CAMBRIDGE INTERNATIONAL EXAMINATIONS
General Certificate of Education Ordinary Level

STATISTICS **4040/02**

Paper 2

October/November 2003

2 hours 15 minutes

Additional Materials: Answer Booklet/Paper
Graph paper (1 sheet)
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** the 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.

MCS-UCB245-S39768-2

University *of* Cambridge
Local Examinations Syndicate

Section A [36 marks]

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

1 The following variables relate to all the trains arriving during one day at a certain railway station. For each variable, state whether it is

discrete or continuous

and whether it is

qualitative or quantitative.

(i) The number of passengers on each train. [2]

(ii) The amount of fuel used by each train during its journey. [2]

(iii) The type of ticket held by each passenger; for example, single, return, adult, child. [2]

2 A charity receives annual donations from each of its members. In 2002 the amounts donated by members had a mean of $60 and a standard deviation of $20.

Because of rising costs the charity has had to ask each member to donate more in 2003. Given that there is no change in membership, calculate the mean and the standard deviation of the amounts donated in 2003

(i) if each member donates $10 more than in 2002, [2]

(ii) if each member donates 10% more than in 2002. [2]

3 Pupils from two different schools took part in a general knowledge quiz. The marks obtained by the pupils from the two schools are summarised in the following table.

School	*Number of pupils*	*Sum of the marks*	*Sum of the squares of the marks*
A	17	225	4495
B	8	165	2992

(i) Calculate the total number of marks obtained by **all** the pupils. [1]

(ii) Calculate the mean mark obtained by **all** the pupils. [2]

(iii) Calculate the sum of the squares of the marks obtained by **all** the pupils. [1]

(iv) Hence calculate the standard deviation of the marks obtained by **all** the pupils. [4]

4 **(a)** The number of spectators attending each of the football matches played in a league on one particular day was recorded, and the data were then summarised in the form of a grouped frequency distribution.
For the class recorded as 1500 – 1999, state

(i) the class mid-point, [1]

(ii) the class interval. [1]

(b) The heights of all the pupils in a school were measured, recorded to the nearest half-centimetre, and then summarised in the form of a grouped frequency distribution.
For the class recorded as 160 – 169.5, state

(i) the class mid-point, [1]

(ii) the class interval. [1]

5 Two events A and B are such that $P(A) = 0.6$, $P(B) = 0.2$, and $P(A \cap B) = 0.1$.

(i) State, in each case giving a reason for your answer,

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

(b) whether A and B are independent events. [2]

(ii) By drawing a Venn diagram, or otherwise, find the probability that neither A nor B occurs. [3]

6 Four students took both written and oral examinations in the language they were all studying. The following table gives their raw marks in the two examinations.

	Student			
	A	B	C	D
Written	40	53	51	56
Oral	35	23	34	28

The mean of the raw written marks is 50, and the mean of the raw oral marks is 30.

The oral marks are to be scaled so that the scaled oral mean is the same as the raw written mean, and the scaled oral standard deviation is twice the raw oral standard deviation.

Each student's position (first, second, third or fourth) in the examination is determined by the sum of their written and oral marks.

(i) Calculate the scaled oral marks for the four students. [4]

(ii) Calculate each student's total mark if **raw** oral marks are used. [1]

(iii) Calculate each student's total mark if **scaled** oral marks are used. [1]

(iv) Which student's position is lower if the **scaled** oral marks are used? [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 The table below gives the number of people living in each of the houses in a certain street, in the form of a frequency distribution.

Number of people	0	1	2	3	4	5	6	7 or more
Number of houses	2	5	5	6	4	2	1	0

(i) Find the total number of houses in the street. [1]

(ii) Calculate the total number of people living in the street. [3]

(iii) A house is chosen at random from the street.
Calculate the probability that

(a) it is not occupied, [1]

(b) it has at least four people living in it. [1]

(iv) A person living in the street is chosen at random.
Calculate the probability that this person lives in a house

(a) in which exactly three people live, [3]

(b) with **at most two other** people. [3]

(v) Two people who live in the street are chosen at random.
Calculate the probability that they both live alone. [4]

8 A biased die in the shape of a pyramid has four triangular faces numbered 1, 2, 3 and 4. When the die is thrown the **score** is the number on the face which comes to rest on the floor. The possible scores and their probabilities are shown in the table below.

Score	1	2	3	4
Probability	$\frac{4}{10}$	$\frac{3}{10}$	$\frac{2}{10}$	$\frac{1}{10}$

(i) Calculate the expected score. [2]

The die is thrown twice. The outcome can be represented as an ordered pair.
For example, (1, 2) represents a score of 1 followed by a score of 2. Given that the variable Y represents the sum of the two scores obtained,

(ii) write down the possible values of Y, [1]

(iii) give, as ordered pairs, the different outcomes producing each value of Y, [3]

(iv) tabulate the possible values of Y and their probabilities, [8]

(v) show that the expected value of Y is 4. [2]

9 The following table shows the number of accidents at work reported in a large industrial company for each quarter of the years 1997–2000. It also shows the calculation of values of a centred four-quarterly moving average.

Year	*Quarter*	*Number of accidents*	*Four-quarterly total*	*Four-quarterly moving average*	*Centred moving average*
1997	1st	86			
	2nd	69			
			304	76	
	3rd	76			75.25
			298	74.5	
	4th	73			74
			294	73.5	
1998	1st	80			72.875
			289	72.25	
	2nd	65			71.625
			w	71	
	3rd	71			70.375
			x	69.75	
	4th	68			69.375
			276	69	
1999	1st	75			68.625
			273	68.25	
	2nd	62			67.875
			270	y	
	3rd	68			z
			268	67	
	4th	65			66.5
			264	66	
2000	1st	73			65.625
			261	65.25	
	2nd	58			65
			259	64.75	
	3rd	65			
	4th	63			

(i) Calculate the values of *w*, *x*, *y* and *z* in the table. [4]

(ii) On graph paper, using a scale of 1cm per quarter on the horizontal axis, and a scale of 2cm to 5 accidents on the vertical axis, starting at 50 accidents, plot the original data and join consecutive points by straight lines. Ensure that the horizontal axis extends to cover the first quarter of 2001. [5]

(iii) Plot the centred moving average values on your graph, and draw a straight line of best fit through the points. [2]

(iv) Give a reason for plotting **centred** moving average values. [1]

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

	1st Quarter	2nd Quarter	3rd Quarter	4th Quarter
Seasonal component	7.0	q	0.7	–1.3

(v) Calculate the value of q. [2]

(vi) Use your trend line, and the appropriate seasonal component, to estimate the number of accidents during the first quarter of 2001. [2]

10 The treasurer of a tennis club is carrying out an analysis of changes in club expenditure. He has summarised expenditure for the year 2000 as follows:

Total cost of maintenance to the courts and clubhouse (including heating and miscellaneous items)	$3000
Average cost of one box of six tennis balls	$5.00
Number of balls purchased during the year	1200
Local taxes for services such as lighting, drainage and refuse collection	$500
Wage rate per hour paid to the club cleaner	$5.00
Number of hours worked by the cleaner during the year	500

(i) Use these data to show that the treasurer should assign weights to the four items, maintenance, balls, local taxes, cleaning in the ratio 6 : 2 : 1 : 5. [4]

In 2001, as compared with 2000,
maintenance costs increased by 3%,
by changing the supplier the cost of balls **decreased** by 10%,
local taxes increased by 2%,
the cleaner's hourly wage rate was increased by 5%.

(ii) Write down price relatives for 2001, taking 2000 as base year, for each of the four items, maintenance, balls, local taxes, cleaning. [3]

(iii) Calculate a weighted aggregate cost index for 2001, taking 2000 as base year. [4]

(iv) Use the index calculated in **(iii)** and the costs for 2000 to estimate the total cost of running the club in 2001. [3]

(v) Give two reasons why your estimate for 2001 may be very different from the true cost in 2001. [2]

11 In a population of size 50, there are 30 men, who are all allocated a different two-digit number in the range 01–30, and 20 women, who are all allocated a different number in the range 31–50. 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 ranges are ignored, and no person may be selected more than once in any one sample.

TWO-DIGIT RANDOM NUMBER TABLE

96	77	56	01	11	11	02	15	26	43	74	49	21	30	48
40	52	36	07	18	99	79	27	36	30	97	14	72	64	82
53	08	66	12	44	38	73	39	52	28	21	05	35	16	50

(i) Starting at the beginning of the first row of the table, and reading along the row, a **simple random sample** is to be selected.

(a) Give a reason why the first three two-digit numbers in the row will not be used. [1]

(b) Give a reason why the sixth two-digit number in the row will not be used. [1]

(c) Write down the two-digit numbers of the five people selected for the sample. [2]

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

(a) Write down the smallest possible two-digit number of the first person selected. [1]

(b) Write down the largest possible two-digit number of the first person selected. [1]

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

(c) Write down the number of the first person selected. [1]

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

(iii) It is believed that gender is an important factor in the survey being carried out, and a **sample stratified by gender** is therefore to be selected.

(a) State how many men and how many women would be selected for such a sample. [1]

(b) Starting at the beginning of the third row of the table, and reading along the row, select a sample stratified by gender. [3]

(iv) State, with a reason, which of the three samples selected is least representative of the population in terms of gender. [2]

(v) Briefly describe how an interviewer might select a **quota sample** of size 5 from this population. [2]