Cambridge O Level

CANDIDATE NAME	

CENTRE NUMBER		CANDIDATE NUMBER	

STATISTICS **4040/23**

Paper 2 **October/November 2014**

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 an HB pencil for any diagrams or graphs.
Do not use staples, paper clips, 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.

[Turn over

Section A [36 marks]

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

1   The number of DVDs bought in a year by each person in a sample of 46 people is given in the following table.

Number of DVDs bought	12	13	14	15	16	17	18
Number of people	2	5	7	6	9	10	7

For example, 6 people each bought 15 DVDs.

(i)   State the modal number of DVDs bought in the year by these people.

....[1]

(ii)   Find the median number of DVDs bought in the year by these people.

....[2]

(iii)   A number, k, of other people, all of whom bought 22 DVDs in the year, have been omitted from the table.

If these people are included, find

(a)   the greatest possible value of k if the median is unchanged,

....[2]

(b)   the greatest possible value of k if the median and the mode are both unchanged.

....[1]

2 Letters posted in the UK may be sent by either 1st or 2nd class post. 40% are sent 1st class.

The following table shows the number of days after posting on which a letter is delivered.

Days after posting	1st class	2nd class
1	80%	
2	20%	50%
3		30%
4		20%

Find the expected number of days for a randomly chosen letter to be delivered.

....[6]

3 A teacher asked each of the 15 boys and 10 girls in her class to estimate the length, in cm, of a piece of string she showed them.

The estimated lengths are summarised in the following table.

	Number of pupils	Sum of lengths	Sum of the squares of the lengths
Boys	15	270	5372
Girls	10	160	2759

  (i) Calculate the total of the estimated lengths of **all** the pupils.

....[1]

  (ii) Calculate the mean of the estimated lengths of **all** the pupils.

....[1]

  (iii) Calculate the sum of the squares of the estimated lengths of **all** the pupils.

....[1]

  (iv) Hence calculate the standard deviation of the estimated lengths of **all** the pupils.

....[3]

4 A variable, *X*, has a mean of 27 and a standard deviation of 12.

   (i) For purposes of comparison with another variable, values of *X* are to be scaled so that they have a mean of 30 and a standard deviation of 6.

   Find the value of *X* which would be unchanged by this scaling.

   ....[3]

   (ii) The highest value of *X* was 51. It is now wished to scale values of *X* to produce a new variable, *Y*, such that the mean of *Y* is 50 and the highest value of *Y* is 100.

   Find the standard deviation of *Y* necessary to achieve this.

   ....[3]

5 A man who goes on a camping holiday each year classifies his expenditure under three headings: Food, Clothing and Equipment.

The amounts, in dollars, which he spent in each of the years 2010 and 2011 are given in the following table.

Year	Food	Clothing	Equipment
2010	135	165	200
2011	124	132	144

(i) On the grid below illustrate the data by a dual bar chart, using one double bar for each item of expenditure.
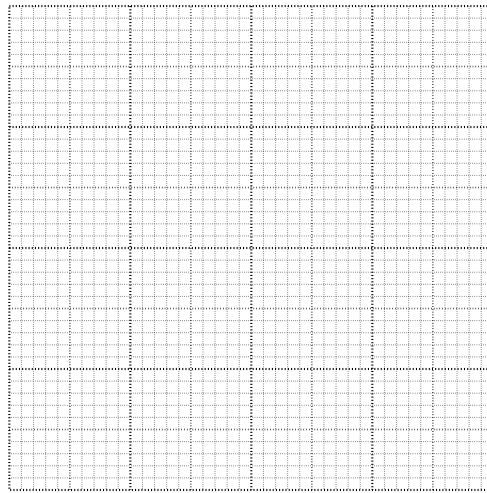
[3]

(ii) On the grid below illustrate the data by a sectional (component) **percentage** bar chart, using one bar for each year.

[3]

6 **(a)** Explain briefly why it is not possible to illustrate a qualitative variable by a histogram.

....

....
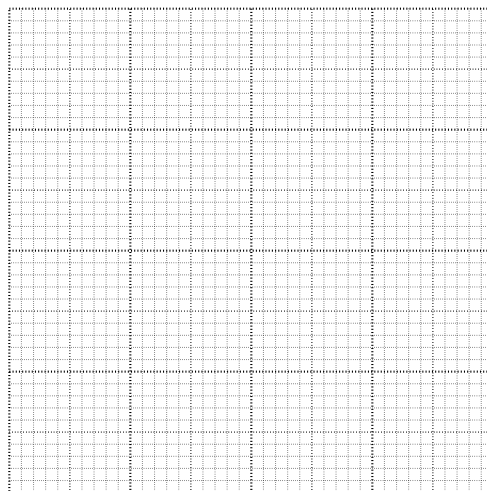
....[2]

(b) A quantitative variable can be either discrete or continuous. Briefly explain the difference between these two types of quantitative variable.

....

....

....

....[2]

(c) A number of athletes all run for one hour and then the numbers of kilometres they have run are formed into a grouped frequency distribution, of which the classes are labelled 12 – 13, 14 – 15, 16 – 17 etc.

State the mid-point of the 14 – 15 class if

(i) the distances have been rounded down to a whole number of kilometres,

....[1]

(ii) the distances have been rounded to the nearest whole number of kilometres.

....[1]

                                                                                                 [Turn over

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 treasurer of a cricket club is carrying out an analysis of changes in club expenditure.
He has summarised expenditure for the year 2012 as follows:

Total cost of maintenance to the grounds and buildings	$10 000
Average cost of one box of three cricket balls	$50
Number of balls purchased	75
Cost of services such as electricity and water supply	$2500
Wage rate paid per hour to the club groundsman	$12.50
Number of hours worked by the groundsman during the year	600

(i) Use these data to show that the treasurer should assign weights to the four categories Maintenance, Balls, Services, Wages in the ratio $8:1:2:6$.

[4]

In 2013, as compared with 2012, the following changes occurred.

Maintenance costs increased by 2%.
By changing the supplier the cost of balls **decreased** by 10%.
Cost of services increased by 5%.
The groundsman's hourly wage rate was increased by 3%.

(ii) Write down price relatives for 2013, taking 2012 as base year, for each of the four categories.

Maintenance ....

Balls ....

Services ....

Wages ....[3]

(iii) Calculate a weighted aggregate cost index for 2013, taking 2012 as base year.

....[4]

(iv) Use the index you have calculated in part **(iii)** and the costs for 2012 to estimate the total cost of running the club in 2013.

.... [3]

(v) Give two reasons why your estimate for 2013 may be very different from the true cost in 2013.

Reason 1 ....

....

Reason 2 ....

....[2]

8 A batch of 500 plastic rods is used as a statistical teaching aid. The following table summarises the lengths of the rods in centimetres.

Length (cm)	Frequency	Cumulative frequency
1 – under 2	12	
2 – under 3	197	
3 – under 4	33	
4 – under 5	13	
5 – under 6	124	
6 – under 7	22	
7 – under 8	11	
8 – under 9	88	

(i) State the modal class.

....[1]

(ii) Find the maximum possible value of the range.

....[1]

(iii) Calculate the cumulative frequencies and insert them in the table. [2]

(iv) Calculate an estimate of the median length of the rods.

....[3]

(v) Calculate estimates of the two quartiles and hence obtain an estimate of the interquartile range.

Lower quartile ....

Upper quartile ....

Interquartile range ....[5]

(vi) (a) Calculate the difference between the median and each of the two quartiles.

....

....[1]

(b) Comment on what your answer to part **(a)** tells you about the shape of this distribution.

....

....[1]

(vii) If a cumulative frequency graph of these data were drawn, state, with a reason, in which part of the graph the gradient would be at its steepest.

....

....

....[2]

 [Turn over

9 **(a)** **(i)** Explain what is meant if two events are described as mutually exclusive.

....

....[1]

(ii) Give an example of two events which are mutually exclusive.

....

....[1]

(iii) Two possible outcomes of an experiment are event *A* and event *B*.

$$P(A) = 0.5 \,, \quad P(B) = 0.6 \,.$$

(a) Explain why, without any further information, it can be stated correctly that *A* and *B* are not mutually exclusive.

....

....[1]

(b) Find the value which $P(A \cap B)$ must have if *A* and *B* are independent.

....[2]

(b) A survey is being carried out at a college as to the 'driving status' of its students.
The following table summarises the responses of a sample of 60 students.

Driving status	Males	Females
Taken driving test and passed	16	8
Taken driving test but failed	7	6
Learning to drive but not yet taken a driving test	10	8
Not started to learn to drive and therefore not yet taken a driving test	2	3

Find the probability that

(i) a student chosen at random has passed a driving test,

....[2]

(ii) a male student chosen at random has taken a driving test,

....[2]

(iii) a female student chosen at random has not yet taken a driving test,

....[2]

(iv) two students chosen at random have both not started to learn to drive,

....[2]

(v) of two students chosen at random, without replacement, the first is male and the second has taken a driving test but failed.

....[3]

10 A statistician wishes to invite 5 people to dinner from among her fifteen closest friends and relatives, and decides to select the 5 by applying a sampling procedure to a list of the fifteen.
The following table lists the fifteen people, numbered 00 to 14, classified by age group and by whether they are a friend or a relative.

Person	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14
Age group	II	I	II	II	I	III	I	I	II	III	I	II	II	I	III
Friend/relative	R	R	R	F	R	F	F	R	F	F	F	F	R	F	F

Age group: I = 18 – 29 years, II = 30 – 49 years, III = 50 – 69 years.
F = friend, R = relative.

You are asked to help the statistician by applying four different sampling procedures to select a sample of size 5 from this population, using the two-digit random number table below. No person may be selected more than once in any one sample.

TWO-DIGIT RANDOM NUMBER TABLE

```
61  12  00  18  12  07  09  53  01  15  74  45  14
91  65  05  78  35  00  26  25  11  32  03  19  21
06  15  09  79  08  59  10  03  04  37  02  99  01
72  24  11  29  13  18  15  10  66  06  02  00  09
```

(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 by starting at the beginning of the second row of the table, and moving along the row.

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

....[1]

  (b) Write down the number of the first person selected.

....[1]

  (c) Write down the numbers of the other four people selected for the systematic sample.

....[1]

(iii) A sample stratified by whether a person is a friend or a relative is to be selected.

(a) State how many friends and how many relatives would be selected for such a sample.

.... friends

.... relatives [1]

(b) Starting at the beginning of the third row of the table, and moving along the row, select this sample. Use every number if the category to which it relates has not yet been fully sampled.

....

....[3]

(iv) A sample stratified by age group is to be selected.

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

.... from age group I

.... from age group II

.... from age group III [1]

(b) Starting at the beginning of the fourth row of the table, and moving along the row, select this sample. Use every number if the age group to which it relates has not yet been fully sampled.

....

....[2]

(v) State, for each of the two stratified samples, with a reason, whether or not it represents the population exactly.

Stratified by friend/relative ....

....

Stratified by age group ....

....[4]

11 The numbers of absences recorded per day in a school over a period of three weeks are shown in the table below.

| Week	Day	Number of absences	5-point moving total	5-point moving average
1	Monday	35		
	Tuesday	20		
	Wednesday	18	140	28.0
	Thursday	24	137	27.4
	Friday	43	133	26.6
2	Monday	32	x = …….	25.4
	Tuesday	16	125	25.0
	Wednesday	12	121	24.2
	Thursday	22	122	24.4
	Friday	39	124	y = …….
3	Monday	33	128	25.6
	Tuesday	18	127	25.4
	Wednesday	16	125	25.0
	Thursday	21		
	Friday	37		

(i) State why it is most appropriate to calculate values of a 5-point moving average in order to analyse these data.

....

....[1]

(ii) Give a reason why the moving average values have not been centred.

....

....[1]

(iii) Plot the numbers of absences on the grid below and describe what they show.

....
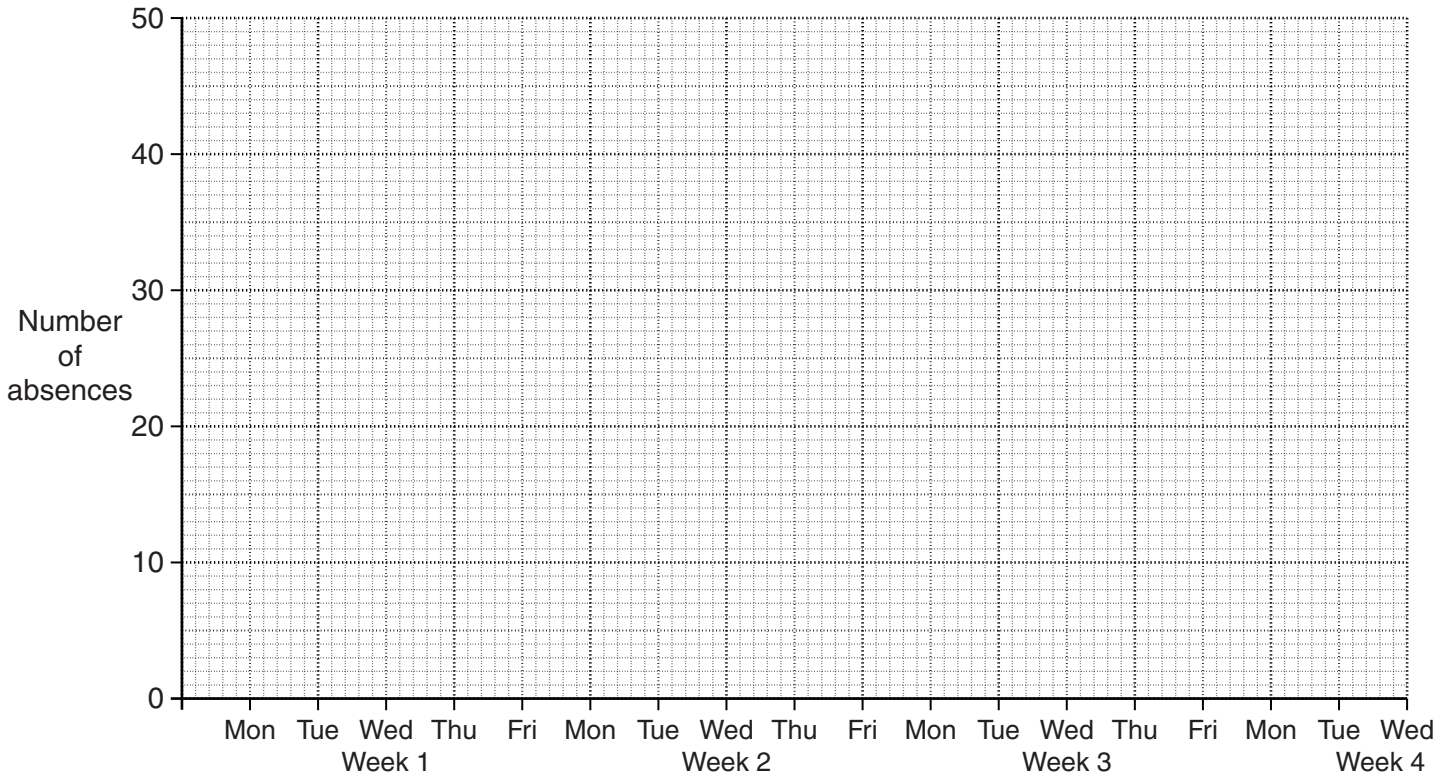
....[3]

(iv) Calculate the values of *x* and *y* and insert them into the table.

[2]

(v) Plot the values of the moving average on your graph. [2]

(vi) State the purpose of calculating moving averages, and indicate whether this appears to have been achieved in this case.

....

....[2]

(vii) Draw an appropriate trend line on your graph. [1]

[Question 11 continues on the next page]

 [Turn over

(viii) The seasonal components for the days of the week are given in the following table.

Day of week	Monday	Tuesday	Wednesday	Thursday	Friday
Seasonal component	7	−8	−11	q	15

Find the value of q.

....[2]

(ix) Use your trend line and the appropriate seasonal component to estimate the number of absences on the Tuesday of Week 4.

....[2]

BLANK PAGE

20

BLANK PAGE