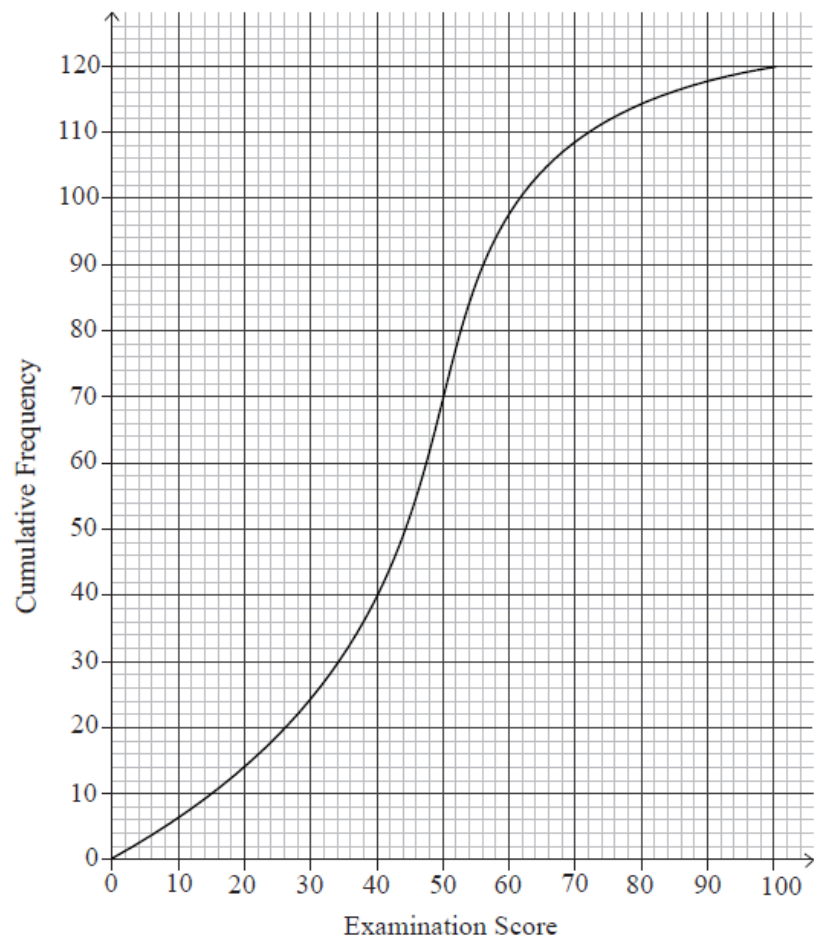


Topic 2 and 4 Review [79 marks]

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.



1a. Complete the grouped frequency table for the students.

[3 marks]

Examination Score $x$ (%)	$0 \leq x \leq 20$	$20 < x \leq 40$	$40 < x \leq 60$	$60 < x \leq 80$	$80 < x \leq 100$
Frequency	14	26			

Markscheme

%	0–20	20–40	40–60	60–80	80–100
F	14	26	58	16	6

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

[3 marks]

1b. Write down the mid-interval value of the  $40 < x \leq 60$  interval.

[1 mark]

## Markscheme

50 (A1) (C1)

[1 mark]

- 1c. Calculate an estimate of the mean examination score of the students.

[2 marks]

## Markscheme

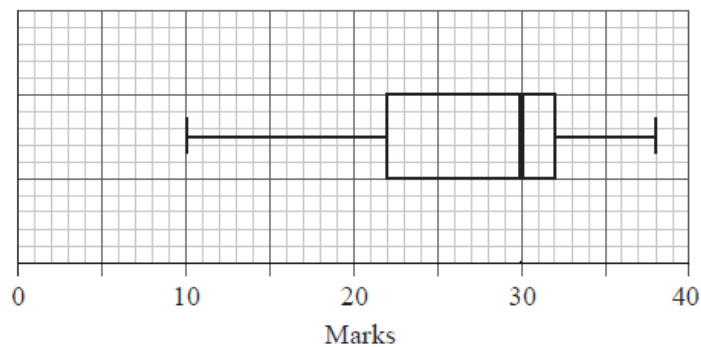
$$\text{Mean} = \frac{10 \times 14 + \dots + 90 \times 6}{120} \quad (M1)$$

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

$$= 45\frac{2}{3} (45.7) \quad (A1)(ft) \quad (C2)$$

[2 marks]

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.



- 2a. Write down the median mark.

[1 mark]

## Markscheme

30 (A1) (C1)

[1 mark]

- 2b. Write down the 75<sup>th</sup> percentile mark.

[1 mark]

## Markscheme

32 (A1) (C1)

[1 mark]

- 2c. Write down the range of marks.

[2 marks]

## Markscheme

$38 - 10 = 28$  (AI)(AI) (C2)

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

[2 marks]

- 2d. Estimate the number of students who achieved a mark greater than 32.

[2 marks]

## Markscheme

$0.25 \times 56 = 14$  (MI)(AI) (C2)

**Note:** Award (MI) for multiplying 0.25 by 56.

[2 marks]

### Part A

A university required all Science students to study one language for one year. A survey was carried out at the university amongst the 150 Science students. These students all studied one of either French, Spanish or Russian. The results of the survey are shown below.

	French	Spanish	Russian
Female	9	29	12
Male	31	40	29

Ludmila decides to use the  $\chi^2$  test at the 5% level of significance to determine whether the choice of language is independent of gender.

- 3a. State Ludmila's null hypothesis.

[1 mark]

## Markscheme

$H_0$  : Choice of language is independent of gender. (AI)

**Notes:** Do not accept "not related" or "not correlated".

[1 mark]

- 3b. Write down the number of degrees of freedom.

[1 mark]

## Markscheme

2 (AI)

[1 mark]

- 3c. Find the expected frequency for the females studying Spanish.

[2 marks]

## Markscheme

$$\frac{50 \times 69}{150} = 23 \quad (M1)(A1)(G2)$$

**Notes:** Award **(M1)** for correct substituted formula, **(A1)** for 23.

[2 marks]

- 3d. Use your graphic display calculator to find the  $\chi^2$  test statistic for this data.

[2 marks]

## Markscheme

$$\chi^2 = 4.77 \quad (G2)$$

**Notes:** If answer is incorrect, award **(M1)** for correct substitution in the correct formula (all terms).

[2 marks]

- 3e. State whether Ludmila accepts the null hypothesis. Give a reason for your answer.

[2 marks]

## Markscheme

Accept  $H_0$  since

$$\chi^2_{calc} < \chi^2_{crit}(5.99) \text{ or } p\text{-value}(0.0923) > 0.05 \quad (R1)(A1)(ft)$$

**Notes:** Do not award **(R0)(A1)**. Follow through from their (d) and (b).

At the end of the year, only seven of the female Science students sat examinations in Science and French. The marks for these seven students are shown in the following table.

<b>Science (<i>S</i>)</b>	23	51	56	62	12	73	72
<b>French (<i>F</i>)</b>	65	45	45	40	70	36	30

- 3f. Draw a labelled scatter diagram for this data. Use a scale of 2 cm to represent 10 marks on the  $x$ -axis ( $S$ ) and 10 marks on the  $y$ -axis ( $F$ ).

[4 marks]

## Markscheme

Award **(A1)** for correct scale and labels.

Award **(A3)** for all seven points plotted correctly, **(A2)** for 5 or 6 points plotted correctly, **(A1)** for 3 or 4 points plotted correctly.

**(A4)**

[4 marks]

- 3g. Use your graphic calculator to find

[2 marks]

- $\bar{S}$ , the mean of  $S$ ;
- $\bar{F}$ , the mean of  $F$ .

## Markscheme

(i)  $\bar{S} = 49.9$ , (GI)

(ii)  $\bar{F} = 47.3$  (GI)

[2 marks]

- 3h. Plot the point M( $\bar{S}$ ,  $\bar{F}$ ) on your scatter diagram.

[1 mark]

## Markscheme

M(49.9, 47.3) plotted on scatter diagram (AI)(ft)

**Notes:** Follow through from (a) and (b).

[1 mark]

- 3i. Use your graphic display calculator to find the equation of the regression line of  $F$  on  $S$ .

[2 marks]

## Markscheme

$F = -0.619S + 78.2$  (GI)(GI)

**Notes:** Award (GI) for  $-0.619S$ , (GI) for 78.2. If the answer is not in the form of an equation, award (GI)(G0). Accept  $y = -0.619x + 78.2$ .

**OR**

$(F - 47.3 = -0.619(S - 49.9))$  (GI)(GI)

**Note:** Award (GI) for  $-0.619$ , (GI) for the coordinates of their midpoint used. Follow through from their values in (b).

[2 marks]

- 3j. Draw the regression line on your scatter diagram.

[2 marks]

## Markscheme

line drawn on scatter diagram (AI)(ft)(AI)(ft)

**Notes:** The drawn line **must** be straight for any marks to be awarded. Award (AI)(ft) passing through their M plotted in (c). Award (AI)(ft) for correct  $y$ -intercept. Follow through from their  $y$ -intercept found in (d).

[2 marks]

- 3k. Carletta's mark on the Science examination was 44. She did not sit the French examination.  
Estimate Carletta's mark for the French examination.

[2 marks]

## Markscheme

$$F = -0.619 \times 44 + 78.2 \quad (M1)$$

$$= 51.0 \text{ (allow 51 or 50.9)} \quad (A1)(ft)(G2)(ft)$$

**Note:** Follow through from their equation.

**OR**

(M1) any indication of an acceptable graphical method. (M1)

(A1)(ft) from their regression line. (A1)(ft)(G2)(ft)

[2 marks]

31. Monique's mark on the Science examination was 85. She did not sit the French examination. Her French teacher wants to use the regression line to estimate Monique's mark. [2 marks]

State whether the mark obtained from the regression line for Monique's French examination is reliable. Justify your answer.

## Markscheme

not reliable (A1)

Monique's score in Science is outside the range of scores used to create the regression line. (R1)

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

[2 marks]

The speed,  $s$ , in  $\text{km h}^{-1}$ , of 120 vehicles passing a point on the road was measured. The results are given below.

Speed, $s$ ( $\text{km h}^{-1}$ )	$0 < s \leq 50$	$50 < s \leq 60$	$60 < s \leq 70$	$70 < s \leq 80$	$80 < s \leq 90$	$90 < s \leq 100$
Number of vehicles	30	46	22	12	8	2

- 4a. Write down the midpoint of the  $60 < s \leq 70$  interval.

[1 mark]

## Markscheme

65 (A1)

[1 mark]

- 4b. Use your graphic display calculator to find an estimate for
- the mean speed of the vehicles;
  - the standard deviation of the speeds of the vehicles.

[3 marks]

## Markscheme

(i) 54 ( $\text{km h}^{-1}$ ) (G2)

**Note:** If the answer to part (b)(i) is consistent with the answer to part (a) then award (G2)(ft) even if no working seen.

(ii) 19.2 (19.2093...) (G1)

**Note:** Accept 19, do not accept 20.

[3 marks]

4c. Write down the number of vehicles whose speed is less than or equal to  $60 \text{ km h}^{-1}$ .

[1 mark]

## Markscheme

76 (A1)

[1 mark]

4d. Consider the cumulative frequency table below.

[2 marks]

Speed, $s$ ( $\text{km h}^{-1}$ )	$s \leq 50$	$s \leq 60$	$s \leq 70$	$s \leq 80$	$s \leq 90$	$s \leq 100$
Number of vehicles	30	$a$	$b$	110	$c$	120

Write down the value of  $a$ , of  $b$  and of  $c$ .

## Markscheme

$a = 76, b = 98$  (A1)(ft)

**Note:** Follow through from their answer to part (c) for  $a$  and  $b = \text{their } a + 22$ .

$c = 118$  (A1)

[2 marks]

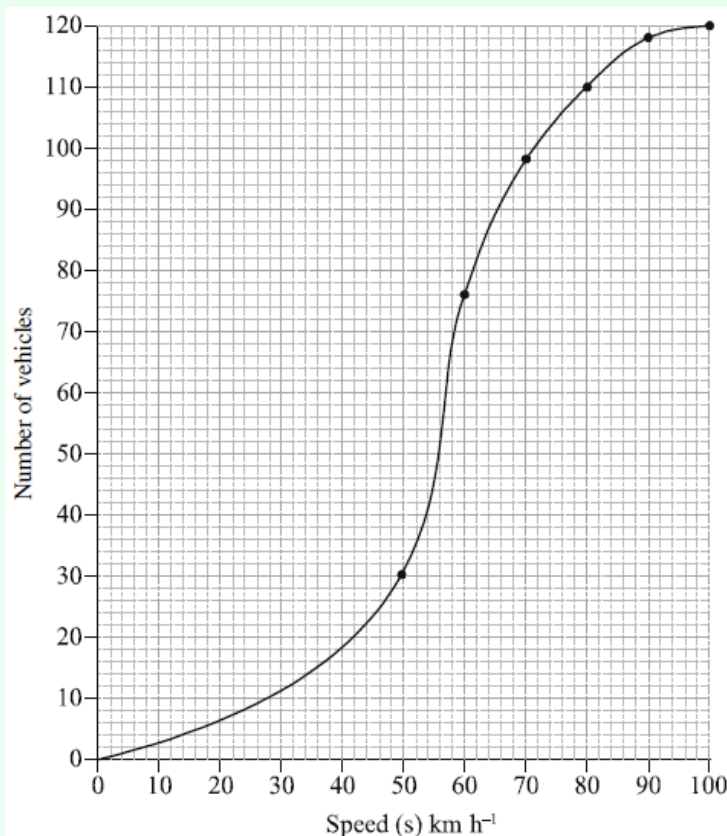
4e. Consider the cumulative frequency table below.

[4 marks]

Speed, $s$ ( $\text{km h}^{-1}$ )	$s \leq 50$	$s \leq 60$	$s \leq 70$	$s \leq 80$	$s \leq 90$	$s \leq 100$
Number of vehicles	30	$a$	$b$	110	$c$	120

Draw a cumulative frequency graph for the information from the table. Use 1 cm to represent  $10 \text{ km h}^{-1}$  on the horizontal axis and 1 cm to represent 10 vehicles on the vertical axis.

## Markscheme



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

**Notes:** Award (A1) for axes labelled and correct scales. If the axes are reversed do not award this mark but follow through. Award (A2)(ft) for **their** 6 points correct, (A1)(ft) for at least 3 of these points correct. Award (A1) for smooth curve drawn through all points **including** (0, 0). If either the  $x$  or the  $y$  axis has a break in it to zero, do not award this final mark.

[4 marks]

4f. Use your cumulative frequency graph to estimate

[4 marks]

- the median speed of the vehicles;
- the number of vehicles that are travelling at a speed less than or equal to  $65 \text{ km h}^{-1}$ .

## Markscheme

- $57 \text{ (km h}^{-1}) (\pm 2)$  (M1)(A1)(ft)(G2)

**Note:** Award (M1) for clear indication of median on their graph. Follow through from their graph. If their answer is consistent with their incorrect graph but there is no working present on graph then no marks are awarded.

- 90 vehicles ( $\pm 2$ ) (M1)(A1)(ft)(G2)

**Note:** Award (M1) for clear indication of method on their graph. Follow through from their graph. If their answer is consistent with their incorrect graph but there is no working present on graph then no marks are awarded.

[4 marks]

4g. All drivers whose vehicle's speed is greater than one standard deviation above the speed limit of  $50 \text{ km h}^{-1}$  will be fined.

[3 marks]

Use your graph to estimate the number of drivers who will be fined.



## Markscheme

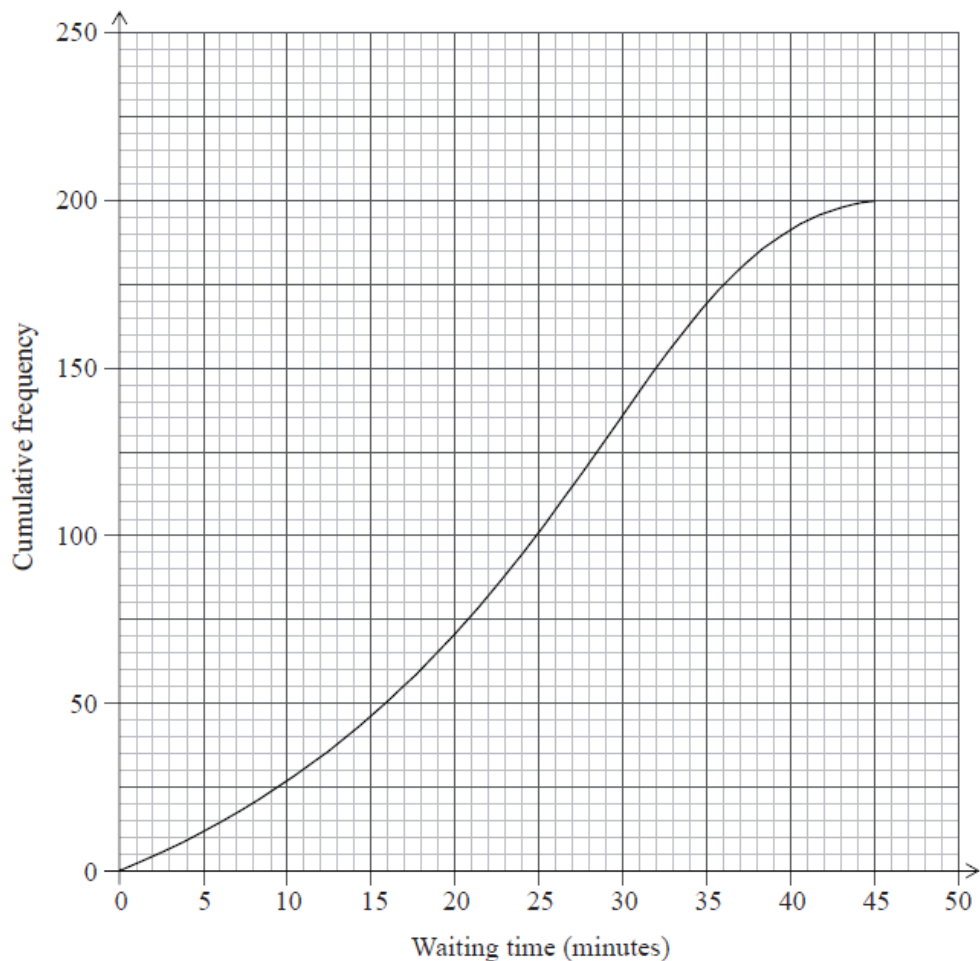
$$50 + 19.2 = 69.2 \quad (AI)(ft)$$

24 ( $\pm 2$ ) drivers will be fined  $(MI)(AI)(ft)(G2)$

**Notes:** Follow through from their graph and from their part (b)(ii). Award  $(MI)$  for indication of method on their graph. If their answer is consistent with their incorrect graph but there is no working present on graph then no marks are awarded.

[3 marks]

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



5a. Write down the median waiting time.

[1 mark]

## Markscheme

Median = 25 mins  $(AI) \quad (CI)$

[1 mark]

5b. Find the interquartile range for the waiting time.

[2 marks]

## Markscheme

$$32 - 16 \quad (A1)$$

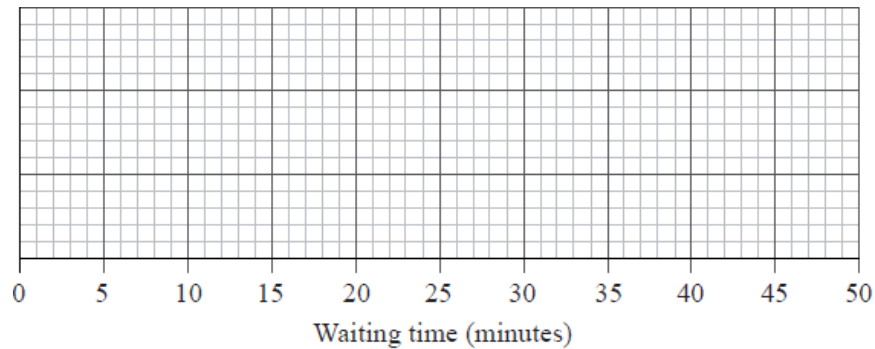
$$= 16 \quad (A1)(ft) \quad (C2)$$

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

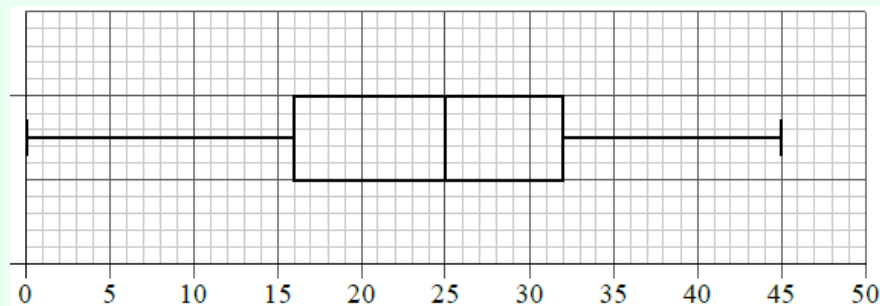
[2 marks]

- 5c. Draw a box and whisker plot on the grid below to represent this information.

[3 marks]



## Markscheme



median shown (A1)(ft)  
 box with ends at their quartiles (A1)(ft)  
 end points at 0 and 45 joined to box with straight lines (A1) (C3)

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

[3 marks]

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

House	A	B	C	D	E	F	G	H
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.

- 6a. The mean is 5.

[1 mark]

## Markscheme

F (A1)

[1 mark]

- 6b. The range is 4. [1 mark]

## Markscheme

T (A1)

[1 mark]

- 6c. The mode is 6. [1 mark]

## Markscheme

F (A1)

[1 mark]

- 6d. The standard deviation is 1.4 correct to 2 significant figures. [1 mark]

## Markscheme

F (A1) (C4)

[1 mark]

- 6e. Calculate the interquartile range for the number of inhabitants per house. [2 marks]

## Markscheme

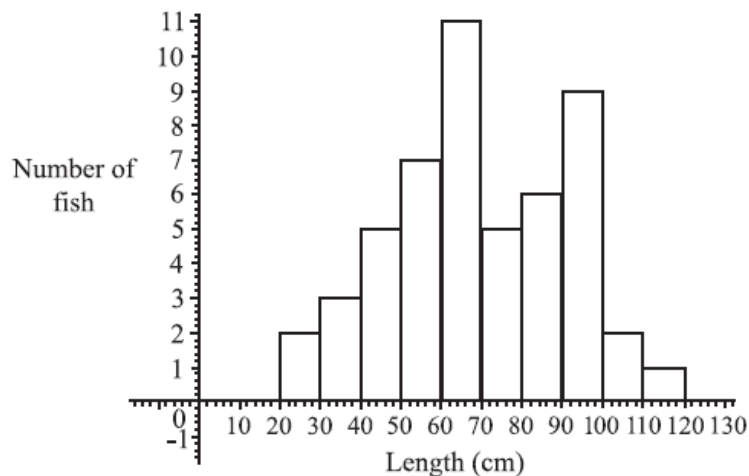
$\text{IQR} = 6 - 4$  (M1)

**Note:** Award (M1) for identifying correct quartiles.

$\text{IQR} = 2$  (A1)(ft) (C2)

[2 marks]

The figure below shows the lengths in centimetres of fish found in the net of a small trawler.



- 7a. Find the total number of fish in the net.

[2 marks]

## Markscheme

Total =  $2 + 3 + 5 + 7 + 11 + 5 + 6 + 9 + 2 + 1$  (MI)

(MI) is for a sum of frequencies.

= 51 (AI)(G2)

[2 marks]

- 7b. Find (i) the modal length interval,  
(ii) the interval containing the median length,  
(iii) an estimate of the mean length.

[5 marks]

## Markscheme

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

- (i) modal interval is 60 – 70

Award (A0) for 65 (AI)

- (ii) median is length of fish no. 26, (MI)(AI)

also 60 – 70 (G2)

Can award (AI)(ft) or (G2)(ft) for 65 if (A0) was awarded for 65 in part (i).

- (iii) mean is  $\frac{2 \times 25 + 3 \times 35 + 5 \times 45 + 7 \times 55 + \dots}{51}$  (MI)

(UP) = 69.5 cm (3sf) (AI)(ft)(G1)

Note: (MI) is for a sum of (frequencies multiplied by midpoint values) divided by candidate's answer from part (a). Accept mid-points 25.5, 35.5 etc or 24.5, 34.5 etc, leading to answers 70.0 or 69.0 (3sf) respectively. Answers of 69.0, 69.5 or 70.0 (3sf) with no working can be awarded (G1).

[5 marks]

- 7c. (i) Write down an estimate for the standard deviation of the lengths.  
(ii) How many fish (if any) have length **greater than** three standard deviations **above** the mean?

[3 marks]

## Markscheme

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

(**UP**) (i) standard deviation is 21.8 cm (**G1**)

For any other answer without working, award (**G0**). If working is present then (**G0**)(**AP**) is possible.

(ii)  $69.5 + 3 \times 21.8 = 134.9 > 120$  (**MI**)

no fish (**A1**)(**ft**)(**G1**)

For 'no fish' without working, award (**G1**) regardless of answer to (c)(i). Follow through from (c)(i) only if method is shown.

[3 marks]

- 7d. The fishing company must pay a fine if more than 10% of the catch have lengths less than 40cm.

[2 marks]

Do a calculation to decide whether the company is fined.

## Markscheme

5 fish are less than 40 cm in length, (**MI**)

Award (**MI**) for any of  $\frac{5}{51}$ ,  $\frac{46}{51}$ , 0.098 or 9.8%, 0.902, 90.2% or 5.1 seen.

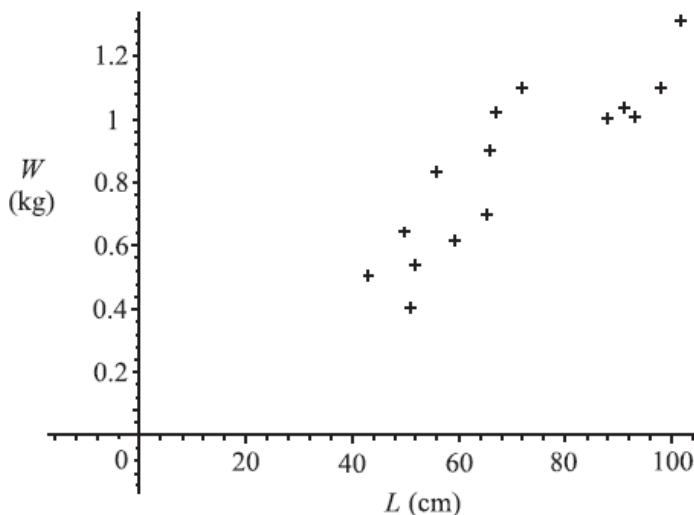
hence no fine. (**A1**)(**ft**)

Note: There is no G mark here and (**M0**)(**A1**) is never allowed. The follow-through is from answer in part (a).

[2 marks]

- 7e. A sample of 15 of the fish was weighed. The weight,  $W$  was plotted against length,  $L$  as shown below.

[2 marks]



Exactly **two** of the following statements about the plot could be correct. Identify the two correct statements.

**Note:** You do **not** need to enter data in a GDC **or** to calculate  $r$  exactly.

- (i) The value of  $r$ , the correlation coefficient, is approximately 0.871.
- (ii) There is an exact linear relation between  $W$  and  $L$ .
- (iii) The line of regression of  $W$  on  $L$  has equation  $W = 0.012L + 0.008$ .
- (iv) There is negative correlation between the length and weight.
- (v) The value of  $r$ , the correlation coefficient, is approximately 0.998.
- (vi) The line of regression of  $W$  on  $L$  has equation  $W = 63.5L + 16.5$ .

## Markscheme

(i) and (iii) are correct. (AI)(AI)

[2 marks]