

Write your name here

Surname					Other names				
Centre Number					Candidate Number				
<div style="border: 1px solid black; width: 100%; height: 100%; display: flex; justify-content: space-around;"> <div style="width: 20%; height: 20px;"></div> <div style="width: 20%; height: 20px;"></div> <div style="width: 20%; height: 20px;"></div> <div style="width: 20%; height: 20px;"></div> <div style="width: 20%; height: 20px;"></div> </div>					<div style="border: 1px solid black; width: 100%; height: 100%; display: flex; justify-content: space-around;"> <div style="width: 20%; height: 20px;"></div> <div style="width: 20%; height: 20px;"></div> <div style="width: 20%; height: 20px;"></div> <div style="width: 20%; height: 20px;"></div> <div style="width: 20%; height: 20px;"></div> </div>				

Edexcel GCSE

Statistics
Paper 1H

SOLUTIONS

Higher Tier

Monday 27 June 2011 – Morning Time: 2 hours	Paper Reference 5ST1H/01
---	------------------------------------

You must have:
Ruler graduated in centimetres and millimetres, protractor, pen
HB pencil, eraser, electronic calculator.

Total Marks

Instructions

- Use **black** ink or ball-point pen.
- **Fill in the boxes** at the top of this page with your name, centre number and candidate number.
- Answer **all** questions.
- Answer the questions in the spaces provided
– *there may be more space than you need.*

Information

- The total mark for this paper is 100.
- The marks for **each** question are shown in brackets
– *use this as a guide as to how much time to spend on each question.*
- Questions labelled with an **asterisk** (*) are ones where the quality of your written communication will be assessed
– *you should take particular care on these questions with your spelling, punctuation and grammar, as well as the clarity of expression.*

Advice

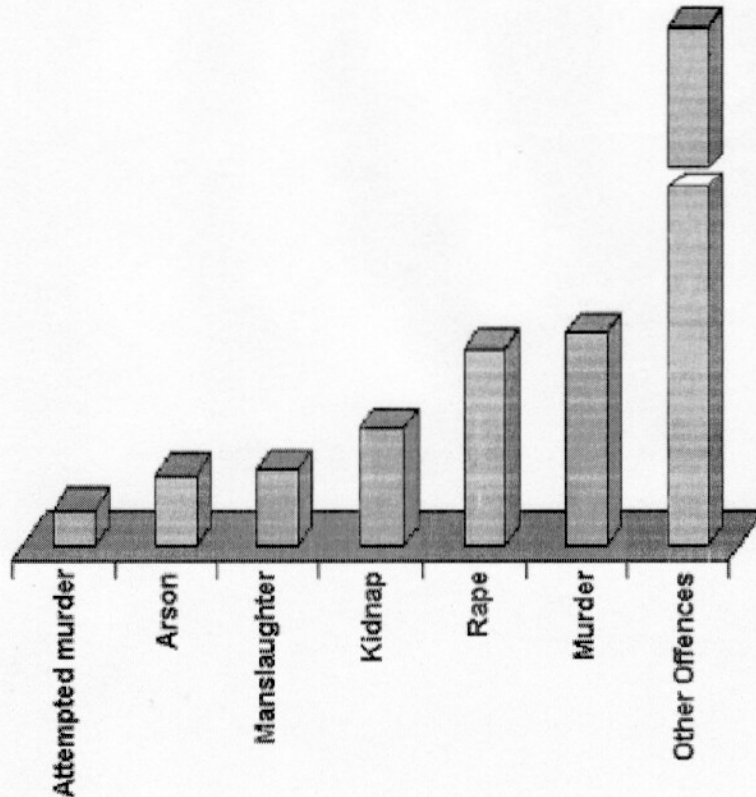
- Read each question carefully before you start to answer it.
- Keep an eye on the time.
- Try to answer every question.
- Check your answers if you have time at the end.

Answer ALL the questions.

You must write down all stages of your working.

1 A national newspaper printed this bar chart.

Re-offending for serious offences in England and Wales 2006–2007



(Source: adapted from Ministry of justice Information)

Write down **three** ways in which this bar chart is misleading.

1 No y-axis scale

2 Gap in the "Other offences" bar

3 3D bars

(Total for Question 1 is 3 marks)

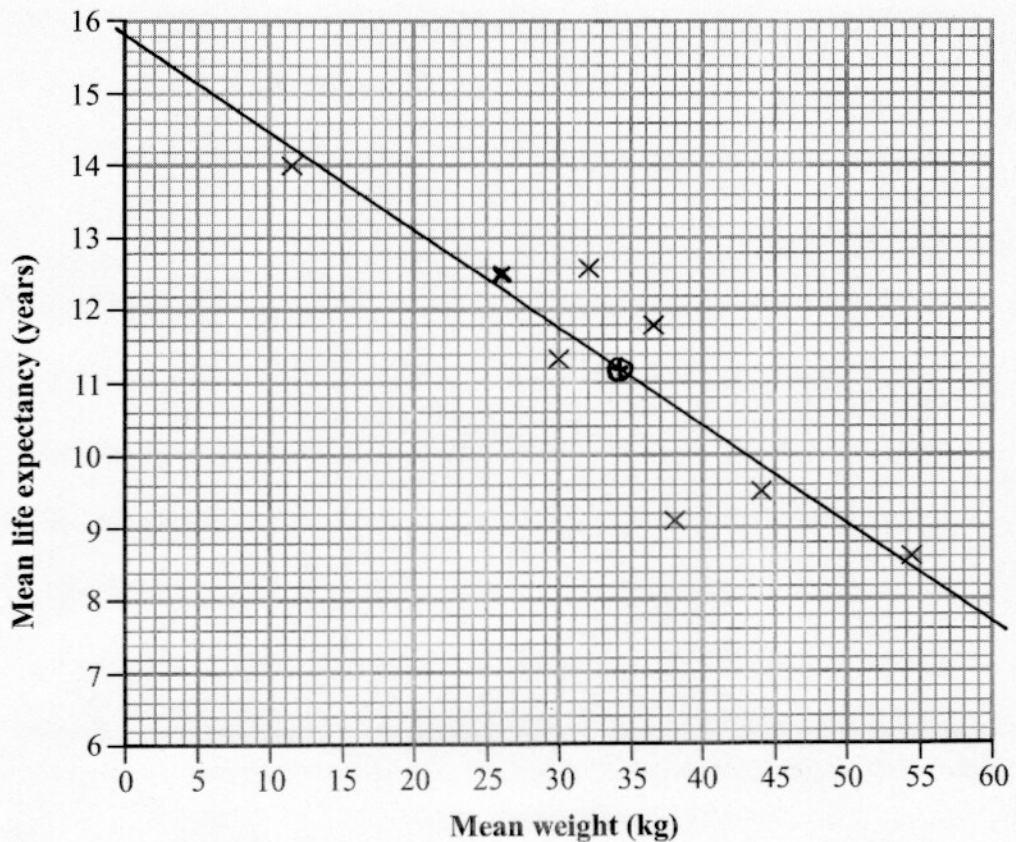
2 Eight breeds of dog are chosen at random.

The table shows the mean weight and the mean life expectancy for each breed.

Breed	Mean weight (x kg)	Mean life expectancy (y years)
Bullmastiff	54.5	8.6
Gordon Setter	30.0	11.3
Labrador Retriever	32.0	12.6
Old English Sheepdog	36.5	11.8
Rhodesian Ridgeback	38.0	9.1
Scottish Deerhound	44.0	9.5
Tibetan Terrier	11.5	14.0
Viszla	26.0	12.5

(Source: RSPCA)

Some of this information is shown on the scatter diagram.



(a) Plot the information for the Vizsla dogs to complete the scatter diagram.

(1)

(b) Describe and interpret the correlation shown by the scatter diagram.

It is a negative correlation.

Heavier ^{breeds of} dogs tend to have a lower life expectancy; lighter breeds have a longer life expectancy.

(2)

These eight breeds of dog have

a mean weight, \bar{x} , of 34.1 kg,

a mean life expectancy, \bar{y} , of 11.2 years.

(c) (i) Plot the mean point (\bar{x}, \bar{y}) .

(ii) Draw a line of best fit through the mean point.

(2)

A dog owner wants to predict the life expectancy of his Border Terrier.

The Border Terrier has a mean weight of 6.5 kg.

Using the line of best fit **may not** be reliable for this prediction.

(d) Explain why.

6.5 kg is lighter than any of the breeds plotted on the graph. Using the line of best fit would be extrapolation, beyond the range of the data.

This is unreliable as there is no evidence the trend continues as a straight line. [Small errors in line gradient will have a large effect, even if a straight line is suitable].

(Total for Question 2 is 6 marks)

3 The table gives information about what full time first degree graduates did after completing their courses in 2002

Destinations of full-time first degree graduates 2002

Area of Study	UK Employment		Overseas Employment	Continuing Education	Unemployed
	Permanent	Temporary			
UK	42.8 %	20.1 %	2.1 %	19.8 %	6.8 %
North East	44.9 %	17.2 %	2.4 %	21.6 %	6.0 %
North West	44.5 %	21.3 %	1.7 %	18.9 %	6.5 %
Yorkshire and the Humber	47.5 %	18.5 %	2.6 %	17.7 %	6.1 %
East Midlands	47.1 %	18.9 %	1.9 %	17.7 %	6.1 %
West Midlands	42.2 %	21.1 %	2.1 %	20.6 %	7.1 %
East	38.9 %	19.1 %	1.9 %	26.5 %	5.6 %
London	40.2 %	19.5 %	1.2 %	19.6 %	9.1 %
South East	42.0 %	21.0 %	2.1 %	19.6 %	6.5 %
South West	45.7 %	19.0 %	2.4 %	16.0 %	6.9 %

(Source: www.gov.uk)

(a) For the graduates who studied in the West Midlands, write down the percentage who are unemployed.

..... 7.1 %
(1)

For the graduates who studied in the South East area, the percentage who went into permanent UK employment is twice the percentage that went into temporary UK employment.

One other area of study had this ratio of permanent to temporary UK employment.

(b) Write down the name of this other area of study.

..... West Midlands
(1)



(c) For graduates who studied in London, work out the total percentage who went into some type of employment.

$$40.2 + 19.5 + 1.2$$

60.9 %
(2)

(d) For graduates who studied in the UK, work out the total percentage represented in this table.

$$42.8 + 20.1 + 2.1 + 19.8 + 6.8$$

91.6 %
(2)

The information in the table was gathered by means of a questionnaire given to all full time first degree students graduating in 2002

(e) The answer to part (d) is not 100%.

Suggest a reason why.

Some of them might never have replied.

(1)

(Total for Question 3 is 7 marks)

4 Some people think that drinking cocoa before bedtime may help to reduce blood pressure.

A university student is going to research this.

(a) Suggest a hypothesis the student could use.

Drinking cocoa before bedtime reduces blood pressure. [Statement not question.]

(1)

The student decides to collect information from students at his university.

The student decides to use a **sample**, not a census.

(b) Write down **two** reasons why.

Reason 1

Quick

Reason 2

Easy to do.

(2)

(c) Describe a sampling frame that the student could use.

A list of all the students at his university

(1)

There are more females than males at the university.

The student wants his sample to show this.

(d) Write down the name of the sampling method he should use.

Stratified

(1)

(e) Explain why the student might use a control group.

He can only know that any effect he sees is due to cocoa if he compares a group drinking cocoa and a group not drinking it, otherwise he cannot tell if the cocoa group always had low blood pressure, or if the b.p. changed because of some other influence eg. exams.

(1)

(Total for Question 4 is 6 marks)

5 A holiday in which people can visit two cities is called a two-centre holiday.

A travel company offers four two-centre holidays.

The two-way table shows the numbers of these holidays booked in 2010

	Venice	Rome	Total
Geneva	24	26	50
Paris	32	12	44
Total	56	38	94

(a) Complete the two-way table.

(2)

(b) Which of the two-centre holidays did most people choose?

Give a reason for your answer.

"Paris and Venice" is the most common holiday (32 bookings in 2010).

(2)

A person taking one of these holidays is chosen at random.

(c) What is the probability that this person visited Rome and Geneva?

$$\frac{26}{94} = \frac{13}{47}$$

(1)

(d) Given that this person visited Venice, work out the probability that they also visited Paris.

$$\frac{32}{56} = \frac{4}{7}$$

(2)

(Total for Question 5 is 7 marks)

6 A town council want to get information about local people's use of recycling facilities.

Two methods of collecting information have been suggested.

Method 1: To ask people using the recycling facilities at a local supermarket.

Method 2: To send a questionnaire to all council tax payers.

(a) Which method is likely to give the most reliable results?

Give a reason for your answer.

Method 2 will be most reliable. The council wants to know about people in general, not just those who already use one particular recycling facility.

Method 2 can get information from more people and will not be biased; method 1 is biased.

(2)

One question on the questionnaire is:

'In what ways do you use the council's recycling facilities?'

This is **not** a good question.

(b) Write down **one** reason why.

It's an open question

(& it assesses use of the facilities, how us
take boxes etc).

(1)

The council wants to find out how many times per month people use the recycling facilities at the supermarket.

(c) Suggest a suitable question they could put on the questionnaire.

How often do you use the recycling facilities?

Never , Once a month

2-3 times per month

4 or more times per month

(2)

(Total for Question 6 is 5 marks)

7 In a New Year sale a shoe shop reduces the prices of pairs of shoes.

The numbers of pairs of shoes sold during the 15 days of the sale were:

~~86~~ ~~84~~ 100 ~~97~~ ~~96~~ ~~88~~ ~~89~~ ~~60~~
~~78~~ ~~99~~ ~~91~~ ~~94~~ ~~79~~ ~~78~~ ~~82~~

(a) For the numbers of pairs of shoes sold in the New Year sale:

(i) find the median,

60 78 78 79 82 84 86 88 89
 91 94 96 97 99 100

$$\left(\frac{15+1}{2}\right)^{\text{th}} = 8^{\text{th}} \text{ value} = 88$$

88

(2)

(ii) find the Lower Quartile and Upper Quartiles.

$$Q_1: \frac{15+1}{4} = \frac{16}{4} = 4^{\text{th}} \text{ value}$$

$$Q_3: \frac{3(15+1)}{4} = 12^{\text{th}} \text{ value}$$

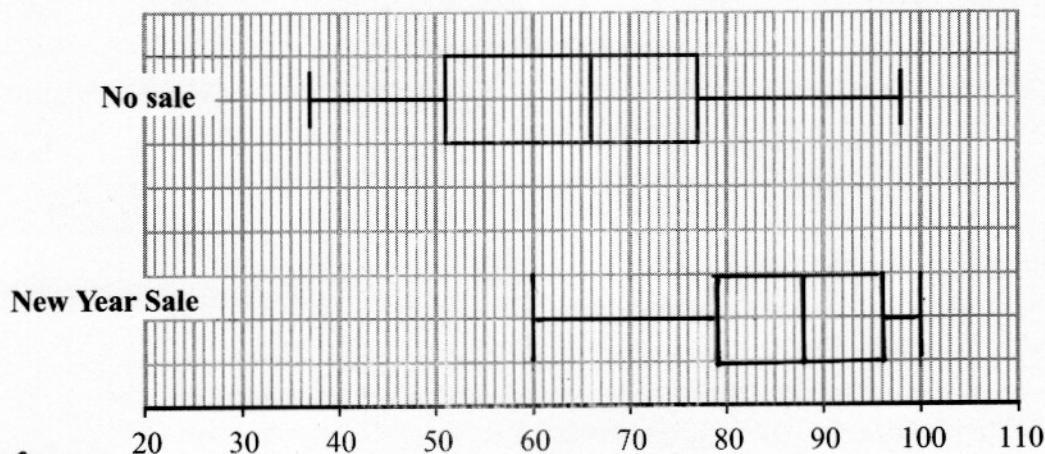
Lower Quartile 79

Upper Quartile 96

(2)

The box plot below shows information about the numbers of pairs of shoes sold when there is no sale on.

(b) On the same diagram draw the box plot for the New Year sale.



Could do:

$$IQR = 96 - 79 = 17$$

$Q_1 - 1.5 \times IQR = 53.5$ so 60 is not an outlier
 - but not needed for 3 marks!

(3)

* (c) Compare the distributions shown by the box plots and comment on how the New Year sale affected the number of pairs of shoes sold.

During the New Year sale the median is higher than at no sale times.

In fact every 'position' parameter (minimum, Q_1 , median, Q_3 , maximum) is higher during the sale.

During the sale the IQR is less.

The 'no sale' data has negative skew (looking at the quartiles ~~to~~ not the whiskers) but the New Year sale data is symmetrical (median centred between quartiles) if one ignores the 60 value.

(4)

(Total for Question 7 is 11 marks)

8 A market research company is going to take a national poll.

They want to find out what people think about the performance of different makes of new cars.

The company thinks about using a telephone poll.

They would choose 10 towns at random.

They would then choose 100 telephone numbers, at random, from each town's phone book.

The company would ring these 1000 numbers.

The people answering the phone would form the sample.

(a) Discuss whether or not this will be a satisfactory sample.

Not satisfactory, possibly biased because it only uses people with landline numbers. Only 10 towns are being used - too few.

1000 people is too few, ~~you~~^{they} will not get many responses for less common makes of cars.

(2)

There are 10 000 names in the phone book of one of these towns.

(b) Describe how the company could take a random sample of 100 people from this book.

Number all the names (from 0 to 9999).
Use a random number generator to make 100 different 4-digit numbers.
Use the people who were assigned these numbers.

(3)

The company finally decides to collect information either using face to face interviews or a questionnaire.

*c) Write down what the company needs to consider when choosing between these two methods of data collection.

You are more likely to get a response from a face to face interview, whereas not everyone will fill in a questionnaire.

An interview gives the chance to explain the questions, get a detailed response and perhaps information you would not have thought to request - this could then be awkward to summarise.

A questionnaire would be faster, ^{cheaper} and less time-consuming.

(3)

(Total for Question 8 is 8 marks)

- 9 There were 64 tsunamis (tidal waves) between the years 2000 and 2009. The table gives information about the maximum wave height, in metres, of these tsunamis.

Wave height h (m)	Frequency	mid-point x	fx	Cum. freq. freq.
$0 < h \leq 0.2$	26	0.1	2.6	26
$0.2 < h \leq 0.5$	8	0.35	2.8	34
$0.5 < h \leq 1.0$	6	0.75	4.5	40
$1.0 < h \leq 3.0$	6	2	12	46
$3.0 < h \leq 5.0$	5	4	20	51
$5.0 < h \leq 10$	8	7.5	60	59
$10 < h \leq 30$	3	20	60	62
$30 < h \leq 60$	2	45	90	64

(Source: National Geophysical Data Centre)

- (a) Work out the class interval that contains the median of these data.

The median will be the average of the 32nd and 33rd values. These are in the $0.2 < h \leq 0.5$ class $0.2 < h \leq 0.5$ metres
(2)

- (b) Calculate an estimate of the mean wave height of the tsunamis.

Give your answer to 1 decimal place.

You may use the blank columns in the table to help with your calculation.

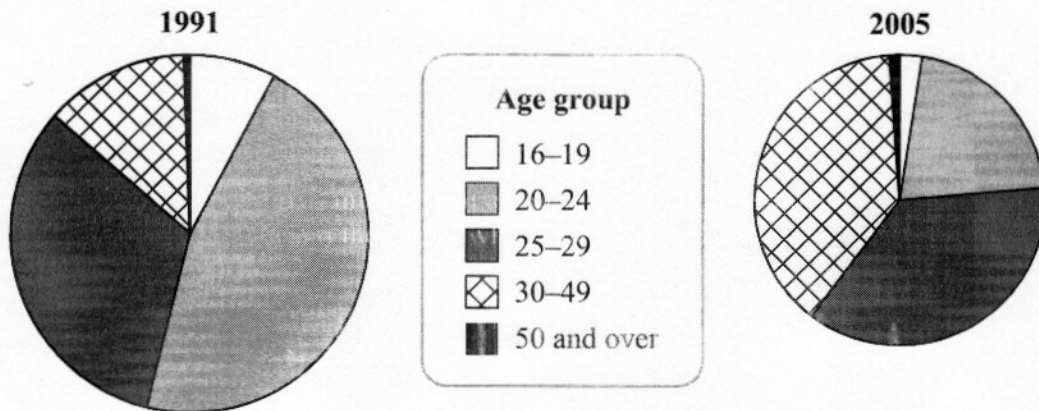
$$\begin{aligned} \sum fx &= 2.6 + 2.8 + 4.5 + 12 + 20 + 60 + 60 + 90 \\ &= 251.9 \end{aligned}$$

$$\text{Mean} = \frac{\sum fx}{n} = \frac{251.9}{64}$$

3.9 metres
(4)

(Total for Question 9 is 6 marks)

10 The comparative pie charts give information about the number of women who got married in 1991 and in 2005 and the age at which they married.



(Source: adapted from www.statistics.gov.uk)

(a) What has happened to the total number of women who married in 2005 compared to the total number in 1991?

Comment on how the pie charts show this.

Fewer women got married in 2005 than in 1991 (the area of the pie chart is smaller).

(2)

(b) Write down the age group with the greatest decrease from 1991 to 2005

(Biggest reduction in area).

20-24

(1)

(c) For the 30-49 age group, describe how the number of women who married changed in 2005 compared to 1991

Give a reason for your answer.

1991: radius 23.8 cm, angle 45°

2005: radius 19.5 cm, angle 135° roughly.

Don't actually need.

Area of sectors: (1991) $\frac{45}{360} \pi r^2 = 220 \text{ cm}^2$

(2005) $\frac{135}{360} \pi r^2 = 450 \text{ cm}^2$

The number of women in this age group who married has increased (bigger area, roughly doubled)

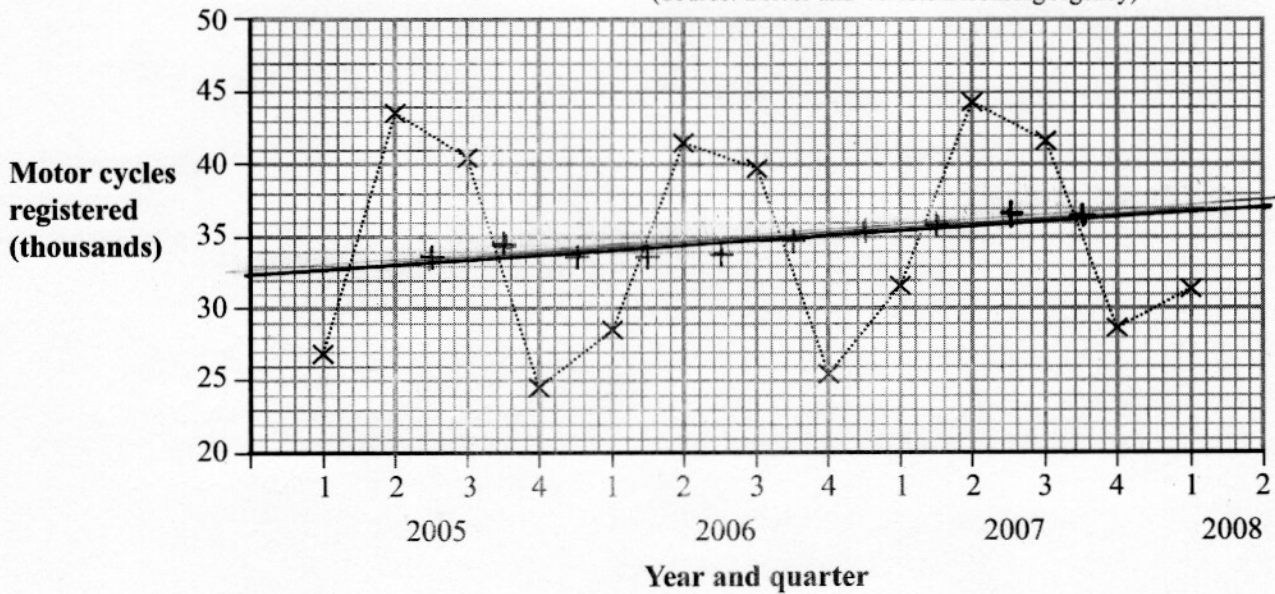
(2)

(Total for Question 10 is 5 marks)

11 The table shows the numbers of motor cycles, in thousands, registered each quarter from the first quarter of 2005 to the first quarter of 2008

	Quarter			
Year	1	2	3	4
2005	27.0	43.5	40.4	24.7
2006	28.7	41.4	39.8	25.3
2007	31.8	44.2	41.8	28.6
2008	31.4			

(Source: Driver and Vehicle Licensing Agency)



These data are plotted as a time series on the graph above.
The 4-point moving averages, except for the last two, are also plotted.

(a) Calculate the last two 4-point moving averages and plot them on the graph.

$$(31.8 + 44.2 + 41.8 + 28.6) / 4 = 36.5 \quad (3)$$

$$(44.2 + 41.8 + 28.6 + 31.4) / 4 = 36.5$$

..... 36.6 thousands and 36.5 thousands

(b) Draw a trend line on the graph.

(1)

(c) Describe and interpret the trend.

It is a rising trend. The number of motorcycles registered each quarter is increasing.

(2)

(d) Write down the quarter with the greatest number of motor cycles registered each year.

2, Second quarter

(1)

(e) Work out the mean seasonal variation for quarter 2
Give your answer to the nearest whole number.
Show your working.

$$\frac{10 + 7 + 8}{3} = \frac{25}{3} = 8\frac{1}{3} \text{ (thousand)}$$
$$= 8333$$
$$\approx 8300 \text{ (2 s.f.)}$$

8300

(2)

(f) Use your answer to part (e) to predict the number of motor cycles registered in quarter 2 of 2008

$$37000 + 8300 = 45300$$
$$\approx 45000$$

45000

(2)

(Total for Question 11 is 11 marks)

12 The table gives information about the year on year % increase in retail prices, and the average mortgage rate (%), in the first week of July for each of 11 years.

Year on year % increase in retail prices	Average mortgage rate (%)	Price increase rank	Mortgage rate rank	d	d ²
8.7	5.71	1	3	-2	4
8.0	5.17	2	7	-5	25
6.1	5.83	3	2	1	1
5.5	5.66	4	6	-2	4
5.4	4.19	5	10	-5	25
5.3	4.24	6	9	-3	9
5.1	3.49	7	11	-4	16
4.9	7.27	8	1	7	49
4.8	5.67	9	5	4	16
3.0	5.70	10	4	6	36
2.9	4.58	11	8	3	9

(Sources: www.statistics.gov.uk and www.nationwide.uk)

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

You may use the blank columns in the table to help with your calculations.

$$\sum d^2 = 4 + 9 + 1 + 9 + 25 + 9 + 16 + 49 + 16 + 36 + 9 = 194$$

$$r = 1 - \frac{6 \sum d^2}{n(n^2-1)} = 1 - \frac{6 \times 194}{11(11^2-1)} = 0.118$$

0.118

(3)

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

No correlation (too weak a correlation to draw reliable conclusions).

RPI and mortgage rate are not associated.

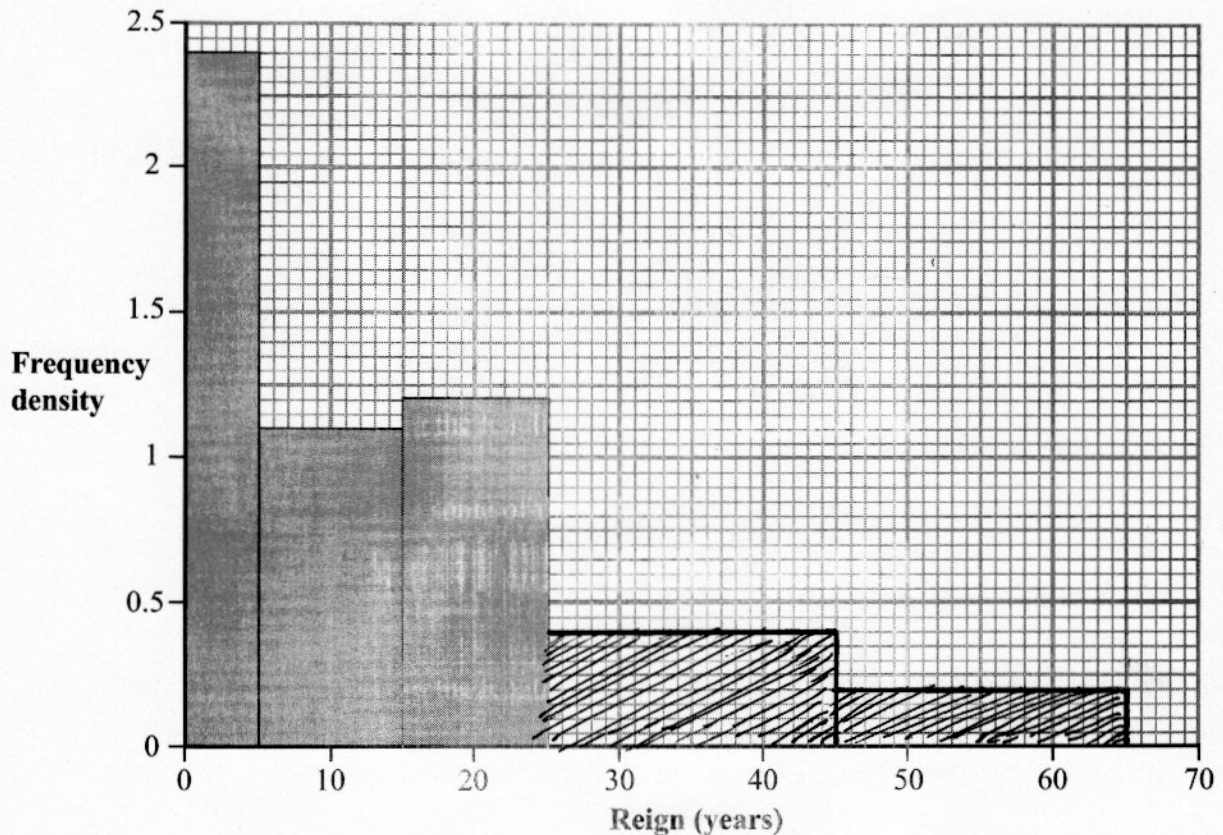
(2)

(Total for Question 12 is 5 marks)

13 The table gives information about the lengths of the reigns, in years, of the English Kings and Queens since the year 1014

Length l (in years)	$0 < l \leq 5$	$5 < l \leq 15$	$15 < l \leq 25$	$25 < l \leq 45$	$45 < l \leq 65$
Frequency	12	11	12	8	4
Class width	5	10	10	20	20
F. density	2.4	1.1	1.2	0.4	0.2

The incomplete histogram shows some of these data.



(a) Complete the histogram.

(3)

(b) Estimate how many English Kings and Queens reigned for between 20 and 50 years.

Area under histogram:

$$= 5 \times 1.2 + 8 + 5 \times 0.2$$

$$= 6 + 8 + 1 = 15 \text{ kings \& queens.}$$

15

(3)

(Total for Question 13 is 6 marks)

14 The probability of a person having an allergy to a particular nut is 0.1

A test is available to see if a person has this allergy.

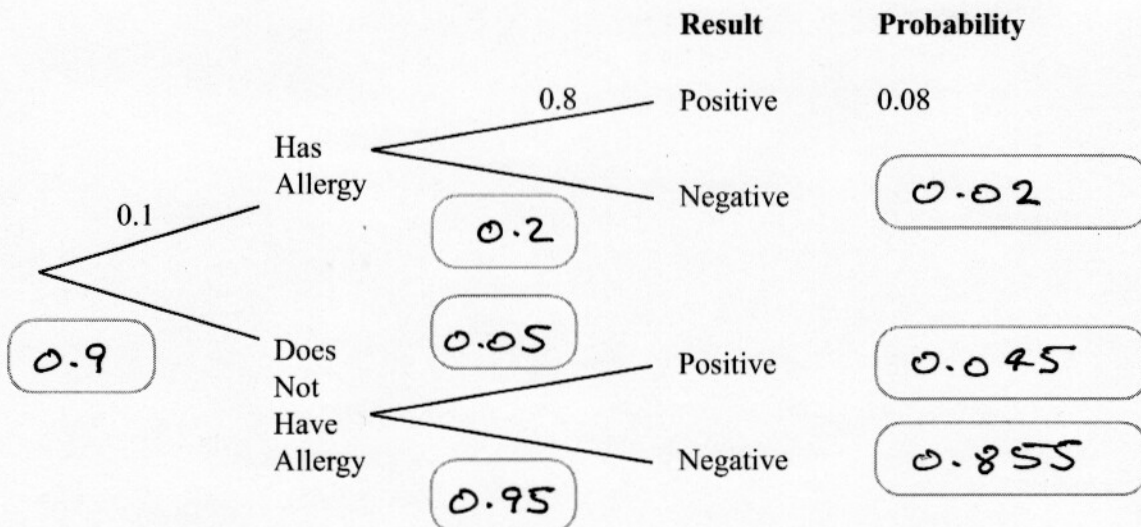
In 80% of the cases where the person has this allergy the test gives a positive result.

If the patient does not have this allergy there is a 0.05 probability of getting a positive result.

(a) Complete the tree diagram for the two events:

A person has the allergy
The test gives a positive result.

(2)



Mular takes the test and gets a positive result.

(b) Work out the probability that Mular has this allergy.

$$\frac{0.08}{0.08 + 0.045} = 0.64$$

0.64

(2)



Five people are selected at random.
 X is the number of these people who have the allergy.

(c) (i) What is the name of the probability distribution that is a suitable model for X ?

Binomial distribution

(1)

(ii) Write down the properties needed, in this context, for this distribution to be a suitable model for X .

The trials must be independent and the probability of having the allergy remains constant (i.e. it is not affected by whether other people have it).

There are a fixed number of trials (5 people).

There are only two possible outcomes for each trial ('has the allergy' or 'does not have it').

(2)

(iii) Work out the probability that out of these five people there will be exactly 3 people who have the allergy.

You may use $(p + q)^5 = p^5 + 5p^4q + 10p^3q^2 + 10p^2q^3 + 5pq^4 + q^5$.

$\begin{matrix} \nearrow & & \searrow & & \searrow & & \searrow \\ \text{All five} & \text{4 have} & & & & & \text{none of them} \\ \text{have it} & \text{it} & & & & & \text{have it.} \\ & & 3 & & 2 & & 1 \end{matrix}$

$$p = \text{probability ("has allergy")} = 0.1$$

$$q = \text{probability ("no allergy")} = 0.9$$

$$P(3 \text{ have allergy}) = 10p^3q^2 = 10 \times 0.1^3 \times 0.9^2 = 0.0081$$

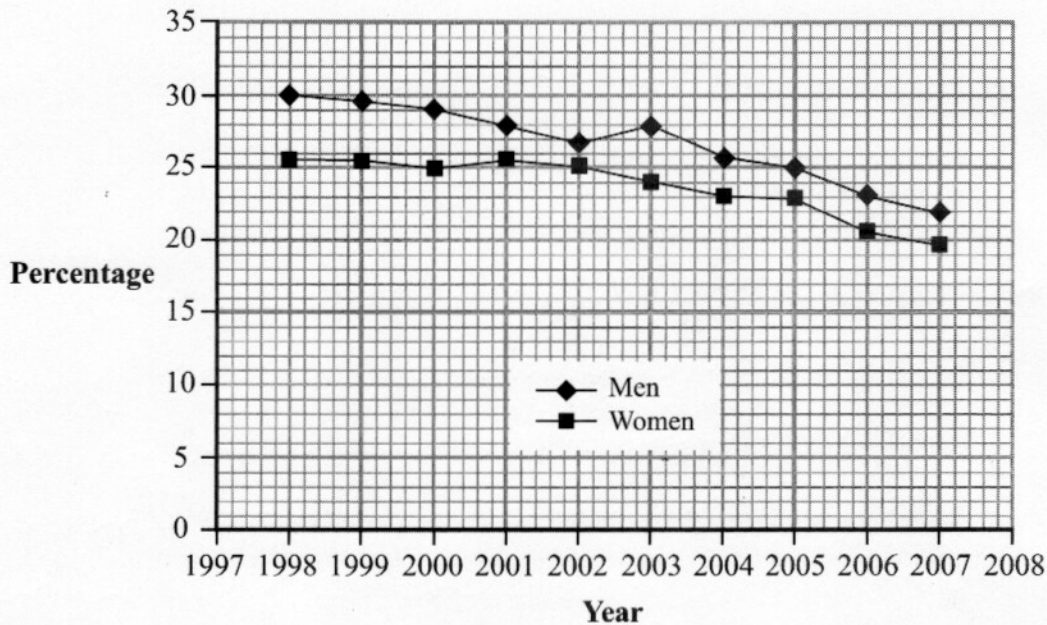
$$\begin{aligned} \text{or use } P(r \text{ "successes" in } n \text{ trials}) &= {}^nC_r p^r q^{n-r} \\ &= {}^5C_3 \times 0.1^3 \times 0.9^{5-3} = 0.0081. \end{aligned}$$

0.0081

(3)

(Total for Question 14 is 10 marks)

15 The time series graphs show the percentages of people aged 16 and over who smoked in the years 1998 to 2007



(Source: statistics.gov.uk)

In 1998 the government published a leaflet 'Smoking Kills'. This had a target of cutting the percentage of people of 16 and over smoking to 24% by 2010

(a) From the graph what conclusion can be made about the success of the leaflet 'Smoking Kills' between 1998 and 2007?

Explain the reasons for your answer.

There is a downwards (falling) trend in the percentage smoking, for both men and women. The 24% target was achieved before 2010 (men 22%, women 20% in 2007).

There is no evidence on this graph that this is a result of the leaflet's publication - there could already have been a long-term trend downwards or it could be due to other causes eg pricing. ⁽²⁾

(b) Compare the two time series graphs.

The women are consistently below the men
(a smaller proportion of women smoked, compared
to men).

The smoking rates for both men and women
have decreased.

The gap between men and women has
decreased.

(2)

(Total for Question 15 is 4 marks)

TOTAL FOR PAPER IS 100 MARKS