A Level Geology

Special Module

Maths for geology

Topics:

1. *Numbers*
2. *Shapes*
3. *Statistics & analysis*

Numbers

Standard form

In geology we often encounter both very large and very small numbers e.g. the age of the Earth and the diameter of a clay mineral. It is impractical to write the age of the Earth as 454 000 000 000 years old or the diameter of a clay mineral as 0.0 00039 m, so standard form is used instead. This is the conventional way of writing both large and small numbers and has the additional advantage of simplifying calculations by enabling the use of the index laws.

To convert large numbers to standard index form, count the zeros in the number. This gives the value of the index of 10. Then make any adjustment required so that the number in front lies between 1 and 10.

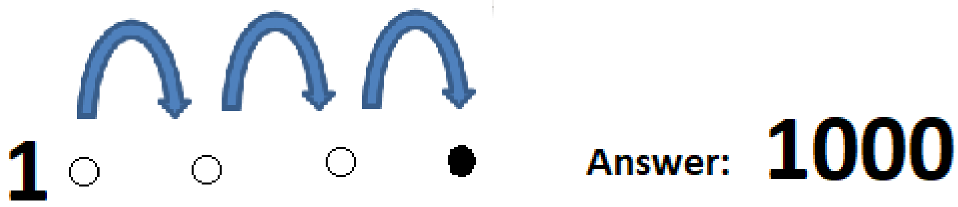
e.g. 4 540 000 000 years is 454 followed by 7 zeros = $454 \times 10^7 = 4.54 \times 10^9$ years

To convert small numbers to standard index form, count the zeros in the number, including the zero before the decimal point. This gives the value of the index of 10, which will be negative if the number is less than one. The digits in front of the index should be written to lie between 1 and 10.

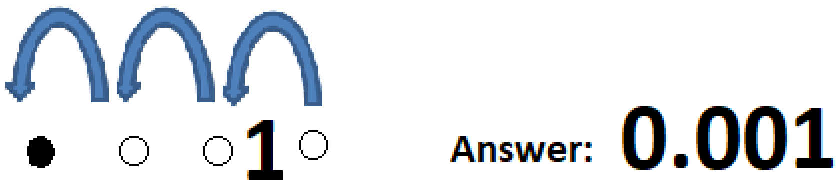
e.g. 0.00000391 m is 6 zeros followed by 391 = 3.91×10^{-6} m

A simple way of doing this:

1 x 10^3 the 3 tells you to make the decimal point jump 3 places to the right. Then fill up the gaps with 0.



1 x 10^{-3} the -3 tells you to make the decimal point jump 3 places to the left. Fill up any gaps with 0.



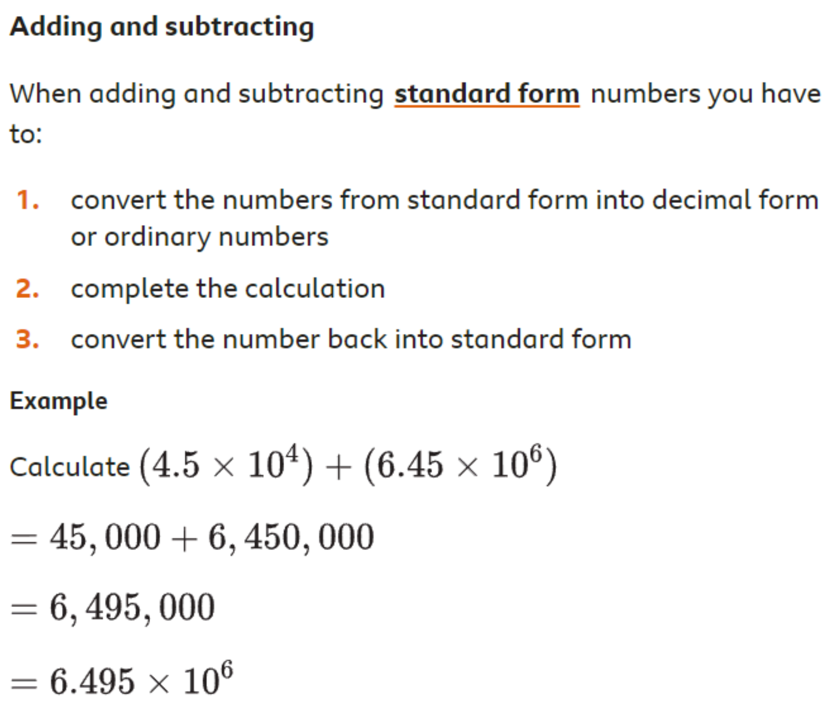
Exercise 1

Each of these objects has been measured in metres. Change from standard form for:

Frog egg 1×10^{-3}

Bacteria 1×10^{-6}

Flu virus 1×10^{-7}



Adding and subtracting

When adding and subtracting **standard form** numbers you have to:

1. convert the numbers from standard form into decimal form or ordinary numbers
2. complete the calculation
3. convert the number back into standard form

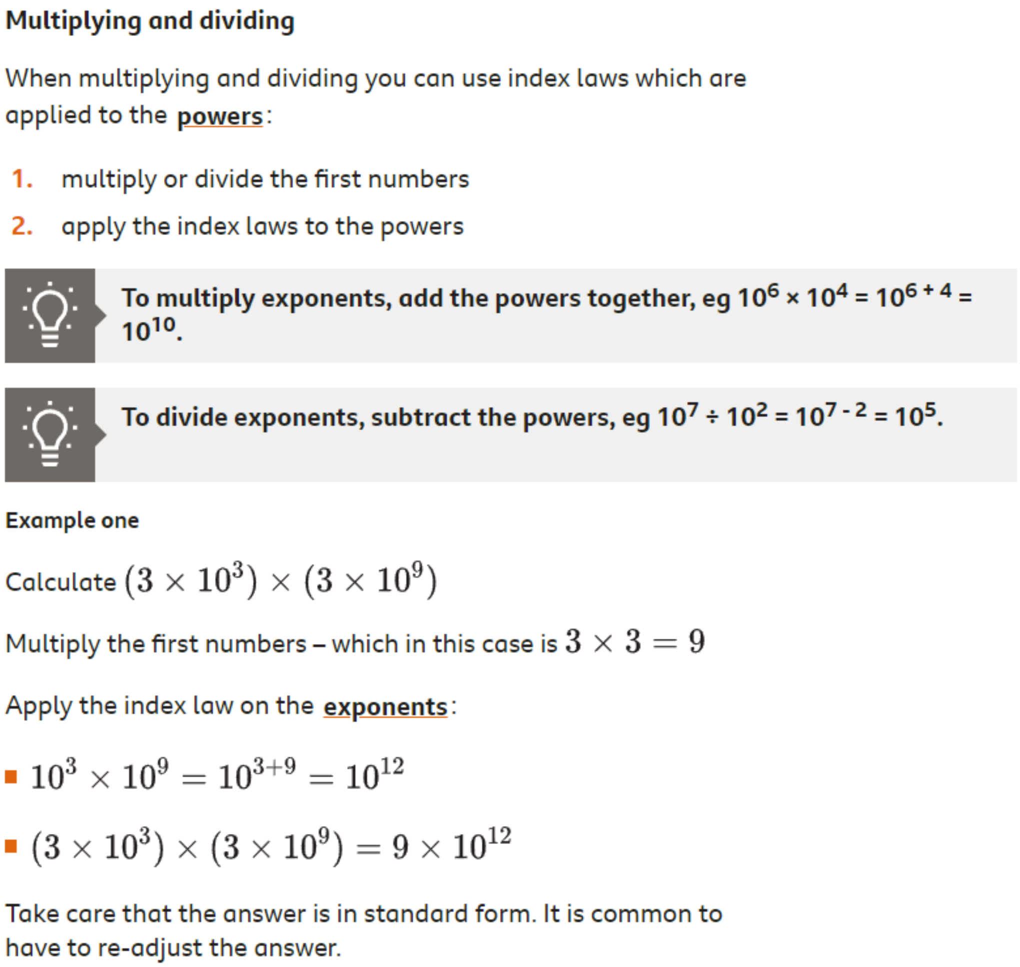
Example

Calculate $(4.5 \times 10^4) + (6.45 \times 10^6)$

$= 45,000 + 6,450,000$

$= 6,495,000$

$= 6.495 \times 10^6$



Multiplying and dividing

When multiplying and dividing you can use index laws which are applied to the **powers**:

1. multiply or divide the first numbers
2. apply the index laws to the powers

To multiply exponents, add the powers together, eg $10^6 \times 10^4 = 10^{6+4} = 10^{10}$.

To divide exponents, subtract the powers, eg $10^7 \div 10^2 = 10^{7-2} = 10^5$.

Example one

Calculate $(3 \times 10^3) \times (3 \times 10^9)$

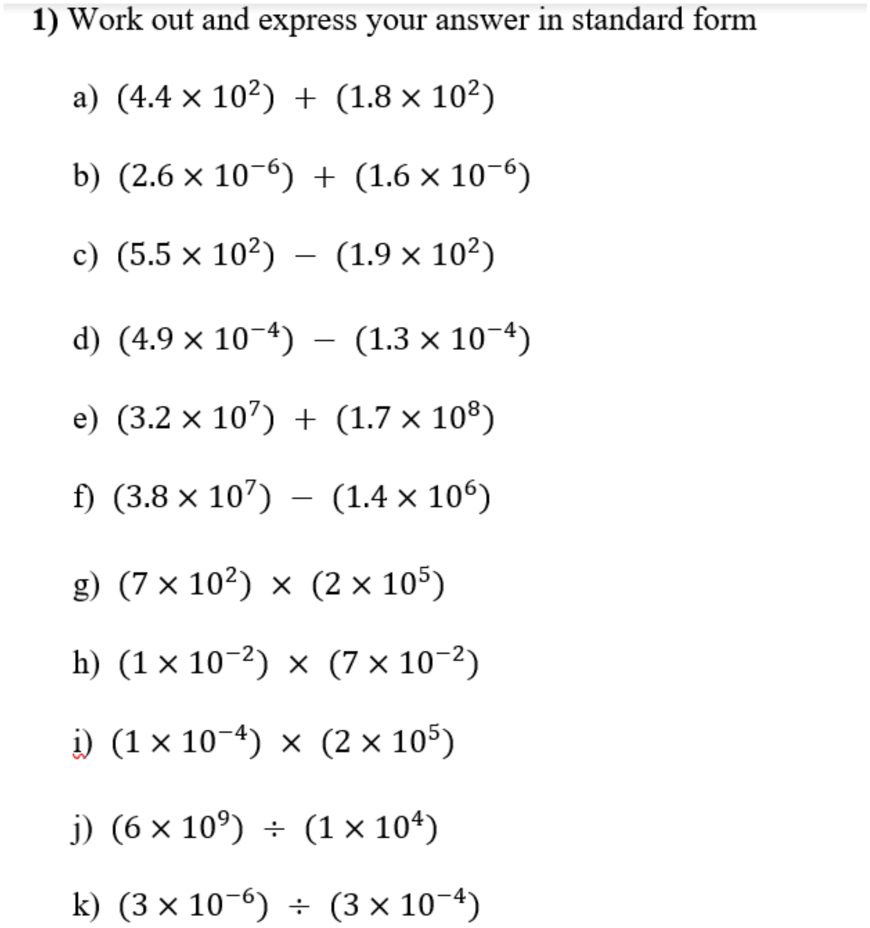
Multiply the first numbers – which in this case is $3 \times 3 = 9$

Apply the index law on the **exponents**:

- $10^3 \times 10^9 = 10^{3+9} = 10^{12}$
- $(3 \times 10^3) \times (3 \times 10^9) = 9 \times 10^{12}$

Take care that the answer is in standard form. It is common to have to re-adjust the answer.

Exercise 2



1) Work out and express your answer in standard form

a) $(4.4 \times 10^{2}) + (1.8 \times 10^{2})$

b) $(2.6 \times 10^{-6}) + (1.6 \times 10^{-6})$

c) $(5.5 \times 10^{2}) - (1.9 \times 10^{2})$

d) $(4.9 \times 10^{-4}) - (1.3 \times 10^{-4})$

e) $(3.2 \times 10^{7}) + (1.7 \times 10^{8})$

f) $(3.8 \times 10^{7}) - (1.4 \times 10^{6})$

g) $(7 \times 10^{2}) \times (2 \times 10^{5})$

h) $(1 \times 10^{-2}) \times (7 \times 10^{-2})$

i) $(1 \times 10^{-4}) \times (2 \times 10^{5})$

j) $(6 \times 10^{9}) \div (1 \times 10^{4})$

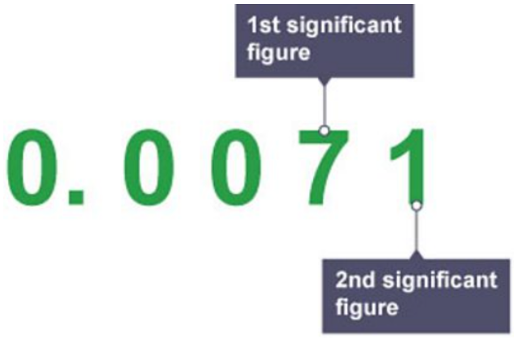
k) $(3 \times 10^{-6}) \div (3 \times 10^{-4})$

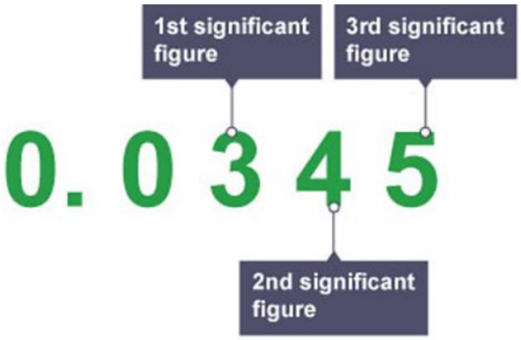
Significant figures

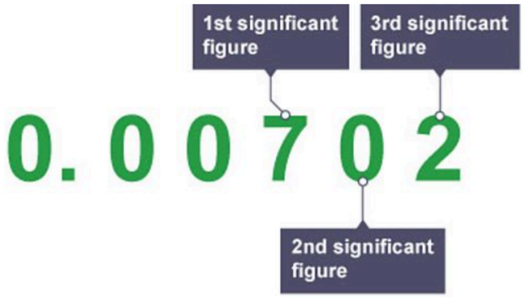
Numbers are often rounded to avoid reporting insignificant figures. For example, it would create false precision to express a measurement as 12.34500 kg (which has seven significant figures) if the scales only measured to the nearest gram and gave a reading of 12.345 kg (which has five significant figures).

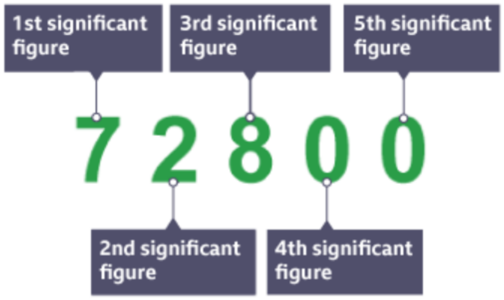
Significant figures start at the **first non-zero number**, so ignore the zeros at the front, but not the ones in between.

Look at the following examples:









Exercise 3

How many significant figures in these numbers:

0.046

4006

7.90

Round 46.603 to 2 significant figures.

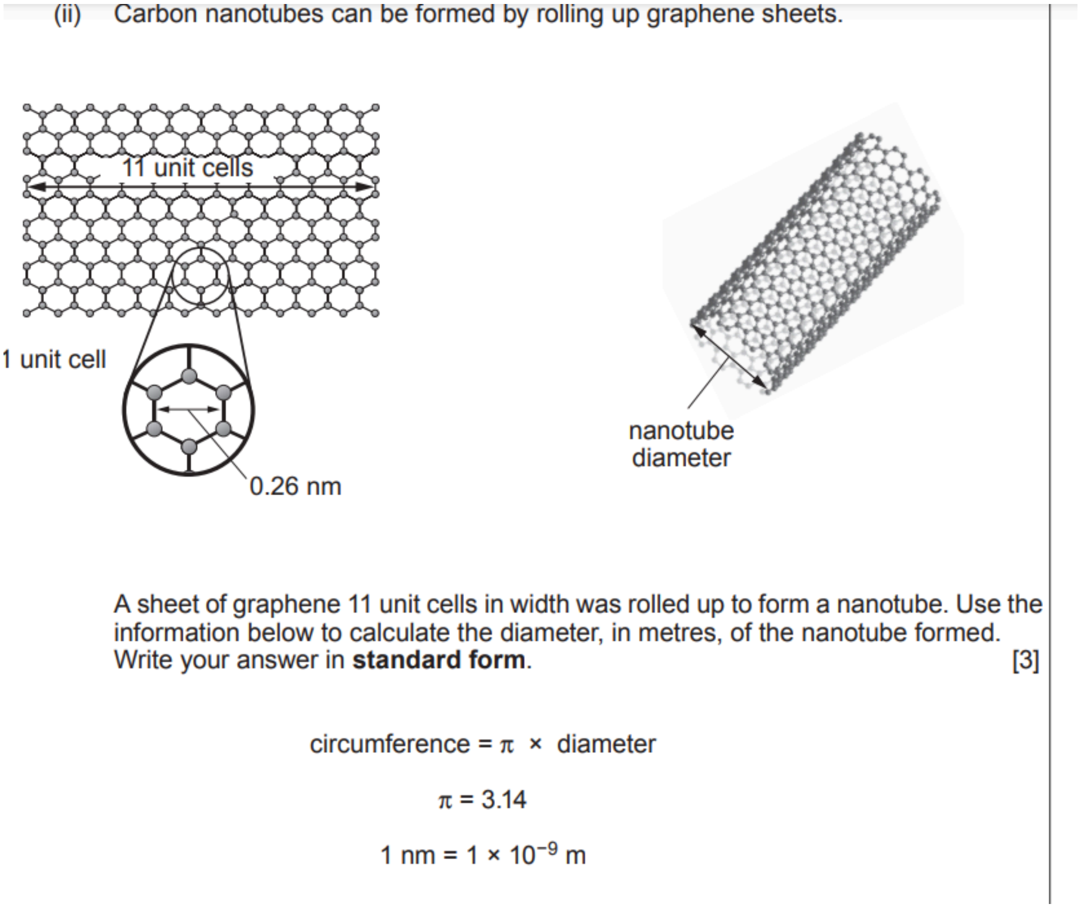
Round 0.04518 to 3 significant figures.

Orders of magnitude

Orders of magnitude can be determined easily when a number is written in standard form. For example, 237 (2.37×10^2) and 823 (8.23×10^2) both have an order of magnitude of 2.

A scanning electron microscope may produce a magnification of up to 40 000 (4×10^4) whereas a typical optical laboratory microscope may produce a magnification of just 40 (4×10^1). A scanning electron microscope therefore produces an image 4 orders of magnitude bigger than the object and 3 orders of magnitude bigger than an optical microscope.

Exercise 4



(ii) Carbon nanotubes can be formed by rolling up graphene sheets.

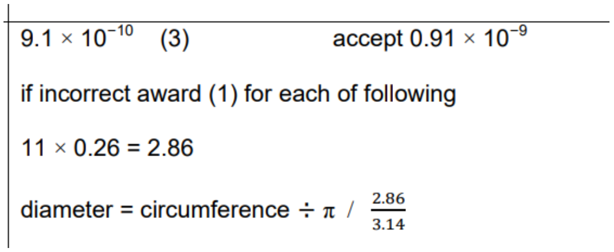
A sheet of graphene 11 unit cells in width was rolled up to form a nanotube. Use the information below to calculate the diameter, in metres, of the nanotube formed. Write your answer in **standard form**. [3]

$$\text{circumference} = \pi \times \text{diameter}$$

$$\pi = 3.14$$

$$1\text{ nm} = 1 \times 10^{-9}\text{ m}$$

Mark scheme



9.1×10^{-10} (3) accept 0.91×10^{-9}

if incorrect award (1) for each of following

$11 \times 0.26 = 2.86$

diameter = circumference ÷ π / $\frac{2.86}{3.14}$

Uncertainties in measurements

When working with a single reading it is recommended that the uncertainty is taken to be ± the smallest measuring division which of course depends on the scale of the measuring instrument. Suggestions for everyday measuring instruments ±1 mm for a metre rule, ±1g for a balance, ± 1 ^{o}C for a laboratory thermometer and ± 1 cm^{3} for a typical measuring cylinder but professional judgement is important.

Occasionally the uncertainty in a single value may be provided e.g. the absolute age of a particular rock unit e.g. the radiocarbon 14C age of 11 750 ± 120 year B.P. for an organic lake sediment.

Percentage uncertainty is a measure of the uncertainty of a measurement compared to the size of the measurement, expressed as a percentage. The calculation is derived by dividing the uncertainty of the experiment into the total value of the measurement and multiplying it by 100.

The percentage uncertainty in the radiocarbon 14C age of 11 750 ± 120 year B.P. for an organic lake sediment is therefore:

$$\% \text{ uncertainty } p = \frac{120}{11750} \times 100 = 1.0\%$$

In geology, it is of course best practice to undertake repeat measurements whenever possible. Suppose the value of a quantity x is measured several times and a series of different values obtained: x1, x2, x3........xn.

Unless there is reason to suspect that one of the results is anomalous, the best thing to do is to calculate the mean of these values. The uncertainty is calculated by taking the range and dividing it by 2.

For example, the following results were obtained for the thickness of a sedimentary bed;
4.5 cm, 4.8 cm, 4.6 cm, 5.1 cm, 5.0 cm.

The mean is:
$(4.5 + 4.8 + 4.6 +5.1+5.0) / 5 = 4.8$ cm

The uncertainty, u, is given by:
$(5.1- 4.5) / 2 = 0.3$ cm

The final answer and uncertainty should be quoted, with units, to the same no. of decimal places, i.e.
Mean bed thickness = 4.8 ± 0.3 cm.

The percentage uncertainty in the mean bed thickness is therefore
$(0.3 / 4.8) \times 100 = 6.3\%$.

Exercise 5

The dip was measured across an area and the results are shown below: 30^o 32^o 35^o 35^o 40^o 42^o

(a) What is the uncertainty on the measurement for 42^o ?

(b) What is the mean dip value? (Quote this to a sensible value)

(c) What is the range?

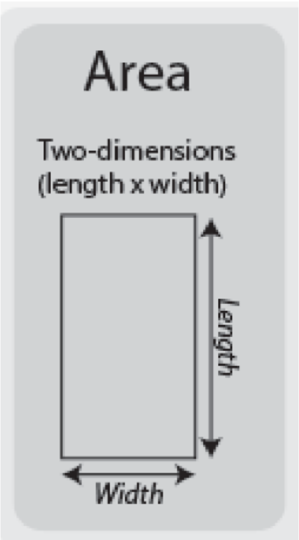
(d) Calculate the uncertainty.

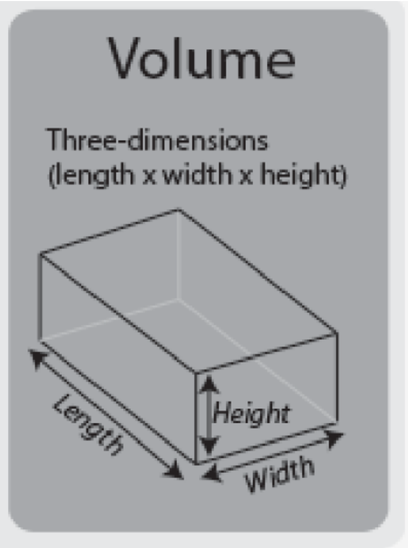
(e) Calculate the percentage uncertainty.

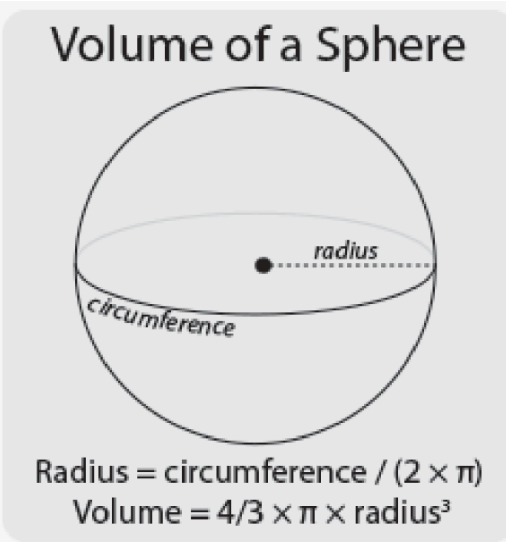
(f) Present your answer for the mean dip in the most appropriate way.

Shapes

Areas and volumes







The perimeter of the rectangle = 2 × (length + width)

The area of the rectangle = length × width

The volume of the cuboid = length × width × height

The surface area of the cuboid = 2 × ((length × width) + (width × height) + (length × height))

The circumference of the circle = $2 \times \pi \times \text{radius}$

The area of the circle = $\pi \times (\text{radius})^2$

The volume of the sphere = $\frac{4}{3} \times \pi \times (\text{radius})^3$

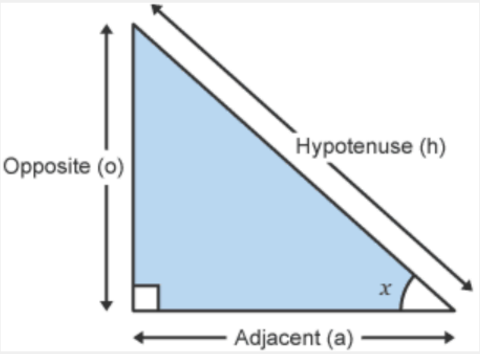
The surface area of the sphere = $4 \times \pi \times (\text{radius})^2$

Trigonometry

The **hypotenuse** (h) is the longest side. It is opposite the right angle.

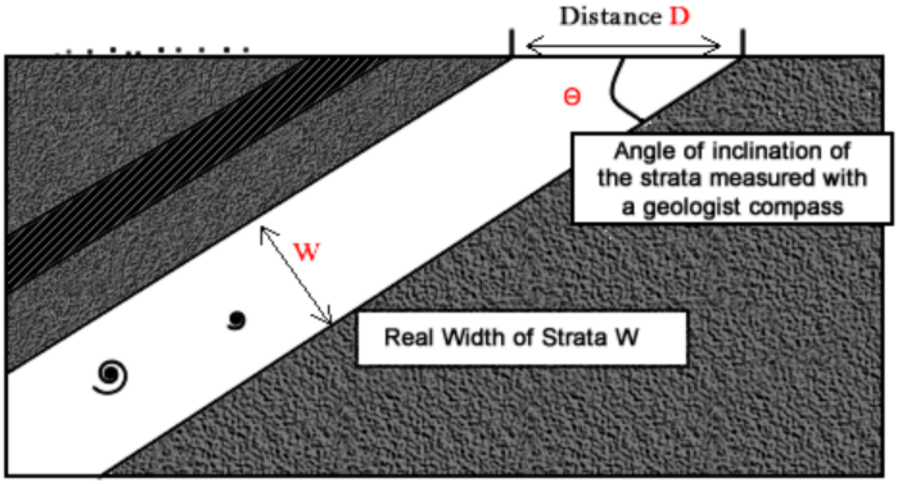
The **opposite side** (o) is opposite the angle in question (x).

The **adjacent side** (a) is next to the angle in question (x).



- $\sin x = \dfrac{\text{opposite}}{\text{hypotenuse}}$
- $\cos x = \dfrac{\text{adjacent}}{\text{hypotenuse}}$
- $\tan x = \dfrac{\text{opposite}}{\text{adjacent}}$

Exercise 6

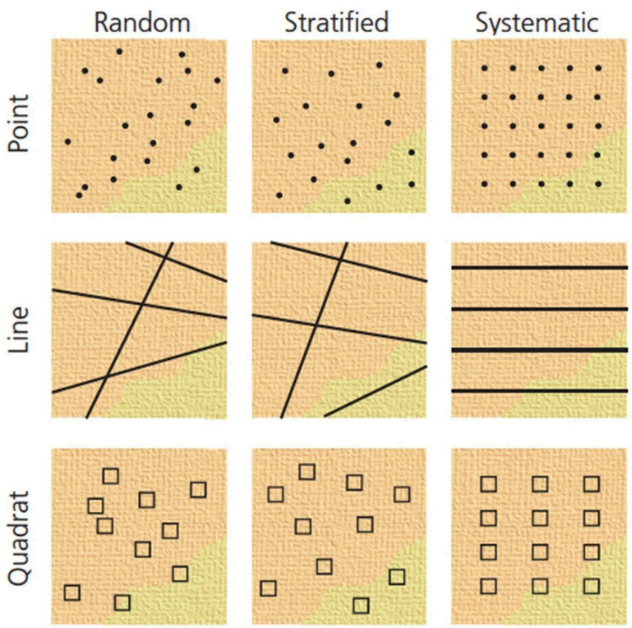


If the outcrop width is 50m and the dip is 30^{o} what is the true thickness?

Statistics and analysis

Sampling

Sampling should be used to collect reliable data. There are three main types of sampling: random, systematic and stratified.



In random sampling every item has an equal chance of being selected. For many studies this is the most desirable approach as there is no bias. The most common way of random sampling is to use a random number table or generator.

In systematic sampling there is some structure or underlying order to the way in which the data is selected.

With stratified sampling the population is purposely split into separate groups/layers (strata). Then each group is further analysed by random or systematic sampling.

Exercise 7

(c) **Figure 1b** is a map showing copper concentrations in soil in parts per million (ppm) in Area **A** on **Figure 1a**.

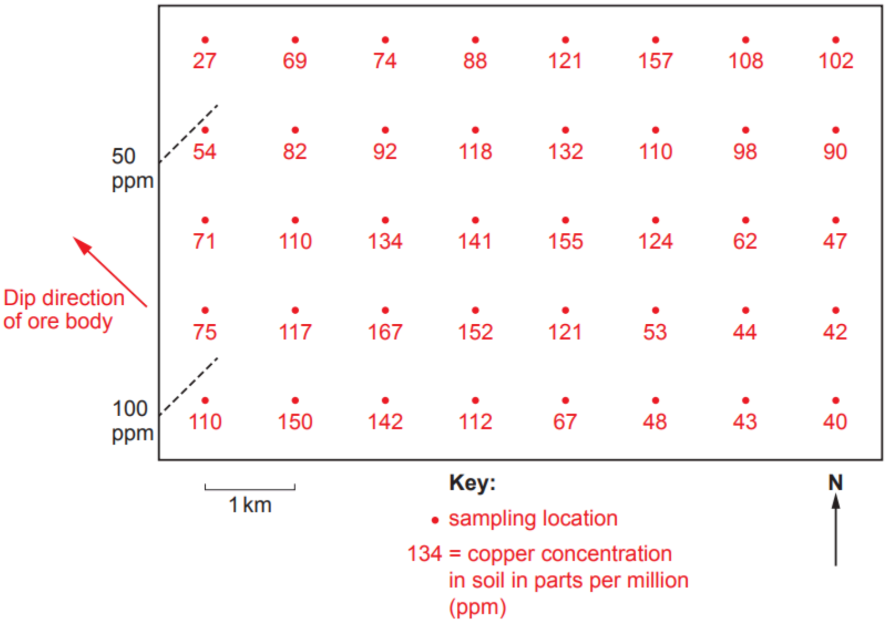


Figure 1b

(i) State which sampling method was used to collect the geochemical data shown in **Figure 1b**. **Tick** (✓) only **one** box. [1]

Cluster sampling ☐

Stratified sampling ☐

Random sampling ☐

Systematic sampling ☐

(ii) Complete **Figure 1b** to show the isolines representing:

- 50 ppm copper concentration in the soil
- 100 ppm copper concentration in the soil. [2]

(i)	Systematic sampling (1)
(ii)	50ppm and 100ppm lines on western side correctly continued (1) 50ppm and 100ppm lines on eastern side drawn (1)

Data analysis - frequency tables

When you collect data you need a method of analysis. Here is a set of results of testing the strength of a particular bed of rock. The results have been obtained by a sampling method.

Number	Schmidt hammer hardness	Number	Schmidt hammer hardness	Number	Schmidt hammer hardness	Number	Schmidt hammer hardness	Number	Schmidt hammer hardness
1	41	*11*	43	*21*	39	*31*	34	*41*	32
2	38	*12*	33	*22*	33	*32*	37	*42*	38
3	44	*13*	32	*23*	43	*33*	36	*43*	36
4	38	*14*	36	*24*	34	*34*	38	*44*	39
5	31	*15*	46	*25*	36	*35*	36	*45*	40
6	37	*16*	35	*26*	36	*36*	40	*46*	38
7	30	*17*	33	*27*	34	*37*	36	*47*	36
8	29	*18*	36	*28*	41	*38*	38	*48*	41
9	40	*19*	31	*29*	38	*39*	37	*49*	42
10	32	*20*	33	*30*	34	*40*	35	*50*	49

The first step is to group the data into classes, to make a frequency table - you can do this by using a tally chart.

Exercise 8

Schmidt hammer hardness (SHH) class			Frequency
1	28<SHH≤31		4
2	31<SHH≤34		11
3	34<SHH≤37		14
4	37<SHH≤40		12
5	40<SHH≤43		6
6	43<SHH≤46		2
7	46<SHH≤49		1

Data analysis - histograms

The next step is to use a suitable graph - for example a histogram:

Exercise 9

Draw a histogram for your frequency table.

Beware of histograms - the area of the bar is proportional to the frequency, so you could draw this histogram like this:

Data analysis - cumulative frequency curves

Graphs are often used in the analyses of sediment grain sizes. A sieve stack is used to separate the grain sizes into different sized fractions. Each fraction is then weighed.

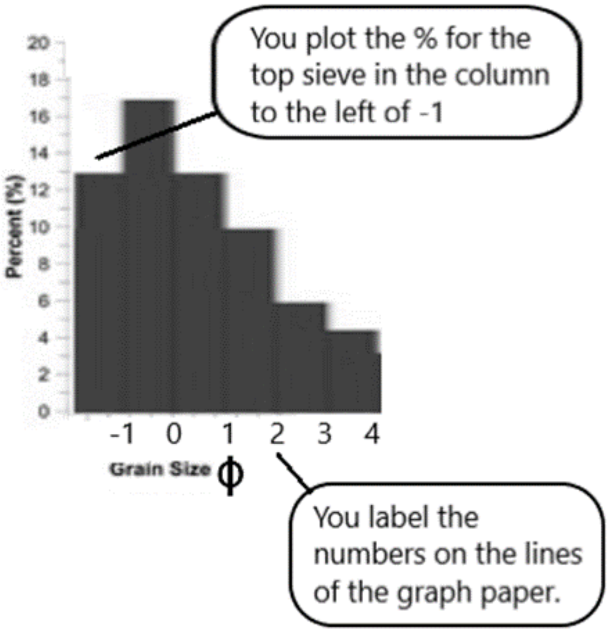
Sieve size (mm)	Phi scale	Mass (g)	Mass % of sample	Cumulative mass (g)	Cumulative %
2	-1	1.0	1.01	1.0	1.01
1	0	3.0	3.03	4.0	4.04
0.5	1	45.0	45.45	49.0	49.49
0.25	2	26.0	26.26	75.0	75.75
0.125	3	21.5	21.72	96.5	97.47
0.063	4	2.5	2.53	99.0	100.00

Mass

Work out the % of each fraction

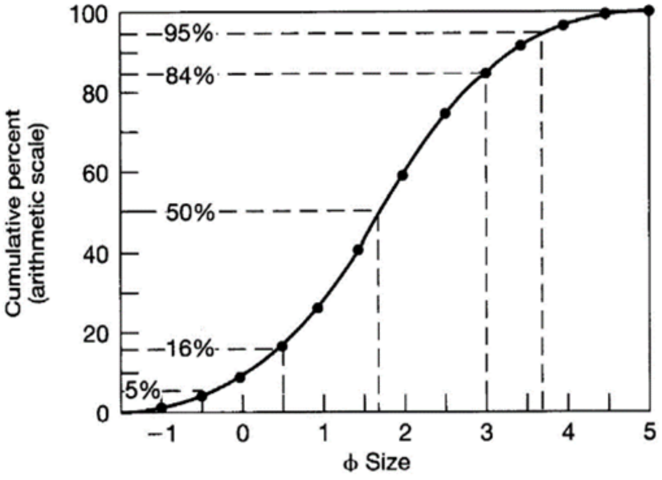
Running total of masses

Running total of %



You plot the % for the top sieve in the column to the left of -1

You label the numbers on the lines of the graph paper.



Exercise 10

Draw a cumulative frequency curve for the results of the sediment analysis above.

Exercise 11

(b) **Table 3** shows the range in size of grains in sediments **A** and **B** in **Figure 3a**. **Figure 3b** shows the cumulative frequency curve for sediment **B**.

	Sediment **A**		Sediment **B**	
Grain Size (mm)	Frequency	Cumulative frequency	Frequency	Cumulative frequency
0.125 – 0.25	15	15	5	•
0.25 – 0.5	25	40	15	•
0.5 – 1	60	100	30	•
1 - 2	0	100	50	•

Table 3

(i) Complete the cumulative frequency column for sediment **B** in **Table 3**. [1]

(ii) Using the information in **Table 3** plot a cumulative frequency curve for sediment **A**. [2]

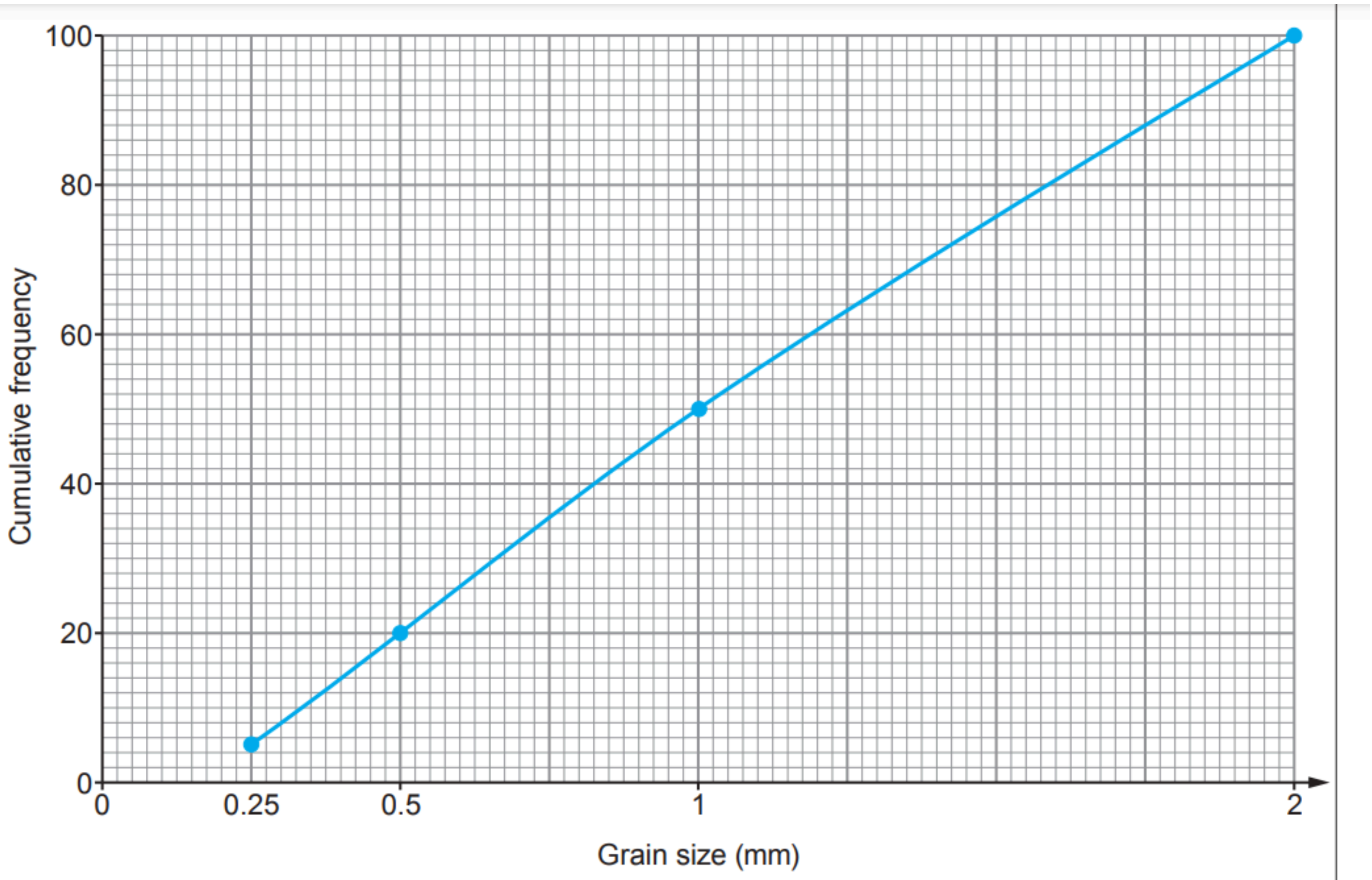


Figure 3b

(i)

	Sediment A		Sediment B	
Size (mm)	Percentage	Cumulative Percentage	Percentage	Cumulative Percentage
0.125 – 0.25	15	15	5	• **5**
0.25 – 0.5	25	40	15	• **20**
0.5 – 1	60	100	30	• **50**
1 - 2	0	100	50	• **100**

All correct (1)

(ii)

4 correct points (2)
3 correct points (1)
2 or 1 correct points (0)

Describing the distribution of your data - averages:

Mean: To find the mean, add up the values in the data set and then divide by the number of values that were added.

Mode: The mode of a sample is the most frequently occurring value. When data is grouped into classes, the modal class is the class containing the greatest number of values. It is possible to have two modes (bimodal). In the example the histogram clearly shows the data is unimodal with the modal class being 35-37 and the mode as 36.

Median: The median of a sample is the value that evenly splits the number of observations into a lower half of smaller observations and an upper half of larger measurements.

Extreme values: The extreme values are the maximum and minimum values in the sample. In the example the minimum value is 29 and the maximum value is 49, hence the range is 49 – 29 = 20.

Describing the distribution of your data - quartiles

Quartiles (including the interquartile range): Split the data into 4 equal-sized groups. In the example, the boundaries between the quartiles are: 34, 36 and 39 - shaded in the table below:

Number	Schmidt hammer hardness	Number	Schmidt hammer hardness	Number	Schmidt hammer hardness	Number	Schmidt hammer hardness	Number	Schmidt hammer hardness
8	29	*22*	33	*26*	36	*4*	38	*45*	40
7	30	*24*	34	*33*	36	*29*	38	*1*	41
5	31	*27*	34	*35*	36	*34*	38	*28*	41
19	31	*30*	34	*37*	36	*38*	38	*48*	41
10	32	*31*	34	*43*	36	*42*	38	*49*	42
13	32	*16*	35	*47*	36	*46*	38	*11*	43
41	32	*40*	35	*6*	37	*21*	39	*23*	43
12	33	*14*	36	*32*	37	*44*	39	*3*	44
17	33	*18*	36	*39*	37	*9*	40	*15*	46
20	33	*25*	36	*2*	38	*36*	40	*50*	49

Interquartile range (IQR): The interquartile range is the difference between the upper and lower quartiles thereby giving a measure of the central spread of the data. A practical rule of thumb is to regard any value deviating more than 1.5 times the IQR from the median as a mild outlier and any value deviating more than 3 times the IQR from the median as an extreme outlier. Outliers are values so markedly different from the rest of the sample that they raise the suspicion that they may be from a different population or may be in error but it is notoriously difficult to show that the values are anomalous.

In the example above the IQR is 39 – 34 = 5.
The mild outlier boundaries are = median ± 1.5 IQR = 36 ± 1.5(5) = 29 and 44.
Therefore only two values (sample 15 and 50) may be considered as mild outliers with sample number 50 being the most extreme.

In comparison to variance/standard deviation (discussed below) the IQR is a more robust method for analysing the central spread of the measurements but, unlike variance/standard deviation, is insensitive to the lower and upper tails. Generally speaking if the median is thought to be the best way in which to describe the data average then the IQR is used as the measure of spread. Conversely if the mean is believed to be the best way in which to describe the data average then the standard deviation is utilised.

Exercise 12

The diameter of the pebbles in a conglomerate are measured and the results are shown below:

Pebble number	**Diameter (mm)**
1	7
2	13
3	5
4	8
5	15
6	3
7	9

Calculate the range.

Exercise 13

There are problems using the range - it doesn't take account of outliers.

1, 1, 3, 12, 2, 4, 5, 2, 1, 1, 6, 3, 4

Work out the range for these data.

Now work out the range by leaving out the outlier.

Exercise 14

Pebble number	**Diameter (mm)**
1	7
2	13
3	5
4	8
5	15
6	3
7	9

Write out the numbers in order of size.

Find the median.

Exercise 15

A different way to determine the median is to use (n+ 1) / 2 where n is the number of numbers.
Find the median in this way.

Exercise 16

Another way to describe the spread of the data is to use quartiles and the interquartile range.

1, 3, 4, 6, 9, 14, 15, 17, 18, 22, 60

Find the lower quartile using (n+1) / 4

Exercise 17

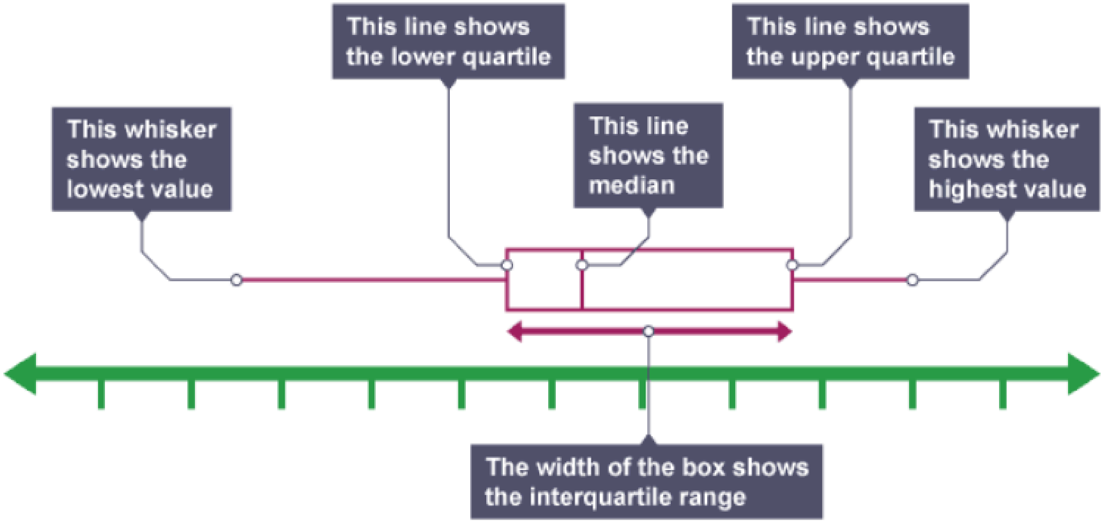
1, 3, 4, 6, 9, 14, 15, 17, 18, 22, 60

Find the upper quartile using 3x (n+1) / 4

Exercise 18

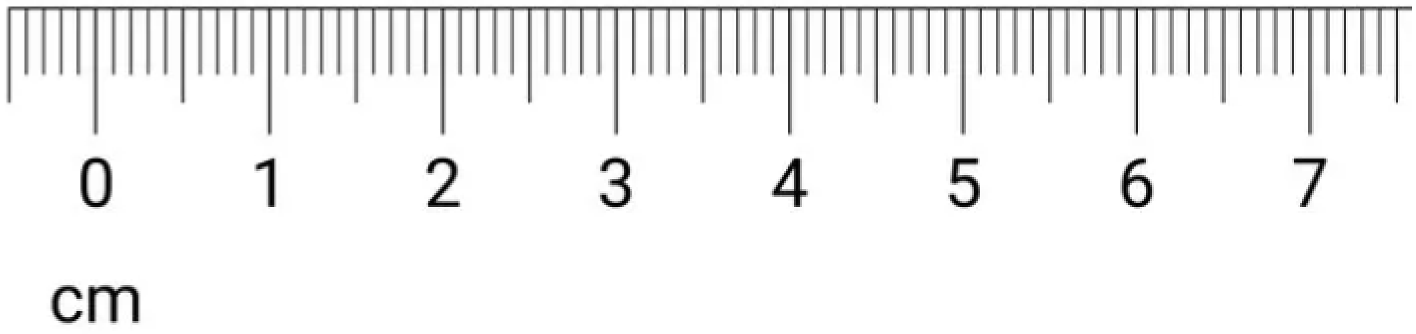
Use your answers to exercise 16 & 17 to find the interquartile range.

The box and whisker plot can be used to show all of the information we have worked out.



Exercise 19

Use the scale below to draw the box and whisker plot for your answers to the previous exercises.



Exercise 20

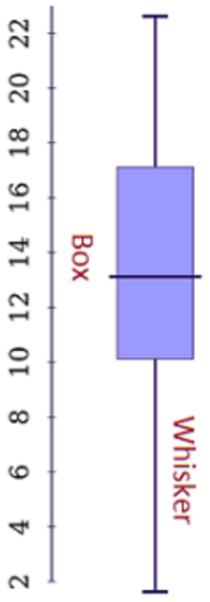
From the box and whisker plot below, determine the

(a) median

(b) upper quartile value

(c) lower quartile value

(d) interquartile range



Describing the distribution of your data - further techniques

Variance: The sample variance, s^2, is another method used to calculate how varied or spread out from the mean a sample is. Sample variance is mathematically defined as the average of the squared differences from the mean. To calculate variance, it is useful to break the calculation down into steps:

Step 1: Calculate the mean (previously discussed).
Step 2: Subtract the mean from each of the values and square the result.
Step 3: Divide by n – 1

$$s^2 = \frac{\Sigma\left(x_i - \bar{x}\right)^2}{n-1}$$

where s^2 is the sample variance

x_i is the individual value

$\bar{x}$ is the sample mean

n is the sample size

Let's calculate the variance of the following data set: 2, 7, 3, 12, 9.

The first step is to calculate the mean. The sum is 33 and there are 5 data points. Therefore, the mean is 33 ÷ 5 = 6.6.

Then you take each value in the data set, subtract the mean and square the difference. For instance, for the first value:

$(2 - 6.6)^2 = 21.16$

The squared differences for all values are added:

21.16 + 0.16 + 12.96 + 29.16 + 5.76 = 69.20

The sum is then divided by the number of data points - 1:

69.20 ÷ 4 = 17.3

The variance is 17.3

Statisticians use variance to see how individual numbers relate to each other within a data set, rather than using broader mathematical techniques such as arranging numbers into quartiles. The advantage of variance is that it treats all deviations from the mean as the same regardless of their direction.

One drawback to variance, though, is that it gives added weight to outliers.

Exercise 21

Calculate the variance for these results:

Pebble number	**Diameter (mm)**
1	7
2	13
3	5
4	8
5	15
6	3
7	9

(a) Calculate the mean

(b) For each pebble, take away the mean and then square the result.

Pebble number	**Diameter (mm)**	$\mathbf{xi - \bar{x}}$	$\mathbf{(x_i - \bar{x})^2}$
1	7		
2	13		
3	5		

4	8		
5	15		
6	3		
7	9		

(c) Work out the sum of these squared values.

(d) Use your answers in the following equation to calculate the variance.

$$s^2 = \frac{\Sigma\left(x_i - \bar{x}\right)^2}{n-1}$$

where s^2 is the sample variance

x_i is the individual value

$\bar{x}$ is the sample mean

n is the sample size

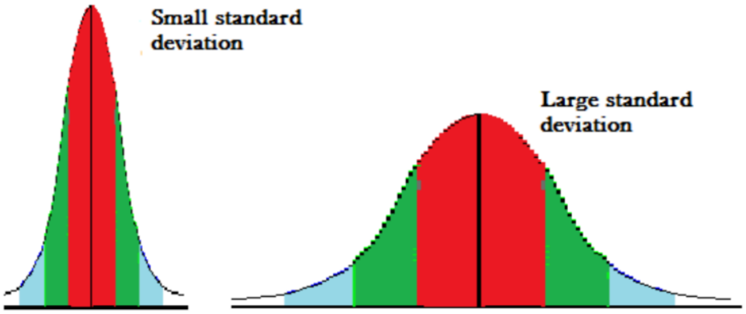
Standard deviation: The standard deviation is the positive square root of the variance.

$$s = \sqrt{\frac{\Sigma\left(x - \bar{x}\right)^2}{n-1}}$$

Exercise 22

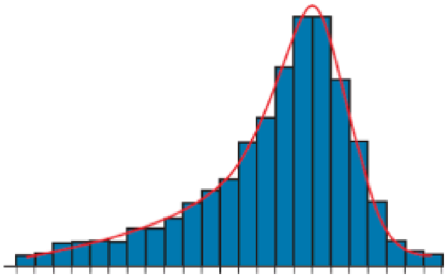
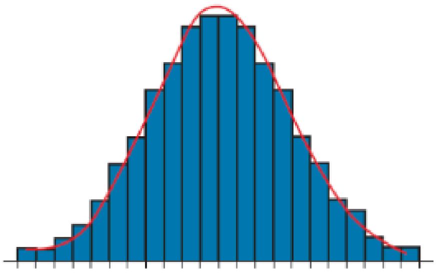
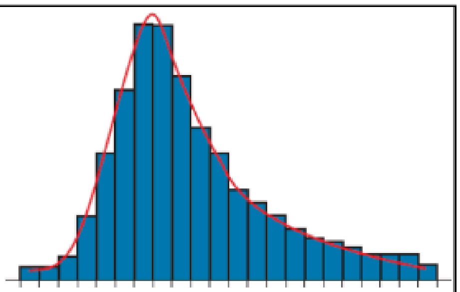
Using your answer for exercise 21, calculate the standard deviation.

Standard deviation gives us a measure of how clustered the data are around the mean. A smaller value of standard deviation indicates that the data is tightly clustered around the mean and vice-versa (see below).



Observing the shape of these two curves shows that they are symmetrical about the centre. This type of curve is called a bell curve and shows that the data is normally distributed about the centre – the mean. In such a normal distribution the mean, mode and median are equal and exactly half the values are to the left of the centre and half the values are to the right. In the standard normal model about 68% of the data falls within one standard deviation of the mean, about 95% of the data falls within two standard deviations of the mean and just over 99% of the data falls within three standard deviations of the mean.

If the graph is not the perfect symmetrical bell shaped curve, you should be able to describe the skewness.

		
A negative skew. The tail of the data extends to the left (in a negative direction). In such a case the median is usually larger than the mean.	No skew. Data is perfectly symmetrical about the mean.	A positive skew. The tail of the data extends to the right (in a positive direction). In such a case the mean is usually larger than the median.

$$\text{skew} = \frac{3\,(\text{mean} - \text{median})}{\text{standard deviation}}$$

Exercise 23

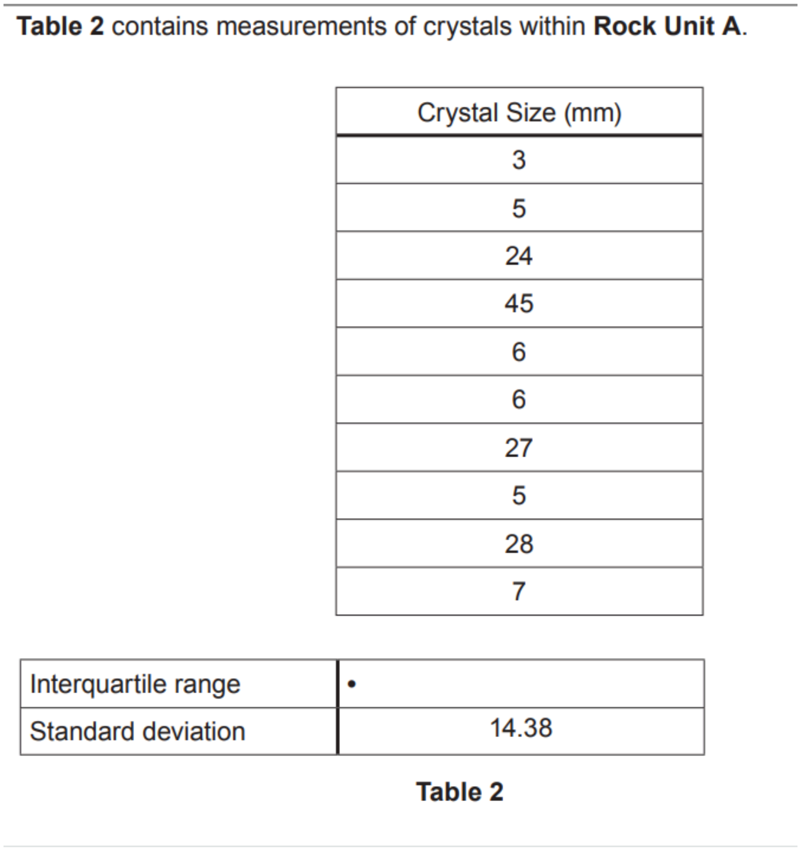
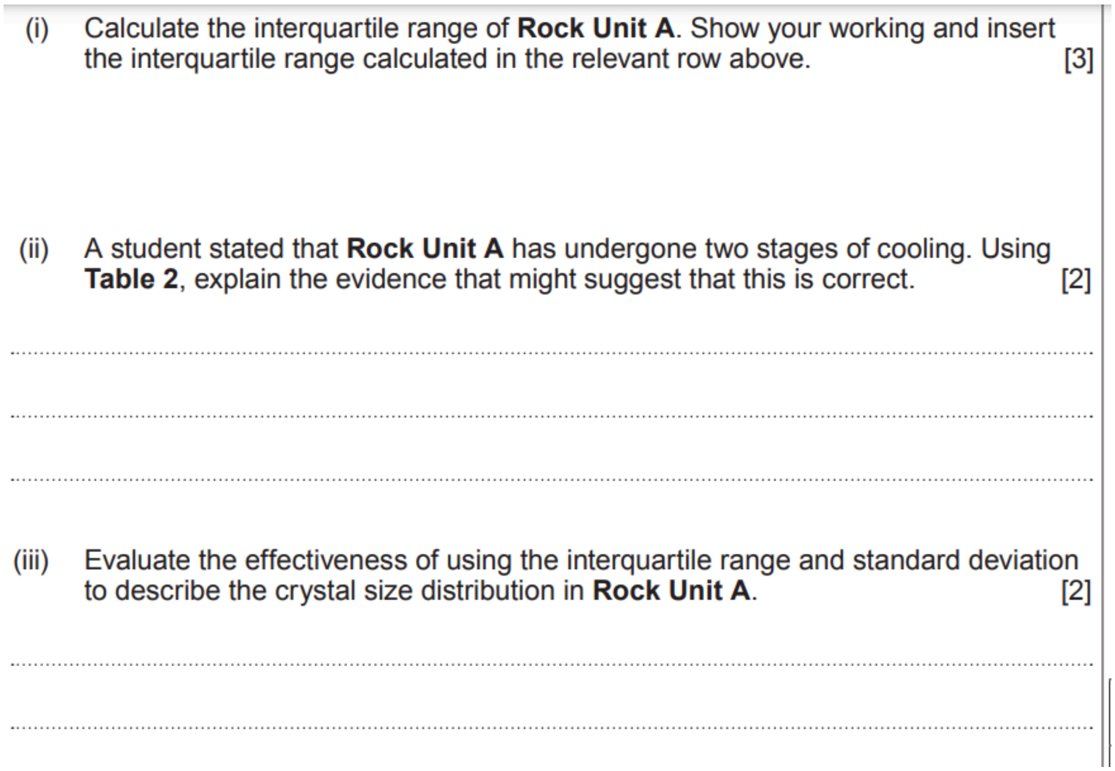


Table 2 contains measurements of crystals within **Rock Unit A**.

Crystal Size (mm)
3
5
24
45
6
6
27
5
28
7

Interquartile range	•
Standard deviation	14.38

Table 2



(i) Calculate the interquartile range of **Rock Unit A**. Show your working and insert the interquartile range calculated in the relevant row above. [3]

(ii) A student stated that **Rock Unit A** has undergone two stages of cooling. Using **Table 2**, explain the evidence that might suggest that this is correct. [2]

....

....

....

(iii) Evaluate the effectiveness of using the interquartile range and standard deviation to describe the crystal size distribution in **Rock Unit A**. [2]

....

....

(i)	Rank the crystal sizes in order (1) Find lower and upper quartile rank (3rd and 8th= 5 and 27) (1) Interquartile range = 22 (1)
(ii)	4 larger crystals and group of smaller crystals/porphyritic (1) Larger crystals take longer to form/smaller crystals form more quickly (1)
(iii)	**Any two x (1) from:** • reference to the ineffectiveness of both • reference to interquartile range being more effective than standard deviation • both demonstrate the distribution of the grain sizes • both hide the bimodal characteristics of Rock Unit A • interquartile range only uses the central 50% of the data so removes potential anomalies • standard deviation shows how much data is clustered around a mean value • standard deviation assumes a normal distribution pattern but rock unit A does not- reference to large phenocryst

Data analysis - statistical tests

1. Spearman's Rank

This is a technique which can be used to test the strength and direction (negative or positive) of a linear relationship between samples of two variables. The result will always be between 1 and minus 1. This is a test that needs at least ten pairs of data.

For example a geologist wants to test the idea that the length of a fault is related to the size of the earthquake. First we have to create a null hypothesis (H_o) to test: *There is no relationship between the length of the fault and the magnitude of the earthquake.*

Fault rupture length (m)	Rank	Earthquake magnitude	Rank	Difference (d)	Difference squared (d^2)
177		7.52			
64		7.24			
245		7.88			
36		7.45			
58		7.41			
31		6.83			
74		7.13			
47		7.51			
89		7.30			
20		5.82			
235		7.46			
55		7.32			
					$\sum d^2 =$

Exercise 24

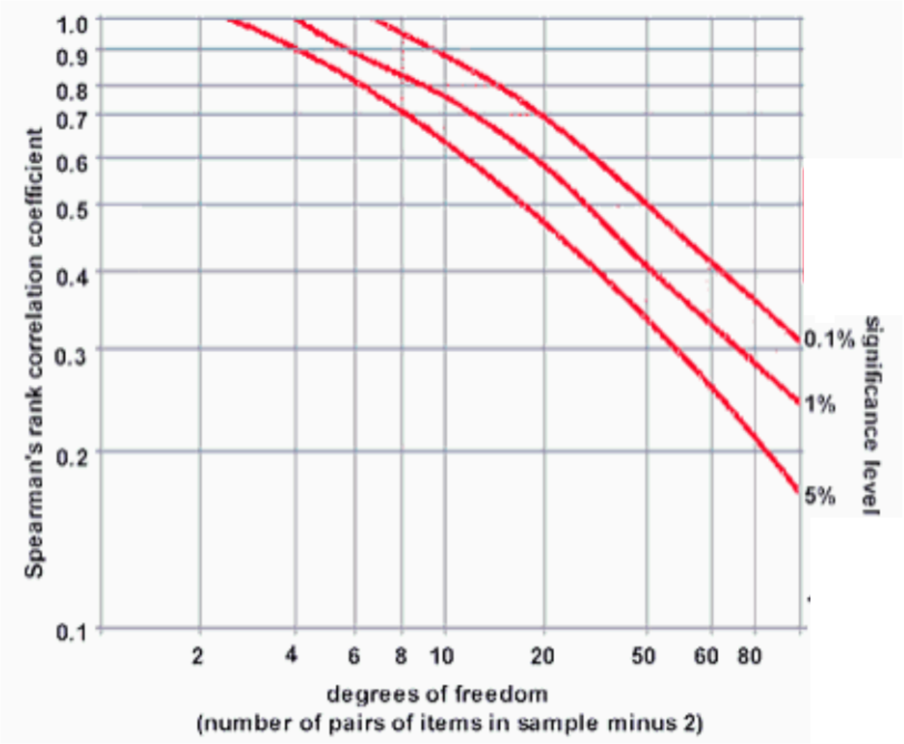
Step 1 - work out the ranks for both the fault lengths and the magnitude.
Step 2 - work out the differences in the ranks.
Step 3 - square the differences.
Step 4 - calculate the total of the differences.
Step 5 - use the result and the equation below to calculate the Spearman’s rank coefficient r_s.

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

where n is the number of pairs of data.

The result you get tells you about the relationship between the two variables. A positive number tells you there is a positive correlation, a negative value indicates a negative correlation.   The closer the value is to 1, the stronger the relationship.

You now have to determine the significance of the correlation. You do this using a confidence interval table.

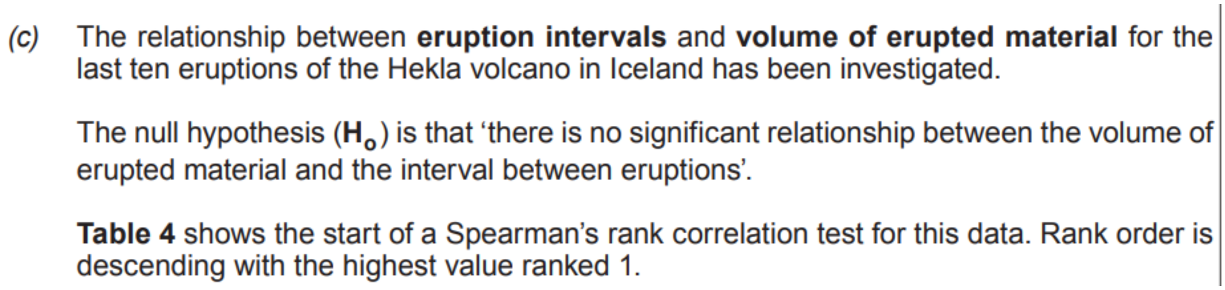


Exercise 25

Work out the number of degrees of freedom (samples size -2).
Find the Spearman’s rank coefficient you calculated in exercise 24 - is it higher than the 5% line?

If the value is BELOW the line, you cannot reject the null hypothesis. In this case, the value is 0.59 which is below the 5% confidence level, so we cannot reject the null hypothesis.

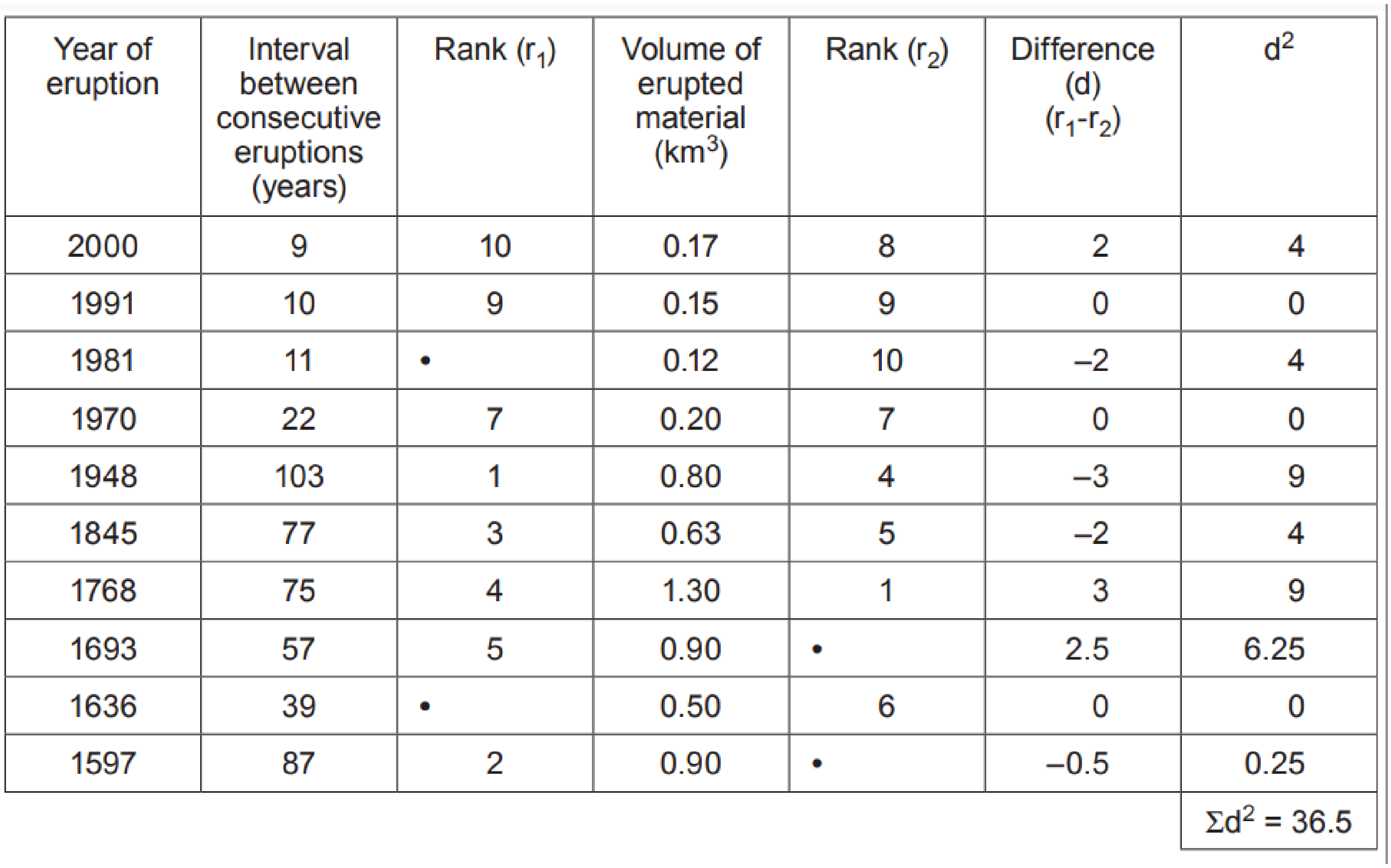
Exercise 26



(c) The relationship between **eruption intervals** and **volume of erupted material** for the last ten eruptions of the Hekla volcano in Iceland has been investigated.

The null hypothesis (**H_o**) is that 'there is no significant relationship between the volume of erupted material and the interval between eruptions'.

Table 4 shows the start of a Spearman's rank correlation test for this data. Rank order is descending with the highest value ranked 1.



Year of eruption	Interval between consecutive eruptions (years)	Rank (r_1)	Volume of erupted material (km^3)	Rank (r_2)	Difference (d) (r_1-r_2)	d^2
2000	9	10	0.17	8	2	4
1991	10	9	0.15	9	0	0
1981	11	•	0.12	10	–2	4
1970	22	7	0.20	7	0	0
1948	103	1	0.80	4	–3	9
1845	77	3	0.63	5	–2	4
1768	75	4	1.30	1	3	9
1693	57	5	0.90	•	2.5	6.25
1636	39	•	0.50	6	0	0
1597	87	2	0.90	•	–0.5	0.25
						$\Sigma d^2 = 36.5$

Correlation coefficient formula: $r_s = 1 - \frac{6\Sigma d^2}{n^3 - n}$

where r_s is the correlation coefficient and n is the number of paired data.

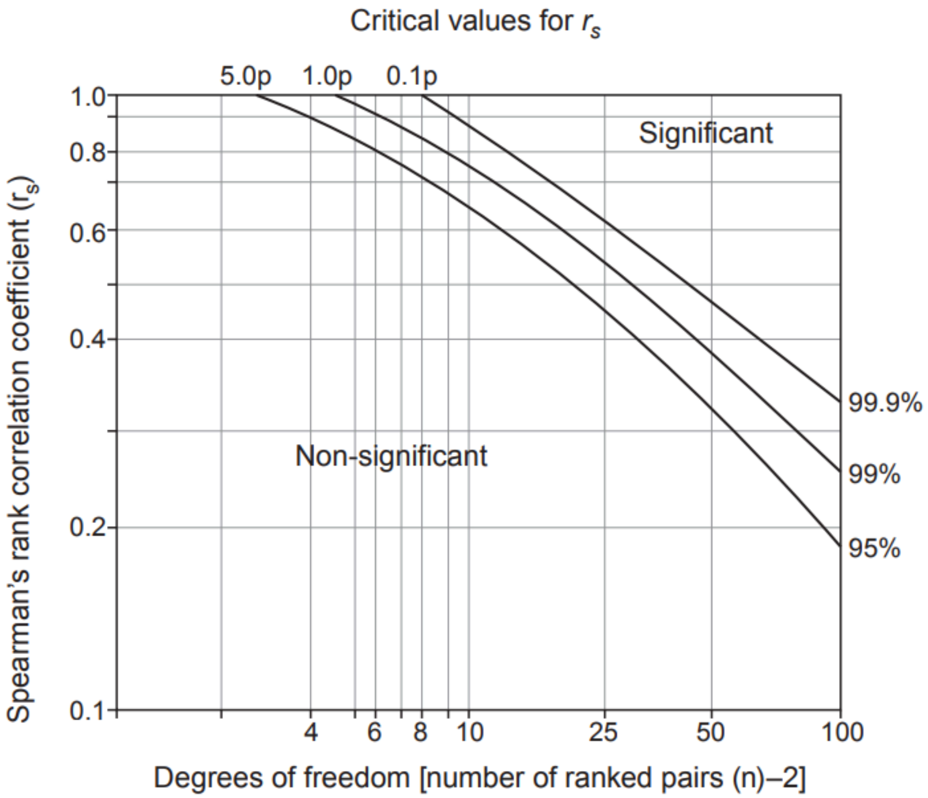
Table 4

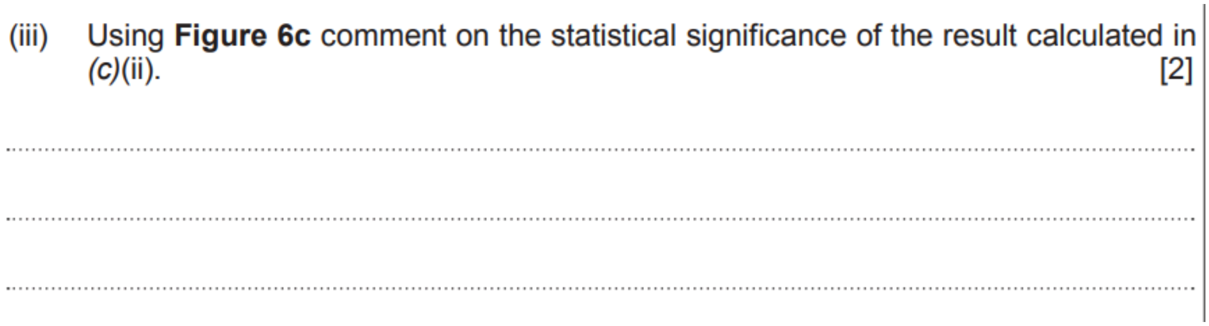
(i) Complete **Table 4** to show the missing values of rank r_1 and r_2. [2]

(ii) Using the formula, calculate the Spearman's rank correlation coefficient (r_s). Show your working. [2]

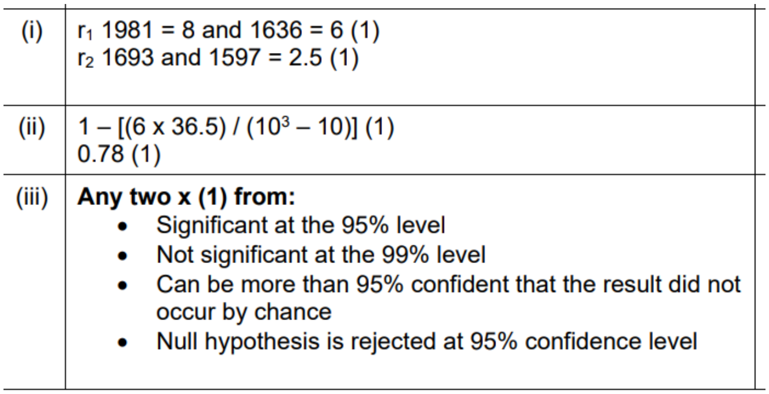
r_s = ...

Figure 6c is a Spearman's rank correlation significance graph.





(iii) Using **Figure 6c** comment on the statistical significance of the result calculated in (*c*)(ii). [2]



(i)	r_1 1981 = 8 and 1636 = 6 (1) r_2 1693 and 1597 = 2.5 (1)
(ii)	1 – [(6 x 36.5) / (10^3 – 10)] (1) 0.78 (1)
(iii)	**Any two x (1) from:** • Significant at the 95% level • Not significant at the 99% level • Can be more than 95% confident that the result did not occur by chance • Null hypothesis is rejected at 95% confidence level

2. Chi Squared Test

The chi-squared test is the most efficient test available to analyse numerical (quantitative) data. This test can only be used on data which has the following characteristics:
i) The data must be in the form of frequencies counted in a number of groups (% cannot be used).
ii) The total number of observations must be > 20.
iii) The observations must be independent (i.e. one observation must not influence another).
iv) The expected frequency in any one category must not normally be < 5. It may be necessary therefore to combine groups.

A geologist was investigating the composition of the clasts in 2 beds of conglomerate. Their null hypothesis (H_o) was *'there is no significant difference in the composition of clasts sampled in the two deposits'*.

Clast Lithology type	Observed Bed 1	Observed Bed 2
Ultramafic	18	32
Gabbro	12	10
Basalt	20	16
Chert	8	12
Chalk	42	30

Exercise 27

Calculate the row and column totals.

Clast Lithology type	Observed Bed 1	Observed Bed 2	Row Total
Ultramafic	18	32	
Gabbro	12	10	
Basalt	20	16	
Chert	8	12	
Chalk	42	30	
Column Total			

The next step is to work out the expected values - these are the values you would get if there is no relationship. You do this using this equation:

$$\text{Expected value} = \frac{\text{column total} \times \text{row total}}{\text{Overall total}}$$

Exercise 28

Complete the table below:

Clast Lithology type	Expected Bed 1	Expected Bed 2	Row Total
Ultramafic	25	25	50
Gabbro			
Basalt			
Chert			
Chalk			
Column Total			

The value of chi squared is calculated using this formula:

$$X^2 = \frac{\Sigma(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

$$X^2 = \frac{(18-25)^2}{25} + \frac{(12-11)^2}{11} + \frac{(20-18)^2}{18} + \frac{(8-10)^2}{10} + \frac{(42-36)^2}{36} + \frac{(32-25)^2}{25} + \frac{(10-11)^2}{11} +$$

$$\frac{(16-18)^2}{18} + \frac{(12-10)^2}{10} + \frac{(30-36)^2}{36}$$

Exercise 29

Calculate the value for chi squared.

To reject the null hypothesis, the value for chi squared must be above the value for 95%.
The number of degrees of freedom is (number of rows - 1) x (number of columns -1)

Degrees of Freedom	Probability										
	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
	Non-significant								Significant		

Exercise 30

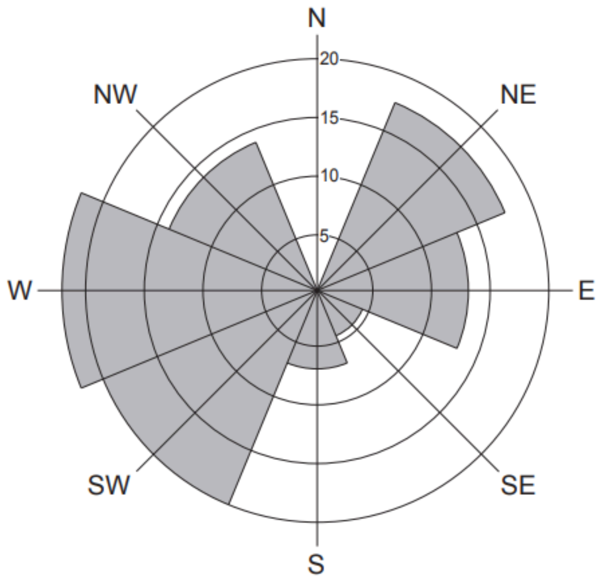
Calculate the number of degrees of freedom.

The value for chi squared was 7.3. This is less than the critical value for 95% which is 9.49, so we CANNOT reject the null hypothesis.

Exercise 31

Figure 6 is a rose diagram showing the distribution of landslides in the Chesterfield region (which includes the area of the **geological map**) based on the direction the landslide slope faces (slope aspect).

A student undertook a field investigation into the orientation of these landslides to test the hypothesis that slope aspect was important in landslide development.

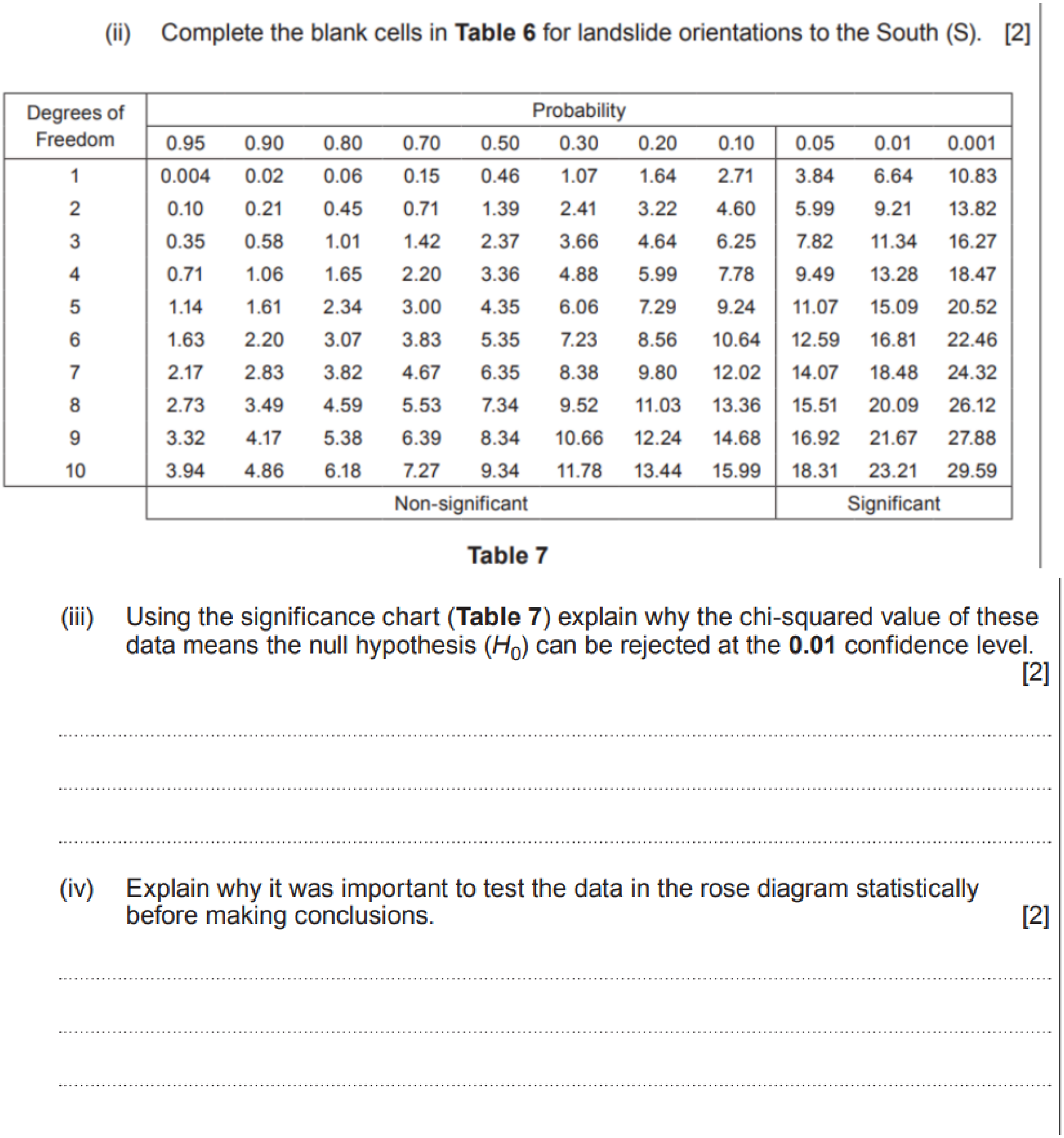


Numbers refer to observed landslide frequency

Table 6 is a partly completed chi-squared test used to test the null hypothesis (H_0) that "there is no significant orientation of the landslides".

Landslide Orientation	Observed frequency (O)	Expected frequency (E)	(O – E)	$(O – E)^2$	$\frac{(O – E)^2}{E}$
N	14	14	0	0	0
NE	18	14	4	16	1.14
E	13	14	–1	1	0.07
SE	4	14	–10	100	7.15
S	7	14	•	•	•
SW	20	14	6	36	2.57
W	22	14	8	64	4.57
NW	14	14	0	0	0
Total	112	112	**chi-squared value**		**19.00**

Table 6



(ii) Complete the blank cells in **Table 6** for landslide orientations to the South (S). [2]

Degrees of Freedom	Probability										
	0.95	0.90	0.80	0.70	0.50	0.30	0.20	0.10	0.05	0.01	0.001
1	0.004	0.02	0.06	0.15	0.46	1.07	1.64	2.71	3.84	6.64	10.83
2	0.10	0.21	0.45	0.71	1.39	2.41	3.22	4.60	5.99	9.21	13.82
3	0.35	0.58	1.01	1.42	2.37	3.66	4.64	6.25	7.82	11.34	16.27
4	0.71	1.06	1.65	2.20	3.36	4.88	5.99	7.78	9.49	13.28	18.47
5	1.14	1.61	2.34	3.00	4.35	6.06	7.29	9.24	11.07	15.09	20.52
6	1.63	2.20	3.07	3.83	5.35	7.23	8.56	10.64	12.59	16.81	22.46
7	2.17	2.83	3.82	4.67	6.35	8.38	9.80	12.02	14.07	18.48	24.32
8	2.73	3.49	4.59	5.53	7.34	9.52	11.03	13.36	15.51	20.09	26.12
9	3.32	4.17	5.38	6.39	8.34	10.66	12.24	14.68	16.92	21.67	27.88
10	3.94	4.86	6.18	7.27	9.34	11.78	13.44	15.99	18.31	23.21	29.59
	Non-significant								Significant		

Table 7

(iii) Using the significance chart (**Table 7**) explain why the chi-squared value of these data means the null hypothesis (H_0) can be rejected at the **0.01** confidence level. [2]

....

....

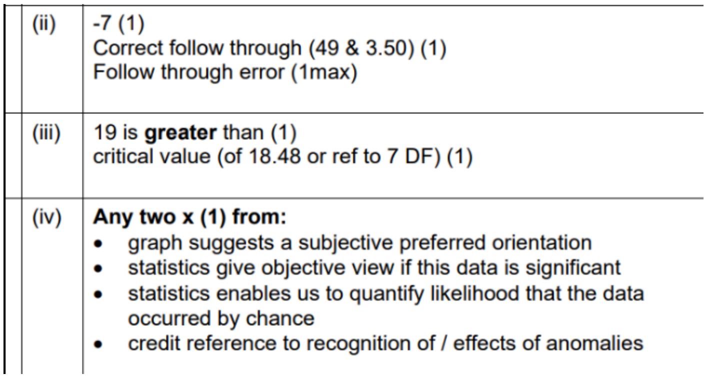
....

(iv) Explain why it was important to test the data in the rose diagram statistically before making conclusions. [2]

....

....

....



(ii)	-7 (1) Correct follow through (49 & 3.50) (1) Follow through error (1max)
(iii)	19 is **greater** than (1) critical value (of 18.48 or ref to 7 DF) (1)
(iv)	**Any two x (1) from:** • graph suggests a subjective preferred orientation • statistics give objective view if this data is significant • statistics enables us to quantify likelihood that the data occurred by chance • credit reference to recognition of / effects of anomalies

3. Mann-Whitney U-test

The Mann-Whitney U-test is another test which is used to analyse the difference between two samples of independent data: more specifically the medians of the two datasets. The data should be quantitative and there should be >5 but ≤ 20 pieces of data in each sample. Additionally, there need not be the same number of observations in each sample. Like other statistical tests, the starting point is a null hypothesis of the form 'H_o: there is no significant difference between the two samples'.

A geologist is investigating the difference in roundess in two sets of pebbles, one set of pebbles are made of hard, igneous basalt and the other set are made of soft chalk. These are the results (the higher the number, the more rounded the pebble).

Cailleux's roundness index	
chalk (n_a)	basalt (n_b)
780	650
640	620
690	570
710	700
550	610
670	490
720	520
660	600
510	590
610	680

Exercise 32

Write a null hypothesis for this investigation.

Exercise 33

The total sample set is then ranked.

Cailleux's roundness index		rank	
chalk (n_a)	basalt (n_b)	R_a	R_b
780	650		
640	620		
690	570		
710	700		
550	610		
670	490		
720	520		
660	600		
510	590		
610	680		
		$\sum R_a =$	$\sum R_b =$

Exercise 34

Calculate the totals for the two rank columns.

For each set of data, calculate the value of U.

$$U_a = n_a n_b + \frac{n_a(n_a+1)}{2} - \Sigma R_a$$

and

$$U_b = n_a n_b + \frac{n_b(n_b+1)}{2} - \Sigma R_b$$

where

U_a and U_b are the Mann Whitney scores for samples a and b respectively

n_a and n_b are the number in samples a and b respectively

and $\sum R_a$ and $\sum R_b$ are the sums of the ranks for samples a and b respectively

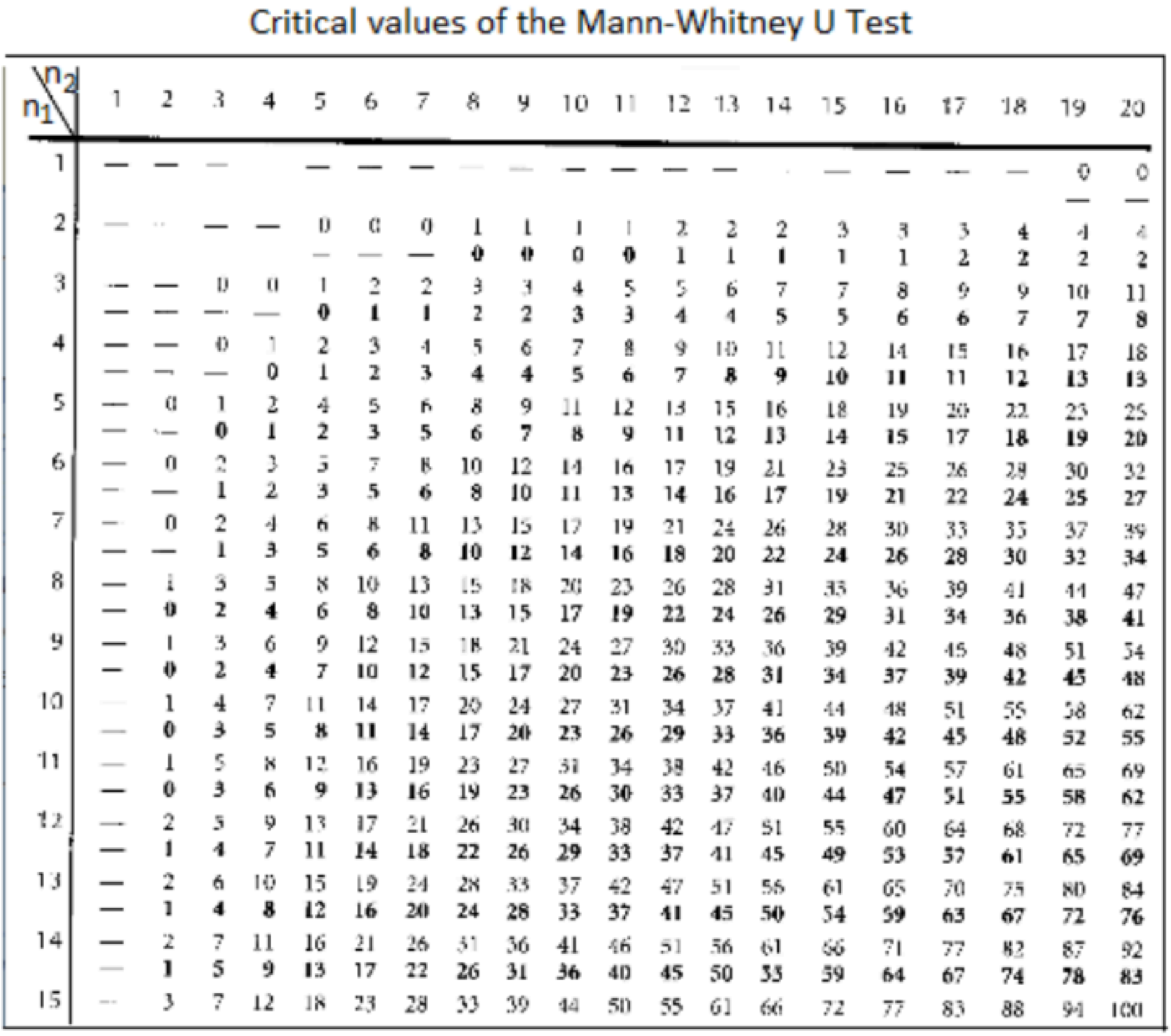
Exercise 35

Complete the calculation below (your answers should add up to 100):

$$U_a = (10 \times 10) + \frac{10(10+1)}{2} - 84.5 =$$

$$U_b = (10 \times 10) + \frac{10(10+1)}{2} - 125.5 =$$

To complete the Mann-Whitney test, you need to use the lower U value. In this case it is U_b. So take this value of 29.5 and compare it to the critical values table - use the bold values in the table. This table is for 95% confidence.



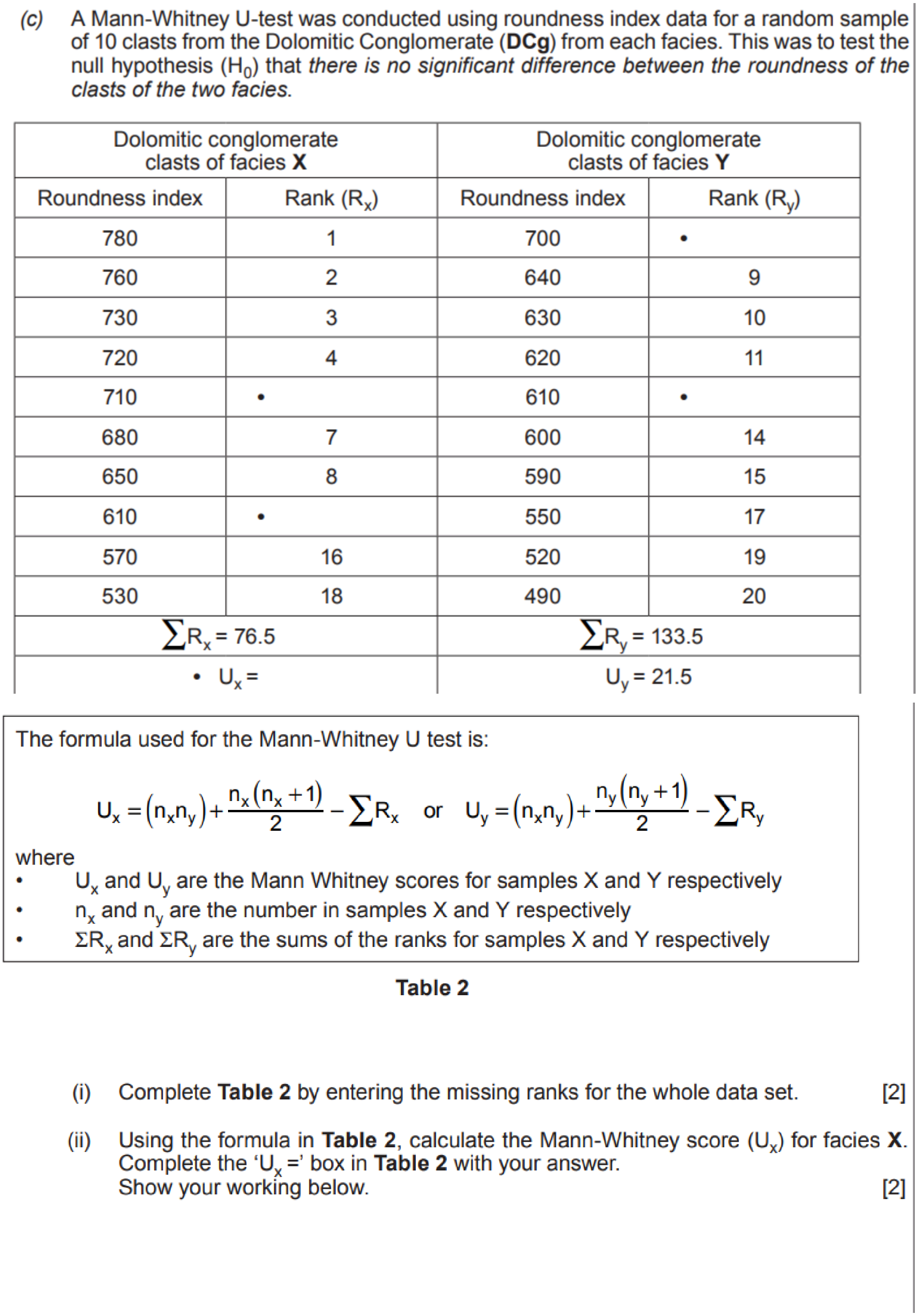
Critical values of the Mann-Whitney U Test

n_1 \ n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	—	—	—		—	—	—	—	—	—	—	—	—		—	—	—	—	0 **—**	0 **—**
2	—	—	—	—	0 **—**	0 **—**	0 **—**	1 **0**	1 **0**	1 **0**	1 **0**	2 **1**	2 **1**	2 **1**	3 **1**	3 **1**	3 **2**	4 **2**	4 **2**	4 **2**
3	— **—**	— **—**	0 **—**	0 **—**	1 **0**	2 **1**	2 **1**	3 **2**	3 **2**	4 **3**	5 **3**	5 **4**	6 **4**	7 **5**	7 **5**	8 **6**	9 **6**	9 **7**	10 **7**	11 **8**
4	— **—**	— **—**	0 **—**	1 **0**	2 **1**	3 **2**	4 **3**	5 **4**	6 **4**	7 **5**	8 **6**	9 **7**	10 **8**	11 **9**	12 **10**	14 **11**	15 **11**	16 **12**	17 **13**	18 **13**
5	— **—**	0 **—**	1 **0**	2 **1**	4 **2**	5 **3**	6 **5**	8 **6**	9 **7**	11 **8**	12 **9**	13 **11**	15 **12**	16 **13**	18 **14**	19 **15**	20 **17**	22 **18**	23 **19**	25 **20**
6	— **—**	0 **—**	2 **1**	3 **2**	5 **3**	7 **5**	8 **6**	10 **8**	12 **10**	14 **11**	16 **13**	17 **14**	19 **16**	21 **17**	23 **19**	25 **21**	26 **22**	28 **24**	30 **25**	32 **27**
7	— **—**	0 **—**	2 **1**	4 **3**	6 **5**	8 **6**	11 **8**	13 **10**	15 **12**	17 **14**	19 **16**	21 **18**	24 **20**	26 **22**	28 **24**	30 **26**	33 **28**	35 **30**	37 **32**	39 **34**
8	— **—**	1 **0**	3 **2**	5 **4**	8 **6**	10 **8**	13 **10**	15 **13**	18 **15**	20 **17**	23 **19**	26 **22**	28 **24**	31 **26**	33 **29**	36 **31**	39 **34**	41 **36**	44 **38**	47 **41**
9	— **—**	1 **0**	3 **2**	6 **4**	9 **7**	12 **10**	15 **12**	18 **15**	21 **17**	24 **20**	27 **23**	30 **26**	33 **28**	36 **31**	39 **34**	42 **37**	45 **39**	48 **42**	51 **45**	54 **48**
10	— **—**	1 **0**	4 **3**	7 **5**	11 **8**	14 **11**	17 **14**	20 **17**	24 **20**	27 **23**	31 **26**	34 **29**	37 **33**	41 **36**	44 **39**	48 **42**	51 **45**	55 **48**	58 **52**	62 **55**
11	— **—**	1 **0**	5 **3**	8 **6**	12 **9**	16 **13**	19 **16**	23 **19**	27 **23**	31 **26**	34 **30**	38 **33**	42 **37**	46 **40**	50 **44**	54 **47**	57 **51**	61 **55**	65 **58**	69 **62**
12	— **—**	2 **1**	5 **4**	9 **7**	13 **11**	17 **14**	21 **18**	26 **22**	30 **26**	34 **29**	38 **33**	42 **37**	47 **41**	51 **45**	55 **49**	60 **53**	64 **57**	68 **61**	72 **65**	77 **69**
13	— **—**	2 **1**	6 **4**	10 **8**	15 **12**	19 **16**	24 **20**	28 **24**	33 **28**	37 **33**	42 **37**	47 **41**	51 **45**	56 **50**	61 **54**	65 **59**	70 **63**	75 **67**	80 **72**	84 **76**
14	— **—**	2 **1**	7 **5**	11 **9**	16 **13**	21 **17**	26 **22**	31 **26**	36 **31**	41 **36**	46 **40**	51 **45**	56 **50**	61 **55**	66 **59**	71 **64**	77 **67**	82 **74**	87 **78**	92 **83**
15	—	3	7	12	18	23	28	33	39	44	50	55	61	66	72	77	83	88	94	100

Exercise 36

Is the calculated value, 29.5 higher than the critical value?
Should the null hypothesis be rejected?

Exercise 37



(c) A Mann-Whitney U-test was conducted using roundness index data for a random sample of 10 clasts from the Dolomitic Conglomerate (**DCg**) from each facies. This was to test the null hypothesis (H_0) that *there is no significant difference between the roundness of the clasts of the two facies.*

Dolomitic conglomerate clasts of facies **X**		Dolomitic conglomerate clasts of facies **Y**	
Roundness index	Rank (R_x)	Roundness index	Rank (R_y)
780	1	700	•
760	2	640	9
730	3	630	10
720	4	620	11
710	•	610	•
680	7	600	14
650	8	590	15
610	•	550	17
570	16	520	19
530	18	490	20
$\sum R_x = 76.5$		$\sum R_y = 133.5$	
• $U_x =$		$U_y = 21.5$	

The formula used for the Mann-Whitney U test is:

$$U_x = (n_x n_y) + \frac{n_x(n_x+1)}{2} - \sum R_x \quad \text{or} \quad U_y = (n_x n_y) + \frac{n_y(n_y+1)}{2} - \sum R_y$$

where

- U_x and U_y are the Mann Whitney scores for samples X and Y respectively
- n_x and n_y are the number in samples X and Y respectively
- ΣR_x and ΣR_y are the sums of the ranks for samples X and Y respectively

Table 2

(i) Complete **Table 2** by entering the missing ranks for the whole data set. [2]

(ii) Using the formula in **Table 2**, calculate the Mann-Whitney score (U_x) for facies **X**. Complete the 'U_x =' box in **Table 2** with your answer.
Show your working below. [2]

(iii) At a significance level of 0.05 (95% confidence level) the critical Mann-Whitney value for U is 23. Using the Mann Whitney value for U_y (21.5), assess the significance of this value in terms of the null hypothesis (H_0). [2]

....

....

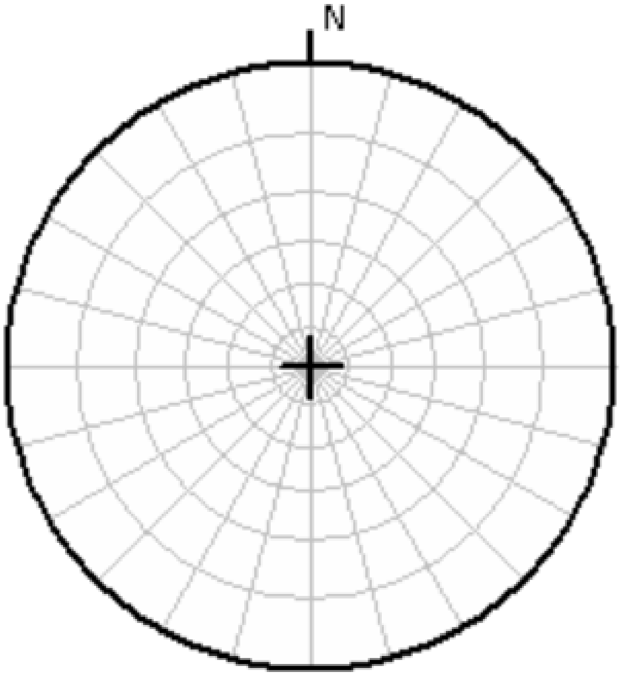
....

(c)	(i)	Rx = 5 and 12.5 Ry = 6 and 12.5 (4 correct = 2 marks, 2 correct = 1 mark)
	(ii)	$Ux = (10x10) + \frac{10(10+1)}{2} = \underset{(1)}{155} - \underset{(1)}{76.5} = 78.5$
	(iii)	**Any two x (1) from:** • Uy = 21.5 is less than the critical value • the null hypothesis is rejected at 95% significance • there is a 95% probability that the roundness of the two facies is different

Stereonets

An appropriate way to present data that involve both direction (azimuth) and dip plotted on the same diagram, e.g. variation in dip angle and direction of a bedding surface, fractures or aligned fragments in a superficial deposit, is to use a polar equal ‘stereonet’. Similar to a rose diagram, this involves a simplified ‘stereonet’ showing polar plots only and not projections or great circles.

Data is collected on both dip angle and direction of dip in the field or from a secondary source (e.g. geological map). The dip angle is plotted as a point from the centre of the graph in the direction of dip (azimuth) according to the radial scale (zero – 90 degrees). It gives a visual display of trends and amounts but has a drawback for folded strata. If this method is used to plot dip directions on either side of a fold or series of folds, the type of fold (antiform or synform) will not be obvious. Although the dip directions of the limbs are shown, there is no indication as to whether the limbs are dipping towards or away from each other unless further annotation is given. An approximation of the orientation of the fold axis can be determined and plotted by eye or calculated from field measurements.



Exercise

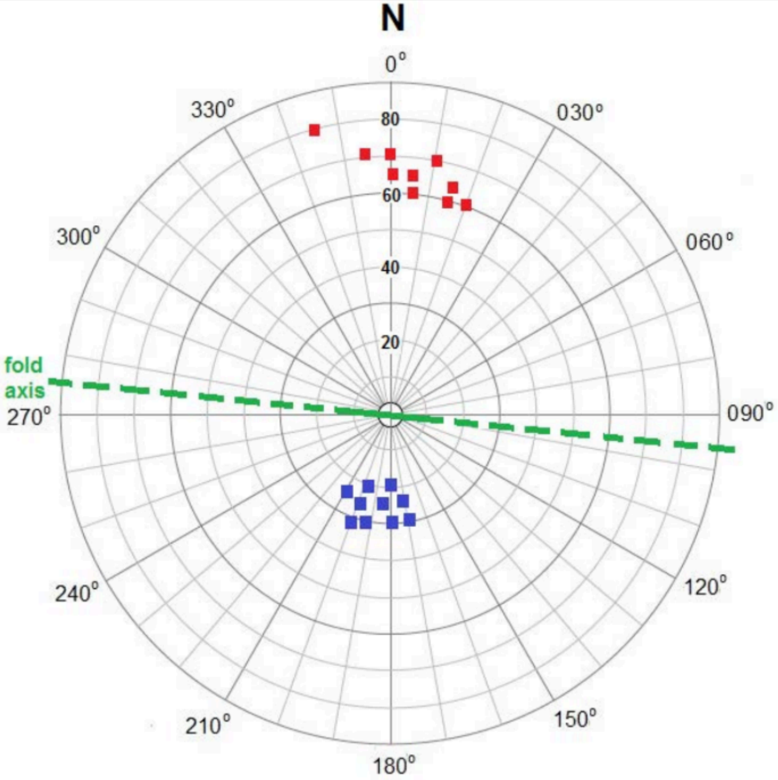
Label the radial lines on the stereonet with azimuths / bearings.
Label the concentric lines with angles up to 90° .

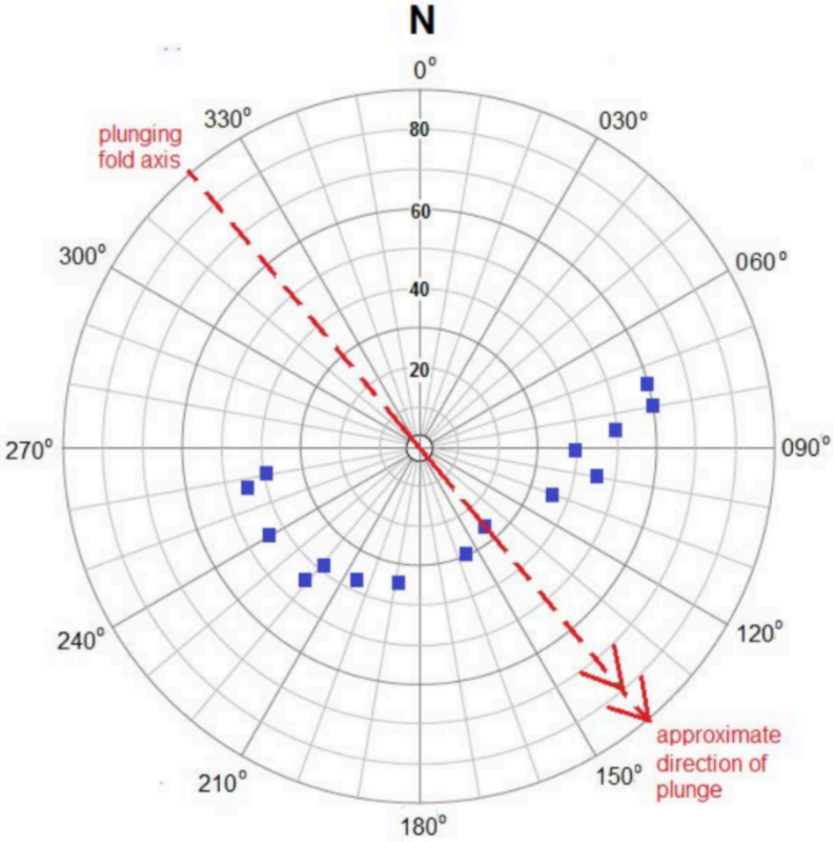
Exercise

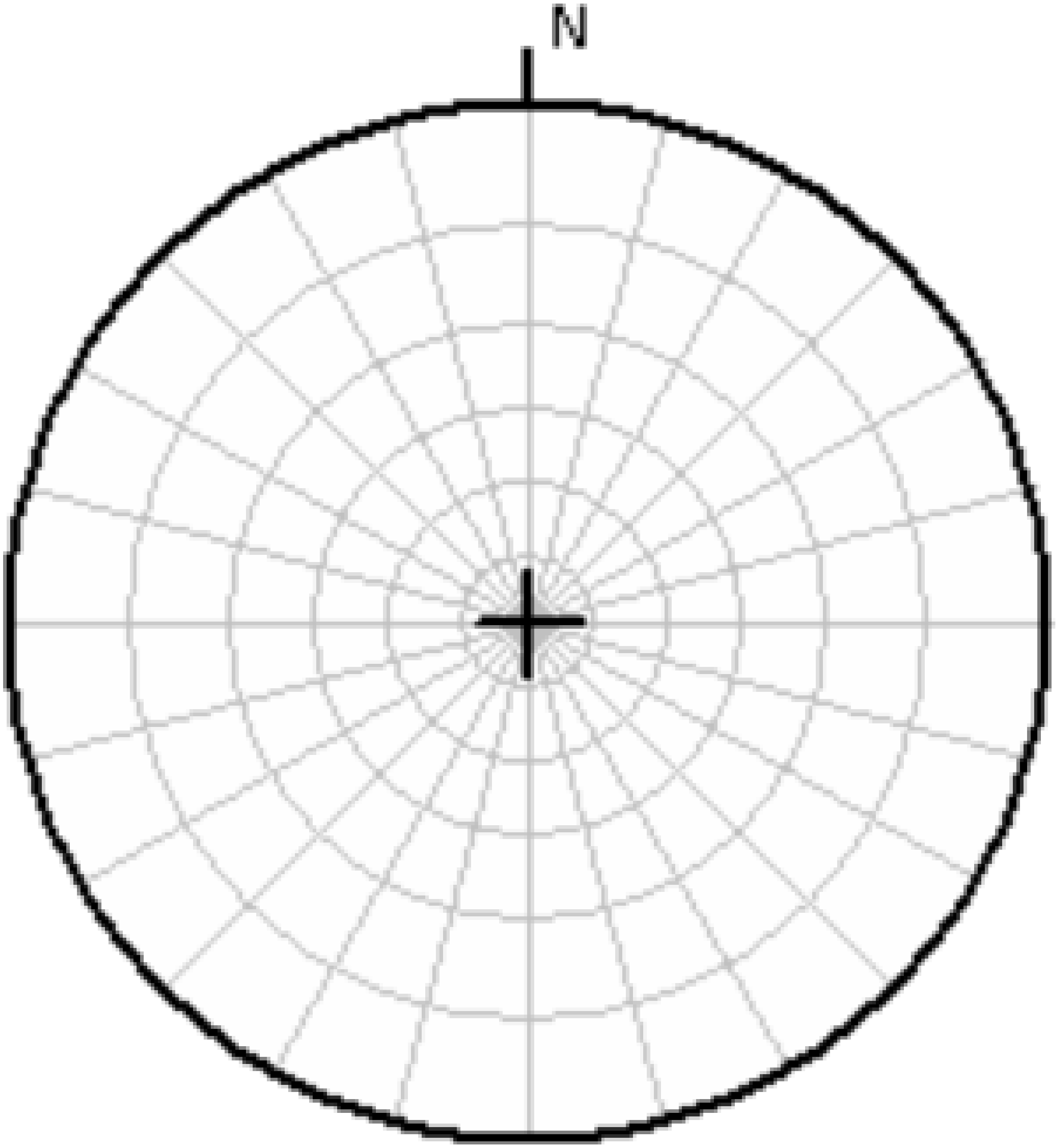
Plot the following dip measurements onto the stereonet as points:
A - Dip 45°, dip direction 030°
B - Dip 60°, dip direction 015°
C- Dip 50°, dip direction 130°

Exercise

What do these stereonets represent?






N

Numeracy skills - checklist

Skill area	**Booklet**	**Website**	**Tick**
Converting numbers to and from standard form	y		
Standard form calculations	y		
Significant figures	y		
Percentage uncertainty	y		
Sampling methods	y	y	
Frequency tables	y		
Histograms	y	y	
Cumulative frequency curves	y	y	
Mean, median, mode, range	y		
Interquartile range	y		
Variance	y		
Standard deviation	y		
Spearman’s rank analysis	y	y	
Chi squared test	y	y	
Mann-Whitney U test	y	y	
Rose diagrams		y	
Stereonets			
Triangular diagrams		y	
Straight lines of best fit y=mx+c			