UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
General Certificate of Education Ordinary Level

CANDIDATE
NAME

CENTRE
NUMBER

CANDIDATE
NUMBER

STATISTICS **4040/13**

Paper 1 **October/November 2013**

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 **19** printed pages and **1** blank page.

UNIVERSITY *of* CAMBRIDGE
International Examinations

[Turn over

Section A [36 marks]

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

1 A survey was carried out to discover whether the quantity of traffic on a busy road was sufficient to justify the installation of a pedestrian crossing. At intervals throughout one day an investigator recorded the number of vehicles passing the proposed location in periods of 30 seconds duration.

The numbers he recorded were:

12 51 64 55 51 61 31 22 ■ 20 15 34 14 69 35

When his record sheet was examined the number shown here as ■ was illegible, but it was certainly a single-digit number.

Although this number is unknown, name, but do **not** calculate,

(i) two measures of central tendency (average) which can still be found,

....

....[2]

(ii) one measure of dispersion which can still be found,

....[1]

(iii) one measure of central tendency (average) which cannot be found,

....[1]

(iv) two measures of dispersion which cannot be found.

....

....[2]
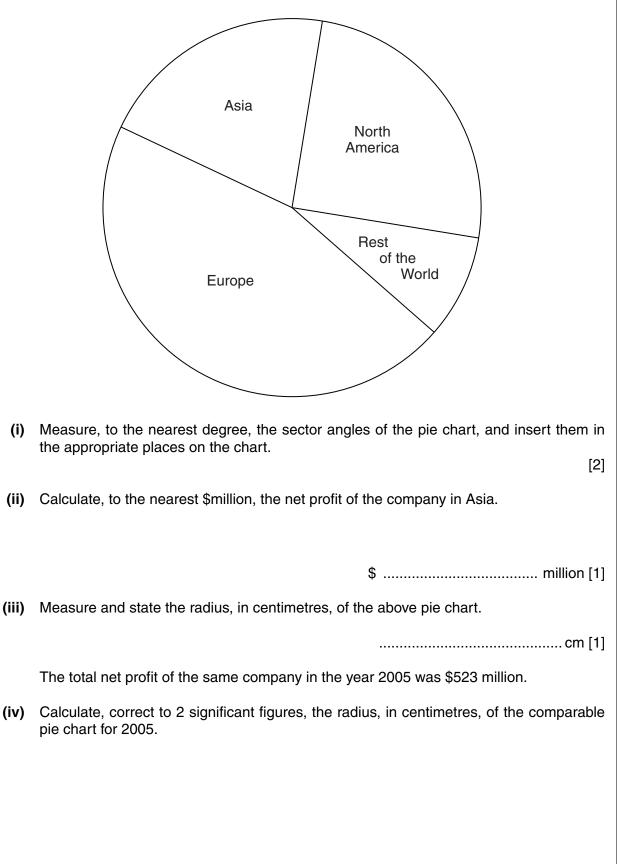
2 The pie chart below illustrates the distribution by location of the total net profit of $787 million earned by an international company in the year 2011.

 (i) Measure, to the nearest degree, the sector angles of the pie chart, and insert them in the appropriate places on the chart.

[2]

 (ii) Calculate, to the nearest $million, the net profit of the company in Asia.

$ .... million [1]

(iii) Measure and state the radius, in centimetres, of the above pie chart.

.... cm [1]

The total net profit of the same company in the year 2005 was $523 million.

(iv) Calculate, correct to 2 significant figures, the radius, in centimetres, of the comparable pie chart for 2005.

.... cm [2]

 [Turn over

3 A factory employs both male and female staff in each of the three categories managerial, inspection and production.

There are altogether 3500 employees, of whom 2150 are male. There are a total of 660 managerial staff, 540 male inspection staff and 785 female production staff.

(i) Insert these values in the appropriate places in the following table.

	Managerial	Inspection	Production	TOTAL
Male				
Female				
TOTAL				

[1]

Two thirds of the managerial staff are female.

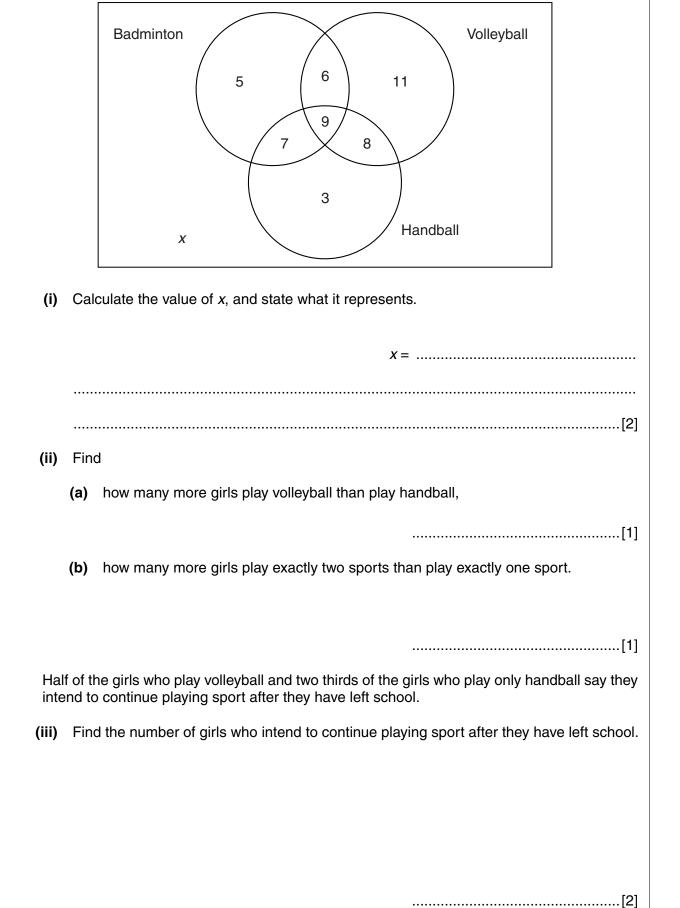**(ii)** Use this further information to complete the table.

[5]

4 There are 50 girls in their final year at a school. The diagram below illustrates the number of the girls who play each of the sports badminton, volleyball and handball.

(i) Calculate the value of *x*, and state what it represents.

x = ...

....

....[2]

(ii) Find

(a) how many more girls play volleyball than play handball,

....[1]

(b) how many more girls play exactly two sports than play exactly one sport.

....[1]

Half of the girls who play volleyball and two thirds of the girls who play only handball say they intend to continue playing sport after they have left school.

(iii) Find the number of girls who intend to continue playing sport after they have left school.

....[2]

5 **In answering this question you are not required to draw a histogram.**

For
Examiner's
Use

The times taken, in minutes, by 174 people to complete an aptitude test are summarised in the following table.

Time (minutes)	Number of people	Height of rectangle (units)
10 – under 30	28	
30 – under 40	36	18
40 – under 45	40	
45 – under 50	32	
50 – under 75	20	
75 – under 120	18	
TOTAL	174	

The times are to be illustrated by a histogram, in which the 30 – under 40 class is represented by a rectangle of height 18 units.

(i) Calculate the height of the rectangle representing the 40 – under 45 class, and insert the value in the table.

[1]

(ii) Calculate the heights of the rectangles representing the remaining four classes, and insert the values in the table.

[3]

(iii) If the final two classes were combined into a single 50 – under 120 class, calculate, to 2 decimal places, the height of the rectangle which would represent the combined class.

....[2]

6 (a) (i) Describe the situation which can lead to the method of systematic sampling producing a biased sample.

....

....

....[1]

(ii) There are 380 students at a college. It is proposed to take a systematic sample of 20 of the students. Explain briefly how this could be achieved.

....

....

....

....

....

....[3]

(b) Briefly explain how a population could be stratified, prior to taking a stratified sample, in order to ascertain the views of members of the public on

(i) a proposed increase in the tax on tobacco products,

....

....[1]

(ii) aircraft noise.

....

....[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 test for a particular disease has a 95% chance of correctly giving a positive result for a person who has the disease, but a 10% chance of incorrectly giving a positive result for a person who does not have the disease.

(i) Find the chance that the test gives a negative result for a person who has the disease, and insert it in the following table.

	Person has the disease	Person does not have the disease
P(test result positive)	0.95	
P(test result negative)		

[1]

(ii) Complete the table.

[1]

15% of the people who are tested are believed to have the disease.

A person is chosen at random and tested.

(iii) Calculate the probability that the test gives a correct result for this person.

....[4]

(b) **Give all probabilities in this part of the question as fractions.**

The following diagram classifies the members of a tennis club as to whether they are male or female, left-handed or right-handed, and whether or not they have represented the club in matches.

	Left-handed	Right-handed
Male	5	7
	Represented club	
	1	5
	3	4
Female	0	8

A member of the club is chosen at random.

(i) Calculate the probability that this member has represented the club in matches.

....[1]

A female member is chosen at random.

(ii) Calculate the probability that she is right-handed.

....[2]

A member who has represented the club in matches is chosen at random.

(iii) Calculate the probability that this member is left-handed.

....[2]

(c) Laura walks to school. On her route she passes two shops, *A* and *B*. The probability that she will go into shop *A* on any morning is 0.2, and into shop *B* is 0.7.
Her decision of whether to go into one of the shops is independent of whether she goes into the other shop. If she goes into either or both shops the probability that she will be late for school is 0.09.

    (i) Calculate the probability that on any morning she will go into exactly one shop and be late for school.

....[3]

Laura has been told that she must aim to be late on no more than 5% of the schooldays on which she goes into exactly one shop.

  (ii) State, with a reason, whether she is likely to achieve this target.

....

....[2]

8 (a) The table below summarises information about the number of GCE O Level subjects passed by different numbers of pupils at a school in the year 2011.

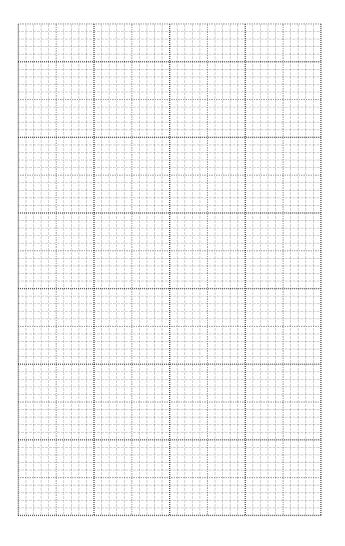| Number of subjects (x)	0	1	2	3	4	5	6
Number of pupils (frequency)	1	3	6	12	15	10	7
Cumulative frequency	1	4	10	22	37	47	54

An **appropriate** cumulative frequency graph is to be drawn to represent these data.

(i) On the grid below, draw and label two axes, the horizontal axis representing the number of subjects passed and the vertical axis representing cumulative frequency.

[2]

(ii) Draw an appropriate cumulative frequency graph to represent these data.

[4]

(b) The cumulative frequency graph below illustrates the lengths of journey times, in minutes, to their homes of a number of students at a college at the end of one particular day.

Use the graph to estimate

(i) the median journey time,

.... minutes [1]

(ii) the lower quartile time,

.... minutes [1]

(iii) the 90th percentile time,

.... minutes [1]

(iv) the number of students whose journey time was longer than 23 minutes,

....[3]

(v) the percentile corresponding to a journey time of 17 minutes.

....[2]

On the next day, due to bad weather, the journey time of all students was 5 minutes longer than the original times illustrated in the graph.

Compared with the original times, state, without further calculation, the effect which the bad weather had on

(vi) the upper quartile journey time,

....[1]

(vii) the interquartile range of journey times.

....[1]

9 The following table gives information about the populations and deaths in two towns, *A* and *B*, during the course of one year, together with the standard population of the area in which both towns are situated.

Age	Town *A*			Town *B*		Standard population
	Population	Deaths	Death rate (per thousand)	Population	Deaths	
0 – under 15	5000	45	*p* =	6000	66	400
15 – under 45	3750	15	4	27000	54	300
45 – under 65	2500	25	10	15000	60	200
65 and over	1250	*q* =	32	2000	30	100

(i) For town *A*, calculate the values of *p* and of *q* and insert them in the table.

[2]

(ii) Calculate the crude death rate of town *B*.

....[4]

(iii) Calculate the standardised death rate of town *B*.

....[4]

(iv) Use the population figures given in the table to state why the crude death rate and the standardised death rate of town *A* are equal.

....

....[2]

The table shows that far more deaths occurred in town *B* than in town *A* during the year, and yet the standardised death rate for town *B* is much lower than that for town *A*.

(v) Give two reasons why this situation has occurred.

....

....

....

....[2]

It was subsequently discovered that a small number of inhabitants of town *B*, none of whom had died during the year, had been misclassified by being included incorrectly in the 45 – under 65 class, when in fact they were all 65 and over.

(vi) State, with a reason, the effect, if any, which correcting this error would have on the crude death rate of town *B*.

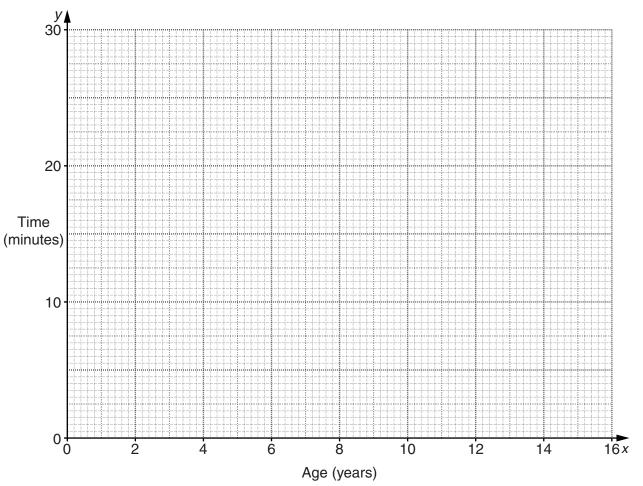..

....[2]

10 The time, in minutes, taken by each of 6 children to walk 1 kilometre, is given in the following table.

Child	A	B	C	D	E	F
Age in years (x)	13	8	7	15	12	9
Time in minutes (y)	12	23	25	11	18	23

(i) Plot these data on the grid below.

Age (years)

[2]

(ii) Calculate the overall mean and the two semi-averages of the data, and plot them on your graph.

[5]

(iii) Use your plotted averages to draw a line of best fit. [1]

(iv) Using any valid method, obtain the equation of your line of best fit, and write it in the form $y = mx + c$.

....[3]

(v) **Use your equation** to estimate, to the nearest minute, the time taken to walk 1 kilometre by a child aged 14 years.

....[1]

(vi) (a) Comment on how well your line of best fit matches the data points.

....

....[1]

(b) From the graph, identify the child for whom your line of best fit most overestimates the time taken.

....[1]

(vii) State, with a reason, whether it would be valid to use your line of best fit to estimate the time taken to walk 1 kilometre by a person whose age is outside the range of values given in the table.

....

....

....

....[2]

18

11 The following table summarises the increase, in dollars, of the annual income of a sample of 200 people between the years 2006 and 2011 (a negative value indicates a decrease).

Increase in annual income (x)	Class mid-point (m)	$y = \dfrac{m-750}{250}$	Frequency (f)	fy	fy²
−2500 – under 0			14		
0 – under 1500			99		
1500 – under 2500			39		
2500 – under 5000			25		
5000 – under 10000			23		
TOTAL			200		

(i) Obtain the mid-point, m, for each of the five classes and insert the values in the table.

[1]

(ii) For each class, obtain the value of the scaled variable, y, where

$$y = \frac{m-750}{250},$$

and insert the values of y in the table.

[2]

(iii) Obtain the values of Σfy and Σfy^2 and use them to estimate the values of the mean of y and the variance of y.

Mean = ...

Variance = ...[7]

(iv) **Use your results from part (iii)** to estimate

    (a) the mean of x,

....[2]

    (b) the variance of x.

....[3]

(v) State the units of the variance of x.

....[1]

20

BLANK PAGE