UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
General Certificate of Education Ordinary Level

CANDIDATE NAME		
CENTRE NUMBER		CANDIDATE NUMBER

6372352199

STATISTICS **4040/23**

Paper 2 **October/November 2012**

2 hours 15 minutes

Candidates answer on the question paper.

Additional Materials: Pair of compasses
Protractor

READ THESE INSTRUCTIONS FIRST

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.
DO **NOT** WRITE IN ANY BARCODES.

Answer **all** questions in Section A and not more than **four** questions from Section B.
If working is needed for any question it must be shown below that question.
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 **17** printed pages and **3** blank pages.

DC (CW/JG) 51084/4

UNIVERSITY *of* CAMBRIDGE
International Examinations

[Turn over

Section A [36 marks]

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

For Examiner's Use

1 **(i)** A and B are two possible outcomes of an experiment, such that

$$P(A) = 0.3 , \quad P(A \cup B) = 0.7 .$$

Find $P(B)$ if A and B are mutually exclusive.

.... [2]

(ii) C and D are two possible outcomes of a second experiment, such that

$$P(C) = 0.5 , \quad P(C \cap D) = 0.4 .$$

Find $P(D)$ if C and D are independent.

.... [2]

(iii) E and F are two possible outcomes of a third experiment.

A student calculated two of the probabilities related to this experiment as

$$P(E) = 0.6 , \quad P(E \cap F) = 0.7 .$$

Comment on the student's results.

....

....

.... [2]

For Examiner's Use

2 In a grouped frequency distribution three consecutive classes are stated as 20 – 24, 25 – 29 and 30 – 34.

Insert, in the table below, the true lower class limit and the true upper class limit of the 25 – 29 class, if the values are

(i) the lengths of the leaves, to the nearest cm, on plants of a particular species,

(ii) the number of candidates sitting the O Level Statistics examination each year at a particular school,

(iii) the **age next birthday**, in years, of people applying for a life insurance policy.

	Lower class limit	*Upper class limit*	
(i)			[2]
(ii)			[2]
(iii)			[2]

3 The students in a class sat an examination, and information about the marks obtained is summarised in the following table.

	Number of pupils	*Sum of marks obtained*	*Sum of squares of marks obtained*
Boys	20	736	30109
Girls	12	480	18147

Calculate

(i) the arithmetic mean of the marks obtained by all the students in the class,

.... [2]

(ii) the standard deviation of the marks obtained by all the students in the class.

.... [4]

4 The table below gives the age distribution, in completed years, of the 120 members of a club.

Age (years)	*Frequency*	*Cumulative Frequency*
18 – 24	20	
25 – 31	35	
32 – 38	25	
39 – 45	18	
46 – 52	12	
53 – 59	7	
60 – 73	3	

(i) Write the cumulative frequencies in the column provided in the table. [1]

(ii) Without drawing a graph, estimate, correct to 1 decimal place, the median age of the club members.

.... [4]

(iii) State with a reason, but without further calculation, which of the lower quartile and the upper quartile you would expect to be closer to the median.

....

.... [1]

For Examiner's Use

5 A motorist recorded the distances, in km to the nearest km, which he drove on each of the 12 journeys he made in his car during the course of one week.

His recorded distances are given below.

7 13 5 6 6 8 14 38 1 1 1 10

(i) For each of the following statements about these recorded distances, state whether it is true or false.

(a) The mode is 1.

.... [1]

(b) Because of the presence of one extreme value, 38, the most appropriate measure of dispersion is the standard deviation.

.... [1]

(c) The range is 3.

.... [1]

(d) The median is 6.5.

.... [1]

(ii) In fact, the journeys had been made on only 6 days of that week, as the motorist had not made any journeys on the Wednesday. A student, Robert, has decided that this means that a value of 0 (for Wednesday) should be included in the recorded distances, thus making 13 observations, and he has used $n = 13$ in all his calculations.

State, with a reason, whether Robert is correct or incorrect in doing this.

....

....

.... [2]

6 Raw values from a distribution with a mean of 50 and a standard deviation of 10 are to be transformed to scaled values in a distribution with a mean of 0 and a standard deviation of 1.

For Examiner's Use

(i) Show that a raw value of 35 corresponds to a scaled value of –1.5.

[1]

Ikram, John and Kofi are all athletes, but compete in different events – the 100 metres, the discus and the decathlon respectively. They each won their respective event in the most recent championships, but wish to compare their performances against those of the other competitors. The following table gives details of the performances of these three athletes, together with the mean and standard deviation of the performances of all the competitors in each of the three events.

Athlete	*Event and units*	*Individual performance*	*Mean of all competitors*	*Standard deviation of all competitors*
Ikram	100 metres (seconds)	12.17	12.42	0.36
John	Discus (metres)	69.21	67.76	3.12
Kofi	Decathlon (points)	8490	8345	217

The performances of Ikram, John and Kofi are to be scaled to a distribution with a mean of 0 and a standard deviation of 1.

(ii) **(a)** Calculate, correct to 2 decimal places, the scaled values of the performances of each of Ikram, John and Kofi.

Ikram ....

John ....

Kofi .... [3]

(b) By comparing the scaled values of the athletes' performances, state, with a reason, which one of the three performed best in relation to the other competitors in his event.

....

....

.... [2]

BLANK PAGE

[Turn over for Section B]

Section B [64 marks]

For Examiner's Use

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

Each question in this section carries 16 marks.

7 In this question, calculate and state all probabilities as fractions.

In a particular game, a player rolls two unbiased six-sided dice, each with faces numbered 1, 2, 3, 4, 5 and 6. The numbers on the uppermost faces are regarded as the numbers shown. The value of the prize a player wins is determined by which of four different outcomes, described below, is achieved.

Outcome	*Description*
A	The numbers shown sum to 3 or 11
B	Both dice show the same number (called a 'double')
C	Neither *A* nor *B* is achieved, but a 6 is shown
D	All other possible outcomes

(i) Calculate P(*A*) and insert it in the table opposite.

[2]

(ii) Calculate P(*B*) and insert it in the table opposite.

[1]

(iii) Show that P(*C*) = 2/9.

[3]

For Examiner's Use

(iv) Calculate P(D) and insert it in the table below.

Outcome	*Probability*	*Prize won* ($)
A		5
B		2
C	2/9	1
D		0

[2]

A player pays an entry fee of $1.50 to play the game once.

(v) Calculate, to the nearest cent, the player's expected profit or loss if he plays the game once.

.... [5]

The game is to be made as fair as possible by altering the prize awarded for obtaining a 'double', but the prize must be a whole number of dollars.

(vi) Calculate the new prize for obtaining a 'double'.

$ .... [3]

For Examiner's Use

8 A school classifies all its expenditure, apart from staffing costs and upkeep of buildings, under three headings – Books & Paper, Equipment, and Consumables (e.g. chemicals, art supplies). The following table gives the mean cost per unit, in dollars, of items bought under these three headings in the years 2000, 2005 and 2010.

	Cost per unit (dollars)		
	2000	2005	2010
Books & Paper	30	45	60
Equipment	50	82	90
Consumables	40	44	52

(i) Using 2000 as base year, calculate the price relative for each heading for 2005.

Books & Paper ....

Equipment ....

Consumables ....[3]

(ii) Hence calculate, correct to 1 decimal place, an unweighted average of relatives index for 2005.

....[3]

The quantities purchased under the three headings in the year 2000 were in the ratio 3 : 4 : 1 respectively.

Weights are to be based on **expenditure** in the year 2000.

(iii) Calculate the weights for the three headings, expressing them as a ratio in its lowest terms.

....[3]

(iv) Calculate, correct to 1 decimal place, a weighted average of relatives price index for 2010, using 2000 as base year.

For Examiner's Use

.... [5]

(v) By referring to your answer to part **(iii)**, give a reason why, in this case, a weighted index is likely to produce a considerably more accurate representation than an unweighted index.

....

....

.... [1]

(vi) Give a reason why the weighted index you have obtained in part **(iv)** may nevertheless not represent the true situation accurately.

....

....

.... [1]

For Examiner's Use

9 Microchips are mass-produced on a production line. The probability of a microchip functioning is 0.9, and all microchips are independent.

(i) For any pair of microchips, calculate the probability that at least one of the pair will function.

.... [2]

Three different designs of an electronic component are to be made using these microchips.

Design *A* contains 3 microchips, all of which must function to enable the component to function.

Design *B* contains 6 microchips arranged in two sets of three. Provided at least one set of three all function, the component will function.

Design *C* contains 6 microchips arranged in three pairs. Provided at least one microchip of each pair functions, the component will function.

(ii) Calculate the probability that a component will function if it is of

(a) design *A*,

.... [2]

(b) design *B*,

.... [3]

(c) design *C*.

.... [3]

(iii) 6000 microchips have been produced. Find which of the three designs would be expected to produce the highest number of functioning components from these microchips, and state what that number is.

For Examiner's Use

Design ....

Number of components ....[6]

10 The 77 students in the Science Department of a college were classified by the subject in which they were specialising and by their home location. The subjects were Biology, Chemistry and Physics. Home location was classified as 'local' (living at home while attending the college), 'national' (from other areas of the country in which the college was situated) and 'international' (from other countries). The following frequency table was produced.

	Local	*National*	*International*	TOTAL
Biology	14	8	0	22
Chemistry	19	8	6	33
Physics	11	6	5	22
TOTAL	44	22	11	77

The students were then each allocated a two-digit random number according to the following table.

	Local	*National*	*International*
Biology	01 – 14	15 – 22	
Chemistry	23 – 41	42 – 49	50 – 55
Physics	56 – 66	67 – 72	73 – 77

Different methods are to be considered for selecting a sample of **size 7** from the students, using the two-digit random number table below. No student may be selected more than once in any one sample, and numbers outside the allocated range are ignored.

TWO-DIGIT RANDOM NUMBER TABLE

67	40	15	82	60	32	02	60	59	99	09	67	01	12	04	36
75	92	41	40	99	03	66	37	59	24	79	75	04	09	15	06
14	14	62	21	03	08	10	61	65	85	78	24	99	48	54	00
12	46	12	14	45	74	13	91	69	89	16	72	88	00	13	01

(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 student 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 numbers of the seven students selected for the systematic sample.

.... [2]

For Examiner's Use

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

(a) State how many students specialising in each subject 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 specialist subject. Use every number if the subject to which it relates has not yet been fully sampled.

.... [2]

(iv) A **sample stratified by home location** is to be selected.

(a) State how many students from each location 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 home location. Use every number if the location to which it relates has not yet been fully sampled.

.... [2]

(v) For each of the four samples you have selected, state how accurately it represents local students specialising in Physics compared to the population as a whole.

....

....

....

.... [3]

(vi) Comment on how accurately any sample of size 7 might be expected to represent this population.

....

....

.... [2]

11 A café opened in a seaside resort at the end of 2008, and over the following three years the owner kept a record of the mean number of drinks sold per day during each quarter of a year. The following table gives these figures for the three-year period, together with appropriate totals and values of a four-point moving average.

Year	*Quarter*	*Mean number of drinks sold per day*	*Four-quarter totals*	*Centred totals*	*Centred moving average values*
2009	I	85			
	II	353			
			1458		
	III	610		2937	367.125
			w = ...		
	IV	410		2976	372
			1497		
2010	I	106		3139	392.375
			1642		
	II	371		3375	421.875
			1733		
	III	755		*x* = ...	440
			1787		
	IV	501		3563	445.375
			1776		
2011	I	160		3608	451
			1832		
	II	360		3676	*y* = ...
			1844		
	III	811			
	IV	513			

(i) Explain why it is necessary to centre the moving average values in this table.

....

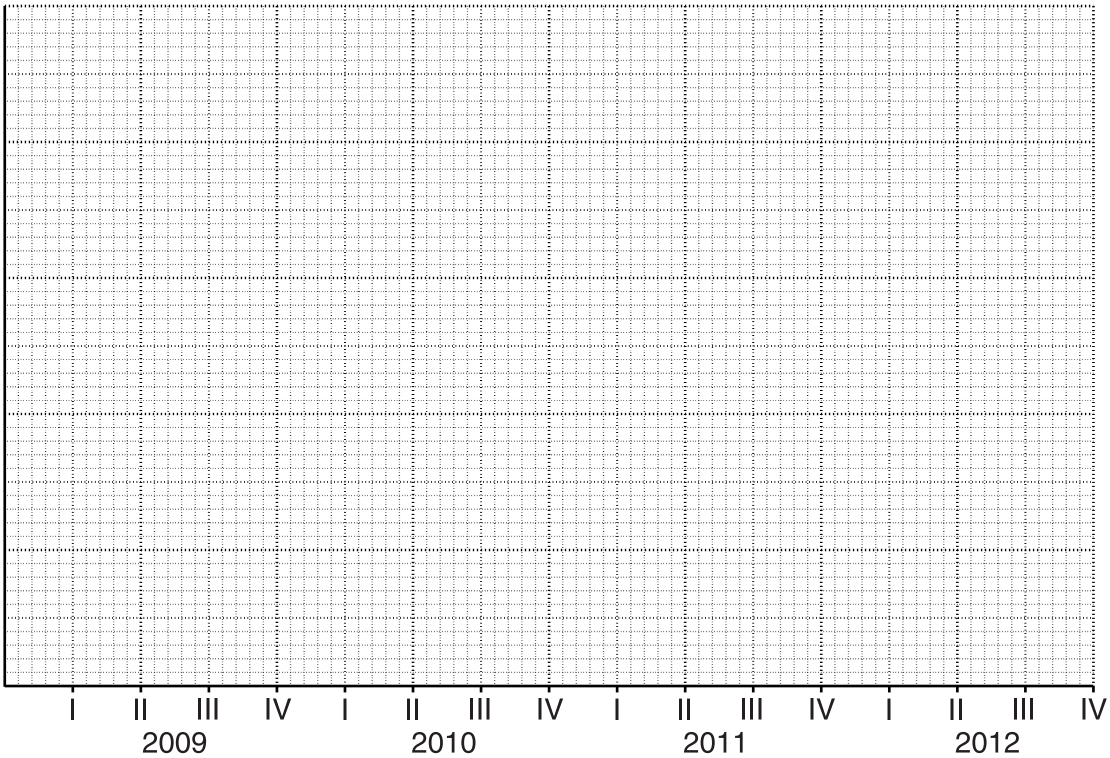
.... [1]

For Examiner's Use

(ii) Calculate the values of w, x and y, and insert them in the table.

[3]

(iii) On the grid below, mark and label an appropriate scale on the vertical axis so that it covers the range 350 to 550.



[2]

(iv) Plot the values of the moving average (**not** the original data) on the grid. [3]

For Examiner's Use

(v) Draw a single straight line on your graph to represent the trend, and comment on what it tells you about sales of drinks at this café.

....

.... [2]

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

Quarter	I	II	III	IV
Quarterly component	–279	–66	q	56

(vi) Calculate the value of q.

q = ... [2]

(vii) Use your trend line and the appropriate quarterly component to estimate the mean number of drinks sold per day in the second quarter of 2012.

.... [2]

(viii) By considering the original data, give a reason why you might have cause for concern about the accuracy of your estimate in part **(vii)**.

....

.... [1]

BLANK PAGE

BLANK PAGE