



UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS  
General Certificate of Education Ordinary Level

CANDIDATE  
NAME

CENTRE  
NUMBER

--	--	--	--	--

CANDIDATE  
NUMBER

--	--	--	--



**STATISTICS**

**4040/13**

Paper 1

**October/November 2012**

**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 a soft pencil for any diagrams or graphs.  
Do not use staples, paper clips, highlighters, glue or correction 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.

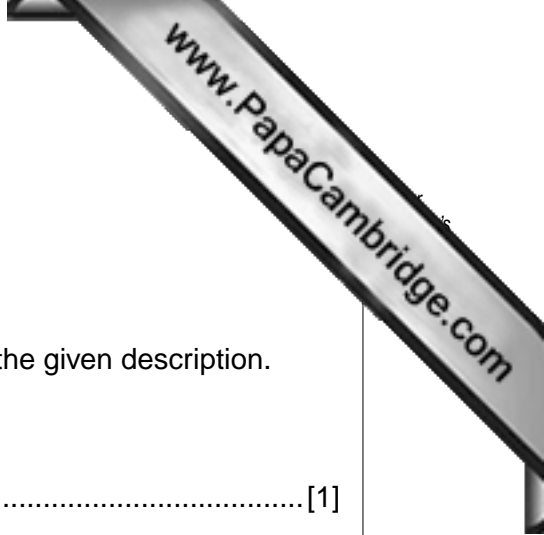
At the end of the examination, fasten all your work securely together.  
The number of marks is given in brackets [ ] at the end of each question or part question.

This document consists of **18** printed pages and **2** blank pages.



**Section A** [36 marks]

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



1 State, in each case, the name of a method of sampling which fits the given description.

(i) A method which never requires a sampling frame.

.....[1]

(ii) A method which only requires the use of random numbers to select the first item in a sample.

.....[1]

(iii) A method which guarantees that the proportions of different categories in a sample will exactly match those in the population being sampled.

.....[1]

(iv) A method which may produce a biased sample if there is some pattern in the population which repeats at regular intervals.

.....[1]

(v) A method in which, at the start of sampling, all items in the population have an equal chance of being selected for the sample.

.....[1]

(vi) A method in which, if a selected person or item is unavailable or cannot be located, a replacement person or item is selected.

.....[1]

- 2 A group of children are given a general knowledge test, and the following table gives, for each of them, their age in months and their score in the test. Each child is identified by a letter in the range A to J.

<i>Child</i>	A	B	C	D	E	F	G	H	I	J
<i>Age (months)</i>	148	174	126	180	160	108	138	192	158	189
<i>Test score</i>	60	72	36	82	61	20	40	86	50	81

The line of best fit method is to be used to estimate a child's test score from his/her age.

- (i) State, with a reason, which of age and test score should be used as the independent variable ( $x$ ) and which as the dependent variable ( $y$ ).

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

- (ii) State, with a reason, the letters corresponding to the children whose ages and scores you would use to calculate the lower semi-average.

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

- (iii) Calculate and state the values of the lower semi-average.

[2]

- 3 It is required to obtain the equation of the line of best fit of two variables,  $x$  and  $y$ , in the form  $y = mx + c$ .

Observations are taken on the two variables and the two semi-averages are calculated.

Lower semi-average	(2.55, 8.05)
Upper semi-average	(9.05, 21.05)

- (i) Use the two semi-averages to obtain the value of  $m$ .

$m = \dots\dots\dots$  [2]

The overall mean is calculated to be (5.80, 14.55).

- (ii) Use your value of  $m$  from part (i) and the overall mean to obtain the value of  $c$ , and hence state the equation of the line of best fit.

$\dots\dots\dots$  [3]

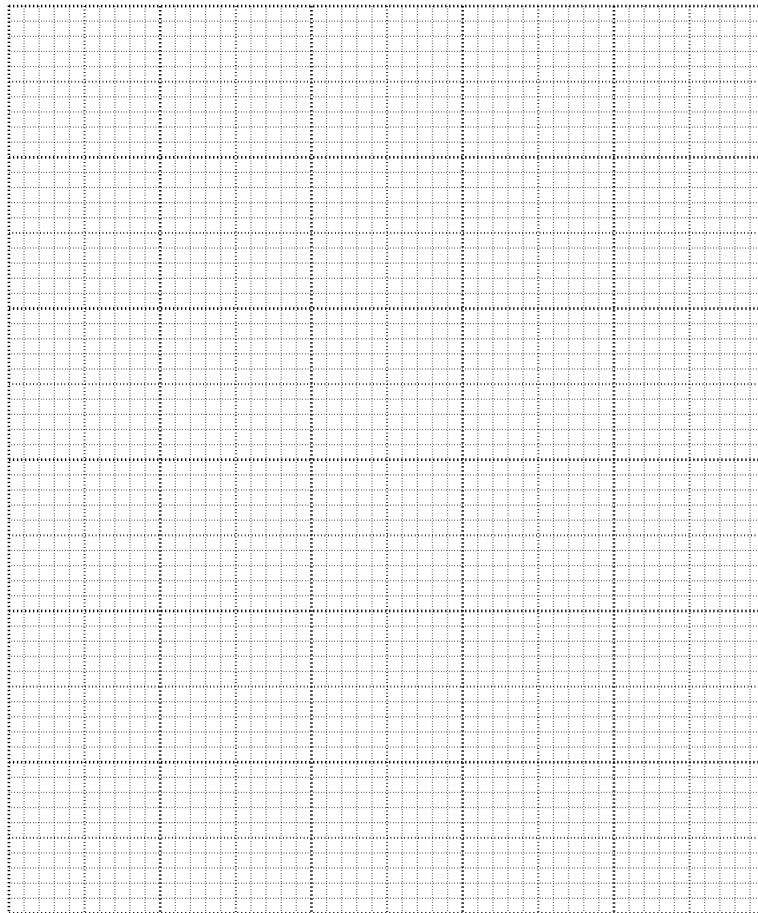
- (iii) State the reason why the same equation would have been obtained if the overall mean and a semi-average had been used to obtain the value of  $m$ .

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

- 4 The table below summarises the number of unsuccessful attempts which a sample of people made before passing their driving test.

<i>Number of unsuccessful attempts</i>	<i>Number of people</i>	<i>Cumulative frequency</i>
0	51	
1	34	
2	10	
3	4	
4	1	

- (i) Calculate the cumulative frequencies and insert them in the table. [1]
- (ii) On the grid below, draw and fully annotate, using suitable scales, an **appropriate** cumulative frequency graph to illustrate these data.



[5]

- 5 Twelve families live in a certain street. The table below summarises the number of girls and boys in each of these families.

		<i>Number of boys in the family</i>			
		0	1	2	3
<i>Number of girls in the family</i>	0	1	1	1	0
	1	1	1	2	1
	2	0	0	2	0
	3	0	1	1	0

For example, there is one family with two boys and three girls.

- (i) Construct the frequency distribution of the number of children per family, and present it in an appropriate table.

[3]

- (ii) Complete the frequency distribution table below to show the number of brothers that the girls have.

<i>Number of brothers</i>	<i>Number of girls (frequency)</i>
0	
1	
2	
3	

- 6 (i) Name one measure of central tendency and one measure of dispersion which would be distorted by the presence of one extreme observation in a set of data.

Measure of central tendency: .....

Measure of dispersion: ..... [2]

- (ii) Name one measure of central tendency and one measure of dispersion which could still be calculated for a set of data even if the largest value in it were unknown.

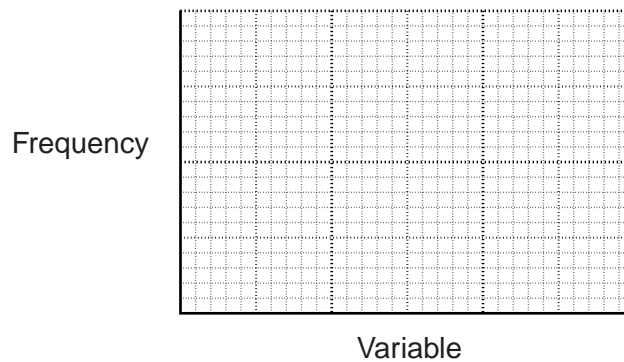
Measure of central tendency: .....

Measure of dispersion: ..... [2]

- (iii) The units in which a particular variable is measured are cm. State the units of the variance of this variable.

..... [1]

- (iv) Draw a sketch of a distribution in which the mean and median are equal, but in which the mode is not equal to them.



[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) A boy has two unbiased coins, one silver and one bronze. In an experiment he tosses the silver coin, and if a tail lands uppermost tail (T) is recorded. If a head lands uppermost he tosses the bronze coin, and whichever face lands uppermost is recorded, head (H) or tail (T).

(i) State the possible **recorded** outcomes and the probability associated with each.

[4]

- (ii) If the experiment is carried out twice, calculate the probability that on both occasions a tail is recorded.

.....[2]

- (b) Of a large bird population near a garden  $\frac{1}{3}$  are thrushes,  $\frac{1}{4}$  are starlings and the rest are robins. Assuming that the birds behave independently of each other, calculate the probability that three birds landing in the garden will

(i) all be thrushes,

.....[1]

- (ii) include one bird of each of the three species.



On another day, two robins, one starling and one thrush are in the garden.

- (iii) Under the same assumption, calculate the probability that they will fly away in alphabetical order.

.....[2]

- (c) Four people, including the car owner and his wife, are going on a journey in a four-seater car. In how many different ways can they be seated in the car if

- (i) the car owner is to drive and his wife is to sit next to him,

.....[1]

- (ii) the car owner is to drive and the other three people may be seated anywhere,

.....[1]

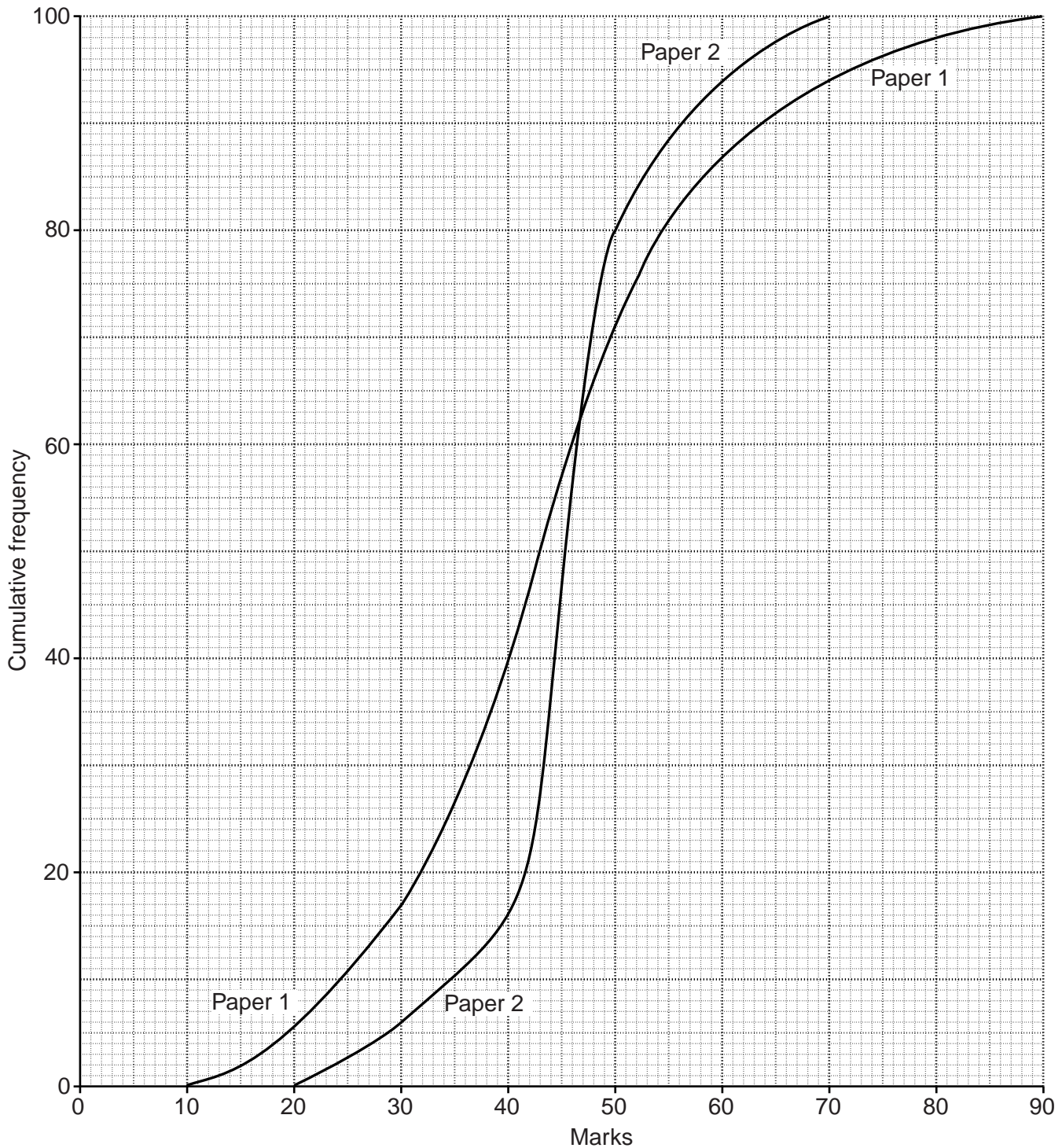
- (iii) any of the four people may drive?

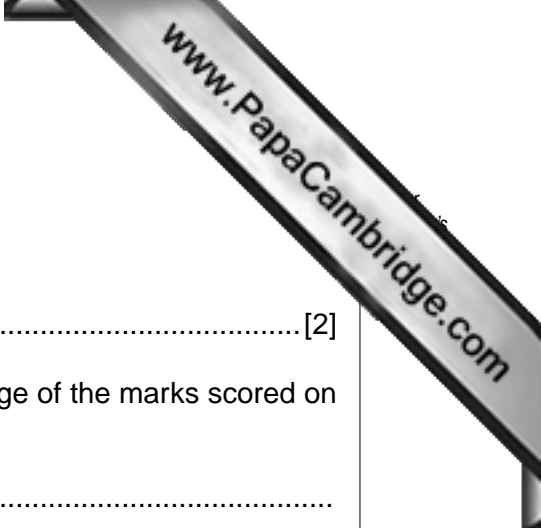
.....[1]



- 8 In answering all parts of this question you are required to show your working, either by the calculations leading to your result, or by drawing appropriate lines on the graph.

The cumulative frequency graph below illustrates the marks obtained by 100 candidates on each of two examination papers.





Use the graph to estimate

(i) the median of the marks scored on **Paper 2**,  
..... [2]

(ii) the lower quartile, the upper quartile and the interquartile range of the marks scored on **Paper 1**.

Lower quartile = .....

Upper quartile = .....

Interquartile range = ..... [4]

The maximum mark is the same on both papers.

State, in each case giving a reason,

(iii) on which paper the strongest candidates generally scored more marks,  
.....  
..... [2]

(iv) on which paper the weakest candidates generally scored more marks,  
.....  
..... [2]

(v) which paper gave the greater discrimination between the strongest and the weakest candidates.  
.....  
..... [2]

For each paper, a distinction is to be awarded to the 7% of candidates who achieve the highest marks on that paper.

(vi) Estimate the mark on Paper 1 and the mark on Paper 2 which would lead to distinctions being awarded.

Mark on Paper 1 = .....

Mark on Paper 2 = ..... [3]

(vii) Interpret the point at which the two lines on the graph intersect.  
.....  
..... [1]

9 In this question give all death rates per thousand, and those that are not exact, to 2 decimal places.

The following table gives the population of a city in the year 2010, together with various statistics relating to deaths in the city during that year, and the standard population of the area in which the city is situated.

<i>Age group</i>	<i>Population (thousands)</i>	<i>Number of deaths</i>	<i>Death rate (per thousand)</i>	<i>Standard population (%)</i>	<i>City population (%)</i>
0 – 4	200	2400	12	9	
5 – 24	$p =$	1260	1.5	29	
25 – 44	940	$q =$	3	30	
45 – 64	800	11 200	$r =$	22	
65 and over	350	38 850	111	10	

(i) Calculate the values of  $p$ ,  $q$  and  $r$  and insert them in the table.

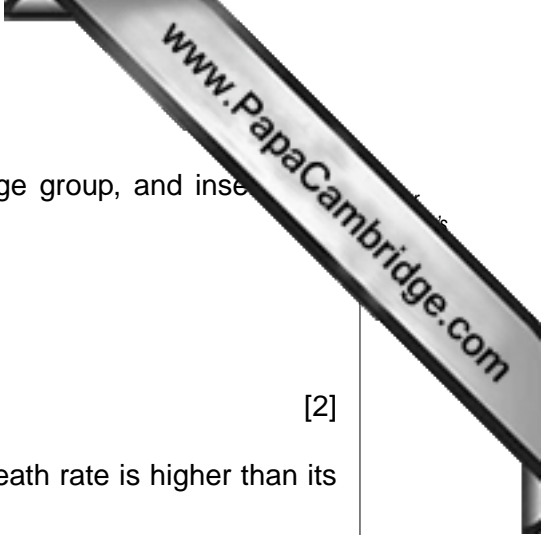
[3]

(ii) Calculate the crude death rate of the city.

.....[4]

(iii) Calculate the standardised death rate of the city.

.....[4]



(iv) Calculate the percentage of the city's population in each age group, and insert your results in the table.

[2]

(v) Using your results in part (iv), explain why the city's crude death rate is higher than its standardised death rate.

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

(vi) If the last two classes in the table were combined into a 45 and over class, state, with a reason, which of the rates in parts (ii) and (iii) would remain unaltered.

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

10 (a) In this part of the question you are NOT required to draw a histogram.

The table below shows the age distribution of the residents of a housing estate.

<i>Age (years)</i>	<i>Number of residents</i>
under 5	8
5 – under 11	9
11 – under 16	15
16 – under 21	23
21 – under 30	36
30 – under 40	35
40 – under 50	25
50 – under 65	22
65 and over	27

The residents in the 30 – under 40 group are to be represented in a histogram by a rectangle of height 35 units.

Calculate the heights of the rectangles in the same histogram representing

(i) the 40 – under 50 group,

..... units [1]

(ii) the 11 – under 16 group,

..... units [2]

(iii) the 50 – under 65 group.

..... units [2]

There are four times as many residents in the 21 – under 30 group as there are in the 5 – under 11 group.

(iv) Give a reason why the rectangle representing the 21 – under 30 group would not be four times as high as that representing the 5 – under 11 group.

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

- (b) A manufacturer of car tyres presented the following data summarising the life (in thousands of distance travelled) of a sample of 100 tyres.

<i>Distance</i>	18	19	20	21	22
<i>Frequency</i>	8	14	26	35	17

A statistician presented with this data assumed that the distances had been measured in thousands of kilometres, correct to the nearest thousand.

- (i) Estimate, to 3 significant figures, the mean and the standard deviation of the distances, under the statistician's assumption.

Mean = ..... thousands of km

Standard deviation = ..... thousands of km [5]

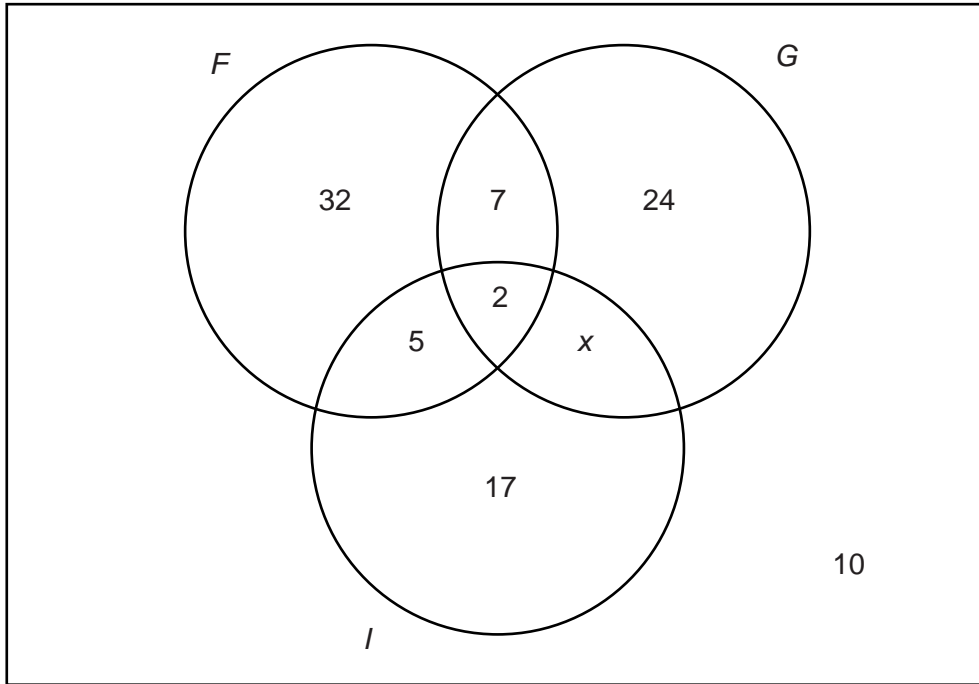
The statistician subsequently discovered that the distances in the table actually represented lower class limits (for example, the 18 represents any distance in the 18 – under 19 range), and also that they were in units of thousands of miles.

- (ii) Taking 1 mile to be equal to 1.6km, estimate the correct mean and standard deviation in metric units.

Mean = ..... thousands of km

Standard deviation = ..... thousands of km [5]

- 11 (a) At the end of the year 2009, 100 students were asked which of the three countries (*F*), Germany (*G*) and Italy (*I*) they had ever visited. The diagram below summarises their replies. Ten of the students had not visited any of the three countries.



- (i) Find the value of  $x$ .

$x = \dots\dots\dots$  [2]

- (ii) State what the value of  $x$  represents.

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



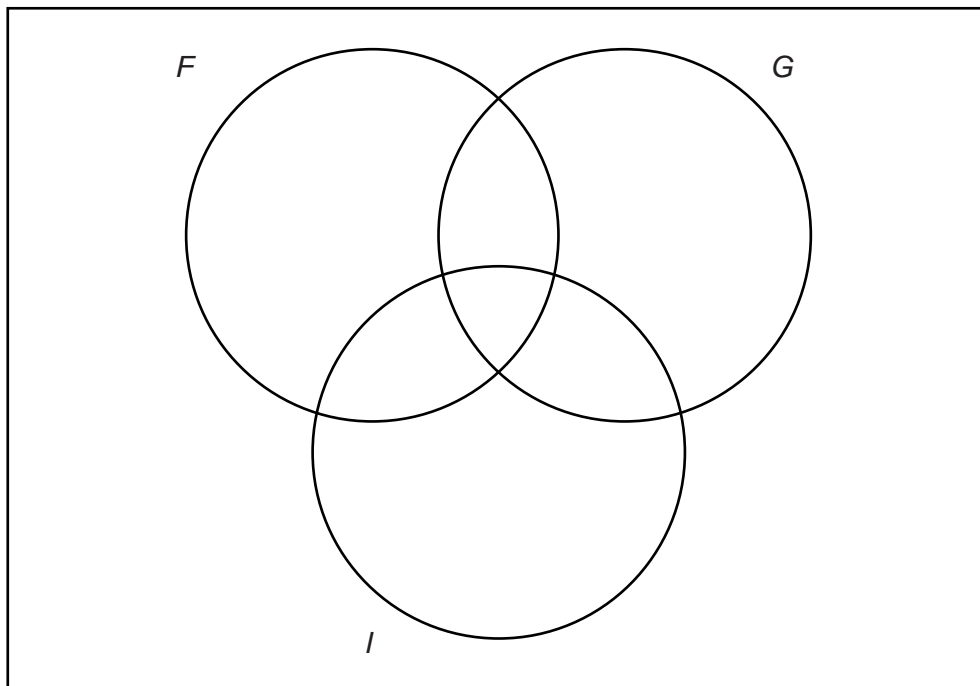
During the year 2010 the following visits took place.

Of the students who had previously visited only France, 2 visited Germany, and a different 1 visited Italy.

Of the students who had previously visited only France and Germany, 3 visited Italy.

Of the 10 students who had previously not visited any of the three countries, 1 visited France, a different 1 visited Italy, and a further 2 visited Germany.

- (iii) Complete the following diagram to illustrate the situation at the end of the year 2010.



[6]

**(b) In this part of the question you are NOT required to draw any pie charts.**

The treasurer of a tennis club classifies all club expenditure under four headings: maintenance of the club house and courts; tennis balls; wages paid to the club cleaner; taxes & insurance. In the year 2010 expenditure on these four items was in the ratio 4:3:2:1 respectively.

These amounts are to be represented by a pie chart of radius 7 cm.

**(i)** Calculate the angle of the sector representing expenditure on tennis balls.

Angle = .....[3]

In the previous year, 2009, total expenditure had been 90% of what it was in 2010.

**(ii)** Calculate, to 1 decimal place, the radius of the pie chart which would be drawn to represent expenditure in 2009.

..... cm [2]

In 2011, because of a large increase in membership, the expenditure on tennis balls was double that in 2010, but all other expenditure remained constant.

**(iii)** Calculate, to the nearest degree, the angle of the sector which would represent the expenditure on tennis balls in a pie chart for the year 2011.

Angle = .....[2]



