CANDIDATE NAME

CENTRE NUMBER

CANDIDATE NUMBER

STATISTICS **4040/22**

Paper 2 **October/November 2013**

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 **20** printed pages.

UNIVERSITY *of* CAMBRIDGE
International Examinations

[Turn over

Section A [36 marks]

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

1 Events *A*, *B*, *C* and *D* are four of the possible outcomes of an experiment such that

$$P(A) = 0.15 , \quad P(B) = 0.2 , \quad P(C) = 0.4 \quad \text{and} \quad P(D) = 0.24 .$$

(i) If events *A* and *B* are independent, find

(a) $P(A \cap B)$,

....[2]

(b) $P(A \cup B)$.

....[2]

(ii) If events *C* and *D* are mutually exclusive, find

(a) $P(C \cap D)$,

....[1]

(b) $P(C \cup D)$.

....[1]

2 (i) The annual salaries of the employees at a company have a mean of m and a standard deviation of s, where $s \neq 0$.
A new employee arrives at the company and is paid an annual salary of m.
The mean and standard deviation of the salaries of the employees are now recalculated to include the salary of the new employee.

For each of the mean and the standard deviation, state whether it will increase, decrease, or stay the same when this new employee's salary is included.

Mean ....

Standard deviation ....[2]

(ii) At another company, at the end of 2011, the employees' annual salaries had a mean of $12\,000$ and a standard deviation of 1000.
During 2012, each of the employees' salaries increased by 5%. At the end of that year they each also received an annual bonus of $200.

Calculate the mean and standard deviation of the annual incomes (salaries plus bonuses) of the employees at the end of 2012.

Mean $ ....

Standard deviation $ ....[4]

3 Ariana and Bella are playing a game.
They each have 4 cards, which are numbered 1, 2, 3 and 4.
Each shuffles her own cards and turns one over at random.

(i) If the cards show the same number, Ariana wins and Bella must pay Ariana $3.
If the cards show different numbers, Bella wins and Ariana must pay Bella $1.

By finding the probabilities of Ariana and Bella winning, show whether or not the game is fair.

[3]

(ii) In a second game the numbers shown on the cards are added together.
If the total is 4 or less, Ariana wins and Bella must pay Ariana $5.
If the total is 5 or more, Bella wins.

If the game is to be fair, how much should Ariana pay Bella if Bella wins?

$ ....[3]

4 The pupils in a class should arrive for registration at 9.00 am. On one particular day, 25 pupils were early, with a mean arrival time of 8.51 am. On the same day, 9 pupils were late with a mean arrival time of 9.21 am, and 2 pupils arrived at 9.00 am exactly.

If x represents the number of minutes a pupil was late (a pupil who was early would have a negative value of x),

(i) find Σx, and hence find the mean arrival time for all 36 pupils.

Σx = ...

Mean = ...[3]

If Σx^2 = 5096 for the 36 pupils,

(ii) find the standard deviation of x, correct to one decimal place.

....[3]

5 The change in a country's annual production (in millions of tonnes) of 4 commodities between 2011 and 2012 is shown in the change chart below.

Change in annual production between 2011 and 2012 (in millions of tonnes)
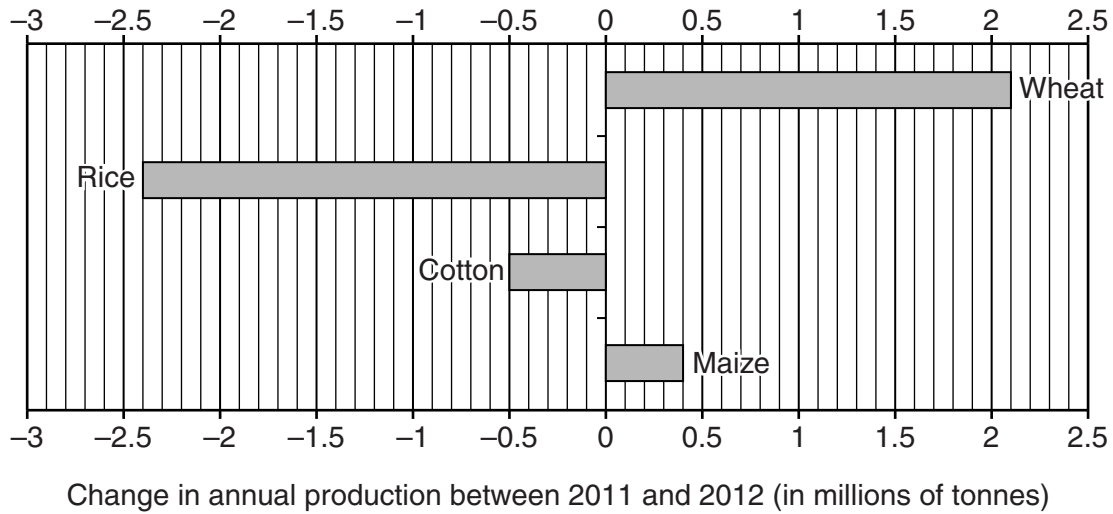
The quantity produced (in millions of tonnes) of the 4 commodities in 2011 in this country is shown in the table below.

Commodity	Quantity produced in 2011 (millions of tonnes)	Quantity produced in 2012 (millions of tonnes)
Wheat	78.6	
Rice	99.2	
Cotton	22.6	
Maize	17.3	

(i) Use these data and the change chart to find the quantities of the commodities produced in 2012 and complete the table. [2]

(ii) On the grid below, draw a dual bar chart to show the quantities produced in 2011 and 2012 of each of the 4 commodities.
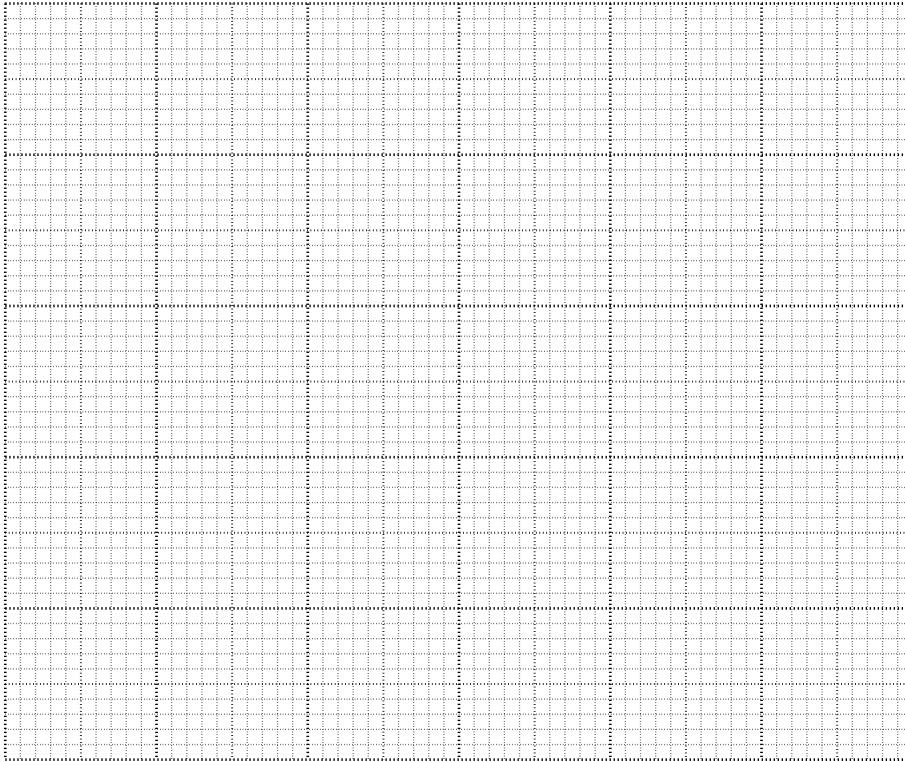
[3]

(iii) State one advantage of a dual bar chart over a change chart.

....

....[1]

6 **(a)** For each of the following state whether the variable is discrete or continuous and whether it is qualitative or quantitative.

	Discrete or Continuous	Qualitative or Quantitative	
(i) the heights of the players in a football competition			[1]
(ii) the towns of birth of the players in a football competition			[1]

(b) A football team used the diagram below to illustrate the number of goals it had scored per match in a season in both the league and cup competitions.
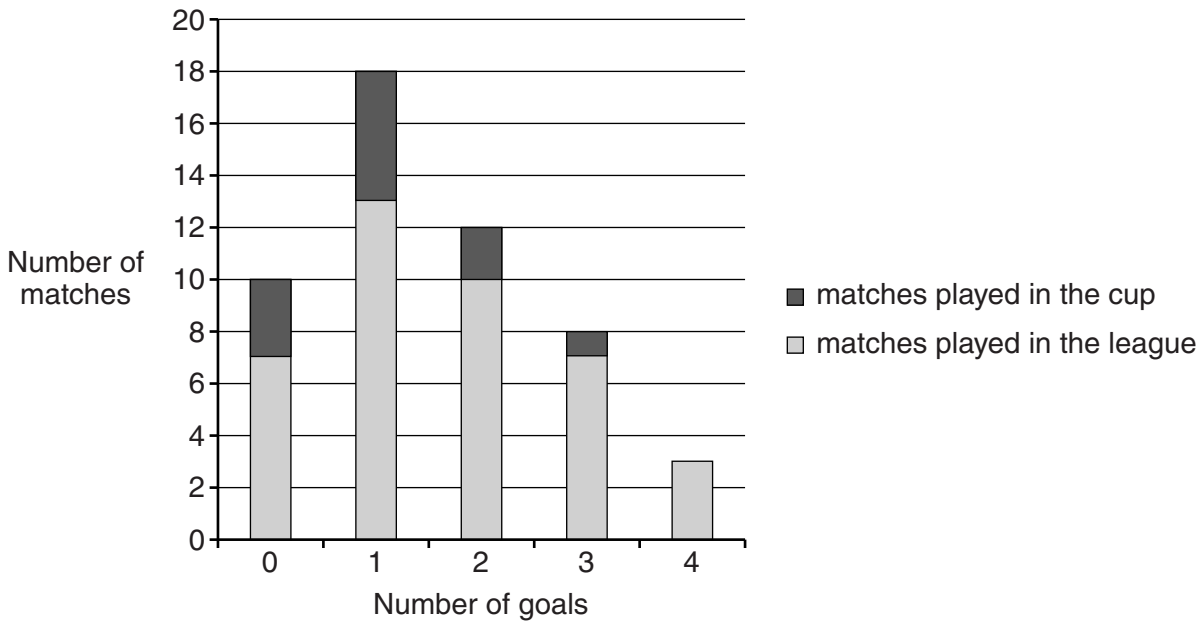
(i) State the full name given to this type of diagram.

....[1]

(ii) Explain why the above diagram is more appropriate than a histogram to illustrate these data.

....

....[1]

(iii) Find the proportion of matches played in the cup in which the team scored 2 or more goals.

....[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 total number of visitors at a tourist attraction has been recorded for every quarter over a three-year period.

   (i) Explain why it might be appropriate to calculate moving average values when establishing the trend in the number of visitors.

   ....

   ....[1]

   (ii) If an *n*-point moving average is to be calculated, state an appropriate value for *n*.

   ....[1]

   (iii) State, with a reason, whether centring would be necessary in this case.

   ....

   ....[2]

(b) A hospital records the number of patients admitted at two-monthly intervals over a period of two years and the results are shown in the table below, together with the 6-point moving average values for these data.

		Number of patients	6-point total	6-point moving average value	Centred moving average value
2010	Jan – Feb	241			
	Mar – Apr	208			
	May – Jun	$x =$			
			1272	212	
	Jul – Aug	185			
			1290	215	
	Sep – Oct	209			
			1290	215	
	Nov – Dec	261			
			1296	216	
2011	Jan – Feb	259			
			$y =$	$z =$	
	Mar – Apr	208			
			1323	220.5	
	May – Jun	174			
			1332	222	
	Jul – Aug	197			
	Sep – Oct	224			
	Nov – Dec	270			

**(i)** Calculate the values of *x*, *y* and *z* and insert them in the table.

[3]

(ii) Calculate the centred moving average values and insert them in the appropriate places in the table.

[3]

(iii) Plot the centred moving average values on the grid below and draw a trend line through the points.
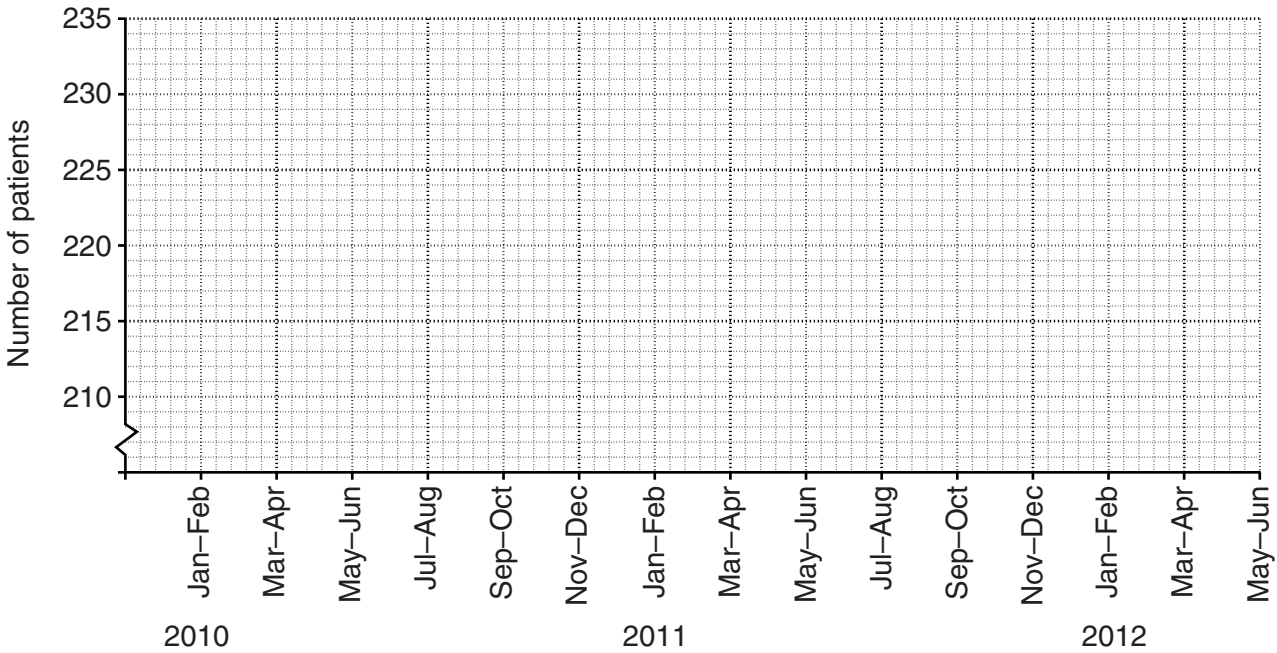
[3]

(iv) Explain what the trend line you have drawn tells you.

....

....[1]

The seasonal component for Mar – Apr is −11.25 .

(v) Estimate the number of patients admitted to the hospital during the period Mar – Apr 2012.

....[2]

8 The students at a college take one of three programmes of study: Physics, Chemistry and Mathematics (PCM) or Physics, Chemistry and Biology (PCB) or Economics, Geography and Mathematics (EGM). The numbers of students who study each programme are shown in the table below.

	PCM	PCB	EGM	TOTAL
Male	60	40	40	140
Female	40	90	30	160
TOTAL	100	130	70	300

(i) Find the probability that a student chosen at random

(a) is a male studying PCM,

....[1]

(b) is female,

....[1]

(c) is studying Physics as part of their programme,

....[1]

(d) is studying PCB, given that they are male.

....[1]

(ii) If two different students are chosen at random, find the probability that they are taking the same programme of study.

....[3]

(iii) If three different students are chosen at random, find the probability that they are each taking a different programme of study.

....[3]

Students are required to buy textbooks for each subject that they study: one textbook for each of Physics, Chemistry and Biology and two textbooks for each of Mathematics, Economics and Geography.

(iv) Find how many textbooks a student taking each programme of study must buy, and complete the table below.

Course	PCM	PCB	EGM
Number of textbooks			

[1]

(v) If one of the textbooks owned by a student at the college is lost at random, find the probability that it

(a) belongs to a student on the PCM programme,

....[3]

(b) is a Mathematics textbook.

....[2]

9 (a) The values of a variable are formed into a grouped frequency distribution, with one of the classes stated as 50 – 60 . State the true class limits of this class if the variable is

	Lower class limit	Upper class limit	
(i) the ages of the residents in a block of flats,			[1]
(ii) the lengths of some rods, measured in mm, to the nearest mm,			[1]
(iii) the lengths of some rods, measured in mm, to the nearest 10 mm.			[1]

(b) A fisherman recorded, in grams (g), to the nearest 100 grams, the masses of 100 fish he had caught in river *A*.

Mass of fish (grams)	Number of fish	Cumulative frequency
100 – 200	12	
300 – 400	31	
500 – 700	29	
800 – 1000	14	
1100 – 1400	8	
1500 – 2000	4	
2100 – 3000	2	

(i) State, with a reason, which of the mean or the median would be the more appropriate measure of central tendency to use in this case.

....

....[2]

(ii) Find the cumulative frequencies and complete the table above. [1]

(iii) Without drawing a graph, calculate an estimate of the interquartile range of the masses of the fish.

....[6]

(iv) The fisherman also recorded the masses of 100 fish caught in river *B* and found the interquartile range of the masses of these fish to be 352 g. Explain what this tells you about the masses of the fish caught in river *B* compared to those caught in river *A*.

....

....[1]

(v) Without drawing a graph, calculate an estimate of the percentage of fish in river *A* with a mass of less than 650 g.

....[3]

10 A hairdresser classifies the expenditure on her business into three categories: Rent, Equipment and Wages.

The cost of Rent has increased from $240 per month in 2010 to $256 per month in 2012.
The price relative of Equipment in 2012 is 110, taking 2010 as base year.
The hourly rate of the Wages of her employees has decreased by 2% between 2010 and 2012.

(i) **(a)** Calculate the price relative, to the nearest whole number, of Rent for 2012, taking 2010 as base year.

....[2]

(b) Explain what the price relative of 110 for Equipment indicates.

....

....

....[3]

(c) State the price relative of Wages for 2012, taking 2010 as base year.

....[1]

(d) Present the price relatives for 2010 and 2012 for each of Rent, Equipment and Wages in a suitable table.

[2]

The hairdresser wishes to calculate a weighted aggregate cost index, using weights calculated in 2010, for the three categories.

(ii) **(a)** Briefly describe how these weights could be calculated.

....

....[1]

The weights in 2010 for Rent, Equipment and Wages were calculated as 7, 2 and 5 respectively.

(b) Calculate, to the nearest integer, a weighted aggregate cost index for 2012, taking 2010 as base year.

....[3]

(c) Her total expenditure on the hairdressing business in 2010 came to $5760. Use your answer to part **(b)** to estimate, to the nearest dollar, her total expenditure on the business in 2012.

....[2]

(d) Give two possible reasons why this estimate might be very inaccurate.

Reason 1 ....

....

Reason 2 ....

....[2]

 [Turn over

11 A small village has a population of 60 people aged 10 and over.
A group of researchers wish to find out what the people of the village think about proposed changes to the timetable for the buses that pass through the village. Each researcher has a list of the population and thinks of a different way to select a sample.

(i) The first researcher plans to stand at the village bus stop at 7 am on a Monday morning and ask the first six people from the population who come to wait for a bus. Explain why this might not produce a reliable sample.

....

....

....[2]

(ii) A second researcher decides to take a simple random sample of size six from the population of 60 people.

(a) Explain what the researcher would need to do with the population list before being able to select the sample from a random number table.

....

....[2]

(b) Use the random number table below, starting at the beginning of the first row and working along the row, to select a simple random sample of size six from the population of 60 people, ensuring that no one is selected more than once.

RANDOM NUMBER TABLE

15 08 73 00 60 15 31 52 86 47 82 99 04 33

23 05 65 27 46 13 81 50 49 34 29 08 94 72

....[2]

(iii) A third researcher decides to take a systematic sample of size six from the population.

  (a) Explain clearly how they should use a random number table to select the first value for such a sample.

  ....

  ....[1]

  (b) Use the random number table below, starting at the beginning of the first row and working along the row, to select a systematic sample of size six.

  RANDOM NUMBER TABLE

    36 04 85 06 63 22 16 64 12 51 25 92 74 43

    35 75 21 44 56 20 83 59 98 35 27 08 14 69

  ....[3]

[Question 11 continues on the next page]

The table below shows the population, split into three different age groups.

	10 – 18 years	19 – 65 years	66 years and over	TOTAL
Number of people	20	30	10	60

(iv) A fourth researcher decides to take a random sample of size six, stratified by age group.

(a) State how many people from each age group would be needed for such a sample.

10 – 18 years ....

19 – 65 years ....

66 years and over ....[1]

(b) Explain clearly what the researcher would need to do before selecting the random sample, stratified by age group, from a random number table.

....

....[2]

(c) Use the random number table below, starting at the beginning of the first row and working along the row, to select a random sample of size six, stratified by age group, ensuring that no one is selected more than once. Use every number if the age group to which it relates has not yet been fully sampled.

RANDOM NUMBER TABLE

17 55 82 25 07 16 35 42 89 37 91 98 24 38

77 29 38 02 47 19 80 53 16 40 28 07 94 73

....[2]

(d) Explain why a random sample, stratified by age group, might be a good idea in this situation.

....

....

....[1]