

PROBABILITY AND STATISTICS -1

9709

(March, June and November series 2020 – 2023 With marking scheme)

Representation of Data

EXERCISE -1

MANJULA BALAJI

1) SP-2020_9709_5Q1

The following back-to-back stem-and-leaf diagram shows the annual salaries of a group of 39 females and 39 males.

Females			Males	
(4)	5 2 0 0	20	3	(1)
(9)	9 8 8 7 6 4 0 0 0	21	0 0 7	(3)
(8)	8 7 5 3 3 1 0 0	22	0 0 4 5 6 6	(6)
(6)	6 4 2 1 0 0	23	0 0 2 3 3 5 6 7 7	(9)
(6)	7 5 4 0 0 0	24	0 1 1 2 5 5 6 8 8 9	(10)
(4)	9 5 0 0	25	3 4 5 7 7 8 9	(7)
(2)	5 0	26	0 4 6	(3)

Key: 2 | 20 | 3 means \$20 200 for females and \$20 300 for males.

- (a) Find the median and the quartiles of the females' salaries. [2]

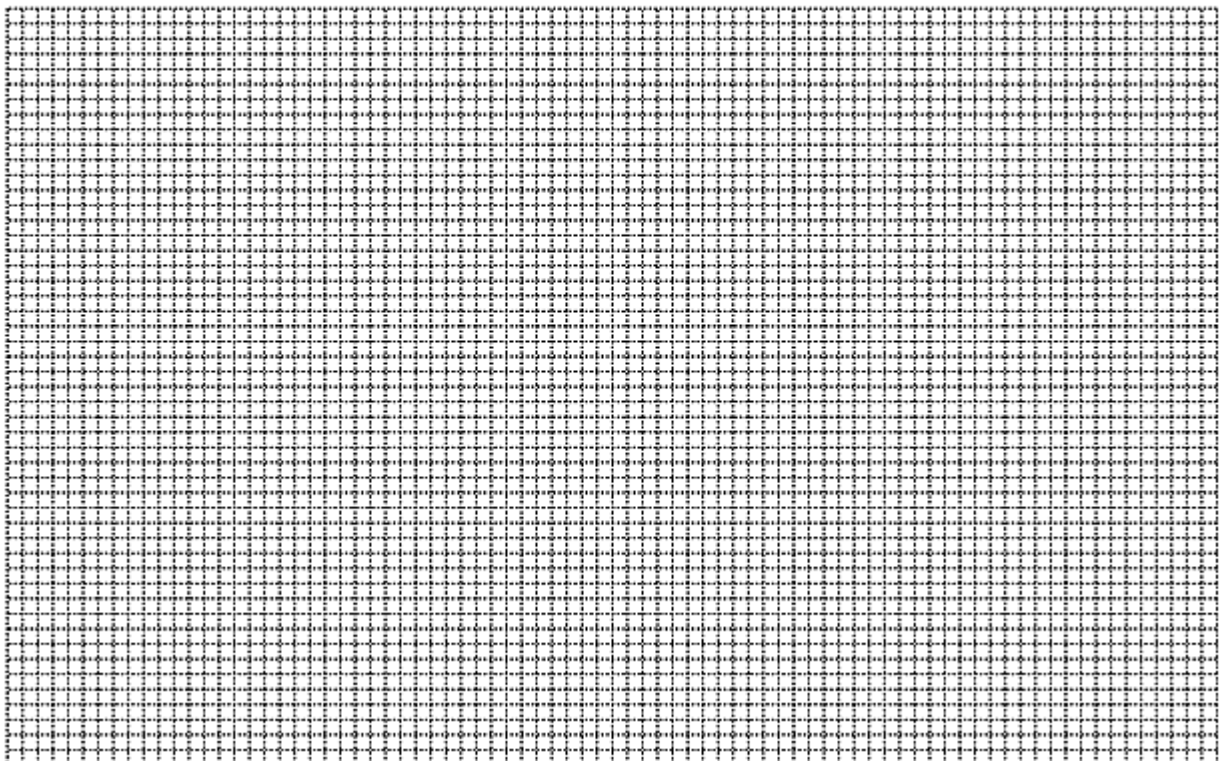
.....

.....

.....

You are given that the median salary of the males is \$24 000, the lower quartile is \$22 600 and the upper quartile is \$25 300.

- (b) Draw a pair of box-and-whisker plots in a single diagram on the grid below to represent the data. [3]



2) SP-2020_9709_5 Q2

A summary of the speeds, x kilometres per hour, of 22 cars passing a certain point gave the following information:

$$\Sigma(x - 50) = 81.4 \quad \text{and} \quad \Sigma(x - 50)^2 = 671.0.$$

Find the variance of the speeds and hence find the value of Σx^2 . [4]

3) MARCH 2020_9709_52 Q7

Helen measures the lengths of 150 fish of a certain species in a large pond. These lengths, correct to the nearest centimetre, are summarised in the following table.

Length (cm)	0 – 9	10 – 14	15 – 19	20 – 30
Frequency	15	48	66	21

(a) Draw a cumulative frequency graph to illustrate the data. [4]

(b) 40% of these fish have a length of d cm or more. Use your graph to estimate the value of d . [2]

The mean length of these 150 fish is 15.295 cm.

(c) Calculate an estimate for the variance of the lengths of the fish. [3]

4) MARCH 2021_9709_52 Q5

A driver records the distance travelled in each of 150 journeys. These distances, correct to the nearest km, are summarised in the following table.

Distance (km)	0 – 4	5 – 10	11 – 20	21 – 30	31 – 40	41 – 60
Frequency	12	16	32	66	20	4

(a) Draw a cumulative frequency graph to illustrate the data. [4]

(b) For 30% of these journeys the distance travelled is d km or more.

Use your graph to estimate the value of d . [2]

(c) Calculate an estimate of the mean distance travelled for the 150 journeys. [3]

5) MARCH 2022_9709_52 Q3

At a summer camp an arithmetic test is taken by 250 children. The times taken, to the nearest minute, to complete the test were recorded. The results are summarised in the table.

Time taken, in minutes	1 – 30	31 – 45	46 – 65	66 – 75	76 – 100
Frequency	21	30	68	86	45

(a) Draw a histogram to represent this information. [4]

(b) State which class interval contains the median. [1]

(c) Given that an estimate of the mean time is 61.05 minutes, state what feature of the distribution accounts for the median and the mean being different. [1]

6) MARCH 2023_9709_52 Q1

Each year the total number of hours, x , of sunshine in Kintoo is recorded during the month of June. The results for the last 60 years are summarised in the table.

x	$30 \leq x < 60$	$60 \leq x < 90$	$90 \leq x < 110$	$110 \leq x < 140$	$140 \leq x < 180$	$180 \leq x \leq 240$
Number of years	4	8	14	25	7	2

- (a) Draw a cumulative frequency graph to illustrate the data. [3]
- (b) Use your graph to estimate the 70th percentile of the data. [2]
- (c) Calculate an estimate for the mean number of hours of sunshine in Kintoo during June over the last 60 years. [3]

7) JUNE 2020_9709_51 Q7

The numbers of chocolate bars sold per day in a cinema over a period of 100 days are summarised in the following table.

Number of chocolate bars sold	1 – 10	11 – 15	16 – 30	31 – 50	51 – 60
Number of days	18	24	30	20	8

- (a) Draw a histogram to represent this information. [5]
- (b) What is the greatest possible value of the interquartile range for the data? [2]
- (c) Calculate estimates of the mean and standard deviation of the number of chocolate bars sold. [4]

8) JUNE 2020_9709_52 Q1

For n values of the variable x , it is given that

$$\Sigma(x - 50) = 144 \quad \text{and} \quad \Sigma x = 944.$$

Find the value of n .

[3]

9) JUNE 2020_9709_52 Q3

Two machines, A and B , produce metal rods of a certain type. The lengths, in metres, of 19 rods produced by machine A and 19 rods produced by machine B are shown in the following back-to-back stem-and-leaf diagram.

A		B
	21	1 2 4
7 6 3 0	22	2 4 5 5 6
8 7 4 3 1 1	23	0 2 6 8 9 9
5 5 5 3 2	24	3 3 4 6
4 3 1 0	25	6

Key: 7 | 22 | 4 means 0.227 m for machine A and 0.224 m for machine B .

- (a) Find the median and the interquartile range for machine A .

[3]

It is given that for machine *B* the median is 0.232 m, the lower quartile is 0.224 m and the upper quartile is 0.243 m.

- (b) Draw box-and-whisker plots for *A* and *B*. [3]
- (c) Hence make two comparisons between the lengths of the rods produced by machine *A* and those produced by machine *B*. [2]

10) JUNE 2020_9709_53 Q6

The annual salaries, in thousands of dollars, for 11 employees at each of two companies *A* and *B* are shown below.

Company <i>A</i>	30	32	35	41	41	42	47	49	52	53	64
Company <i>B</i>	26	47	30	52	41	38	35	42	49	31	42

- (a) Represent the data by drawing a back-to-back stem-and-leaf diagram with company *A* on the left-hand side of the diagram. [4]
- (b) Find the median and the interquartile range of the salaries of the employees in company *A*. [3]
A new employee joins company *B*. The mean salary of the 12 employees is now \$38 500.
- (c) Find the salary of the new employee. [3]

11) JUNE 2021_9709_51 Q5

The times taken by 200 players to solve a computer puzzle are summarised in the following table.

Time (<i>t</i> seconds)	$0 \leq t < 10$	$10 \leq t < 20$	$20 \leq t < 40$	$40 \leq t < 60$	$60 \leq t < 100$
Number of players	16	54	78	32	20

- (a) Draw a histogram to represent this information. [4]
- (b) Calculate an estimate of the mean time taken by these 200 players. [2]
- (c) Find the greatest possible value of the interquartile range of these times. [2]

12) JUNE 2021_9709_52 Q7

The heights, in cm, of the 11 basketball players in each of two clubs, the Amazons and the Giants, are shown below.

Amazons	205	198	181	182	190	215	201	178	202	196	184
Giants	175	182	184	187	189	192	193	195	195	195	204

- (a) State an advantage of using a stem-and-leaf diagram compared to a box-and-whisker plot to illustrate this information. [1]
- (b) Represent the data by drawing a back-to-back stem-and-leaf diagram with Amazons on the left-hand side of the diagram. [4]

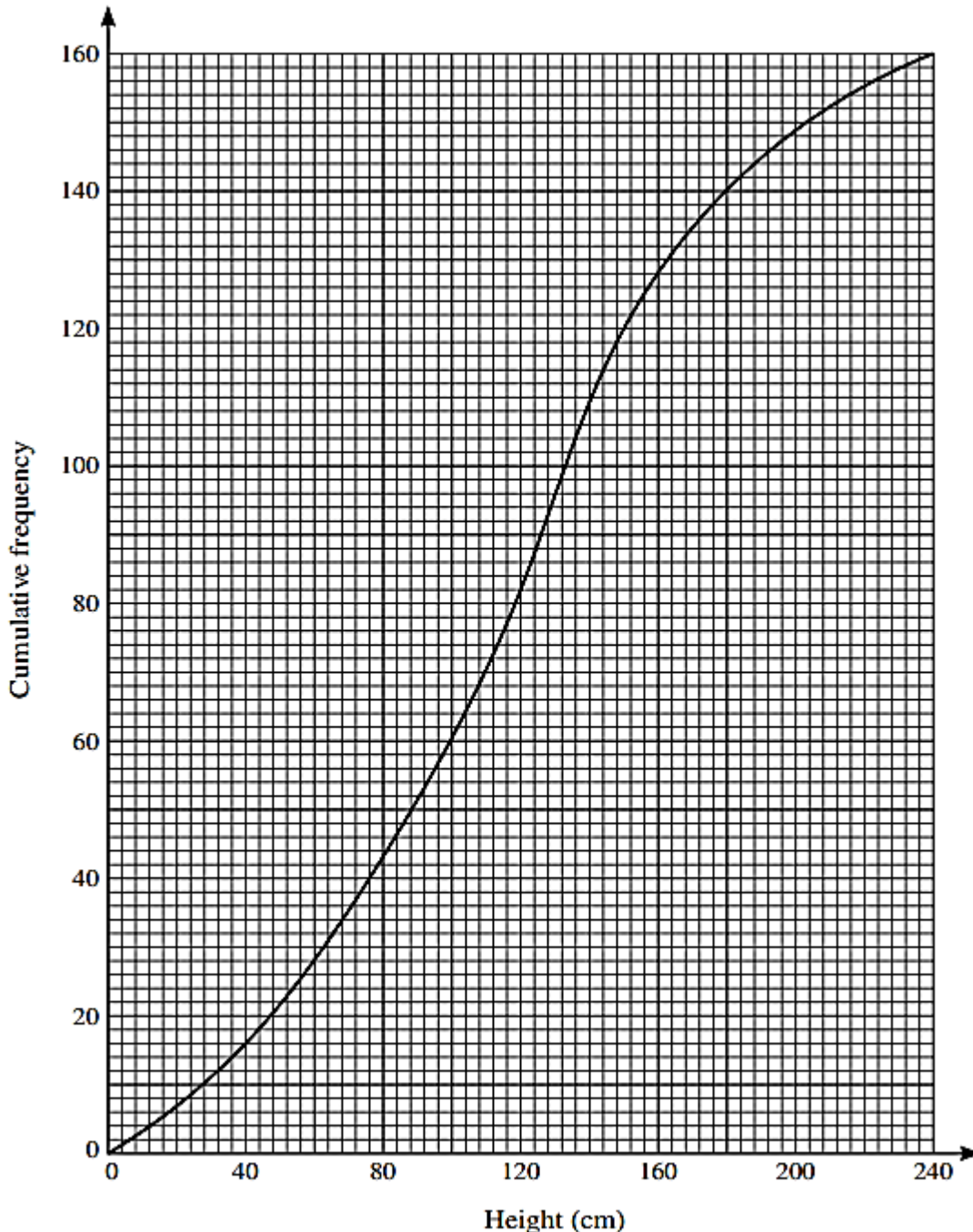
(c) Find the interquartile range of the heights of the players in the Amazons. [2]

Four new players join the Amazons. The mean height of the 15 players in the Amazons is now 191.2 cm. The heights of three of the new players are 180 cm, 185 cm and 190 cm.

(d) Find the height of the fourth new player. [3]

13) JUNE 2021_9709_53 Q1

The heights in cm of 160 sunflower plants were measured. The results are summarised on the following cumulative frequency curve.



(a) Use the graph to estimate the number of plants with heights less than 100 cm. [1]

(b) Use the graph to estimate the 65th percentile of the distribution. [2]

(c) Use the graph to estimate the interquartile range of the heights of these plants. [2]

14) JUNE 2021_9709_53 Q3

A sports club has a volleyball team and a hockey team. The heights of the 6 members of the volleyball team are summarised by $\Sigma x = 1050$ and $\Sigma x^2 = 193\,700$, where x is the height of a member in cm. The heights of the 11 members of the hockey team are summarised by $\Sigma y = 1991$ and $\Sigma y^2 = 366\,400$, where y is the height of a member in cm.

- (a) Find the mean height of all 17 members of the club. [2]
- (b) Find the standard deviation of the heights of all 17 members of the club. [3]

15) JUNE 2022_9709_51 Q3

The times taken to travel to college by 2500 students are summarised in the table.

Time taken (t minutes)	$0 \leq t < 20$	$20 \leq t < 30$	$30 \leq t < 40$	$40 \leq t < 60$	$60 \leq t < 90$
Frequency	440	720	920	300	120

- (a) Draw a histogram to represent this information. [4]
- From the data, the estimate of the mean value of t is 31.44.
- (b) Calculate an estimate of the standard deviation of the times taken to travel to college. [3]
- (c) In which class interval does the upper quartile lie? [1]

It was later discovered that the times taken to travel to college by two students were incorrectly recorded. One student's time was recorded as 15 instead of 5 and the other's time was recorded as 65 instead of 75.

- (d) Without doing any further calculations, state with a reason whether the estimate of the standard deviation in part (b) would be increased, decreased or stay the same. [1]

16) JUNE 2022_9709_52 Q1

For n values of the variable x , it is given that

$$\Sigma(x - 200) = 446 \quad \text{and} \quad \Sigma x = 6846.$$

- Find the value of n . [3]

17) JUNE 2022_9709_52 Q3

The back-to-back stem-and-leaf diagram shows the diameters, in cm, of 19 cylindrical pipes produced by each of two companies, *A* and *B*.

Company <i>A</i>						Company <i>B</i>				
				4	33	1	2	8		
9	8	3	2	0	34	1	6	8	9	9
8	7	5	4	1	35	1	2	2	3	
		9	6	5	36	5	6			
			4	3	37	0	3	4		
					38	2	8			

Key: 1 | 35 | 3 means the pipe diameter from company *A* is 0.351 cm and from company *B* is 0.353 cm.

(a) Find the median and interquartile range of the pipes produced by company *A*. [3]

It is given that for the pipes produced by company *B* the lower quartile, median and upper quartile are 0.346 cm, 0.352 cm and 0.370 cm respectively.

(b) Draw box-and-whisker plots for companies *A* and *B* on the grid below. [3]

(c) Make one comparison between the diameters of the pipes produced by companies *A* and *B*. [1]

18) JUNE 2022_9709_53 Q1

The time taken, t minutes, to complete a puzzle was recorded for each of 150 students. These times are summarised in the table.

Time taken (t minutes)	$t \leq 25$	$t \leq 50$	$t \leq 75$	$t \leq 100$	$t \leq 150$	$t \leq 200$
Cumulative frequency	16	44	86	104	132	150

(a) Draw a cumulative frequency graph to illustrate the data. [2]

(b) Use your graph to estimate the 20th percentile of the data. [1]

19) JUNE 2022_9709_53 Q2

Twenty children were asked to estimate the height of a particular tree. Their estimates, in metres, were as follows.

4.1	4.2	4.4	4.5	4.6	4.8	5.0	5.2	5.3	5.4
5.5	5.8	6.0	6.2	6.3	6.4	6.6	6.8	6.9	19.4

(a) Find the mean of the estimated heights. [1]

(b) Find the median of the estimated heights. [1]

(c) Give a reason why the median is likely to be more suitable than the mean as a measure of the central tendency for this information. [1]

20) JUNE 2023_9709_51 Q1

A summary of 50 values of x gives

$$\Sigma(x - q) = 700, \quad \Sigma(x - q)^2 = 14\,235,$$

where q is a constant.

(a) Find the standard deviation of these values of x . [2]

(b) Given that $\Sigma x = 2865$, find the value of q . [2]

21) JUNE 2023_9709_51 Q5

The populations of 150 villages in the UK, to the nearest hundred, are summarised in the table.

Population	100 – 800	900 – 1200	1300 – 2000	2100 – 3200	3300 – 4800
Number of villages	8	12	50	48	32

(a) Draw a histogram to represent this information. [4]

(b) Write down the class interval which contains the median for this information. [1]

(c) Find the greatest possible value of the interquartile range for the populations of the 150 villages. [2]

22) JUNE 2023_9709_52 Q3

The following back-to-back stem-and-leaf diagram represents the monthly salaries, in dollars, of 27 employees at each of two companies, A and B .

Company A							Company B							
		5	4	1	1	25	4	4	5	6	6	7		
9	9	8	7	2	1	26	0	1	3	5	5	7	9	9
	8	6	4	2	1	27	1	3	4	6	6	8	8	
		6	5	4	2	28	0	1	2	2	2			
			9	8	5	29								
				1		30	9							

Key: 1 | 27 | 6 means \$2710 for company A and \$2760 for company B

(a) Find the median and the interquartile range of the monthly salaries of employees in company A . [3]

The lower quartile, median and upper quartile for company B are \$2600, \$2690 and \$2780 respectively.

(b) Draw two box-and-whisker plots in a single diagram to represent the information for the salaries of employees at companies A and B . [3]

(c) Comment on whether the mean would be a more appropriate measure than the median for comparing the given information for the two companies. [1]

23) JUNE 2023_9709_53 Q4

The times taken, in minutes, to complete a cycle race by 19 cyclists from each of two clubs, the Cheetahs and the Panthers, are represented in the following back-to-back stem-and-leaf diagram.

Cheetahs			Panthers
9 8	7		4
8 7 3 2 0	8		6 8
9 8 7	9		1 7 8 9 9
6 5 3 3 1	10		2 3 4 4 5 6
9 8 2	11		1 2 8
4	12		0 6

Key: 7 | 9 | 1 means 97 minutes for Cheetahs and 91 minutes for Panthers

- (a) Find the median and the interquartile range of the times of the Cheetahs. [3]

Another cyclist, Kenny, from the Cheetahs also took part in the race. The mean time taken by the 20 cyclists from the Cheetahs was 99 minutes.

- (c) Find the time taken by Kenny to complete the race. [3]

24) OCT 2020_9709_51 Q6

The times, t minutes, taken by 150 students to complete a particular challenge are summarised in the following cumulative frequency table.

Time taken (t minutes)	$t \leq 20$	$t \leq 30$	$t \leq 40$	$t \leq 60$	$t \leq 100$
Cumulative frequency	12	48	106	134	150

- (a) Draw a cumulative frequency graph to illustrate the data. [2]
- (b) 24% of the students take k minutes or longer to complete the challenge. Use your graph to estimate the value of k . [2]
- (c) Calculate estimates of the mean and the standard deviation of the time taken to complete the challenge. [6]

25) OCT2020_9709_52 Q5

The following table gives the weekly snowfall, in centimetres, for 11 weeks in 2018 at two ski resorts, Dados and Linva.

Dados	6	8	12	15	10	36	42	28	10	22	16
Linva	2	11	15	16	0	32	36	40	10	12	9

- (a) Represent the information in a back-to-back stem-and-leaf diagram. [4]
- (b) Find the median and the interquartile range for the weekly snowfall in Dados. [3]
- (c) The median, lower quartile and upper quartile of the weekly snowfall for Linva are 12, 9 and 32 cm respectively. Use this information and your answers to part (b) to compare the central tendency and the spread of the weekly snowfall in Dados and Linva. [2]

26) OCT 2020_9709_53 Q7

A particular piece of music was played by 91 pianists and for each pianist, the number of incorrect notes was recorded. The results are summarised in the table.

Number of incorrect notes	1 – 5	6 – 10	11 – 20	21 – 40	41 – 70
Frequency	10	5	26	32	18

- (a) Draw a histogram to represent this information. [5]
- (b) State which class interval contains the lower quartile and which class interval contains the upper quartile.
Hence find the greatest possible value of the interquartile range. [2]
- (c) Calculate an estimate for the mean number of incorrect notes. [3]

27) OCT 2021_9709_51 Q2

A summary of 40 values of x gives the following information:

$$\Sigma(x - k) = 520, \quad \Sigma(x - k)^2 = 9640,$$

where k is a constant.

- (a) Given that the mean of these 40 values of x is 34, find the value of k . [2]
- (b) Find the variance of these 40 values of x . [2]

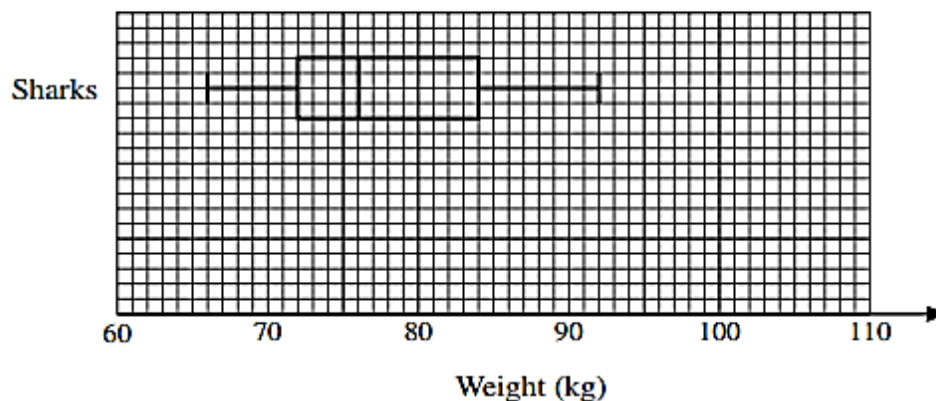
28) OCT 2021_9709_51 Q6

The weights, in kg, of 15 rugby players in the Rebels club and 15 soccer players in the Sharks club are shown below.

Rebels	75	78	79	80	82	82	83	84	85	86	89	93	95	99	102
Sharks	66	68	71	72	74	75	75	76	78	83	83	84	85	86	92

- (a) Represent the data by drawing a back-to-back stem-and-leaf diagram with Rebels on the left-hand side of the diagram. [4]
- (b) Find the median and the interquartile range for the Rebels. [3]

A box-and-whisker plot for the Sharks is shown below.



- (c) On the same diagram, draw a box-and-whisker plot for the Rebels. [2]
- (d) Make one comparison between the weights of the players in the Rebels club and the weights of the players in the Sharks club. [1]

29) OCT 2021_9709_52 Q7

The distances, x m, travelled to school by 140 children were recorded. The results are summarised in the table below.

Distance, x m	$x \leq 200$	$x \leq 300$	$x \leq 500$	$x \leq 900$	$x \leq 1200$	$x \leq 1600$
Cumulative frequency	16	46	88	122	134	140

- (a) On the grid, draw a cumulative frequency graph to represent these results. [2]
- (b) Use your graph to estimate the interquartile range of the distances. [2]
- (c) Calculate estimates of the mean and standard deviation of the distances. [6]

30) OCT 2021_9709_53 Q2

Lakeview and Riverside are two schools. The pupils at both schools took part in a competition to see how far they could throw a ball. The distances thrown, to the nearest metre, by 11 pupils from each school are shown in the following table.

Lakeview	10	14	19	22	26	27	28	30	32	33	41
Riverside	23	36	21	18	37	25	18	20	24	30	25

(a) Draw a back-to-back stem-and-leaf diagram to represent this information, with Lakeview on the left-hand side. [4]

(b) Find the interquartile range of the distances thrown by the 11 pupils at Lakeview school. [2]

31) OCT 2021_9709_53 Q3

The times taken, in minutes, by 360 employees at a large company to travel from home to work are summarised in the following table.

Time, t minutes	$0 \leq t < 5$	$5 \leq t < 10$	$10 \leq t < 20$	$20 \leq t < 30$	$30 \leq t < 50$
Frequency	23	102	135	76	24

(a) Draw a histogram to represent this information. [4]

(b) Calculate an estimate of the mean time taken by an employee to travel to work. [2]

32) OCT 2022_9709_51 Q3

The Lions and the Tigers are two basketball clubs. The heights, in cm, of the 11 players in each of their first team squads are given in the table.

Lions	178	186	181	187	179	190	189	190	180	169	196
Tigers	194	179	187	190	183	201	184	180	195	191	197

(a) Draw a back-to-back stem-and-leaf diagram to represent this information, with the Lions on the left. [4]

(b) Find the median and the interquartile range of the heights of the Lions first team squad. [3]

It is given that for the Tigers, the lower quartile is 183 cm, the median is 190 cm and the upper quartile is 195 cm.

(c) Make two comparisons between the heights of the players in the Lions first team squad and the heights of the players in the Tigers first team squad. [2]

33) OCT 2022_9709_52 Q4

The times taken, in minutes, to complete a word processing task by 250 employees at a particular company are summarised in the table.

Time taken (t minutes)	$0 \leq t < 20$	$20 \leq t < 40$	$40 \leq t < 50$	$50 \leq t < 60$	$60 \leq t < 100$
Frequency	32	46	96	52	24

(a) Draw a histogram to represent this information. [4]

From the data, the estimate of the mean time taken by these 250 employees is 43.2 minutes.

(b) Calculate an estimate for the standard deviation of these times. [3]

34) OCT 2022_9709_53 Q1

50 values of the variable x are summarised by

$$\Sigma(x - 20) = 35 \quad \text{and} \quad \Sigma x^2 = 25\,036.$$

Find the variance of these 50 values. [3]

35) OCT 2022_9709_53 Q3

The times, t minutes, taken to complete a walking challenge by 250 members of a club are summarised in the table.

Time taken (t minutes)	$t \leq 20$	$t \leq 30$	$t \leq 35$	$t \leq 40$	$t \leq 50$	$t \leq 60$
Cumulative frequency	32	66	112	178	228	250

(a) Draw a cumulative frequency graph to illustrate the data. [2]

(b) Use your graph to estimate the 60th percentile of the data. [1]

It is given that an estimate for the mean time taken to complete the challenge by these 250 members is 34.4 minutes.

(c) Calculate an estimate for the standard deviation of the times taken to complete the challenge by these 250 members. [4]

MARKING SCHEME

1) SP-2020_9709_5 Q1

1(a)	females: median \$22 700	1	B1	
	Lower quartile \$21 700 Upper quartile \$24 000	1	B1	Both correct
		2		
1(b)	Uniform scale and labels	1	B1	Must see Salary, \$000
		2	B1B1FT	B1 for correct graph for males CAO B1 for correct graph for females FT their quartiles. Horizontal line not through box
		3		

2) SP-2020_9709_5 Q2

	Coded mean = $\frac{81.4}{22} = 3.7$	1	M1	Attempt to find variance using coding in both, correct use of formula
	Var = $\frac{671}{22} - 3.7^2$			
	Var = 16.81	1	A1	Accept 16.8
	$16.81 = \frac{\sum x^2}{22} - 53.7^2$	1	M1	using their variance and their mean with uncoded formula for both
	$\sum x^2 = 63 811$	1	A1	Accept 63 800
	Available marks	4		

3) MARCH 2020_9709_52 Q7

(a)	15, 63, 129, 150	B1	Correct cumulative frequencies seen (may be on graph)
		B1	$0 \leq$ Horizontal axis ≤ 30 , $0 \leq$ vertical axis ≤ 150 Labels correct: length cm, cf
		M1	At least 3 points plotted at upper end points (e.g. allow 9, 9.5, 10) with a linear horizontal scale.
		A1	Linear vertical scale, all points at correct upper end points (9.5 etc.), curve drawn accurately, joined to (0,0) (condone (-0.5, 0))
		4	
(b)	60% of 150 = 90	M1	90 seen or implied by use on graph
	Approx. 16.5 [cm]	A1FT	FT <i>their</i> increasing cumulative frequency graph, Use of graph must be seen. If no clear evidence of use of graph SCB1FT correct value from <i>their</i> graph
		2	

4) MARCH 2021_9709_52 Q5

(a)	<table border="1"> <tr> <td>Distance</td> <td>0-4</td> <td>5-10</td> <td>11-20</td> <td>21-30</td> <td>31-40</td> <td>41-60</td> </tr> <tr> <td>Upper boundary</td> <td>4.5</td> <td>10.5</td> <td>20.5</td> <td>30.5</td> <td>40.5</td> <td>60.5</td> </tr> <tr> <td>Cumulative frequency</td> <td>12</td> <td>28</td> <td>60</td> <td>126</td> <td>146</td> <td>150</td> </tr> </table>	Distance	0-4	5-10	11-20	21-30	31-40	41-60	Upper boundary	4.5	10.5	20.5	30.5	40.5	60.5	Cumulative frequency	12	28	60	126	146	150	B1	Correct cumulative frequencies seen (may be by table or plotted accurately on graph), condone 12 not stated.
	Distance	0-4	5-10	11-20	21-30	31-40	41-60																	
Upper boundary	4.5	10.5	20.5	30.5	40.5	60.5																		
Cumulative frequency	12	28	60	126	146	150																		
	B1	Axes labelled 'distance (or d) [in] km' from 0 to 60 and 'cumulative frequency' (or cf) from 0 to 150.																						
	M1	At least 5 points plotted at upper end points for d (allow upper boundary ± 0.5) with a linear scale for distance, condone 0 – 4 interval inaccurate, no scale break on axis. Not bar graph/histogram unless clear indication of upper end point only of each bar.																						
	A1	All plotted correctly at correct upper end points (4.5 etc.) with both scales linear ($0 \leq d \leq 60$, $0 \leq cf \leq 150$), curve drawn accurately joined to (0,0), cf line > 150 , no daylight if > 150 .																						
		4																						
(b)	70% of 150 = 105	M1	105 seen or implied by indication on grid.																					
	Approx. 27	A1 FT	Strict FT <i>their</i> increasing cumulative frequency graph, use of graph must be seen. If no clear evidence of use of graph: SC B1 FT correct value from <i>their</i> increasing cumulative frequency graph.																					
		2																						
(c)	Midpoints: 2.25, 7.5, 15.5, 25.5, 35.5, 50.5	B1	At least 5 correct midpoints seen.																					
	$\text{Mean} = \frac{2.25 \times 12 + 7.5 \times 16 + 15.5 \times 32 + 25.5 \times 66 + 35.5 \times 20 + 50.5 \times 4}{150}$ $= \frac{27 + 120 + 496 + 1683 + 710 + 202}{150}$	M1	Using 6 midpoint attempts (e.g. 2.25 ± 0.5), condone one error not omission, multiplied by frequency, accept unevaluated, denominator either correct or <i>their</i> Σ frequencies.																					
	$\left[= \frac{3238}{150} \right] = 21.6, 21 \frac{44}{75}$	A1	Evaluated, WWW, accept $21.5[866\dots]$.																					
		3																						

5) MARCH 2022_9709_52 Q3

(a)	<table border="1"> <tr> <td>Class Width</td> <td>30</td> <td>15</td> <td>20</td> <td>10</td> <td>25</td> </tr> <tr> <td>Frequency Density</td> <td>0.7</td> <td>2</td> <td>3.4</td> <td>8.6</td> <td>1.8</td> </tr> </table>	Class Width	30	15	20	10	25	Frequency Density	0.7	2	3.4	8.6	1.8	M1	At least 4 frequency densities calculated
	Class Width	30	15	20	10	25									
Frequency Density	0.7	2	3.4	8.6	1.8										
		A1	All heights correct on graph												
		B1	Bar ends at 0.5, 30.5, 45.5, 65.5, 75.5, 100.5 (at axis), 5 bars drawn, condone 0 in first bar $0.5 \leq \text{time axis} \leq 100.5$, linear scale with at least 3 values indicated.												
		B1	Axes labelled: Frequency density (fd), time (t) and mins (or appropriate title). Linear fd scale, with at least 3 values indicated $0 \leq \text{fd axis} \leq 8.6$												
		4													
(b)	66 – 75	B1	Condone 65.5 – 75.5												
		1													
(c)	Distribution is not symmetrical	B1	Or skewed, ignore nature of skew												
		1													

6) MARCH 2023_9709_52 Q1

(a)	<table border="1"> <tr> <td>Upper value</td> <td>60</td> <td>90</td> <td>110</td> <td>140</td> <td>180</td> <td>240</td> </tr> <tr> <td>cf</td> <td>4</td> <td>12</td> <td>26</td> <td>51</td> <td>58</td> <td>60</td> </tr> </table>	Upper value	60	90	110	140	180	240	cf	4	12	26	51	58	60	B1	All cumulative frequencies stated. May be under data table, condone omission of 4. May be read accurately from graph, must include 4.
	Upper value	60	90	110	140	180	240										
cf	4	12	26	51	58	60											
		M1	At least 5 points plotted at class upper end points, daylight rule tolerance. Linear cf scale $0 \leq cf \leq 60$, linear time scale $30 \leq \text{time} \leq 240$ with at least 3 values identified on each axis.														
		A1	All points plotted correctly. Curve drawn (within tolerance), no ruled segments, and joined to (30, 0). Axes labelled 'cumulative frequency' and 'hours [of sunshine]' (OE including appropriate title).														
		3															
(b)	$[60 \times 0.7 =] 42$	M1	42 may be implied by clear use on graph.														
	126	A1 FT	Must be clear evidence on graph of use of 42, e.g. an appropriate mark on either axis, appropriate mark on curve. FT from increasing cf graph only read at 42 only.														
(c)	Midpoints: 45, 75, 100, 125, 160, 210	2															
	$[\text{Mean} =] \frac{4 \times 45 + 8 \times 75 + 14 \times 100 + 25 \times 125 + 7 \times 160 + 2 \times 210}{60}$ $[= \frac{6845}{60}]$	M1	Correct mean formula using their 6 midpoints (must be within class, not upper bound, lower bound), condone 1 data error If correct midpoints seen accept $\frac{180 + 600 + 1400 + 3125 + 1120 + 420}{60}$														
	$= 114, 114 \frac{1}{12}$	A1	Accept 114.1, 114.08[3...] If A1 not awarded, SC B1 for $114, 114 \frac{1}{12}, 114.1$ or 114.08[3...].														
		3															

7) JUNE 2020_9709_51 Q7

a)	Class widths: 10, 5, 15, 20, 10	M1
	Frequency density = frequency/their class width: 1.8, 4.8, 2, 1, 0.8	M1
	All heights correct on diagram (using a linear scale)	A1
	Correct bar ends	B1
	Bar ends: 10.5, 15.5, 30.5, 50.5, 60.5	B1
		5
b)	11 – 15 and 31 – 50	B1
	Greatest IQR = 50 – 11 = 39	B1
		2
c)	$\text{Mean} = \frac{18 \times 5.5 + 24 \times 13 + 30 \times 23 + 20 \times 40.5 + 8 \times 55.5}{100} = \frac{2355}{100} = 23.6$	B1
	$\text{Var} = \frac{18 \times 5.5^2 + 24 \times 13^2 + 30 \times 23^2 + 20 \times 40.5^2 + 8 \times 55.5^2}{100} - \text{mean}^2$	M1
	$\frac{77917.5}{100} - \text{mean}^2 = 224.57$	A1
	Standard deviation = 15.0 (FT their variance)	A1 FT
	4	

8) JUNE 2020_9709_52 Q1

$\Sigma x - 50n = 144$	B1
$50n + 144 = 944$	M1
$n = 16$	A1
	3

9) JUNE 2020_9709_52 Q3

a)	Median = 0.238	B1
	UQ = 0.245, LQ = 0.231, So IQR = 0.245 - 0.231	M1
	0.014	A1
		3

b)	<table border="1"> <thead> <tr> <th></th> <th></th> <th>LQ</th> <th>M</th> <th>UQ</th> <th></th> </tr> </thead> <tbody> <tr> <td>A</td> <td>0.220</td> <td>0.231 FT</td> <td>0.238 FT</td> <td>0.245 FT</td> <td>0.254</td> </tr> <tr> <td>B</td> <td>0.211</td> <td>0.224</td> <td>0.232</td> <td>0.243</td> <td>0.256</td> </tr> </tbody> </table>			LQ	M	UQ		A	0.220	0.231 FT	0.238 FT	0.245 FT	0.254	B	0.211	0.224	0.232	0.243	0.256	
		LQ	M	UQ																
A	0.220	0.231 FT	0.238 FT	0.245 FT	0.254															
B	0.211	0.224	0.232	0.243	0.256															
	Medians and quartiles correctly plotted for <i>A</i> or <i>B</i>	B1																		
	End points correct for <i>A</i> or <i>B</i>	B1																		
	Completely correct, including scale	B1																		
		3																		

c)	Lengths of rods produced by machine <i>A</i> are longer. (B1 for comparison of central tendency)	B1
	Lengths of rods produced by machine <i>A</i> are less spread out (B1 for comparison of spread)	B1
		2

10) JUNE 2020_9709_53 Q6

(a)	<table border="1"> <thead> <tr> <th>A</th> <th></th> <th>B</th> </tr> </thead> <tbody> <tr> <td></td> <td>2</td> <td>6</td> </tr> <tr> <td>5 2 0</td> <td>3</td> <td>0 1 5 8</td> </tr> <tr> <td>9 7 2 1 1</td> <td>4</td> <td>1 2 2 7 9</td> </tr> <tr> <td>3 2</td> <td>5</td> <td>2</td> </tr> <tr> <td>4</td> <td>6</td> <td></td> </tr> </tbody> </table> <p>KEY 1 4 2 means \$41 000 for A and \$42 000 for B</p>	A		B		2	6	5 2 0	3	0 1 5 8	9 7 2 1 1	4	1 2 2 7 9	3 2	5	2	4	6		
A		B																		
	2	6																		
5 2 0	3	0 1 5 8																		
9 7 2 1 1	4	1 2 2 7 9																		
3 2	5	2																		
4	6																			
	Correct stem	B1																		
	Correct A on LHS	B1																		
	Correct B on same diagram	B1																		
	Correct key for <i>their</i> diagram, both companies identified and correct units	B1																		
		4																		
(b)	Median = [\$]42 000	B1																		
	LQ = [\$]35 000 UQ = [\$]52 000	B1																		
	IQR = [\$]17 000 (FT if $49000 \leq UQ \leq 53000 - 32000 \leq LQ \leq 41000$)	B1 FT																		
		3																		
(c)	Sum of given 11 numbers is 433 000	M1																		
	Sum of 12 numbers, including new = $38\,500 \times 12 = 462\,000$	M1																		
	Difference = new salary = [\$]29 000	A1																		
		3																		

11) JUNE 2021_9709_51 Q5

(a)	<table border="1"> <tbody> <tr> <td>Class width</td> <td>10</td> <td>10</td> <td>20</td> <td>20</td> <td>40</td> </tr> <tr> <td>Frequency Density</td> <td>1.6</td> <td>5.4</td> <td>3.9</td> <td>1.6</td> <td>0.5</td> </tr> </tbody> </table> 	Class width	10	10	20	20	40	Frequency Density	1.6	5.4	3.9	1.6	0.5	<p>M1 At least 4 frequency densities calculated, accept unsimplified. May be read from graph using <i>their</i> scale, 3SF or correct</p> <p>A1 All heights correct on graph</p> <p>B1 Bar ends at 0, 10, 20 ..., etc. with a horizontal linear scale with at least 3 values indicated, $0 \leq \text{horizontal axis} \leq 100$</p> <p>B1 Axes labelled: Frequency density (<i>fd</i>), time (<i>t</i>) and seconds. Linear vertical scale, with at least 3 values indicated $0 \leq \text{vertical axis} \leq 5.4$</p>
Class width	10	10	20	20	40									
Frequency Density	1.6	5.4	3.9	1.6	0.5									
		4												

b)	$\text{Mean} = \left[\frac{16 \times 5 + 54 \times 15 + 78 \times 30 + 32 \times 50 + 20 \times 80}{200} \right]$ $= \frac{80 + 810 + 2340 + 1600 + 1600}{200}$	MI	Uses at least 4 midpoint attempts (e.g. 5 ± 0.5). Accept unsimplified expression, denominator either correct or <i>their</i> Σ frequencies
	$\left[\frac{6430}{200} = \right] 32 \frac{3}{20} \text{ or } 32.15$	A1	Accept 32.2
		2	
c)	A value in correct UQ (40–60) – a value in correct LQ (10–20)	MI	
	Greatest possible value is $60 - 10 = 50$	A1	Condone 49.9
		2	

12) JUNE 2021_9709_52 Q7

a)	Includes all data	B1	Reference to <i>either</i> including all/raw data or further statistical processes are possible that cannot be found using data from box-and-whisker, eg frequency, mean, mode or standard deviation not only median, IQR, range or spread which can be found from both.																								
		1																									
b)	<table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th colspan="2">Amazons</th> <th colspan="2">Giants</th> </tr> </thead> <tbody> <tr> <td></td> <td>8</td> <td>17</td> <td>5</td> </tr> <tr> <td>4</td> <td>2 1</td> <td>18</td> <td>2 4 7 9</td> </tr> <tr> <td>8</td> <td>6 0</td> <td>19</td> <td>2 3 5 5 5</td> </tr> <tr> <td>5</td> <td>2 1</td> <td>20</td> <td>4</td> </tr> <tr> <td></td> <td>5</td> <td>21</td> <td></td> </tr> </tbody> </table> <p>Key: 1 18 2 means 181 cm for Amazons and 182 cm for Giants</p>	Amazons		Giants			8	17	5	4	2 1	18	2 4 7 9	8	6 0	19	2 3 5 5 5	5	2 1	20	4		5	21		B1	Correct stem can be upside down, ignore extra values
Amazons		Giants																									
	8	17	5																								
4	2 1	18	2 4 7 9																								
8	6 0	19	2 3 5 5 5																								
5	2 1	20	4																								
	5	21																									
		B1	Correct Amazons labelled on left, leaves in order from right to left and lined up vertically (less than halfway to next column), no commas or other punctuation.																								
		B1	Correct Giants labelled on same diagram, leaves in order and lined up vertically (less than halfway to next column), no commas or other punctuation.																								
		B1	Correct single key for their diagram, need both teams identified and 'cm' stated at least once here or in leaf headings or title. SC for if 2 separate diagrams drawn, award SCB1 if both keys meet these criteria (Max B1, B0, B0, B1)																								
		4																									
c)	[UQ = 202 (cm), LQ = 182 (cm)] [IQR =] $202 - 182 = 20$ (cm)	M1	$201 \leq \text{UQ} \leq 205 - 181 \leq \text{LQ} \leq 184$																								
		A1	WWW																								
		2																									
d)	$[\Sigma_{11} = 2132$ $\Sigma_{15} = 191.2 \times 15 = 2868]$	B1	Both Σ_{11} and Σ_{15} found. Accept unevaluated.																								
	<i>their</i> $2868 = \text{their } 2132 + (180 + 185 + 190) + h$	M1	Forming an equation for the height using <i>their</i> Σ_{11} and Σ_{15} .																								
	181 (cm)	A1																									
Alternative method for Question 7(d)																											
	$[\Sigma_{15} = 191.2 \times 15 = 2868$ $\Sigma_{15} = 2687 + h]$	B1	Σ_{15} found using the mean and raw data methods. Accept unevaluated.																								
	<i>their</i> $2868 = \text{their } 2687 + h$	M1	Forming an equation for the height using <i>their</i> Σ_{15} expressions.																								
	181 (cm)	A1																									
Alternative method for Question 7(d)																											
	$[\Sigma_{15} = 2687 + h$ $\frac{\Sigma_{15}}{15} = 191.2]$	B1	Σ_{15} found using raw data method and statement on calculating new mean. Accept unevaluated.																								
	<i>their</i> $\frac{2687 + h}{15} = 191.2$	M1	Forming an equation for the height using <i>their</i> Σ_{15} expressions																								
	181 (cm)	A1																									
		3	N.B. All methods can be presented as a logical numerical argument which can be condoned if clear.																								

13) JUNE 2021_9709_53 Q1

a)	60	B1	Accept 60 or 61. No decimals
		1	
b)	65% of 160 = 104	M1	0.65 × 160 (=104) seen unsimplified or implied by use on graph
	136 (cm)	A1	Use of graph must be seen. SCB1 correct value (136 only) if neither 104 nor use of graph are evident
		2	
c)	UQ: 150 LQ: 76 IQR = 150 - 76 = 74 [cm]	M1	UQ - LQ ; 148 ≤ UQ ≤ 152; 74 ≤ LQ ≤ 78.
		A1	Must be from 150 - 76
		2	

14) JUNE 2021_9709_53 Q3

a)	Mean height = $\frac{\Sigma x + \Sigma y}{6+11} = \frac{1050+1991}{6+11} = \frac{3041}{17}$	M1	Use of appropriate formula with values substituted, accept unsimplified.
	178.9	A1	Allow 178.88, $178\frac{15}{17}$, 179
		2	
b)	$\frac{\Sigma x^2 + \Sigma y^2}{6+11} = \frac{193700+366400}{6+11}$	M1	Use of appropriate formula with values substituted, accept unsimplified.
	$Sd^2 = \frac{560100}{17} - \text{their } 178.88^2 [= 948.289]$	M1	Appropriate variance formula using <i>their</i> mean ² , accept unsimplified expression.
	Standard deviation = 30.8	A1	Accept 30.7
		3	

15) JUNE 2022_9709_51 Q3

a)	<table border="1"> <tr> <td>Class width</td> <td>20</td> <td>10</td> <td>10</td> <td>20</td> <td>30</td> </tr> <tr> <td>Frequency density</td> <td>22</td> <td>72</td> <td>92</td> <td>15</td> <td>4</td> </tr> </table>	Class width	20	10	10	20	30	Frequency density	22	72	92	15	4	M1	At least 4 frequency densities calculated (Frequency ÷ class width, e.g. $\frac{440}{20}$ (condone $\frac{440}{19.5}$, $\frac{440}{20.5}$) Accept unsimplified, may be read from graph using <i>their</i> scale
Class width	20	10	10	20	30										
Frequency density	22	72	92	15	4										
		A1	All heights correct on graph NOT FT												
		B1	Bar ends at {0,} 20, 30, 40, 60, 90 at axis with a horizontal linear scale with at least 3 values indicated. 0 ≤ horizontal scale ≤ 90												
		B1	Axes labelled frequency density (fd), time (t) and minutes (mins) or in a title. Linear vertical scale, with at least 3 values indicated 0 ≤ vertical axes ≤ 92 (condone 90 used).												
		4													

(b)	<table border="1"> <tr> <td>Midpoints</td> <td>10</td> <td>25</td> <td>35</td> <td>50</td> <td>75</td> </tr> </table>	Midpoints	10	25	35	50	75	B1	At least 4 correct midpoints seen
Midpoints	10	25	35	50	75				
	<p>[Mean = 31.44 given]</p> <p>[Variance = $\frac{440 \times 10^2 + 720 \times 25^2 + 920 \times 35^2 + 300 \times 50^2 + 120 \times 75^2}{2500} - 31.44^2$]</p> <p>= $\frac{44000 + 450000 + 1127000 + 750000 + 675000}{2500} - 31.44^2$</p> <p>[= $\frac{3046000}{2500} - 31.44^2 = 229.9264$]</p> <p>Or</p> <p>Variance =</p> <p>$\frac{440(10 - 31.44)^2 + 720(25 - 31.44)^2 + 920(35 - 31.44)^2 + 300(50 - 31.44)^2 + 120(75 - 31.44)^2}{2500}$</p> <p>= $\frac{202256 + 29860 + 11659 + 103342 + 227697}{2500} = \frac{574814}{2500} = 229.9264$</p>	M1	Correct formula for variance or standard deviation ($- \text{mean}^2$ included with <i>their</i> midpoints (not upper bound, lower bound, class width, frequency density, frequency or cumulative frequency) and <i>their</i> $\sum f$ if calculated. Condone 1 data error.						
	Standard deviation = 15.2	A1	WWW, allow 15.16[3...]						
		3							
(c)	30–40	B1							
		1							
(d)	Stays the same, data still in same intervals	B1	Frequencies unchanged						
		1							

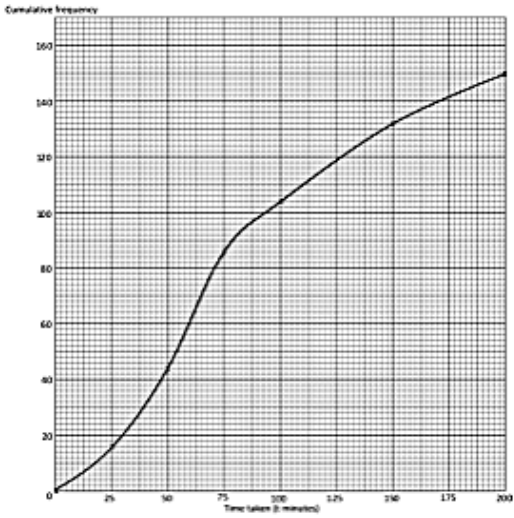
16) JUNE 2022_9709_52 Q1

$\Sigma x - \Sigma 200 = \Sigma(x - 200)$	B1	Forming a correct 3-term (linear) equation from Σx , $\Sigma 200$ and $\Sigma(x - 200)$. Accept $6846 - 200n = 446$ OE. Condone 1 sign error.
$\Sigma 200 = 200n$	B1	SOI
$[200n = 6846 - 446 = 6400] n = 32$	B1	WWW
	3	

17) JUNE 2022_9709_52 Q3

(a)	Median = 0.355	B1	Identified condone Q2.
	[IQR =] 0.366 – 0.348	M1	$0.365 \leq \text{UQ} \leq 0.369 - 0.343 \leq \text{LQ} \leq 0.349$. Subtraction may be implied by answer.
	0.018	A1	If 0/3 scored SC B1 for figs Median = 355 IQR = 18.
		3	
(b)	<p>Box-and-whisker plot on provided grid</p>	B1	All 5 key values for <i>B</i> plotted accurately in standard format using <i>their</i> scale. Labelled <i>B</i> . Check accuracy in the middle of vertical line.
		B1 FT	All 5 key values for <i>A</i> , FT from part 3(a), plotted in standard format accurately using <i>their</i> scale. Labelled <i>A</i> . Check accuracy in the middle of vertical line.
		B1	Whiskers not through box for both, not drawn at corners of boxes, single linear scale with at least 3 values stated, covering at least 0.34 to 0.38 and labelled diameter (<i>d</i> etc) and cm. Accept as a title.
		3	If both plots attempted and plot(s) not labelled, SC B1 for at least 1 fully correct set of values plotted.
(c)	A comparison in context	B1	Single comment comparing spread or central tendency in context. Must reference either diameter or pipes. Not a simple numerical comparison of statistical values such as median, range, IQR or min/max.
		1	

18) JUNE 2022_9709_53 Q1

(a)	Cumulative frequency (cf) graph	M1	At least 3 points plotted accurately at class upper end points (25,16) (50,44) (75,86) (100,104) (150, 132) (200, 150). Linear cf scale $0 \leq cf \leq 150$ and linear time scale $0 \leq \text{time}(\text{mins}) \leq 200$ with at least 3 values identified on each axis.
		A1	All points plotted correctly, curve drawn (within tolerance) and joined to (0,0). Axes labelled cumulative frequency (cf), time (t) and minutes (min), or a suitable title.
		2	
(b)	Line from cumulative frequency = 30 to meet graph at t is between 37.5 and 42	B1 FT	Not from wrong working. Must be an increasing cumulative frequency graph.
		1	

19) JUNE 2022_9709_53 Q2

(a)	$\left[\frac{123.4}{20} = \right] 6.17$	B1	Accept 6 m 17 cm, $\frac{1234}{200}$.
		1	
(b)	$\frac{10\text{th} + 11\text{th}}{2} = \frac{5.4 + 5.5}{2} = 5.45$ (m)	B1	Accept 5 m 45 cm.
		1	
(c)	The mean is unduly influenced by an extreme value, 19.4.	B1	Comment must be within context.
		1	

20) JUNE 2023_9709_51 Q1

(a)	$\text{Var} = \left[\frac{\Sigma(x-q)^2}{50} - \left(\frac{\Sigma(x-q)}{50} \right)^2 \right] = \left[\frac{14235}{50} - \left(\frac{700}{50} \right)^2 \right]$ $[= 284.7 - 196 = 88.7]$	M1	$\frac{14235}{a} - \left(\frac{700}{a} \right)^2$; where $a = 49, 50, 51$.
[sd = $\sqrt{88.7}$ =] 9.42		A1	9.4180677 rounded to at least 3SF.
		2	
(b)	$\Sigma x - 50q = 700$ $[2865 - 50q = 700]$	M1	Forming equation with Σx , $50q$ and 700.
$q = 43.3, 43 \frac{3}{10}$		A1	If M0 scored, SC B1 for 43.3 WWW.
		2	

21) JUNE 2023_9709_51 Q5

(a)	<table border="1"> <tr> <td>cw</td> <td>800</td> <td>400</td> <td>800</td> <td>1200</td> <td>1600</td> </tr> <tr> <td>fd</td> <td>0.01</td> <td>0.03</td> <td>0.0625</td> <td>0.04</td> <td>0.02</td> </tr> </table>	cw	800	400	800	1200	1600	fd	0.01	0.03	0.0625	0.04	0.02	<p>M1 At least 4 frequency densities calculated (F/cw, e.g. $\frac{8}{800}$ (condone $\frac{8}{n}$, $799 \leq n \leq 801$)) Accept unsimplified, may be read from graph using <i>their</i> scale.</p>
	cw	800	400	800	1200	1600								
	fd	0.01	0.03	0.0625	0.04	0.02								
		<p>A1 All heights correct on graph.</p>												
	<p>B1 Bar ends at 50, 850, 1250, 2050, 3250, 4850 read at the axis with a horizontal linear scale with at least 3 values indicated. 50 \leq horizontal scale \leq 4850.</p>													
	<p>B1 Axes labelled frequency density (fd) and population (pop) OE, or in a title. Linear vertical scale, with at least 3 values indicated. Vertical axis must cover at least the range $0 \leq$ vertical axis \leq 0.0625. Axes may be reversed.</p>													
		4												
(b)	2100 – 3200	<p>B1 Accept 2050 – 3250 OE. Condone '4th interval'.</p>												
		1												
(c)	3249 – 1250	<p>M1 2050 \leq UQ \leq 3250 – 1250 \leq LQ \leq 2050.</p>												
	1999	<p>A1 Condone 3250 – 1250 = 2000.</p>												
		2												

22) JUNE 2023_9709_52 Q3

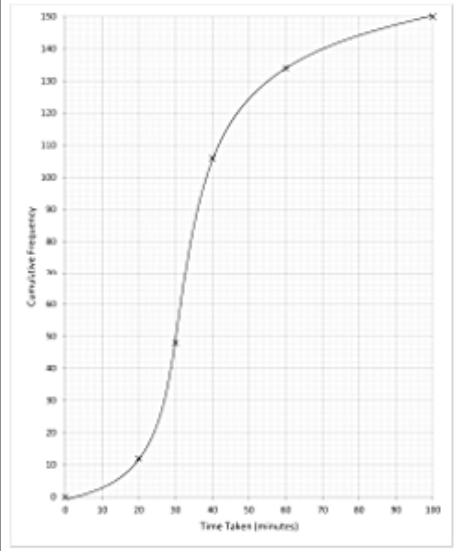
a)	Median = 2710	<p>B1 Must be identified, condone Q2. Ignore units throughout.</p>
	2840 – 2610	<p>M1 2820 \leq UQ \leq 2850 – 2600 \leq LQ \leq 2620.</p>
	230	<p>A1 www If M0 scored SC B1 for 230 www. If key ignored consistently: B0 Median = 271 SC M1 282 \leq UQ \leq 285 – 260 \leq LQ \leq 262 SC A1 23.</p>
		3
b)	Box-and-whisker plot on provided grid.	<p>B1 All 5 key values for <i>B</i> plotted accurately in standard format using a linear scale with 3 identified values. Labelled <i>B</i>. Scale at least 1 cm = \$100.</p>
	<p>B: 2540 2600 2690 2780 3090 A: 2500 2610 2710 2840 3010</p>	<p>B1FT All 5 key values for <i>A</i>, FT from (a), plotted accurately in standard format using a linear scale with 3 identified values. Labelled <i>A</i>. Scale at least 1cm = \$100</p>
		<p>B1 Whiskers not through box for both, not drawn at corners of boxes, single linear scale for the diagram and labelled 'salaries' (oe) and \$.</p>
		3

c)	Examples: Mean less appropriate than median because of extreme value for company B [at \$3090]. No, extreme value in company B. No, \$3090 is an anomaly.	B1	Must refer to company B, may be implied by appropriate use of \$3090. Must include an indication that the mean is not appropriate. No contradictory statements can be present, e.g. acceptable comment with 'but mean could be used for company A'. Condone reference to \$309.
		1	

23) JUNE 2023_9709_53 Q4

i(a)	Median = 99 [minutes]	B1	
	[IQR =] 106 – 83	M1	$105 \leq UQ \leq 112 - 82 \leq LQ \leq 87$.
	23 [minutes]	A1	www. If M0 scored SC B1 for 23 www.
		3	
i(b)	The times for the Cheetahs are faster than the times for the Panthers	B1	Correct statement comparing central tendency in context.
	The times for the Cheetahs are more spread than the times for the Panthers	B1	Correct statement comparing range/IQR in context.
		2	
i(c)	[Total time including Kenny = $99 \times 20 =$] 1980	B1	Accept unsimplified.
	[Kenny's time =] 1980 – 1862	M1	For <i>their</i> 1980 – <i>their</i> 1862.
	= 118 [minutes]	A1	Accept 1 hour 58 mins.
	Alternative Method for Question 4(c)		
	$\frac{1862 + \textit{their Kenny's time}}{20} = 99$	B1	$\frac{1862 + \textit{their Kenny's time}}{20} = 99$ seen.
	[Kenny's time = $99 \times 20 - 1862$]	M1	For <i>their</i> $99 \times 20 - \textit{their}$ 1862.
	= 118 [minutes]	A1	Accept 1 hour 58 mins.
	3		

24) OCT 2020_9709_51 Q6

a)		M1	At least 4 points plotted at upper end points, with both scales linear with at least 3 values indicated
	Correct cumulative frequency curve	A1	All plotted correctly with curve drawn joined to (0, 0), axes labelled cumulative frequency, time, minutes
		2	
b)	$150 \times 0.76 = 114$	M1	114 SOI, may be on graph
	$k = 45$ (mins)	A1 FT	Clear indication that <i>their</i> graph has been used, tolerance ± 1 mm
		2	

(c)	Frequencies: 12 36 58 28 16	B1	Correct frequencies seen
	Mean = $\frac{10 \times 12 + 25 \times 36 + 35 \times 58 + 50 \times 28 + 80 \times 16}{150}$	B1	At least 4 correct midpoints seen and used
	$\frac{120 + 900 + 2030 + 1400 + 1280}{150}$	M1	Correct formula with <i>their</i> midpoints (not upper boundary, lower boundary, class width or frequency density).
	38.2, $38\frac{1}{5}$	A1	
	Variance = $\frac{12 \times 10^2 + 36 \times 25^2 + 58 \times 35^2 + 28 \times 50^2 + 16 \times 80^2}{150} - \text{mean}^2$ = $\frac{1200 + 22500 + 71050 + 70000 + 102400}{150} - \text{mean}^2$	M1	Substitute <i>their</i> midpoints and frequencies (condone use of cumulative frequency) in correct variance formula, must have '- <i>their</i> mean ² '
	(Standard deviation = $\sqrt{321.76}$) = 17.9	A1	
		6	

25) OCT2020_9709_52 Q5

(a)	<table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th style="text-align: center;">Dados</th> <th style="text-align: center;"> </th> <th style="text-align: center;">Linva</th> </tr> </thead> <tbody> <tr> <td style="text-align: center;">8 6 0</td> <td style="text-align: center;"> </td> <td style="text-align: center;">0 2 9</td> </tr> <tr> <td style="text-align: center;">6 5 2 0 0</td> <td style="text-align: center;"> </td> <td style="text-align: center;">1 0 1 2 5 6</td> </tr> <tr> <td style="text-align: center;">8 2 2</td> <td style="text-align: center;"> </td> <td style="text-align: center;"></td> </tr> <tr> <td style="text-align: center;">6 3</td> <td style="text-align: center;"> </td> <td style="text-align: center;">2 6</td> </tr> <tr> <td style="text-align: center;">2 4</td> <td style="text-align: center;"> </td> <td style="text-align: center;">0</td> </tr> </tbody> </table> <p>KEY 6 3 2 means 36 cm (snow) in Dados and 32 cm (snow) in Linva</p>	Dados		Linva	8 6 0		0 2 9	6 5 2 0 0		1 0 1 2 5 6	8 2 2			6 3		2 6	2 4		0	B1	Correct stem can be upside down, ignore extra values
Dados		Linva																			
8 6 0		0 2 9																			
6 5 2 0 0		1 0 1 2 5 6																			
8 2 2																					
6 3		2 6																			
2 4		0																			
		B1	Correct Dados labelled, leaves in order and lined up vertically (less than midway to next column), no commas etc, no extra terms																		
		B1	Correct Linva on opposite side of stem labelled, leaves in order and lined up vertically (less than midway to next column), no commas etc, no extra terms																		
		B1	Correct single key for their diagram, need both resorts identified and 'cm' stated at least once here or in leaf headings or title. SC If 2 separate diagrams drawn, SCB1 if both keys meet these criteria B0B1B0SCB1 max.																		
		4																			
(b)	Median or Q2 = 15 (cm)	B1	Correct																		
	UQ or Q3 = 28 cm, LQ or Q1 = 10 cm IQR = 28 - 10	M1	$22 \leq \text{UQ} \leq 36 - 8 \leq \text{LQ} \leq 10$																		
	18 (cm)	A1	WWW																		
		3																			
(c)	On average the snowfall in Davos is higher	B1 FT	FT from <i>their</i> 5(b) values for Dados. Statement comparing central tendency in context																		
	The amount of snowfall in Linva varies more than in Davos	B1 FT	Statement comparing spread in context Note: simply stating and comparing the values is not sufficient.																		
		2																			

26) OCT 2020_9709_53 Q7

a)	Class widths: 5, 5, 10, 20, 30 Frequency density: 2, 1, 2.6, 1.6, 0.6	M1	At least 3 class widths correct and used in a calculation
		M1	At least 3 correct frequency densities unsimplified – FT <i>their</i> class widths
		A1	All correct heights on a histogram using a linear vertical scale from zero – no FT
		B1	Correct upper bar ends (5.5, 10.5, 20.5, 40.5, 70.5) and 4 correct lower bar ends of 5.5, 10.5, 20.5, 40.5. Condone 0 or 1.
		B1	Linear scales with at least 3 values indicated on each axis, vertical scale from 0, axes labelled 'fd' and 'no. of (incorrect) notes', or better.
		5	
b)	LQ: 11 – 20 UQ: 21 – 40	B1	Both UQ and LQ correct
	Greatest IQR = 40 – 11 = 29	B1 FT	Subtract lower end of <i>their</i> LQ interval from upper end of <i>their</i> UQ interval
		2	
c)	Midpoints: 3 8 15.5 30.5 55.5	M1	At least 4 midpoints correct and used
	$\text{Mean} = \frac{3 \times 10 + 8 \times 5 + 15.5 \times 26 + 30.5 \times 32 + 55.5 \times 18}{91}$ $= \frac{30 + 40 + 403 + 976 + 999}{91}$ $= \frac{2448}{91}$	M1	Correct formula with <i>their</i> midpoints (not upper boundary, lower boundary, class width, frequency density, frequency or cumulative frequency)
	26.9, $26\frac{82}{91}$	A1	Accept 26 or 27
		3	

27) OCT 2021_9709_51 Q2

a)	$\left[\frac{\sum x}{40} - k = \frac{\sum (x-k)}{40} \right]$ $\frac{40 \times 34}{40} - k = \frac{520}{40}$	M1	Forms an equation involving $\sum x$, $\sum(x-k)$ and k . Accept at a numeric stage with k .
	$k [= 34 - 13] = 21$	A1	Evaluated.
		2	
b)	$\text{Var} = \left[\frac{\sum (x-k)^2}{40} - \left(\frac{\sum (x-k)}{40} \right)^2 \right] = \frac{9640}{40} - \left(\frac{520}{40} \right)^2 = [241 - 13^2 =]$	M1	Values substituted into an appropriate variance formula, accept unsimplified.
	72	A1	
		2	

28) OCT 2021_9709_51 Q6

(a)	<table border="1" style="display: inline-table; border-collapse: collapse;"> <thead> <tr> <th style="text-align: center;">Rebels</th> <th style="text-align: center;">Sharks</th> </tr> </thead> <tbody> <tr> <td style="text-align: right;">9 8 5</td> <td style="text-align: left;">6 8</td> </tr> <tr> <td style="text-align: right;">2 2 0</td> <td style="text-align: left;">1 2 4 5 5 6 8</td> </tr> <tr> <td style="text-align: right;">9 5 3</td> <td style="text-align: left;">3 3 4 5 6</td> </tr> <tr> <td style="text-align: right;">2</td> <td style="text-align: left;">9 2</td> </tr> <tr> <td style="text-align: right;">2</td> <td style="text-align: left;">10</td> </tr> </tbody> </table>	Rebels	Sharks	9 8 5	6 8	2 2 0	1 2 4 5 5 6 8	9 5 3	3 3 4 5 6	2	9 2	2	10	B1	Correct stem, ignore extra values (not in reverse).
	Rebels	Sharks													
	9 8 5	6 8													
	2 2 0	1 2 4 5 5 6 8													
9 5 3	3 3 4 5 6														
2	9 2														
2	10														
Key: 8 7 2 means 78 kg for Rebels and 72 kg for Sharks	B1	Correct Rebels labelled on left, leaves in order from right to left and lined up vertically, no commas.													
	B1	Correct Sharks labelled on same diagram, leaves in order and lined up vertically, no commas.													
	B1	Correct key for their diagram, need both teams identified and 'kg' stated at least once here or in leaf headings or title. SC If 2 separate diagrams drawn, SC B1 if both keys meet these criteria.													
	4														
(b)	Median = 84 (kg)	B1													
	[UQ = 93, LQ = 80] 93 – 80	M1	$95 \leq UQ \leq 89 - 79 \leq LQ \leq 82$												
	[IQR =] 13 (kg)	A1	WWW												
		3													
(c)	Box and whisker with end points 75 and 102	B1	Whiskers drawn to correct end points not through box, not joining at top or bottom of box.												
	Median and quartiles plotted as found in (b)	B1 FT	Quartiles and median plotted as box graph.												
		2													
(d)	e.g. Average weight of Rebels is higher than average weight of Sharks	B1	Acceptable answers refer to: Range, skew, central tendency within context. E.g. range of Rebels is greater B0. Range of weights of the rebels is greater B1. Simple value comparison insufficient.												
		1													

29) OCT 2021_9709_52 Q7

(a)	Cumulative frequency graph drawn	B1	Axes labelled 'cumulative frequency' (or cf) from 0 to at least 140 and 'distance (or d) [in] m' from 0 to at least 1600, linear scales with at least 3 values stated.
		B1	All plotted correctly at correct upper end points (200 etc.) curve drawn accurately joined to (0, 0) (straight line segments B0) but no daylight above 140. Cf scale no less than 2 cm = 20 children .
		2	
(b)	[UQ at 75% of 140 = 105, LQ at 25% of 140 = 35] [IQR:] 700 – 260	M1	Accept $660 \leq UQ \leq 720 - 240 \leq LQ \leq 290$. If values are outside our range, FT providing scales linear and increasing cf drawn.
	440	A1	Accept correct evaluation of $660 \leq \text{their } UQ \leq 720 - 240 \leq \text{their } LQ \leq 290$ with clear indication that graph has been used for at least one of 105 or 35.
		2	

30) OCT 2021_9709_53 Q2

(a)	Lakeview	Riverside	B1 Correct stem, ignore extra values.	
	9 4 0 1	8 8		B1 Correct Lakeview labelled on left, leaves in order from right to left and lined up vertically, no commas.
	8 7 6 2 2	0 1 3 4 5 5		
3 2 0 3	0 6 7	B1 Correct Riverside labelled on same diagram, leaves in order and lined up vertically, no commas.		
	1 4		B1 Correct key for their diagram, need both teams identified and 'm' stated at least once here or in leaf headings or title. SC If 2 separate diagrams drawn: SC B1 if both keys meet these criteria.	
	Key: 6 2 3 means 26m for Lakeview and 23m for Riverside		4	
(b)	UQ = 32, LQ = 19		M1 $(30 \leq UQ \leq 33) - (14 \leq LQ \leq 22)$	
	IQR = 32 - 19 = 13		A1 WWW	
			2	

31) OCT 2021_9709_53 Q3

(a)	Cw: 5 5 10 10 20	M1 At least 4 frequency densities calculated (f/cw), accept unsimplified and class widths ± 1 of true values. May be implied by graph.
	Fd: 4.6 20.4 13.5 7.6 1.2	A1 All heights correct on graph NOT FT
		B1 Bar ends at 0, 5, 10, 20, 30, 50 clear intention not to draw at 4.5 or 5.5 etc. B1 Axes labelled: Frequency density (fd), time (t) and mins (or appropriate title). Linear scales between 0 and 20.4 or above on vertical axis, and 0 and 50 or above on the horizontal axis. (Axes may be reversed.)
		4
(b)	$\frac{2.5 \times 23 + 7.5 \times 102 + 15 \times 135 + 25 \times 76 + 40 \times 24}{360}$	M1 Uses at least 4 midpoint attempts (e.g. 2.5 ± 0.5) in correct formula, accept unsimplified expression, denominator either correct or <i>their</i> Σ frequencies.
	$\left[\frac{5707.5}{360} = \right] 15.9, 15\frac{41}{48}$	A1 Evaluated.

$\sum x - 50 \times 20 = 35; \sum x = 1035$ or $\bar{x} = \frac{35}{50} + 20 = \frac{1035}{50} [= 20.7]$	B1	Correct value for $\sum x$ or \bar{x} .
$\frac{25036}{50} - \left(\frac{\sum x}{50}\right)^2 = \frac{25036}{50} - \left(\frac{1035}{50}\right)^2$	M1	$\frac{25036}{50} - \left(\text{their} \left(\frac{\sum x}{50}\right)^2\right)$
72.23	A1	Exact answer only SC B1 for 72.23 with no substitution in formula.
	3	

35) OCT 2022_9709_53 Q3

(a)	<p>Cumulative frequency graph</p>	M1	At least 3 points plotted accurately at class upper end points: (20, 32), (30, 66), (35, 112), (40, 178), (50, 228), (60, 250). Linear cf scale $0 \leq cf \leq 250$ and linear time scale $0 \leq \text{time} \leq 60$ with at least 3 values identified on each.
		A1	All points plotted correct, curve drawn (within tolerance) and joined to (0,0). Axes labelled cumulative frequency (cf), time (t) and minutes (min or m) – or a suitable title. Axes can be the other way round.
		2	
(b)	Line drawn from 150 on cf axis to meet graph at about $t = 38$ minutes	B1 FT	Must be an increasing cf graph with correct upper bounds. Use of graph must be seen. Expect an answer in range $37 \leq t \leq 39$ for a correct graph
		1	