

CANDIDATE  
NAME

--

CENTRE  
NUMBER

--	--	--	--	--

CANDIDATE  
NUMBER

--	--	--	--

**STATISTICS**

**4040/01**

Paper 1

**For Examination from 2018**

SPECIMEN PAPER

**2 hours 15 minutes**

Candidates answer on the Question Paper.

Additional Materials:      Pair of compasses  
   Protractor  
   Electronic calculator

**READ THESE INSTRUCTIONS FIRST**

Write your Centre number, candidate number and name in the spaces at the top of this page.  
Write in dark blue or black pen.  
You may use an HB pencil for any diagrams or graphs.  
Do not use staples, paper clips, glue or correction fluid.  
**DO NOT WRITE IN ANY BARCODES.**

Answer **all** questions.  
If working is needed for any question it must be shown below that question.  
Essential working must be shown for full marks to be awarded.  
Electronic calculators should be used.

At the end of the examination, fasten all your work securely together.  
The number of marks is given in brackets [ ] at the end of each question or part question.  
The total number of marks for this paper is 100.

This document consists of **22** printed pages.



**1** Six statistical measures are

mean,  
 median,  
 mode,  
 range,  
 interquartile range,  
 and standard deviation.

In each of the following situations, one of these measures is to be found by the person described. State the appropriate measure in each case.

**(a)** A doctor finds the most common age of her patients.

..... [1]

**(b)** An athlete who competes in the 100 metres sprint finds the difference between his slowest and quickest practice times.

..... [1]

**(c)** A graduate who seeks employment with a company finds a measure of central tendency for the salaries of the company's employees. The company has twenty employees, of whom three are managers earning salaries very much higher than the other employees.

..... [1]

**(d)** A teacher finds a measure of dispersion for the scores of her pupils in a test, in which no pupil scored an exceptionally high mark, and no pupil scored an exceptionally low mark.

..... [1]

2 A large keep fit class for women is held at a sports club once every week. The manager of the club asks the class instructor to select a sample of 10 women from the class.

(a) State the method of sampling used if the instructor decides to select

(i) the first 10 women to arrive at the class,

..... [1]

(ii) women at regular intervals from the class register.

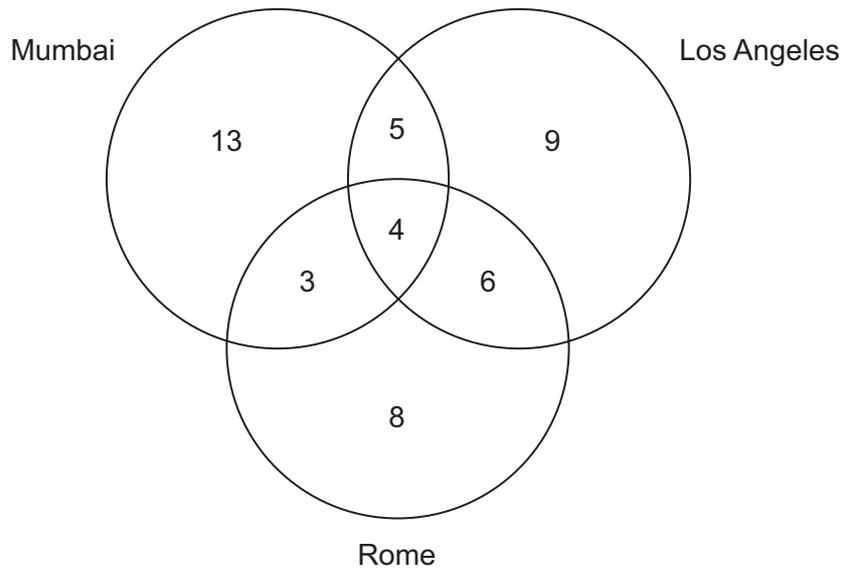
..... [1]

The sample is required to obtain responses to a proposal to change the time of the class from Monday evening to Monday afternoon. For class members the only items of data presently available to the instructor are name and age.

(b) State, and justify, one other item of data relating to class members which the instructor needs to know when selecting the sample in order to avoid bias in responses. You are **not** required to describe how the sample is selected.

.....  
.....  
.....  
..... [2]

- 3 The diagram below shows the number of actors at a film festival who have worked in one or more of the cities Mumbai, Los Angeles and Rome.



- (a) Find the number of actors who have worked in Mumbai.

..... [1]

- (b) Interpret the value 6 in the diagram.

.....  
 ..... [1]

A journalist selects one of these actors at random for interview.

Find the probability of selecting an actor who has worked in

- (c) Mumbai or Los Angeles or both,

..... [2]

- (d) Rome, given that the actor has worked in Mumbai and Los Angeles.

..... [1]

- 4 In a photographic equipment store a record was kept of the number of cameras sold each day. The values, for eleven consecutive days, were as follows.

6 0 8 2 1 6 0 9 6 4 1

For these values find

- (a) the mode,

..... [1]

- (b) the mean, correct to one decimal place,

..... [2]

- (c) the median.

..... [2]

The values recorded for the next three days were  $x$ ,  $x + 1$ , and  $x + 2$ .

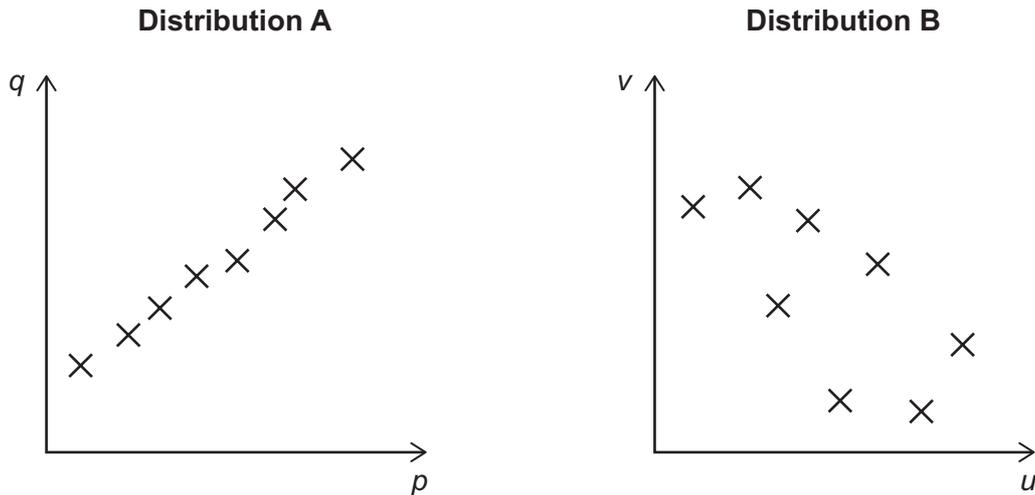
- (d) If the median for the entire fourteen-day period was the same as the median for the first eleven days, find  $x$ .

$x =$  ..... [1]

- 5 (a) Explain briefly what is meant by the statement that there is positive correlation between the two variables in a bivariate distribution.

.....  
 ..... [1]

The variables in two bivariate distributions were measured and plotted. The following scatter diagrams were obtained.



- (b) For each of these distributions, state whether the correlation shown is either positive or negative, and is either strong or weak.

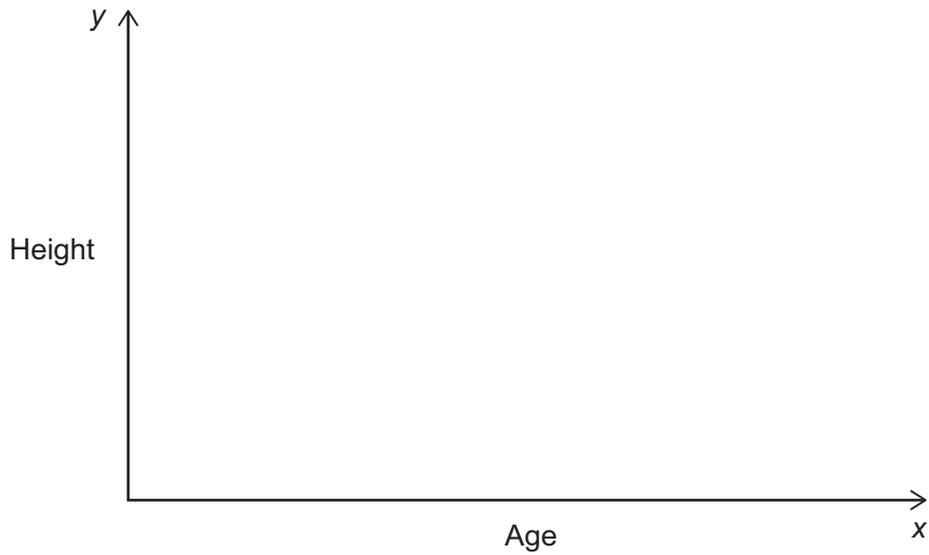
For Distribution A, it is ..... and .....

For Distribution B, it is ..... and .....

[2]

The age,  $x$ , and the height,  $y$ , of eight children of widely different ages were measured and recorded.

- (c) Sketch, using the axes below, the scatter diagram you would expect to obtain if the recorded values of  $x$  and  $y$  were plotted. You are **not** required to scale the axes.



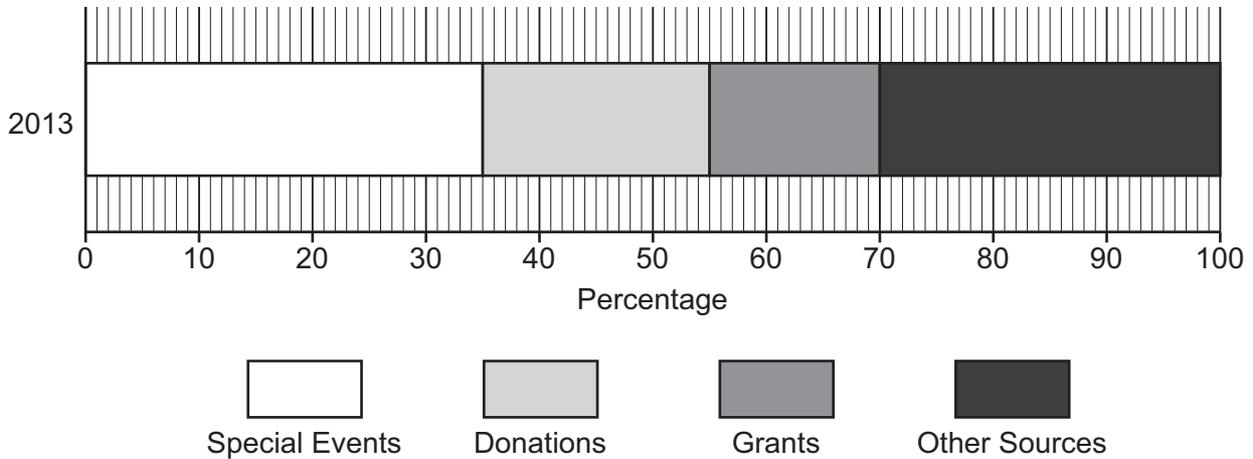
[2]

- (d) Describe fully the correlation shown by your sketch in part (c).

..... [1]

**6 In this question you are not required to draw any charts.**

A charity, Camfam, classifies the income it receives under the headings Special Events, Donations, Grants, and Other Sources. In Camfam's report for 2013, the following percentage bar chart was given, which represents a total income of \$80 million.



(a) Find the income which Camfam received in 2013 from Grants.

\$ ..... [2]

(b) If a pie chart were to be drawn to represent this information, find the angle which would represent the sector for Special Events.

..... [2]

Camfam's total income in 2014 was \$60 million.

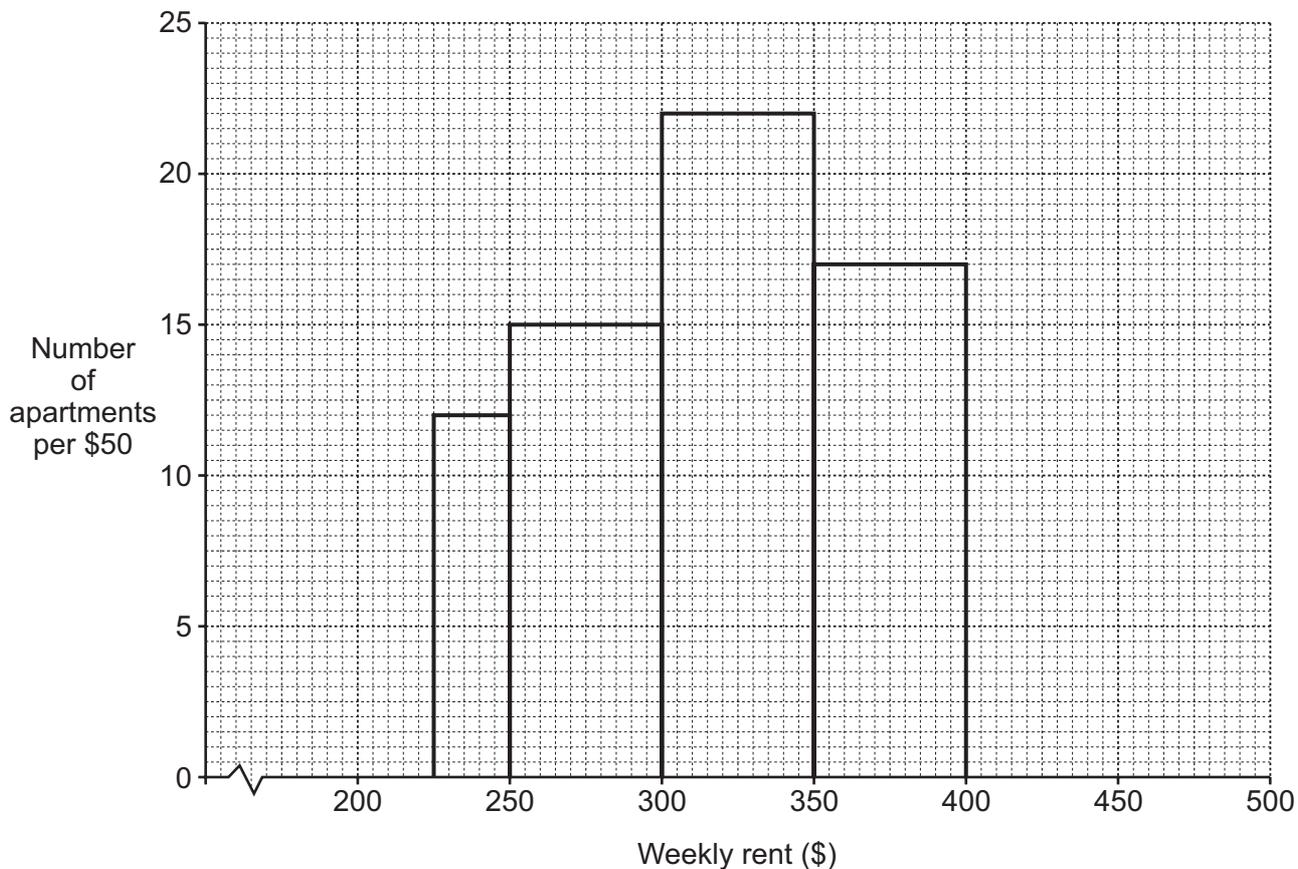
Two pie charts, one for 2013 and one for 2014, are to be presented together in a new report.

(c) If the radius of the chart representing 2013 is 4 cm, find, correct to one decimal place, the radius of the chart representing 2014.

..... [3]

**[Turn over for Question 7]**

- 7 In a large residential building there are 70 rental apartments. The weekly rents, in dollars, charged for these apartments are represented in the histogram below. One rectangle, representing the \$400 to under \$500 class, has been omitted from the histogram.



Use the histogram to find the number of apartments for which the weekly rent is

- (a) from \$250 to under \$400,

..... [2]

- (b) from \$225 to under \$250.

..... [2]

There are 10 apartments for which the weekly rent is from \$400 to under \$500.

- (c) Complete the histogram by drawing on the grid the rectangle representing the \$400 to under \$500 class.

[1]

- (d) Write down the term used to describe the \$300 to under \$350 class.

..... [1]

Two of the apartments in the building are chosen at random for safety checks.

- (e) Find the probability that both have a weekly rent of less than \$400.

..... [2]

- 8 The presenter of a radio programme plans his programme. The radio programme includes recordings of popular songs. For each song chosen he writes down the song length, in terms of time, in minutes, taken to play the song. The following table summarises the song lengths.

<i>Song length, <math>x</math> (minutes)</i>	<i>Number of songs, <math>f</math></i>	
$2.8 \leq x < 3.2$	3	
$3.2 \leq x < 3.4$	5	
$3.4 \leq x < 3.6$	9	
$3.6 \leq x < 3.8$	8	
$3.8 \leq x < 4.0$	7	
$4.0 \leq x < 4.2$	4	
$4.2 \leq x < 4.6$	2	

- (a) Estimate, in minutes, the mean and standard deviation of these song lengths. Give your answers to 3 significant figures.

Mean = .....

Standard deviation = ..... [8]

Information about five of the presenter's earlier programmes is shown below.

<i>Programme</i>	<i>Number of songs played</i>	<i>Mean of song lengths (minutes)</i>	<i>Standard deviation of song lengths (minutes)</i>
P	38	3.70	0.339
Q	39	3.52	0.328
R	42	3.69	0.294
S	37	3.83	0.305
T	38	3.74	0.291

**(b)** State in which of the programmes P, Q, R, S or T, songs were generally

**(i)** shortest in length,

..... [1]

**(ii)** most similar in length.

..... [1]

All the presenter's programmes are three hours in duration. Songs are not played continuously throughout each programme; for some of the time the presenter talks about the songs and the singers.

A listener switched on programme P at a random time during its transmission.

**(c)** Find the probability that a song was not being played at that moment.

..... [4]

- 9 In this question the fertility rate of a population is defined as the number of births per 1000 females.

The table below gives information about the female population and fertility rates by age group in a particular city for the year 2014, together with the standard population of females in the area in which the city is situated.

<i>Age group of females</i>	<i>Births</i>	<i>Population of females in age group</i>	<i>Fertility rate</i>	<i>Standard population of females (%)</i>
Under 20		2900	50	18
20–29		4500	184	22
30–39		5250	136	25
Over 39		5800	15	35

- (a) Calculate, to 1 decimal place, the standardised fertility rate for the city.

..... [4]

- (b) Calculate the number of births for each age group and insert the values in the table above.

[2]

(c) Calculate, to 1 decimal place, the crude fertility rate for the city.

..... [4]

There are equal numbers of males and females in the city and in the standard population.  
The standardised and crude death rates for the city in 2014 were 8.5 and 7.8 per thousand of the population respectively.

(d) Using **one** of these values, and any other appropriate values from parts (a), (b) and (c), find the increase in the population of the city in 2014 due to births and deaths.

..... [5]

- 10 The mid-day temperature at a particular location in a city was measured every day throughout the year 2014. The following table summarises the results obtained.

<i>Temperature (°C)</i>	<i>Number of days</i>	<i>Cumulative frequency</i>
0 – under 5	8	
5 – under 10	25	
10 – under 15	52	
15 – under 20	81	
20 – under 25	79	
25 – under 30	68	
30 – under 35	37	
35 – under 40	15	

- (a) Complete the cumulative frequency column in the above table. [2]

- (b) Plot the cumulative frequencies on the grid opposite, joining the points by a smooth curve. [3]

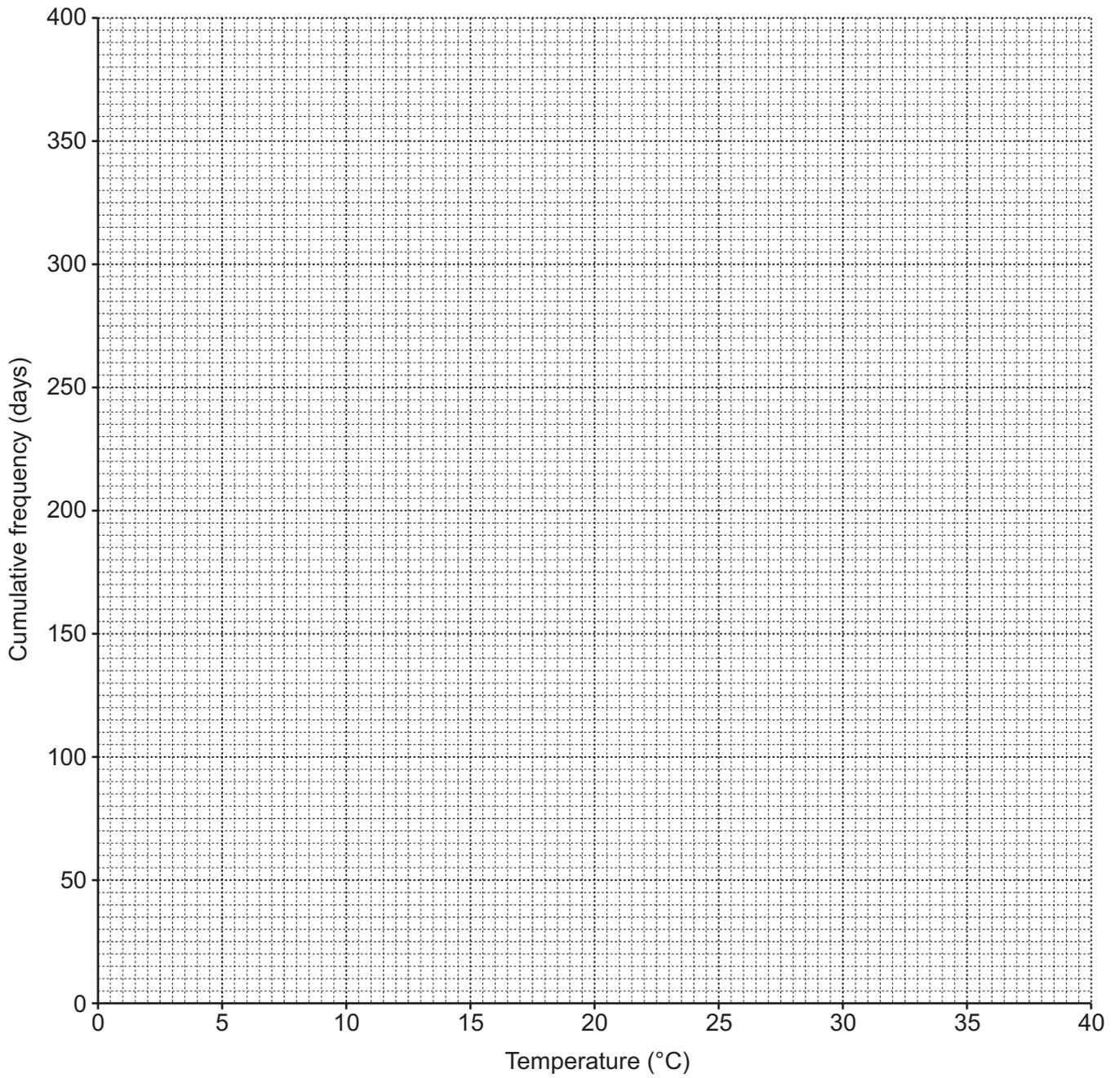
- (c) Use your graph to estimate

- (i) the median of these temperatures,

..... °C [1]

- (ii) the interquartile range of these temperatures.

..... °C [4]



[Question 10 continues on the next page]

When the results were obtained, a scientist predicted that, because of climate change, temperatures in the city would increase at the rate of  $0.5^{\circ}\text{C}$  every ten years. Assume that this prediction is accurate.

For this particular location,

**(d)** use your answers to part **(c)** to estimate, for the year 2050,

**(i)** the median of the mid-day temperatures,

.....  $^{\circ}\text{C}$  [2]

**(ii)** the interquartile range of the mid-day temperatures,

.....  $^{\circ}\text{C}$  [1]

**(e)** use your graph to estimate, for the year 2050, the number of days with a mid-day temperature of more than  $36^{\circ}\text{C}$ .

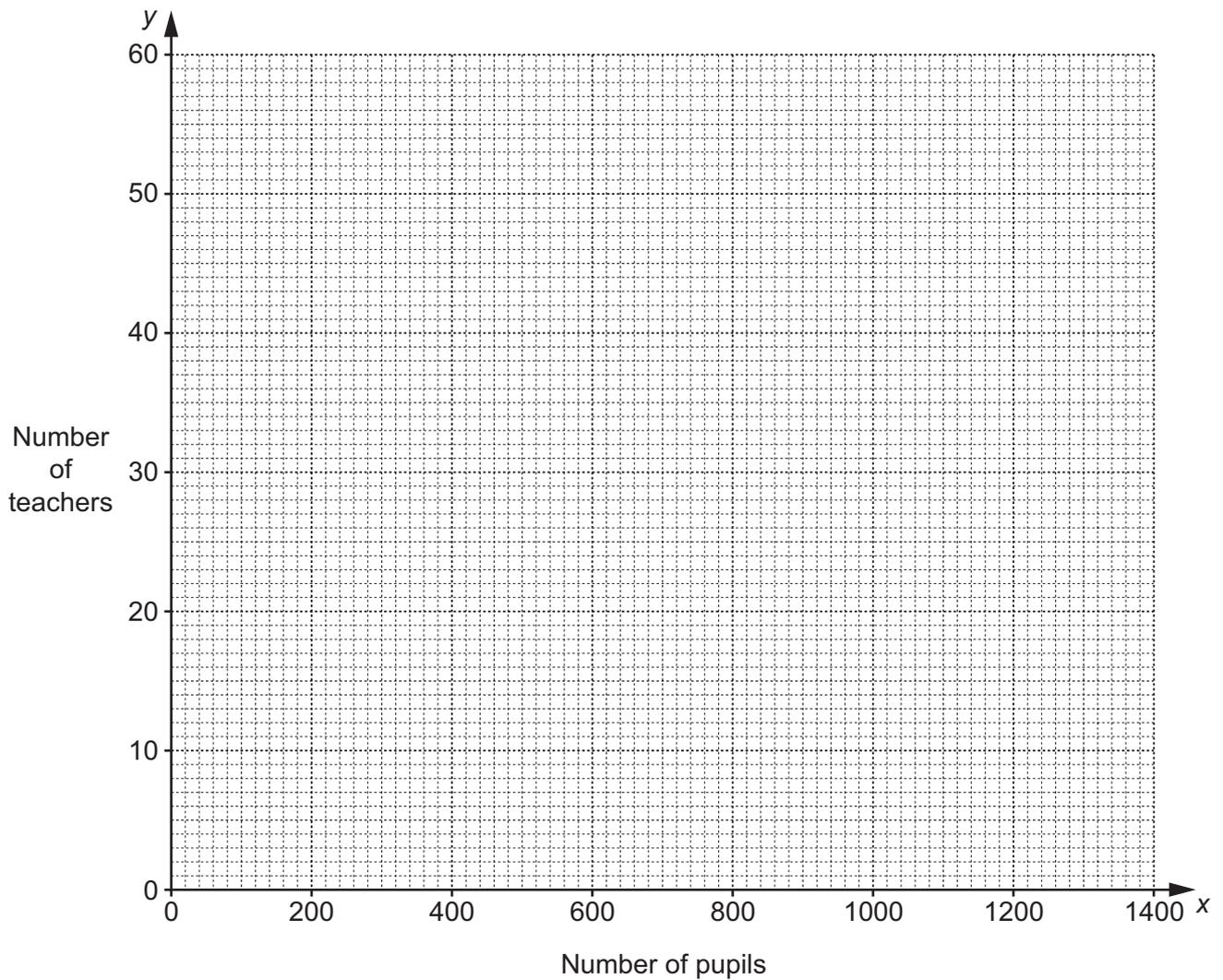
..... [2]

**[Turn over for Question 11]**

- 11 Emilie, a student teacher, conducted research on the number of pupils and the number of teachers in the schools in the town of Astra, where she lives. The schools supplied the following data.

School	A	B	C	D	E	F	G	H
Number of pupils, $x$	760	1219	927	470	1361	628	381	1085
Number of teachers, $y$	29	44	33	34	52	24	16	40

- (a) Plot these data on the grid below.



[2]

The data have an overall mean of (853.875, 34) and an upper semi-average of (1148, 42.25).

(b) Show how the value 1148 is calculated.

[2]

(c) Find the lower semi-average.

..... [2]

(d) **Without plotting the averages, and without drawing the line**, find the equation of the line of best fit in the form  $y = mx + c$ .

..... [3]

(e) Explain briefly why the value of  $c$  which you found in part (d) does not give useful information in this case.

.....  
..... [1]

Emilie discovered later that the data supplied by one of the schools gave, incorrectly, the total number of people employed by the school, and not the number of teachers.

(f) Ignoring the point representing the school which supplied incorrect data, draw, by eye, on the grid in part (a), a line of best fit through the remaining seven points. [1]

(g) Use the line you have drawn in part (f) to find its equation in the form  $y = mx + c$ .

..... [3]

[Question 11 continues on the next page]

Emilie repeated the research for schools in the nearby town of Belport, for which she found the equation of the line of best fit to be  $y = 0.0431x + 1.72$ .

- (h) Using this equation, and your answer to part (g), state in which of the two towns a pupil might choose to be educated, if free to choose. Explain your answer briefly.

.....

.....

..... [2]

---

Permission to reproduce items where third-party owned material protected by copyright is included has been sought and cleared where possible. Every reasonable effort has been made by the publisher (UCLES) to trace copyright holders, but if any items requiring clearance have unwittingly been included, the publisher will be pleased to make amends at the earliest possible opportunity.

Cambridge International Examinations is part of the Cambridge Assessment Group. Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.