



UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
General Certificate of Education Ordinary Level

STATISTICS

4040/02

Paper 2

October/November 2008

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 For each of the following variables, state whether it is
discrete or continuous
and whether it is
quantitative or qualitative.
- (i) The number of goals scored by a football team in each game during the course of one season. [1]
- (ii) The colours displayed by traffic lights at a road junction at randomly chosen points in time. [1]
- (iii) The birth weight of babies. [1]
- 2 In an examination, a student wrote the following answers to three parts of a question about the method of moving averages.
- A To obtain an estimate for a future season, the appropriate seasonal component is subtracted from the value given by the trend line.
- B Cyclical variation is an approximately regular pattern repeated in a series of data, but not always over the same period of time.
- C The method of moving averages removes the long term trend from a series of data.
- (i) State which **one** of *A*, *B* or *C* is a correct statement. [1]
- (ii) For each of the other two statements, give one reason why it is incorrect. [2]

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

The net profit, in \$millions, to the nearest \$million, reported by a large company in each of the years 1993 to 2006 is shown in the following table.

<i>Year</i>	1993	1994	1995	1996	1997	1998	1999
<i>Profit (\$million)</i>	18	21	25	22	20	23	29

<i>Year</i>	2000	2001	2002	2003	2004	2005	2006
<i>Profit (\$million)</i>	33	30	26	32	38	40	37

- (i) Using a scale of 1 cm to represent 1 year on the horizontal axis, and a scale of 2 cm to represent \$5million on the vertical axis, plot these values on a graph.

Label the graph, and join consecutive plotted points by straight lines. [5]

- (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 Two events, A and B , are such that

$$P(A) = 0.5,$$

$$P(B) = 0.4,$$

$$P(A \cup B) = 0.7.$$

- (i) Calculate $P(A \cap B)$. [2]

- (ii) State, with a reason in each case, whether A and B are

(a) mutually exclusive, [2]

(b) independent. [2]

- 5 At the start of 2003 the students in a random sample of new students arriving at a college were asked to keep a record of their expenditure on different items during the three years of their course. The data they collected were then used to calculate the price relatives for five different items of expenditure during those three years. These are given in the following table, together with weights allocated to the items by the college from information provided by the students.

<i>Item</i>	<i>Weight</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>
Food	4	100	104	106
Clothing	5	100	101	103
Transport	2	100	103	104
Books	3	100	102	104
Other items	6	100	105	105

- (i) What do the values 100 and 104 in the row headed 'Food' tell you? [2]
- (ii) What do the two values of 105 in the row headed 'Other items' tell you? [1]
- (iii) Calculate a weighted aggregate price index for 2005, using 2003 as base year. [4]
- (iv) Give a reason why this index would be inappropriate to use for the national population. [1]
- 6 The following table summarises the distances, in km, to the nearest km, travelled to work each day by the staff of a large college.

<i>Distance travelled (km)</i>	<i>Number of staff</i>
0 – 4	13
5 – 9	19
10 – 14	23
15 – 19	29
20 – 29	28
30 – 39	18
40 – 59	7
TOTAL	137

- (i) Explain why the class width (class interval) of the 10 – 14 class is 5 km. [2]
- (ii) State the cumulative frequencies for the classes in the table. [1]
- (iii) Without drawing a graph, calculate, giving your answers to 1 decimal place, an estimate of
- (a) the lower quartile, [2]
- (b) the 83rd percentile [3]
- of the distances travelled.

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 house in a small village, in the form of a frequency distribution.

<i>Number of people</i>	0	1	2	3	4	5	6 or more
<i>Number of houses</i>	2	7	8	10	8	5	0

For example, there are 7 houses in each of which only 1 person lives.

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

- (ii) Calculate the total number of people living in the village. [2]

In the remainder of this question express all final answers as either exact fractions, or correct to 3 significant figures.

- (iii) A house in the village is chosen at random. Find the probability that
- (a) it has exactly 5 people living in it, [1]
- (b) it has at most 3 people living in it. [1]
- (iv) A person living in the village is chosen at random. Find the probability that this person lives in a house in which exactly 3 people live. [3]
- (v) Two residents of the village are chosen at random. Find the probability that they live in the same house with no other people. [4]

A house in which at least 4 people live may be referred to as a large household. Residents of the village are chosen one at a time, at random and without replacement, **from those who do not live alone.**

- (vi) Find the probability that the third resident chosen is the first who does not live in a large household. [4]

- 8 (a) Two teams of athletes, *A* and *B*, took part in a competition in which each athlete was awarded a number of points for his or her performance. The following table summarises information about the number of points obtained by the athletes in the two teams.

<i>Team</i>	<i>Number of athletes in the team</i>	<i>Sum of the points scored</i>	<i>Sum of the squares of the points scored</i>
<i>A</i>	22	305	6493
<i>B</i>	13	205	2773

- (i) Calculate the total number of points scored by **all** the athletes. [1]
- (ii) Calculate, correct to 1 decimal place, the mean number of points scored by **all** the athletes. [2]
- (iii) Calculate the sum of the squares of the points scored by **all** the athletes. [1]
- (iv) Hence calculate, correct to 1 decimal place, the standard deviation of the points scored by **all** the athletes. [4]

- (b) A school French examination consists of Paper 1, Paper 2 and Paper 3.

The following table gives the raw marks in the three papers for three pupils, together with the raw means and standard deviations for **all** pupils who took the examination. It also gives the corresponding values after the marks for each paper have been scaled to a mean of 100 and a standard deviation of 12.

<i>Pupil</i>	<i>Paper 1</i>		<i>Paper 2</i>		<i>Paper 3</i>	
	<i>Raw</i>	<i>Scaled</i>	<i>Raw</i>	<i>Scaled</i>	<i>Raw</i>	<i>Scaled</i>
Xavier	69	<i>r</i>	75	103	52	106
Yolande	63	92	83	<i>t</i>	47	100
Zinedine	<i>s</i>	116	69	94	52	106
Mean	67	100	73	100	<i>u</i>	<i>p</i>
Standard deviation	6	12	8	<i>q</i>	10	12

- (i) Find the values of *p*, *q*, *r*, *s*, *t* and *u*. [5]

For each pupil, calculate

- (ii) the sum of the raw marks for the three papers, [1]
- (iii) the sum of the scaled marks for the three papers. [1]

A pupil's final position in the examination is determined by the sum of his or her marks for the three papers.

- (iv) State which pupil would benefit from the use of scaled marks rather than the use of raw marks. [1]

- 9 In a population of size 60, there are 20 men, who are all allocated a different two-digit number in the range 01–20, and 40 women, who are all allocated a different number in the range 21–60. Different methods are to be considered for selecting a sample of size 6 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

05	61	53	50	12	82	25	53	37	83	44	21	66	80	35
28	46	27	41	79	03	63	72	97	99	81	70	63	25	04
84	03	12	94	47	34	62	51	20	11	11	10	55	77	69

- (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) Write down the two-digit numbers of the six people selected for the sample. [2]
- (b) For each number in the row which you did not use before your sample was complete, give the reason why you did not use it. [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 five people selected for the systematic sample. [2]
- (iii) It is believed that gender (male or female) 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 should 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) Compare the three samples you have selected as regards how well they represent the population in terms of gender. [1]
- (v) An interviewer was told to select a sample of the required size by the method of **quota sampling**. Briefly explain why the interviewer would not need to use a random number table to select the people to be interviewed. [2]

- 10** A game at a funfair requires a player to roll an unbiased dice **and** to toss an unbiased coin **and** to pick a card at random from a pack of 52 cards. Before doing this, the player must predict the three outcomes: the number (1, 2, 3, 4, 5 or 6) landing uppermost on the dice, whether the coin lands showing heads or tails, and the suit (hearts, clubs, diamonds or spades) of the chosen card.

Give all your answers in part (i) as exact fractions.

- (i) Assuming that all predictions are made at random, calculate the probability that a player gets
- (a) all three predictions correct, [1]
- (b) all three predictions **incorrect**, [1]
- (c) exactly one of the three predictions correct, [3]
- (d) exactly two of the three predictions correct. [3]
- (ii) State which is the most likely outcome of the game. [1]

Players each pay \$3 to enter the game. The following table shows the prizes which the owner of the game pays for the different outcomes.

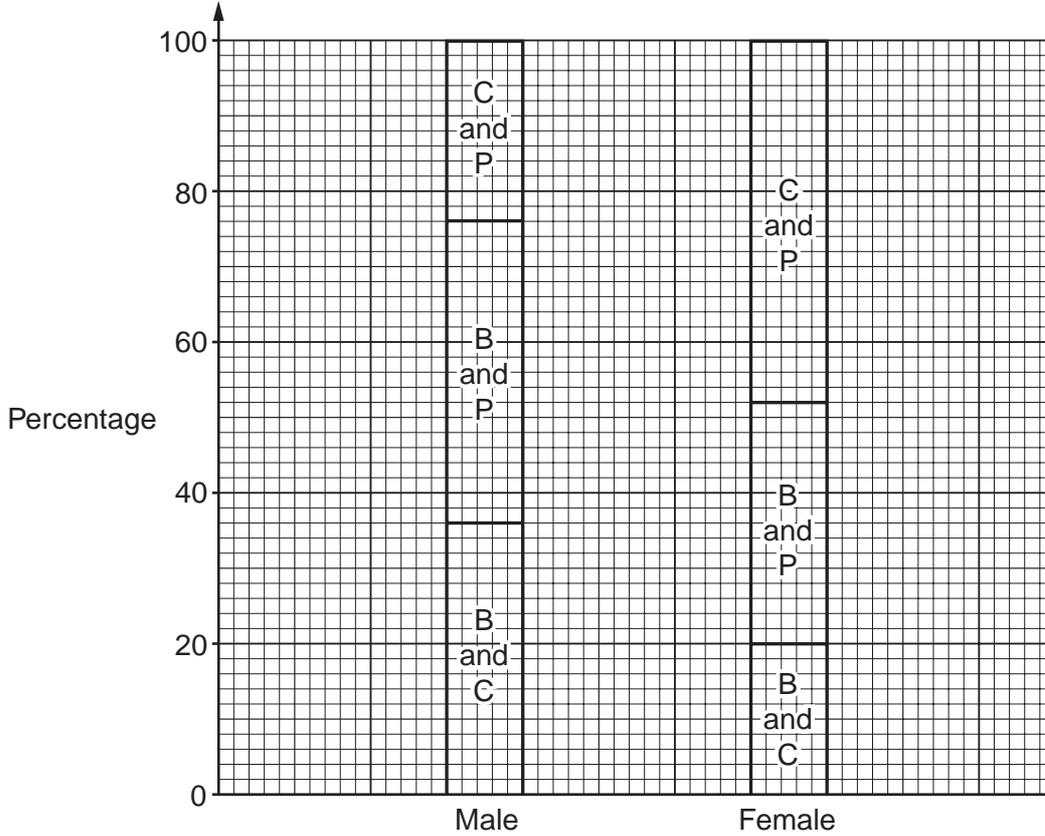
<i>Number of correct predictions</i>	<i>Prize (in \$)</i>
0	No prize
1	1
2	5
3	40

- (iii) Calculate the owner's expected gain for one game. [3]
- (iv) How many times must the game be played if the owner wishes to make an expected gain of \$300? [2]

The owner is considering changing the rules of the game. Instead of predicting the suit of a card, a player would be required to predict its face value (ace, king, queen, jack, 10, 9, 8, etc). Everything else, including the prizes paid, would remain unchanged.

- (v) **Without carrying out any calculations**, state, with a reason, whether the proposed change would be to the advantage of the owner or a player. [2]

- 11 At a particular university students taking a 'double-honours' science degree course study two of the three subjects Biology (B), Chemistry (C) and Physics (P). The sectional (component) percentage bar chart below shows the percentages of male and female students studying each combination of subjects.



- (i) Write down the percentage of males studying each of the combinations. [1]
- (ii) Write down the percentage of females studying each of the combinations. [1]

There are altogether 150 male and 125 female students on the course.

- (iii) Calculate separately the number of males and the number of females who are studying each of the three subjects, Biology, Chemistry and Physics. Present your results in a suitable table. [5]
- (iv) Find, correct to 1 decimal place, the percentage of all students who are studying Biology. [3]
- (v) Using a scale of 2 cm to represent 20 students on the vertical axis, draw a dual bar chart to compare the numbers of males and females studying each of the three subjects. [5]
- (vi) Briefly explain why the total frequency indicated by your bar chart is different from the number of students on the course. [1]

Permission to reproduce items where third-party owned material protected by copyright is included has been sought and cleared where possible. Every reasonable effort has been made by the publisher (UCLES) to trace copyright holders, but if any items requiring clearance have unwittingly been included, the publisher will be pleased to make amends at the earliest possible opportunity.

University of Cambridge International Examinations is part of the Cambridge Assessment Group. Cambridge Assessment is the brand name of University of Cambridge Local Examinations Syndicate (UCLES), which is itself a department of the University of Cambridge.