

# Cumulative frequency

This chapter will show you how to

- ✓ construct cumulative frequency tables
- ✓ find quartiles and interquartile ranges
- ✓ draw box plots
- ✓ compare and interpret distributions

You will need to know

- how to construct frequency tables

## 16.1 Cumulative frequency

You can construct a **cumulative frequency table** from a **grouped frequency table** by calculating the running total of the frequency up to the end of each class interval.

You can use the table to plot a **cumulative frequency diagram**.

You can draw cumulative frequency diagrams for both discrete and continuous data sets.



### EXAMPLE 1

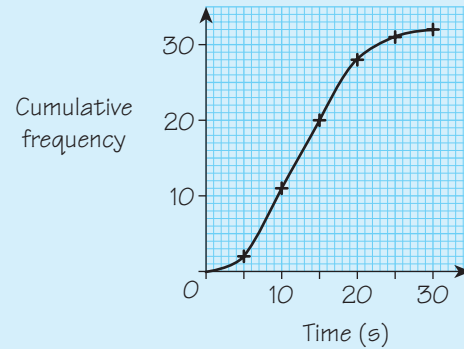
The following grouped frequency distribution shows the time taken for students to solve a puzzle. Using this data, draw a cumulative frequency diagram.

| Time taken (seconds) | Frequency |
|----------------------|-----------|
| $0 < t \leq 5$       | 2         |
| $5 < t \leq 10$      | 9         |
| $10 < t \leq 15$     | 9         |
| $15 < t \leq 20$     | 8         |
| $20 < t \leq 25$     | 3         |
| $25 < t \leq 30$     | 1         |

continued ▼

Class intervals (such as  $0 < t \leq 5$ ) were discussed in Chapter 15.

| Time taken (seconds) | Cumulative frequency                  |
|----------------------|---------------------------------------|
| $0 < t \leq 5$       | <b>2</b>                              |
| $5 < t \leq 10$      | $2 + 9 = 11$                          |
| $10 < t \leq 15$     | $2 + 9 + 9 = \mathbf{20}$             |
| $15 < t \leq 20$     | $2 + 9 + 9 + 8 = \mathbf{28}$         |
| $20 < t \leq 25$     | $2 + 9 + 9 + 8 + 3 = \mathbf{31}$     |
| $25 < t \leq 30$     | $2 + 9 + 9 + 8 + 3 + 1 = \mathbf{32}$ |



Draw a smooth curve through the points (to give a cumulative frequency curve), as shown.

The shape of the cumulative frequency diagram reflects the characteristics of the data and how this data is spread or distributed within the range. This characteristic S shape (called an ogive) appears in nearly all cumulative frequency diagrams.

## Cumulative frequency graphs

- Choose a suitable scale for each axis.
- Always plot the cumulative frequencies on the vertical axis (the y-axis).
- Plot the points using the upper class boundary e.g. (5, 2), (10, 11), (15, 20) and so on.
- Draw a smooth curve through the points (points can also be joined with straight lines) and include the origin.
- Check that you have plotted your points correctly.
- Check that your graph is an S-shape.

Begin by creating a cumulative frequency table. This cumulative frequency column is a running total of the frequency so far.

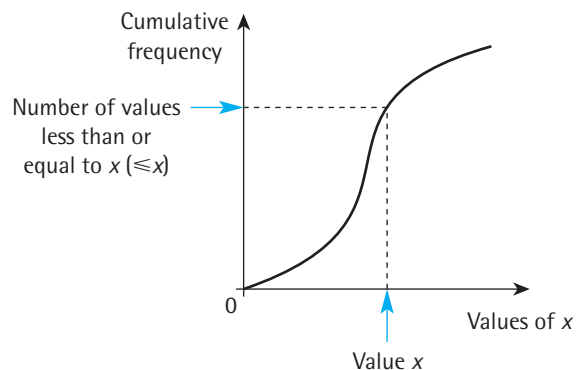
The figures in bold are the values of the cumulative frequency. The class intervals ( $t \leq$ ) are also changed to show the values 'up to and including' i.e. the **upper class boundary** value.

You can now use this data to draw a cumulative frequency diagram by plotting the cumulative frequency against the upper class boundary of each class interval.

Cumulative frequency is always plotted on the vertical axis. The maximum value on the y-axis (the cumulative frequency axis) should be the total number of values in the data set, in this case 32.

## Estimating values using a cumulative frequency graph

Once you have drawn your cumulative frequency diagram you can use it to estimate values for the data.



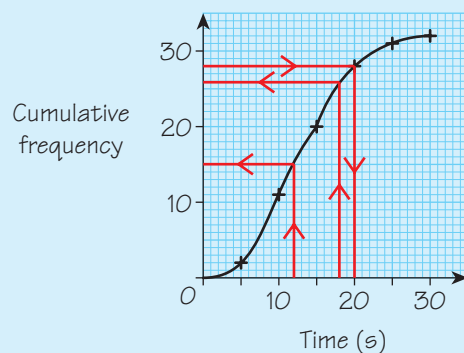
For example, given a value of  $x$  you can estimate the number of values less than or greater than  $x$  by drawing a straight line that meets the cumulative frequency curve and then drawing a corresponding line to meet the other axis.



### EXAMPLE 2

Use the cumulative frequency curve from Example 1 to estimate

- (a) how many students solved the puzzle  
 (i) within 12 seconds      (ii) in more than 18 seconds  
 (b) the time by which 28 students had solved the puzzle.



- (a) (i) 15 students solved the puzzle within 12 seconds.  
 (ii)  $32 - 26 = 6$   
 6 students took longer than 18 seconds.  
 (b)  $x = 20$   
 28 students solved the puzzle within 20 seconds.

Always draw lines on your graph to show how you obtained your answers. Do not rub them out.

(a) Draw a vertical line for  $x = 12$  to intersect the cumulative frequency curve. Now draw a line to meet the cumulative frequency axis. The value obtained is 15.

(b) The line at  $x = 18$  seconds gives a value of 26. This means that  $32 - 26 = 6$  students took longer than 18 seconds to solve the puzzle.

(c) Draw a line from the cumulative frequency value of 28 to intersect the curve. Now draw a vertical line to meet the  $x$ -axis at the point  $x = 20$ . This is the time by which 28 students solved the puzzle.



## EXERCISE 16A

- 1 (a) Copy and complete the cumulative frequency table showing the distribution of test marks for a group of 32 students.

| Mark (%) | Number of students | Mark (%)  | Cumulative frequency |
|----------|--------------------|-----------|----------------------|
| 1–10     | 1                  | $\leq 10$ | 1                    |
| 11–20    | 2                  | $\leq 20$ | 3                    |
| 21–30    | 4                  |           |                      |
| 31–40    | 7                  |           |                      |
| 41–50    | 5                  |           |                      |
| 51–60    | 8                  |           |                      |
| 61–70    | 2                  |           |                      |
| 71–80    | 2                  |           |                      |
| 81–90    | 1                  |           |                      |
| 91–100   | 0                  |           |                      |

Your x-axis (mark) should run from 0 to 100, and your y-axis from 0 to 35.

- (b) Plot the results in a cumulative frequency graph.
- 2 The table below shows the frequency distribution of test marks for 120 students.

| Mark (%) | Number of students |
|----------|--------------------|
| 1–10     | 1                  |
| 11–20    | 6                  |
| 21–30    | 8                  |
| 31–40    | 15                 |
| 41–50    | 17                 |
| 51–60    | 24                 |
| 61–70    | 22                 |
| 71–80    | 15                 |
| 81–90    | 9                  |
| 91–100   | 3                  |

Construct a cumulative frequency table (taking the first class interval to be  $\leq 10$  and the last interval to be  $\leq 100$ ). Draw the corresponding cumulative frequency graph for this distribution.

- 3 The results for the long jump at a school sports day are shown in the table.

Draw a cumulative frequency diagram for this distribution.

Estimate how many students jumped over 2.35 m.

| Distance $x$ (m)     | Frequency |
|----------------------|-----------|
| $1.70 < x \leq 1.80$ | 2         |
| $1.80 < x \leq 1.90$ | 6         |
| $1.90 < x \leq 2.00$ | 9         |
| $2.00 < x \leq 2.10$ | 7         |
| $2.10 < x \leq 2.20$ | 15        |
| $2.20 < x \leq 2.30$ | 8         |
| $2.30 < x \leq 2.40$ | 8         |
| $2.40 < x \leq 2.50$ | 2         |

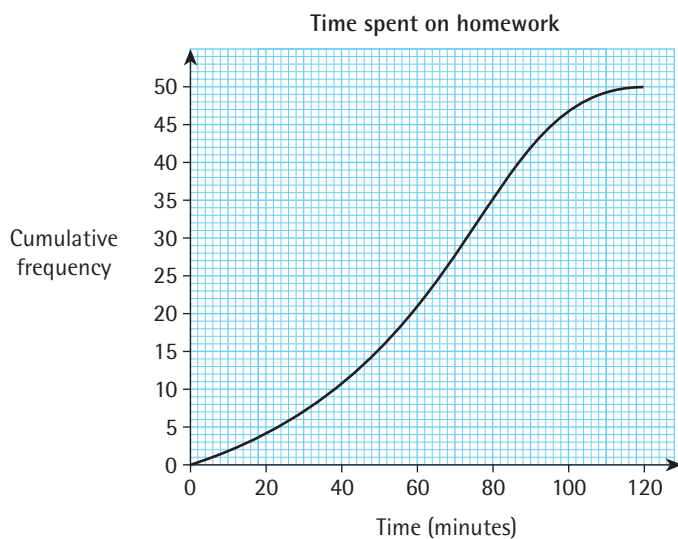
- 4 The temperature in  $^{\circ}\text{C}$  recorded over a 66-day period is shown in the table.

Draw a cumulative frequency diagram and estimate the number of days that the temperature was above  $18^{\circ}\text{C}$ .

| Temperature $t$ ( $^{\circ}\text{C}$ ) | Number of days |
|--|----------------|
| $0 < t \leq 3$                         | 1              |
| $3 < t \leq 7$                         | 7              |
| $7 < t \leq 11$                        | 18             |
| $11 < t \leq 15$                       | 20             |
| $15 < t \leq 19$                       | 17             |
| $19 < t \leq 23$                       | 2              |
| $23 < t \leq 27$                       | 1              |
| $2.40 < x \leq 2.50$                   | 2              |

- 5 The cumulative frequency curve for the amount of time spent on a homework task is shown below.

Construct a cumulative frequency table and use this to estimate the number of students who spent more than 1 hour on the task.



## 16.2 The median, quartiles and inter-quartile range

You can obtain important statistical measures from a cumulative frequency graph. These are useful when you want to compare two or more data sets.

- The **median** is the middle value of the distribution. This is at  $\frac{1}{2}(n + 1)$  on the cumulative frequency axis.
- The **lower quartile (LQ)** is at  $\frac{1}{4}(n + 1)$  on the cumulative frequency axis.
- The **upper quartile (UQ)** is at  $\frac{3}{4}(n + 1)$  on the cumulative frequency axis.
- The **interquartile range (IQR)** gives an improved measure of the spread of the data and is given by  
Inter-quartile range = upper quartile – lower quartile

$n$  is the total frequency.

The answers for the median, the lower quartile and the upper quartile are read off on the *horizontal* axis. The first step in this process is to locate the appropriate points on the cumulative frequency axis.

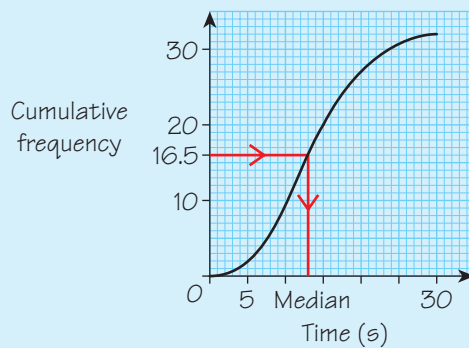
The IQR is a measure of spread of the middle 50% of the data, so it excludes extreme values.



### EXAMPLE 3

Use the cumulative frequency graph of Example 1 to find an estimate for

- |                        |                              |
|------------------------|------------------------------|
| (a) the median         | (b) the lower quartile       |
| (c) the upper quartile | (d) the interquartile range. |

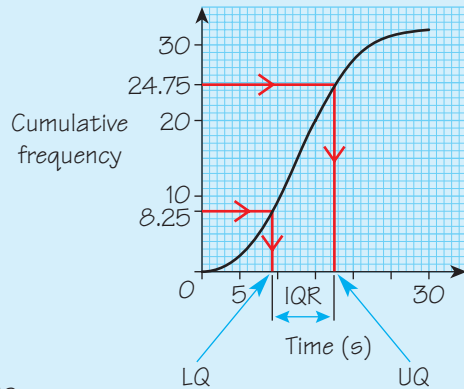


continued ▼

$n = 32$   
The middle value is  
 $\frac{1}{2}(32 + 1) = 16\frac{1}{2}$ th value

Draw a horizontal line from this point to meet the curve. The median value is where this vertical line meets the  $x$ -axis, as shown.

(a) The median = 13 seconds.



(b) Total = 32

$$\frac{1}{4} \text{ of } (32 + 1) = 8.25$$

$$\text{LQ} = 9.5 \text{ seconds}$$

(c)  $\frac{3}{4}$  of  $(32 + 1) = 24.75$

$$\text{UQ} = 24.5 \text{ seconds}$$

(d)  $\text{IQR} = 17.5 - 9.5 = 8 \text{ seconds}$

(b) To find the lower quartile (LQ), first locate a point at  $\frac{1}{4}(32 + 1) = 8.25$  on the frequency axis and draw a horizontal line to intersect the curve. Draw a vertical line to meet the x-axis. This gives the lower quartile.

(c) Find the upper quartile (UQ) in a similar way, but this time use a point at  $\frac{3}{4}(32 + 1) = 24.75$  on the cumulative frequency axis.

(d) Remember  $\text{IQR} = \text{UQ} - \text{LQ}$

The cumulative graph can be divided up into other fractions, not just quarters.

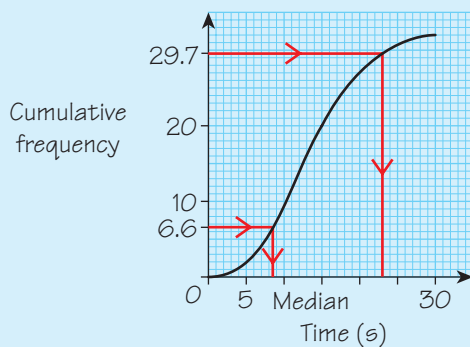
Tenths are quite common and these are called **deciles**.

**Percentiles** divide the cumulative frequency into hundredths.

The 70th percentile will be at  $\frac{70}{100}(n + 1)$  on the cumulative frequency (cf) axis.



### EXAMPLE 4



Use the cumulative frequency curve to find an estimate for these.

(a) the 20th percentile

(b) the 90th percentile

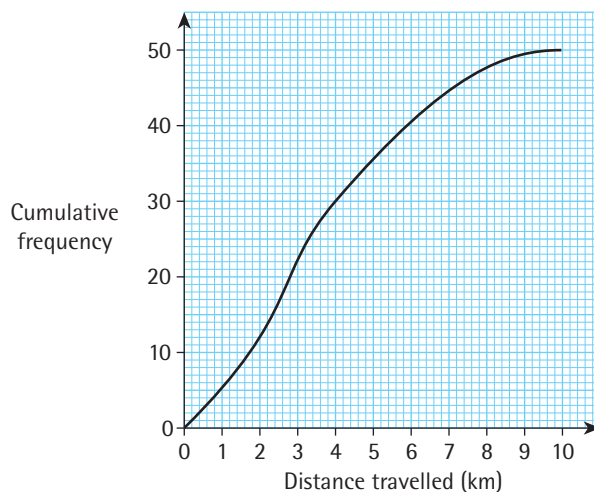
continued ▼

- (a) The 20th percentile will be at  $\frac{20}{100}(32 + 1) = 6.6$   
 Draw a horizontal line at 6.6 to the curve and then a vertical line down to the time axis.  
 20th percentile = 9 seconds
- (b) The 90th percentile will be at  $\frac{90}{100}(32 + 1) = 29.7$   
 Repeat the process of (a) at 29.7  
 90th percentile = 23 seconds



## EXERCISE 16B

- 1 The cumulative frequency graph for the distances travelled to school by 50 students is shown below.



From the graph, estimate

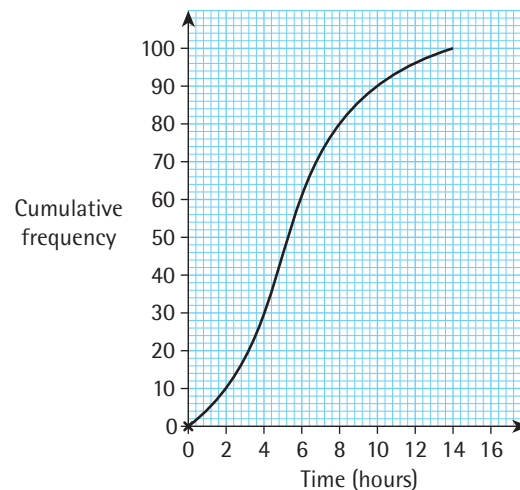
- (a) the median distance      (b) the lower quartile distance  
 (c) the upper quartile distance    (d) the inter-quartile range.
- 2 A survey was conducted to find the amount of time students spent on homework each week. The results are shown in the cumulative frequency graph.

From the graph, estimate

- (a) the median time  
 (b) the lower quartile time  
 (c) the upper quartile time  
 (d) the IQR.

Give your answers in hours and minutes.

- (e) Estimate what percentage of students spent 5 hours or less on homework each week.

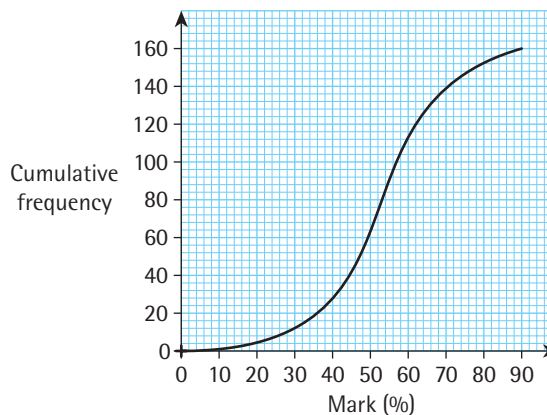




- 3 The cumulative frequency diagram shows the marks gained in a geography exam.

Use the graph to find mark estimates for

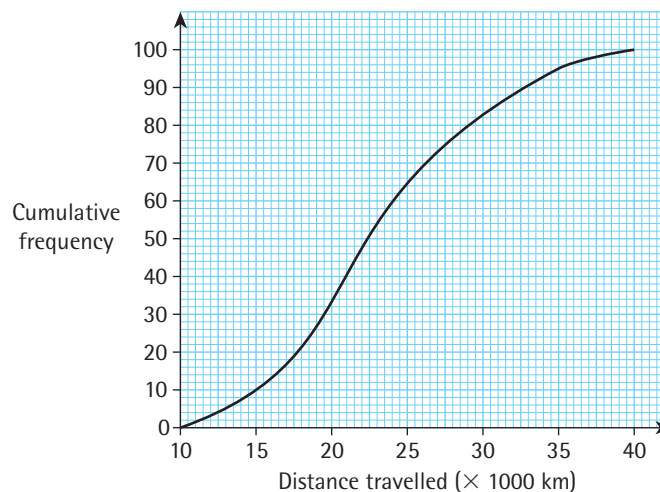
- (a) the median
- (b) LQ
- (c) UQ
- (d) IQR
- (e) the 70th percentile.



- 4 A tyre company carried out a survey to find how far cars travel before they need new tyres. The results for 100 cars is shown in the cumulative frequency diagram.

Use the graph to estimate

- (a) the median distance travelled
- (b) the lower quartile distance travelled
- (c) the upper quartile distance travelled
- (d) the inter-quartile range.
- (e) what percentage of cars travelled a distance of more than 30 000 kilometres before the tyres needed to be changed?



- 5 During a busy day at a doctor's surgery the amount of waiting time for patients was recorded as shown in the table.

The shortest waiting time was  $1\frac{1}{2}$  mins. and the longest time was  $9\frac{1}{2}$  mins.

- (a) Construct a cumulative frequency table.
- (b) Draw a cumulative frequency curve.
- (c) Give an estimate of the median waiting time.
- (d) Find the 60th percentile.

| Waiting time, $t$ (minutes) | Frequency ( $f$ ) |
|-----------------------------|-------------------|
| $0 < t \leq 1$              | 0                 |
| $1 < t \leq 2$              | 4                 |
| $2 < t \leq 3$              | 23                |
| $3 < t \leq 4$              | 43                |
| $4 < t \leq 5$              | 58                |
| $5 < t \leq 6$              | 37                |
| $6 < t \leq 7$              | 11                |
| $7 < t \leq 8$              | 3                 |
| $8 < t \leq 9$              | 0                 |
| $9 < t \leq 10$             | 1                 |

- 6 The local sports centre conducted a survey on the age distribution of its 800 members and the results are shown in the table. The youngest person was 6 years old and the oldest was 78.

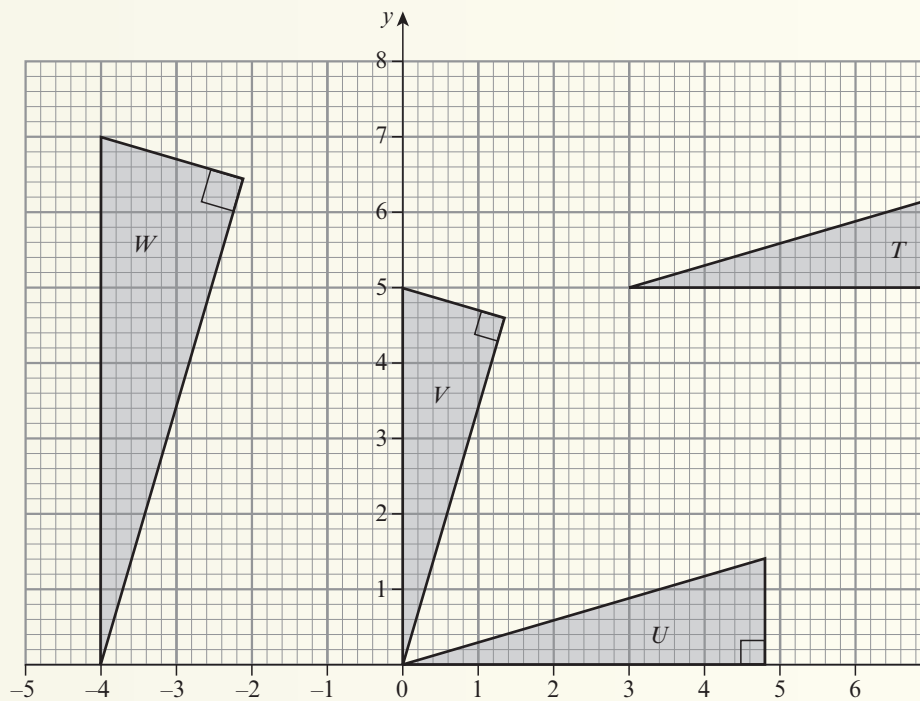
| Age, $a$ (years) | Frequency ( $f$ ) |
|------------------|-------------------|
| $0 < a \leq 10$  | 41                |
| $10 < a \leq 20$ | 138               |
| $20 < a \leq 30$ | 168               |
| $30 < a \leq 40$ | 192               |
| $40 < a \leq 50$ | 126               |
| $50 < a \leq 60$ | 85                |
| $60 < a \leq 70$ | 39                |
| $70 < a \leq 80$ | 11                |

- (a) Construct a cumulative frequency table for this data.  
(b) Draw a cumulative frequency graph.  
(c) Use the curve to estimate the median age of the leisure centre members.  
(d) Find an estimate of the lower and upper quartiles.  
(e) Find the 95th percentile.
-

## EXAMINATION QUESTIONS

- 1 500 eggs were sorted by mass into five different sizes.

|             | Mass ( $m$ grams) | Frequency |
|-------------|-------------------|-----------|
| Small       | $35 \leq m < 40$  | 20        |
| Medium      | $40 \leq m < 50$  | 60        |
| Standard    | $50 \leq m < 60$  | 200       |
| Large       | $60 \leq m < 75$  | 180       |
| Extra large | $75 \leq m < 80$  | 40        |

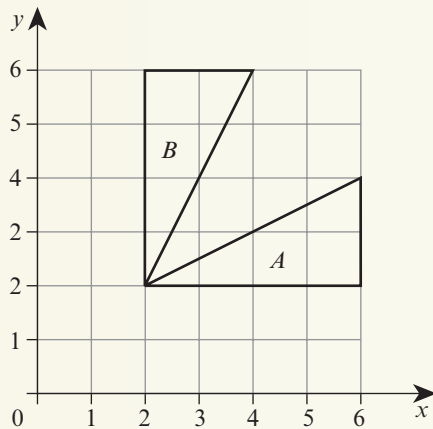


This cumulative frequency curve has been drawn using the information in the table above.

- (a) Explain why the point  $(60, 280)$  is on the curve. [2]  
 (b) Estimate the median mass of the eggs. [1]  
 (c) Estimate the inter-quartile range of the masses of the eggs. [2]

(CIE Paper 4, Jun 2000)

2



The cumulative frequency diagram shows the heights of plants measured in an experiment. From the diagram, estimate

- (a) (i) the lower quartile, [1]  
 (ii) the inter-quartile range. [1]  
 (b) the number of plants with a height greater than 25cm. [1]

(CIE Paper 2, Nov 2000)

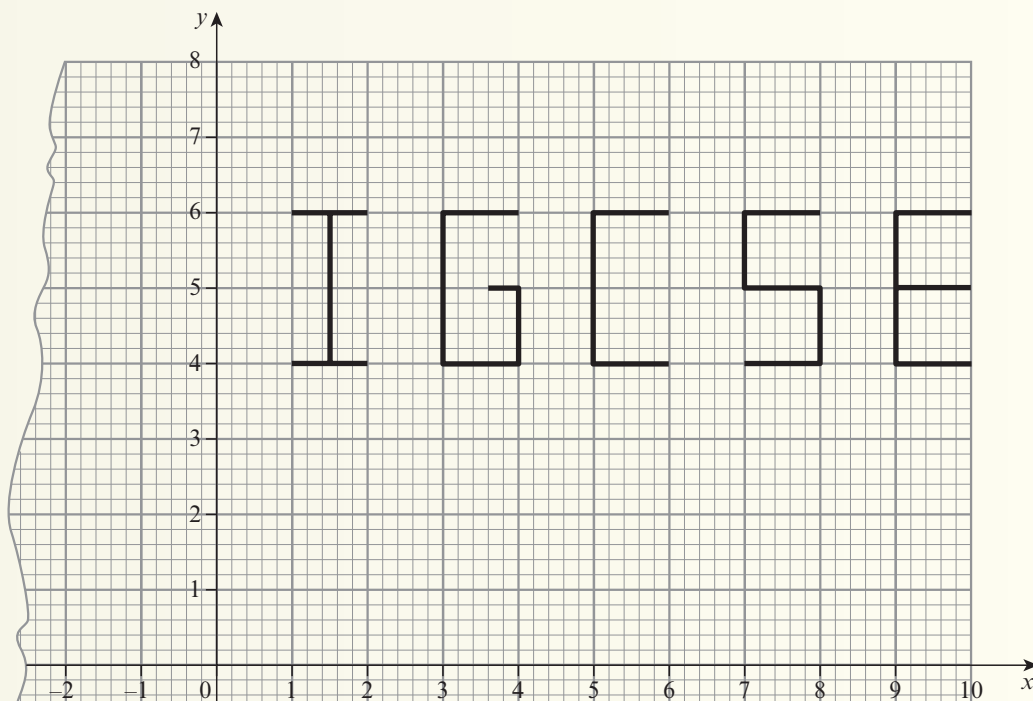
- 3 In a survey, 200 shoppers were asked how much they had just spent at a supermarket. The results are shown in the table.

| Amount (\$x)       | $0 < x \leq 20$ | $20 < x \leq 40$ | $40 < x \leq 60$ | $60 < x \leq 80$ | $80 < x \leq 100$ | $100 < x \leq 140$ |
|--------------------|-----------------|------------------|------------------|------------------|-------------------|--------------------|
| Number of shoppers | 10              | 32               | 48               | 54               | 36                | 20                 |

- (a) (i) Make a cumulative frequency table for these 200 shoppers. [2]  
 (ii) Using a scale of 2 cm to represent \$20 on the horizontal axis and 2 cm to represent 20 shoppers on the vertical axis, draw a cumulative frequency diagram for this data. [4]  
 (b) Use your cumulative frequency diagram to find  
 (i) the median amount, [1]  
 (ii) the upper quartile, [1]  
 (iii) the inter-quartile range, [1]  
 (iv) how many shoppers spent at least \$75. [2]

(CIE Paper 4, Jun 2003)

- 4 The depth,  $d$  centimetres, of a river was recorded each day during a period of one year (365 days). The results are shown by the cumulative frequency curve.



- (a) Use the cumulative frequency curve to find
- (i) the median depth, [1]
  - (ii) the inter-quartile range, [2]
  - (iii) the depth at the 40th percentile, [2]
  - (iv) the number of days when the depth of the river was **at least** 25 cm. [2]

(b)

| $d$            | $0 < d \leq 10$ | $10 < d \leq 20$ | $20 < d \leq 30$ | $30 < d \leq 40$ | $40 < d \leq 50$ | $50 < d \leq 60$ | $60 < d \leq 70$ |
|----------------|-----------------|------------------|------------------|------------------|------------------|------------------|------------------|
| Number of days | 17              | 41               | 62               | 98               | 85               | $p$              | $q$              |

Show that  $p = 47$  and  $q = 15$ . [2]

(CIE Paper 4, Jun 2004)