

**OCR**

**A Level**

# A Level Mathematics

## Outliers and Cleaning Data

Name:

**M M E**

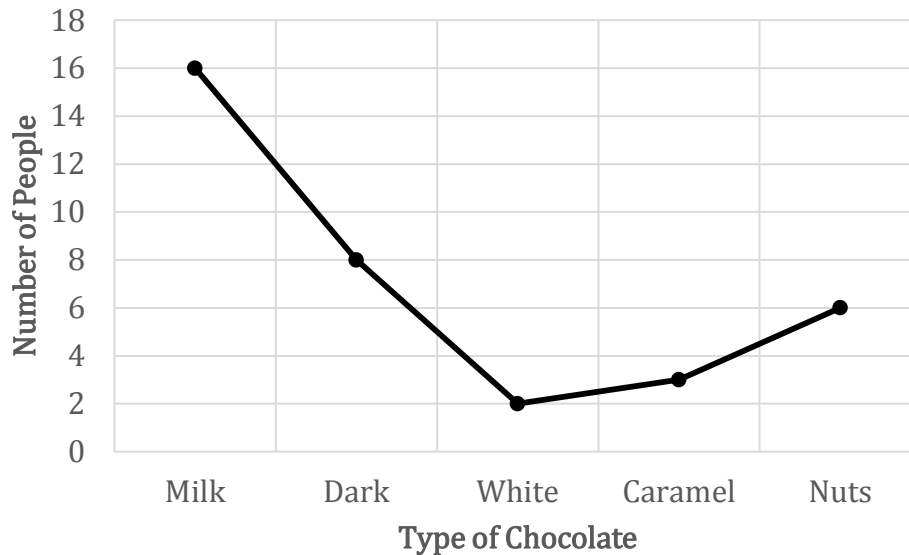
Mathsmadeeasy.co.uk

Total Marks:

## L4- Outliers and Cleaning Data- Questions

OCR

- 1) Kevin has the results of 34 people's favourite chocolate. He has represented it in a graph and calculated the mode as milk, the median as dark, the range as 13, and the mean as 6.8.



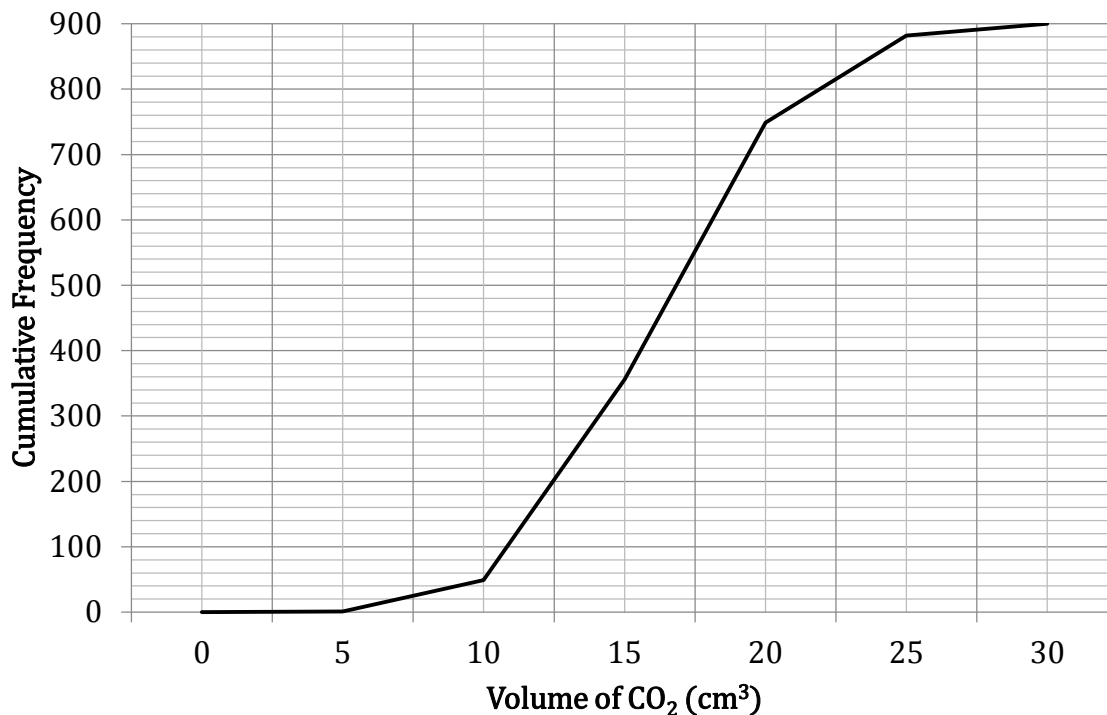
Critique the following:

- i) Kevin's choice of graph. [1]
  - ii) His summary statistics. [4]
  - iii) His decision to remove *Fruit* because he considered it to be outlier. [1]
- 2) The number of spots on insects at a nature park was recorded for one day in July. The summary table of this experiment is shown in the grouped frequency table.

Number of Spots ( $s$ )	$0 < s \leq 2$	$2 < s \leq 4$	$4 < s \leq 6$	$6 < s \leq 8$	$8 < s \leq 10$
Frequency	3	8	13	14	6

- i) Calculate the mean and range number of spots. [4]
  - ii) Draw an appropriate chart for the data, and use this chart to obtain an estimate for the median [4]
- 3) Salaries of 30 people at a company are normally distributed, with a mean of 25,000 and standard deviation of 2000. For each of the following scenarios state what will happen if:
- i) One person, with a salary  $10^2$  larger than the mean is added. [2]
  - ii) One person, with a salary of  $10^0$  larger than the mean is added. [2]

- 4) The volume of  $CO_2$  a machine produces was measured 900 times in June. These volumes are displayed in the cumulative frequency graph below.



- i) Estimate the median and interquartile range of the data

[1 mark]

An estimate of the median lies around 16.

[1 mark]

An estimate of the interquartile range is  $18$  (75%)  $-$   $13$  (25%)  $= 5$ .

- ii) Identify the outliers, if there are any.

The interquartile range is 5. Estimates of outliers are  $1.5 \times$  *interquartile range* above the upper quartile and similarly, below the lower quartile.

[1 mark]

$$upQ + 1.5(IQR) = 18 + 7.5 = 22.5$$

[1 mark]

$\therefore$  all values above 22.5 could be considered outliers, of which there are about 125 readings.

[1 mark]

$$lowQ - 1.5(IQR) = 13 - 7.5 = 6.5$$

[1 mark]

$\therefore$  all values below 6.5 are outliers, about 20 readings.

**iii) Discuss the whether these outliers should be removed.**

[1 mark]

There is no reason to remove them unless the recording machine was faulty. Knowing the upper limits of these emissions might prove useful.

**iv) On the next six occasions, the machine reported negative values. Discuss how adding these into the data would alter the graph.**

[1 mark]

No. These outliers are because of an error in the machine, or the person recording the reading. You cannot have a negative volume.