Guidance on the use of codes for this mark scheme	
M	Method mark
A	Accuracy mark
B	Working mark
cao	Correct answer only
oe	Or equivalent
ft	Follow through

Question	Working	Answer	Mark	AO	Notes	Grade
1 a	$\frac{112}{200} \times 50\,000$	28 000 would vote for Party A.	M1	2	M1 for multiplication of total population by proportion of sample	B
b	28 000 × 0.9 = 25 200	This is over 50% of voters so Party A should be confident of winning.	M1 A1 **3**		M1 for 28 000 × 0.9 A1 for explanation that this represents over half the whole population and so a majority is likely	
2	$\frac{7}{28} \times 1260 = 315$	315 students are in the sample.	M1 A1 **2**	2	M1 for multiplication of total population by proportion of sample A1 cao	B
3		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.	B1 B1 **2**	2 3	B1 for explanation of general method of sampling B1 for specific example related to the concert crowd	B

<table>
<tr><td>4 a

b</td><td></td><td>Define what she means 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).

Data to be collected would depend on what the aunt felt had changed (or what has actually changed).This could be data from the train company about 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 it would be subjective and may not be possible, depending on time frame.</td><td>B1
B1

B1

B1

B1

5</td><td>3</td><td>B1 for explanation of subjective nature of this enquiry
B1 for description of what might be seen as ‘good’

B1 for explanation of possible data to be collected with reason
B1 for appreciation of this being a subjective enquiry over time
B1 for a correct description of a data-collection process</td><td>B</td></tr>
<tr><td>5</td><td>2.5 × 20 = 50
So there were 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 (E)</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>Frequency (f)</td><td>2</td><td>3</td><td>4</td><td>6</td><td>4</td><td>1</td></tr><tr><td>$E \times f$</td><td>0</td><td>3</td><td>8</td><td>18</td><td>16</td><td>5</td></tr></table>
<table><tr><td>Eggs (E)</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr><tr><td>Frequency (f)</td><td>2</td><td>3</td><td>4</td><td>2</td><td>7</td><td>1</td></tr><tr><td>$E \times f$</td><td>0</td><td>3</td><td>8</td><td>6</td><td>28</td><td>5</td></tr></table>
Could be 3 eggs, frequency 6 and 4 eggs, frequency 4.

Could be 3 eggs, frequency 2 and 4 eggs, frequency 7.</td><td>M1
M1

M1

M1
A1
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
A1 for clear description or table oe</td><td>B</td></tr>
</table>

6 a	1 + 12 + 24 + 15 + 13 + 9 + 5 = 79	79 patients	M1	3.1	M1 for addition of frequencies	B
b	Total number of minutes = sum of (midpoint × frequency) for each class = 2754 Mean time = $\frac{2754}{79}$ = 34.86… minutes	<table><tr><td>**Time (minutes)**</td><td>0–10</td><td>11–20</td><td>21–30</td><td>31–40</td></tr><tr><td>**Midpoint (M)**</td><td>5</td><td>15.5</td><td>25.5</td><td>35.5</td></tr><tr><td>**Frequency (f)**</td><td>1</td><td>12</td><td>24</td><td>15</td></tr><tr><td>$M \times f$</td><td>5</td><td>186</td><td>612</td><td>532.5</td></tr></table>	M1 M1 M1		M1 for midpoints × frequencies M1 for total minutes M1 for division of total minutes by total patients (ft)	
c	Estimated mean = 35 minutes	<table><tr><td>**Time (minutes)**</td><td>41–50</td><td>51–60</td><td>61–70</td></tr><tr><td>**Midpoint (M)**</td><td>45.5</td><td>55.5</td><td>65.5</td></tr><tr><td>**Frequency (f)**</td><td>13</td><td>9</td><td>5</td></tr><tr><td>$M \times f$</td><td>591.5</td><td>499.5</td><td>327.5</td></tr></table> Estimate mean waiting time = 35 minutes	B1		B1 for appropriate rounding to whole minutes	
d	Modal time = 21–30 minutes $\frac{(79+1)}{2}$ = 40 The median is the 40th person. The 40th person is in the 31–40 group. Median time is 31–40 minutes For the group who waited over one hour (61–70), there were 5 patients. So 74 patients were seen in one hour or less. $\frac{74}{79}$ = 0.9367…. So 93.7% of patients saw a doctor within one hour.	The hospital would use the modal time. 93.7% of patients saw a doctor within one hour.	M1 M1 A1 M1 A1		M1 for mode M1 for median A1 for correct, shortest, average waiting time M1 for calculation of $\frac{74}{79}$ or 1 – ($\frac{5}{79}$) as a percentage A1 cao	
			10			

<table>
<tr><td>7 a

b</td><td>Estimate:
Often: 90° (= 25%)
Very often: 75° (≈ 20%)
Rarely: 100° (≈ 30%)
Never: 75° (≈ 20%)
Always: 10° (≈ 5%)</td><td>No. It only shows proportions

Summary and interpretation of available information and data</td><td>M1

A1

B2

4</td><td>3</td><td>M1 for reference to proportions

A1 for commenting on the proportions of people giving the range of answers, for example: 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.
B2 for summary, for example: So approximately half the sample consider their health to some degree and half do not.</td><td>B</td></tr>
<tr><td>8</td><td></td><td>The 5 minutes spent waiting is halfway between the 4-minute and the 6-minute groups. These people are in that band, but maybe no one had to wait exactly 5 minutes.</td><td>M1

1</td><td>3</td><td>M1 for awareness that the 5 minutes represents the midpoint of a class of grouped data, so people may have been waiting for any time between 4 and 6 minutes</td><td>B</td></tr>
<tr><td>9</td><td></td><td>Divide the frequency of the class interval by the width of the class interval to find the frequency density.</td><td>M1

1</td><td>3</td><td>M1 for correct method for calculation of frequency density</td><td>B</td></tr>
<tr><td>10</td><td>Kathy’s mean (scores in order):
(8 + 7 + 6 + 5 + 5 + 5 + 4) ÷ 7 = 5.7
Connie’s mean (scores in order):
(9 + 8 + 7 + 2 + 2) ÷ 5 = 5.6
Evie’s mean (scores in order):
(8 + 8 + 3 + 2 + 1) ÷ 5 = 4.4</td><td>Kathy would choose her mean score of 5.7.

Connie would choose her median score of 7.

Evie would choose her modal score of 8.

<table><tr><td>Dancer</td><td>Scores</td><td>Mean</td><td>Median</td><td>Mode</td></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>M1
M1
M1
A1

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
A1 for correct average identified in all three cases</td><td>M</td></tr>
</table>

11	Working in thousands of pounds: $\text{Range}_A = 86 - 18 = 70$ $\text{Range}_B = 45 - 22 = 23$ $\text{Mean}_A = \frac{(86 + 62 + 23 + 23 + (5 \times 18))}{9}$ $= 31.555\ldots$ $\text{Mean}_B = \frac{(45 + 36 + (4 \times 36) + (3 \times 22))}{9}$ $= 27.888\ldots..$	Valid advice based on the salary scales provided (see notes), for example: Jasmin should join firm B if she is prepared to accept that the highest salary is not as high as that offered by firm A. However, the starting salary is higher and Jasmin can progress more quickly to a higher salary with firm B than with firm A. There is a greater range of salaries in firm A than in 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.	M1 M1 M1 M1 M1 A1 **6**	2	**Firm** | Range | Mean | Median | Mode **A** | £70 000 | £31 556 | £18 000 | £18 000 **B** | £23 000 | £27 889 | £26 000 | £26 000 M1 for calculation of ranges M1 for calculation of means M1 for identification of medians M1 for identification of modes M1 for justification of either firm A or firm B A1 for use of table oe for comparison	M

12 a

Set of grouped data with an estimated range of 26.

B1

2
3

B1 for correct example (class widths do not need to be equal), e.g. $0 \leq x < 5$; $5 \leq x < 25$; $25 \leq x < 32$
Estimated range (from midpoints) = 28.5 – 2.5

M

b

Set of grouped data an estimated median of 46.

B1

B1 for correct example (class widths do not need to be equal), e.g.

Class	$0 \leq x < 45$	$45 \leq x < 47$	$47 \leq x < 50$
Midpoint		46	
Frequency	5	1	5
cf	5	6	11

n = 11 hence the sixth data point is the median
Estimated median (from midpoint) = 46

c

Set of grouped data an estimated median of 22.5 and an estimated range of 62

B1
B1

B1 for correct example of estimated median
B1 for correct example of estimated range (class widths do not need to be equal), e.g.

Class	$0 \leq x < 20$	$20 \leq x < 25$	$25 \leq x < 119$
Midpoint	10	22.5	72
Frequency	5	1	5
cf	5	6	11

n =11 hence the sixth data point is the median
Estimated median (from midpoint) = 22.5
Estimated range (from midpoints) = 72 – 10 = 62

d

Set of grouped data an estimated mean of 36 (to one decimal place).

B1

B1

B1 for correct division:
(sum of frequencies × midpoint) ÷ (total frequency)
B1 for correct example of estimated mean (ft)
(class widths do not need to be equal), e.g.

Class	Frequency (f)	Midpoint (m)	$f \times m$
$0 \leq x < 6$	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

Estimated mean= 1296 ÷ 36

6

<table>
<tr>
<td>13</td>
<td></td>
<td>
Median = $\frac{n+1}{2}$ th value

Lower quartile = $\frac{n+1}{4}$ th value

Upper quartile = $\frac{3(n+1)}{2}$ th 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="2"></td></tr>
<tr><td>1</td><td></td></tr>
<tr><td>2</td><td></td><td rowspan="8">Difference between 14 and 7 equals 7 as required.</td></tr>
<tr><td>3</td><td>7</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="2"></td></tr>
<tr><td>1</td><td></td></tr>
<tr><td>2</td><td></td><td rowspan="2"></td></tr>
<tr><td>3</td><td></td></tr>
<tr><td>4</td><td>7</td><td rowspan="8">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>
</table>
</td>
<td>B1

B1

B1

M1
A1</td>
<td>2</td>
<td>B1 for correct formula for median

B1 for correct formula for LQ

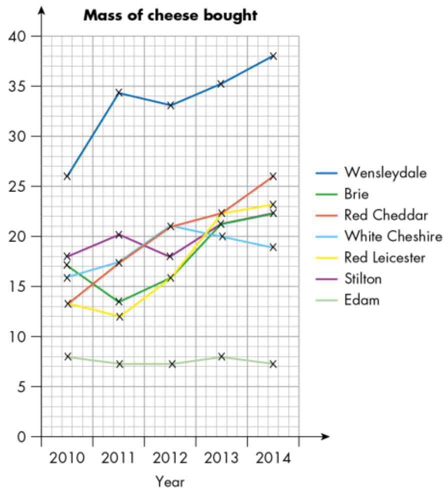
B1 for correct formula for UQ

M1, for values for UQ and LQ with a difference of 7
A1 for explanation of difference in relation to range in this context</td>
<td>M</td>
</tr>
</table>

		<table><tr><td>11</td><td></td><td rowspan="2"></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>	**5**			
14		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!).	M1 A1 **2**	2	M1 for choice of one person with justification A1 for further consideration of second person	M
15 a **bi** **ii**		The statement: 'There is strong evidence to support my hypothesis' is more precise. In reality it is usually not possible to prove a hypothesis just to gather evidence to support it. Because you may not have gathered enough evidence. Because you haven't gathered enough evidence or you're using a 'null' hypothesis.	M1 A1 A1 **3**	2	M1 for selection of 'There is strong evidence to support my hypothesis' with explanation about practicalities of proving an hypothesis A1 for explanation of having sufficiently large sample / evidence A1 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	M
16 a **b** **c**		Valid comparative statement such as: 'The means are very similar but there is evidence that the broadsheets use a broader range of length of words.' No, it provides some evidence, but only assuming length of words is a suitable measure of difficulty. The length of words may not be the thing that readers find makes a paper hard to read. For example: How long did it take to read the sample or a specific question to challenge the readers' understanding of the sample.' Could also look at the modal word length to see if there is evidence to support the	B1 B1 B1		B1 for comparative statement oe B1 for explanation that this is insufficient evidence to answer the question oe B1 for a suitable further question oe	H

<table>
<tr><td></td><td></td><td>broadsheet using a greater range of words, for example: Are most of the words in the tabloid four letters long?</td><td>3</td><td></td><td></td><td></td></tr>
<tr><td>17</td><td>Mean$_1$ = $\frac{455}{40}$
= 11.375
There is a difference of 0.625 so mean$_2$ is 11.375 ± 0.625.
So it is either 12 or 10.75.
The new total is either:
40 × 12 = 480 (increase of 25)
or
40 × 10.75 = 430 (decrease of 25)
By trial and improvement:
15 and 10 are reversed.</td><td>The 15 and the 10 are the wrong way round.</td><td>M1
M1
M1

A1
A1
5</td><td>3</td><td><table>
<tr><td>Mass</td><td>$0 < m \leq 5$</td><td>$5 < m \leq 10$</td><td>$10 < m \leq 15$</td><td>$15 < m \leq 20$</td><td>$20 < m \leq 25$</td><td>Total</td></tr>
<tr><td>freq</td><td>4</td><td>15</td><td>10</td><td>8</td><td>3</td><td>40</td></tr>
<tr><td>mp</td><td>2.5</td><td>7.5</td><td>12.5</td><td>17.5</td><td>22.5</td><td></td></tr>
<tr><td>$m \times f$</td><td>10</td><td>112.5</td><td>125</td><td>140</td><td>67.5</td><td>455</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>freq</td><td>4</td><td>10</td><td>15</td><td>8</td><td>3</td><td>40</td></tr>
<tr><td>$m \times f$</td><td>10</td><td>75</td><td>187.5</td><td>140</td><td>67.5</td><td>480</td></tr>
</table>
M1 for calculation of new mean
M1 for calculation of new totals
M1 for trial and improvement with information gathered oe
A1 for correct identification of 15 and 10
A1 for table to show corrected data oe</td><td>H</td></tr>
</table>

<table>
<tr>
<td>18 a</td>
<td></td>
<td>The diagram chosen needs to show how the mass of each cheese varies from one year to the next. A line graph, as follows, is a good choice when there is variation over time.
<table>
<tr><th></th><th>Mean</th><th>Median</th><th>Mode</th><th>Range</th></tr>
<tr><td>W</td><td>33.2</td><td>34</td><td>-</td><td>12</td></tr>
<tr><td>B</td><td>17.5</td><td>17</td><td>-</td><td>9</td></tr>
<tr><td>RC</td><td>19.8</td><td>21</td><td>-</td><td>13</td></tr>
<tr><td>WC</td><td>18.6</td><td>19</td><td>-</td><td>5</td></tr>
<tr><td>RL</td><td>17.2</td><td>16</td><td>-</td><td>11</td></tr>
<tr><td>S</td><td>19.8</td><td>20</td><td>18</td><td>4</td></tr>
<tr><td>E</td><td>7.4</td><td>7</td><td>7</td><td>1</td></tr>
</table>


</td>
<td>M1

A1

M1

M1
A1

5</td>
<td>3</td>
<td>M1 for report that includes an awareness of change over time, and which cheeses are increasing and which are decreasing in popularity
A1 for appropriate diagram or graph oe
M1 for calculation of ranges to show which cheeses are most popular oe
M1 for calculation of mean or median to show average mass of cheese purchased oe
A1 for interpretation of calculations within the report</td>
<td>H</td>
</tr>
</table>

<table>
<tr>
<td>19</td>
<td></td>
<td>Students' diagrams, measures and reports will vary.
<table>
<tr><th></th><th>Mean</th><th>Median</th><th>Mode</th><th>Range</th></tr>
<tr><td>C</td><td>7.8</td><td>8</td><td>8</td><td>3</td></tr>
<tr><td>PG</td><td>3.8</td><td>4</td><td>5</td><td>3</td></tr>
<tr><td>SB</td><td>4.6</td><td>5</td><td>5</td><td>3</td></tr>
<tr><td>M</td><td>4.4</td><td>4</td><td>4</td><td>1</td></tr>
<tr><td>S</td><td>4</td><td>4</td><td>3,5</td><td>2</td></tr>
<tr><td>CS</td><td>4.6</td><td>5</td><td>5</td><td>1</td></tr>
<tr><td>R</td><td>2</td><td>2</td><td>2</td><td>0</td></tr>
</table>
</td>
<td>B1

B1
M1

M1

A1

5</td>
<td>3</td>
<td>B1 for report that includes an awareness of change over time and which wines are increasing and which are decreasing in popularity
B1 for appropriate diagram / graph oe
M1 for calculation of ranges to show which wines are most popular oe.
M1 for calculation of mean, median or mode to show average number of cases of wine purchased purchased oe

A1 for interpretation of calculations within the report</td>
<td>H</td>
</tr>
<tr>
<td>20</td>
<td>Estimated mean = $\frac{2405}{87}$
= 27.6… minutes</td>
<td>From the cumulative frequency (cf) graph (some variance depending on students' reading of unmarked intervals on graph):
<table>
<tr><th>mp</th><th>c.freq</th><th>freq</th><th>f×mp</th></tr>
<tr><td>7.5</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>12.5</td><td>8</td><td>8</td><td>100</td></tr>
<tr><td>17.5</td><td>15</td><td>7</td><td>105</td></tr>
<tr><td>22.5</td><td>25</td><td>10</td><td>225</td></tr>
<tr><td>27.5</td><td>45</td><td>20</td><td>550</td></tr>
<tr><td>32.5</td><td>75</td><td>30</td><td>975</td></tr>
<tr><td>37.5</td><td>87</td><td>12</td><td>450</td></tr>
<tr><td>Totals</td><td>87</td><td>86</td><td>2405</td></tr>
</table>
Estimate: 28 minutes (to the nearest minute).</td>
<td>M1
M1

A1

A1
4</td>
<td>2
3</td>
<td>M1 for identification of frequencies from cf graph
M1 for identification of mid-points of classes from cf graph
A1 for table to show data derived from cf graph oe

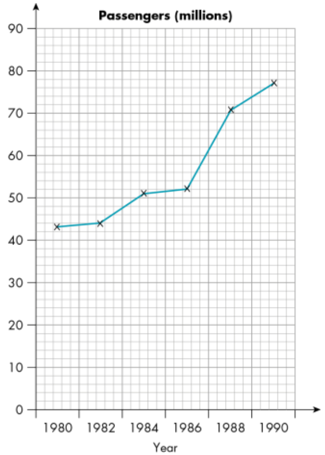
A1 for their estimated mean (ft)
Allow range of 26–29 minutes</td>
<td>H</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>Find the top 15% on the cumulative frequency scale, read along to the graph and read down to the marks.
The mark seen will be the minimum mark needed for this top grade.</td>
<td>M1
A1

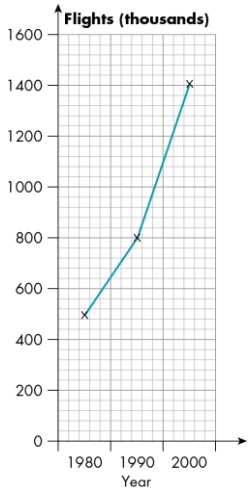
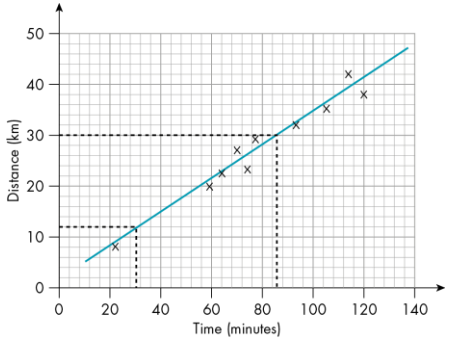
2</td>
<td>2</td>
<td>M1 for explanation of use of graph
A1 for explanation of where to find the minimum grade</td>
<td>H</td>
</tr>
</table>

22	<table><tr><td>Class width</td><td>fd</td><td>f</td></tr><tr><td>20</td><td>0.5</td><td>10</td></tr><tr><td>20</td><td>1.5</td><td>30</td></tr><tr><td>10</td><td>6</td><td>60</td></tr><tr><td>10</td><td>4.5</td><td>45</td></tr><tr><td>10</td><td>4</td><td>40</td></tr><tr><td>20</td><td>1.5</td><td>30</td></tr><tr><td></td><td></td><td>215</td></tr></table> Frequency density (fd) = frequency (f) divided by class width Number of students getting top grade is 10% of 215 = 21.5. 30 students got 80–100 $\frac{21.5}{30} = 0.7167$ Class width = 20 $20 \times (1 - 0.716) = 5.667$ Mark for top grade = 80 + 5.667…	The mark for the top grade is 85.7 (or 86 to the nearest integer score).	M1 M1 A1 M1 M1 A1 **6**	2 3	M1 for identification of class widths and frequency densities M1 for calculation of frequencies using f = fd × w A1 for clear table showing w, fd and f M1 for division to get 0.716 oe M1 for class width × 1 – 0.716 oe A1 cao	H
23	Work out frequency density from $\text{frequency} = \frac{\text{frequency}}{\text{class width}}$ <table><tr><td>Class width</td><td>fd</td><td>f</td></tr><tr><td>2</td><td>22</td><td>44</td></tr><tr><td>2</td><td>25</td><td>50</td></tr><tr><td>2</td><td>11</td><td>22</td></tr><tr><td>4</td><td>5</td><td>20</td></tr><tr><td></td><td></td><td>136</td></tr></table> 20 people travel more than 4 km to church. $\frac{20}{136} = \frac{5}{34} = 0.147 = 0.15$ to 2 dp	The probability is 0.15 (or $\frac{5}{34}$).	M1 M1 M1 A1 **4**	2 3	M1 for calculation of frequency of 20 for class travelling over 4 km to church M1 for 20 ÷ 136 oe M1 for rounding to one dp A1 cao Notes: Any one in 4–6 class could travel exactly 4 km to church so only consider the 8–10 class	H

24	<table><tr><td></td><td>B</td><td>C</td></tr><tr><td>Max</td><td>10</td><td>14</td></tr><tr><td>Min</td><td>4</td><td>2</td></tr><tr><td>LQ</td><td>6</td><td>4</td></tr><tr><td>M</td><td>7</td><td>5</td></tr><tr><td>UQ</td><td>9</td><td>8</td></tr><tr><td>R</td><td>6</td><td>12</td></tr><tr><td>IQR</td><td>3</td><td>4</td></tr></table>	Gabriel could see either doctor, but students should provide a plausible reason, for example, Dr Ball because patients never have to wait longer than 10 minutes, whereas they may have to wait up to 14 minutes for Dr Charlton; or Dr Charlton because the average waiting time is less than for Dr Ball.	M1 B1 B1 **3**	3	M1 for interpretation of data from box plots B1 for choice of doctor with reason, e.g. average waiting time oe B1 for choice of doctor with further reason, e.g. reference to range or inter-quartile range	H
25	<table><tr><td></td><td>G</td><td>B</td></tr><tr><td>Max</td><td>95</td><td>95</td></tr><tr><td>Min</td><td>15</td><td>25</td></tr><tr><td>LQ</td><td>40</td><td>50</td></tr><tr><td>M</td><td>55</td><td>65</td></tr><tr><td>UQ</td><td>65</td><td>75</td></tr><tr><td>R</td><td>80</td><td>70</td></tr><tr><td>IQR</td><td>25</td><td>25</td></tr></table>	There are quite a few possible correct answers, for example, boys' results have higher median and higher quartiles; or girls got lower marks than the boys, interquartile range is same for boys and girls.	M1 B1 B1 **3**	2	M1 for interpretation of data from box plots B1 for choice of either boys of girls with one justification, using their data from box plots oe B1 for choice of either boys of girls with one further justification, using their data from box plots oe	H
26		Many different possibilities. Each should contain no specific data – only general data such as: 'Scarborough generally had more sunshine than Blackpool', 'Blackpool tended to have more settled weather than Scarborough' or 'Scarborough had more sunshine on any one day'.	B1 B1 **2**	2	B1 for first correct comparison between Scarborough and Blackpool B1 for a further correct comparison between Scarborough and Blackpool	H
27	<table><tr><td></td><td>Box</td><td>CF</td></tr><tr><td>Max</td><td>90</td><td>100</td></tr><tr><td>Min</td><td>10</td><td>20</td></tr><tr><td>LQ</td><td>30</td><td>58</td></tr><tr><td>M</td><td>34</td><td>61</td></tr><tr><td>UQ</td><td>45</td><td>66</td></tr><tr><td>R</td><td>80</td><td>80</td></tr><tr><td>IQR</td><td>15</td><td>8</td></tr></table>	These are representations from different distributions as no data points match for measures of spread or tendency.	B1 B1 **2**	2	B1 for interpretation of data from box plots B1 for comparison of measures of spread or tendency to arrive at a 'no' answer	H
28		The measures of central tendency are likely to be different. The range is likely to be bigger for 7 to 13 than for 12 to 13 only. The greatest value (for the tallest student) will be the same and there may be similar differences between boys and girls.	B1 B1 **2**	2	B1 for mention of comparisons of range of heights oe B1 for consideration of gender differences in height with comparison of different age ranges, for example, girls are often taller than boys in Y7 and this trend is reversed by the time they reach Y13	H

29 a		Suitable statements about the distributions with justification for why the representations are from different distributions.	B1	2	B1 for comparison statements such as D1 has data for ages 18–11, which D2 does not; D2 has data for ages 30–32, which D1 does not; D1 has a positive skew so it has a low mode with a long right tail and D2 has a slight negative skew, so it has a higher mode but more consistency across a greater range oe	H
b		Distribution 1 is for Africa, distribution 2 is for Europe, plus observations such as: there is a peak at 21; distribution 1 is positively skewed; there is a more constant frequency across the range for distribution 2 with more women marrying older in distribution 2.	B1 **2**		B1 for observation of skewness and students' comments on and perception of possible reasons such as: women in Africa tend to marry at a younger age; European women have better health care or careers and so marry later; there is more gender equality and independent women in Europe oe	
30 a		Examples such as: positive skew – income distribution negative skew – age at death in developed countries.	B1 B1	2	B1 for positive skew context where there is a long tail to the right (higher values) and the bulk of the distribution is to the left (lower values) B1 for negative skew context where there is a long tail to the left (lower values) and the bulk of the distribution is to the right (higher values)	H
b		The box will be shifted to be in line with the bulk of the distribution. It will veer right for negative skew, left for positive and central for no skew.	B1		B1 for description of negative skew that includes a comment on the distribution being focused to the right (lower values) oe	
c		Suitable box plot for each distribution. Positive skew (long right tail, median to the left)  No skew  Negative skew (long left tail, median to the right) 	B1 B1 B1 **6**		B1 for suitable positive skew box plot B1 for suitable no skew box plot B1 for suitable negative skew box plot Positive skew (long right tail, median to the left)	

31 a		Examples of diagrams to illustrate changes over a ten-year period could either be number of flights or number of passengers as line graphs oe.	M1 A1	2	M1 for graph or diagram to illustrate either flights or passengers from 1980 to1990 oe B1 for clear diagram and appropriate labels	H
b		Students' research may vary, as will their diagrams. Examples of diagrams to show changes over the period 1980, 1990, 2000 could be either numbers of flights or numbers of passengers as line graphs oe.  	M1 A1 A1		M1 for graph or diagram to illustrate either flights or passengers from 1980 to 2000 oe A1 for clear diagram and appropriate labels A1 for appropriate research	
c		Extrapolating the data in the table for the period between 1980 and 1990 gives: Passengers in 2010 – approximately 145 (million) Flights in 2010 – approximately 1400 (thousand) or 1.4 million. However, looking at current data on the internet, the above 2010 estimates were already being approached in 2000 (cheap flights were introduced). Extrapolating the later data (up to 1999) gives: Passengers in 2010 – approximately 200 (million)	A1 A1		A1 for their extrapolation figure for number of passengers in 2010 A1 for their extrapolation figure for number of flights in 2010	

		Flights in 2010 – approximately 2000 (thousand) or 2.0 million.  		**7**			
32 a **b** **c**		 12 km 86 minutes	M1 A1 A1 **3**	2	M1 for appropriate distance–time graph A1 for ft answer from their graph A1 for ft answer from their graph	H

33 a		Harry may have misinterpreted the information because owning multiple TVs is more likely to be a measure of wealth rather than of health. In countries of high wealth, where people are likely to own more than one TV, the healthcare system is probably much better than in poorer countries, or nutrition is better than in poorer countries. It may also be a measure of accessibility to electricity.	B1	2	B1 for clear evidence of awareness of other variables that could contribute to the results shown on the scatter graph	H
b		A more likely explanation for the relationship is that people who live in a more affluent country will have more access to high-value goods, such as TVs, a better supply of electricity and better healthcare and nutrition.	B1 **2**		B1 for appropriate suggestions as described	