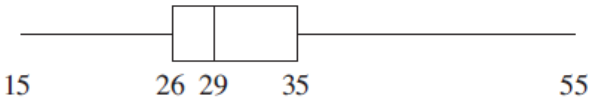
The times taken, in minutes, by 80 people to complete a crossword puzzle are summarised by the box and whisker plot below.

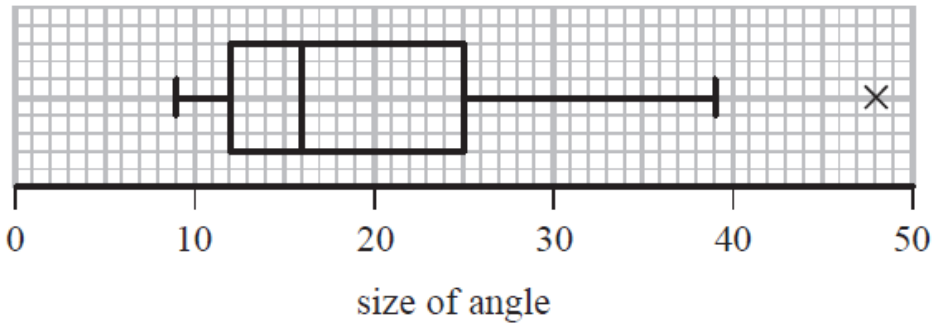


(i) Write down the range and the interquartile range of the times. [2]

(ii) Determine whether any of the times can be regarded as outliers. [3]

(iii) Describe the shape of the distribution of the times. [1]

Each of 60 students was asked to draw a 20° angle without using a protractor. The size of each angle drawn was measured. 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)

Q 1 **(i)**	The range = 55 – 15 = 40 The interquartile range = 35 – 26 = 9	B1 CAO B1 CAO	 **2**
(ii)	 35 + 1.5 x 9 = 48.5 26 – 1.5 x 9 = 12.5 Any value > 48.5 is an outlier (so 55 will be an outlier).	 M1 for 48.5 oe M1 for 12.5 oe A1 (FT their IQR in (i))	 **3**
(iii)	One valid comment such as eg: Positively skewed Middle 50% of data is closely bunched	E1	 **1**
		TOTAL	**6**

1. **(a)**	[Range = 48 – 9] = **39**	B1 (1)
(b)	[IQR = 25 – 12]= **13**	B1 (1)

Every day, George attempts the quiz in a national newspaper. The quiz always consists of 7 questions. In the first 25 days of January, the numbers of questions George answers correctly each day are summarised in the table below.

Number correct	1	2	3	4	5	6	7
Frequency	1	2	3	3	4	7	5

(i) Draw a vertical line chart to illustrate the data. [2]

(ii) State the type of skewness shown by your diagram. [1]

(iii) Calculate the mean and the mean squared deviation of the data. [3]

(iv) How many correct answers would George need to average over the next 6 days if he is to achieve an average of 5 correct answers for all 31 days of January? [2]

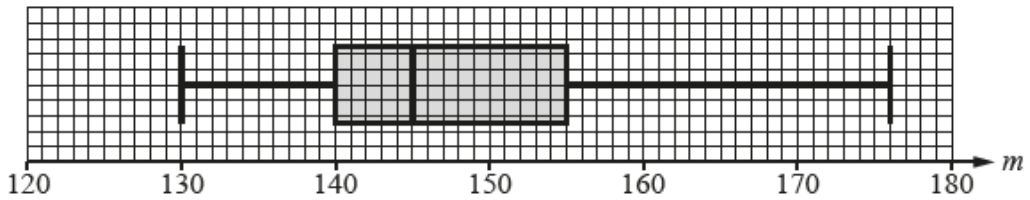
3 The masses, m grams, of 52 apples of a certain variety were found and summarised as follows.

$$n = 52 \qquad \Sigma(m - 150) = -182 \qquad \Sigma(m - 150)^2 = 1768$$

(i) Find the mean and variance of the masses of these 52 apples. **[5]**

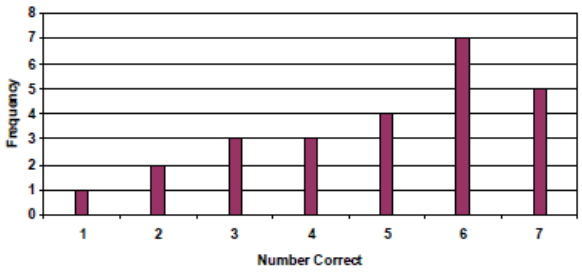
(ii) Use your answers from part **(i)** to find the exact value of Σm^2. **[3]**

The masses of the apples are illustrated in the box-and-whisker plot below.



(iii) How many apples have masses in the interval $130 \leqslant m < 140$? **[2]**

(iv) An 'outlier' is a data item that lies more than 1.5 times the interquartile range above the upper quartile, or more than 1.5 times the interquartile range below the lower quartile. Explain whether any of the masses of these apples are outliers. **[3]**

Q1 **(i)**		G1 Labelled linear scales G1 Height of lines	**2**
(ii)	Negative (skewness)	B1	**1**
(iii)	Σfx = 123 so mean = 123/25 = 4.92 o.e. $S_{xx} = 681 - \frac{123^2}{25} = 75.84$ M.s.d = $\frac{75.84}{25} = 3.034$	B1 M1 for S_{xx} attempted A1 FT their 4.92	**3**
(iv)	Total for 25 days is 123 and totals for 31 days is 155. Hence total for next 6 days is 32 and so mean = 5.33	M1 31 x 5 – 25xtheir 4.92 A1 FT their 123	**2**
		TOTAL	**8**

3	NB in (i) and (ii) 1768 + 150^2 × 52 = 1171768 is incorrect and scores no marks in either part, except possible ft in (ii) .					
3	i		$\frac{-182}{52}$ or –3.5 seen or implied Mean = 150 – "3.5" = 146.5 or 147 $\frac{1768}{52} - ("-3.5")^2$ alone, eg not if + 150 = 21.75 or 21.8	B1 M1 A1 M1 A1 [5]	or $\frac{-182+150\times52}{52}$ or $\frac{7800-182}{52}$ B1M1 Allow within √ sign Not ISW, eg $\sqrt{21.75}$ (or 4.66) M1A0 ans 4.66, no working, M1A0 NB $\frac{1768}{52} - "146.5"^2$ or 1768 -('-3.5²) M0A0	Σm = 150x52 - 182 or 7618 B1 "7618" ÷ 52 M1 =146.5 A1 (Σ(m-150)²=1768 Σm²-300Σm+150²x52=1768 Σm²=1768+300x"7618"-150²x52 = 1117168) $\frac{1768+300\times"7618"-150^2\times52}{52} - "146.5'^2$ or $\frac{'1117168'}{52} - "146.5'^2$ fully correct method M1 = 21.75 A1
	ii		$\frac{\Sigma m^2}{52} - "146.5"^2 = "21.75"$ or Σm^2 = ('21.75' + '146.5²') × 52 ft their mean & +ve var from (i) for M2 Σm^2 = 1117168 ISW	M2 A1 [3]	Allow M1 for $\frac{\Sigma m^2}{52} - "3.5"^2 = "21.75"$ or Σm^2 = ('21.75' + '3.5²') × 52 Exact; no ft from (i) eg 147 or 21.8	Σ(m-150)²=1768 Σm²-300Σm+150²x52=1768 ≥ 2 terms correct M1 Σm²=1768+300x"7618"-150²x52 correct method M1 =1117168 A1 Correct ans, no wking M1M1A1 If incorrect ans given with no wking, possibly M1M1 for (ii) may be obtained by correct method seen in (i), However M1M0 or M0M0 is more likely.
3(iii)	The correct method is in the 1st column. However, most candidates will give the allowed method in the middle column and score both marks. NB 3rd column					
	iii		(52 + 1) ÷ 4 = 13.25 or (26+1) ÷ 2 = 13.5 (⇒ 13th apple has mass < 140) ⇒ (no. below 140 =) 13	M1 A1 [2]	Allow 52÷ 4 or 26 ÷2 (= 13) M1 ⇒ (no. below 140 =) 13 A1	Allow 52÷ 4 or 26 ÷2 (= 13) M1 (⇒ 13th apple has mass 140) ⇒ (no. below 140 =) 12 A0
	iv		IQR = 15 seen or implied 155+1.5×15 = 177.5 (or > 176) or 140–1.5×15 = 117.5 (or < 130) No outliers	B1 B1 B1 [3]	or 22.5 seen or implied 176-155 = 21 (or < 22.5) or 140-130=10 (or < 22.5) Ignore method	$\frac{176-155}{15}$ =1.4 (or <1.5) or $\frac{140-130}{15} = \frac{2}{3}$ (or <1.5) Equivalent correct methods may be seen For 2nd B1 allow 14 ≤ IQR ≤ 16

The times taken for 480 university students to travel from their accommodation to lectures are summarised below.

Time (t minutes)	$0 \leqslant t < 5$	$5 \leqslant t < 10$	$10 \leqslant t < 20$	$20 \leqslant t < 30$	$30 \leqslant t < 40$	$40 \leqslant t < 60$
Frequency	34	153	188	73	27	5

(i) Illustrate these data by means of a histogram. [5]

(ii) Identify the type of skewness of the distribution. [1]

2 40 people were asked to guess the length of a certain road. Each person gave their guess, l km, correct to the nearest kilometre. The results are summarised below.

l	10–12	13–15	16–20	21–30
Frequency	1	13	20	6

(i) **(a)** Use appropriate formulae to calculate estimates of the mean and standard deviation of l. **[6]**

(b) Explain why your answers are only estimates. **[1]**

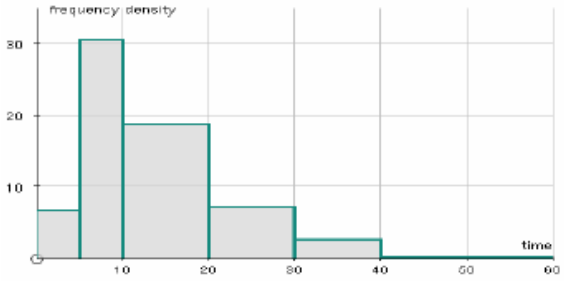
(ii) A histogram is to be drawn to illustrate the data. Calculate the frequency density of the block for the 16–20 class. **[2]**

(iii) Explain which class contains the median value of l. **[2]**

(iv) Later, the person whose guess was between 10 km and 12 km changed his guess to between 13 km and 15 km. Without calculation state whether the following will increase, decrease or remain the same:

(a) the mean of l, **[1]**

(b) the standard deviation of l. **[1]**

Q 3 (i)	time / freq / width / f dens 0- / 34 / 5 / 6.8 5- / 153 / 5 / 30.6 10- / 188 / 10 / 18.8 20- / 73 / 10 / 7.3 30- / 27 / 10 / 2.7 40- / 5 / 20 / 0.25  	M1 for fds A1 CAO Accept any suitable unit for fd such as eg freq per 5 mins. G1 linear scales on both axes and label G1 width of bars G1 height of bars	**5**
(ii)	Positive skewness	B1 CAO (indep)	**1**
		TOTAL	**6**

2 (i) (a)	Use of correct midpts		B1	11,14,18,25.5
	$\Sigma lf \div \Sigma f$	(= 706 ÷ 40)	M1	l within class, ≥ three lf seen
	= 17.65		A1	[17.575,17.7]
	$\Sigma l^2 f$	(= 13050.5)	M1	≥ three $l^2 f$ seen
	$\sqrt{\frac{"13050.5"}{40} - "17.65"^2}$	(= √14.74)	M1	÷40,-mean², √.Dep>0. ∑(l-17.65)²f, at least 3 M1,÷40,√
	= 3.84 (3 sfs)		A1 6	M1,3.84 A1. ÷ 4 ⇒ max B1M0A0M1M0A0
(b)	mid pts used or data grouped or exact values unknown oe		B1 1	not "orig values were guesses"
(ii)	20 ÷ 5		M1	condone 20 ÷ [4,5] or ans 5
	= 4		A1 2	
(iii)	20.5th value requ'd and		M1	condone 20th
	1st two classes contain 14 values			oe
	16 – 20		B1 2	or third class oe
(iv) (a)	increase		B1 1	
(b)	decrease		B1 1	
Total			[13]	

The marks of a group of female students in a statistics test are summarised in Figure 1

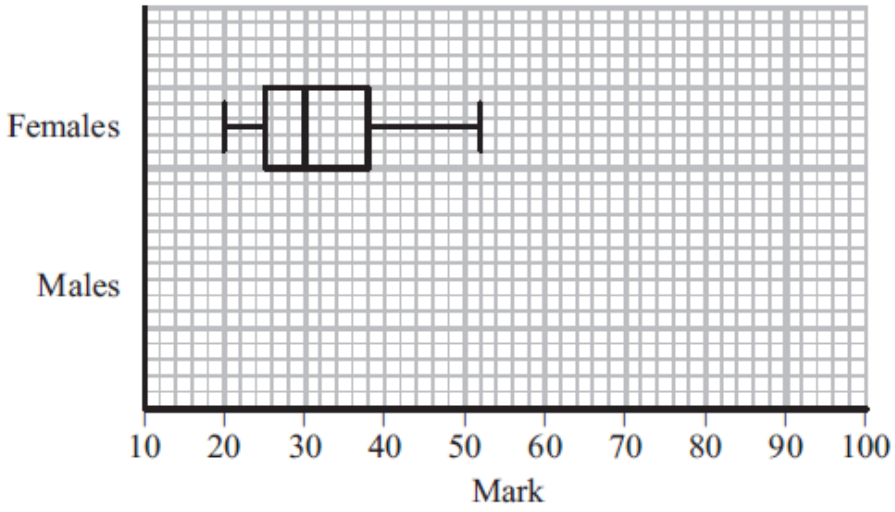


Figure 1

(a) Write down the mark which is exceeded by 75% of the female students.

(1)

The marks of a group of male students in the same statistics test are summarised by the stem and leaf diagram below.

Mark	(2|6 means 26)	Totals
1	4	(1)
2	6	(1)
3	4 4 7	(3)
4	0 6 6 7 7 8	(6)
5	0 0 1 1 1 3 6 7 7	(9)
6	2 2 3 3 3 8	(6)
7	0 0 8	(3)
8	5	(1)
9	0	(1)

(b) Find the median and interquartile range of the marks of the male students.

(3)

An outlier is a mark that is

either more than 1.5 × interquartile range above the upper quartile

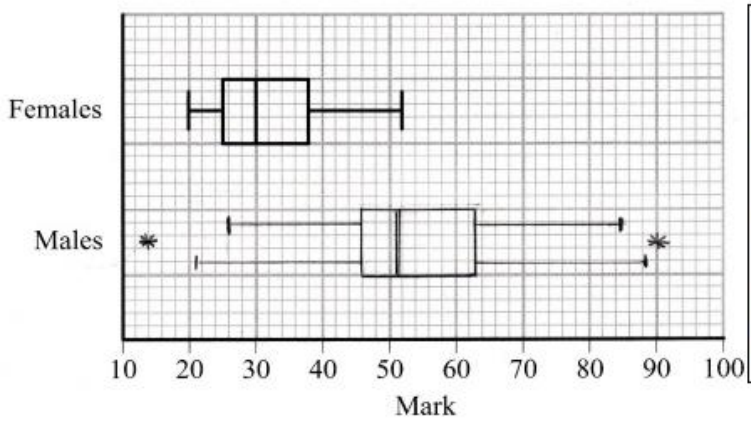
or more than 1.5 × interquartile range below the lower quartile.

(c) In the space provided on Figure 1 draw a box plot to represent the marks of the male students, indicating clearly any outliers.

(5)

(d) Compare and contrast the marks of the male and the female students.

(2)

2.	**(a)**	25 (allow any x where $24 < x < 26$)	B1 **(1)**
	(b)	$\mathbf{Q_2}$ (or median or m) = **51** **IQR** = 63 – 46 ,= **17** (or $Q_3 - Q_1 = 17$)	B1 M1, A1 **(3)**
	(c)	Outliers given by $46 - 1.5 \times 17 = 20.5$ or $63 + 1.5 \times 17 = 88.5$ Outliers limits are **20.5** and **88.5**  Allow lower whisker to 20.5 and upper whisker to 88.5 Do **not** allow a mix of whiskers e.g 20.5 and 85 Do **not** allow both sets of whiskers	M1 A1 M1 A1ft B1 **(5)**
	(d)	**Medians:** Median for females lower than males **IQR:** IQR for females smaller than males. Allow "lower/higher" but not "wider" **Range:** Range of females is less than males **Skewness:** Male and female marks are both positively skew Ignore other statements about average, spread, mean, st. Dev, variation, outliers etc	B1ft B1ft **(2)** **(11 marks)**

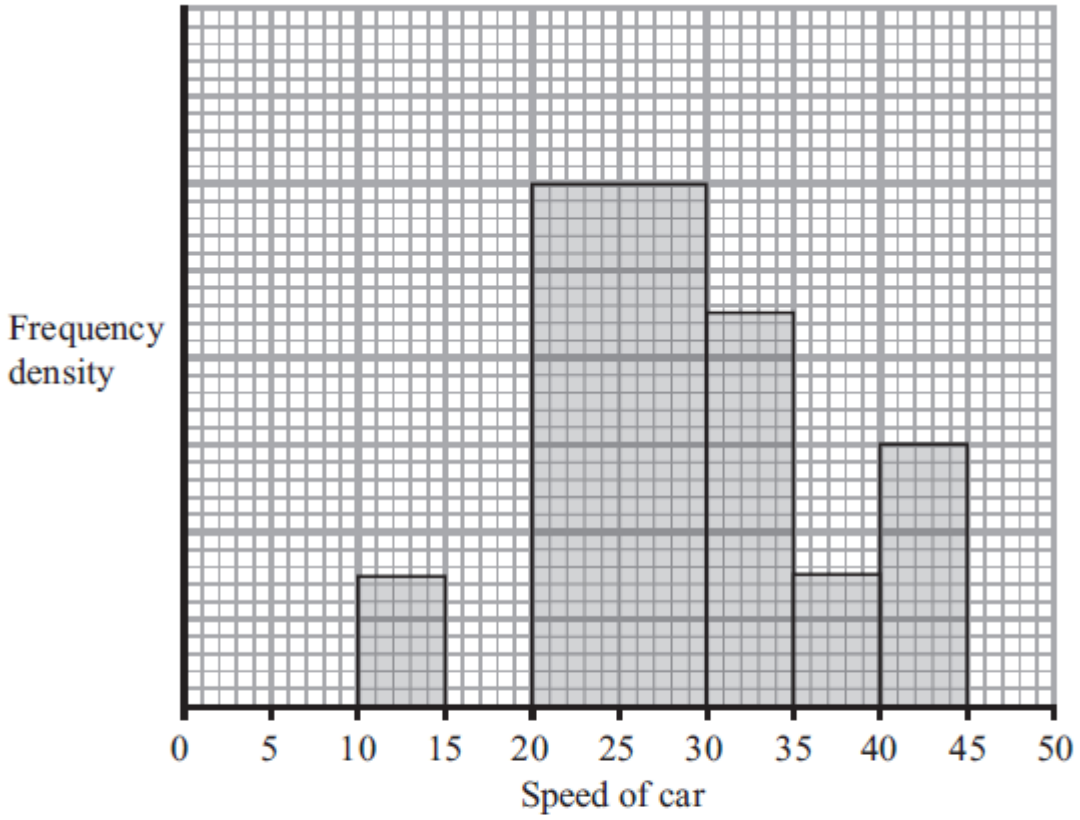


Figure 2

A policeman records the speed of the traffic on a busy road with a 30 mph speed limit. He records the speeds of a sample of 450 cars. The histogram in Figure 2 represents the results.

(a) Calculate the number of cars that were exceeding the speed limit by at least 5 mph in the sample.

(4)

(b) Estimate the value of the mean speed of the cars in the sample.

(3)

(c) Estimate, to 1 decimal place, the value of the median speed of the cars in the sample.

(2)

(d) Comment on the shape of the distribution. Give a reason for your answer.

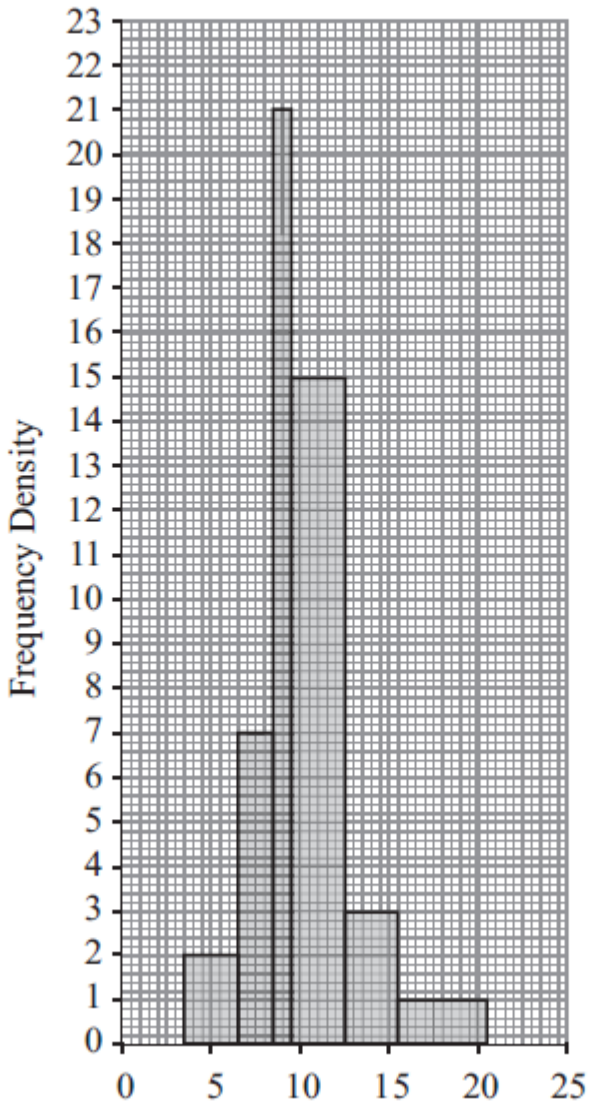
(2)

(e) State, with a reason, whether the estimate of the mean or the median is a better representation of the average speed of the traffic on the road.

(2)

5.	**(a)**	One large square $= \frac{450}{"22.5"}$ or one small square $= \frac{450}{"562.5"}$ (o.e. e.g. $\frac{"562.5"}{450}$)	M1	
		One large square = 20 cars or one small square = 0.8 cars or 1 car = 1.25 squares	A1	
		No. > 35 mph is: $4.5\times"20"$ or $112.5\times"0.8"$ (or equivalent e.g. using fd)	dM1	
		= **90** (cars)	A1	(4)
	(b)	$[\bar{x}] = \frac{30\times12.5+240\times25+90\times32.5+30\times37.5+60\times42.5}{450}$ $\left[=\frac{12975}{450}\right]$	M1 M1	
		$= 28.83...$ or $\frac{173}{6}$ **awrt 28.8**	A1	(3)
	(c)	$[Q_2 =]\ 20+\frac{195}{240}\times10$ (o.e.) [Allow use of $(n+1)$ giving 195.5 instead of 195]	M1	
		$= 28.125$ [Use of $(n+1)$ gives 28.145…] **awrt 28.1**	A1	(2)
	(d)	$Q_2 < \bar{x}$ So positive skew | [Condone $Q_2 \approx \bar{x}$] [so (almost) symmetric]	B1ft dB1ft	(2)
	(e)	[If chose skew in (d)] **median** (Q_2) Since the data is skewed or median not affected by extreme values | [If chose symmetric in (d)] **mean** ($\bar{x}$) Since it uses all the data	B1 dB1	(2)
				[13]

The histogram in Figure 1 shows the time, to the nearest minute, that a random sample of 100 motorists were delayed by roadworks on a stretch of motorway.



(a) Complete the table.

Delay (minutes)	Number of motorists
4 – 6	6
7 – 8	
9	21
10 – 12	45
13 – 15	9
16 – 20	

(2)

(b) Estimate the number of motorists who were delayed between 8.5 and 13.5 minutes by the roadworks.

(2)

1 (a)	14, 5	M1 A1 **(2)**
(b)	$21 + 45 + 3 = 69$	M1 A1 **(2)**
		Total 4

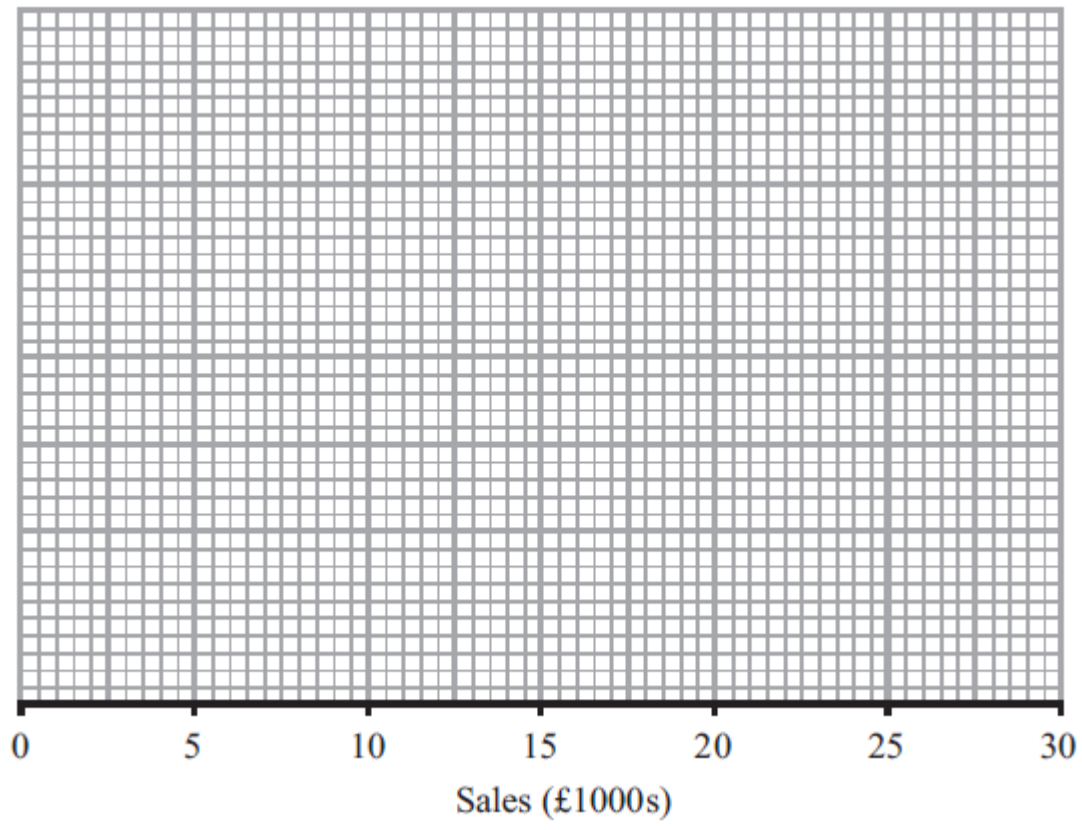
Over a long period of time a small company recorded the amount it received in sales per month. The results are summarised below.

	Amount received in sales (£1000s)
Two lowest values	3, 4
Lower quartile	7
Median	12
Upper quartile	14
Two highest values	20, 25

An outlier is an observation that falls
either 1.5 × interquartile range above the upper quartile
or 1.5 × interquartile range below the lower quartile.

(a) On the graph paper below, draw a box plot to represent these data, indicating clearly any outliers.

(5)

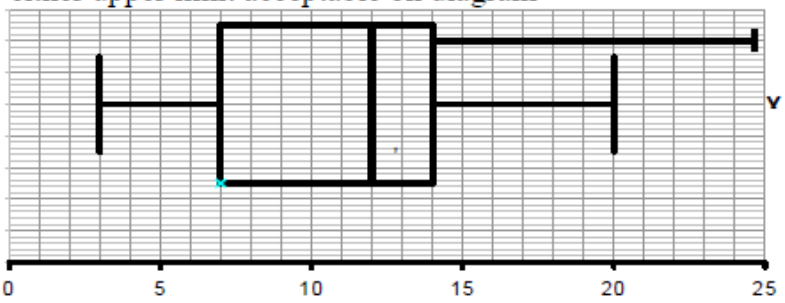


(b) State the skewness of the distribution of the amount of sales received. Justify your answer.

(2)

(c) The company claims that for 75 % of the months, the amount received per month is greater than £10000. Comment on this claim, giving a reason for your answer.

(2)

3.			
	(a)	Outliers $14 + 1.5\times(14-7) = 24.5$ $7 - 1.5\times(14-7) = -3.5$ Outlier 25 either upper limit acceptable on diagram  	M1 A1 M1 A1ft B1 **(5)**
	(b)	Since $Q_3 - Q_2 < Q_2 - Q_1$. Allow written explanation negatively skew	B1 dB1 **(2)**
	(c)	not true since the lower quartile is 7000 and therefore 75% above 7000 not 10000 or 10 is inside the box or any other sensible comment	B1 dB1 **(2)** **[9]**

The 19 employees of a company take an aptitude test. The scores out of 40 are illustrated in the stem and leaf diagram below.

	2|6 means a score of 26	
0	7	(1)
1	88	(2)
2	4468	(4)
3	2333459	(7)
4	00000	(5)

Find

(a) the median score,

(1)

(b) the interquartile range.

(3)

The company director decides that any employees whose scores are so low that they are outliers will undergo retraining.

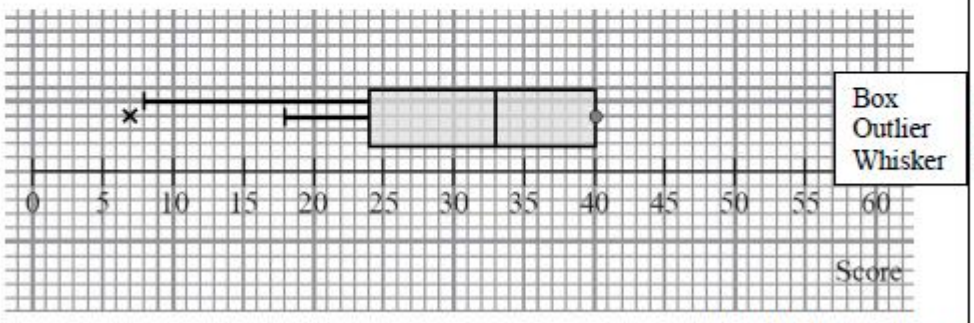
An outlier is an observation whose value is less than the lower quartile minus 1.0 times the interquartile range.

(c) Explain why there is only one employee who will undergo retraining.

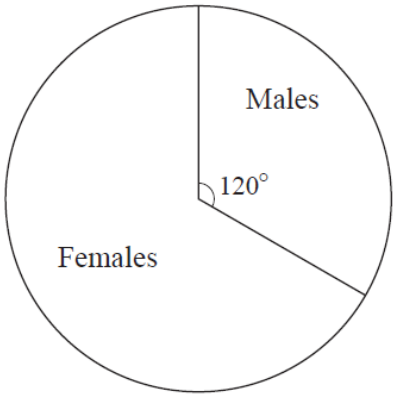
(2)

(d) On the graph paper on page 5, draw a box plot to illustrate the employees' scores.

(3)

Q2	(a)	Median is 33	B1 (1)
	(b)	$Q_1 = 24, Q_3 = 40, \text{IQR} = 16$	B1 B1 B1ft (3)
	(c)	$Q_1 - \text{IQR} = 24 - 16 = 8$ So 7 is only outlier	M1 A1ft (2)
	(d)	 (accept either whisker)	B1ft B1 B1ft (3)
			Total [9]

6 **(i)** The numbers of males and females in Year 12 at a school are illustrated in the pie chart. The number of males in Year 12 is 128.

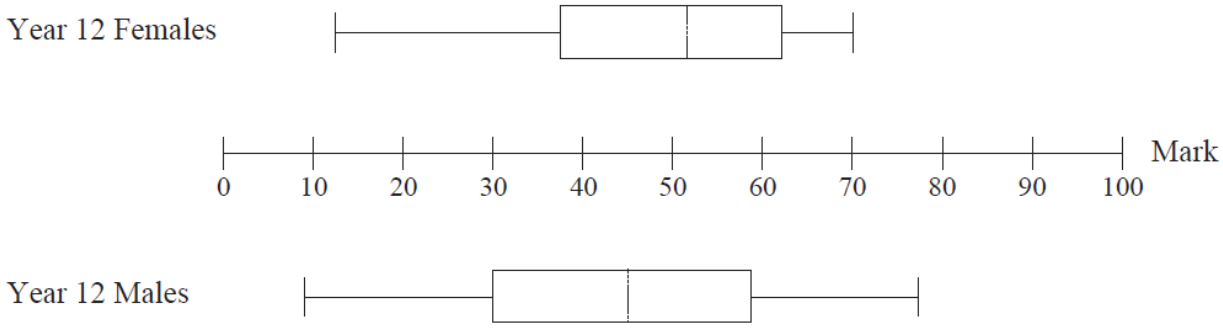


Year 12

(a) Find the number of females in Year 12. [1]

(b) On a corresponding pie chart for Year 13, the angle of the sector representing males is 150°. Explain why this does not necessarily mean that the number of males in Year 13 is more than 128. [1]

(ii) All the Year 12 students took a General Studies examination. The results are illustrated in the box-and-whisker plots.



(a) One student said "The Year 12 pie chart shows that there are more females than males, but the box-and-whisker plots show that there are more males than females."

Comment on this statement. [1]

(b) Give two comparisons between the overall performance of the females and the males in the General Studies examination. [2]

(c) Give one advantage and one disadvantage of using box-and-whisker plots rather than histograms to display the results. [2]

(iii) The mean mark for 102 of the male students was 51. The mean mark for the remaining 26 male students was 59. Calculate the mean mark for all 128 male students. [3]

6(i)(a)	256	B1 1	
			(i)(b) & (ii)(abc): ISW ie if correct seen, ignore extras
(b)	Total unknown or totals poss diff or Y13 may be smaller or similar or size of pie chart may differ	B1 1	pie chart shows only proportions oe or no. of students per degree may differ not "no. of F may be less" not "Y13 may be larger"
(ii)(a)	B&W does not show frequencies oe	B1 1	or B&W shows spread or shows mks or M lger range
(b)			1 mk about overall standard, based on median or on F's IQR being "higher" 1 mk about spread (or range or IQR) or about skewness. must be overall, not indiv mks must be comparison, not just figures Examples:
	F generally higher or median higher F higher on average or F better mks F IQR is above M IQR	B1	not F higher mean
	F more compact M wide(r) range or gter IQR or gter variation or gter variance or more spread or less consistent M evenly spread or F skewed	B1 2	not M have hiest and lowest mks condone F +ve skew
(c)	Advantage: B&W shows med or Qs or IQR or range or hiest & lowest or key values	B1	not B&W shows skewness not B&W shows info at a glance not B&W easier to compare data sets not B&W shows mean not B&W shows spread not B&W easier to calculate or easier to read
	Disadvantage: B&W loses info' B&W shows less info' B&W not show freqs B&W not show mode B&W: outlier can give false impression hist shows more info hist shows freqs or fds hist shows modal class (allow mode) hist shows distribution better can calc mean from hist	B1 2	not B&W does not give indiv (or raw) data not B&W does not show mean not hist shows freq for each mark not hist shows all the results not hist shows total allow adv of hist as disadv of B&W
(iii)	102 x 51 + 26 x 59 ÷ 128 = 52.6 (3 sfs)	M1 M1dep A1 3	or 5202 + 1534 or 6736
Total		**10**	