CANDIDATE NAME	

CENTRE NUMBER						CANDIDATE NUMBER				

STATISTICS **4040/22**

Paper 2 **October/November 2011**

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 **20** printed pages.

[Turn over

Section A [36 marks]

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

1 A college has asked a statistician to carry out a survey of its catering facilities. As part of his survey the statistician measures the time that students, chosen randomly, spend in a queue in the refectory, from the moment they join the queue until the moment they leave the serving area. The following times, **in completed minutes**, are those recorded for twelve of the chosen students.

7 8 7 0 1 1 6 5 3 7 4 6

For each of the following state whether it is true or false about this sample of data.

(i) The mode of the times is 8 minutes.

....[1]

(ii) The median time is 5.5 minutes.

....[1]

(iii) The range is a suitable measure of dispersion for this data.

....[1]

(iv) The mean waiting time is 5 minutes.

....[1]

(v) The mean waiting time is 4.58 minutes (to 2 decimal places).

....[1]

(vi) The value 0 must be incorrect because a person has to spend at least a short time being served.

....[1]

2   **(a)** Give a brief explanation of each of the following terms in the context of moving averages.

    (i) Time series ....

    ....

    ....[1]

    (ii) Seasonal variation ....

    ....

    ....[1]

    (iii) Trend ....

    ....

    ....[1]

    (b) **(i)** Give an example of a variable which you would expect to be subject to seasonal variation.

    ....[1]

    (ii) State what would be a 'season' for your chosen variable.

    ....[1]

    (iii) State, with a reason, whether moving average values based on your chosen variable would need to be centred.

    ....

    ....[1]

3 The following table gives the exchange rates between the United States dollar and a number of other currencies on 1 January 2009, and on 1 January 2010. For example, on 1 January 2009, one dollar could be exchanged for 0.690 pounds. The final column of the table is to show the percentage change in each exchange rate between the two dates.

| Currency	Exchange rate on 1/1/09	Exchange rate on 1/1/10	Percentage change in exchange rate
Pound	0.690	0.620	−10.1
Yen	90.459	92.586	+2.4
Euro	0.712	0.696	
Rupee	49.900	46.899	
Rouble	29.158	30.314	

(i) Give the calculations to show that the percentage change in the pound exchange rate is −10.1%.

[1]

(ii) Calculate the percentage changes in the exchange rates of each of the euro, rupee and rouble, and insert them in the table.

[1]

(iii) Complete the change chart below to illustrate the percentage changes for all five currencies.
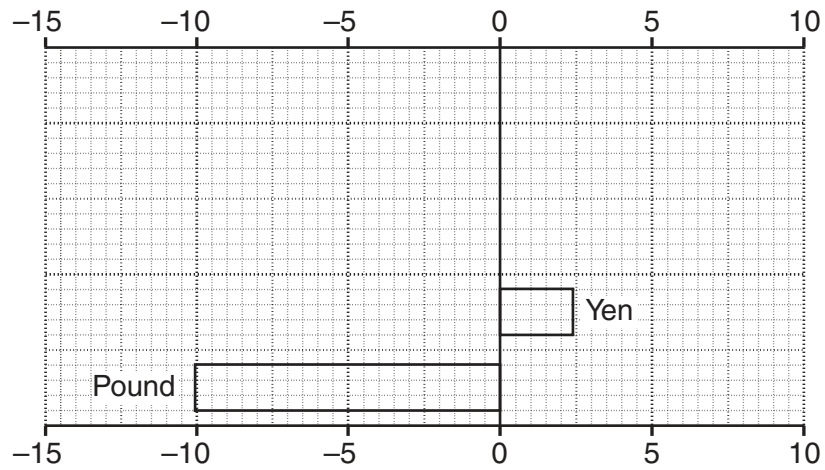
Percentage change in exchange rates 2009–2010

[2]

(iv) Give one advantage and one disadvantage of displaying data in a change chart.

Advantage..

....

....

Disadvantage..

....

....[2]

 [Turn over

4 Values of a variable, Y, were measured for each of the four quarters of three consecutive years. Using these measurements, values of an appropriate centred moving average were calculated. These were, in turn, used to estimate the four seasonal components, and to draw the trend line of values of Y from which the seasonal components had been removed.

The estimated seasonal components are given in the following table.

Season	I	II	III	IV
Seasonal Component	12	q	−16	−2

(i) State, giving a reason, whether you would expect the value of Y for quarter III of any year to be above or below the trend line.
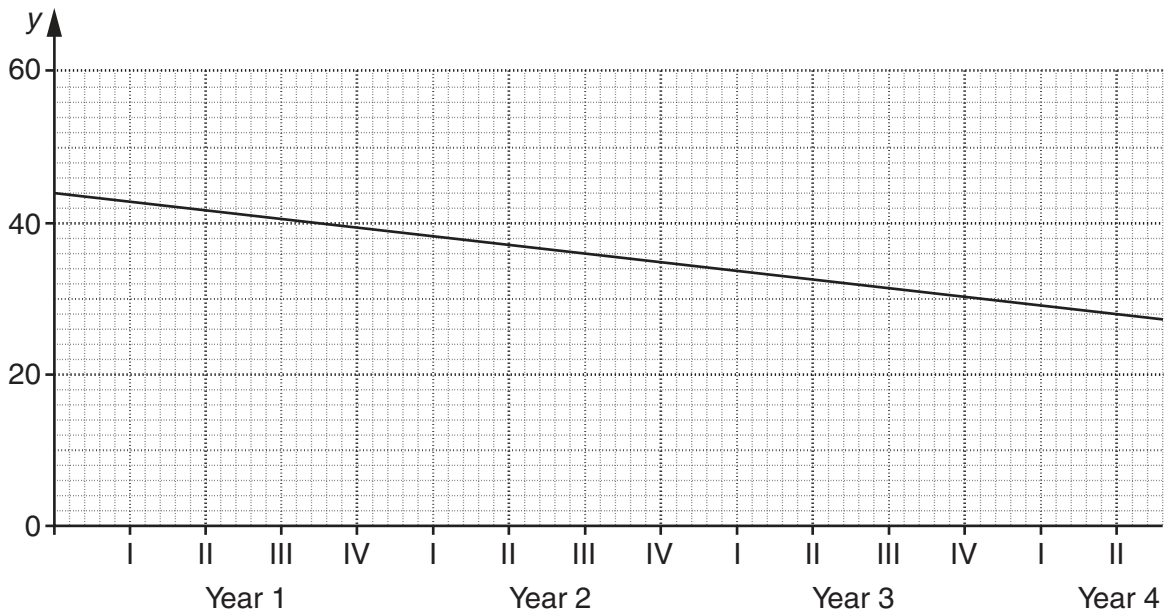
....

....

....[2]

(ii) Find the value of q.

$q =$ .... [2]
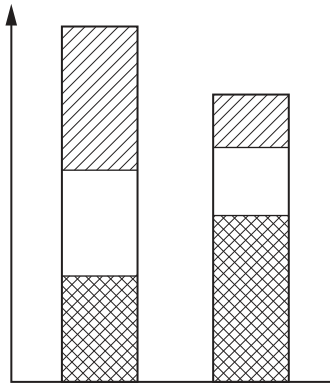
The grid below shows the graph of the trend line.

(iii) Use the graph to estimate the value of Y for quarter I of year 4.

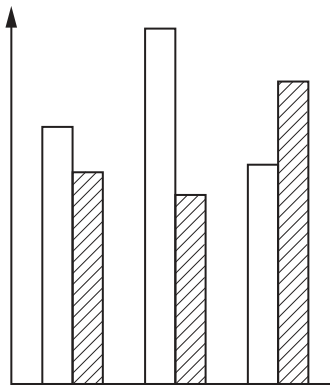..[2]

5 (i) Name the type of bar chart illustrated in the diagram below.

....[1]

(ii) Name the type of bar chart illustrated in the diagram below.

....[1]

(iii) Both diagrams are incomplete. List four items of information which need to be added to each of them to make them meaningful.

1 ....

2 ....

3 ....

4 ....[2]

(iv) A diagram of the type in part **(ii)** is to illustrate the quantities of different blends of tea drunk in the United Kingdom in three consecutive years. State one advantage and one disadvantage of using this type of diagram **for this purpose**.

Advantage...

....

....

Disadvantage...

....

....[2]

6 The table below shows the marks obtained in two tests, Mathematics and Statistics, by each of two students, Peter and John. It also shows the mean and the standard deviation of the marks in these two tests of all the students in the class of which Peter and John are members.

	Mathematics	*Statistics*
Peter	40	43
John	30	48
Class mean	20	44
Class standard deviation	10	4

The marks in both subjects are to be scaled to a mean of 50 and a standard deviation of 16.

(i) Obtain the scaled marks for both subjects for both students, and present your results in a suitable table.

[4]

Each student's total mark is obtained by adding the two test marks together.

(ii) Show that one student has the higher total raw mark, but the other has the higher total scaled mark.

[1]

(iii) Explain briefly what has caused the situation described in part **(ii)**.

....

....

....[1]

Section B [64 marks]

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

Each question in this section carries 16 marks.

7 In a game, a player throws an unbiased dice, with faces numbered 1, 2, 3, 4, 5 and 6, three times. On any throw, if the number on the upper face is 6, or if the number on the upper face on the previous throw was 6, then she scores the number on the upper face, otherwise she scores 0. The player's total is then the sum of her three scores.

For example,

the sequence 4 6 3 gives a total of 0 + 6 + 3 = 9,
the sequence 6 1 6 gives a total of 6 + 1 + 6 = 13,
the sequence 6 2 5 gives a total of 6 + 2 + 0 = 8.

Possible totals are 0, and all integers from 6 to 18 inclusive.

(i) State one possible sequence of three throws which would give a total of 0.

....[1]

(ii) Calculate the probability that a player's total is 0.

....[2]

The entry fee to play the game once is $1. A prize of $8 is paid for a total greater than 15, and $2 is paid for a total between 10 and 15 inclusive. Nothing is paid for a total of less than 10.

(iii) Calculate the probability of winning a prize of $8.

....[5]

The probability of failing to win a prize is $\frac{5}{6}$.

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

....[6]

(v) If the entry fee and other prizes remain unchanged, calculate, to the nearest cent, the prize which should be paid for a total of 10 to 15 inclusive to make the game fair.

....[2]

 [Turn over

8 The table below shows, under five headings, the expenditure, to the nearest $1000, of a small company in the year 2007, together with the price relatives in the year 2008 for the same headings, taking 2007 as the base year.

	2007 expenditure ($'000)	2008 price relatives
Wages	375	108.8
Equipment	72	116.7
Buildings	100	107.0
Fuel	52	140.4
Telephone	131	100.0

(i) State what can be deduced from the value of 100.0 for the telephone price relative.

....

....

....[1]

(ii) Calculate, to 1 decimal place, an 'unweighted average of price relatives' index for 2008, taking 2007 as base year.

Unweighted index = ... [2]

(iii) Using the expenditure for 2007 as weights, calculate, to 1 decimal place, a weighted aggregate cost index for 2008.

Weighted index = ... [5]

(iv) State, with a reason, which of the two indices you have calculated in parts **(ii)** and **(iii)** is more likely to reflect the actual increase in expenditure.

....

....[2]

(v) Use the total 2007 expenditure, and the two indices, to obtain an estimate, for each index, of the total 2008 expenditure.

Estimate for unweighted index = ...

Estimate for weighted index = ... [3]

When the company's accounts for 2008 were published, they showed that actual expenditure in the year had been $807 000.

(vi) State, with a reason, whether this figure justifies your answer to part **(iv)**.

....

....

....[1]

(vii) Using the information in the table, suggest a reason why one of the two indices is far too high.

....

....

....[2]

9 **(a)** A standard pack of 52 playing cards is shuffled, and one card placed face down on a table.

    (i) State the probability that this card is an ace.

....[1]

  You are then told that this card is a heart.

   (ii) What is now the probability that this card is an ace?

....[1]

  (iii) State, with a reason, what the probabilities you have given in parts **(i)** and **(ii)** tell you about the events 'the card is an ace' and 'the card is a heart'.

....

....

....[2]

  (iv) State, with a reason, whether the events 'the card is an ace' and 'the card is a heart' are mutually exclusive.

....

....

....[2]

(b) The events A and B are two possible outcomes of a statistical experiment such that $P(A) = 0.7$, $P(B) = 0.3$, $P(A \cap B) = 0.21$.

   (i) State, with a reason, whether events A and B are mutually exclusive.

....

....

....[2]

(ii) State, with a reason, whether events A and B are independent.

....

....

....[2]

(iii) Calculate the quantity $P(A \cup B)$, and interpret in words what it represents.

....

....[4]

(iv) Calculate the quantity $P(A) + P(B) - 2P(A \cap B)$, and interpret in words what it represents.

....

....[2]

10 A publisher of transport literature published 70 new books last year, 40 on rail transport, 20 on road transport and 10 on sea transport.

You have been asked to carry out a survey of these books, but have only enough resources of time and money to investigate a 10% sample.

(i) State the number of books you would need to include in your sample.

....[1]

(ii) If you were told that the purpose of the survey was not related to different types of transport, how might this affect your choice of which method of sampling to use?

....

....

....[1]

You have allocated the two-digit random numbers 00 to 39 to the rail books, 40 to 59 to the road books, and 60 to 69 to the sea books, and are going to use the random number table below to select various samples. Numbers outside the allocated range are to be ignored, and no book may appear more than once in any one sample.

TWO-DIGIT RANDOM NUMBER TABLE

```
20  74  21  20  30  25  38  35  83  39
87  99  19  85  30  95  39  06  13  04
02  33  45  60  55  41  85  66  10  36
```

(iii) (a) 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.

....[1]

(b) State, giving reasons, which two of the **first seven** numbers in the first row of the table would not be included in your sample.

....

....

....[2]

(iv) A **systematic sample** is to be selected.

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

....[1]

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

....[2]

(v) A **sample stratified by type of transport** is to be selected.

(a) State how many books on each type of transport 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 type of transport. Use every number if the type of transport to which it relates has not yet been fully sampled.

....[4]

(vi) Explain whether or not the composition of the **simple random sample** you have selected in part **(iii)(a)** would give you cause for concern.

....

....

....[2]

11 This question must be answered by calculation; graphical solutions will not be awarded any marks.

The following table summarises the net value, in dollars, of the 690 orders received in one week by a mail order company. The net value is the total cost of the goods purchased before taxes and any other additional charges are added.

Net value of order ($)	Number of orders	Cumulative frequency
5 – under 10	85	
10 – under 20	120	
20 – under 30	225	
30 – under 50	135	
50 – under 100	105	
100 – under 150	20	

(i) Without carrying out any calculations, explain why the mean net value of these orders would be expected to be greater than the median.

....

....

....[2]

(ii) State, with a reason, which measure of dispersion would be most appropriate to represent this data.

....

....

....[2]

(iii) Calculate the cumulative frequencies, and write them in the third column of the above table.

[1]

Because it finds small orders uneconomic to process, the company is considering not accepting orders with a net value below $12.50.

(iv) Estimate, correct to 1 decimal place, the percentage of these orders which would not be accepted if this limit had been in place.

....[3]

A voucher giving a discount off a future order is included with the 15% of orders which have the highest net value.

(v) Estimate, to the nearest cent, the lowest net value of these orders which will qualify for a voucher.

....[4]

The gross value of an order is calculated by adding a tax of 20% to the net value, and **then** adding $5 for packing and delivery.

(vi) (a) If the voucher were to be given to the 15% of orders with the highest gross value, estimate, to the nearest cent, the minimum gross value which would qualify for a voucher.

....[2]

[Question 11 continues on the next page]

(b) The standard deviation of the net orders is $26.04.

Estimate, to the nearest cent, the standard deviation of the gross orders.

....[2]