STATISTICS   **4040/01**

Paper 1   **October/November 2008**

2 hours 15 minutes

Additional Materials: Answer Booklet/Paper
                          Graph paper (3 sheets)
                          Mathematical tables
                          Pair of compasses
                          Protractor

READ THESE INSTRUCTIONS FIRST

If you have been given an Answer Booklet, follow the instructions on the front cover of the Booklet.
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.

Answer **all** questions in Section A and not more than **four** questions from Section B.
Write your answers on the separate Answer Booklet/Paper provided.
All working must be clearly shown.
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 **10** printed pages and **2** blank pages.

UNIVERSITY *of* CAMBRIDGE
International Examinations

[Turn over

Section A [36 marks]

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

1 A school wishes to take a 10% sample of its 500 pupils in order to investigate their opinions about its catering facilities. Four different suggestions are put forward as to how the sample should be selected. For each suggestion, state the method of sampling which would be used.

(i) Interviewers stand at the school gate at the start of a day, and interview every tenth pupil to arrive. [1]

(ii) The names of all pupils are listed alphabetically, allocated an individual number, and then a table of random numbers is used to select the sample. [1]

(iii) Interviewers stand at the school gate to interview pupils as they arrive, having been instructed to interview 25 girls and 25 boys. [1]

(iv) Pupils are chosen randomly using class registers, with the number selected from each class being in proportion to the class size. [1]

2 A man kept a record each year of the amount of money he spent travelling to and from work, and the amount spent travelling to and from leisure activities.

The following pie charts, which are **not** drawn to scale, illustrate these amounts for the years 2000 and 2005. The radii of the two charts are 2 cm and 3 cm respectively. In the year 2000, the **total** amount the man spent on work and leisure travel was $2317.
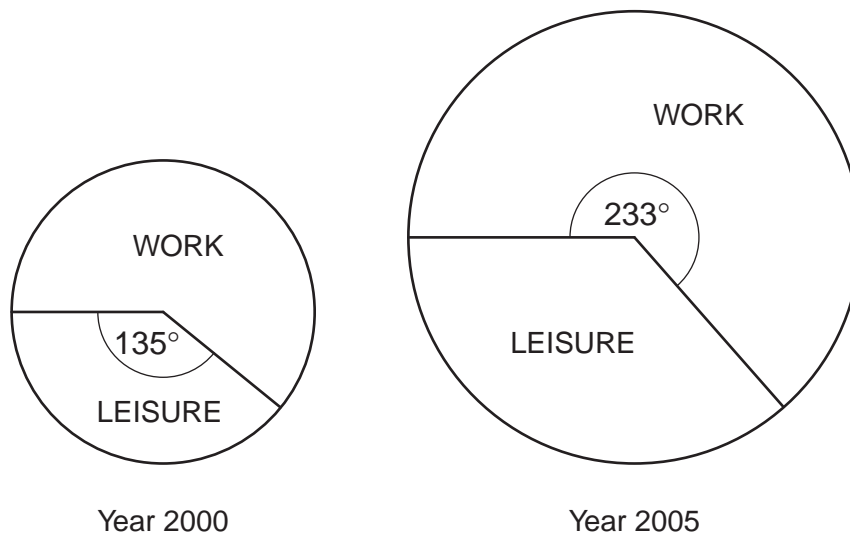
Year 2000   Year 2005

Calculate, in $, and giving your answers to 2 decimal places,

(i) the amount spent on work travel in the year 2000, [2]

(ii) the **total** amount spent on work and leisure travel in the year 2005, [2]

(iii) the amount spent on leisure travel in the year 2005. [2]

3 All of 100 households own at least one of a car, a motorbike and a bicycle. The following diagram illustrates how many of these 100 households own each of these methods of transport.
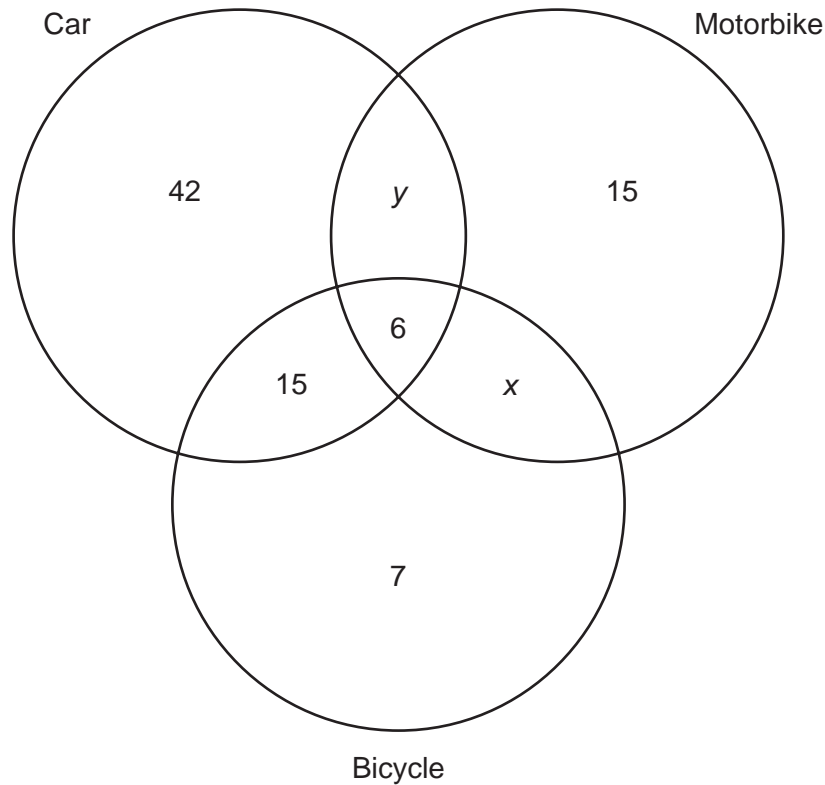
(i) A total of 31 households own a bicycle. Find the value of *x*. [1]

(ii) State what the value of *x* represents. [1]

(iii) Find the value of *y*. [2]

(iv) What percentage of households which own a car also own at least one of the other two methods of transport? [2]

4 The table below summarises the number of eggs in birds' nests observed during a survey carried out during the egg-laying season.

Number of eggs	0	1	2	3	4	5	6
Number of nests	9	11	10	6	20	5	3

For example, 10 nests each contained 2 eggs.

(i) Calculate the cumulative frequencies for these data. [2]

(ii) Using a scale of 2 cm to represent 1 egg on the horizontal axis, and 2 cm to represent a cumulative frequency of 10 on the vertical axis, draw an **appropriate** cumulative frequency graph to illustrate these data. [4]

5 Data has been collected on two variables, *x* and *y*. The following table gives the overall means and semi-averages calculated from these data.

	x	*y*
Overall mean	42	43
Lower semi-average	17	26
Upper semi-average	67	60

   (i) Use the values in the table to obtain the equation of the line of best fit for *x* and *y*, in the form $y = mx + c$. [4]

   (ii) Use your equation to estimate the value of *y* when $x = 33$. [2]

6 A man kept a record, (in the order Sunday, Monday, Tuesday, etc), of the number of journeys he made in his car on each day during the course of one particular week. The values, in the order in which he recorded them, were as follows.

$$2 \qquad 3 \qquad 5 \qquad 4 \qquad 1 \qquad 16 \qquad 1$$

   (i) Briefly explain why, although it is the middle value in the above list, 4 is not the median of this set of data. [1]

   (ii) State the value of the median. [1]

   (iii) State the value of the mode. [1]

   (iv) Calculate the value of the mean. [2]

   (v) State which of the median, mode or mean it would be most appropriate to use as a measure of average (central tendency) to represent these data. [1]

   (vi) For each of the two measures of average you did not choose in **(v)**, give a reason why you did not choose it. [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 The following table summarises the weight loss, x kg, of people who attended a course to assist people who wished to lose weight. A negative value of x indicates a gain in weight.

Weight loss (x kg)	Number of people
$-10 \leqslant x < 0$	6
$0 \leqslant x < 5$	9
$5 \leqslant x < 10$	15
$10 \leqslant x < 20$	16
$20 \leqslant x < 30$	8
$30 \leqslant x < 50$	4
TOTAL	58

(i) **(a)** Using a scale of 2 cm to represent 10 kg, starting at −10 kg, on the horizontal axis, and a scale of 2 cm to represent 2 people per 5 kg on the vertical axis, draw on graph paper a histogram to illustrate these data. [6]

(b) If the first two classes in the table were to be merged into a single class, $-10 \leqslant x < 5$, calculate the height in cm of the rectangle which would represent this merged class in the histogram. [2]

(ii) Estimate, using the figures in the table, and giving your answers to 3 significant figures,

(a) the mean, [3]

(b) the standard deviation, [5]

of the weight loss of the people who attended the course.

8 (a) Abdul, Brian and Claude, share a flat. Whenever a visitor calls at the flat it is to see only one of them. However, it is not certain that any of them will be at home at the time a visitor calls.

The following table gives the probabilities that a visitor calls to see a particular person, and the probabilities (assumed to be independent) that each of them will be at home when a visitor calls.

Person	*Probability that a visitor has called to see this person*	*Probability that person is at home when a visitor calls*
Abdul	0.3	0.8
Brian	0.3	0.7
Claude	0.4	0.6

(i) Explain why the probabilities in the second column of the above table must sum to 1, but why those in the third column do not have to do so. [2]

Calculate the probability that

(ii) a visitor calls to see Brian but finds that Brian is not at home, [2]

(iii) the person who the visitor has called to see is not at home, [2]

(iv) a visitor calls to see Claude but Claude is the only one of the three not at home. [2]

(b) A doctor is consulted by 140 of his patients who believe they have a particular disease. In these 140 people there are 55 females. 25 of these females have the disease and 30 do not. 20 males have the disease.

(i) **Copy the following table onto your answer paper** and then complete it using the above information. [3]

	Male	*Female*	TOTAL
Have disease			
Do not have disease			
TOTAL			

There is a test which indicates whether a person is likely to have the disease. Of people who have the disease, 95% react positively to the test, but so do 10% of those who do not have it. The doctor gives the test to all 140 patients.

(ii) Calculate the probability that a patient chosen randomly reacts positively to the test. [3]

A patient chosen randomly has reacted positively to the test.

(iii) Calculate the probability that this patient has the disease. [2]

9 The data in the following table refers to two areas, Highshire and Lowshire, in the same country.

Age group	Highshire			Lowshire Population	Standard Population
	Population	Number of deaths	Death rate (per thousand)		
0 – 29	13 000	91	X	11 000	44%
30 – 49	Y	60	12	6000	24%
50 and over	7000	Z	55	8000	32%

(i) Find the values represented in the table by *X, Y* and *Z*. [3]

(ii) Calculate, per thousand,

   (a) the crude death rate for Highshire, [3]

   (b) the standardised death rate for Highshire. [4]

For Lowshire, the crude death rate and the standardised death rate are both exactly 22.25 per thousand.

(iii) Give the reason why, for Lowshire, the crude death rate and the standardised death rate are identical. [1]

(iv) State, giving a reason, whether the crude death rates or standardised death rates should be used to determine which area offers the better chance of a longer life. [2]

(v) Use your answer to **(iv)** to state, giving a reason, whether Highshire or Lowshire offers the better chance of a longer life. [1]

Soon after the data in the above table was produced, a number of the '50 and over' inhabitants of Highshire moved to live in Lowshire.

(vi) Assuming that those who moved were a representative cross-section of the '50 and over' age group in Highshire, state, **without carrying out any calculations**, what effect this movement of the population would have on

   (a) the crude death rate for Highshire, [1]

   (b) the standardised death rate for Highshire. [1]

10 Two doctors, Dr Smith and Dr Jones, hold daily surgeries at a medical centre. Over the course of one week a record was kept of the time each patient who arrived at the centre had to wait before being able to see a doctor. These times are illustrated in the following cumulative frequency graph, which shows the figures for each doctor separately. During the week, each doctor saw 70 patients.
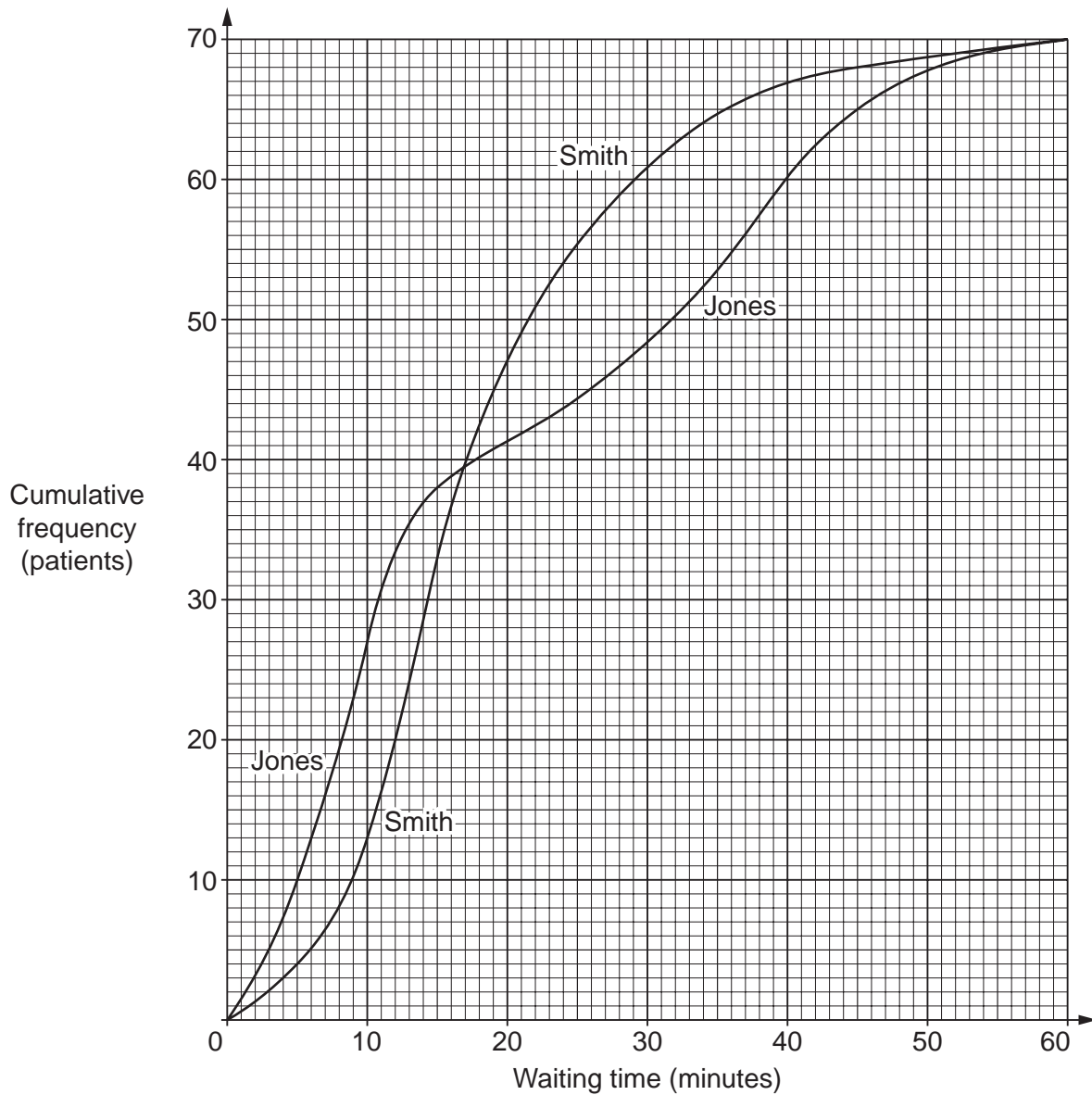
(i) State the median waiting time for each doctor. [2]

(ii) State the upper and lower quartile waiting times for each doctor, and hence find for which doctor the interquartile waiting time is longer, and by how much. [6]

(iii) For which waiting time is there the greatest difference between the number of patients who have already been seen by each of the two doctors? [1]

(iv) Find the number of patients who waited for 40 minutes or more. [3]

A patient arriving at the medical centre is given a choice of which doctor he or she wishes to see.

(v) Assuming that the week illustrated in the graph is a typical week at the centre, state, giving a reason in each case, which doctor should be chosen by a patient who wishes

   (a) to maximise the chance of having to wait for less than 12 minutes, [2]

   (b) to minimise the chance of having to wait for 33 minutes or more. [2]

[Question 11 is printed on the next page.]

11 A flat surface, *PQ*, is hinged at the point *Q*, so that it can be raised at any angle, *x*°, to a horizontal floor. A sphere is released at *P* and allowed to roll down the sloping surface and along the floor. The distance, *y* cm, which it travels along the floor before coming to rest, is measured.
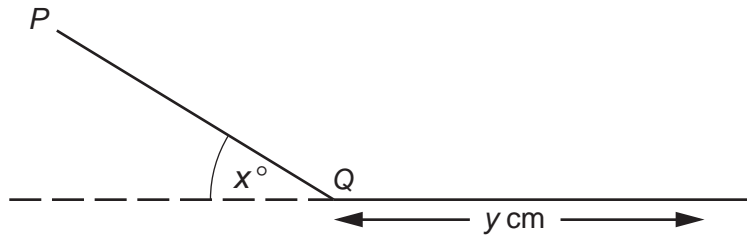
The following table gives the values of *y* which were measured for different values of *x*, together with a reference letter for each pair of values.

Reference letter	A	B	C	D	E	F	G	H
Angle (*x*°)	6.0	26.5	29.5	17.0	20.0	11.0	28.0	14.0
Distance (*y* cm)	39	129	147	82	99	96	137	71

(i) State why you would not use the first four points in the table to calculate one semi-average, and the last four points to calculate the other semi-average. [1]

(ii) State the reference letters of the four points you would use to calculate one of the semi-averages. [1]

(iii) Calculate the overall mean and the two semi-averages for these data. [5]

(iv) Using a scale of 2 cm to represent 5° on the horizontal axis, and 2 cm to represent a distance of 10 cm, starting at 30 cm, on the vertical axis, plot the eight points given in the table, and the three points you have calculated in **(iii)**, on a scatter diagram. [4]

(v) Draw the line of best fit on your scatter diagram. You are **not** required to obtain the equation of this line. [1]

It was later discovered that one of the pairs of values had been obtained using a different sphere.

(vi) Suggest which pair of values had been obtained using the different sphere. [1]

(vii) Ignoring the point you have suggested in **(vi)**, and **without carrying out any further calculations**, draw, by eye, a line of best fit through the remaining points. [1]

(viii) The line drawn in **(v)** is used to estimate a value of *y* when *x* = 9. State, with a reason, whether this estimate would be higher or lower than the true value. [2]

BLANK PAGE

12

BLANK PAGE