Year 9 2025 Half Yearly Maths (Data Analysis A + Area, Surface Area A and Volume A)

# Data Analysis A

## Centre and Spread

We can use the summary statistics of **mean**, **median** or **mode** to describe the **centre** of a data set and the **range** to describe the **spread**.

**Mean**

* Often called the average of a data set, this is the sum of the data values divided by the number of data values.

**Median**

* The middle value of a data set ranked in order.

**Mode**

* The data value that occurs most frequently. We can describe data sets using their number of modes: no modes (**uniform**), one mode (**unimodal**), two modes (**bimodal**) or more than two modes (**multimodal**).

**Range**

* The difference between the smallest and largest scores in the set.

### Outliers

Removing **outliers** will have the following effects on the summary statistics:

| Remove a low outlier | Remove a high outlier |
| --- | --- |
| The range will decrease. | The range will decrease |
| The median might increase. | The median might decrease. |
| The mean will increase. | The mean will decrease |
| The mode will not change. | The mode will not change. |

## Standard Deviation

### *MAD*

*The MAD is the average distance between each data point and the mean. Recall distance is always positive, so we find the distances by taking the absolute value of the difference between the mean and each data point. The MAD is the average of these absolute differences.*

*Mean absolute deviation*

*Sample size*

*A particular data value*

*Sample mean*

*A small MAD indicates that most scores are close to the mean, while a large MAD indicates that the scores are more spread out.*

### Standard Deviation

For both population and sample standard deviation, use your calculator.

#### Population Standard Deviation

The **standard deviation** is another way of showing how spread out numbers in a data set are. Unlike the MAD, standard deviation considers the square of the distances from the mean, making it more sensitive to extreme values.

Similar to the MAD, a small standard deviation indicates that most scores are close to the mean, while a large standard deviation indicates that the scores are more spread out.

Population standard deviation

Population size

A particular data value

Population mean

The process for calculating standard deviation is time consuming, so we will be using our calculator to find the standard deviation. In statistics mode on a calculator, the symbol ​ may be used.

#### Sample Standard Deviation

There is a second type of standard deviation for when you are working with a sample and not a population. This is the sample standard deviation, with the symbol .

Sample standard deviation

Sample size

A particular data value

Sample mean

As you can see from the formula, they are calculated almost the same way, but using the sample versus population means and dividing by instead of . So and will be generally fairly close.

## Comparisons Using Standard Deviation

The measures of centre and measures of spread tell us different things about a data set.

Measures of centre summarise a data set and tell us about where the majority of the data lies.

|  | Benefits | Drawbacks |
| --- | --- | --- |
| Mean | Includes all of the data in the calculation, widely used | Heavily impacted by outliers |
| Median | Tells us the middle, not impacted by outliers | Does not include all data values |
| Mode | Quick to identify, tells us about the most frequent value(s) | Does not include all data value, is not necessarily in the middle |

Measures of spread tell us about the variability and consistency.

* Range
* Interquartile range (IQR)
* Standard deviation
* Mean absolute deviation

Previously, we saw that an **outlier** can skew the data. Outliers have an impact on certain measures of centre and spread.

The range, MAD, and the standard deviation are all affected by outliers. This makes sense as the mean is used as part of the calculation for the both MAD and the standard deviation, and the mean is affected by outliers.

|  | Benefits | Drawbacks |
| --- | --- | --- |
| Mean | Includes all of the data in the calculation, widely used | Heavily impacted by outliers |
| Median | Tells us the middle, not impacted by outliers | Does not include all data values |
| Mode | Quick to identify, tells us about the most frequent value(s) | Does not include all data value, is not necessarily in the middle |
| Range | Easy to calculate, tells about the extremes of the data | Heavily impacted by outliers |
| Interquartile range | Tells us about the middle half of the data, not impacted by outliers | Does not include all data values |
| MAD | Tells us about how far values are from the mean | Impacted by outliers, best to use technology to calculate |
| Standard deviation | Tells us about how far values are from the mean, widely used in other areas of statistics | Impacted by outliers, best to use technology to calculate |

When analysing sets of data, if the data has outliers, it is best to use the IQR to describe the spread of the data since the IQR is more resistant to outliers.

## Cumulative Frequency

We have seen that a **frequency table** allows us to group data into intervals and tells us how many data values are in that interval.

**Cumulative frequency** is a ‘running total’ of the frequencies. To calculate it, we add an additional column to the frequency distribution table:

| Class interval | Frequency | Cumulative frequency |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

* The first value in the cumulative frequency column will always be the same as the first value in the frequency column.
* To get the second cumulative frequency value we add the second frequency to the first cumulative frequency value,
* The second cumulative frequency value tells us there are  values in the interval .
* The third cumulative frequency value tells us there are  values in the interval , and so on.
* The final cumulative frequency value is always equal to the sum of the frequencies. In this case, there are  values in the entire data set, represented by .

### Cumulative Frequency Graphs

We have already seen how frequency tables can be used to create a histogram. The cumulative frequencies can also be plotted to create another type of chart, called a cumulative frequency graph. This graph is made up of a cumulative frequency histogram and polygon.

A **cumulative frequency histogram** will always increase in size from left to right. The frequency represented by any particular column will be equal to the difference in height between that column and the one before it.

A **cumulative frequency polygon**, is a line graph connecting cumulative frequencies at the upper endpoint of each class interval.

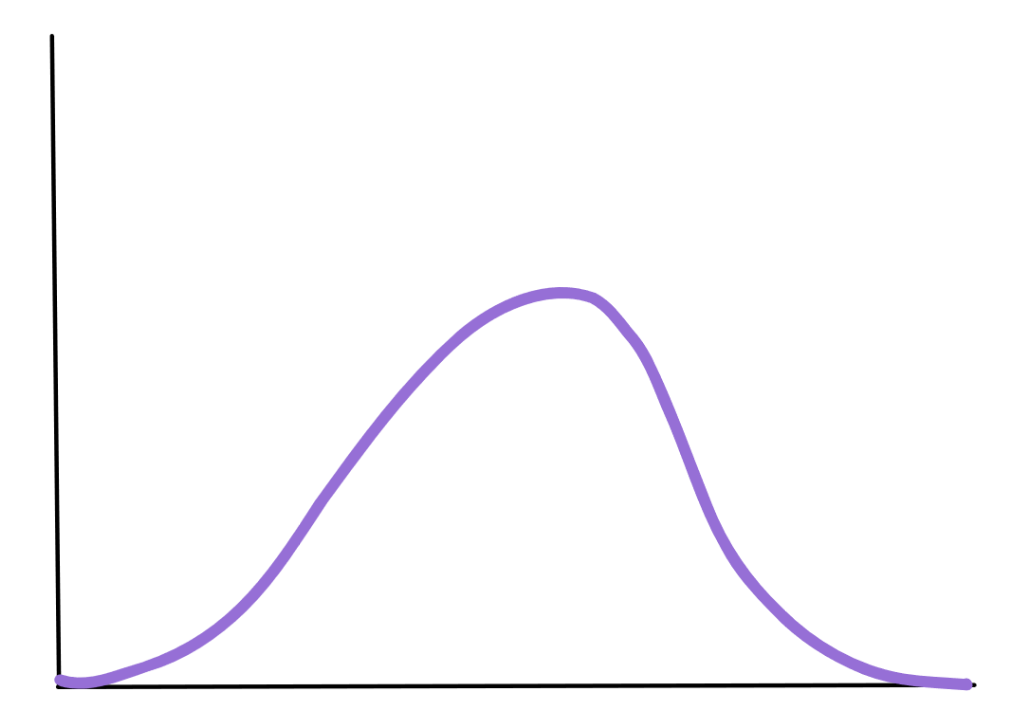
We can use a cumulative frequency polygon to find values such as the median and quartiles which are at % % and % of the cumulative frequencies.

## The Shape of Data

### Symmetry and Skew

Data may be described as **symmetrical** or **skewed**.

This is the shape that symmetrical data would look like. It is similar to the shape of a bell. This is often referred to as a normal distribution.

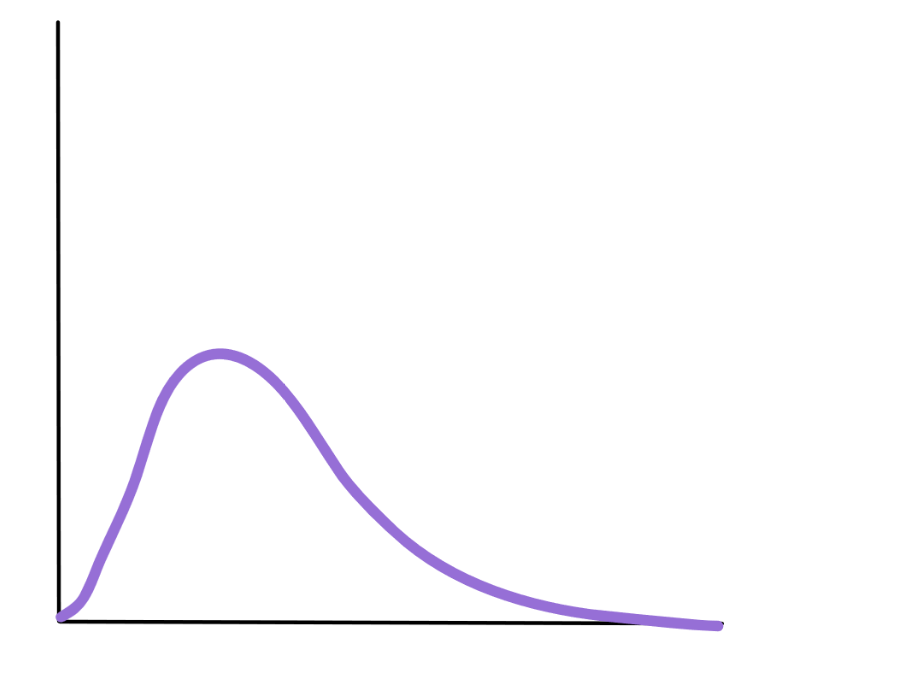


Normal Distribution|394x284

There are a few features that are related to **symmetrical** data:

* The data tends to be around a **central** value with **no bias** left or right.
* Roughly 50% of scores will be above the mean and 50% of scores will be below the mean.
* The **mean**, **median** and **mode** should **roughly coincide**.

This is the shape of data that is **positively skewed**. Notice the tail at the positive end.



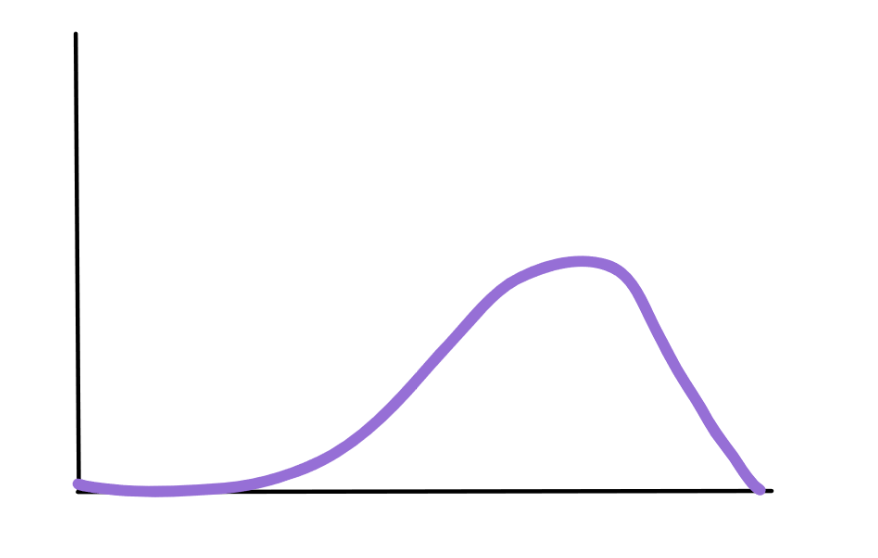
Positively Skewed|366x280

There are a few features that are related to **positively** skewed data:

* The data is bias to the **right**.
* The **mean** is usually **greater** than the **median**. The **median** is usually **greater** than the **mode**.

A **negatively skewed** graph has this general shape with left side stretched out.

This is the general shape shown over a histogram of positively skewed data.



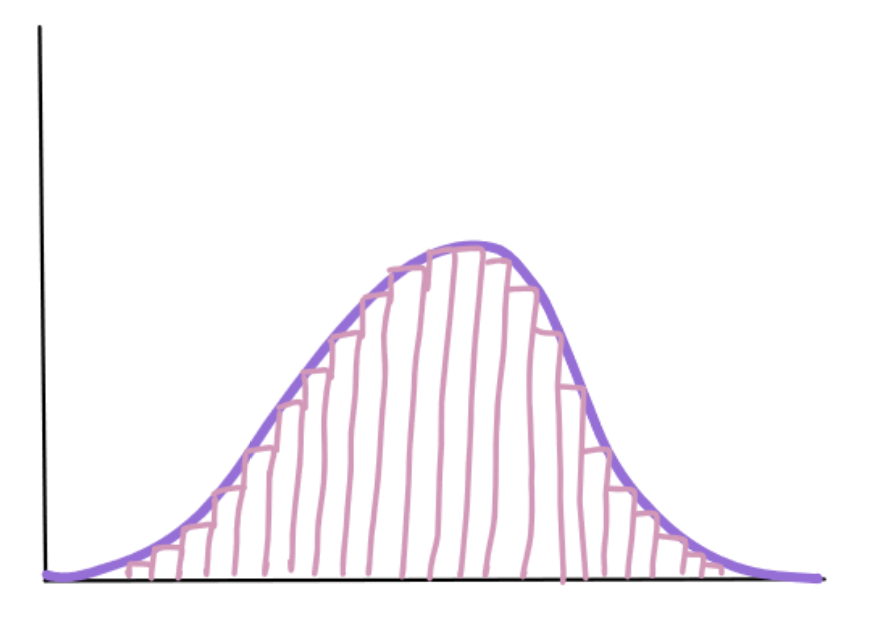
Negatively Skewed|343x218

There are a few features that are related to **negatively** skewed data:

* The data is bias to the **left**.
* The mean is usually **less** than the **median**. The **median** is usually **less** than the **mode**.

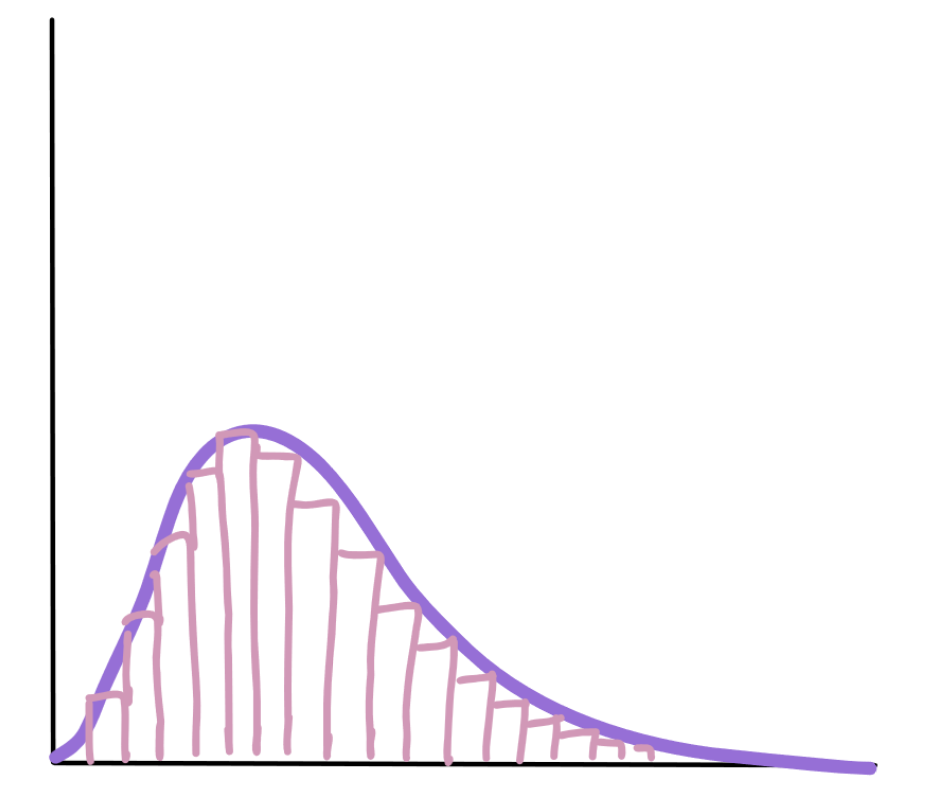
This is the general shape of a frequency histogram with a symmetrical shape.

The purple line shows the symmetrical curve that can be drawn over the histogram that the data roughly follows.



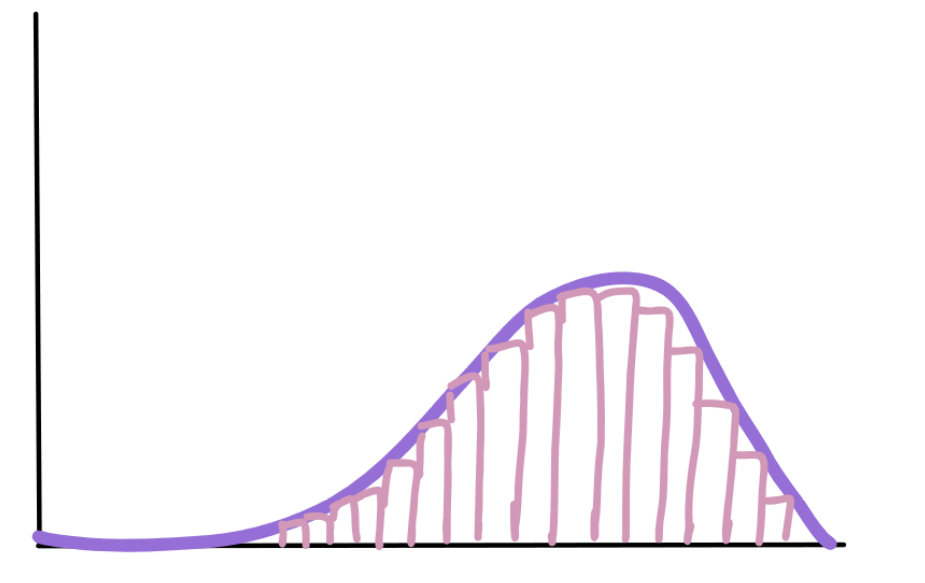
Normal Distribution with Freq.|267x191

This is the general shape shown over a histogram of **positively** skewed data.



Positive Skew with Freq.|281x239

This is the general shape shown over a histogram of **negatively** skewed data.

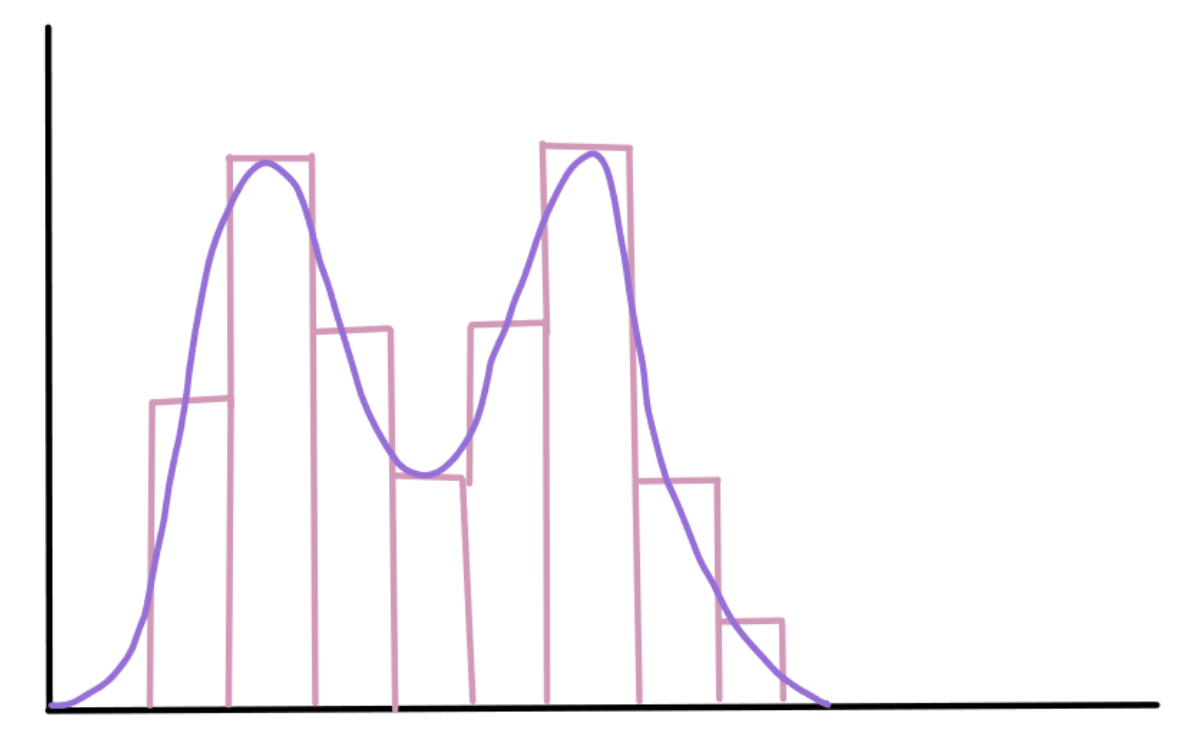


Negative Skew with Freq.|280x176

### Clusters

In a set of data, a **cluster** occurs when a large number of the scores are grouped together within a small range. Clustering may occur at a single location or several locations.

If the data has two clear peaks then the shape is called **bimodal**.



Example of Bimodal Data|300x184

If the shape of data has more than 22 peaks (modes) than it is described as **multimodal**.

## Stem and Leaf Dot Plots

## Quartiles

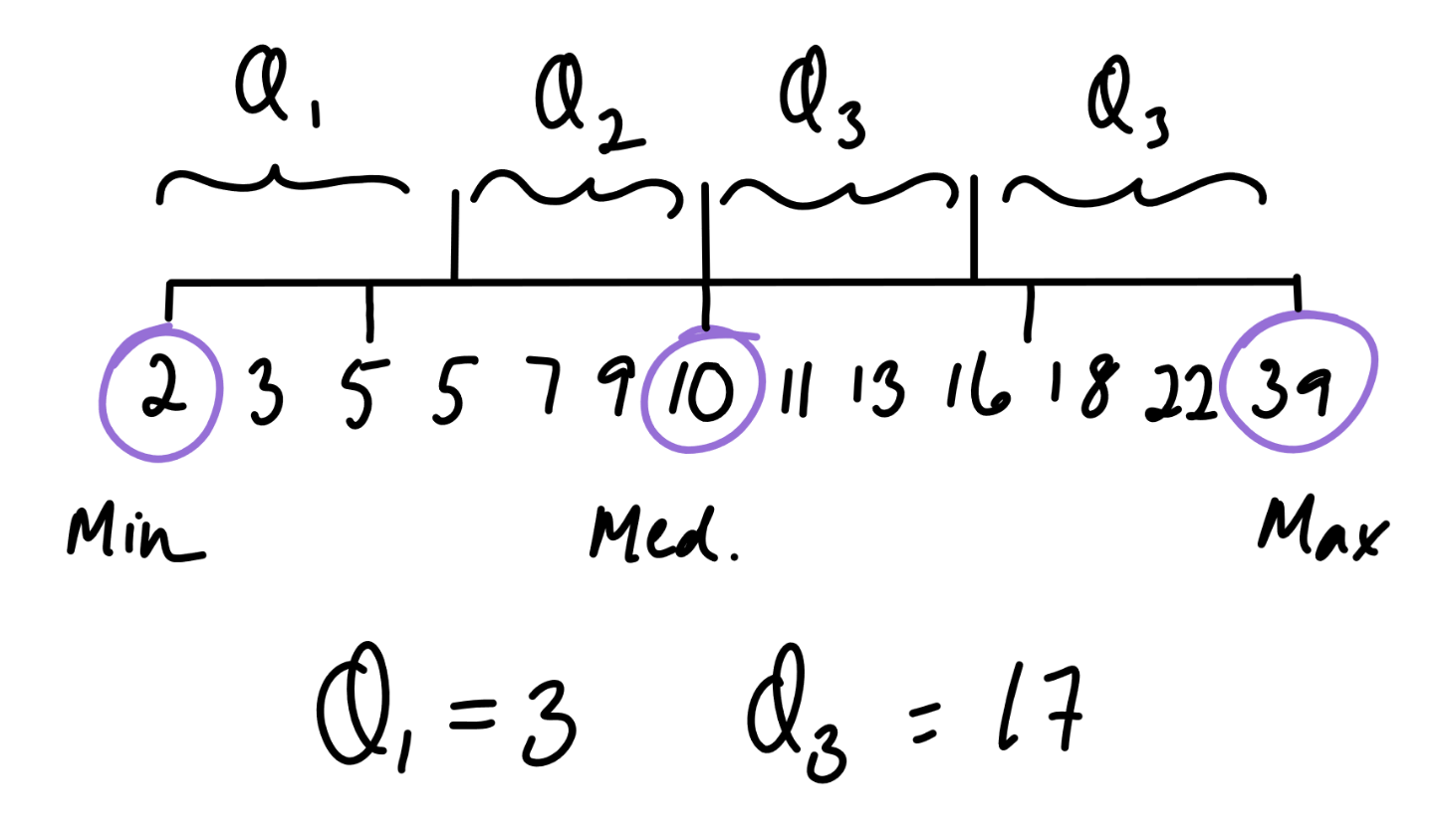
The **range** is a **measure of spread** based on the minimum and maximum in a data set, but does not tell us about the spread of data within these two values.

Range = Maximum - Minimum

Maximum The highest value in the data set Minimum The lowest value in the data set

The **median** is a **measure of centre**, and tells us where the middle of the data set is. For a data set with  items.

To get a better picture of the internal spread in a data set, we can find the **quartiles** of a data set, which can be used for a measure of spread called **interquartile range (IQR)**.



500x293

Quartiles are scores at particular locations in the data set-similar to the median, but instead of dividing a data set into halves, they divide a data set into quarters.

* Minimum: Lowest value
* **Lower quartile, ​**: At most 25% of the data is **below** this value
* Median: 50%50% of the data lies on **either** side of this value
* **Upper quartile, ​**: At most 25% of the data is **above** this value
* Maximum: highest value

Each quartile represents approximately 25% of the data set. For a data set with  items,

### Interquartile Range (IQR)

The **interquartile range** (IQR) is the difference between the third quartile and the first quartile. 50%50% of scores lie within the IQR.

Since it focuses on the middle 50%50% of the data set, the interquartile range often gives a better indication of the internal spread than the range does, and it is less affected by outliers.

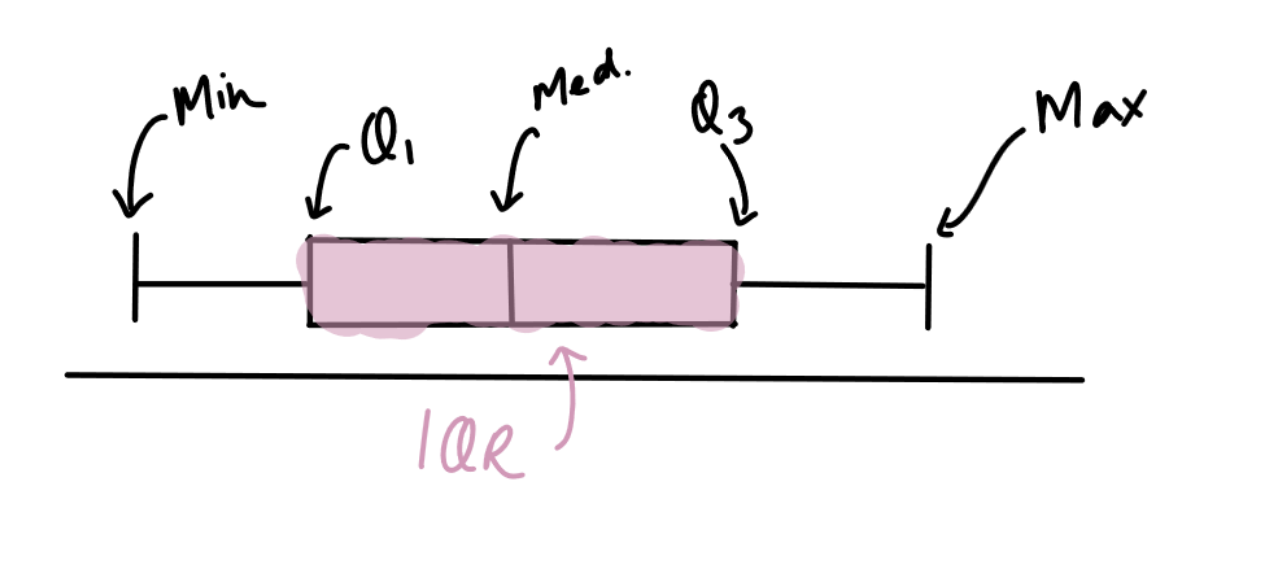
= Lower Quartile

= Upper Quartile

## Box Plots

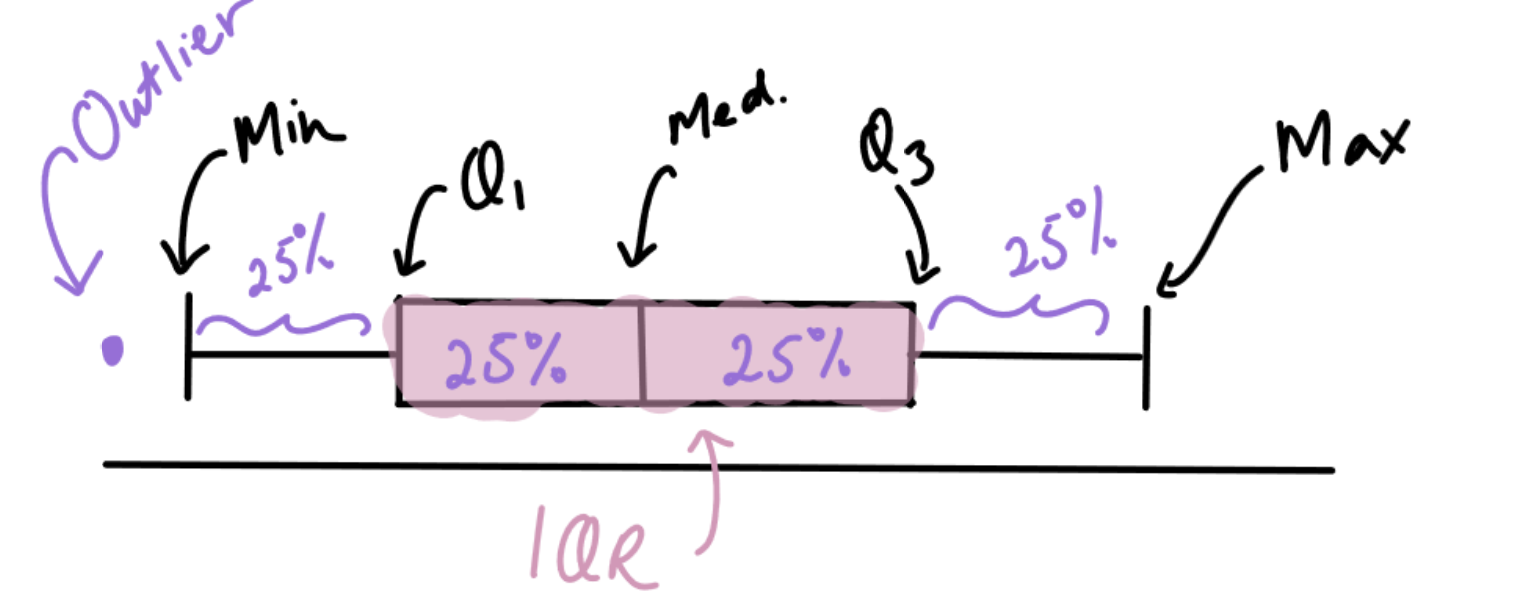
**Box plots** (also known as box-and-whisker plots) are a way of showing the five-number summary for a data set. The **five-number summary** consists of the following five statistics:

| Five Number Summary: |
| --- |
| Minimum |
| Lower quartile |
| Median |
| Upper quartile |
| Maximum |



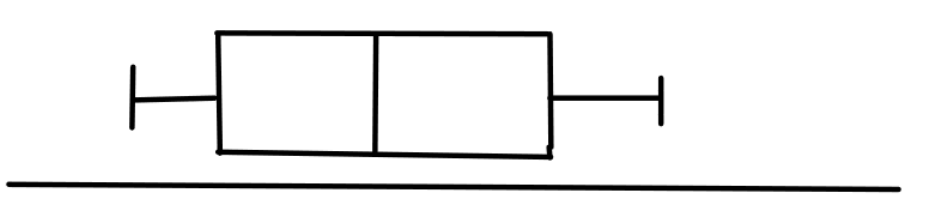
500x225

As you can see the box plot consists of a number line, a rectangle with a line inside (the box), and two horizontal lines (the whiskers). The box represents the middle 50% of the scores and its size tells us the interquartile range. About 25% of the scores lie in each quartile



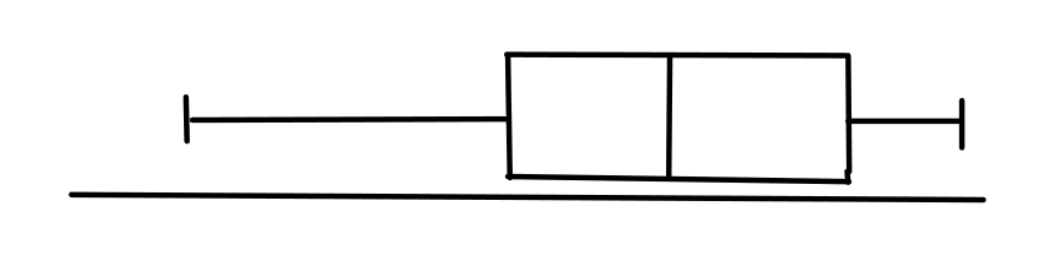
500x201

**Outliers** are data points which are very far away from the rest of the data set. On a box plot, the outliers are shown as points or crosses. When there are outliers, the end of the whiskers may not be the minimum or maximum.



500x117

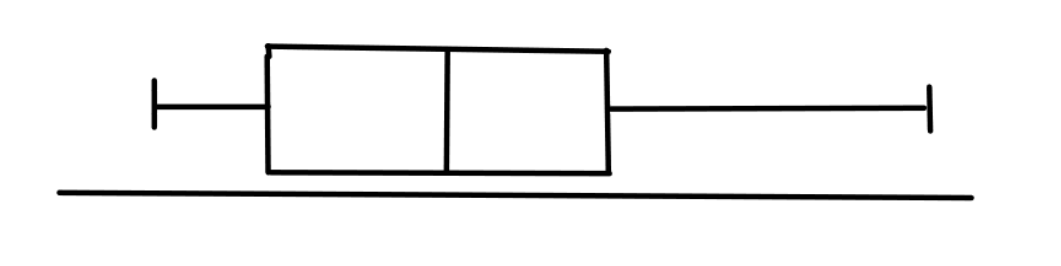
A **symmetrical** data set is distributed around the centre with a similar frequency on the left and right.



500x123

A **negatively skewed** data set is where the majority of the data points have higher values, with some data points at lower values. It is sometimes called a left skew.

The mean is less than the median.



500x125

A **positively skewed** data set is where the majority of the data points have lower values, with some data points at higher values. It is sometimes called a right skew.

The mean is greater than the median.

# Area A, Surface Area A and Volume A

## Area A

### Area of Polygons

**Area** is the amount of space inside a two-dimensional shape.

From smallest to largest, the common units used for measuring area are:

* square millimetres ()
* square centimetres ()
* square metres ()
* square kilometres ()

A hectare is a special unit of area often used to measure the area of a piece of land.

1 ha=10000 m21 ha=10000 m2.

A summary of the area formulas for common polygons:

A table showing the area formulas for common polygons. Ask your teacher for more information.

### Area of Circles

We can calculate the area of a **circle** using the formula:

When you are required to give the exact value of the area of a circle, you will leave your answer in terms of π.

To find the area of a semicircle (half a circle) we find the total area, then find a half of that by dividing by 2.

To find the area of quarter of a circle we find the total area, then find a quarter of that by dividing by 4.

### Area of Composite Shapes

There are two main methods for finding the area of a composite shape: the **addition method** and the **subtraction method.**

The addition method involves identifying the basic shapes, calculating the area of the basic shapes and adding them all together to find the total area of the composite shape.

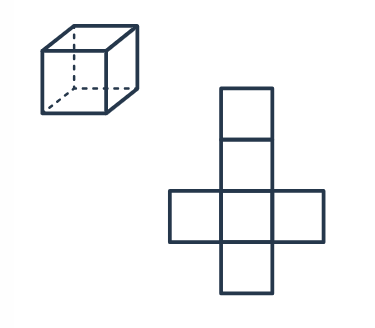
The subtraction method involves identifying the basic shapes, calculating the area of the basic shapes and subtracting the area of the smaller shapes (that were cut out) from the area of the larger shape to find the total area of the composite shape.

### Nets of Prisms

A **prism** is a three-dimensional figure that has a constant (uniform) **cross-section**.

The cross-section is parallel to the two identical ends.

A **net** of a prism is a representation of the faces arranged in such a way that it can be folded to form the three-dimensional prism. A prism can have many different nets. Nets can be useful to find the total surface area of a prism by finding the areas of all its faces.



A cube and its net|196x171

## Surface Area A

### Surface Area of Prisms

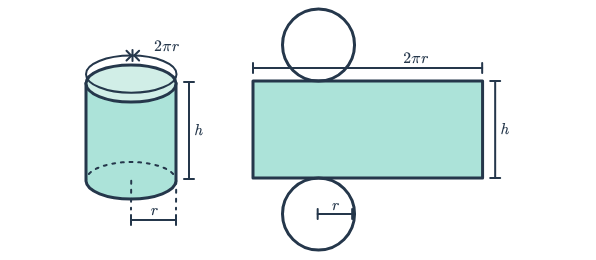
The **surface area** of a prism is the sum of the areas of all the faces.

To find the surface area of a prism, we can determine the basic shapes we need to add together. The **net** of a prism will show all of the faces as 2D shapes.

Another way to calculate the surface area for rectangular prisms and cubes is to use the formula. For any faces that are the same, we can find the area of one face and multiply it by how many times that face appears.

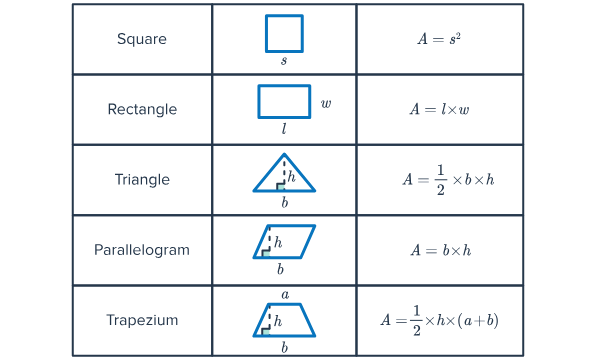
A **rectangular prism** has surface area:

### Surface Area of a Cylinder

A **cylinder** has three faces: two identical circular bases and a curved surface that joins the two bases together. The surface area of a cylinder is the sum of the areas of these three faces. The area of a circle is . Now we need to find the area of the curved face that connects the two circles. When unrolled the curved surface is a rectangle. One side length is equal to the height of the cylinder and the other side length is the circumference of the base circle. The area of the curved part of a cylinder is , where  is the radius and  is the height.  To find the surface area of the whole cylinder, we need to add the area of the top and bottom circles to the area of the curved part.

the radius of the clynders the height of thue cynlinder ## Volume A ### Volume of Prisms **Volume** is the amount of space a three-dimensional object takes up.

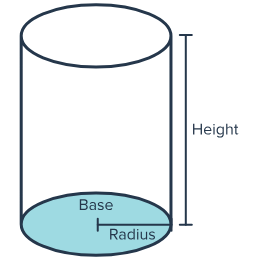
A prism is a three-dimensional object with two identically shaped ends joined by rectangular faces. For all prisms:

The first step in the calculation is to find the area of the base, which may be a basic shape or a composite shape.  ### Units for Volume and Capacity Volume is the amount of space a solid takes up. **Capacity** is the amount that something can hold.

Volume is measured using units such as:

* cubic millimetres ()
* cubic centimetres ()
* cubic metres ()
* cubic kilometres () The units of capacity include megalitres (), kilolitres (), litres () and millilitres (). When we consider liquids, the relationship between capacity and volume can be expressed as:

### Volume of Cylinders

A **cylinder** is very similar to a prism with a circular base.  The volume of a cylinder can be found in a similar way:

Remember the area of a circle is equal to . So, the volume of the cylinder can be written as:

where is the radius.

#### Volume of a part of Acylinder

For prisms and cylinders, we have seen that:

volume of the solid area of the end face height of the solid

When we are asked to find the volume of half or a quarter of a cylinder, we multiply the full volume by the fraction we are trying to find.

To find the volume of a tube, we can find the area of the end face (base) by subtracting the smaller circle from the larger one. We can then use the area of the base to find the volume of the solid.

### Volume of Composite Solids

**Composite solids** are three-dimensional objects composed of two or more simple geometric solids. We can find the volume of composite solids by either adding or subtracting the volumes of the individual shapes.