

Integrating skills – Practical approaches

Geographical skills in relation to both an equal weighting of quantitative and qualitative skills are required for A level learners, and the list in Appendix A of the specification indicates those selected for study for all components in this specification. All the skills need to be addressed within these components but not all will apply to fieldwork. The four required days of fieldwork should contribute to learners building a holistic and balanced understanding of quantitative and qualitative skills related to fieldwork and the six-stage enquiry process.

Definitions:

- Quantitative research is “explaining phenomena by collecting numerical data that are analysed using mathematically based methods (in particular statistics).”
- Qualitative research seeks to answer questions about why and how people behave in the way that they do. It provides in-depth information about human behaviour.

Source: <http://www.skillsyouneed.com/learn/quantitative-and-qualitative.html>

Additional guidance:

Qualitative skills

- RGS article
(<https://www.rgs.org/OurWork/Schools/Fieldwork%2Band%2Blocal%2Blearning/Planning%2Byour%2Bfieldtrip/Fieldwork%2Bideas/Qualitative%2Bfieldwork.htm>)
- Learning to analyse qualitative data – online tutorial
(<http://archive.learnhigher.ac.uk/analysethis/main/qualitative.html>)

Quantitative skills

- I-Use project (<http://i-use.eu/resources.html>) Including ‘How to’ video guides and student tasks
- GA ‘How to guides’ Conducting statistical tests in fieldwork
(<http://www.geography.org.uk/resources/conductingstatisticaltestsforfieldwork/>)

Learners need to develop competence in using the geographical skills specified in the DfE Geography GCE AS and A level Subject Content (December 2014), as shown in the 'Integrating geographical skills in delivery of the core themes' tables in Appendix A of the specification.

Below are some examples of tasks that can be used to integrate skills into lesson delivery.

Coastal Landscapes

1.1.1

- use of numerical data to calculate sediment budgets

Constructing accurate sediment budgets is a time consuming process which involves measurement and monitoring of the rates all the major sediment transport processes and storage zones. Complete sediment budgets are therefore relatively unusual; however this example of a sediment budget approach to coastal erosion in South Carolina usefully exemplifies the application of the approach in a coastal management

context: <http://pubs.usgs.gov/of/2008/1206/html/processes1.html>

1.1.2

- measures of central tendency (mean, mode) – mean wave frequency

To estimate wave frequency, count the number of waves over a 10 minute period and divide the total by the number of minutes to determine the mean number of waves per minute.

1.1.3

- scale

Trace a 30–40 km coastline at a range of scales (1: 1000 000, 1: 50 000 and 1:25 000), and comment on the influence of scale on the plan of the coastline.

- landscape system identification

Classification of coastal landscapes according to landscape character type (LCT)

Holmes, D. (2013) Assessing landscapes. *Geography Review* 27 (2) pp.34–36

- digital and geo-located data

Comparisons of characteristics of rocky, sandy and estuarine coastal environments using GIS mapping of the variety of coastal (rocky, sandy and estuarine) landscapes both for and beyond the UK <https://www.arcgis.com/home/>

NASA's Visible Earth Programme is a source of satellite photographs of coasts

www.tinyurl.com/kk5cq32

<http://visibleearth.nasa.gov/>

1.1.4

- distance and area

Calculate the maximum fetch using an atlas. Work out the maximum fetch for the following locations:

Aberdeen in north-east Scotland has a fetch of _____ km

Rhossili in south-west Wales has a fetch of _____ km

Dover in south-east England has a fetch of _____ km

Use the formula $H = 0.36\sqrt{VF}$ to calculate the maximum possible wave height at these locations, as determined by fetch.

- rose/star/radial diagrams

Draw a wind rose of the tabulated data to show the prevailing wind direction shown below:

Mean percentage frequency of winds in the British Isles											
N	NNE	ENE	E	ESE	SSE	S	SSW	WSW	W	WNW	NNW
7	5	6	7	5	8	9	9	14	15	8	7

1.1.5

- field sketches of cliff profiles

A key field skill for geomorphologists is observation. The ability to observe landforms in the field, to systematically record those observations and then apply classroom knowledge of the environment and process to explain the genesis of the forms observed is central. Producing annotated field sketches is a great way to formalise this process. Annotating photographs in the field using appropriate apps such as Skitch, which helps with the labelling and annotation of photographs, is another option (see Holmes, 2013).

Holmes, D. (2013) Fieldwork of the future. *Geography Review* 26 (4) pp.25–27

1.1.6

- sampling

Sampling of beach pebbles, including the ability to identify sources of error in data, measurement errors and misuse of data

<http://geographyfieldwork.com/MinimumSampleSize.htm>

Article covering coastal fieldwork on a beach

www.thegeographeronline.net/uploads/2/6/6/2/26629356/gf551.pdf

See Holmes, D. (2013) Are your data reliable, accurate and valid? *Geography Review* 26 (3) pp.34–36

- data sets

Samples of beach pebbles see Holmes, D. (2010) Beach profiles. *Geography Review* 23 (3) pp.5–7

- frequencies

Recorded frequencies of shape of beach pebbles using Power's scale shown in the table below:

Power's scale	Frequency	Power's scale	Frequency	Power's scale	Frequency
Sample of 20 beach pebbles taken from western location on Pwll Du beach		Sample of 20 beach pebbles taken from central location on Pwll Du beach		Sample of 20 beach pebbles taken from eastern location on Pwll Du beach	
0	0	0	0	0	0
1	0	1	0	1	0
2	8	2	4	2	0
3	11	3	6	3	0
4	1	4	8	4	7
5	0	5	2	5	9
6	0	6	0	6	4

- measures of central tendency (mode)

Identify the modal Power's scale for each of the 3 samples tabulated above

- measures of dispersion (range, standard deviation, interquartile range)

<http://geographyfieldwork.com/MinimumSampleSize.htm>

Smaller standard deviations reflect more clustered data. More clustered data means less extreme values. A data set with less extreme values has a more reliable mean. The standard deviation is therefore a good measure of the reliability of the mean value. The formula is as follows:

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

σ = standard deviation

\sum = sum of

x = each value in the data set

\bar{x} = mean of all values in the data set

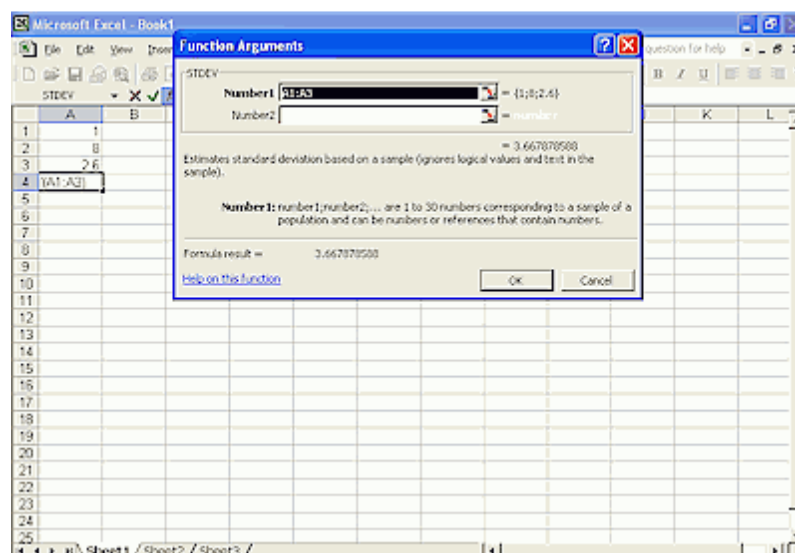
n = number of value in the data set

Is there an easy way to calculate it?

The Microsoft Excel programme will automatically calculate the standard deviation and mean for a set of data listed in a spreadsheet column.

Method:

- List data set in a single column
- Click on the empty cell below the last data item
- Open INSERT menu > FUNCTION > STDEV > click OK
- The standard deviation is then shown and will appear in the empty cell.
- The excel screen example below is for a data set of 3 items



Pebble data set of 30 pebble long axes from Site 1 Stiges beach, Spain. Calculate the range, standard deviation and interquartile range of the sample.

Pebble number	Long Axis (cm)
1	10
2	9
3	8
4	8
5	16
6	12
7	8.5
8	10
9	12
10	9
11	13
12	14
13	10
14	14
15	17
16	12
17	6
18	17
19	9
20	5
21	10

22	7.5
23	13
24	13
25	7.5
26	15
27	12
28	8
29	22
30	16

Mean	11.20 cm
Standard Deviation	3.81 cm
Range	5 cm – 22 cm = 17 cm
Interquartile range	8.5 cm – 14 cm = 5.5 cm

Samples of beach pebbles were taken at 12 locations spread west to east along Pwll Du beach, Gower, at intervals of 25 m as shown in the table below:

Sampling point	1 (W)	2	3	4	5	6	7	8	9	10	11	12 (E)
Distance (m)	0	25	50	75	100	125	150	175	200	225	250	275
Particles < 10 mm x-axis (%)	3	4	12	16	9	26	42	34	60	73	71	78

- Draw a scatter plot to show the relationship between distance (west to east) along Pwll Du beach and the % particles with x-axis < 10 mm size of particles
- Draw a line of best fit
- Analyse the statistical significance of the relationship using Spearman's Rank Correlation Coefficient

- Spearman's Rank Correlation Coefficient

This is a technique which can be used to summarise the strength and direction (negative or positive) of a relationship between two variables. The result will always be between 1 and minus 1.

Method – calculating the coefficient:

- Create a table from your data.
- Rank the two data sets. Ranking is achieved by giving the ranking '1' to the biggest value in a column, '2' to the second biggest value and so on. The smallest value in the column will get the lowest ranking. This should be done for both sets of measurements.
- Tied scores are given the mean (average) rank.
- Find the difference in the ranks (d): This is the difference between the ranks of the two values on each row of the table. The rank of the second value (% particles with x -axis < 10 mm) is subtracted from the rank of the first (sampling point).
- Square the differences (d^2) to remove negative values and then sum them ($\sum d^2$).

Distance (m)	Rank	% particles with x -axis < 10 mm	Rank	Difference (d)	Difference squared (d^2)
0	12	3	12	0	0
25	11	4	11	0	0
50	10	12	9	-1	1
75	9	16	8	-1	1
100	8	9	10	2	4
125	7	26	7	0	0
150	6	42	5	-1	1
175	5	34	6	1	1
200	4	60	4	0	0
225	3	73	2	-1	1
250	2	71	3	1	1
275	1	78	1	0	0
$\sum d^2 = 10$					

- Calculate the coefficient (R) using the formula below. The answer will always be between 1.0 (a perfect positive correlation) and -1.0 (a perfect negative correlation).

When written in mathematical notation, the Spearman Rank formula looks like this:

$$(R) = 1 - \frac{6 \sum d^2}{n^3 - n}$$

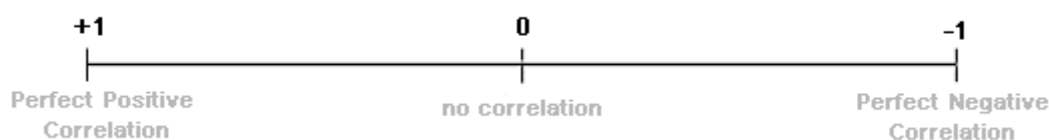
Now put all these values into the formula.

- Find the value of all the d^2 values by adding up all the values in the Difference squared (d^2) column. In our example, this is 10. Multiplying this by 6 gives 60.
- Now for the bottom line of the equation. The value n is the number of sites at which you took measurements. In our example, this is 12. Substituting these values into $n^3 - n$ we get $1728 - 12$
- We now have the formula: $R = 1 - (60/1716)$ which gives a value for R :

$$1 - 0.03 = 0.97$$

What does this R value of 0.97 mean?

The closer R is to +1 or -1, the stronger the likely correlation. A perfect positive correlation is +1 and a perfect negative correlation is -1. The R value of 0.97 suggests a very strong positive relationship.



A further technique is now required to test the significance of the relationship.

The R value of 0.97 must be looked up on the Spearman Rank significance table below as follows:

- Work out the 'degrees of freedom' you need to use. This is the number of pairs in your sample minus 2 ($n - 2$). In the example, it is 10 ($12 - 2$).
- Now plot your result on the table.
- If it is below the line marked 5%, then it is possible your result was the product of chance and you must reject the hypothesis.
- If it is above the 0.1% significance level, then we can be 99.9% confident the correlation has not occurred by chance.
- If it is above 1%, but below 0.1%, you can say you are 99% confident.
- If it is above 5%, but below 1%, you can say you are 95% confident (i.e. statistically there is a 5% likelihood the result occurred by chance).

In the example, the value 0.97 gives a significance level of more than 0.1%. That means that you can be 99.9% certain that your hypothesis is correct.

[The significance of the Spearman's Rank Correlation Coefficients and degrees of freedom.](#)

- The fact that two variables correlate cannot prove anything – only further research can actually prove that one thing affects the other.
- Data reliability is related to the size of the sample. The more data you collect, the more reliable your result.

Click [Spearman's Rank Significance Graph](#) for a blank copy of the significance graph.

- inferential statistics, including Chi-square

The Chi-squared test (χ^2) is used to test whether there is a significant difference between data. It can only be used on data which has the following characteristics:

- The data must be in the form of frequencies counted in a number of groups (% cannot be used).
- The total number of observations must be > 20 .
- The observations must be independent (i.e. one observation must not influence another).
- The expected frequency in any one category must not normally be > 5 .

Method – calculating χ^2 :

- State the hypothesis being tested – there is a significant difference between sample groups. It is convention to give a null hypothesis – no significant difference between the samples.
- Tabulate the data as shown in the example below. The data being tested for significance is the 'observed' frequency and the column headed 'O'.
- Calculate the 'expected' number of frequencies that you would expect to find in the column headed 'E'.
- Calculate the statistic using the formula $\chi^2 = \sum (\text{Observed} - \text{Expected})^2 \div \text{Expected}$
- Calculate the degrees of freedom.
- Compare the calculated figure with the critical values in the significance tables using the appropriate degrees of freedom. Read off the probability that the data frequencies you are testing could have occurred by chance.

Example (using one sample):

In an investigation into the size of material deposited on a beach it was noticed there were differences with increasing distance along the beach, with pebbles appearing to become smaller. χ^2 can be used to test if the variations in pebble size are significant or random. The data in the table below shows the number of pebbles over 5 cm long in a quadrat at 4 sites along a beach between 2 groynes.

Beach site	Observed number of pebbles > 5 cm long
1	40
2	15
3	5
4	12

1. The null hypothesis (H_0) states that there is no significant difference in the sizes of pebbles sampled along the beach. The alternative hypothesis (H_1) is that there is a significant difference in the sizes of pebbles sampled along the beach.
2. If there is no difference in the sizes of pebbles, the sites should all have approximately the same frequency of pebbles > 5cm.
3. Place the data into a table (see below).

	O	E	(O-E)	(O-E) ²	(O-E) ² /E
Beach site	Number of pebbles > 5 cm long	Mean number of pebbles > 5 cm long			
1	40	18	22	484	20.89
2	15	18	3	9	0.5
3	5	18	13	169	9.39
4	12	18	6	36	2
					Σ 38.78

4. Calculate the degrees of freedom (df) = number of rows – 1 = (4 – 1) = 3
5. The critical values for 3 df are:
0.05 (95% confidence level) = 7.82
0.01 (99% confidence level) = 11.34
6. As the calculated value of 38.78 exceeds the tabulated figure at 3 degrees of freedom at the 99% confidence (11.34), it can be stated with 99% confidence that there is a statistically significant difference in pebble size along this stretch of beach.
7. The next stage is to explain the result.

Example (using two samples):

The following figures provide data on the distribution of pebbles of different shapes on the foreshore (intertidal) and storm ridge (top of beach) zones of Pwll Du beach. Pwll Du is a shingle beach on the south-eastern coast of the Gower peninsular.

Sediment shape influences sediment movement. By using the dimensions of the 3 axes, larger beach particles can be placed in one of the 4 shape categories:

Disc – flat and round

Sphere – like a ball

Rod – long and thin

Blade – long and flat

The 3 dimensional shape of a particle influences its movement. Rod and sphere shaped particles roll more easily. Blades can roll, but not as well as rods and spheres, and they are not thrown as effectively as discs.

Pebble shape					ROW TOTAL
	Observed foreshore	Expected foreshore	Observed storm ridge	Expected storm ridge	
Discs	6	23	40		46
Blades	12		31		43
Rods	29		17		46
Spheres	53		12		65
Column Total	100		100		200

For each cell, calculate the Expected value (E) by multiplying the row total by the column total and dividing your result by the overall total. For example, multiply the row total for discs (46) by the column total of pebbles sampled from the foreshore (100), and divide this figure by the total number of 200 pebbles. This gives an expected value of 23.

Pebble shape					ROW TOTAL
	Observed foreshore	Expected foreshore	Observed storm ridge	Expected storm ridge	
Discs	6	23	40	23	46
Blades	12	21.5	31	21.5	43
Rods	29	23	17	23	46
Spheres	53	32.5	12	32.5	65
Column Total	100		100		200

$$\chi^2 = \sum (\text{Observed} - \text{Expected})^2 \div \text{Expected}$$

$$\chi^2 = (6 - 23)^2 \div 23 + (12 - 21.5)^2 \div 21.5 + (29 - 23)^2 \div 23 + (53 - 32.5)^2 \div 32.5 +$$

$$(40 - 23)^2 \div 23 + (31 - 21.5)^2 \div 21.5 + (17 - 23)^2 \div 23 + (12 - 32.5)^2 \div 32.5$$

$$\chi^2 = 12.56 + 4.20 + 1.56 + 12.93 + 12.56 + 4.20 + 1.56 + 12.93$$

$$\chi^2 = 62.5$$

$$\text{Degrees of Freedom} = \text{number of rows} - 1 \times \text{columns} - 1 = 3 \times 1 = 3$$

The tabulated figure at 3 degrees of freedom at the 99% confidence level is 11.34. As the calculated value of 62.5 is above this, it can be stated with 99% confidence that there is a statistically significant difference between the distribution of observed pebble shapes between foreshore and storm ridge.

Table: Chi-Square Probabilities

The areas given across the top are the areas to the right of the critical value. To look up an area on the left, subtract it from one, and then look it up (i.e. 0.05 on the left is 0.95 on the right).

df	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757

Suggest reasons for the result of the Chi-square test:

- cross-sections and long profiles

Beach transects

- Holmes, D. (2010) Beach profiles. *Geography Review* 23 (3) pp.5–7

1.1.7

- cross-sections and long profiles

Cross-section of sand dune *Geography Review* January 2003

Holmes, D. (2003) Investigating coastal sand dunes. *Geography Review* 16 (3) pp.16–20

1.1.8

Ordnance Survey maps (1:25 000) map interpretation of a distinctive landform indicating past sea level change

Glacial Landscapes

1.2.1

- glacier mass balance

Year	Winter (metres of water equivalent)	Summer (metres of water equivalent)	Net glacier budget (metres of water equivalent)
1985	2.18	−3.38	−1.20
1986	2.45	−3.06	−0.61
1987	2.04	−4.10	
1988	2.44	−3.78	
1989	2.43	−3.34	
1990	2.60	−2.71	
1991	3.54	−3.47	
1992	1.91	−3.92	
1993	1.98	−3.21	
1994	2.39	−3.99	
1995	2.86	−3.55	
1996	2.94	−2.84	
1997	3.71	−3.08	
1998	2.76	−4.62	
1999	3.59	−2.57	
2000	3.32	−2.94	
2001	1.90	−3.47	
2002	4.02	−3.47	
2003	2.66	−4.76	
2004	2.08	−3.73	
2005	1.97	−4.42	

- Complete the table by calculating the figures for the net glacial budget column.
- Present the data in the form of a line graph.
- Describe the trends shown by the graph.
- To what extent does the graph support the evidence of glacial retreat?

Use GIS and aerial photo interpretation to calculate mean rates of glacial retreat
<https://nsidc.org/glims/glaciers/melt/>

Retreat of the Sierra de Sangra Glaciers
<http://visibleearth.nasa.gov/view.php?id=87541>

Monitoring of changing areal extent of Swiss glaciers
<http://glaciology.ethz.ch/messnetz/massbalance.html>

1.2.2

1.2.3

1.2.4

- distance and area
- scale

Comparison of past and present distribution of glaciated landscapes: table summarising their volume and extent.

Ice sheet	Area	Present volume	Devensian maximum
Laurentide ice sheet (North America)	$10.2 - 11.3 \times 10^6 \text{ km}^2$	0	$34.8 \times 10^6 \text{ km}^3$
Greenland ice sheet	$1.7 \times 10^6 \text{ km}^2$	$2.4 \times 10^6 \text{ km}^3$	$3.5 \times 10^6 \text{ km}^3$
Antarctica	$14 \times 10^6 \text{ km}^2$	$30 \times 10^6 \text{ km}^3$	$34 \times 10^6 \text{ km}^3$

Types of ice mass at a range of scales

Monitoring of changing area of Swiss glaciers
<http://glaciology.ethz.ch/messnetz/massbalance.html>

NASA's Visible Earth Programme is a source of satellite photographs of landscapes shaped by glaciers
http://disc.sci.gsfc.nasa.gov/geomorphology/GEO_9

1.2.5

- cross-section

Geography Advanced Topic Masters: *Glaciation & Periglaciation*. Author: Jane Knight 144pp • 978 1 844 89617 2

OS map cross-section of Nant Ffrancon valley p.36

- Ordnance Survey maps

Cirque orientation analysis using OS maps and rose diagram (see below).

- field sketches of landforms of glacial erosion

A key field skill for geomorphologists is observation. The ability to observe landforms in the field, to systematically record those observations and then apply classroom knowledge of the environment and process to explain the genesis of the forms observed is central. Producing annotated field sketches is a great way to formalise this process. Annotating photographs in the field using appropriate apps such as Skitch, which helps with the labelling and annotation of photographs, is another option (see Holmes, 2013).

Holmes, D. (2013) Fieldwork of the future. *Geography Review* 26 (4) pp.25–27

1.2.6

- sampling

Sampling of glacial clasts, including the ability to identify sources of error in data, measurement errors and misuse of data

<http://geographyfieldwork.com/MinimumSampleSize.htm>

- data sets

Samples of glacial clasts, see Swain, L. & Kedwards, D. (2007) Analysing glacial deposits. *Geography Review* 20 (5) pp.26–30

- frequencies
- measures of central tendency (mode)

Recorded frequencies of shape of samples of glacial clasts using Power's scale shown in the table below:

Power's scale Frequency		Power's scale Frequency		Power's scale Frequency	
0	8	0	39	0	0
1	37	1	35	1	5
2	41	2	16	2	19
3	11	3	5	3	30
4	3	4	3	4	21
5	0	5	2	5	15
6	0	6	0	6	10

Identify the modal Power's scale for each of the 3 samples tabulated above.

- measures of dispersion (range, standard deviation, interquartile range)

Calculate the range, standard deviation and interquartile range from a sample of glacial clasts.

Smaller standard deviations reflect more clustered data. More clustered data means less extreme values. A data set with less extreme values has a more reliable mean. The standard deviation is therefore a good measure of the reliability of the mean value. The formula is as follows:

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

σ = standard deviation

\sum = sum of

x = each value in the data set

\bar{x} = mean of all values in the data set

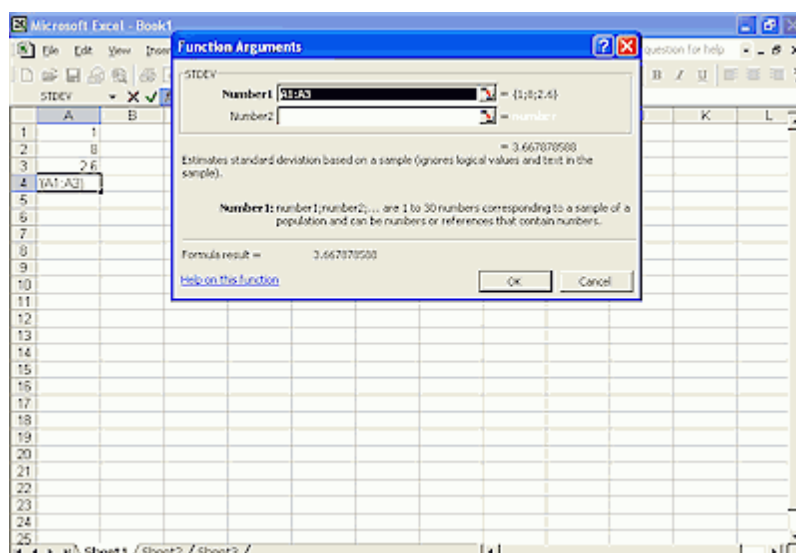
n = number of value in the data set

Is there an easy way to calculate it?

The Microsoft Excel programme will automatically calculate the standard deviation and mean for a set of data listed in a spreadsheet column.

Method:

- List data set in a single column
- Click on the empty cell below the last data item
- Open INSERT menu > FUNCTION > STDEV > click OK
- The standard deviation is then shown and will appear in the empty cell.
- The excel screen example below is for a data set of 3 items



Data set of 30 long axes measurements from a sample of 30 glacial clasts. Calculate the range, standard deviation and interquartile range of the sample.

Pebble number	Long Axis (cm)
1	10
2	9
3	8
4	8
5	16
6	12
7	8.5
8	10
9	12
10	9
11	13
12	14
13	10
14	14
15	17
16	12
17	6
18	17
19	9
20	5

21	10
22	7.5
23	13
24	13
25	7.5
26	15
27	12
28	8
29	22
30	16

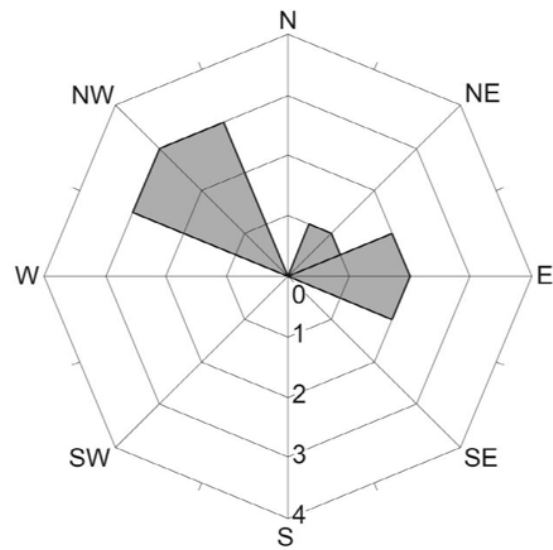
Mean	11.20 cm
Standard Deviation	3.81 cm
Range	5 cm – 22 cm = 17 cm
Interquartile range	8.5 cm – 14 cm = 5.5 cm

Table 11 of named cirques and their orientation

Cirque	Name of cirque	Orientation
A	Craig Maesglas	NE
B	Craig Portas	N
C	Glaslyn	E
D	Llyn Bochlwyd	N
E	Llyn Cau	E
F	Llyn Coch	NW
G	Llyn Du'r Arddu	NW
H	Llyn Gafr	NW
I	Llyn Llydaw	N
J	Llyn y Gadair	N

Use the data in the table above to complete the rose diagram for the orientation of cirques in Wales.

Key: Number of cirques (1 to 4)



1.2.7

- landscape system identification

Classification of glacial landscapes according to landscape character type (LCT)

Holmes, D. (2013) Assessing landscapes. *Geography Review* 27 (2) pp.34–36

1.2.8

Samples of scree deposits were taken at 12 locations along a transect from the top to the base of a scree shown in the table below:

Sampling point along transect	1 (Top of scree)	2	3	4	5	6	7	8	9	10	11	12 (Base of scree)
Distance (m)	0	5	10	15	20	25	30	35	40	45	50	55
Mean length of x-axis (cm)	13	14	22	26	19	27	42	34	60	73	71	78

- Draw a scatter plot to show the relationship between distance (top to bottom of scree) and the mean length of x-axis (cm)
- Draw a line of best fit
- Analyse the statistical significance of the relationship using Spearman's Rank Correlation Coefficient.

The scatter graph shows the possibility of a positive correlation between the two variables and the Spearman's Rank Correlation technique should be used to see if there is indeed a correlation, and to test the strength of the relationship.

- Spearman's Rank Correlation Coefficient

A correlation can easily be drawn as a [scatter graph](#), but the most precise way to compare several pairs of data is to use a statistical test – this establishes whether the correlation is really significant or if it could have been the result of chance alone. Spearman's Rank Correlation Coefficient is a technique which can be used to summarise the strength and direction (negative or positive) of a relationship between two variables.

The result will always be between 1 and minus 1.

Method – calculating the coefficient:

- Create a table from your data.
- Rank the two data sets. Ranking is achieved by giving the ranking '1' to the biggest value in a column, '2' to the second biggest value and so on. The smallest value in the column will get the lowest ranking. This should be done for both sets of measurements.
- Tied scores are given the mean (average) rank.
- Find the difference in the ranks (d): This is the difference between the ranks of the two values on each row of the table. The rank of the second value is subtracted from the rank of the first (distance).
- Square the differences (d^2) to remove negative values and then sum them ($\sum d^2$).

Distance (m)	Rank	Length of x-axis (cm)	Rank	Difference (d)	Difference squared (d^2)
0	12	13	12	0	0
5	11	14	11	0	0
10	10	22	9	-1	1
15	9	26	8	-1	1
20	8	19	10	2	4
25	7	27	7	0	0
30	6	42	5	-1	1
35	5	34	6	1	1
40	4	60	4	0	0
45	3	73	2	-1	1
50	2	71	3	1	1
55	1	78	1	0	0
$\sum d^2 = 10$					

- Calculate the coefficient (R) using the formula below. The answer will always be between 1.0 (a perfect positive correlation) and -1.0 (a perfect negative correlation).

When written in mathematical notation, the Spearman Rank formula looks like this:

$$(R) = 1 - \frac{6 \sum d^2}{n^3 - n}$$

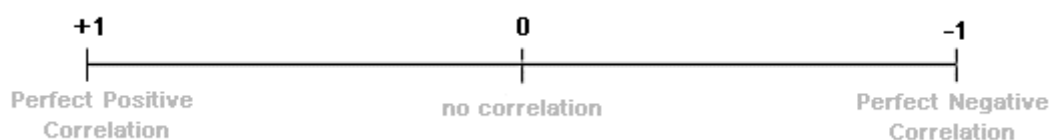
Now put all these values into the formula.

- Find the value of all the d^2 values by adding up all the values in the Difference squared (d^2) column. In our example, this is 10. Multiplying this by 6 gives 60.
- Now for the bottom line of the equation. The value n is the number of sites at which you took measurements. In our example, this is 12. Substituting these values into $n^3 - n$ we get $1728 - 12$
- We now have the formula: $R = 1 - (60/1716)$ which gives a value for R :

$$1 - 0.03 = 0.97$$

What does this R value of 0.97 mean?

The closer R is to +1 or -1, the stronger the likely correlation. A perfect positive correlation is +1 and a perfect negative correlation is -1. The R value of 0.97 suggests a very strong positive relationship.



A further technique is now required to test the significance of the relationship.

The R value of 0.97 must be looked up on the Spearman Rank significance table below as follows:

- Work out the 'degrees of freedom' you need to use. This is the number of pairs in your sample minus 2 ($n - 2$). In the example, it is 10 ($12 - 2$).
- Now plot your result on the table.
- If it is below the line marked 5%, then it is possible your result was the product of chance and you must reject the hypothesis.
- If it is above the 0.1% significance level, then we can be 99.9% confident the correlation has not occurred by chance.
- If it is above 1%, but below 0.1%, you can say you are 99% confident.
- If it is above 5%, but below 1%, you can say you are 95% confident (i.e. statistically there is a 5% likelihood the result occurred by chance).

In the example, the value 0.97 gives a significance level of more than 0.1%. That means that you can be 99.9% certain that your hypothesis is correct.

- The fact that two variables correlate cannot prove anything – only further research can actually prove that one thing affects the other.
- Data reliability is related to the size of the sample. The more data you collect, the more reliable your result.

Click [Spearman's Rank Significance Graph](#) for a blank copy of the above significance graph.

- inferential statistics, including Chi-square

The Chi-squared test (χ^2) is used to test whether there is a significant difference between data. It can only be used on data which has the following characteristics:

- The data must be in the form of frequencies counted in a number of groups (% cannot be used).
- The total number of observations must be > 20 .
- The observations must be independent (i.e. one observation must not influence another).
- The expected frequency in any one category must not normally be > 5 .

Method – calculating χ^2 :

- State the hypothesis being tested – there is a significant difference between sample groups. It is convention to give a null hypothesis – no significant difference between the samples.
- Tabulate the data as shown in the example below. The data being tested for significance is the 'observed' frequency and the column headed 'O'.
- Calculate the 'expected' number of frequencies that you would expect to find in the column headed 'E'.
- Calculate the statistic using the formula $\chi^2 = \sum (\text{Observed} - \text{Expected})^2 \div \text{Expected}$
- Calculate the degrees of freedom.
- Compare the calculated figure with the critical values in the significance tables using the appropriate degrees of freedom. Read off the probability that the data frequencies you are testing could have occurred by chance.

Example (using one sample):

The following figures provide data on the number of cirques and their orientation.

Orientation of cirques	Number of cirques
NE	40
SE	15
SW	5
NW	12

1. The null hypothesis (H_0) states that there is no significant difference in the orientation of cirques sampled. The alternative hypothesis (H_1) is that there is a significant difference in the orientation of cirques.
2. If there is no difference in the orientation of cirques, they should all have approximately the same frequency.

3. Place the data into a table (see below).

	O	E	(O-E)	(O-E) ²	(O-E) ² /E
Orientation of cirques	Number of cirques	Mean number of cirques			
NE	40	18	22	484	20.89
SE	15	18	3	9	0.5
SW	5	18	13	169	9.39
NW	12	18	6	36	2
					Σ 38.78

4. Calculate the degrees of freedom (df) = number of rows – 1 = (4 – 1) = 3

5. The critical values for 3 df are:

0.05 (95% confidence level) = 7.82

0.01 99% confidence level) = 11.34

6. As the calculated value of 38.78 exceeds the tabulated figure at 3 degree of freedom at the 99% confidence (11.34), it can be stated with 99% confidence that there is a statistically significant difference in the frequency of cirques and their orientation.

7. The next stage is to explain the result.

Example (using two samples):

The following figures provide data on the distribution of scree deposits of different sizes on the upper part and lower part of the scree at Mewslade. Mewslade is a dry valley on the south-western coast of the Gower peninsular.

Scree size (long axis cm)					ROW TOTAL
	Observed	Expected	Observed	Expected	
	20–24.9 m from free face	lower scree	0–4.9 m from free face	upper scree	
0–50	6	23	40		46
51–100	12		31		43
101–150	29		17		46
151–200	53		12		65
Column Total	100		100		200

For each cell, calculate the Expected value (E) by multiplying the row total by the column total and dividing your result by the overall total. For example, multiply the row total for scree particles between 0–50 cm 20–24.9 m from the free face (46) by the column total of scree particles 20–24.9 m from the free face (100), and divide this figure by the total number of scree particles (200). This gives an expected value of 23.

Scree size (long axis cm)					ROW TOTAL
	Observed 20–24.9 m from free face	Expected lower scree	Observed 0–4.9 m from free face	Expected upper scree	
0–50	6	23	40	23	46
51–100	12	21.5	31	21.5	43
101–150	29	23	17	23	46
151–200	53	32.5	12	32.5	65
Column Total	100		100		200

$$\chi^2 = \sum (\text{Observed} - \text{Expected})^2 \div \text{Expected}$$

$$\chi^2 = (6 - 23)^2 \div 23 + (12 - 21.5)^2 \div 21.5 + (29 - 23)^2 \div 23 + (53 - 32.5)^2 \div 32.5 +$$

$$(40 - 23)^2 \div 23 + (31 - 21.5)^2 \div 21.5 + (17 - 23)^2 \div 23 + (12 - 32.5)^2 \div 32.5$$

$$\chi^2 = 12.56 + 4.20 + 1.56 + 12.93 + 12.56 + 4.20 + 1.56 + 12.93$$

$$\chi^2 = 62.5$$

$$\text{Degrees of Freedom} = \text{number of rows} - 1 \times \text{columns} - 1 = 3 \times 1 = 3$$

The tabulated figure at 3 degrees of freedom at the 99% confidence level is 11.34. As the calculated value of 62.5 is above this, it can be stated with 99% confidence that there is a statistically significant difference between the size of scree deposits on the upper part and lower part of the scree.

Table: Chi-Square Probabilities

The areas given across the top are the areas to the right of the critical value. To look up an area on the left, subtract it from one, and then look it up (i.e. 0.05 on the left is 0.95 on the right).

df	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	---	---	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757

Suggest reasons for the result of the Chi-square test.

Spare tables re: glacial clasts below:

Orientation (°)	Frequency	Orientation (°)	Frequency	Orientation (°)	Frequency
1–30	5	1–30	4	1–30	3
31–60	10	31–60	2	31–60	3
61–90	14	61–90	3	61–90	8

91–120	11	91–120	11	91–120	15
121–150	15	121–150	19	121–150	17
151–180	3	151–180	5	151–180	8
181–210	6	181–210	4	181–210	2
211–240	2	211–240	1	211–240	3
241–270	14	241–270	6	241–270	5
271–300	11	271–300	11	271–300	12
301–330	4	301–330	19	301–330	17
331–360	5	331–360	15	331–360	7

	Till sample 1		Till sample 2	
Clast number	Orientation (°)	Length of A-axis (cm)	Orientation (°)	Length of A-axis (cm)
1	100	4.8	162	3.9
2	70	6.9	17	3.5
3	95	5.0	51	7.6
4	54	20.5	121	7.0
5	70	22.0	126	8.0
6	85	11.5	32	7.2
7	227	6.0	32	7.2
8	225	7.5	14	10.9
9	232	10.5	139	3.6
10	170	4.5	120	11.0
11	80	7.0	156	9.6
12	100	6.0	89	10.0

13	121	15.5	18	4.8
14	120	12.0	58	12.2
15	152	9.8	149	7.0
16	104	6.0	100	7.0
17	166	5.6	100	11.4
18	100	7.0	61	31.0
19	120	6.5	72	4.0
20	120	4.0	140	5.9