SL Paper 1

In a particular week, the number of eggs laid by each hen on a farm was counted. The results are summarized in the following table.

Number of eggs	1	2	3	4	5	6
Frequency	4	7	12	10	14	13

- a. State whether these data are discrete or continuous.
- b. Write down
 - (i) the number of hens on the farm;
 - (ii) the modal number of eggs laid.

c. Calculate

- (i) the mean number of eggs laid;
- (ii) the standard deviation.

Markscheme

a. discrete (A1) (C1)

- b. (i) 60 **(A1)**
 - (ii) 5 **(A1) (C2)**
- c. (i) $\frac{1 \times 4 + 2 \times 7 + 3 \times 12...}{60}$ (M1)

Notes: Award (*M1*) for an attempt to substitute into the "mean of a set of data" formula, with at least three correct terms in the numerator. Denominator must be 60.

Denominator must be 60.

Follow through from part (b)(i), only if work is seen.

= 4.03 (4.03333...) (A1)

Notes: Award at most (M1)(A0) for an answer of 4 but only if working seen.

(ii) 1.54 (1.53803...) (A1) (C3)

Examiners report

- a. ^[N/A]
- b. [N/A]
- c. ^[N/A]

[1]

[2]

[3]

The resting pulse rates of a group of 10 students who exercise regularly are given below.

65, 62, 75, 63, 69, 58, 65, 67, 55, 60

a.	Find the median resting pulse rate of the students.	[2]
b.	Find the mean resting pulse rate of the students.	[2]
c.	A new student joins the class and the mean resting pulse rate of the group of 11 students becomes 65.	[2]
	Find the resting pulse rate of the student who joined the group.	

Markscheme

a. Attempt to order set of numbers (M1)

64 (A1) (C2)

[2 marks]

```
b. \frac{639}{10} (M1)
```

Note: Award (M1) for their sum divided by 10.

63.9 (A1) (C2)

[2 marks]

c. $\frac{(639+x)}{11} = 65$ or equivalent *(M1)*

x = 76 (A1)(ft) (C2)

Notes: Award (*M1*) for setting up an equation (*their* part (b) $\times 10 + x$)/11 = 65. Follow through from their sum seen in part (b). Accept correct alternative methods but not trial and error.

[2 marks]

Examiners report

- a. This question proved to be relatively easy for most candidates. They could find the median, mean and also the pulse rate of the student who joined the group. Where mistakes were made, they were in not ordering the list of numbers. Part (c) presented the most challenge for weaker candidates.
- b. This question proved to be relatively easy for most candidates. They could find the median, mean and also the pulse rate of the student who joined the group. Where mistakes were made, they were in not ordering the list of numbers. Part (c) presented the most challenge for weaker candidates.
- c. This question proved to be relatively easy for most candidates. They could find the median, mean and also the pulse rate of the student who joined the group. Where mistakes were made, they were in not ordering the list of numbers. Part (c) presented the most challenge for weaker candidates.

The time, in minutes, that students in a school spend on their homework per day is presented in the following box-and-whisker diagram.



a. Find

- (i) the longest amount of time spent on homework per day;
- (ii) the interquartile range.
- b. State the statistical term corresponding to the value of 140 minutes.
- c. Find the percentage of students who spend
 - (i) between 100 and 140 minutes per day on their homework;
 - (ii) more than 100 minutes per day on their homework.

Markscheme

a. (i) 300 (minutes) OR 5 hours (A1)

Note: If answer given in hours, the unit must be seen.

(ii) 220 - 100 (M1)

Notes: Award (M1) for the two quartiles seen.

= 120 (minutes) OR 2 hours (A1) (C3)

Note: If answer given in hours, the unit must be seen.

b. median (time spent on homework per day) (A1) (C1)

Note: Do not accept middle or medium etc.

c. (i) 25 (A1)

Examiners report

- a [N/A]
- b. ^[N/A]
- c. ^[N/A]

[1]

[2]

⁽ii) 75 (A1) (C2)

The grades obtained by a group of 13 students are listed below.

$5\ 3\ 6\ 5\ 7\ 3\ 2\ 6\ 4\ 6\ 6\ 6\ 4$

a.	Write down the modal grade.	[1]
b.	Find the mean grade.	[2]
c.	Write down the standard deviation.	[1]
d.	Find the interguartile range.	[2]

Markscheme

a. 6 (A1) (C1)

[1 mark]

b. $\frac{5+3+6+\ldots+4}{13}$ (M1)

Note: Award (M1) for correctly substituted mean formula, division by 13 must be seen.

$$=4.85\left(rac{63}{13}
ight)$$
 (4.84615...) (A1) (C2)

[2 marks]

c. 1.46 (1.4595...) (A1) (C1)

[1 mark]

d. 6 - 3.5 (M1)

```
= 2.5 (A1) (C2)
```

Note: Award (*M1*) for their quartiles seen or a correct ordered list. Accept a correct ordered list from any previous part of the question. [2 marks]

Examiners report

a. Whilst many knew what the mode was, there was some confusion by weaker candidates who interpreted the required value as the maximum value in the list, namely 7. Many candidates gave the correct value of the mean in part (b) but a surprising number seemed to have selected the incorrect value from their calculator display for the standard deviation. In mathematical studies, the smaller value of the standard deviation, σ_x , should be given. In part (d), ordered lists were often seen but it proved problematic for a significant number of candidates to find Q_1 and Q_3 . Common mistakes such as $Q_3 = \frac{6+6}{2} = 6.5$ and $Q_1 = 3$ were seen on many scripts.

- b. Whilst many knew what the mode was, there was some confusion by weaker candidates who interpreted the required value as the maximum value in the list, namely 7. Many candidates gave the correct value of the mean in part (b) but a surprising number seemed to have selected the incorrect value from their calculator display for the standard deviation. In mathematical studies, the smaller value of the standard deviation, σ_x , should be given. In part (d), ordered lists were often seen but it proved problematic for a significant number of candidates to find Q_1 and Q_3 . Common mistakes such as $Q_3 = \frac{6+6}{2} = 6.5$ and $Q_1 = 3$ were seen on many scripts.
- c. Whilst many knew what the mode was, there was some confusion by weaker candidates who interpreted the required value as the maximum value in the list, namely 7. Many candidates gave the correct value of the mean in part (b) but a surprising number seemed to have selected the incorrect value from their calculator display for the standard deviation. In mathematical studies, the smaller value of the standard deviation, σ_x , should be given. In part (d), ordered lists were often seen but it proved problematic for a significant number of candidates to find Q_1 and Q_3 . Common mistakes such as $Q_3 = \frac{6+6}{2} = 6.5$ and $Q_1 = 3$ were seen on many scripts.
- d. Whilst many knew what the mode was, there was some confusion by weaker candidates who interpreted the required value as the maximum value in the list, namely 7. Many candidates gave the correct value of the mean in part (b) but a surprising number seemed to have selected the incorrect value from their calculator display for the standard deviation. In mathematical studies, the smaller value of the standard deviation, σ_x , should be given. In part (d), ordered lists were often seen but it proved problematic for a significant number of candidates to find Q_1 and Q_3 . Common mistakes such as $Q_3 = \frac{6+6}{2} = 6.5$ and $Q_1 = 3$ were seen on many scripts.

The mean of the ten numbers listed below is 6.8.

a. Write down an equation in terms of <i>p</i> and <i>q</i> .	[2]
b.i. The mode of these ten numbers is five and p is less than q .	[1]
Write down the value of <i>p</i> .	
b.ii. The mode of these ten numbers is five and p is less than q .	[1]
Write down the value of <i>q</i> .	
c. Find the median of the ten numbers.	[2]

Markscheme

a. $\frac{8+5+5+10+8+4+9+7+p+q}{10} = 6.8$ or equivalent (M1)(A1) (C2)

Note: Award (M1) for correct substituted mean formula, (A1) for correct substitution.

b.i.p = 5 (A1)(ft)

[1 mark]

b.ii*q* = 7 (A1)(ft) (C2)

Note: Follow through from their answers to parts (a) and (b) (i).

[1 mark]

c. 7 (M1)(A1)(ft) (C2)

Notes: Award *(M1)* for an attempt to order their numbers. Follow through from their answers to parts (b)(i) and (ii).

[2 marks]

Examiners report

- a. A large number of candidates gained full marks on this question. Many correct variations of the equation were given and the values of p, q and the median could then be found. Some candidates neglected the extra information of p less than q and lost a mark for having these values the wrong way around. Follow through marks could be awarded for the median, if working was shown, with incorrect values of p and q. It was pleasing to see that most candidates realised that a list had to be ordered, before finding the middle value.
- b.i. A large number of candidates gained full marks on this question. Many correct variations of the equation were given and the values of p, q and the median could then be found. Some candidates neglected the extra information of p less than q and lost a mark for having these values the wrong way around. Follow through marks could be awarded for the median, if working was shown, with incorrect values of p and q. It was pleasing to see that most candidates realised that a list had to be ordered, before finding the middle value.
- b.ii A large number of candidates gained full marks on this question. Many correct variations of the equation were given and the values of p, q and the median could then be found. Some candidates neglected the extra information of p less than q and lost a mark for having these values the wrong way around. Follow through marks could be awarded for the median, if working was shown, with incorrect values of p and q. It was pleasing to see that most candidates realised that a list had to be ordered, before finding the middle value.
- c. A large number of candidates gained full marks on this question. Many correct variations of the equation were given and the values of p, q and the median could then be found. Some candidates neglected the extra information of p less than q and lost a mark for having these values the wrong way around. Follow through marks could be awarded for the median, if working was shown, with incorrect values of p and q. It was pleasing to see that most candidates realised that a list had to be ordered, before finding the middle value.

The distribution of rainfall in a town over 80 days is displayed on the following box-and-whisker diagram.



[1]

[1]

[2]

[2]

- a. Write down the median rainfall.
- b. Write down the minimum rainfall.
- c. Find the interquartile range.
- d. Write down the number of days the rainfall will be
 - (i) between 43 mm and 48 mm;
 - (ii) between 20 mm and 59 mm.

Markscheme

a. 43 (mm) (A1) (C1)
[1 mark]
b. 10 (mm) (A1) (C1)
[1 mark]
c. 48 - 20 (A1)

= 28 (A1) (C2)

Note: Award (A1) for identifying correct quartiles, (A1) for correct subtraction of the quartiles.

[2 marks]

d. (i) 20 (days) **(A1)**

(ii) 60 (days) (A1) (C2)

[2 marks]

Examiners report

a. [N/A]

b. [N/A]

c. [N/A]

Eight houses in a street are inhabited by different numbers of people, as shown in the table below.

House	А	В	С	D	Е	F	G	Η
Number of inhabitants	5	4	7	6	4	3	6	4

The following statements refer to the number of inhabitants per house. Write down true (T) or false (F) for each.

a, i.The mean is 5.	[1]
a, iiThe range is 4.	[1]
a, ii ī. he mode is 6.	[1]
a, iThe standard deviation is 1.4 correct to 2 significant figures.	[1]
b. Calculate the interquartile range for the number of inhabitants per house.	[2]

Markscheme

a, i.F	(A1)
[1	mark]
a, iiT	(A1)
[1	mark]
a, iiF.	(A1)
[1	mark]
a, i ⊮ .	(A1) (C4)
[1	mark]
b. IC	QR = 6 – 4 <i>(M1)</i>

Note: Award (M1) for identifying correct quartiles.

IQR = 2 (A1)(ft) (C2)

[2 marks]

Examiners report

a, iIt was expected that candidates would use the GDC for this question.

- a, iiIt was expected that candidates would use the GDC for this question.
- a, iiIt was expected that candidates would use the GDC for this question.
- a, iVt was expected that candidates would use the GDC for this question.

It was pleasing to note that most candidates were able to choose the appropriate standard deviation.

b. It was expected that candidates would use the GDC for this question.

A common error was writing the IQR as an interval.

The cumulative frequency graph represents the speed, s, in km h^{-1} , of 80 cars passing a speed camera.



a. Write down the number of cars passing the camera with speed of less than or equal to 50 ${\rm km}\,{
m h}^{-1}.$

$s(\mathbf{km} \mathbf{h}^{-1})$	0 < <i>s</i> ≤ 50	50 < <i>s</i> ≤ 70	70 < <i>s</i> ≤ 80	80 < <i>s</i> ≤ 90	90 < <i>s</i> ≤100
Frequency			34		4

- c. Write down the mid-interval value of the $50 < s \leqslant 70$ interval.
- d. Use your graphic display calculator to find an estimate of
 - (i) the mean speed of the cars passing the camera;
 - (ii) the standard deviation of the speed of the cars passing the camera.

Markscheme

a. 10 (A1) (C1)

[1 mark]

b.	$s(\text{km}\text{h}^{-1})$	0 < <i>s</i> ≤ 50	$50 < s \le 70$	$70 < s \le 80$	$80 < s \le 90$	$90 < s \le 100$	<i>(Δ1)(</i> ft)	(C1)
	Frequency	10	20	34	12	4	(~ /)(1)	(01)

Note: Follow through from their answer to part (a).

[1 mark]

c. 60 (A1) (C1)

[1 mark]

d. (i) $67.5 \, \left({\rm km \, h}^{-1} \right)$ (A2)(ft)

Notes: Award (M1) for an attempt to use the formula for the mean with at least two midpoint values consistent with their answer to part (c). Follow through from their table in part (b).

(ii) 18.6(18.6413...) (A1)(ft) (C3)

Note: Follow through from their table in part (b).

[3 marks]

Examiners report

- a. ^[N/A]
- b. ^[N/A]
- c. [N/A]
- d. [N/A]

[1]

[3]

The following six integers are arranged from smallest to largest

1, x, 3, y, 14, z

The mode is 1, the median is 5 and the mean is 7.

a.	F	ir	۱d	х	;	

b. Find *y* ;

c. Find z.

Markscheme

a. x = 1 (A1) (C1)

[1 mark]

b. $rac{3+y}{2}=5$ (M1)

Note: Award (M1) for setting the correct equation or equivalent.

y = 7 **(A1) (C2)**

[2 marks]

c. $1 + 1 + 3 + 7 + 14 + z = 6 \times 7$ (M1)(M1)

Note: Award (M1) for the sum of their 5 numbers and z, and (M1) for 6×7 .

z = 16 (A1)(ft) (C3)

Note: Follow through from their x and y found in part (b) provided y is a positive integer less than 14.

[3 marks]

Examiners report

a. In part (a), the majority of candidates were able to identify the correct value for *x*. In part (b), many candidates seemed to think that the median was the same as the value of *y*. Consequently the value 5 proved to be a popular, but incorrect, answer. The majority of candidates wrote down the correct value of 7 and went on to give the required answer of 16 in part (c). A cautionary note here though: A common set of responses to this question was 1, 5 and 18. Without working this earned 1 mark, with working (in part (c)) this earned 4 marks.

[1]

[2]

[3]

b. In part (a), the majority of candidates were able to identify the correct value for *x*. In part (b), many candidates seemed to think that the median was the same as the value of *y*. Consequently the value 5 proved to be a popular, but incorrect, answer. The majority of candidates wrote down the correct value of 7 and went on to give the required answer of 16 in part (c). A cautionary note here though: A common set of responses to this question was 1, 5 and 18. Without working this earned 1 mark, with working (in part (c)) this earned 4 marks.

c. In part (a), the majority of candidates were able to identify the correct value for *x*. In part (b), many candidates seemed to think that the median was the same as the value of *y*. Consequently the value 5 proved to be a popular, but incorrect, answer. The majority of candidates wrote down the correct value of 7 and went on to give the required answer of 16 in part (c). A cautionary note here though: A common set of responses to this question was 1, 5 and 18. Without working this earned 1 mark, with working (in part (c)) this earned 4 marks.

For a study, a researcher collected 200 leaves from oak trees. After measuring the lengths of the leaves, in cm, she produced the following cumulative frequency graph.



The researcher finds that 10% of the leaves have a length greater than $k \, {\rm cm}$.

a. Write down the median length of these leaves.	[1]
b. Write down the number of leaves with a length less than or equal to 8 cm.	[1]
c.i. Use the graph to find the value of k .	[2]
c.ii.Before measuring, the researcher estimated k to be approximately 9.5 cm. Find the percentage error in her estimate.	[2]

Markscheme

a. 9 (cm) (A1) (C1)

[1 mark]

b. 40 (leaves) (A1) (C1)

[1 mark]

c.i. $(200 \times 0.90 =)$ 180 or equivalent (M1)

Note: Award *(M1)* for a horizontal line drawn through the cumulative frequency value of 180 and meeting the curve (or the corresponding vertical line from 10.5 cm).

(k =) 10.5 (cm) (A1) (C2)

Note: Accept an error of ±0.1.

[2 marks]

c.ii. $\left|rac{9.5-10.5}{10.5}
ight| imes 100\%$ (M1)

Notes: Award (M1) for their correct substitution into the percentage error formula.

```
9.52~(\%)~(9.52380\ldots~(\%)) (A1)(ft) (C2)
```

```
Notes: Follow through from their answer to part (c)(i).
```

Award (A1)(A0) for an answer of -9.52 with or without working.

[2 marks]

Examiners report

a. [N/A] b. [N/A] c.i. [N/A] c.ii.[N/A]

The following table shows the number of errors per page in a 100 page document.

Number of errors	0	1	2	3	4
Number of pages	28	24	20	17	11

a. State whether the data is discrete, continuous or neither.

- b. Find the mean number of errors per page.
- c. Find the median number of errors per page.
- d. Write down the mode.

[1]

[2]

[2]

[1]

Markscheme

a. Discrete (A1) (C1)

[1 mark]

b. $\frac{0+24+40+51+44}{100} = \frac{159}{100} = 1.59$ (M1)(A1) (C2)

Notes: Award *(M1)* for correctly substituted formula. Award *(M1)(A1)* for 1 or 2 if 1.59 is seen. Award *(M0)(A0)* for 1 or 2 seen with no working.

[2 marks]

c. 1 (M1)(A1) (C2)

Note: Award (*M1*) for attempt to order raw data (if frequency table not used) or (*M1*) for indicating halfway between 50th and 51st result or (*M1*) for 50th percentile seen.

[2 marks]

d. 0 **(A1) (C1)**

[1 mark]

Examiners report

- a. This question was not well answered by a number of candidates. In this part, many did not know that the data was discrete.
- b. This question was not well answered by a number of candidates. The most common error in this part was (0 + 1 + 2 + 3 + 4) / 5 = 2.
- c. This question was not well answered by a number of candidates.
- d. This question was not well answered by a number of candidates.

Many candidates did not know how to find the median and the most common incorrect answer for this part was 11.

A survey was conducted of the number of bedrooms in 208 randomly chosen houses. The results are shown in the following table.

Number of bedrooms	1	2	3	4	5	6
Number of houses	41	60	52	32	15	8

- a. State whether the data is discrete or continuous.
- b. Write down the mean number of bedrooms per house.
- c. Write down the standard deviation of the number of bedrooms per house.

[1]

[2]

[1]

d. Find how many houses have a number of bedrooms greater than one standard deviation above the mean.

Markscheme

a. Discrete (A1) (C1)

[1 mark]

b. For attempting to find $\sum fx / \sum f$ (M1)

2.73 (A1) (C2)

Note: for (b) and (c), if both mean and standard deviation given to 2 significant figures. Award **(C1)(C0)(AP)** for 2.7. Award **(A1)(ft)** for 1.3 (**(AP)** already deducted).

[2 marks]

```
c. 1.34 (A1) (C1)
```

Note: for (b) and (c), if both mean and standard deviation given to 2 significant figures. Award **(C1)(C0)(AP)** for 2.7. Award **(A1)(ft)** for 1.3 (**(AP)** already deducted).

[1 mark]

d. Attempt to find their mean + their standard deviation (can be implied) (M1)

23, (ft) their mean and standard deviation. (A1)(ft) (C2)

[2 marks]

Examiners report

- a. Many candidates did not use the calculator correctly to find the mean and standard deviation, trying to do all the calculations by hand with a significant number not realizing it was a frequency table this could have been the error with those using the calculator also as few received full marks on this question. When the candidate had made an error follow through marks could be awarded for the final part *provided working had been shown*. Most knew the data was discrete.
- b. Many candidates did not use the calculator correctly to find the mean and standard deviation, trying to do all the calculations by hand with a significant number not realizing it was a frequency table this could have been the error with those using the calculator also as few received full marks on this question. When the candidate had made an error follow through marks could be awarded for the final part *provided working had been shown*. Most knew the data was discrete.
- c. Many candidates did not use the calculator correctly to find the mean and standard deviation, trying to do all the calculations by hand with a significant number not realizing it was a frequency table this could have been the error with those using the calculator also as few received full marks on this question. When the candidate had made an error follow through marks could be awarded for the final part *provided working had been shown*. Most knew the data was discrete.
- d. Many candidates did not use the calculator correctly to find the mean and standard deviation, trying to do all the calculations by hand with a significant number not realizing it was a frequency table this could have been the error with those using the calculator also as few received full marks on this question. When the candidate had made an error follow through marks could be awarded for the final part *provided working*

The weights, in kg, of 60 adolescent females were collected and are summarized in the box and whisker diagram shown below.



a.	Write down the median weight of the females.	[1]
b.	. Calculate the range.	[2]
c.	Estimate the probability that the weight of a randomly chosen female is more than 50 kg.	[1]

d. Use the box and whisker diagram to determine if the mean weight of the females is less than the median weight. Give a reason for your answer. [2]

Markscheme

a. 42 kg (A1) (C1)

Note: The units are required.

b. 58 - 33 (A1)

Note: Award (A1) for correct maximum and minimum seen.

= 25 (A1) (C2)

- c. $\frac{1}{4}(0.25, 25\%)$ (A1) (C1)
- d. Mean weight is more than the median weight. (A1)

The upper half of the distribution is wider (more dispersed) or data is positively (or right) skewed or equivalent reason. (R1) OR

 $\left(ext{The mean is calculated } ar{x} = rac{35.5 imes 15 + 40 imes 15 + 54 imes 15}{60}
ight)$ $ar{x} = 43.875 \ (kg) \quad ext{(R1)} \quad ext{(C2)}$

Note: Do not award (A1)(R0).

Examiners report

- a. Many candidates omitted the "kg" units that were required for the median weight. It is not only area and volume answers where marks may be lost for either missing or incorrect units. Candidates confused IQR with range. Only the very strongest candidates were able to deduce from a box and whisker plot that the data was asymmetric (with a positive skew) hence the mean was greater than the median. This was one of two reasoning marks in the paper and only the very strongest candidates wrote down a correct reason.
- b. Many candidates omitted the "kg" units that were required for the median weight. It is not only area and volume answers where marks may be lost for either missing or incorrect units. Candidates confused IQR with range. Only the very strongest candidates were able to deduce from a box and whisker plot that the data was asymmetric (with a positive skew) hence the mean was greater than the median. This was one of two reasoning marks in the paper and only the very strongest candidates wrote down a correct reason.
- c. Many candidates omitted the "kg" units that were required for the median weight. It is not only area and volume answers where marks may be lost for either missing or incorrect units. Candidates confused IQR with range. Only the very strongest candidates were able to deduce from a box and whisker plot that the data was asymmetric (with a positive skew) hence the mean was greater than the median. This was one of two reasoning marks in the paper and only the very strongest candidates wrote down a correct reason.
- d. Many candidates omitted the "kg" units that were required for the median weight. It is not only area and volume answers where marks may be lost for either missing or incorrect units. Candidates confused IQR with range. Only the very strongest candidates were able to deduce from a box and whisker plot that the data was asymmetric (with a positive skew) hence the mean was greater than the median. This was one of two reasoning marks in the paper and only the very strongest candidates wrote down a correct reason.

6 8 12 14 22 10 16 18 20 Daily rainfall (mm)

The daily rainfall for the town of St. Anna is collected over a 20-day period of time. The collected data are represented in the box and whisker plot below.

a. Write down

(i) the lowest daily rainfall;

(ii) the highest daily rainfall.

[2]

b. State what the value of 12 mm represents on the given diagram.

- c. Find the interquartile range.
- d. Write down the percentage of the data which is less than the upper quartile.

Markscheme

a. (i) 6 (mm) (A1)

(ii) 20 (mm) (A1) (C2)

[2 marks]

b. Median (A1) (C1)

Note: Award (A1) for Q₂ or 50th percentile.

[1 mark]

c. 14 – 9 **(A1)**

Note: Award (A1) for 9 and 14 seen.

5 (mm) **(A1) (C2)**

[2 marks]

d. 75 (%) **(A1) (C1)**

[1 mark]

Examiners report

- a. Parts (a) and (b) proved to be very well done with many correct answers seen. On a few scripts however, candidates who seemed unsure of the correct average, wrote down, average, mean or even medium.
- b. Parts (a) and (b) proved to be very well done with many correct answers seen. On a few scripts however, candidates who seemed unsure of the correct average, wrote down, average, mean or even medium.
- c. Part (c) was generally well done with many candidates correctly identifying Q1 and Q3 and many correct answers of 5 were seen.
- d. 75% proved to be an elusive answer on many scripts for part (d) as a significant number of candidates did not seem to understand the meaning of quartiles. Indeed, a popular, but erroneous answer seen was 57.1% which was arrived at from the calculation $\frac{14-6}{14} \times 100$.

240 cars were tested to see how far they travelled on 10 litres of fuel. The graph shows the cumulative frequency distribution of the results.

[1]



a. Find the median distance travelled by the cars.

b. Calculate the interquartile range of the distance travelled by the cars.

c. Find the number of cars that travelled more than $130\ \rm km.$

Markscheme

a. $Q_2=119~{
m km}$ (M1)(A1) (C2)

Note: Award (M1) for indication on graph of correct position of median.

[2 marks]

b. $Q_1=114 \mbox{ and } Q_3=123$ $\mbox{ (A1)}$

Note: Award (A1) for correct quartiles seen.

9 **(A1) (C2)**

[2 marks]

c. 240 - 220 (M1)

[2]

[2]

= 20 (A1) (C2)

[2 marks]

Examiners report

- a. Many candidates gained full marks on this question although a significant number could not find the interquartile range.
- b. Many candidates gained full marks on this question although a significant number could not find the interquartile range.
- c. Many candidates gained full marks on this question although a significant number could not find the interquartile range.

80 matches were played in a football tournament. The following table shows the number of goals scored in all matches.

Number of goals	0	1	2	3	4	5
Number of matches	16	22	19	17	1	5

a.	Find the mean number of goals scored per match.	[2]
b.	Find the median number of goals scored per match.	[2]
c.	A local newspaper claims that the mean number of goals scored per match is two. Calculate the percentage error in the local newspaper's	[2]

c. A local newspaper claims that the mean number of goals scored per match is two. Calculate the percentage error in the local newspaper claim.

Markscheme

a. $\frac{0 \times 16 + 1 \times 22 + 2 \times 19...}{80}$ (M1)

Note: Award (M1) for substituting correct values into mean formula.

1.75 (A1) (C2)

[2 marks]

b. An attempt to enumerate the number of goals scored. (M1)

2 (A1) (C2)

[2 marks]

c. $rac{2-1.75}{1.75} imes 100$ (M1)

14.3% (A1)(ft) (C2)

Notes: Award (*M1*) for correctly substituted % error formula. % sign not required. Follow through from their answer to part (a). If 100 is missing and answer incorrect award (*M0*)(*A0*). If 100 is missing and answer incorrectly rounded award (*M1*)(*A1*)(ft)(*AP*).

[2 marks]

Examiners report

- a. In parts (a) and (b), 2.5 was a common incorrect error for both parts as some candidates were confused as to the concept of both the mean and the median from tabular data and simply looked at the mean and median of the *Number of goals*, ignoring the weighting of the number of matches. Candidates faired a little better with part (c) and many correct answers (many as follow through answers) were seen in this part of the question.
- b. In parts (a) and (b), 2.5 was a common incorrect error for both parts as some candidates were confused as to the concept of both the mean and the median from tabular data and simply looked at the mean and median of the *Number of goals*, ignoring the weighting of the number of matches. Candidates faired a little better with part (c) and many correct answers (many as follow through answers) were seen in this part of the question.
- c. In parts (a) and (b), 2.5 was a common incorrect error for both parts as some candidates were confused as to the concept of both the mean and the median from tabular data and simply looked at the mean and median of the *Number of goals*, ignoring the weighting of the number of matches. Candidates faired a little better with part (c) and many correct answers (many as follow through answers) were seen in this part of the question.

Five pipes labelled, "6 metres in length", were delivered to a building site. The contractor measured each pipe to check its length (in metres) and recorded the following;

[3]

[3]

5.96, 5.95, 6.02, 5.95, 5.99.

a. (i) Find the mean of the contractor's measurements.

(ii) Calculate the percentage error between the mean and the stated, approximate length of 6 metres.

b. Calculate $\sqrt{3.87^5 - 8.73^{-0.5}}$, giving your answer

(i) correct to the nearest integer,

(ii) in the form $a imes 10^k$, where 1 \leq a < 10, $k\in\mathbb{Z}$.

Markscheme

a. (i) Mean = (5.96 + 5.95 + 6.02 + 5.95 + 5.99) / 5 = 5.974 (5.97) (A1)

(ii) $\% \text{ error} = \frac{error}{actualvalue} \times 100\%$ = $\frac{6-5.974}{5.974} \times 100\% = 0.435\%$ (M1)(A1)(ft)

(M1) for correctly substituted formula.

Allow 0.503% as follow through from 5.97

Note: An answer of 0.433% is incorrect. (C3)

[3 marks]

```
b. number is 29.45728613
(i) Nearest integer = 29 (A1)
(ii) Standard form = 2.95 × 10<sup>1</sup> (accept 2.9 × 10<sup>1</sup>) (A1)(ft)(A1)
Award (A1) for each correct term
Award (A1)(A0) for 2.95 × 10 (C3)
```

[3 marks]

Examiners report

- a. a) Almost all candidates calculated the mean correctly but less than half were able to find the % error, many dividing by 6. This was despite the boldening of 'approximate' in the question.
- b) Main errors were giving the answer correct to 1 significant figure (30) or 1 decimal place. Some candidates just counted the number of figures on the calculator to determine the index for the standard form, giving 10⁹ instead of 10¹.

Each month the number of days of rain in Cardiff is recorded.

The following data was collected over a period of 10 months.

11 13 8 11 8 7 8 14 *x* 15

For these data the median number of days of rain per month is 10.

a. Find the value of x.	[2]
b.i.Find the standard deviation	[2]
b.iiFind the interquartile range.	[2]

Markscheme

a. $\frac{x+11}{2} = 10$ (M1)

Note: Award (M1) for correct substitution into median formula or for arranging all 9 values into ascending/descending order.

(x =) 9 (A1) (C2)

[2 marks]

b.i.2.69 (2.69072...) (A2)(ft)

Note: Follow through from part (a).

[2 marks]

b.ii.13 - 8 (M1)

Note: Award (M1) for 13 and 8 seen.

= 5 (A1)(ft) (C4) Note: Follow through from part (a).

[2 marks]

Examiners report

a. ^[N/A] b.i.^[N/A] b.ii^[N/A]

The lengths of trout in a fisherman's catch were recorded over one month, and are represented in the following histogram.



a. Complete the following table.

Length of trout	Frequency
$20\mathrm{cm} < \mathrm{trout}\ \mathrm{length} \leq 30\mathrm{cm}$	0
$30\mathrm{cm}$ < trout length \leq $40\mathrm{cm}$	
$40\mathrm{cm}$ < trout length \leq 50 cm	
$50\mathrm{cm} < \text{trout length} \le 60\mathrm{cm}$	
$60\mathrm{cm} < \text{trout length} \leq 70\mathrm{cm}$	
$70\mathrm{cm} < \mathrm{trout}\ \mathrm{length} \le 80\mathrm{cm}$	1

b. State whether length of trout is a continuous or discrete variable.

- c. Write down the modal class.
- d. Any trout with length 40 cm or less is returned to the lake.

[2]

[1]

[1]

[2]

Calculate the percentage of the fisherman's catch that is returned to the lake.

Markscheme

a.

Length of trout	Frequency		
$20 \text{ cm} < \text{trout length} \le 30 \text{ cm}$	0		
$30 \text{ cm} < \text{trout length} \le 40 \text{ cm}$	4		
$40 \text{ cm} < \text{trout length} \le 50 \text{ cm}$	2	(A2)	(C2)
$50 \mathrm{cm} < \mathrm{trout} \ \mathrm{length} \le 60 \mathrm{cm}$	7		
$60 \text{ cm} < \text{trout length} \le 70 \text{ cm}$	8		
$70 \text{ cm} < \text{trout length} \le 80 \text{ cm}$	1		

Note: Award (A2) for all correct entries, (A1) for 3 correct entries.

[2 marks]

b. continuous (A1) (C1)

[1 mark]

c. $60~(\mathrm{cm}) < \mathrm{trout}~\mathrm{length} \leqslant 70~(\mathrm{cm})$ (A1) (C1)

Note: Accept equivalent notation such as (60, 70] or]60, 70]. Award **(A0)** for "60-70" (incorrect notation).

[1 mark]

d. $rac{4}{22} imes 100$ (M1)

Note: Award (M1) for their 4 divided by their 22.

 $= 18.2 \ (18.1818\ldots)$ (A1)(ft) (C2)

Note: Follow through from their part (a). Do not accept 0.181818....

[2 marks]

Examiners report

- a. [N/A]
- b. [N/A]
- c. [N/A]
- d. ^[N/A]





a.	Write down the median weight of the males.	[1]
b.	Calculate the interquartile range.	[2]
c.	Estimate the number of males who weigh between 61 kg and 66 kg .	[1]
d.	Estimate the mean weight of the lightest 40 males.	[2]

Markscheme

a. 61 kg (A1) (C1)

[1 mark]

```
b. 66 - 52 (A1)
```

= 14 (A1)(ft) (C2)

Note: Award (A1) for identifying quartiles, (A1)(ft) for correct subtraction of their quartiles.

[2 marks]

c. 20 (A1) (C1)

[1 marks]

d. $\frac{49.5 \times 20 + 56.5 \times 20}{40}$ (M1)

Note: Award (M1) for multiplication of midpoints by frequencies.

 $= 53 \ {
m kg}$ (A1) (C2)

[2 marks]

Examiners report

- a. Question 14 parts a and b were well done.
- b. Question 14 parts a and b were well done.

- c. Parts c and d were omitted or incorrectly answered more frequently than any other question on the exam paper.
- d. Parts c and d were omitted or incorrectly answered more frequently than any other question on the exam paper.

The IB grades attained by a group of students are listed as follows.

 $6 \ 4 \ 5 \ 3 \ 7 \ 3 \ 5 \ 4 \ 2 \ 5$

[2]

[2]

[2]

- a. Find the median grade.b. Calculate the interquartile range.
- c. Find the probability that a student chosen at random from the group scored at least a grade 4.

Markscheme

a. 2 3 3 4 4 5 5 5 6 7 (M1)

Note: Award (M1) for correct ordered set.

(Median =) 4.5 (A1) (C2)

b. 5 – 3 (M1)

Note: Award (M1) for correct quartiles seen.

```
= 2 (A1) (C2)
```

c. $\frac{7}{10}$ (0.7, 70%) (A2) (C2)

Examiners report

- a. Part (a) was generally well done although some candidates seemed to be confused between the mean and median.
- b. In part (b) it was not unusual to see an upper quartile of 5.5 (resulting from (5+6)/2).
- c. A significant number of candidates had difficulty with "at least four" in part (c), answering 2/10 which resulted from calculating the probability of a grade equal to 4 and not at least 4.

The weights of 90 students in a school were recorded. The information is displayed in the following table.

Weight (kg)	Number of students
$40 \le w < 50$	7
$50 \le w < 60$	28
$60 \le w < 70$	35
$70 \le w < 80$	20

[1]

[2]

[1]

[2]

a. Write down the mid interval value for the interval $50\leqslant w\leqslant 60.$

b.i. Use your graphic display calculator to find an estimate for the mean weight.

b.iiUse your graphic display calculator to find an estimate for the standard deviation.

c. Find the weight that is 3 standard deviations below the mean.

Markscheme

a. 55 (A1) (C1)

[1 mark]

b.i.62.5 (62.6) (A2)(ft) (C2)

[2 marks]

b.ii8.86 (A1) (C1)

Note: Follow through from their answer to part (a).

[1 mark]

c. 62.6 - 3 × 8.86 = 36.0 (M1)(A1)(ft) (C2)

Note: Accept 36.

Follow through from their values in part (b) only if working is seen.

[2 marks]

Examiners report

- a. This question was not well answered. Many candidates could not find the mid interval value, and used their graphic display calculator incorrectly to find the mean and standard deviation.
- b.i. This question was not well answered. Many candidates could not find the mid interval value, and used their graphic display calculator incorrectly to find the mean and standard deviation.

- b.ii. This question was not well answered. Many candidates could not find the mid interval value, and used their graphic display calculator incorrectly to find the mean and standard deviation.
- c. This question was not well answered. Many candidates could not find the mid interval value, and used their graphic display calculator incorrectly to find the mean and standard deviation. The candidates who showed working and correct method in part c) were awarded the final two marks. If working was not shown then these marks could not be awarded.

The table below shows the frequency distribution of the number of dental fillings for a group of 25 children.

Number of fillings	0	1	2	3	4	5
Frequency	4	3	8	\overline{q}	4	1

- a. Find the value of q.
- b. Use your graphic display calculator to find
 - (i) the mean number of fillings;
 - (ii) the median number of fillings;
 - (iii) the standard deviation of the number of fillings.

Markscheme

a. q = 25 - (4 + 3 + 8 + 4 + 1) (M1)

Note: Award (M1) for subtraction from $25 \mbox{ of all values from the table.}$

= 5 (A1) (C2)

[2 marks]

b. (i) 2.2 (A2)(ft) (C2)

Note: Award (M1) for use of mean formula with correct substitution. Follow through from part (a), irrespective of whether working is shown.

(ii) 2 (A1) (C1)

(iii) 1.39 (A1)(ft) (C1)

Note: Follow through from part (a), irrespective of whether working is shown. Award (A1) for 1.38.

[4 marks]

Examiners report

[2] [4]

- a. Part b was not well answered by the majority of candidates, indicating that the use of the GDC is not a natural tool for answering this type of question. Many students ignored the frequencies when finding the mean, median, and standard deviation.
- b. Part b was not well answered by the majority of candidates, indicating that the use of the GDC is not a natural tool for answering this type of question. Many students ignored the frequencies when finding the mean, median, and standard deviation.

A group of 20 students travelled to a gymnastics tournament together. Their ages, in years, are given in the following table.

Age (years)	14	15	16	17	18	19	20	22
Frequency	1	2	7	1	4	1	1	3

The lower quartile of the ages is 16 and the upper quartile is 18.5.

a.i. For the students in this group find the mean age;	[2]
a.ii.For the students in this group write down the median age.	[1]
b. Draw a box-and-whisker diagram, for these students' ages, on the following grid.	[3]

b. Draw a box-and-whisker diagram, for these students' ages, on the following grid.



Markscheme

a.i. $\frac{14+2\times15+7\times16+17+4\times18+19+20+3\times22}{2}$ (M1) 20

Award (M1) for correct substitutions into mean formula. Note:

(=) 17.5(C2) (A1)

[2 marks]

a.ii.16.5 (A1) (C1)

[1 mark]



Note: Award (A1) for correct endpoints, (A1) for correct quartiles, (A1)(ft) for their median. Follow through from part (a)(ii), but only if median is between 16 and 18.5. If a horizontal line goes through the box, award at most (A1)(A1)(A0). Award at most (A0)(A1)(A1) if a ruler has not been used.

[3 marks]

Examiners report

a.i. ^[N/A] a.ii.^[N/A] b. ^[N/A]

A tetrahedral (four-sided) die has written on it the numbers 1, 2, 3 and 4. The die is rolled many times and the scores are noted. The table below shows

the resulting frequency distribution.

Score	1	2	3	4
Frequency	18	x	У	22

[1]

[2]

[3]

The die was rolled a total of 100 times.

The mean score is 2.71.

a. Write down an equation, in terms of x and y, for the total number of times the die was rolled.

b. Using the mean score, write down a second equation in terms of x and y.

c. Find the value of x and of y.

Markscheme

a. 18 + x + y + 22 = 100 or equivalent (A1) (C1)

[1 mark]

b. $\frac{18+2x+3y+88}{100}=2.71$ or equivalent *(M1)(A1)* (C2)

Note: Award (M1) for a sum including x and y, divided by 100 and equated to 2.71, (A1) for a correct equation.

[2 marks]

c. x + y = 60 and 2x + 3y = 165 (M1)

Note: Award *(M1)* for obtaining a correct linear equation in one variable from their (a) and their (b). This may be implied if seen in part (a) or part (b).

 $x=15;\;y=45$ (A1)(ft)(A1)(ft) (C3)

Notes: Follow through from parts (a) and (b), irrespective of working seen provided the answers are positive integers.

[3 marks]

Examiners report

a. ^[N/A] b. ^[N/A]

с. ^[N/A]

56 students were given a test out of 40 marks. The teacher used the following box and whisker plot to represent the marks of the students.



[1]

[1]

[2]

[2]

a.i. Write down the median mark.

a.ii.Write down the 75th percentile mark.

a.iiiWrite down the range of marks.

b. Estimate the number of students who achieved a mark greater than 32.

Markscheme

a.i.30 (A1) (C1)

[1 mark]

a.ii.32 (A1) (C1)

[1 mark]

a.iii38 - 10 = 28 (A1)(A1) (C2)

Note: Award (A1) for 10 and 38 seen, (A1) for correct answer only.

[2 marks]

b. 0.25 × 56 = 14 (M1)(A1) (C2)

Note: Award (M1) for multiplying 0.25 by 56.

[2 marks]

Examiners report

- a.i. Many students had difficulty with reading the box and whisker plot and interpreting this question. Some candidates had difficulty with finding the range in part (a)(iii). Many wrote down the end points for the required range of data instead of writing the difference between the largest and smallest values. A number of candidates had problems estimating the number of students who achieved a mark greater than 32. Many students used the number 40 instead of the total number of student 56 for the estimation in part b).
- a.ii.Many students had difficulty with reading the box and whisker plot and interpreting this question. Some candidates had difficulty with finding the range in part (a)(iii). Many wrote down the end points for the required range of data instead of writing the difference between the largest and smallest values. A number of candidates had problems estimating the number of students who achieved a mark greater than 32. Many students used the number 40 instead of the total number of student 56 for the estimation in part b).
- a.iiiMany students had difficulty with reading the box and whisker plot and interpreting this question. Some candidates had difficulty with finding the range in part (a)(iii). Many wrote down the end points for the required range of data instead of writing the difference between the largest and smallest values. A number of candidates had problems estimating the number of students who achieved a mark greater than 32. Many students used the number 40 instead of the total number of student 56 for the estimation in part b).
- b. Many students had difficulty with reading the box and whisker plot and interpreting this question. Some candidates had difficulty with finding the range in part (a)(iii). Many wrote down the end points for the required range of data instead of writing the difference between the largest and smallest values. A number of candidates had problems estimating the number of students who achieved a mark greater than 32. Many students used the number 40 instead of the total number of student 56 for the estimation in part b).

A class of 13 Mathematics students received the following grades in their final IB examination.

3 5 3 4 7 3 2 7 5 6 5 3 4

a.	For these grades, find the mode;	[1]
b.	For these grades, find the median;	[2]
c.	For these grades, find the upper quartile;	[1]
d.	For these grades, find the interquartile range.	[2]

Markscheme

```
a. 3 (A1) (C1)
```

[1 mark]

```
b. 4 (M1)(A1) (C2)
```

Note: Award (M1) for ordered list of numbers seen.

[2 marks]

c. 5.5 (A1) (C1)

```
[1 mark]
```

```
d. 5.5 – 3 (M1)
```

Note: Award (M1) for 3 and their 5.5 seen.

= 2.5 (A1)(ft) (C2)

Note: Follow through from their answer to part (c).

[2 marks]

Examiners report

a. ^[N/A]

b. [N/A]

c. [N/A]

d. ^[N/A]

(i) Speeds of cars travelling along a road.

a. State which of the following sets of data are discrete.

- (ii) Numbers of members in families.
- (iii) Maximum daily temperatures.
- (iv) Heights of people in a class measured to the nearest cm.
- (v) Daily intake of protein by members of a sporting team.
- b. The boxplot below shows the statistics for a set of data.



(iii) the minimum value present

c. Write down three different integers whose mean is 10.

Markscheme

a. (ii) and (iv) are discrete. (A1)(A1)

Award (A1)(A0) for both correct and one incorrect.

Award (A1)(A0) for one correct and two incorrect.

Otherwise, (A0)(A0). (C2)

[2 marks]

b. (i) Median = 10 (A1)

(ii) Q₃ = 12 *(A1)*

(iii) Min value = 1 (±0.2) (A1) (C3)

[3 marks]

c. Any three different integers whose mean is 10 e.g. 9, 10, 11. (A1) (C1)

[1 mark]

Examiners report

a. a) Very few candidates obtained both marks for part (a), though a majority did gain one mark.

[3]

[1]

b. b) Most candidates answered (i) and (ii) well. Quite a few could not interpret the scale for (iii).

c. c) An easy mark for candidates. It was very disappointing to see some candidates getting this wrong.

A survey was carried out on a road to determine the number of passengers in each car (excluding the driver). The table shows the results of the survey.

Number of passengers	0	1	2	3	4
Number of cars	37	23	36	15	9

[1]

[1]

[4]

a. State whether the data is discrete or continuous.

```
b. Write down the mode.
```

- c. Use your graphic display calculator to find
 - (i) the mean number of passengers per car;
 - (ii) the median number of passengers per car;
 - (iii) the standard deviation.

Markscheme

a. discrete (A1) (C1)

[1 mark]

b. 0 (A1) (C1)

[1 mark]

c. (i) 1.47 (1.46666...) (A2)

Note: Award (*M1*) for $\frac{176}{120}$ seen.

Accept 1 or 2 as a final answer if 1.4666... or 1.47 seen.

(ii) 1.5 **(A1)**

(iii) 1.25 (1.25122...) (A1) (C4)

[4 marks]

Examiners report

- a. Most candidates could state whether the data was discrete or continuous and find the mode however the calculations to find the mean, median and standard deviation appeared problematic for some candidates. A significant number of candidates gave the mode as 37 rather than 0. Many did not appear to use their graph and some obtained the incorrect value of 1.47 from their graphic display calculator.
- b. Most candidates could state whether the data was discrete or continuous and find the mode however the calculations to find the mean, median and standard deviation appeared problematic for some candidates. A significant number of candidates gave the mode as 37 rather than 0. Many did not appear to use their graph and some obtained the incorrect value of 1.47 from their graphic display calculator.
- c. Most candidates could state whether the data was discrete or continuous and find the mode however the calculations to find the mean, median and standard deviation appeared problematic for some candidates. A significant number of candidates gave the mode as 37 rather than 0. Many did not appear to use their graph and some obtained the incorrect value of 1.47 from their graphic display calculator.

[1]

[1]

[2]

[2]

Toronto's annual snowfall, x, in cm, has been recorded for the past 176 years. The results are shown in the table.

Snowfall (cm)	$2 \le x < 6$	$6 \le x < 10$	$10 \le x < 14$	$14 \le x < 18$	$18 \le x < 22$	$22 \le x < 26$	$26 \le x < 30$
Frequency	30	26	29	32	18	27	14

- a. Write down the modal class.
- b. Write down the mid interval value for the class $6 \le x < 10$.
- c. Calculate an estimate of the mean annual snowfall.
- d. Find the number of years for which the annual snowfall was at least 18 cm.

Markscheme

a. $14 \le x < 18$ (A1) (C1)

[1 mark]

b. 8 (A1) (C1)

[1 mark]

c. $\frac{4 \times 30 + 8 \times 26 + 12 \times 29 + 16 \times 32 + 20 \times 18 + 24 \times 27 + 28 \times 14}{176}$ (M1)

Notes: Award *(M1)* for an attempt to substitute their mid-interval values (consistent with their answer to part (b)) into the formula for the mean. Award *(M1)* where a table is constructed with their (consistent) mid-interval values listed along with the frequencies.

= 14.7 (cm) (14.7045...) (A1)(ft) (C2)

Notes: Follow through from their answer to part (b). If a final incorrect answer that is consistent with their (b) is given award (M1)(A1)(ft) even if no working is seen.

[2 marks]
d. 18 + 27 + 14 (M1)

Note: Award (M1) for adding 18, 27 and 14.

= 59 (A1) (C2) [2 marks]

Examiners report

- a. Part (a) was generally well done but, in part (b), writing down the mid-interval value of a class proved difficult for some candidates and many incorrect answers of 7.5 were seen.
- b. Part (a) was generally well done but, in part (b), writing down the mid-interval value of a class proved difficult for some candidates and many incorrect answers of 7.5 were seen.
- c. Popular, but erroneous answers, seen in part (c) were 15.5 and 16. These seemed to be as a result of adding their mid-class values together and dividing by 7 rather than the total of the frequencies.
- d. There was much confusion over the meaning of the phrase 'at least' in part (d) and, as a consequence, there were as many wrong answers of 117 (30 + 26 + 29 + 32) seen as there were correct answers.

The temperatures in °C, at midday in Geneva, were measured for eight days and the results are recorded below.

7, 4, 5, 4, 8, *T*, 14, 4

The mean temperature was found to be 7 °C.

a.	Find the value of <i>T</i> .	[3]
b.	Write down the mode.	[1]
c.	Find the median.	[2]

Markscheme

a. $\frac{7+4+5+4+8+T+14+4}{8} = 7$ (A1)(A1)

Note: Award (A1) for sum + T, (A1) for 56 or 7×8 or 8 in the denominator and 7 seen.

T = 10 (A1) (C3)

[3 marks]

b. 4 (A1) (C1)

[1 mark]

c. 4, 4, 4, 5, 7, 8, 10, 14 *(M1)*

Note: Award (M1) for arranging their numbers in order.

Median = 6 (A1)(ft) (C2)

[2 marks]

Examiners report

a. This question was well answered by most candidates.

b. This question was well answered by most candidates.

c. This question was well answered by most candidates.

The number of passengers in the first ten carriages of a train is listed below.

6, 8, 6, 3, 8, 4, 8, 5, *p*, *p*

The mean number of passengers per carriage is 5.6.

a.	Calculate the value of <i>p</i> .	[2]
b.	Find the median number of passengers per carriage.	[2]
c.	If the passengers in the eleventh carriage are also included, the mean number of passengers per carriage increases to 6.0.	[2]
	Determine the number of passengers in the eleventh carriage of the train.	

Markscheme

a. $\frac{48+2p}{10}=5.6$ (M1)

Notes: Accept equivalent forms. Award (M1) for correct substitutions in mean formula.

4 (A1) (C2)

b. Correctly rearranging the list with their *p* (M1)

5.5 (A1)(ft) (C2)

Note: Follow through from their value of p in part (a).

c. $\frac{56+x}{11} = 6.0$ (M1)

Notes: Accept equivalent forms. Award (M1) for correct substitutions in mean formula.

OR

 $\frac{48+2\times \text{their part (a)}+x}{11}$ (M1) 10 (A1)(ft) (C2)

Note: Follow through from their answer to part (a).

Examiners report

a. [N/A]

b. [N/A]

c. [N/A]

The grades obtained by a group of 20 IB students are listed below:

6	2	5	3	5	5	6	2	6	1
7	6	2	4	2	4	3	4	5	6

a. Complete the following table for the grades obtained by the students.

Grade	Frequency
1	
2	
3	2
4	
5	4
6	
7	1

- b. Write down the modal grade obtained by the students.
- c. Calculate the median grade obtained by the students.
- d. One student is chosen at random from the group.

Find the probability that this student obtained either grade $4 \mbox{ or grade } 5.$

Markscheme

a.	Grade	Frequency		
	1	1		
	2	4		
	3	(2)	(42)	(C2)
	4	3		(02)
	5	(4)		
	6	5		
	7	(1)		

[2]

[1]

[2]

[1]

Notes: Award (A1) for three correct. Award (A0) for two or fewer correct.

[2 marks]

b. Mode = 6 (A1)(ft) (C1)

[1 mark]

c. Median = 4.5 (M1)(A1)(ft) (C2)

Note: (M1) for attempt to order raw data (if frequency table not used) or (M1) halfway between 10th and 11th result.

[2 marks]

d. $\frac{7}{20}$ (0.35, 35%) (A1)(ft) (C1)

[1 mark]

Examiners report

- a. Parts (a) and (b) were well done by the vast majority of candidates.
- b. Parts (a) and (b) were well done by the vast majority of candidates.
- c. Part (c) caused problems to many with (1) the mean of the two grades not being taken (2) the mean being calculated instead of the median.
- d. Part (d) was successfully completed by those candidates who did the question by counting. Those who tried to use the probability laws were not successful.

Much of the question could have been checked by inputting the data into the GDC.

The following histogram shows the weights of a number of frozen chickens in a supermarket. The weights are grouped such that $1 \leq \text{weight} < 2$, $2 \leq \text{weight} < 3$ and so on.



- b. Find the total number of chickens.
- c. Write down the modal group.
- d. Gabriel chooses a chicken at random.

Find the probability that this chicken weighs less than $4\ kg.$

Markscheme

b. 96 (A1) (C1)

[1 mark]

c. $3 \leqslant {
m weight} < 4~{
m kg}$. Accept $3-4~{
m kg}$ (A1) (C1)

[1 mark]

d. For adding three heights or subtracting $14 \mbox{ from } 96 \mbox{ (M1)}$

 $\frac{82}{96}$ (0.854 or $\frac{41}{48}$, 85.4%) (ft) from (b). (A1)(ft) (C2)

[2 marks]

Examiners report

b. Very few candidates could draw a frequency polygon correctly. The word 'Draw' means that a ruler should be used. Many managed to draw from the mid-point of the bar but did not extend it to 0.5 or 5.5. Most could answer the probability part of the question.

[1] [1]

- c. Very few candidates could draw a frequency polygon correctly. The word 'Draw' means that a ruler should be used. Many managed to draw from the mid-point of the bar but did not extend it to 0.5 or 5.5. Most could answer the probability part of the question.
- d. Very few candidates could draw a frequency polygon correctly. The word 'Draw' means that a ruler should be used. Many managed to draw from the mid-point of the bar but did not extend it to 0.5 or 5.5. Most could answer the probability part of the question.

[1]

[1]

[2]

[2]



The cumulative frequency graph shows the heights, in cm, of 80 young trees.

- a. Write down the median height of the trees.
- b. Write down the 75th percentile.
- c. Find the interquartile range.
- d. Estimate the number of trees that are more than 40 cm in height.

Markscheme

a. 54 (cm) (A1) (C1)

b. 58 (A1) (C1)

c. 58, 45 (A1)

Note: Award (A1) for 45 seen.

58 – 45

= 13 (A1)(ft) (C2)

Note: Follow through from part (b).

```
d. 80 – 14 (M1)
```

Note: Award (M1) for 14 seen.

= 66 (A1) (C2)

Examiners report

- a. The reading of values from a cumulative frequency was difficult for candidates and a notable number of candidates left this question unanswered or scored zero.
- b. The reading of values from a cumulative frequency was difficult for candidates and a notable number of candidates left this question unanswered or scored zero.
- c. The reading of values from a cumulative frequency was difficult for candidates and a notable number of candidates left this question unanswered or scored zero.
- d. The reading of values from a cumulative frequency was difficult for candidates and a notable number of candidates left this question unanswered or scored zero.

A sample of 120 oranges was tested for Vitamin C content. The cumulative frequency curve below represents the Vitamin C content, in milligrams, of these oranges.



The minimum level of Vitamin C content of an orange in the sample was 30.1 milligrams. The maximum level of Vitamin C content of an orange in the sample was 35.0 milligrams.

- a. Giving your answer to one decimal place, write down the value of
 - (i) the median level of Vitamin C content of the oranges in the sample;
 - (ii) the lower quartile;
 - (iii) the upper quartile.
- b. Draw a box-and-whisker diagram on the grid below to represent the Vitamin C content, in milligrams, for this sample.



Markscheme

- a. (i) 32.5 **(A1)**
 - (ii) 31.9 **(A1)**
 - (iii) 33.1 **(A1) (C3)**

[3]

[3]



Note: Award (A1)(ft) for correct median, (A1)(ft) for correct quartiles and box, (A1) for correct end points of whiskers and straight whiskers. Award at most (A1)(A1)(A0) if a horizontal line goes right through the box or if the whiskers are not well aligned with the midpoint of the box. Follow through from part (a).

[3 marks]

Examiners report

a. ^[N/A] b. ^[N/A]

The length, in cm, of six baseball bats was measured. The lengths are given below.

104.5, 105.1, 104.8, 105.2, 104.9, 104.9

a.	Calculate the exact value of the mean length.	[2]
b.	Write your answer to part (a) in the form $a imes 10^k$ where $1\leq a<10$ and $k\in\mathbb{Z}.$	[2]
c.	Marian calculates the mean length and finds it to be 105 cm.	[2]

Calculate the percentage error made by Marian.

Markscheme

```
a. \left(\frac{104.5+105.1+...}{6}\right) (M1)
```

Note: Award (M1) for use of mean formula.

= 104.9 (cm) (A1) (C2)

[2 marks]

b. 1.049 × 10² (A1)(ft)(A1)(ft) (C2)

Notes: Award (A1)(ft) for 1.049, (A1)(ft) for 10². Follow through from their part (a).

[2 marks]

c. $\frac{105-104.9}{104.9} imes 100$ (%) (M1)

Notes: Award (M1) for their correctly substituted % error formula.

% error = 0.0953 (%) (0.0953288...) (A1)(ft) (C2) Notes: A 2sf answer of 0.095 following $\frac{105-104.9}{105} \times 100$ working is awarded no marks. Follow through from their part (a), provided it is not 105. Do not accept a negative answer. % sign not required.

[2 marks]

Examiners report

- a. Another well answered question with candidates showing a good understanding of standard form and many correct answers were seen in parts
 (a) and (b). Whilst the formula is given for percentage error, there were still a minority of candidates who divided by 105 rather than the required value of 104.9.
- Another well answered question with candidates showing a good understanding of standard form and many correct answers were seen in parts
 (a) and (b). Whilst the formula is given for percentage error, there were still a minority of candidates who divided by 105 rather than the required value of 104.9.
- c. Another well answered question with candidates showing a good understanding of standard form and many correct answers were seen in parts (a) and (b). Whilst the formula is given for percentage error, there were still a minority of candidates who divided by 105 rather than the required value of 104.9.

The table below shows the number of words in the extended essays of an IB class.

Number of words	$3200 \le w \le 3400$	$3400 \le w \le 3600$	$3600 \le w \le 3800$	$3800 \le w \le 4000$	$4000 \le w \le 4200$
Frequency	2	5	8	17	3

a. Draw a histogram on the grid below for the data in this table.



- b. Write down the modal group.
- c. The maximum word count is $4000 \ {\rm words}.$

Write down the probability that a student chosen at random is on or over the word count.

Markscheme



Notes: (*A3*) for correct histogram, (*A2*) for one error, (*A1*) for two errors, (*A0*) for more than two errors. Award maximum (*A2*) if lines do not appear to be drawn with a ruler. Award maximum (*A2*) if a frequency polygon is drawn.

[3 marks]

b. Modal group = $3800 \le w < 4000$ (A1) (C1)

[1 mark]

c. Probability = $\frac{3}{35}$ (0.0857, 8.57%) (A1)(A1) (C2)

Note: (A1) for correct numerator (A1) for correct denominator.

[2 marks]

Examiners report

[1]

- a. A surprising number of the candidates did not appear to have brought a ruler/straight edge and so lost a mark in this question as they were asked to draw a histogram which means the lines must be drawn using a ruler/straight edge. Some candidates drew a frequency polygon. Parts (b) and (c) were generally answered well though 20/35 was seen occasionally in part (c).
- b. A surprising number of the candidates did not appear to have brought a ruler/straight edge and so lost a mark in this question as they were asked to draw a histogram which means the lines must be drawn using a ruler/straight edge. Some candidates drew a frequency polygon. Parts (b) and (c) were generally answered well though 20/35 was seen occasionally in part (c).
- c. A surprising number of the candidates did not appear to have brought a ruler/straight edge and so lost a mark in this question as they were asked to draw a histogram which means the lines must be drawn using a ruler/straight edge. Some candidates drew a frequency polygon. Parts (b) and (c) were generally answered well though 20/35 was seen occasionally in part (c).



Consider the frequency histogram for the distribution of the time, t, in minutes of telephone calls that Helen made last week.

a. Complete the frequency table for this distribution.

Time (minutes)	Number of telephone calls
$0 < t \le 5$	
$5 < t \le 10$	
$10 < t \le 15$	
$15 < t \le 20$	

- c. Write down the mid interval value of the $10 < t \leqslant 15$ class.
- d. Use your graphic display calculator to find an estimate for the mean time.

Markscheme

a.	Time (minutes)	Number of telephone calls		
	0 <i><t≤</i> 5	12		
	5 <i><t< i="">≤10</t<></i>	4	(A2)	(C2)
	10 <i><t< i="">≤15</t<></i>	6		
	15 <i><t< i="">≤20</t<></i>	8		

Note: Award (A2) for four correct entries, (A1) for three correct entries, (A0) otherwise.

- b. $0 < t \leqslant 5$ (A1) (C1)
- c. 12.5 (A1) (C1)

d.
$$\frac{275}{30}$$
 (M1)

Note: Award (M1) for division with 275 seen.

= 9.17 (9.16666...) (A1)(ft) (C2)

Note: Follow through from their parts (a) and (c), irrespective of whether working is shown.

Examiners report

- a. ^[N/A]
- b. ^[N/A]
- c. [N/A]
- d. ^[N/A]

In an international competition, participants can answer questions in **only one** of the three following languages: Portuguese, Mandarin or Hindi. 80

participants took part in the competition. The number of participants answering in Portuguese, Mandarin or Hindi is shown in the table.

		l			
		Portuguese	Mandarin	Hindi	Total
Participants	Boys	20	18	5	43
	Girls	18	7	12	37
	Total	38	25	17	80

a.	State the number of boys who answered questions in Portuguese.	[1]
b.	Find the probability that the boy answered questions in Hindi.	[2]
c.	Two girls are selected at random.	[3]

Calculate the probability that one girl answered questions in Mandarin and the other answered questions in Hindi.

Markscheme

a. 20 (A1) (C1)

[1 mark]

b. $\frac{5}{43}$ (0.11627..., 11.6279...%) (A1)(A1) (C2)

Note: Award (A1) for correct numerator, (A1) for correct denominator.

[2 marks]

c. $\frac{7}{37} \times \frac{12}{36} + \frac{12}{37} \times \frac{7}{36}$ (A1)(M1)

Note: Award (A1) for first or second correct product seen, (M1) for adding their two products or for multiplying their product by two.

```
=rac{14}{111}\;(\,0.12612\ldots,\;12.6126\,\%\,) (A1) (C3)
```

[3 marks]

Examiners report

a. ^[N/A]

- b. [N/A] c. [N/A]

a.	Write down the following numbers in increasing order.	[3]
	$3.5, 1.6 imes 10^{-19}$, 60730, 6.073 $ imes 10^5$, 0.006073 $ imes 10^6$, π , 9.8 $ imes 10^{-18}$.	
b.	Write down the median of the numbers in part (a).	[1]
c.	State which of the numbers in part (a) is irrational.	[1]

Markscheme

a. $1.6 imes 10^{-19}$, $9.8 imes 10^{-18}$, π , 3.5, $0.006073 imes 10^6$, 60730, $6.073 imes 10^5$ (A4)

Award (A1) for π before 3.5

Award (A1) for $1.6 imes 10^{-19}$ before $9.8 imes 10^{-18}$

Award (A1) for the three numbers containing 6073 in the correct order.

Award (A1) for the pair with negative indices placed before 3.5 and π and the remaining three numbers placed after (independently of the other three marks).

(C4)

```
Award (A3) for numbers given in correct decreasing order.
   Award (A2) for decreasing order with at most 1 error
   [3 marks]
b. The median is 3.5.
                       (A1)(ft)
   Follow through from candidate's list.
                                         (C1)
   [1 mark]
c. \pi is irrational. (A1) (C1)
```

[1 mark]

Examiners report

- a. This was the best-answered question on the paper with most candidates achieving 5 or 6 marks. The main errors were finding the mean instead of the median in part (b) and giving numbers with negative indices as irrational numbers for part (c). Some candidates gave the list in reverse order (which lost them one mark).
- b. This was the best-answered question on the paper with most candidates achieving 5 or 6 marks. The main errors were finding the mean instead of the median in part (b) and giving numbers with negative indices as irrational numbers for part (c). Some candidates gave the list in reverse order (which lost them one mark).
- c. This was the best-answered question on the paper with most candidates achieving 5 or 6 marks. The main errors were finding the mean instead of the median in part (b) and giving numbers with negative indices as irrational numbers for part (c). Some candidates gave the list in reverse order (which lost them one mark).

Complete the following table of values for the height and weight of seven students.

	Values	Mode	Median	Mean	Standard deviation
Height (cm)	151, 158, 171, 163, 184, 148, 171			164	11.7
Weight (kg)	53, 61, 58, 82, 45, 72, 82	82	61		

Markscheme

Mode = 171 (A1)

Median 148, 151, 158, 163, 171, 171, 184

= 163 (A1)

 $\begin{aligned} \text{Mean} &= 64.7 \quad \textbf{(A1)} \\ \text{Standard deviation} &= 13.3 \quad \textbf{(A1)} \quad \textbf{(C4)} \\ \textbf{Note:} \text{ If both mean and standard deviation given to 2 significant figures} \\ \text{Mean } 65, \textbf{(A0)(AP)} \\ \text{Standard deviation } 13 \textbf{(A1)(ft)} \textbf{((AP) already deducted)}. \end{aligned}$

[4 marks]

Examiners report

Part (a) was well answered, albeit with some choosing the wrong value (14.3) for the standard deviation. The majority of candidates had the correct stem and leaf plot, though many lost a mark through not including a key and a few used a single digit in the stem. Some tried to draw two stem and leaf plots for the heights and weights in part (a) without realizing that part (b) had the ages written above.

The distribution of the weights, correct to the nearest kilogram, of the members of a football club is shown in the following table.

Weight (kg)	40 - 49	50 - 59	60 - 69	70 - 79
Frequency	6	18	14	4

a. On the grid below draw a histogram to show the above weight distribution.



b. Write down the mid-interval value for the 40-49 interval.

c. Find an estimate of the mean weight of the members of the club.

d. Write down an estimate of the standard deviation of their weights.

Markscheme

[1]

[2]

[1]



Notes: (*A1*) for all correct heights, (*A1*) for all correct end points (39.5, 49.5 etc.). Histogram must be drawn with a ruler (straight edge) and endpoints must be clear. Award (*A1*) only if both correct histogram and correct frequency polygon drawn.

[2 marks]

b. 44.5 (A1) (C1)

Note: If (b) is given as 45 then award (b) 45 (A0) (c) 58.8 kg (M1)(A1)(ft) or (C2)(ft) if no working seen. (d) 8.44 (C1)

[1 mark]

c. Unit penalty (UP) applies in this question.

Mean = $\frac{44.5 \times 6 + 54.5 \times 18 + ...}{42}$ (M1)

Note: (M1) for a sum of frequencies multiplied by midpoint values divided by 42.

 $= 58.3 \, {
m kg}$ (A1)(ft) (C2)

Note: Award (A1)(A0)(AP) for 58.

Note: If (b) is given as 45 then award (b) 45 (A0) (c) 58.8 kg (M1)(A1)(ft) or (C2)(ft) if no working seen. (d) 8.44 (C1)

[2 marks]

d. Standard deviation = 8.44 (A1) (C1)

Note: If (b) is given as 45 then award (b) 45 (A0) (c) 58.8 kg (M1)(A1)(ft) or (C2)(ft) if no working seen. (d) 8.44 (C1)

[1 mark]

Examiners report

a. The class boundaries needed to be correctly identified to permit full credit to be given. Weight being a continuous variable and given to the nearest kg meant that the lowest class boundary was 39.5. Thereafter, the use of midpoints is standard.

(a) The endpoints of the bars caused problems for all but a very few candidates. Diagrams drawn without a ruler were also penalized.

b. The class boundaries needed to be correctly identified to permit full credit to be given. Weight being a continuous variable and given to the nearest kg meant that the lowest class boundary was 39.5. Thereafter, the use of midpoints is standard.

(b) This was well attempted by the majority; it acted as a prompt for the following parts.

c. The class boundaries needed to be correctly identified to permit full credit to be given. Weight being a continuous variable and given to the nearest kg meant that the lowest class boundary was 39.5. Thereafter, the use of midpoints is standard.

(c) (d) resulted in many incorrect answers; it was expected that the GDC would be used for these parts of the question, though a number calculated the mean by hand.

d. The class boundaries needed to be correctly identified to permit full credit to be given. Weight being a continuous variable and given to the

nearest kg meant that the lowest class boundary was 39.5. Thereafter, the use of midpoints is standard.

(c) (d) resulted in many incorrect answers; it was expected that the GDC would be used for these parts of the question, though a number calculated the mean by hand.



A random sample of 200 females measured the length of their hair in cm. The results are displayed in the cumulative frequency curve below.

- a. Write down the median length of hair in the sample.
- b. Find the interquartile range for the length of hair in the sample.
- c. Given that the shortest length was 6 cm and the longest 47 cm, draw and label a box and whisker plot for the data on the grid provided below. [3]

[1]



Markscheme

a. Unit penalty (UP) is applicable where indicated in the left hand column.

(UP) 26cm (A1) (C1) [1 mark]

b. Unit penalty (UP) is applicable where indicated in the left hand column.

33-19 for identifying correct quartiles. (A1)

(UP) $= 14 \ \mathrm{cm}$. (A1)(ft) (C2)

Note: (ft) on their quartiles.

[2 marks]



Examiners report

a. This was done well by most candidates. Some did not realise what the inter-quartile range was, and just quoted the quartiles.

b. This was done well by most candidates. Some did not realise what the inter-quartile range was, and just quoted the quartiles.

c. This was done well by most candidates. Some did not realise what the inter-quartile range was, and just quoted the quartiles.

The cumulative frequency curve shows the percentage marks, given correct to the nearest integer, gained by 500 students in an examination.



The passing grades were determined as given below.

85 to 100 %, grade A
66 to 84 %, grade B
57 to 65 %, grade C
50 to 56 %, grade D

Those scoring less than 50 % failed the examination.

Find the lowest mark required to be eligible for further study.

a.	Find the number of students who failed the examination.	[2]
b.	Find the number of students who were awarded grade C or better.	[2]
c.	The top 20 % of the students are eligible for further study.	[2]

Markscheme

a. 200 (students) (M1)(A1) (C2)

Note: Award (M1) for line drawn on the graph connecting 50 % with 200 or any indication (cross or dash) at the required point on the graph, (A1) for correct answer.

[2 marks]

b. 500 - 350 (M1)

Notes: Award (M1) for 350 seen or for a line on the graph from 57 % up to the curve showing number of students. An indication (cross or dash) at the required point on the graph is sufficient for method.

= 150 (A1) (C2)

[2 marks]

c. 60 (%) (M1)(A1) (C2)

Notes: Award (M1) for 400 or a line on the graph at 400 seen, (A1) for correct answer. % sign not required. An indication (cross or dash) at the required point on the graph is sufficient for method.

[2 marks]

Examiners report

a. Candidates who drew vertical or horizontal lines at correct positions on the graph were able to pick up the three method marks for this question and, for parts (a) and (b), a range of answers were accepted. The most common error on this question seemed to be in part (a) where the vertical line was drawn at 49% leading to a value outside the acceptable range of 190–200. Candidates are expected to read values off a continuous cumulative frequency curve at the given critical values (in this instance at 50%).

There was a mistake in the Spanish translation of question 5, which was discovered prior to marking. The principal examiner was informed and this unfortunate situation was addressed during the marking and awarding in order to ensure no candidate was disadvantaged.

b. Candidates who drew vertical or horizontal lines at correct positions on the graph were able to pick up the three method marks for this question and, for parts (a) and (b), a range of answers were accepted. In part (b), a common incorrect answer seen was 350 which was simply the number of candidates who were awarded less than a grade C rather than those with a grade C or higher. On a minority of scripts, an answer of

90 reflected the candidate's misinterpretation of the requirement of the question as 'grade C only'.

There was a mistake in the Spanish translation of question 5, which was discovered prior to marking. The principal examiner was informed and this unfortunate situation was addressed during the marking and awarding in order to ensure no candidate was disadvantaged.

c. Candidates who drew vertical or horizontal lines at correct positions on the graph were able to pick up the three method marks for this question and, for parts (a) and (b), a range of answers were accepted. In part (c), a numerical answer of 60 was required, with 'grade C' on its own losing this last mark.

There was a mistake in the Spanish translation of question 5, which was discovered prior to marking. The principal examiner was informed and this unfortunate situation was addressed during the marking and awarding in order to ensure no candidate was disadvantaged.

There are 120 teachers in a school. Their ages are represented by the cumulative frequency graph below.



a. Write down the median age.

- b. Find the interquartile range for the ages.
- c. Given that the youngest teacher is 21 years old and the oldest is 72 years old, represent the information on a box and whisker plot using the [3]

scale below.



Markscheme

a. Median = 45 (A1) (C1)

Accept 45.5

[1 mark]

b. 53 - 37 for identifying correct quartiles (A1)

=16 for correct answer to subtraction (A1)(ft) (C2)

(ft) on their quartiles

[2 marks]

[1]



Median marked correctly. (A1)(ft) Box with ends at candidate's quartiles. (A1)(ft) End points at 21 and 72 joined to box with straight lines. (A1) (C3)

Note: Award (A0) if lines go right through the box.

[3 marks]

Examiners report

- a. Accuracy reading the graph in part (a) was a problem with little allowance given. A number read the maximum as 130 instead of 120.
- b. Part (b) was poorly done. Many candidates did not know how to find the quartile values from the graph.
- c. Accuracy reading the graph in part (a) was a problem with little allowance given. A number read the maximum as 130 instead of 120. Part (b) was poorly done. Many candidates did not know how to find the quartile values from the graph. Many candidates gained follow through marks for the box and whisker plot by writing down their values for part (b). Some candidates did lose a mark for drawing the whiskers through the box.

The diagram below shows the cumulative frequency distribution of the heights in metres of 600 trees in a wood.



a.	Write down the median height of the trees.	[1]
b.	Calculate the interquartile range of the heights of the trees.	[2]
c.	Given that the smallest tree in the wood is $3~{ m m}$ high and the tallest tree is $28~{ m m}$ high, draw the box and whisker plot on the grid below that	[3]





Markscheme

a. Median = 11 m (A1) (C1)

Note: Award A0 for "11" without units; correct units must be included for the A1 to be awarded.

[1 mark]

b. Interquartile range = 14 - 10 (A1)

= 4 (A1)(ft) (C2)

Note: (M1) for taking a sensible difference or for both correct quartile values seen.

[2 marks]



Examiners report

a. Candidates showed less facility in this question compared to question 2.

(a) was generally answered well. There were a number of inaccurate readings from the graph.

b. Candidates showed less facility in this question compared to question 2.

(b) Errors came from candidates who either used the x coordinates for the quartiles or who wrote the quartiles as an interval, rather than subtracting these.

- c. Candidates showed less facility in this question compared to question 2.
 - (c) was well attempted from the candidates' (a) and (b).

Consider the following set of data which is plotted on the scatter diagram below.

x	2	4	7	12	4	8	9	2
у	5	8	12	26	12	14	20	7





[2]

[2]

[2]



c. Draw the regression line for y on x on the set of axes above.

Markscheme

- a. (6, 13) (A1)(A1) (C2)
- b. 0.952 (0.95202...) (A2) (C2)



(A1) (ft) line passes through their mean point (A1)(A1)(ft) (C2)

Examiners report

a. ^[N/A]

- b. ^[N/A]
- c. [N/A]
- a. A group of students were asked how long they spend practising mathematics during the week. The results are shown in the following table. [3]

Time, <i>t</i> (hours)	Number of students
$0 \le t < 1$	35
$1 \le t < 2$	30
$2 \le t < 3$	а
$3 \le t < 4$	52
$4 \le t < 5$	43

It is known that 35 < a < 52 .

Write down

- i) the modal class;
- ii) the mid-interval value of the modal class;
- iii) the class in which the median lies.
- b. For this group of students, the estimated mean number of hours spent practising mathematics is 2.69.

Calculate the value of a .

Markscheme

a. i) $3\leqslant t<4$ (A1) (C1)

Note: Accept equivalent notation: [3, 4) or [3, 4[.

ii) 3.5 (A1)(ft) (C1)

Note: Follow through from part (a)(i).

iii) $2\leqslant t<3$ (A1)(ft) (C1)

Note: Follow through from part (a)(i), for consistent misuse of inequality. Accept equivalent notation: [2, 3] or [2, 3].

b. $\frac{3.5 \times 0.5 + 30 \times 1.5 + a \times 2.5 + 52 \times 3.5 + 43 \times 4.5}{35 + 30 + a + 52 + 43} = 2.69$ (M1)(A1)(ft)

Notes: Award (*M1*) for substitution into mean formula and equating to 2.69, (*A1*)(ft) for correct substitutions. Follow through from their mid-interval value in part (a)(ii).

(a =) 40 (A1)(ft) (C3)

Note: The final (A1)(ft) is awarded only if a is an integer and 35 < a < 52. Follow through from part (a)(ii).

Examiners report

a. Question 14: Grouped frequency table.

Candidates were able to identify the modal class and the class in which the median lies but few were able to find a missing value from the grouped frequency table given the estimated mean.

b. Question 14: Grouped frequency table.

Candidates were able to identify the modal class and the class in which the median lies but few were able to find a missing value from the grouped frequency table given the estimated mean.

The cumulative frequency graph shows the amount of time in minutes, 200 students spend waiting for their train on a particular morning.



a. Write down the median waiting time.

b. Find the interquartile range for the waiting time.

c. Draw a box and whisker plot on the grid below to represent this information.

[1] [2]



Markscheme

a. Median = 25 mins (A1) (C1)

[1 mark]

b. 32 - 16 (A1)

= 16 (A1)(ft) (C2)

Notes: Award (A1) for identifying correct quartiles, (A1)(ft) for correct answer to subtraction of their quartiles.



Note: Award (A1)(ft)(A1)(ft)(A0) if lines go right through the box.

[3 marks]

Examiners report

- a. This question was well answered with many candidates gaining full marks. Some received a unit penalty in part (a) for omitting the minutes.
- b. This question was well answered with many candidates gaining full marks.
- c. This question was well answered with many candidates gaining full marks. Most of the candidates knew how to draw the box and whisker plot.

A mark was deducted if the whiskers were drawn all the way through the box.

Two groups of 40 students were asked how many books they have read in the last two months. The results for **the first group** are shown in the

following table.

Number of books read	Frequency
2	5
3	8
4	13
5	7
6	4
7	2
8	1

The quartiles for these results are 3 and 5.

- a. Write down the value of the median for these results.
- b. Draw a box-and-whisker diagram for these results on the following grid.



c. The results for the second group of 40 students are shown in the following box-and-whisker diagram.



Estimate the number of students in the second group who have read at least 6 books.

Markscheme

a. 4 **(A1)(C1)**

[2]

[1]

[3]



Notes: Award (A1)(ft) for correct median, (A1) for correct quartiles and box, (A1) for endpoints 2 and 8 joined by a straight line that does not cross the box. Follow through from their median from part (a).

c. 40×0.25 (M1)

Notes: Award (M1) for $40 \times 25\%$ OR $40 - 40 \times 75\%$.

10 (A1) (C2)

Examiners report

a. ^[N/A]

b. ^[N/A]

c. [N/A]

120 Mathematics students in a school sat an examination. Their scores (given as a percentage) were summarized on a cumulative frequency

diagram. This diagram is given below.



a. Complete the grouped frequency table for the students.

Examination Score <i>x</i> (%)	$0 \le x \le 20$	$20 < x \le 40$	$40 < x \le 60$	$60 < x \le 80$	$80 < x \le 100$
Frequency	14	26			

b. Write down the mid-interval value of the $40 < x \leqslant 60$ interval.

c. Calculate an estimate of the mean examination score of the students.

Markscheme



[1 mark]

[1]

c. Mean = $\frac{10 \times 14 + \dots + 90 \times 6}{120}$ (M1)

Note: Award (M1) for correct substitution of their values from (a) in mean formula.

$$=45\frac{2}{2}(45.7)$$
 (A1)(ft) (C2)

[2 marks]

Examiners report

- a. This question was poorly answered by many of the candidates. A number of students could not identify the specified frequencies from the graph in part a). Most could not give the midinterval value although surprisingly many of these candidates then went on and used the correct mid-interval value in the mean formula. A number did not understand the meaning of 'an estimate of the mean' and just wrote down a number read from the diagram.
- b. This question was poorly answered by many of the candidates. A number of students could not identify the specified frequencies from the graph in part a). Most could not give the midinterval value although surprisingly many of these candidates then went on and used the correct mid-interval value in the mean formula. A number did not understand the meaning of 'an estimate of the mean' and just wrote down a number read from the diagram.
- c. This question was poorly answered by many of the candidates. A number of students could not identify the specified frequencies from the graph in part a). Most could not give the midinterval value although surprisingly many of these candidates then went on and used the correct mid-interval value in the mean formula. A number did not understand the meaning of 'an estimate of the mean' and just wrote down a number read from the diagram.

In a high school, 160 students completed a questionnaire which asked for the number of people they are following on a social media website. The results were recorded in the following box-and-whisker diagram.



The following incomplete table shows the distribution of the responses from these 160 students.

Number of people they are following (x)	Number of high school students
$0 \le x \le 50$	4
$50 < x \le 100$	
$100 < x \le 150$	34
$150 < x \le 200$	46
$200 < x \le 250$	
$250 < x \le 300$	16

a.	Write down the median.	[1]
b.	Complete the table.	[2]
c.i	Write down the mid-interval value for the $100 < x \le 150$ group.	[1]

c.ii.Using the table, calculate an estimate for the mean number of people being followed on the social media website by these 160 students. [2]

Markscheme

a. 180 (A1) (C1)

[1 mark]

```
b. 36, 24 (A1)(A1) (C2)
```

Note: Award (A0)(A1) for two incorrect values that add up to 60.

[2 marks]

c.i. 125 (accept 125.5) (A1)

c.ii. $\frac{4 \times 25 + 36 \times 75 + 34 \times 125 + 46 \times 175 + 24 \times 225 + 16 \times 275}{160}$ (M1)

Note: Award (M1) for correct substitution of their mid-interval values, multiplied by their frequencies, into mean formula.

=156 (155.625) (A1)(ft) (C3)

Note: Follow through from parts (b) and (c)(i).

[3 marks]

Examiners report

a. [N/A]

b. [N/A]

c.i. [N/A]

c.ii.^[N/A]

A cumulative frequency graph is given below which shows the height of students in a school.



a.	Write down the median height of the students.	[1]
b	Write down the 25 th percentile.	[1]
c.	Write down the 75 th percentile.	[1]
d	The height of the tallest student is 195 cm and the height of the shortest student is 136 cm.	[3]
	Draw a box and whisker plot on the grid below to represent the heights of the students in the school.	



Markscheme

a. 170 (A1) (C1)

[1 mark]

b. 163 (A1) (C1)

[1 mark]

c. 172 (A1) (C1)



Notes: Award (A1)(ft) for correct median, (A1)(ft) for correct quartiles and box (A1) for correct end points of whiskers.

Award at most (A1)(A1)(A0) if lines go right through the box.

[3 marks]

Examiners report

- a. The box and whisker plot was well done, even when the students had incorrect values. Most candidates found the correct median but a few could not find the 25th and 75th percentiles.
- b. The box and whisker plot was well done, even when the students had incorrect values. Most candidates found the correct median but a few could not find the 25th and 75th percentiles.
- c. The box and whisker plot was well done, even when the students had incorrect values. Most candidates found the correct median but a few could not find the 25th and 75th percentiles.
- d. The box and whisker plot was well done, even when the students had incorrect values. Most candidates found the correct median but a few could not find the 25th and 75th percentiles.

The histogram shows the time, *t*, in minutes, that it takes the customers of a restaurant to eat their lunch on one particular day. Each customer took less than 25 minutes.



The histogram is incomplete, and only shows data for $0 \le t < 20$.

The mean time it took **all** customers to eat their lunch was estimated to be 12 minutes.

It was found that k customers took between 20 and 25 minutes to eat their lunch.

a. Write down the mid-interval value for $10 \le t < 15$.	[1]
b.i.Write down the total number of customers in terms of <i>k</i> .	[1]
b.iiCalculate the value of <i>k</i> .	[3]
c. Hence, complete the histogram.	[1]

Markscheme

a. 12.5 (A1) (C1)

[1 mark]

b.i.33 + k **OR** 10 + 8 + 5 + 10 + k (A1)

Note: Award **(A1)** for "number of customers = 33 + k".

[1 mark]

b.ii. $rac{2.5 imes 10+7.5 imes 8+ \ldots + 22.5 imes k}{33+k} = 12$ (M1)(A1)(ft)

Note: Award (M1) for substitution into the mean formula and equating to 12, (A1)(ft) for their correct substitutions.

(k =) 7 (A1)(ft) (C4)

Note: Follow through from part (b)(i) and their mid-interval values, consistent with part (a). Do not award final (A1) if answer is not an integer.

[3 marks]



Note: Follow through from their part (b)(ii) but only if the value is between 1 and 10, inclusive.

[1 mark]

Examiners report

a. [N/A] b.i.[N/A] b.ii.[N/A] b.ii. c. [N/A] The table shows the number of bicycles owned by 50 households.

Number of bicycles per household	Frequency (number of households)	Cumulative frequency
0	3	3
1	7	10
2	12	22
3	14	36
4	4	40
5	t	w
6	2	50

a. Write down the value of

(i) t;

(ii) w .

b. Indicate with a tick (\checkmark) whether the following statements are True or False.

Statement	True	False
Every household owns at least 1 bicycle.		
The median number of bicycles per household is 3.		
The 25 th percentile is 1 bicycle per household.		
There are 10 households with at most 1 bicycle.		

Markscheme

a. (i) 8 (A1)

b.

(ii) 48 (A1)(ft) (C2)

Note: Follow through from their *t*, even if no workings seen as long as w < 50.

Statement	True	False
Every household owns at least 1 bicycle.		1
The median number of bicycles per household is 3.	1	
The 25 th percentile is 1 bicycle per household.		1
There are 10 households with at most 1 bicycle.	~	

(A1)(A1)(A1)(A1) (C4)

[2]

[4]

a. ^[N/A] b. ^[N/A]

Examiners report

a. In a school 160 students sat a mathematics examination. Their scores, given as marks out of 90, are summarized on the cumulative frequency [1]





Write down the median score.

b. The lower quartile of these scores is 40.

Find the interquartile range.

c. The lowest score was 6 marks and the highest score was 90 marks.

Draw a box-and-whisker diagram on the grid below to represent the students' examination scores.



Markscheme

- a. 48 (A1) (C1)
- b. 58 40 (M1)

Note: Award (M1) for 58 and 40 seen.



(A1)(A1)(ft)(A1)(ft) (C3)

Note: Award (A1) for the correct maximum and minimum, (A1)(ft) for their correct median and (A1)(ft) for 40 and their upper quartile.

Follow through from parts (a) and (b).

Award a maximum of (A1)(A1)(A1)(A0) if the horizontal line goes through the box or if a ruler has clearly not been used.

Examiners report

a. Question 3: Statistics - boxplot

Whilst there were many fine attempts, the number of poor responses to this question was surprising. The interquartile range was little understood,

yet quartiles were correctly plotted. The lack of precision - due mainly to not using a ruler - in drawing the boxplot was also disappointing.

b. Question 3: Statistics - boxplot

Whilst there were many fine attempts, the number of poor responses to this question was surprising. The interquartile range was little understood, yet quartiles were correctly plotted. The lack of precision – due mainly to not using a ruler – in drawing the boxplot was also disappointing.

c. Question 3: Statistics - boxplot

Whilst there were many fine attempts, the number of poor responses to this question was surprising. The interquartile range was little understood, yet quartiles were correctly plotted. The lack of precision – due mainly to not using a ruler – in drawing the boxplot was also disappointing.