UNIVERSITY OF CAMBRIDGE INTERNATIONAL EXAMINATIONS
General Certificate of Education Ordinary Level

CANDIDATE NAME

CENTRE NUMBER

CANDIDATE NUMBER

STATISTICS **4040/22**

Paper 2 **October/November 2010**

2 hours 15 minutes

Candidates answer on the question paper.

Additional Materials: Mathematical tables
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 **15** printed pages and **1** blank page.

Section A [36 marks]

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

1 **(i)** Two events A and B are mutually exclusive. State a probability equation which A and B must satisfy.

.....[1]

(ii) Two events C and D are independent. State a probability equation which C and D must satisfy.

.....[1]

(iii) Two events E and F are such that

$$P(E) = 0.6, \quad P(F) = 0.3, \quad P(E \cup F) = 0.8.$$

Find, in each case giving a reason, whether E and F are mutually exclusive and whether E and F are independent.

[4]

2 In a grouped frequency table, values of the variable are given in classes labelled 50 – under 60, 60 – under 70, 70 – under 80, etc.
Insert, in the table below, the true lower and upper class limits of the 60 – under 70 class, if the values are

(i) masses measured to the nearest kg,

(ii) ages expressed in number of **complete** years,

(iii) the number of cars in a car park at the same time each day.

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

3 The following table shows the number of cars of different colours in two different car parks.

Colour	*Number of cars in car park A*	*Percentage of cars in car park A*	*Number of cars in car park B*	*Percentage of cars in car park B*
Silver	40		80	
Black	36		70	
Grey	20		10	
Other colours	64		40	

(i) Explain briefly why a bar chart and not a histogram may be used to illustrate these data.

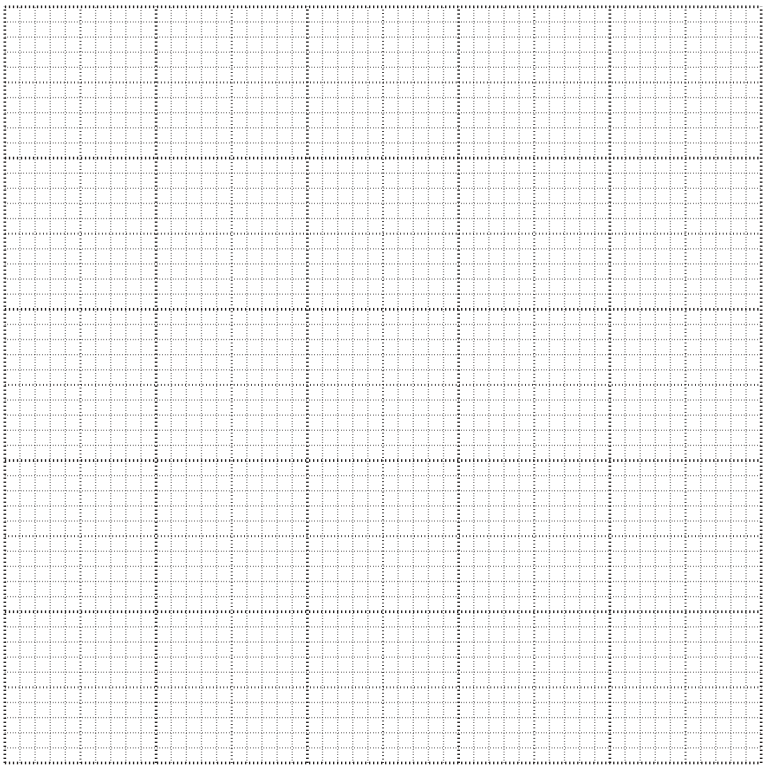
.....

.....[1]

(ii) Complete the above table to show the percentage of cars of the different colours in each of the two car parks.

[2]

(iii) On the grid below, draw, and fully annotate, a sectional (component) bar chart to illustrate the percentages you have calculated in **(ii)**.



[3]

4 48 values of a variable, X, are obtained and are found to have a mean of 2.5.
Two further values of X are then obtained, both of which are 5.

(i) Calculate the mean of all 50 values of X.

.....[4]

(ii) State, with a reason, whether or not it is possible to tell if the standard deviation of X will have changed as a result of the two further values being included.

.....

.....[2]

5 **(i)** The raw marks in a test have a mean of 15 and a standard deviation of 5. They are scaled so that they have a mean of m and a standard deviation of 3.
If a raw mark of 5 corresponds to a scaled mark of 25, calculate the value of m.

.....[2]

(ii) In another test the raw marks have a mean of 56 and a standard deviation of 12. The marks are scaled so that they have a mean of 50 and a standard deviation of 16.
Find the mark which remains unchanged by this scaling.

.....[4]

6 In a game, competitors are required to throw up to 3 darts at a target. A competitor stops throwing after a hit, or after 3 throws if no hits have been recorded.

Prizes are paid for a hit as given in the following table.

Outcome	*Prize ($)*
Hit on first throw	10
Hit on second or third throw	5

The probability that one competitor, Raoul, will hit the target on any throw is 0.2, and all his throws are independent.

(i) Complete the following table of probabilities when Raoul enters the game.

Throw on which hit is recorded	*Probability*
First	
Second	
Third	

[3]

(ii) Calculate the fee which Raoul should pay to enter the game if it is to be fair.

$.....[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 family wished to investigate changes in their cost of living. They chose five items, as given in the table below, from a normal week's groceries, and recorded the price per unit of each item every three months for a year.
The price relatives obtained, taking the prices on January 1st as base, are given in the following table, together with the weights for each item.

Item	*Weight*	*March 31st*	*June 30th*	*September 30th*	*December 31st*
Meat	6	106	108	107	109
Bread	4	103	104	105	107
Milk	5	103	109	110	113
Coffee	2	105	107	109	110
Tea	3	102	105	107	106

(i) **(a)** Calculate a simple average of relatives index for December 31st, taking January 1st as base.

.....[2]

(b) State one disadvantage of using this as an index number.

.....

.....[1]

(ii) Calculate, to the nearest integer, a weighted aggregate price index for December 31st, using January 1st as base.

.....[5]

(iii) Show that the price relative for milk for December 31st, taking June 30th as base, is 104, to the nearest integer.

[2]

(iv) State, with a reason, to which of the following you would expect the weights to be proportional:

A the numbers of each item bought, in a week, by the family,

B the prices of the different items on January 1st,

C the amount of money spent on each item by the family during the first week in January.

.....

.....[2]

(v) A second weighted aggregate price index, using the same weights, was calculated for December 31st taking June 30th as base. The value calculated was 102.

(a) Give a reason for the difference between this value and your answer to **(ii)**.

.....

.....

.....[2]

(b) State a disadvantage of using this second weighted aggregate price index as an index number.

.....

.....

.....[2]

8 **(i)** State why moving averages with an even number of observations per cycle need to be centred, but those with an odd number of observations per cycle do not.

.....

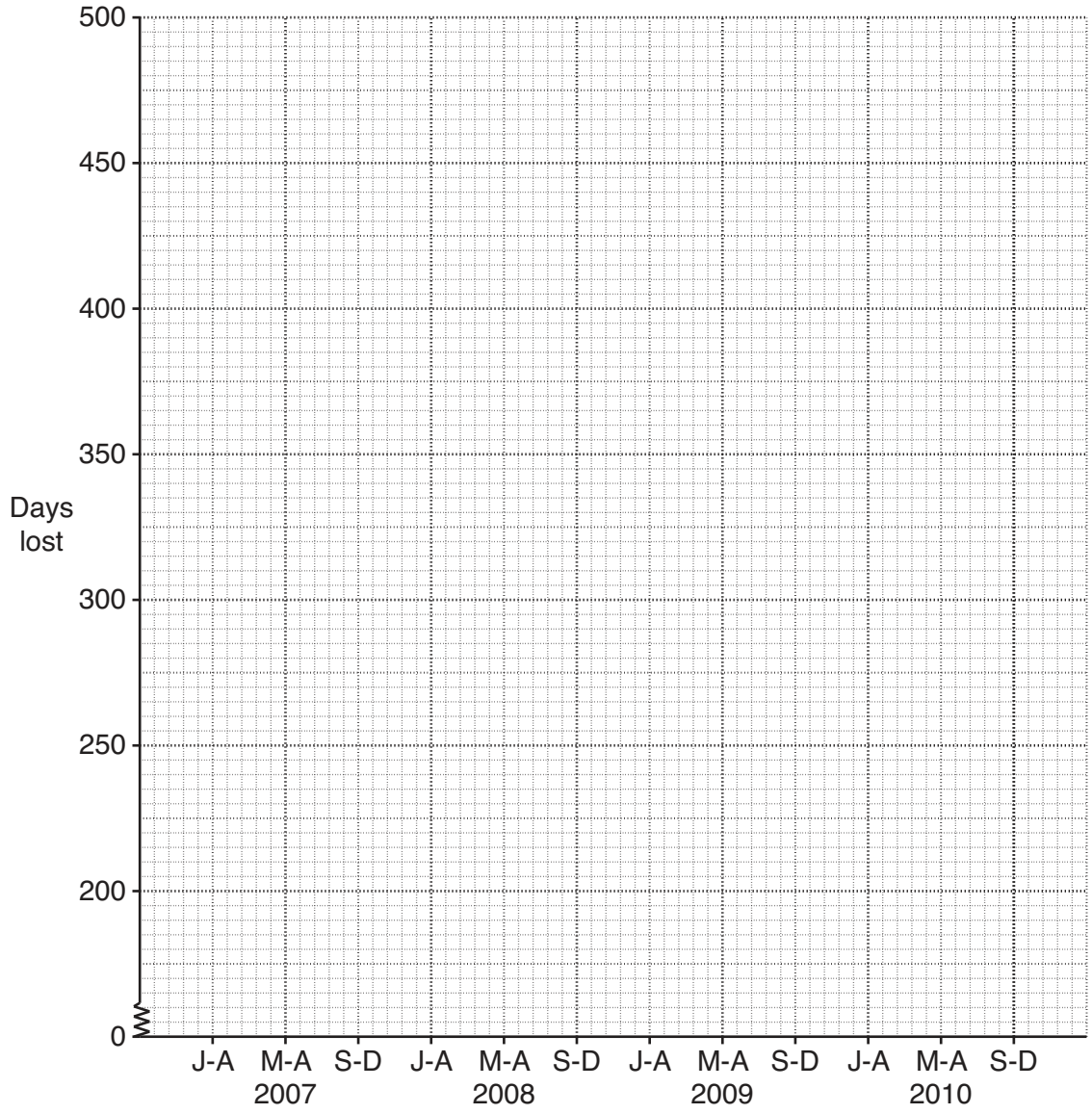
.....

.....[1]

(ii) An industrial organisation keeps a record of the number of days work lost through illness by its employees. The records for three years are summarised in the table below, each figure representing the number of days work lost during a four-month period.

Year	*Number of days work lost*		
	Jan–Apr	*May–Aug*	*Sep–Dec*
2007	459	279	261
2008	468	300	267
2009	477	339	282

Plot these values on the grid below, joining consecutive points by straight lines.



[3]

(iii) Calculate values of a three-point moving average for these figures, and insert them in the appropriate places in the following table. In the space below the table show the full calculation details for **at least one** value of your moving average.

Year	*Three-point moving average*		
	Jan–Apr	*May–Aug*	*Sep–Dec*
2007			
2008			
2009			

[5]

(iv) **(a)** Plot the moving average values on your graph and draw the trend line through them. [2]

(b) Comment briefly on what the trend line shows.

.....

.....[1]

(v) The 'seasonal components' for these data are given in the following table.

Seasonal component		
Jan–Apr	*May–Aug*	*Sep–Dec*
123	*y*	–76

Calculate the value of *y*.

y = ...[2]

(vi) Use the graph and your value of *y* to estimate the number of days work lost during May–August 2010.

.....[2]

9 Give all answers to this question as exact fractions or as decimals correct to 3 decimal places.

The population of a small village is classified by age group and gender as shown in the following table.

	Male	*Female*	TOTALS
Young	16	24	
Middle-aged	29	21	
Old	15	25	
TOTALS			

(i) Insert the row totals, column totals and overall total in the table. [1]

(ii) One person from the village is selected at random. State the probability that the chosen person is

(a) male,

.....[1]

(b) male or not old,

.....[1]

(c) not middle-aged, given that the person is male.

.....[1]

(iii) Three people are selected from the village at random **with replacement**.
Calculate the probability that there will be one person selected from each of the three age groups.

.....[3]

(iv) Three people are selected from the village at random **without replacement**. Calculate the probability that they are all female when

(a) one is selected from each age group,

.....[2]

(b) all are selected from the entire population.

.....[3]

(v) 45% of the young people in the village own a bicycle, as do 30% of the middle-aged, and 35% of the old.
One person from the village is selected at random. Calculate the probability that this person

(a) owns a bicycle,

.....[2]

(b) is middle-aged, given that he/she owns a bicycle.

.....[2]

10 A sample of **size 9** is to be taken from the adult residents of three streets. Each resident has already been allocated a two-digit random number, as shown in the following table.

Name of street	*Number of adult residents*	*Random numbers allocated*
Jamaica Drive	24	00–23
Liberia Avenue	12	24–35
Malawi Road	18	36–53

Using this information, and the two-digit random number table below, you are to use different methods to select the 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

91	00	14	50	10	10	29	11	01	43	45	98	06	30	01	66
65	65	53	99	00	92	59	94	70	07	87	99	01	63	92	75
64	69	25	89	07	50	82	22	11	76	05	26	11	07	40	46

List the residents in each sample in the order in which you select them.

(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.

.....

.....[3]

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

Jamaica Drive .....

Liberia Avenue .....

Malawi Road .....[1]

(ii) A **systematic sample** is to be selected by starting at the beginning of the second row of the table, and moving along the row.

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

..... [1]

(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]

Because of a number of differences between the streets, it is thought that a sample **stratified by street** might be the most appropriate to take.

(iii) (a) State how many residents in each street would be selected for such a sample.

Jamaica Drive .....

Liberia Avenue .....

Malawi Road .....[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.

.....

.....

.....[4]

(iv) Compare the three samples you have selected as regards how accurately they represent the number of residents in each street.

.....

.....

.....[2]

(v) Briefly explain the meaning of

(a) an unbiased sample,

.....

.....

.....[1]

(b) an unbiased method of sampling.

.....

.....

.....[1]

11 This question must be answered by calculation. A graphical solution will receive no marks.

A sales manager drew up the following grouped frequency table of sales from 700 invoices.

Amount on invoice ($)	*Number of invoices*	*Cumulative frequency*
0 – under 10	44	
10 – under 20	194	
20 – under 50	157	
50 – under 100	131	
100 – under 150	69	
150 – under 200	40	
200 – under 500	58	
500 – under 750	7	

(i) State, with a reason, whether the mean or the median is the more appropriate measure of central tendency (average) to represent these data.

.....

.....[2]

(ii) Calculate the cumulative frequencies of the data and insert them in the final column of the table.

[1]

(iii) For the invoice amounts, estimate, correct to two decimal places,

(a) the median,

.....[4]

(b) the interquartile range.

.....[6]

(iv) Estimate, to one decimal place, the percentage of these invoices which are for an amount greater than $160.

.....[3]

BLANK PAGE