

9 Data Analysis

9.3 Calculations with the Mean

1. The mean shoe size of 12 shoes is 7.5. An extra shoe of size 8 is added to the group of shoes. What is the new mean shoe size?
2. After 9 games, the mean number of goals scored by a football team is 3. If after one more game
 - (a) they score 5 goals, what is the new mean value of goals scored?
 - (b) they score no goals, what is the new mean value of goals scored?
 - (c) the new mean value of goals scored is 2.9, how many goals did they score in this game?
3. The mean number of sandwiches eaten at a party by 20 people was 2.8. How many sandwiches would you need to order for a similar party for 35 people?
4. The first seven of eight judges in a skating competition gave the competitor an average score of 5.8. If the competitor wants to score at least 5.7, what is the least score the eighth judge has to give the competitor?
5. The mean of 7 numbers is 5. When an extra number is added the mean is 5.5. What is the extra number?
6. When 8 is added to a set of 4 numbers, the mean changes to 9.6. What was the mean of the original numbers?
7. On a plane there are 20 business class passengers and 123 tourist class passengers. The mean weight of baggage for the business class passengers was 17.5 kg, and for the tourist class was 9.4 kg.
 - (a) What is the mean weight of baggage for all passengers?
 - (b) If the plane is allowed to carry 2000 kg of luggage, how much extra luggage could have been carried?
8. The table below shows the number of people in each of the 100 cars passing a particular place.

<i>No. of people in each car</i>	1	2	3	4
<i>No. of cars</i>	x	50	y	16

- (a) Find the value of $x + y$.
- (b) If the mean number of people per car is 2.4, show that $x + 3y = 76$.
- (c) Find the value of x and of y by solving appropriate equations.
- (d) State the modal number of people per car.

9. (a) The median of a set of eight numbers is $4\frac{1}{2}$. Given that seven of the numbers are 9, 2, 3, 4, 12, 13 and 1, find the eighth number.
- (b) The mean of a set of six numbers is 2 and the mean of another set of ten numbers is m . If the mean of the combined set of sixteen numbers is 7, find the value of m .

9.4 Mean, Median and Mode for Grouped Data

1. 100 sentences were taken from a book and the number of words per sentence was counted.
- (a) Copy and complete the following table .

<i>No. of words per sentence</i>	<i>Mid-value (x)</i>	<i>No. of sentences (f)</i>
1 - 5		16
6 - 10		22
11 - 15		18
16 - 20		11
21 - 25	23	12
26 - 30		9
31 - 35		8
36 - 40		2
41 - 45		2
		$\sum f = 100$

- (b) Hence, estimate the mean number of words per sentence.
- (c) Given that the mean number of words per sentence of the next 50 sentences is 17.3, estimate the mean number of words per sentence of all 150 sentences.
2. The daily wages of 100 construction workers are displayed in the table below.

<i>Daily wage (£)</i>	$16 \leq x < 18$	$18 \leq x < 20$	$20 \leq x < 22$	$22 \leq x < 24$	$24 \leq x < 26$	$26 \leq x < 28$	$28 \leq x < 30$
<i>No. of workers</i>	8	10	18	30	22	7	5

- (a) Write down the modal class.
- (b) Estimate the mean daily wage.

3. A school librarian recorded the number of books borrowed weekly by pupils in one particular class during 40 successive weeks. The results are shown in the table below.

<i>No. of books borrowed</i>	10 - 14	15 - 19	20 - 24	25 - 29	30 - 34	35 - 39
<i>No. of weeks</i>	5	7	9	8	5	6

- (a) Write down the modal class.
- (b) Draw a bar chart to illustrate this information.
- (c) Estimate the mean number of books borrowed per week.
4. In an experimental farm, 30 hectare plots of land growing potatoes produced yields in tonnes as shown in the frequency table below.

<i>Yield (x tonnes)</i>	$3.4 \leq x < 3.6$	$3.6 \leq x < 4.0$	$4.0 \leq x < 4.4$	$4.4 \leq x < 4.8$	$4.8 \leq x < 5.0$	$5.0 \leq x < 5.6$
<i>No. of plots</i>	3	6	8	5	6	2

- (a) Draw a histogram to represent the results.
- (b) Estimate the mean yield per plot.
5. (a) The diameters of 50 ball bearings produced by a factory measured in mm (correct to 2 significant figures) are given in the table below.

<i>Diameter (mm)</i>	5.0 - 5.2	5.3 - 5.5	5.6 - 5.8	5.9 - 6.1	6.2 - 6.4	6.5 - 6.7
<i>Frequency</i>	6	8	12	11	7	6

- (i) State the modal class.
- (ii) Estimate the mean diameter.
- (b) The diameters of 50 test tubes produced by a glass factory measured in mm (correct to 2 significant figures) are given in the table below.

<i>Diameter (mm)</i>	5.8 - 6.0	6.1 - 6.3	6.4 - 6.6	6.7 - 6.9	7.0 - 7.2	7.3 - 7.5
<i>Frequency</i>	6	8	12	11	7	6

- (i) State the modal class.
- (ii) Using the answer in (a) (ii), or otherwise, estimate the mean diameter of the 50 test tubes.
6. The ages of a group of 25 artists are given below.

16	17	15	15	8
26	16	14	9	13
9	16	20	19	22
18	11	15	14	21
12	20	w21	16	17

- (a) Without grouping, find the mean age.
- (b) Arrange the data in classes, 8 - 12, 13 - 17, and so on. Estimate the mean age.
- (c) Find the difference between the estimated mean age in (b) and the actual mean age in (a) and express this difference as a percentage of the actual mean age.

7. The following data show the places of wedding ceremony against length of engagement (in months), for a sample of 250 couples.

<i>Length of Engagement</i>	0 - 6	6 - 12	12 - 24	24 - 36	36 - 42
<i>Registry Office</i>	39	20	14	10	2
<i>Church</i>	16	13	35	62	39

- (a) Using separate diagrams, draw histograms to illustrate the above information.
 - (b) Calculate the mean length of engagement for each place of marriage.
 - (c) What conclusion can you draw from the answers in (b)?
8. (a) The ages of 30 men convicted for the first time of violent crime in Country X gave the following figures.

22 32 29 28 22 16 19 17 17 16
 19 18 18 30 20 20 28 28 20 23
 23 35 19 22 21 17 32 23 30 21

- (i) Arrange the data in classes of 15 - 19, 20 - 24, . . . , 35 - 39.
 - (ii) Estimate the mean age.
- (b) The grouped frequency distribution of the ages of a group of men convicted for the first time of violent crime from Country Y is as shown in the table below.

<i>Age (x years)</i>	$16 \leq x < 18$	$18 \leq x < 20$	$20 \leq x < 25$	$25 \leq x < 28$	$28 \leq x < 30$	$30 \leq x < 40$
<i>Frequency</i>	12	10	23	15	8	12

- (i) Draw a histogram to represent this information.
- (ii) Estimate the mean age.
- (iii) Comparing the answers in (a) (ii) and (b) (ii), draw a conclusion concerning the ages of men convicted for the first time of violent crime from Country X and Country Y.

9. The heights of 30 children are given in the table below.

<i>Height in cm</i>	<i>Frequency</i>
$150 \leq x < 155$	2
$155 \leq x < 160$	5
$160 \leq x < 165$	8
$165 \leq x < 170$	10
$170 \leq x < 175$	5

- (a) Calculate an estimate of the mean height.
- (b) The class teacher said she expected the average height to be about 165 cm. How was the teacher able to do this?

(SEG)

10. Andrew is a checkout operator at the local supermarket.

At the end of a shift, he looked at the total amounts of money that people had spent.

<i>Amount spent (£x)</i>	<i>Number of people</i>
$0 < x \leq 20$	25
$20 < x \leq 40$	9
$40 < x \leq 60$	10
$60 < x \leq 80$	15
$80 < x \leq 100$	8

- (a) Calculate an estimate of the mean amount spent by his customers during that shift.
- (b) The manager of the supermarket decides to give a bonus to the most efficient checkout operator. She decides that this will be the person who works at the fastest rate.

Here is some information about the three checkout operators after their shift.

<i>Operator</i>	<i>Number of items sold</i>	<i>Time worked</i>
Andrew	10 500	$7\frac{1}{2}$ hours
Barbara	6400	4 hours 15 min
Colin	9120	6 hours

Who should get the bonus?

Give a reason for your answer and show all your working.

(NEAB)

11. Vicki investigated the times taken to serve 120 customers at Supermarket A. Her results are shown below.

<i>Time (seconds)</i>	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
<i>Number of customers</i>	4	17	48	16	35

- (a) (i) Calculate an estimate of the mean time to serve the customers.
(ii) Write down the modal class for the serving times.

Vicki decided to extend her investigation to Supermarket, B.

She obtained the times taken to serve 120 customers at Supermarket B. Her extended table is shown below.

<i>Time (seconds)</i>	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
<i>Number of customers at Supermarket A</i>	4	17	48	16	35
<i>Number of customers at Supermarket B</i>	5	20	54	36	5

- (b) Vicki correctly worked out the mean and modal class for the times at Supermarket B. She also worked out correctly the median of the times for each supermarket.

Use your answers to part (a) to complete the table below.

	<i>Supermarket A</i>	<i>Supermarket B</i>
<i>Median</i>	48.1	46.5
<i>Modal class</i>		40 - 50
<i>Mean</i>		46.3

- (c) Which average in this table represents the data most fairly?
Give a reason for your answer.

(MEG)

12. The table shows the weights of 100 children in year 7.

Weight, w (kg)	Frequency
$20 < w \leq 30$	12
$30 < w \leq 40$	21
$40 < w \leq 50$	38
$50 < w \leq 60$	a
$60 < w \leq 70$	b

An estimate of the mean weight of the children is calculated as 44 kg.

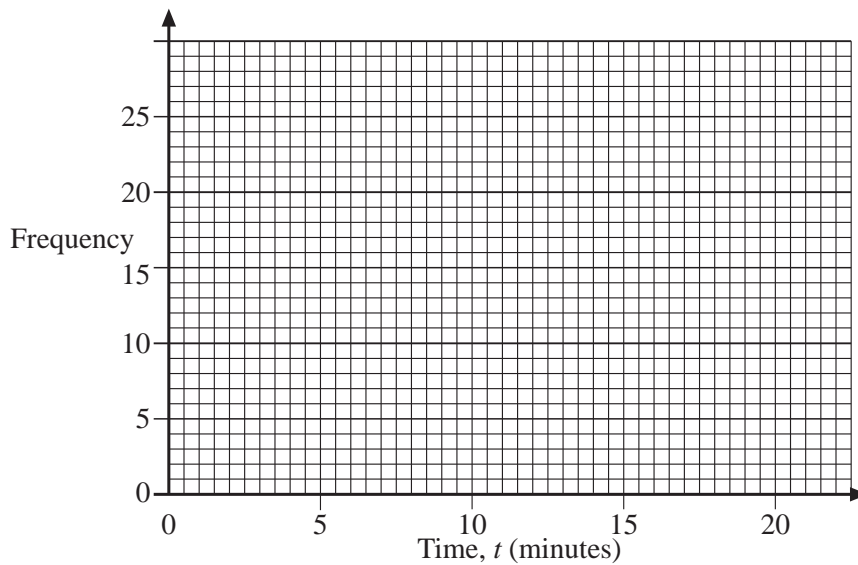
Calculate the values of a and b .

(AQA)

13. 50 people were asked how long they had to wait for a bus.
The table shows the results.

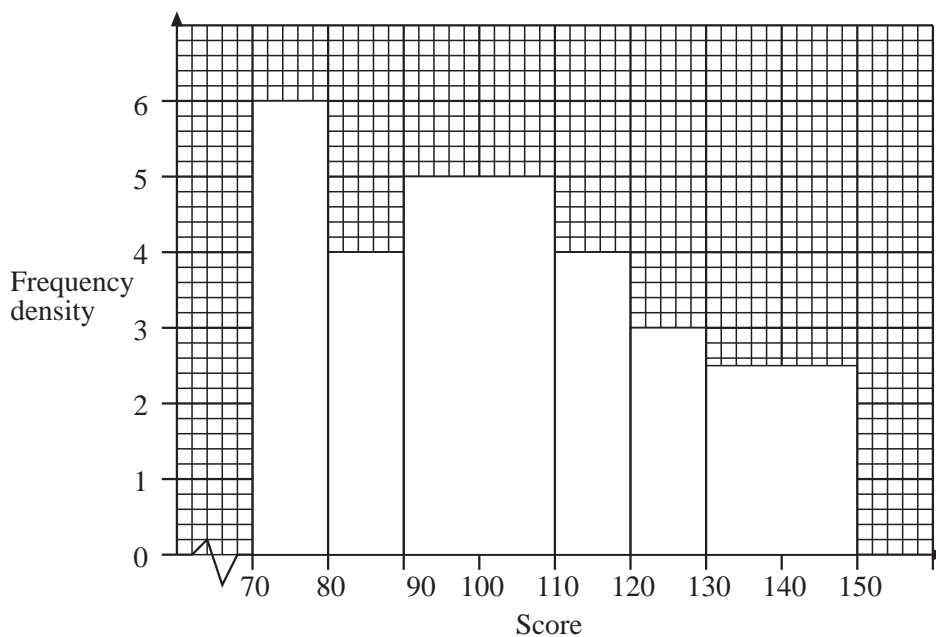
Time, t (minutes)	Frequency	Mid-point	
$0 < t \leq 5$	16		
$5 < t \leq 10$	21		
$10 < t \leq 15$	10		
$15 < t \leq 20$	3		

- (a) Calculate an estimate of the average time they had to wait.
(b) On a copy of the grid below, draw a frequency diagram to represent the data.



(AQA)

14. The histogram shows the test scores of 320 children in a school.



- (a) Find the median score.
 (b) Find the interquartile range of the scores.

(AQA)

15. The table shows the weight of the luggage for passengers on one plane.

Weight, w (kg)	Frequency
$0 < w \leq 5$	14
$5 < w \leq 10$	28
$10 < w \leq 15$	12
$15 < w \leq 20$	9
$20 < w \leq 25$	2

- (a) What was the modal class?
 (b) One of the passengers is selected at random.
 What is the probability that this passenger's luggage weighs 15 kg or less?
 (c) Draw a frequency diagram for this distribution.
 (d) Calculate an estimate of the mean weight of luggage for these passengers.

(OCR)

9.5 Cumulative Frequency

1. A check was made on the speeds of vehicles travelling along a motorway.

<i>Speed in mph (x)</i>	<i>Number of Vehicles</i>
$45 < x \leq 55$	4
$55 < x \leq 65$	9
$65 < x \leq 75$	10
$75 < x \leq 85$	14
$85 < x \leq 95$	11
$95 < x \leq 105$	8
$105 < x \leq 115$	7

Construct a cumulative frequency table and answer the questions that follow.

- How many vehicles were travelling at 85 mph or less?
 - How many vehicles were travelling at 75 mph or less?
 - How many vehicles were travelling at more than 75 mph?
2. The following frequency table shows marks scored by a class of pupils in a test.

<i>Mark (x)</i>	<i>Number of Pupils</i>
$0 \leq x < 20$	4
$20 \leq x < 40$	8
$40 \leq x < 60$	18
$60 \leq x < 80$	8
$80 \leq x < 100$	2

Construct a cumulative frequency table and answer the questions that follow.

- How many pupils scored less than 80 marks?
 - If the pass mark was 40,
 - how many pupils failed the test?
 - what percentage of pupils passed the test?
3. The life spans of 40 batteries are tested using an electric toy by recording the length of time the toy operates before each battery fails. The results are recorded in the following cumulative frequency table.

<i>Life span in hours (x)</i>	10	11	12	13	14
<i>Number of batteries having life spans of less than x</i>	4	12	25	35	40

Using a vertical scale of 1 cm to 5 batteries and a horizontal scale of 2 cm to 1 hour, draw a cumulative frequency curve. Use your graph to answer the following questions.

- (a) How many batteries have life spans less than $11\frac{1}{2}$ h?
- (b) A battery is graded as 'super' if its life span is $13\frac{1}{2}$ h or longer. Find the percentage of 'super' batteries in this batch of 40 batteries.
4. The heights of 80 plants of the same species were measured. The results were tabulated as follows.

<i>Height in cm (x)</i>	<i>Number of Plants</i>
$18 < x \leq 21$	15
$21 < x \leq 24$	16
$24 < x \leq 27$	21
$27 < x \leq 30$	20
$30 < x \leq 33$	8

The table was then reorganised to read:

<i>Height in cm (x)</i>	21	24	27	30	33
<i>Number of plants of this height x or less</i>	15	A	52	72	B

- (a) Find the values of A and B.
- (b) How many plants have heights 21 cm or less?
- (c) How many plants have heights 27 cm or less?
- (d) If all plants with heights greater than 27 cm are exported, how many plants are exported? What percentage of the total is exported?
5. A survey was carried out on 100 pupils to find out the distance of each of their houses from school. The results are shown in the table below.

<i>Distance in miles (x)</i>	2	4	6	8	10
<i>Number of pupils whose house is x miles or less from school</i>	18	50	80	96	100

Using a vertical scale of 2 cm to 10 children and a horizontal scale of 1 cm to 1 mile, draw a cumulative frequency curve. Use your graph to estimate

- (a) the median distance,
- (b) the inter-quartile range.

6. A check was made on the speeds of 100 vehicles travelling along a motorway. The following frequency table shows the results.

<i>Speed in mph (x)</i>	<i>Number of Vehicles</i>
$30 < x \leq 40$	4
$40 < x \leq 50$	5
$50 < x \leq 60$	8
$60 < x \leq 70$	16
$70 < x \leq 80$	23
$80 < x \leq 90$	25
$90 < x \leq 100$	12
$100 < x \leq 110$	6
$110 < x \leq 120$	1

- (a) Copy and complete the following cumulative frequency table.

<i>Speed in mph (x)</i>	40	50	60	70	80	90	100	110	120
<i>Number of vehicles with this speed x or less</i>	4	9							100

- (b) Using a horizontal scale of 2 cm to 10 mph and a vertical scale of 2 cm to 10 vehicles, draw a cumulative frequency curve.
- (c) Use your graph to estimate
- the median speed,
 - the inter-quartile range.
- (d) If driving at a speed above 85 mph is considered speeding, what percentage of vehicles were speeding? Give your answer correct to the nearest whole number.
7. As part of his Geography fieldwork, Tony took measurements of the steepness of slopes. The steepness was measured as the angle the slope made with the horizontal.

Tony's results are shown below.

15°	16°	9°	21°	32°
37°	25°	36°	40°	8°
13°	21°	32°	29°	32°
7°	4°	18°	17°	32°

Tony decided to group the data into 4 equal class intervals on an observation sheet.

- (a) Copy and complete the observation sheet below, using 4 equal class intervals.

<i>Class interval (steepness °)</i>	<i>Tally</i>	<i>Frequency</i>

- (b) Use the completed observation sheet to draw a frequency diagram of the data.

(LON)

8. A group of people took a fitness test.

They exercised hard.

Then they were timed to see how long their pulses took to return to normal.

The time taken for a pulse to return to normal is called the *recovery time*.

The recovery times for the group are shown in the table below.

<i>Recovery Time (seconds)</i>	<i>Frequency</i>	<i>Cumulative Frequency</i>
0 up to but not including 20	0	0
20 up to but not including 40	7	7
40 up to but not including 60	9	16
60 up to but not including 80	18	34
80 up to but not including 90	13	47
90 up to but not including 100	12	59
100 up to but not including 120	9	68
120 up to but not including 140	6	74

- (a) Use the figures in the table to draw a cumulative frequency curve.

- (b) Use your cumulative frequency curve to estimate the value of
- the median,
 - the inter-quartile range.

A second group of people took the fitness test. The recovery times of people in this group had a median of 61 seconds and an inter-quartile range of 22 seconds.

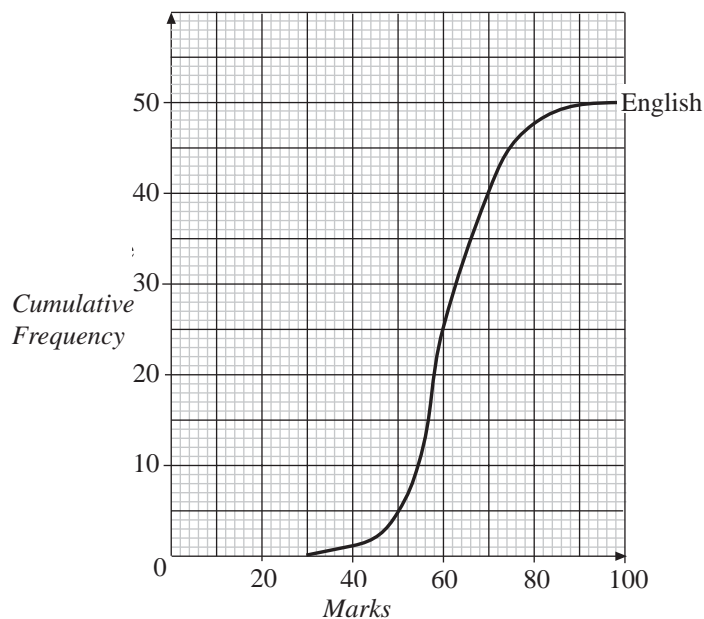
- (c) Compare the fitness results of these two groups.

(LON)

9. (a) 50 pupils take an English exam and a Maths exam. The distribution of the marks they obtained is shown in the table below.

<i>Mark</i>		21 - 30	31 - 40	41 - 50	51 - 60	61 - 70	71 - 80	81 - 90	91 - 100
<i>Number of pupils</i>	<i>English exam</i>	0	1	4	20	14	8	2	1
	<i>Maths exam</i>	2	3	6	10	12	10	4	3

The following graph shows the cumulative frequency for the English marks.



- On a copy of the the graph, show the cumulative frequency for the Maths marks.
- Copy and complete the table below.

	<i>English</i>	<i>Maths</i>
<i>Median</i>	60	
<i>Inter-quartile range</i>	14	

- Use the information in the table to comment on the differences between the two distributions of marks.

- (b) Of the 50 pupils, 30 pass the Maths exam at the first attempt. From past performance it is known that, if a pupil fails at the first attempt, the probability of passing at the second attempt is 0.7. Calculate the probability that a pupil, chosen at random from 50 pupils, will pass the maths exam at either the first or second attempt.

(MEG)

10. Pete wanted to find out the length of time cars were left in a car park. His results, to the nearest minute, are given in the table.

<i>Length of stay (minutes)</i>	<i>Number of cars (frequency)</i>	<i>Cumulative frequency</i>
$0 < t \leq 15$	0	
$15 < t \leq 30$	23	
$30 < t \leq 45$	35	
$45 < t \leq 60$	41	
$60 < t \leq 75$	63	
$75 < t \leq 90$	21	
$90 < t \leq 120$	10	
$120 < t \leq 135$	7	

- (a) Copy and complete the table.
 (b) Draw a cumulative frequency diagram for the data.
 (c) Use your diagram to estimate the inter-quartile range.
 (d) The owners of the car park think that about two thirds of the cars are parked for between 40 and 80 minutes.

Do Pete's results support this?

Give a reason for your answer.

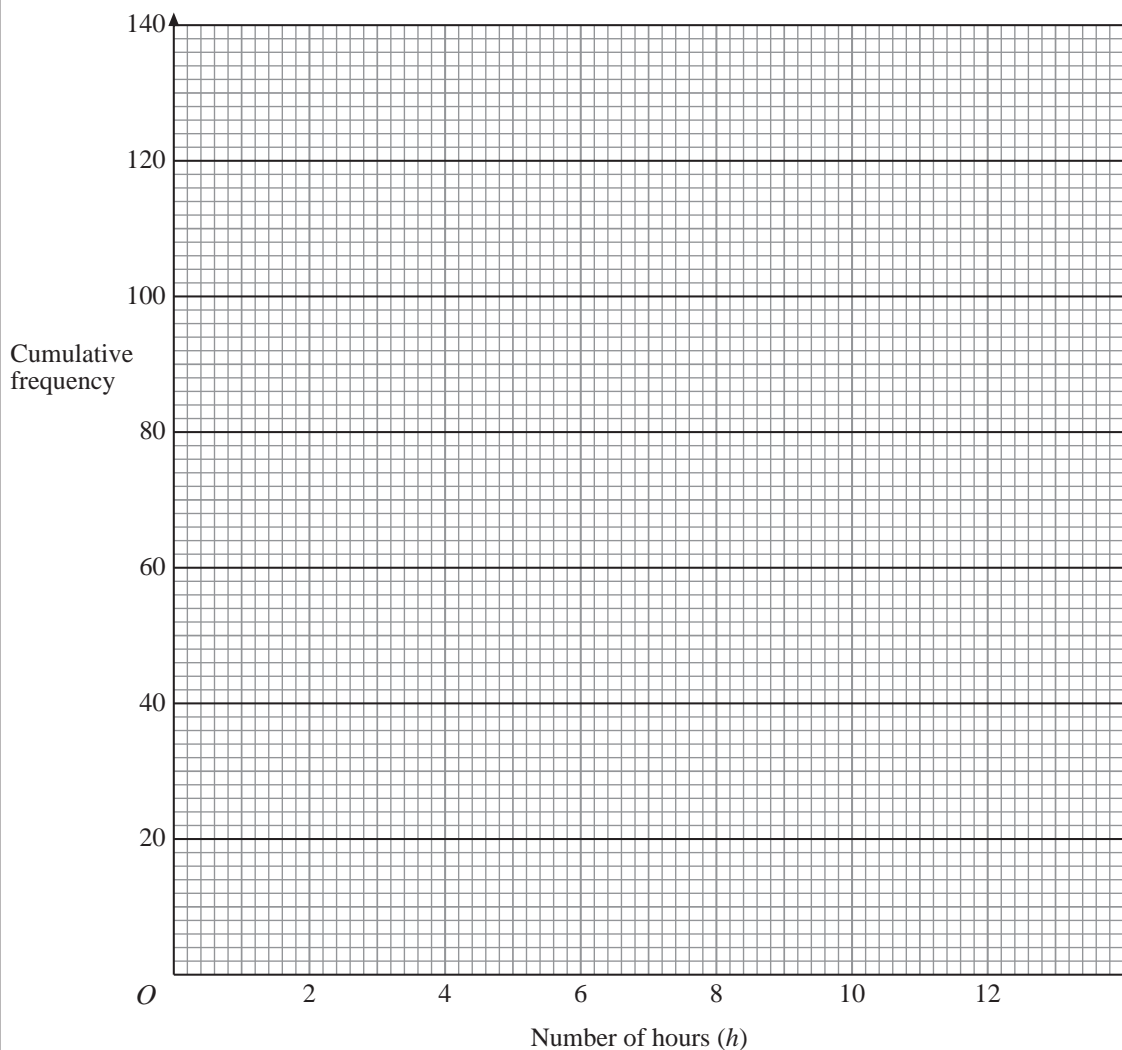
(NEAB)

11. The table shows information about the number of hours that 120 children watched television last week.

Number of hours (<i>h</i>)	Frequency
$0 < h \leq 2$	10
$2 < h \leq 4$	20
$4 < h \leq 6$	25
$6 < h \leq 8$	40
$8 < h \leq 10$	15
$10 < h \leq 12$	10

- (a) Work out an estimate for the mean number of hours that the children watched television last week.
- (b) Copy and complete the cumulative frequency table.

Number of hours (h)	Frequency
$0 < h \leq 2$	10
$0 < h \leq 4$	
$0 < h \leq 6$	
$0 < h \leq 8$	
$0 < h \leq 10$	
$0 < h \leq 12$	



- (c) On a copy of the grid, draw a cumulative frequency graph for your table.
- (d) Use your graph to find an estimate for the number of children who watched television for **fewer** than 5 hours last week.

(Edexcel)

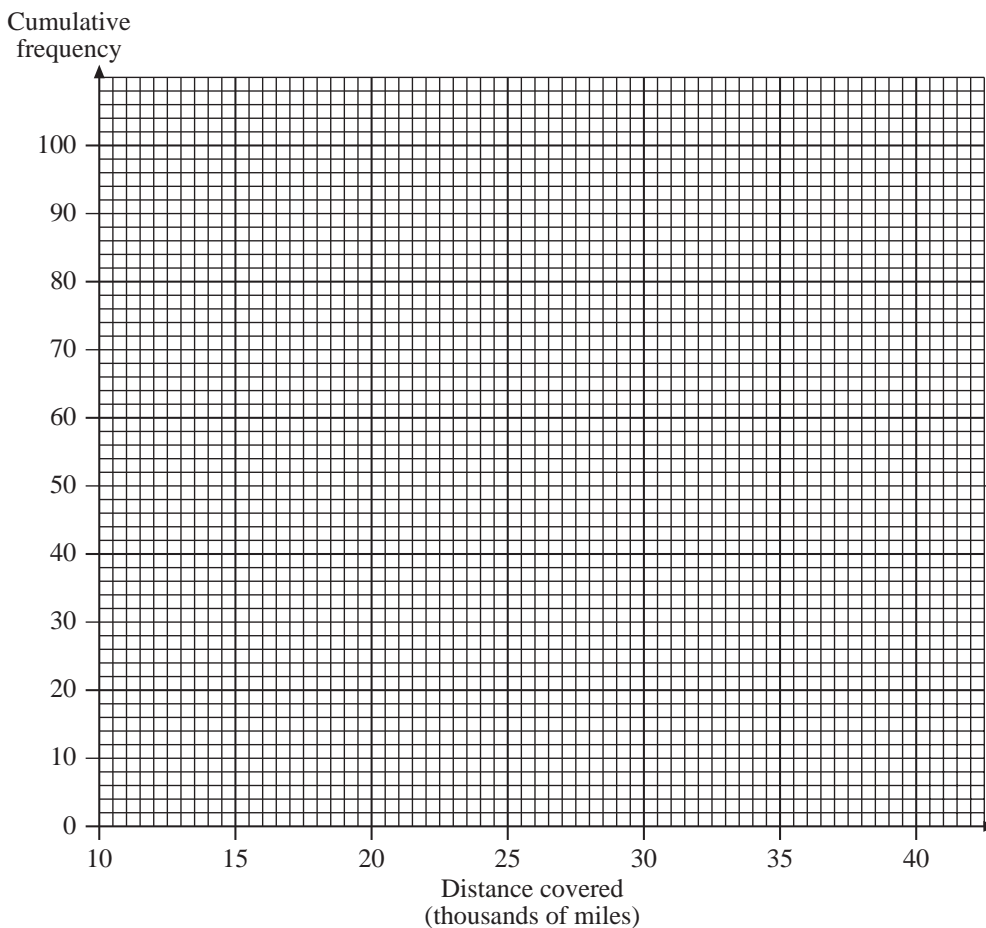
12. A manufacturer investigates how far a car travels before it needs new tyres. The distances covered by 100 cars before they needed new tyres is shown in the table below.

Distance covered (x thousands of miles)	Number of cars
$10 < x \leq 15$	10
$15 < x \leq 20$	23
$20 < x \leq 25$	31
$25 < x \leq 30$	19
$30 < x \leq 35$	12
$35 < x \leq 40$	5

- (a) Complete a copy of the cumulative frequency table for the 100 cars.

Distance covered (x thousand miles)	$x \leq 15$	$x \leq 20$	$x \leq 25$	$x \leq 30$	$x \leq 35$	$x \leq 40$
Cumulative frequency	10					

- (b) Draw the cumulative frequency diagram on a copy of the grid below.



- (c) Use your cumulative frequency diagram to estimate the median distance covered.
- (d) Use your diagram to estimate how many cars travelled less than 23 000 miles before needing new tyres.

(OCR)

9.6 Standard Deviation

1. Find the mean and standard deviation of these sets of numbers
- (a) 8, 9, 10, 11, 12.
- (b) 1, 2, 10, 18, 19.
- (c) 10, 10, 10, 10, 10.
- (d) 1, 10, 10, 10, 19.
2. Without doing all calculations, decide which of the following sets of numbers has
- (a) the smallest
- (b) the largest
- standard deviation.

Set A : 10, 15, 20, 25, 30.

Set B : 5, 10, 20, 30, 35.

Set C : 18, 19, 20, 21, 22.

Set D : 18, 18, 20, 22, 22.

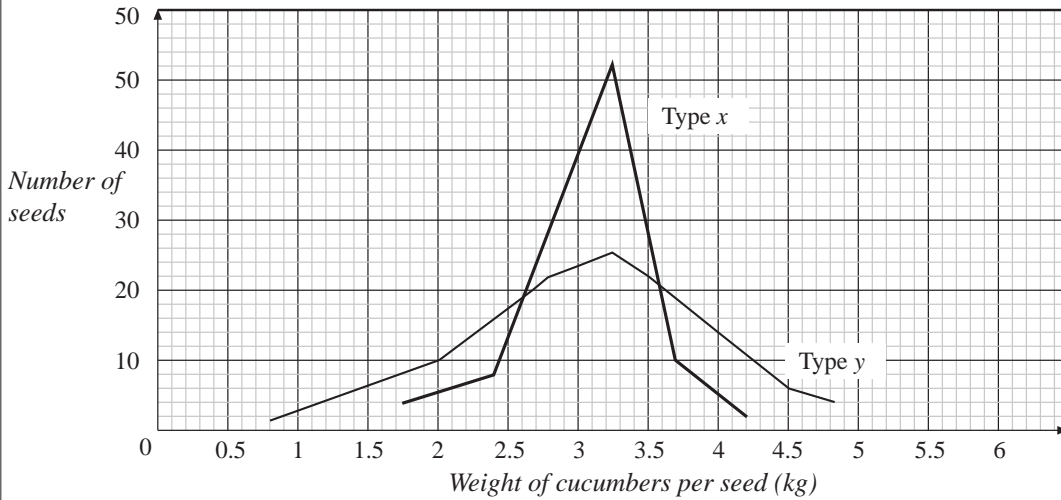
Check your answers by actually calculating the standard deviation.

3. The number of minutes late was noted for a sample of journeys on the 'Golden Hind' Intercity train from Penzance to London Paddington. Some of these journeys took place before privatisation, others after privatisation.

<i>Before</i>	0,	3,	0,	25,	0,	0,	5,	55,	0
<i>After</i>	10,	5,	0,	0,	3,	2,	0,	15,	4

Find the mean and standard deviation of each set of times. What can you conclude from this data? Is this a fair conclusion?

4. The following graph gives information about the weight of cucumbers produced from 100 seeds of two different varieties, type x and type y .
- (a) Which variety of seed has more variation in the weight of cucumbers produced?
- (b) Give a reason for your answer.



(SEG)

5. A researcher is trying out two types of compost, A and B. The yield of each tomato plant, grown with each compost, in similar conditions, is given below.

A (kg)	25, 22, 30, 19, 17, 26, 12, 9
B (kg)	22, 18, 21, 20, 19, 23, 20, 17

Find the mean and standard deviation of each set of results. Which compost should the researcher recommend for commercial use?

6. A group of 20 pupils took two mock exams from different exam boards. Each score was out of a total of 100 marks. The scores obtained are given below.

Exam	Mark
First	76, 52, 10, 27, 34, 53, 72, 64, 55, 67, 42, 45, 91, 28, 52, 61, 70, 34, 96, 15
Second	43, 52, 61, 55, 72, 36, 37, 53, 62, 71, 32, 25, 74, 67, 69, 55, 57, 43, 62, 70

Find the mean and standard deviation for each exam.

Compare and contrast the results.

If the pass marks for each exam were the same, which exam do you think is the easier?

7. Normally, a superstore sells bottles of champagne with a mean of 10 bottles per day and standard deviation of 2.5.

It is introducing a new brand, and over the first 20 days sells the following quantities:

17, 15, 12, 7, 5, 12, 12, 10, 7, 15,
12, 11, 6, 8, 11, 13, 10, 14, 12, 11

Find the mean and standard deviation of this data.

Would you vote this new brand as a success? Give your reasons.

8. A football manager wants to buy a new striker for his team. He has two players in mind. The goal scoring record of these players for their last 14 games is given in the table below.

<i>Player 1</i>	0,	1,	0,	0,	0,	3,	1,	0,	1,	1,	3,	0,	0,	3
<i>Player 2</i>	1,	1,	0,	2,	1,	1,	0,	1,	1,	0,	2,	1,	1,	1

Find the mean and standard deviation of each player's scoring record.

Which striker do you think the manager should buy? Give your reasons.

9. The numbers of students travelling on the school bus each day over the past term have mean value 20.4 and standard deviation 3.2.

For the new term the bus fares have been reduced. The loadings for the first 15 days of operation were:

22, 25, 24, 22, 26,

27, 25, 24, 26, 27,

28, 27, 25, 29, 28

Find the mean and standard deviation of this data.

Has the reduction in fares been successful? Explain your answer.