



UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
General Certificate of Education Ordinary Level

STATISTICS

4040/02

Paper 2

October/November 2007

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.

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 **9** printed pages and **3** blank pages.



Section A [36 marks]

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

- 1 The traffic at a particular road junction is controlled by lights which can show either red or green. Vehicles arriving at lights showing red have to wait until the lights show green. For each of the following, state which one of the options *A*, *B*, *C*, *D* or *E* is a correct description.
- (i) The colours shown by the lights at randomly chosen points in time. [1]
 - (ii) The number of vehicles waiting at lights showing red. [1]
 - (iii) The length of time between consecutive vehicles arriving at a red light. [1]
 - (iv) The number of lights at the junction. [1]

Options

- A* Discrete qualitative variable.
- B* Discrete quantitative variable.
- C* Continuous qualitative variable.
- D* Continuous quantitative variable.
- E* Not a variable.

- 2 *A* and *B* are two events such that

$$P(A) = 0.3,$$

$$P(B) = 0.6,$$

$$P(A \cap B) = 0.18.$$

- (i) State, with a reason,
 - (a) whether *A* and *B* are mutually exclusive, [1]
 - (b) whether *A* and *B* are independent. [2]
- (ii) Calculate the probability that either *A* or *B* but **not both** occurs. [3]

- 3 The values of a variable, X , are formed into a grouped frequency distribution, with class widths of 5, and class boundaries as $0 - 4$, $5 - 9$, $10 - 14$, etc.
For the $5 - 9$ class, give the true class limits and the class mid-point in each of the two cases.

(i) x represents the number of cars parked in a car park at noon on each day during one month. [2]

(ii) x represents the length, rounded to the nearest millimetre, of the leaves on a particular plant. [2]

(iii) x represents the number of **completed** kilometres run in a fixed amount of time by each of a number of athletes. [2]

- 4 The lengths, x mm, of a sample of 99 metal rods produced by one particular machine were measured accurately. The results are summarised in the following table.

<i>Length of rod (x mm)</i>	<i>Number of rods</i>
$160 \leq x < 170$	10
$170 \leq x < 180$	23
$180 \leq x < 185$	29
$185 \leq x < 190$	26
$190 \leq x < 200$	11
TOTAL	99

(i) Calculate the cumulative frequencies for the classes in this table. [1]

(ii) **Calculate**, to 1 decimal place, an estimate of

(a) the median, [2]

(b) the lower quartile, [2]

(c) the 90th percentile [3]

for the lengths of the rods.

- 5 The numbers of male and female students at three different colleges, A, B and C, over the academic year are given in the following table.

	College			TOTAL
	A	B	C	
<i>Number of males</i>	321	91	222	634
<i>Number of females</i>	155	61	381	597

Using the results of any calculations to the nearest integer, illustrate the percentages of males and of females attending each of the three colleges by drawing, on graph paper, appropriate compound bar charts. Use a scale of 2 cm to represent 10% on the vertical axis. [6]

- 6 A population of size 70 contains 30 men, who are each allocated a different two-digit number in the range 00 – 29, and 40 women, who are each allocated a different two-digit number in the range 30 – 69.

EXTRACT FROM RANDOM NUMBER TABLE

48 64 35 14 74 30 52 27 02 99 44 15 15

- (i) Use the above extract from a random number table to select a sample of **size 7, stratified by gender**, from this population. Start at the beginning of the row of numbers and move along the row. Use every number if the gender to which it relates has not yet been fully sampled. List the items in your sample in the order in which they have been selected. [4]
- (ii) For each number which you have rejected before your sample was complete, state the reason for rejecting it. [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) The mean and standard deviation of the three numbers a , b and c are 5 and 2 respectively.

State the mean and standard deviation of the three numbers

(i) $4a$, $4b$, $4c$, [2]

(ii) $(a - 3)$, $(b - 3)$, $(c - 3)$, [2]

(iii) $(3a + 7)$, $(3b + 7)$, $(3c + 7)$. [2]

- (b) Rowena's class sat examinations in French and Spanish. The marks for both examinations are to be scaled to a mean of 50 and a standard deviation of 20. The following table summarises information about Rowena's raw and scaled marks, and the class raw marks.

<i>Subject</i>	<i>Rowena's raw mark</i>	<i>Rowena's scaled mark</i>	<i>Class raw mean</i>	<i>Class raw standard deviation</i>	<i>Scaled mean</i>	<i>Scaled standard deviation</i>
French	82	x	64	12	50	20
Spanish	51	18	y	15	50	20

- (i) Calculate the value of x , Rowena's scaled mark in French. [2]

- (ii) Calculate the value of y , the class raw mean in Spanish. [2]

- (iii) Calculate the French mark which would remain unchanged by this scaling. [2]

- (c) Another class sat a Mathematics examination, for which the class raw mean was 52, and the raw standard deviation 6. The highest individual raw mark was 72. It is required to scale the marks so that the scaled mean is 50 and the scaled highest mark is 100. Calculate the scaled standard deviation which must be used to achieve this. [4]

- 8 A small company classified its expenditure into four categories: Fuel, Labour, Raw Materials and Other Costs. The price relatives for each of these four categories in the years 2002, 2004 and 2006 are given in the table below.

Category	2002	2004	2006
Fuel	100	103	106
Labour	100	105	105
Raw Materials	100	100	102
Other Costs	100	104	103

- (i) Explain what each of the following tells you:

- (a) the 100's in the column headed 2002, [2]
 (b) the 100 for raw materials in 2004, relative to 2002, [1]
 (c) the 105 for labour in 2006, relative to 2004, [2]
 (d) the 103 for other costs in 2006, relative to 2004. [1]

- (ii) Calculate, to 2 decimal places, the price relative for fuel for 2006, taking 2004 as base year. [2]

The company wishes to calculate a weighted aggregate cost index, using weights for the four categories which were calculated in 2002.

- (iii) (a) Briefly describe how these weights could have been calculated. [1]

The weights for fuel, labour, raw materials and other costs were calculated as 33, 17, 19 and 11 respectively.

- (b) Calculate, to 1 decimal place, a weighted aggregate cost index for 2006, **taking 2004 as base year**, and using the weights calculated in 2002. [6]
 (c) Give one reason why the result in (iii)(b) may not give an accurate indication of the company's costs in 2006. [1]

9 In this question give all probabilities as exact fractions.

Two boxes contain counters of different colours, as shown in Table 1.

Table 1

<i>Box</i>	<i>Gold</i>	<i>Silver</i>	<i>Bronze</i>
<i>A</i>	2	3	7
<i>B</i>	3	4	9

Contestants enter a game, in which they pick one counter at random from each box, and are paid a prize if the two counters are of the same colour, as shown in Table 2.

Table 2

<i>Counters</i>	<i>Prize (in dollars)</i>
Two gold	10
Two silver	5
Two bronze	2

- (i) Calculate the probability of winning each of the possible prizes. [4]
- (ii) Calculate the amount in dollars, to 2 decimal places, which contestants should be charged for entry to the game if it is to be fair. [3]
- (iii) The entry charge to the game is \$2. Calculate a contestant's expected gain or loss for the game. [2]

In a new game, all the counters, 5 gold, 7 silver and 16 bronze, are put in one box. Contestants pick two counters from it, at random and without replacement, and contestants are again paid a prize if the two counters are of the same colour, as shown in Table 2.

- (iv) Calculate the probability of winning each prize in this new game. [4]
- (v) Calculate the amount in dollars, to 2 decimal places, which contestants should be charged for entry to this new game if it is to be fair. [2]
- (vi) The entry charge for this new game is also \$2. State, with a reason, whether the new game is to the advantage, or to the disadvantage, of the contestants, compared to the original game. [1]

- 10 The table below gives the number of absences recorded on each of the five schooldays from Monday to Friday, in the first three weeks of a school year at a large school, together with the 5-day totals, and five-day moving average values for the absences.

	Day of week	Number of absences	5-day totals	5-day moving average
Week 1	Monday	47		
	Tuesday	24		
	Wednesday	21	182	36.4
	Thursday	30	186	37.2
	Friday	w	187	37.4
Week 2	Monday	x	185	37.0
	Tuesday	25	190	38.0
	Wednesday	19	192	38.4
	Thursday	35	y	38.6
	Friday	62	196	39.2
Week 3	Monday	52	204	40.8
	Tuesday	28	205	z
	Wednesday	27	206	41.2
	Thursday	36		
	Friday	63		

- (i) Calculate the values of w , x , y and z . [4]
- (ii) On graph paper, using a scale of 1 cm to represent 1 schoolday on the horizontal axis, and a scale of 2 cm to represent 10 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 Tuesday of Week 4. [4]
- (iii) Plot the moving average values on your graph, and draw a trend line through the points. [3]
- (iv) Comment on what the trend line shows, and state, with a reason, whether you would expect the trend to continue over a long period of time. [2]

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

	Monday	Tuesday	Wednesday	Thursday	Friday
'Seasonal' component	12.5	t	-16.5	-5.5	22.6

- (v) Calculate the value of t . [1]
- (vi) Using your trend line and the value of t , estimate the number of absences on the Tuesday of Week 4. [2]

- 11 In this question give your answers either as exact fractions or as decimals significant figures.

Of the 30 workers in a small factory, 13 travel to work by car, 7 travel by bus, 6 cycle and 4 walk.

- (i) Three workers are selected at random. Find the probability that
- (a) they all travel to work by car, [3]
 - (b) none of them travels to work by bus. [4]
- (ii) Two workers are selected at random from those who **do not walk** to work. Find the probability that they travel by different methods of transport. [5]
- (iii) Workers are selected at random, one at a time and without replacement. Find the probability that the fourth worker selected will be the first cyclist selected. [4]

