UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
General Certificate of Education Ordinary Level

STATISTICS **4040/01**

Paper 1 **October/November 2007**

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 **11** 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 The table below gives the population (in millions, correct to 1 decimal place) of each of the four countries of the United Kingdom at the Census in the year 2001.

Country	Population (millions)
England	49.1
Scotland	5.1
Wales	2.9
Northern Ireland	1.7
TOTAL	58.8

The data are to be illustrated by a pie chart.

(i) Calculate, each to the nearest degree, the sector angles of the pie chart. [2]

(ii) Using a circle of radius 5 cm, draw the pie chart. [2]

In the Census in the year 1951 the population of the United Kingdom was 50.3 million (correct to 1 decimal place).

(iii) Calculate, to 2 significant figures, the radius of the comparable pie chart which could be used to represent the population in 1951. [2]

2 A bus company was carrying out an investigation into the demand for its services. An observer standing at one bus stop recorded the number of people waiting in the queue when each bus arrived. The numbers recorded for nine consecutive buses on one day were as follows.

15   15   17   3*   20   18   17   15   13

The largest number, shown here as 3*, had not been recorded clearly, although it was definitely a two-digit number with first digit 3.

For these data,

(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 be found. [2]

3   The Science Department of a college offers courses in three subjects, Biology, C̶ and Physics. There are 170 students who take courses in at least one of these subjects. Th̶ diagram gives, for these 170 students, information on the number of students taking the ̶ subjects.
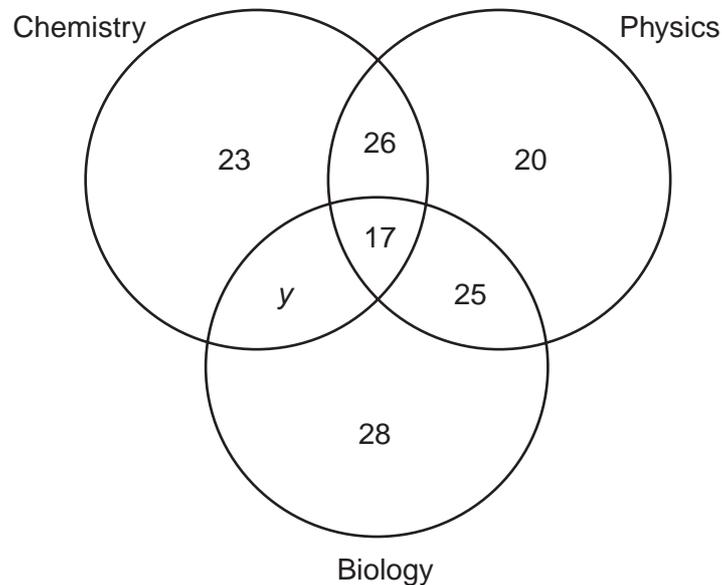
(i)   Find the value of y.   [2]

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

(iii) Calculate how many of the students do not study Biology.   [2]

(iv) Calculate the total number of students who study Chemistry.   [1]

4   **(a)** Give one advantage which quota sampling has over simple random sampling, and one advantage which simple random sampling has over quota sampling.   [2]

(b) The names of the 100 pupils in a small school are arranged in alphabetical order and then two-digit numbers in the range 00 to 99 are allocated, one to each pupil, in that order.

(i)   It is required to select a **systematic sample of size 5** from the pupils. The number of the first pupil to be selected is obtained from a random number table and is found to be 17. Write down the numbers of the other pupils selected for the sample.   [2]

(ii)   Briefly explain the situation which can lead to the **method** of systematic sampling being biased, and state, with a reason, whether it is likely to occur in this case.   [2]

5 The table below summarises how many O level subjects at grade C were obtained by the 120 pupils who sat the examinations at one school in a particular year.

Number of subjects	0	1	2	3	4	5	6	7	8	9
Number of pupils	2	2	11	17	24	25	22	12	4	1

For example, 17 pupils each obtained 3 subjects at grade C.

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

(ii) Using a scale of 2 cm to represent 1 subject 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]

6 **Give your answers in this question either as exact fractions or as decimals correct to 3 significant figures.**

Holly has three cards numbered 1, 3 and 5.

Akinyi has three cards numbered 2, 3 and 4.

(i) Each girl selects one of her three cards at random, and then the numbers on the two selected cards are added together to give a total.

Find the probability that the total is

(a) 7, [1]

(b) at most 6. [2]

(ii) Each girl selects one of her three cards at random, and then the numbers on the two selected cards are multiplied together to give a product.

Find the probability that the product is

(a) 20, [1]

(b) more than 6. [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 man found an old weighing machine and, in order to test its accuracy, he placed a number of loads of known mass on it and recorded the readings which it showed. The results are given in the table below.

Reading of weighing machine (x kg)	1.1	2.0	3.6	4.9	6.8	8.0
Actual mass of load (y kg)	0.5	1.4	3.5	5.0	7.6	9.0

  (i) Using 2 cm to represent 1 kg on both axes, and starting the y-axis at the value −1, plot on graph paper a scatter diagram of the above data. [2]

  (ii) By calculating and plotting $(\bar{x}, \bar{y})$ and the two semi-averages, draw the line of best fit. [6]

  (iii) **Using your graph**, obtain the equation of the line of best fit in the form $y = mx + c$. [4]

  (iv) **Use your equation** to estimate, to 1 decimal place, the actual mass of an item for which the reading on the machine was 6 kg. [1]

  (v) By drawing an appropriate line on your graph, estimate the value for which the reading on the machine would equal the actual mass. [2]

  (vi) Use the two lines you have drawn to comment on the accuracy of the machine. [1]

8 A large number of seeds of the same variety of flower were sown on the same day. Some time later, the heights, h mm, of the 80 surviving flowers were measured accurately. The histogram below illustrates the data obtained.
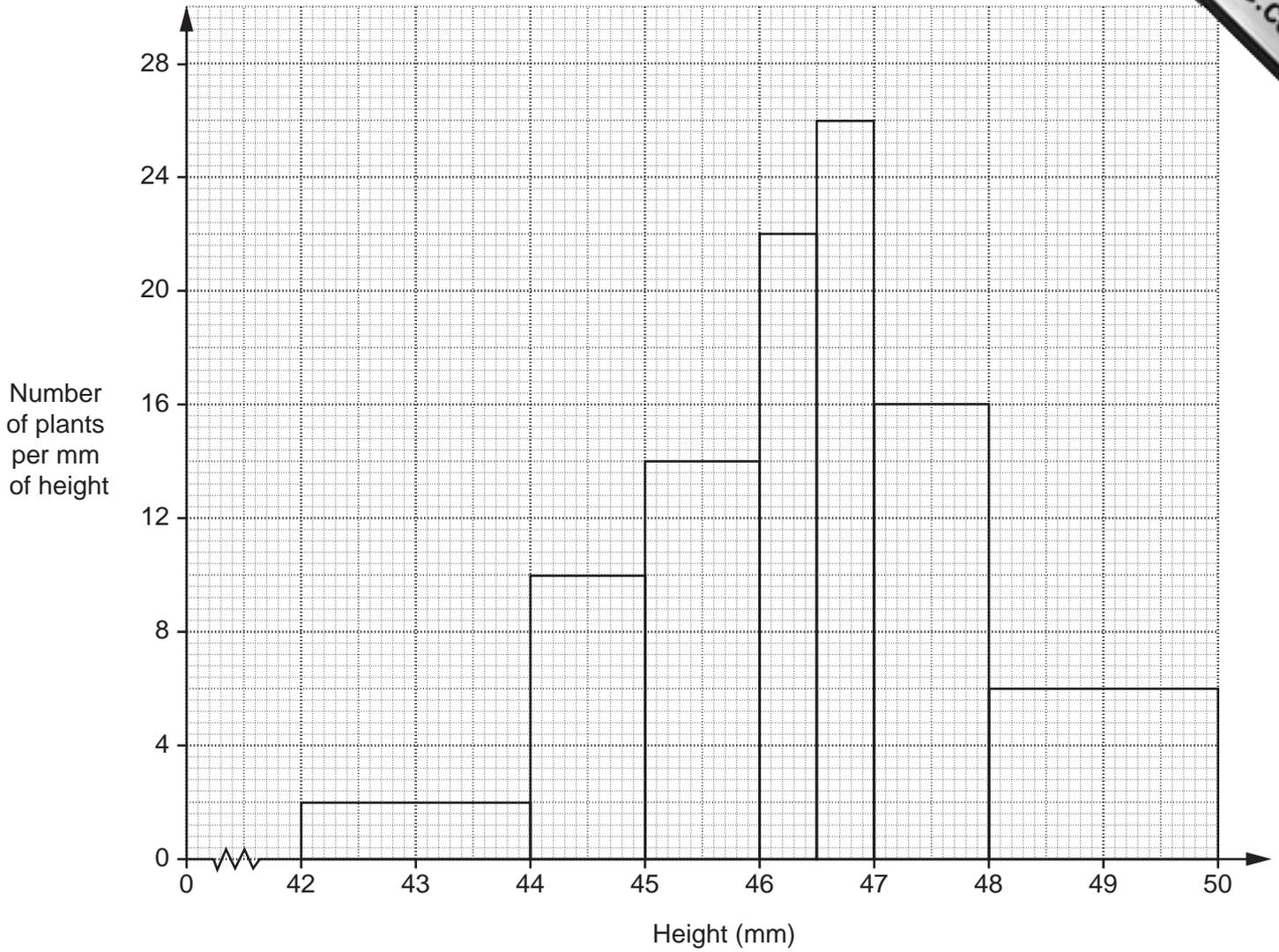
(i) Use the histogram to **copy and complete** the following grouped frequency table.

| Height (*h* mm)	Frequency
$42 \leqslant h < 44$	
$44 \leqslant h < 45$	
$45 \leqslant h < 46$	
$46 \leqslant h < 46.5$	
$46.5 \leqslant h < 47$	
$47 \leqslant h < 48$	
$48 \leqslant h < 50$	
TOTAL	80

[6]

(ii) Hence construct a simpler grouped frequency table, using the class intervals

$$42 \leqslant h < 44, \quad 44 \leqslant h < 46, \quad 46 \leqslant h < 48, \quad 48 \leqslant h < 50.$$ [5]

(iii) Draw, on graph paper, a histogram to illustrate the data in your table in **(ii)**, using the same scales and labelling on both axes as in the given histogram. [3]

(iv) **Use the histogram you have drawn in (iii)** to estimate the modal length to the nearest 0.1 mm. [2]

9 **(a)** A circular dart board, of radius 15 cm, is mounted centrally on a square piece o~~f~~~~cork~~ ~~plate~~
50 cm. A dart is thrown at random and sticks in either the dart board or the cork s~~urrounding~~
the dart board.
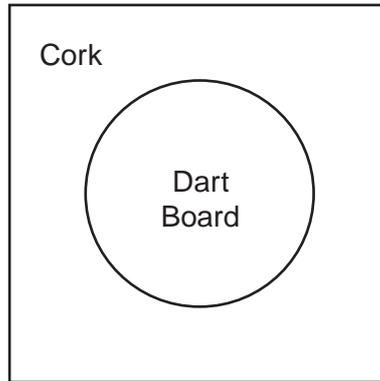
Calculate, to 3 significant figures, the probability that the dart sticks in the cork.
(Take the value of π to be 3.14.) [6]

(b) A hospital monitored the number of patients admitted during the course of one year and
suffering from a variety of illnesses. For 3 of these illnesses the records are summarised in
the following table.

Illness	Male	Female
Thrombosis	10	20
Pneumonia	21	27
Appendicitis	18	14

One of these patients is selected at random. Find the probability that the patient

(i) is female and suffers from thrombosis, [1]

(ii) suffers from pneumonia, [2]

(iii) suffers from appendicitis, given that he is male. [2]

Two of the patients are chosen at random.

(iv) Find the probability that both of them are male and suffering from thrombosis. [3]

(c) A biased coin is such that when it is tossed, the probability of a head being obtained is $\frac{2}{3}$.
Calculate the probability that exactly one head is obtained when the coin is tossed twice.
[2]

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

The table below gives information about the population and deaths in the town of Brown... the year 2005, together with the standard population of the area in which Brownville is situate...

| Age group	Deaths	Population in age group	Standard population (%)
Under 15	2	750	30
15 – 40	10	2000	30
41 – 65	53	5000	25
Over 65	85	1500	15

(i) For Brownville in the year 2005,

   (a) calculate the crude death rate, [4]

   (b) calculate the death rate for each age group, [2]

   (c) use your results in **(i)(b)** to calculate the standardised death rate. [4]

The table below gives information about Greentown, another town in the same area, also for the year 2005.

Age group	Death rate per thousand	Population in age group (%)
Under 15	4	18.75
15 – 40	5.5	43.75
41 – 65	11.2	31.25
Over 65	80	6.25

(ii) Calculate the **crude** death rate for Greentown in the year 2005. [3]

(iii) Use your results in **(i)(b)** and the information in the table for Greentown to compare the age group death rates for the two towns. [1]

(iv) State, giving a reason, an age group for which one of the towns appears to have a far less healthy environment. [2]

11   The cumulative frequency graph below illustrates the journey times to work, t minutes,   0
     workers in a factory.
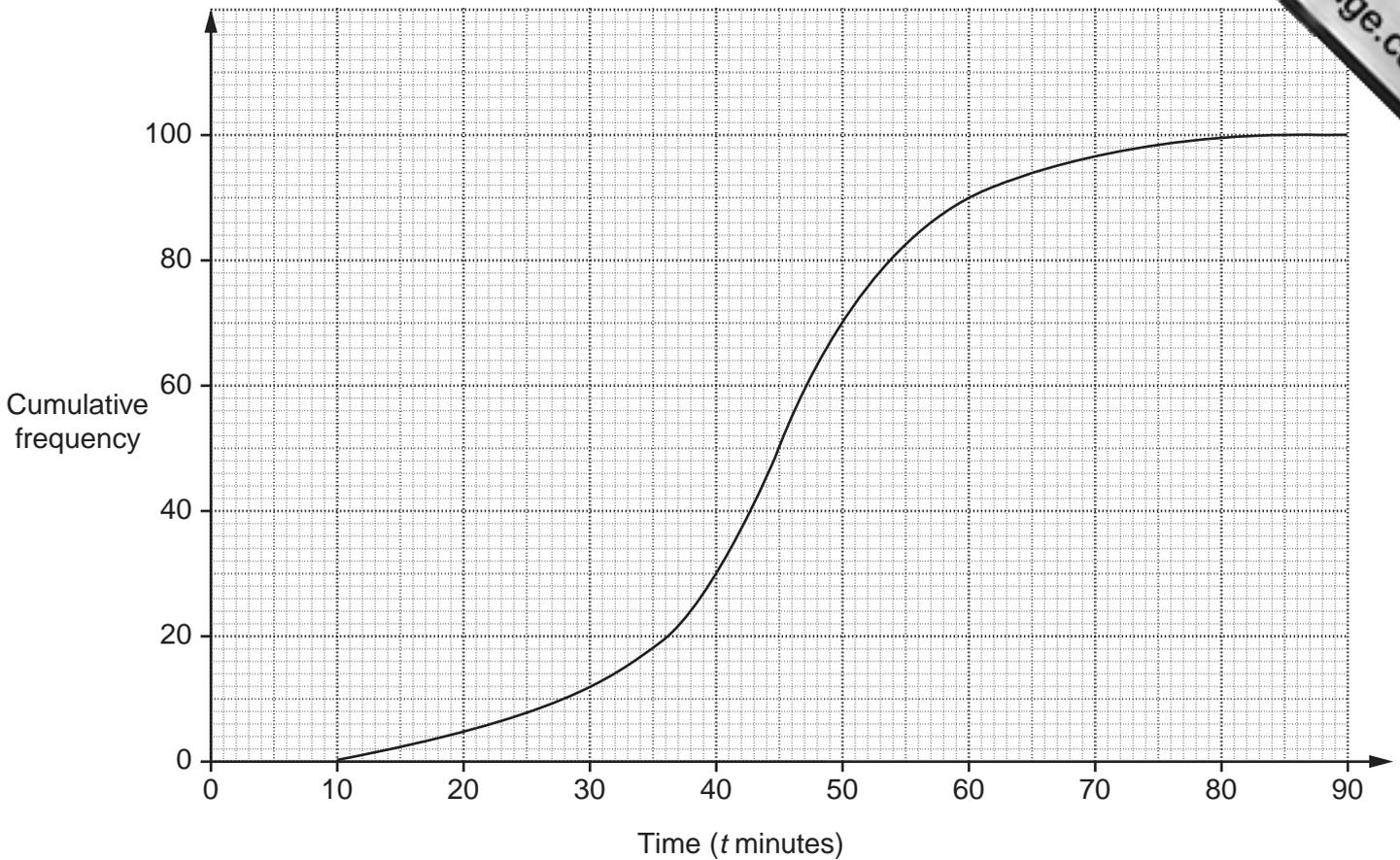
Time (t minutes)

   (i)   Use the graph to estimate

   (a)   the median,   [1]

   (b)   the interquartile range   [4]

   of the journey times.

(ii) Use the graph to **copy and complete** the following table.

Journey time (t minutes)	Cumulative frequency	Frequency
$10 \leqslant t < 30$		
$30 \leqslant t < 40$		
$40 \leqslant t < 50$		
$50 \leqslant t < 60$		
$60 \leqslant t < 90$		
TOTAL		100

[3]

(iii) Use the frequencies you have obtained to estimate

   (a) the mean, [4]

   (b) the standard deviation [2]

   of the journey times, giving your answers to 1 decimal place.

(iv) Compare the values of the median and the mean, giving an explanation for your answer. [2]

BLANK PAGE