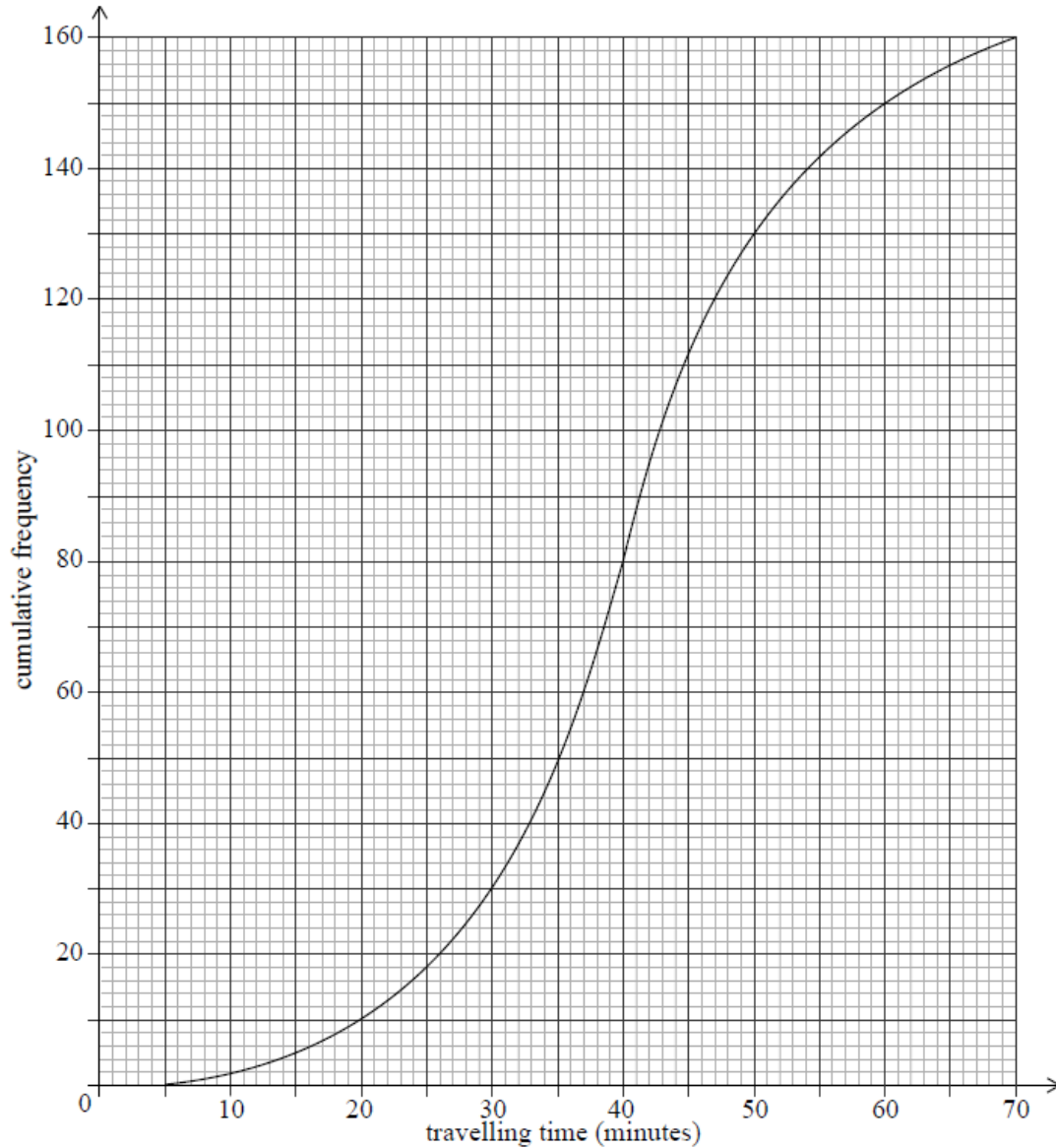


Stats - review (TL) [97 marks]

1. [Maximum mark: 15]

SPM.1.SL.TZ0.7

A large company surveyed 160 of its employees to find out how much time they spend traveling to work on a given day. The results of the survey are shown in the following cumulative frequency diagram.



(a) Find the median number of minutes spent traveling to work.

[2]

Markscheme

evidence of median position (M1)

80th employee

40 minutes **A1**

[2 marks]

- (b) Find the number of employees whose travelling time is within 15 minutes of the median. [3]

Markscheme

valid attempt to find interval (25–55) **(M1)**

18 (employees), 142 (employees) **A1**

124 **A1**

[3 marks]

Only 10% of the employees spent more than k minutes traveling to work.

- (c) Find the value of k . [3]

Markscheme

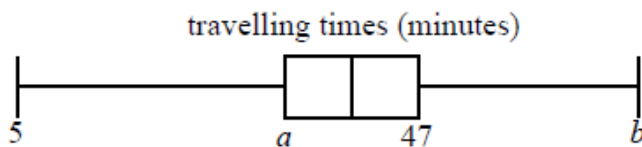
recognising that there are 16 employees in the top 10% **(M1)**

144 employees travelled more than k minutes **(A1)**

$k = 56$ **A1**

[3 marks]

The results of the survey can also be displayed on the following box-and-whisker diagram.



- (d) Write down the value of b . [1]

Markscheme

$$b = 70 \quad A1$$

[1 mark]

(e.i) Find the value of a .

[2]

Markscheme

recognizing a is first quartile value (M1)

40 employees

$$a = 33 \quad A1$$

[2 marks]

(e.ii) Hence, find the interquartile range.

[2]

Markscheme

$$47 - 33 \quad (M1)$$

$$\text{IQR} = 14 \quad A1$$

[2 marks]

(f) Travelling times of less than p minutes are considered outliers.

Find the value of p .

[2]

Markscheme

attempt to find $1.5 \times$ **their** IQR (M1)

$$33 - 21$$

$$12 \quad (A1)$$

[2 marks]

2. [Maximum mark: 4]

EXN.2.SL.TZ0.1

A data set consisting of 16 test scores has mean 14.5. One test score of 9 requires a second marking and is removed from the data set.

Find the mean of the remaining 15 test scores.

[4]

Markscheme

* This sample question was produced by experienced DP mathematics senior examiners to aid teachers in preparing for external assessment in the new MAA course. There may be minor differences in formatting compared to formal exam papers.

$$\frac{\sum_{i=1}^{16} x_i}{16} = 14.5 \quad (\mathbf{M1})$$

Note: Award **M1** for use of $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$.

$$\Rightarrow \sum_{i=1}^{16} x_i = 232 \quad (\mathbf{A1})$$

$$\text{new } \bar{x} = \frac{232-9}{15} \quad (\mathbf{A1})$$

$$= 14.9 \quad \left(= 14.8\bar{6}, = \frac{223}{15} \right) \quad \mathbf{A1}$$

Note: Do not accept 15.

[4 marks]

3. [Maximum mark: 8]

EXN.2.SL.TZ0.4

The following table shows the systolic blood pressures, p mmHg, and the ages, t years, of 6 male patients at a medical clinic.

Patient	P1	P2	P3	P4	P5	P6
t (years)	40	72	35	47	21	61
p (mmHg)	105	145	100	130	95	132

(a.i) Determine the value of Pearson's product-moment correlation coefficient, r , for these data.

[2]

Markscheme

* This sample question was produced by experienced DP mathematics senior examiners to aid teachers in preparing for external assessment in the new MAA course. There may be minor differences in formatting compared to formal exam papers.

$$r = 0.946 \quad \mathbf{A2}$$

[2 marks]

(a.ii) Interpret, in context, the value of r found in part (a) (i).

[1]

Markscheme

the value of r shows a (very) strong positive correlation between age and (systolic) blood pressure **A1**

[1 mark]

The relationship between t and p can be modelled by the regression line of p on t with equation $p = at + b$.

(b) Find the equation of the regression line of p on t .

[2]

Markscheme

$$p = 1.05t + 69.3 \quad \mathbf{A1A1}$$

Note: Only award marks for an equation. Award **A1** for $a = 1.05$ and **A1** for $b = 69.3$. Award **A1A0** for $y = 1.05x + 69.3$.

[2 marks]

A 50-year-old male patient enters the medical clinic for his appointment.

- (c) Use the regression equation from part (b) to predict this patient's systolic blood pressure.

[2]

Markscheme

122 (mmHg) **(M1)A1**

[2 marks]

- (d) A 16-year-old male patient enters the medical clinic for his appointment.

Explain why the regression equation from part (b) should not be used to predict this patient's systolic blood pressure.

[1]

Markscheme

the regression equation should not be used because it involves extrapolation **A1**

[1 mark]

4. [Maximum mark: 5]

22N.2.SL.TZ0.1

The following table shows the Mathematics test scores (x) and the Science test scores (y) for a group of eight students.

Mathematics scores (x)	64	68	72	75	80	82	85	86
Science scores (y)	67	72	77	76	84	83	89	91

The regression line of y on x for this data can be written in the form $y = ax + b$.

(a) Find the value of a and the value of b .

[2]

Markscheme

1.01206..., 2.45230...

$a = 1.01$, $b = 2.45$ ($1.01x + 2.45$) **A1A1**

[2 marks]

(b) Write down the value of the Pearson's product-moment correlation coefficient, r .

[1]

Markscheme

0.981464...

$r = 0.981$ **A1**

Note: A common error is to enter the data incorrectly into the GDC, and obtain the answers $a = 1.01700\dots$, $b = 2.09814\dots$ and $r = 0.980888\dots$. Some candidates may write the 3 sf answers, ie. $a = 1.02$, $b = 2.10$ and $r = 0.981$ or 2 sf answers, ie. $a = 1.0$, $b = 2.1$ and $r = 0.98$. In these cases award **A0A0** for part (a) and **A0** for part (b). Even though some values round to an accepted answer, they come from incorrect working.

[1 mark]

- (c) Use the equation of your regression line to predict the Science test score for a student who has a score of 78 on the Mathematics test. Express your answer to the nearest integer.

[2]

Markscheme

correct substitution of 78 into **their** regression equation (M1)

81. 3930 . . . 81. 23 from 3 sf answer

81 A1

[2 marks]

5. [Maximum mark: 4]

22M.2.SL.TZ1.2

The number of hours spent exercising each week by a group of students is shown in the following table.

Exercising time (in hours)	Number of students
2	5
3	1
4	4
5	3
6	x

The median is 4.5 hours.

(a) Find the value of x .

[2]

Markscheme

EITHER

recognising that half the total frequency is 10 (may be seen in an ordered list or indicated on the frequency table) (A1)

OR

$$5 + 1 + 4 = 3 + x \quad (A1)$$

OR

$$\sum f = 20 \quad (A1)$$

THEN

$$x = 7 \quad A1$$

[2 marks]

(b) Find the standard deviation.

[2]

Markscheme

METHOD 1

1.58429...

1.58 A2

METHOD 2

EITHER

$$\sigma^2 = \frac{5 \times (2-4.3)^2 + 1 \times (3-4.3)^2 + 4 \times (4-4.3)^2 + 3 \times (5-4.3)^2 + 7 \times (6-4.3)^2}{20} \quad (= 2.51)$$

(A1)

OR

$$\sigma^2 = \frac{5 \times 2^2 + 1 \times 3^2 + 4 \times 4^2 + 3 \times 5^2 + 7 \times 6^2}{20} - 4.3^2 \quad (= 2.51) \quad (A1)$$

THEN

$$\sigma = \sqrt{2.51} = 1.58429 \dots$$

= 1.58 A1

[2 marks]

6. [Maximum mark: 6]

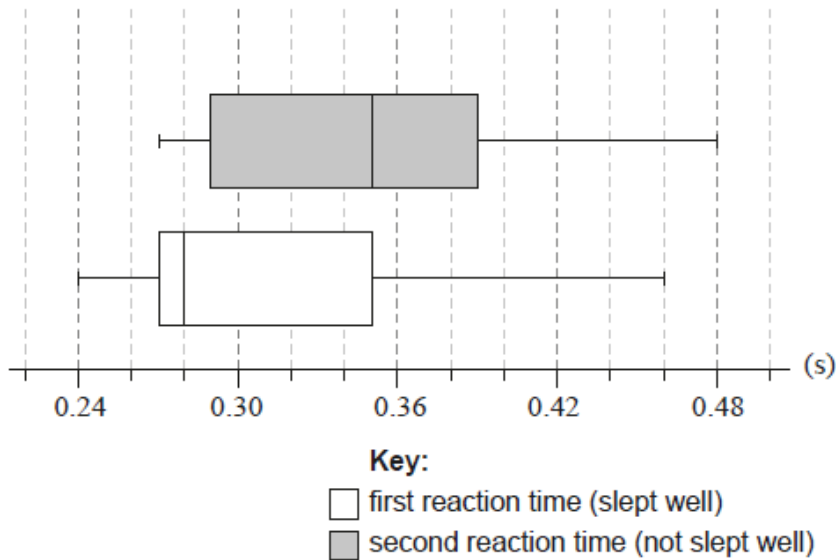
22M.2.SL.TZ2.5

A random sample of nine adults were selected to see whether sleeping well affected their reaction times to a visual stimulus. Each adult's reaction time was measured twice.

The first measurement for reaction time was taken on a morning after the adult had slept well.

The second measurement was taken on a morning after the same adult had not slept well.

The box and whisker diagrams for the reaction times, measured in seconds, are shown below.



Consider the box and whisker diagram representing the reaction times after sleeping well.

(a) State the median reaction time after sleeping well.

[1]

Markscheme

0.28 (s) **A1**

[1 mark]

(b) Verify that the measurement of 0.46 seconds is not an outlier.

[3]

Markscheme

$\text{IQR} = 0.35 - 0.27 (= 0.08) \text{ (s)}$ **(A1)**

substituting **their IQR** into correct expression for upper fence **(A1)**

$$0.35 + 1.5 \times 0.08 (= 0.47) \text{ (s)}$$

$$0.46 < 0.47 \quad \mathbf{R1}$$

so 0.46 (s) is not an outlier **AG**

[3 marks]

- (c) State why it appears that the mean reaction time is greater than the median reaction time.

[1]

Markscheme

EITHER

the median is closer to the lower quartile (positively skewed) **R1**

OR

The distribution is positively skewed **R1**

OR

the range of reaction times below the median is smaller than the range of reaction times above the median **R1**

Note: These are sample answers from a range of acceptable correct answers. Award **R1** for any correct statement that explains this.

Do not award **R1** if there is also an incorrect statement, even if another statement in the answer is correct. Accept a correctly and clearly labelled diagram.

[1 mark]

- (d) Now consider the two box and whisker diagrams.

Comment on whether these box and whisker diagrams provide any evidence that might suggest that not sleeping well causes an increase in reaction time.

[1]

Markscheme

EITHER

the distribution for 'not sleeping well' is centred at a higher reaction time *R1*

OR

The median reaction time after not sleeping well is equal to the upper quartile reaction time after sleeping well *R1*

OR

75% of reaction times are < 0.35 seconds after sleeping well, compared with 50% after not sleeping well *R1*

OR

the sample size of 9 is too small to draw any conclusions *R1*

Note: These are sample answers from a range of acceptable correct answers. Accept any relevant correct statement **that relates to the median and/or quartiles shown in the box plots. Do not accept** a comparison of means. Do not award *R1* if there is also an incorrect statement, even if another statement in the answer is correct.

Award *R0* to "correlation does not imply causation".

[1 mark]

7. [Maximum mark: 5]

21N.2.SL.TZ0.1

In Lucy's music academy, eight students took their piano diploma examination and achieved scores out of 150. For her records, Lucy decided to record the average number of hours per week each student reported practising in the weeks prior to their examination. These results are summarized in the table below.

Average weekly practice time (h)	28	13	45	33	17	29	39	36
Diploma score (D)	115	82	120	116	79	101	110	121

(a) Find Pearson's product-moment correlation coefficient, r , for these data.

[2]

Markscheme

use of GDC to give $(M1)$

$$r = 0.883529\dots$$

$$r = 0.884 \quad A1$$

Note: Award the $(M1)$ for any correct value of r, a, b or $r^2 = 0.780624\dots$ seen in part (a) or part (b).

[2 marks]

(b) The relationship between the variables can be modelled by the regression equation $D = ah + b$. Write down the value of a and the value of b .

[1]

Markscheme

$$a = 1.36609\dots, b = 64.5171\dots$$

$$a = 1.37, b = 64.5 \quad A1$$

[1 mark]

- (c) One of these eight students was disappointed with her result and wished she had practised more. Based on the given data, determine how her score could have been expected to alter had she practised an extra five hours per week.

[2]

Markscheme

attempt to find their difference (M1)

$5 \times 1.36609 \dots$ OR

$1.36609 \dots (h + 5) + 64.5171 \dots - (1.36609 \dots h + 64.5171 \dots)$

$6.83045 \dots$

$= 6.83$ (6.85 from 1.37)

the student could have expected her score to increase by 7 marks. A1

Note: Accept an increase of 6, 6.83 or 6.85.

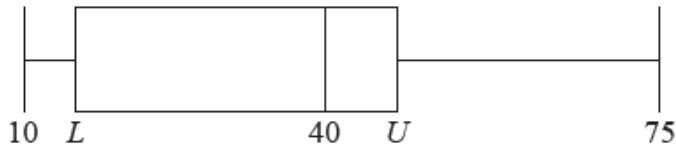
[2 marks]

8. [Maximum mark: 5]

21M.1.SL.TZ1.4

A research student weighed lizard eggs in grams and recorded the results. The following box and whisker diagram shows a summary of the results where L and U are the lower and upper quartiles respectively.

diagram not to scale



The interquartile range is 20 grams and there are no outliers in the results.

(a) Find the minimum possible value of U .

[3]

Markscheme

attempt to use definition of outlier

$$1.5 \times 20 + Q_3 \quad (M1)$$

$$1.5 \times 20 + U \geq 75 \quad (\Rightarrow U \geq 45, \text{ accept } U > 45) \quad \text{OR} \quad 1.5 \times 20 + Q_3 = 75$$

A1

$$\text{minimum value of } U = 45 \quad A1$$

[3 marks]

(b) Hence, find the minimum possible value of L .

[2]

Markscheme

attempt to use interquartile range (M1)

$$U - L = 20 \quad (\text{may be seen in part (a)}) \quad \text{OR} \quad L \geq 25 \quad (\text{accept } L > 25)$$

$$\text{minimum value of } L = 25 \quad A1$$

[2 marks]

9. [Maximum mark: 6]

21M.2.SL.TZ2.1

At a café, the waiting time between ordering and receiving a cup of coffee is dependent upon the number of customers who have already ordered their coffee and are waiting to receive it.

Sarah, a regular customer, visited the café on five consecutive days. The following table shows the number of customers, x , ahead of Sarah who have already ordered and are waiting to receive their coffee and Sarah's waiting time, y minutes.

Number of customers (x)	3	9	11	10	5
Sarah's waiting time (y)	6	10	12	11	6

The relationship between x and y can be modelled by the regression line of y on x with equation $y = ax + b$.

(a.i) Find the value of a and the value of b .

[2]

Markscheme

$$a = 0.805084\dots \text{ and } b = 2.88135\dots$$

$$a = 0.805 \text{ and } b = 2.88 \quad \mathbf{A1A1}$$

[2 marks]

(a.ii) Write down the value of Pearson's product-moment correlation coefficient, r .

[1]

Markscheme

$$r = 0.97777\dots$$

$$r = 0.978 \quad \mathbf{A1}$$

[1 mark]

(b) Interpret, in context, the value of a found in part (a)(i).

[1]

Markscheme

a represents the (average) increase in waiting time (0.805 mins) per additional customer (waiting to receive their coffee) **R1**

[1 mark]

- (c) On another day, Sarah visits the café to order a coffee. Seven customers have already ordered their coffee and are waiting to receive it.

Use the result from part (a)(i) to estimate Sarah's waiting time to receive her coffee.

[2]

Markscheme

attempt to substitute $x = 7$ into their equation **(M1)**

8.51693...

8.52 (mins) **A1**

[2 marks]

10. [Maximum mark: 6]

20N.1.SL.TZ0.T_3

Hafizah harvested 49 mangoes from her farm. The weights of the mangoes, w , in grams, are shown in the following grouped frequency table.

Weight (g)	$100 \leq w < 200$	$200 \leq w < 300$	$300 \leq w < 400$	$400 \leq w < 500$	$500 \leq w < 600$
Frequency	4	7	14	16	8

(a) Write down the modal group for these data.

[1]

Markscheme

* This question is from an exam for a previous syllabus, and may contain minor differences in marking or structure. It appeared in a paper that permitted the use of a calculator, and so might not be suitable for all forms of practice.

$400 \leq w < 500$ (A1) (C1)

Note: Accept alternative notation $[400, 500)$ or $[400, 500[$.
Do not accept "400-500".

[1 mark]

(b) Use your graphic display calculator to find an estimate of the standard deviation of the weights of mangoes from this harvest.

[2]

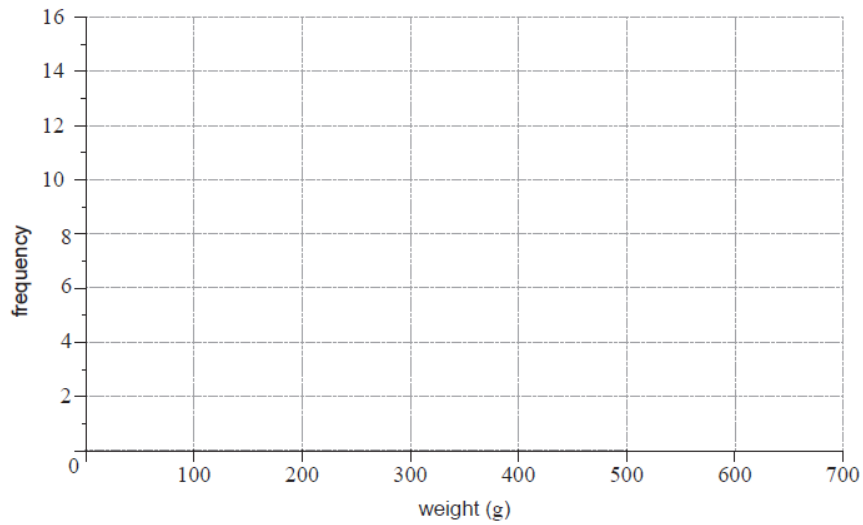
Markscheme

115 (115.265... (g)) (A2) (C2)

Note: Award (A1)(A0) for an answer of 116 (116.459...).

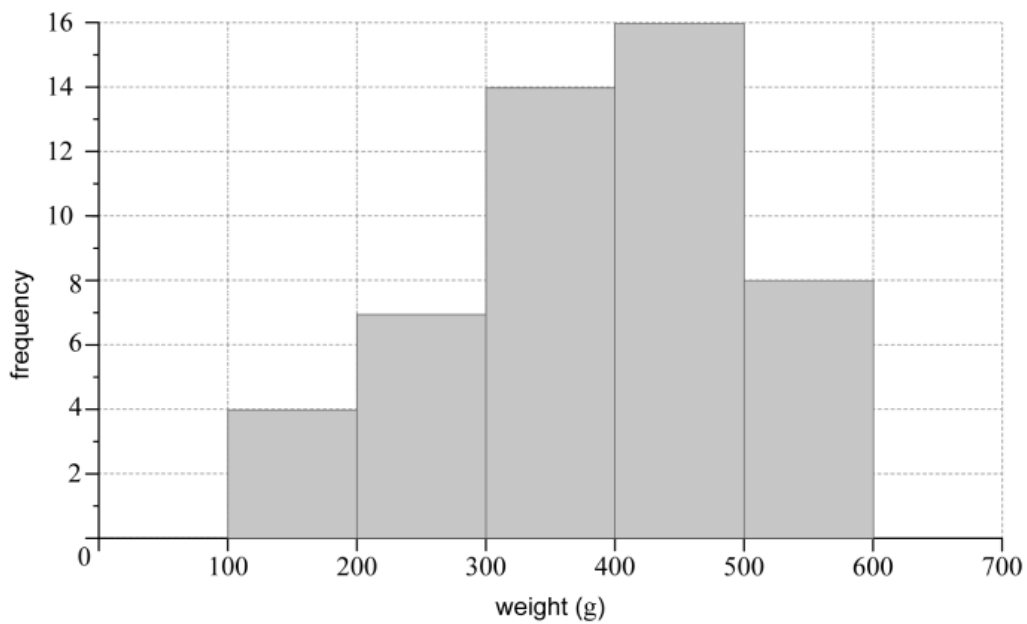
[2 marks]

(c) On the grid below, draw a histogram for the data in the table.



[3]

Markscheme



(A2)(A1) (C3)

Note: Award (A2) for all correct heights of bars or (A1) for three or four correct heights of bars.

Award (A1) for rectangular bars all with correct left and right end points (100, 200, 300, 400, 500 and 600) and for no gaps; the bars do **not** have to be shaded.

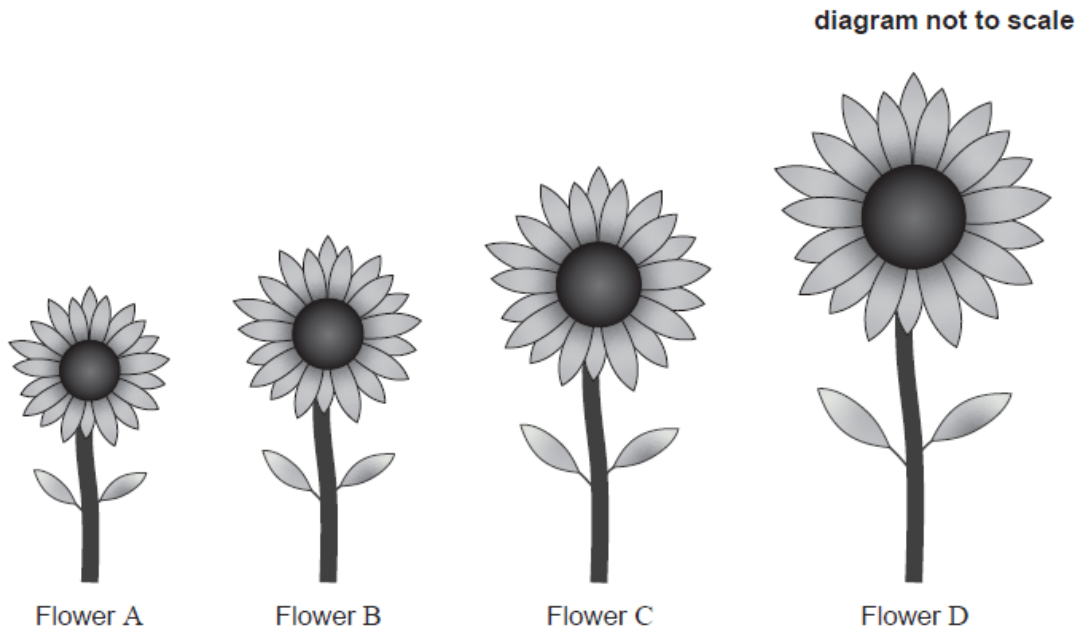
Award at most (A2)(A0) if a ruler is not used for all lines.

[3 marks]

11. [Maximum mark: 6]

20N.1.SL.TZ0.T_7

Anne-Marie planted four sunflowers in order of height, from shortest to tallest.



Flower C is 32 cm tall.

The median height of the flowers is 24 cm.

(a) Find the height of Flower A.

[2]

Markscheme

* This question is from an exam for a previous syllabus, and may contain minor differences in marking or structure. It appeared in a paper that permitted the use of a calculator, and so might not be suitable for all forms of practice.

$$24 - 8 \text{ OR } 24 - (32 - 24) \text{ OR } 24 = \frac{32+h}{2} \quad (M1)$$

Note: Award (M1) for subtracting 8 from the median, or equivalent.

16 (cm) (A1) (C2)

[2 marks]

The range of the heights is 50 cm. The height of Flower A is p cm and the height of Flower D is q cm.

(b) Using this information, write down an equation in p and q .

[1]

Markscheme

$$q - p = 50 \text{ (or equivalent) } \textbf{(A1) (C1)}$$

[1 mark]

The mean height of the flowers is 27 cm.

(c) Write down a second equation in p and q .

[1]

Markscheme

$$\frac{p+16+32+q}{4} = 27 \textbf{ OR } p + q = 60 \text{ (or equivalent) } \textbf{(A1)(ft) (C1)}$$

Note: Follow through from part (a).

[1 mark]

(d.i) Using your answers to **parts (b) and (c)**, find the height of Flower A.

[1]

Markscheme

$$5 \text{ (cm) } \textbf{(A1)(ft) (C1)}$$

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

[1 mark]

(d.ii) Using your answers to **parts (b) and (c)**, find the height of Flower D.

[1]

Markscheme

55 (cm) (A1)(ft) (C1)

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

[1 mark]

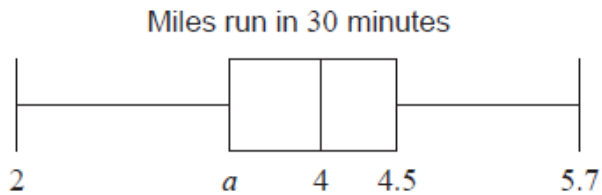
12. [Maximum mark: 15]

20N.1.SL.TZ0.S_8

Each athlete on a running team recorded the distance (M miles) they ran in 30 minutes.

The median distance is 4 miles and the interquartile range is 1.1 miles.

This information is shown in the following box-and-whisker plot.



(a) Find the value of a .

[2]

Markscheme

*This question is from an exam for a previous syllabus, and may contain minor differences in marking or structure.

valid approach (M1)

$$\text{eg } Q_3 - Q_1, Q_3 - 1.1, 4.5 - a = 1.1$$

$$a = 3.4 \quad \text{A1 N2}$$

[2 marks]

The distance in miles, M , can be converted to the distance in kilometres, K , using the formula $K = \frac{8}{5}M$.

(b) Write down the value of the median distance in kilometres (km).

[1]

Markscheme

$$\frac{32}{5} (= 6.4) \text{ (km)} \quad \text{A1 N1}$$

[1 mark]

The variance of the distances run by the athletes is $\frac{16}{9} \text{ km}^2$.

The standard deviation of the distances is b miles.

(c) Find the value of b .

[4]

Markscheme

METHOD 1 (standard deviation first)

valid approach (M1)

$$\text{eg standard deviation} = \sqrt{\text{variance}}, \sqrt{\frac{16}{9}}$$

$$\text{standard deviation} = \frac{4}{3} \text{ (km)} \quad (\text{A1})$$

valid approach to convert **their** standard deviation (M1)

$$\text{eg } \frac{4}{3} \times \frac{5}{8}, \sqrt{\frac{16}{9}} = \frac{8}{5} M$$

$$\frac{20}{24} \text{ (miles)} \left(= \frac{5}{6} \right) \quad \text{A1 N3}$$

Note: If no working shown, award **M1A1MOA0** for the value $\frac{4}{3}$.

If working shown, and candidate's final answer is $\frac{4}{3}$, award **M1A1MOA0**.

METHOD 2 (variance first)

valid approach to convert variance (M1)

$$\text{eg } \left(\frac{5}{8}\right)^2, \frac{64}{25}, \frac{16}{9} \times \left(\frac{5}{8}\right)^2$$

$$\text{variance} = \frac{25}{36} \quad (\text{A1})$$

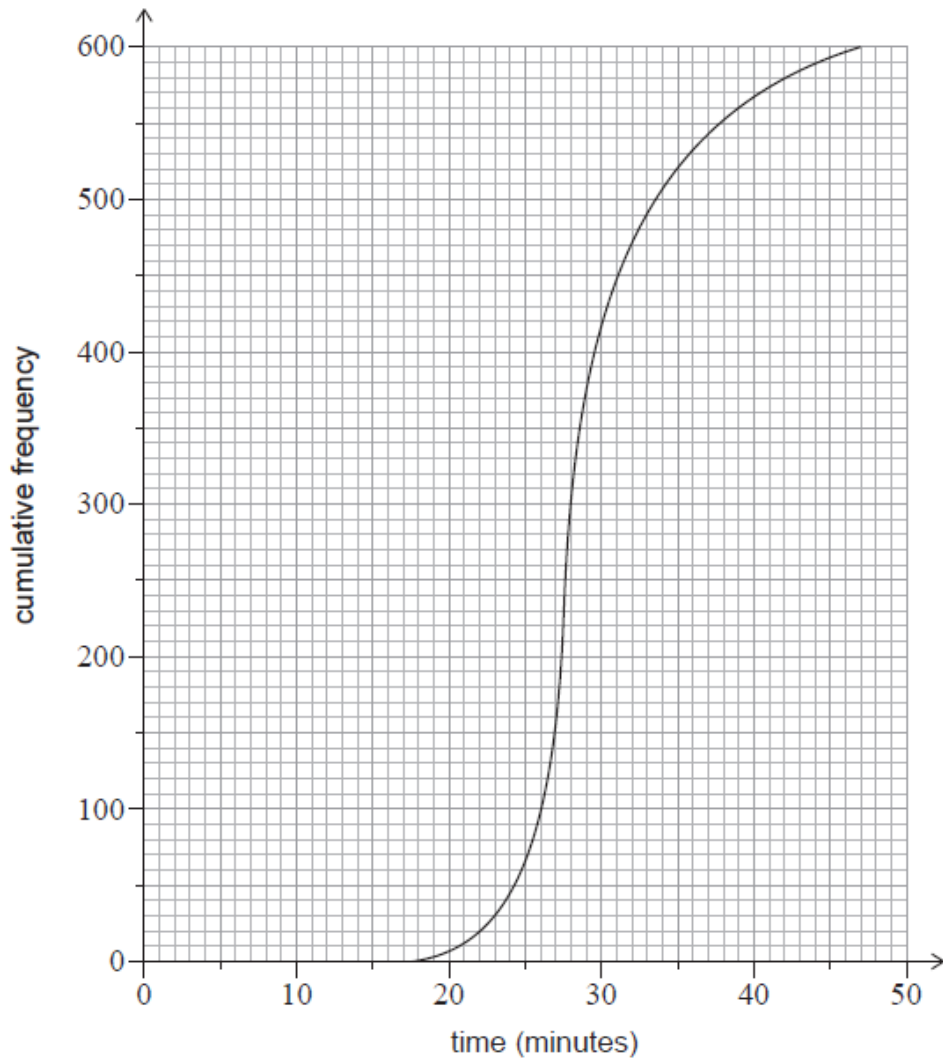
valid approach (M1)

$$\text{eg standard deviation} = \sqrt{\text{variance}}, \sqrt{\frac{25}{36}}, \sqrt{\frac{16}{9} \times \left(\frac{5}{8}\right)^2}$$

$$\frac{20}{24} \text{ (miles)} \left(= \frac{5}{6} \right) \quad \text{A1 N3}$$

[4 marks]

A total of 600 athletes from different teams compete in a 5 km race. The times the 600 athletes took to run the 5 km race are shown in the following cumulative frequency graph.



There were 400 athletes who took between 22 and m minutes to complete the 5 km race.

(d) Find m .

[3]

Markscheme

correct frequency for 22 minutes (A1)

eg 20

adding **their** frequency (do not accept $22 + 400$) (M1)

eg $20 + 400$, 420 athletes

$m = 30$ (minutes) A1 N3

[3 marks]

- (e) The first 150 athletes that completed the race won a prize.

Given that an athlete took between 22 and m minutes to complete the 5 km race, calculate the probability that they won a prize.

[5]

Markscheme

27 (minutes) (A1)

correct working (A1)

eg 130 athletes between 22 and 27 minutes, $P(22 < t < 27) = \frac{150-20}{600}$, $\frac{13}{60}$

evidence of conditional probability or reduced sample space (M1)

eg $P(A|B)$, $P(t < 27 | 22 < t < 30)$, $\frac{P(22 < t < 27)}{P(22 < t < m)}$, $\frac{150}{400}$

correct working (A1)

eg $\frac{\frac{130}{600}}{\frac{400}{600}}$, $\frac{150-20}{400}$

$\frac{130}{400}$ ($\frac{13}{40} = \frac{78000}{240000} = \frac{390}{1200} = 0.325$) A1 N5

Note: If no other working is shown, award **AOAOM1AOAO** for answer of $\frac{150}{400}$.

Award **NO** for answer of $\frac{3}{8}$ with no other working shown.

[5 marks]

13. [Maximum mark: 6]

19M.1.SL.TZ1.T_2

The fastest recorded speeds of eight animals are shown in the following table.

Animal	Speed (km h ⁻¹)
Golden eagle	300
Swordfish	97
Hare	80
Lion	80
Horse	71
Zebra	64
Komodo dragon	21
Tiger beetle	6

(a) State whether **speed** is a continuous or discrete variable.

[1]

Markscheme

* This question is from an exam for a previous syllabus, and may contain minor differences in marking or structure.

continuous (A1)(C1)

[1 mark]

(b) Write down the median speed for these animals.

[1]

Markscheme

75.5 (km h⁻¹) (A1)(C1)

Note: Answer must be exact.

[1 mark]

(c) Write down the range of the animal speeds.

[1]

Markscheme

294 (km h⁻¹) (A1)(C1)

[1 mark]

(d.i) For these eight animals find the mean speed.

[2]

Markscheme

$$\frac{300+97+80+80+71+64+21+6}{8} \text{ OR } \frac{719}{8} \quad (M1)$$

Note: Award (M1) for correct sum divided by 8.

89.9 (89.875)(km h⁻¹) (A1)(C2)

[2 marks]

(d.ii) For these eight animals write down the standard deviation.

[1]

Markscheme

84.6 (84.5597...)(km h⁻¹) (A1)(C1)

Note: If the response to part (d)(i) is awarded zero marks, a correct response to part (d)(ii) is awarded (C2).

[1 mark]

14. [Maximum mark: 6]

19M.1.SL.TZ2.T_12

University students were surveyed and asked how many hours, h , they worked each month. The results are shown in the following table.

Hours per month, h	Frequency	Cumulative frequency
$0 < h \leq 10$	3	3
$10 < h \leq 20$	7	10
$20 < h \leq 30$	10	20
$30 < h \leq 40$	14	34
$40 < h \leq 50$	p	44
$50 < h \leq 60$	6	50
$60 < h \leq 70$	4	54
$70 < h \leq 80$	2	q

Use the table to find the following values.

(a.i) p .

[1]

Markscheme

$$p = 10 \quad (A1) (C1)$$

Note: Award (A1) for each correct value.

[1 mark]

(a.ii) q .

[1]

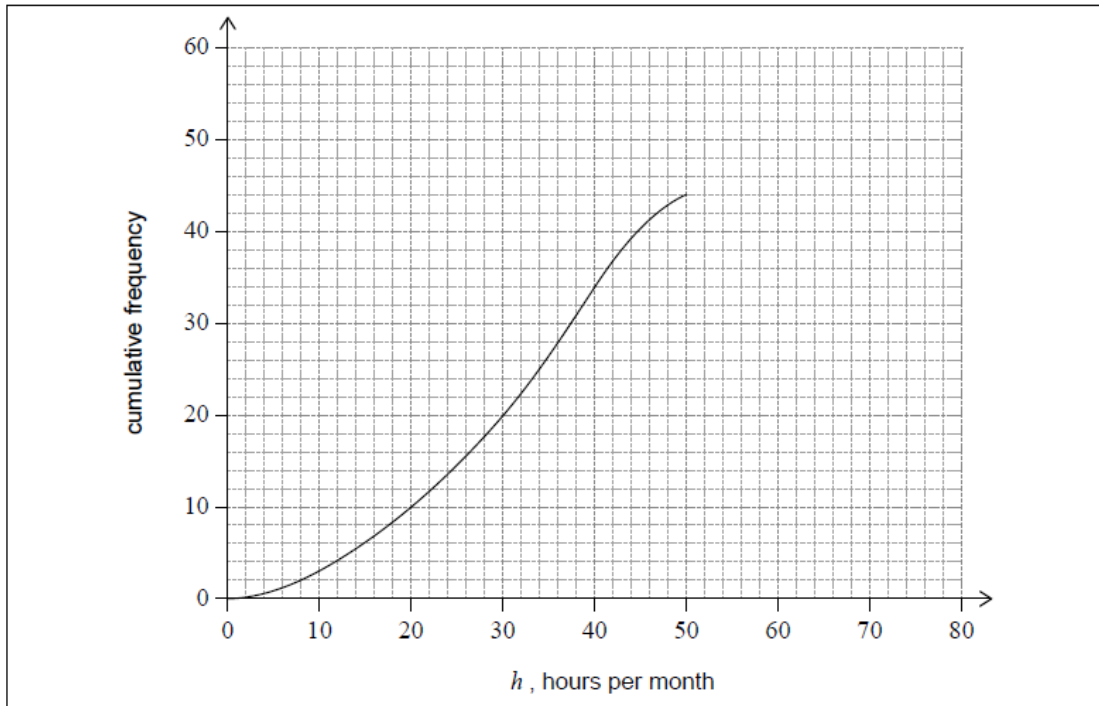
Markscheme

$$q = 56 \quad (A1) (C1)$$

Note: Award (A1) for each correct value.

[1 mark]

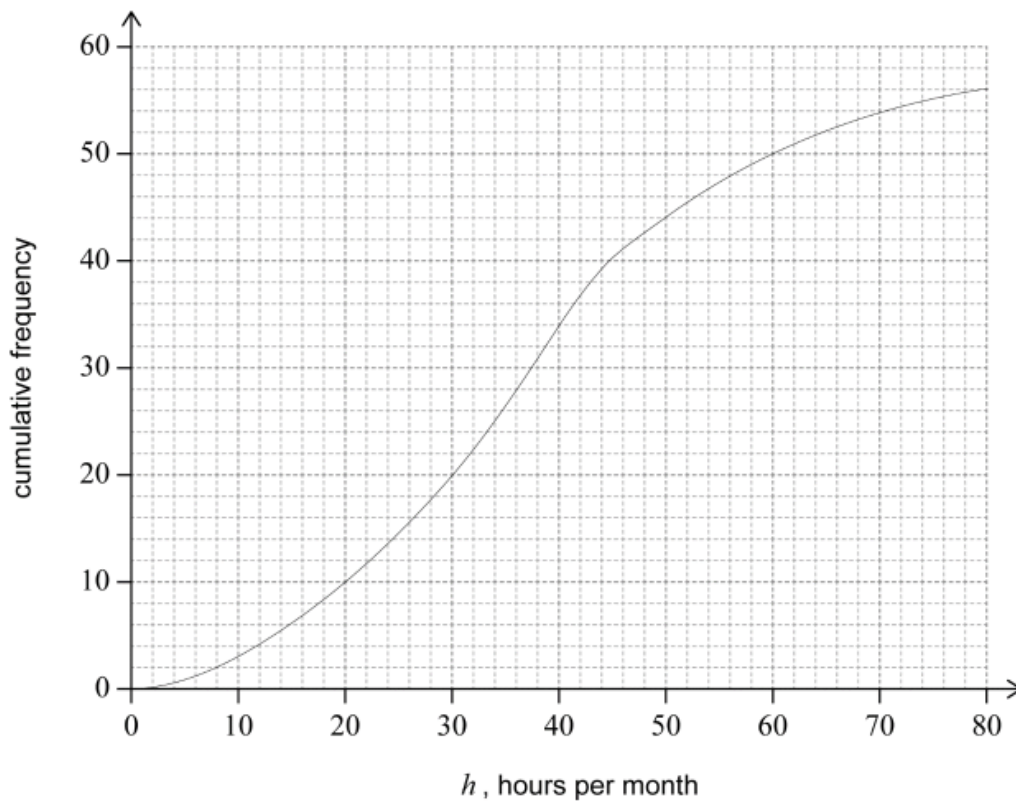
The first five class intervals, indicated in the table, have been used to draw part of a cumulative frequency curve as shown.



(b) On the same grid, complete the cumulative frequency curve for these data.

[2]

Markscheme



(A1)(A1) (C2)

Note: Award (A1)(ft) for their 3 correctly plotted points; award (A1)(ft) for completing diagram with a smooth curve through their points. The second (A1)(ft) can follow through from incorrect points, provided the gradient of the curve is never negative. Award (C2) for a completely correct smooth curve that goes through the correct points.

[2 marks]

- (c) Use the cumulative frequency curve to find an estimate for the number of students who worked at most 35 hours per month.

[2]

Markscheme

a straight vertical line drawn at 35 (accept 35 ± 1) (M1)

26 (students) (A1) (C2)

Note: Accept values between 25 and 27 inclusive.

[2 marks]

