GCSE Maths – Statistics

Grouped Discrete Data and Continuous Data
(Higher Only)

Notes

WORKSHEET

Discrete Data

Discrete data – data that must take one of a set of certain values. For example, the number of people in a room cannot be 5.25, it can only take a whole number value such as 5.

Grouped discrete data – data points are given as values within a group. We call this group a **class**. For example, there are 3 people who own 1-3 pairs of shoes, 2 people who own 4-6 pairs of shoes, and so on.

Continuous Data

Continuous data – numerical data which can take any value in a given range. Examples include age, height and weight.

Grouped continuous data – data points are given as values within a group. For example, there might be 7 netball players with height in the class 150cm < height ≤ 160cm, 4 netball players with height in the class 160cm < height ≤ 170cm.

Histograms

We use histograms to display **continuous grouped data**.

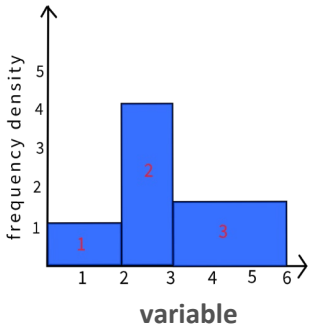
Histograms look similar to bar graphs. However, the height of each bar does not represent the frequency. Instead,

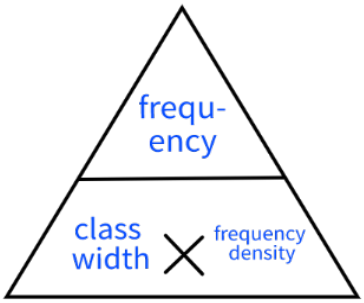
$$\textbf{\textit{Area of Bar}} = \textbf{\textit{Frequency}}$$

The area of the bar is its height multiplied by its width.

The **class interval or class width** is the size of each class (group of data). Widths might be equal or unequal. This is plotted along the x-axis.

Frequency density is the frequency divided by the class width. Frequency density is plotted on the y-axis and is equal to the height of each bar.





The frequency density, frequency and class width are related, as shown in the formula triangle. Using the triangle, we can easily find the formula for each quantity by covering that section.

For example, to find class width, cover 'class width' to see that:

$$Class\ Width = \frac{Frequency}{Frequency\ Density}$$

Example: 100 people were timed while completing a puzzle. Their results are shown in the table. Create a histogram displaying their times.

Time (minutes)	Frequency
$0 < t \leq 10$	10
$10 < t \leq 15$	15
$15 < t \leq 20$	30
$20 < t \leq 30$	20
$30 < t \leq 45$	25

1. **Add columns to the table for class width and frequency density.**

   Calculate class width and use this value along with frequency to calculate each frequency density.

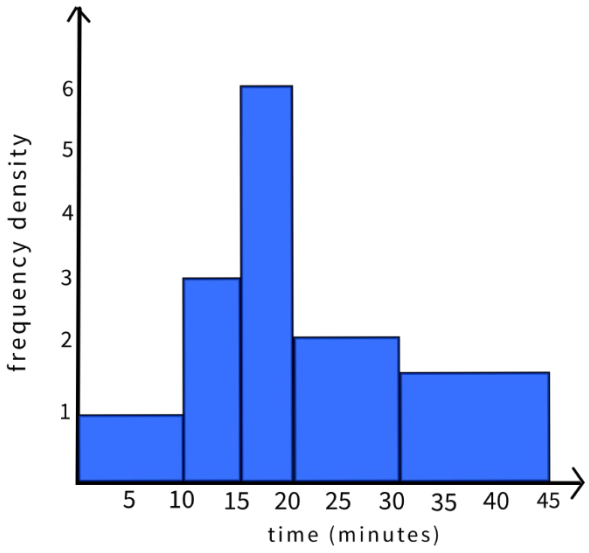
Time (minutes)	Frequency	Class width	Frequency Density
$0 < t \leq 10$	10	10	$10 \div 10 = 1$
$10 < t \leq 15$	15	5	$15 \div 5 = 3$
$15 < t \leq 20$	30	5	$30 \div 5 = 6$
$20 < t \leq 30$	20	10	$20 \div 10 = 2$
$30 < t \leq 45$	25	15	$25 \div 15 = 1.67$

2. **Use the class widths and frequency densities to plot the histogram.**

   Draw the x- and y-axis. Label the x-axis with the investigating variable (time) and the y-axis with frequency density.

   Draw each bar by plotting the length of time against the frequency density.

   For example, the first group is 10 minutes long, and has a frequency density of 1. The bar that represents it should be 10 units wide and 1 unit tall.



Cumulative frequency graphs

Cumulative frequency graphs show the **total frequencies** of grouped data at fixed points. The variable measured is on the x-axis, and cumulative frequency on the y-axis. The cumulative frequency is a running total of all recorded frequencies so far, and increases as you add each grouped frequency. You can use cumulative frequency graphs to find the **median** and **interquartile range**.

Example: The hop length of 100 footballers was collected and recorded in a table.

Length (metres)	Frequency
$0 < m \leq 1$	10
$1 < m \leq 2$	20
$2 < m \leq 2.5$	30
$2.5 < m \leq 3$	30
$3 < m \leq 4$	10

Draw a cumulative frequency graph to display this data.

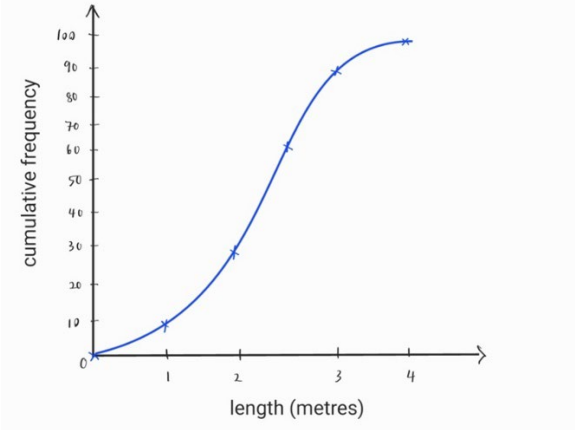
1. **Add a column to the table for cumulative frequency and upper boundary.**

   The cumulative frequency is a running total taken at the end of each group whereas the upper boundary is the highest value within the class group.

Length (metres)	Frequency	Cumulative frequency	Upper boundary
$0 < m \leq 1$	10	10	1
$1 < m \leq 2$	20	20 + 10 = 30	2
$2 < m \leq 2.5$	30	30 + 30 = 60	2.5
$2.5 < m \leq 3$	30	60 + 30 = 90	3
$3 < m \leq 4$	10	90 + 10 = 100	4

2. **Plot cumulative frequency graph**

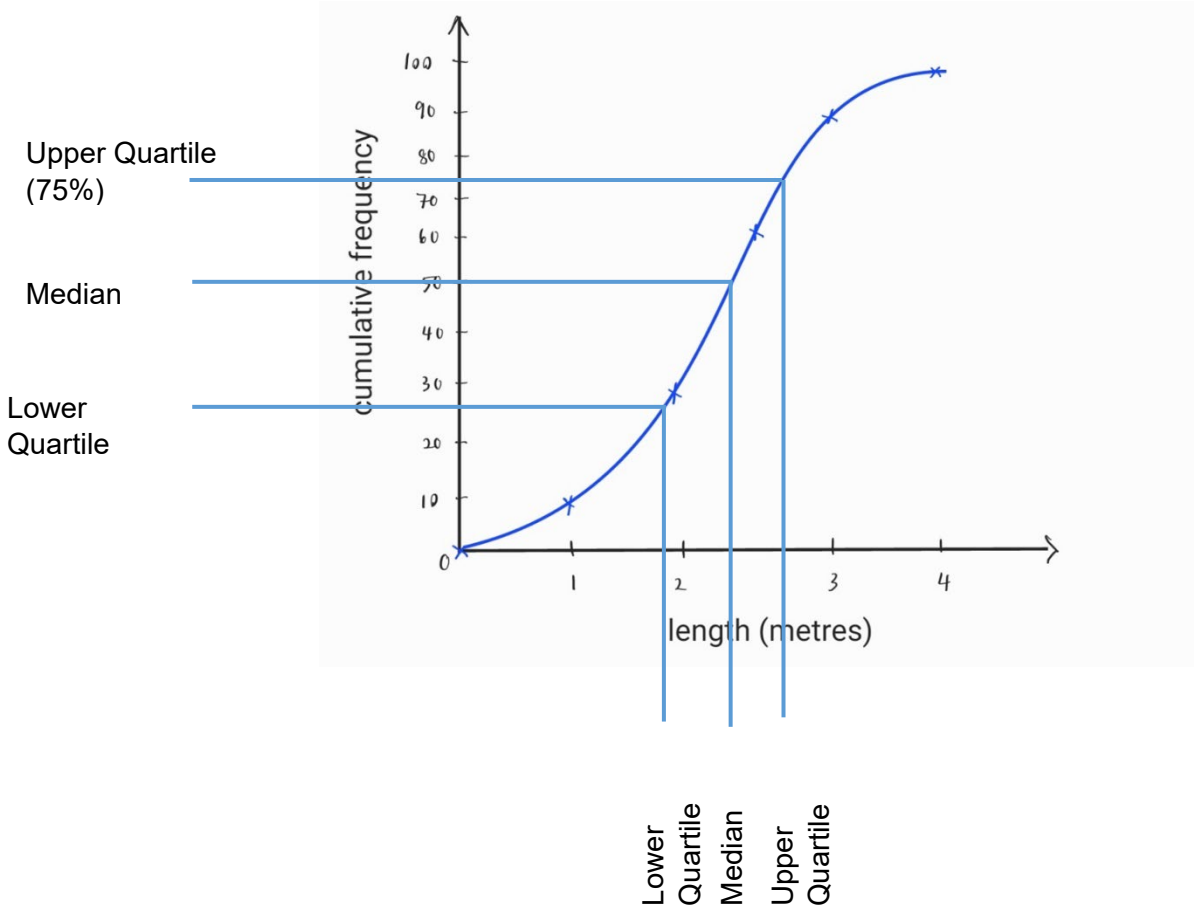
   Plot a point showing cumulative frequencies on the upper boundary of each group. For example, the first point would be at 10 (the cumulative frequency) and 1 (the upper bound of the first group). Draw this point on the graph at $(10, 1)$. *Join all the points with a smooth curve, making sure your line passed through the origin* $(0,0)$.



Interpreting cumulative frequency graphs

To find the **median** value from a cumulative frequency graph, draw a line across from the point on the y-axis which is **half of the total frequency**, and draw a line down to the x-axis to get a reading. This value from the x-axis is the median of the data.

To find the **interquartile range**, use the same process to find the lower quartile (Q1) which is at 25% of cumulative frequency, and upper quartile (Q3) which is 75% of the cumulative frequency. The interquartile range is Q3 – Q1.



Grouped Discrete Data and Continuous Data (Higher Only) - Practice Questions

1. Below is a grouped frequency table showing the results of an English test.

   a) Draw a histogram showing the results of the test.

Score, p (%)	Frequency
$40 < p \leq 50$	5
$50 < p \leq 60$	12
$60 < p \leq 70$	8
$70 < p \leq 90$	15
$90 < p \leq 100$	10

   b) Calculate an estimate for the median score using the histogram.

2. The table below shows information about how long it takes a group of children to get to school.

   Draw a histogram to show this data.

Time, t (minutes)	Frequency
$0 < t \leq 5$	4
$5 < t \leq 15$	18
$15 < t \leq 20$	10
$20 < t \leq 30$	6
$30 < t \leq 40$	2

3. Below is a frequency table showing the amount of time people spent on social media per day. Plot a cumulative frequency graph to show this data.

Time, t (minutes)	Frequency
$0 < t \leq 5$	10
$5 < t \leq 15$	25
$15 < t \leq 30$	20
$30 < t \leq 50$	15
$50 < t \leq 60$	20

4. Below is a frequency table showing the price brackets that some laptops on sale fall into. Plot a cumulative frequency graph to show this data.

Price, p (£)	Frequency
$50 < p \leq 100$	2
$100 < p \leq 150$	15
$150 < p \leq 200$	9
$200 < p \leq 300$	19
$300 < p \leq 450$	5

Worked solutions for the practice questions can be found amongst the worked solutions for the corresponding worksheet file.