UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
General Certificate of Education Ordinary Level

STATISTICS **4040/02**

Paper 2

October/November 2006

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.
The use of an electronic calculator is expected in this paper.

The number of marks is given in brackets [ ] at the end of each question or part question.
At the end of the examination, fasten all your work securely together.

Section A [36 marks]

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

1 **(a)** A white dice and a blue dice are thrown. Each dice is unbiased and has faces numbered 1, 2, 3, 4, 5 and 6.

Events A, B, C and D are defined below.

A: 3 is scored on the white dice.

B: 3 is scored on the blue dice.

C: 3 is scored on both dice.

D: 4 is scored on both dice.

(i) Name **two** of these events which are *mutually exclusive*. [1]

(ii) Name **two** of these events which are *independent*. [1]

(b) Given that events E and F are independent, state, for each of the following, whether or not it **must** be true.

(i) $P(E) = P(F)$. [1]

(ii) $P(E \text{ and } F) = P(E) + P(F)$. [1]

(iii) $P(E \text{ and } F) = P(E) \times P(F)$. [1]

(iv) $P(E \text{ and } F) = 0$. [1]

2 Students in two different classes, *A* and *B*, sat the same examination. The marks obtained by the students are summarised in the following table.

	Number of students in the class	*Mean mark obtained by students in the class*
Class A	21	54
Class B	12	57.5

(i) Calculate, to 1 decimal place, the mean mark obtained by **all** the students in **both** classes. [4]

The sum of the squares of the marks obtained by all the 33 students is 109 528.

(ii) Calculate, to 1 decimal place, the standard deviation of the marks obtained by all the 33 students. [2]

3 The masses of the 300 eggs laid on one particular day by the hens on a small poultry farm were recorded to the nearest gram. The results are summarised in the following table.

Mass (g)	36 – 40	41 – 45	46 – 50	51 – 55	56 – 60
Number of eggs	59	70	71	64	36

For the 51 – 55 class, state

(i) the class boundaries, [2]

(ii) the class interval. [1]

Eggs weighing 44 grams and under as classed as small.

(iii) Estimate the percentage of the day's production of eggs which was classed as small. [3]

4 A machine at the exit to a car park is able to calculate, from information on a ticket inserted into it by each driver, the length of time, t, in completed minutes, for which a car has been in the car park. The following table summarises the information about the length of stay of cars in the car park during one particular day.

t (minutes)	*Number of cars*
$t < 30$	24
$30 \leqslant t < 60$	25
$60 \leqslant t < 90$	41
$90 \leqslant t < 120$	46
$120 \leqslant t < 180$	51
$180 \leqslant t < 240$	42
$240 \leqslant t < 300$	34
$t \geqslant 300$	12
TOTAL	275

Calculate, in minutes, to the nearest minute, the interquartile range of the times for which cars were in the car park on that day. [6]

5 During their weekly gymnastics class, all the pupils in one particular school spend two minutes exercising on a rowing machine. The machine records the number of strokes which each pupil completes during that time. The number of strokes completed, r, is found to have a mean of 54 and a standard deviation of 8.

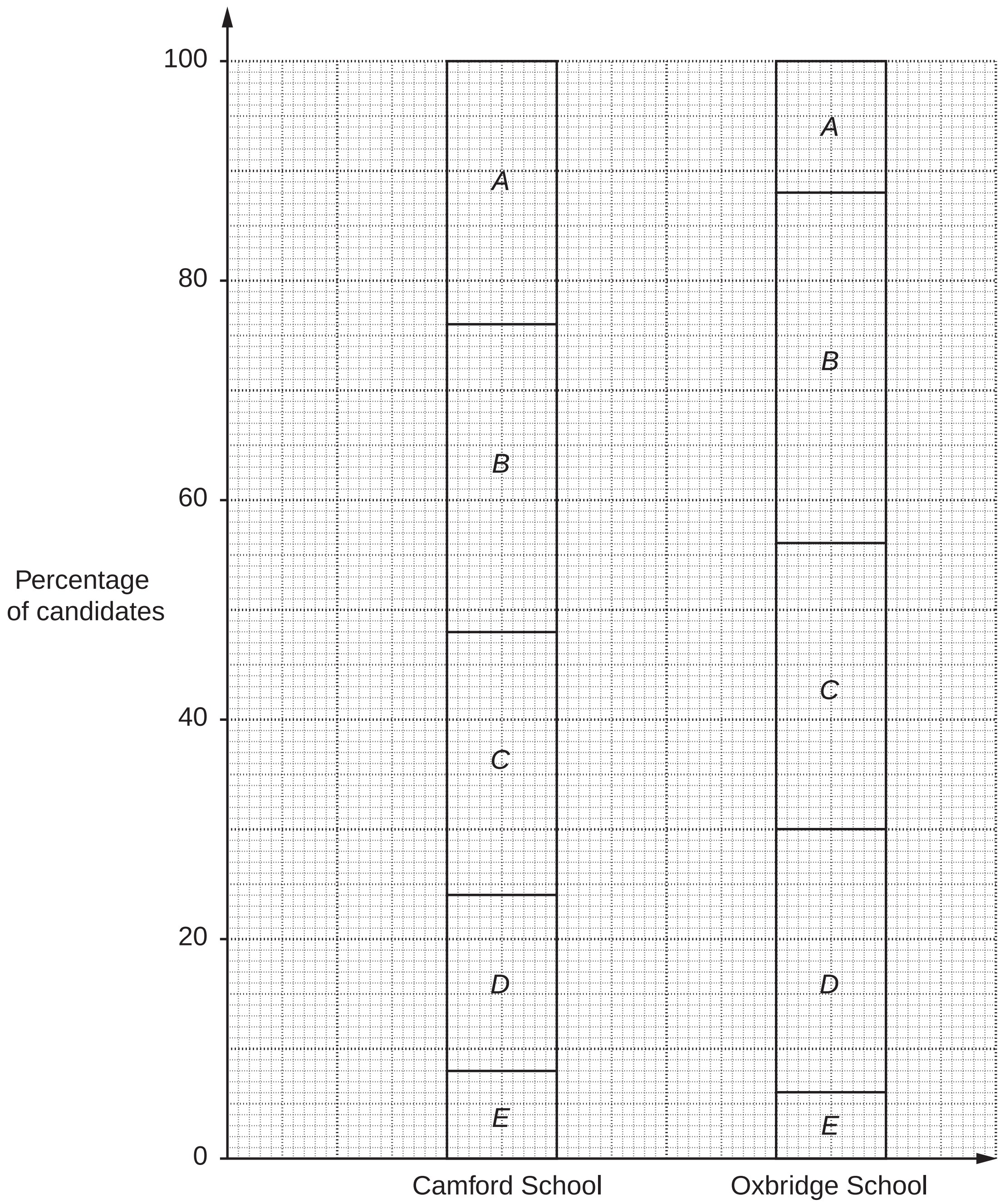
To enable a comparison to be made with pupils' performances in other gymnastic activities, the number of strokes completed are to be transformed to a scaled value, s, which has a mean of 100 and a standard deviation of 10.

(i) Calculate the scaled value for a pupil who completed 78 strokes on the machine during two minutes. [2]

(ii) Calculate the number of strokes completed in two minutes by a pupil whose scaled value was 80. [2]

(iii) Find an expression for the scaled value, s, in terms of the corresponding value of r, and simplify it. [2]

6 The diagram below shows the examination results in O level Mathematics, by grades A, B, C, D and E, of the candidates from two schools in 2004.



(i) State the name given to this type of diagram. [1]

There were 75 candidates from Camford School and 50 candidates from Oxbridge School.

(ii) Showing all your working, find which school had more candidates achieving a grade C in this examination. [5]

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 pupils at a particular school may play either, both or neither of two sports, basketball and tennis. Some of those who play each sport are chosen to represent the school in that sport. The following table summarises the information about the pupils at the school this year.

		Tennis			TOTAL
		Play and represent the school	*Play but do not represent the school*	*Do not play*	
Basketball	*Play and represent the school*	11	22	6	39
	Play but do not represent the school	26	325	41	392
	Do not play	9	52	38	99
TOTAL		46	399	85	530

(i) Interpret the value 9 in the above table. [2]

(ii) One pupil is chosen at random. Calculate the probability that this pupil

(a) plays basketball, but does not represent the school at that sport, [1]

(b) does not play basketball, [1]

(c) plays tennis. [2]

(iii) A pupil, chosen at random, is found to be one who represents the school at basketball. Calculate the probability that this pupil does not play tennis. [3]

(iv) A pupil, chosen at random, is found to be one who does not play tennis. Calculate, to 3 significant figures, the probability that this pupil represents the school at basketball. [4]

(v) Two pupils are chosen at random. Calculate the probability that they both play both sports, but do not represent the school at either sport. [3]

8 A company which runs a large number of supermarkets has four categories of employee: managerial, clerical, skilled manual and unskilled manual. The table below shows the wage rate, in $ per hour, for each of the categories in 1999 and in 2004, together with weights which were determined for the categories in 1999.

Category	*Weight*	*1999 Wage rate* ($ per hour)	*2004 Wage rate* ($ per hour)
Managerial	5	40	50
Clerical	12	22	26
Skilled manual	8	27	32
Unskilled manual	25	11	14

(i) Suggest two different factors which may have been used to determine the weights. [2]

(ii) Calculate wage rate relatives for 2004, taking 1999 as base year, for each of the four categories of employee. [4]

(iii) Calculate a weighted aggregate wage rate index for 2004, taking 1999 as base year. [4]

The total wage bill for a particular one of the company's supermarkets in 1999 was $2.5 million.

(iv) Use the index calculated in **(iii)** to estimate, in $million and to 1 decimal place, the total wage bill for this supermarket in 2004. [3]

(v) Suggest **three** reasons why the estimate obtained in **(iv)** may be very different from the actual wage bill for the supermarket in 2004. [3]

9 A new housing estate contains three streets. Suppose you have been given the task of interviewing a sample of the adult residents of the estate to find out their views on which public facilities are most needed. Each resident has already been allocated a two-digit random number and the following table summarises the information available to you.

Name of street	*Number of adult residents*	*Random numbers allocated*
Albion Road	40	00 – 39
Briony Avenue	30	40 – 69
Cherry Boulevard	20	70 – 89

Using this information, and the two-digit random number table below, you are to use different methods to select a sample of **size 9** from the residents.

Numbers outside the allocated range are to be ignored, and no resident may be selected more than once in any one sample.

TWO-DIGIT RANDOM NUMBER TABLE

66	01	30	06	98	45	43	01	11	29	10	10	50	14	00	91
75	92	63	01	99	87	07	70	94	59	92	00	99	53	65	65
07	11	63	90	26	05	76	11	22	84	50	07	89	25	69	64

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

(b) State how many of the residents in the sample live in each of the three streets. [1]

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

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

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

It is believed that the residents of the different streets may have differing views. A **sample stratified by street** is therefore to be selected.

(iii) **(a)** State how many residents in each street 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 street. Use every number if the street to which it relates has not yet been fully sampled. [3]

(iv) Compare the three samples you have selected as regards how accurately they represent the residents of the different streets. [2]

An alternative method of obtaining the required views would be to sample a number of houses and then interview one resident from each chosen house.

(v) **(a)** Give one advantage and one disadvantage of using this method, rather than sampling residents individually. [2]

(b) Two suggestions for selecting the one resident to be interviewed from each chosen house are: to call at the house and interview whoever answers the door, and to telephone and interview whoever answers the phone. Give one criticism of each of these suggestions. [2]

10 The Information Centre in a winter sports resort has kept records of how many enquiries (in thousands) it received during each four-monthly period for the years 2001 to 2004. These records are summarised in the following table.

Year	*Jan – Apr*	*May – Aug*	*Sep – Dec*
2001	7.5	3.8	15.4
2002	9.1	4.4	18.3
2003	9.3	5.2	17.5
2004	10.9	6.2	19.0

(i) **(a)** State why there is no need to centre moving average values for these data. [1]

(b) Calculate, to 1 decimal place, the appropriate moving average values for these data. Present your results in a suitable tabular form. [7]

(ii) On graph paper, using a scale of 1 cm per four-month period on the horizontal (short) axis, and a scale of 2 cm to 500 enquiries, starting at 8000 enquiries, on the vertical (long) axis, plot the moving average values you have calculated. [You are **not** required to plot the original data.] [3]

(iii) Draw a trend line through the points you have plotted. [1]

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

	Jan – Apr	*May – Aug*	*Sep – Dec*
Seasonal component	–0.9	q	6.6

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

(v) Use your trend line, and the appropriate seasonal component, to estimate the number of enquiries received during the period January – April 2005. [2]

11 Chantelle and Maryann each roll an unbiased cubical dice with faces numbered 1, 2, 3, 4, 5 and 6. Every time they roll their dice, the score, X, is the difference between the numbers shown on the faces which land uppermost. If the numbers on the faces are unequal, the difference is always calculated as being equal to the larger number minus the smaller number.

(i) Copy and complete the following table, inserting, in each cell in the table, the appropriate value of X. [2]

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

(ii) List the six possible values which X can take. [1]

(iii) Tabulate the possible values of X and their probabilities. [6]

Chantelle and Maryann consider playing a number of different games.

Game 1: If the value of X is odd, Chantelle wins the game. If it is even (including 0) then Maryann wins.

(iv) Explain why Game 1 is fair. [1]

Game 2: If the value of X is less than 3, Chantelle gives Maryann \$1. If the value of X is 3 or more, Maryann gives Chantelle \$2.

(v) Explain why Game 2 is fair. [2]

Game 3: If the numbers on the two dice are equal, Chantelle wins. If the numbers on the two dice are unequal, Chantelle gives Maryann a number of dollars equal to the value of X.

(vi) Calculate, in dollars, and to 2 decimal places, the amount which Maryann must give to Chantelle if the numbers are equal, in order to make this game fair. [4]

BLANK PAGE