

UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
General Certificate of Education Ordinary Level

STATISTICS

4040/02

Paper 2

October/November 2004

2 hours 15 minutes

Additional Materials: Answer Booklet/Paper
Graph paper
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 **10** printed pages and **2** blank pages.



Section A [36 marks]

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

1 The set of six numbers

2, 5, 7, 12, 19, n , where n is a positive integer,

has a mean of m and a standard deviation of s .

Use these two results to obtain, in terms of m and s , the mean and standard deviation of the set of numbers

(i) $-1, 2, 4, 9, 16, n-3$, [2]

(ii) $4, 10, 14, 24, 38, 2n$. [2]

2 Four definitions A, B, C and D are given below.

A An approximate regular pattern, repeated in a series of data, but often over different periods of time.

B Data recorded at intervals of time.

C A pattern of fluctuations in a series of data which recurs at definite intervals of time.

D The general course or direction which data follows over a long period of time.

For each of the following terms, write down which of A, B, C and D best describes it.

(i) Time series.

(ii) Seasonal variation.

(iii) Cyclic variation.

(iv) Trend.

[4]

3 You are not required to carry out any calculations in answering this question.

The following table gives the value, for ten consecutive years, of an index of the output of a certain industry.

<i>Year</i>	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
<i>Value of index</i>	127	134	126	133	143	135	138	146	141	145

(i) Plot points representing these data on a graph. Use a scale of 2 cm to represent 1 year on the horizontal axis, and 2 cm to represent 5 units of the value of the index on the vertical axis. Start the vertical axis at a value of 120, and label your graph. Join consecutive plotted points by straight lines. [4]

(ii) Use your graph to state which type of moving average you would need to calculate to remove variation from these data. Give a reason for your answer. [2]

(iii) State, **giving a reason**, whether you would need to centre the values of your chosen type of moving average. [1]

4 The lengths of all the 168 plums in a consignment were measured to the nearest millimetre. The measurements were then summarised in the following grouped frequency table.

<i>Length to nearest mm</i>	<i>Number of plums</i>
30–32	7
33–35	18
36–38	27
39–41	30
42–44	41
45–47	21
48–50	16
51–56	5
57–62	3

(i) **Calculate** an estimate of the median length of the 168 plums, giving your answer in mm, correct to 1 decimal place. [3]

(ii) **Calculate** an estimate of the interquartile range of the lengths of the plums, giving your answer in mm, correct to 1 decimal place. [5]

- 5 A group of 39 pupils was given an examination, which was marked out of 50. The modal mark for the examination was 30. State whether each of the following statements is true, or is false, or there is not enough information to decide whether it is true or false. Give a clear reason for each of your answers.
- (i) More pupils scored 30 than any other single mark. [2]
 - (ii) 30 pupils scored the same mark as each other. [2]
 - (iii) 20 pupils scored 25 marks. [2]
- 6 (a) John has been asked to scale a value of 30, taken from a set of data with a mean of 35 and a standard deviation of 5, to a value from a set with mean 40 and standard deviation 6. He gives his result as 46. State, **without carrying out any calculations**, why John's answer **must** be incorrect. [2]
- (b) A set of data, S , has a mean of 60 and a standard deviation of 10.
- (i) The set S is scaled so that the mean is μ and the standard deviation is 8. When a value of 56 taken from S is scaled, the result is 72. Calculate the value of μ . [2]
 - (ii) The set S is now scaled so that the mean is 50 and the standard deviation is σ . When a value of 40 taken from S is scaled, the result is 42. Calculate the value of σ . [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 A catalogue of new books on transport history contains 90 books. 40 are classified as having a large page size, 30 are classified as having a medium page size, and 20 are classified as having a small page size. Each book in these categories is allocated a different two-digit number in the range 01–40, 41–70 and 71–90 respectively. Different methods are to be considered for selecting a sample **of size 9** from the catalogue, using the two-digit random number table below. Numbers outside the allocated ranges are ignored, and no book may be selected more than once in any one sample.

TWO-DIGIT RANDOM NUMBER TABLE

36	70	10	85	46	41	35	59	70	55	60	52	45	33	02
66	15	53	85	32	04	13	06	39	95	30	85	19	99	87
14	36	48	09	63	35	38	25	81	60	50	99	21	74	50

- (i) (a) 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]
- (b) Give a reason why one of the first nine two-digit numbers in that row has not been used. [1]
- (ii) A **systematic sample** is to be selected.
- (a) Write down the smallest possible and largest possible two-digit numbers of the first book 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 book selected. [1]
- (c) Write down the numbers of the other eight books selected for the systematic sample. [2]
- (iii) A **sample stratified by page size** is to be selected.
- (a) State how many books of each page size 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 page size. Use every number if the page size to which it relates has not yet been fully sampled. [3]
- (c) For every number which you have not used in the third row of the table, state the reason why it has not been used. [3]
- (iv) State which of the samples is least representative of the population in terms of page size, and explain in what way it is not representative. [2]

- 8 The head teacher of a school analysed the school's expenditure on the three stationery items, paper, books and 'other stationery'. The following table gives certain information relating to these items for the three years 1990 (taken as base year), 2000 and 2002.

	<i>Price relatives</i>			<i>1990 weights</i>
	1990	2000	2002	
Paper	100	120	125	4
Books	100	150	160	7
Other stationery	100	125	125	5

- (i) Calculate, to one decimal place, a weighted aggregate index for these costs for 2002, taking 1990 as base year and using the weights 4, 7 and 5 which were determined in 1990. [4]
- (ii) Suggest a way in which these weights may have been determined. [1]
- (iii) The total amount spent on the three items in 1990 was \$28 000. Use the index number you calculated in (i) to estimate the total amount spent in 2002, giving your result to the nearest \$1000. [2]
- (iv) State what you can deduce from the values of the price relative for the 'other stationery' in the years 2000 and 2002. [1]
- (v) Calculate new price relatives for each of the three items in 2002, taking 2000 as base year, and giving your results to the nearest whole number. [3]

The table below shows the actual amounts spent in the year 2000.

<i>Item</i>	<i>Amount (\$)</i>
Paper	15 000
Books	20 000
Other stationery	30 000

- (vi) Using these amounts as weights, and your results from (v), show that the weighted aggregate index for 2002, taking 2000 as base year, is 103.1. [3]
- (vii) Use your answer to (vi) to estimate the total amount spent on the three items in 2002, giving your result to the nearest \$1000. [1]
- (viii) Suggest a reason for the difference in your two estimates of expenditure in 2002, using your results from (iii) and (vii). [1]

- 9 A 'sibling' is a brother or a sister. So, a person with one brother and one sister has two siblings, a person with two brothers and three sisters has five siblings. A school register shows, for the 140 pupils in their final year, how many brothers and sisters each pupil has. No pupil in their final year has a brother or sister in the same year. The number of pupils in each category is given in the following table.

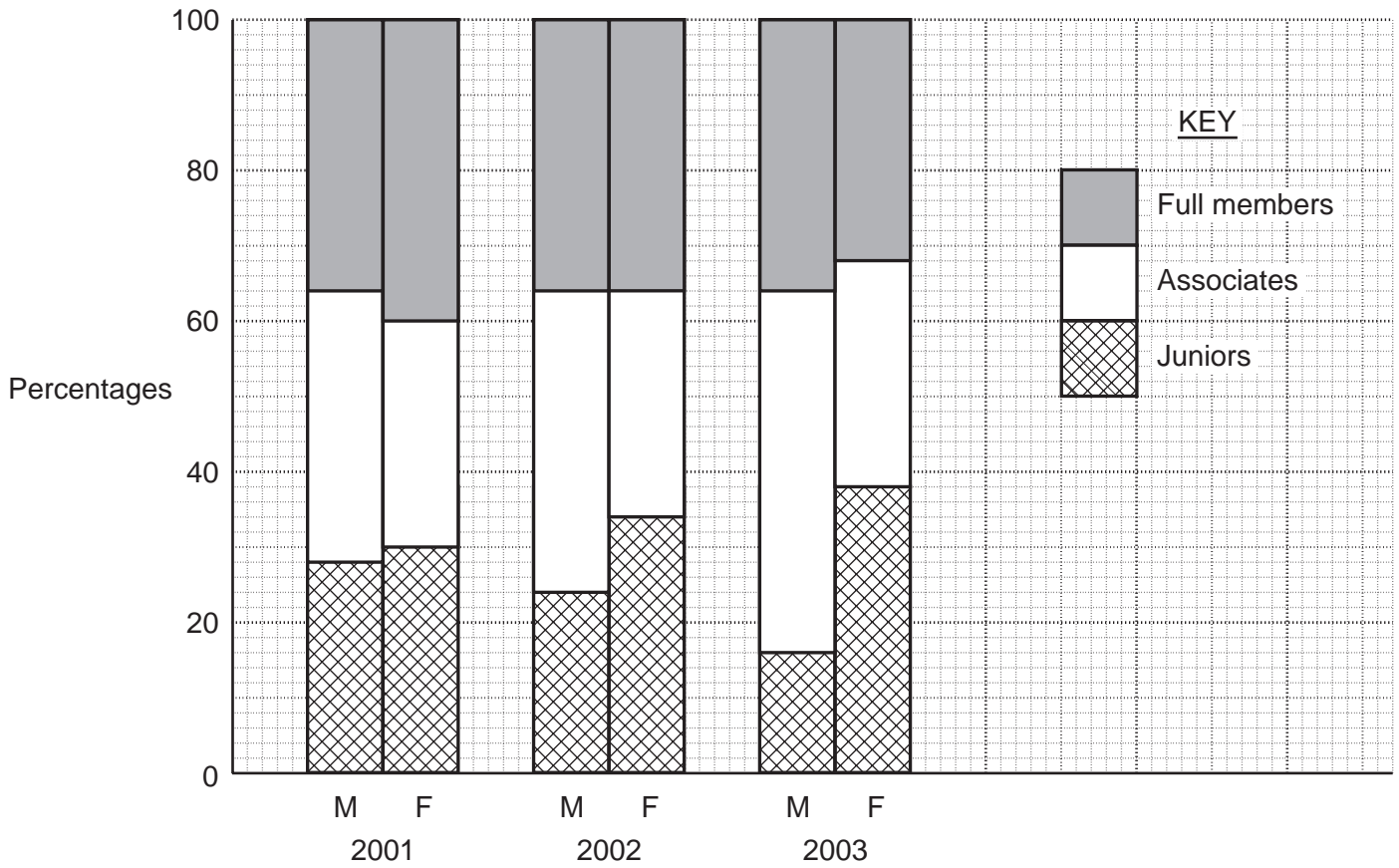
<i>Number of siblings</i>	<i>Boys</i>	<i>Girls</i>	TOTALS
0 brothers, 0 sisters	3	5	8
0 brothers, 1 sister	4	6	10
1 brother, 0 sisters	7	5	12
0 brothers, 2 sisters	10	12	22
1 brother, 1 sister	13	11	24
2 brothers, 0 sisters	10	11	21
1 brother, 2 sisters	7	8	15
2 brothers, 1 sister	6	10	16
2 brothers, 2 sisters	5	7	12
TOTALS	65	75	140

Thus, for example, 13 boys have 1 brother and 1 sister.

Giving your results as either exact fractions, or decimals correct to 3 significant figures, calculate the probability that

- (i) a pupil chosen at random is a boy, [1]
- (ii) a girl chosen at random has no siblings, [3]
- (iii) a boy chosen at random has two siblings, [3]
- (iv) a pupil chosen at random has the same number (including 0) of brothers as sisters, [2]
- (v) a pupil, chosen at random from those who have at least one sibling, has more brothers than sisters, [3]
- (vi) two pupils, chosen at random from those who have more sisters than brothers, are both girls. [4]

- 10 A sports club has three types of membership, Full members, Associates, and Juniors. The bar chart below illustrates the percentages (rounded to the nearest even integer) of each type of membership for male (M) and female (F) members in the years 2001, 2002 and 2003.



- (i) Use the information given in the bar chart to copy and complete the following table.

Percentage of members of each type

	2001	2001	2002	2002	2003	2003
	M	F	M	F	M	F
Full members						
Associates						
Juniors						

[6]

(ii) Use the given information (with the given level of accuracy) to select which three of the following statements **must** be true, giving a reason in each case. [6]

- A Among males in 2003, there were three times as many Associates as Juniors.
- B The number of male Full members was the same in all three years.
- C There were fewer male Juniors than female Juniors in all three years.
- D Associates formed the same percentage of membership among females in all three years.
- E There were fewer male Juniors in 2003 than in 2002.
- F There was the same number of male Full members and male Associates in 2001.
- G There was the same number of female Full members in 2001 as male Associates in 2002.

(iii) Further information is made available in the following table.

Total number of members

	2001	2002	2003
Males	72	60	75
Females	54	56	56

Using this further information, for each of the four statements which you did **not** select in (ii), state with a reason whether it is true or false. [4]

[Question 11 is printed on the next page.]

- 11 (a) A biased six-sided die bearing the numbers 1, 2, 3, 4, 5 and 6, is rolled once. The score is the number on the uppermost face of the die. Each even score is twice as likely as each odd score.

- (i) Show that the probability of scoring 4 is $\frac{2}{9}$. [2]
- (ii) Copy and complete the following table.

Score	1	2	3	4	5	6
Probability						

[1]

- (iii) A game requires a player to pay an entry fee. He then rolls the die, and is paid a number of dollars equal to the score if the score is odd, but nothing if the score is even.

Find how much the player should pay to enter the game if it is to be fair. [2]

- (b) In another game, a player rolls the die twice and his total score is the sum of the two scores.

- (i) Show that the probability of a total score of 3 is $\frac{4}{81}$. [2]
- (ii) Show that the probability of a total score of 6 is $\frac{11}{81}$. [2]
- (iii) Copy and complete the following table. [3]

Total score	2	3	4	5	6	7	8	9	10	11	12
Probability											

- (iv) This game also requires a player to pay an entry fee. Before rolling the die twice, a player selects two of the possible total scores and, if either occurs, he is paid \$1 multiplied by his total score, but nothing otherwise. Determine which two total scores the player should select if he is to maximise his expected profit. [2]
- (v) If the player selects the two total scores you have found in (iv), find, correct to 2 decimal places, how much he should pay to enter the game if it is to be fair. [2]

