

Chapter 1-3:

Mr Faruk

Teacher of Mathematics
BSc/MSc/PGCE Mathematics

✉ cieigcsolutions@gmail.com



4. A researcher recorded the time, t minutes, spent using a mobile phone during a particular afternoon, for each child in a club.

The researcher coded the data using $v = \frac{t - 5}{10}$ and the results are summarised in the table below.

Coded Time (v)	Frequency (f)	Coded Time Midpoint (m)
$0 \leq v < 5$	20	2.5
$5 \leq v < 10$	24	a
$10 \leq v < 15$	16	12.5
$15 \leq v < 20$	14	17.5
$20 \leq v < 30$	6	b

(You may use $\sum fm = 825$ and $\sum fm^2 = 12\,012.5$)

- (a) Write down the value of a and the value of b . (1)
- (b) Calculate an estimate of the mean of v . (1)
- (c) Calculate an estimate of the standard deviation of v . (2)
- (d) Use linear interpolation to estimate the median of v . (2)
- (e) Hence describe the skewness of the distribution. Give a reason for your answer. (2)
- (f) Calculate estimates of the mean and the standard deviation of the time spent using a mobile phone during the afternoon by the children in this club. (4)

Organised by Mr Omar Faruk

4. A group of 100 adults recorded the amount of time, t minutes, they spent exercising each day. Their results are summarised in the table below.

Time (t minutes)	Frequency (f)	Time midpoint (x)
$0 \leq t < 15$	25	7.5
$15 \leq t < 30$	17	22.5
$30 \leq t < 60$	28	45
$60 \leq t < 120$	24	90
$120 \leq t \leq 240$	6	180

[You may use $\sum fx^2 = 455\,512.5$]

A histogram is drawn to represent these data.

The bar representing the time $0 \leq t < 15$ has width 0.5 cm and height 6 cm.

- (a) Calculate the width and height of the bar representing a time of $60 \leq t < 120$ (3)
- (b) Use linear interpolation to estimate the median time spent exercising by these adults each day. (2)
- (c) Find an estimate of the mean time spent exercising by these adults each day. (2)
- (d) Calculate an estimate for the standard deviation of these times. (2)
- (e) Describe, giving a reason, the skewness of these data. (1)

Further analysis of the above data revealed that 18 of the 25 adults in the $0 \leq t < 15$ group took no exercise each day.

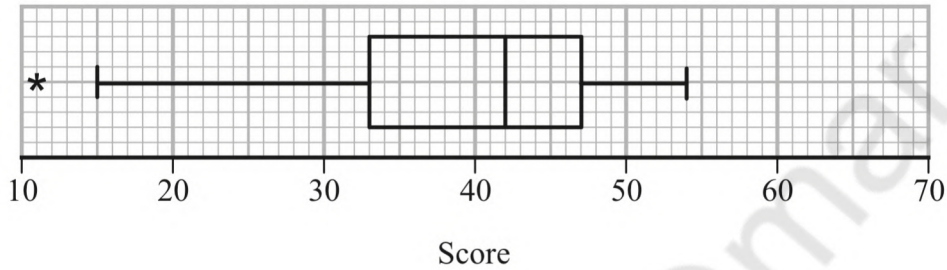
- (f) State, giving a reason, what effect, if any, this new information would have on your answers to
- (i) the estimate of the median in part (b),
- (ii) the estimate of the mean in part (c),
- (iii) the estimate of the standard deviation in part (d). (3)

Organised by Mr Omar Faruk

2. Chi wanted to summarise the scores of the 39 competitors in a village quiz. He started to produce the following stem and leaf diagram

Score	Key: 2 5 is a score of 25
1 1 5 8 9	
2 0 2 5 8 9	
3 3 5 5 7 8 9 ...	

He did not complete the stem and leaf diagram but instead produced the following box plot.



Chi defined an outlier as a value that is

$$\text{greater than } Q_3 + 1.5 \times (Q_3 - Q_1)$$

or

$$\text{less than } Q_1 - 1.5 \times (Q_3 - Q_1)$$

(a) Find

- (i) the interquartile range
- (ii) the range.

(2)

(b) Describe, giving a reason, the skewness of the distribution of scores.

(2)

Albert and Beth asked for their scores to be checked.

Albert's score was changed from 25 to 37

Beth's score was changed from 54 to 60

(c) On the grid on page 5, draw an updated box plot. Show clearly any calculations that you used.

(7)

Some of the competitors complained that the questions were biased towards the younger generation. The product moment correlation coefficient between the age of the competitors and their score in the quiz is -0.187

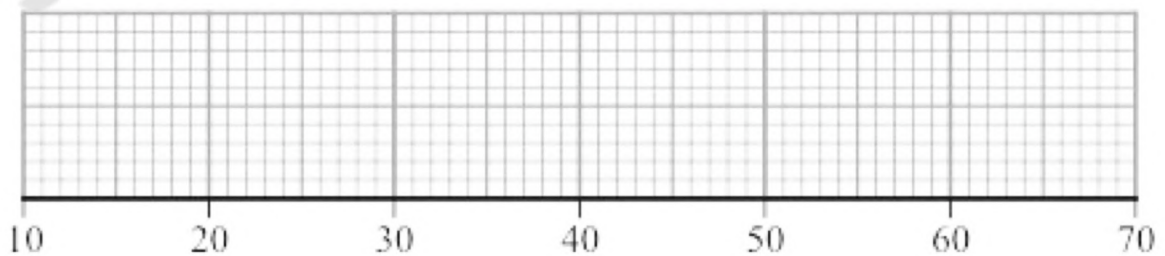
(d) State, giving a reason, whether or not the complaint is supported by this statistic.

(2)

Organised by Mr Omar Faruk

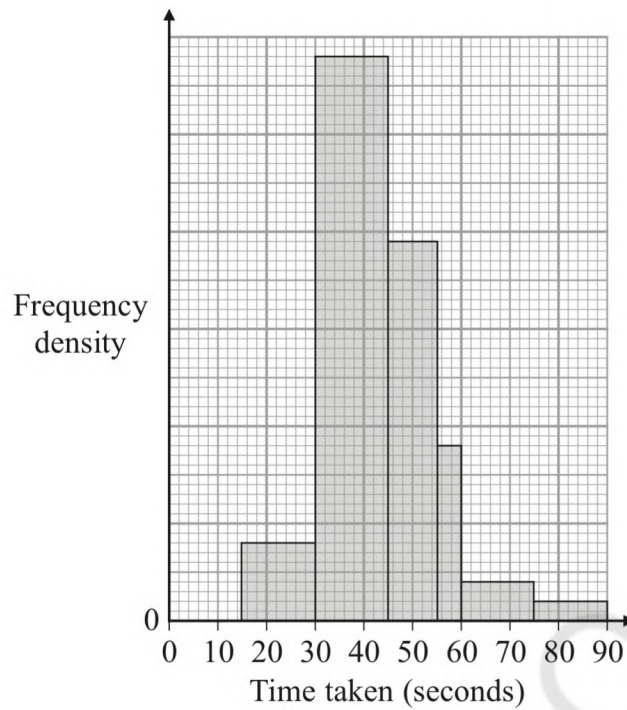
Blank grid for drawing a box plot.

Only use this grid if you need to redraw your box plot.



Score

2. The histogram shows the times taken, in seconds, by each of 260 people to complete a puzzle.



- (a) Use the histogram to complete the frequency table for the times taken to complete the puzzle.

Time taken (seconds)	15–30	30–45	45–55	55–60	60–75	75–90
Frequency (f)	20	145			10	5
Time midpoint (t seconds)	22.5	37.5	50	57.5	67.5	82.5

(3)

Given that $\sum ft = 11\,087.5$ and $\sum ft^2 = 505\,718.75$

- (b) find an estimate for
- (i) the mean time taken to complete the puzzle, (1)
 - (ii) the standard deviation of the times taken to complete the puzzle. (2)
- (c) Use linear interpolation to estimate the median time taken to complete the puzzle. (2)
- (d) Describe the skewness of these data. Give a reason for your answer. (1)

Three of the 260 people are chosen at random.

- (e) Estimate the probability that all 3 of their times are less than 36 seconds. (4)

Organised by Mr Omar Faruk

Organised by Mr Omar Faruk

4. A researcher is studying the birth weights of babies. A random sample of 98 babies was taken and their birth weights, w kg, are summarised in the table below.

Birth weight (w kg)	Frequency (f)	Birth weight midpoint (x)
$1.50 \leq w < 2.50$	16	2.00
$2.50 \leq w < 3.00$	24	2.75
$3.00 \leq w < 3.50$	32	3.25
$3.50 \leq w < 4.00$	14	3.75
$4.00 \leq w < 5.50$	12	4.75

(You may use $\sum fx = 311.5$ and $\sum fx^2 = 1051.125$)

A histogram is drawn to represent these data.

The bar representing the birth weight $1.50 \leq w < 2.50$ has a width of 1 cm and a height of 4 cm.

- (a) Calculate the width and height of the bar representing birth weight $3.50 \leq w < 4.00$ (3)

- (b) Use linear interpolation to estimate the lower quartile of the birth weights of the 98 babies. (2)

The researcher estimated the median to be 3.14 kg and the upper quartile to be 3.55 kg.

- (c) Use the median and quartiles to describe the skewness of these data. (2)

- (d) Find an estimate for (i) the mean birth weight
(ii) the standard deviation of the birth weights. (3)

- (e) Use the formula

$$\text{skewness} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

to estimate a value for the skewness of these data. Give your answer to 2 significant figures. (2)

The researcher read that birth weights should be approximately normally distributed and decides to split the class $3.00 \leq w < 3.50$

The frequency for $3.00 \leq w < 3.25$ is 9 and the frequency for $3.25 \leq w < 3.50$ is 23

- (f) (i) State, giving a reason, what the effect would be on the estimate of the median.
(ii) Without carrying out any further calculations state, giving a reason, what the effect of this change would be on the estimate of the mean. (2)

Organised by Mr Omar Faruk

Organised by Mr Omar Faruk

4. A group of students took some tests. A teacher is analysing the average mark for each student. Each student obtained a different average mark.

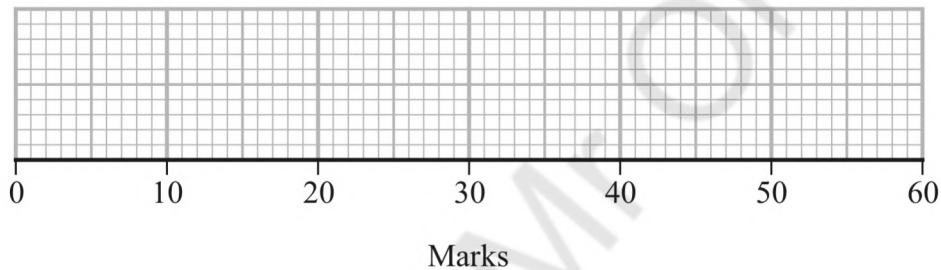
For these average marks, the lower quartile is 24, the median is 30 and the interquartile range (IQR) is 10

The three lowest average marks are 8, 10 and 15.5 and the three highest average marks are 45, 52.5 and 56

The teacher defines an outlier to be a value that is either

more than $1.5 \times \text{IQR}$ below the lower quartile or
more than $1.5 \times \text{IQR}$ above the upper quartile

- (a) Determine any outliers in these data. (4)
- (b) On the grid below draw a box plot for these data, indicating clearly any outliers. (3)

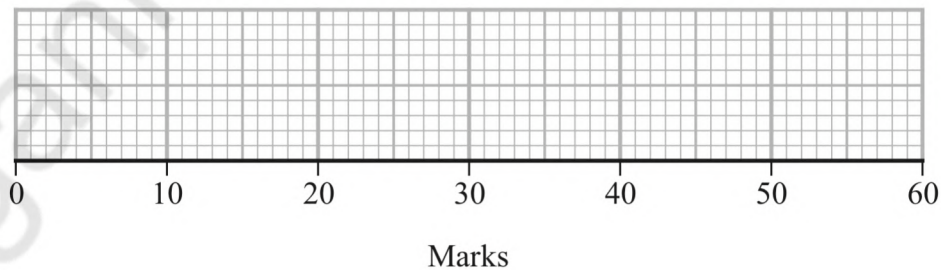


- (c) Use the quartiles to describe the skewness of these data.
Give a reason for your answer. (2)

Two more students also took the tests. Their average marks, which were both less than 45, are added to the data and the box plot redrawn.

The median and the upper quartile are the same but the lower quartile is now 26

- (d) Redraw the box plot on the grid below. (3)



- (e) Give ranges of values within which each of these students' average marks must lie. (2)

Organised by Mr Omar Faruk

2. The stem and leaf diagram below shows the ages (in years) of the residents in a care home.

Age		Key: 4 3 is an age of 43
4	3	(1)
5	4	(1)
6	2 3 5 6 8 8 8 9 9	(9)
7	1 1 3 4 4 6 6 6 8 8 9	(11)
8	0 0 2 7 8 8 9	(7)
9	3 7	(2)

- (a) Find the median age of the residents. (1)

- (b) Find the interquartile range (IQR) of the ages of the residents. (2)

An outlier is defined as a value that is either

more than $1.5 \times (\text{IQR})$ below the lower quartile or

more than $1.5 \times (\text{IQR})$ above the upper quartile.

- (c) Determine any outliers in these data. Show clearly any calculations that you use. (3)

- (d) On the grid on page 5, draw a box plot to summarise these data. (3)

Organised by Mr Omar Faruk

3. A random sample of 100 carrots is taken from a farm and their lengths, L cm, recorded. The data are summarised in the following table.

Length, L cm	Frequency, f	Class mid point, x cm
$5 \leq L < 8$	5	6.5
$8 \leq L < 10$	13	9
$10 \leq L < 12$	16	11
$12 \leq L < 15$	25	13.5
$15 \leq L < 20$	30	17.5
$20 \leq L < 28$	11	24

A histogram is drawn to represent these data.

The bar representing the class $5 \leq L < 8$ is 1.5 cm wide and 1 cm high.

- (a) Find the width and height of the bar representing the class $15 \leq L < 20$ (3)

- (b) Use linear interpolation to estimate the median length of these carrots. (2)

- (c) Estimate

- (i) the mean length of these carrots, (2)

- (ii) the standard deviation of the lengths of these carrots. (3)

A supermarket will only buy carrots with length between 9 cm and 22 cm.

- (d) Estimate the proportion of carrots from the farm that the supermarket will buy. (2)

Any carrots that the supermarket does not buy are sold as animal feed.

The farm makes a profit of 2.2 pence on each carrot sold to the supermarket, a profit of 0.8 pence on each carrot longer than 22 cm and a loss of 1.2 pence on each carrot shorter than 9 cm.

- (e) Find an estimate of the mean profit per carrot made by the farm. (2)

Organised by Mr Omar Faruk

Organised by Mr Omar Faruk

3. The stem and leaf diagram shows the ages of the 35 male passengers on a cruise.

Age	Key: 1 3 represents an age of 13 years
1	3 (1)
2	7 9 (2)
3	1 2 8 8 (4)
4	5 5 6 7 8 8 9 (7)
5	2 2 3 3 4 4 5 6 6 8 (10)
6	0 1 1 4 4 4 7 (7)
7	3 6 (2)
8	7 8 (2)

- (a) Find the median age of the male passengers. (1)

- (b) Show that the interquartile range (IQR) of these ages is 16 (2)

An outlier is defined as a value that is more than

$1.5 \times \text{IQR}$ above the upper quartile

or

$1.5 \times \text{IQR}$ below the lower quartile

- (c) Show that there are 3 outliers amongst these ages. (3)

- (d) On the grid in Figure 1 on page 9, draw a box plot for the ages of the male passengers on the cruise. (4)

Figure 1 on page 9 also shows a box plot for the ages of the female passengers on the cruise.

- (e) Comment on any difference in the distributions of ages of male and female passengers on the cruise.
State the values of any statistics you have used to support your comment. (1)

Anja, along with her 2 daughters and a granddaughter, now join the cruise.

Anja's granddaughter is younger than both of Anja's daughters.

Anja had her 23rd birthday on the day her eldest daughter was born.

When their 4 ages are included with the other female passengers on the cruise, the box plot does not change.

- (f) State, giving reasons, what you can say about (3)
- (i) the granddaughter's age
 - (ii) Anja's age.

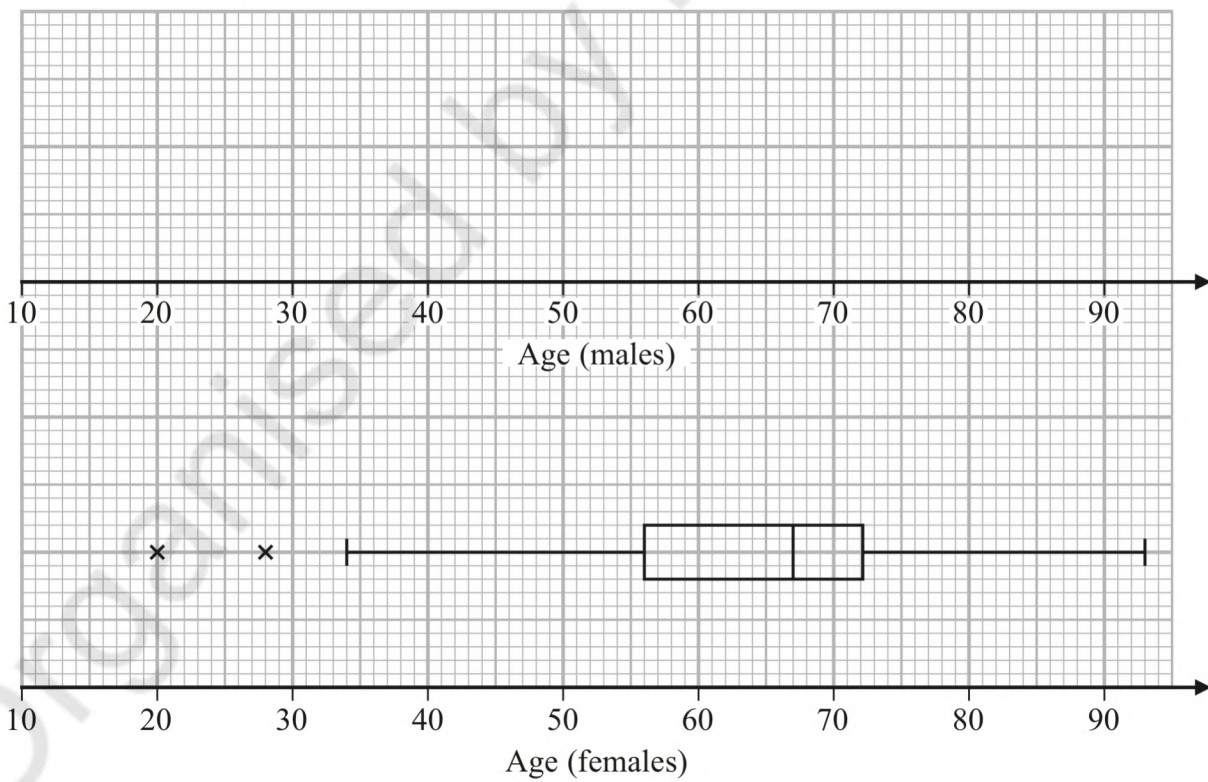


Figure 1

Turn over for a spare grid if you need to redraw your box plot.

3. The stem and leaf diagram shows the number of deliveries made by Pat each day for 24 days

Key: 10 | 8 represents 108 deliveries

10	8	9									(2)	
11	0	3	6	6	6	8	8	9	9	9	9	(11)
12	4	5	5	5	5	5	5	8				(8)
13	<i>a</i>	<i>b</i>	<i>c</i>									(3)

where a , b and c are positive integers with $a < b < c$

An outlier is defined as any value greater than $1.5 \times$ interquartile range above the upper quartile.

Given that there is only one outlier for these data,

- (a) show that $c = 9$

(3)

The number of deliveries made by Pat each day is represented by d

The data in the stem and leaf diagram are coded using

$$x = d - 125$$

and the following summary statistics are obtained

$$\sum x = -96 \quad \text{and} \quad \sum (x - \bar{x})^2 = 1306$$

- (b) Find the mean number of deliveries.

(3)

- (c) Find the standard deviation of the number of deliveries.

(2)

One of these 24 days is selected at random. The random variable D represents the number of deliveries made by Pat on this day.

The random variable $X = D - 125$

- (d) Find $P(D > 118 | X < 0)$

(2)

Organised by Mr Omar Faruk

Organised by Mr Omar Faruk

1. The company *Seafield* requires contractors to record the number of hours they work each week. A random sample of 38 weeks is taken and the number of hours worked per week by contractor Kiana is summarised in the stem and leaf diagram below.

Stem	Leaf	
1	4 4 4 5 5 5 6 6 9 9 9	(11)
2	1 2 2 3 3 4 4 4 w 9	(10)
3	2 3 4 4 5 6 7 7 7 9	(10)
4	1 1 2 3	(4)
5	1 9	(2)
6	4	(1)

Key : 3|2 means 32

The quartiles for this distribution are summarised in the table below.

Q_1	Q_2	Q_3
x	26	y

- (a) Find the values of w , x and y (3)

Kiana is looking for outliers in the data. She decides to classify as outliers any observations greater than

$$Q_3 + 1.0 \times (Q_3 - Q_1)$$

- (b) Showing your working clearly, identify any outliers that Kiana finds. (2)

- (c) Draw a box plot for these data in the space provided on the grid opposite. (3)

- (d) Use the formula

$$\text{skewness} = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}$$

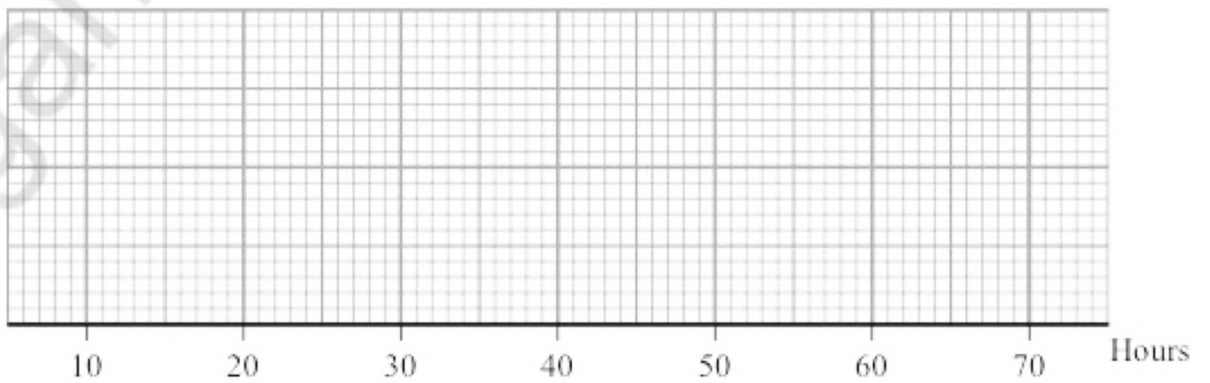
to find the skewness of these data. Give your answer to 2 significant figures. (2)

Kiana's new employer, *Landacre*, wishes to know the average number of hours per week she worked during her employment at *Seafield* to help calculate the cost of employing her.

- (e) Explain why *Landacre* might prefer to know Kiana's mean, rather than median, number of hours worked per week. (1)

Organised by Mr Omar Faruk

Lined writing area for notes or calculations.



Turn over for a spare grid if you need to redraw your box plot.

3. Gill buys a bag of logs to use in her stove. The lengths, l cm, of the 88 logs in the bag are summarised in the table below.

Length (l)	Frequency (f)
$15 < l \leq 20$	19
$20 < l \leq 25$	35
$25 < l \leq 27$	16
$27 < l \leq 30$	15
$30 < l \leq 40$	3

A histogram is drawn to represent these data.

The bar representing logs with length $27 < l \leq 30$ has a width of 1.5 cm and a height of 4 cm.

- (a) Calculate the width and height of the bar representing log lengths of $20 < l \leq 25$ (3)

- (b) Use linear interpolation to estimate the median of l (2)

The maximum length of log Gill can use in her stove is 26 cm.

Gill estimates, using linear interpolation, that x logs from the bag will fit into her stove.

- (c) Show that $x = 62$ (1)

Gill randomly selects 4 logs from the bag.

- (d) Using $x = 62$, find the probability that all 4 logs will fit into her stove. (2)

The weights, W grams, of the logs in the bag are coded using $y = 0.5w - 255$ and summarised by

$$n = 88 \quad \sum y = 924 \quad \sum y^2 = 12862$$

- (e) Calculate
- (i) the mean of W (3)
- (ii) the variance of W (3)

Organised by Mr Omar Faruk

Organised by Mr Omar Faruk

1. The stem lengths of a sample of 120 tulips are recorded in the grouped frequency table below.

Stem length (cm)	Frequency
$40 \leq x < 42$	12
$42 \leq x < 45$	18
$45 \leq x < 50$	23
$50 \leq x < 55$	35
$55 \leq x < 58$	24
$58 \leq x < 60$	8

A histogram is drawn to represent these data.

The area of the bar representing the $40 \leq x < 42$ class is 16.5 cm^2

- (a) Calculate the exact area of the bar representing the $42 \leq x < 45$ class.

(2)

The height of the tallest bar in the histogram is 10 cm.

- (b) Find the exact height of the second tallest bar.

(3)

Q_1 for these data is 45 cm.

- (c) Use linear interpolation to find an estimate for

(i) Q_2

(ii) the interquartile range.

(4)

One measure of skewness is given by

$$\frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$$

- (d) By calculating this measure, describe the skewness of these data.

(2)

Organised by Mr Omar Faruk

3. Morgan is investigating the body length, b centimetres, of squirrels.

A random sample of 8 squirrels is taken and the data for each squirrel is coded using

$$x = \frac{b - 21}{2}$$

The results for the coded data are summarised below

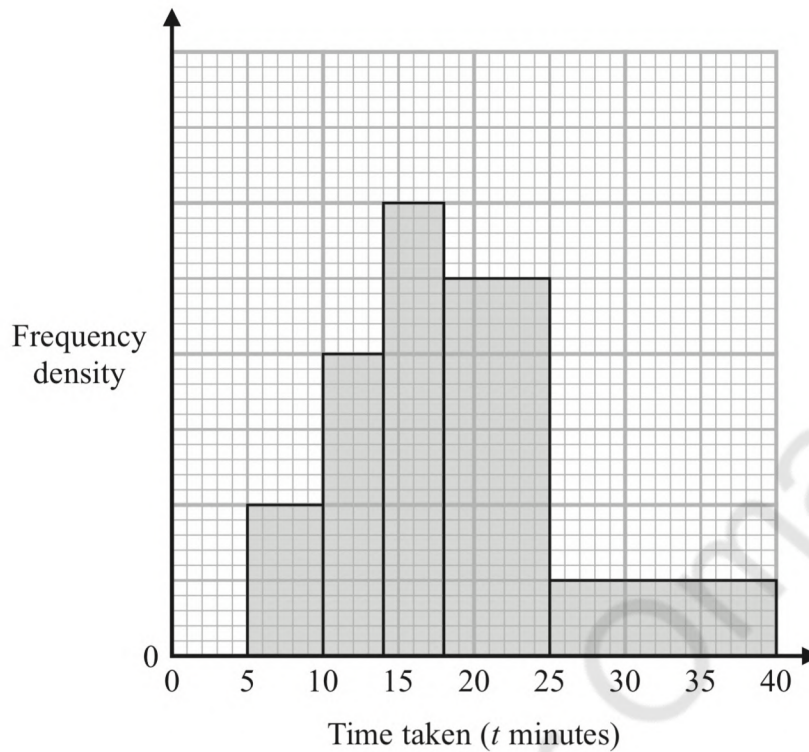
$$\sum x = -1.2 \quad \sum x^2 = 5.1$$

- (a) Find the mean of b (3)
- (b) Find the standard deviation of b (3)

A 9th squirrel is added to the sample. Given that for all 9 squirrels $\sum x = 0$

- (c) find
 - (i) the body length of the 9th squirrel, (2)
 - (ii) the standard deviation of x for all 9 squirrels. (2)

1. The histogram shows the times taken, t minutes, by each of 100 people to swim 500 metres.



- (a) Use the histogram to complete the frequency table for the times taken by the 100 people to swim 500 metres.

Time taken (t minutes)	5 – 10	10 – 14	14 – 18	18 – 25	25 – 40
Frequency (f)	10	16	24		

(1)

- (b) Estimate the number of people who took less than 16 minutes to swim 500 metres.

(2)

- (c) Find an estimate for the mean time taken to swim 500 metres.

(2)

Given that $\sum ft^2 = 41\,033$

- (d) find an estimate for the standard deviation of the times taken to swim 500 metres.

(2)

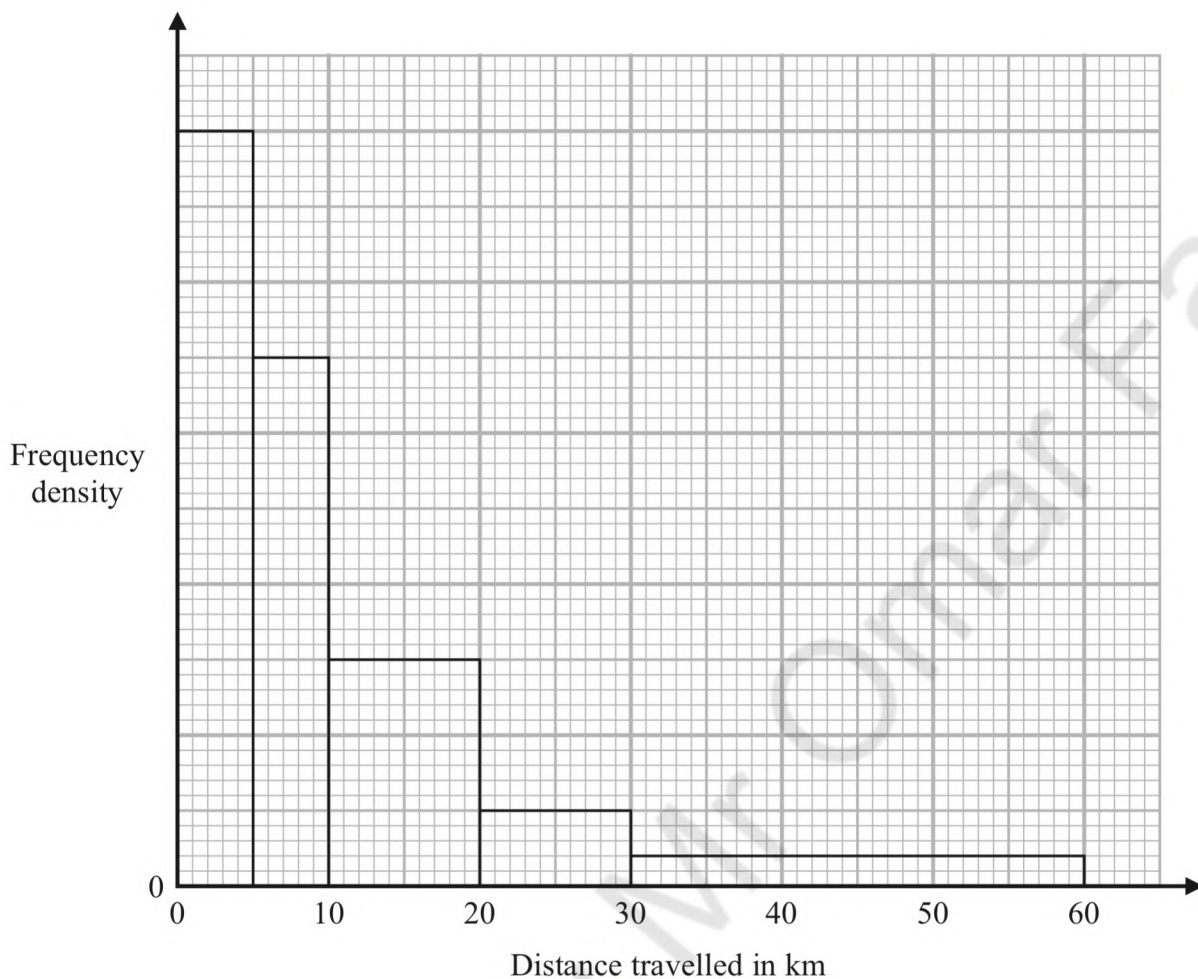
Given that $Q_3 = 23$

- (e) use linear interpolation to estimate the interquartile range of the times taken to swim 500 metres.

Organised by Mr Omar Faruk

Organised by Mr Omar Faruk

1. The histogram shows the distances, in km, that 274 people travel to work.



Given that 60 of these people travel between 10 km and 20 km to work, estimate

- (a) the number of people who travel between 22 km and 45 km to work, (3)
- (b) the median distance travelled to work by these 274 people, (2)
- (c) the mean distance travelled to work by these 274 people. (3)

Organised by Mr Omar Faruk

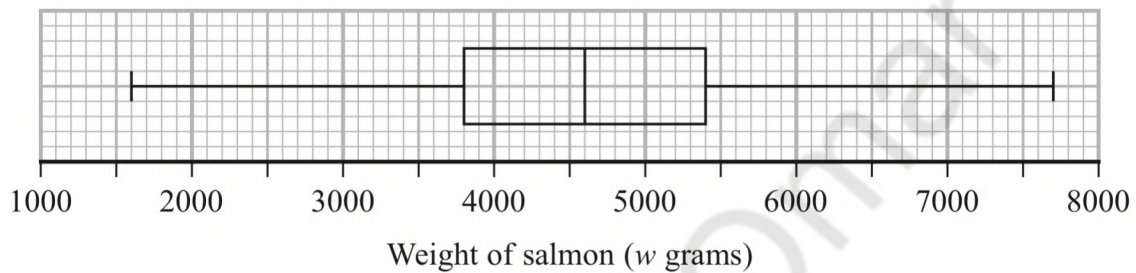
Organised by Mr Omar Faruk

3. Jim records the length, l mm, of 81 salmon. The data are coded using $x = l - 600$ and the following summary statistics are obtained.

$$n = 81 \qquad \sum x = 3711 \qquad \sum x^2 = 475181$$

- (a) Find the mean length of these salmon. (3)
- (b) Find the variance of the lengths of these salmon. (2)

The weight, w grams, of each of the 81 salmon is recorded to the nearest gram. The recorded results for the 81 salmon are summarised in the box plot below.



- (c) Find the maximum number of salmon that have weights in the interval $4600 < w \leq 7700$ (1)

Raj says that the box plot is incorrect as Jim has not included outliers.

For these data an outlier is defined as a value that is more than

$$1.5 \times \text{IQR above the upper quartile} \quad \text{or} \quad 1.5 \times \text{IQR below the lower quartile}$$

- (d) Show that there are no outliers. (3)

Organised by Mr Omar Faruk

Organised by Mr Omar Faruk

2. The weights, to the nearest kilogram, of a sample of 33 red kangaroos taken in December are summarised in the stem and leaf diagram below.

	Weight (kg)	Totals	Key: 3 2 represents 32 kg
1	6	(1)	
2	3 6	(2)	
3	2 4 6	(3)	
4	2 5 5 6 6 7 8	(7)	
5	3 4 7 7 7 8 9 9	(8)	
6	0 2 2 3 3 7 8	(7)	
7	2 8	(2)	
8	2 6	(2)	
9	4	(1)	

(a) Find

- (i) the value of the median
- (ii) the value of Q_1 and the value of Q_3
for the weights of these red kangaroos.

(3)

For these data an outlier is defined as a value that is

greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$

or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$

(b) Show that there are 2 outliers for these data.

(3)

Figure 1 on page 7 shows a box plot for the weights of the same 33 red kangaroos taken in February, earlier in the year.

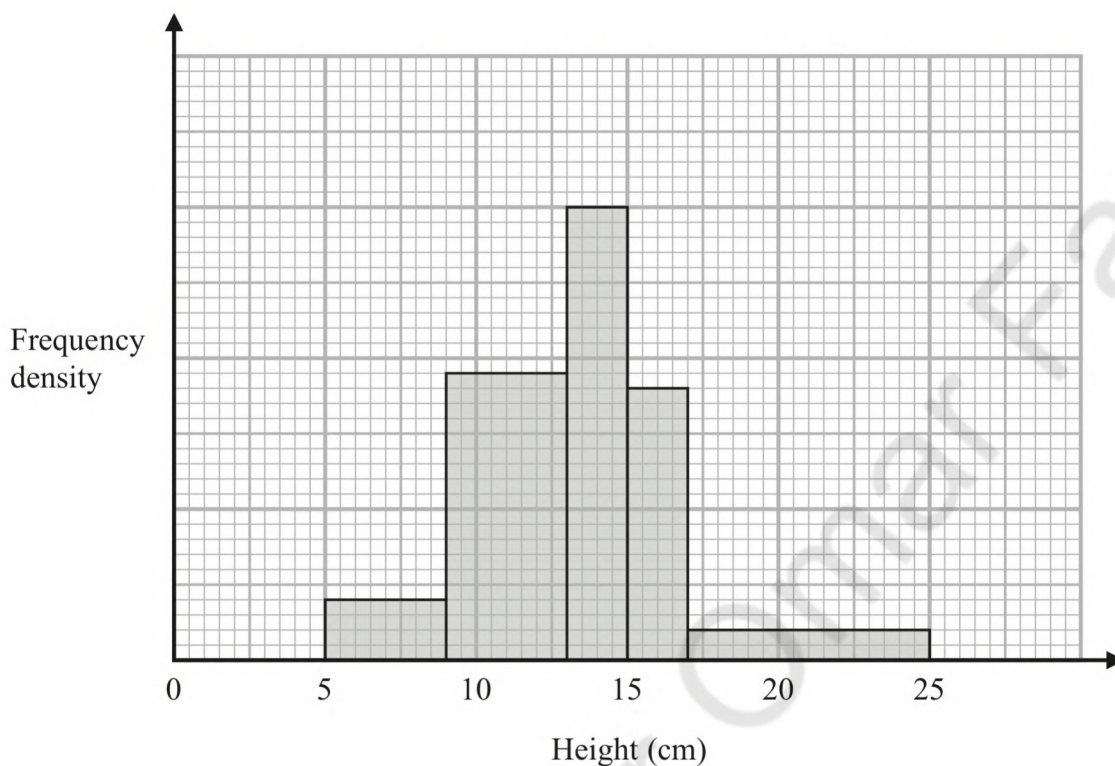
(c) In the space on Figure 1, draw a box plot to represent the weights of these red kangaroos in December.

(4)

(d) Compare the distribution of the weights of red kangaroos taken in February with the distribution of the weights of red kangaroos taken in December of the same year. You should interpret your comparisons in the context of the question.

(3)

1. The histogram below shows the distribution of the heights, to the nearest cm, of 408 plants.



- (a) Use the histogram to complete the following table.

Height (h cm)	$5 \leq h < 9$	$9 \leq h < 13$	$13 \leq h < 15$	$15 \leq h < 17$	$17 \leq h < 25$
Frequency	32	152	120		

(2)

- (b) Use interpolation to estimate the median.

(2)

The mean height of these plants is 13.2 cm correct to one decimal place.

- (c) Describe the skew of these data. Give a reason for your answer.

(1)

Two of these plants are chosen at random.

- (d) Estimate the probability that both of their heights are between 8 cm and 14 cm

(3)

Organised by Mr Omar Faruk

4. A French test and a Spanish test were sat by 11 students.

The table below shows their marks.

Student	A	B	C	D	E	F	G	H	I	J	K
French mark (f)	24	30	32	32	36	36	40	44	50	60	68
Spanish mark (s)	16	90	24	28	32	36	38	44	48	48	68

Greg says that if these points were plotted on a scatter diagram, then the point (30, 90) would be an outlier because 90 is an outlier for the Spanish marks.

An outlier is defined as a value that is

$$\text{greater than } Q_3 + 1.5 \times (Q_3 - Q_1) \text{ or smaller than } Q_1 - 1.5 \times (Q_3 - Q_1)$$

(a) Show that 90 is an outlier for the Spanish marks.

(3)

1. A researcher is investigating the growth of two types of tree, Birch and Maple. The height, to the nearest cm, a seedling grows in one year is recorded for 35 Birch trees and 32 Maple trees. The results are summarised in the back-to-back stem and leaf diagram below.

Totals	Birch		Maple	Totals	
(2)		9 8	2	5 7 7 8 9	(5)
(8)	9 9 9 6 5 3 1 1		3	0 2 6 6 8 9 9	(7)
(9)	9 8 8 7 6 3 1 1 1		4	1 1 1 <i>k</i> 7 8	(6)
(9)	7 7 7 5 4 3 2 1 0		5	0 1 2 3 4 4 4	(7)
(3)		7 6 5	6	3 4 6	(3)
(3)		6 5 4	7	0 7	(2)
(1)		5	8	0 0	(2)

Key: 5 | 6 | 3 means 65 cm for a Birch tree and 63 cm for a Maple tree

The median height that these **Maple** trees grow in one year is 45 cm.

- (a) Find the value of *k*, used in the stem and leaf diagram. (1)
- (b) Find the lower quartile and the upper quartile of the height grown in one year for these **Birch** trees. (2)

The researcher defines an outlier as an observation that is

$$\text{greater than } Q_3 + 1.5 \times (Q_3 - Q_1) \text{ or less than } Q_1 - 1.5 \times (Q_3 - Q_1)$$

- (c) Show that there is only one outlier amongst the Birch trees. (2)

The grid on page 3 shows a box plot for the heights that the Maple trees grow in one year.

- (d) On the same grid draw a box plot for the heights that the Birch trees grow in one year. (4)
- (e) Comment on any difference in the distributions of the growth of these Birch trees and the growth of these Maple trees.
State the values of any statistics you have used to support your comment. (1)

The researcher realises he has missed out 4 pieces of data for the **Maple** trees. The heights each seedling grows in one year, to the nearest cm, in ascending order, for these 4 Maple trees are 27 cm, *a* cm, 48 cm, 2*a* cm.

Given that there is no change to the box plot for the **Maple** trees given on page 3

- (f) find the range of possible values for *a*
Show your working clearly. (3)

Organised by Mr Omar Faruk

3. The lengths, x mm, of 50 pebbles are summarised in the table below.

Length	Frequency
$20 \leq x < 30$	2
$30 \leq x < 32$	16
$32 \leq x < 36$	20
$36 \leq x < 40$	8
$40 \leq x < 45$	3
$45 \leq x < 50$	1

A histogram is drawn to represent these data.

The bar representing the class $32 \leq x < 36$ is 2.5 cm wide and 7.5 cm tall.

- (a) Calculate the width and the height of the bar representing the class $30 \leq x < 32$ (3)
- (b) Using linear interpolation, estimate the median of x (2)

The weight, w grams, of each of the 50 pebbles is coded using $10y = w - 20$
These coded data are summarised by

$$\sum y = 104 \qquad \sum y^2 = 233.54$$

- (c) Show that the mean of w is 40.8 (2)
- (d) Calculate the standard deviation of w (4)

The weight of a pebble recorded as 40.8 grams is added to the sample.

- (e) Without carrying out any further calculations, state, giving a reason, what effect this would have on the value of (3)
- (i) the mean of w
- (ii) the standard deviation of w

Organised by Mr Omar Faruk

1. The back-to-back stem and leaf diagram on page 3 shows information about the running times of 31 Action films and 31 Comedy films.
The running times are given to the nearest minute.

(a) Write down the modal running time for these Action films.

(1)

Some of the quartiles for these two distributions are shown in the table below.

	Action films	Comedy films
Lower quartile	121	<i>a</i>
Median	<i>b</i>	117
Upper quartile	138	<i>c</i>

(b) Find the value of *a*, the value of *b* and the value of *c*

(3)

(c) For these Action films find, to one decimal place,

- (i) the mean running time,
- (ii) the standard deviation of the running times.

(You may use $\sum x = 4016$ and $\sum x^2 = 525056$ where *x* is the running time, in minutes, of an Action film.)

(3)

One measure of skewness is found using

$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}$$

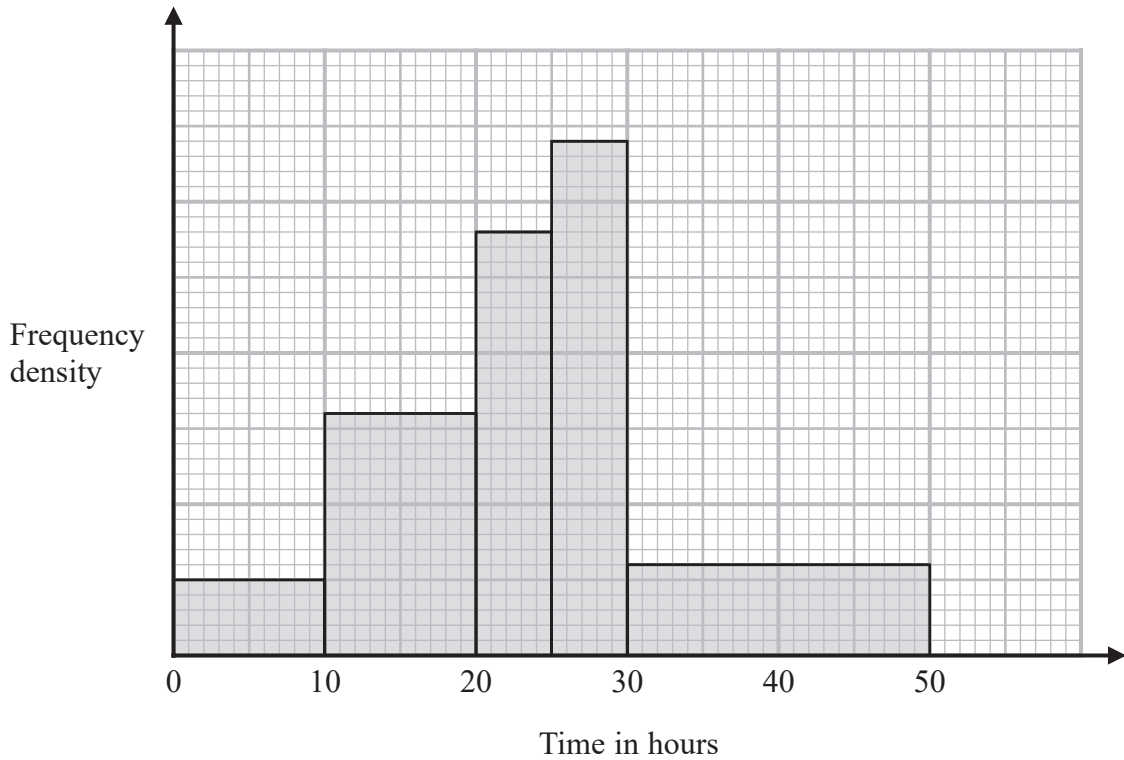
(d) Evaluate this measure and describe the skewness for the running times of these Action films.

(2)

(e) Comment on one difference between the distribution of the running times of these Action films and the distribution of the running times of these Comedy films.
State the values of any statistics you have used to support your comment.

(1)

5.



The histogram shows the number of hours worked in a given week by a group of 64 freelance photographers.

(a) Give a reason to justify the use of a histogram to represent these data. (1)

Given that 16 of these freelance photographers spent between 10 and 20 hours working in this week,

(b) estimate the number that spent between 12 and 24 hours working in this week. (3)

(c) Find an estimate for the median time spent working in this week by these 64 freelance photographers. (2)

Charlie decides to model these data using a normal distribution.
 Charlie calculates an estimate of the mean to be 23.9 hours to one decimal place.

(d) Comment on Charlie’s decision to use a normal distribution.
 Give a justification for your answer. (2)

2. As part of an investigation, Bobby collects a sample of 47 observations, x

The results are shown in the following stem and leaf diagram, where a is a constant.

Stem	Leaf	
2	1 2 5 7 7	(5)
3	0 2 2 3 4 5 5 5 9 9	(10)
4	0 0 1 4 4 5 7 8 8 9 9	(11)
5	3 3 5 6 7 9	(6)
6	0 2 a a a 7 8	(7)
7	1 2 3 6 8	(5)
8	0 6 7	(3)

Key: 3 | 2 means 0.32

(a) Find the range of these observations. (1)

(b) Find the value of the median of these observations. (1)

Given that the interquartile range of these observations is 0.31

(c) find the value of a (3)

Bobby calculates the following statistics from these observations

$$\sum x = 23.72 \qquad \sum x^2 = 13.4228$$

(d) Show that the standard deviation of these observations is 0.176 to 3 significant figures. (2)

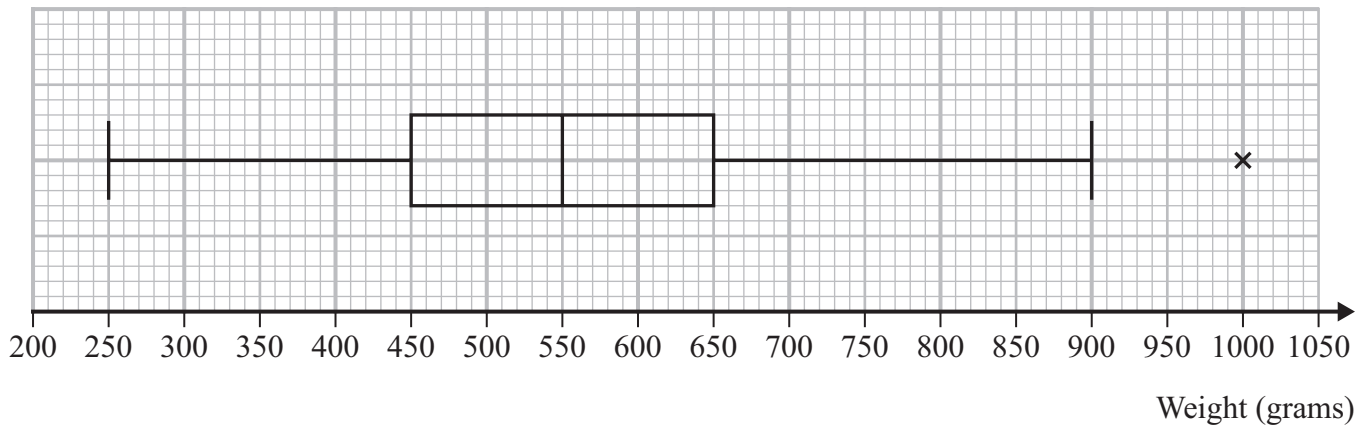
Bobby now collects 18 more observations, y , from the same investigation.

(e) Using all 65 observations, the sample mean is 0.502 and the sample standard deviation is 0.204

(i) Show that $\sum y = 8.91$ (2)

(ii) Showing your working, calculate $\sum y^2$ (3)

7. A farmer supplies cabbages to a local supermarket. The supermarket manager takes a random sample of 200 of these cabbages and the weight of each cabbage is recorded. The weights of the cabbages in the sample are summarised in the following box plot.



- (a) Use linear interpolation to estimate the probability that a randomly chosen cabbage from this sample has a weight of less than 570 grams. (2)

The supermarket manager takes the quartiles from this sample and now decides to classify any cabbage whose weight is at least $Q_3 + 1.5 \times (Q_3 - Q_1)$ as extra-large.

- (b) Find the minimum weight of a cabbage that the supermarket manager would classify as extra-large. (1)

The supermarket manager assumes that the weights of cabbages supplied to the supermarket may be modelled by a normal distribution.

- (c) Explain whether or not the box plot supports this assumption. (1)

The farmer records the weight of every cabbage harvested from a field and classifies any cabbage whose weight is more than 2 standard deviations above the mean as an outlier.

Assuming that the weights of the cabbages from this field may be modelled by a normal distribution,

- (d) find the probability that a randomly selected cabbage harvested from the field would be classified as an outlier using the farmer's method. (2)

The mean weight of the cabbages harvested from the field is 560 grams.

Given that a cabbage with a weight of 1000 grams is classified as an outlier by the farmer,

- (e) find the maximum possible value of the standard deviation of the weights of cabbages harvested from the field. (2)

