Answers

1.2 Data

1. Discrete, Continuous, Discrete

2. (a) Undertake a survey. Other valid answers possible. (b) Secondary data
   (c) May not be reliable - discuss with students

3. (a) Discrete (b) Qualitative (c) Continuous (d) Qualitative

1.3 Questionnaires and Surveys

1.- 6. Answers should be discussed with students.

7. Question (A) gives genuine answers – although they may be difficult to analyse.

   Question (B) will give precise answers to just three aspects of life on this housing estate.

8. (a) Cycling and Swimming - discuss with students

9. Answers should be discussed with students.

10. (a) The percentage late is given by:

Bus	Cycle	Car	Walk
26.7%	20%	22%	12.5%

   so her conclusion is supported by the data.

   (b) Choose pupils randomly; do not always use Tuesdays; increase sample size. Allow other valid answers.

   (c) Wrong – because you do not know whether there was an equal number of Y7 and Y8 pupils in the survey, and because there is no breakdown of lates for Years 7 and 8 separately.

11. Students' answers should be discussed.

12. Sample is not random; sample value of 19% is not necessarily true for the whole population; they may only have travelled once on a bus in the last week – not each day. Allow other valid answers.

1.4 Experimental Design

1. You would need a large sample to have proof! Indeed, for *proof*, you would need to survey *all* women.

2. (a) 220 (b) 180

   (c) 180, since it came from a larger sample.

   (d) Add a species of fish not present at the moment and then apply the same technique as in the jelly baby problem.

3. (a) Find the number of woodlice in similar areas of the garden, but some areas being damp and others not. Allow other valid answers.

   (b) Survey the number of woodlice in dark, damp conditions and in light damp conditions. Allow other valid answers.

1.5 Sampling

1. (a) Sample too small; samples may not be independent of each other. Allow other valid answers.

   (b) Easier and cheaper than investigating the whole population. Allow other valid answers.

2. Too small a sample, and not representative (as they are in the same family). Allow other valid answers.

3. Nobody under age 18 interviewed; too many males (particularly in the 26-40, 41-64 age groups). Allow other valid answers.

4. (a) 600 (b) (i) 72% (ii) 88%

   (c) The problem is not the sample size, but the large number of non-respondents. Allow other valid answers.

   (d) Undertake interviews (either face to face or by telephone). Allow other valid answers.

5. (a) (i) Label 001 to 500 all pupils, and take random digits, three at a time, ignoring those over 500 or repeats. Allow other valid answers. Discuss with students.

   (ii) Choose a number between 01 and 50 at random, say x; then choose $x,\ x+50,\ x+100,\ \ldots\ x+450$ as the sample. Allow other valid answers. Discuss with students.

   (b) (i) BOY/GIRL or by age groups or by Year groups. Allow other valid answers.

   (ii) You are more likely to get a representative sample. Allow other valid answers.

6. (a) Method 2 - it should include pupils from all year groups, although, being random, there is a chance that it may miss one year group altogether.

   (b) Method 1 would only sample pupils who arrive early.

   Method 3 might not have enough pupils from some year groups.

   (c) Question is posed in a negative way and is leading. Discuss with students - double negative in question - grammatical error.

   (d) e.g. "Should school uniform be worn? (YES/NO)". Discuss with students.

7. (a) Only people travelling by train would be sampled.

   Sample would consist mainly of commuters on their way home - not representative.

   'First 100' may not be a representative sample.

   Only one day sampled. Discuss with students.

   (b) Use electoral register as sampling frame, and choose sample randomly. Allow other valid answers.

   (c) (i) It is not viable to test the lifetime of all batteries (testing to destruction).

   (ii) Too expensive and too time consuming to take a census. Allow other valid answers.

   (iii) Check for quality control at regular intervals; too expensive to test all. Allow other valid answers.

8. (a) Pupils of different age and gender will have different spending patterns.

   (b) (i) There may be an unequal number of boys and girls so stratified sampling is an attempt to avoid bias.

   (ii) Number all boys in sampling frame, and choose 20 at random. Allow other valid answers.

9. (a) Possible answers: (i) C, F, I, L, O, R, U, X. (ii) D, H, L, P, T, X.

   (iii) F, L, R, X. Discuss with students.

   (b) Discuss with students.

   (c) Number 24 balls and place in a bag, then draw out the relevant number of balls to correspond with the number of machines. Allow other valid answers.

10. Y7:13; Y8:13; Y9:14; Y10:14; Y11:13; Y12:7; Y13:6 (taking the integer part of each answer plus the two greatest remainders to give a sample of size 80)

11. (a) Managers: 1; Supervisors: 2; Administrators: 1;
    Manual Staff: 13; Delivery Staff: 3 (b) Complete sample: 12

12. Change 5 to 6 for Group C. The calculated number of people to be selected for Group D is 4.5, which must be rounded to either 4 or 5 depending on whether the overall size of the sample is to be 27 or 28.

13. (a) Working either across the rows or down the columns, choose the 10th square in each case. Allow other valid answers. Discuss with students.

   (b) For example, using 4th row generates the sample 4H, 4G, 4D, 5F, 6E, 4F, 3F, 4B, 5I, 6A. Using the 6th column generates the sample 1B, 4F, 6E, 4B, 7H, 6A, 9C, 1H, 2F, 6I.

   These two samples have 4 squares in common. Allow other valid answers. Discuss with students.

14. Use pupil responses as the basis for discussion; possible points might include

   (a) People changing telephone numbers, so not contactable;
   inability to contact people not on the telephone so sample is biased.
   Allow other valid answers.

   (b) Sample may be very large and therefore unmanageable;
   people tend to enter in groups so identifying every 3rd person is not easy;
   selected people may refuse to take part;
   different entrances to arcade need to be manned. Allow other valid answers.

   (c) Sample is not representative;
   those sampled may object to being interviewed if it disrupts their leisure activity;
   responses may be influenced by how well the team are playing.
   Allow other valid answers.

   (d) Sample is not representative;
   sample may be fairly small;
   time of day has not been taken into account, nor day of the week.
   Allow other valid answers.

15. Use pupil responses as the basis for discussion.

16. (a) Identify adults with children at local primary schools, then select a sample stratified by the sizes of the primary schools. Allow other valid answers. Discuss with students.

1.5

(b) Random telephone poll, quota sample of people at a large shopping centre, etc.

(c) Sample of people using the bowling green (sample taken at various times of the day over, say, a 7-day period). Allow other valid answers. Discuss with students.

(d) Sample of people using public transport (buses, trains, etc.) conducted over a suitable time period (e.g. 2 weeks), at various times of the day, at suitable embarkation points (stations, bus stops), etc. Extension of survey to non-users of public transport to find out why. Allow other valid answers. Discuss with students.

2.1 Data Tables

1. (a) 0800 (b) 0923

   (c) He should catch the next train and get off at Exeter St. Davids and walk.

   (d) For each journey - catch the 0723 from Paignton and change at either Newton Abbot, Exeter St. Davids or Exeter Central - arriving at Bristol at 0932 or London at 1110.

2. (a) 2106 (b) Yes (c) The 2027 train from Reading.

3. (a) No (b) 1235 (c) 1141

4. (a) (i) 109 miles (ii) 34 miles (iii) 78 miles (b) (i) 149 miles

   (ii) The travelling distance would be reduced by 29 miles because Manchester is *en route* travelling from Birmingham to Leeds.

   (c) The route via Sheffield

5. (a) St. Malo (b) 3984 km (c) 2781 km

   (d) Quimper is closest to St. Malo, and Ile de Re is closest to Calais.

6. (a) 9 (b) 11 (c) 'D' grade.

7. (a) 7 (b) 6 (c) Years 7, 8, 9 and 11

   (d) (i) Year 10 (ii) 43 students

8. (a) £305 (b) £156 (c) £276 (d) (i) £260 (ii) £45

9. (a)

	Male	*Female*	*Total*
Standard	64	48	112
Senior	20	8	28
Total	84	56	140

   (b) Men are more likely to become senior conductors than women. Allow other valid answers.

10. (a) 51 (b) 11

    (c) 52 outdoor and 50 indoor shows a roughly even split (only marginally in favour of outdoor sports). Discuss with students.

11. (a) 4 (b) 19

    (c) People seem to achieve similar or better English results than French. Discuss with students.

12. (a) 31 (b) (i) 90 (ii) 30%

2.2 Stem and Leaf Diagrams

1. (a)

1	5 6 8
2	1 2 4 4 6 8 9
3	1 1 2 5 6 7 8
4	0 1 2 6 9
5	1 2 7

(b) (i) 42 (ii) 32

2. (a)

1	5 7 8 8 8 9 9
2	0 1 1 1 1 2 3 3 3 4 4 5 6 7 7 8
3	1 2 3 4 5 5 6 7
4	0 3 5 5 8
5	1 7 8
6	1 2 3 5 5 9
7	1 3
8	2

(b) The age category most commonly involved in fatal accidents is 20 – 29.
Fatal accidents are more likely to involve drivers under 40 years of age.
There is also a significant number of accidents involving drivers in their sixties.
Allow other valid answers. Discuss with students.

3. (a)

Volume 1				*Volume 2*			
			17	0	4	9	9
			18	2	3	5	
			19	0	5	5	
2	3	3	20	1	2		
		5	21	0			
		1	22				
		7	23				
0	6	9	24	0			
	3	4	25				
2	6	6	26	3			
		3	27				
	3	8	28				

(b) The tracks on Volume 1 are generally longer than those on Volume 2.
There is greater variation in the lengths of the tracks on Volume 2 than on Volume 1.
Altogether (assuming you like that type of music), Volume 1 is better value for money than Volume 2. Allow other valid answers. Discuss with students.

2.3 Pictograms and Bar Charts

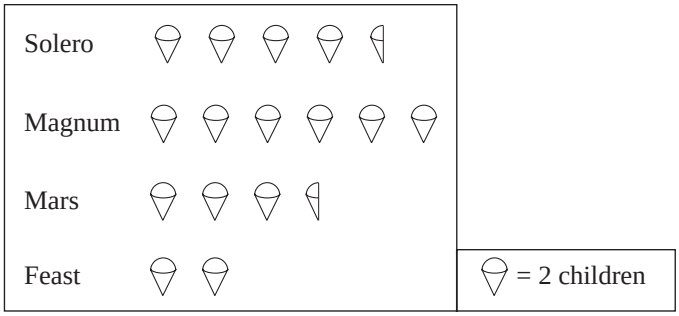
1. (a) 2002 (b) 2001

(c) (i) 10 (ii) 7 (iii) 12 (iv) 14; slight increase as she is generally recycling more each year and more publicity is being given to recycling (or other reasonable answer). Discuss with students.

2. (a) 400 (b) 250 (c) 700

(d) $5\frac{1}{2}$ (e) 3300 (f) 750 or 800

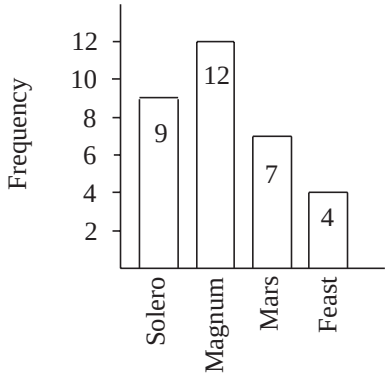
3. (a)



Allow other valid answers.

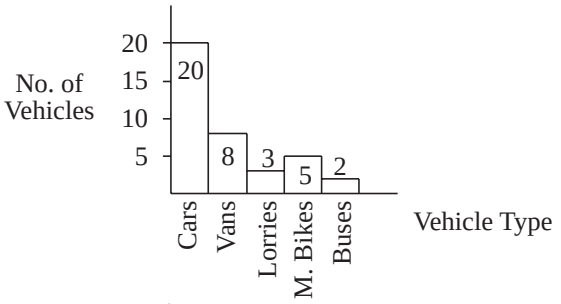
2.3

(b)



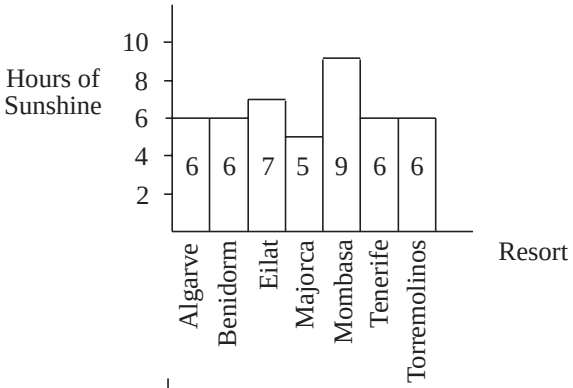
(c) Both allow easy visual comparison of popularity. For the pictogram, the key has to be used to determine the numbers of children voting for each brand whereas these can be read direct from the axis of the bar chart. Allow other valid answers.

4.

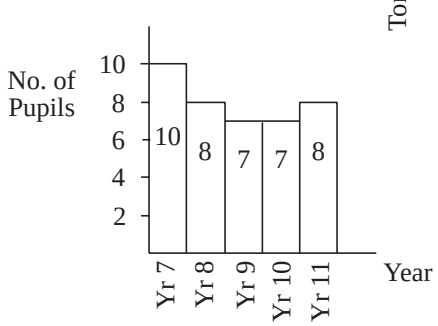


The majority of the traffic passing the school consisted of cars and vans. There was relatively little traffic of other types such as lorries and buses. Allow other valid answers.

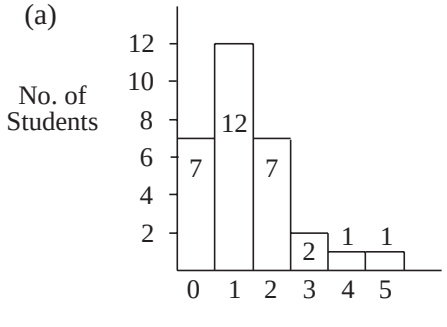
5.



6.



7. (a)



(b) To predict future pupil intake to the school and to plan for rises and falls in pupil numbers. Allow other valid answers.

2.3

8. (a) (i) 90% (ii) About 97% (b) (i) 0% (ii) About 22%

   (c) The percentage of households with some sort of TV has changed little, but there has been an increase in the number of homes receiving satellite or cable TV. Allow other valid answers.

9. (a) 50 (b) 35

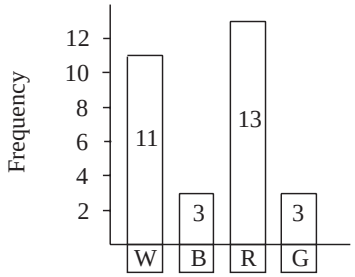
10. (a) 23 (b) We would expect there to be many more boys with shoe sizes around 8 and 9 than for 5 or 12, so the results are surprising.

11. (a) £14 (b) £375 (c) 15-16 coins

    (d) Only whole coins are used, and so the number is rounded up/down. Allow other valid answers.

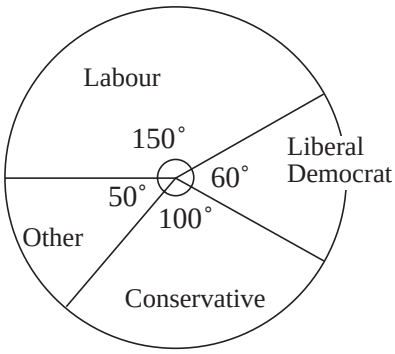
12. (a)

Colour	*Tally*	*Frequency*
White	𝍸 𝍸 I	11
Blue	III	3
Red	𝍸 𝍸 III	13
Green	III	3

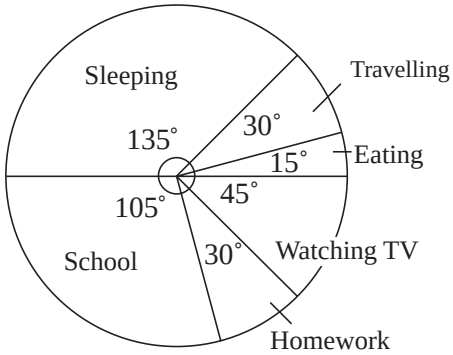


2.4 Pie Charts

1.

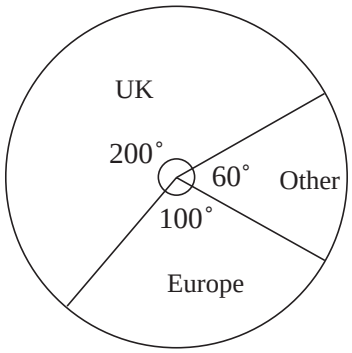


2.(a)



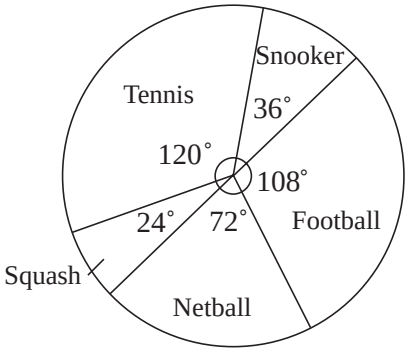
(b) Allow other valid answers. Discuss with students.

3.



2.4

4.



5. (a) 2 hours (b) 5 hours

6. (a) £10 (b) £15 (c) £35

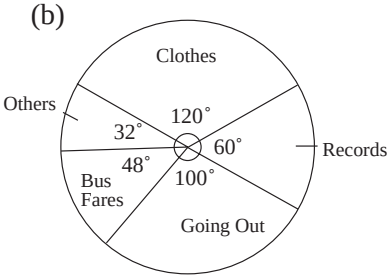
7. (a) 96° (b) 6 (c) 30 (d) 13

8. Airmail: 50, First Class: 320, Second Class: 350

9. (a)

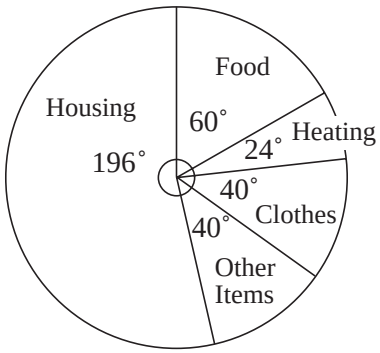
Items	Angle of Sector
Bus fares	48°
Going out	100°
Clothes	120°
Records	60°
Others	32°
Total of angles	360°

(b)



(c) $\frac{1}{3}$

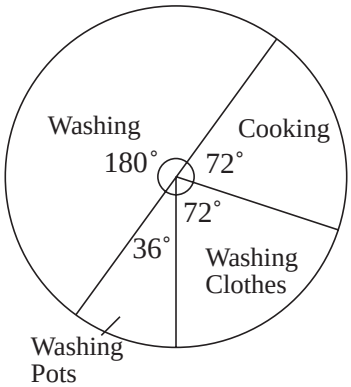
10. (a)



(b) Arthur spends more money than average on housing, less than average on food, less than average on clothes. Allow other valid answers.

11. (a) (i) 81 litres (ii) $\frac{15}{100} = \frac{3}{20}$

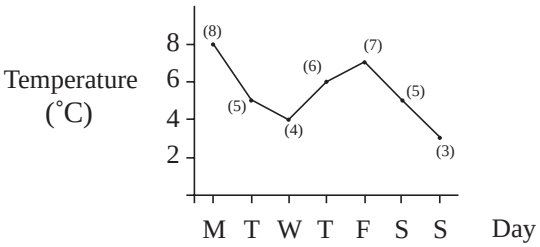
(b)

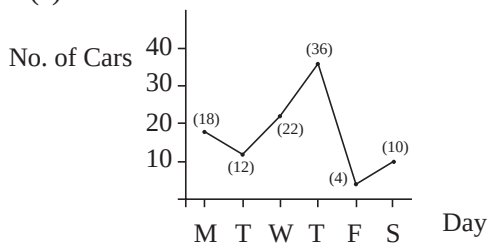


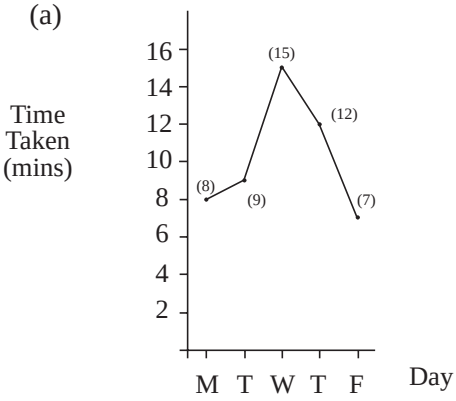
12. (a) (i) Five hundred thousand
(ii) 200 000

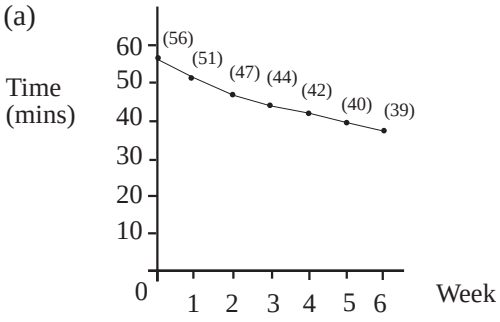
(b) (i) 39% (ii) Over 21 year olds (iii) 0.61 or $\frac{61}{100}$

2.5 Line Graphs

1. (a) 40°C (b) 80°C (c) 20 mins after filling the mug (d) 25 mins
   (e) Approximately 20° to 25° (slightly above room temperature)
   (f) The rate of fall in temperature is decreasing with time. Allow other valid answers.
   (g) No, it will approach room temperature.
2. Discuss all answers to Q2 with students.
   (a) 8 cm (b) 22 cm (c) 84 cm
   (d) Between 85 cm and 90 cm (e) 3 weeks
3. 

4. (a) 

   (b) Use pupil responses as the basis for discussion; possible points include different times when she starts off to work, road works, traffic congestion, flexible working (on Friday), shoppers (on Saturday), etc.
5. (a) 

   (b) Use pupil responses as the basis for discussion; possible points might include poor weather on Wednesday and Thursday, waiting for a friend, going to a shop *en route* to school, etc.
6. (a) 

   (b) His training times are coming down as he becomes more fit.
   (c) Approximately 35 minutes. Allow other valid answers.

2.6 Obtaining Data from Graphs and Charts

1. (a) 40 (b) 30 (c) 180 (d) 9 (e) £7500

2. (a) About 260 kilojoules (b) About 10 000 kilojoules

   (c) As people grow older they need less energy. Allow other valid answers.

3. (a) 1 800 000 (b) UK (c) 200 000

   (d) There are far fewer teachers in Somalia than in Vietnam.

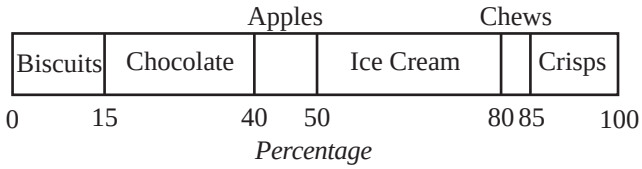
4. (a) 28/29 (b) 6 (c) 6

5. (a) 370 000 (million) (b) 15 days

   (c) China has a very large population.

6. (a) 6 (b) 36 (c) 120

   (d)



   (15%, 25%, 10%, 30%, 5%, 15%)

7. (a) 65 - 70 (b) (i) 10-15 (ii) about 14% (c) about 6.5%

   (d) e.g. In India, very few people live to over 70; in UK, significantly more females than males live beyond 70. Allow other valid answers.

2.7 Misleading Diagrams

1. (a) £8562

   (b) Each disc for Fairplan is worth £7132, which is far less than £8562.

2. Vertical axis does not start at zero; houses not in proportion to % values. Allow other valid answers.

3. (a) The vertical axis starts at 5 which gives a misleading impression of the overall trends; also the 'Western Europe' group looks similar to the other groups whereas it is in fact much larger. Also vertical axis is labelled 'Number of Visitors' rather than 'Number of Visitors in millions'. Allow other valid answers.

   (b) 1985 (c) 7 (million) (d) 1989

4. (a) 1983 (b) Just over 6 bn Fr.

   (c) All improving with SNCF continuing to be more productive than DB, and DB more productive than BR. Allow other valid answers.

   (d) The vertical scale does not start at zero, and hence exaggerates productivity, especially for SNCF, making it appear far more productive (compared to DB and BR) than it actually is. The 3-D perspective further distorts the picture and the comparisons to be drawn from the diagram. Allow other valid answers.

2.8 Frequency Graphs

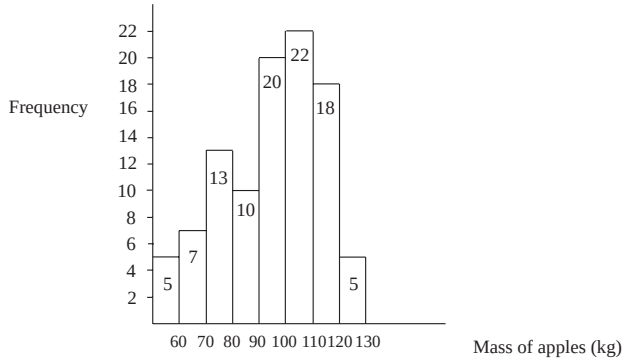
1. (a) 5 (b) 55 (c) 15 (d) 81 pupils in year group

2. (a) 10 (b) 49 (c) 73 (d) largest = £549.99, smallest = £50

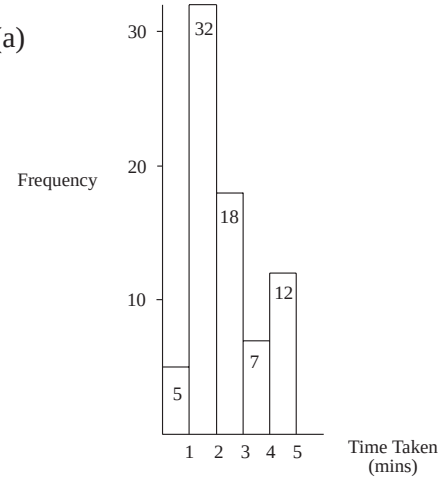
   (e) Shop floor workers, middle management, senior managers

2.8

3.

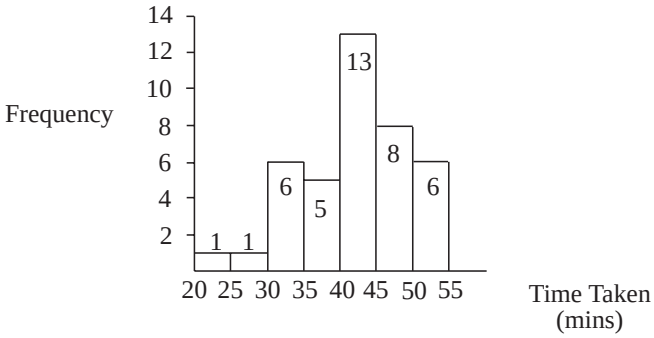


4. (a)

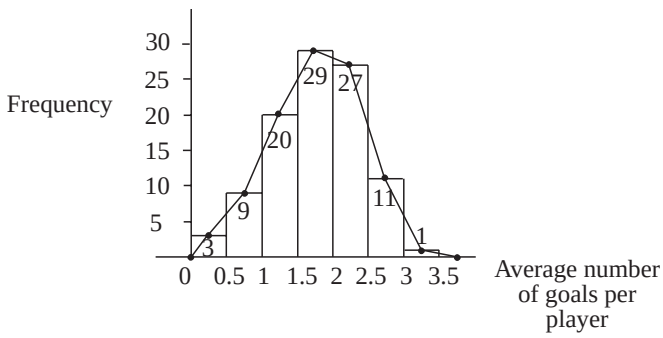


(b) The diagram shows that the people generally solved the puzzle reasonably quickly (under 2 minutes) but some found it challenging (4 minutes or more). Allow other valid answers.

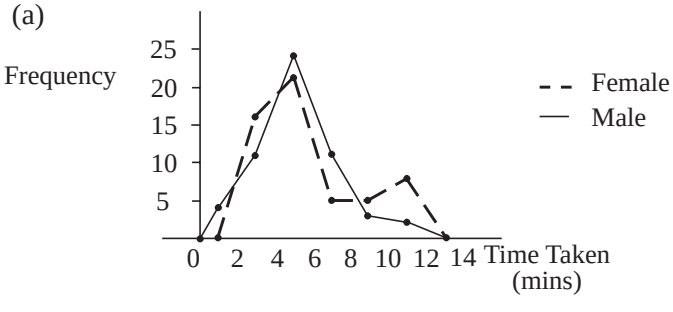
5.



6.



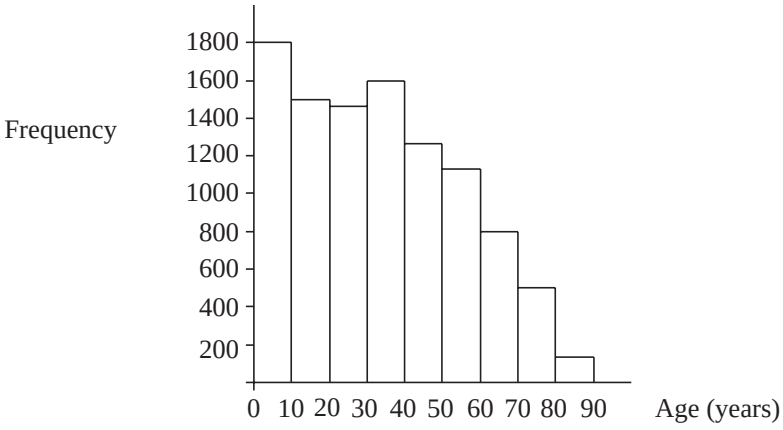
7. (a)



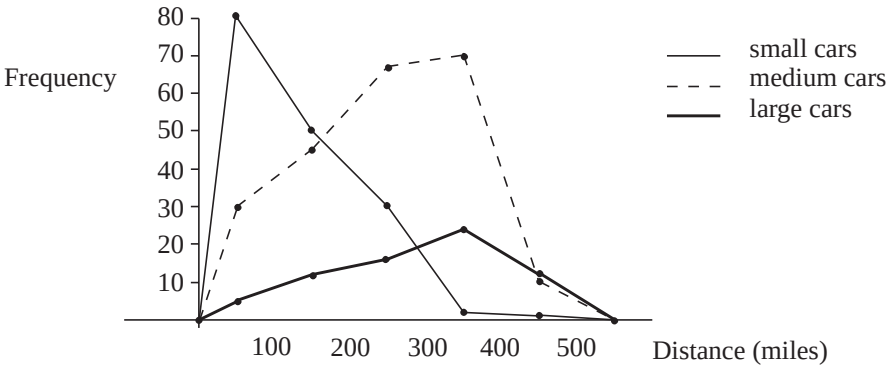
(b) Any reasonable answer (e.g. larger number of females taking longer than 8 minutes to find a space and park; no females parking very quickly (2 minutes or quicker).

2.8

8.

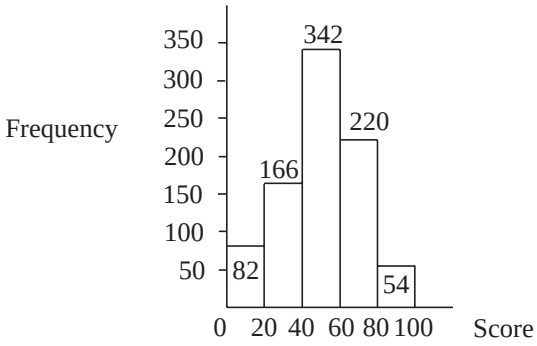


9. (a)



(b) People tend to use small cars for short journeys and larger cars for long distances; medium cars are used for all types of journey. Allow other valid answers.

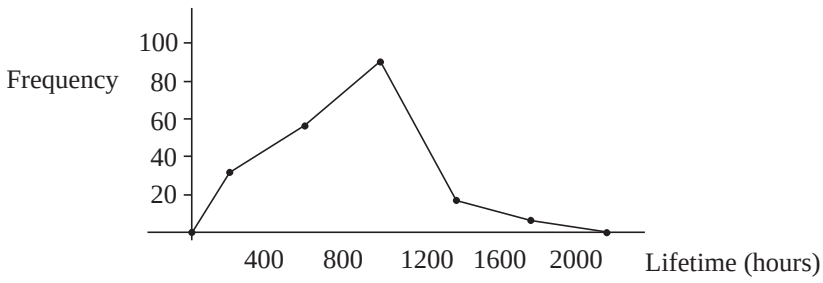
10.



11. $62 + 70 = 132$ pupils

12. (a) (i) probability $= \frac{32}{200} = 0.16$ (ii) prob. $= \frac{106}{200} = 0.53$

(b)



2.8

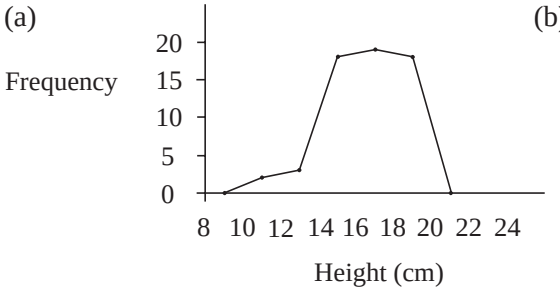
13. (a)

Height	*Frequency*
120 - 130	7
130 - 140	22
140 - 150	20
150 - 160	4

(b) The columns should be adjacent (no gaps).
The height of the second column $(130 \le h < 140)$ should be 22 not 24.

(c) The pupil should be placed in the last category, i.e. $150 \text{ cm} \le h < 160 \text{ cm}$.

14. (a)



(b) Type B has an even spread of plants heights between 10 cm and 20 cm, whereas Type A plants tend to be mainly towards the top end of the same height range.

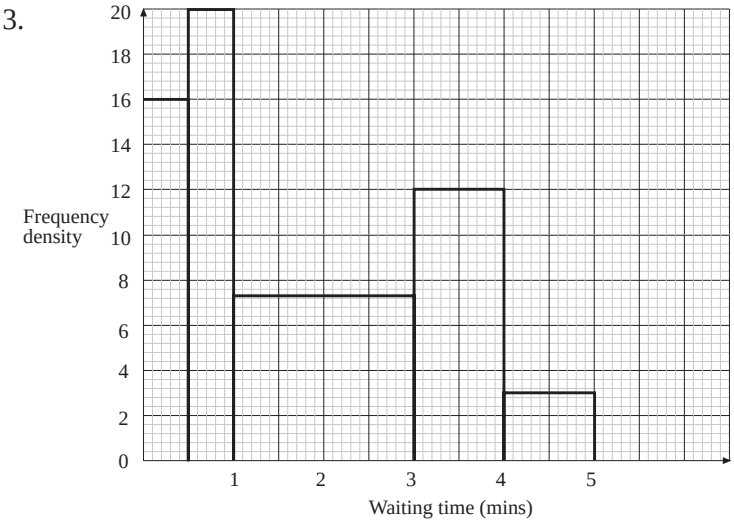
Typical Type B plants have height 15 cm whereas typical Type A plants have height 17 cm, i.e. Type A plants are generally 2 cm taller. Allow other valid answers.

2.9 Histograms

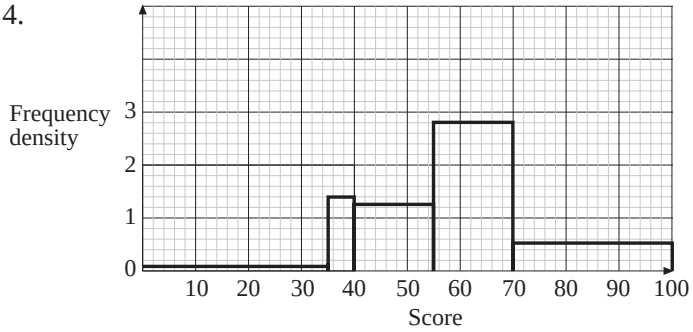
1. (a) 9 (b) 4 (c) 23

2. The area for the 20 - 25 interval is $0.9 \times 5 = 4.5$. This should be a whole number.

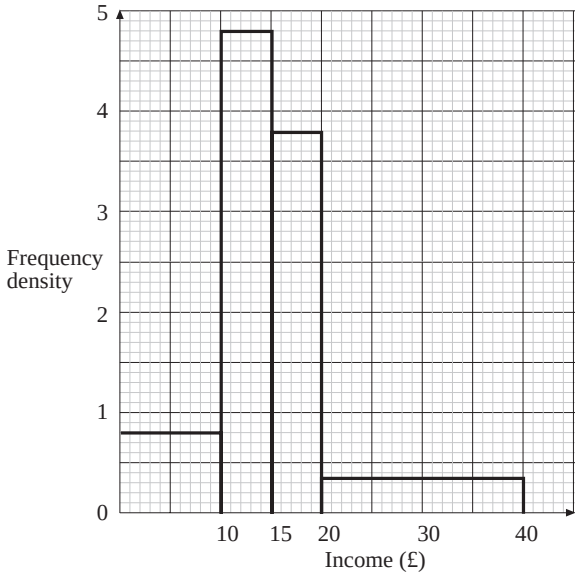
3.



4.



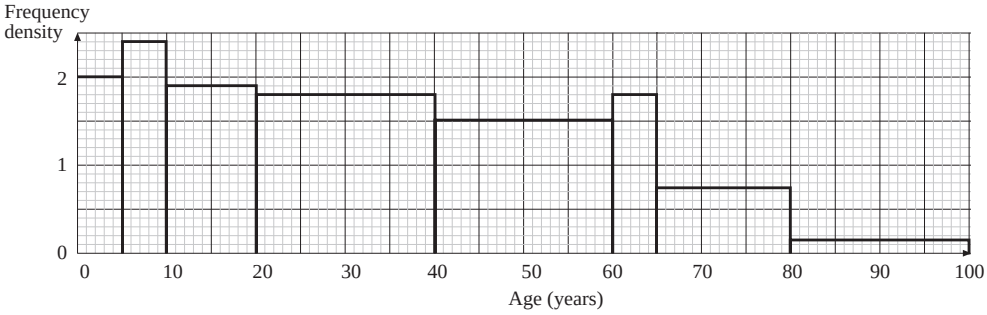
5.



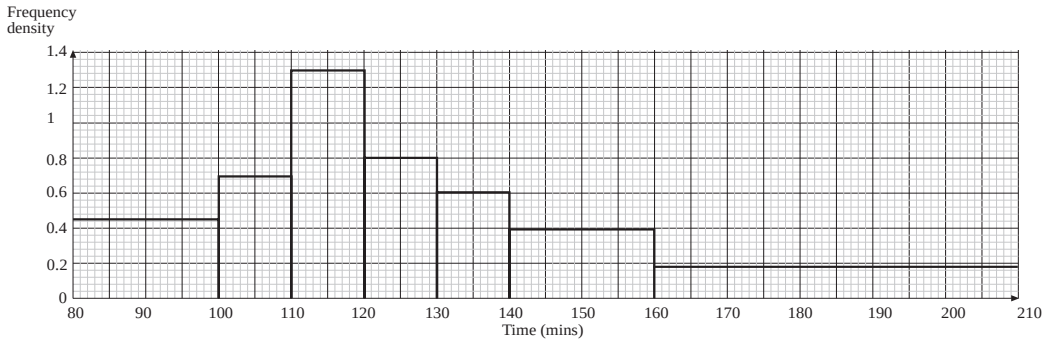
6. (a) The width is 5 because age 0–4 is actually age 0 up to age 5, i.e. $0 \leq \text{age} < 5$ years, giving class width 5. Allow other valid answers. Discuss with students.

(b) The widths are (from the 2nd interval) - 5, 10, 20, 20, 5, 15, 20.

(c)



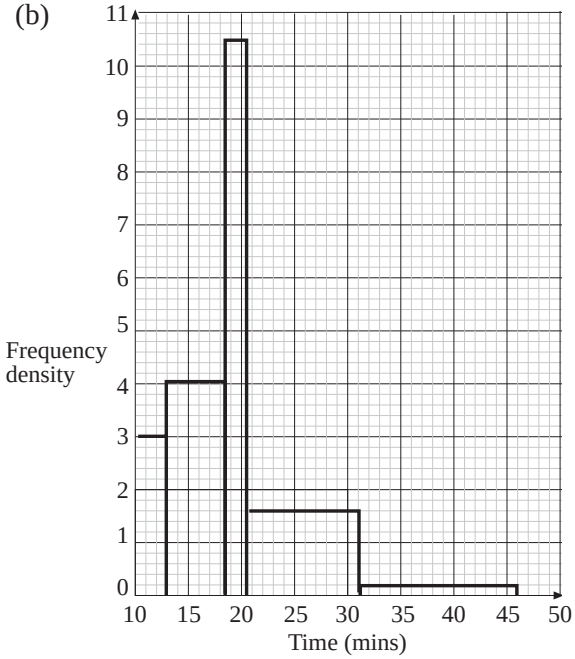
7.



Discuss with students.

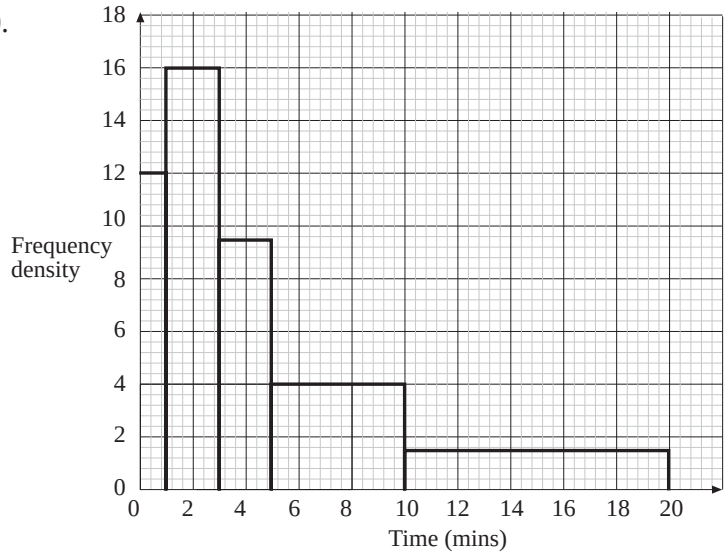
8. (a) Width of class intervals: 3, 6, 2, 10, 15, e.g. the first class is 9.5 min ≤ time < 12.5 min).

(b)



9. (a) 4 people (b) 23 people

10.



11. (a) Frequencies- 4, 6, 7, 4, 6, 1 (b) 28 matches

3.1 Mean, Median, Mode and Range

1. (a) Mean = 5 Median = 4.5 Mode = 3 Range = 8
   (b) Mean = 9.9 Median = 9.5 Mode = 7 Range = 11
   (c) Mean = 16.25 Median = 16 5 Mode = 17 Range = 10
   (d) Mean = 107.6 Median = 108 Mode = 108 Range = 13
   (e) Mean = 63 Median = 64 Mode = 61 Range = 10
   (f) Mean = 21.75 Median = 21.5 Mode = 16 Range = 14
2. (a) 6.8 (b) 6.75 (c) 6 (d) $5\frac{1}{2}$
3. (a) (i) £4.875, i.e. £4.88 to the nearest penny (ii) £3.50 (iii) £3
   (b) (i) Mean (ii) Mode (c) Range = £13
4. (a) Fred: mean = 20.8, range = 2. Harry: mean = 24.2, range = 24.
   (b) Fred (c) Harry
5. (a) A: mean = 15.9, median = 20, mode = 20
   B: mean = 16.9, median = 17, mode = 18
   (b) Mode suggests A (c) Mean suggests B (d) Range: A = 15, B = 3.
6. (a) Mean = 68, median = 68.5 (b) His mean increases to 68.4.
   (c) The median. It increases from 68.5 to 70, whereas the mean only increases by 0.4.
7. (a) He objects because the mode = 0 = median.
   (b) Mean = 2.30, range = 18 (c) 15 fish
8. (a) 268.4 cars (b) The mean decreases (to 266.9).
9. (a) Mean = 2.036, median = 2, mode = 2.
   (b) Either median or mode (whole numbers). Allow other valid answers.
10. The mean will increase.
11. (a) 19°C (b) Archangel (c) 27°C
12. 225 grams
13. (a) Mean = 2 (b) Range = 4
14. (a) 77 kg (b) Hereward House, because they have a much heavier team. Discuss with students.
15. (a) Modal class = 24 pupils (b) Mean class = 26 pupils
    (c) There is a more consistent class size in Year 9.
16. (a) Pat: mean = 25.3, range = 18
    (b) They both have approximately the same mean, but Kim's scores are more consistent as shown by the smaller range; thus Kim should be selected. Discuss with students.
17. (a) (i) 5.6 (ii) 0.5 (b) Leaving out the two extreme marks probably gives a less biased measure of performance. Discuss with students.
    (c) Maximum score = minimum + range which must be less than mean + range = 5.2 + 0.6 = 5.8, therefore no mark could possibly exceed 5.8.

3.1

18. (a) 41 (b) 41 (c) 53

(d) Overall, he scored lower. Discuss with students.

19. (a)

3	6 7 7 7
4	2 2 6 8 9
5	1 1 3 3 4 5 6 7 7 8
6	0 2 3 4 8
7	0

(b) (i) 53 (ii) 37 (iii) $70 - 36 = 34$

(c) The median better represents the average, the mode being an unrepresentative low value. Discuss with students.

20. (a)

Number of passengers	*Frequency*
5 – 9	2
10 – 14	3
15 – 19	2
20 – 24	5
25 – 29	2
30 – 34	1

(b) $\frac{2}{15}$ (c) $\frac{8}{15}$

3.2 Finding the Mean from Tables and Tally Charts

1. Mean = 1.25 cars per household
2. Mean = 1.93 TV sets per household
3. Mean = 4.08 calls per day
4. (a) Mean = 3.56 runs per over (b) About 142 runs
5. (a) Mean = 1.95 worms (b) 22 times. Discuss with students.
6. Mean number = 1.15 trains
7. Mean = 2.295
8. (a) 6 (b) 2.04
9. Missing frequencies are 1, 5, 0.
   Missing frequencies $\times$ number of tickets are 0, 20, 21, 10, 0.
   Mean = 2.8 (1 d.p.)
10. (a) 3 (b) Frequency = 21, total = 48, mean = 2.29.

    (c) The number of children per family has decreased on average (2.29 < 2.7), and there is less variation from family to family (today's range is 3, whilst in 1960 it was 7). Discuss with students.

3.3 Mean, Median and Mode for Grouped Data

1. (a) 32.09 (b) 30 - 39 (c) 30 - 39

2. (a) 40.80 (b) $40 \le w < 45$ kg (c) $40 \le w < 45$

3. (a) $20 \le t < 25$ seconds

   (b) Yes. Mode and median in class interval $20 \le t < 25$ seconds. Discuss with students.

   (c) 20.54 seconds

4. (a) No. Modal class $0 \le d < 0.5$, median class $0.5 \le d < 1.0$. Discuss with students.

   (b) Mean = 0.78.

5. 11.95 years

6. (a) 11.95 cm (b) 11 - 15 cm

7. (a) 9.65 nights (b) 6 - 10 nights (c) 11 - 15 nights

8. (a) (i) 26.8 (ii) 21 - 30 (iii) 21 - 30

   (b) (i) 21.5 (ii) 21 - 30 (iii) 21 - 30

   (c) The second class have a lower mean so the first class did better on the test. Allow other valid answers. Discuss with students.

9. (a) 0 - £1.00 (b) £1.43

10. (a) 21, 7, 2

    (b) People would spend more time watching television than in summer. Allow other valid answers. (c) 24.04 (24 hours)

11. (a) 23.5 - 24.5 cm (b) 24.32 cm

    (c) The first group tended to give higher estimates for the length of the rod (modes 25 cm and 23 cm). The first group were more consistent in their estimates (ranges 7 cm and 13 cm). Allow other valid answers.

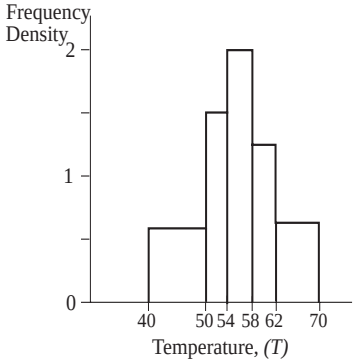
12. (a)

Temperature , T	*Mid-point*	*Frequency*
$40 < T \le 50$	45	6
$50 < T \le 54$	52	6
$54 < T \le 58$	56	8
$58 < T \le 62$	60	5
$62 < T \le 70$	66	5

(b) $\frac{1660}{30} \approx 55.3$

(c)

Interval	*Freq. Density*
40-50	$\frac{6}{10} = 0.6$
50-54	$\frac{6}{4} = 1.5$
54-58	$\frac{8}{4} = 2$
58-62	$\frac{5}{4} = 1.25$
62-70	$\frac{5}{8} = 0.625$

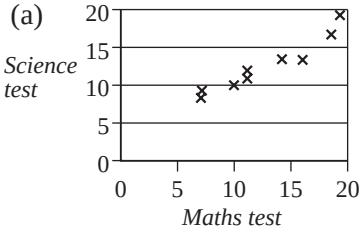


3.4 Calculations with the Mean

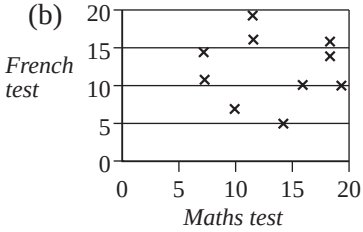
1. Mean = 161.55
2. Mean = 2
3. Mean = 4
4. Mean = 60.15 kg
5. 7
6. 84%
7. £6000
8. 9.5
9. Mean $\approx$ 4.47
10. 320

3.5 Scatter Plots and Lines of Best Fit

1. (a)



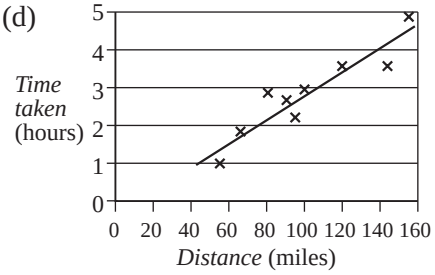
(b)



(c) Maths and Science

(d) Not for Maths and French because there is no evidence of correlation.

2. (a) and (d)

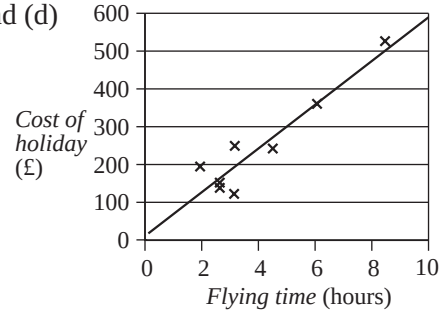


(b) Positive correlation. This means that the further the journey, the longer it takes; the shorter the journey, the less time it takes. (c) 100 miles, 2.81 hours

(e) 70 miles (f) 3.9 hours

(g) Because 300 miles is beyond the range of the original data and such a long journey would involve rest periods.

3. (a) and (d)

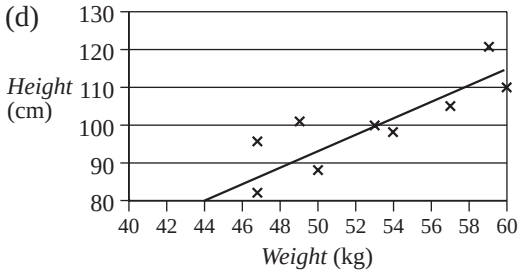


(b) Positive correlation. This means that holidays with longer journey times tend to cost more.

(c) 3.83 hours, £235 (e) £300 (f) 6.5 hours

(g) Because £1000 is well beyond the range of the original data and such a holiday might be to an exotic location or have special features.

4. (a) and (d)

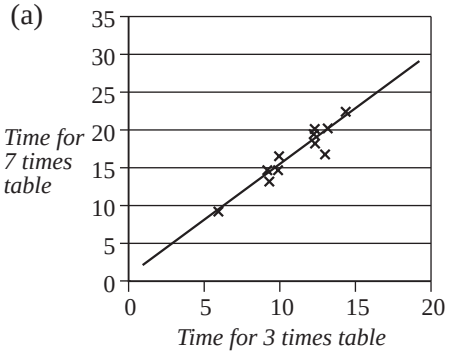


(b) Positive correlation. This means that taller children tend to weigh more, shorter children tend to weigh less.

(c) 52.2 kg, 97.9 cm

(e) The line of best fit can be applied to this data because there is evidence of positive correlation. (f) 110 - 115 cm (g) 57 - 60 kg

5. (a)



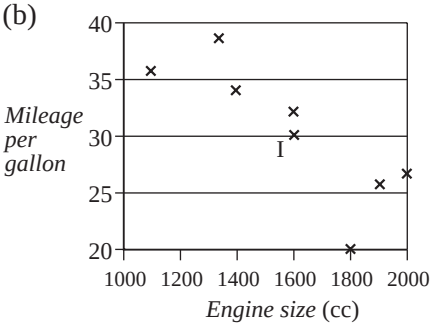
(b) 16 - 17 seconds

(c) 7 - 8 seconds

(d) Because the estimates would be based on a larger sample which should mean a more accurate line of best fit.

6. (a) E : 1600 cc ; F : 26 miles per gallon ; H : 2000 cc, 27 miles per gallon

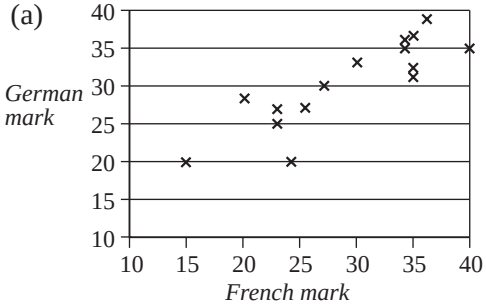
(b)



(c) G, because it is well below any line of best fit.

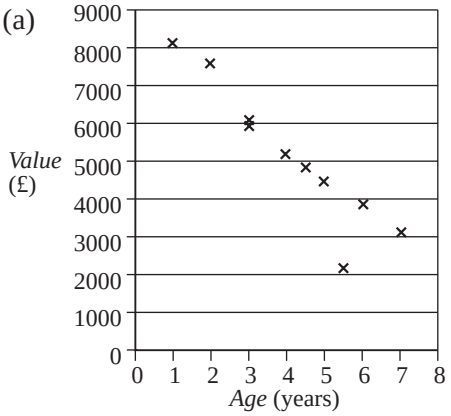
7. Any reasonable answers - use them to promote class discussion.

8. (a)



(b) There is evidence of positive correlation, i.e. those who did well/ poorly in one test tended to do well/poorly in the other test.

9. (a)

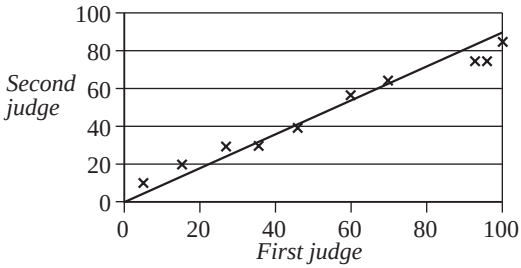


(b) There is evidence of negative correlation, i.e, the older the car the lower its value.

(c) The car that is $5\frac{1}{2}$ years old and valued at £2200 does not follow the trend.

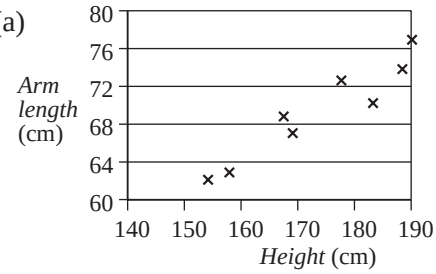
(d) Its low value for its age may be because it has very high mileage or because it has been involved in a major accident. Allow other valid answers.

10. (a) (i) and (ii)



(b) About 67 - 68 marks

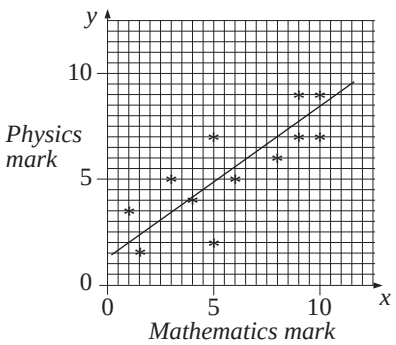
11. (a)



(b) (i) About 68.5 cm

(ii) The scatter diagram shows evidence of positive correlation but not strong positive correlation. The estimate obtained from a line of best fit only gives a guide as to Peter's arm length. There will be considerable variation in the body dimensions of children all of whom have the same height. Discuss with students.

12. (a) and (c) (i)



(b) Yes; the scatter diagram shows positive correlation, which is what we would expect because students who are good/poor at Mathematics are generally good/poor at Physics.

(c) (ii) 6 or 7 marks

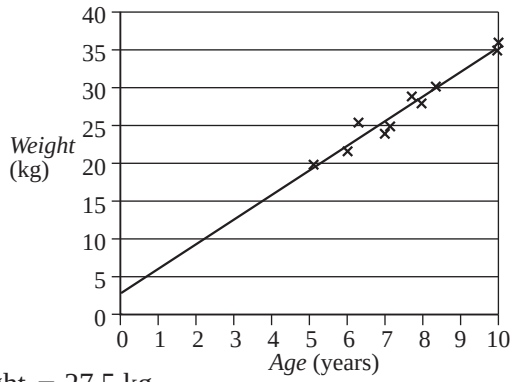
(iii) The student who scored 10 for Mathematics and 9 for Physics.

13. (a) Negative correlation: i.e. the higher the average temperature, the lower the total rainfall; the lower the average temperature, the greater the total rainfall.

(b) 115 - 120 cm

3.6 Equations of Lines of Best Fit

1. (a) and (c)



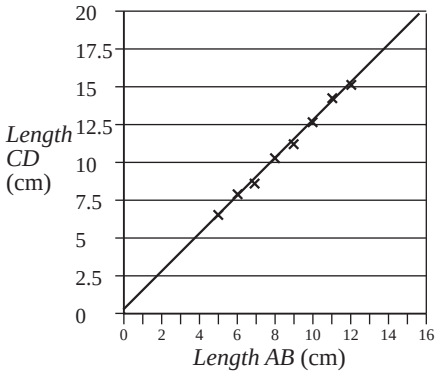
(b) Mean weight $= 27.5$ kg

(d) $y = 3.2x + 3$ (N.B. The calculated regression line has equation $y = 3.23x + 2.952$)

(e) 28.6 kg (N.B. The regression line gives 28.8 kg)

(f) Because 12 years of age is beyond the range of the data used to generate this line of best fit. Also because other physical characteristics (e.g. puberty) start to affect weight around the age of 12 years.

2. (a) and (c)



(b) Mean length of AB $= 8.5$ cm.

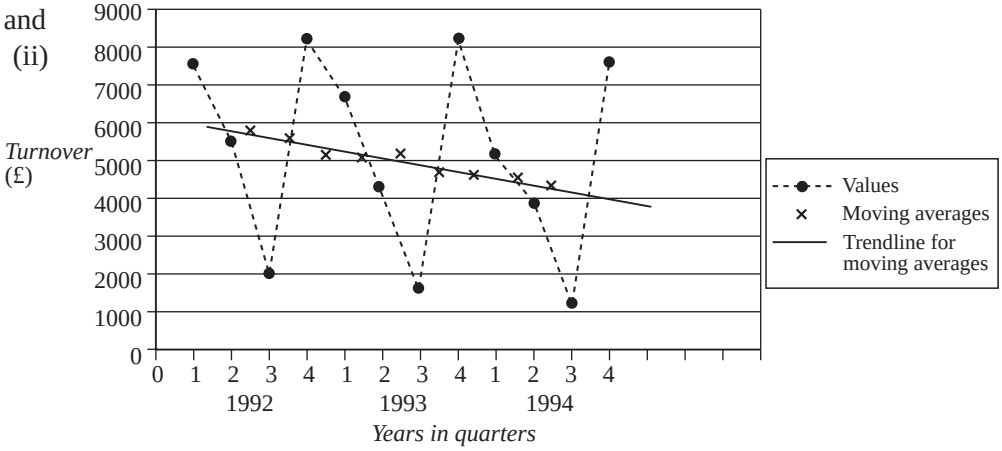
(d) $y = 1.3x$ (N.B. The calculated regression line has equation $y = 1.257x + 0.139$)

(e) 8.45 cm (N.B. The regression line gives 8.3 cm)

3.7 Moving Averages

1. (a) 1992 (b) Closed for summer holidays

(c) and
(d) (ii)

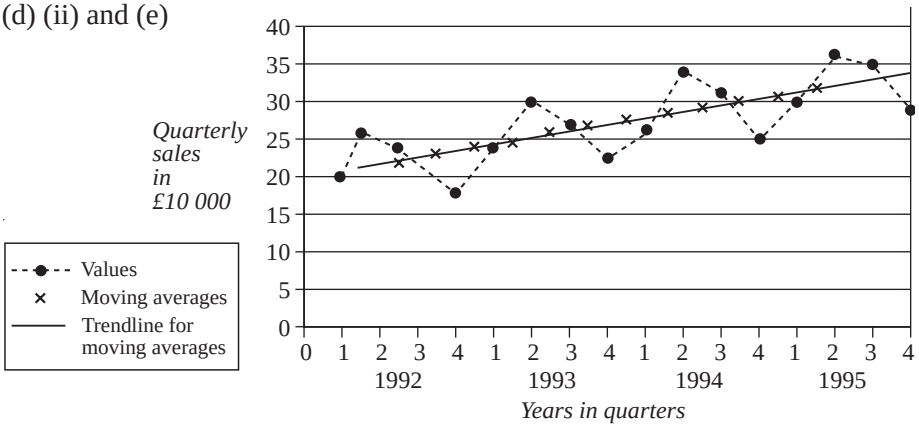


(d) (i)

Quarter	Turnover	Moving average
1	7500	
2	5500	
3	2000	5800
4	8200	5600
1	6700	5300
2	4300	5200
3	1600	5200
4	8200	4800
1	5100	4700
2	3900	4600
3	1200	4425
4	7500	

(iii) The trendline indicates that the underlying trend is downward.

2. (a), (d) (ii) and (e)



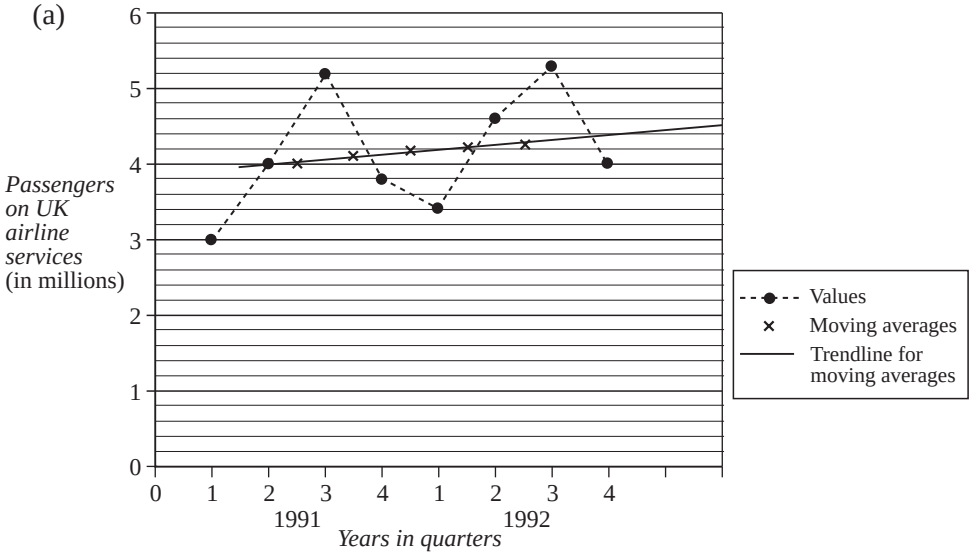
(b) Spring/summer time gives a boost to sales.

(c)

Year	Quarter	Sales £10 000's	Four-point moving average
1992	1	20	
	2	26	
	3	24	22
	4	18	23
1993	1	24	~~24~~
	2	30	24.75
	3	27	26
	4	23	26.5
1994	1	26	~~27.5~~
	2	34	28.5
	3	31	29
	4	25	30
1995	1	30	~~30.5~~
	2	36	31.5
	3	35	32.5
	4	29	

(f) £320 000 to £330 000 (Explanation: the trendline predicts a value of 33.25 for the next value of the moving average (for quarters 2, 3 and 4 of 1995 and quarter 1 of 1996). Hence the predicted total for that period is $4 \times 33.25 = 133$. Subtracting the figures for quarters 2, 3 and 4 of 1995 gives a prediction of $133 - (36 + 35 + 29) = 33$ for the first quarter of 1996, meaning sales of £330 000.)

3. (a)



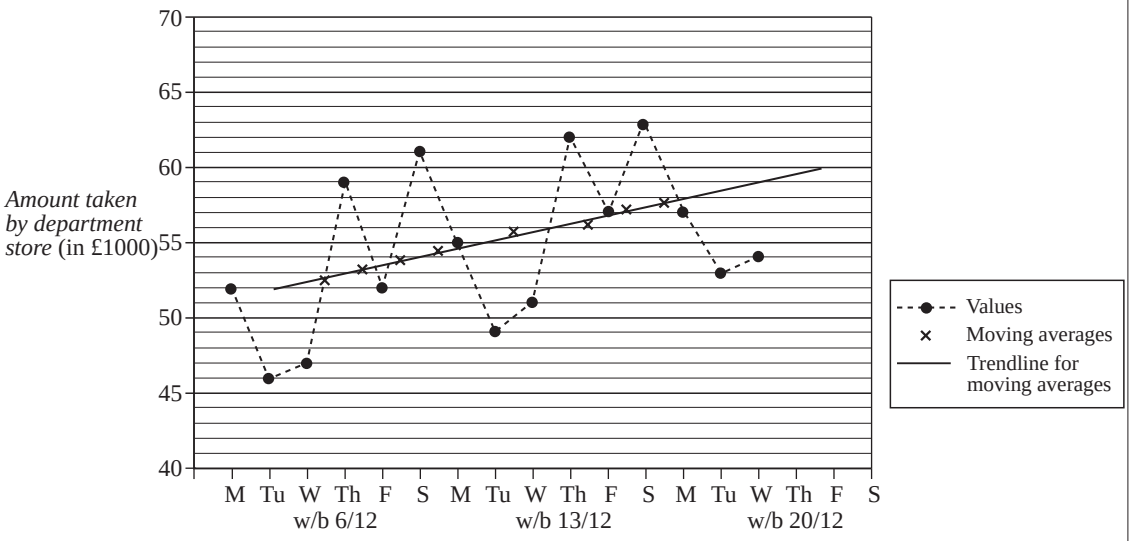
(b) The first quarter is always the lowest figure (people often prefer to stay at home in the winter months); the third quarter is always the highest figure (summer holidays); the second quarter (spring) is always higher than the fourth quarter (autumn). Allow other valid answers.

(c) The trendline for the moving average predicts passenger numbers of 4.36, 4.43 and 4.50 (million) in the first, second and third quarters of 1993. Hence the predicted total for quarters 2, 3 and 4 of 1992 and quarter 1 of 1993 is $4 \times 4.36 = 17.44$. Subtracting the figures for quarters 2, 3 and 4 of 1992 gives a prediction of $17.44 - (4.4 + 5.3 + 4.0) = 3.74$ for the first quarter of 1993. Repeating this process for the next quarter gives a predicted total for quarters 3 and 4 of 1992 and quarters 1 and 2 of 1993 $= 4 \times 4.43 = 17.72$. Subtracting the figures for quarters 3 and 4 of 1992 and the predicted figure of 3.74 for quarter 1 of 1993 gives a prediction of $17.72 - (5.3 + 4.0 + 3.74) = 4.68$ for the second quarter of 1993. Repeating this process one more time gives a predicted total for quarter 4 of 1992 and quarters 1, 2 and 3 of 1993 $= 4 \times 4.50 = 18$. Subtracting the figure for quarter 4 of 1992, and the predicted figures for quarters 1 and 2 of 1993, gives a prediction of $18 - (4.0 + 3.74 + 4.68) = 5.58$ for the third quarter of 1993.

Therefore, the predicted passenger numbers for the third quarter of 1993 (i.e. for July - September 1993) is 5.6 million. One third of this figure exceeds the quoted figure of 1 800 000 for the single month of July. Furthermore, it is likely that there will be more passengers in July and August than in September because of school holidays, so we would expect July's figure to comfortably exceed one third of the predicted figure. The newspaper claim is therefore incorrect. Allow other valid answers. Discuss with pupils.

3.7

4. (a) and (e)



(b) In the first two weeks, the amounts taken on Thursday are higher than the amounts taken on each of the other weekdays apart from Saturday when shops do not normally open for late-night shopping. Allow other valid answers.

(c) The 6-day moving averages (to 1 decimal place) are 52.8, 53.3, 53.8, 54.5, 55.0, 55.8, 56.2, 56.5, 57.2 and 57.7.

(d) A 6-point moving average is appropriate in this case because the store is only open 6 days a week and we would therefore expect store takings to follow a 6-day cycle. Allow other valid answers.

(f) The graph of the moving average shows that the store's takings are rising reasonably steadily during the run-up to Christmas.

(g) The trendline predicts takings of 58.3 for the next value of the moving average. Hence the predicted total for Friday and Saturday of the week beginning 13 December and Monday to Thursday of the week beginning 20 December is $6 \times 58.3 = 349.8$. Subtracting the figures for the 5 days mentioned above gives a prediction of $349.8 - (57 + 63 + 57 + 53 + 54) = 65.8$. Since the figures in the original table have been given as integers, we round this value to 66. The predicted takings for the store is £66 000 on Thursday 23 December. Discuss with students.

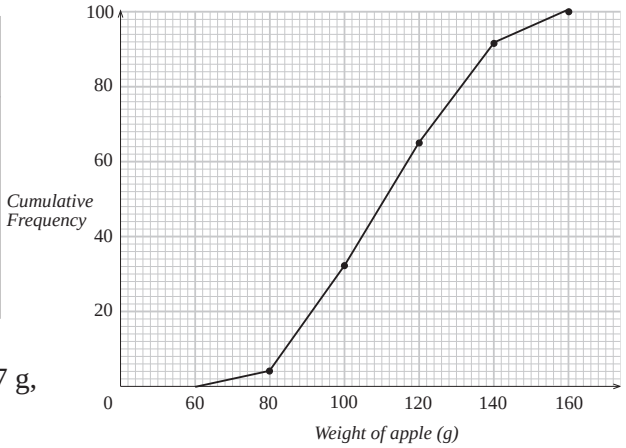
(h) It would be unwise to use moving averages to predict the amount taken by the store on Friday 24 December because that is Christmas Eve and there are a lot of last minute shoppers panic buying, so we would expect the takings on that day to be much higher than on a normal Friday, or perhaps lower, if the shop closes earlier than usual. Allow other valid answers.

The store is likely to be closed on Saturday 25 December, meaning zero takings, which the trendline for the moving average would not predict.

3.8 Cumulative Frequency

1. (a)

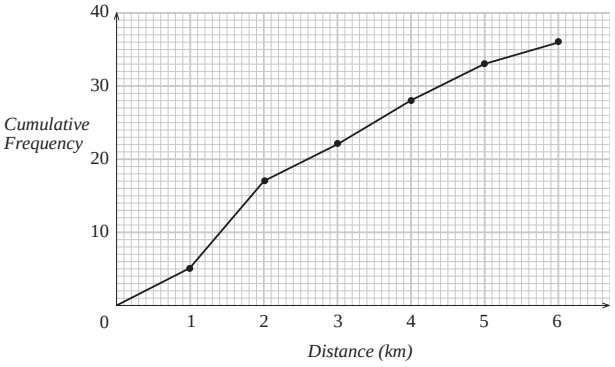
Weight in grams (w)	*Cumulative Frequency*
$w \le 80$	4
$w \le 100$	32
$w \le 120$	65
$w \le 140$	92
$w \le 160$	100



Median = 111 g, LQ = 95 g, UQ = 127 g,
Inter-quartile range = 32 g

(b)

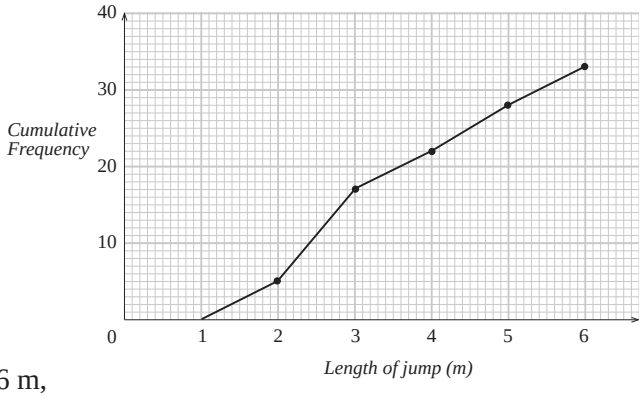
Distance in km (d)	*Cumulative Frequency*
$d \le 1$	5
$d \le 2$	17
$d \le 3$	22
$d \le 4$	28
$d \le 5$	33
$d \le 6$	36



Median = 2.2 km, LQ = 1.3 km, UQ = 3.8 km,
Inter-quartile range = 2.5 km

(c)

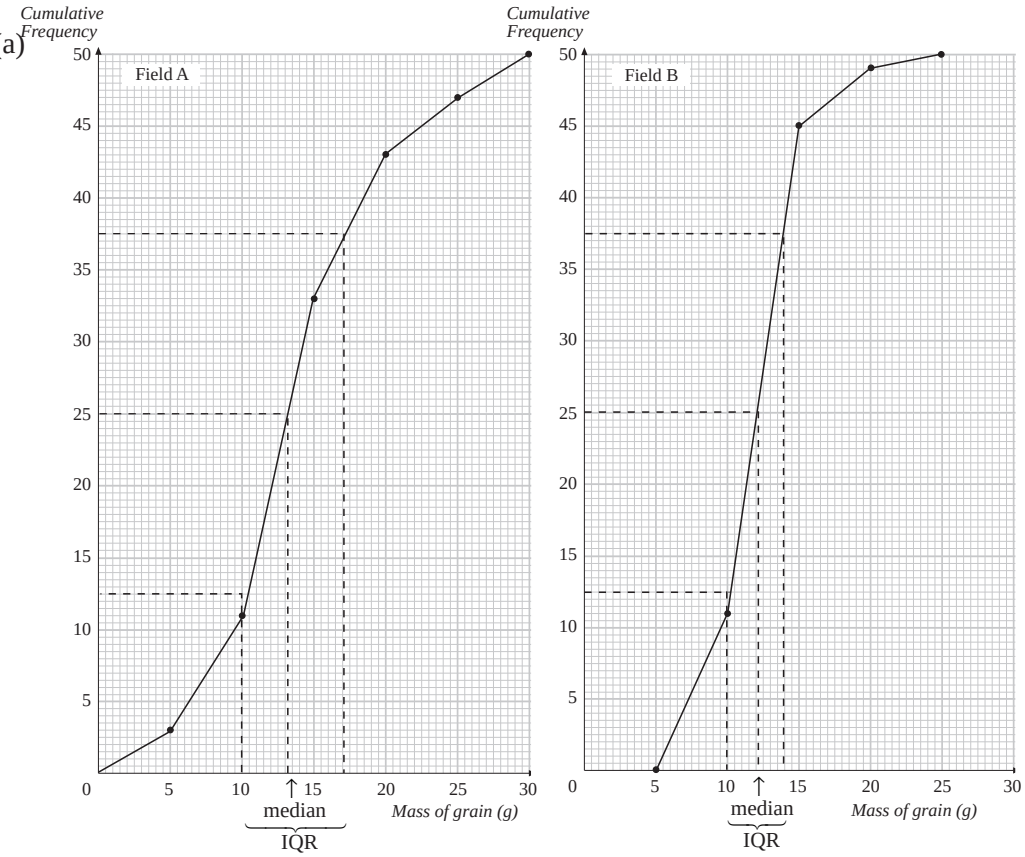
Length in metres (d)	*Cumulative Frequency*
$d \le 2$	5
$d \le 3$	17
$d \le 4$	22
$d \le 5$	28
$d \le 6$	33



Median = 3 m, LQ = 2.3 m, UQ = 4.6 m,
Inter-quartile range = 2.3 m

3.8

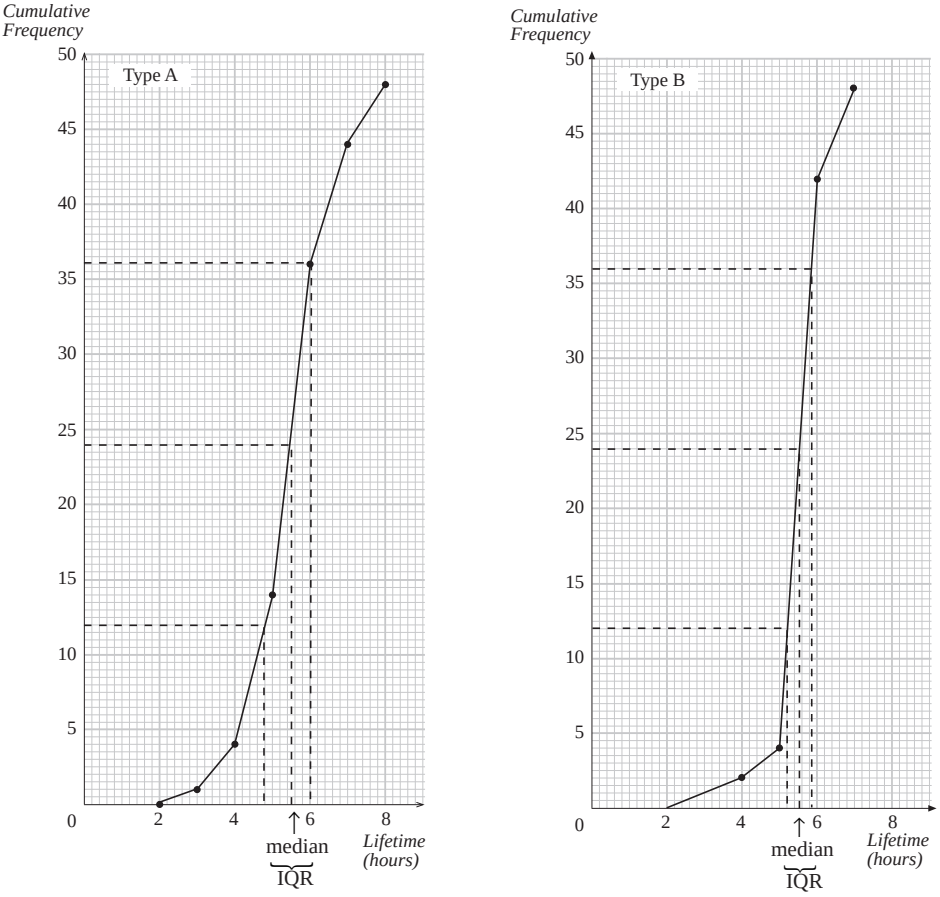
2. (a)



(b) The median of field A is 13.2 g, the median of field B is 12.1 g.
The inter-quartile range for field A is around 6.9 g, for B it is around 3.7 g.

(c) Field B is more reliable than field A (its inter-quartile range is narrower), although it is generally less productive (its median is lower than the median of field B). Allow other valid answers. Discuss with students.

3. (a)

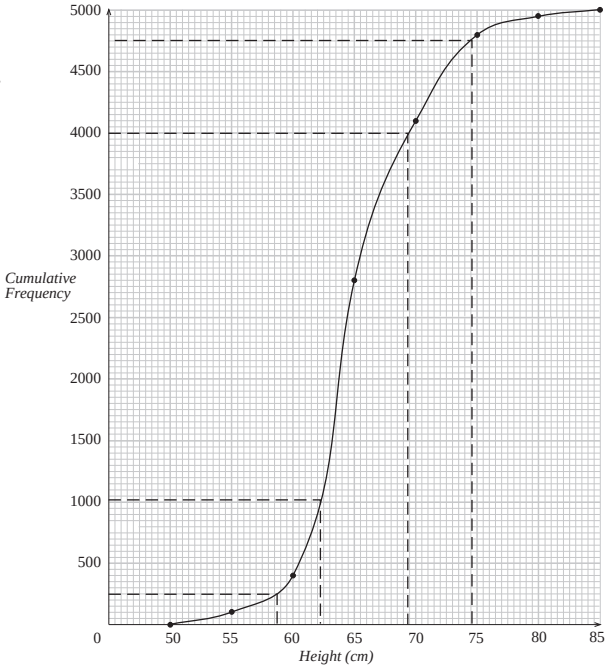


Median for type A = 5.5 hours, Inter-quartile range for type A = 1.2 hours.
Median for type B = 5.5 hours, Inter-quartile range for type B = 0.6 hours.

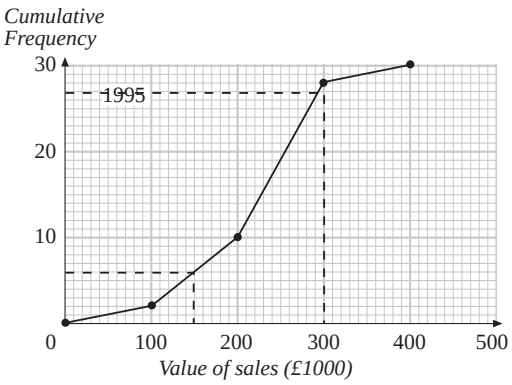
(b) Type B - both types last, on average, for 5.5 hours, but there is less variation in the lifetime for type B, so recommend type B because it is more reliable. Accept other valid answers.

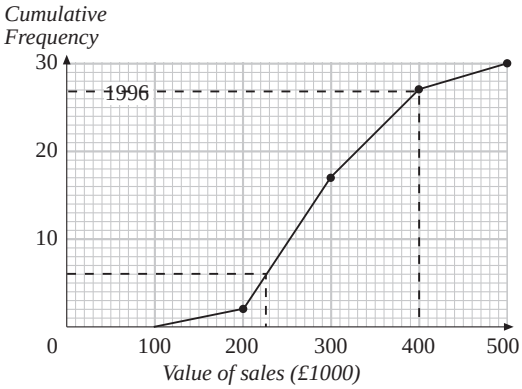
4. N.B. The sample is very large so here we draw a cumulative curve. The heights of children in each category (to the nearest cm) are:

Very tall	-	$75 < h \leq 85$
Tall	-	$70 < h \leq 75$
Normal	-	$62 < h \leq 70$
Short	-	$58 < h \leq 62$
Very short	-	$50 < h \leq 58$



5.





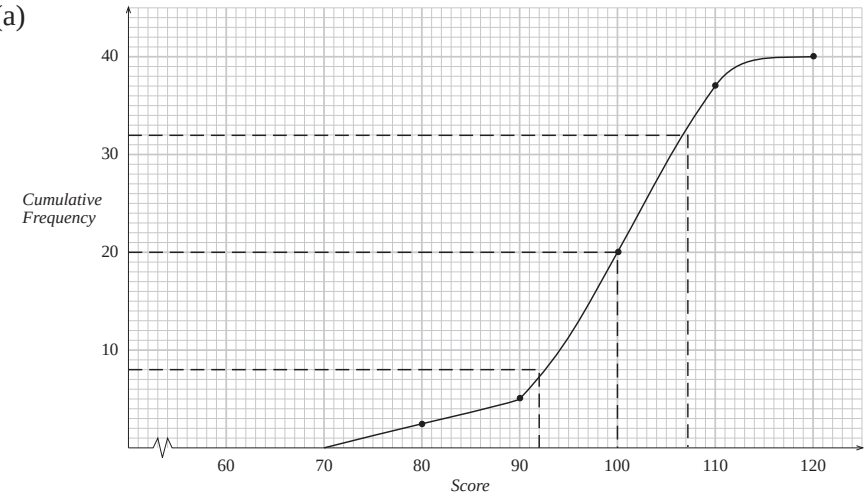
For 1995:

Bonus	*Value of sales*
£50	$0 < V \leq 150$
£250	$150 < V \leq 294$
£500	$294 < V \leq 500$

For 1996:

Bonus	*Value of sales*
£50	$100 < V \leq 227$
£250	$227 < V \leq 400$
£500	$400 < V \leq 500$

6. (a)



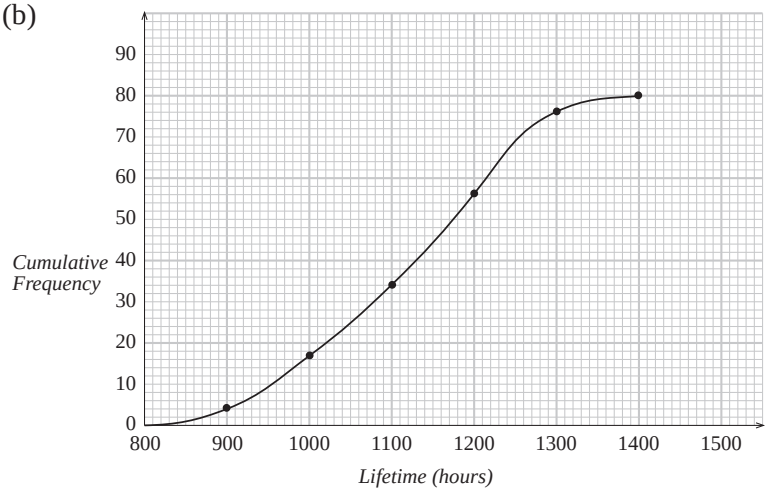
(b) (i) Laura's median = 100 (ii) LQ = 93, UQ = 106,
Inter-quartile range = 13.

(c) (i) Joy was the more consistent player because her inter-quartile range was lower.

(ii) Laura was the better player. Her median (i.e. average) score of 100 was lower than Joy's median score of 103.

7. (a) Cumulative frequencies - 34, 56, 76, 80, 80.

(b)



(c) 58 bulbs (d) LQ = 1020 hours, UQ = 1220 hours,
Inter-quartile range = 200 hours.

(e) The bulbs from the second sample are more reliable than those from the first.

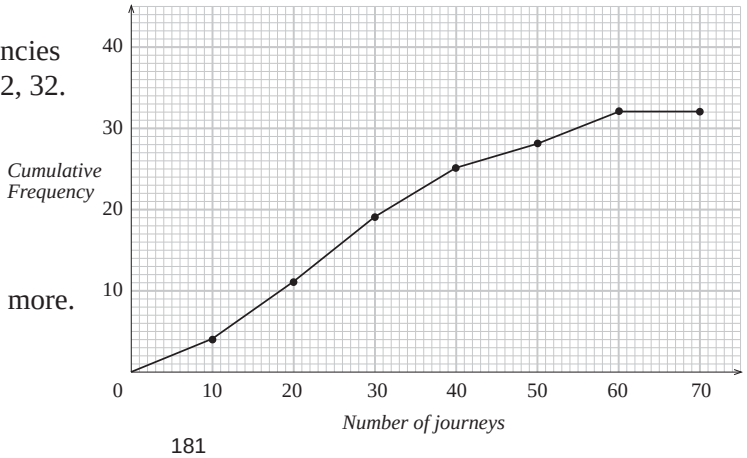
8. (a) Cumulative frequencies
4, 11, 19, 25, 28, 32, 32.

(b) (i) As graph.

(ii) Median = 26

(iii) 6 people

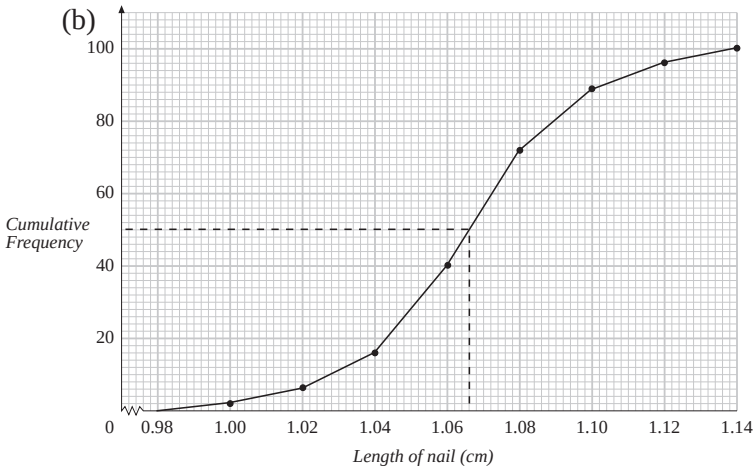
(c) The second group generally travelled more. Accept other valid answers.



3.8

9. (a) Cumulative frequencies: 2, 6, 16, 40, 72, 89, 96, 100.

(b)



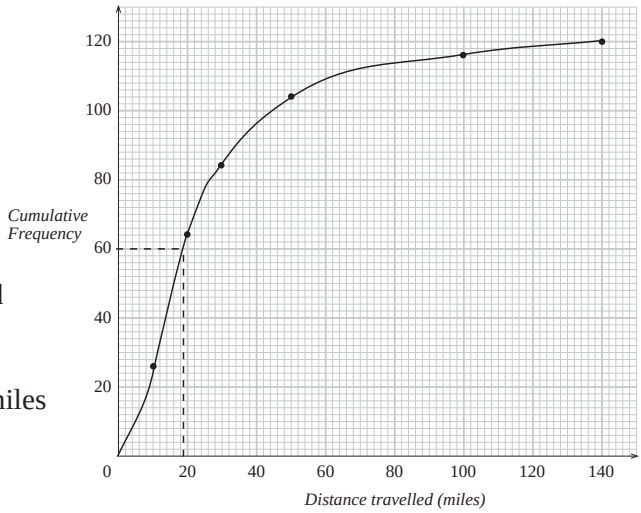
(i) Median = 1.066 cm

(ii) LQ = 1.048 cm, UQ = 1.084 cm, Inter-quartile range = 0.036 cm.

10. (a) Mean distance = 28.17 miles (to 2 d.p.)

(b) (i) Number of guests: 26, 64, 84, 104, 116.

(ii) as graph



(c) (i) Median = 19 miles.

(ii) The range is very large (140 miles), and 36 people (who make up almost a third) travel more than 30 miles (above the median). These upper values influenced the mean which is far above the median. Accept other valid answers. Discuss with students.

3.9 Box and Whisker Plots

1. (b)

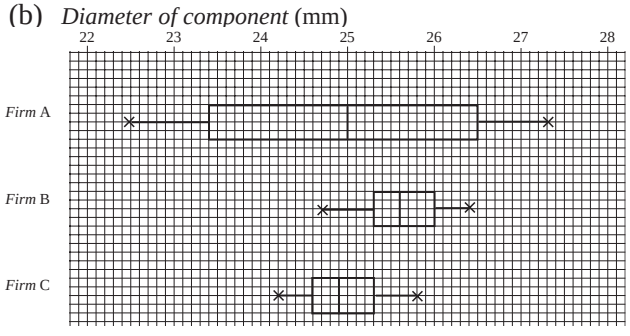
0	3 5 7 8 9 9 9 9
1	0 0 0 0 1 2 3 3 4 5 5 6 6 7
2	0 1 2 3 4

(c) It is easy to construct: it shows the main features of the distribution. Allow other valid answers. Discuss with students.

(d) Minimum value = 3; maximum value = 24; range 21; median = 12; LQ = 9; UQ = 16.

2. (a) Range = 26.4 – 24.7 = 1.7 mm. (ii) IQR = 25.3 – 24.6 = 0.7 mm.

(b)



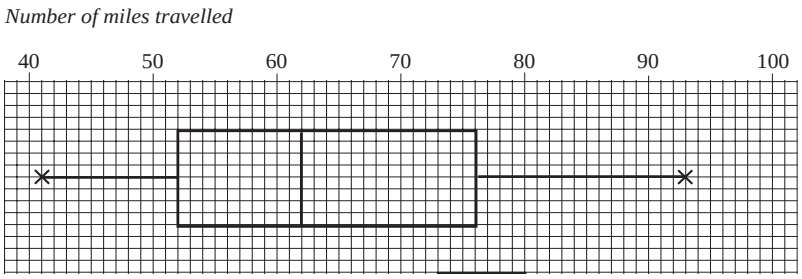
(c) Firm C – as it is the only one to meet the requirements. Discuss with students.

3. (a)

4	1 2 3 4 6 8 8
5	2 2 3 4 6 8 8
6	0 2 6 7
7	0 0 2 4 4 6 7 8
8	0 5 5 6
9	2 3

(b) 64 ; 52, 76

(c)

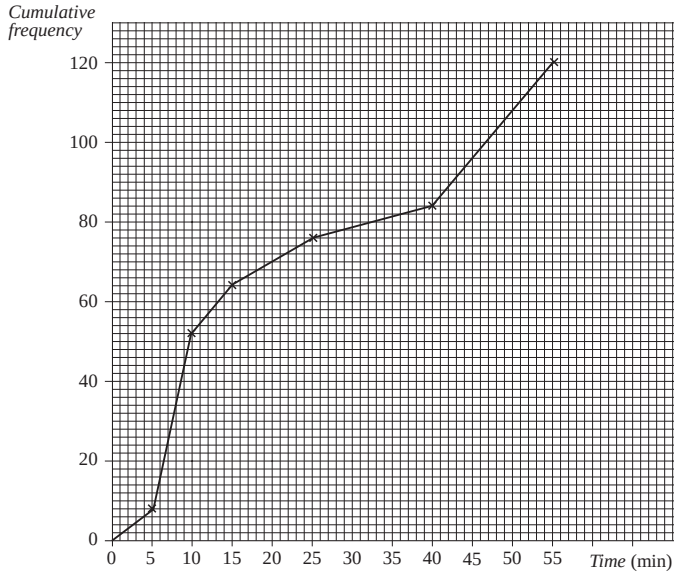


(d) Stem and leaf diagram is easy to construct.
Box plot quickly shows you the main statistical features.
Both diagrams show the way the data is distributed. Allow other valid answers.

4. (a) (i) 9.5 min (ii) $20 - 7 = 13$ min (iii) About 8%

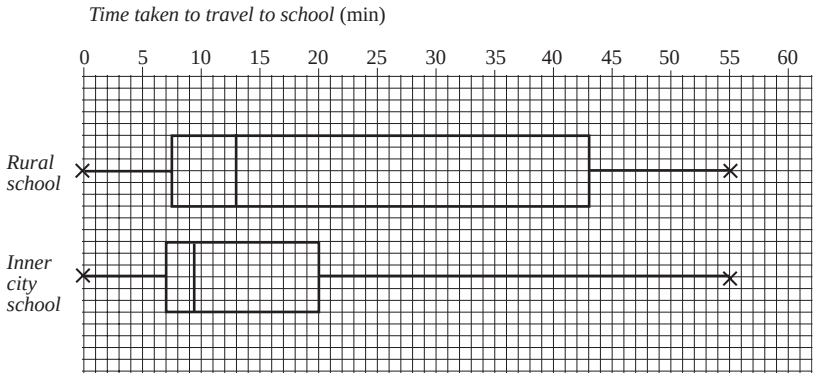
(b) (i) 8, 52, 67, 76, 83, 120

(ii)



(iii) Median = 13 min,
LQ = 7.5 min, UQ = 43 min, IQR = 35.5 min

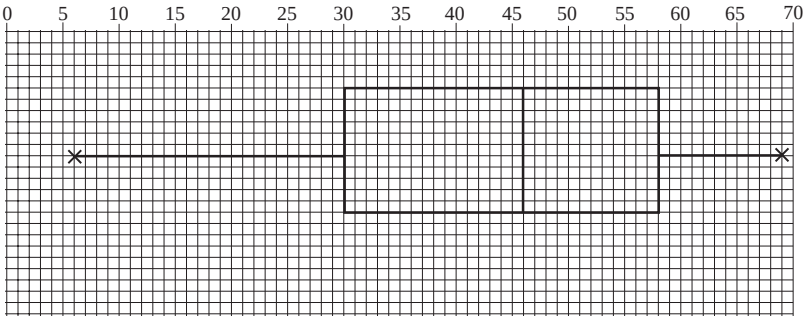
(c)



In general, pupils in the rural school took longer to travel to school than pupils in the city centre. Comparing the widths of the boxes (i.e. the interquartile ranges) shows that there is greater variation in the travel times for pupils in the rural school. Allow other valid answers.

5. (a) 19 (b) 2.53% (c) (i) 46 (ii) 30, 58

(d) *Age at death* (years)



(e) The box and whisker plot shows that 75% of the deaths that year were for people aged 58 or younger. This suggests that there is not a large proportion of elderly residents in the village. Allow other valid answers.