimal places:
Step 3:	$1 - \frac{14733}{7980}$	$r = -0.85$
Step 4:	$1 - 1.846$	

Additional data sets for Spearman's Rank practice

1. Total alkalinity and dissolved inorganic carbon

Sample date	Total alkalinity (umol/kg)	Rank	Dissolved inorganic carbon (umol/kg)	Rank	d	d^2
17/03/10	2137		2003			
19/03/10	2114		2000			
21/03/10	2134		2006			
23/03/10	2129		2001			
25/03/10	2123		2001			
27/03/10	2117		2002			
29/03/10	2142		2008			
31/03/10	2109		2002			
02/04/10	2115		2009			
04/04/10	2121		2007			
06/04/10	2121		2013			
08/04/10	2128		2008			
10/04/10	2129		2010			
12/04/10	2155		2010			
14/04/10	2142		2023			
16/04/10	2144		2018			
18/04/10	2138		2029			
20/04/10	2134		2025			
22/04/10	2143		2024			
24/04/10	2140		2025			

Σd^2	
--------------	--

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

2. Temperature and salinity

Sample date	Temperature (°C)	Rank	Salinity (parts per thousand ‰)	Rank	<i>d</i>	<i>d</i> ²
17/03/10	-1.644		29.94			
19/03/10	-1.644		29.94			
21/03/10	-1.644		30.04			
23/03/10	-1.643		30.23			
25/03/10	-1.641		30.46			
27/03/10	-1.640		29.98			
29/03/10	-1.641		29.99			
31/03/10	-1.642		29.97			
02/04/10	-1.647		30.11			
04/04/10	-1.646		30.08			
06/04/10	-1.649		30.30			
08/04/10	-1.649		30.12			
10/04/10	-1.645		30.16			
12/04/10	-1.651		30.16			
14/04/10	-1.652		30.23			
16/04/10	-1.652		30.19			
18/04/10	-1.652		30.26			
20/04/10	-1.654		30.26			
22/04/10	-1.650		30.34			
24/04/10	-1.654		30.28			

Σd^2	
--------------	--

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

Additional data sets for Spearman's Rank practice - ANSWERS

3. Total alkalinity and dissolved inorganic carbon

Sample date	Total alkalinity (umol/kg)	Rank	Dissolved inorganic carbon (umol/kg)	Rank	d	d^2
17/03/10	2137	8	2003	15	7	49
19/03/10	2114	19	2000	19	0	0
21/03/10	2134	9.5	2006	14	4.5	29.25
23/03/10	2129	11.5	2001	18.5	7	49
25/03/10	2123	14	2001	18.5	4.5	20.25
27/03/10	2117	17	2002	16.5	-0.5	0.25
29/03/10	2142	4.5	2008	11.5	7	49
31/03/10	2109	20	2002	16.5	-3.5	12.25
02/04/10	2115	18	2009	10	-8	64
04/04/10	2121	15.5	2007	13	-2.5	6.25
06/04/10	2121	15.5	2013	7	-8.5	72.25
08/04/10	2128	13	2008	11.5	-1.5	2.25
10/04/10	2129	11.5	2010	8.5	-3	9
12/04/10	2155	1	2010	8.5	7.5	56.25
14/04/10	2142	4.5	2023	5	1.5	2.25
16/04/10	2144	2	2018	6	4	16
18/04/10	2138	7	2029	1	-6	36
20/04/10	2134	9.5	2025	2.5	-7	49
22/04/10	2143	3	2024	4	1	1
24/04/10	2140	6	2025	2.5	-3.5	12.25

Σd^2	526.5
--------------	-------

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

$$r = 1 - \frac{6 \times 526.5}{8000 - 20}$$

$$r = 1 - \frac{3159}{7800}$$

$$r = 1 - 0.396$$

$r = 0.60$

4. Temperature and salinity

Sample date	Temperature (°C)	Rank	Salinity (parts per thousand ‰)	Rank	d	d^2
17/03/10	-1.644	7	29.94	19.5	12.5	156025
19/03/10	-1.644	7	29.94	19.5	12.5	156025
21/03/10	-1.644	7	30.04	15	8	64
23/03/10	-1.643	5	30.23	7.5	2.5	6.25
25/03/10	-1.641	2.5	30.46	1	-1.5	2.25
27/03/10	-1.640	1	29.98	17	16	256
29/03/10	-1.641	2.5	29.99	16	13.5	182.25
31/03/10	-1.642	4	29.97	18	14	196
02/04/10	-1.647	11	30.11	13	2	4
04/04/10	-1.646	10	30.08	14	4	16
06/04/10	-1.649	12.5	30.30	3	-9.5	90.25
08/04/10	-1.649	12.5	30.12	12	-0.5	0.25
10/04/10	-1.645	9	30.16	10.5	1.5	2.25
12/04/10	-1.651	15	30.16	10.5	-4.5	20.25
14/04/10	-1.652	17	30.23	7.5	-9.5	90.25
16/04/10	-1.652	17	30.19	9	-8	64
18/04/10	-1.652	17	30.26	5.5	-11.5	132.25
20/04/10	-1.654	19.5	30.26	5.5	-14	196
22/04/10	-1.650	14	30.34	2	-12	144
24/04/10	-1.654	19.5	30.28	4	-15.5	240.25

Σd^2	2019
--------------	------

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

$$r = 1 - \frac{6 \times 2019}{8000 - 20}$$

$$r = 1 - \frac{12114}{7800}$$

$$r = 1 - 1.518$$

$r = -0.52$

Spearman's Rank Correlation Coefficient – Excel Guide

What is Spearman's Rank correlation coefficient?

Spearman's Rank correlation coefficient is used to identify and test the strength of a relationship between two sets of data. It is often used as a statistical method to aid with either proving or disproving a hypothesis e.g. the depth of a river does not progressively increase the further from the river bank.

The formula used to calculate Spearman's Rank is shown below.

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

NB. Sometimes $n^3 - n$ is written as $n(n^2 - 1)$. Both mean the same thing.

How can the calculation be carried out in Excel?

Once the data has been collected, Excel can be used to calculate and graph Spearman's Rank correlation to discover if a relationship exists between the two sets of data, and how strong this relationship is. Please note this example uses a dataset of 10 samples, but your dataset should include a minimum of 15 to be valid.

Step 1: Create a table in Excel and enter your data sets.

Sample	Width (cm)	Width Rank	Depth (cm)	Depth Rank
1	0		0	
2	50		10	
3	150		28	
4	200		42	
5	250		59	
6	300		51	
7	350		73	
8	400		85	
9	450		104	
10	500		96	

1

¹ Data for this example has been taken from www.geography-site.co.uk/pages/skills/fieldwork/stats/spear.html

Step 2: Rank each set of data (width rank and depth rank). Rank 1 will be given to the **largest** number in column 2. Continue ranking till all widths have been ranked. Once all the widths have been ranked then do exactly the same for depth.

Sample	Width (cm)	Width Rank	Depth (cm)	Depth Rank
1	0	10	0	10
2	50	9	10	9
3	150	8	28	8
4	200	7	42	7
5	250	6	59	5
6	300	5	51	6
7	350	4	73	4
8	400	3	85	3
9	450	2	104	1
10	500	1	96	2

If there are two samples with the same value, the mean (average) rank should be used – for example if there were 3 samples all with the same depth ranked 6th in order you would add the rank values together (6+7+8=21) then divide this number by the number of samples with the same depth number, in this case 3 (21/3=7) so they would all receive a rank of 7. The next greatest depth would be given a value of 9.

Step 3: The next stage is to find d (the difference in **rank** between the width and depth). First, add a new column to your table, and then calculate d by subtracting the depth rank column (column 5) from the width rank column (column 3). For example, for sample 6 width rank is 5 and the depth rank is 6 so $d = 5 - 6 = -1$. To calculate d in Excel, select the cell you wish to enter the information into and type =. Now click on the width rank cell you want to use and type -. Finally, click on the depth rank cell and press enter. The value of d should appear in the first box you selected.

Sample	Width (cm)	Width Rank	Depth (cm)	Depth Rank	d
1	0	10	0	10	0
2	50	9	10	9	0
3	150	8	28	8	0
4	200	7	42	7	0
5	250	6	59	5	1
6	300	5	51	6	-1
7	350	4	73	4	0
8	400	3	85	3	0
9	450	2	104	1	1
10	500	1	96	2	-1

Step 4: The next step is to calculate d^2 . Add another column to your table and label it. To calculate d^2 type in the first cell =POWER(number,power). In this case the number is the value of d and the power is 2 as we are trying to find the square value e.g. for sample 6 the value of d is -1 so you would enter into the cell =POWER(-1,2) then press enter and the value you should get is 1.

Sample	Width (cm)	Width Rank	Depth (cm)	Depth Rank	d	d^2
1	0	10	0	10	0	
2	50	9	10	9	0	
3	150	8	28	8	0	
4	200	7	42	7	0	
5	250	6	59	5	1	
6	300	5	51	6	-1	1
7	350	4	73	4	0	
8	400	3	85	3	0	
9	450	2	104	1	1	
10	500	1	96	2	-1	

Repeat the same process until all of your samples have a value of d^2 . Once all the d^2 values have been calculated add them together to calculate $\sum d^2$. The quickest way to do this in Excel is to click on the cell underneath your last entry into the d^2 column, click on the autosum symbol Σ (which you can find on the tool bar at the top of the page), and press enter. (Depending on which version of Excel you are using, you may have to select the column you wish to add together before you press enter.)

Step 5: Now we have the $\sum d^2$ values, but to complete the equation we still need to calculate $n^3 - n$. n is the number of samples, so in this case is 10. As in step 4, type into the cell you wish to use =POWER(number,power) which will give you a value for n^3 . Remember this time that 'number' is the number of samples and 'power' is 3 as you are cubing not squaring the value. Once n^3 has been calculated, subtract the value of n from it.

Sample	Width (cm)	Width Rank	Depth (cm)	Depth Rank	d	d^2
1	0	10	0	10	0	0
2	50	9	10	9	0	0
3	150	8	28	8	0	0
4	200	7	42	7	0	0
5	250	6	59	5	1	1
6	300	5	51	6	-1	1
7	350	4	73	4	0	0
8	400	3	85	3	0	0
9	450	2	104	1	1	1
10	500	1	96	2	-1	1
					$\sum d^2$	4
					n	10
					n^3	1000
					$n^3 - n$	990

Step 6: All that is left to do now is to insert the values into the equation to calculate r .

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

The formula you should enter into a cell in this case would be =1-((6*4)/990)

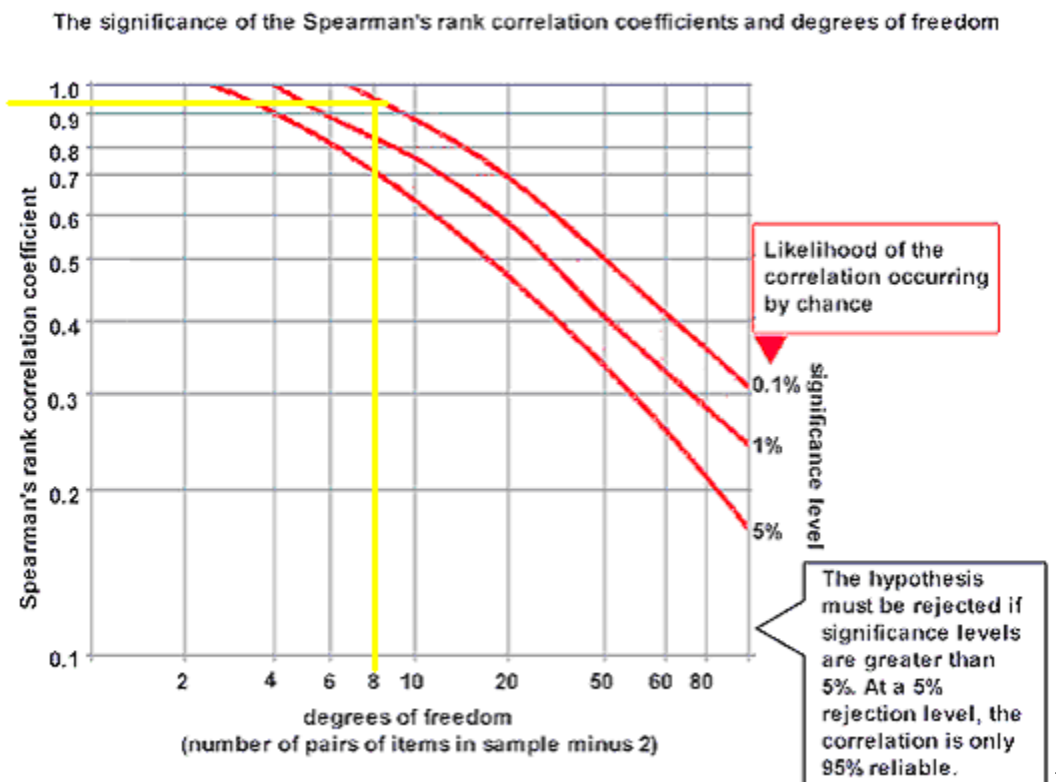
(N.B * is the symbol for multiply)

End result

The result at the end should always be between the value of +1 and -1. In this case the value is 0.9757 to 4 decimal places, or 0.98 to 2 decimal places.

Step 7: To see if this r value is significant, a Spearman's Rank significance table or graph must be used. To do this you need to calculate the degrees of freedom.

This is calculated using the sum $n - 2$, which in this example = 8 (remember n is the number of samples). Using this number and your value for r you can use the table below to work out the significance level of your data (the yellow lines).



As you can see the yellow line meets the red line at 0.1% meaning that there is a greater than 99% chance that the relationship is significant and not random. Thus we can **reject** the hypothesis and conclude that in this case the depth of a river does progressively increase the further the distance from the river bank.

² Table taken from <http://geographyfieldwork.com/SpearmansRankSignificance.htm>

Carbon chemistry correlation analysis

Helen Findlay

Table 1: The values to the left of the grey boxes represent the product moment correlation coefficients (r) for each parameter set. The values to the right of the grey boxes represent the associated probability values (p).

	TA	DIC	pH	Salinity	Temp.
TA		0.0001	0.0001	0.0055	0.0005
DIC	0.585336		0.848	0.0001	0.0001
pH	0.788581	-0.02974		0.989	0.7279
Salinity	0.41148	0.584703	-0.00213		0.0019
Temp.	-0.5065	-0.77745	-0.05396	-0.45464	

Temperature and salinity show a negative correlation, this means that as salinity increases temperature decreases. They correlate because seawater has associated properties depending on where the seawater has come from and therefore what has happened to it as its been moved around the ocean. This dataset represents the top 10 m of a column of water under sea ice. This 'surface layer' is therefore highly influenced by the ice lying on top. The seawater below the ice is cold and salty, but as the ice starts to warm up it begins to melt. Because the ice is actually fresher than the seawater (salts were extruded from the sea ice as it formed) it has a freezing point (melting point) at a higher (warmer) temperature than the freezing point of the more salty seawater. So the melted ice water is warmer and less salty than the underlying seawater. So the seawater we have sampled here has an ice melt signal of warmer, less saline water.

Total alkalinity and pH have a significant positive correlation such that as TA increases, pH increases. pH is the measure of acidity. It is actually the negative log of the hydrogen ion concentration. So a low pH means there are lots of hydrogen ions present, which make the solution very acidic. A high pH means there are few hydrogen ions (H^+) present, more hydroxide ions (OH^-), which makes the solution less acidic, more basic. Total alkalinity is a measure of all the ions that are present in seawater that are able to counter any change in H^+ . So a high TA concentration means there are lots of ions in the seawater which are available to buffer changes in H^+ . Freshwater, for example, has a low TA concentration, because it contains very few ions other than OH^- . This meant that when you add acid (H^+) there are not many ions available to soak up the protons and the water becomes more acidic. We would expect a solution with lots of negative ions (BOH^- , NO_3^- , PO_4^- , SiO_4^- , CO_3^{2-} and HCO_3^- etc, as is the case with seawater) to have a high alkalinity and therefore a high pH because these negative ions soak up any addition of acid.

DIC, on the other hand, is the total concentration of all dissolved inorganic carbon compounds in seawater. This includes only $CO_{2(aq)}$, CO_3^{2-} and HCO_3^- . So while it does include the carbonate and bicarbonate ions that could potentially buffer pH, it also contains CO_2 which can react with water to form more acid (and decrease the pH). We would therefore not expect there to be any correlation between pH and DIC.

Temperature and salinity can influence the amount of ions that are present in seawater by diluting or concentrating the ions in the seawater (through processes such as evaporation, precipitation, ice freezing and melting). As TA and DIC are both measures of the concentration of these various ions that are present

in seawater we would expect that TA and DIC are therefore influenced by, and therefore correlate with, both temperature and salinity.

The correlations between temperature and DIC or TA are negative. So as temperature increases, DIC or TA decrease. Correlations between salinity and DIC or TA are positive, so as salinity increases, DIC or TA increase.

As discussed briefly above, temperature and salinity can be used as markers of where the water has come from and what processes have occurred within it. So the fresh (less saline), warm water that was added from the melting ice, also has low concentrations of DIC and TA because when the seawater froze the ions within the seawater were also extruded from the sea ice.

It is perhaps a little bit surprising that pH does not have a significant correlation with salinity or temperature, because TA and pH correlate, and TA correlates with both temperature and salinity. The most likely explanation is that there are other processes going on that contribute to pH more strongly than changes in temperature and salinity, which are not captured just by measuring the carbonate system parameters.

TA and DIC show a significant positive correlation. This is not likely to be a cause and effect relationship but an unrelated correlation; a coincidence. We have seen that both DIC and TA are correlated with temperature and salinity in the same manner; therefore it is likely that the changes in temperature and salinity that were found in these water samples are responsible for the changes in DIC and TA, which respond in the same pattern.