SL Paper 2

As part of his IB Biology field work, Barry was asked to measure the circumference of trees, in centimetres, that were growing at different distances, in metres, from a river bank. His results are summarized in the following table.

Distance, x (metres)	5	12	17	21	24	30	34	44	47
Circumference, y (centimetres)	82	76	70	68	67	60	62	50	50

a. State whether *distance from the river bank* is a continuous **or** discrete variable. [1]

b. **On graph paper,** draw a scatter diagram to show Barry's results. Use a scale of 1 cm to represent 5 m on the *x*-axis and 1 cm to represent 10 cm on the *y*-axis. [4]

c. Write down [2]

(i) the mean distance, $\bar{x}$, of the trees from the river bank;

(ii) the mean circumference, $\bar{y}$, of the trees.

d. Plot and label the point $\mathrm{M}(\bar{x},\ \bar{y})$ on your graph. [2]

e. Write down [4]

(i) the Pearson's product–moment correlation coefficient, r, for Barry's results;

(ii) the equation of the regression line y on x, for Barry's results.

f. Draw the regression line y on x on your graph. [2]

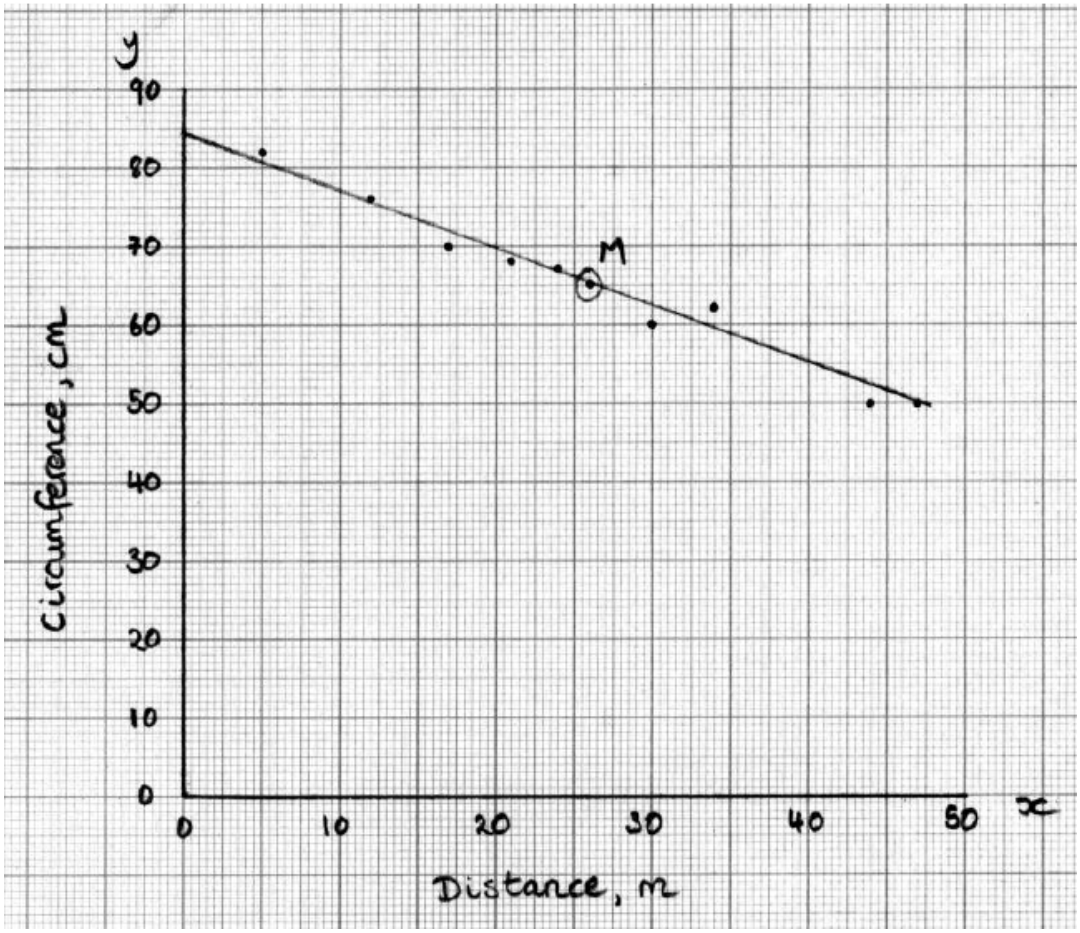
g. **Use the equation of the regression line** y on x to estimate the circumference of a tree that is 40 m from the river bank. [2]

Markscheme

a. continuous ***(A1)***

[1 mark]

b.



(A1)(A1)(A1)(A1)

Notes: Award ***(A1)*** for labelled axes and correct scales; if axes are reversed award ***(A0)*** and follow through for their points. Award ***(A1)*** for at least 3 correct points, ***(A2)*** for at least 6 correct points, ***(A3)*** for all 9 correct points. If scales are too small or graph paper has not been used, accuracy cannot be determined; award ***(A0)***. Do not penalize if extra points are seen.

[4 marks]

c. (i) 26 (m) ***(A1)***

(ii) 65 (cm) ***(A1)***

[2 marks]

d. point M labelled, in correct position ***(A1)(A1)*****(ft)**

Notes: Award ***(A1)*****(ft)** for point plotted in correct position, ***(A1)*** for point labelled M or $(\bar{x},\ \bar{y})$. Follow through from their answers to part (c).

[2 marks]

e. (i) $-0.988\ \ (-0.988432\ldots)$ ***(G2)***

Note: Award ***(G2)*** for -0.99. Award ***(G1)*** for -0.990.

Award ***(A1)(A0)*** if minus sign is omitted.

(ii) $y = -0.756x + 84.7 \quad (y = -0.756281\ldots x + 84.6633\ldots)$ ***(G2)***

Notes: Award ***(A1)*** for $-0.756x$, ***(A1)*** for 84.7. If the answer is not given as an equation, award a maximum of ***(A1)(A0)***.

[4 marks]

f. regression **line** through their M ***(A1)*****(ft)**

regression **line** through their $(0, 85)$ (accept 85 ± 1) ***(A1)*****(ft)**

Notes: Follow through from part (d). Award a maximum of ***(A1)(A0)*** if the line is not straight. Do not penalize if either the line does not meet the *y*-axis or extends into quadrants other than the first.

If M is not plotted or labelled, then follow through from part (c).

Follow through from their *y*-intercept in part (e)(ii).

[2 marks]

g. $-0.756281(40) + 84.6633$ ***(M1)***

$= 54.4 \text{ (cm)} \ (54.4120\ldots)$ ***(A1)*****(ft)*****(G2)***

Notes: Accept 54.5 (54.46) for use of 3 sf. Accept 54.3 from use of -0.76 and 84.7.

Follow through from their equation in part (e)(ii) **irrespective of working shown**; the final answer seen must be consistent with that equation for the final ***(A1)*** to be awarded.

Do not accept answers taken from the graph.

[2 marks]

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]
d. [N/A]
e. [N/A]
f. [N/A]
g. [N/A]

The table below shows the distribution of test grades for 50 IB students at Greendale School.

Test grade	1	2	3	4	5	6	7
Frequency	1	3	7	13	11	10	5

A student is chosen at random from these 50 students.

A second student is chosen at random from these 50 students.

The number of minutes that the 50 students spent preparing for the test was normally distributed with a mean of 105 minutes and a standard deviation of 20 minutes.

a.i. Calculate the mean test grade of the students; [2]

a.ii. Calculate the standard deviation. [1]

b. Find the median test grade of the students. [1]

c. Find the interquartile range. [2]

d. Find the probability that this student scored a grade 5 or higher. [2]

e. Given that the first student chosen at random scored a grade 5 or higher, find the probability that both students scored a grade 6. [3]

f.i. Calculate the probability that a student chosen at random spent at least 90 minutes preparing for the test. [2]

f.ii. Calculate the expected number of students that spent at least 90 minutes preparing for the test. [2]

Markscheme

a.i. $\frac{1(1)+3(2)+7(3)+13(4)+11(5)+10(6)+5(7)}{50} = \frac{230}{50}$ ***(M1)***

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

$= 4.6$ ***(A1)*** ***(G2)***

[2 marks]

a.ii. $1.46\ (1.45602\ldots)$ ***(G1)***

[1 mark]

b. 5 ***(A1)***

[1 mark]

c. $6 - 4$ ***(M1)***

Note: Award ***(M1)*** for 6 and 4 seen.

$= 2$ ***(A1)*** ***(G2)***

[2 marks]

d. $\frac{11+10+5}{50}$ ***(M1)***

Note: Award ***(M1)*** for $11 + 10 + 5$ seen.

$= \frac{26}{50}\ \left(\frac{13}{25},\ 0.52,\ 52\%\right)$ ***(A1)*** ***(G2)***

[2 marks]

e. $\frac{10}{\text{their } 26} \times \frac{9}{49}$ ***(M1)(M1)***

Note: Award ***(M1)*** for $\frac{10}{\text{their } 26}$ seen, ***(M1)*** for multiplying their first probability by $\frac{9}{49}$.

OR

$\frac{\frac{10}{50} \times \frac{9}{49}}{\frac{26}{50}}$

Note: Award ***(M1)*** for $\frac{10}{50} \times \frac{9}{49}$ seen, ***(M1)*** for dividing their first probability by $\frac{\text{their } 26}{50}$.

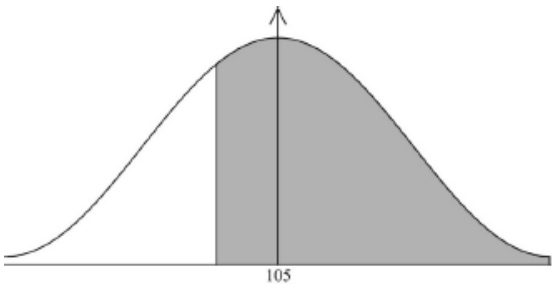
$= \frac{45}{637}\ (0.0706,\ 0.0706436\ldots,\ 7.06436\ldots\%)$ ***(A1)*****(ft)** ***(G3)***

Note: Follow through from part (d).

[3 marks]

f.i. $\mathrm{P}(X \geqslant 90)$ ***(M1)***

OR



(M1)

Note: Award ***(M1)*** for a diagram showing the correct shaded region (> 0.5).

$0.773\ (0.773372\ldots)\ 0.773\ (0.773372\ldots,\ 77.3372\ldots\%)$ ***(A1)*** ***(G2)***

[2 marks]

f.ii. $0.773372\ldots \times 50$ ***(M1)***

$= 38.7\ (38.6686\ldots)$ ***(A1)*****(ft)** ***(G2)***

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

[2 marks]

Examiners report

a.i. [N/A]
a.ii. [N/A]
b. [N/A]
c. [N/A]
d. [N/A]
e. [N/A]
f.i. [N/A]
f.ii. [N/A]

In the month before their IB Diploma examinations, eight male students recorded the number of hours they spent on social media.

For each student, the number of hours spent on social media (x) and the number of IB Diploma points obtained (y) are shown in the following table.

Hours on social media (x)	6	15	26	12	13	40	33	23
IB Diploma points (y)	43	33	27	36	39	17	20	33

Use your graphic display calculator to find

Ten female students also recorded the number of hours they spent on social media in the month before their IB Diploma examinations. Each of these female students spent between 3 and 30 hours on social media.

The equation of the regression line *y* on *x* for these ten female students is

$$y = -\frac{2}{3}x + \frac{125}{3}.$$

An eleventh girl spent 34 hours on social media in the month before her IB Diploma examinations.

a. On graph paper, draw a scatter diagram for these data. Use a scale of 2 cm to represent 5 hours on the x-axis and 2 cm to represent 10 points on the y-axis. [4]

b. (i) $\bar{x}$, the mean number of hours spent on social media; [2]

(ii) $\bar{y}$, the mean number of IB Diploma points.

c. Plot the point $(\bar{x},\ \bar{y})$ on your scatter diagram and label this point M. [2]

d. Write down the value of r, the Pearson's product–moment correlation coefficient, for these data. [2]

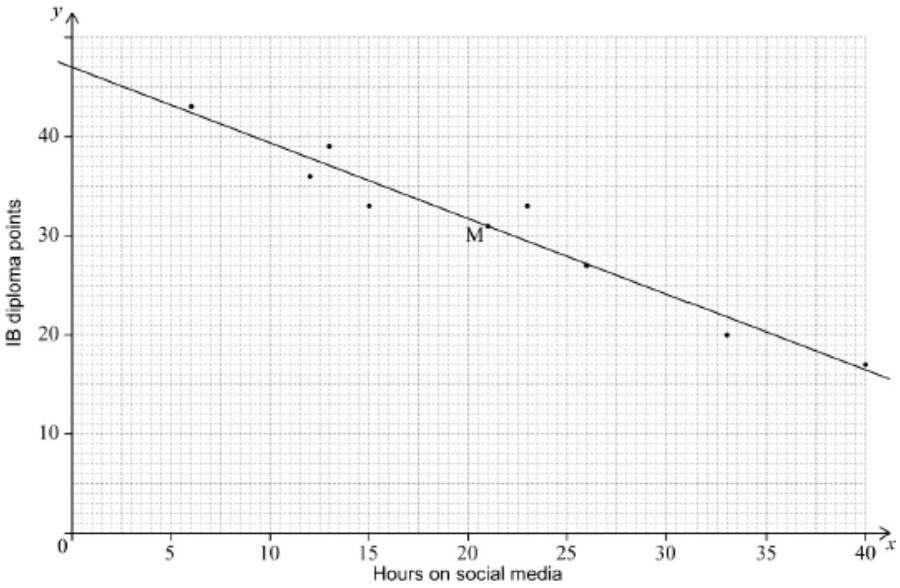
e. Write down the equation of the regression line y on x for these eight male students. [2]

f. Draw the regression line, from part (e), on your scatter diagram. [2]

g. Use the given equation of the regression line to estimate the number of IB Diploma points that this girl obtained. [2]

h. Write down a reason why this estimate is not reliable. [1]

Markscheme

a. 

(A4)

Notes: Award ***(A1)*** for correct scale and labelled axes.

Award ***(A3)*** for 7 or 8 points correctly plotted,

(A2) for 5 or 6 points correctly plotted,

(A1) for 3 or 4 points correctly plotted.

Award at most ***(A0)(A3)*** if axes reversed.

Accept x and y sufficient for labelling.

If graph paper is not used, award ***(A0)***.

If an inconsistent scale is used, award ***(A0)***. Candidates' points should be read from this scale **where possible** and awarded accordingly.

A scale which is too small to be meaningful (ie mm instead of cm) earns ***(A0)*** for plotted points.

[4 marks]

b. (i) $\bar{x} = 21$ ***(A1)***

(ii) $\bar{y} = 31$ ***(A1)***

[2 marks]

c. $(\bar{x},\ \bar{y})$ correctly plotted on graph ***(A1)*****(ft)**

this point labelled M ***(A1)***

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

Only accept M for labelling.

[2 marks]

d. $-0.973\ (-0.973388\ldots)$ ***(G2)***

Note: Award ***(G1)*** for 0.973, without minus sign.

[2 marks]

e. $y = -0.761x + 47.0$ $(y = -0.760638\ldots x + 46.9734\ldots)$ ***(A1)(A1)(G2)***

Notes: Award ***(A1)*** for $-0.761x$ and ***(A1)*** $+47.0$. Award a maximum of ***(A1)(A0)*** if answer is not an equation.

[2 marks]

f. line on graph ***(A1)*****(ft)*****(A1)*****(ft)**

Notes: Award ***(A1)*****(ft)** for **straight line** that passes through their M, ***(A1)*****(ft)** for line (extrapolated if necessary) that passes through $(0,\ 47.0)$.

If M is not plotted or labelled, follow through from part (e).

[2 marks]

g. $y = -\frac{2}{3}(34) + \frac{125}{3}$ ***(M1)***

Note: Award ***(M1)*** for correct substitution.

19 (points) ***(A1)(G2)***

[2 marks]

h. extrapolation ***(R1)***

OR

34 hours is outside the given range of data ***(R1)***

Note: Do not accept 'outlier'.

[1 mark]

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]
d. [N/A]
e. [N/A]
f. [N/A]
g. [N/A]
h. [N/A]

A survey was conducted to determine the length of time, t, in minutes, people took to drink their coffee in a café. The information is shown in the following grouped frequency table.

Time, t (minutes)	Number of People
$0 < t \leq 5$	3
$5 < t \leq 10$	5
$10 < t \leq 15$	12
$15 < t \leq 20$	14
$20 < t \leq 25$	16
$25 < t \leq 30$	10

a. Write down the total number of people who were surveyed. [1]

b. Write down the mid-interval value for the $10 < t \leqslant 15$ group. [1]

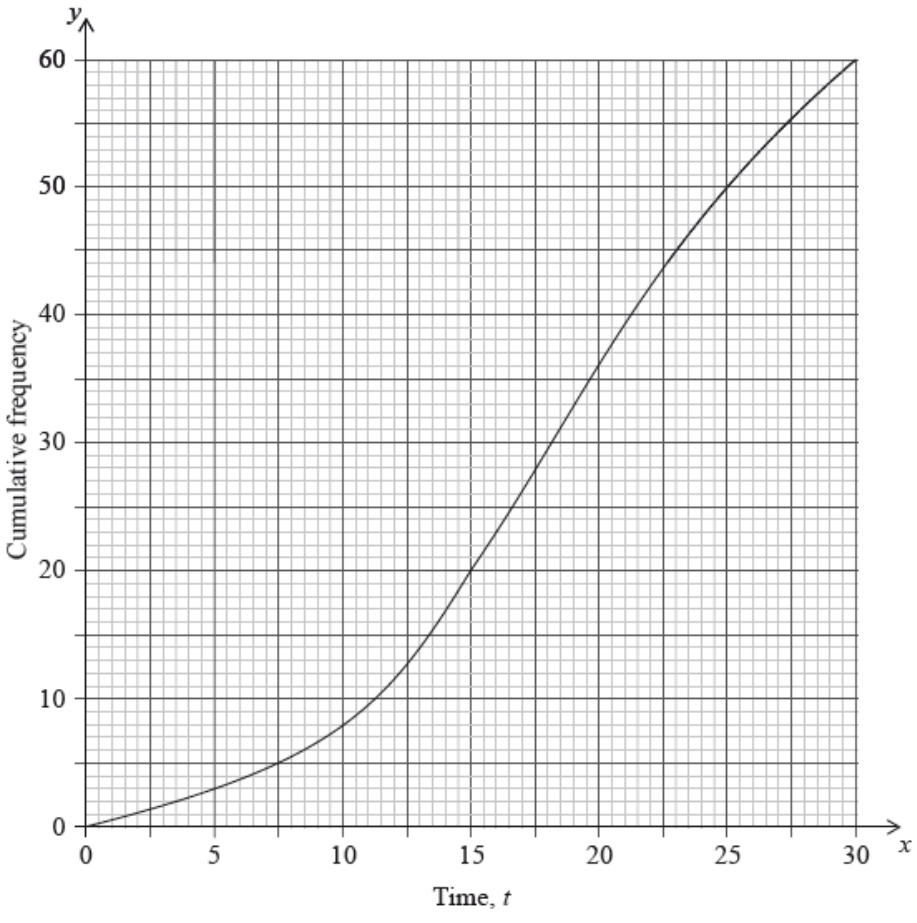
c. Find an estimate of the mean time people took to drink their coffee. [2]

d. The information above has been rewritten as a cumulative frequency table. [2]

Time, t (minutes)	$t \leq 5$	$t \leq 10$	$t \leq 15$	$t \leq 20$	$t \leq 25$	$t \leq 30$
Cumulative frequency	3	8	20	a	50	b

Write down the value of a and the value of b.

e. This information is shown in the following cumulative frequency graph. [4]



For the people who were surveyed, use the graph to estimate

(i) the time taken for the first 40 people to drink their coffee;

(ii) the number of people who take less than 8 minutes to drink their coffee;

(iii) the number of people who take more than 23 minutes to drink their coffee.

Markscheme

a. 60 ***(A1)***

[1 mark]

b. 12.5 ***(A1)***

[1 mark]

c. $\frac{3\times2.5+5\times7.5+\ldots+10\times27.5}{60}$ ***(M1)***

Note: 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.

$= \frac{1075}{60}\ \left(\frac{215}{12},\ 17.9, 17.9166\ldots\right)$ ***(A1)*****(ft)*****(G2)***

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

[2 marks]

d. $a = 34,\ b = 60$ ***(A1)(A1)***

[2 marks]

e. (i) $\leqslant 21.25\ \text{minutes}$ ***(A1)***

Note: Accept 21.25.

Accept any answer between 21 and 21.5.

(Accept 21.5, but do not accept 21.)

(ii) 5 ***(A1)***

Note: Accept < 6. Do not accept 6.

Answer must be an integer.

(iii) $60 - 45$ ***(M1)***

$= 15$ ***(A1)(G2)***

Notes: Award ***(M1)*** for subtraction from 60. Accept 15 ± 1.

Answer must be an integer.

[4 marks]

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]
d. [N/A]
e. [N/A]

A group of 800 students answered 40 questions on a category of their choice out of History, Science and Literature.

For each student the category and the number of correct answers, N, was recorded. The results obtained are represented in the following table.

		Number of correct answers, N				
		$1 \leq N \leq 10$	$11 \leq N \leq 20$	$21 \leq N \leq 30$	$31 \leq N \leq 40$	**Total number of students**
Category	History	46	80	68	39	**233**
	Science	37	82	85	56	**260**
	Literature	31	110	104	62	**307**
	Total number of students	**114**	**272**	**257**	**157**	**800**

A χ^2 test at the 5% significance level is carried out on the results. The critical value for this test is 12.592.

a. State whether N is a discrete or a continuous variable. [1]

b.i. Write down, for N, the modal class; [1]

b.ii. Write down, for N, the mid-interval value of the modal class. [1]

c.i. Use your graphic display calculator to estimate the mean of N; [2]

c.ii. Use your graphic display calculator to estimate the standard deviation of N. [1]

d. Find the expected frequency of students choosing the Science category and obtaining 31 to 40 correct answers. [2]

e.i. Write down the null hypothesis for this test; [1]

e.ii. Write down the number of degrees of freedom. [1]

f.i. Write down the p-value for the test; [1]

f.ii. Write down the χ^2 statistic. [2]

g. State the result of the test. Give a reason for your answer. [2]

Markscheme

a. discrete ***(A1)***

[1 mark]

b.i. $11 \leqslant N \leqslant 20$ ***(A1)***

[1 mark]

b.ii. 15.5 ***(A1)*****(ft)**

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

[1 mark]

c.i. $21.2\ (21.2125)$ ***(G2)***

[2 marks]

c.ii. $9.60\ (9.60428\ldots)$ ***(G1)***

[1 marks]

d. $\frac{260}{800} \times \frac{157}{800} \times 800$ **OR** $\frac{260 \times 157}{800}$ ***(M1)***

Note: Award ***(M1)*** for correct substitution into expected frequency formula.

$= 51.0\ (51.025)$ ***(A1)(G2)***

[2 marks]

e.i. choice of category and number of correct answers are independent ***(A1)***

Notes: Accept “no association” between (choice of) category and number of correct answers. Do not accept “not related” or “not correlated” or “influenced”.

[1 mark]

e.ii. 6 ***(A1)***

[1 mark]

f.i. $0.0644\ (0.0644123\ldots)$ ***(G1)***

[1 mark]

f.ii. $11.9\ (11.8924\ldots)$ ***(G2)***

[2 marks]

g. the null hypothesis is not rejected (the null hypothesis is accepted) ***(A1)*****(ft)**

OR

(choice of) category and number of correct answers are independent ***(A1)*****(ft)**

as $11.9 < 12.592$ **OR** $0.0644 > 0.05$ ***(R1)***

Notes: Award ***(R1)*** for a correct comparison of either their χ^2 statistic to the χ^2 critical value or their p-value to the significance level. Award ***(A1)*****(ft)** from that comparison.

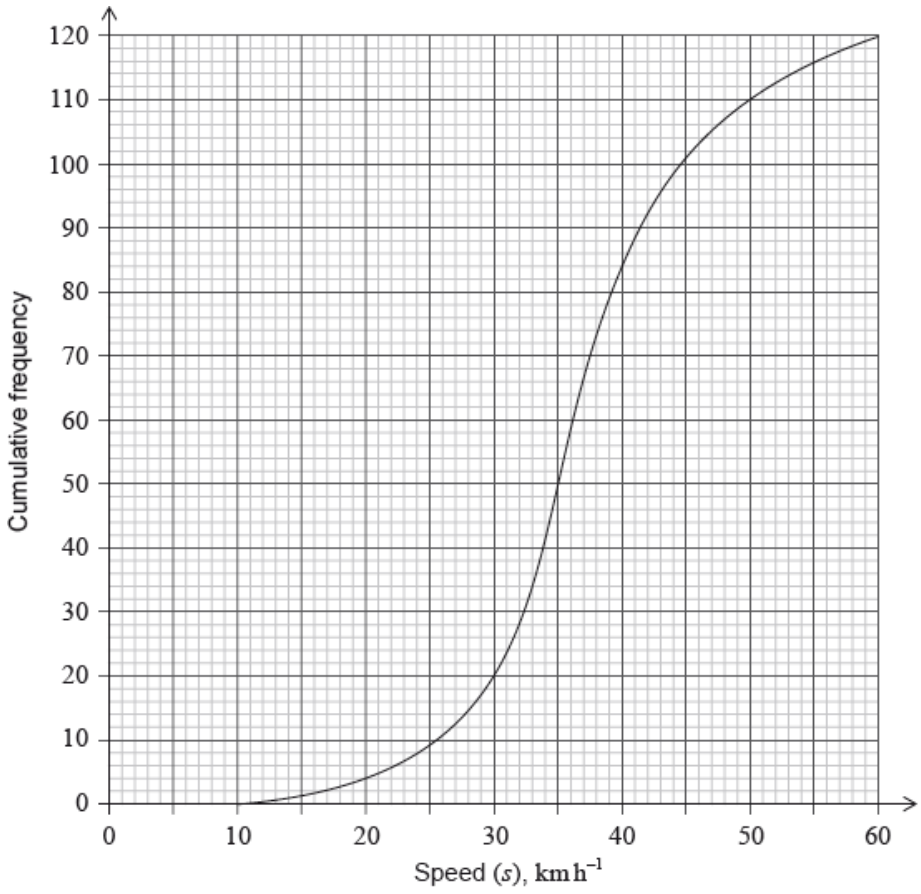
Follow through from part (f). Do not award ***(A1)*****(ft)*****(R0)***.

[2 marks]

Examiners report

a. [N/A]
b.i. [N/A]
b.ii. [N/A]
c.i. [N/A]
c.ii. [N/A]
d. [N/A]
e.i. [N/A]
e.ii. [N/A]
f.i. [N/A]
f.ii. [N/A]
g. [N/A]

The cumulative frequency graph shows the speed, s, in $\mathrm{km\,h^{-1}}$, of 120 vehicles passing a hospital gate.



a. Estimate the minimum possible speed of one of these vehicles passing the hospital gate. [1]

b. Find the median speed of the vehicles. [2]

c. Write down the 75^{th} percentile. [1]

d. Calculate the interquartile range. [2]

e. The speed limit past the hospital gate is $50\ \text{km}\,\text{h}^{-1}$. [2]

Find the number of these vehicles that exceed the speed limit.

f. The table shows the speeds of these vehicles travelling past the hospital gate. [2]

Speed of Vehicles	Number of Vehicles
$0 < s \leq 10$	0
$10 < s \leq 20$	p
$20 < s \leq 30$	16
$30 < s \leq 40$	64
$40 < s \leq 50$	26
$50 < s \leq 60$	q

Find the value of p and of q.

g. The table shows the speeds of these vehicles travelling past the hospital gate. [2]

Speed of Vehicles	Number of Vehicles
$0 < s \leq 10$	0
$10 < s \leq 20$	p
$20 < s \leq 30$	16
$30 < s \leq 40$	64
$40 < s \leq 50$	26
$50 < s \leq 60$	q

(i) Write down the modal class.

(ii) Write down the mid-interval value for this class.

h. The table shows the speeds of these vehicles travelling past the hospital gate. [3]

Speed of Vehicles	Number of Vehicles
$0 < s \leq 10$	0
$10 < s \leq 20$	p
$20 < s \leq 30$	16
$30 < s \leq 40$	64
$40 < s \leq 50$	26
$50 < s \leq 60$	q

Use your graphic display calculator to calculate an estimate of

(i) the mean speed of these vehicles;

(ii) the standard deviation.

i. It is proposed that the speed limit past the hospital gate is reduced to $40\ \text{km h}^{-1}$ from the current $50\ \text{km h}^{-1}$. [2]

Find the percentage of these vehicles passing the hospital gate that **do not** exceed the current speed limit but **would** exceed the new speed limit.

Markscheme

a. $10\ (\text{km h}^{-1})$ ***(A1)***

b. 36 ***(G2)***

c. 41.5 ***(G1)***

d. $41.5 - 32.5$ ***(M1)***

$= 9\ (\pm 1)$ ***(A1)*****(ft)*****(G2)***

Notes: Award ***(M1)*** for quartiles seen. Follow through from part (c).

e. $120 - 110$ ***(M1)***

$= 10$ ***(A1)(G2)***

Note: Award ***(M1)*** for 110 seen.

f. $p = 4 \quad q = 10$ ***(A1)*****(ft)*****(A1)*****(ft)**

Note: Follow through from part (e).

g. (i) $30 < s \leqslant 40$ ***(A1)***

(ii) 35 ***(A1)*****(ft)**

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

h. (i) $36.8\ (\text{km h}^{-1}) \quad (36.8333)$ ***(G2)*****(ft)**

Notes: Follow through from part (f).

(ii) $8.85 \quad (8.84904\ldots)$ ***(G1)*****(ft)**

Note: Follow through from part (f), irrespective of working seen.

i. $\frac{26}{120} \times 100$ ***(M1)***

Note: Award ***(M1)*** for $\frac{26}{120} \times 100$ seen.

$= 21.7\ (\%) \quad \left(21.6666\ldots,\ 21\frac{2}{3},\ \frac{65}{3}\right)$ ***(A1)(G2)***

Examiners report

a. For the great majority, this was a straightforward and accessible question. There were many, however, who had no appreciation of medians, percentiles and quartiles – all straightforward concepts. Most were able to read from the graph, using correctly the scales; only the weakest misinterpreting these. Calculation of the mean and standard deviation are expected to be completed using the graphic display calculator (GDC) – formulae are no longer required and the covariance will **not** be given in questions. Many candidates, however, were unable to calculate the mean and standard deviation of a (grouped) frequency distribution, instead treating the data as raw; comments on the G2 forms from schools indicated that some teachers were also unable to do this and advice must be sought.

b. For the great majority, this was a straightforward and accessible question. There were many, however, who had no appreciation of medians, percentiles and quartiles – all straightforward concepts. Most were able to read from the graph, using correctly the scales; only the weakest misinterpreting these. Calculation of the mean and standard deviation are expected to be completed using the graphic display calculator (GDC) – formulae are no longer required and the covariance will **not** be given in questions. Many candidates, however, were unable to calculate the mean and standard deviation of a (grouped) frequency distribution, instead treating the data as raw; comments on the G2 forms from schools indicated that some teachers were also unable to do this and advice must be sought.

c. For the great majority, this was a straightforward and accessible question. There were many, however, who had no appreciation of medians, percentiles and quartiles – all straightforward concepts. Most were able to read from the graph, using correctly the scales; only the weakest misinterpreting these. Calculation of the mean and standard deviation are expected to be completed using the graphic display calculator (GDC) – formulae are no longer required and the covariance will **not** be given in questions. Many candidates, however, were unable to calculate the mean and standard deviation of a (grouped) frequency distribution, instead treating the data as raw; comments on the G2 forms from schools indicated that some teachers were also unable to do this and advice must be sought.

d. For the great majority, this was a straightforward and accessible question. There were many, however, who had no appreciation of medians, percentiles and quartiles – all straightforward concepts. Most were able to read from the graph, using correctly the scales; only the weakest misinterpreting these. Calculation of the mean and standard deviation are expected to be completed using the graphic display calculator (GDC) – formulae are no longer required and the covariance will **not** be given in questions. Many candidates, however, were unable to calculate the mean and standard deviation of a (grouped) frequency distribution, instead treating the data as raw; comments on the G2 forms from schools indicated that some teachers were also unable to do this and advice must be sought.

e. For the great majority, this was a straightforward and accessible question. There were many, however, who had no appreciation of medians, percentiles and quartiles – all straightforward concepts. Most were able to read from the graph, using correctly the scales; only the weakest misinterpreting these. Calculation of the mean and standard deviation are expected to be completed using the graphic display calculator (GDC) – formulae are no longer required and the covariance will **not** be given in questions. Many candidates, however, were unable to calculate the mean and standard deviation of a (grouped) frequency distribution, instead treating the data as raw; comments on the G2 forms from schools indicated that some teachers were also unable to do this and advice must be sought.

f. For the great majority, this was a straightforward and accessible question. There were many, however, who had no appreciation of medians, percentiles and quartiles – all straightforward concepts. Most were able to read from the graph, using correctly the scales; only the weakest misinterpreting these. Calculation of the mean and standard deviation are expected to be completed using the graphic display calculator (GDC) – formulae are no longer required and the covariance will **not** be given in questions. Many candidates, however, were unable to calculate the mean and standard deviation of a (grouped) frequency distribution, instead treating the data as raw; comments on the G2 forms from schools indicated that some teachers were also unable to do this and advice must be sought.

g. For the great majority, this was a straightforward and accessible question. There were many, however, who had no appreciation of medians, percentiles and quartiles – all straightforward concepts. Most were able to read from the graph, using correctly the scales; only the weakest misinterpreting these. Calculation of the mean and standard deviation are expected to be completed using the graphic display calculator (GDC) – formulae are no longer required and the covariance will **not** be given in questions. Many candidates, however, were unable to calculate the mean and standard deviation of a (grouped) frequency distribution, instead treating the data as raw; comments on the G2 forms from schools indicated that some teachers were also unable to do this and advice must be sought.

h. For the great majority, this was a straightforward and accessible question. There were many, however, who had no appreciation of medians, percentiles and quartiles – all straightforward concepts. Most were able to read from the graph, using correctly the scales; only the weakest misinterpreting these. Calculation of the mean and standard deviation are expected to be completed using the graphic display calculator (GDC) – formulae are no longer required and the covariance will **not** be given in questions. Many candidates, however, were unable to calculate the mean and standard deviation of a (grouped) frequency distribution, instead treating the data as raw; comments on the G2 forms from schools indicated that some teachers were also unable to do this and advice must be sought.

i. For the great majority, this was a straightforward and accessible question. There were many, however, who had no appreciation of medians, percentiles and quartiles – all straightforward concepts. Most were able to read from the graph, using correctly the scales; only the weakest misinterpreting these. Calculation of the mean and standard deviation are expected to be completed using the graphic display calculator (GDC) – formulae are no longer required and the covariance will **not** be given in questions. Many candidates, however, were unable to calculate the mean and standard deviation of a (grouped) frequency distribution, instead treating the data as raw; comments on the G2 forms from schools indicated that some teachers were also unable to do this and advice must be sought.

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

a. (i) Write down the mean score in Round 1. [5]

(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.

b. Write down the correlation coefficient, *r*. [2]

c. Write down the equation of the regression line of *y* on *x*. [2]

d. Another golfer scored 70 in Round 1. [2]

Calculate an estimate of his score in Round 2.

e. Another golfer scored 89 in Round 1. [2]

Determine whether you can use the equation of the regression line to estimate his score in Round 2. Give a reason for your answer.

Markscheme

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

$72.4\left(72.4166..., \frac{869}{12}\right)$ ***(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]

b. 0.990 (0.99014…) ***(G2)***

[2 marks]

c. $y = 1.01x + 0.816$ ($y = 1.01404...x + 0.81618...$) ***(G1)(G1)***

Notes: Award ***(G1)*** for $1.01x$ and ***(G1)*** for 0.816. If the answer is not an equation award a maximum of ***(G1)(G0)***.

OR

$y - 74.25 = 1.01(x - 72.4)$($y - 74.25 = 1.01404...(x - 72.4166...)$) ***(A1)(A1)***

Notes: Award ***(A1)*** for 1.01 correctly substituted in the equation, and ***(A1)*****(ft)** for correct substitution of (72.4, 74.25) in the equation. Follow through from their part (a)(i). If the final answer is not an equation award a maximum of ***(A1)(A0)***.

[2 marks]

d. $y = 1.01404... \times 70 + 0.81618...$ ***(M1)***

Note: Award ***(M1)*** for substitution of 70 into their regression line equation from part (c).

$y = 72$ (71.7989...) ***(A1)*****(ft)*****(G2)***

Note: Follow through from their part (c).

[2 marks]

e. No, equation cannot be (reliably) used as 89 is outside the data range. ***(A1)(R1)***

OR

Yes, but the result is not valid/not reliable as 89 is outside the data range/as we extrapolate ***(A1)(R1)***

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

[2 marks]

Examiners report

a. 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.

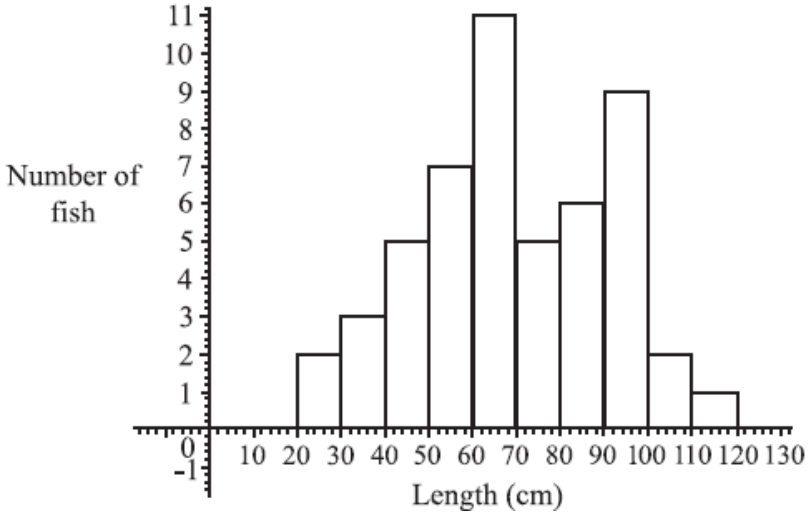
b. The question was for the most part approached by almost all candidates and answered relatively well. The question in part (e) related to the use of the equation of the regression line for predicting, although regularly asked on exams, was still found to be a difficult one by some candidates. Some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

c. 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.

d. 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.

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



a. Find the total number of fish in the net. [2]

b. Find (i) the modal length interval, [5]

(ii) the interval containing the median length,

(iii) an estimate of the mean length.

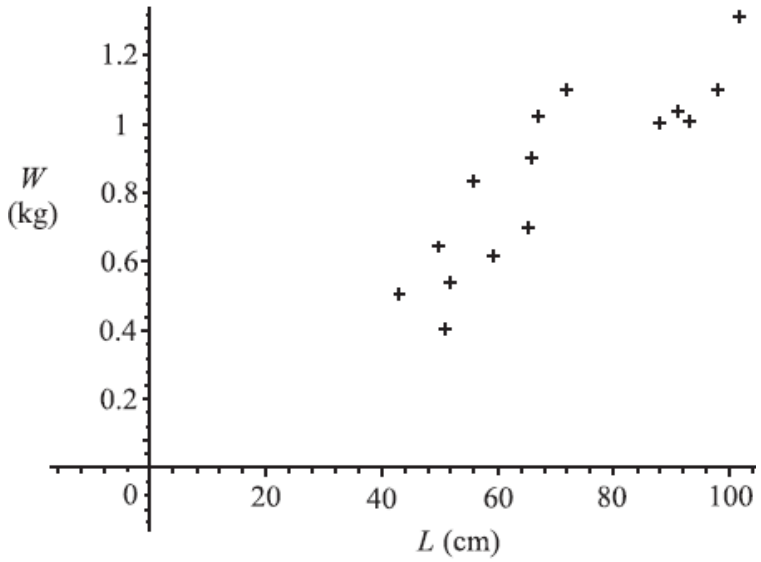
c. (i) Write down an estimate for the standard deviation of the lengths. [3]

(ii) How many fish (if any) have length **greater than** three standard deviations **above** the mean?

d. The fishing company must pay a fine if more than 10% of the catch have lengths less than 40cm. [2]

Do a calculation to decide whether the company is fined.

e. A sample of 15 of the fish was weighed. The weight, W was plotted against length, L as shown below. [2]



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

a. Total = 2 + 3 + 5 + 7 + 11 + 5 + 6 + 9 + 2 + 1 ***(M1)***

(M1) *is for a sum of frequencies.*

= 51 ***(A1)(G2)***

[2 marks]

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

(i) modal interval is 60 – 70

Award ***(A0)*** *for 65* ***(A1)***

(ii) median is length of fish no. 26, ***(M1)(A1)***

also 60 – 70 ***(G2)***

Can award ***(A1)*****(ft)** *or* ***(G2)*****(ft)** *for 65 if* ***(A0)*** *was awarded for 65 in part (i).*

(iii) mean is $\frac{2\times25+3\times35+5\times45+7\times55+...}{51}$ ***(M1)***

(UP) = 69.5 cm (3sf) ***(A1)*****(ft)*****(G1)***

Note: ***(M1)*** *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]

c. *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) $69.5 + 3 \times 21.8 = 134.9 > 120$ ***(M1)***

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]

d. 5 fish are less than 40 cm in length, ***(M1)***

Award ***(M1)*** *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]

e. (i) and (iii) are correct. ***(A1)(A1)***

[2 marks]

Examiners report

a. a) b), c) There was much confusion about how to present the intervals. Often the mid-point only was seen. (eg. 65 instead of 60-70).

Understanding of mode, median and mean was usually good but too many candidates wasted time calculating standard deviation by hand and got it wrong. In c(ii) 'greater than three' caused no problems but 'above the mean' was often ignored.

b. a) b), c) There was much confusion about how to present the intervals. Often the mid-point only was seen. (eg. 65 instead of 60-70).

Understanding of mode, median and mean was usually good but too many candidates wasted time calculating standard deviation by hand and got it wrong. In c(ii) 'greater than three' caused no problems but 'above the mean' was often ignored.

c. a) b), c) There was much confusion about how to present the intervals. Often the mid-point only was seen. (eg. 65 instead of 60-70).

Understanding of mode, median and mean was usually good but too many candidates wasted time calculating standard deviation by hand and got it wrong. In c(ii) 'greater than three' caused no problems but 'above the mean' was often ignored.

d. d) This was often well done, even if earlier parts were poorly done.

e. e) Rather mixed performance here. It was hard to identify any consistency in the errors made.

Too much time was spent on this question. It was only worth two marks and candidates ought to have realised that it relied on a general pictorial understanding of the concepts, possibly supplemented by a little elementary arithmetic only, to compare (iii) and (vi). With good understanding, many of the options could be ruled out in a few seconds.

Daniel grows apples and chooses at random a sample of 100 apples from his harvest.

He measures the diameters of the apples to the nearest cm. The following table shows the distribution of the diameters.

Diameter (to the nearest cm)	5	6	7	8	9
Frequency	15	27	33	17	8

a. Using your graphic display calculator, write down the value of [3]

(i) the mean of the diameters in this sample;

(ii) the standard deviation of the diameters in this sample.

b. Daniel assumes that the diameters of all of the apples from his harvest are normally distributed with a mean of 7 cm and a standard deviation of [3]

1.2 cm. He classifies the apples according to their diameters as shown in the following table.

Classification	Diameter (cm)
Small	Diameter < 6.5
Medium	6.5 ≤ Diameter < a
Large	Diameter ≥ a

Calculate the percentage of **small** apples in Daniel's harvest.

c. Daniel assumes that the diameters of all of the apples from his harvest are normally distributed with a mean of 7 cm and a standard deviation of 1.2 cm. He classifies the apples according to their diameters as shown in the following table. [2]

Classification	Diameter (cm)
Small	Diameter < 6.5
Medium	6.5 ≤ Diameter < a
Large	Diameter ≥ a

Of the apples harvested, 5% are **large** apples.

Find the value of a.

d. Daniel assumes that the diameters of all of the apples from his harvest are normally distributed with a mean of 7 cm and a standard deviation of 1.2 cm. He classifies the apples according to their diameters as shown in the following table. [2]

Classification	Diameter (cm)
Small	Diameter < 6.5
Medium	6.5 ≤ Diameter < a
Large	Diameter ≥ a

Find the percentage of **medium** apples.

e. Daniel assumes that the diameters of all of the apples from his harvest are normally distributed with a mean of 7 cm and a standard deviation of 1.2 cm. He classifies the apples according to their diameters as shown in the following table. [2]

Classification	Diameter (cm)
Small	Diameter < 6.5
Medium	6.5 ≤ Diameter < a
Large	Diameter ≥ a

This year, Daniel estimates that he will grow $100\,000$ apples.

Estimate the number of **large** apples that Daniel will grow this year.

Markscheme

a. (i) $6.76\ (\text{cm})$ ***(G2)***

Notes: Award ***(M1)*** for an attempt to use the formula for the mean with a least two rows from the table.

(ii) $1.14\ (\text{cm})\quad (1.14122\ldots\ (\text{cm}))$ ***(G1)***

b. $\mathrm{P}(\text{diameter} < 6.5) = 0.338\quad (0.338461)$ ***(M1)(A1)***

Notes: Award ***(M1)*** for attempting to use the normal distribution to find the probability **or** for correct region indicated on labelled diagram. Award ***(A1)*** for correct probability.

$33.8(\%)$ ***(A1)*(ft)*(G3)***

Notes: Award ***(A1)*(ft)** for converting their probability into a percentage.

c. $\mathrm{P}(\text{diameter} \geqslant a) = 0.05$ ***(M1)***

Note: Award ***(M1)*** for attempting to use the normal distribution to find the probability **or** for correct region indicated on labelled diagram.

$a = 8.97\ (\text{cm})\quad (8.97382\ldots)$ ***(A1)(G2)***

d. $100 - (5 + 33.8461\ldots)$ ***(M1)***

Note: Award ***(M1)*** for subtracting "$5+$ their part (b)" from 100 **or** ***(M1)*** for attempting to use the normal distribution to find the probability $\mathrm{P}\left(6.5 \leqslant \text{diameter} < \text{their part (c)}\right)$ **or** for correct region indicated on labelled diagram.

$= 61.2(\%)\quad (61.1538\ldots(\%))$ ***(A1)*(ft)*(G2)***

Notes: Follow through from their answer to part (b). Percentage symbol is not required. Accept $61.1(\%)$ $(61.1209\ldots(\%))$ if 8.97 used.

e. $100\,000 \times 0.05$ ***(M1)***

Note: Award ***(M1)*** for multiplying by 0.05 (or 5%).

$= 5000$ ***(A1)(G2)***

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]
d. [N/A]
e. [N/A]

The table shows the distance, in km, of eight regional railway stations from a city centre terminus and the price, in $, of a return ticket from each regional station to the terminus.

Distance in km (x)	3	15	23	42	56	62	74	93
Price in $ (y)	5	24	43	56	68	74	86	100

a. Draw a scatter diagram for the above data. Use a scale of 1 cm to represent 10 km on the x-axis and 1 cm to represent $\$10$ on the y-axis. [4]

b. Use your graphic display calculator to find [2]

(i) $\bar{x}$, the mean of the distances;

(ii) $\bar{y}$, the mean of the prices.

c. Plot and label the point $\mathrm{M}\ (\bar{x},\ \bar{y})$ on your scatter diagram. [1]

d. Use your graphic display calculator to find [3]

(i) the product–moment correlation coefficient, r ;

(ii) the equation of the regression line y on x.

e. Draw the regression line y on x on your scatter diagram. [2]

f. A ninth regional station is 76 km from the city centre terminus. [3]

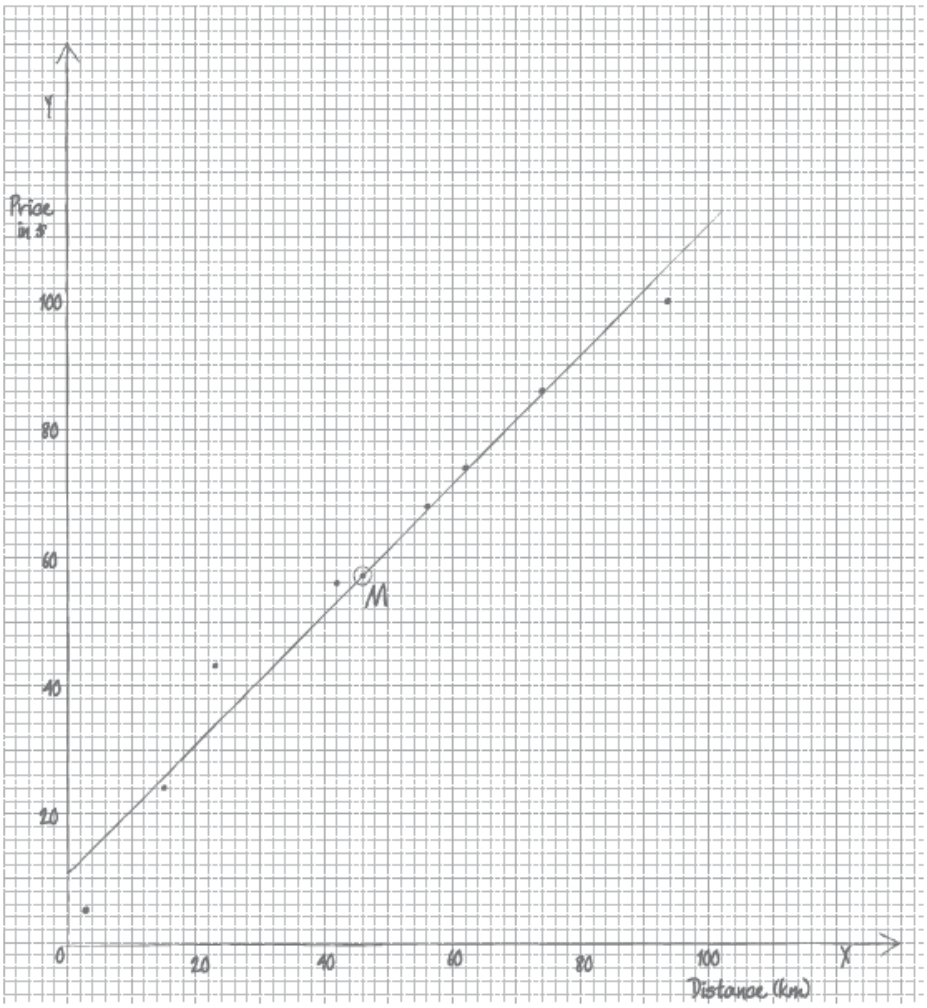
Use the equation of the regression line to estimate the price of a return ticket to the city centre terminus from this regional station. **Give your answer correct to the nearest $\$$.**

g. Give a reason why it is valid to use your regression line to estimate the price of this return ticket. [1]

h. The actual price of the return ticket is $\$80$. [2]

Using your answer to part (f), calculate the percentage error in the estimated price of the ticket.

Markscheme

a. 

(A4)

Notes: Award ***(A1)*** for correct scale and labels (accept x and y).

Award ***(A3)*** for 7 or 8 points plotted correctly.

Award ***(A2)*** for 5 or 6 points plotted correctly.

Award ***(A1)*** for 3 or 4 points plotted correctly.

Award at most ***(A1)(A2)*** if points are joined up.

If axes are reversed, award at most ***(A0)(A3)***.

If graph paper is not used, award at most ***(A1)(A0)***.

[4 marks]

b. (i) $(\bar{x} =) \, 46$ ***(G1)***

(ii) $(\bar{y} =) \, 57$ ***(G1)***

[2 marks]

c. $\mathrm{M}(46, 57)$ plotted and labelled on the scatter diagram ***(A1)*****(ft)**

Notes: Follow through from their part (b).

Accept $(\bar{x}, \, \bar{y})$ as the label.

[1 mark]

d. (i) $0.986 \quad (0.986322...)$ ***(G1)***

(ii) $y = 1.01x + 10.3 \quad (y = 1.01431\ldots x + 10.3412\ldots)$ ***(G1)(G1)***

Notes: Award ***(G1)*** for $1.01x$, ***(G1)*** for 10.3.

Award ***(G1)(G0)*** if not written in the form of an equation.

OR

$(y - 57) = 1.01(x - 46) \quad (y - 57 = 1.01431...(x - 46))$ ***(G1)(G1)*****(ft)**

Note: Award ***(G1)*** for 1.01, ***(G1)*** for their 57 and 46.

[3 marks]

e. straight line drawn on the scatter diagram ***(A1)*****(ft)*****(A1)*****(ft)**

Notes: The line must be straight for either of the two marks to be awarded.

Award ***(A1)*****(ft)** passing through their M plotted in (c).

Award ***(A1)*****(ft)** for correct y-intercept (between 9 and 12).

Follow through from their y-intercept found in part (d).

If part (d) is used, award ***(A1)*****(ft)** for their intercept (± 1).

[2 marks]

f. $y = 1.01431... \times 76 + 10.3412\ldots$ ***(M1)***

Note: Award ***(M1)*** for substitution of 76 into their regression line.

$= 87.4295\ldots$ ***(A1)*****(ft)**

Note: Follow through from part (d). If 3 sf values are used the value is 87.06.

$\$87$ ***(A1)*****(ft)*****(G2)***

Notes: The final ***(A1)*** is awarded for their answer given correct to the nearest dollar.

Method, followed by the answer of 87 earns ***(M1)(G2).*** It is not necessary to see the interim step.

Where the candidate uses their graph instead of the equation, and arrives at an answer other than 87, award, at most, ***(G1)*****(ft)**.

If the candidate uses their graph and arrives at the required answer of 87, award ***(G2)*****(ft)**.

[3 marks]

g. 76 is within the range of distances given in the data **OR** the correlation coefficient is close to 1. ***(R1)***

Notes: Award ***(R1)*** if **either** condition is given.

Sufficient to indicate that 76 is 'within the data range' and the correlation is 'strong'.

Allow r^2 close to 1.

Do **not** accept "within the range of prices".

[1 mark]

h. $\text{Percentage error} = \frac{87-80}{80} \times 100$ ***(M1)***

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

8.75% ***(A1)*****(ft)*****(G2)***

Notes: Follow through from their answer to part (f).

Accept either the rounded or unrounded answer to part (f).

If no integer value seen in part (f), follow through from their unrounded answer to part (f).

Answer must be positive.

[2 marks]

Examiners report

a. This question was very well attempted by a significant majority of candidates. Many good and accurate attempts at plotting a scatter diagram were seen in part (a). However, a minority of candidates chose not to use graph paper but instead used their answer book. These candidates achieved, at most, one mark for that part question. Many correct answers were seen in parts (b) and (d) reflecting good use of the graphic display calculator. Whilst many candidates realized that the line of regression passes through the point *M*, a significant number of candidates seemed to draw their line 'by eye' rather than using the equation found in part (d) and, as a consequence for many, their straight line (or projected line) did not fall within the required tolerances for the second mark. Many candidates understood the requirements for part (f) and full marks were seen on a majority of scripts. Those candidates, however, who used their graph instead scored, at most, two marks here. Many candidates seemed to be well-drilled in giving a suitable reason in part (f) and 'within the data range' or a 'strong correlation' were frequently seen. Percentage error caused very few problems for candidates and many correct answers were seen in part (h).

b. This question was very well attempted by a significant majority of candidates. Many good and accurate attempts at plotting a scatter diagram were seen in part (a). However, a minority of candidates chose not to use graph paper but instead used their answer book. These candidates achieved, at most, one mark for that part question. Many correct answers were seen in parts (b) and (d) reflecting good use of the graphic display calculator. Whilst many candidates realized that the line of regression passes through the point *M*, a significant number of candidates seemed to draw their line 'by eye' rather than using the equation found in part (d) and, as a consequence for many, their straight line (or projected line) did not fall within the required tolerances for the second mark. Many candidates understood the requirements for part (f) and

full marks were seen on a majority of scripts. Those candidates, however, who used their graph instead scored, at most, two marks here. Many candidates seemed to be well-drilled in giving a suitable reason in part (f) and 'within the data range' or a 'strong correlation' were frequently seen. Percentage error caused very few problems for candidates and many correct answers were seen in part (h).

c. This question was very well attempted by a significant majority of candidates. Many good and accurate attempts at plotting a scatter diagram were seen in part (a). However, a minority of candidates chose not to use graph paper but instead used their answer book. These candidates achieved, at most, one mark for that part question. Many correct answers were seen in parts (b) and (d) reflecting good use of the graphic display calculator. Whilst many candidates realized that the line of regression passes through the point M, a significant number of candidates seemed to draw their line 'by eye' rather than using the equation found in part (d) and, as a consequence for many, their straight line (or projected line) did not fall within the required tolerances for the second mark. Many candidates understood the requirements for part (f) and full marks were seen on a majority of scripts. Those candidates, however, who used their graph instead scored, at most, two marks here. Many candidates seemed to be well-drilled in giving a suitable reason in part (f) and 'within the data range' or a 'strong correlation' were frequently seen. Percentage error caused very few problems for candidates and many correct answers were seen in part (h).

d. This question was very well attempted by a significant majority of candidates. Many good and accurate attempts at plotting a scatter diagram were seen in part (a). However, a minority of candidates chose not to use graph paper but instead used their answer book. These candidates achieved, at most, one mark for that part question. Many correct answers were seen in parts (b) and (d) reflecting good use of the graphic display calculator. Whilst many candidates realized that the line of regression passes through the point M, a significant number of candidates seemed to draw their line 'by eye' rather than using the equation found in part (d) and, as a consequence for many, their straight line (or projected line) did not fall within the required tolerances for the second mark. Many candidates understood the requirements for part (f) and full marks were seen on a majority of scripts. Those candidates, however, who used their graph instead scored, at most, two marks here. Many candidates seemed to be well-drilled in giving a suitable reason in part (f) and 'within the data range' or a 'strong correlation' were frequently seen. Percentage error caused very few problems for candidates and many correct answers were seen in part (h).

e. This question was very well attempted by a significant majority of candidates. Many good and accurate attempts at plotting a scatter diagram were seen in part (a). However, a minority of candidates chose not to use graph paper but instead used their answer book. These candidates achieved, at most, one mark for that part question. Many correct answers were seen in parts (b) and (d) reflecting good use of the graphic display calculator. Whilst many candidates realized that the line of regression passes through the point M, a significant number of candidates seemed to draw their line 'by eye' rather than using the equation found in part (d) and, as a consequence for many, their straight line (or projected line) did not fall within the required tolerances for the second mark. Many candidates understood the requirements for part (f) and full marks were seen on a majority of scripts. Those candidates, however, who used their graph instead scored, at most, two marks here. Many candidates seemed to be well-drilled in giving a suitable reason in part (f) and 'within the data range' or a 'strong correlation' were frequently seen. Percentage error caused very few problems for candidates and many correct answers were seen in part (h).

f. This question was very well attempted by a significant majority of candidates. Many good and accurate attempts at plotting a scatter diagram were seen in part (a). However, a minority of candidates chose not to use graph paper but instead used their answer book. These candidates achieved, at most, one mark for that part question. Many correct answers were seen in parts (b) and (d) reflecting good use of the graphic display calculator. Whilst many candidates realized that the line of regression passes through the point M, a significant number of candidates

seemed to draw their line ‘by eye’ rather than using the equation found in part (d) and, as a consequence for many, their straight line (or projected line) did not fall within the required tolerances for the second mark. Many candidates understood the requirements for part (f) and full marks were seen on a majority of scripts. Those candidates, however, who used their graph instead scored, at most, two marks here. Many candidates seemed to be well-drilled in giving a suitable reason in part (f) and ‘within the data range’ or a ‘strong correlation’ were frequently seen. Percentage error caused very few problems for candidates and many correct answers were seen in part (h).

g. This question was very well attempted by a significant majority of candidates. Many good and accurate attempts at plotting a scatter diagram were seen in part (a). However, a minority of candidates chose not to use graph paper but instead used their answer book. These candidates achieved, at most, one mark for that part question. Many correct answers were seen in parts (b) and (d) reflecting good use of the graphic display calculator. Whilst many candidates realized that the line of regression passes through the point M, a significant number of candidates seemed to draw their line ‘by eye’ rather than using the equation found in part (d) and, as a consequence for many, their straight line (or projected line) did not fall within the required tolerances for the second mark. Many candidates understood the requirements for part (f) and full marks were seen on a majority of scripts. Those candidates, however, who used their graph instead scored, at most, two marks here. Many candidates seemed to be well-drilled in giving a suitable reason in part (f) and ‘within the data range’ or a ‘strong correlation’ were frequently seen. Percentage error caused very few problems for candidates and many correct answers were seen in part (h).

h. This question was very well attempted by a significant majority of candidates. Many good and accurate attempts at plotting a scatter diagram were seen in part (a). However, a minority of candidates chose not to use graph paper but instead used their answer book. These candidates achieved, at most, one mark for that part question. Many correct answers were seen in parts (b) and (d) reflecting good use of the graphic display calculator. Whilst many candidates realized that the line of regression passes through the point M, a significant number of candidates seemed to draw their line ‘by eye’ rather than using the equation found in part (d) and, as a consequence for many, their straight line (or projected line) did not fall within the required tolerances for the second mark. Many candidates understood the requirements for part (f) and full marks were seen on a majority of scripts. Those candidates, however, who used their graph instead scored, at most, two marks here. Many candidates seemed to be well-drilled in giving a suitable reason in part (f) and ‘within the data range’ or a ‘strong correlation’ were frequently seen. Percentage error caused very few problems for candidates and many correct answers were seen in part (h).

The lengths (l) in centimetres of 100 copper pipes at a local building supplier were measured. The results are listed in the table below.

Length l (cm)	Frequency
17.5	12
32.5	26
47.5	32
62.5	21
77.5	9

a. Write down the mode. [1]

b. Using your graphic display calculator, write down the value of [4]

(i) the mean;

(ii) the standard deviation;

(iii) the median.

c. Find the interquartile range. [2]

d. Draw a box and whisker diagram for this data, on graph paper, using a scale of $1\ \text{cm}$ to represent $5\ \text{cm}$. [4]

e. Sam estimated the value of the mean of the measured lengths to be $43\ \text{cm}$. [2]

Find the percentage error of Sam's estimated mean.

Markscheme

a. $47.5\ (\text{cm})$ ***(A1)***

b. (i) $45.85\ (\text{cm})$ ***(G2)***

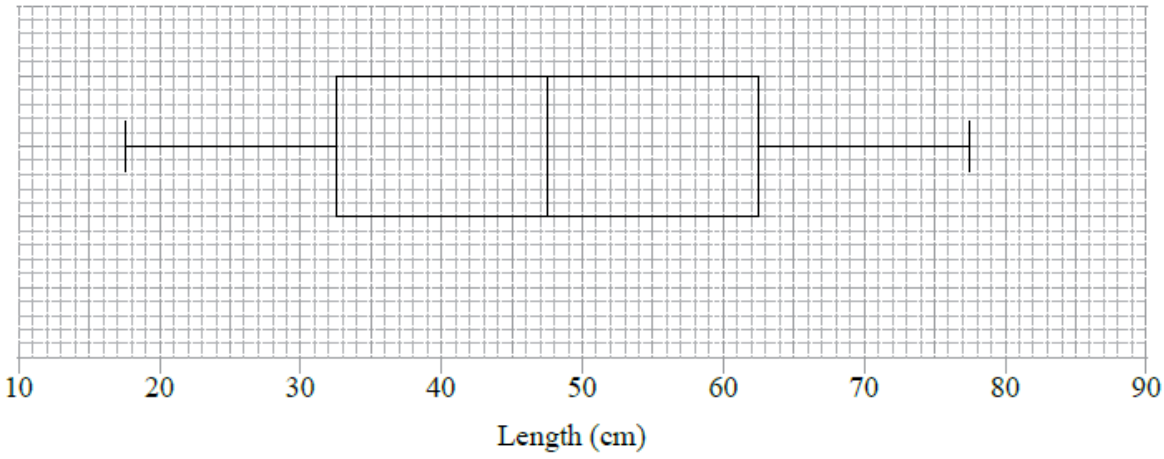
Note: Accept 45.9 .

(ii) $17.1\ (17.0888\ldots)$ ***(G1)***
(iii) $47.5\ (\text{cm})$ ***(G1)***

c. $62.5 - 32.5 = 30$ ***(M1)(A1)(G2)***

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

d.



(A1) for correct label and scale
(A1)**(ft)** for correct median
(A1)**(ft)** for correct quartiles and box
(A1) for endpoints at 17.5 and 77.5 joined to box by straight lines ***(A1)(A1)*****(ft)*****(A1)*****(ft)*****(A1)***

Notes: The final ***(A1)*** is lost if the lines go through the box. Follow through from their parts (b) and (c).

e. $\varepsilon = \left|\frac{43-45.85}{45.85}\right| \times 100\%$ **(M1)**

Note: Award ***(M1)*** for their correct substitution in $\%$ error formula.

$= 6.22\% \ (6.21592\ldots)$ ***(A1)*****(ft)*****(G2)***

Notes: Follow through from their answer to part (b)(i). Accept 6.32% with use of 45.9 .

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]
d. [N/A]
e. [N/A]

Alex and Kris are riding their bicycles together along a bicycle trail and note the following distance markers at the given times.

Time (t hours)	1	2	3	4	5	6	7
Distance (d km)	57	65	72	81	89	97	107

a. Draw a scatter diagram of the data. Use 1 cm to represent 1 hour and 1 cm to represent 10 km. [3]

b.i. Write down for this set of data the mean time, $\bar{t}$. [1]

b.ii. Write down for this set of data the mean distance, $\bar{d}$. [1]

c. Mark and label the point $M(\bar{t},\ \bar{d})$ on your scatter diagram. [2]

d. Draw the line of best fit on your scatter diagram. [2]

e. **Using your graph**, estimate the time when Alex and Kris pass the 85 km distance marker. Give your answer correct to **one decimal place**. [2]

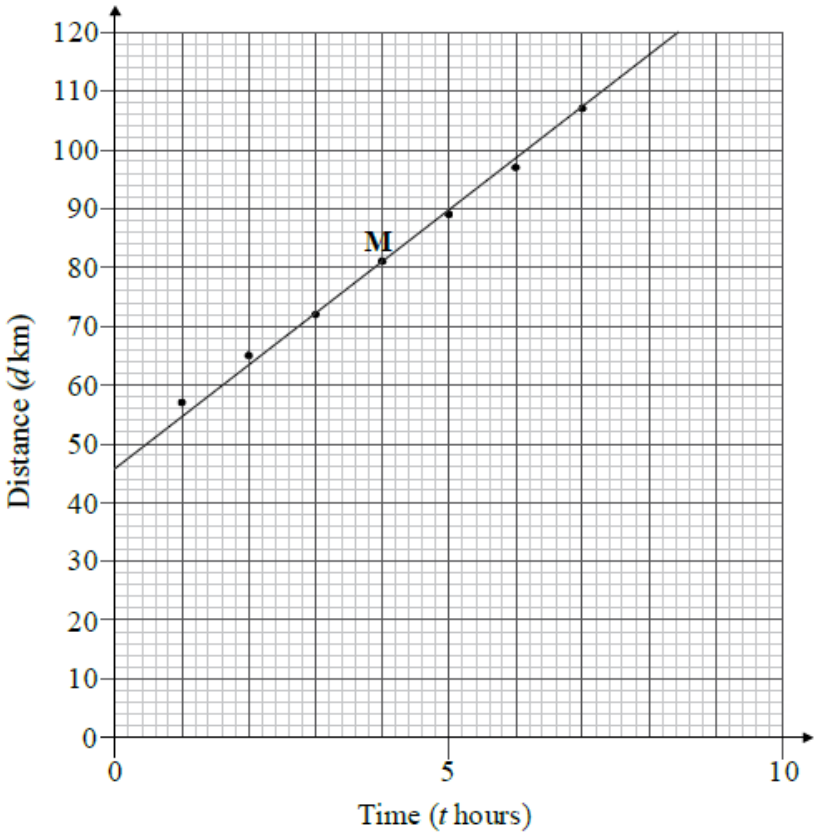
f. Write down the equation of the regression line for the data given. [2]

g.i. **Using your equation** calculate the distance marker passed by the cyclists at 10.3 hours. [2]

g.ii. Is this estimate of the distance reliable? Give a reason for your answer. [2]

Markscheme

a.



(A1)(A2)

Notes: Award ***(A1)*** for axes labelled with d and t and correct scale, ***(A2)*** for 6 or 7 points correctly plotted, ***(A1)*** for 4 or 5 points, ***(A0)*** for 3 or less points correctly plotted. Award at most ***(A1)(A1)*** if points are joined up. If axes are reversed award at most ***(A0)(A2)***

[3 marks]

b.i. $\bar{t} = 4$ ***(G1)***

[1 mark]

b.ii.

$\bar{d} = 81.1\left(\frac{568}{7}\right)$ ***(G1)***

Note: If answers are the wrong way around award in (i) ***(G0)*** and in (ii) ***(G1)*****(ft)**.

[1 mark]

c. Point marked and labelled with M or $\bar{t}$, $\bar{d}$ on their graph ***(A1)*****(ft)*****(A1)*****(ft)**

[2 marks]

d. Line of best fit drawn that passes through their M and (0, 48) ***(A1)*****(ft)*****(A1)*****(ft)**

Notes: Award ***(A1)*****(ft)** for straight line that passes through their M, ***(A1)*** for line (extrapolated if necessary) that passes through (0, 48).

Accept error of ±3. If ruler not used award a maximum of ***(A1)*****(ft)*****(A0)***.

[2 marks]

e. 4.5h (their answer ±0.2) ***(M1)(A1)*(ft)*(G2)***

Note: Follow through from their graph. If method shown by some indication on graph of point but answer is incorrect, award ***(M1)(A0)***.

[2 marks]

f. $d = 8.25t + 48.1$ ***(G1)(G1)***

Notes: Award ***(G1)*** for 8.25, ***(G1)*** for 48.1.

Award at most ***(G1)(G0)*** if $d =$ (*or* $y =$) is not seen.

Accept $d - 81.1 = 8.25(t - 4)$ or equivalent.

[2 marks]

g.i. $d = 8.25 \times 10.3 + 48.1$ ***(M1)***

$d = 133$ km ***(A1)*(ft)*(G2)***

[2 marks]

g.ii. No ***(A1)***

Outside the set of values of t or equivalent. ***(R1)***

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

[2 marks]

Examiners report

a. This question was well answered by most of the candidates. Diagrams were in general well drawn except for some students that reversed the axes or did not use the stated scales. They were able to use the GDC to find the means and the equation of the regression line. Very few students could take the correct decision in (g) (ii) by stating that the value was outside the range of the data set. The majority inclined their answers towards the context of the question and forgot what they had been taught about how wrong extrapolation can be.

b.i. This question was well answered by most of the candidates. Diagrams were in general well drawn except for some students that reversed the axes or did not use the stated scales. They were able to use the GDC to find the means and the equation of the regression line. Very few students could take the correct decision in (g) (ii) by stating that the value was outside the range of the data set. The majority inclined their answers towards the context of the question and forgot what they had been taught about how wrong extrapolation can be.

b.ii. This question was well answered by most of the candidates. Diagrams were in general well drawn except for some students that reversed the axes or did not use the stated scales. They were able to use the GDC to find the means and the equation of the regression line. Very few students could take the correct decision in (g) (ii) by stating that the value was outside the range of the data set. The majority inclined their answers towards the context of the question and forgot what they had been taught about how wrong extrapolation can be.

c. This question was well answered by most of the candidates. Diagrams were in general well drawn except for some students that reversed the axes or did not use the stated scales. They were able to use the GDC to find the means and the equation of the regression line. Very few students could take the correct decision in (g) (ii) by stating that the value was outside the range of the data set. The majority inclined their answers towards the context of the question and forgot what they had been taught about how wrong extrapolation can be.

d. This question was well answered by most of the candidates. Diagrams were in general well drawn except for some students that reversed the axes or did not use the stated scales. They were able to use the GDC to find the means and the equation of the regression line. Very few students could take the correct decision in (g) (ii) by stating that the value was outside the range of the data set. The majority inclined their answers towards the context of the question and forgot what they had been taught about how wrong extrapolation can be.

e. This question was well answered by most of the candidates. Diagrams were in general well drawn except for some students that reversed the axes or did not use the stated scales. They were able to use the GDC to find the means and the equation of the regression line. Very few students could take the correct decision in (g) (ii) by stating that the value was outside the range of the data set. The majority inclined their answers towards the context of the question and forgot what they had been taught about how wrong extrapolation can be.

f. This question was well answered by most of the candidates. Diagrams were in general well drawn except for some students that reversed the axes or did not use the stated scales. They were able to use the GDC to find the means and the equation of the regression line. Very few students could take the correct decision in (g) (ii) by stating that the value was outside the range of the data set. The majority inclined their answers towards the context of the question and forgot what they had been taught about how wrong extrapolation can be.

g.i. This question was well answered by most of the candidates. Diagrams were in general well drawn except for some students that reversed the axes or did not use the stated scales. They were able to use the GDC to find the means and the equation of the regression line. Very few students could take the correct decision in (g) (ii) by stating that the value was outside the range of the data set. The majority inclined their answers towards the context of the question and forgot what they had been taught about how wrong extrapolation can be.

g.ii. This question was well answered by most of the candidates. Diagrams were in general well drawn except for some students that reversed the axes or did not use the stated scales. They were able to use the GDC to find the means and the equation of the regression line. Very few students could take the correct decision in (g) (ii) by stating that the value was outside the range of the data set. The majority inclined their answers towards the context of the question and forgot what they had been taught about how wrong extrapolation can be.

In a mountain region there appears to be a relationship between the number of trees growing in the region and the depth of snow in winter. A set of 10 areas was chosen, and in each area the number of trees was counted and the depth of snow measured. The results are given in the table below.

Number of trees (x)	Depth of snow in cm (y)
45	30
75	50
66	40
27	25
44	30
28	5
60	35
35	20
73	45
47	25

In a study on 100 students there seemed to be a difference between males and females in their choice of favourite car colour. The results are given in the table below. A χ^2 test was conducted.

	Blue	Red	Green
Males	14	6	8
Females	31	24	17

A, a(i) Use your graphic display calculator to find the mean number of trees. [1]

A, a(ii) Use your graphic display calculator to find the mean depth of snow. [1]

A, a(iii) Use your graphic display calculator to find the standard deviation of the depth of snow. [1]

A, b The covariance, $S_{xy} = 188.5$. [2]

Write down the product-moment correlation coefficient, r.

A, c Write down the equation of the regression line of y on x. [2]

A, d(i) If the number of trees in an area is 55, estimate the depth of snow. [2]

A, e(i) Use the equation of the regression line to estimate the depth of snow in an area with 100 trees. [1]

A, e(ii) Decide whether the answer in (e)(i) is a valid estimate of the depth of snow in the area. Give a reason for your answer. [2]

B, a Write down the total number of male students. [1]

B, b Show that the expected frequency for males, whose favourite car colour is blue, is 12.6. [2]

B, c The calculated value of χ^2 is 1.367 and the critical value of χ^2 is 5.99 at the 5% significance level. [1]

Write down the null hypothesis for this test.

B, d(i) The calculated value of χ^2 is 1.367 and the critical value of χ^2 is 5.99 at the 5% significance level. [1]

Write down the number of degrees of freedom.

B, d(ii) The calculated value of χ^2 is 1.367 and the critical value of χ^2 is 5.99 at the 5% significance level. [2]

Determine whether the null hypothesis should be accepted at the 5% significance level. Give a reason for your answer.

Markscheme

A, a.i 50. ***(G1)***

[1 mark]

A, a.ii 30.5 ***(G1)***

[1 mark]

A, a.iii 12.3 ***(G1)***

Note: Award ***(A1)*****(ft)** for 13.0 in (iv) but only if 17.7 seen in (a)(ii).

[1 mark]

A, b. $= \frac{188.5}{(16.79 \times 12.33)}$ ***(M1)***

Note: Award ***(M1)*** for using their values in the correct formula.

= 0.911 (accept 0.912, 0.910) ***(A1)*****(ft)*****(G2)***

[2 marks]

A, c. $y = 0.669x - 2.95$ ***(G1)(G1)***

Note: Award ***(G1)*** for $0.669x$, ***(G1)*** for -2.95. If the answer is not in the form of an equation, award at most ***(G1)(G0)***.

[2 marks]

A, d. Depth = 0.669 × 55 − 2.95 ***(M1)***

= 33.8 ***(A1)*****(ft)*****(G2)*****(ft)**

Note: Follow through from their (c) even if no working seen.

[2 marks]

A, e. 64.0 (accept 63.95, 63.9) ***(A1)*****(ft)*****(G1)*****(ft)**

Note: Follow through from their (c) even if no working seen.

[1 mark]

A, f. It is not valid. It lies too far outside the values that are given. *Or equivalent.* ***(A1)(R1)***

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

[2 marks]

B, a. 28 ***(A1)***

[1 mark]

B, b. $\frac{28\times45}{100}\left(\frac{28}{100}\times\frac{45}{100}\times100\right)$ ***(M1)(A1)(ft)***

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

= 12.6 ***(AG)***

Note: Do not award ***(A1)*** unless 12.6 seen.

[2 marks]

B, c. i. the favourite car colour is **independent** of gender. ***(A1)***

Note: Accept there is no association between gender and favourite car colour.

Do not accept 'not related' or 'not correlated'.

[1 mark]

B, c. ii. ***(A1)***

[1 marks]

B, c. iii. Accept the null hypothesis since $1.367 < 5.991$ ***(A1)*(ft)*(R1)***

Note: Allow "Do not reject". Follow through from their null hypothesis and their critical value.

Full credit for use of p-values from GDC [$p = 0.505$].

Do not award ***(A1)(R0)***. Award ***(R1)*** for valid comparison.

[2 marks]

Examiners report

A, a. i. A straightforward question that saw many fine attempts. Given its nature – where much of the work was done on the GDC – it must be emphasised to candidates that incorrect entry of data into the calculator will result in considerable penalties; they must check their data entry most carefully.

The use of the inappropriate standard deviation was seen, but infrequently.

A, a. ii. A straightforward question that saw many fine attempts. Given its nature – where much of the work was done on the GDC – it must be emphasised to candidates that incorrect entry of data into the calculator will result in considerable penalties; they must check their data entry most carefully.

The use of the inappropriate standard deviation was seen, but infrequently.

A, a(i) A straightforward question that saw many fine attempts. Given its nature – where much of the work was done on the GDC – it must be emphasised to candidates that incorrect entry of data into the calculator will result in considerable penalties; they must check their data entry most carefully.

The use of the inappropriate standard deviation was seen, but infrequently.

A, b A straightforward question that saw many fine attempts. Given its nature – where much of the work was done on the GDC – it must be emphasised to candidates that incorrect entry of data into the calculator will result in considerable penalties; they must check their data entry most carefully.

It is expected that the GDC is used to calculate the correlation coefficient; the covariance was given to aid those candidates for whom the reset process removes this function from the display. It is anticipated that this hint will not be given in future papers.

A, c A straightforward question that saw many fine attempts. Given its nature – where much of the work was done on the GDC – it must be emphasised to candidates that incorrect entry of data into the calculator will result in considerable penalties; they must check their data entry most carefully.

A, d A straightforward question that saw many fine attempts. Given its nature – where much of the work was done on the GDC – it must be emphasised to candidates that incorrect entry of data into the calculator will result in considerable penalties; they must check their data entry most carefully.

A, e(i) A straightforward question that saw many fine attempts. Given its nature – where much of the work was done on the GDC – it must be emphasised to candidates that incorrect entry of data into the calculator will result in considerable penalties; they must check their data entry most carefully.

A, e(ii) A straightforward question that saw many fine attempts. Given its nature – where much of the work was done on the GDC – it must be emphasised to candidates that incorrect entry of data into the calculator will result in considerable penalties; they must check their data entry most carefully.

The dangers of extrapolation should be clearly explained to students.

B, a Once again, a straightforward question on chi-squared testing that was either highly successful (for the majority) or showed a lack of syllabus coverage.

B, b Once again, a straightforward question on chi-squared testing that was either highly successful (for the majority) or showed a lack of syllabus coverage. A surprising number of candidates lacked knowledge of the theory underlying the test and were thus unable to attempt (b).

B, c(i) Once again, a straightforward question on chi-squared testing that was either highly successful (for the majority) or showed a lack of syllabus coverage. In (c)(i) it is worth stressing that the test is for the mathematical **independence** of two characteristics and this determines the null hypothesis.

B, c(ii) Once again, a straightforward question on chi-squared testing that was either highly successful (for the majority) or showed a lack of syllabus coverage.

B, Once again, a straightforward question on chi-squared testing that was either highly successful (for the majority) or showed a lack of syllabus coverage. A number of candidates confuse the critical value and p-value approach to the test and thus lost marks in (c)(iv).

The number of bottles of water sold at a railway station on each day is given in the following table.

Day	0	1	2	3	4	5	6	7	8	9	10	11	12
Temperature ($T°$)	21	20.7	20	19	18	17.3	17	17.3	18	19	20	20.7	21
Number of bottles sold (n)	150	141	126	125	98	101	93	99	116	121	119	134	141

a. Write down [2]

(i) the mean temperature;

(ii) the standard deviation of the temperatures.

b. Write down the correlation coefficient, r, for the variables n and T. [1]

c. Comment on your value for r. [2]

d. The equation of the line of regression for n on T is $n = dT - 100$. [2]

(i) Write down the value of d.

(ii) Estimate how many bottles of water will be sold when the temperature is $19.6°$.

e. On a day when the temperature was $36°$ Peter calculates that 314 bottles would be sold. Give one reason why his answer might be unreliable. [1]

Markscheme

a. (i) 19.2 ***(G1)***

(ii) 1.45 ***(G1)***

[2 marks]

b. $r = 0.942$ ***(G1)***

[1 mark]

c. Strong, positive correlation. ***(A1)*****(ft)*****(A1)*****(ft)**

[2 marks]

d. (i) $d = 11.5$ ***(G1)***

(ii) $n = 11.5 \times 19.6 - 100$

$= 125$ (*accept* 126) ***(A1)*****(ft)**

Note: Answer must be a whole number.

[2 marks]

e. It is unreliable to extrapolate outside the values given (outlier). ***(R1)***

[1 mark]

Examiners report

a. (i) Generally well done but many lost an AP here

(ii) Only correct if the candidate knew how to use their GDC and even then several gave the wrong standard deviation.

b. Again, only correct if the candidate could use their GDC. Many answers given were greater than 1 and the candidates did not see anything wrong with this.

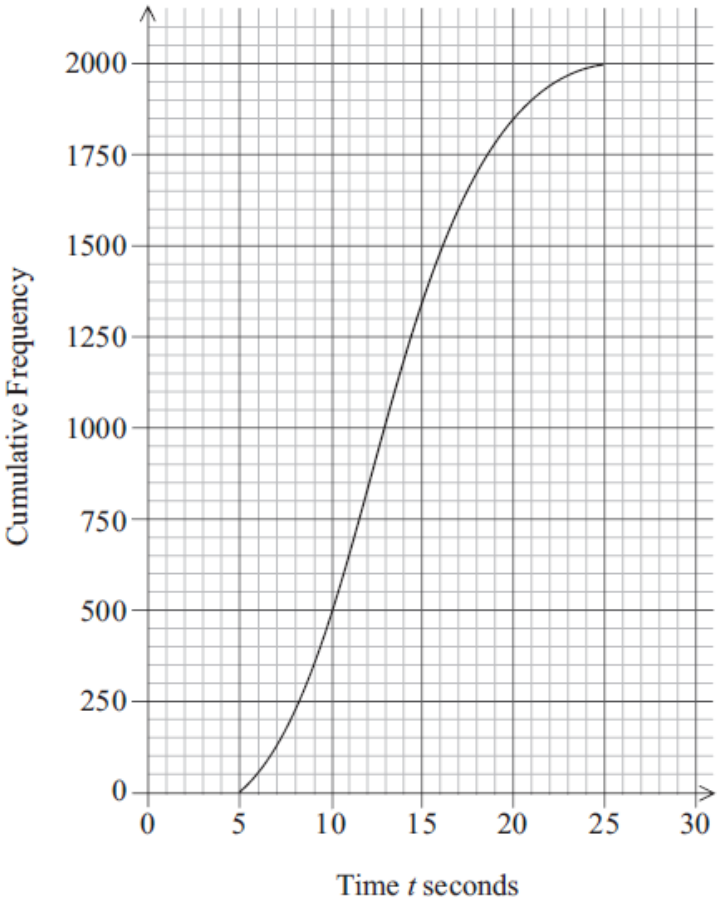
c. Many received a ft mark for this part. The word "positive" was often omitted.

d. (i) Most candidates substituted the first set of points into the equation instead of finding the regression line on their GDC.

(ii) Most managed to score a ft point here. But some did not give their answer as a whole number.

e. Not many candidates mentioned the idea of an outlier. Most came up with some creative reason, albeit wrong, as to why the answer might be unreliable. Some of them made interesting reading.

The diagram shows the cumulative frequency graph for the time t taken to perform a certain task by 2000 men.



a, i. Use the diagram to estimate the median time. [1]

a, ii. Use the diagram to estimate the upper quartile and the lower quartile. [2]

a, iii. Use the diagram to estimate the interquartile range. [1]

b. Find the number of men who take **more than** 11 seconds to perform the task. [3]

c. 55 % of the men took less than p seconds to perform the task. Find p. [2]

d, i. The times taken for the 2000 men were grouped as shown in the table below. [1]

Time	Frequency
$5 \leq t < 10$	500
$10 \leq t < 15$	850
$15 \leq t < 20$	a
$20 \leq t < 25$	b

Write down the value of a.

d, ii. The times taken for the 2000 men were grouped as shown in the table below. [1]

onbekend.png

Write down the value of b.

e, i. Use your graphic display calculator to find an estimate of the mean time. [2]

e, ii. Use your graphic display calculator to find an estimate of the standard deviation of the time. [1]

f. Everyone who performs the task in **less than** one standard deviation **below** the mean will receive a bonus. Pedro takes 9.5 seconds to perform the task. [3]

Does Pedro receive the bonus? Justify your answer.

Markscheme

a, i. ***Unit penalty (UP) applies in this part***

(UP) median = 13 seconds ***(A1)***

[1 mark]

a, ii. ***Unit penalty (UP) applies in this part***

(UP) 16 seconds and 10 seconds ***(A1)(A1)***

Note: Accept 16.1 or 16.2 for the upper quartile value.

[2 marks]

a, iiIQR = 6 seconds ***(A1)*****(ft)**

Note: (ft) from reasonable answers to (ii).

[1 mark]

b. value seen 650 ***(A1)***

2000 – value = 2000 – 650 ***(M1)***

= 1350 ***(A1)(G2)***

[3 marks]

c. 55 % of 2000 = 1100 ***(A1)***

$p = 13.5$ ***(A1)(G2)***

[2 marks]

d, i$a = 500$ ***(A1)***

[1 mark]

d, ii$b = 150$ ***(A1)***

[1 mark]

e, i***Unit penalty (UP) applies in this part***

(UP) $\bar{t} = 13.25$ seconds (13.3 seconds) ***(G2)***

OR

$\bar{t} = \frac{7.5 \times 500 + 12.5 \times 850 + 17.5 \times \text{their } a + 22.5 \times \text{their } b}{2000}$ ***(M1)***

(UP) $\bar{t} = 13.25$ seconds (13.3 seconds) ***(A1)*****(ft)**

Note: Award **(ft)** from their *a* and their *b* only if working is seen.

[2 marks]

e, ii$\sigma = 4.41$ seconds ***(G1)***

[1 mark]

f. $\bar{t} - \sigma = 8.84$ ***(A1)*****(ft)**

Their $\bar{t} - \sigma$ compared to 9.5 ***(R1)***

Pedro does not receive the bonus ***(A1)*****(ft)**

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

[3 marks]

Examiners report

a, i. Most of the students knew the definition of the median, quartiles and inter-quartile range though some confused variables and worked with the frequencies instead. Few could use their calculator to estimate the mean and standard deviation from grouped data. It cannot be said that the calculator was misused but that frequencies and midpoints were ignored when doing the calculations.

a, ii. Most of the students knew the definition of the median, quartiles and inter-quartile range though some confused variables and worked with the frequencies instead. Few could use their calculator to estimate the mean and standard deviation from grouped data. It cannot be said that the calculator was misused but that frequencies and midpoints were ignored when doing the calculations.

a, iii. Most of the students knew the definition of the median, quartiles and inter-quartile range though some confused variables and worked with the frequencies instead. Few could use their calculator to estimate the mean and standard deviation from grouped data. It cannot be said that the calculator was misused but that frequencies and midpoints were ignored when doing the calculations.

b. Most of the students knew the definition of the median, quartiles and inter-quartile range though some confused variables and worked with the frequencies instead. Few could use their calculator to estimate the mean and standard deviation from grouped data. It cannot be said that the calculator was misused but that frequencies and midpoints were ignored when doing the calculations.

c. Most of the students knew the definition of the median, quartiles and inter-quartile range though some confused variables and worked with the frequencies instead. Few could use their calculator to estimate the mean and standard deviation from grouped data. It cannot be said that the calculator was misused but that frequencies and midpoints were ignored when doing the calculations.

d, i. Most of the students knew the definition of the median, quartiles and inter-quartile range though some confused variables and worked with the frequencies instead. Few could use their calculator to estimate the mean and standard deviation from grouped data. It cannot be said that the calculator was misused but that frequencies and midpoints were ignored when doing the calculations.

d, ii. Most of the students knew the definition of the median, quartiles and inter-quartile range though some confused variables and worked with the frequencies instead. Few could use their calculator to estimate the mean and standard deviation from grouped data. It cannot be said that the calculator was misused but that frequencies and midpoints were ignored when doing the calculations.

e, i. Most of the students knew the definition of the median, quartiles and inter-quartile range though some confused variables and worked with the frequencies instead. Few could use their calculator to estimate the mean and standard deviation from grouped data. It cannot be said that the calculator was misused but that frequencies and midpoints were ignored when doing the calculations.

e, ii. Most of the students knew the definition of the median, quartiles and inter-quartile range though some confused variables and worked with the frequencies instead. Few could use their calculator to estimate the mean and standard deviation from grouped data. It cannot be said that the calculator was misused but that frequencies and midpoints were ignored when doing the calculations.

f. Most of the students knew the definition of the median, quartiles and inter-quartile range though some confused variables and worked with the frequencies instead. Few could use their calculator to estimate the mean and standard deviation from grouped data. It cannot be said that the calculator was misused but that frequencies and midpoints were ignored when doing the calculations. Part (f) acted as a good discriminator. Follow through marks were awarded in (f) when working was shown.

On one day 180 flights arrived at a particular airport. The distance travelled and the arrival status for each incoming flight was recorded. The flight was then classified as on time, slightly delayed, or heavily delayed.

The results are shown in the following table.

		Distance travelled			
		At most 500 km	Between 500 km and 5000 km	At least 5000 km	TOTAL
Arrival Status	On time	19	17	16	52
	Slightly delayed	13	18	14	45
	Heavily delayed	28	15	40	83
	TOTAL	60	50	70	180

A χ^2 test is carried out at the 10 % significance level to determine whether the arrival status of incoming flights is independent of the distance travelled.

The critical value for this test is 7.779.

A flight is chosen at random from the 180 recorded flights.

a. State the alternative hypothesis. [1]

b. Calculate the expected frequency of flights travelling at most 500 km and arriving slightly delayed. [2]

c. Write down the number of degrees of freedom. [1]

d.i. Write down the χ^2 statistic. [2]

d.ii. Write down the associated p-value. [1]

e. State, with a reason, whether you would reject the null hypothesis. [2]

f. Write down the probability that this flight arrived on time. [2]

g. Given that this flight was not heavily delayed, find the probability that it travelled between 500 km and 5000 km. [2]

h. Two flights are chosen at random from those which were slightly delayed. [3]

Find the probability that each of these flights travelled at least 5000 km.

Markscheme

a. The arrival status is dependent on the distance travelled by the incoming flight ***(A1)***

Note: Accept “associated” or “not independent”.

[1 mark]

b. $\frac{60\times45}{180}$ **OR** $\frac{60}{180}\times\frac{45}{180}\times180$ ***(M1)***

Note: Award ***(M1)*** for correct substitution into expected value formula.

= 15 ***(A1) (G2)***

[2 marks]

c. 4 ***(A1)***

Note: Award ***(A0)*** if "2 + 2 = 4" is seen.

[1 mark]

d.i. 9.55 (9.54671…) ***(G2)***

Note: Award ***(G1)*** for an answer of 9.54.

[2 marks]

d.ii. 0.0488 (0.0487961…) ***(G1)***

[1 mark]

e. Reject the Null Hypothesis ***(A1)*****(ft)**

Note: Follow through from their hypothesis in part (a).

9.55 (9.54671…) > 7.779 ***(R1)*****(ft)**

OR

0.0488 (0.0487961…) < 0.1 ***(R1)*****(ft)**

Note: Do not award ***(A1)*****(ft)*****(R0)*****(ft)**. Follow through from part (d). Award ***(R1)*****(ft)** for a correct comparison, ***(A1)*****(ft)** for a consistent conclusion with the answers to parts (a) and (d). Award ***(R1)*****(ft)** for $\chi^2_{calc} > \chi^2_{crit}$, provided the calculated value is explicitly seen in part (d)(i).

[2 marks]

f. $\frac{52}{180}\ \left(0.289,\ \frac{13}{45},\ 28.9\,\%\right)$ ***(A1)(A1) (G2)***

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

[2 marks]

g. $\frac{35}{97}\ (0.361,\ 36.1\,\%)$ ***(A1)(A1) (G2)***

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

[2 marks]

h. $\frac{14}{45}\times\frac{13}{44}$ ***(A1)(M1)***

Note: Award ***(A1)*** for two correct fractions and ***(M1)*** for multiplying their two fractions.

$=\frac{182}{1980}\ \left(0.0919,\ \frac{91}{990},\ 0.091919\ldots,\ 9.19\,\%\right)$ ***(A1) (G2)***

[3 marks]

Examiners report

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

f. [N/A]
g. [N/A]
h. [N/A]

A random sample of 167 people who own mobile phones was used to collect data on the amount of time they spent per day using their phones.

The results are displayed in the table below.

Time spent per day (t minutes)	$0 \leq t < 15$	$15 \leq t < 30$	$30 \leq t < 45$	$45 \leq t < 60$	$60 \leq t < 75$	$75 \leq t < 90$
Number of people	21	32	35	41	27	11

Manuel conducts a survey on a random sample of 751 people to see which television programme type they watch most from the following: Drama, Comedy, Film, News. The results are as follows.

	Drama	Comedy	Film	News
Males under 25	22	65	90	35
Males 25 and over	36	54	67	17
Females under 25	22	59	82	15
Females 25 and over	64	39	38	46

Manuel decides to ignore the ages and to test at the 5 % level of significance whether the most watched programme type is independent of **gender.**

i.a. State the modal group. [1]

i.b. Use your graphic display calculator to calculate approximate values of the mean and standard deviation of the time spent per day on these mobile phones. [3]

i.c. On graph paper, draw a fully labelled histogram to represent the data. [4]

ii.a. Draw a table with 2 rows and 4 columns of data so that Manuel can perform a chi-squared test. [3]

ii.b. State Manuel's null hypothesis and alternative hypothesis. [1]

ii.c. Find the expected frequency for the number of females who had 'Comedy' as their most-watched programme type. Give your answer to the nearest whole number. [2]

ii.d. Using your graphic display calculator, or otherwise, find the chi-squared statistic for Manuel's data. [3]

ii.e. (i) State the number of degrees of freedom available for this calculation. [3]

(ii) State his conclusion.

Markscheme

i.a. $45 \leqslant t < 60$ ***(A1)***

[1 mark]

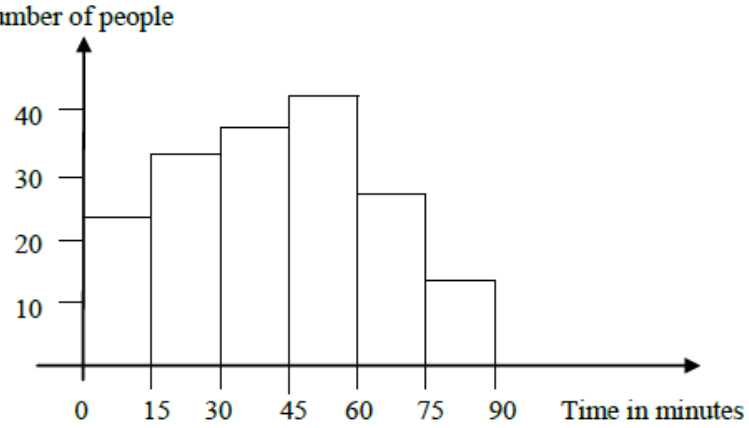
i.b. *Unit penalty **(UP)** is applicable in question part (i)(b) **only**.*

(UP) 42.4 minutes ***(G2)***

21.6 minutes ***(G1)***

[3 marks]

i.c.



(A4)

[4 marks]

ii.a.

	Drama	Comedy	Film	News
Males	58	119	157	52
Females	86	98	120	61

(M1)(M1)(A1)

[3 marks]

ii.b. H_0: favourite TV programme is independent of gender or no association between favourite TV programme and gender

H_1: favourite TV programme is dependent on gender *(must have both)* ***(A1)***

[1 mark]

ii.c. $\frac{365\times217}{751}$ ***(M1)***

$= 105$ ***(A1)*****(ft)*****(G2)***

[2 marks]

ii.d. 12.6 *(accept* 12.558*)* ***(G3)***

[3 marks]

ii.e. (i) 3 ***(A1)***

(ii) reject H_0 *or equivalent statement (e.g. accept H_1)* ***(A1)*****(ft)**

[3 marks]

Examiners report

i.a. Many candidates who had survived the previous two unit penalties, fell here with omission of units for the mean and standard deviation. The modal group was answered well. Part (b), finding the mean and standard deviation by GDC, was answered very poorly. Most did put the midpoints in one list and the frequencies in a second list but then either used the 2-Var stats button or 1-var stats button but only named L1 instead of L1, L2. Candidates who showed midpoints in their working did at least score a method mark.

i.b. Many candidates who had survived the previous two unit penalties, fell here with omission of units for the mean and standard deviation. The modal group was answered well. Part (b), finding the mean and standard deviation by GDC, was answered very poorly. Most did put the midpoints in one list and the frequencies in a second list but then either used the 2-Var stats button or 1-var stats button but only named L1 instead of L1, L2. Candidates who showed midpoints in their working did at least score a method mark.

i.c. Many candidates who had survived the previous two unit penalties, fell here with omission of units for the mean and standard deviation. The modal group was answered well. Part (b), finding the mean and standard deviation by GDC, was answered very poorly. Most did put the midpoints in one list and the frequencies in a second list but then either used the 2-Var stats button or 1-var stats button but only named L1 instead of L1, L2. Candidates who showed midpoints in their working did at least score a method mark.

ii.a. The chi-squared question was answered well by the majority of candidates and almost all found the chi-squared statistic correctly by GDC, though many could not look up the correct critical value.

ii.b. The chi-squared question was answered well by the majority of candidates and almost all found the chi-squared statistic correctly by GDC, though many could not look up the correct critical value.

ii.c. The chi-squared question was answered well by the majority of candidates and almost all found the chi-squared statistic correctly by GDC, though many could not look up the correct critical value.

ii.d. The chi-squared question was answered well by the majority of candidates and almost all found the chi-squared statistic correctly by GDC, though many could not look up the correct critical value.

ii.e. The chi-squared question was answered well by the majority of candidates and almost all found the chi-squared statistic correctly by GDC, though many could not look up the correct critical value.

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

Speed, s (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

a. Write down the midpoint of the $60 < s \leqslant 70$ interval. [1]

b. Use your graphic display calculator to find an estimate for [3]

(i) the mean speed of the vehicles;

(ii) the standard deviation of the speeds of the vehicles.

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

d. Consider the cumulative frequency table below. [2]

Speed, s (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.

e. Consider the cumulative frequency table below. [4]

Speed, s (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.

f. Use your cumulative frequency graph to estimate [4]

(i) the median speed of the vehicles;

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

g. 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]

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

Markscheme

a. 65 ***(A1)***

[1 mark]

b. (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) $19.2\ (19.2093\ldots)$ ***(G1)***

Note: Accept 19, do not accept 20.

[3 marks]

c. 76 ***(A1)***

[1 mark]

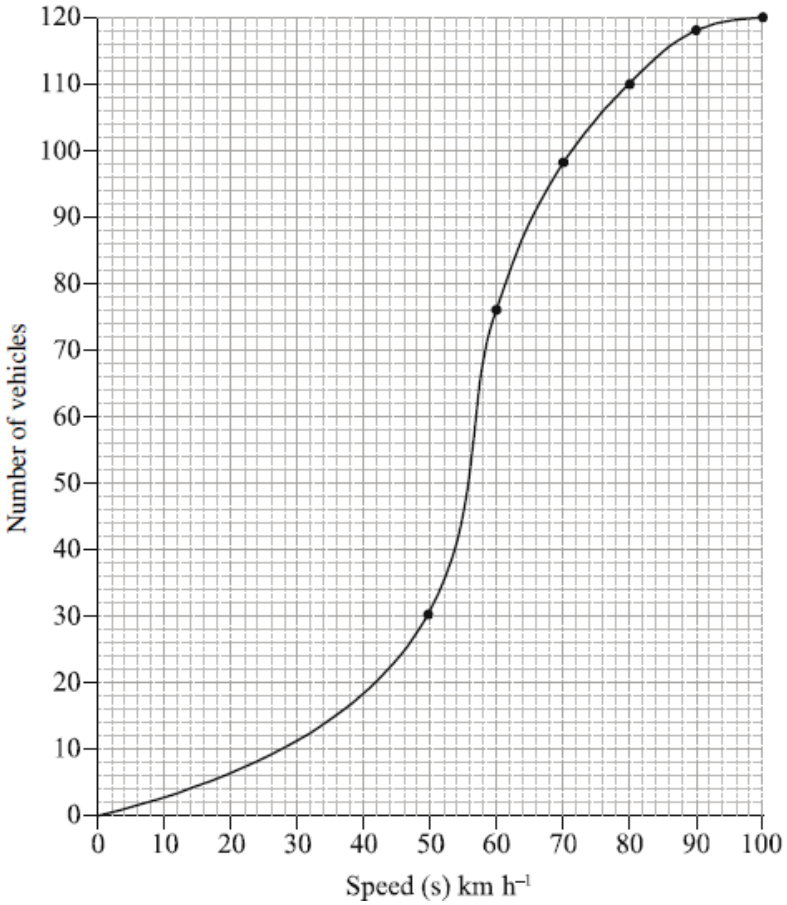
d. $a = 76, b = 98$ ***(A1)*****(ft)**

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

$c = 118$ ***(A1)***

[2 marks]

e.



(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]

f. (i) $57\ (\text{km h}^{-1})\ (\pm 2)$ ***(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.

(ii) 90 vehicles (± 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]

g. $50 + 19.2 = 69.2$ ***(A1)*(ft)**

$24\ (\pm 2)$ drivers will be fined ***(M1)(A1)*(ft)*(G2)***

Notes: Follow through from their graph and from their part (b)(ii). Award ***(M1)*** 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]

Examiners report

a. The great majority of candidates found this question to be a good start to the paper. The different class widths seemed to cause more problems to the teachers commenting on the G2 forms than to the students. However, 1(b) was a discriminator at the grade 4 level. Most candidates were successful in drawing the cumulative frequency curve or attempting to do so. There were a small number who clearly had never had any experience with this type of graph. A common error was the incorrect plotting of points at the interval midpoints. Weaker candidates plotted bar charts.

   A small number of candidates did not use the graph paper provided, preferring instead to use lined paper. This is to be strongly discouraged since no judgment will be made about the scale used or the accuracy of the plotted points. Similarly, the graph will not be used to benefit students whose answers lie outside accepted tolerances but who have shown working. Drawing an accurate graph requires the use of graph paper.

b. The great majority of candidates found this question to be a good start to the paper. The different class widths seemed to cause more problems to the teachers commenting on the G2 forms than to the students. However, 1(b) was a discriminator at the grade 4 level. Most candidates were successful in drawing the cumulative frequency curve or attempting to do so. There were a small number who clearly had never had any experience with this type of graph. A common error was the incorrect plotting of points at the interval midpoints. Weaker candidates plotted bar charts.

   A small number of candidates did not use the graph paper provided, preferring instead to use lined paper. This is to be strongly discouraged since no judgment will be made about the scale used or the accuracy of the plotted points. Similarly, the graph will not be used to benefit students whose answers lie outside accepted tolerances but who have shown working. Drawing an accurate graph requires the use of graph paper.

c. The great majority of candidates found this question to be a good start to the paper. The different class widths seemed to cause more problems to the teachers commenting on the G2 forms than to the students. However, 1(b) was a discriminator at the grade 4 level. Most candidates were successful in drawing the cumulative frequency curve or attempting to do so. There were a small number who clearly had never had any experience with this type of graph. A common error was the incorrect plotting of points at the interval midpoints. Weaker candidates plotted bar charts.

   A small number of candidates did not use the graph paper provided, preferring instead to use lined paper. This is to be strongly discouraged since no judgment will be made about the scale used or the accuracy of the plotted points. Similarly, the graph will not be used to benefit students whose answers lie outside accepted tolerances but who have shown working. Drawing an accurate graph requires the use of graph paper.

d. The great majority of candidates found this question to be a good start to the paper. The different class widths seemed to cause more problems to the teachers commenting on the G2 forms than to the students. However, 1(b) was a discriminator at the grade 4 level. Most candidates were successful in drawing the cumulative frequency curve or attempting to do so. There were a small number who clearly had never had any experience with this type of graph. A common error was the incorrect plotting of points at the interval midpoints. Weaker candidates plotted bar charts.

   A small number of candidates did not use the graph paper provided, preferring instead to use lined paper. This is to be strongly discouraged since no judgment will be made about the scale used or the accuracy of the plotted points. Similarly, the graph will not be used to benefit students whose answers lie outside accepted tolerances but who have shown working. Drawing an accurate graph requires the use of graph paper.

e. The great majority of candidates found this question to be a good start to the paper. The different class widths seemed to cause more problems to the teachers commenting on the G2 forms than to the students. However, 1(b) was a discriminator at the grade 4 level. Most candidates were successful in drawing the cumulative frequency curve or attempting to do so. There were a small number who clearly had never had any

experience with this type of graph. A common error was the incorrect plotting of points at the interval midpoints. Weaker candidates plotted bar charts.

A small number of candidates did not use the graph paper provided, preferring instead to use lined paper. This is to be strongly discouraged since no judgment will be made about the scale used or the accuracy of the plotted points. Similarly, the graph will not be used to benefit students whose answers lie outside accepted tolerances but who have shown working. Drawing an accurate graph requires the use of graph paper.

f. The great majority of candidates found this question to be a good start to the paper. The different class widths seemed to cause more problems to the teachers commenting on the G2 forms than to the students. However, 1(b) was a discriminator at the grade 4 level. Most candidates were successful in drawing the cumulative frequency curve or attempting to do so. There were a small number who clearly had never had any experience with this type of graph. A common error was the incorrect plotting of points at the interval midpoints. Weaker candidates plotted bar charts.

A small number of candidates did not use the graph paper provided, preferring instead to use lined paper. This is to be strongly discouraged since no judgment will be made about the scale used or the accuracy of the plotted points. Similarly, the graph will not be used to benefit students whose answers lie outside accepted tolerances but who have shown working. Drawing an accurate graph requires the use of graph paper.

g. The great majority of candidates found this question to be a good start to the paper. The different class widths seemed to cause more problems to the teachers commenting on the G2 forms than to the students. However, 1(b) was a discriminator at the grade 4 level. Most candidates were successful in drawing the cumulative frequency curve or attempting to do so. There were a small number who clearly had never had any experience with this type of graph. A common error was the incorrect plotting of points at the interval midpoints. Weaker candidates plotted bar charts.

A small number of candidates did not use the graph paper provided, preferring instead to use lined paper. This is to be strongly discouraged since no judgment will be made about the scale used or the accuracy of the plotted points. Similarly, the graph will not be used to benefit students whose answers lie outside accepted tolerances but who have shown working. Drawing an accurate graph requires the use of graph paper.

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

a. Draw a scatter diagram to show the relationship between the mass of a chicken and its cooking time. Use 2 cm to represent 0.5 kg on the horizontal axis and 1 cm to represent 10 minutes on the vertical axis. [4]

b. Write down for this set of data [2]

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

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

c. Label the point $\mathrm{M}(\bar{m}, \bar{t})$ on the scatter diagram. [1]

d. Draw the line of best fit on the scatter diagram. [2]

e. Using your line of best fit, estimate the cooking time, in minutes, for a 1.7 kg chicken. [2]

f. Write down the Pearson's product–moment correlation coefficient, *r* . [2]

g. Using your value for *r* , comment on the correlation. [2]

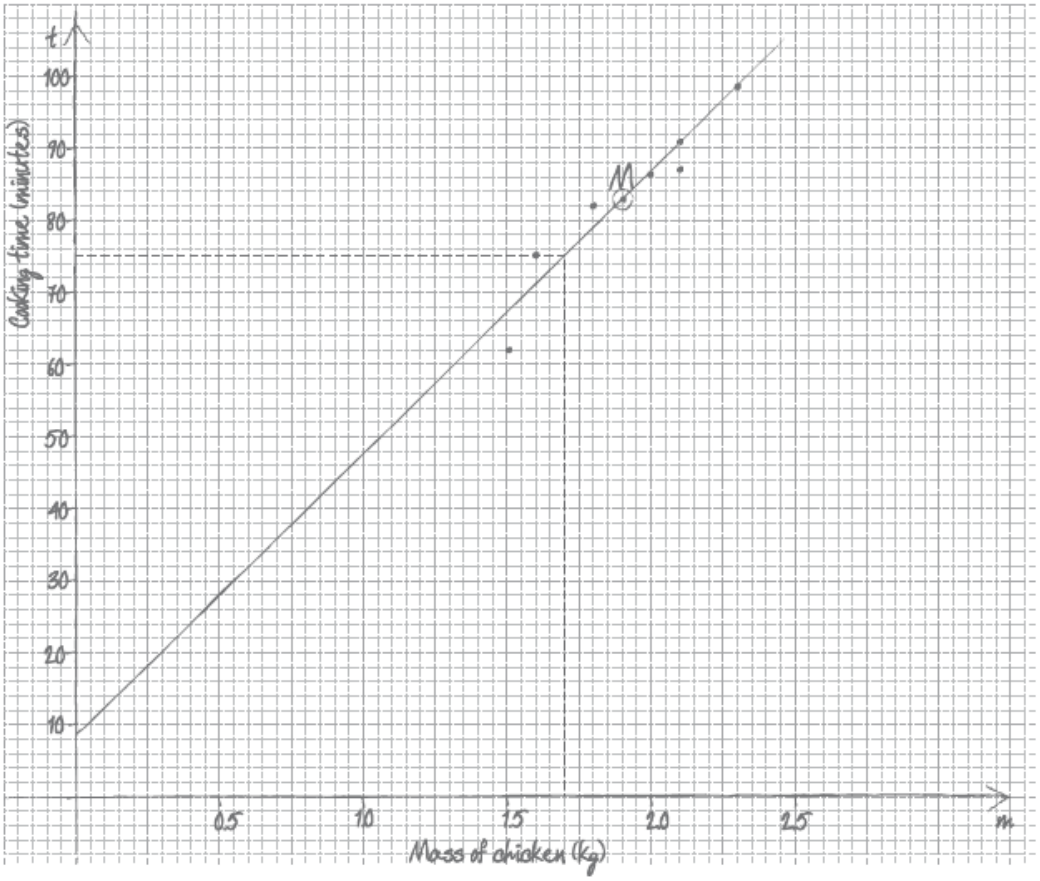
h. The cooking time of an additional 2.0 kg chicken is recorded. If the mass and cooking time of this chicken is included in the data, the correlation is weak. [2]

(i) Explain how the cooking time of this additional chicken might differ from that of the other eight chickens.

(ii) Explain how a new line of best fit might differ from that drawn in part (d).

Markscheme

a.



(A1) *for correct scales and labels (mass or m on the horizontals axis, time or t on the vertical axis)*

(A3) *for 7 or 8 correctly placed data points*

(A2) *for 5 or 6 correctly placed data points*

(A1) *for 3 or 4 correctly placed data points,* ***(A0)*** *otherwise.* ***(A4)***

Note: If axes reversed award at most ***(A0)(A3)*****(ft)**. If graph paper not used, award at most ***(A1)(A0)***.

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

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

c. Their mean point labelled. ***(A1)*****(ft)**

Note: Follow through from part (b). Accept any clear indication of the mean point. For example: circle around point, (m, t), M , etc.

d. Line of best fit drawn on scatter diagram. ***(A1)*****(ft)*****(A1)*****(ft)**

Notes:Award ***(A1)*****(ft)** for straight line through their mean point, ***(A1)*****(ft)** for line of best fit with intercept 9(±2) . The second ***(A1)*****(ft)** can be awarded even if the line does not reach the t-axis but, if extended, the t-intercept is correct.

e. 75 ***(M1)(A1)*****(ft)*****(G2)***

Notes: Accept 74.77 from the regression line equation. Award ***(M1)*** for indication of the use of their graph to get an estimate **OR** for correct substitution of 1.7 in the correct regression line equation $t = 38.5m + 9.32$.

f. 0.960 (0.959614...) ***(G2)***

Note: Award ***(G0)(G1)*****(ft)** for 0.95, 0.959

g. Strong and positive ***(A1)*****(ft)*****(A1)*****(ft)**

Note: Follow through from their correlation coefficient in part (f).

h. (i) Cooking time is much larger (or smaller) than the other eight ***(A1)***

(ii) The gradient of the new line of best fit will be larger (or smaller) ***(A1)***

Note: Some acceptable explanations may include but are not limited to:

The line of best fit may be further away from the plotted points
It may be steeper than the previous line (as the mean would change)
The t-intercept of the new line is smaller (larger)

Do not accept vague explanations, like:

The new line would vary
It would not go through all points
It would not fit the patterns
The line may be slightly tilted

Examiners report

a. [N/A]
b. [N/A]
c. [N/A]
d. [N/A]
e. [N/A]
f. [N/A]

g. [N/A]
h. [N/A]

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 χ^2 test at the 5% level of significance to determine whether the choice of language is independent of gender.

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.

Science (S)	23	51	56	62	12	73	72
French (F)	65	45	45	40	70	36	30

A.aState Ludmila's null hypothesis. [1]

A.bWrite down the number of degrees of freedom. [1]

A.cFind the expected frequency for the females studying Spanish. [2]

A.dUse your graphic display calculator to find the χ^2 test statistic for this data. [2]

A.eState whether Ludmila accepts the null hypothesis. Give a reason for your answer. [2]

B.aDraw a labelled scatter diagram for this data. Use a scale of $2\ \text{cm}$ to represent $10\ \text{marks}$ on the x-axis (S) and $10\ \text{marks}$ on the y-axis (F). [4]

B.bUse your graphic calculator to find [2]

(i) $\bar{S}$, the mean of S ;

(ii) $\bar{F}$, the mean of F .

B.cPlot the point $\text{M}(\bar{S}, \bar{F})$ on your scatter diagram. [1]

B.dUse your graphic display calculator to find the equation of the regression line of F on S . [2]

B.eDraw the regression line on your scatter diagram. [2]

B.f.Carletta's mark on the Science examination was 44. She did not sit the French examination. [2]

Estimate Carletta's mark for the French examination.

B.g. 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]

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

Markscheme

A.a.

H_0 : Choice of language is independent of gender. ***(A1)***

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

[1 mark]

A.b.

2 ***(A1)***

[1 mark]

A.c.

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

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

[2 marks]

A.d.

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

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

[2 marks]

A.e.

Accept H_0 since

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

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

B.a.

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]

B.b. (i) $\bar{S} = 49.9,$ ***(G1)***

(ii) $\bar{F} = 47.3$ ***(G1)***

[2 marks]

B.c. $\mathrm{M}(49.9, 47.3)$ plotted on scatter diagram ***(A1)*****(ft)**

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

[1 mark]

B.d.

$F = -0.619S + 78.2$ ***(G1)(G1)***

Notes: Award ***(G1)*** for $-0.619S$, ***(G1)*** for 78.2. If the answer is not in the form of an equation, award ***(G1)(G0)***. Accept $y = -0.619x + 78.2$.

OR

(F - 47.3 = - 0.619(S - 49.9)) ***(G1)(G1)***

Note: Award ***(G1)*** for -0.619, ***(G1)*** for the coordinates of their midpoint used. Follow through from their values in (b).

[2 marks]

B.e. line drawn on scatter diagram ***(A1)*****(ft)*****(A1)*****(ft)**

Notes: The drawn line **must** be straight for any marks to be awarded. Award ***(A1)*****(ft)** passing through their M plotted in (c). Award ***(A1)*****(ft)** for correct y-intercept. Follow through from their y-intercept found in (d).

[2 marks]

B.f. $F = -0.619 \times 44 + 78.2$ ***(M1)***

$= 51.0$ (allow 51 or 50.9) ***(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]

B.g. 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]

Examiners report

A.a. **Part A: Chi-square test**

This question part was answered well by most candidates. The null hypothesis and degrees of freedom were mostly correct. Some candidates offered a conclusion supported by good justifications, but others still showed lack of the necessary knowledge to do that. Some responses to part d) incurred an accuracy penalty for not adhering to the required accuracy level.

A.b. **Part A: Chi-square test**

This question part was answered well by most candidates. The null hypothesis and degrees of freedom were mostly correct. Some candidates offered a conclusion supported by good justifications, but others still showed lack of the necessary knowledge to do that. Some responses to part d) incurred an accuracy penalty for not adhering to the required accuracy level.

A.c **Part A: Chi-square test**

This question part was answered well by most candidates. The null hypothesis and degrees of freedom were mostly correct. Some candidates offered a conclusion supported by good justifications, but others still showed lack of the necessary knowledge to do that. Some responses to part d) incurred an accuracy penalty for not adhering to the required accuracy level.

A.d **Part A: Chi-square test**

This question part was answered well by most candidates. The null hypothesis and degrees of freedom were mostly correct. Some candidates offered a conclusion supported by good justifications, but others still showed lack of the necessary knowledge to do that. Some responses to part d) incurred an accuracy penalty for not adhering to the required accuracy level.

A.e **Part A: Chi-square test**

This question part was answered well by most candidates. The null hypothesis and degrees of freedom were mostly correct. Some candidates offered a conclusion supported by good justifications, but others still showed lack of the necessary knowledge to do that. Some responses to part d) incurred an accuracy penalty for not adhering to the required accuracy level.

B.a **Part B: Scatter plot and Regression line**

Many candidates reversed the axes in a), but the points were mostly plotted well. The values of the coefficients of the equation of the regression line $y = ax + b$ were often given not to the required 3 significant figure accuracy, and incurred a penalty. The regression line was often drawn not passing through point M and the y-intercept. The responses to the last part of the question were particularly weak, and many candidates were not able to offer a satisfactory reason to support their conclusion.

B.b **Part B: Scatter plot and Regression line**

Many candidates reversed the axes in a), but the points were mostly plotted well. The values of the coefficients of the equation of the regression line $y = ax + b$ were often given not to the required 3 significant figure accuracy, and incurred a penalty. The regression line was often drawn not passing through point M and the y-intercept. The responses to the last part of the question were particularly weak, and many candidates were not able to offer a satisfactory reason to support their conclusion.

B.c **Part B: Scatter plot and Regression line**

Many candidates reversed the axes in a), but the points were mostly plotted well. The values of the coefficients of the equation of the regression line $y = ax + b$ were often given not to the required 3 significant figure accuracy, and incurred a penalty. The regression line was often drawn not passing through point M and the y-intercept. The responses to the last part of the question were particularly weak, and many candidates were not able to offer a satisfactory reason to support their conclusion.

B.d **Part B: Scatter plot and Regression line**

Many candidates reversed the axes in a), but the points were mostly plotted well. The values of the coefficients of the equation of the regression line $y = ax + b$ were often given not to the required 3 significant figure accuracy, and incurred a penalty. The regression line was often drawn not passing through point M and the y-intercept. The responses to the last part of the question were particularly weak, and many candidates were not able to offer a satisfactory reason to support their conclusion.

B.e **Part B: Scatter plot and Regression line**

Many candidates reversed the axes in a), but the points were mostly plotted well. The values of the coefficients of the equation of the regression line $y = ax + b$ were often given not to the required 3 significant figure accuracy, and incurred a penalty. The regression line was often drawn not passing through point M and the y-intercept. The responses to the last part of the question were particularly weak, and many candidates were not able to offer a satisfactory reason to support their conclusion.

B.f. **Part B: Scatter plot and Regression line**

Many candidates reversed the axes in a), but the points were mostly plotted well. The values of the coefficients of the equation of the regression line $y = ax + b$ were often given not to the required 3 significant figure accuracy, and incurred a penalty. The regression line was often drawn not passing through point M and the y-intercept. The responses to the last part of the question were particularly weak, and many candidates were not able to offer a satisfactory reason to support their conclusion.

B.g **Part B: Scatter plot and Regression line**

Many candidates reversed the axes in a), but the points were mostly plotted well. The values of the coefficients of the equation of the regression line $y = ax + b$ were often given not to the required 3 significant figure accuracy, and incurred a penalty. The regression line was often drawn not passing through point M and the y-intercept. The responses to the last part of the question were particularly weak, and many candidates were not able to offer a satisfactory reason to support their conclusion.

The diagram below shows a square based right pyramid. ABCD is a square of side 10 cm. VX is the perpendicular height of 8 cm. M is the midpoint of BC.

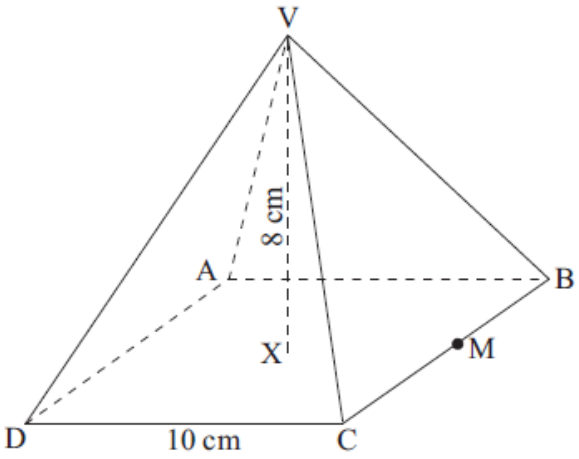


diagram not to scale

In a mountain region there appears to be a relationship between the number of trees growing in the region and the depth of snow in winter. A set of 10 areas was chosen, and in each area the number of trees was counted and the depth of snow measured. The results are given in the table below.

Number of trees (x)	Depth of snow in cm (y)
45	30
75	50
66	40
27	25
44	30
28	5
60	35
35	20
73	45
47	25

A path goes around a forest so that it forms the three sides of a triangle. The lengths of two sides are 550 m and 290 m. These two sides meet at an angle of 115°. A diagram is shown below.

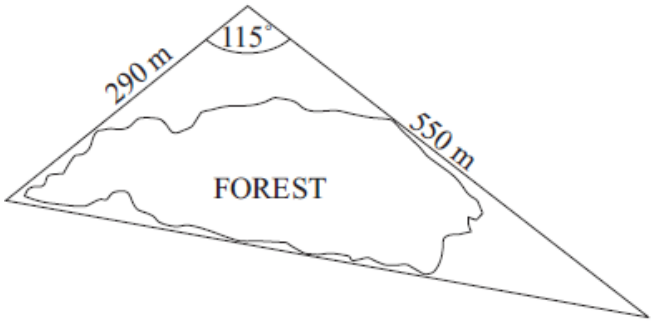


diagram not to scale

A, a, i. Write down the length of XM. [1]

A, a, ii. Use your graphic display calculator to find the standard deviation of the number of trees. [1]

A, b. Calculate the length of VM. [2]

A, c. Calculate the angle between VM and ABCD. [2]

B, a. Calculate the length of the third side of the triangle. Give your answer correct to the nearest 10 m. [4]

B, b. Calculate the area enclosed by the path that goes around the forest. [3]

B, c. Inside the forest a second path forms the three sides of another triangle named ABC. Angle BAC is 53°, AC is 180 m and BC is 230 m. [4]

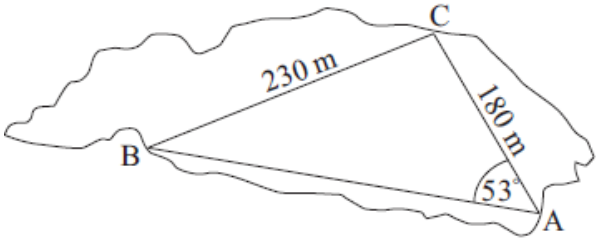


diagram not to scale

Calculate the size of angle ACB.

Markscheme

A, a. ***UP applies in this question***

(UP) XM = 5 cm ***(A1)***

[1 mark]

A, a, i. 6.18 ***(G1)***

[1 mark]

A, b. ***UP applies in this question***

$VM^2 = 5^2 + 8^2$ ***(M1)***

Note: Award ***(M1)*** for correct use of Pythagoras Theorem.

(UP) $VM = \sqrt{89} = 9.43$ cm ***(A1)*(ft)*(G2)***

[2 marks]

A, d. $\tan \mathrm{VMX} = \frac{8}{5}$ ***(M1)***

Note: Other trigonometric ratios may be used.

$\hat{\mathrm{VMX}} = 58.0^\circ$ ***(A1)*****(ft)*****(G2)***

[2 marks]

B, a. ***UP applies in this question***

$l^2 = 290^2 + 550^2 - 2 \times 290 \times 550 \times \cos 115^\circ$ ***(M1)(A1)***

Note: Award ***(M1)*** for substituted cosine rule formula, ***(A1)*** for correct substitution.

$l = 722$ ***(A1)(G2)***

(UP) = 720 m ***(A1)***

Note: If 720 m seen without working award ***(G3)***.

The final ***(A1)*** is awarded for the correct rounding of their answer.

[4 marks]

B, b. ***UP applies in this question***

$\text{Area} = \frac{1}{2} \times 290 \times 550 \times \sin 115$ ***(M1)(A1)***

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

(UP) $= 72\,300\ \text{m}^2$ ***(A1)(G2)***

[3 marks]

B, c. $\frac{180}{\sin \mathrm{B}} = \frac{230}{\sin 53}$ ***(M1)(A1)***

Note: Award ***(M1)*** for substituted sine rule formula, ***(A1)*** for correct substitution.

$\mathrm{B} = 38.7^\circ$ ***(A1)(G2)***

$\hat{\mathrm{ACB}} = 180 - (53^\circ + 38.7^\circ)$

$= 88.3^\circ$ ***(A1)*****(ft)**

[4 marks]

Examiners report

A, a. This part proved accessible to the great majority of candidates. The common errors were (1) the inversion of the tangent ratio (2) the omission of the units and (3) the incorrect rounding of the answer; with 58° being all too commonly seen.

A, a A straightforward question that saw many fine attempts. Given its nature – where much of the work was done on the GDC – it must be emphasised to candidates that incorrect entry of data into the calculator will result in considerable penalties; they must check their data entry most carefully.

The use of the inappropriate standard deviation was seen, but infrequently.

A, b This part proved accessible to the great majority of candidates. The common errors were (1) the inversion of the tangent ratio (2) the omission of the units and (3) the incorrect rounding of the answer; with 58° being all too commonly seen.

A, c This part proved accessible to the great majority of candidates. The common errors were (1) the inversion of the tangent ratio (2) the omission of the units and (3) the incorrect rounding of the answer; with 58° being all too commonly seen.

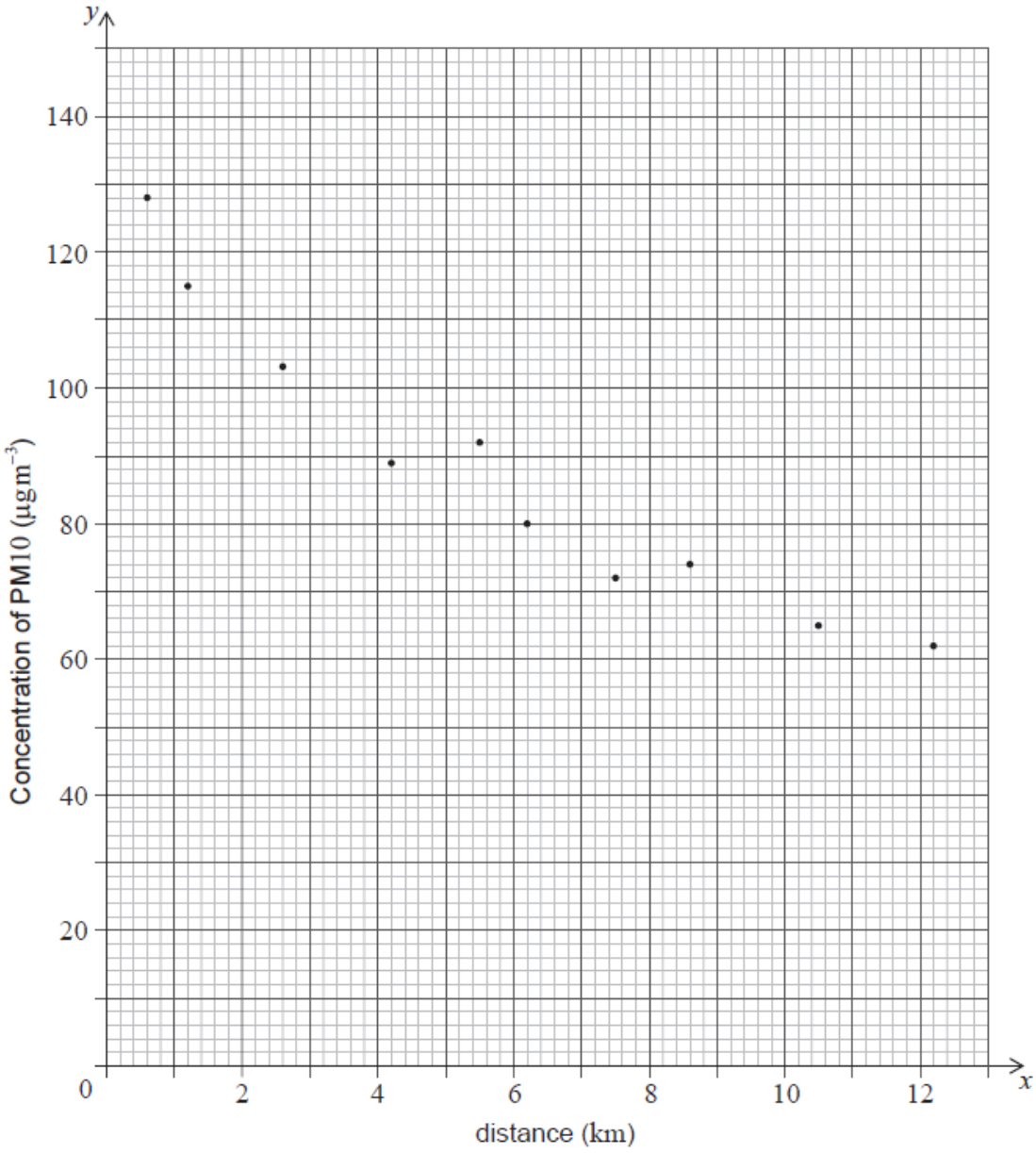
B, a Again, this part proved accessible to the majority with a large number of candidates attaining full marks. However, there were also a number of candidates who seemed not to have been prepared in the use of trigonometry in non right-angled triangles. Also, failing to round the answer in (a) to the nearest 10m was a common omission.

B, b Again, this part proved accessible to the majority with a large number of candidates attaining full marks. However, there were also a number of candidates who seemed not to have been prepared in the use of trigonometry in non right-angled triangles. Also, failing to round the answer in (a) to the nearest 10 m was a common omission.

B, c Again, this part proved accessible to the majority with a large number of candidates attaining full marks. However, there were also a number of candidates who seemed not to have been prepared in the use of trigonometry in non right-angled triangles. Also, failing to round the answer in (a) to the nearest 10 m was a common omission.

---

a. For an ecological study, Ernesto measured the average concentration (y) of the fine dust, $\mathrm{PM10}$, in the air at different distances (x) from a power plant. His data are represented on the following scatter diagram. The concentration of $\mathrm{PM10}$ is measured in micrograms per cubic metre and the distance is measured in kilometres. [2]



His data are also listed in the following table.

Distance (x)	0.6	1.2	2.6	a	5.5	6.2	7.5	8.6	10.5	12.2
Concentration of PM10 (y)	128	115	103	89	92	80	72	b	65	62

Use the scatter diagram to find the value of a and of b in the table.

b. Calculate [4]

i) $\bar{x}$, the mean distance from the power plant;

ii) $\bar{y}$, the mean concentration of $\mathrm{PM10}$;

iii) r , the Pearson's product–moment correlation coefficient.

c. Write down the equation of the regression line y on x . [2]

d. Ernesto's school is located $14\,\mathrm{km}$ from the power plant. He uses the equation of the regression line to estimate the concentration of $\mathrm{PM10}$ in the air at his school. [4]

i) Calculate the value of Ernesto's estimate.

ii) State whether Ernesto's estimate is reliable. Justify your answer.

Markscheme

a. $a = 4.2\,;\ b = 74$ ***(A1)(A1)***

b. i) $5.91\,(\text{km})$ ***(A1)*****(ft)**

ii) 88 (micrograms per cubic metre) ***(A1)*****(ft)**

Note: Follow through from part (a) irrespective of working seen.

iii) $-0.956\ \ (-0.955528...)$ ***(G2)*****(ft)**

Note: Follow through from part (a) irrespective of working seen.

c. $y = -5.39x + 120\ \ (y = -5.38955...x + 119.852...)$ ***(A1)*****(ft)*****(A1)*****(ft)**

Note: Award ***(A1)*****(ft)** for -5.39. Award ***(A1)*****(ft)** for 120. If answer is not an equation award at most ***(A1)*****(ft)*****(A0)***. Follow through from part (a) irrespective of working seen.

d. i) $-5.38955... \times 14 + 119.852...$ ***(M1)***

Note: Award ***(M1)*** for correct substitution into their regression line.

$= 44.4\ (44.3984...)$ ***(A1)*****(ft)*****(G2)***

Note: Follow through from part (c). Accept $44.5\ (44.54)$ from use of 3 significant figure values.

ii) Ernesto's estimate is not reliable ***(A1)***

this is extrapolation ***(R1)***

OR

$14\,\text{km}$ is not within the range (outside the domain) of distances given ***(R1)***

Note: Do not accept "14 is too high" or "14 is an outlier" or "result not valid/not reliable" if explanation not given. Do not award ***(A1)(R0)***. Do not accept reasoning based on the strength of r.

Examiners report

a. Question 1: Reading scatter diagram, mean, correlation and regression line.

The majority of the candidates scored very well on this question. There were only a few candidates who read the diagram incorrectly. The most common mistake in parts (b), (c) and (d)(i) were rounding errors, sometimes resulting in candidates losing follow-through marks when working was not presented. Part (d)(ii) was answered incorrectly by most candidates. The most common incorrect answer was based on strong correlation. Some commented on the trend of decreasing PM10 values for increasing distances, showing lack of understanding about extrapolation.

b. Question 1: Reading scatter diagram, mean, correlation and regression line.

The majority of the candidates scored very well on this question. There were only a few candidates who read the diagram incorrectly. The most common mistake in parts (b), (c) and (d)(i) were rounding errors, sometimes resulting in candidates losing follow-through marks when working was not presented. Part (d)(ii) was answered incorrectly by most candidates. The most common incorrect answer was based on strong correlation. Some commented on the trend of decreasing PM10 values for increasing distances, showing lack of understanding about extrapolation.

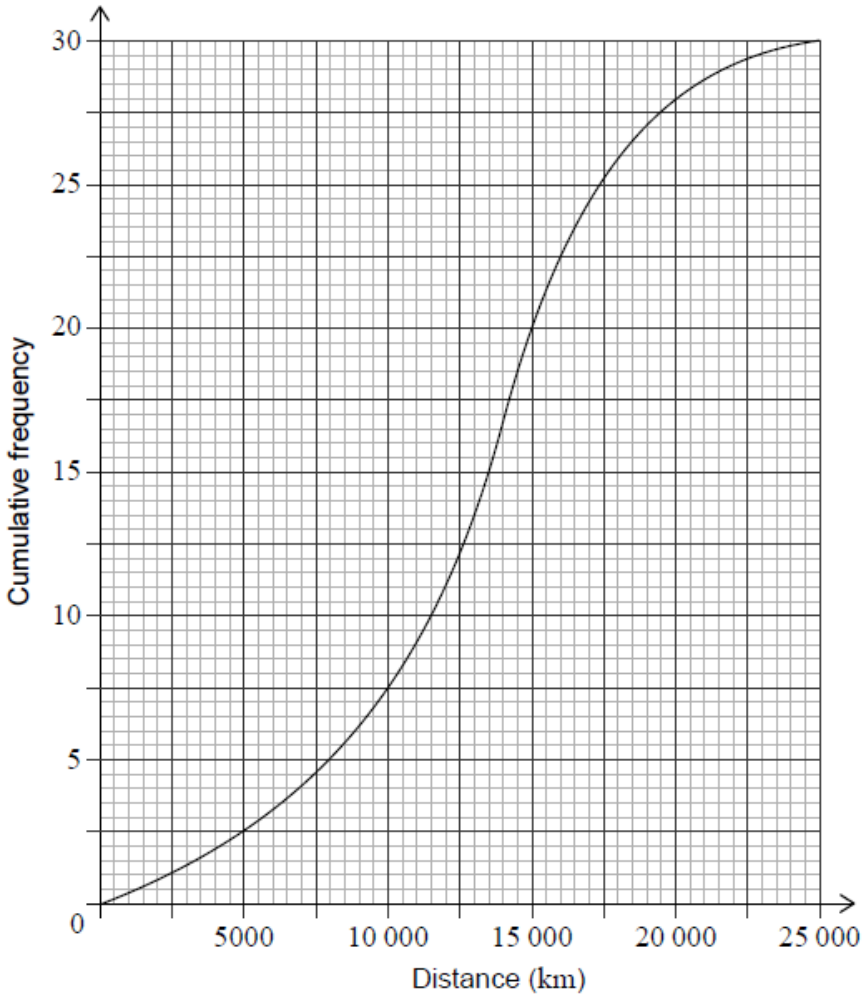
c. Question 1: Reading scatter diagram, mean, correlation and regression line.

The majority of the candidates scored very well on this question. There were only a few candidates who read the diagram incorrectly. The most common mistake in parts (b), (c) and (d)(i) were rounding errors, sometimes resulting in candidates losing follow-through marks when working was not presented. Part (d)(ii) was answered incorrectly by most candidates. The most common incorrect answer was based on strong correlation. Some commented on the trend of decreasing PM10 values for increasing distances, showing lack of understanding about extrapolation.

d. Question 1: Reading scatter diagram, mean, correlation and regression line.

The majority of the candidates scored very well on this question. There were only a few candidates who read the diagram incorrectly. The most common mistake in parts (b), (c) and (d)(i) were rounding errors, sometimes resulting in candidates losing follow-through marks when working was not presented. Part (d)(ii) was answered incorrectly by most candidates. The most common incorrect answer was based on strong correlation. Some commented on the trend of decreasing PM10 values for increasing distances, showing lack of understanding about extrapolation.

A transportation company owns 30 buses. The distance that each bus has travelled since being purchased by the company is recorded. The cumulative frequency curve for these data is shown.



It is known that 8 buses travelled more than m kilometres.

a. Find the number of buses that travelled a distance between 15000 and 20000 kilometres. [2]

b.i. Use the cumulative frequency curve to find the median distance. [2]

b.ii. Use the cumulative frequency curve to find the lower quartile. [1]

b.iii. Use the cumulative frequency curve to find the upper quartile. [1]

c. Hence write down the interquartile range. [1]

d. Write down the percentage of buses that travelled a distance greater than the upper quartile. [1]

e. Find the number of buses that travelled a distance less than or equal to 12000 km. [1]

f. Find the value of m. [2]

g. The smallest distance travelled by one of the buses was 2500 km. [4]

The longest distance travelled by one of the buses was 23 000 km.

On graph paper, draw a box-and-whisker diagram for these data. Use a scale of 2 cm to represent 5000 km.

Markscheme

a. 28 – 20 ***(A1)***

Note: Award ***(A1)*** for 28 and 20 seen.

8 ***(A1)(G2)***

[2 marks]

b.i. 13500 ***(G2)***

Note: Accept an answer in the range 13500 to 13750.

[2 marks]

b.ii. 10000 ***(G1)***

Note: Accept an answer in the range 10000 to 10250.

[1 mark]

b.iii. 16000 ***(G1)***

Note: Accept an answer in the range 16000 to 16250.

[1 mark]

c. 6000 ***(A1)*****(ft)**

Note: Follow through from their part (b)(ii) and (iii).

[1 mark]

d. 25% ***(A1)***

[1 mark]

e. 11 ***(G1)***

[1 mark]

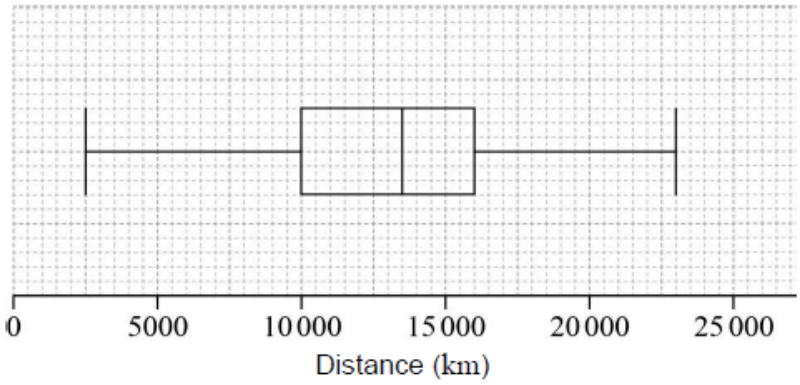
f. 30 – 8 **OR** 22 ***(M1)***

Note: Award ***(M1)*** for subtracting 30 – 8 or 22 seen.

15750 ***(A1)(G2)***

Note: Accept 15750 ± 250.

[2 marks]

g. 

(A1)(A1)(A1)(A1)

Note: Award ***(A1)*** for correct label and scale; accept "distance" or "km" for label.

(A1)**(ft)** for correct median,
(A1)**(ft)** for correct quartiles and box,
(A1) for endpoints at 2500 and 23 000 joined to box by straight lines.
Accept ±250 for the median, quartiles and endpoints.
Follow through from their part (b).
The final ***(A1)*** is not awarded if the line goes through the box.

[4 marks]

Examiners report

a. [N/A]
b.i. [N/A]
b.ii. [N/A]
b.iii. [N/A]
c. [N/A]
d. [N/A]
e. [N/A]
f. [N/A]
g. [N/A]

The heat output in thermal units from burning 1 kg of wood changes according to the wood's percentage moisture content. The moisture content and heat output of 10 blocks of the same type of wood each weighing 1 kg were measured. These are shown in the table.

Moisture content % (x)	8	15	22	30	34	45	50	60	74	82
Heat output (y)	80	77	74	69	68	61	61	55	50	45

a. Draw a scatter diagram to show the above data. Use a scale of 2 cm to represent 10% on the x-axis and a scale of 2 cm to represent 10 thermal units on the y-axis. [4]

b. Write down [2]

(i) the mean percentage moisture content, $\bar{x}$;

(ii) the mean heat output, $\bar{y}$.

c. Plot the point $(\bar{x}, \bar{y})$ on your scatter diagram and label this point M . [2]

d. Write down the product-moment correlation coefficient, r . [2]

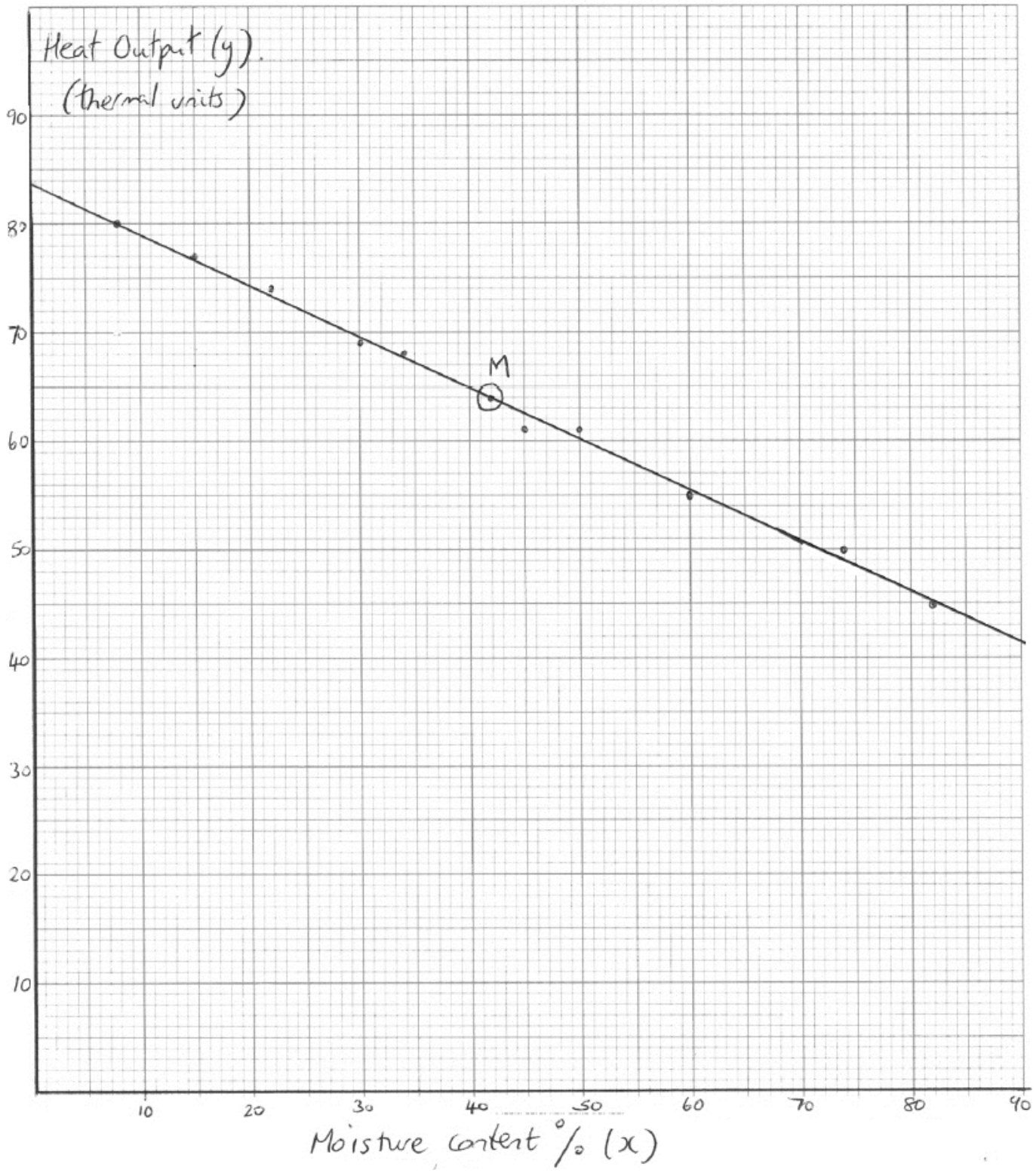
e. The equation of the regression line y on x is $y = -0.470x + 83.7$. Draw the regression line y on x on your scatter diagram. [2]

f. The equation of the regression line y on x is $y = -0.470x + 83.7$. Estimate the heat output in thermal units of a $1\ \text{kg}$ block of wood that has [2]
25% moisture content.

g. The equation of the regression line y on x is $y = -0.470x + 83.7$. State, with a reason, whether it is appropriate to use the regression line y [2]
on x to estimate the heat output in part (f).

Markscheme

a.



(A1) for correct scales and labels
(A3) for all ten points plotted correctly
(A2) for eight or nine points plotted correctly
(A1) for six or seven points plotted correctly ***(A4)***

Note: Award at most ***(A0)(A3)*** if axes reversed.

[4 marks]

b. (i) $\bar{x} = 42$ ***(A1)***

(ii) $\bar{y} = 64$ ***(A1)***

[2 marks]

c. $(\bar{x}, \bar{y})$ plotted on graph and labelled, M ***(A1)*****(ft)*****(A1)***

Note: Award ***(A1)*****(ft)** for position, ***(A1)*** for label.

[2 marks]

d. -0.998 ***(G2)***

Note: Award ***(G1)*** for correct sign, ***(G1)*** for correct absolute value.

[1 mark]

e. line on graph ***(A1)*****(ft)*****(A1)***

Notes: Award ***(A1)*****(ft)** for line through their M, **(A1)** for approximately correct intercept (allow between 83 and 85). It is not necessary that the line is seen to intersect the y-axis. The line must be straight for any mark to be awarded.

[2 marks]

f. $y = -0.470(25) + 83.7$ ***(M1)***

Note: Award ***(M1)*** for substitution into formula or some indication of method on their graph. $y = -0.470(0.25) + 83.7$ is incorrect.

$= 72.0$ (accept 71.95 and 72) ***(A1)*****(ft)*****(G2)***

Note: Follow through from graph only if they show working on their graph. Accept 72 ± 0.5 .

[2 marks]

g. Yes since 25% lies within the data set and r is close to -1 ***(R1)(A1)***

Note: Accept Yes, since r is close to -1

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

[2 marks]

Examiners report

a. The great majority of candidates found this question to be a good start to the paper. The common errors were (1) incorrect scales being used; SI units are standard in this course and candidates are expected to know the difference between centimetres and millimetres (2) the lack of r on the GDC (3) not knowing that the regression line y on x passes through the mean point and (4) not realising that the value of r determines the validity of using the regression line y on x .

b. The great majority of candidates found this question to be a good start to the paper. The common errors were (1) incorrect scales being used; SI units are standard in this course and candidates are expected to know the difference between centimetres and millimetres (2) the lack of r on the GDC (3) not knowing that the regression line y on x passes through the mean point and (4) not realising that the value of r determines the validity of using the regression line y on x .

c. The great majority of candidates found this question to be a good start to the paper. The common errors were (1) incorrect scales being used; SI units are standard in this course and candidates are expected to know the difference between centimetres and millimetres (2) the lack of r on the GDC (3) not knowing that the regression line y on x passes through the mean point and (4) not realising that the value of r determines the

validity of using the regression line y on x .

d. The great majority of candidates found this question to be a good start to the paper. The common errors were (1) incorrect scales being used; SI units are standard in this course and candidates are expected to know the difference between centimetres and millimetres (2) the lack of r on the GDC (3) not knowing that the regression line y on x passes through the mean point and (4) not realising that the value of r determines the validity of using the regression line y on x .

e. The great majority of candidates found this question to be a good start to the paper. The common errors were (1) incorrect scales being used; SI units are standard in this course and candidates are expected to know the difference between centimetres and millimetres (2) the lack of r on the GDC (3) not knowing that the regression line y on x passes through the mean point and (4) not realising that the value of r determines the validity of using the regression line y on x .

f. The great majority of candidates found this question to be a good start to the paper. The common errors were (1) incorrect scales being used; SI units are standard in this course and candidates are expected to know the difference between centimetres and millimetres (2) the lack of r on the GDC (3) not knowing that the regression line y on x passes through the mean point and (4) not realising that the value of r determines the validity of using the regression line y on x .

g. The great majority of candidates found this question to be a good start to the paper. The common errors were (1) incorrect scales being used; SI units are standard in this course and candidates are expected to know the difference between centimetres and millimetres (2) the lack of r on the GDC (3) not knowing that the regression line y on x passes through the mean point and (4) not realising that the value of r determines the validity of using the regression line y on x .

In an environmental study of plant diversity around a lake, a biologist collected data about the number of different plant species (y) that were growing at different distances (x) in metres from the lake shore.

Distance (x)	2	5	8	10	13	17	23	35	40
Plant species (y)	35	34	30	29	24	19	15	13	8

a. Draw a scatter diagram to show the data. Use a scale of 2 cm to represent 10 metres on the x-axis and 2 cm to represent 10 plant species on the y-axis. [4]

b. Using your scatter diagram, describe the correlation between the number of different plant species and the distance from the lake shore. [1]

c.i. Use your graphic display calculator to write down $\bar{x}$, the mean of the distances from the lake shore. [1]

c.ii. Use your graphic display calculator to write down $\bar{y}$, the mean number of plant species. [1]

d. Plot the point $(\bar{x}, \bar{y})$ on your scatter diagram. **Label this point M.** [2]

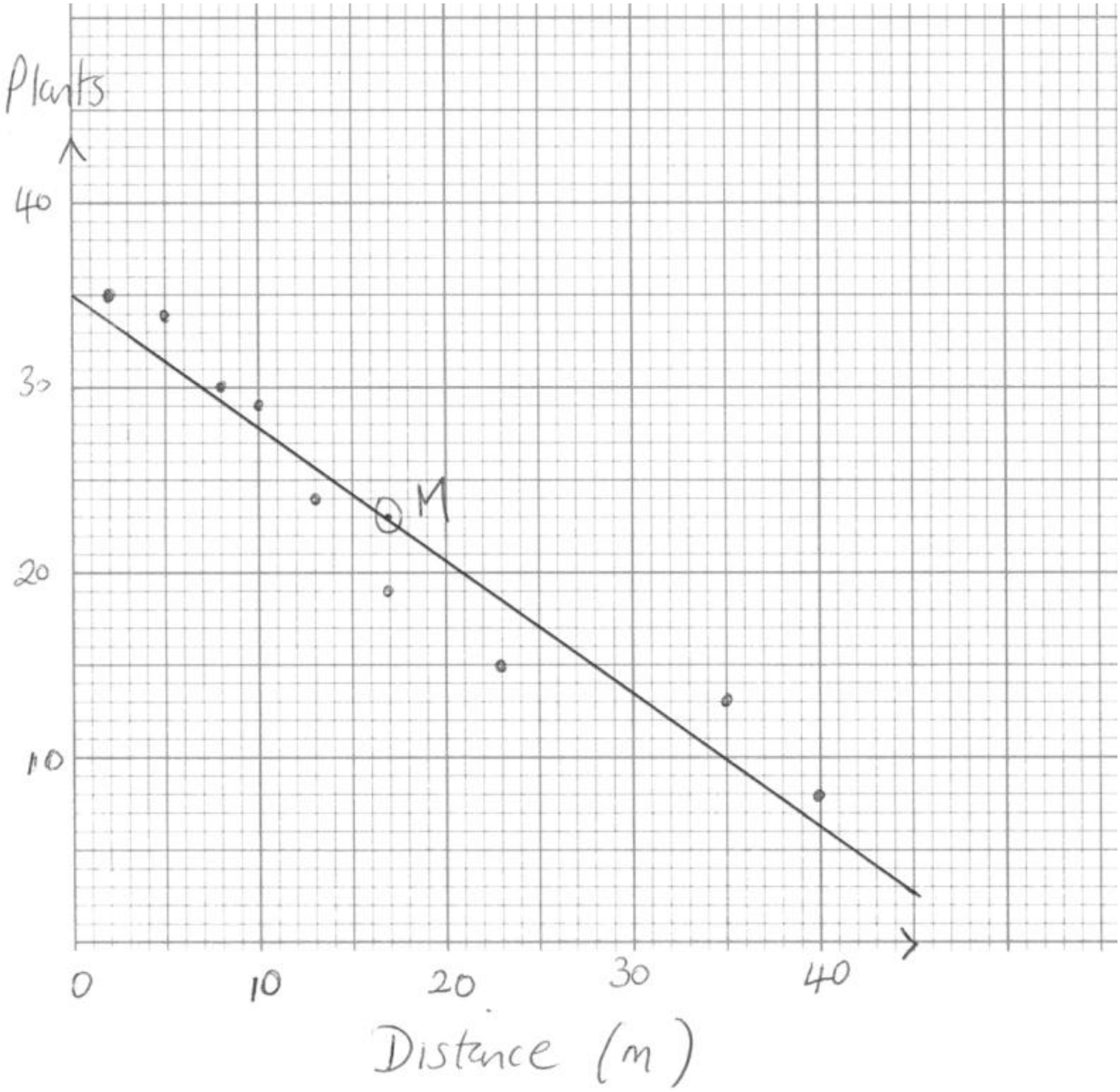
e. Write down the equation of the regression line y on x for the above data. [2]

f. Draw the regression line y on x on your scatter diagram. [2]

g. Estimate the number of plant species growing 30 metres from the lake shore. [2]

Markscheme

a.



(A1)(A3)

Notes: Award ***(A1)*** for scales and labels (accept *x*/*y*).

Award ***(A3)*** for all points correct.

Award ***(A2)*** for 7 or 8 points correct.

Award ***(A1)*** for 5 or 6 points correct.

Award at most ***(A1)(A2)*** if points are joined up.

If axes are reversed award at most ***(A0)(A3)*****(ft)**.

[4 marks]

b. Negative ***(A1)***

[1 mark]

c.i. 17 ***(G1)***

[1 mark]

c.ii. 23 ***(G1)***

[1 mark]

d. Point correctly placed and labelled M ***(A1)*****(ft)*****(A1)***

Note: Accept an error of ±0.5.

[2 marks]

e. $y = -0.708x + 35.0$ ***(G1)(G1)***

Note: Award at most ***(G1)(G0)*** if $y =$ not seen. Accept 35.

[2 marks]

f. Regression line drawn that passes through M and (0, 35) ***(A1)*****(ft)*****(A1)*****(ft)**

Note: Award ***(A1)*** for straight line that passes through M, ***(A1)*** for line (extrapolated if necessary) that passes through (0, 35) (accept error of ±1). If ruler not used, award a maximum of ***(A1)(A0)***.

[2 marks]

g. $y = -0.708(30) + 35.0$ ***(M1)***

$= 14$ *(Accept 13)* ***(A1)*****(ft)*****(G2)***

OR

Using graph: ***(M1)*** for some indication on graph of point, ***(A1)*****(ft)** for answers. Final answer must be consistent with their graph. ***(M1)(A1)*****(ft)*****(G2)***

Note: The final answer must be an integer.

[2 marks]

Examiners report

a. This question, by far, was the most accessible to the great majority of candidates. However, far too many candidates do not (1) use the scale as required by the question, (2) use a scale at all, (3) either draw or label axes, (4) use a ruler at all (5) use the provided graph paper. Accurate plotting of points can not be assessed unless graph paper has been used; the diagram is not a graph.

Many candidates did not seem aware that the regression line must pass through the mean point. Others, though they had obtained the equation of the regression line, did not use it to identify its y intercept.

b. This question, by far, was the most accessible to the great majority of candidates. However, far too many candidates do not (1) use the scale as required by the question, (2) use a scale at all, (3) either draw or label axes, (4) use a ruler at all (5) use the provided graph paper. Accurate plotting of points can not be assessed unless graph paper has been used; the diagram is not a graph.

Many candidates did not seem aware that the regression line must pass through the mean point. Others, though they had obtained the equation of the regression line, did not use it to identify its *y* intercept.

c.i. This question, by far, was the most accessible to the great majority of candidates. However, far too many candidates do not (1) use the scale as required by the question, (2) use a scale at all, (3) either draw or label axes, (4) use a ruler at all (5) use the provided graph paper. Accurate plotting of points can not be assessed unless graph paper has been used; the diagram is not a graph.

Many candidates did not seem aware that the regression line must pass through the mean point. Others, though they had obtained the equation of the regression line, did not use it to identify its *y* intercept.

c.ii. This question, by far, was the most accessible to the great majority of candidates. However, far too many candidates do not (1) use the scale as required by the question, (2) use a scale at all, (3) either draw or label axes, (4) use a ruler at all (5) use the provided graph paper. Accurate plotting of points can not be assessed unless graph paper has been used; the diagram is not a graph.

Many candidates did not seem aware that the regression line must pass through the mean point. Others, though they had obtained the equation of the regression line, did not use it to identify its *y* intercept.

d. This question, by far, was the most accessible to the great majority of candidates. However, far too many candidates do not (1) use the scale as required by the question, (2) use a scale at all, (3) either draw or label axes, (4) use a ruler at all (5) use the provided graph paper. Accurate plotting of points can not be assessed unless graph paper has been used; the diagram is not a graph.

Many candidates did not seem aware that the regression line must pass through the mean point. Others, though they had obtained the equation of the regression line, did not use it to identify its *y* intercept.

e. This question, by far, was the most accessible to the great majority of candidates. However, far too many candidates do not (1) use the scale as required by the question, (2) use a scale at all, (3) either draw or label axes, (4) use a ruler at all (5) use the provided graph paper. Accurate plotting of points can not be assessed unless graph paper has been used; the diagram is not a graph.

Many candidates did not seem aware that the regression line must pass through the mean point. Others, though they had obtained the equation of the regression line, did not use it to identify its *y* intercept.

f. This question, by far, was the most accessible to the great majority of candidates. However, far too many candidates do not (1) use the scale as required by the question, (2) use a scale at all, (3) either draw or label axes, (4) use a ruler at all (5) use the provided graph paper. Accurate plotting of points can not be assessed unless graph paper has been used; the diagram is not a graph.

Many candidates did not seem aware that the regression line must pass through the mean point. Others, though they had obtained the equation of the regression line, did not use it to identify its *y* intercept.

g. This question, by far, was the most accessible to the great majority of candidates. However, far too many candidates do not (1) use the scale as required by the question, (2) use a scale at all, (3) either draw or label axes, (4) use a ruler at all (5) use the provided graph paper. Accurate plotting of points can not be assessed unless graph paper has been used; the diagram is not a graph.

Many candidates did not seem aware that the regression line must pass through the mean point. Others, though they had obtained the equation of the regression line, did not use it to identify its *y* intercept.

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

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

b. State the modal group. [1]

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

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

(i) the mean;

(ii) the standard deviation.

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

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

Write down the value of

(i) q;

(ii) r.

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

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.

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

(i) the median;

(ii) the 90th percentile of the results;

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

Markscheme

a. continuous ***(A1)***

[1 mark]

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

[1 mark]

c. 15 ***(A1)***

[1 mark]

d. (i) 21.5 ***(G2)***

(ii) 9.21 (9.20597…) ***(G1)***

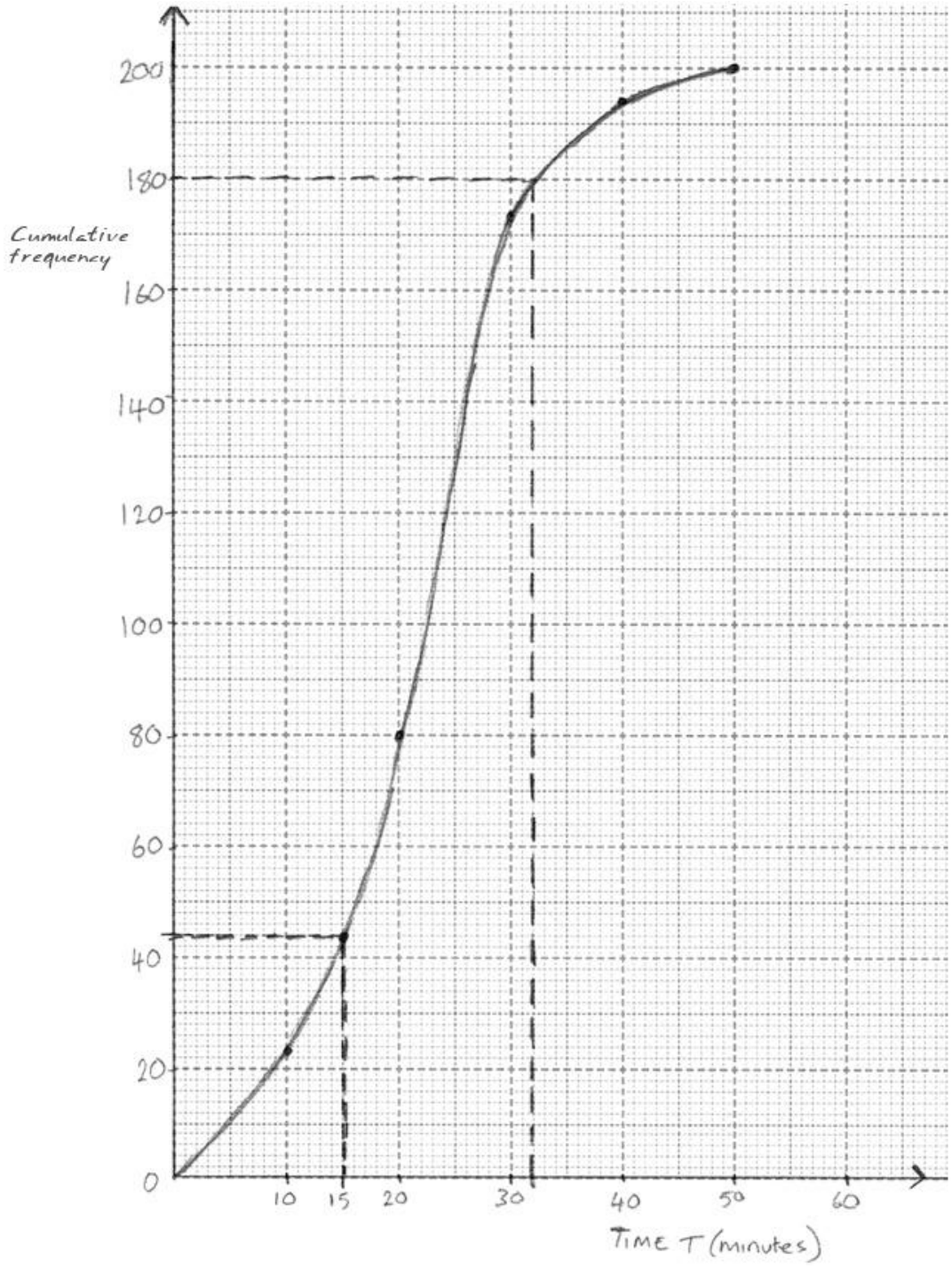
[3 marks]

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

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

[2 marks]

f.



(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]

g. (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

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

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

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

d. (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.

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

f. (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.

g. (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.