

CANDIDATE
NAME

--

CENTRE
NUMBER

--	--	--	--	--

CANDIDATE
NUMBER

--	--	--	--



STATISTICS

Paper 1

4040/13

October/November 2014

2 hours 15 minutes

Candidates answer on the question paper.

Additional Materials: Pair of compasses
 Protractor

READ THESE INSTRUCTIONS FIRST

Write your Centre number, candidate number and name on all the work you hand in.

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 corrections fluid.

DO NOT WRITE IN ANY BARCODES.

Answer **all** questions in Section A and not more than **four** questions from Section B.

If working is needed for any question it must be shown below that question.

The use of an electronic calculator is expected in this paper.

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.

This document consists of **17** printed pages and **3** blank pages.

Section A [36 marks]

Answer **all** of the questions 1 to 6.

1 The number of passengers on each of 13 consecutive buses arriving at a terminus were

7 8 16 10 20 5 8 9 8 2 26 9 15 .

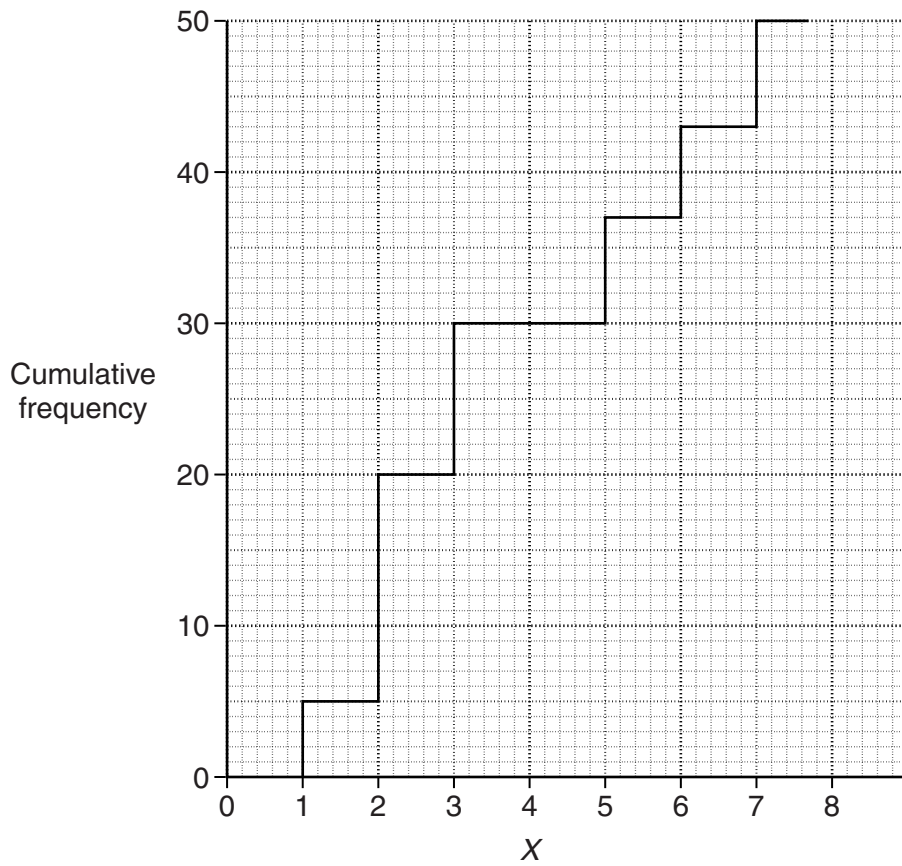
Three different measures of central tendency (average) of these numbers are 8, 9 and 11.

Complete the following table by giving, for each of these three measures, its name and a brief explanation of how its value has been obtained.

<i>Measure</i>	<i>Name</i>	<i>How obtained</i>
8
9
11

[6]

2 The variable X is illustrated in the cumulative frequency graph below.



(i) State, with a reason, whether the variable X is continuous or discrete.

.....
 [2]

(ii) State for which of the integer values shown in the graph the frequency of X is 0.

..... [2]

(iii) Complete the following table.

x	0	1	2	3	4	5	6	7
<i>Frequency</i>								

[2]

- 3 (a) The population of a town is tabulated in different age groups. A research organisation wishes to interview, from the population, a sample which represents it in terms of age. It proposes to do this using either stratified random sampling or quota sampling.

State one way in which the use of these sampling methods would be similar, and one way in which it would be different.

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

- (b) It is wished to obtain an estimate of the mean number of words on each page of a book. For each of the following methods state, with a reason, whether a sample obtained using it would be likely to be biased or unbiased:

- (i) counting the number of words on the last page of each chapter;

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

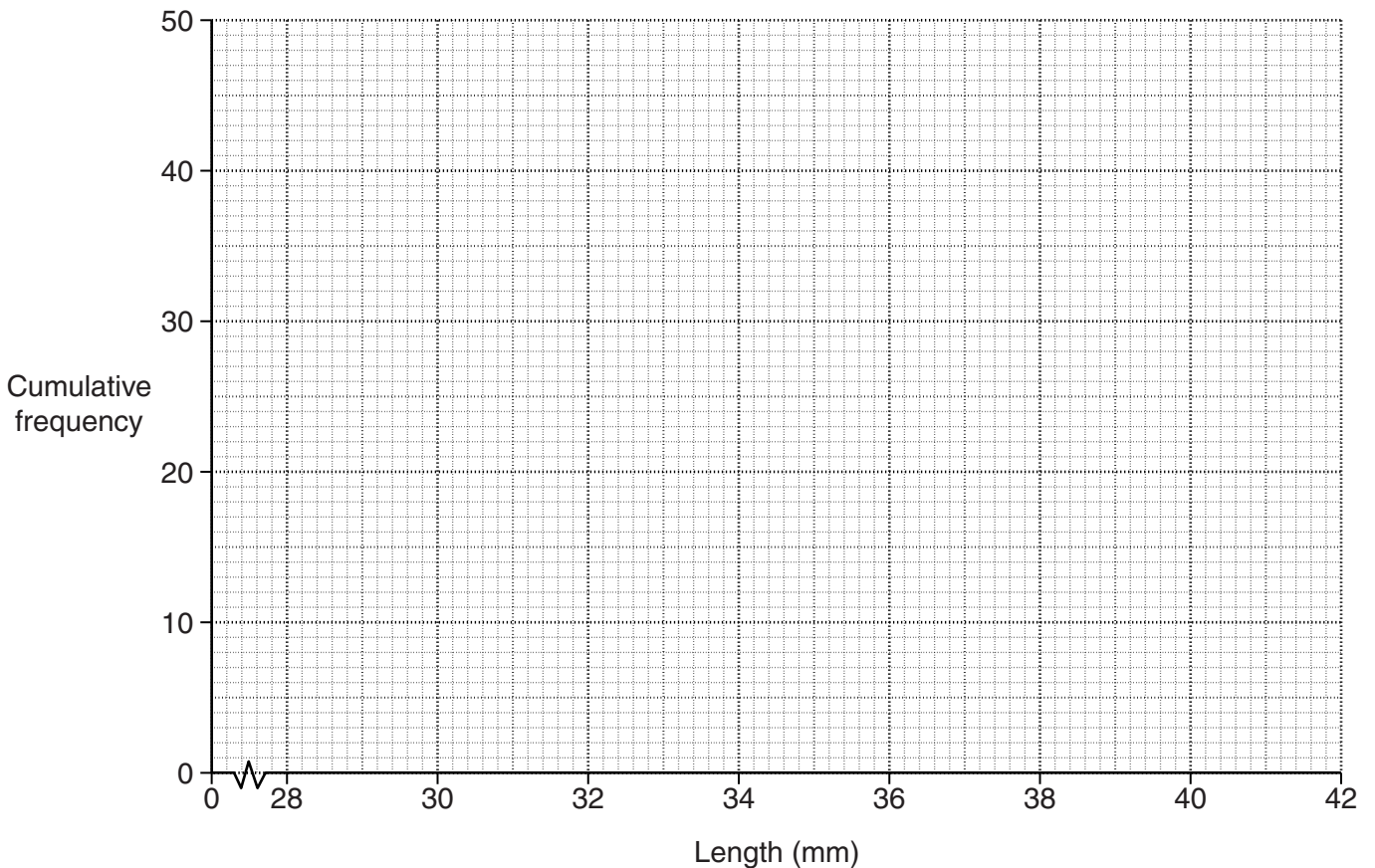
- (ii) counting the number of words on a systematic sample of pages.

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

- 4 The table below summarises the lengths, in millimetres, of a random sample of 50 leaves taken from a bush.

<i>Length (mm)</i>	<i>Frequency</i>	<i>Cumulative frequency</i>
Under 30	0	
30 – under 32	8	
32 – under 34	10	
34 – under 36	17	
36 – under 38	11	
38 – under 40	4	

- (i) Calculate the cumulative frequencies and insert them in the table. [1]
- (ii) Plot the cumulative frequencies on the grid below and draw a smooth curve through the plotted points. [2]



- (iii) Use your graph to estimate
- (a) the lower quartile length, mm [1]
- (b) the percentage of leaves that have a length greater than 37.2 mm.
..... [2]

5 A company which produces different sizes of sawn wood wishes to display information about the amount of sawn wood it produces in each of two consecutive years.

(i) State one advantage and one disadvantage of using a dual bar chart, as opposed to a percentage bar chart, to illustrate the amount produced in the two years.

Advantage.....
.....

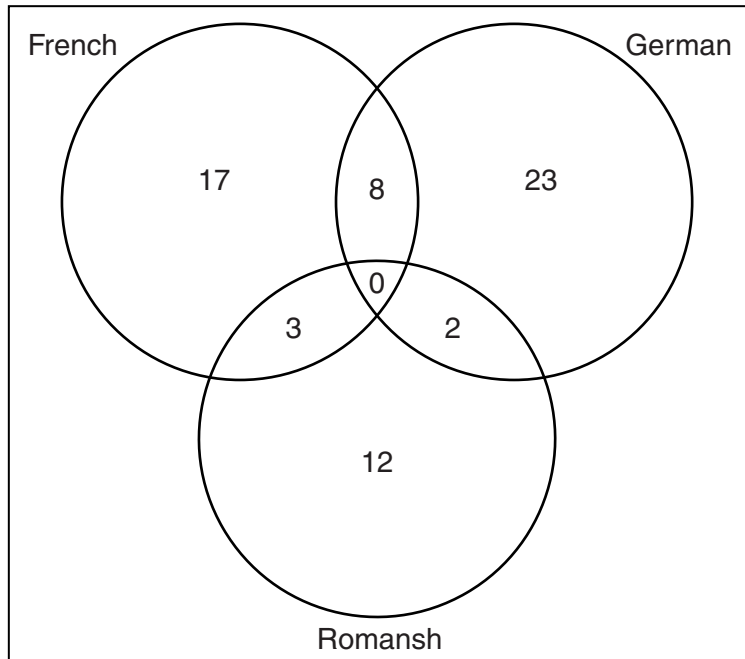
Disadvantage.....
.....[2]

(ii) Name a quantity which neither a dual bar chart nor a percentage bar chart would show.
.....[1]

(iii) State the names of two types of diagram which will give a relative indication of both the amount of different sizes of sawn wood and the total amount of sawn wood produced in each year.
.....
.....[2]

(iv) State the name of the type of diagram which will give a direct indication of the differences in the total amount of sawn wood produced from one year to the next.
.....[1]

- 6 Three of the official languages of Switzerland are French, German and Romansh. The diagram below illustrates which of these languages are spoken by a random sample of 70 Swiss citizens.



- (i) Find the value which should be written inside the box but outside the circles in order to complete the diagram.

.....[2]

- (ii) Interpret the value 0 in the diagram.

.....
[1]

- (iii) State, with a reason, in each of the following cases, whether the value 0 would be changed if the person described learned to speak Romansh.

- (a) One of the people denoted in the diagram by the value 17.

.....
[1]

- (b) One of the people denoted in the diagram by the value 8.

.....
[1]

- (c) One of the people denoted in the diagram by your answer to part (i).

.....
[1]

Section B [64 marks]

Answer not more than **four** of the questions 7 to 11.

Each question in this section carries 16 marks.

- 7 (a) Mr Hassan can travel to work by either car or train. The probability that on any day he travels by train is $\frac{4}{7}$. If he travels by car the probability that he will be late for work is $\frac{1}{9}$, but by train it is $\frac{1}{5}$.

Calculate the probability that on any randomly chosen day he is **not** late for work.

.....[4]

- (b) Three children are to be chosen at random from a group of seven, consisting of four boys, Ian, James, Michael and Nathan, and three girls, Karen, Lucy and Olive.

- (i) Calculate the probability that Ian, Lucy and Nathan are the three chosen.

.....[2]

Two of the seven are a brother and sister.

- (ii) Calculate the probability that the brother and sister will both be among the three chosen.

.....[3]

- (c) Sammy and Pecos each have a bag containing a number of blue balls and white balls. Each selects one ball from his bag at random. If the selected balls are of the same colour, Sammy puts them both in his bag; if they are of different colours, Pecos puts them both in his bag.

Originally, Sammy's bag contains 2 blue and 6 white balls, while Pecos' bag contains 3 blue and 5 white balls.

- (i) Calculate the probability that both selected balls are of the same colour.

.....[3]

- (ii) If, on the first selection, the balls were of the same colour (so both were put in Sammy's bag before a second selection), calculate the probability that on the second selection the balls are of different colours.

.....[4]

- 8 The following table summarises the times, x minutes, which the visitors to an art gallery during one day spent in the gallery. The first row of the table gives the column numbers.

(1)	(2)	(3)	(4)	(5)	(6)
<i>Time, x (minutes)</i>	<i>Frequency, f</i>	<i>m</i>	<i>y</i>	<i>fy</i>	<i>fy²</i>
0 – under 30	6				
30 – under 35	11				
35 – under 40	4				
40 – under 50	40				
50 – under 60	26				
60 – under 70	14				
70 – under 100	4				
TOTAL	105				

- (i) Insert in column (3) of the table the mid-points, m , of each of the classes.

[1]

- (ii) Values of a variable, y , are given by $y = \frac{m - 45}{2.5}$.

Calculate the value of y for each class and insert the values in column (4) of the table.

[2]

- (iii) For each class, calculate the value of the product fy , and insert the values in column (5) of the table. [1]

- (iv) For each class, calculate the value of fy^2 , and insert the values in column (6) of the table.

[1]

- (v) Calculate Σfy and Σfy^2 and insert their values in the table.

[1]

(vi) Estimate the mean of y .

.....[2]

(vii) Estimate the standard deviation of y .

.....[2]

(viii) Use your results to parts (vi) and (vii) to estimate

(a) the mean of x ,

.....[2]

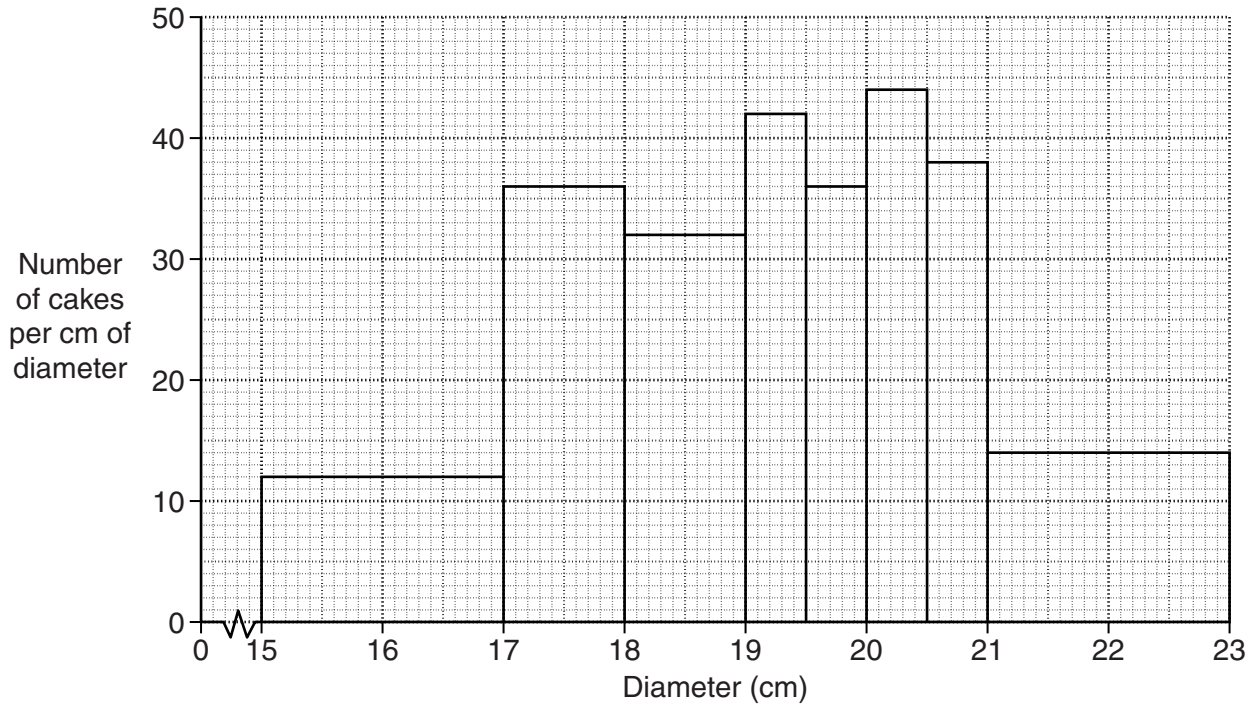
(b) the standard deviation of x .

.....[2]

(ix) Comment on whether or not, for these data, the interquartile range would be a more appropriate measure of dispersion than the standard deviation.

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

- 9 A bakery kept a record of the diameters, d centimetres, of the cakes it produced during one week. The results are summarised in the histogram below.



- (i) Use the histogram to complete the following grouped frequency table for d .

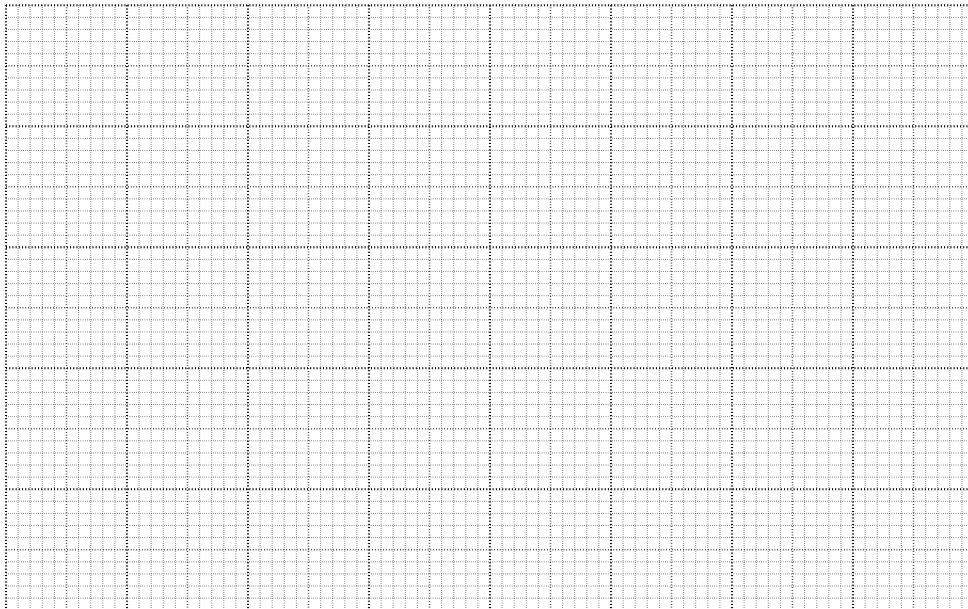
<i>Diameter, d (cm)</i>	<i>Frequency</i>
15 – under 17	
17 – under 18	
18 – under 19	
19 – under 19.5	
19.5 – under 20	
20 – under 20.5	
20.5 – under 21	
21 – under 23	

[4]

- (ii) Use the frequencies you have obtained to produce a simpler grouped frequency distribution, having four classes of equal width between 15 cm and 23 cm, and present your distribution in a table.

[3]

- (iii) On the grid below illustrate your simpler grouped frequency distribution by a histogram.



[3]

- (iv) Use the histogram you have drawn in part (iii) to estimate the modal diameter.

..... cm [2]

- (v) Cakes with a diameter between 16.5 cm and 22 cm can be sold in the bakery's shop. Find the percentage of this week's cakes which can be sold in the shop.

..... [4]

10 In this question calculate all death rates per thousand, and to 2 decimal places.

The first table below gives certain information about the population and deaths in a town, Eastbury, for the year 2012, together with the standard population of the area in which Eastbury is situated.

<i>Age group</i>	<i>Deaths</i>	<i>Population in age group</i>	<i>Standard population (%)</i>
0 – 14	25	4500	20
15 – 34	x	7000	35
35 – 59	47	6000	25
60 and over	83	7000	20

- (i) The death rate for the 15 – 34 age group is 3.00 per thousand.
Show that $x = 21$.

[1]

- (ii) Calculate the crude death rate for Eastbury.

.....[4]

- (iii) Calculate the death rates for the other three age groups.

0 – 14 age group.....
 35 – 59 age group.....
 60 and over age group.....[2]

- (iv) Using the given rate for the 15 – 34 age group, and the rates you have calculated in part (iii), calculate the standardised death rate for Eastbury.

.....[4]

The table below gives information about Westville, another town in the same area, for the year 2012.

The crude death rate for Westville in 2012 was 6.62 per thousand.

<i>Age group</i>	<i>Death rate per thousand</i>	<i>Population in age group (%)</i>
0 – 14	4	35
15 – 34	3	25
35 – 59	5	27
60 and over	24	13

- (v) Calculate the standardised death rate for Westville, using the same standard population as for Eastbury.

.....[2]

One of the two towns has a higher crude death rate, but the other has a higher standardised death rate.

- (vi) Give a brief explanation of why such a situation can occur.

.....
[1]

- (vii) State, with a reason, which of the two towns would appear to have the healthier environment.

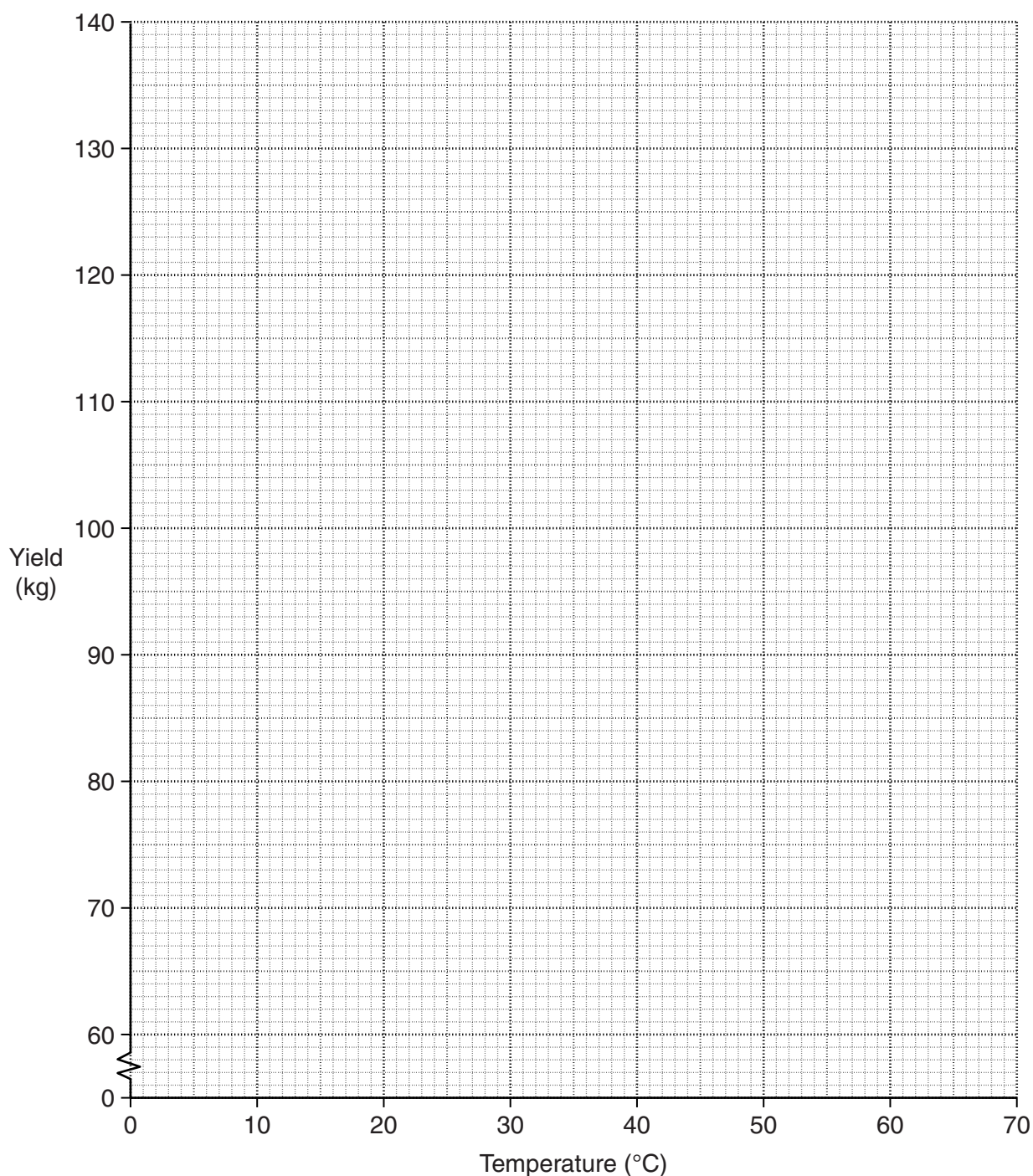
.....

[2]

- 11 Three trainee technicians, A, B and C, carried out laboratory trials to examine the effect of temperature, x , in $^{\circ}\text{C}$, on the yield, y , in kg, of an industrial process. The following table shows the results obtained by each technician.

<i>Technician</i>	A	B	C	A	B	C	A	B	C	A	B	C
<i>Temperature, x ($^{\circ}\text{C}$)</i>	10	15	20	25	30	35	40	45	50	55	60	65
<i>Yield, y (kg)</i>	80	106	75	90	117	118	97	127	80	109	140	115

- (i) Plot the points representing these results on the grid below and label each point A, B or C according to which technician carried out the trial.



[3]

(ii) Calculate and plot the overall mean.

[3]

The two semi-averages are (22.5, 97.7) and (52.5, 111.3).

(iii) Plot the semi-averages and use the three plotted averages to draw the line of best fit. [3]

It is known that over this range of temperatures the relationship between yield and temperature is approximately linear.

(iv) Comment on the results obtained by the three trainees.

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

An experienced and reliable technician carried out a trial at a temperature of 40°C and obtained a yield of 125 kg.

(v) Plot the experienced technician’s result on the graph. [1]

(vi) What might this extra information tell you about the performance of the trainees?

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

(vii) Use the extra information to draw, by eye, a revised line of best fit. [1]

(viii) Use this revised line of best fit to estimate the yield for a temperature of 52°C.

..... kg [1]

BLANK PAGE

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.