

Write your name here						
Surname SOLUTIONS	Other names					
Edexcel GCSE	Centre Number <table border="1" style="display: inline-table; border-collapse: collapse; text-align: center;"> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> </table>					
	Candidate Number <table border="1" style="display: inline-table; border-collapse: collapse; text-align: center;"> <tr><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td><td style="width: 20px; height: 20px;"></td></tr> </table>					
<h1 style="margin: 0;">Statistics</h1> <h2 style="margin: 0;">Paper 1H</h2>						
Higher Tier						
Monday 18 June 2012 – Afternoon 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.
- **Calculators may not be used.**

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 with your spelling, punctuation and grammar, as well as the clarity of expression, on these questions.*

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. Write your answers in the spaces provided.
You must write down all stages of your working.

1. The table gives information about the reasons given for authorised and unauthorised absences from school, of boys and girls, in the UK.

	State funded Secondary Schools		
	Boys %	Girls %	Total %
Percentage of absent sessions due to:			
Authorised Absence			
Illness (NOT medical or dental appointments)	54.05	56.52	55.30
Medical/dental appointments	5.61	6.32	5.97
Religious observance	1.25	1.14	1.19
Study leave	2.01	1.97	1.99
Traveller absence	0.12	0.10	0.11
Agreed family holiday	4.66	4.64	4.65
Agreed extended family holiday	0.09	0.08	0.08
Excluded, no alternative provision	3.52	1.29	2.39
Other authorised circumstances	8.03	7.46	7.74
Total Authorised Absence	79.33	79.52	79.43
Percentage of absent sessions due to:			
Unauthorised Absence			
Family holiday not agreed	1.67	1.74	1.70
Arrived late	1.12	1.10	1.11
Other unauthorised circumstances	13.65	13.63	13.64
No reason yet	4.22	4.02	4.12
Total Unauthorised Absence	20.67	20.48	20.57
Total Overall Absence	100.0	100.0	100.0

(Data source: dcsf.gov.uk)

- (a) Write down the main reason given for

- (i) authorised absence,

..... *illness*

- (ii) unauthorised absence.

..... *other unauthorised circumstances*

(2)

The percentage of boys' absences was more than twice the percentage of girls' absences for one of the reasons listed in the table.

- (b) Which reason?

..... *Excluded, no alternative provision.*

(1)

(Total for Question 1 is 3 marks)

2. Archaeologists divided a field into 36 squares of equal size. The number of Roman roof tiles found in each square is shown below.

Numbers of Roman roof tiles

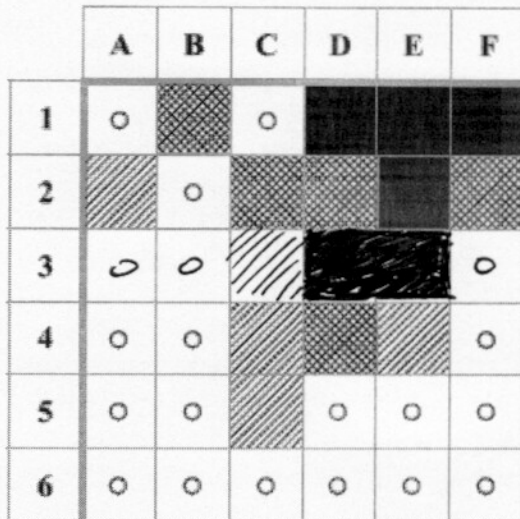
	A	B	C	D	E	F
1	3	9	4	18	16	24
2	5	3	11	14	20	12
3	1	0	8	16	18	4
4	2	2	5	10	5	3
5	1	4	6	4	3	2
6	2	3	4	1	2	2

Key

3	Means 3 roof tiles in square AI
---	---------------------------------

- (a) Use the information above to complete the choropleth map.

(2)



Key

Number of tiles

○	Less than 5
▨	5-8
▨	9-15
■	16-24

- (b) Use the choropleth map to describe the area in which they found the greatest number of roof tiles. Give a reason for your answer.

Top right of the grid - where the squares have the darkest shading.

(2)

(Total for Question 2 is 4 marks)

3. Collis Town Council wants to produce a magazine for adults living in Collis.

The council wants to find out the topics the adults are interested in.
They decide to ask a sample of the adults.

(a) Write down **two** advantages of using a sample rather than a census.

Advantage 1

Cheaper

Advantage 2

Quicker

(2)

(b) Describe a suitable sampling frame that could be used.

The electoral roll.

(1)

The council sends a questionnaire to all the adults living in North Street.

(c) Discuss whether or not this would be a good sample. *people are*

*Not a good sample - not selected at random
from the population, North Street may
not be representative.*

(2)

(d) Here are two questions on the questionnaire.

Explain what is wrong with each question.

(i) It is a good idea to have an article on recycling isn't it?

Yes No Don't know

This is a leading question - it tries to persuade you to agree. The results may be biased.

(ii) How much would you be willing to pay for a Collis magazine?

£0 Up to £1 £1 to £2 £2 or more

£2 could be in "£1 to £2" or "£2 or more" - categories overlap.

No time period - is this per week, per month?

(2)

(Total for Question 3 is 7 marks)

4. The table shows information from a 2008 survey about the ages of women having their first child.

Age (a) years	Frequency
$16 \leq a < 20$	7
$20 \leq a < 25$	19
$25 \leq a < 30$	27
$30 \leq a < 35$	27
$35 \leq a < 40$	17
$40 \leq a < 45$	3

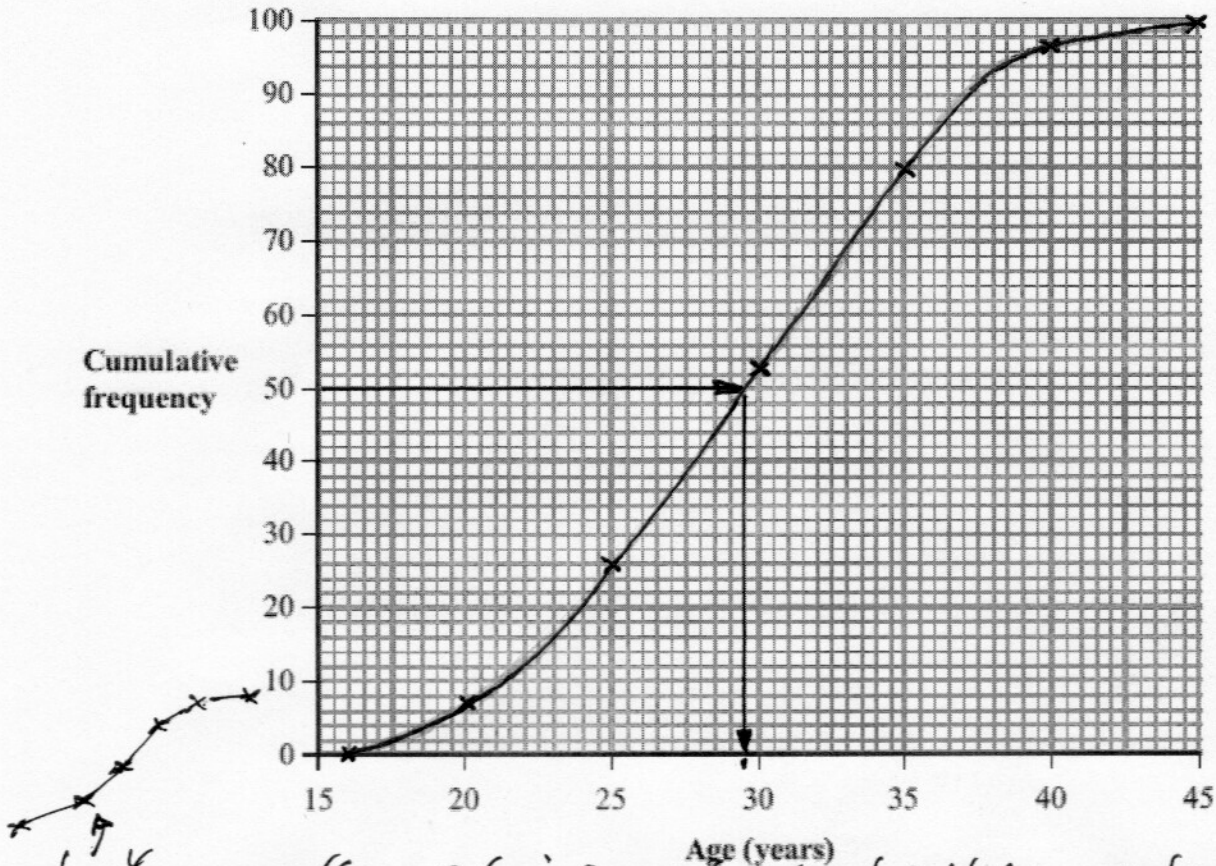
(Data source: Office for National Statistics)

- (a) Complete the cumulative frequency table below for these data.

Age (a) years	Frequency
$16 \leq a < 20$	7
$20 \leq a < 25$	$7+19 = 26$
$25 \leq a < 30$	$26+27 = 53$
$30 \leq a < 35$	$53+27 = 80$
$35 \leq a < 40$	$80+17 = 97$
$40 \leq a < 45$	$97+3 = 100$

(1)

(b) Draw a cumulative frequency diagram for this information.



nb. You are allowed to join points with straight lines instead of a curve

(c) Use your cumulative frequency diagram to find an estimate for the median age.

For grouped data, don't use $(\frac{n+1}{2})^{\text{th}}$ point,

use $\frac{n}{2} = \frac{100}{2} = 50^{\text{th}}$

29.5

(2)

In 1990, the median age for women having their first child was 25

(d) Using this information and your answer to (c) what can you say about the age for women having their first child?

The age at which women had their first child

went up between 1990 and 2008

↑ must show years to get mark

(1)

(Total for Question 4 is 7 marks)

5. There are 16 children at a party.

Here are their weights in kg.

14.8 18.5 15.1 15.4 15.8 17.3 16.4 16.7
 16.7 14.9 17.0 17.1 16.3 16.9 17.5 16.5

(a) Complete the ordered stem and leaf diagram for the weights.

	<u>Unordered:</u>
14	8 9
15	1 4 8
16	4 7 7 3 9 5
17	3 0 1 5
18	5

	<u>Ordered:</u>
14	8 9
15	1 4 8
16	3 4 5 7 7 9
17	0 1 3 5
18	5

Key
 16 | 3 = 16.3

(2)

(b) Find the median weight.

$$\text{Median} = \frac{8^{th} + 9^{th}}{2} = \frac{16.5 + 16.7}{2}$$

..... 16.6 kg
 (1)

(c) Work out the mean weight.

$$(14.8 + 18.5 + \dots + 16.5) \div 16 = \frac{262.9}{16} = 16.43125$$

..... 16.4 kg
 (2)

Another child arrives at the party.

This child weighs 16.6 kg.

(d) Use words to describe the effect this will have on

(i) the median,

When added to the sorted list (stem + leaf diagram) this child is between 16.5 and 16.7, hence is 9th place out of 17 children. $\left(\frac{17+1}{2}\right)^{th} = 9^{th}$ for median
 \therefore median = 16.6, unchanged.

(ii) the mean.

16.6 is above the mean so the mean will go up slightly.
[Actually goes up to 16.441, still 16.4 to 1 d.p.]
Either is OK

(2)

One child is picked at random from the children now at the party.

(e) Work out the probability that this child weighs less than 16 kg.

There are 17 children at the party and 5 of them weigh < 16 kg, so $\frac{5}{17}$

$$\frac{5}{17}$$

(2)

(Total for Question 5 is 9 marks)

6. A government statistician plans to do a survey on unemployment.

He is going to compare the number of months of male unemployment with the number of months of female unemployment.

(a) Write down a hypothesis he can use.

Men tend to have been unemployed for longer than women.

(1)

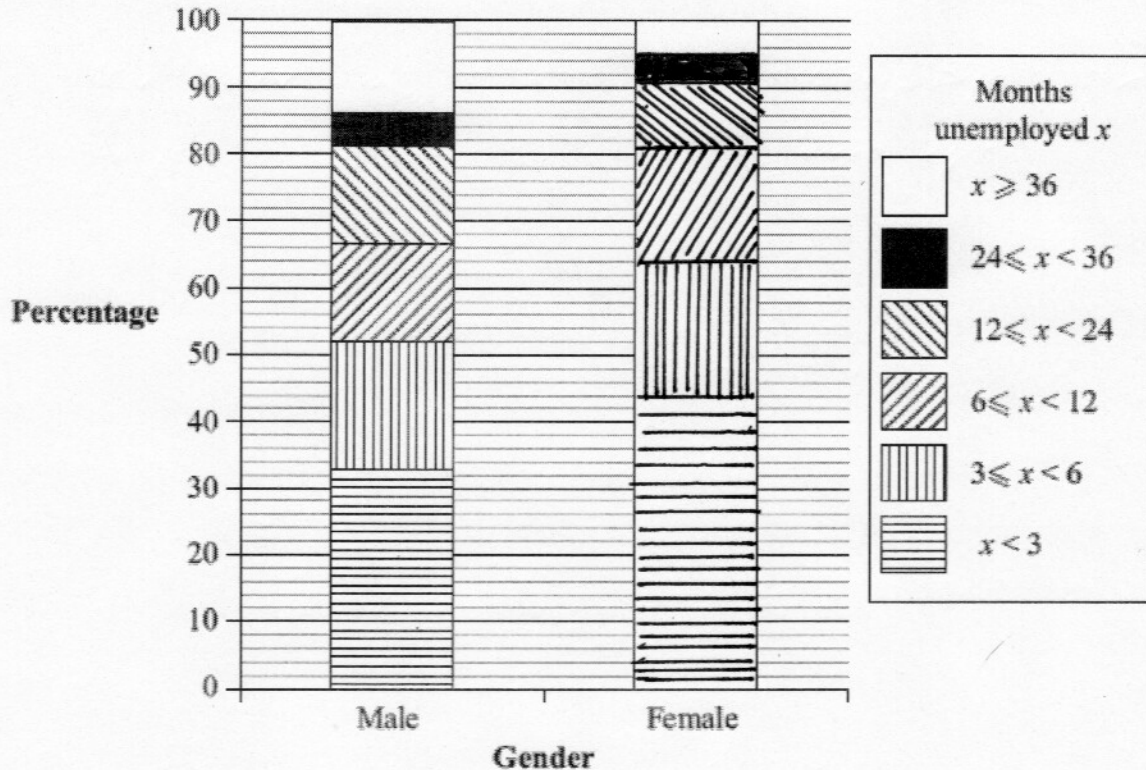
The table shows some information about the number of months of unemployment of males and females aged 16 and over at the end of 2007.

Number of months of unemployment (x)	Males %	Females %
$x < 3$	33%	44%
$3 \leq x < 6$	19%	20%
$6 \leq x < 12$	15%	17%
$12 \leq x < 24$	14%	10%
$24 \leq x < 36$	5%	4%
$x \geq 36$	14%	5%
Total	100%	100%

(Data source: Labour force survey, Office for National Statistics)

(b) Use the information in the table to complete the composite bar chart.

Percentages of unemployed males and unemployed females in different age groups for a given number of months



(3)

(c) Using information in the composite bar chart, what conclusions can you make about your hypothesis?

Give a reason for your answer.

A larger percentage of women have only been unemployed for short periods (eg 68% of women for 6 months or less), compared with 52% of men). A larger percentage of men have been unemployed for long periods (23.5% of men for > 12 months versus 19% women).

The hypothesis is correct.

(2)

(Total for Question 6 is 6 marks)

7. A random sample was taken of 10 women tennis players. The table shows data about each player.

Player	Country	Speed of serve (mph)	Position in world
V. Williams	USA	124	1
S. Lisicki	GER	123	49
S. Stosur	AUS	119	7
N. Petrova	RUS	118	20
L. Hradecka	CZE	117	87
J. Goerges	GER	116	77
C. Wozniacki	DEN	114	3
E. Dementieva	RUS	113	5
E. Baltacha	GBR	112	62
E. Vesnina	RUS	111	44

(Data source: 2009.wimbledon.org, sonyericsonwtatour.com)

- (a) **quantitative** **response** **qualitative**

Choose one of the words above to describe each of

- (i) country,

Qualitative
.....

- (ii) speed of serve,

Quantitative
.....

- (iii) position in world.

Quantitative
.....

(3)

- (b) Which data is continuous?

Speed of serve
.....

(1)

- (c) Work out Spearman's rank correlation coefficient between the speed of serve and the position in world tennis.
(You may use the blank columns in the table to help with your calculations.)

Player	Speed of serve (mph)	Rank for speed of serve	Position in world	Rank for position in world	d	d^2
V. Williams	124	1	1	1	0	0
S. Lisicki	123	2	49	7	-5	25
S. Stosur	119	3	7	4	-1	1
N. Petrova	118	4	20	5	-1	1
L. Hradecka	117	5	87	10	-5	25
J. Goerges	116	6	77	9	-3	9
C. Wozniacki	114	7	3	2	5	25
E. Dementieva	113	8	5	3	5	25
E. Baltacha	112	9	62	8	1	1
E. Vesnina	111	10	44	6	4	16

$$\sum d^2 = 0 + 25 + 1 + 1 + 25 + 9 + 25 + 25 + 1 + 16 = 128$$

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

← check, use formula on page 2 of paper

$$n = 10$$

$$\therefore r = 1 - \frac{6 \times 128}{10 \times 99} = 0.224$$

0.224

(4)

- (d) Describe and interpret your answer to part (c).

This is a very weak positive correlation. There is no clear evidence of a relationship between serve speed and position in the world.

OR

"Weak positive" correlation and "slight evidence that the faster the serve, the higher the position".

(2)

(Total for Question 7 is 10 marks)

*8. The head teacher wants to change the school starting time.
She wants school to start half an hour earlier.
She wants to find out what the students think about the change.

She is going to ask a sample of students instead of using a census.

(a) Give **two** disadvantages of using a census.

1. *Time consuming*

2. *Expensive*

The head teacher wants to use a representative sample of 100 students.
There are a total of 800 students in the school.

The table shows the number of students in each year at the school.

Year	7	8	9	10	11	12	13
Number of students	144	144	152	96	96	80	88

(b) Describe in detail how she should take this sample.

She should use stratified sampling to pick an appropriate number from each year.

Each student should be given a number (from 0 to 799).

For a total sample size of 100, she needs $\frac{100}{800} = \frac{1}{8}$ of the students from each year:

Year: 7 8 9 10 11 12 13
pick: $\frac{144}{8} = 18, 18, 19, 12, 12, 10, 11$

She should use a random number generator to make 3 digit random numbers and use these to select the students who will be consulted, until she has the required number from each year.

(5)

(Total for Question 8 is 7 marks)

9. A sample was taken of 100 people who had an accident when driving a car.

The table gives some information about the ages of these people.

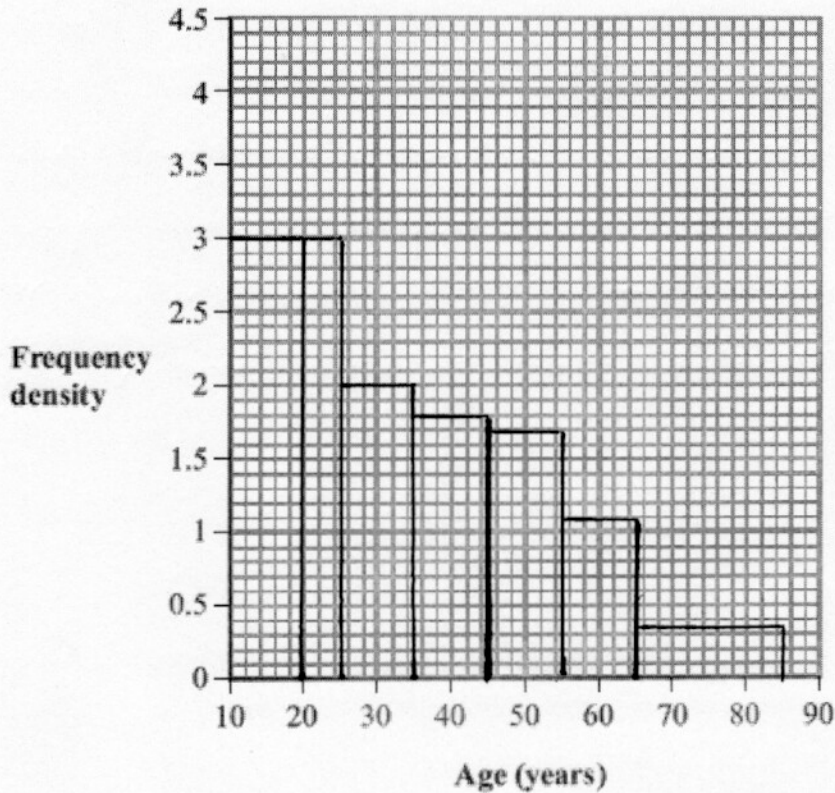
Age (x years)	Frequency	Class width	Frequency density
$16 \leq x < 20$	12	4	3
$20 \leq x < 25$	15	5	3
$25 \leq x < 35$	20	10	2
$35 \leq x < 45$	18	10	1.8
$45 \leq x < 55$	17	10	1.7
$55 \leq x < 65$	11	10	1.1
$65 \leq x < 85$	7	20	0.35

(Data source: census.gov)

(a) Complete the table.

(3)

(b) Draw a histogram to represent these data.



(2)

Young drivers pay more for their motor insurance than older drivers.

(c) Use your histogram to explain why.

The histogram has positive skew - the highest accident rates are for younger drivers, while older drivers have a lower risk of having an accident.

The insurance premium is related to risk - safer (older) people pay less than young drivers.

(2)

(Total for Question 9 is 7 marks)

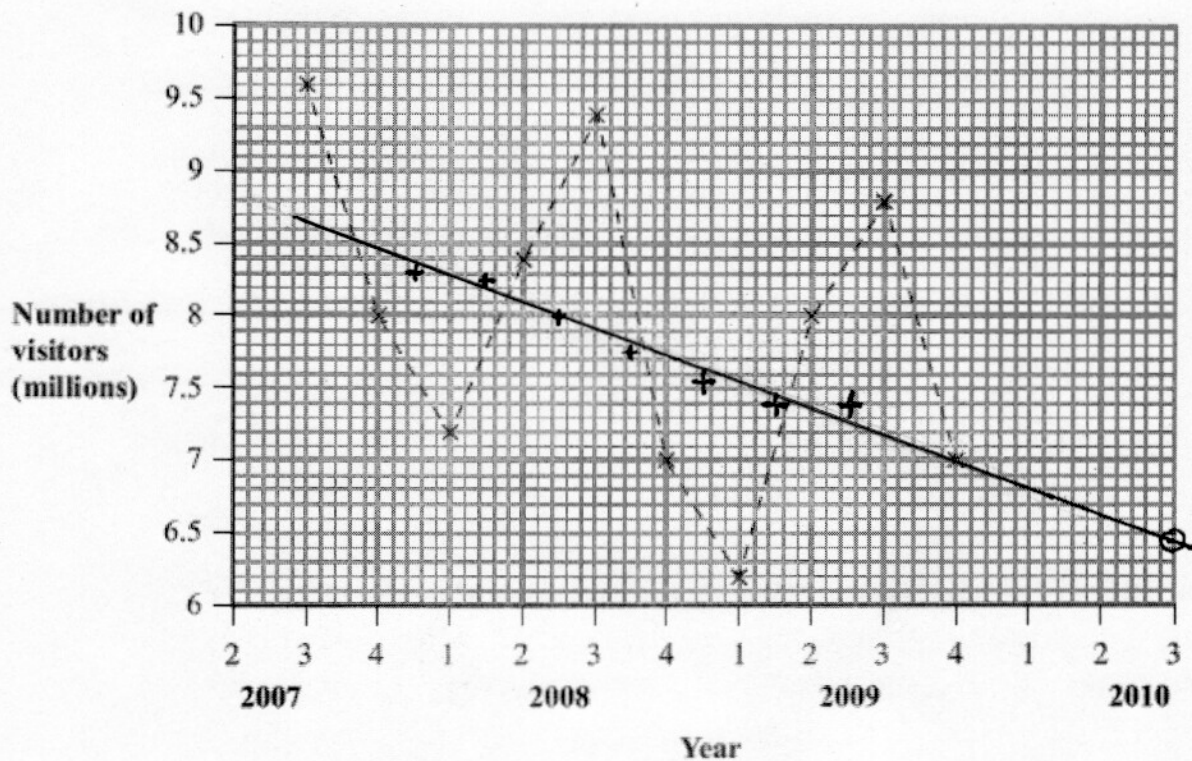
10. The table gives information about the numbers, in millions, of visitors to Great Britain during the years 2007 to 2009.

Quarter \ Year	Winter 1	Spring 2	Summer 3	Autumn 4
2007			9.6	8.0
2008	7.2	8.4	9.4	7.0
2009	6.2	8.0	8.8	7.0

(Data source: visitbritain.org)

The time series graph shows this information.

The first four 4-point moving averages have been plotted on the time series graph.



- (a) (i) Calculate the last **three** 4-point moving averages.

$$(9.8 + 7.0 + 6.2 + 8.0) / 4 = \underline{\quad 7.65 \quad}$$

$$(7 + 6.2 + 8 + 8.8) / 4 = \underline{\quad 7.5 \quad}$$

$$(6.2 + 8 + 8.8 + 7) / 4 = \underline{\quad 7.5 \quad}$$

- (ii) Plot these moving averages on the time series graph.

(4)

(b) Draw a trend line for the moving averages.

(1)

(c) Describe the trend.

A falling trend.

(1)

(d) Work out the average seasonal effect for quarter 3.

Seasonal variations from the graph:

2007 Q3	0.95	} average 1.32 = $\frac{0.95 + 1.5 + 1.6}{3}$
2008 Q3	1.5	
2009 Q3	1.6	

1.3

million

(2)

(e) Use your answer to (d) to estimate the number of visitors in quarter 3 of 2010.

The trend line hits 6.45 million in Q3 2010

$$6.45 + 1.32 = 7.77 \quad \text{(Must show working)}$$

= 7.8 to 2 significant figures

7.8

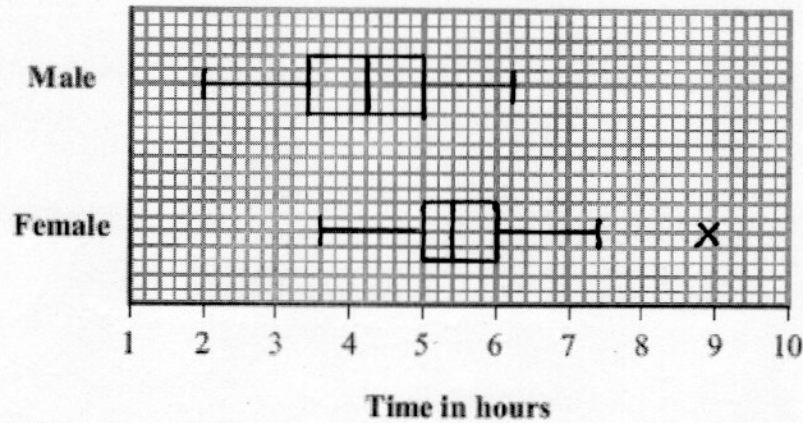
million

[7.7 to 8.9 ok for final answer] (3)

(Total for Question 10 is 11 marks)

11. The box plots give information about the time, in hours, taken by a random sample of male runners and female runners to complete a marathon.

Times taken to complete a Marathon



(Data source: Marathonguide data)

- (a) Write down an estimate of the time by which 50% of the male runners had finished the race.

.....^{4.2}.....hours
 (= 4h 12 minutes). (1)

- (b) Write down an estimate of the time by which 75% of the male runners had finished the race.

.....⁵.....hours
 (1)

- (c) (i) What does the × represent on the female box plot?

.....^{An outlier}.....

- (ii) Show why point × is in this category.

From the female box plot, IQR = 6 - 5 = 1 hour
 $Q_3 = 6$ hours.
 Values above $Q_3 + 1.5 \times IQR$ are outliers
 $6 + 1.5 \times 1 = 7.5$ hours, 9 hours is beyond this,
 hence it is an outlier. (4)

- (d) Compare the distributions of the times of the male runners and the times of the female runners.

Male runners have a lower median (so typically faster than female runners).

Male runners have higher interquartile range than female runners (more spread of times) and a higher range also.

The male runners have a symmetrical distribution; the female runners have slight positive skew.

(4)

(Total for Question 11 is 10 marks)

12. There were 20 people in a figure-skating competition.

The mark (x) for each person was recorded.

(a) Given that $\sum x = 280$ and $\sum x^2 = 4220$,

(i) Show that the mean mark is 14

$$n = 20$$

$$\text{Mean} = \frac{\sum x}{n} = \frac{280}{20} = 14$$

(ii) Show that the standard deviation of the marks is 3.9 to one decimal place.

$$\begin{aligned} \text{SD} &= \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{4220}{20} - 14^2} \\ &= \sqrt{211 - 196} = \sqrt{15} = 3.873 \\ &= 3.9 \text{ to 1 d.p.} \end{aligned}$$

(3)

Melvin got a mark of 12 in the figure-skating competition.

(b) Use your answers from parts (a)(i) and (a)(ii) to find Melvin's standardised score for the figure-skating competition.

$$\begin{aligned} \text{Standardised score} &= \frac{x - \mu}{\sigma} = \frac{12 - 14}{3.9} \\ &= -0.516 \\ &= -0.5 \text{ to 1 d.p.} \\ &= -0.5 \end{aligned}$$

MUST SHOW WORKING

Final answer is only 1 mark!!

..... (3)

Melvin got a standardised score of -1.1 in a free-skating competition.

(c) Did Melvin do better in the figure-skating competition or in the free-skating competition?

Give a reason for your answer.

Melvin's standardised scores:

Figure skating -0.5

Free " -1.1

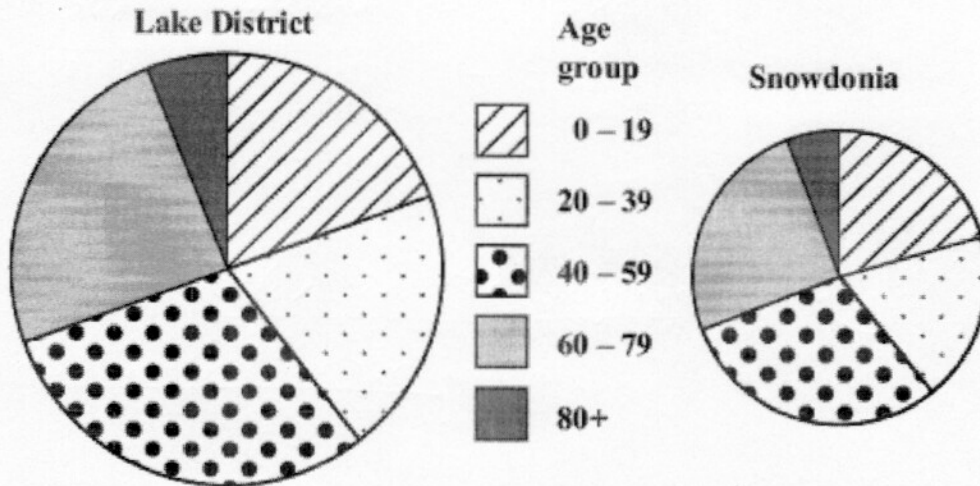
He did better in the figure skating (higher standardised score).

(2)

(Total for Question 12 is 8 marks)

13. The comparative pie charts give information about the numbers of people living in two national parks. They also show the age groups of the people.

The numbers of people living in two national parks



(Data source: Office for National Statistics)

- (a) Compare the total number of people living in the Lake District with the total number of people living in Snowdonia.

Give reasons for your answer.

There are more people living in the Lake District than in Snowdonia because the Lake District pie chart has a larger area.

(2)

The largest number of people living in the Lake District is in the same age group as the largest number of people living in Snowdonia.

- (b) Write down the age group.

40-59

(1)

The angles at the centres of the sectors for the 40 – 59 age group are the same on both pie charts.

- (c) Describe how the number of people aged 40 – 59 years in the Lake District compares with the number of people aged 40 – 59 years in Snowdonia.

Give a reason for your answer.

There are more 40-59 year olds living in the Lake District than in Snowdonia. The sector area is the same fraction of a larger pie chart \therefore larger area.

(2)

(Total for Question 13 is 5 marks)

14. There are 100 students at a music school.

12 of the students play the piano and violin only.

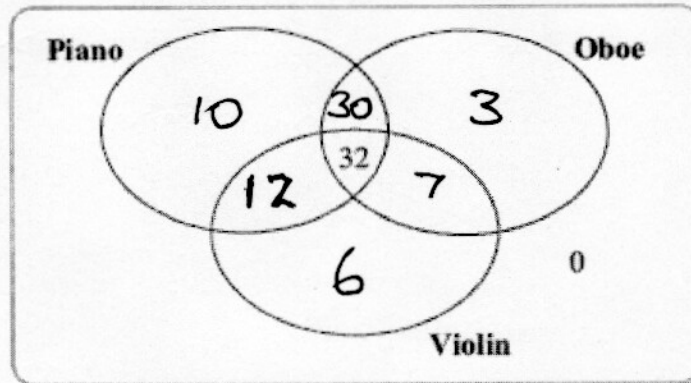
7 of the students play the violin and oboe only.

30 of the students play the piano and oboe only.

84 of the students in total play the piano.

57 of the students in total play the violin.

(a) Complete the Venn diagram using this information.



(3)

Fill in the 12, 7 and 30 first.

$$84 - (12 + 32 + 30) = 10$$

$$57 - (12 + 32 + 7) = 6$$

$$100 - (84 + 6 + 7 + 0) = 3$$

One of the 100 students is picked at random.

(b) Write down the probability that this student plays the piano and the oboe.

$$\frac{30 + 32}{100} = \frac{\cancel{62}}{\cancel{100}} = \frac{62}{100}$$

$$\frac{31}{50}$$

(1)

Given that the student plays the piano,

(c) find the probability that this student also plays the violin.

$$\frac{12+32}{84} = \frac{44}{84} = \frac{11}{21}$$

$$\frac{11}{21} \dots\dots\dots (2)$$

(Total Question 14 is 6 marks)

TOTAL FOR PAPER IS 100 MARKS

END