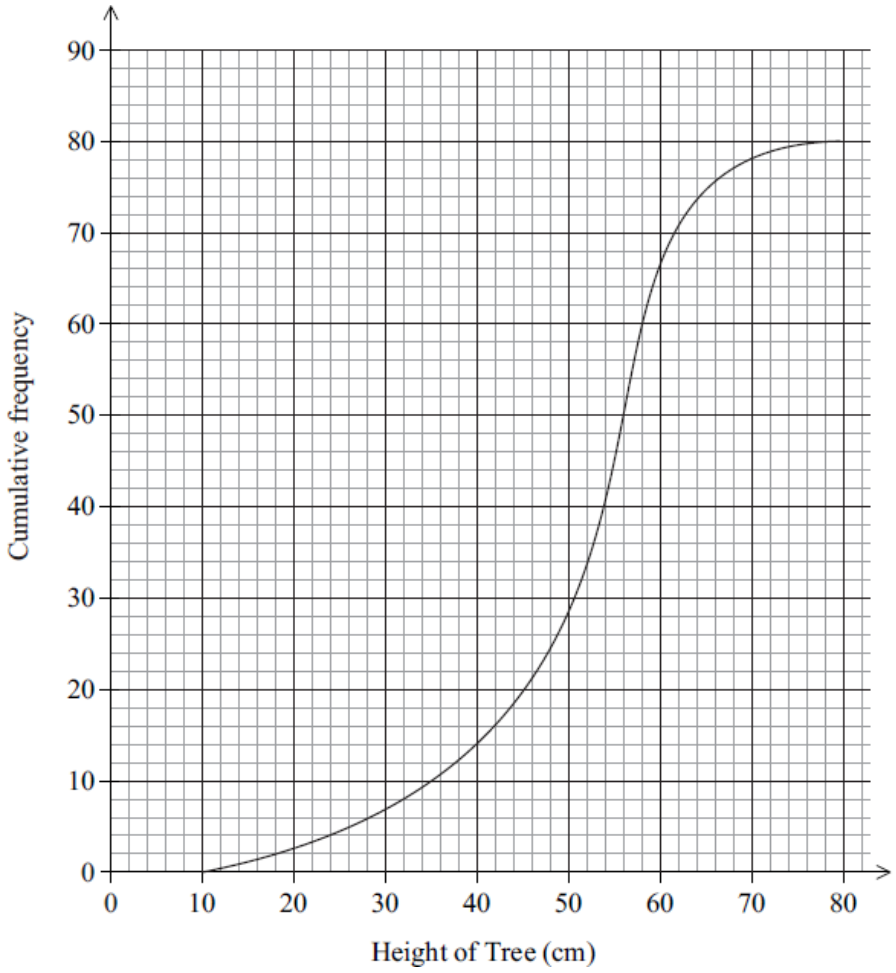
IB review questions Sept 1 *[75 marks]*

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



1a. Write down the median height of the trees. *[1 mark]*

Markscheme

54 (cm) ***(A1)*** ***(C1)***

Examiners report

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.

1b. Write down the 75^{th} percentile. *[1 mark]*

Markscheme

58 ***(A1)*** ***(C1)***

Examiners report

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.

1c. Find the interquartile range. *[2 marks]*

Markscheme

58, 45 ***(A1)***

Note: Award ***(A1)*** for 45 seen.

58 – 45

= 13 ***(A1)*****(ft)** ***(C2)***

Note: Follow through from part (b).

Examiners report

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.

1d. Estimate the number of trees that are more than 40 cm in height. *[2 marks]*

Markscheme

80 – 14 ***(M1)***

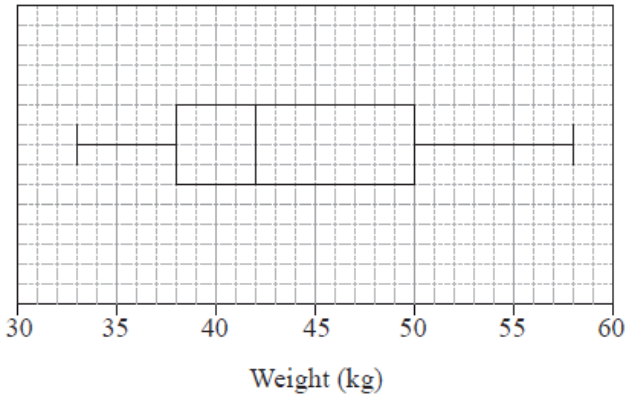
Note: Award ***(M1)*** for 14 seen.

= 66 ***(A1)*** ***(C2)***

Examiners report

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.

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



2a. Write down the median weight of the females. *[1 mark]*

Markscheme

42 kg *(A1)* *(C1)*

Note: The units are required.

Examiners report

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.

2b. Calculate the range. *[2 marks]*

Markscheme

58 – 33 *(A1)*

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

= 25 *(A1)* *(C2)*

Examiners report

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.

2c. Estimate the probability that the weight of a randomly chosen female is more than 50 kg. *[1 mark]*

Markscheme

$\frac{1}{4}(0.25, 25\%)$ *(A1)* *(C1)*

Examiners report

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.

2d. 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 marks]*

Markscheme

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

$\left(\text{The mean is calculated } \bar{x} = \frac{35.5\times15+40\times15+54\times15}{60}\right)$

$\bar{x} = 43.875\ (kg)$ *(R1)* *(C2)*

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

Examiners report

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.

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

3a. Write down the value of *[2 marks]*

(i) t ;

(ii) w .

Markscheme

(i) 8 *(A1)*

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

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

Examiners report

[N/A]

3b. Indicate with a tick (✓) whether the following statements are True or False. *[4 marks]*

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

Markscheme

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

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

Examiners report

[N/A]

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.

4a. Calculate the value of p. *[2 marks]*

Markscheme

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

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

4 *(A1)* *(C2)*

Examiners report

[N/A]

4b. Find the median number of passengers per carriage. *[2 marks]*

Markscheme

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

5.5 *(A1)*(ft) *(C2)*

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

Examiners report

[N/A]

4c. If the passengers in the eleventh carriage are also included, the mean number of passengers per carriage increases to 6.0. *[2 marks]*

Determine the number of passengers in the eleventh carriage of the train.

Markscheme

$\frac{56+x}{11} = 6.0$ ***(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

[N/A]

Francesca is a chef in a restaurant. She cooks eight chickens and records their masses and cooking times. The mass m of each chicken, in kg, and its cooking time t, in minutes, are shown in the following table.

Mass m (kg)	Cooking time t (minutes)
1.5	62
1.6	75
1.8	82
1.9	83
2.0	86
2.1	87
2.1	91
2.3	98

5. Write down for this set of data *[2 marks]*

(i) the mean mass, $\bar{m}$;

(ii) the mean cooking time, $\bar{t}$.

Markscheme

(i) 1.91 (kg) (1.9125 kg) ***(G1)***

(ii) 83 (minutes) ***(G1)***

Examiners report

[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

6. Calculate the **exact value** of the mean length. *[2 marks]*

Markscheme

$\left(\frac{104.5+105.1+...}{6}\right)$ ***(M1)***

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

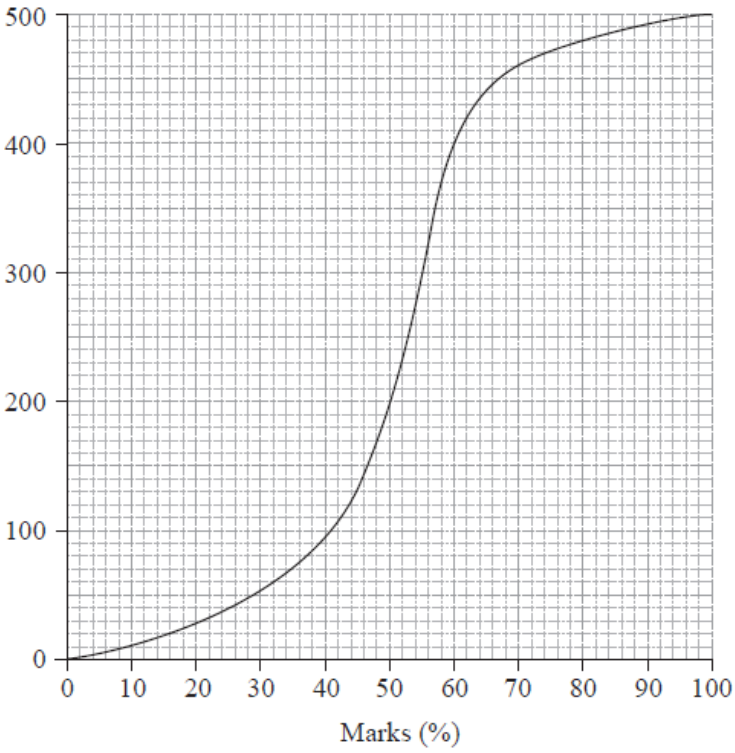
= 104.9 (cm) ***(A1)*** ***(C2)***

[2 marks]

Examiners report

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 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.

7a. Find the number of students who failed the examination. *[2 marks]*

Markscheme

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]

Examiners report

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.

7b. Find the number of students who were awarded grade C or better. *[2 marks]*

Markscheme

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]

Examiners report

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.

7c. The top 20 % of the students are eligible for further study. *[2 marks]*

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

Examiners report

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.

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

8a. Write down the modal class. *[1 mark]*

Markscheme

$14 \le x < 18$ ***(A1)*** ***(C1)***

[1 mark]

Examiners report

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.

8b. Write down the mid interval value for the class $6 \le x < 10$. *[1 mark]*

Markscheme

8 ***(A1)*** ***(C1)***

[1 mark]

Examiners report

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.

8c. Calculate an estimate of the mean annual snowfall. *[2 marks]*

Markscheme

$\frac{4\times30+8\times26+12\times29+16\times32+20\times18+24\times27+28\times14}{176}$ ***(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]

Examiners report

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.

8d. Find the number of years for which the annual snowfall was at least 18 cm. *[2 marks]*

Markscheme

18 + 27 + 14 ***(M1)***

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

= 59 ***(A1)*** ***(C2)***

[2 marks]

Examiners report

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 table below shows the scores for 12 golfers for their first two rounds in a local golf tournament.

Round 1 (x)	71	79	66	73	69	76	68	75	82	67	69	74
Round 2 (y)	73	81	68	75	70	79	69	77	83	68	72	76

9. (i) Write down the mean score in Round 1. *[5 marks]*

(ii) Write down the standard deviation in Round 1.

(iii) Find the number of these golfers that had a score of more than one standard deviation above the mean in Round 1.

Markscheme

(i) $\frac{71+79+\ldots}{12}$ ***(M1)***

$72.4\left(72.4166\ldots, \frac{869}{12}\right)$ ***(A1)(G2)***

Note: Award ***(M1)*** for correct substitution into the mean formula.

(ii) 4.77 (4.76896…) ***(G1)***

(iii) 72.4 + 4.77 = 77.17 ***(M1)***

Note: Award ***(M1)*** for adding their mean to their standard deviation.

Two golfers ***(A1)*****(ft)*****(G2)***

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

[5 marks]

Examiners report

The question was for the most part approached by almost all candidates and answered relatively well. The question in part (e) related to the use of the equation of the regression line for predicting, although regularly asked on exams, was still found to be a difficult one by some candidates. Some answers still suggested mathematical thinking and language unaccustomed to drawing conclusions and providing justifications.

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.

10a. Find *[1 mark]*

x ;

Markscheme

$x = 1$ ***(A1)*** ***(C1)***

[1 mark]

Examiners report

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.

10b. Find *[2 marks]*

y ;

Markscheme

$\frac{3+y}{2} = 5$ ***(M1)***

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

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

[2 marks]

Examiners report

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.

10c. Find *[3 marks]*

z.

Markscheme

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

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

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

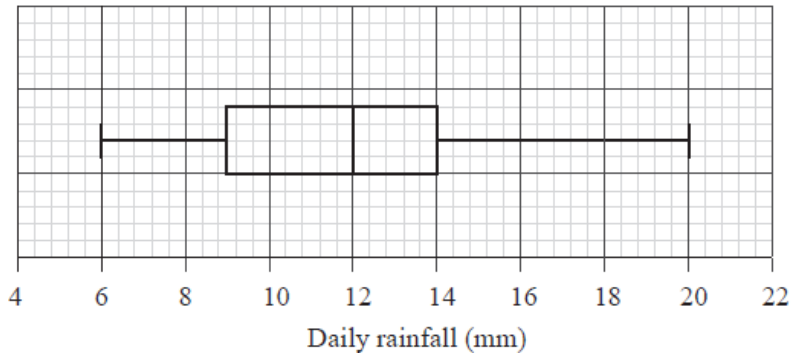
Note: Follow through from their x and y found in part (b) provided y is a positive integer less than 14.

[3 marks]

Examiners report

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.

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.



11a. Write down *[2 marks]*

(i) the lowest daily rainfall;

(ii) the highest daily rainfall.

Markscheme

(i) 6 (mm) ***(A1)***

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

[2 marks]

Examiners report

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.

11b. State what the value of 12 mm represents on the given diagram. *[1 mark]*

Markscheme

Median ***(A1)*** ***(C1)***

Note: Award ***(A1)*** for Q_2 or 50^{th} percentile.

[1 mark]

Examiners report

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.

11c. Find the interquartile range. *[2 marks]*

Markscheme

14 – 9 *(A1)*

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

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

[2 marks]

Examiners report

Part (c) was generally well done with many candidates correctly identifying Q1 and Q3 and many correct answers of 5 were seen.

11d. Write down the percentage of the data which is less than the upper quartile. *[1 mark]*

Markscheme

75 (%) *(A1)* *(C1)*

[1 mark]

Examiners report

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 $\frac{14-6}{14} \times 100$.

200 people were asked the amount of time T (minutes) they had spent in the supermarket. The results are represented in the table below.

Time (T)	$0 < T \leq 10$	$10 < T \leq 20$	$20 < T \leq 30$	$30 < T \leq 40$	$40 < T \leq 50$
Number of people	23	57	93	21	6

12a. State if the data is discrete or continuous. *[1 mark]*

Markscheme

continuous *(A1)*

[1 mark]

Examiners report

(a) Many candidates thought that this was discrete data.

12b. State the modal group. *[1 mark]*

Markscheme

$20 < T \leq 30$ *(A1)*

[1 mark]

Examiners report

(b) This part was very well done with the occasional candidate writing down the number rather than the group.

12c. Write down the midpoint of the interval $10 < T \leq 20$. *[1 mark]*

Markscheme

15 ***(A1)***

[1 mark]

Examiners report

(c) Fairly well done although 15.5 was seen quite often.

12d. Use your graphic display calculator to find an estimate for *[3 marks]*

(i) the mean;

(ii) the standard deviation.

Markscheme

(i) 21.5 ***(G2)***

(ii) 9.21 (9.20597…) ***(G1)***

[3 marks]

Examiners report

(d) This was really badly done with most candidates only putting the midpoints into their GDC or only putting the frequencies into their GDC. Perhaps they did not know how to use their GDC correctly.

12e. The results are represented in the cumulative frequency table below, with upper class boundaries of 10, 20, 30, 40, 50. *[2 marks]*

Upper class boundaries	10	20	30	40	50
Cumulative frequency	23	80	173	q	r

Write down the value of

(i) q;

(ii) r.

Markscheme

(i) $q = 194$ ***(A1)***

(ii) $r = 200$ ***(A1)***

[2 marks]

Examiners report

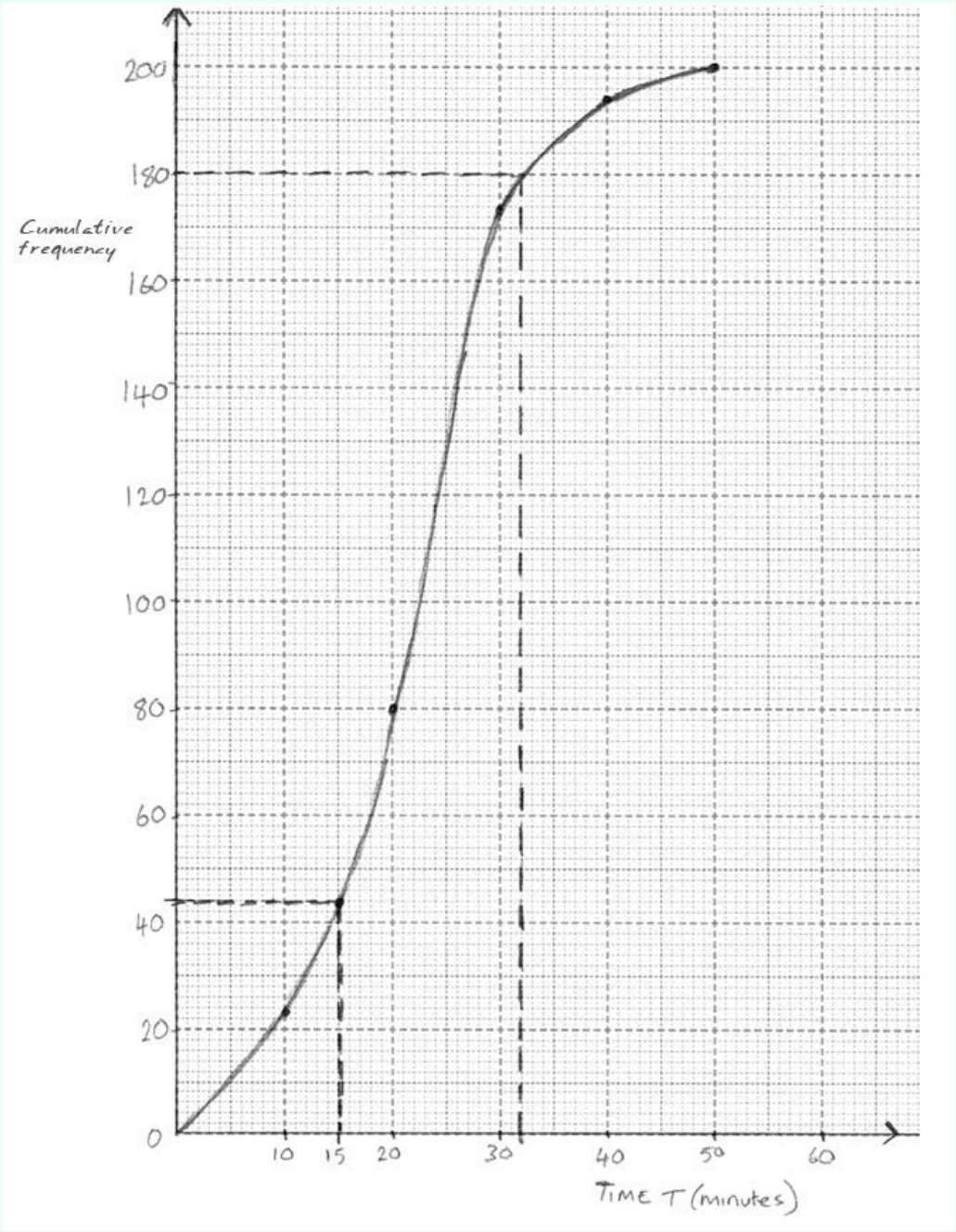
(e) The values of q and r were mostly correct.

12f. The results are represented in the cumulative frequency table below, with upper class boundaries of 10, 20, 30, 40, 50. *[4 marks]*

Upper class boundaries	10	20	30	40	50
Cumulative frequency	23	80	173	q	r

On graph paper, draw a cumulative frequency graph, using a scale of 2 cm to represent 10 minutes (T) on the horizontal axis and 1 cm to represent 10 people on the vertical axis.

Markscheme



(A1)(A2)*(ft)*(A1)

Notes: Award ***(A1)*** for scale and axis labels, ***(A2)*(ft)** for 5 correct points, ***(A1)*(ft)** for 4 or 3 correct points, ***(A0)*** for less than 3 correct points, ***(A1)*** for smooth curve through their points, starting at (0, 0). Follow through from their answers to parts (e)(i) and (e)(ii).

[4 marks]

Examiners report

(f) Most candidates plotted the points correctly. Some had problems plotting the 23 and 173. A few candidates used the midpoints instead of the end points and some drew bar charts.

12g. **Use your graph** from part (f) to estimate *[6 marks]*

(i) the median;

(ii) the 90^{th} percentile of the results;

(iii) the number of people who shopped at the supermarket for more than 15 minutes.

Markscheme

(i) 22.5 ± 2 ***(A1)***

(ii) 32 ± 2 ***(M1)(A1)*****(ft)*****(G2)***

Note: Award ***(M1)*** for lines drawn on graph or some indication of method, follow through from their graph if working is shown.

(iii) 44 ± 2 ***(A1)*****(ft)**

Note: Follow through from their graph if working is shown.

200 – 44 = 156 ***(M1)(A1)*****(ft)*****(G2)***

Note: Award ***(M1)*** for subtraction from 200, follow through from their graph if working is shown.

[6 marks]

Examiners report

(g) There was a lot of follow through marks gained here by those candidates who drew lines or put marks on their graphs in the correct places.