1. A teacher selects a random sample of 56 students and records, to the nearest hour, the time spent watching television in a particular week.

Hours	1–10	11–20	21–25	26–30	31–40	41–59
Frequency	6	15	11	13	8	3
Mid-point	5.5	15.5		28		50

(a) Find the mid-points of the 21–25 hour and 31–40 hour groups.

(2)

A histogram was drawn to represent these data. The 11–20 group was represented by a bar of width 4 cm and height 6 cm.

(b) Find the width and height of the 26–30 group.

(3)

(c) Estimate the mean and standard deviation of the time spent watching television by these students.

(5)

(d) Use linear interpolation to estimate the median length of time spent watching television by these students.

(2)

The teacher estimated the lower quartile and the upper quartile of the time spent watching television to be 15.8 and 29.3 respectively.

(e) State, giving a reason, the skewness of these data.

(2)
(Total 14 marks)

2. 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)

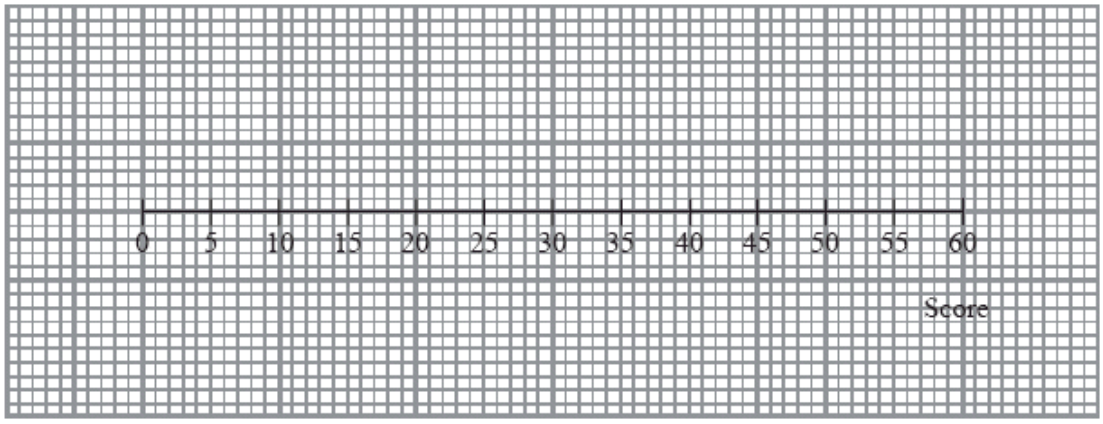
The company director decides that any employees whose scores are so low that they are outliers will undergo retraining.

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 below, draw a box plot to illustrate the employees' scores.



(3)
(Total 9 marks)

3. The birth weights, in kg, of 1500 babies are summarised in the table below.

Weight (kg)	Midpoint, x kg	Frequency, f
0.0 – 1.0	0.50	1
1.0 – 2.0	1.50	6
2.0 – 2.5	2.25	60
2.5 – 3.0		280
3.0 – 3.5	3.25	820
3.5 – 4.0	3.75	320
4.0 – 5.0	4.50	10
5.0 – 6.0		3

[You may use $\sum fx = 4841$ and $\sum fx^2 = 15\,889.5$]

(a) Write down the missing midpoints in the table above.

(2)

(b) Calculate an estimate of the mean birth weight.

(2)

(c) Calculate an estimate of the standard deviation of the birth weight.

(3)

(d) Use interpolation to estimate the median birth weight.

(2)

(e) Describe the skewness of the distribution. Give a reason for your answer.

(2)

(Total 11 marks)

4. There are 180 students at a college following a general course in computing. Students on this course can choose to take up to three extra options.

112 take systems support,
70 take developing software,
81 take networking,
35 take developing software and systems support,
28 take networking and developing software,
40 take systems support and networking,
4 take all three extra options.

(a) Draw a Venn diagram to represent this information.

(5)

A student from the course is chosen at random.

Find the probability that this student takes

(b) none of the three extra options,

(1)

(c) networking only.

(1)

Students who want to become technicians take systems support and networking. Given that a randomly chosen student wants to become a technician,

(d) find the probability that this student takes all three extra options.

(2)

(Total 9 marks)

5. The variable x was measured to the nearest whole number. Forty observations are given in the table below.

x	10 – 15	16 – 18	19 –
Frequency	15	9	16

A histogram was drawn and the bar representing the 10 – 15 class has a width of 2 cm and a height of 5 cm. For the 16 – 18 class find

(a) the width,

(1)

(b) the height

(2)

of the bar representing this class.

(Total 3 marks)

6. A researcher measured the foot lengths of a random sample of 120 ten-year-old children. The lengths are summarised in the table below.

Foot length, l, (cm)	**Number of children**
$10 \leq l < 12$	5
$12 \leq l < 17$	53
$17 \leq l < 19$	29
$19 \leq l < 21$	15
$21 \leq l < 23$	11
$23 \leq l < 25$	7

(a) Use interpolation to estimate the median of this distribution.

(2)

(b) Calculate estimates for the mean and the standard deviation of these data.

(6)

One measure of skewness is given by

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

(c) Evaluate this coefficient and comment on the skewness of these data.

(3)

Greg suggests that a normal distribution is a suitable model for the foot lengths of ten-year-old children.

(d) Using the value found in part (c), comment on Greg's suggestion, giving a reason for your answer.

(2)
(Total 13 marks)

7. In a study of how students use their mobile telephones, the phone usage of a random sample of 11 students was examined for a particular week.

The total length of calls, y minutes, for the 11 students were

17, 23, 35, 36, 51, 53, 54, 55, 60, 77, 110

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

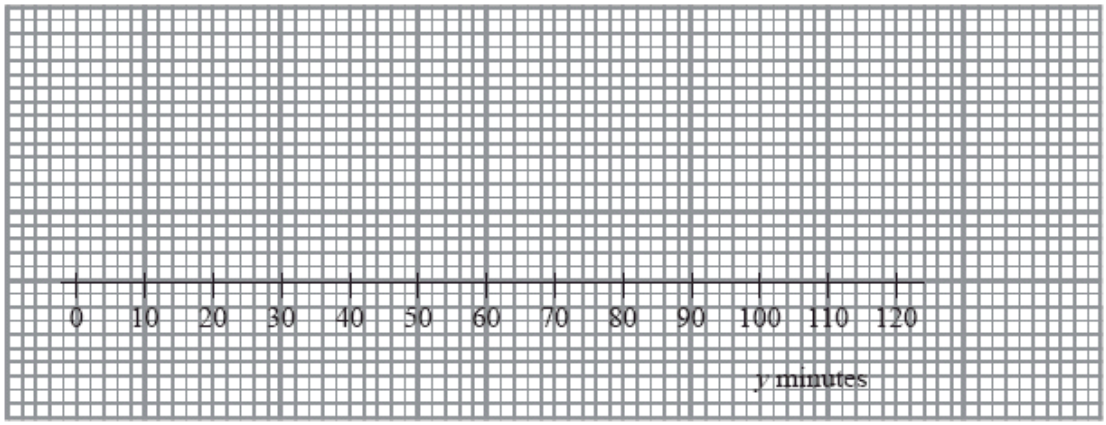
(3)

A value that is greater than $Q_3 + 1.5 \times (Q_3 - Q_1)$ or smaller than $Q_1 - 1.5 \times (Q_3 - Q_1)$ is defined as an outlier.

(b) Show that 110 is the only outlier.

(2)

(c) Using the graph below draw a box plot for these data indicating clearly the position of the outlier.



(3)

The value of 110 is omitted.

(d) Show that S_{yy} for the remaining 10 students is 2966.9

(3)

These 10 students were each asked how many text messages, x, they sent in the same week.

The values of S_{xx} and S_{xy} for these 10 students are $S_{xx} = 3463.6$ and $S_{xy} = -18.3$.

(e) Calculate the product moment correlation coefficient between the number of text messages sent and the total length of calls for these 10 students.

(2)

A parent believes that a student who sends a large number of text messages will spend fewer minutes on calls.

(f) Comment on this belief in the light of your calculation in part (e).

(1)
(Total 14 marks)

8. In a shopping survey a random sample of 104 teenagers were asked how many hours, to the nearest hour, they spent shopping in the last month. The results are summarised in the table below.

Number of hours	Mid-point	Frequency
0 – 5	2.75	20
6 – 7	6.5	16
8 – 10	9	18
11 – 15	13	25
16 – 25	20.5	15
26 – 50	38	10

A histogram was drawn and the group (8 – 10) hours was represented by a rectangle that was 1.5 cm wide and 3 cm high.

(a) Calculate the width and height of the rectangle representing the group (16 – 25) hours.

(3)

(b) Use linear interpolation to estimate the median and interquartile range.

(5)

(c) Estimate the mean and standard deviation of the number of hours spent shopping.

(4)

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

(2)

(e) State, giving a reason, which average and measure of dispersion you would recommend to use to summarise these data.

(2)

(Total 16 marks)

9. A disease is known to be present in 2% of a population. A test is developed to help determine whether or not someone has the disease.

Given that a person has the disease, the test is positive with probability 0.95

Given that a person does not have the disease, the test is positive with probability 0.03

(a) Draw a tree diagram to represent this information.

(3)

A person is selected at random from the population and tested for this disease.

(b) Find the probability that the test is positive.

(3)

A doctor randomly selects a person from the population and tests him for the disease.
Given that the test is positive,

(c) find the probability that he does not have the disease.

(2)

(d) Comment on the usefulness of this test.

(1)
(Total 9 marks)

10. The age in years of the residents of two hotels are shown in the back to back stem and leaf diagram below.

Abbey Hotel	8|5|0 means 58 years in Abbey hotel and 50 years in Balmoral hotel		Balmoral Hotel	
(1)	2	0		
(4)	9751	1		
(4)	9831	2	6	(1)
(11)	99997665332	3	447	(3)
(6)	987750	4	005569	(6)
(1)	0	5	000013667	(9)
		6	233457	(6)
		7	015	(3)

For the Balmoral Hotel,

(a) write down the mode of the age of the residents,

(1)

(b) find the values of the lower quartile, the median and the upper quartile.

(3)

(c) (i) Find the mean, $\bar{x}$, of the age of the residents.

(ii) Given that $\sum x^2 = 81213$ find the standard deviation of the age of the residents.

(4)

One measure of skewness is found using

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

(d) Evaluate this measure for the Balmoral Hotel.

(2)

For the Abbey Hotel, the mode is 39, the mean is 33.2, the standard deviation is 12.7 and the measure of skewness is – 0.454

(e) Compare the two age distributions of the residents of each hotel.

(3)
(Total 13 marks)

11. A person's blood group is determined by whether or not it contains any of 3 substances *A*, *B* and *C*.

A doctor surveyed 300 patients' blood and produced the table below

Blood Contains	No. of Patients
only *C*	100
A and *C* but not *B*	100
only *A*	30
B and *C* but not *A*	25
only *B*	12
A, B and *C*	10
A and *B* but not *C*	3

(a) Draw a Venn diagram to represent this information.

(4)

(b) Find the probability that a randomly chosen patient's blood contains substance *C*.

(2)

Harry is one of the patients. Given that his blood contains substance *A*,

(c) find the probability that his blood contains all 3 substances.

(2)

Patients whose blood contains none of these substances are called universal blood donors.

(d) Find the probability that a randomly chosen patient is a universal blood donor.

(2)
(Total 10 marks)

12. Cotinine is a chemical that is made by the body from nicotine which is found in cigarette smoke. A doctor tested the blood of 12 patients, who claimed to smoke a packet of cigarettes a day, for cotinine. The results, in appropriate units, are shown below.

Patient	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	K	L
Cotinine level, x	160	390	169	175	125	420	171	250	210	258	186	243

[You may use $\sum x^2 = 724961$]

(a) Find the mean and standard deviation of the level of cotinine in a patient's blood.

(4)

(b) Find the median, upper and lower quartiles of these data.

(3)

A doctor suspects that some of his patients have been smoking more than a packet of cigarettes per day. He decides to use $Q_3 + 1.5(Q_3 - Q_1)$ to determine if any of the cotinine results are far enough away from the upper quartile to be outliers.

(c) Identify which patient(s) may have been smoking more than a packet of cigarettes a day. Show your working clearly.

(4)

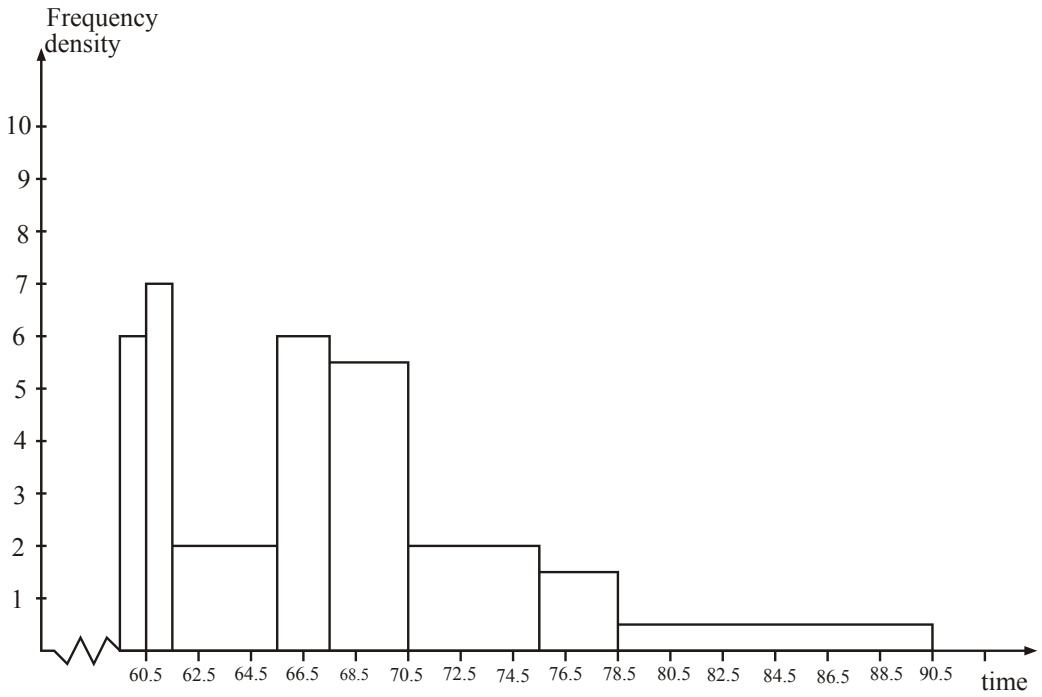
Research suggests that cotinine levels in the blood form a skewed distribution.

One measure of skewness is found using $\dfrac{(Q_1 - 2Q_2 + Q_3)}{(Q_3 - Q_1)}$

(d) Evaluate this measure and describe the skewness of these data.

(3)
(Total 14 marks)

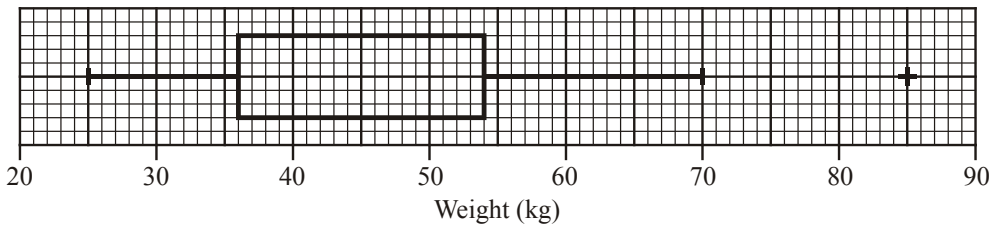
13. The histogram below shows the time taken, to the nearest minute, for 140 runners to complete a fun run.



Use the histogram to calculate the number of runners who took between 78.5 and 90.5 minutes to complete the fun run.

(Total 5 marks)

14. The box plot shown below shows a summary of the weights of the luggage, in kg, for each musician in an orchestra on an overseas tour.



The airline's recommended weight limit for each musician's luggage was 45 kg.
Given that none of the musicians' luggage weighed exactly 45 kg,

(a) state the proportion of the musicians whose luggage was below the recommended weight limit.

(1)

A quarter of the musicians had to pay a charge for taking heavy luggage.

(b) State the smallest weight for which the charge was made.

(1)

(c) Explain what you understand by the + on the box plot in the diagram above, and suggest an instrument that the owner of this luggage might play.

(2)

(d) Describe the skewness of this distribution. Give a reason for your answer.

(2)

One musician of the orchestra suggests that the weights of luggage, in kg, can be modelled by a normal distribution with quartiles as given in the diagram above.

(e) Find the standard deviation of this normal distribution.

(4)
(Total 10 marks)

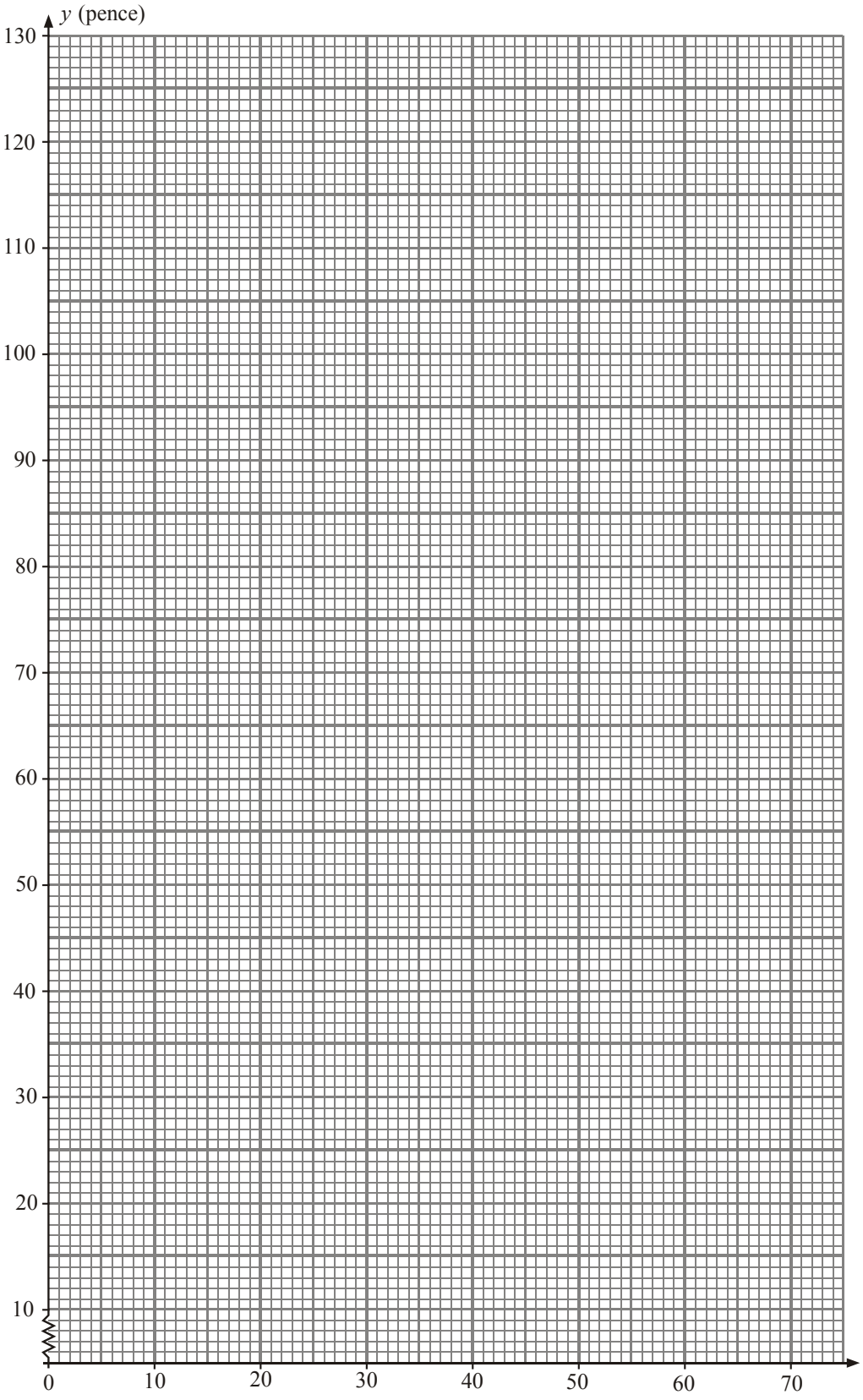
15. A student is investigating the relationship between the price (y pence) of 100g of chocolate and the percentage (x%) of cocoa solids in the chocolate.
The following data is obtained

Chocolate brand

Chocolate brand	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
x (% cocoa)	10	20	30	35	40	50	60	70
y(pence)	35	55	40	100	60	90	110	130

(You may use: $\sum x = 315$, $\sum x^2 = 15\,225$, $\sum y = 620$, $\sum y^2 = 56\,550$, $\sum xy = 28\,750$)

(a) On the graph paper below draw a scatter diagram to represent these data.


y (pence)
130
120
110
100
90
80
70
60
50
40
30
20
10
0
10
20
30
40
50
60
70

(2)

(b) Show that $S_{xy} = 4337.5$ and find S_{xx}.

(3)

The student believes that a linear relationship of the form $y = a + bx$ could be used to describe these data.

(c) Use linear regression to find the value of a and the value of b, giving your answers to 1 decimal place.

(4)

(d) Draw the regression line on your scatter diagram.

(2)

The student believes that one brand of chocolate is overpriced.

(e) Use the scatter diagram to

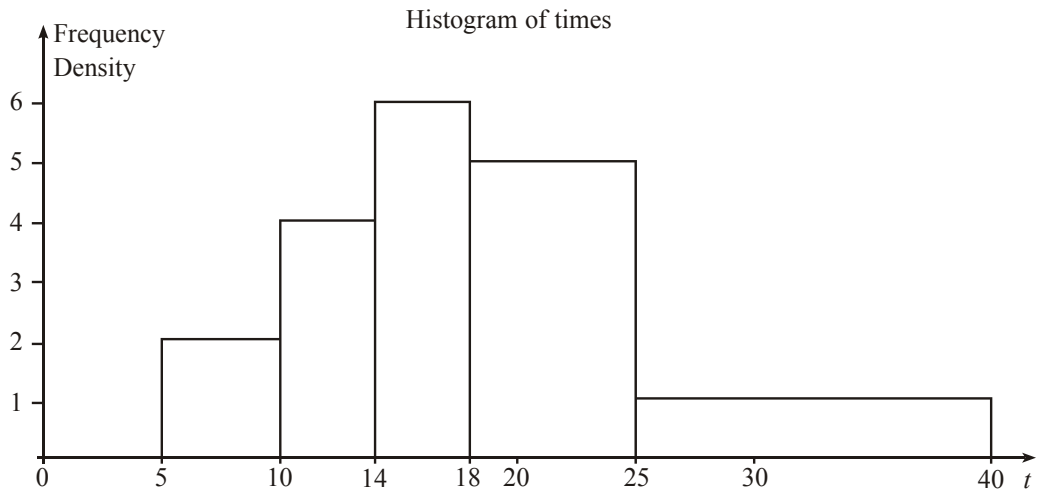
(i) state which brand is overpriced,

(ii) suggest a fair price for this brand.

Give reasons for both your answers.

(4)
(Total 15 marks)

16.



The diagram above shows a histogram for the variable t which represents the time taken, in minutes, by a group of people to swim 500m.

(a) Complete the frequency table for t.

t	5–10	10–14	14–18	18–25	25–40
Frequency	10	16	24		

(2)

(b) Estimate the number of people who took longer than 20 minutes to swim 500m.

(2)

(c) Find an estimate of the mean time taken.

(4)

(d) Find an estimate for the standard deviation of t.

(3)

(e) Find the median and quartiles for t.

(4)

One measure of skewness is found using $\frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$.

(f) Evaluate this measure and describe the skewness of these data.

(2)
(Total 17 marks)

17. Summarised below are the distances, to the nearest mile, travelled to work by a random sample of 120 commuters.

Distance (to the nearest mile)	Number of commuters
0–9	10
10–19	19
20–29	43
30–39	25
40–49	8
50–59	6

60–69	5
70–79	3
80–89	1

For this distribution,

(a) describe its shape,

(1)

(b) use linear interpolation to estimate its median.

(2)

The mid-point of each class was represented by x and its corresponding frequency by f giving

$$\Sigma fx = 3550 \text{ and } \Sigma fx^2 = 138020$$

(c) Estimate the mean and the standard deviation of this distribution.

(3)

One coefficient of skewness is given by

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

(d) Evaluate this coefficient for this distribution.

(3)

(e) State whether or not the value of your coefficient is consistent with your description in part (a). Justify your answer.

(2)

(f) State, with a reason, whether you should use the mean or the median to represent the data in this distribution.

(2)

(g) State the circumstance under which it would not matter whether you used the mean or the median to represent a set of data.

(1)

(Total 14 marks)

18. A teacher recorded, to the nearest hour, the time spent watching television during a particular week by each child in a random sample. The times were summarised in a grouped frequency table and represented by a histogram.

One of the classes in the grouped frequency distribution was 20–29 and its associated frequency was 9. On the histogram the height of the rectangle representing that class was 3.6 cm and the width was 2 cm.

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

(1)

(b) Write down the underlying feature associated with each of the bars in a histogram.

(1)

(c) Show that on this histogram each child was represented by 0.8 cm^2.

(3)

The total area under the histogram was 24 cm^2.

(d) Find the total number of children in the group.

(2)
(Total 7 marks)

19. (a) Describe the main features and uses of a box plot.

.....

.....

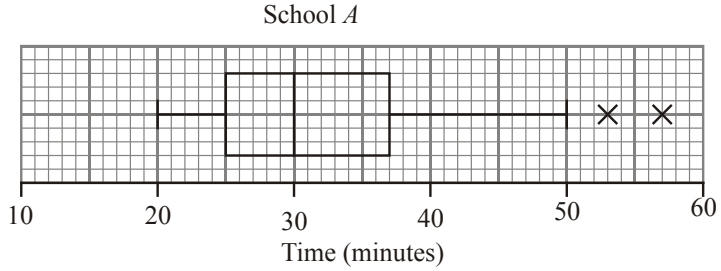
.....

.....

.....

(3)

Children from school *A* and *B* took part in a fun run for charity. The times to the nearest minute, taken by the children from school *A* are summarised in the figure below.



(b) (i) Write down the time by which 75% of the children in school *A* had completed the run.

.....

.....

(ii) State the name given to this value.

.....

(2)

(c) Explain what you understand by the two crosses (✕) on the figure above.

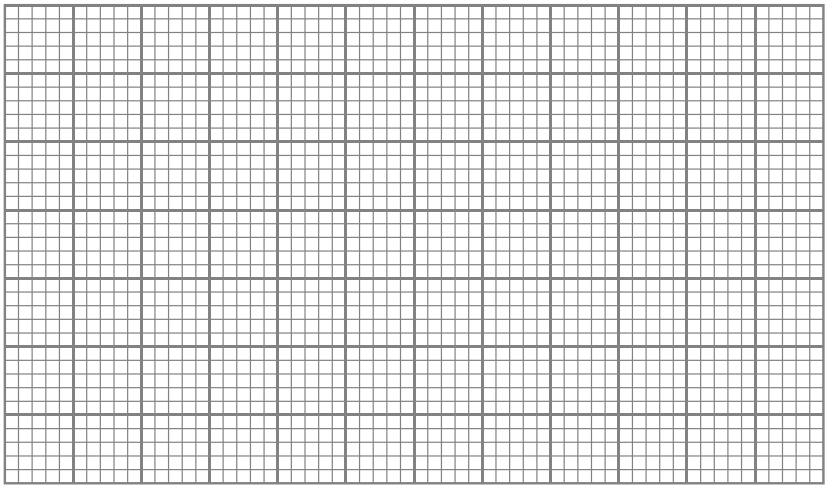
.....

.....

(2)

For school *B* the least time taken by any of the children was 25 minutes and the longest time was 55 minutes. The three quartiles were 30, 37 and 50 respectively.

(d) Draw a box plot to represent the data from school *B*.



(4)

(e) Compare and contrast these two box plots.

.....

.....

.....

.....

.....

(4)
(Total 15 marks)

20. Sunita and Shelley talk to one another once a week on the telephone. Over many weeks they recorded, to the nearest minute, the number of minutes spent in conversation on each occasion. The following table summarises their results.

Time (to the nearest minute)	Number of Conversations
5–9	2
10–14	9
15–19	20
20–24	13
25–29	8
30–34	3

Two of the conversations were chosen at random.

(a) Find the probability that both of them were longer than 24.5 minutes.

(2)

The mid-point of each class was represented by x and its corresponding frequency by f, giving $\Sigma fx = 1060$.

(b) Calculate an estimate of the mean time spent on their conversations.

(2)

During the following 25 weeks they monitored their weekly conversations and found that at the end of the 80 weeks their overall mean length of conversation was 21 minutes.

(c) Find the mean time spent in conversation during these 25 weeks.

(4)

(d) Comment on these two mean values.

(2)

(Total 10 marks)

21. Over a period of time, the number of people x leaving a hotel each morning was recorded. These data are summarised in the stem and leaf diagram below.

Number leaving								3 | 2 means 32	Totals
2	7	9	9						(3)
3	2	2	3	5	6				(5)
4	0	1	4	8	9				(5)
5	2	3	3	6	6	6	8		(7)
6	0	1	4	5					(4)
7	2	3							(2)
8	1								(1)

For these data,

(a) write down the mode,

(1)

(b) find the values of the three quartiles.

(3)

Given that $\Sigma x = 1335$ and $\Sigma x^2 = 71\ 801$ find

(c) the mean and the standard deviation of these data.

(4)

One measure of skewness is found using

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

(d) Evaluate this measure to show that these data are negatively skewed.

(2)

(e) Give two other reasons why these data are negatively skewed.

(4)

(Total 14 marks)

22. The following table summarises the distances, to the nearest km, that 134 examiners travelled to attend a meeting in London.

Distance (km)	Number of examiners
41–45	4
46–50	19
51–60	53
61–70	37
71–90	15
91–150	6

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

(1)

(b) Calculate the frequency densities needed to draw a histogram for these data.
(DO NOT DRAW THE HISTOGRAM)

(2)

(c) Use interpolation to estimate the median Q_2, the lower quartile Q_1, and the upper quartile Q_3 of these data.

(4)

The mid-point of each class is represented by x and the corresponding frequency by f.
Calculations then give the following values

$$\sum fx = 8379.5 \qquad \text{and} \qquad \sum fx^2 = 557489.75$$

(d) Calculate an estimate of the mean and an estimate of the standard deviation for these data.

(4)

One coefficient of skewness is given by

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

(e) Evaluate this coefficient and comment on the skewness of these data.

(4)

(f) Give another justification of your comment in part (e).

(1)
(Total 16 marks)

23. Aeroplanes fly from City *A* to City *B*. Over a long period of time the number of minutes delay in take-off from City *A* was recorded. The minimum delay was 5 minutes and the maximum delay was 63 minutes. A quarter of all delays were at most 12 minutes, half were at most 17 minutes and 75% were at most 28 minutes. Only one of the delays was longer than 45 minutes.

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 graph paper, draw a box plot to represent these data.

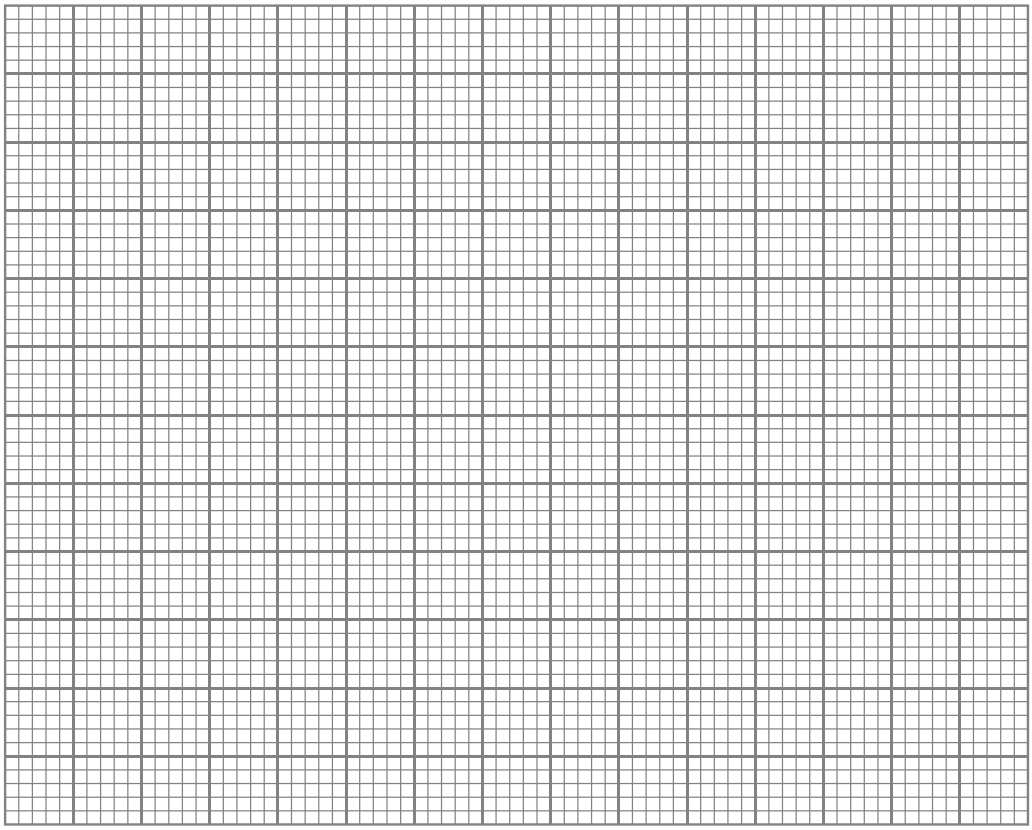
(7)

(b) Comment on the distribution of delays. Justify your answer.

(2)

(c) Suggest how the distribution might be interpreted by a passenger who frequently flies from City *A* to City *B*.

(1)



(Total 10 marks)

24. The number of caravans on Seaview caravan site on each night in August last year is summarised in the following stem and leaf diagram.

Caravans 1 | 0 means 10

		Totals
1	0 5	(2)
2	1 2 4 8	(4)
3	0 3 3 3 4 7 8 8	(8)
4	1 1 3 5 8 8 8 9 9	(9)
5	2 3 6 6 7	(5)
6	2 3 4	(3)

(a) Find the three quartiles of these data.

(3)

During the same month, the least number of caravans on Northcliffe caravan site was 31. The maximum number of caravans on this site on any night that month was 72. The three quartiles for this site were 38, 45 and 52 respectively.

(b) On graph paper and using the same scale, draw box plots to represent the data for both caravan sites. You may assume that there are no outliers.

(*One sheet of graph paper to be provided*)

(6)

(c) Compare and contrast these two box plots.

(3)

(d) Give an interpretation to the upper quartiles of these two distributions.

(2)
(Total 14 marks)

25. As part of their job, taxi drivers record the number of miles they travel each day. A random sample of the mileages recorded by taxi drivers Keith and Asif are summarised in the back-to-back stem and leaf diagram below.

-

Totals	Keith											Asif								Totals
(9)			8	7	7	4	3	2	1	1	0	18	4	4	5	7				(4)
(11)	9	9	8	7	6	5	4	3	3	1	1	19	5	7	8	9	9			(5)
(6)						8	7	4	2	2	0	20	0	2	2	4	4	8		(6)
(6)						9	4	3	1	0	0	21	2	3	5	6	6	7	9	(7)
(4)								6	4	1	1	22	1	1	2	4	5	5	8	(7)
(2)										2	0	23	1	1	3	4	6	6	7	8	(8)
(2)										7	1	24	2	4	8	9				(4)
(1)											9	25	4							(1)
(2)										9	3	26								(0)

Key: 0 ⏐18⏐ 4 means 180 for Keith and 184 for Asif

The quartiles for these two distributions are summarised in the table below.

	Keith	Asif
Lower quartile	191	a
Median	b	218
Upper quartile	221	c

(a) Find the values of a, b and c.

(3)

Outliers are values that lie outside the limits

$$Q_1 - 1.5(Q_3 - Q_1) \quad \text{and} \quad Q_3 + 1.5(Q_3 - Q_1).$$

(b) On graph paper, and showing your scale clearly, draw a box plot to represent Keith's data.

(8)

(c) Comment on the skewness of the two distributions.

(3)
(Total 14 marks)

26. A college organised a 'fun run'. The times, to the nearest minute, of a random sample of 100 students who took part are summarised in the table below.

Time	Number of students
40–44	10
45–47	15
48	23
49–51	21
52–55	16
56–60	15

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

(1)

(b) Write down the upper class boundary and the lower class boundary of the class 40–44.

(1)

(c) On graph paper, draw a histogram to represent these data.

(4)
(Total 6 marks)

27. The attendance at college of a group of 18 students was recorded for a 4-week period.

The number of students actually attending each of 16 classes are shown below.

18	18	17	17
16	17	16	18
18	14	17	18
15	17	18	16

(a) (i) Calculate the mean and the standard deviation of the number of students attending these classes.

(ii) Express the mean as a percentage of the 18 students in the group.

(5)

In the same 4-week period, the attendance of a different group of 20, students is shown below.

20	16	18	19
15	14	14	15
18	15	16	17
16	18	15	14

(b) Construct a back-to-back stem and leaf diagram to represent the attendance in both groups.

(5)

(c) Find the mode, median and inter-quartile range for each group of students.

(6)

The mean percentage attendance and standard deviation for the second group of students are 81.25 and 1.82 respectively.

(d) Compare and contrast the attendance of these 2 groups of students.

(3)
(Total 19 marks)

28. The values of daily sales, to the nearest £, taken at a newsagents last year are summarised in the table below.

Sales	Number of days
1 – 200	166
201 – 400	100
401 – 700	59
701 – 1000	30
1001 – 1500	5

(a) Draw a histogram to represent these data.

(*One sheet of graph paper to be provided*).

(5)

(b) Use interpolation to estimate the median and inter-quartile range of daily sales.

(5)

(c) Estimate the mean and the standard deviation of these data.

(6)

The newsagent wants to compare last year's sales with other years.

(d) State whether the newsagent should use the median and the inter-quartile range or the mean and the standard deviation to compare daily sales. Give a reason for your answer.

(2)
(Total 18 marks)

29. A travel agent sells holidays from his shop. The price, in £, of 15 holidays sold on a particular day are shown below.

299	1050	2315	999	485
350	169	1015	650	830
99	2100	689	550	475

For these data, find

(a) the mean and the standard deviation,

(3)

(b) the median and the inter-quartile range.

(4)

An outlier is an observation that falls either more than 1.5 × (inter-quartile range) above the upper quartile or more than 1.5 × (inter-quartile range) below the lower quartile.

(c) Determine if any of the prices are outliers.

(3)

The travel agent also sells holidays from a website on the Internet. On the same day, he recorded the price, £x, of each of 20 holidays sold on the website. The cheapest holiday sold was £98, the most expensive was £2400 and the quartiles of these data were £305, £1379 and £1805. There were no outliers.

(d) On graph paper, and using the same scale, draw box plots for the holidays sold in the shop and the holidays sold on the website.

(4)

(e) Compare and contrast sales from the shop and sales from the website.

(2)

(Total 16 marks)

30. In a particular week, a dentist treats 100 patients. The length of time, to the nearest minute, for each patient's treatment is summarised in the table below.

Time (minutes)	4 – 7	8	9 – 10	11	12 – 16	17 – 20
Number of patients	12	20	18	22	15	13

Draw a histogram to illustrate these data.

(Total 5 marks)

31. The number of bags of potato crisps sold per day in a bar was recorded over a two-week period. The results are shown below.

20, 15, 10, 30, 33, 40, 5, 11, 13, 20, 25, 42, 31, 17

(a) Calculate the mean of these data.

(2)

(b) Draw a stem and leaf diagram to represent these data.

(3)

(c) Find the median and the quartiles of these data.

(3)

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.

(d) Determine whether or not any items of data are outliers.

(3)

(e) On graph paper draw a box plot to represent these data. Show your scale clearly.

(3)

(f) Comment on the skewness of the distribution of bags of crisps sold per day. Justify your answer.

(2)

(Total 16 marks)

32. The total amount of time a secretary spent on the telephone in a working day was recorded to the nearest minute. The data collected over 40 days are summarised in the table below.

Time (mins)	90–139	140–149	150–159	160–169	170–179	180–229
No. of days	8	10	10	4	4	4

Draw a histogram to illustrate these data.

(4)

33. A restaurant owner is concerned about the amount of time customers have to wait before being served. He collects data on the waiting times, to the nearest minute, of 20 customers. These data are listed below.

15, 14, 16, 15, 17, 16, 15, 14, 15, 16,
17, 16, 15, 14, 16, 17, 15, 25, 18, 16

(a) Find the median and inter-quartile range of the waiting times.

(5)

An outlier is an observation that falls either 1.5 × (inter-quartile range) above the upper quartile or 1.5 × (inter-quartile range) below the lower quartile.

(b) Draw a boxplot to represent these data, clearly indicating any outliers.

(7)

(c) Find the mean of these data.

(2)

(d) Comment on the skewness of these data. Justify your answer.

(2)

(Total 16 marks)

34. A botany student counted the number of daisies in each of 42 randomly chosen areas of 1 m by 1 m in a large field. The results are summarised in the following stem and leaf diagram.

Number of daisies									1 | 1 means 11
1	1	2	2	3	4	4	4		(7)
1	5	5	6	7	8	9	9		(7)
2	0	0	1	3	3	3	3	4	(8)
2	5	5	6	7	9	9	9		(7)
3	0	0	1	2	4	4			(6)
3	6	6	7	8	8				(5)
4	1	3							(2)

(a) Write down the modal value of these data.

(1)

(b) Find the median and the quartiles of these data.

(4)

(c) On graph paper and showing your scale clearly, draw a box plot to represent these data.

(4)

(d) Comment on the skewness of this distribution.

(1)

The student moved to another field and collected similar data from that field.

(e) Comment on how the student might summarise both sets of raw data before drawing box plots.

(1)
(Total 11 marks)

35. Data relating to the lifetimes (to the nearest hour) of a random sample of 200 light bulbs from the production line of a manufacturer were summarised in a group frequency table. The mid-point of each group in the table was represented by x and the corresponding frequency for that group by f. The data were then coded using $y = \frac{(x - 755.0)}{2.5}$ and summarised as follows:

$$\Sigma fy = -467, \quad \Sigma fy^2 = 9179.$$

(a) Calculate estimates of the mean and the standard deviation of the lifetimes of this sample of bulbs.

(9)

An estimate of the interquartile range for these data was 27.7 hours.

(b) Explain, giving a reason, whether you would recommend the manufacturer to use the interquartile range or the standard deviation to represent the spread of lifetimes of the bulbs from this production line.

(2)
(Total 11 marks)