UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
General Certificate of Education Ordinary Level

STATISTICS **4040/02**

Paper 2

October/November 2005

2 hours 15 minutes

Additional Materials: Answer Booklet/Paper
Graph paper (2 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 on both sides of the paper.
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.
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.
The use of an electronic calculator is expected in this paper.

This document consists of **9** printed pages and **3** blank pages.

UNIVERSITY of CAMBRIDGE
International Examinations

[Turn over

Section A [36 marks]

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

1 The set of five numbers

$$1 \quad 5 \quad 9 \quad 13 \quad n$$

has a mean m and a standard deviation s.

State, in terms of m and s, the mean and standard deviation of each of the following sets of five numbers

(i) 2 10 18 26 $2n$, [2]

(ii) –3 1 5 9 $(n - 4)$, [2]

(iii) 8 20 32 44 $(3n + 5)$. [2]

2 The following statements refer to the books in a school library during a particular year.

A The subject (e.g. History, Geography, Fiction) of different books.

B The number of books on different subjects.

C The colour of the covers of different books.

D The number of Statistics books on the shelves when the library opened on 15th January.

E The number of times in the year in which different books were taken out on loan.

F The number of books which were never taken out on loan during that particular year.

(i) State which two of A, B, C, D, E and F **cannot** be considered as defining a variable. [2]

(ii) State which of the remaining four statements are

(a) qualitative variables,

(b) quantitative variables.

[4]

3 **(i)** Give a brief explanation of the purpose of calculating moving averages. [2]

(ii) If moving averages were calculated for each of the following, state, **with a reason in each case**, whether or not the moving averages would need to be centred.

(a) The total attendance each term of pupils in a school where each year is divided into three terms.

(b) The total attendance each semester of students at a college where each year is divided into two semesters.

(c) The total takings each day in a shop which is open on every day of the week.

(d) The total takings each day in a shop which is open on every day of the week except Sunday.

[4]

4 A and B are two events, where

$$P(A) = p,$$

$$P(B) = 0.5,$$

$$P(A \text{ or } B \text{ or both}) = 0.8.$$

Find the value of p if

(i) A and B are mutually exclusive events, [2]

(ii) A and B are independent events. [4]

5 The table below gives the number of eggs laid in a particular week by some of the hens on a small farm.

Number of eggs	5	6	7	8	9	10	11
Number of hens	8	12	9	7	8	6	3

For example, 3 hens each laid 11 eggs.

(i) State the modal number of eggs laid by these hens. [1]

(ii) Find the median number of eggs laid by these hens. [2]

There are another N hens on the farm, all of which laid 14 eggs during that week.

(iii) When these hens are included in the table, the median is unchanged. Find the greatest possible value of N. [3]

6 A farmer is investigating the annual cost of running his farm, and finds that his four largest items of expenditure are animal food, labour, veterinary services and fuel. On the basis of his spending in the year 2004 he allocates the following weights to these categories.

Animal food	13
Labour	5
Veterinary services	2
Fuel	3

Taking 2003 as base year, the price relatives for each of these categories for the years 2004 and 2005 are

	2004	2005
Animal food	126	131
Labour	109	120
Veterinary services	109	117
Fuel	101	103

(i) Calculate a weighted aggregate price index for 2004 taking 2003 as base year, giving your answer correct to 1 decimal place. [3]

(ii) Taking 2004 as base year, calculate

(a) the price relative for animal food in 2005,

(b) the price relative for labour in 2003.

[3]

Section B [64 marks]

Answer not more than **four** of the questions 7 to 11.

Each question in this section carries 16 marks.

7 Coloured sweets are sold in packets each containing 12 sweets. Some of these packets contain 5 red and 7 yellow sweets, and all the other packets contain 4 yellow and 8 green sweets.

Jane, Kumar, Lucy, Mohamed and Ninette are each given one packet containing 5 red and 7 yellow sweets and one packet containing 4 yellow and 8 green sweets. In the following questions assume that all choices are made at random.

(i) Jane chooses one of her two packets and then chooses one sweet from it. Find the probability that the chosen sweet is yellow. [3]

(ii) Kumar chooses two sweets from his packet containing 4 yellow and 8 green sweets. Find the probability that the two sweets are of different colours. [4]

(iii) Lucy chooses one sweet from each of her two packets. Find the probability that they are of the same colour. [3]

(iv) Mohamed chooses one of his two packets, and then chooses two sweets from it. Find the probability that both sweets are red. [3]

(v) Ninette puts all the sweets from her two packets together in a bag, and then chooses two sweets from the bag. Find the probability that both sweets are green. [3]

[Turn over

8 A large factory is in operation continuously, the 24 hours in each day being divided into three 8-hour shifts, Early, Late and Night. The following table gives the number of workers absent through illness for each of the shifts during the first five days of one week. It also shows the calculation of a three-point moving average, correct to one decimal place where necessary.

Day	*Shift*	*Number of absences*	*Three-point total*	*Three-point moving average*
1	Early	14		
	Late	12	55	18.3
	Night	29	56	18.7
2	Early	15	55	18.3
	Late	11	*w*	19
	Night	31	*x*	19.7
3	Early	17	64	21.3
	Late	16	65	21.7
	Night	32	66	22
4	Early	18	68	22.7
	Late	18	73	*y*
	Night	37	76	25.3
5	Early	21	79	26.3
	Late	21	80	26.7
	Night	38	*z*	

(i) Calculate the values of *w*, *x* and *y* in the table. [3]

(ii) Explain why it is not possible to calculate the value of *z* in the table. [1]

(iii) On graph paper, using a scale of 1 cm to 1 shift on the horizontal axis, and a scale of 2 cm to 5 absences on the vertical axis, plot the original data, and join consecutive points by straight lines. Ensure that the horizontal axis extends to cover the Early shift of Day 6. [5]

(iv) Plot the moving average values on your graph, and draw a single straight line to represent the trend. [2]

(v) Comment on the extent to which the moving average values have removed variation from the data. [1]

The shift components for these data are summarised in the following table.

Shift	Early	Late	Night
Shift component	–4.4	q	10.8

(vi) Calculate the value of q. [1]

(vii) Use your trend line, and the appropriate shift component, to estimate the number of workers who are absent during the Early shift of Day 6. [3]

9 The **age last birthday** of each of the 100 employees of a small company was noted on 1st January this year, and these ages are summarised in the following table.

Age last birthday	*Number of employees*
15 - 19	5
20 - 24	8
25 - 34	13
35 - 44	20
45 - 54	40
55 - 64	14
TOTAL	100

(i) Calculate an estimate of the median age in years of the employees of the company. [3]

(ii) Calculate an estimate of the interquartile range of the ages of the employees, giving your answer in years, correct to 1 decimal place. [5]

(iii) Using an assumed mean of 30 years,

(a) calculate an estimate of the mean age of the employees, giving your answer in years, correct to 1 decimal place,

(b) show that an estimate of the standard deviation of the ages of the employees, correct to 1 decimal place, is 12.2 years.

[8]

[Turn over

10 The table below lists a population of fifteen children, numbered 01 - 15, and classified by gender and age group.

Child	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15
Gender	F	F	M	F	F	F	F	M	F	F	M	F	M	M	M
Age group	II	I	III	III	I	II	III	I	I	II	I	III	III	I	III

Gender: F = Female, M = Male
Age group: I = 0 - 4 years, II = 5 - 9 years, III = 10 - 14 years.

Different methods are to be considered for selecting a sample **of size 5** from the population, using the two-digit random number table below. Numbers outside the allocated range are ignored, and no child may be selected more than once in any one sample.

TWO-DIGIT RANDOM NUMBER TABLE

82	60	12	02	60	69	99	09	67	01	12	04	36
40	99	02	66	37	59	24	79	35	04	09	15	06
21	02	08	10	91	65	05	78	09	99	15	14	00
14	45	74	15	01	53	09	07	12	18	00	13	61

(i) Starting at the beginning of the first row of the table, and moving along the row, select a **simple random sample** of the required size. [2]

(ii) A **systematic sample** is to be selected.

(a) Write down the smallest possible and largest possible two-digit numbers of the first child selected. [1]

The systematic sample is selected by starting at the beginning of the second row of the table, and moving along the row.

(b) Write down the number of the first child selected. [1]

(c) Write down the numbers of the other four children selected for the systematic sample. [1]

(iii) A **sample stratified by gender** is to be selected.

(a) State how many children of each gender would be selected for such a sample. [1]

(b) Starting at the beginning of the third row of the table, and moving along the row, select a sample stratified by gender. Use every number if the gender to which it relates has not yet been fully sampled. [3]

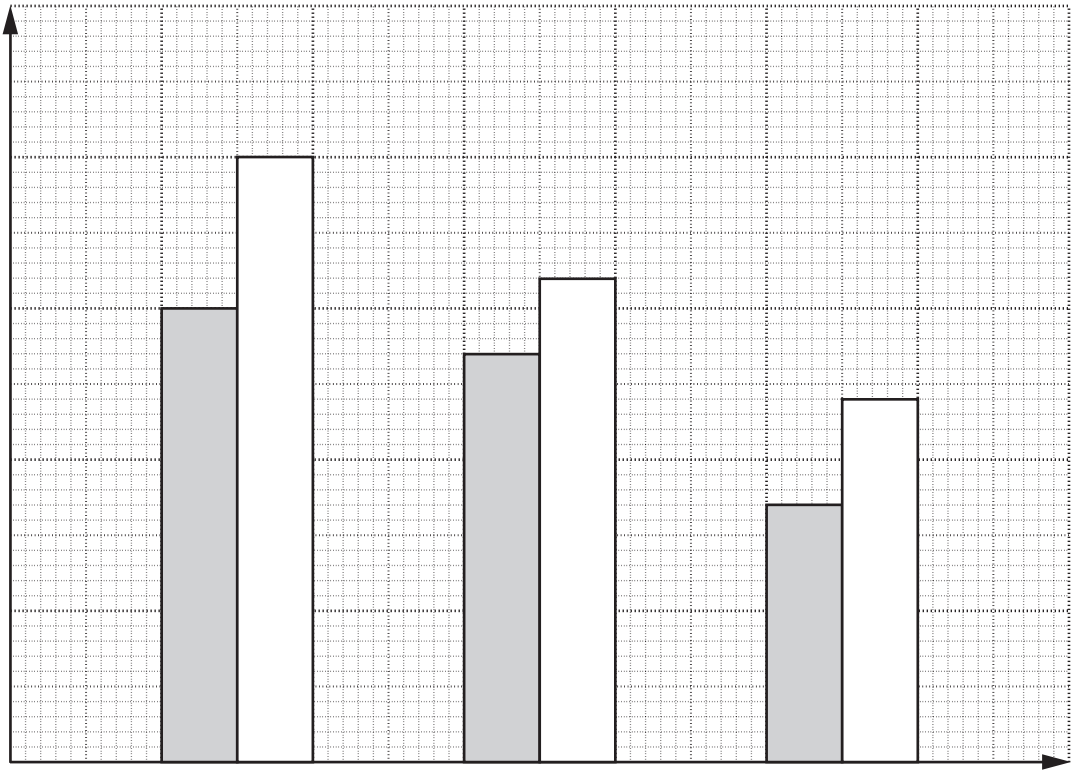
(iv) A **sample stratified by age group** is to be selected.

(a) State how many children of each age group would be selected for such a sample. [1]

(b) Starting at the beginning of the fourth row of the table, and moving along the row, select a sample stratified by age group. Use every number if the age group to which it relates has not yet been fully sampled. [2]

(v) By examining the gender and age group of the children in each of the samples you have selected, state, with a reason for each sample, whether it represents the population exactly in terms of both gender and age group, and is therefore unbiased, or is in some way biased. [4]

11 (a)



(i) Name the type of bar chart illustrated in the above diagram. [1]

(ii) This diagram is incomplete. State the five items of information which need to be included on it before it may be considered to be complete. [5]

(iii) State two advantages and two disadvantages of this type of bar chart compared with a percentage sectional bar chart illustrating the same data. [4]

(b) The following table gives the home region of male and female students attending one particular course at a university.

Home region	*Number of male students*	*Number of female students*
South	74	57
Midlands	70	48
North	56	45
TOTAL	200	150

Illustrate these data on graph paper by means of a percentage sectional bar chart. [6]

BLANK PAGE

BLANK PAGE

BLANK PAGE