

Centre No.						Paper Reference						Surname	Initial(s)	
Candidate No.						1	3	8	9	/	1	H	Signature	

Paper Reference(s)

**1389/1H**

**Edexcel GCSE**

**Statistics**

Paper 1H

**Higher Tier**

Thursday 21 June 2007 – Morning

Time: 2 hours 30 minutes

Examiner's use only

--	--	--

Team Leader's use only

--	--	--

**Materials required for examination**

Ruler graduated in centimetres and millimetres, protractor, compasses, pen, HB pencil, eraser, electronic calculator.

**Items included with question papers**

Nil

**Instructions to Candidates**

In the boxes above, write your centre number, candidate number, your surname, initials and signature. Check that you have the correct question paper.

Answer ALL the questions. Write your answers in the spaces provided in this question paper.

**You must NOT write on the formulae page or any blank pages. Anything you write on these pages will gain NO credit.**

If you need more space to complete your answer to any question, use additional answer sheets.

**Information for Candidates**

The marks for individual questions and the parts of questions are shown in round brackets: e.g. (2).

This question paper has 8 questions in Section A and 6 questions in Section B. The total mark for this paper is 100.

There are 24 pages in this question paper. Any blank pages are indicated.

**Advice to Candidates**

Show all stages in any calculations.

Work steadily through the paper. Do not spend too long on one question.

If you cannot answer a question, leave it and attempt the next one.

Return at the end to those you have left out.

This publication may be reproduced only in accordance with Edexcel Limited copyright policy. ©2007 Edexcel Limited.

Printer's Log. No.  
**N25839A**

W850/R1389/57570 6/6/6/4/3/



*Turn over*

**edexcel**   
advancing learning, changing lives

**GCSE Statistics 1389**

Higher Tier Formulae

**You must not write on this page.  
Anything you write on this page will gain NO credit.**

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

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

Variance  $= \frac{\sum (x - \bar{x})^2}{n}$

Standard deviation (set of numbers)  $\sqrt{\left[ \frac{\sum x^2}{n} - \left( \frac{\sum x}{n} \right)^2 \right]}$

or  $\sqrt{\left[ \frac{\sum (x - \bar{x})^2}{n} \right]}$

where  $\bar{x}$  is the mean set of values.

Standard deviation (discrete frequency distribution)  $\sqrt{\left[ \frac{\sum fx^2}{\sum f} - \left( \frac{\sum fx}{\sum f} \right)^2 \right]}$

or  $\sqrt{\left[ \frac{\sum f(x - \bar{x})^2}{\sum f} \right]}$

Spearman's Rank Correlation Coefficient  $1 - \frac{6 \sum d^2}{n(n^2 - 1)}$



**SECTION A**

**Answer ALL the questions. Write your answers in the spaces provided.**

**You must write down all stages in your working.**

1. The table shows information about the number of energy units supplied in Great Britain.

**Number of energy units and their source**

Year	Total	Coal	Petroleum	Natural gas	Nuclear	Natural flow hydro	Net imports
1995	223.5	51.4	76.6	72.3	21.28	0.47	1.40
1996	227.0	46.9	75.6	80.6	22.10	0.32	1.44
1997	229.2	43.3	74.7	87.3	22.02	0.41	1.43
1998	235.2	43.3	76.5	90.4	23.39	0.52	1.07
1999	235.7	38.1	77.7	95.8	22.24	0.54	1.22
2000	238.0	40.0	77.8	98.8	19.66	0.51	1.22
2001	238.1	43.0	76.0	96.9	20.84	0.44	0.89
2002	234.2	39.8	73.5	99.5	20.09	0.56	0.72
2003	237.3	42.3	75.7	98.2	20.49	0.40	0.19

(Data source: Department of Trade and Industry)

- (a) How many units of energy did Nuclear supply in 1998?

.....units  
**(1)**

- (b) Which source supplied the greatest number of energy units

(i) in 1995,

.....

(ii) in 2003?

.....

**(2)**

- (c) Describe the trend in the Net imports of energy units between the years 2000 and 2003.

.....

.....

**(1)**

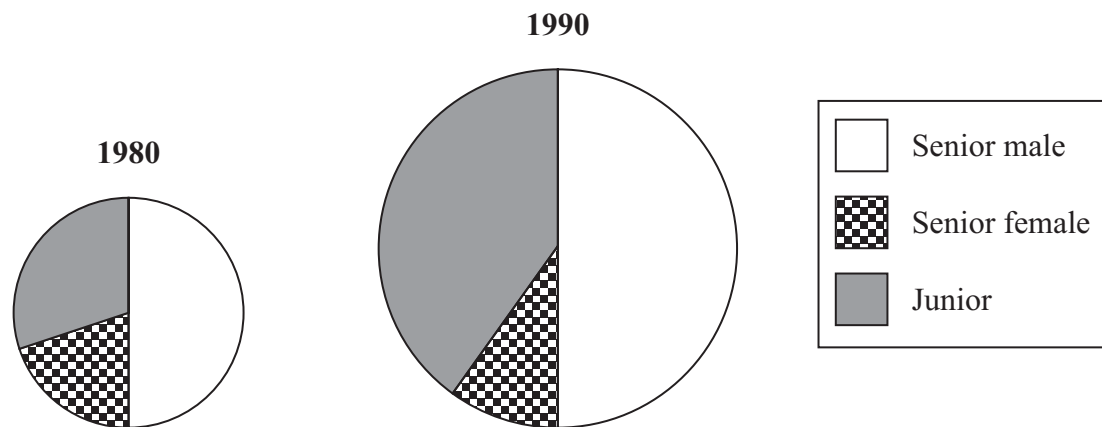
**(Total 4 marks)**

**Q1**



2. The comparative pie charts show some information about the players at Seaton squash club in 1980 and in 1990.

The three types of players at Seaton squash club are Senior male, Senior female and Junior.



(Data source: *Seaton squash club*)

(a) What has happened to the number of Senior male players at Seaton squash club between 1980 and 1990? Give a reason for your answer.

.....  
 .....

(2)

The table shows the numbers of players in 2005.

	Number of players
Senior male	197
Senior female	108
Junior	81

A stratified sample of 40 players was taken from the 386 players of the club.

(b) Explain why a **stratified** sample was taken.

.....  
 .....  
 .....

(1)



<p>(c) Work out the number of Junior players who were selected for the stratified sample.</p> <p>.....</p> <p style="text-align: right;"><b>(1)</b></p> <p style="text-align: right;"><b>(Total 4 marks)</b></p>	<p>Leave blank</p> <p><b>Q2</b></p> <input type="text"/>
<p><b>3.</b> Mary wants to estimate the number of fish in a lake. She catches 40 fish, marks them and puts them back in the lake. Later she catches another 40 fish and finds that 5 of them are marked.</p> <p>(a) Work out an estimate for the number of fish in the lake.</p> <p>.....</p> <p style="text-align: right;"><b>(2)</b></p> <p>(b) Write down <b>one</b> assumption you have made about the population of fish in the lake.</p> <p>.....</p> <p>.....</p> <p style="text-align: right;"><b>(1)</b></p> <p style="text-align: right;"><b>(Total 3 marks)</b></p>	<p><b>Q3</b></p> <input type="text"/>



Leave blank

4. Storm Engineering employs 200 workers.

Storm Engineering wish to consult their workers about a new pension scheme.

(a) Give **one** advantage of taking a census of the workers rather than a sample.

.....  
.....  
(1)

Storm Engineering will use a questionnaire.

They will use closed questions.

(b) Give **one** advantage of using closed questions on a questionnaire.

.....  
.....  
(1)

Before Storm Engineering gives the questionnaire to all their workers they are advised to do something to check that the questionnaire is suitable.

(c) Write down what Storm Engineering need to do.

.....  
.....  
(1)

One question suggested for the questionnaire was

‘Do you agree that the proposed pension scheme should be compulsory?’

This question is leading.

(d) Rewrite this question so that it is not leading. You should include response boxes.

.....  
.....  
.....  
.....  
.....  
(2)

(Total 5 marks)

Q4



Leave blank

5. The Gross Domestic Product per person, or GDP per capita, is a measure of a country's wealth.  
The greater the GDP per capita, the greater the country's wealth.  
The table shows the GDP per capita and the life expectancy at birth for each of nine countries.

Country	GDP per capita (\$)	Life expectancy at birth (years)				
Luxembourg	58 198	78.7				
UK	29 483	78.4				
Monaco	26 844	79.6				
Uruguay	14 423	76.1				
Seychelles	7711	71.8				
Grenada	4916	64.5				
St. Helena	2413	77.8				
Haiti	1484	52.9				
Comoros	657	62.0				

(Data source: [www.nationmaster.com](http://www.nationmaster.com))

- (a) Work out Spearman's rank correlation coefficient for these data.

.....  
(3)

- (b) Interpret your answer to part (a).

.....  
.....  
.....  
(2)

(Total 5 marks)

Q5



Leave blank

6. Waterside garage is doing a survey on the colours of vans.

An employee is going to count how many vans of each colour pass the garage.

(a) Use the best word from the list to complete each sentence below.

**qualitative    quantitative    continuous    secondary**

(i) The colours of the vans are ..... data.

(ii) The number of vans is ..... data. **(2)**

The manager has three suggestions for when the employee should do the survey.

A: between 8.00 am and 10.00 am on a weekday

B: between 10.00 am and midday on a Saturday

C: at a randomly selected 1 hour period between 8.00 am and 5.00 pm every day for a week

(b) Which suggestion is best? Give a reason for your answer.

.....  
.....  
.....

**(1)**

The manager has three suggestions for how the data could be recorded.

X: to write down the colour of each van as it passes, e.g. red, white, blue

Y: to write down just the first letter of the colour e.g. R, W, B

Z: to fill in a tally chart

(c) Which suggestion is best? Give a reason for your answer.

.....  
.....  
.....

**(1)**

**Q6**

**(Total 4 marks)**





Leave  
blank

7. The table gives the mean and standard deviation of the times, in seconds, for each of two races in the Olympic Games of 2004. The times in each of these races are normally distributed.

Race	Mean (seconds)	Standard deviation (seconds)
100 m	10.95	0.238
400 m	49.62	1.290

(Data source: *www.athens2004.com*)

Roman Sebrle ran a time of 10.85 seconds in the 100 m race and a time of 48.36 seconds in the 400 m race.

- (a) Calculate Roman Sebrle's standardised score for his time in each of these two races.

Standardised 100 m score .....

Standardised 400 m score .....

(3)

- (b) Interpret your answers to part (a).

.....  
.....  
.....  
.....

(2)

(Total 5 marks)

Q7



Leave  
blank

8. A machine fills packets with sugar.  
Every hour Peter takes a sample of 10 packets.

(a) Write down the name given to this method of sampling.

.....  
(1)

Peter calculates the mean weight for each sample of 10 packets.  
These mean weights are normally distributed with a mean of 505 g and a standard deviation  
of 1.6 g.

(b) Write down the percentage of the samples that have a mean weight  
within  $\pm 2$  standard deviations of 505 g.

..... %  
(1)

The allowable limits for the mean weights of the samples are  $505 \text{ g} \pm 3$  standard deviations.

Peter takes a sample, it has a mean weight of 500 g.

(c) What should Peter do? Show your working.

.....  
.....  
(3)

(Total 5 marks)

Q8

TOTAL FOR SECTION A: 35 MARKS



**BLANK PAGE**



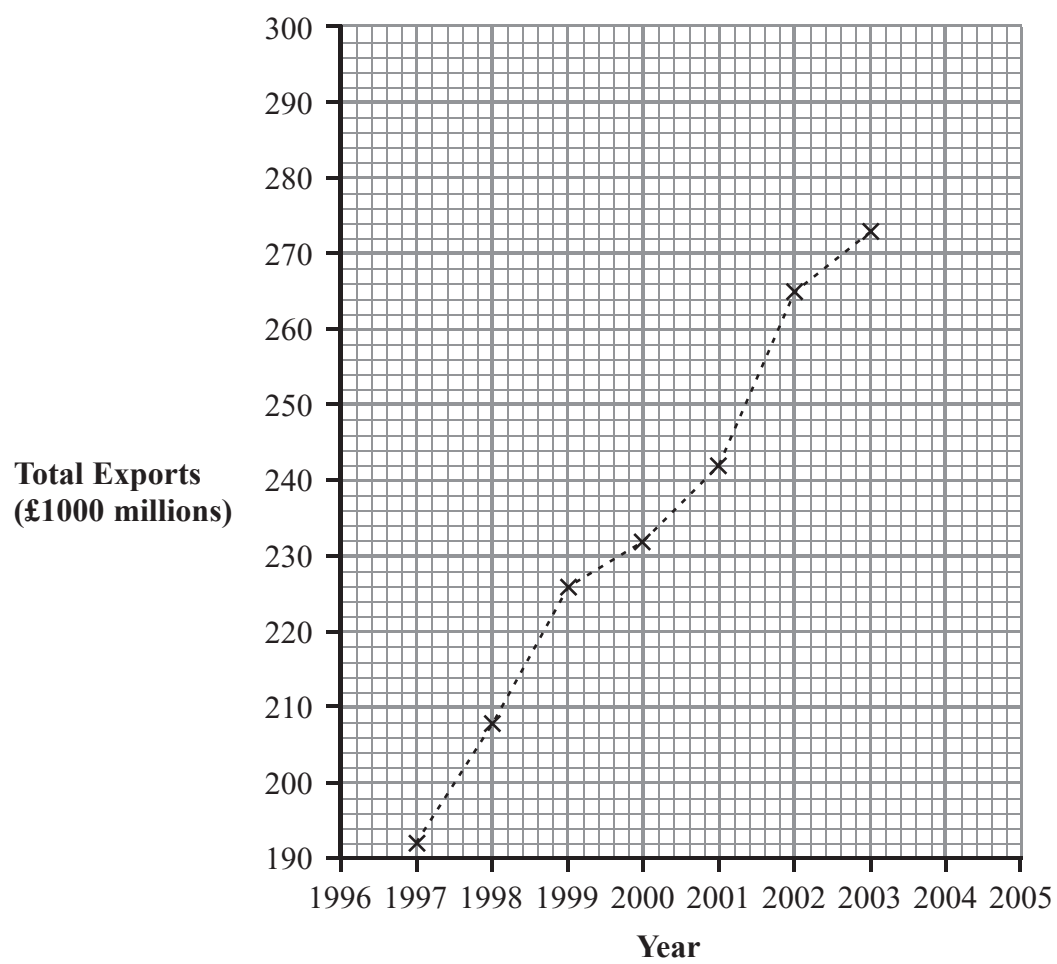
**SECTION B**

**Answer ALL the questions. Write your answers in the spaces provided.**

**You must write down all stages in your working.**

1. The time series graph shows the values, to the nearest £1000 million, of the Total Exports from the United Kingdom between 1997 and 2003.

**Total United Kingdom Exports**



(Data source: Office for National Statistics)

- (a) Draw a trend line on the time series graph. **(1)**

- (b) What does the trend line show about the Total United Kingdom Exports between 1997 and 2003?

.....  
 .....

**(1)**

- (c) Use your trend line to predict the Total United Kingdom Exports in 2004.

.....

**(1)**



Leave  
blank

(d) Why might this prediction for the Total United Kingdom Exports in 2004 be unreliable?

.....  
.....

**(1)**

In 2005 the Total United Kingdom Exports was 273 thousand million pounds.

(e) Does this figure follow the overall trend shown by the trend line on the time series graph?

Give a reason for your answer.

.....  
.....  
.....

**(2)**

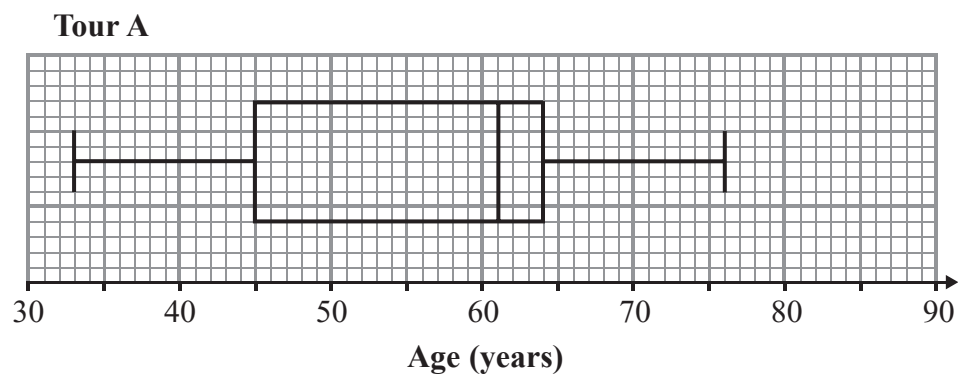
**Q1**

**(Total 6 marks)**



2. A holiday company recorded the ages, in years, of the people on each of two tours.

The ages of the people on tour A are summarised in this box plot.



(Data source: *Blue Lagoon*)

(a) Describe the skewness of the distribution of the ages of the people on tour A.

.....  
(1)

(b) What percentage of these ages lie between the upper and lower quartiles?

.....  
(1)

Here are the ages of the people on tour B.

32	33	35	39	43	44	44	47	48	48	50	51
52	54	55	55	57	58	60	65	68	70	86	

(c) For the ages of the people on tour B, find

(i) the median,

.....

(ii) the lower quartile,

.....

(iii) the upper quartile.

.....

(3)



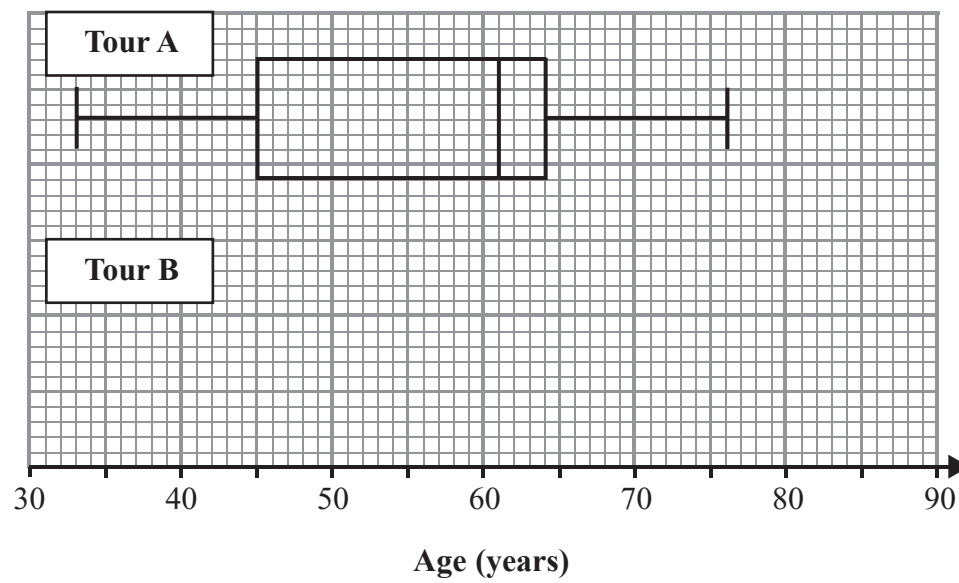
Leave blank

(d) Show that 86 is an outlier for the ages of the people on tour B.

(3)

There are no other outliers.

(e) On the grid, draw a box plot to show the distribution of the ages of the people on tour B.



(3)

(f) Compare the distributions of the ages of the people on the two tours.

.....

.....

.....

.....

.....

(2)

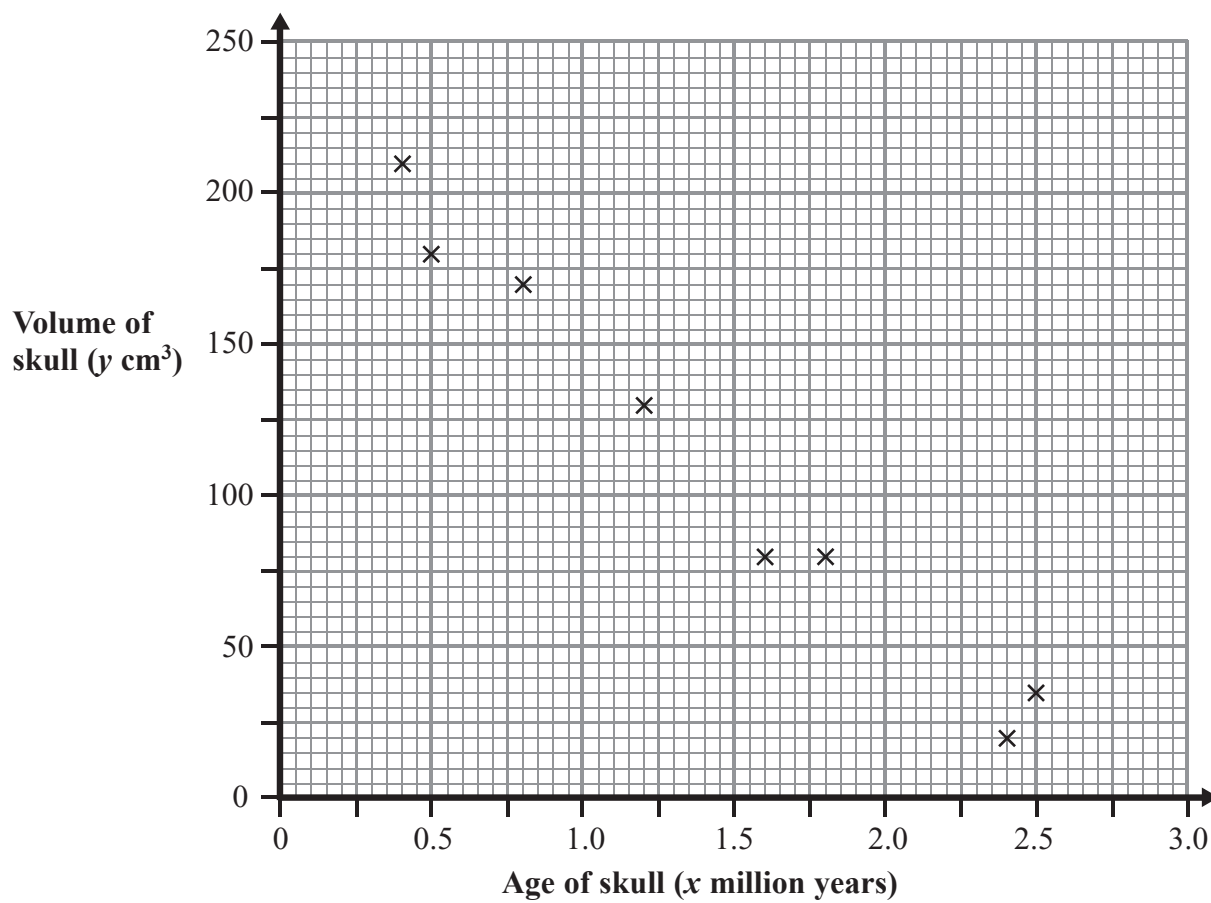
(Total 13 marks)

Q2



Leave blank

3. The scatter diagram shows the ages,  $x$  million years, and the volumes,  $y$  cm<sup>3</sup>, of eight skulls of a particular type of ape.



- (a) Describe the correlation between the volume of a skull and the age of a skull for this type of ape.

..... (1)

The table gives the ages,  $x$  million years, and the volumes,  $y$  cm<sup>3</sup>, of the eight skulls shown in the scatter diagram.

$x$	2.5	1.8	0.8	2.4	1.6	0.5	0.4	1.2
$y$	34	80	170	20	80	180	210	130

- (b) Calculate the coordinates of the mean point  $(\bar{x}, \bar{y})$  for these data.

( ..... , ..... ) (2)





Leave blank

- (c) On the scatter diagram
- (i) plot the point  $(\bar{x}, \bar{y})$ ,
  - (ii) draw the line of best fit. (2)

A skull of this particular type of ape has an age of one million years.

- (d) Find an estimate for the volume of this skull.
- .....cm<sup>3</sup>  
(1)

The skull of another ape is to be classified. It has an age of 1 million years and a volume of 75 cm<sup>3</sup>.

- (e) Discuss whether this skull is likely to be from the same type of ape.
- .....  
.....  
.....  
.....  
(2)

- (f) Give a reason why your line of best fit may not be used reliably to predict the volume of a skull with an age of 3 million years.
- .....  
.....  
(1)

The equation of the line of best fit has the form  $y = ax + b$ .

- (g) Use your line of best fit to find the value of  $a$  and the value of  $b$ .
- $a =$  .....  
 $b =$  .....  
(3)

- (h) Give a practical interpretation of the meaning of  $a$ .
- .....  
.....  
(2)

(Total 14 marks)

Q3

--	--



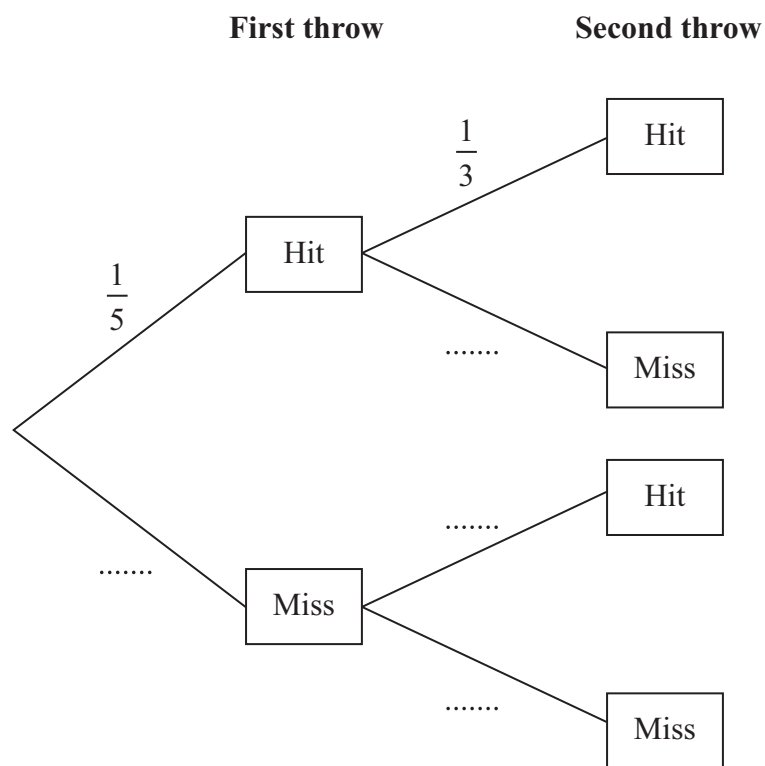
4. Michael throws a stone at a target. He then throws a second stone at the target.

The probability he hits the target on his first throw is  $\frac{1}{5}$

If he hits the target on his first throw, the probability he will hit the target on his second throw is  $\frac{1}{3}$

If he misses the target on his first throw, the probability he will hit the target on his second throw is  $\frac{1}{4}$

(a) Complete the probability tree diagram.



(3)

(b) Work out the probability that Michael will hit the target on both throws.

.....  
(2)

(c) Work out the probability that Michael will hit the target only once in the two throws.

.....  
(3)



Leave  
blank

Gordon is going to throw five stones in turn at the target.  
Gordon can hit the target with a probability of 0.8 with any one of these five stones.

- (d) Name the probability distribution that models the number of times he will hit the target in the five throws.

.....  
(1)

- (e) Work out the probability that he will hit the target with only one of the five stones.  
[You may use  $(p + q)^5 = p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$ ]

.....  
(2)

- (f) Work out the most likely number of times he will hit the target.  
[You may use  $(p + q)^5 = p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$ ]

.....  
(2)

(Total 13 marks)

Q4

--	--

19

Turn over



N 2 5 8 3 9 A 0 1 9 2 4

Leave  
blank

5. The manager of a supermarket records the times,  $t$  minutes, to serve 200 customers.

The results are summarised in the frequency table.

Time $t$ (minutes)	Frequency
$0 < t \leq 3$	9
$3 < t \leq 5$	47
$5 < t \leq 6$	38
$6 < t \leq 8$	60
$8 < t \leq 12$	46

(Data source: *Supa store Dec 2004*)

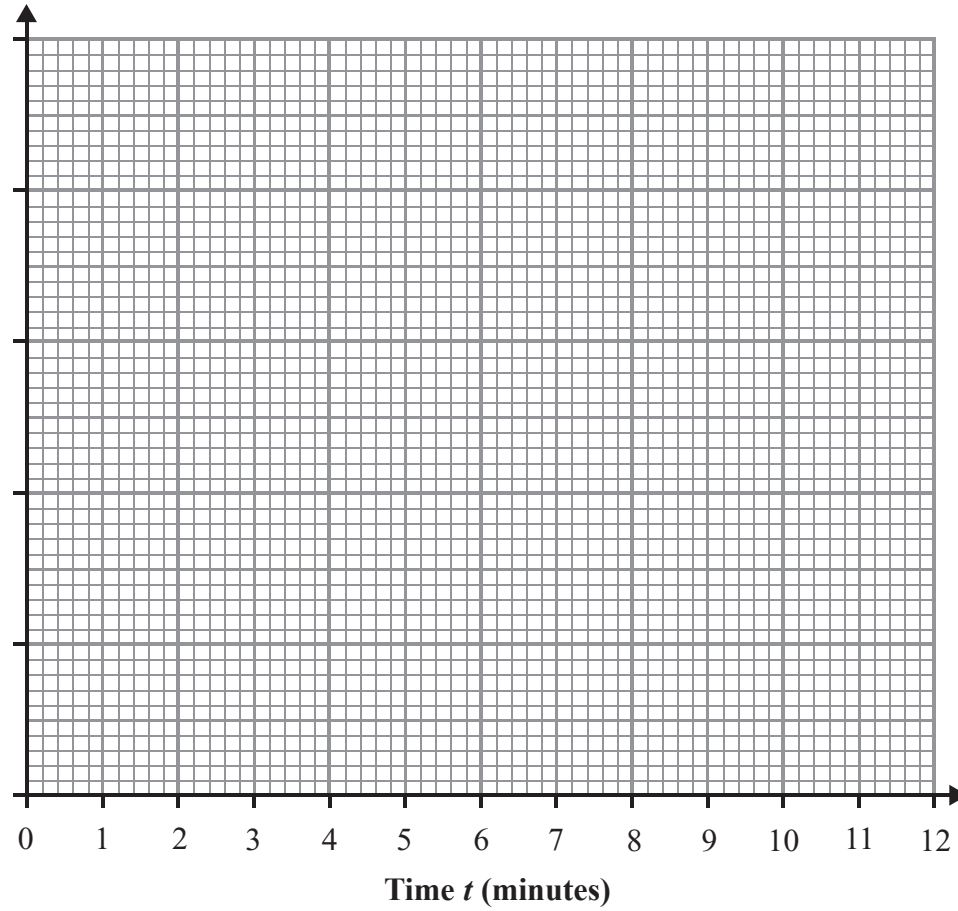
(a) Calculate an estimate for the mean time to serve these customers.

..... minutes  
**(3)**



Leave blank

(b) Draw a histogram for this information.



(3)

(c) Calculate an estimate for the median time to serve these customers.

..... minutes  
(2)

It is claimed that these times are normally distributed.

(d) Discuss whether the claim is true or false.

.....  
.....  
.....

(2)  
(Total 10 marks)

Q5

21

Turn over



N 2 5 8 3 9 A 0 2 1 2 4

6. Nikki uses old editions of her local newspaper to collect information about the average price of the houses in her area.

Her results for the years 2001 to 2005 are summarised in the table.

Year	2001	2002	2003	2004	2005
Average price (£)	120 000	128 000	137 000	145 500	156 500

(Data source: *Western Gazette*)

Nikki is using secondary data.

- (a) Write down **one** advantage and **one** disadvantage of using secondary data.

Advantage .....

.....

Disadvantage .....

.....

(2)

Nikki calculates the chain base index numbers for the average price of these houses for the years 2002, 2003, 2004 and 2005.

Her chain base index number for 2002 is shown in the table below.

Year	2002	2003	2004	2005
Chain base index number	106.7			

- (b) Work out the chain base index numbers for 2003, 2004 and 2005 and complete the table.

(3)



(c) Work out the geometric mean of the chain base index numbers for 2002 to 2005.

Leave  
blank

.....  
(2)

(d) Interpret your answer to part (c).

.....  
.....  
.....  
(2)

Q6

(Total 9 marks)

**TOTAL FOR SECTION B: 65 MARKS**  
**TOTAL FOR PAPER: 100 MARKS**

**END**



**BLANK PAGE**

