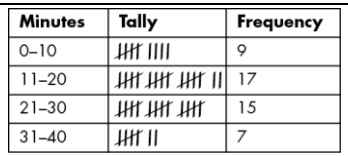
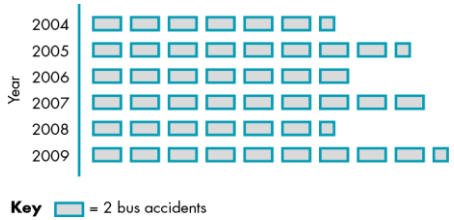
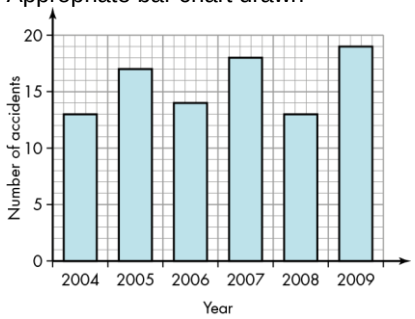
Guidance on the use of codes for this mark scheme	
M	Method mark
A	Accuracy mark
B	Mark awarded independent of method
C	Communication mark
P	Proof, process or justification mark
cao	Correct answer only
oe	Or equivalent
ft	Follow through

Question	Working	Answer	Mark	AO	Notes	Grade
1 a		Three appropriate questions using statistical measure for the data provided. e.g. What is the range of heights of the students? What is the median height of students? What is the mean average height of students?	C3	2	C1 for each question	B
b		e.g. stem and leaf The number of data points makes it difficult to represent. Data all to the nearest cm makes it easy.	C1 C1 **5**		C1 for stem and leaf oe C1 for appropriate evaluation of what makes it easy/hard to represent this data	
2		Total number of data points and how many in each category.	C1 **1**	2	C1 for correct description of two pieces of required information	B
3		 Range between 4 and 38 minutes, so choose at least four groups for data.	M1 M1 P1 **3**	2	M1 for appropriate grouping of time ranges M1 for tally chart with frequencies identified P1 for evaluation of/reasoning for choice of time ranges.	B
4		e.g. The classes overlap: where would you put a drink costing 40p or 60p? A drink could cost over £1.	C2 **2**	2 3	C1 for each correct reason	B
5		Response identifying that the pie charts represent proportions not actual numbers.	P1 **1**	2	P1 for correct reasoning and selection of 'The proportion of people buying fruit in town is greater than in the village'	B
6		Response explaining that Kevin may have data to show that crowds at the cricket ground always exceed 18 000. If there is no time when the crowds are below 18 000 then there is no point showing this part of the graph as it will be empty.	C1 C1 **2**	2	C1 for comment that data may exist to show that the minimum crowd is 18 000 C1 for comment oe	B

7	**a**		Example of a small set of data that has a mode of 6 with a valid explanation. e.g. 6, 6, 6, 5	M1	2 3	M1 for any data set with a mode of 6 and an explanation that more 6s than anything else were included	B
	b		Example of a small set of data that has a mode of 6 and a range of 15 with a valid explanation e.g. 1, 6, 6, 6, 6, 16	M1		M1 for any data set with a mode of 6 and an explanation that more 6s than anything else were included and the difference between the greatest and least value is 15	
	c		Two different small sets of data that have the same mode and range with explanation. e.g. 1, 5, 5, 5, 8 2, 5, 5, 5, 9	M1 C1		M1 for two correct data sets C1 for describing method for creating different ranges For example, shifting the greatest and least values by the same amount to keep the difference the same oe	
	d		Set with a mean of 4 and a median of 3 with explanation e.g. 1, 3, 3, 9	M1 M1 **6**		M1 for correct median M1 for data points summing to correct value (in the example, 4 data points must sum to 16 as 4 × 4 = 16)	
8	**a**	October: 16 boys and 12 girls December: 18 boys and 10 girls Both months have 28 competitors	October and December	A1	2	A1 for accurate interpretation of data from graph	B
	b	Number of boys = $\frac{2}{5}$ × 25 = 10 Number of girls = $\frac{2}{3}$ × 18 = 12	Total number taking part is 22. So this would be their lowest month for participation.	M1 M1 M1 C1 **5**		M1 for correct calculation of number of boys M1 for correct calculation of number of girls M1 for bar chart completed correctly C1 for comment explaining relatively low participation	
9	**a**		Shows the proportion not the number is greater. For example, if there were a larger number of men in the company, then their pie chart would represent greater numbers within each section of the chart.	P1	2 3	P1 for comment that proportionality shown on pie charts	B
	b		The pie charts show proportion but not the actual numbers. He needs to use proportion in the justification or he needs to provide the numbers of men and women involved in the survey.	P1 **2**		P1 for an understanding that some further clarification is needed if pie charts are used OR a suitable suggestion of an alternative such as a bar chart (which would need some numerical data) oe	

10	**a**		Pictogram drawn with key  	M1 M1	2	M1 for pictogram M1 for key	B
	b		Appropriate bar chart drawn  	M1 M1		M1 for bar chart with at least four bars correctly drawn M1 for appropriate scale on y-axis e.g. going up in 2s	
	c		Choice of either with justification such as the bar chart shows the increasing trend more clearly' or 'the pictogram is easier for readers to understand'.	P1 **5**		P1 for justification of choice	
11	**a**		Internet search, secondary data	M1	2	M1 for each correct description Data handling cycle: Plan the data collection Collect the data Choose the best way to process and represent the data. Interpret the data and draw conclusions Primary data: collected by you Secondary data: used by you but collected by a third party	B
	b		Experiment, primary data	M1			
	c		Internet search (or questionnaire), secondary (or primary) data	M1			
	d		Internet search (or questionnaire), secondary (or primary) data	M1			
	e		Questionnaire, primary data	M1 **5**			

12		Data set of 10 points plus reasons as to why this would not make a good stem-and-leaf (wide range; variance in place value) e.g. how far in kilometres ten people live from a Blackpool Tower: this could have a theoretical range of 0–1000 km (or greater), with any values in between, making it meaningless when selecting representation for the stem and the leaf. 5.6, 6.7, 2.3, 10.9, 12.6, 15.0, 15.0, 57.9, 178.2, 897.4	C1 P1 **2**	2	C1 for appropriate example where the range of data values is too varied to make a stem-and-leaf diagram have meaning P1 for appropriate reasoning	M
13 a		Statements that identify that pie charts represents proportions, e.g. What proportion or percentage of people are in the East? What is the fraction of people in the South? What is the probability that someone is from the North?	C1	2	C1 for appropriate question	M
b		Total, range and mode are difficult to use as they are numerical and the pie chart shows proportions.	P1 **2**		P1 for justification as to why numerical measures are difficult to use in this case	
14 a		The data is discrete/cannot take any values between the ones shown. It should be plotted as a bar chart.	P1 B1	2 3	P1 for correct identification of discrete data oe B1 for bar chart	M
b		Continuous data, e.g. cost of living over time.	C1 **2**		C1 for appropriate example oe	
15		e.g. a line graph is better for comparing trends over time as trends in the data are more obvious.	P1 P1 **2**	3	P1 for selection of line graph with appropriate reasoning (link to 'over time') P1 for further explanation (link to clarity of diagram) oe	M
16	5, 5, 5, 5 (mean = 5) 10, 10, 10, 10 (mean = 10)	True plus suitable explanation e.g. n, n, n, n (mean = n) $2n$, $2n$, $2n$, $2n$ (mean = $2n$)	M1 P1 **2**	2	M1 for true with specific example to back up explanation P1 for more general example to demonstrate statement is true	M

<table>
<tr>
<td>17</td>
<td>$\frac{26+1}{2}$ = 26.5th position

So 3 days

<table>
<tr><td>Days d</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>Freq f</td><td>17</td><td>2</td><td>4</td><td>13</td><td>15</td><td>1</td></tr>
<tr><td>Cumulative frequency</td><td>17</td><td>19</td><td>23</td><td>36</td><td>51</td><td>52</td></tr>
</table>
</td>
<td>3 days</td>
<td>M1

M1
A1

3</td>
<td>2</td>
<td>M1 for correct method to find median value (tabulated calculations) oe
M1 for correct calculation to find position of median oe
A1 cao</td>
<td>M</td>
</tr>
<tr>
<td>18</td>
<td colspan="2">Estimated mean = total profit ÷ total frequency
= 15 418.5 ÷ 52
= 296.5096

<table>
<tr><td>Profit, p £</td><td>0-200</td><td>201-400</td><td>401-600</td><td>601-800</td><td></td></tr>
<tr><td>Freq,f</td><td>15</td><td>26</td><td>8</td><td>3</td><td>52</td></tr>
<tr><td>Midpoint, m</td><td>100</td><td>300.5</td><td>500.5</td><td>700.5</td><td></td></tr>
<tr><td>$F \times m$</td><td>1500</td><td>7813</td><td>4004</td><td>2101.5</td><td>15 418.5</td></tr>
</table>
Answer: £297</td>
<td>M1

M1

A1
3</td>
<td>2</td>
<td>M1 for correct method to find the estimated mean (tabulated calculations) oe
M1 for correct final calculation with appropriate rounding

A1 cao</td>
<td>M</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>Need to design a questionnaire and ask students what they have for breakfast.</td>
<td>C1
1</td>
<td>2</td>
<td>C1 for descriptive comment about the data needed and/or method of collection oe</td>
<td>M</td>
</tr>
<tr>
<td>20 a

b</td>
<td>Start with 3, 3, 3 and adjust at both ends.

Start with 3 in the centre and find a pair for the ends which sum to 6 with a difference of 3.</td>
<td>Any three values that sum to 9 with a difference between greatest and least of 3, e.g. 2, 2, 5

Any three values with a sum of 9, a difference between greatest and least of 3 and a median of 3, e.g. 1.5, 3 , 4.5</td>
<td>M1
A1

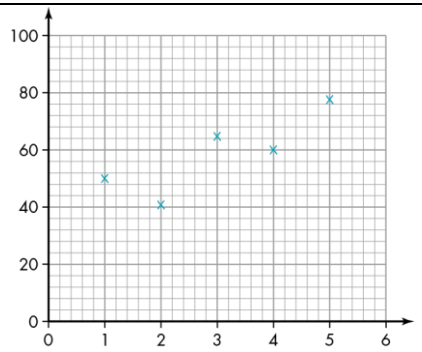
M1
A1

4</td>
<td>2
3</td>
<td>M1 for either correct range or mean
A1 cao

M1 for either correct range or mean
A1 cao
Note: this can be done by solving a pair of linear equations
a, 3, b such that $b + a = 6$ and $b - a = 3$; $b = 4.5$ and $a = 1.5$</td>
<td>M</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>They could have been referring to two different averages.
For example: in a class of 10 students with the following marks
10, 17, 28, 28, 28, 28, 28, 35, 57, 61
The median and modal mark is 28 but the mean mark is 32.</td>
<td>C1

P1

2</td>
<td>3</td>
<td>C1 for correct comment regarding two different averages
P1 for example to justify</td>
<td>M</td>
</tr>
</table>

<table>
<tr>
<td>22</td>
<td> 
</td>
<td>A scatter graph with positive correlation shows a trend with a positive gradient/ as one quantity increases the other also increases.</td>
<td>C1
C1

2</td>
<td>2</td>
<td>C1 for diagram to support explanation
C1 for comment oe</td>
<td></td>
</tr>
<tr>
<td>23 a

b

c</td>
<td>Total number of male students = 61
11 ÷ 61 = 0.18 (to 2 dp)

Total number of female students = 31
4 ÷ 31 = 0.13 (to 2 dp)</td>
<td>18% to nearest percent

13% to nearest percent

Find the mid-point of each class
Multiply mid-point of class by frequency of class and add them all together.
Divide the total earnings by the total frequency.</td>
<td>A1

A1

C1
C1

C1
5</td>
<td>2
3

</td>
<td>A1 cao

A1 cao

C1 for explanation
C1 for explanation

C1 for explanation (no credit for actual calculation as an explanation as this is specified in the question)</td>
<td>M</td>
</tr>
<tr>
<td>24</td>
<td>Mean for boys =
204 ÷ 34 = 6
<table>
<tr><td>Texts, T</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td><td></td></tr>
<tr><td>f</td><td>5</td><td>6</td><td>4</td><td>5</td><td>4</td><td>5</td><td>3</td><td>2</td><td>34</td></tr>
<tr><td>T × f</td><td>15</td><td>24</td><td>20</td><td>30</td><td>28</td><td>40</td><td>27</td><td>20</td><td>204</td></tr>
</table></td>
<td>Mean for boys is 6, so based on the information provided and this particular statistical measure then the answer would be no. However, the information is limited for example the sample of boys is small and you are not told how many girls are sampled. The mean could be based on a single girl.</td>
<td>M1

P1
2</td>
<td>2
3

</td>
<td>M1 for calculation of mean number of boys text messages for comparison

P1 for appreciation of sample size oe</td>
<td>M</td>
</tr>
<tr>
<td>25 a

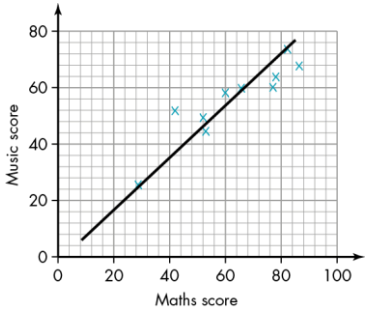
b</td>
<td></td>
<td>Demonstrate the use of data to produce a time series graph to predict the life expectancy in both countries in 2025.
UK: 82–83, Ukraine: anything between 65 and 72

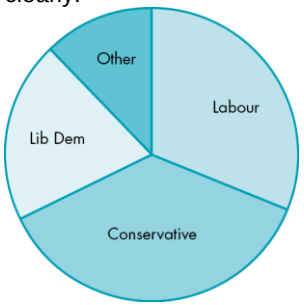
UK may be easier to predict as historically it has been more consistent.</td>
<td>M1

P1
2</td>
<td>2
3

</td>
<td>M1 for data interpretation either by time-series graph or by spreadsheet

P1 for understanding of consistency making a trend easier to identify</td>
<td>M</td>
</tr>
</table>

26 **a/b**		Scatter diagram completed with line of best fit  	M1 M1	3	M1 for correct points plotted on scatter diagram M1 for line of best fit	M
c		Ben is most likely to be the student who was sick when they took maths as his score is much lower than his music score.	C1		C1 for choice of student with valid reason oe	
d		Kris is likely to have scored around 40 marks in music as all students except Ben scored similar but less in music than in maths (see trend line)	P1		P1 for score for music based on their trend line (ft)	
e		Lex could have scored 85 in maths with a 78 score in music (see trend line)	P1		P1 for score for maths based on their trend line (ft)	
			5			

<table>
<tr>
<td>27</td>
<td>Labour:
$\frac{32 + 32 + 30 + 30}{4} = 31$
Conservative:
$\frac{37 + 37 + 35 + 37}{4} = 37$
Lib Dem:
$\frac{19 + 21 + 21 + 20}{4} = 20$
Other:
$\frac{12 + 9 + 14 + 13}{4} = 12$
Angles:
Labour: $\frac{31}{100} \times 360° = 112°$
Conservative: $\frac{37}{100} \times 360° = 133°$
Lib Dem: $\frac{20}{100} \times 360° = 72°$
Other: $\frac{12}{100} \times 360° = 43°$</td>
<td>Pie chart constructed accurately and clearly.
 
</td>
<td>M1
M1
M1
M1
4</td>
<td></td>
<td>M1 for calculation of average, rounded to the nearest percentage
M1 for correct degree calculations
M1 for accurate measurement of sectors
M1 for correct labelling of sectors</td>
<td>M</td>
</tr>
<tr>
<td>28</td>
<td></td>
<td>Suitable example
Algebraic explanation:
$\frac{x_1 + x_2 + x_3 + x_4}{4} = 2 \times \text{mode}$
$x_1 + x_2 + x_3 + x_4 = 8 \times \text{mode}$
e.g., if the mode is 2 the sum of the data would have to be 16 so that the mean was 4. One possible set could be: 2, 2, 2, 10. Mode is 2, mean is
$\frac{2 + 2 + 2 + 10}{4} = 16 \div 4 = 4$ .Mean is twice mode, as required.</td>
<td>C1
P1
2</td>
<td>2</td>
<td>C1 for specific example with suitable explanation.
P1 for general solution (could be presented algebraically or in words)</td>
<td>M</td>
</tr>
<tr>
<td>29</td>
<td>
<table>
<tr><th></th><th>Mean</th><th>Median</th><th>Mode</th><th>Range</th></tr>
<tr><td>Girls' scores %</td><td>52.5</td><td>50.5</td><td>24, 47</td><td>71</td></tr>
<tr><td>Boys' scores %</td><td>63</td><td>66</td><td>82</td><td>74</td></tr>
</table>
</td>
<td>Comparison of different statistical measures that provide counterargument.
e.g. Boys have a higher mean, median and modal score.
The range is very similar.</td>
<td>M1
M1
P1
3</td>
<td>2</td>
<td>M1 for calculation of at least two averages or one average and the range to compare
M1 for calculation of further averages to support argument in favour of boys
P1 for evaluation and interpretation of averages to provide counter argument for boys</td>
<td>M</td>
</tr>
</table>

30 a	<table><tr><th></th><th>Median</th><th>Mode</th><th>Range</th><th>Max</th><th>Min</th></tr><tr><td>Male</td><td>37m 20s</td><td>None</td><td>05 m 04 s</td><td>38m 42s</td><td>33 m 38 s</td></tr><tr><td>Female</td><td>42m 46s</td><td>39m 08 s</td><td>07 m 57 s</td><td>46m 37s</td><td>38m 40s</td></tr></table>	Valid comparison of the distributions for male and female runners. e.g. all men except one ran faster than the women (max and min values) Or there was less variance in men's running times than women's (range) e.g. median values	C1 C1	2 3	C1 for first comparison C1 for further comparison such as the median values	M
b		Valid couple of sentences to summarise findings suitable for a national running magazine.	C1 C1 **4**		C1 for each of two summary sentences based on their interpretations of the distributions (ft)	
31 a	112 ÷ 200 × 50 000	28 000	M1	2	M1 for multiplication of total population by proportion of sample	M
b	28 000 × 0.9 = 25 200	Over 50% of voters so should be confident of winning.	M1 P1 **3**		M1 for 28 000 × 0.9 P1 for explanation that his represents over half the whole population and so a majority is likely	
32	7 ÷ 28 × 1260 = 315	315	M1 A1 **2**	2	M1 for multiplication of total population by proportion of sample A1 cao	M
33		At four different points in the concert, entrances or bars, each point to conduct the survey on approximately 100 people each, trying to ask as many males as females.	P1 C1 **2**	2 3	P1 for explanation of general method of sampling C1 for specific example related to the concert crowd	H

<table>
<tr><td>34 a</td><td></td><td>Define what is meant by a ‘good’ train service. Could be number of trains per hour or number of trains that are on time in a specified period or customer satisfaction regarding the quality of various aspects of the journey such as level of cleanliness, availability of sets etc.
To find out if the statement is true or not, the first step is to see what has changed since the aunt has been travelling (see above)</td><td>P1
C1</td><td>3</td><td>P1 for explanation of subjective nature of this enquiry
C1 for description of what might be seen as ‘good’</td><td>H</td></tr>
<tr><td>b</td><td></td><td>Data to be collected would be dependent on what the aunt felt had changed (or what has actually changed).This could be data from the train company on volume of trains, amount of rolling stock at one time, number of complaints received and about what.
This could also be data collected from a customer survey. However, this would be real time and would be hard to gather retrospectively as this would be subjective and may not be possible, depending on time frame.</td><td>P1
C1
C1
5</td><td></td><td>P1 for explanation of possible data to be collected with reason
C1 for appreciation of this being a subjective enquiry over time
C1 for a correct description of a data collection process, such as constructing a questionnaire</td><td></td></tr>
<tr><td>35</td><td>2.5 × 20 = 50
So 50 eggs in total

0 × 2 = 0
1 × 3 = 3
2 × 4 = 8
5 × 1 = 5
0 + 3 + 8 + 5 = 16 eggs
50 – 16 = 34 eggs to find
Hens could lay 3 or 4 eggs.
(3 × 6) + (4 × 4) = 18 + 16 = 34
(3 × 2) + (4 × 7) = 6 + 28 = 34</td><td><table><tr><td>Eggs</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>Frequency</td><td>2</td><td>3</td><td>4</td><td>6</td><td>4</td><td>1</td></tr><tr><td>Egg × frequency</td><td>0</td><td>3</td><td>8</td><td>18</td><td>16</td><td>5</td></tr></table>
Could be 3 eggs, frequency 6 and 4 eggs, frequency 4
Or could be 3 eggs, frequency 2 and 4 eggs, frequency 7</td><td>M1
M1

M1
M1
C1
5</td><td>2
3</td><td>M1 for total eggs (50)
M1 for multiplication of eggs by frequency plus addition and subtraction to find how many eggs are left to find oe
M1 for one solution
M1 for a second solution
C1 for clear description or table oe</td><td>H</td></tr>
</table>

<table>
<tr>
<td>36 a</td>
<td>1 + 12 + 24 + 15 + 13 + 9 + 5 = 79</td>
<td>79 patients</td>
<td>M1</td>
<td>3</td>
<td>M1 for addition of frequencies</td>
<td>H</td>
</tr>
<tr>
<td>b</td>
<td>$\frac{2754}{79}$ = 34.86…..</td>
<td>
<table>
<tr><td>Time</td><td>0–10</td><td>11–20</td><td>21–30</td><td>31–40</td></tr>
<tr><td>Mid point</td><td>5</td><td>15.5</td><td>25.5</td><td>35.5</td></tr>
<tr><td>Frequency</td><td>1</td><td>12</td><td>24</td><td>15</td></tr>
<tr><td>Mid-point × frequency</td><td>5</td><td>186</td><td>612</td><td>532.5</td></tr>
</table>
<table>
<tr><td>Time</td><td>41–50</td><td>51–60</td><td>61–70</td></tr>
<tr><td>Mid point</td><td>45.5</td><td>55.5</td><td>65.5</td></tr>
<tr><td>Frequency</td><td>13</td><td>9</td><td>5</td></tr>
<tr><td>Mid-point × frequency</td><td>591.5</td><td>499.5</td><td>327.5</td></tr>
</table>
Sum of mid-point × frequencies = total number of minutes = 2754
Estimate mean waiting time = 35 minutes</td>
<td>M1
M1
M1
C1</td>
<td></td>
<td>M1 for mid-points × frequencies
M1 for total minutes
M1 for division of total minutes by total patients (ft)
C1 for appropriate rounding to whole minutes</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>Estimated mean = 35 minutes
Modal time = 21–30 minutes
Median time: 31–40 minutes
$\frac{79+1}{2}$ = 40
40th person is the median. 40th person is in the 31–40 group</td>
<td>Mode</td>
<td>M1
M1
P1</td>
<td></td>
<td>M1 for mode
M1 for median
P1 for correct , shortest, average waiting time</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>Over one hour (61–70) is 5 patients
$\frac{5}{79}$ =0.00632….
So 6% of patients were over an hour
100 – 6 = 94%</td>
<td>94%</td>
<td>M1
A1
10</td>
<td></td>
<td>M1 for calculation of 1 – ($\frac{5}{79}$) or $\frac{74}{79}$ as a percentage
A1 cao</td>
<td></td>
</tr>
</table>

<table>
<tr><td>37 a

b</td><td>Estimate:
Often = 90° (=25%)
Very often = 80° (≈20%)
Rarely = 100° (≈30%)
Never = 80° (≈20%)
Always= 10° (≈5%)</td><td>No, it only shows proportions

e.g. All percentages are estimates based on the approximate angle of the sector of the pie chart.
25% of people sampled said that they often considered their health when planning a diet, with 5% saying that this was always the case and 20% saying very often.
30% rarely consider their health when planning a diet and 20% never do.
e.g. So approximately half the sample consider their health to some degree and half do not.</td><td>P1

C1

C2

3</td><td>3</td><td>P1 for reference to proportions

C1 for comment on the proportions of people giving the range of answers

C2 for summary</td><td>H</td></tr>
<tr><td>38</td><td></td><td>The 5 minutes spent waiting is halfway between the 4- and 6-minute groups. These people are in that band, but maybe no one had to wait exactly 5 minutes.</td><td>P1

1</td><td>3</td><td>P1 for awareness that the 5 minutes represents the mid-point of grouped data. People may have been waiting for any time between 4 and 6 minutes</td><td>H</td></tr>
<tr><td>39</td><td></td><td>Divide the frequency of the class interval by the width of the class interval.</td><td>C1
1</td><td>3</td><td>C1 for correct method for calculations of frequency density</td><td>H</td></tr>
<tr><td>40</td><td>Kathy’s mean
= (8 + 7 + 6 + 5 + 5 + 5 + 4) ÷ 9 = 5.7
Connie’s mean
= (9 + 8 + 7 + 2 + 2) ÷ 5 = 5.6
Evie’s mean
= (8 + 8 + 3 + 2 + 1) ÷ 5 = 4.4
<table><tr><th>Dancer</th><th>Scores</th><th>Mean</th><th>Median</th><th>Mode</th></tr><tr><td>Kathy</td><td>8, 7, 6, 5, 5, 5, 4</td><td>5.7</td><td>5</td><td>5</td></tr><tr><td>Connie</td><td>9, 8, 7, 2, 2</td><td>5.6</td><td>7</td><td>2</td></tr><tr><td>Evie</td><td>8, 8, 3, 2, 1</td><td>4.4</td><td>3</td><td>8</td></tr></table></td><td>Kathy: mean score of 5.7
Connie: median score of 7
Evie: modal score of 8</td><td>M1
M1
M1
C1

4</td><td>2</td><td>M1 for calculation of all three means
M1 for identification of all three modes
M1 for identification of all three medians
C1 for correct average identified in all three cases</td><td>H</td></tr>
</table>

<table>
<tr>
<td>41</td>
<td>$\text{Range}_A = 86 - 18 = 70$
$\text{Range}_B = 45 - 22 = 23$

$\text{Mean}_A = \frac{86 + 62 + 23 + 23 + (5 \times 18)}{9} =$
31.555…
$\text{Mean}_B = \frac{(45 + 36 + (4 \times 36) + (3 \times 22)}{9} =$
27.888…

<table>
<tr><th>Firm</th><th>Range</th><th>Mean</th><th>Median</th><th>Mode</th></tr>
<tr><td>A</td><td>£70 000</td><td>£31 556</td><td>£18 000</td><td>£18 000</td></tr>
<tr><td>B</td><td>£23 000</td><td>£27 889</td><td>£26 000</td><td>£26 000</td></tr>
</table>
</td>
<td>Valid advice based on the salary scales provided (see notes)
e.g. Jasmin should join firm B if she is prepared to accept that the higher final salary is not as high as A. However, the starting salary is higher and Jasmin can progress more quickly to a higher salary with B than A. There is a greater range of salaries in firm A than firm B
Jasmin should join firm A if she aspires to being the boss.

Alternatively she could start with firm B and move to firm A after 7 salary increments for maximum earnings.</td>
<td>M1
M1
M1
M1
P1
C1

6</td>
<td>2</td>
<td>M1 for calculation of ranges
M1 for calculation of means
M1 for identification of medians
M1 for identification of modes
P1 for justification of either firm A or firm B
C1 for use of table oe for comparison</td>
<td>H</td>
</tr>
</table>

42 a e.g. $0 \leq x < 5$; $5 \leq x < 25$; $25 \leq x < 32$
Estimated range (from mid-points) = 28.5–2.5

M1 | 2 3 | M1 for correct example (class widths do not need to be equal) | H

b e.g.

Class	$0 \leq x < 45$	$45 \leq x < 47$	$47 \leq x < 50$
Mid-point		46	
Frequency	5	1	5
Cumulative frequency	5	6	11

M1 | | M1 for correct example (class widths do not need to be equal)

n = 11 hence 6th data point is the median
Estimated median (from mid-point) = 46

c e.g.

Class	$0 \leq x < 20$	$20 \leq x < 25$	$25 \leq x < 119$
Mid-point	10	22.5	72
Frequency	5	1	5
Cumulative frequency	5	6	11

M1 | | M1 for correct example of estimated median
M1 | | M1 for correct example of estimated range(class widths do not need to be equal)

n = 11 hence 6th data point is the median
Estimated median (from mid-point) = 22.5
Estimated range (from mid-points) = 72–10 = 62

d e.g.

Class	Frequency	Mid-point	Frequency × mid-point
$0 \leq x < 6$	0	3	0
$6 \leq x < 10$	2	8	16
$10 \leq x < 20$	5	15	75
$20 \leq x < 30$	5	25	125
$30 \leq x < 40$	9	35	315
$40 \leq x < 62$	15	51	765
Totals	36		1296

M1 | | M1 for correct division: (sum of frequencies × mid-point) ÷ (total frequency)
M1 | | M1 for correct example of estimated mean (ft) (class widths do not need to be equal)

Estimated mean = 1296 ÷ 36 = 36 (to one decimal place)

6

<table>
<tr>
<td>43</td>
<td colspan="2">Median = $\frac{n+1}{2}$ value Lower quartile = $\frac{n+1}{4}$ value Upper quartile = $\frac{3(n+1)}{4}$ value
So if you have a table with 12 values these will be the 6th, 3rd and 9th values.
<table>
<tr><td>Order</td><td>Value</td><td rowspan="3"></td></tr>
<tr><td>1</td><td></td></tr>
<tr><td>2</td><td></td></tr>
<tr><td>3</td><td>7</td><td rowspan="7">Difference between 14 and 7 equals 7 as required.</td></tr>
<tr><td>4</td><td></td></tr>
<tr><td>5</td><td></td></tr>
<tr><td>6</td><td>10</td></tr>
<tr><td>7</td><td></td></tr>
<tr><td>8</td><td></td></tr>
<tr><td>9</td><td>14</td></tr>
<tr><td>10</td><td></td><td rowspan="2"></td></tr>
<tr><td>11</td><td></td></tr>
</table>
These can be adapted as long as rows are added equally in each section (see example below)
<table>
<tr><td>Order</td><td>Value</td><td rowspan="3"></td></tr>
<tr><td>1</td><td></td></tr>
<tr><td>2</td><td></td></tr>
<tr><td>3</td><td></td><td></td></tr>
<tr><td>4</td><td>7</td><td rowspan="9">Difference between 14 and 7 equals 7 as required.</td></tr>
<tr><td>5</td><td></td></tr>
<tr><td>6</td><td></td></tr>
<tr><td>7</td><td></td></tr>
<tr><td>8</td><td>10</td></tr>
<tr><td>9</td><td></td></tr>
<tr><td>10</td><td></td></tr>
<tr><td>11</td><td></td></tr>
<tr><td>12</td><td>14</td></tr>
<tr><td>13</td><td></td><td rowspan="3"></td></tr>
<tr><td>14</td><td></td></tr>
<tr><td>15</td><td></td></tr>
</table>
</td>
<td>C1
C1
C1

M1
P1

5</td>
<td>2</td>
<td>C1 for correct formula for median
C1 for correct formula for lower quartile
C1 for correct formula for upper quartile

M1, for values for upper and lower quartile with a difference of 7
P1 for explanation of difference in relation to range in this context</td>
<td>H</td>
</tr>
<tr>
<td>44</td>
<td></td>
<td>Either Harold, as he had bigger tomatoes, or Connie, as she had more tomatoes (depending on if you want lots of tomatoes or large tomatoes!)</td>
<td>P1
C1

2</td>
<td>2</td>
<td>P1 for choice of one person with justification
C1 for further consideration of second person</td>
<td>H</td>
</tr>
</table>

45 a		The following statement is more precise: 'There is strong evidence to support my hypothesis'. In reality it is usually not possible to prove a hypothesis just to gather evidence to support it.	P1	2	P1 for selection of 'There is strong evidence to support my hypothesis' with explanation about practicalities of proving an hypothesis	H
b i		Because you may not have gathered enough evidence.	P1		P1 for explanation of having sufficiently large sample/evidence	
ii		Because you haven't gathered enough evidence or you're using a 'null' hypothesis.	P1		P1 for giving same reason / an awareness of the practical difficulties of getting all other variables equal or unchanged so that they do not affect the success or otherwise of your evidence gathering oe	
			3			