Y10 Maths Knowledge Organiser Higher Tier: Frequency Distributions and Charts

What must I be able to do?	Key vocabulary	
New content: ☐ Understand different methods of sampling ➢ Mathswatch 152 and 176 (GCSE) ☐ Draw and interpret a stem and leaf diagram ➢ Mathswatch 128b (GCSE) ☐ Draw and interpret frequency polygons ➢ Mathswatch 65b (GCSE) ☐ Draw and interpret cumulative frequency graphs and boxplots ➢ Mathswatch 186 and 187 (GCSE) ☐ Draw and interpret histograms ➢ Mathswatch 205 (GCSE)	**Population**	The whole group of people or items being studied.
	Sample	A selection taken from the population.
	Bias	A built in error that makes values incorrect.
	Cumulative frequency	The total of a frequency and all frequencies so far in a distribution.
	Quartiles	A quartile divides data into 4 quarters, the lower quartile (25%), median (50%) and upper quartile (75%).
	Frequency Density	Frequency density = frequency ÷ class width.

Sampling

Random samples are where each item in the population has an equal chance to be picked. The most common method is assigning each value in the population a number, then randomly picking numbers out of a hat or using a random number generator.

Stratified samples use sub-groups in the population sampled in the same proportion as in the population e.g. If a population of Y7s has 80 girls and 40 boys, the sample of 10% (12 students) will have twice as many girls as boys to retain the ratio of girls to boys, therefore 8 girls and 4 boys.

Sampling can be used to **approximate the size of a population** by doing a capture/recapture method:

e.g. There are an unknown number of birds in a colony. 30 birds are captured and have a tag fitted, then released. The following week a further 30 birds are captured and only 8 have a tag. Approximately how many birds are in the colony?

$\frac{30}{n} = \frac{8}{30}$

Sample divided by population (n)

Tagged in 2nd sample ÷ total tagged

multiplying by n and by 30 gives us $900 = 8n$

Therefore $n = 900 \div 8 = 112.5$ = approximately 113 birds.

Stem and Leaf diagrams

A stem and leaf diagram is used to represent an ordered set of data. It must contain a **key**. The stem goes vertically downwards and the leaves go horizontal.

e.g. Put the following data into a stem and leaf diagram.

4, 14, 17, 17, 24, 25, 26, 30, 31, 33, 34, 34, 35.

0	4
1	4, 7, 7
2	4, 5, 6
3	0, 1, 3, 4, 4, 5

Key

1 | 4 = 14

As the data in the table is in order it is easy to find the median and quartiles of the data.

0	4
1	4, 7, 7
2	4, 5, 6
3	0, 1, 3, 4, 4, 5

Lower quartile = 17 (middle of 17 and 17 as it is the median of the first half of the data)

Median = 26 (middle of all of the data)

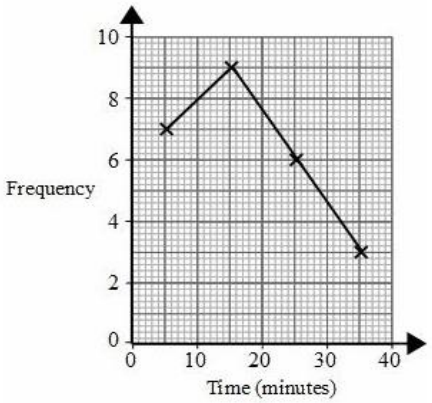
Upper quartile = 33.5 (middle of 33 and 34 as it is the median of the second half of the data)

Frequency Polygons

A frequency polygon is a line graph which is joined using straight lines. Frequency is plotted on the vertical axis and if the data is **grouped**, plot the **midpoint** on the horizontal axis. The horizontal axis should be a linear scale, not grouped and the vertical axis should start from 0.

e.g.

Time	Frequency
$0 \leq t < 10$	7
$10 \leq t < 20$	9
$20 \leq t < 30$	6
$30 \leq t < 40$	3



Histograms

A histogram is similar to a bar chart, but where a bar chart is used for categorical or discrete data, we use a histogram for continuous data e.g. heights, weights, time etc.

Key features:

- There are no gaps between bars and bars may be different widths
- The horizontal scale is linear and not grouped
- The vertical axis is labelled **frequency density**
- The **frequency** is represented by the **area of each bar** rather than the height of each bar

e.g. Draw a histogram of the following data

Length of time	Frequency
$0 \leq t < 10$	60
$0 \leq t < 15$	40
$15 \leq t < 20$	75
$20 \leq t < 50$	150

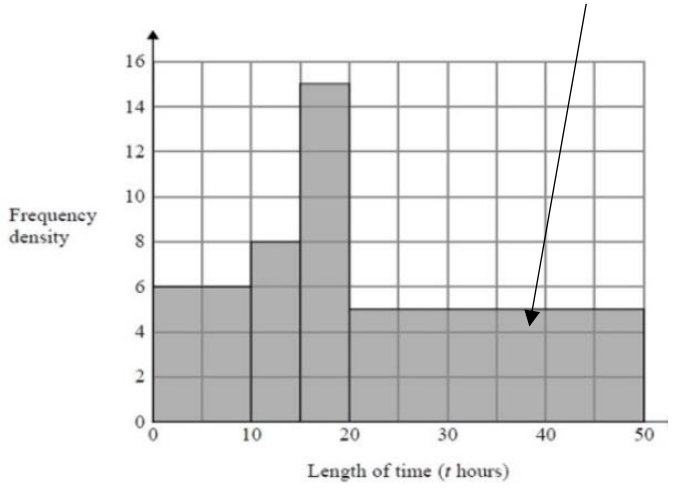
First we need to calculate the frequency density

$$\text{Frequency density} = \frac{\text{Frequency}}{\text{Class width}}$$

Length of time	Frequency	Frequency density
$0 \leq t < 10$	60	$60 \div 10 = 6$
$10 \leq t < 15$	40	$40 \div 5 = 8$
$15 \leq t < 20$	75	$75 \div 5 = 15$
$20 \leq t < 50$	150	$150 \div 30 = 5$

Class width is the difference between the 2 bounds so this one is $50 - 20 = 30$

The area of this bar is 150 which is the frequency of the group $20 \leq t < 50$



Cumulative Frequency graphs

A cumulative frequency table shows a running total of the frequencies. A cumulative frequency diagram or graph, is drawn by **plotting** the **cumulative frequency** against the **upper boundary** of the class interval and then joined together.

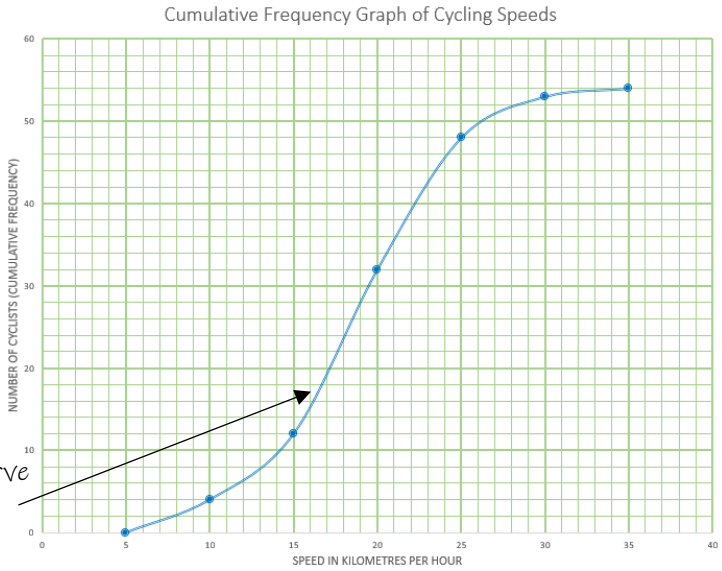
e.g. Plot a cumulative frequency diagram of the following data

Speed of cyclists	Frequency	Cumulative frequency
$5 \le t < 10$	4	4
$10 \le t < 15$	8	4 + 8 = 12
$15 \le t < 20$	20	12 + 20 = 32
$20 \le t < 25$	16	32 + 16 = 48
$25 \le t < 30$	5	48 + 5 = 53
$30 \le t < 35$	1	53 + 1 = 54

Start by calculating the cumulative frequency

Plot the upper bound of the groups against cumulative frequency so (10, 4), (15, 12) and so on.

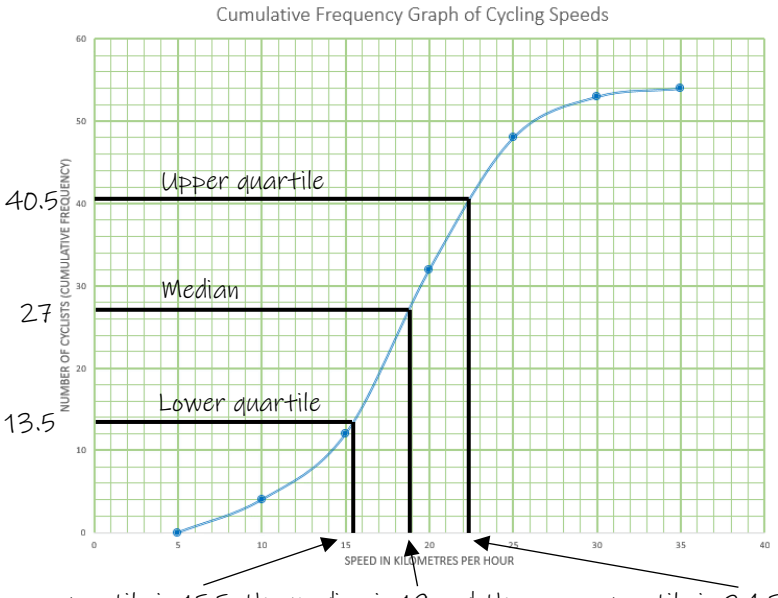
Join as a smooth curve or as straight lines between each point



A cumulative frequency graph can be used to estimate the lower quartile, median and upper quartile of grouped data.

Find one quarter of the total cumulative frequency, (in this case 54 ÷ 4 = 13.5), one half of the total cumulative frequency (54 ÷ 2 = 27) and three quarters of the total cumulative frequency (13.5 x 3 = 40.5).

Draw a line across from each of these points until they hit the curve, then go down to the horizontal scale and read off.

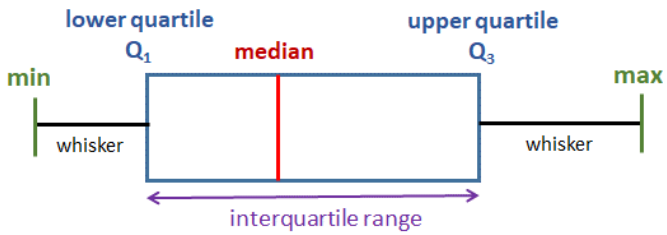


The lower quartile is 15.5, the median is 19 and the upper quartile is 24.5

Boxplots (or box and whisker diagrams)

A boxplot is used as a visual representation of the spread of data. It shows the smallest value, largest value, lower quartile, upper quartile and median. The actual box represents the spread of the middle 50% of the data which is known as the **interquartile range**. The first 25% of the data is the first whisker and the final 25% of the data is the second.

Interquartile range = upper quartile – lower quartile.



A box plot is drawn on top of a normal linear scale so that values can be read off and compared.

GLUE

HERE