AI HL 29.09 [92 marks]

1. [Maximum mark: 6] 18M.1.SL.TZ1.T_2

Each month the number of days of rain in Cardiff is recorded.
The following data was collected over a period of 10 months.

11 13 8 11 8 7 8 14 x 15

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

(a) Find the value of x. [2]

Markscheme

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

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

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

$(x =)\ 9$ ***(A1)(C2)***

[2 marks]

(b.i) Find the standard deviation [2]

Markscheme

2.69 (2.69072…) ***(A2)*****(ft)**

Note: Follow through from part (a).

[2 marks]

(b.ii) Find the interquartile range. [2]

Markscheme

13 – 8 ***(M1)***

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

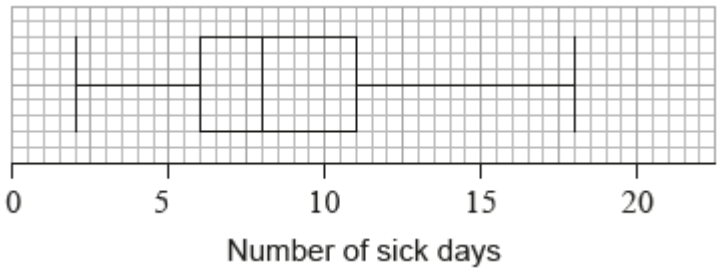
= 5 ***(A1)*****(ft)** ***(C4)***

Note: Follow through from part (a).

[2 marks]

2. [Maximum mark: 5] 21M.1.SL.TZ2.5

The number of sick days taken by each employee in a company during a year was recorded. The data was organized in a box and whisker diagram as shown below:



For this data, write down

(a.i) the minimum number of sick days taken during the year. [1]

Markscheme

2 ***A1***

[1 mark]

(a.ii) the lower quartile. [1]

Markscheme

6 ***A1***

[1 mark]

(a.iii) the median. [1]

Markscheme

8 ***A1***

[1 mark]

(b) Paul claims that this box and whisker diagram can be used to infer that the percentage of employees who took fewer than six sick days is smaller than the percentage of employees who took more than eleven sick days.

State whether Paul is correct. Justify your answer. [2]

Markscheme

EITHER

Each of these percentages represent approximately 25% of the employees. ***R1***

OR

The diagram is not explicit enough to show what is happening at the quartiles regarding 6 and 11 / we do not have the data points ***R1***

OR

Discrete data not clear how to interpret "fewer". ***R1***

THEN

Hence, Paul is not correct (**OR** no such inference can be made). ***A1***

Note: Do not award ***R0A1***.

[2 marks]

3. [Maximum mark: 6] 18M.1.SL.TZ1.S_2

The following box-and-whisker plot shows the number of text messages sent by students in a school on a particular day.

(a) Find the value of the interquartile range. [2]

Markscheme

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

recognizing Q_1 or Q_3 (seen anywhere) ***(M1)***

eg 4,11 , indicated on diagram

IQR = 7 ***A1 N2***

[2 marks]

(b) One student sent k text messages, where $k > 11$. Given that k is an outlier, find the least value of k. [4]

Markscheme

recognizing the need to find 1.5 IQR ***(M1)***

eg $1.5 \times \text{IQR}$, 1.5×7

valid approach to find k ***(M1)***

eg $10.5 + 11$, $1.5 \times \text{IQR} + Q_3$

21.5 ***(A1)***

$k = 22$ ***A1 N3***

Note: If no working shown, award ***N2*** for an answer of 21.5.

[4 marks]

4. [Maximum mark: 6] 19N.1.SL.TZ0.T_13

Chicken eggs are classified by grade $(4, 5, 6, 7$ or $8)$, based on weight. A mixed carton contains 12 eggs and could include eggs from any grade. As part of the science project, Rocky buys 9 mixed cartons and sorts the eggs according to their weight.

Grade	Weight, w (grams)	Frequency
4	$40 \leq w < 50$	3
5	$50 \leq w < 60$	30
6	$60 \leq w < 70$	45
7	$70 \leq w < 80$	25
8	$80 \leq w < 90$	5

(a) State whether the weight of the eggs is a continuous or discrete variable. [1]

Markscheme

continuous ***(A1) (C1)***

[1 mark]

(b) Write down the modal grade of the eggs. [1]

Markscheme

6 ***(A1) (C1)***

Note: Award ***(A0)*** for an answer of $60 \leq w < 70$.

[1 mark]

(c) Use your graphic display calculator to find an estimate for the standard deviation of the weight of the eggs. [2]

Markscheme

8.97 $(8.97479\ldots)$ (g) ***(A2)*** ***(C2)***

[2 marks]

(d) The mean weight of these eggs is 64.9 grams, correct to three significant figures.

Use the table and your answer to part (c) to find the **smallest possible** number of eggs that could be within one standard deviation of the mean. [2]

Markscheme

$[55.9,\ 73.9]$ **OR** $55.9252\ldots \leq w \leq 73.8747\ldots$ ***(M1)***

Note: Award ***(M1)*** for correct endpoints seen. If the answer to part (c) is $14.1421\ldots$, award ***(M1)*** for endpoints of $[50.7578\ldots,\ 79.0421\ldots]$.

45 ***(A1)*****(ft)** ***(C2)***

Note: Follow through from their part (c). For a standard deviation between 0 and 5 inclusive, the ***FT*** answer is 0.

[2 marks]

5. [Maximum mark: 6] 19N.1.SL.TZ0.T_6

Galois Airways has flights from Hong Kong International Airport to different destinations. The following table shows the distance, x kilometres, between Hong Kong and the different destinations and the corresponding airfare, y, in Hong Kong dollars (HKD).

Destination	Bali, Indonesia	Sydney, Australia	Bengaluru, India	Singapore	Auckland, New Zealand	Bangkok, Thailand
Distance x, (km)	3400	7400	4000	2600	9200	1700
Airfare y, (HKD)	1550	3600	2800	1300	4000	1400

The Pearson's product–moment correlation coefficient for this data is 0.948, correct to three significant figures.

(a) Use your graphic display calculator to find the equation of the regression line y on x. [2]

Markscheme

$y = 0.384x + 629$

$y = (0.384221\ldots)\,x + (629.421\ldots)$ ***(A1)(A1) (C2)***

Note: Award ***(A1)*** for $0.384x$, ***(A1)*** for 629. If the answer is not given as an equation, award a maximum of ***(A1)(A0)***.

[2 marks]

The distance from Hong Kong to Tokyo is 2900 km.

(b) Use your regression equation to estimate the cost of a flight from Hong Kong to Tokyo with Galois Airways. [2]

Markscheme

$y = 0.384221\ldots \times 2900 + 629.421\ldots$ ***(M1)***

Note: Award ***(M1)*** for substitution into *their* regression equation.

1740 $(1744, 1743.66\ldots)$ (HKD) ***(A1)*****(ft)** ***(C2)***

Note: Follow through from part (a).

[2 marks]

(c) Explain why it is valid to use the regression equation to estimate the airfare between Hong Kong and Tokyo. [2]

Markscheme

the correlation is (very) strong ***(R1)***

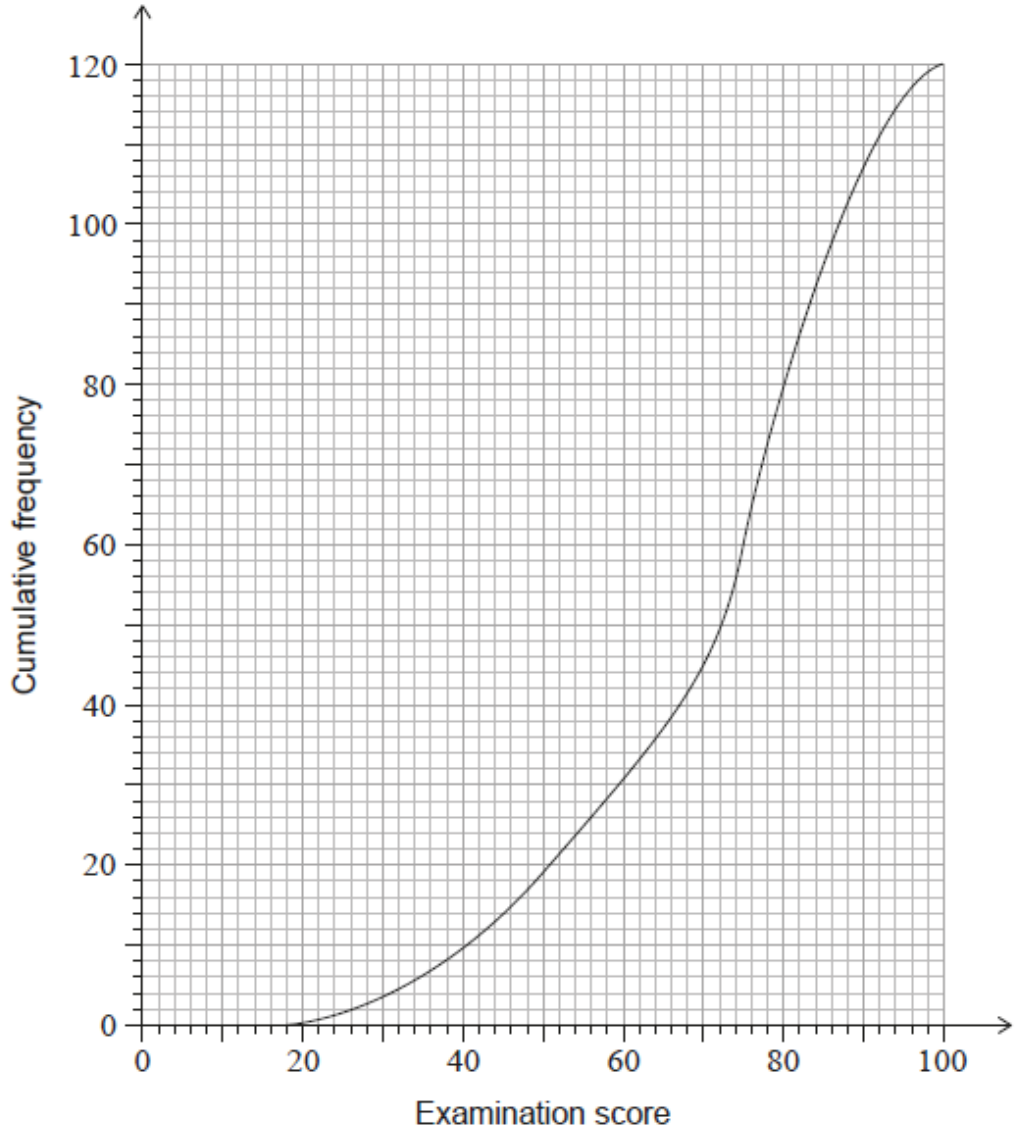
2900 (km) is within the given data range (interpolation) ***(R1)*** ***(C2)***

Note: Two correct reasons are required for the awarding of ***(C2)***.

[2 marks]

6. [Maximum mark: 8] 21N.1.SL.TZ0.9

A group of 120 students sat a history exam. The cumulative frequency graph shows the scores obtained by the students.



(a) Find the median of the scores obtained. [1]

Markscheme
75 ***A1***

[1 mark]

The students were awarded a grade from 1 to 5, depending on the score obtained in the exam. The number of students receiving each grade is shown in the following table.

Grade	1	2	3	4	5
Number of students	6	13	26	a	b

(b) Find an expression for a in terms of b. [2]

Markscheme

recognition that all entries add up to 120 ***(M1)***

$a = 120 - 6 - 13 - 26 - b$ **OR** $a = 75 - b$ ***A1***

[2 marks]

The mean grade for these students is 3.65.

(c.i) Find the number of students who obtained a grade 5. [3]

Markscheme

$\frac{6\times1+13\times2+26\times3+(75-b)\times4+b\times5}{120} = 3.65$ ***(M1)(A1)***

Note: Award ***(M1)*** for attempt to substitute into mean formula, LHS expression is sufficient for the ***M*** mark. Award ***(A1)*** for correct substitutions in one variable **OR** in two variables, followed by evidence of solving simultaneously with $a + b = 75$.

$(b =)\ 28$ ***A1***

[3 marks]

(c.ii) Find the minimum score needed to obtain a grade 5. [2]

Markscheme

$120 -$ their part (c)(i) seen (e.g. 92 indicated on graph) ***(M1)***

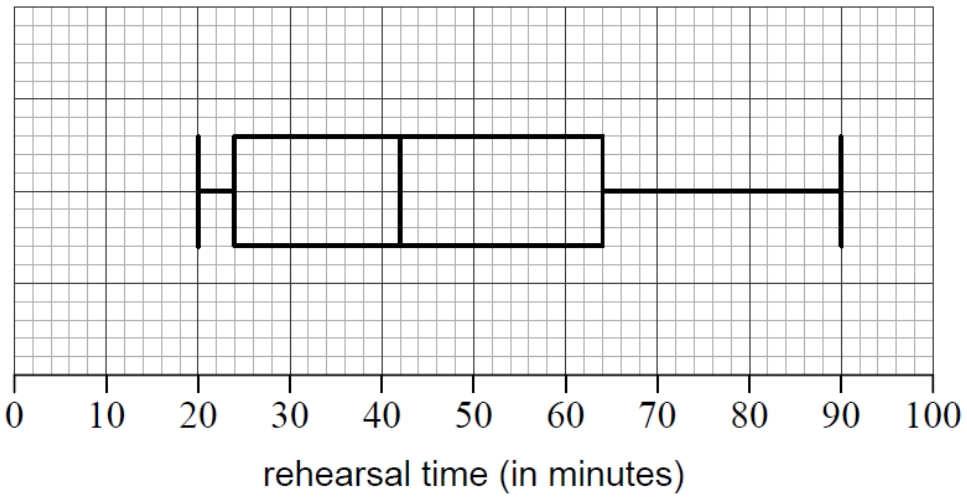
84 ***A1***

[2 marks]

7. [Maximum mark: 6] 19N.1.SL.TZ0.T_15

Stephen was invited to perform a piano recital. In preparation for the event, Stephen recorded the amount of time, in minutes, that he rehearsed each day for the piano recital.

Stephen rehearsed for 32 days and data for all these days is displayed in the following box-and-whisker diagram.



(a) Write down the median rehearsal time. [1]

Markscheme

42 (minutes) ***(A1)*** ***(C1)***

[1 mark]

Stephen states that he rehearsed on each of the 32 days.

(b) State whether Stephen is correct. Give a reason for your answer. [2]

Markscheme

Stephen is correct. ***(A1)***

the minimum rehearsal time is greater than zero *(R1)*

OR

he rehearsed at least 20 minutes every day *(R1)(C2)*

Note: Do not award *(A1)(R0)*. Accept equivalent reasoning based on the box-and-whisker diagram.

[2 marks]

(c) On k days, Stephen practiced exactly 24 minutes.

Find the possible values of k. [3]

Markscheme

$0,\ 2,\ 3,\ 4,\ \ldots,\ 15$ *(A1)(A1)(A1) (C3)*

Note: Award *(A1)(A1)* for each correct endpoint of the interval, *(A1)* for indication of integer values, except 1, between *their* endpoints.

[3 marks]

8. [Maximum mark: 5] 22M.1.SL.TZ2.7

A college runs a mathematics course in the morning. Scores for a test from this class are shown below.

$25 \quad 33 \quad 51 \quad 62 \quad 63 \quad 63 \quad 70 \quad 74 \quad 79 \quad 79 \quad 81 \quad 88 \quad 90 \quad 90 \quad 98$

For these data, the lower quartile is 62 and the upper quartile is 88.

(a) Show that the test score of 25 would not be considered an outlier. [3]

Markscheme

$(88 - 62) \times 1.5$ **OR** 26×1.5 seen anywhere **OR** 39 seen anywhere ***(M1)***

$62 - 39$

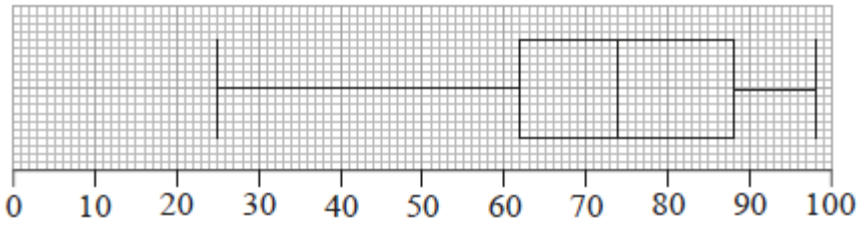
23 ***A1***

$25 > 23$ ***R1***

so is not an outlier ***AG***

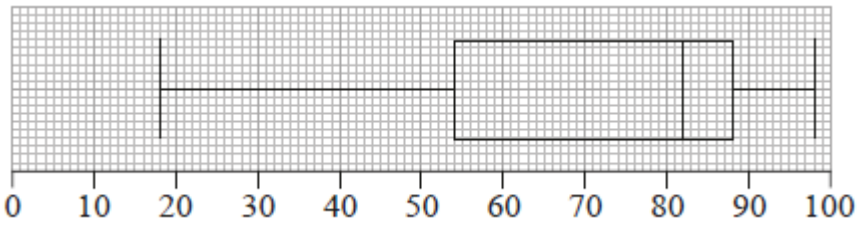
[3 marks]

The box and whisker diagram showing these scores is given below.



Test scores

Another mathematics class is run by the college during the evening. A box and whisker diagram showing the scores from this class for the same test is given below.



Test scores

A researcher reviews the box and whisker diagrams and believes that the evening class performed better than the morning class.

(b) With reference to the box and whisker diagrams, state one aspect that may support the researcher's opinion and one aspect that may counter it. [2]

Markscheme

The median score for the evening class is higher than the median score for the morning class. ***A1***

THEN

but the scores are more spread out in the evening class than in the morning class ***A1***

OR

the scores are more inconsistent in the evening class ***A1***

OR

the lowest scores are in the evening class ***A1***

OR

the interquartile range is lower in the morning class ***A1***

OR

the lower quartile is lower in the evening class ***A1***

Note: If an incorrect comparison is also made, award at most ***A1A0***.

Award ***A0*** for a comparison that references "the mean score" unless working is shown for the estimated means of the data sets, calculated from the mid-points of the 4 intervals. The estimated mean for the morning class is 71.375 and the estimated mean for the evening class is 70.5.

[2 marks]

9. [Maximum mark: 12] 17M.2.SL.TZ1.T_1

The following table shows the average body weight, x, and the average weight of the brain, y, of seven species of mammal. Both measured in kilograms (kg).

Species	Average body weight, x (kg)	Average weight of the brain, y (kg)
Cat	3	0.026
Cow	465	0.423
Donkey	187	0.419
Giraffe	529	0.680
Goat	28	0.115
Jaguar	100	0.157
Sheep	56	0.175

(a) Find the range of the average body weights for these seven species of mammal. [2]

Markscheme

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

$529 - 3$ ***(M1)***

$= 526\ (\text{kg})$ ***(A1)(G2)***

[2 marks]

(b.i) For the data from these seven species calculate r, the Pearson's product–moment correlation coefficient; [2]

Markscheme

$0.922\ (0.921857\ldots)$ ***(G2)***

[2 marks]

(b.ii) For the data from these seven species describe the correlation between the average body weight and the average weight of the brain. [2]

Markscheme

(very) strong, positive ***(A1)*****(ft)*****(A1)*****(ft)**

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

[2 marks]

(c) Write down the equation of the regression line y on x, in the form $y = mx + c$. [2]

Markscheme

$y = 0.000986x + 0.0923$ $(y = 0.000985837\ldots x + 0.0923391\ldots)$
(A1)(A1)

Note: Award ***(A1)*** for $0.000986x$, ***(A1)*** for 0.0923.

Award a maximum of ***(A1)(A0)*** if the answer is not an equation in the form $y = mx + c$.

[2 marks]

The average body weight of grey wolves is 36 kg.

(d) Use your regression line to estimate the average weight of the brain of grey wolves. [2]

Markscheme

$0.000985837\ldots(36) + 0.0923391\ldots$ ***(M1)***

Note: Award ***(M1)*** for substituting 36 into their equation.

$0.128\ (\text{kg})\ \ (0.127829\ldots\ (\text{kg}))$ ***(A1)*****(ft)*****(G2)***

Note: Follow through from part (c). The final ***(A1)*** is awarded only if their answer is positive.

[2 marks]

In fact, the average weight of the brain of grey wolves is 0.120 kg.

(e) Find the percentage error in your estimate in part (d). [2]

Markscheme

$\left|\frac{0.127829\ldots - 0.120}{0.120}\right| \times 100$ ***(M1)***

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

$6.52\ (\%)\ (6.52442...\ (\%))$ ***(A1)*****(ft)*****(G2)***

Note: Follow through from part (d). Do not accept a negative answer.

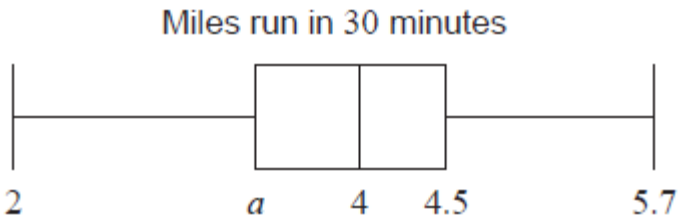
[2 marks]

10. [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)***

eg $Q_3 - Q_1$, $Q_3 - 1.1$, $4.5 - a = 1.1$

$a = 3.4$ ***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

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

[1 mark]

The variance of the distances run by the athletes is $\frac{16}{9}$ 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)***

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

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

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

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

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

Note: If no working shown, award ***M1A1M0A0*** for the value $\frac{4}{3}$.
If working shown, and candidate's final answer is $\frac{4}{3}$, award ***M1A1M0A0***.

METHOD 2 (variance first)

valid approach to convert variance ***(M1)***

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

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

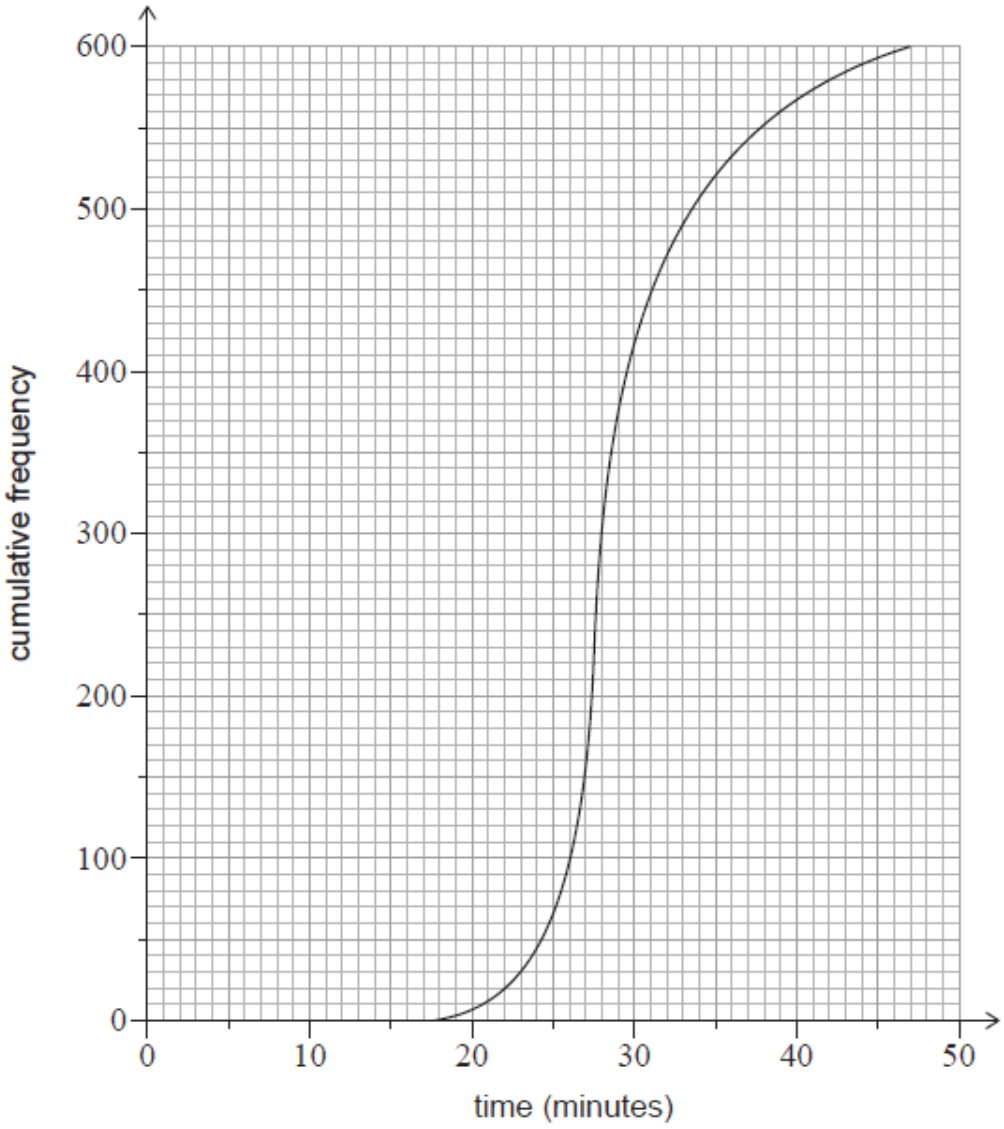
valid approach ***(M1)***

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

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

[4 marks]

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



There were 400 athletes who took between 22 and m minutes to complete the $5\ \text{km}$ 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,
$\mathrm{P}(22 < t < 27) = \frac{150-20}{600}$, $\frac{13}{60}$

evidence of conditional probability or reduced sample space ***(M1)***

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

correct working ***(A1)***

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

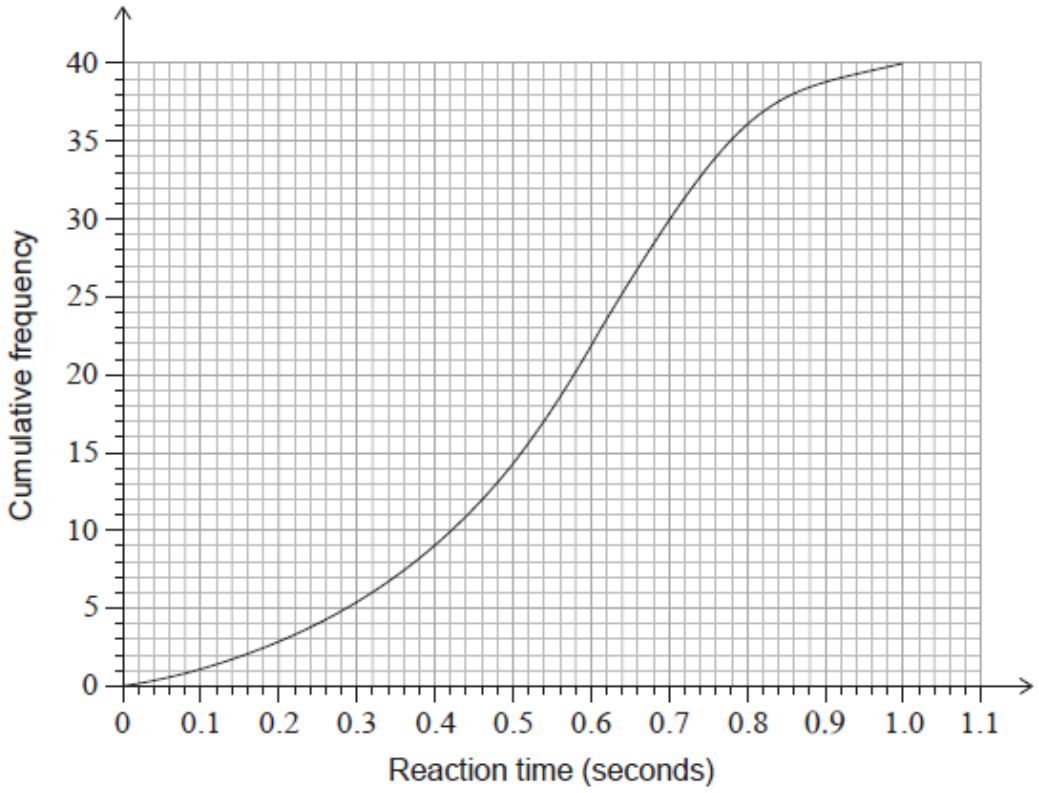
$\frac{130}{400}$ $\left(\frac{13}{40} = \frac{78000}{240000} = \frac{390}{1200} = 0.325\right)$ ***A1 N5***

Note: If no other working is shown, award ***A0A0M1A0A0*** for answer of $\frac{150}{400}$. Award ***N0*** for answer of $\frac{3}{8}$ with no other working shown.

[5 marks]

11. [Maximum mark: 17] 22M.2.SL.TZ2.1

Mackenzie conducted an experiment on the reaction times of teenagers. The results of the experiment are displayed in the following cumulative frequency graph.



Use the graph to estimate the

(a.i) median reaction time. [1]

Markscheme

$0.58\,(\text{s})$ ***A1***

[1 mark]

(a.ii) interquartile range of the reaction times. [3]

Markscheme

$0.7 - 0.42$ ***(A1)(M1)***

Note: Award ***A1*** for correct quartiles seen, ***M1*** for subtraction of their quartiles.

$0.28\,(\text{s})$ ***A1***

[3 marks]

(b) Find the estimated number of teenagers who have a reaction time greater than 0.4 seconds. [2]

Markscheme

9 (people have reaction time ≤ 0.4) ***(A1)***

31 (people have reaction time > 0.4) ***A1***

[2 marks]

(c) Determine the 90th percentile of the reaction times from the cumulative frequency graph. [2]

Markscheme

$(90\% \times 40 =)\ 36$ **OR** 4 ***(A1)***

$0.8\ s$ ***A1***

[2 marks]

Mackenzie created the cumulative frequency graph using the following grouped frequency table.

Reaction time, t (s)	Frequency
$0 < t \leq 0.2$	3
$0.2 < t \leq 0.4$	a
$0.4 < t \leq 0.6$	13
$0.6 < t \leq 0.8$	14
$0.8 < t \leq 1.0$	b

(d.i) Write down the value of a. [1]

Markscheme

$(a =)\ 6$ ***A1***

[1 mark]

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

Markscheme

$(b =)\ 4$ ***A1***

[1 mark]

(e) Write down the modal class from the table. [1]

Markscheme

$0.6 < t \leq 0.8$ **A1**

[1 mark]

(f) Use your graphic display calculator to find an estimate of the mean reaction time. [2]

Markscheme

$0.55\ \text{s}$ **A2**

[2 marks]

Upon completion of the experiment, Mackenzie realized that some values were grouped incorrectly in the frequency table. Some reaction times recorded in the interval $0 < t \leq 0.2$ should have been recorded in the interval $0.2 < t \leq 0.4$.

(g) Suggest how, if at all, the estimated mean and estimated median reaction times will change if the errors are corrected. Justify your response. [4]

Markscheme

the mean will increase **A1**

because the incorrect reaction times are moving from a lower interval to a higher interval which will increase the numerator of the mean calculation **R1**

the median will stay the same ***A1***

because the median or middle of the data is greater than both intervals being changed ***R1***

Note: Do not award ***A1R0***.

[4 marks]