Surname		Other Names	
Centre Number			Candidate Number									
Candidate Signature												

For Examiner's Use

General Certificate of Secondary Education
June 2007

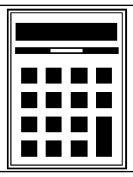
STATISTICS
Higher Tier

3311/H

H

Thursday 21 June 2007 9.00 am to 11.30 am

For this paper you must have:
- a calculator
- mathematical instruments.



Time allowed: 2 hours 30 minutes

Instructions

- Use blue or black ink or ball-point pen. Draw diagrams in pencil.
- Fill in the boxes at the top of this page.
- Answer **all** questions in the spaces provided.
- Do all rough work in this book.

Information

- The maximum mark for this paper is 120.
- Mark allocations are shown in brackets.
- Additional answer paper and graph paper will be issued on request and must be tagged securely to this answer book.
- You are expected to use a calculator where appropriate.

Advice

- In all calculations, show clearly how you work out your answer.

For Examiner's Use	
Pages	Mark
3	
4–5	
6–7	
8–9	
10–11	
12–13	
14–15	
16–17	
18–19	
20–21	
22–23	
24–25	
26–27	
28–29	
TOTAL	
Examiner's Initials	

You may need to use the following formulae:

Mean of a frequency distribution $= \dfrac{\sum fx}{\sum f}$

Mean of a grouped frequency distribution $= \dfrac{\sum fx}{\sum f}$, where x is the mid-interval value.

Standard deviation for a set of numbers $x_1, x_2, \ldots, x_n$ having a mean value of $\bar{x}$ is given by

$$\sqrt{\frac{\sum (x - \bar{x})^2}{n}} \text{ or } \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$$

Standard deviation for a frequency distribution

$$\sqrt{\frac{\sum f(x - \bar{x})^2}{\sum f}} \text{ or } \sqrt{\frac{\sum fx^2}{\sum f} - \bar{x}^2}$$

The same formula applies to the standard deviation of a grouped frequency distribution where x is the mid-interval value.

Spearman's rank correlation coefficient $= 1 - \dfrac{6\sum d^2}{n(n^2 - 1)}$

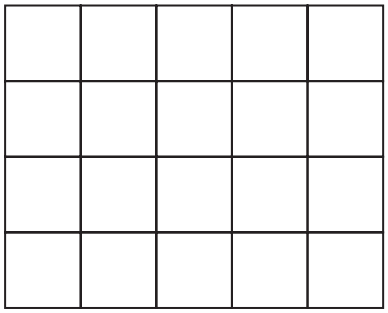
Answer **all** questions in the spaces provided.

1 The diagram shows a rectangular field divided into small square areas.

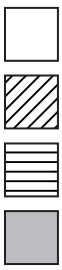
The number in each square shows the number of sheep in that square.

5	7	11	6	3
3	6	7	4	2
0	2	4	3	1
0	0	0	2	2

(a) Use the Key provided to produce a shading (choropleth) map on the blank copy of the field below.



Key:



0 sheep

1–4 sheep

5–8 sheep

9 or more sheep

(3 marks)

(b) A sheepdog is sitting in the field.
Mark with a 'D' the likely position of the sheepdog in the field.
Give a reason for your answer.

.....

.....

(1 mark)

4

2 Rodney is considering opening a small restaurant in the village where he lives.

To find out the views of local people he delivers a questionnaire to every house in the village.

(a) Included in the questionnaire is a closed question asking for people's age.

(i) Explain what is meant by a *closed question*.

.....

.....

(1 mark)

(ii) Give one advantage of using a closed question for age.

.....

.....

(1 mark)

(b) Only 12% of the questionnaires are returned to Rodney.

How might Rodney have improved the response rate?

.....

.....

(1 mark)

(c) The returned questionnaires showed that some of his questions had been badly worded.

What should Rodney have done before he delivered his questionnaire to avoid this problem?

.....

.....

(1 mark)

(d) One of Rodney's questions was

"How often do you eat out at a pub or restaurant?"

Give two criticisms of this question.

Criticism 1 .....

.....

(1 mark)

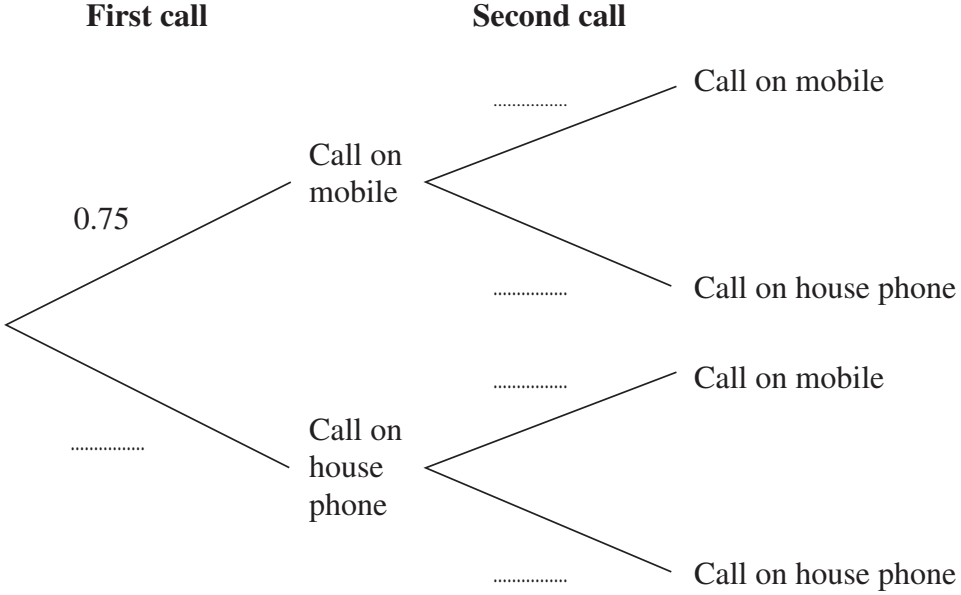
Criticism 2 .....

.....

(1 mark)

3 Danielle lives in a house which has a phone.
She also has a mobile phone.
Danielle receives 75% of all her calls on her mobile phone.

(a) (i) Complete the tree diagram to show the probabilities for the next two calls.
Assume that all calls are independent.



(3 marks)

(ii) Explain what is meant by the phrase in part (a)(i) that 'all calls are independent'.

.....

.....

(1 mark)

(b) Use the tree diagram to find the probability that the next two calls for Danielle are on her mobile phone.

.....

.....

Answer ..... *(2 marks)*

12

4 When motorists call a particular road breakdown service, they are put into one of three categories by the operator at the switchboard.

These categories are Emergency (E), Urgent(U) and Non-Urgent(N).

Emergency and Non-Urgent categories are equally likely.
The Urgent category is four times more likely than the Emergency category.

The breakdown service wish to carry out a simulation of 20 calls to their switchboard using a fair dice.

(a) Describe how they could allocate the numbers 1, 2, 3, 4, 5 and 6 on the dice to a particular category of call.

.....

.....

.....

.....

(3 marks)

(b) Using your answer to part (a) list the type of call simulated by the following numbers on the dice using the letters E, U or N.

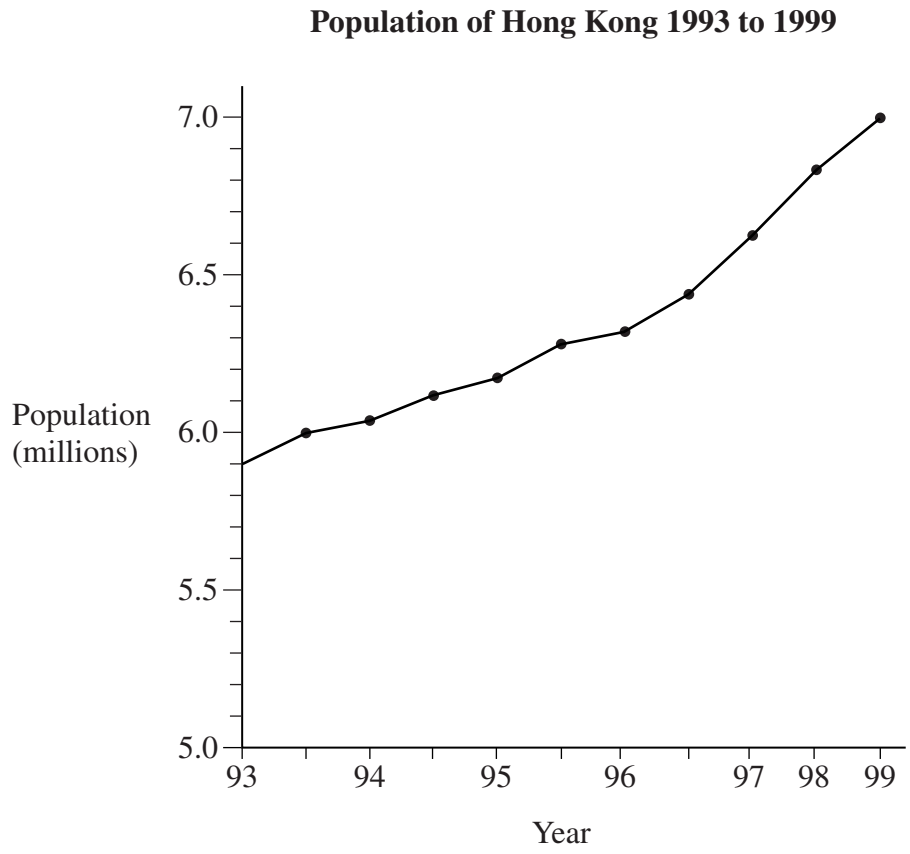
.....

.....

5	1	4	6	4	4	2	3	2	6
1	1	3	4	2	6	6	5	4	5

(2 marks)

5 (a) The diagram shows the population of Hong Kong from 1993 to 1999.



Give two reasons why this diagram is misleading.

Reason 1 .....

.....

.....

Reason 2 .....

.....

.....

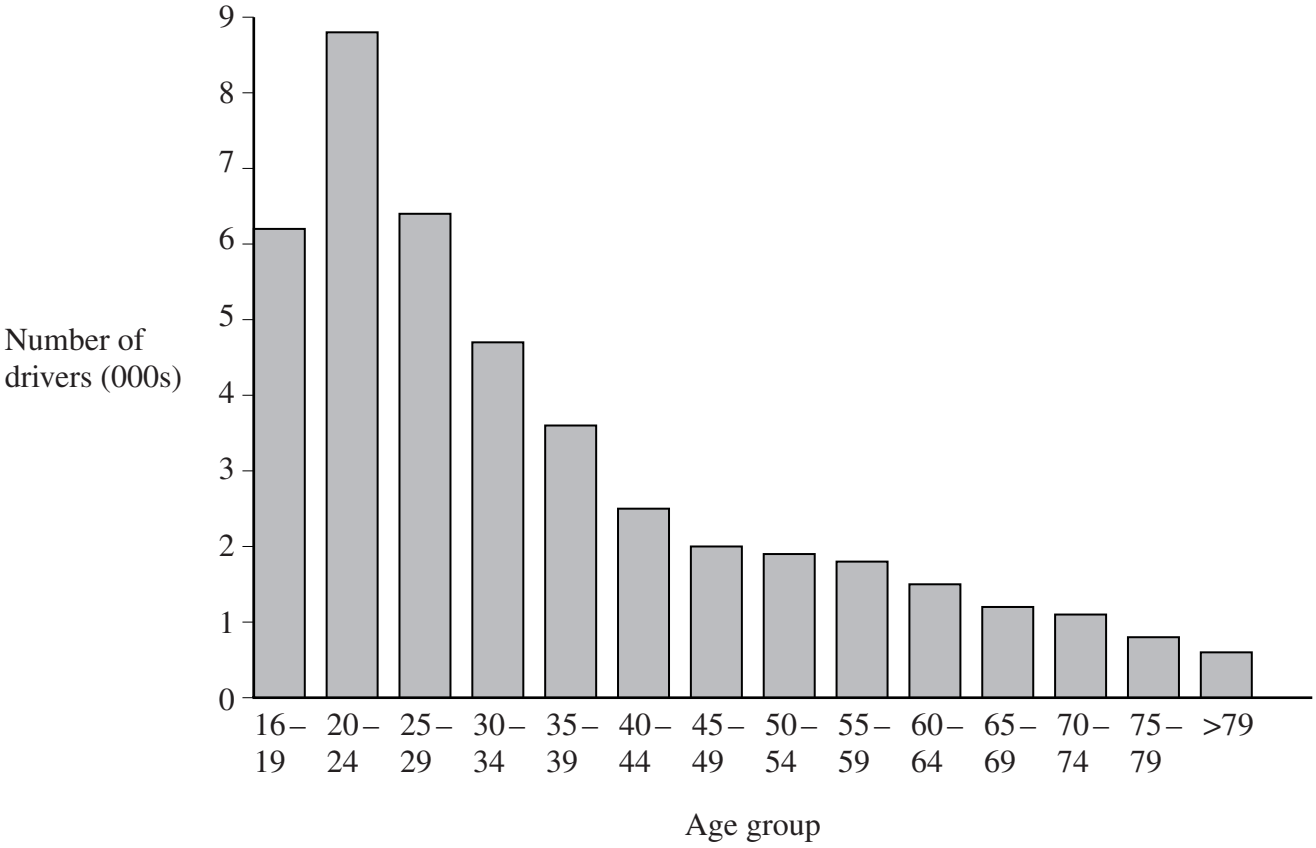
(2 marks)

Question 5 continues on the next page

7

(b) Explain why this graph does **not** show that drivers aged over 79 years are the safest on the roads.

Number of drivers in fatal crashes 1988



.....

.....

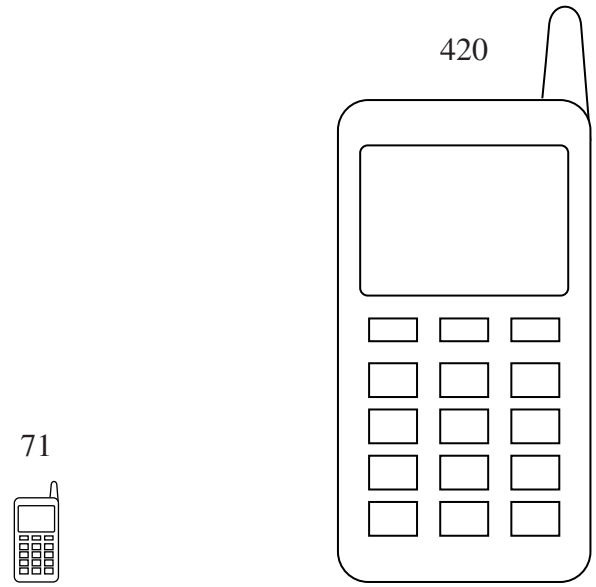
.....

.....

(2 marks)

(c) The diagram shows mobile phone ownership in Hong Kong in 1994 and 1998.

Mobile phones per 1000 population



Give one criticism of this diagram.

.....

.....

(1 mark)

Turn over for the next question

3

6 The tables gives the size of households in Great Britain between 1971 and 2001. For example, in 1991, 34% of households consisted of two people.

	1971	1981	1991	2001
One person	17	22	27	28
Two people	33	32	34	35
Three people	19	17	16	16
Four people	17	18	16	14
Five or more people	14	11	8	7

Source: *Adapted from Social Trends 2005*

(a) What percentage of households in 1981 consisted of four people?

Answer ..... % *(1 mark)*

(b) Throughout the period 1971 - 2001 what size household accounted for about a third of the households?

Answer ..... *(1 mark)*

(c) The total of the percentages for 1991 is 101%.

Give a possible reason for this.

.....

.....

(1 mark)

(d) The composite bar chart shows the data from the table for 1971 - 1991.

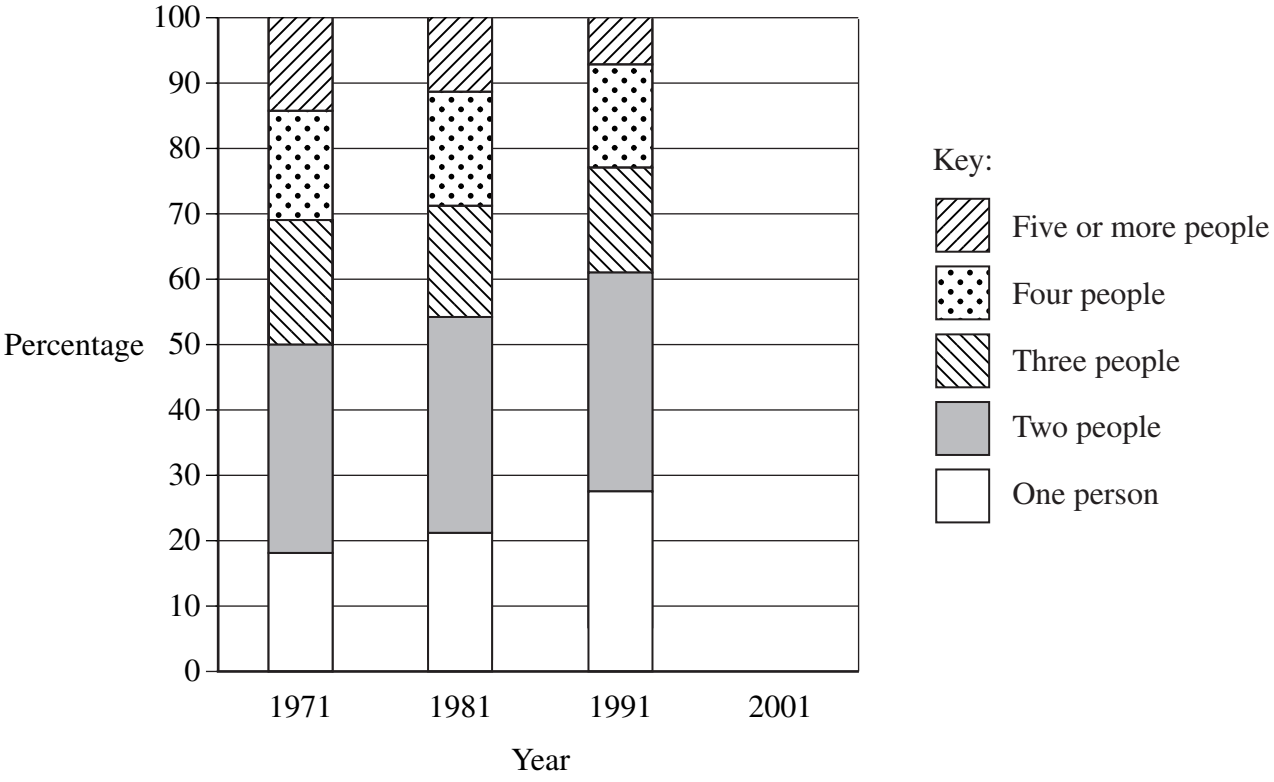
Complete the chart by drawing the bar for 2001.

.....

.....

(3 marks)

Composite bar chart for household size



(e) Use the composite bar chart to identify one similarity and one difference in the data for the year 1971 and the year 2001.

Similarity .....

.....

Difference .....

.....

(2 marks)

8

7 The table shows the number of visits to America by UK residents each quarter from 2002 to 2004. Some of the 4-point moving averages are shown in the table.

Year	Quarter	Visits (tens of thousands)	4-point moving average
2002	Q1	88	
	Q2	100	105
	Q3	118	107.5
	Q4	114	109
2003	Q1	98	110.5
	Q2	106	111.5
	Q3	124	
	Q4	118	
2004	Q1	102	
	Q2	116	
	Q3	134	
	Q4	126	

Source: *Adapted from Social Trends 2005*

(a) The data for number of visits each quarter for 2002 has been plotted on the time series graph on the grid opposite.

Complete the graph.

(2 marks)

(b) Describe two different patterns in the data.

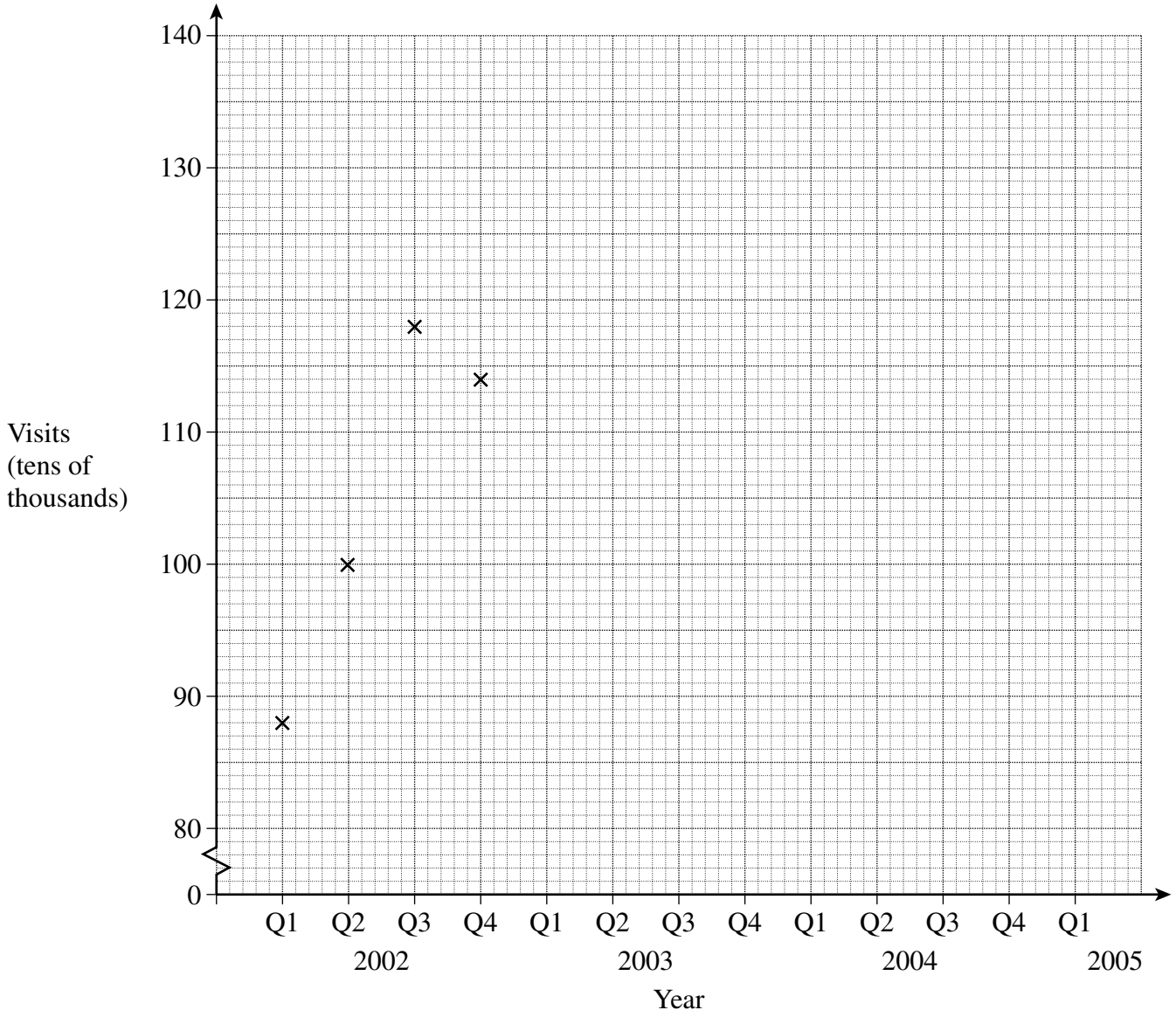
Pattern 1 .....

.....

Pattern 2 .....

.....

(2 marks)



Question 7 continues on the next page

4

(c) Calculate the remaining 4-point moving averages.

Write your answers in the table.

.....

.....

.....

.....

(4 marks)

(d) (i) Plot the moving averages on your graph. *(2 marks)*

(ii) Draw a trend line. *(1 mark)*

(e) Work out the value of the average seasonal effect for Q1, the first quarter.

.....

.....

.....

Answer ..... *(3 marks)*

(f) Use the trend line and your answer to part (e) to estimate the number of visits to America by UK residents in the first quarter of 2005.

.....

.....

.....

Answer ..... *(3 marks)*

LEAVE MARGIN BLANK

Turn over for the next question

13

8 A survey was carried out at eight retail stores on one morning.

Information was recorded about the number of employees and the percentage of employees absent.

The information is shown in the table.

Store	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
Number of employees	60	75	100	50	100	130	200	80
Percentage of employees absent	5	8	4	2	12	10	8.5	2.5

(a) Calculate the value of Spearman's rank correlation coefficient for the two sets of data.

.....

.....

.....

Answer ..... *(6 marks)*

(b) Use your answer to part (a) to comment on the statement

'The absentee rate increases with the size of workforce.'

.....

.....

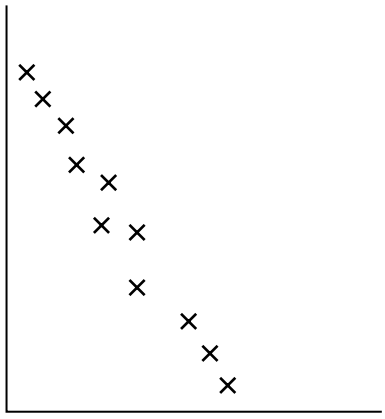
.....

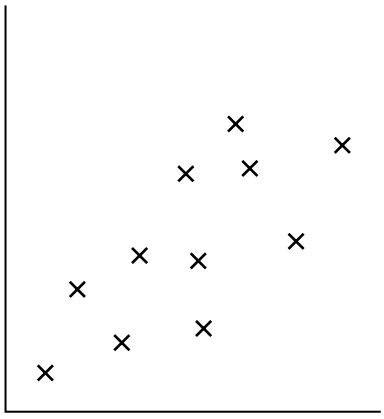
(1 mark)

(c) Write down the most likely value of Spearman's rank correlation coefficient for each of the following graphs.

Select from the following list:

–0.45 0.75 –1.2 –0.98 0.96





Answer ..... Answer .....

(2 marks)

Turn over for the next question

9

9 The table shows the traffic using the Channel Tunnel during the period 1994 – 2000.

Channel Tunnel Traffic (Thousands)							
Year	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**
Passenger vehicles	82	1246	2135	2383	2864	3448	3342
Passengers on trains	315	7018	12749	14613	18496	17424	17152
Freight vehicles	65	391	519	268	705	1133	939

Source: *Adapted from Monthly Digest of Statistics 2004*

(a) Give one similarity and one difference between the annual number of passenger vehicles and freight vehicles over the seven year period.

Similarity .....

.....

.....

.....

Difference .....

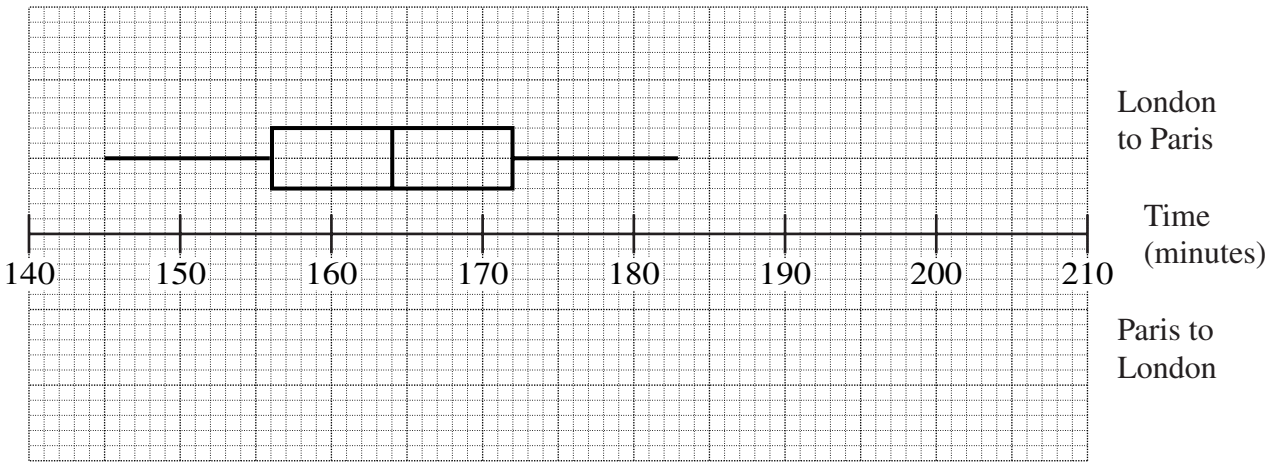
.....

.....

.....

(2 marks)

(b) The times taken, in minutes, by a sample of passenger trains to travel from London to Paris over a period of one month are summarised in the box and whisker diagram.



The times taken, in minutes, by a sample of passenger trains to travel from Paris to London over the same period of one month are summarised in the table.

Minimum	Lower quartile	Median	Upper quartile	Maximum
148	150	155	195	202

(i) Use these values to draw a box and whisker diagram on the same grid.

(3 marks)

(ii) Describe the shape of the two distributions.

London to Paris .....

.....

Paris to London .....

.....

(2 marks)

(iii) State the least journey time beyond which any **large** outliers could occur for the Paris to London distribution.

.....

.....

.....

Answer ..... minutes to ..... minutes *(3 marks)*

10

10 (a) A bank kept a daily record of the number of cheques with errors that were presented for payment.

The results from a random sample of 150 days are shown in the table.

Number of cheques with errors	Number of days
0	4
1	21
2	38
3	38
4	22
5	15
6	8
7	4

Calculate the mean, variance and standard deviation of the number of cheques per day with errors.

.....

.....

.....

.....

.....

.....

Answer Mean .....

Variance .....

Standard deviation ..... *(5 marks)*

(b) Another bank recorded the times taken to process an equal number of cheque errors at two of its branches, A and B.

The times taken in minutes, at each of the two branches were normally distributed with mean and standard deviation shown in the table.

	Mean (minutes)	Standard deviation (minutes)
Branch A	16.5	2.8
Branch B	14	4.5

To compare the performance of both branches it was agreed to standardise the times taken at each branch.

(i) What would be the standardised value for a cheque from Branch A taking 21 minutes to process?

.....

.....

.....

Answer ..... *(2 marks)*

(ii) A cheque processed at Branch B had a standardised time of 2.4

What was the actual processing time?

.....

.....

.....

Answer ..... minutes *(3 marks)*

(iii) Between what limits would you expect approximately 99.9% of the cheque processing time for Branch A to lie?

.....

.....

.....

Answer ..... minutes *(3 marks)*

13

11 The table shows the number of GCSE passes for 30 Year 11 students.

The gender of the students is also shown.

Student	Gender	Number of passes	Student	Gender	Number of passes
01	M	4	16	M	3
02	M	7	17	M	5
03	F	9	18	F	10
04	M	8	19	M	2
05	F	6	20	M	4
06	F	6	21	F	8
07	M	9	22	M	3
08	M	2	23	F	7
09	F	7	24	M	2
10	M	4	25	M	1
11	F	8	26	M	4
12	M	7	27	F	9
13	F	5	28	M	5
14	M	5	29	M	6
15	M	4	30	M	2

Here is a table of random numbers from 01 to 50.

Line 1	15	41	01	15	20	16
Line 2	32	22	33	30	19	08
Line 3	04	31	49	29	13	29
Line 4	14	23	37	11	24	29

(a) (i) Starting with the first number on Line 1 of the random number table and reading across from left to right, select a random sample of size eight.

Write your answers in the table.

Student number								
Gender								
Number of passes								

(3 marks)

(ii) Calculate the mean number of passes per student for this sample.

.....

Answer ..... *(1 mark)*

(b) (i) Starting with the first number on Line 3 of the random number table and reading from left to right, select a random sample of size six stratified by gender.

Write your answers in the table.

Student number						
Gender						
Number of passes						

.....

.....

(4 marks)

(ii) Calculate the mean number of passes per student for this sample.

.....

Answer ..... *(1 mark)*

Question 11 continues on page 24

9

(c) Give a reason why the sampling method used in part (b)(i) is better than the sampling method used in part (a)(i).

.....

.....

.....

(1 mark)

(d) After the examination period the students were given a questionnaire to complete.

One of the questions contained the following statement.

‘Exams are getting easier’

…… ☐

…… ☐

…… ☐

…… ☐

…… ☐

(please tick a box)

Complete the response box headings.

(2 marks)

Turn over for the next question

3

12 The table shows the breakdown of the Retail Price Index at July 2004.

	Group	Weight	Index (1987 = 100)
1	Food	111	152.9
2	Catering	49	231.8
3	Leisure goods	46	99.1
4	Leisure services	70	251.3
5	Housing	209	263.8
6	Fuel and light	28	139.4
7	Household goods	71	145.3
8	Household services	59	180.4
9	Clothing and footwear	51	99.0
10	Personal goods and services	42	200.0
11	Motoring expenditure	146	184.5
12	Fares and other travel costs	21	215.6
13	Alcoholic drink	68	204.3
14	Tobacco	29	313.7

(Based on the Retail Price Index, Monthly Digest of Statistics, July 2004)

The 'All groups' index for July 2004 is 198.4

(a) (i) Calculate an 'All groups' index, **excluding** expenditure on both Alcoholic drink and Tobacco.

Give your answer to one decimal place.

.....

.....

.....

.....

.....

Answer ..... *(4 marks)*

(ii) Explain why the 'All groups' index has decreased.

.....

.....

.....

(2 marks)

(b) The table shows the annual cost of Rashid's car insurance for the past five years.

Year	2002	2003	2004	2005	2006
Annual cost (£)	500	546	670	640	625

(i) Use the chain base method to calculate index numbers for the years 2003 to 2006 inclusive.

Give your answers to one decimal place.

.....

.....

.....

.....

Answer 2003 .....

2004 .....

2005 .....

2006 ..... *(4 marks)*

(ii) Describe what these chain base index numbers show.

.....

.....

.....

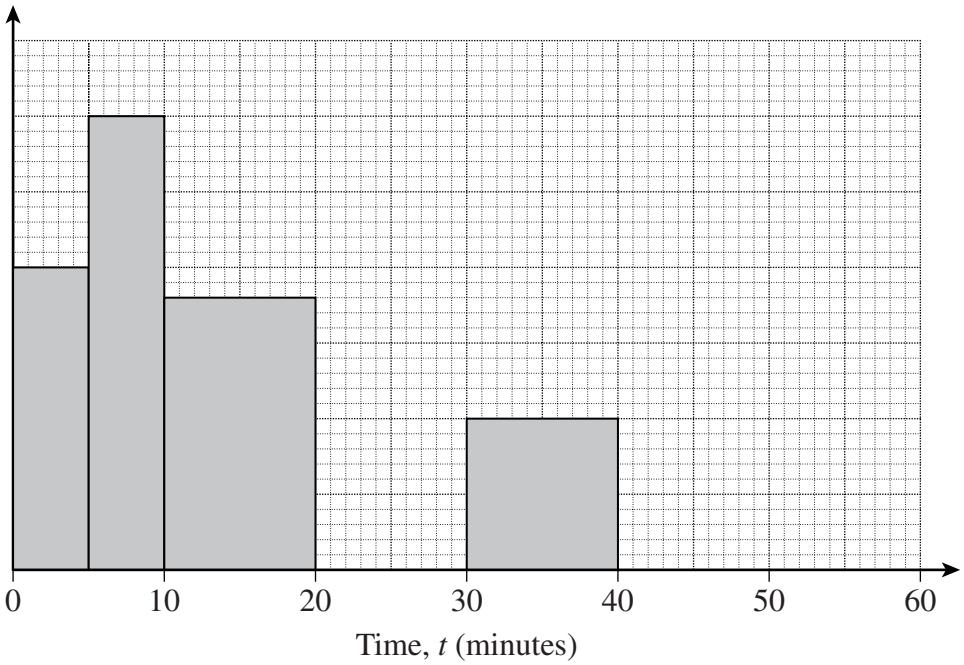
.....

(2 marks)

12

13 The incomplete table and histogram show, some of the times taken by a number of students to complete a mathematical puzzle.

Time, t (minutes)	**Frequency**
$0 < t \leqslant 5$	20
$5 < t \leqslant 10$	
$10 < t \leqslant 20$	
$20 < t \leqslant 30$	16
$30 < t \leqslant 40$	
$40 < t \leqslant 60$	8



(a) Use the information in the table to complete the histogram.

.....

.....

.....

(3 marks)

(b) Use the information in the histogram to complete the table.

.....

.....

.....

(2 marks)

(c) State the modal class.

Answer ..... *(1 mark)*

(d) Two students are selected at random.

Calculate the probability that one student completes the puzzle in five minutes or less and the other student completes the puzzle in more than 40 minutes.

.....

.....

.....

.....

Answer ..... *(4 marks)*

(e) Calculate the probability that four students selected at random, each take 10 minutes or less to complete the puzzle.

.....

.....

.....

.....

Answer ..... *(3 marks)*

END OF QUESTIONS

There are no questions printed on this page

There are no questions printed on this page

There are no questions printed on this page