# 2014-2018 Chapter 1-3:

Mr Faruk
Teacher of Mathematics
BSc/MSc/PGCE Mathematics
cieigcsesolutions@gmail.com

Leave blank

**2.** A rugby club coach uses club records to take a random sample of 15 players from 1990 and an independent random sample of 15 players from 2010. The body weight of each player was recorded to the nearest kg and the results from 2010 are summarised in the table below.

| **Body weight (kg)** | 75–79 | 80–84 | 85–89 | 90–94 | 95–99 | 100–104 | 105–109 |
|---|---|---|---|---|---|---|---|
| **Number of Players (2010)** | 1 | 2 | 2 | 4 | 3 | 2 | 1 |

(a) Find the estimated values in kg of the summary statistics *a*, *b* and *c* in the table below.

| | **Estimate in 1990** | **Estimate in 2010** |
|---|---|---|
| **Mean** | 83.0 | *a* |
| **Median** | 82.0 | *b* |
| **Variance** | 44.0 | *c* |

Give your answers to 3 significant figures.

**(6)**

The rugby coach claims that players' body weight increased between 1990 and 2010.

(b) Using the table in part (a), comment on the rugby coach's claim.

**(2)**

Leave blank

**Question 2 continued**

**Q2**

**(Total 8 marks)**

Leave blank

**8.** A manager records the number of hours of overtime claimed by 40 staff in a month.

The histogram in Figure 1 represents the results.

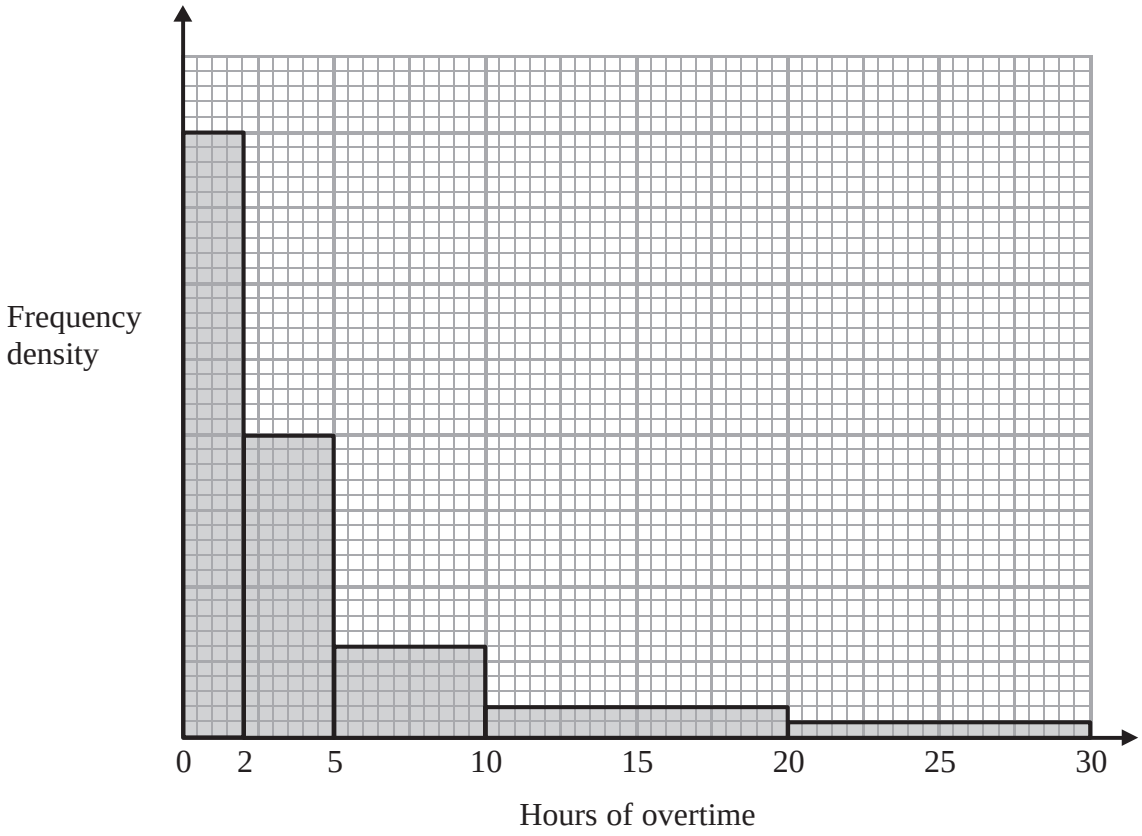



**Figure 1**

(a) Calculate the number of staff who have claimed less than 10 hours of overtime in the month.

**(4)**

(b) Estimate the median number of hours of overtime claimed by these 40 staff in the month.

**(2)**

(c) Estimate the mean number of hours of overtime claimed by these 40 staff in the month.

**(2)**

The manager wants to compare these data with overtime data he collected earlier to find out if the overtime claimed by staff has decreased.

(d) State, giving a reason, whether the manager should use the median or the mean to compare the overtime claimed by staff.

**(2)**

Leave blank

**Question 8 continued**

Leave blank

**2.** A sports teacher recorded the number of press-ups done by his students in two minutes. He recorded this information for a Year 7 class and for a Year 11 class.

The back-to-back stem and leaf diagram shows this information.

| Totals | Year 7 class | | Year 11 class | Totals |
|---|---|---|---|---|
| (6) | 8 7 6 5 5 4 | 1 | | |
| (10) | 9 7 7 6 5 4 4 4 2 2 | 2 | 0 5 6 9 | (4) |
| (7) | 8 7 5 4 3 3 0 | 3 | 3 4 5 8 8 | (5) |
| (5) | 9 9 7 2 2 | 4 | 0 5 6 7 9 | (5) |
| (3) | 8 4 0 | 5 | 0 3 5 5 6 6 7 7 7 9 9 | (11) |
| | | 6 | 0 3 3 3 3 4 8 | (7) |

Key: 2|4|0 means 42 press-ups for a Year 7 student and 40 press-ups for a Year 11 student

(a) Find the median number of press-ups for each class.

**(2)**

For the Year 11 class, the lower quartile is 38 and the upper quartile is 59

(b) Find the lower quartile and the upper quartile for the Year 7 class.

**(2)**

(c) Use the medians and quartiles to describe the skewness of each of the two distributions.

**(3)**

(d) Give two reasons why the normal distribution should not be used to model the number of press-ups done by the Year 11 class.

**(2)**

Leave blank

**Question 2 continued**

**Q2**

**(Total 9 marks)**

Leave blank

**2.**

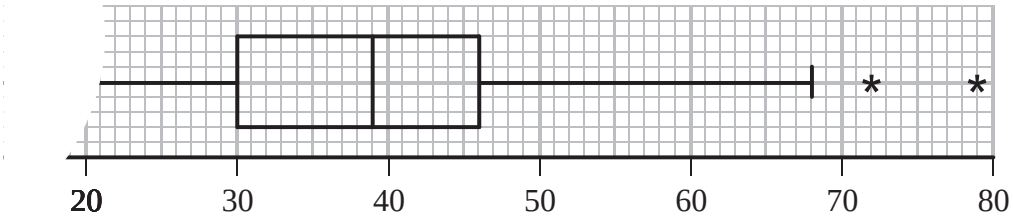



**Figure 1**

Figure 1 shows part of a box and whisker plot for the marks in an examination with a large number of candidates. Part of the lower whisker has been torn off.

(a) Given that 75% of the candidates passed the examination, state the lowest mark for the award of a pass.

**(1)**

(b) Given that the top 25% of the candidates achieved a merit grade, state the lowest mark for the award of a merit grade.

**(1)**

An outlier is defined as any value greater than $c$ or any value less than $d$ where

$$c = Q_3 + 1.5(Q_3 - Q_1)$$
$$d = Q_1 - 1.5(Q_3 - Q_1)$$

(c) Find the value of $c$ and the value of $d$.

**(2)**

(d) Write down the 3 highest marks scored in the examination.

**(2)**

The 3 lowest marks in the examination were 5, 10 and 15

(e) On the diagram on page 7, complete the box and whisker plot.

**(3)**

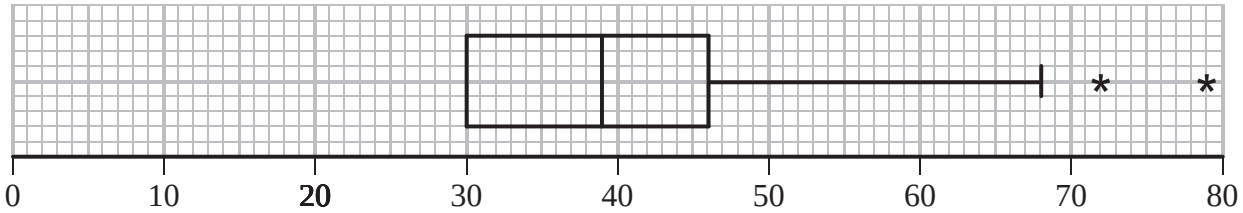
Three candidates are selected at random from those who took this examination.

(f) Find the probability that all 3 of these candidates passed the examination but only 2 achieved a merit grade.

**(3)**

Leave blank

**Question 2 continued**




**Turn over for a spare diagram if you need to redraw your plot.**

Leave blank

**6.** Yujie is investigating the weights of 10 young rabbits. She records the weight, $x$ grams, of each rabbit and the results are summarised below.

$$\sum x = 8360 \quad \text{and} \quad \sum (x - \bar{x})^2 = 63840$$

(a) Calculate the mean and the standard deviation of the weights of these rabbits.

**(3)**

Given that the median weight of these rabbits is 815 grams,

(b) describe, giving a reason, the skewness of these data.

**(2)**

Two more rabbits weighing 776 grams and 896 grams are added to make a group of 12 rabbits.

(c) State, giving a reason, how the inclusion of these two rabbits would affect the mean.

**(2)**

(d) By considering the change in $\sum (x - \bar{x})^2$, state what effect the inclusion of these two rabbits would have on the standard deviation.

**(2)**

Leave blank

**Question 6 continued**

**(Total 9 marks)**

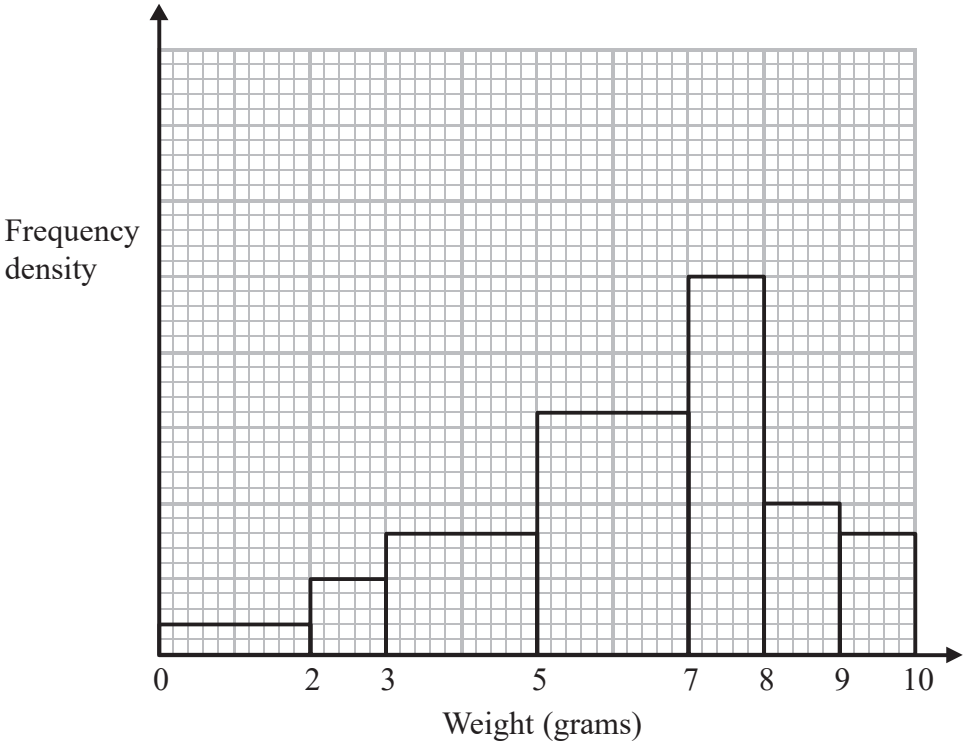
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

**1.** Ralph records the weights, in grams, of 100 tomatoes. This information is displayed in the histogram below.




Given that 5 of the tomatoes have a weight between 2 and 3 grams,

(a) find the number of tomatoes with a weight between 0 and 2 grams.

**(2)**

One of the tomatoes is selected at random.

(b) Find the probability that it weighs more than 3 grams.

**(2)**

(c) Estimate the proportion of the tomatoes with a weight greater than 6.25 grams.

**(2)**

(d) Using your answer to part (c), explain whether or not the median is greater than 6.25 grams.

**(1)**

Given that the mean weight of these tomatoes is 6.25 grams and using your answer to part (d),

(e) describe the skewness of the distribution of the weights of these tomatoes. Give a reason for your answer.

**(1)**

Two of these 100 tomatoes are selected at random.

(f) Estimate the probability that both tomatoes weigh within 0.75 grams of the mean.

**(4)**

Leave blank

**Question 1 continued**

Leave blank

**1.** Two classes of students, class *A* and class *B*, sat a test.

Class *A* has 10 students. Class *B* has 15 students.

Each student achieved a score, $x$, on the test and their scores are summarised in the table below.

| | $n$ | $\sum x$ | $\sum x^2$ |
|---|---|---|---|
| Class *A* | 10 | 770 | 59610 |
| Class *B* | 15 | $t$ | 58035 |

The mean score for Class *A* is 77 and the mean score for Class *B* is 61

(a) Find the value of $t$

**(1)**

(b) Calculate the variance of the test scores for each class.

**(3)**

The highest score on the test was 95 and the lowest score was 45

These were each scored by students from the same class.

(c) State, with a reason, which class you believe they were from.

**(1)**

The two classes are combined into one group of 25 students.

(d) (i) Find the mean test score for all 25 students.

(ii) Find the variance of the test scores for all 25 students.

**(4)**

The teacher of class *A* later realises that he added up the test scores for his class incorrectly. Each student's test score in class *A* should be increased by 3

(e) Without further calculations, state, with a reason, the effect this will have on

(i) the variance of the test scores for class *A*

(ii) the mean test score for all 25 students

(iii) the variance of the test scores for all 25 students.

**(3)**

Leave blank

**Question 1 continued**

Leave blank

**2.** The table below shows the distances (to the nearest km) travelled to work by the 50 employees in an office.

| Distance (km) | Frequency (f) | Distance midpoint ($x$) |
|---|---|---|
| 0 – 2 | 16 | 1.25 |
| 3 – 5 | 12 | 4 |
| 6 – 10 | 10 | 8 |
| 11 – 20 | 8 | 15.5 |
| 21 – 40 | 4 | 30.5 |

[You may use $\sum fx = 394$, $\sum fx^2 = 6500$]

A histogram has been drawn to represent these data.

The bar representing the distance of 3 – 5 has a width of 1.5 cm and a height of 6 cm.

(a) Calculate the width and height of the bar representing the distance of 6 – 10

**(3)**

(b) Use linear interpolation to estimate the median distance travelled to work.

**(2)**

(c) (i) Show that an estimate of the mean distance travelled to work is 7.88 km.

(ii) Estimate the standard deviation of the distances travelled to work.

**(4)**

(d) Describe, giving a reason, the skewness of these data.

**(2)**

Peng starts to work in this office as the 51st employee.

She travels a distance of 7.88 km to work.

(e) Without carrying out any further calculations, state, giving a reason, what effect Peng's addition to the workforce would have on your estimates of the

(i) mean,

(ii) median,

(iii) standard deviation

of the distances travelled to work.

**(3)**

Leave blank

**Question 2 continued**

Leave blank

**8.**

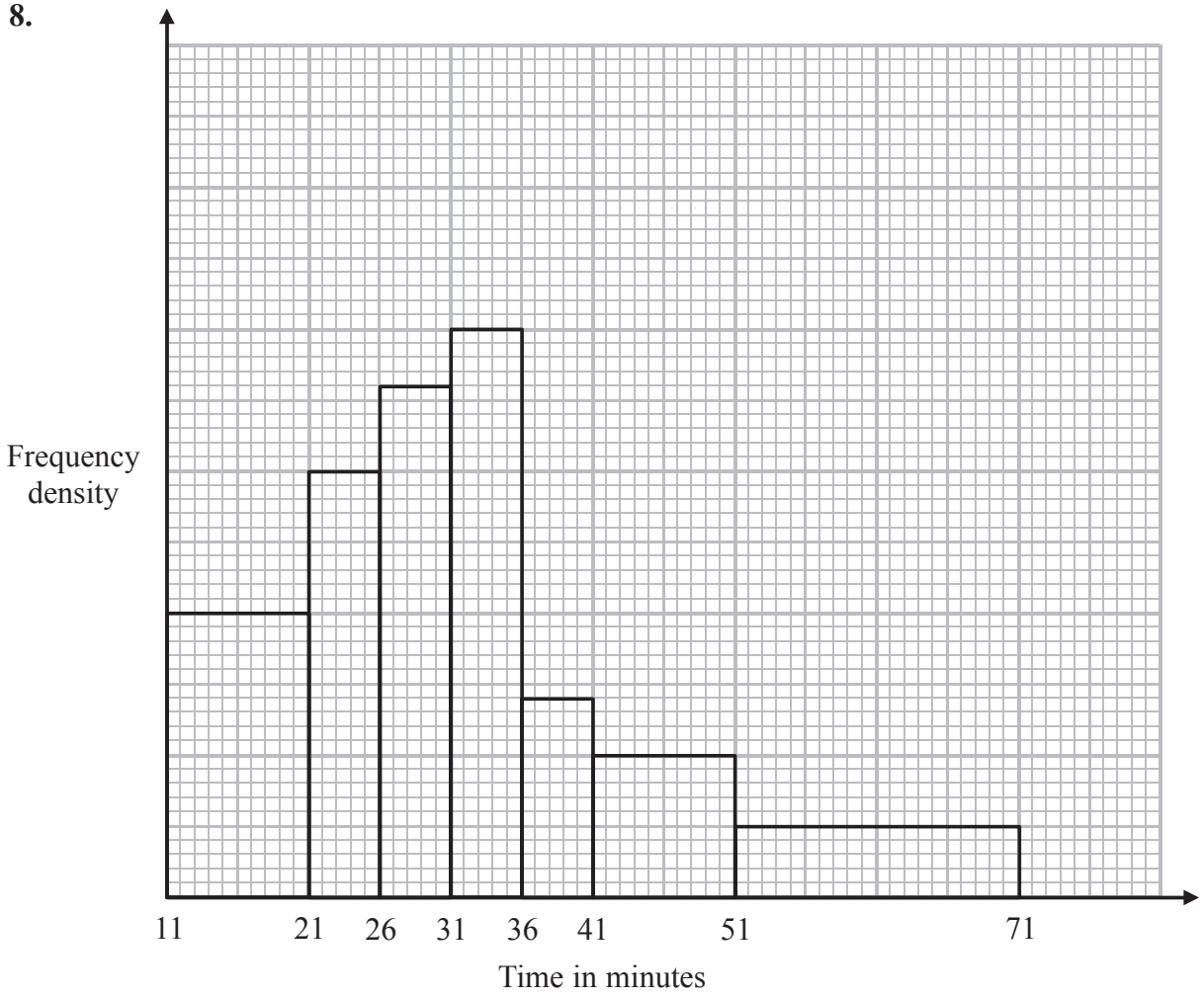



**Figure 1**

The histogram in Figure 1 summarises the times, in minutes, that 200 people spent shopping in a supermarket.

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

**(1)**

Given that 40 people spent between 11 and 21 minutes shopping in the supermarket, estimate

(b) the number of people that spent between 18 and 25 minutes shopping in the supermarket,

**(3)**

(c) the median time spent shopping in the supermarket by these 200 people.

**(2)**

The mid-point of each bar is represented by $x$ and the corresponding frequency by f.

(d) Show that $\sum fx = 6390$

**(2)**

Leave blank

**Question 8 continued**

Given that $\sum f x^2 = 238\,430$

(e) for the data shown in the histogram, calculate estimates of

(i) the mean,

(ii) the standard deviation.

**(3)**

A coefficient of skewness is given by $\dfrac{3(\text{mean} - \text{median})}{\text{standard deviation}}$

(f) Calculate this coefficient of skewness for these data.

**(1)**

The manager of the supermarket decides to model these data with a normal distribution.

(g) Comment on the manager's decision. Give a justification for your answer.

**(2)**

Leave blank

**Question 8 continued**

Leave blank

**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 ($m$) |
|---|---|---|
| $0 \leqslant v < 5$ | 20 | 2.5 |
| $5 \leqslant v < 10$ | 24 | $a$ |
| $10 \leqslant v < 15$ | 16 | 12.5 |
| $15 \leqslant v < 20$ | 14 | 17.5 |
| $20 \leqslant v < 30$ | 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)**

**Question 4 continued**

Leave blank

Leave blank

**1.** Nina weighed a random sample of 50 carrots from her shop and recorded the weight, in grams to the nearest gram, for each carrot. The results are summarised below.

| Weight of carrot | Frequency (f) | Weight midpoint ($x$ grams) |
|---|---|---|
| 45–54 | 5 | 49.5 |
| 55–59 | 10 | 57 |
| 60–64 | 22 | 62 |
| 65–74 | 13 | 69.5 |

(You may use $\sum fx^2 = 192\,102.5$)

(a) Use linear interpolation to estimate the median weight of these carrots.

**(2)**

(b) Find an estimate for the mean weight of these carrots.

**(2)**

(c) Find an estimate for the standard deviation of the weights of these carrots.

**(2)**

A carrot is selected at random from Nina's shop.

(d) Estimate the probability that the weight of this carrot is more than 70 grams.

**(2)**

Leave blank

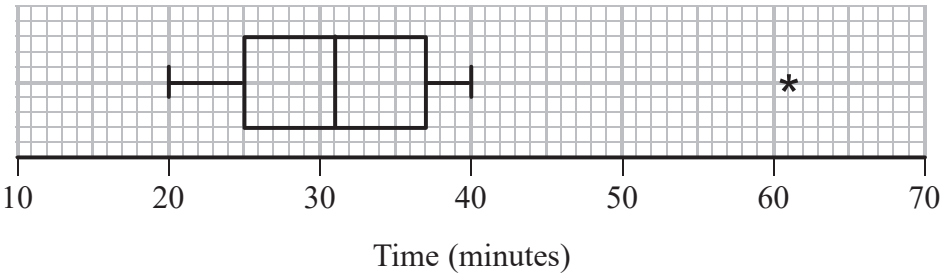
**Question 1 continued**

**Q1**

**(Total 8 marks)**

Leave blank

**2.** The box plot shows the times, $t$ minutes, it takes a group of office workers to travel to work.




(a) Find the range of the times.

**(1)**

(b) Find the interquartile range of the times.

**(1)**

(c) Using the quartiles, describe the skewness of these data. Give a reason for your answer.

**(2)**

Chetna believes that house prices will be higher if the time to travel to work is shorter. She asks a random sample of these office workers for their house prices £$x$, where $x$ is measured in thousands, and obtains the following statistics

$$S_{xx} = 5514 \qquad S_{xt} = 10 \qquad S_{tt} = 1145.6$$

(d) Calculate the product moment correlation coefficient between $x$ and $t$.

**(2)**

(e) State, giving a reason, whether or not your correlation coefficient supports Chetna's belief.

**(2)**

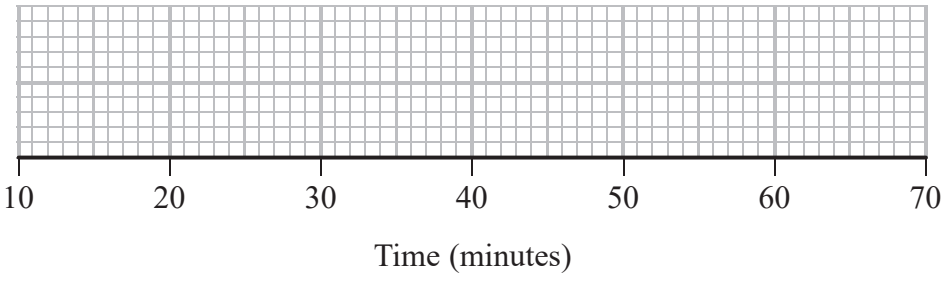
Adam and Betty are part of the group of office workers and they have both moved house.
Adam's time to travel to work changes from 32 minutes to 36 minutes.
Betty's time to travel to work changes from 38 minutes to 58 minutes.
Outliers are defined as values that are more than 1.5 times the interquartile range above the upper quartile.

(f) Showing all necessary calculations, determine how the box plot of times to travel to work will change and draw a new box plot on the grid on page 5.

**(3)**

Leave blank

**Question 2 continued**




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

Leave blank

**2.** Two youth clubs, *Eastyou* and *Westyou*, decided to raise money for charity by running a 5 km race. All the members of the youth clubs took part and the time, in minutes, taken for each member to run the 5 km was recorded.

The times for the *Westyou* members are summarised in Figure 1.

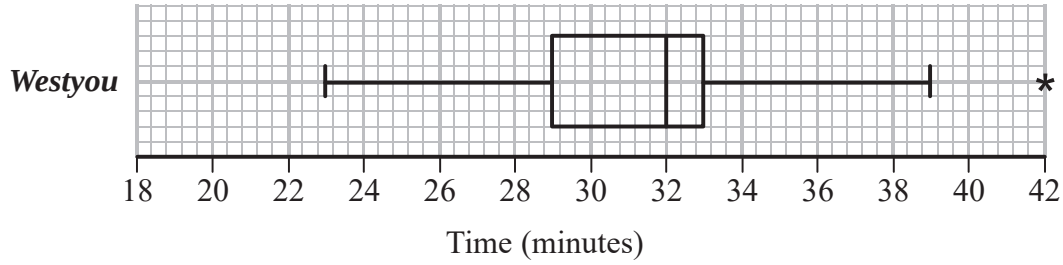



**Figure 1**

(a) Write down the time that is exceeded by 75% of *Westyou* members.

**(1)**

The times for the *Eastyou* members are summarised by the stem and leaf diagram below.

| Stem | Leaf | |
|---|---|---|
| 2 | 0 2 3 4 | (4) |
| 2 | 5 6 8 8 8 9 9 | (7) |
| 3 | 0 0 0 0 0 1 1 1 2 2 2 2 3 4 | (14) |
| 3 | 5 5 5 7 9 | (5) |

**Key**: 2|0 means 20 minutes

(b) Find the value of the median and interquartile range for the *Eastyou* members.

**(3)**

An outlier is a value that falls either

more than $1.5 \times (Q_3 - Q_1)$ above $Q_3$

or more than $1.5 \times (Q_3 - Q_1)$ below $Q_1$

(c) On the grid on page 7, draw a box plot to represent the times of the *Eastyou* members.

**(4)**

(d) State the skewness of each distribution. Give reasons for your answers.

**(3)**

DO NOT WRITE IN THIS AREA

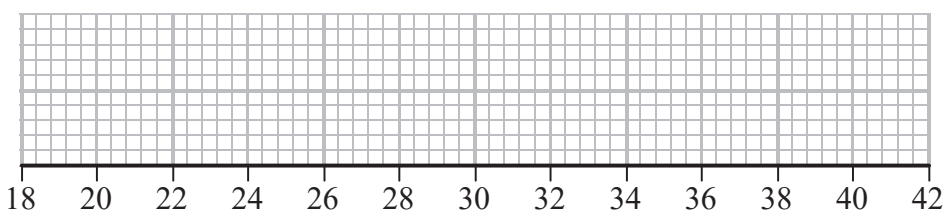
DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Leave blank

**Question 2 continued**

***Eastyou***




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

Leave blank

**5.** The weights, in grams, of a random sample of 48 broad beans are summarised in the table.

| Weight in grams ($x$) | Frequency (f) | Class midpoint ($y$) |
|---|---|---|
| $0.9 < x \leqslant 1.1$ | 9 | 1.0 |
| $1.1 < x \leqslant 1.3$ | 12 | 1.2 |
| $1.3 < x \leqslant 1.5$ | 11 | 1.4 |
| $1.5 < x \leqslant 1.7$ | 8 | 1.6 |
| $1.7 < x \leqslant 1.9$ | 3 | 1.8 |
| $1.9 < x \leqslant 2.1$ | 3 | 2.0 |
| $2.1 < x \leqslant 2.7$ | 2 | 2.4 |

(You may assume $\sum fy^2 = 101.56$)

A histogram was drawn to represent these data. The $2.1 < x \leqslant 2.7$ class was represented by a bar of width 1.5 cm and height 1 cm.

(a) Find the width and height of the $0.9 < x \leqslant 1.1$ class.

**(3)**

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

**(1)**

(c) Estimate the mean and the standard deviation of the weights of these broad beans.

**(4)**

(d) Use linear interpolation to estimate the median of the weights of these broad beans.

**(2)**

One of these broad beans is selected at random.

(e) Estimate the probability that its weight lies between 1.1 grams and 1.6 grams.

**(1)**

One of these broad beans having a recorded weight of 0.95 grams was incorrectly weighed. The correct weight is 1.4 grams.

(f) State, giving a reason, the effect this would have on your answers to part (c). Do not carry out any further calculations.

**(2)**

**Question 5 continued**

Leave blank

Leave blank

**6.** The stem and leaf diagram gives the blood pressure, $x$ mmHg, for a random sample of 19 female patients.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 1 | 2 | | | | | | | | (2) |
| 11 | 2 | 7 | 7 | 8 | 8 | | | | | (5) |
| 12 | 0 | 2 | 2 | 3 | 4 | 4 | 5 | 5 | 7 | (9) |
| 13 | 1 | 2 | 9 | | | | | | | (3) |

Key: 10 | 1 means blood pressure of 101 mmHg

(a) Find the median and the quartiles for these data.

**(3)**

(b) Find the interquartile range $(Q_3 - Q_1)$

**(1)**

An outlier is a value that is greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or less than $Q_1 - 1.5 \times (Q_3 - Q_1)$

(c) Showing your working clearly, identify any outliers for these data.

**(3)**

(d) On the grid on page 21 draw a box and whisker plot to represent these data. Show any outliers clearly.

**(4)**

The above data can be summarised by

$$\sum x = 2299 \quad \text{and} \quad \sum x^2 = 279\,709$$

(e) Calculate the mean and the standard deviation for these data.

**(3)**

For a random sample taken from a normal distribution, a rule for determining outliers is:

an outlier is more than 2.7 × standard deviation above or below the mean.

(f) Find the limits to determine outliers using this rule.

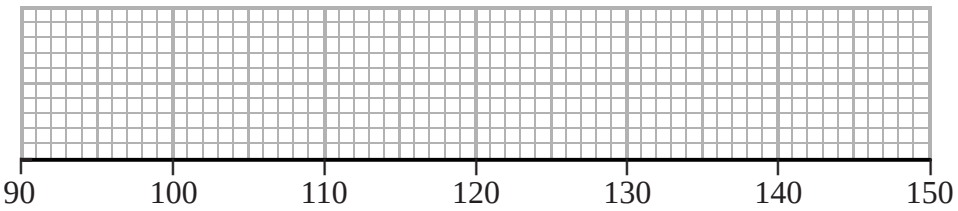
**(2)**

(g) State, giving a reason based on some of the above calculations, whether or not a normal distribution is a suitable model for these data.

**(1)**

Leave blank

**Question 6 continued**




**Turn over for a spare diagram if you need to redraw your plot.**

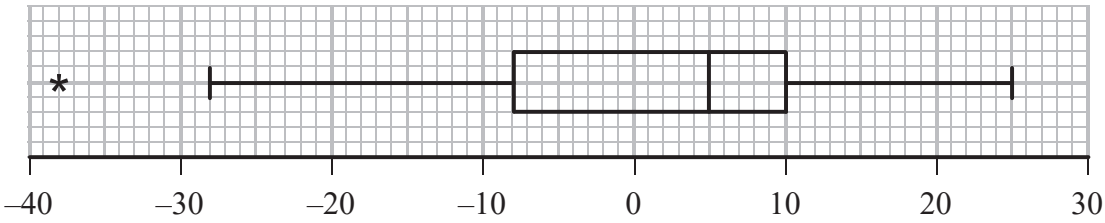
Leave blank

**Question 6 continued**


DO NOT WRITE IN THIS AREA
DO NOT WRITE IN THIS AREA

Leave blank

**1.** At the start of a course, an instructor asked a group of 80 apprentices to estimate the length of a piece of pipe. The error (true length – estimated length) was recorded in centimetres. The results are summarised in the box plot below.




(a) Find the range for these data.

**(1)**

(b) Find the interquartile range for these data.

**(1)**

One month later, the instructor asked the 80 apprentices to estimate the length of a different piece of pipe and recorded their errors. The results are summarised in the table below.

| Error ($e$ cm) | Number of apprentices |
|---|---|
| $-40 < e \leqslant -16$ | 2 |
| $-16 < e \leqslant -8$ | 18 |
| $-8 < e \leqslant 0$ | 33 |
| $0 < e \leqslant 8$ | 14 |
| $8 < e \leqslant 16$ | 10 |
| $16 < e \leqslant 40$ | 3 |

(c) Use linear interpolation to estimate the median error for these data.

**(2)**

(d) Show that the upper quartile for these data, to the nearest centimetre, is 4.

**(1)**

For these data, the lower quartile is –8 and the five worst errors were –25, –21, 18, 23, 28

An outlier is a value that falls either
more than 1.5 × (interquartile range) above the upper quartile or
more than 1.5 × (interquartile range) below the lower quartile.

(e) (i) Show that there are only 2 outliers for these data.

(ii) Draw a box plot for these data on the grid on page 3.

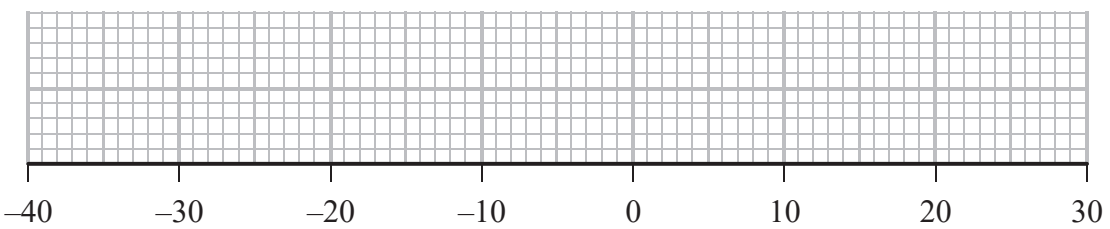
**(6)**

(f) State, giving reasons, whether or not the apprentices' ability to estimate the length of a piece of pipe has improved over the first month of the course.

**(3)**

Leave blank

**Question 1 continued**




**Turn over for a spare diagram if you need to redraw your plot.**

Leave blank

**5.** A company wants to pay its employees according to their performance at work. Last year's performance score $x$ and annual salary $y$, in thousands of dollars, were recorded for a random sample of 10 employees of the company.

The performance scores were

15 24 32 39 41 18 16 22 34 42

(You may use $\sum x^2 = 9011$)

(a) Find the mean and the variance of these performance scores.

**(3)**

The corresponding $y$ values for these 10 employees are summarised by

$$\sum y = 306.1 \quad \text{and} \quad S_{yy} = 546.3$$

(b) Find the mean and the variance of these $y$ values.

**(2)**

The regression line of $y$ on $x$ based on this sample is

$$y = 12.0 + 0.659x$$

(c) Find the product moment correlation coefficient for these data.

**(5)**

(d) State, giving a reason, whether or not the value of the product moment correlation coefficient supports the use of a regression line to model the relationship between performance score and annual salary.

**(1)**

The company decides to use this regression model to determine future salaries.

(e) Find the proposed annual salary, in dollars, for an employee who has a performance score of 35

**(2)**

**Question 5 continued**

Leave blank

Leave blank

**2.** The weights, to the nearest kilogram, of a sample of 33 female spotted hyenas living in the Serengeti are summarised in the stem and leaf diagram below.

| Weight (kg) | | Totals |
|---|---|---|
| 3 | 2 3 7 | (3) |
| 4 | 1 3 3 4 5 5 6 9 | (8) |
| 5 | 1 2 2 3 4 4 5 5 5 7 8 8 9 9 9 | (15) |
| 6 | 2 3 3 | (3) |
| 7 | 1 4 7 | (3) |
| 8 | 4 | (1) |

**Key**: 3|2 means 32

(a) Find the median and quartiles for the weights of the female spotted hyenas.

**(3)**

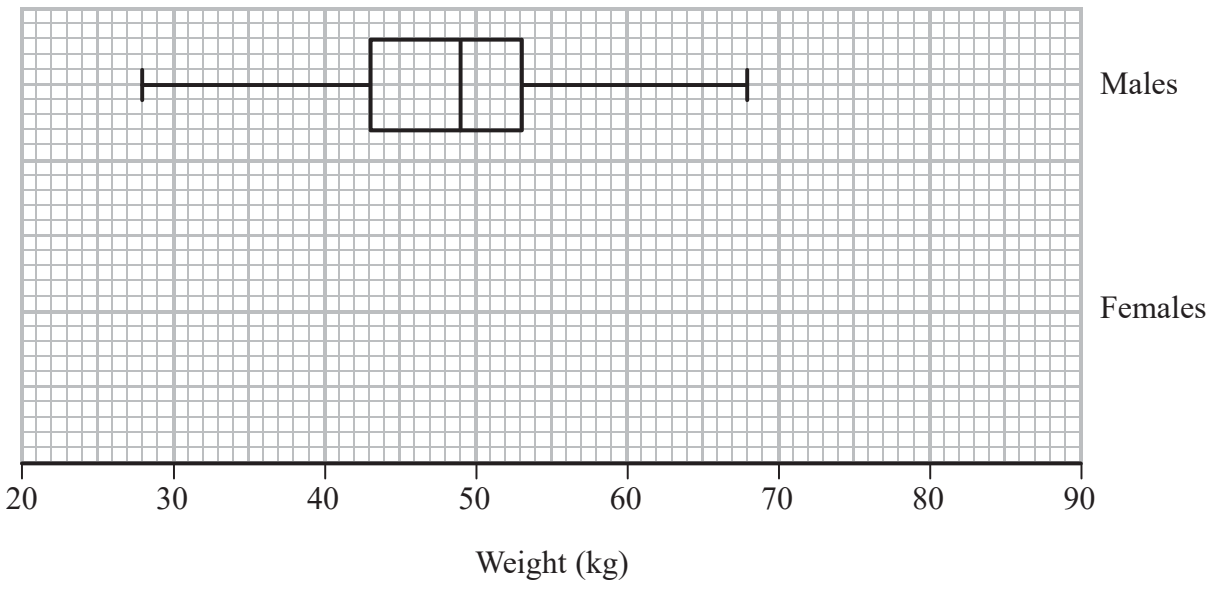
An outlier is defined as any value greater than $c$ or any value less than $d$ where

$$c = Q_3 + 1.5(Q_3 - Q_1)$$
$$d = Q_1 - 1.5(Q_3 - Q_1)$$

(b) Showing your working clearly, identify any outliers for these data.

**(3)**

The weights, to the nearest kilogram, of a sample of male spotted hyenas living in the Serengeti are summarised below.




(c) In the space provided in the grid above, draw a box and whisker plot to represent the weights of female spotted hyenas living in the Serengeti. Indicate clearly any outliers. **(A copy of this grid is on page 9 if you need to redraw your box and whisker plot.)**

**(3)**

(d) Compare the weights of male and female spotted hyenas living in the Serengeti.

**(2)**

**Question 2 continued**

Leave blank

Leave blank

**3.** The parking times, $t$ hours, for cars in a car park are summarised below.

| Time ($t$ hours) | Frequency (f) | Time midpoint ($m$) |
|---|---|---|
| $0 \leqslant t < 1$ | 10 | 0.5 |
| $1 \leqslant t < 2$ | 18 | 1.5 |
| $2 \leqslant t < 4$ | 15 | 3 |
| $4 \leqslant t < 6$ | 12 | 5 |
| $6 \leqslant t < 12$ | 5 | 9 |

(You may use $\sum fm = 182$ and $\sum fm^2 = 883$)

A histogram is drawn to represent these data.

The bar representing the time $1 \leqslant t < 2$ has a width of 1.5 cm and a height of 6 cm.

(a) Calculate the width and the height of the bar representing the time $4 \leqslant t < 6$

**(3)**

(b) Use linear interpolation to estimate the median parking time for the cars in the car park.

**(2)**

(c) Estimate the mean and the standard deviation of the parking time for the cars in the car park.

**(3)**

(d) Describe, giving a reason, the skewness of the data.

**(2)**

One of these cars is selected at random.

(e) Estimate the probability that this car is parked for more than 75 minutes.

**(3)**

**Question 3 continued**

Leave blank