UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
General Certificate of Education Ordinary Level

CANDIDATE NAME

CENTRE NUMBER

CANDIDATE NUMBER

STATISTICS **4040/13**

Paper 1 **October/November 2011**

2 hours 15 minutes

Candidates answer on the question paper.

Additional Materials: Mathematical tables
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.

DC (RCL (PL/CGW)) 34787/5

UNIVERSITY of CAMBRIDGE International Examinations

[Turn over

Section A [36 marks]

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

1 The number of patients in the waiting room of a medical surgery was recorded at noon on each of eleven different days. The recorded numbers were

3 5 7 8 9 1■ 7 5 5 6 4

The largest number, shown as 1■, was partially illegible, although it was clearly a two-digit number with first digit 1.

(i) Name and calculate **two** measures of average (central tendency) which can still be found.

[4]

(ii) Name and calculate **one** measure of dispersion which can still be found.

..... [2]

2 **(a)** Describe briefly the **difference** between the method of obtaining a simple random sample and that of obtaining a stratified random sample.

.....

.....

.....

.....[2]

(b) Describe briefly how you could obtain a systematic sample of the pupils in a school.

.....

.....

.....

.....[2]

(c) The results of a statistical survey may be subject to either or both of *bias* and *error*. Explain the difference between *bias* and *error* in this context.

.....

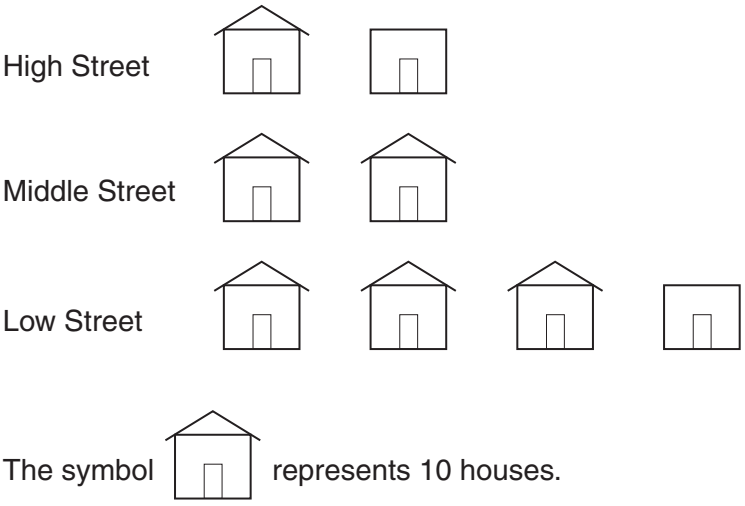
.....

.....

.....[2]

[Turn over

3 The pictogram below illustrates the number of houses in each of three streets in a town.



(i) State the number of houses in Middle Street.

.....[1]

(ii) State how many **more** houses there are in Low Street than in High Street.

.....[1]

In these three streets there are altogether 74 houses.

(iii) Find the number of houses represented by the symbol 

.....[2]

Two of the houses in these three streets are chosen at random.

(iv) Calculate, as a fraction in its lowest terms, the probability that both chosen houses are in Middle Street.

.....[2]

4 A survey was conducted in 47 households to discover how many cats and how many dogs each owned. No household owned more than 3 of either animal. The results are to be shown in the following table.

		Number of dogs				
		0	1	2	3	TOTAL
Number of cats	0	13				29
	1					12
	2					4
	3					2
	TOTAL	24	13	7	3	47

The given frequency shows that 13 households owned neither a cat nor a dog.

(i) Insert the appropriate frequencies into the table for each of the following pieces of information.

(a) 1 household owned two cats and one dog, and 2 households owned one cat and two dogs. [1]

(b) All other households which owned more than one of either animal owned none of the other. [1]

(ii) Complete the table.

[4]

5 During the course of one week a man recorded how much time he spent carrying out various tasks in his garden. He spent $\frac{1}{3}$ of the total time weeding and $\frac{1}{4}$ of the total time planting flowers and vegetables. The remaining time was split equally between mowing grass and all other work.

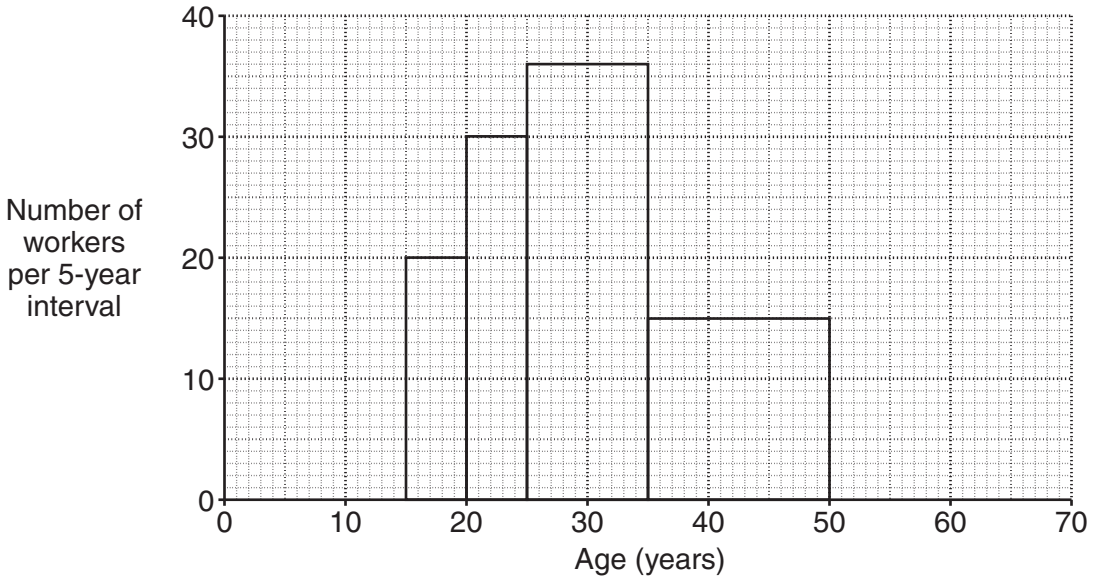
(i) Draw and label a pie chart of radius 5 cm to illustrate the proportion of the time spent by the man on each of these four tasks.

[4]

(ii) During the following week the man spent 50% **more** time working in his garden than in the first week. Calculate, to 2 decimal places, the radius of a comparative pie chart that would represent the second week. (You are **not** required to draw a second pie chart.)

.....cm [2]

6 The histogram below represents the age distribution of the workers in a factory.



Find

(i) the number of workers who were aged between 35 and 50,

.....[2]

(ii) the number of workers who were aged at most 35 years.

.....[2]

There were altogether 191 workers in the factory. All those not yet represented were aged between 50 and 70 years.

(iii) Draw a rectangle on the histogram to represent the 50 – 70 group.

[2]

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)** Oliver and Pierre are playing a game which requires that initially a fair coin is tossed to decide which of the two has the first turn. A turn involves rolling an unbiased six-sided dice with faces numbered 1, 2, 3, 4, 5 and 6. They keep rolling the dice alternately until one of them rolls a 4, at which point he is declared the winner.

(i) Calculate the probability, before the coin is tossed, that Pierre will win on his first turn.

.....[3]

(ii) Given that the coin has been tossed, and Oliver is to roll the dice first, calculate the probability that he will win on his second turn.

.....[2]

(b) A box contains 5 red discs and 7 blue discs. A girl draws a disc, notes its colour, and **replaces** it in the box. A boy then draws a disc and notes its colour.

Calculate the probability that

(i) both discs drawn are red,

.....[1]

(ii) the discs drawn are one of each colour.

.....[2]

The experiment is then repeated, using the same original contents of the box, except that the girl's disc is **not replaced** before the boy draws.

Calculate the probability that

(iii) both discs drawn are red,

.....[2]

(iv) the discs drawn are one of each colour.

.....[1]

(c) In a group of 100 people, each is classified by gender and according to whether or not he or she wears glasses. The results are tabulated below.

	Wears glasses	*Does not wear glasses*
Male	12	45
Female	10	33

One person is chosen at random from the group. Denote the event 'the chosen person is female' by A, and the event 'the chosen person does not wear glasses' by B.

(i) Define the events B' and $(A' \cup B)$.

.....

.....

.....

.....[2]

(ii) Calculate the probability of each of the events in part **(i)**.

$P(B') =$ .....

$P(A' \cup B) =$ .....[2]

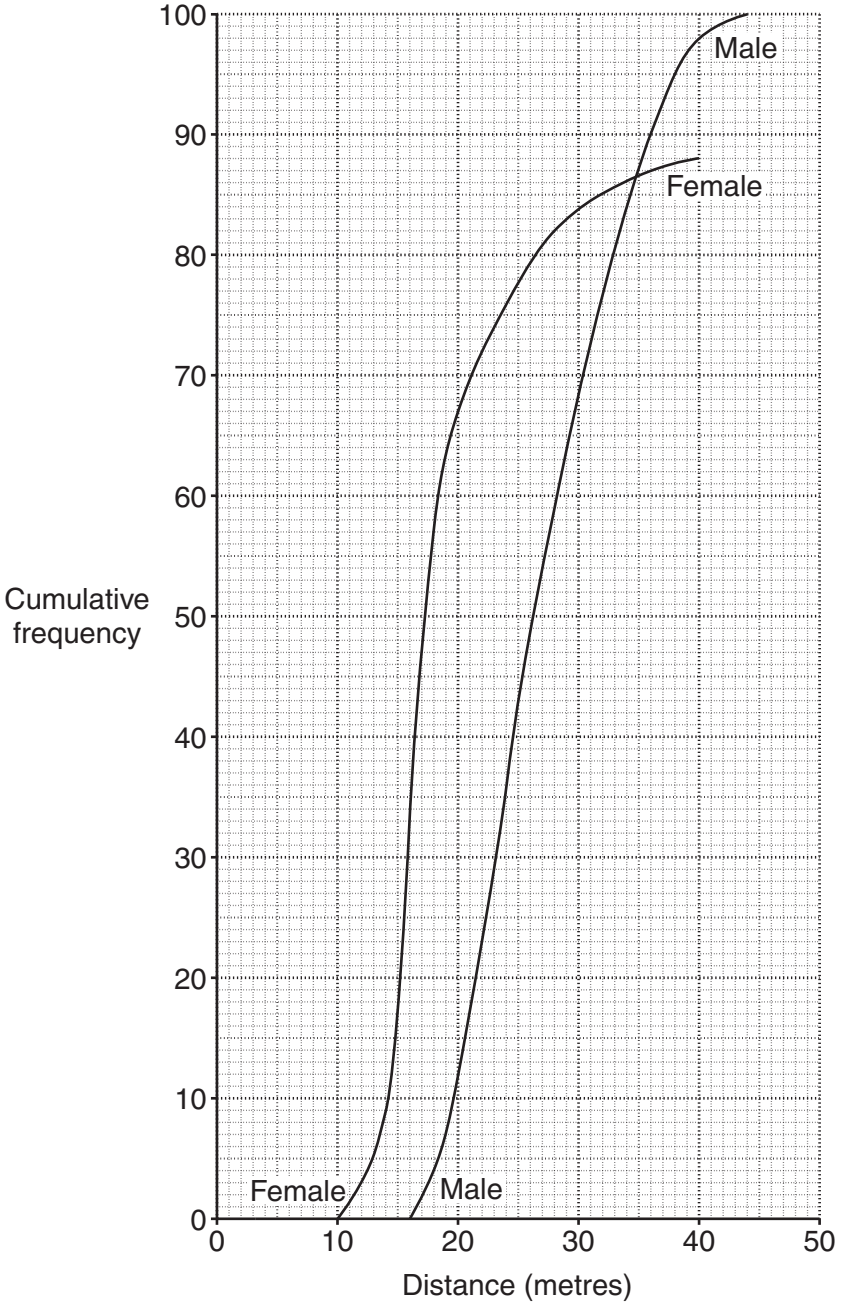
(iii) Given that the chosen person is male, calculate the probability that he does not wear glasses.

.....[1]

BLANK PAGE

[Turn over for Question 8]

8 A school wishes to establish a system of awarding points in each event in the school athletics championships. As a trial experiment, it has been decided to analyse data from past years for one particular throwing event. The distances thrown, in metres, have been summarised into two grouped frequency distributions, one for male athletes and one for female athletes. The cumulative frequency graphs illustrating these two distributions are shown below.



The further the throw, the more points are to be awarded. Each gender is to have its own points scale.

(i) For male athletes the shortest 10% of throws are to be awarded the minimum of 1 point. Estimate the shortest distance which a male athlete must throw to be awarded 2 points.

.....metres [2]

(ii) For male athletes, the central 40% of throws are to be awarded 3 points. Estimate the distances between which a male athlete must throw to be awarded 3 points.

..... metres and ..... metres [4]

(iii) Estimate the median, upper quartile and lower quartile distances thrown by **female** athletes.

Median = ... metres

Upper quartile = ... metres

Lower quartile = ... metres [5]

A quantity called the inter-quartile ratio is obtained by expressing the inter-quartile range as a percentage of the median.

(iv) Using your results from part **(iii)**, estimate, to 1 decimal place, the inter-quartile ratio for female athletes.

.....% [3]

(v) Are the positions of the two curves, relative to each other, what you would expect them to be? Give a reason.

.....

.....

.....[2]

9 **(a)** For summarising its census data, a country divides its population into three age groups, 0 to 10 years, 11 to 30 years, and 31 years and older. The *marriage* rates for the three groups at the last census were mistakenly published in the wrong order as 12 per thousand, 5 per thousand and 0 per thousand.

State, with a reason in each case, which marriage rate relates to

(i) the '0 to 10' group,

.....

.....[1]

(ii) the '11 to 30' group,

.....

.....[1]

(iii) the '31 and over' group.

.....

.....[1]

(b) The table below gives information about two neighbouring towns, *A* and *B*, together with the standard population for the area in which the towns are situated.

Age group	*Standard population*	*Town A*			*Town B population*
		Population	*Number of deaths*	*Death rate* (per thousand)	
0 – 24	3500	3000	45	*P*	2500
25 – 49	3000	2500	*Q*	8	1500
50 and over	2500	*R*	30	20	3000

For Town *A*, calculate

(i) the values of *P*, *Q* and *R*,

P = ...

Q = ...

R = ...[3]

(ii) the crude death rate per thousand, to 2 decimal places,

.....[2]

(iii) the standardised death rate per thousand, to 2 decimal places.

.....[4]

For Town *B* the crude death rate is 14.0 per thousand, and the standardised death rate is 12.5 per thousand. One of these rates is larger than the corresponding rate for Town *A*, but the other is smaller.

(iv) State, with a reason, which rate should be used to compare the chances of survival in the two towns, and which town gives the better chance.

.....

.....

.....

.....[3]

(v) State what causes the other rate to lead to an incorrect conclusion **in this case**.

.....

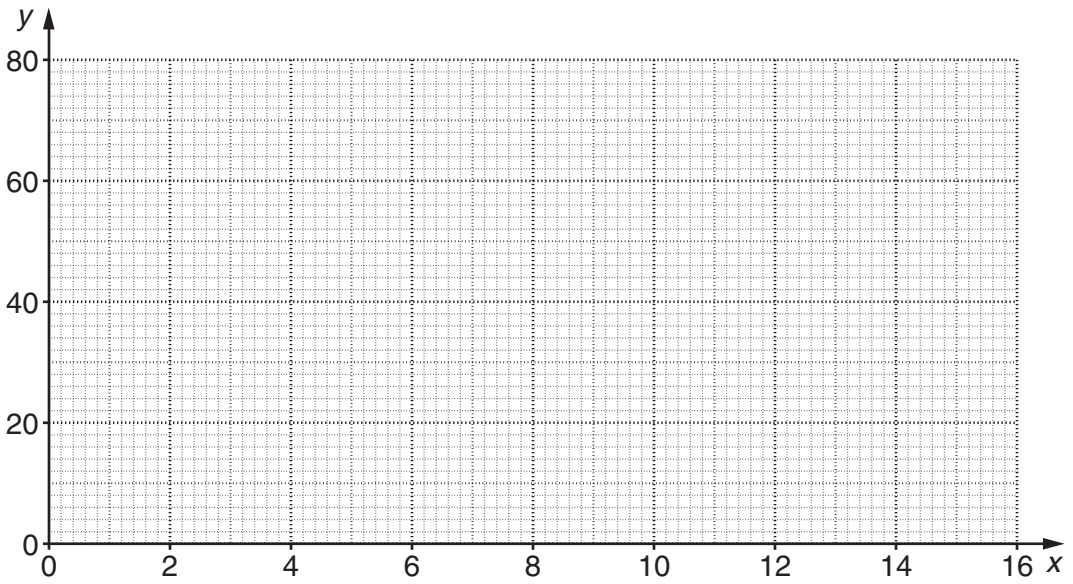
.....[1]

10 The following table gives data for eight geographical areas derived from the 1991 UK Census.

Area	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
X	12.8	7.3	7.6	15.2	3.1	6.6	5.0	7.9
Y	77	52	57	73	33	49	32	51

X is the percentage of the population holding university qualifications.
Y is the percentage of houses owned by the occupiers.

(i) Plot points representing these eight areas on the grid below. [3]



(ii) Calculate the overall mean point and plot it on the grid.

[2]

(iii) Explain why, in calculating the semi-averages, you should use areas *B*, *E*, *F* and *G* for one average, and areas *A*, *C*, *D* and *H* for the other.

.....

.....[1]

(iv) Obtain the two semi-averages and plot them on the grid.

[3]

(v) Draw the line of best fit using your plotted averages. [1]

(vi) Use your graph to obtain the equation of the line of best fit, giving it in the form $y = mx + c$.

$y =$ .....[3]

(vii) **(a)** Comment on how well the line you have drawn fits the points you have plotted.

.....

.....[1]

(b) Explain why the variable labelled *X* has been considered to be the independent variable.

.....

.....[1]

(c) If the variable labelled *Y* had been the independent variable, state which four areas would have been used to obtain the lower semi-average. (You are **not** required to calculate this semi-average.)

.....

.....[1]

11 A company doctor wished to assess whether there was an association between the amount of sleep employees achieved during nights prior to working days and their productivity. Fifty employees were asked to record, **to the nearest hour**, the total number of hours they slept during such nights in one particular week. The results are tabulated below.

31	36	37	48	38	34	37	36	44	36
38	37	38	29	42	36	40	26	37	34
39	38	37	41	33	38	38	36	37	39
36	37	35	37	36	37	37	43	36	43
36	37	36	32	38	35	37	38	41	40

The total number of hours slept by the fifty employees was 1857.

(i) Calculate, to 2 decimal places, the mean number of hours slept per employee.

.....[1]

The table below, from which some values have been omitted, summarises the data in the form of a grouped frequency distribution.

Hours of sleep (to nearest hour)	*Class mid-points (x)*	*Number of employees* (frequency)		
25 – 29		2		
30 –	31			
– 35	34	5		
36	36			
37		12		
38				
39 –	40	6		
– 44	43			
45 – 49		1		

(ii) Insert in the table all the missing values of class limits, class mid-points and frequencies. [5]

(iii) Estimate, to 2 decimal places, the mean of the data in your table.

.....[2]

(iv) Explain why the values for the mean number of hours slept which you have obtained in parts **(i)** and **(iii)** are different.

.....

.....

.....[2]

(v) Express the difference between the two means as a percentage of the true value, correct to 3 significant figures.

.....[2]

(vi) Using the data **in your table**, estimate, to 2 decimal places, the standard deviation of the number of hours slept by these employees.

.....[4]

BLANK PAGE