Write your name here

Surname	Other names

Pearson Edexcel GCSE

Centre Number	Candidate Number

Statistics

Paper 1H

Higher Tier

Monday 27 June 2016 – 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.

Turn over ▶

P46488A

6/6/6/

P 4 6 4 8 8 A 0 1 2 4

PEARSON

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 $= \dfrac{\sum fx}{\sum f}$

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

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

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

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

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

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

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

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

Answer ALL the questions.

Write your answers in the spaces provided.

You must write down all stages in your working.

1 John wants to know which pets are more popular, dogs or cats.

(a) Write down a question that John could use on a questionnaire to investigate this.

(1)

Many types of animal can be pets.
Valerie thinks that the gender of a person may affect their choice of pet.
She plans to investigate this with a face-to-face survey.

(b) Design a data collection sheet for Valerie to use.

(2)

(c) Valerie will not be able to use a scatter diagram to show her results.

Explain why not.

.....

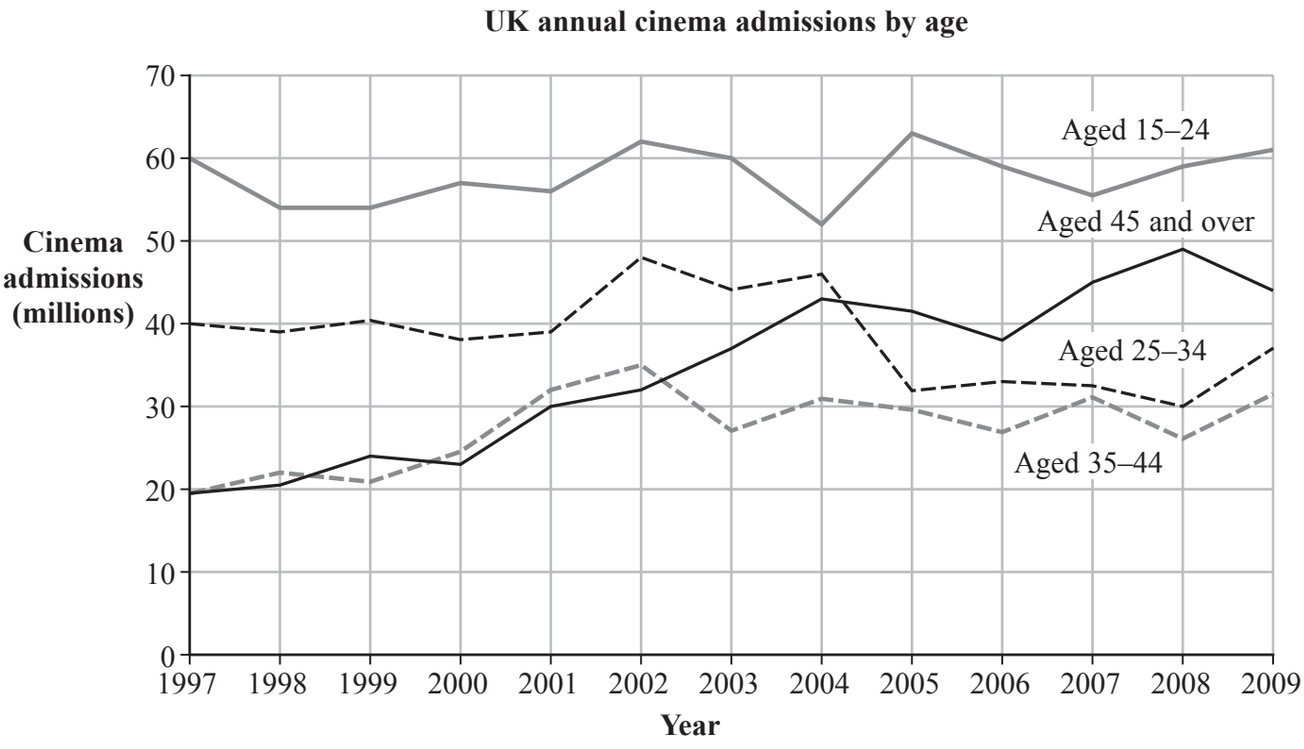
.....

.....

(1)

(Total for Question 1 is 4 marks)

2 The graph shows some information about annual cinema admissions in the UK from 1997 to 2009



(*Data source: UK Film Council*)

(a) Write down the age group which had the least number of cinema admissions in 2003

.....

(1)

(b) Write down the number of cinema admissions for people aged 25–34 in 2008

..... million

(1)

(c) Describe the trend in annual cinema admissions for people aged 45 and over between 1997 and 2009

.....

(1)

Jonathan says that the graph does not show all the cinema admissions from 1997 to 2009

(d) Explain why Jonathan is correct.

.....

.....

(1)

(Total for Question 2 is 4 marks)

3 Here are 10 cards.

1	2	2	2	2	3	3	4	4	4

Devika picks at random one of these cards.

(a) (i) Find the probability that this card has a 2 or a 4 on it.

.....

(ii) Find the probability that this card does **not** have a 3 on it.

.....

(2)

Getting an even number and getting an odd number are mutually exclusive events.

(b) Explain what is meant by the term mutually exclusive.

.....

.....

(1)

Sam picks at random one of the 10 cards.
He looks at the number on the card and then replaces the card.
Sam again picks at random one of the cards and looks at the number on the card.

(c) Find the probability that both of the numbers are odd numbers.

.....

(2)

(Total for Question 3 is 5 marks)

4 The table shows the number of students in each year in the mathematics department of a university.

Year	first year	second year	third year	**Total**
Number of students	90	78	72	240

Amanda wants to find out what the students think about the mathematics department.

She decides to take a sample of 40 of these students, stratified by year.

(a) Show that there should be 15 first year students in the sample.

(1)

Amanda uses a computer to generate the following list of random numbers.

47	12	53	53	26	06	03	89
27	04	44	49	11	24	33	14

(b) Explain how she can use these numbers to select the 15 first year students in the sample.

(3)

Amanda chooses the sample of 15 first year students.

7 of these students said they were happy with the mathematics department.

(c) Work out an estimate for the total number of first year students who are happy with the mathematics department.

.....

(2)

(Total for Question 4 is 6 marks)

5 The BabyDrive company is designing a car seat for babies.

The company has to decide what size to make the seat.
The company asks for data about the babies born in five UK hospitals in the last year.
The company does this by carrying out a survey.

(a) (i) Write down one variable the company should use in their survey.

.....

.....

(ii) Circle the word in the list below that best describes this variable.

Qualitative Discrete Continuous

(2)

(b) (i) State whether the data in the survey is primary or secondary.
Give a reason for your answer.

.....

.....

.....

(ii) Give one advantage and one disadvantage of using this type of data.

Advantage .....

.....

.....

Disadvantage .....

.....

.....

(3)

(Total for Question 5 is 5 marks)

6 A record company wants to estimate the proportion of people who downloaded music illegally during the last month.

They want people to be able to give honest answers so they designed the survey below.

Flip a fair coin. Keep the result to yourself.

- If you get Heads on the coin, tick box A.
- If you get Tails on the coin, answer this question.

  Have you downloaded music illegally during the last month?

  If **yes**, tick box A. If **no**, tick box B.

☐ A ☐ B

1000 people completed the survey.

(a) Estimate the number of people who got Heads on the coin.

.....

(1)

560 of the 1000 people ticked box A.

(b) Estimate the proportion of people who downloaded music illegally during the last month.

.....

(2)

(Total for Question 6 is 3 marks)

7 Mobile phones use one of a number of Operating Systems.

The table shows information about worldwide sales of mobile phones during Quarter 3 of 2012 and Quarter 3 of 2013

Operating System	Quarter 3 of 2012		Quarter 3 of 2013	
	Sales (millions)	Market Share (%)	Sales (millions)	Market Share (%)
Android	124.5	72.6	205.0	81.9
iOS	24.6	14.3	30.3	12.1
Microsoft	4.0	2.3	8.9	3.6
Blackberry	8.9	5.2	4.4	1.8
Other	9.5	5.5	1.6	0.6
Total	171.6	100.0	250.2	100.0

(*Data source: adapted from Gartner, November 2013*)

(a) Describe what happened to the total sales of mobile phones between Quarter 3 of 2012 and Quarter 3 of 2013

.....

(1)

(b) Which Operating System more than doubled its number of sales?

.....

(1)

(c) Which Operating System had the biggest change in Market Share percentage?

.....

(1)

The Market Share (%) column for Quarter 3 of 2012 adds up to 99.9 but the total is given as 100.0

(d) Give a reason why.

.....

.....

(1)

Debra plans to draw comparative pie charts to show the information in the table for each year.

(e) Give a reason why she might do this instead of drawing multiple bar charts.

.....

.....

(1)

(f) Find the size of the angle she should use for iOS in the 2012 pie chart.

..... °

(2)

Debra uses a radius of 5 cm for the 2012 pie chart.

(g) Work out the radius she should use for the 2013 pie chart.

..... cm

(2)

(Total for Question 7 is 9 marks)

8 A large company has 60 offices in different towns.

The directors of the company want to find out the opinions of their employees on a planned change to working hours.

They decide to choose at random 10 offices and survey all the employees in these offices.

(a) State one advantage and one disadvantage of using this sampling method.

Advantage .....

.....

Disadvantage .....

.....

(2)

(b) Circle the word below that best describes this sampling method.

Random Quota Systematic Cluster Stratified

(1)

The directors plan to use one of two options for the survey.

Option 1: each employee completes a questionnaire anonymously.

Option 2: each employee is interviewed by their office manager.

(c) What might you say to the directors to help them decide between these two options?

.....

.....

.....

.....

.....

.....

.....

.....

(2)

(Total for Question 8 is 5 marks)

9 The table shows the average cost of comprehensive motor insurance bought online each January from 2010 to 2013

January 2010	January 2011	January 2012	January 2013
£ 501.75	£ 618.59	£ 651.32	£ 595.66

(*Data source: AA British Insurance Premium Index*)

(a) Using 2010 as the base year, find the price index (price relative) for motor insurance in January 2011

.....

(2)

*(b) (i) Calculate the value of the **chain base** index number for motor insurance in January 2013

.....

(ii) Interpret your answer.

.....

.....

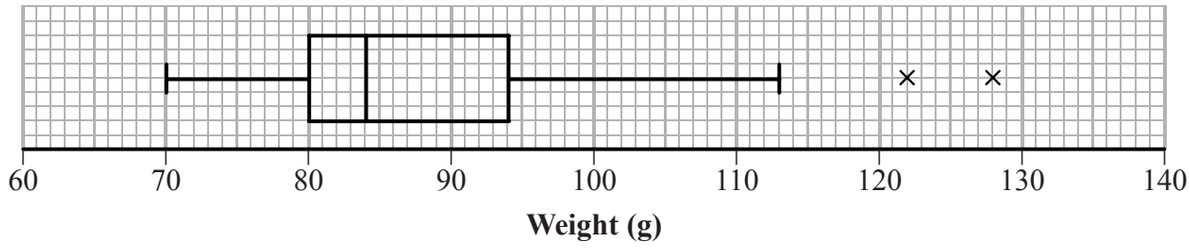
.....

.....

(4)

(Total for Question 9 is 6 marks)

10 The box plot shows the distribution of weights of a sample of salad potatoes.



(a) Find the median and the interquartile range (IQR).

median = ... g

IQR = ... g

(3)

The mean and standard deviation of this sample are

mean = 85.5 g standard deviation = 9.4 g

A sample of new potatoes has the following summary statistics

median	34 g	mean	35 g
IQR	7 g	standard deviation	4.4 g

Sally wants to compare the distributions of the weights of the sample of salad potatoes and the sample of new potatoes.

For the comparison she can use one of two options.

Option 1: the values of median and IQR.

Option 2: the values of mean and standard deviation.

*(b) Explain which of these options would be best for Sally to use to compare these distributions.

.....

.....

.....

(2)

*(c) Compare the distributions of the weights of the samples of salad potatoes and new potatoes.

.....

.....

.....

.....

.....

(2)

(Total for Question 10 is 7 marks)

11 Seb thinks that the number of medals won by a country in the Olympic Games is affected by the wealth of the country.

(a) Suggest a hypothesis you could use to investigate this.

.....

.....

.....

(1)

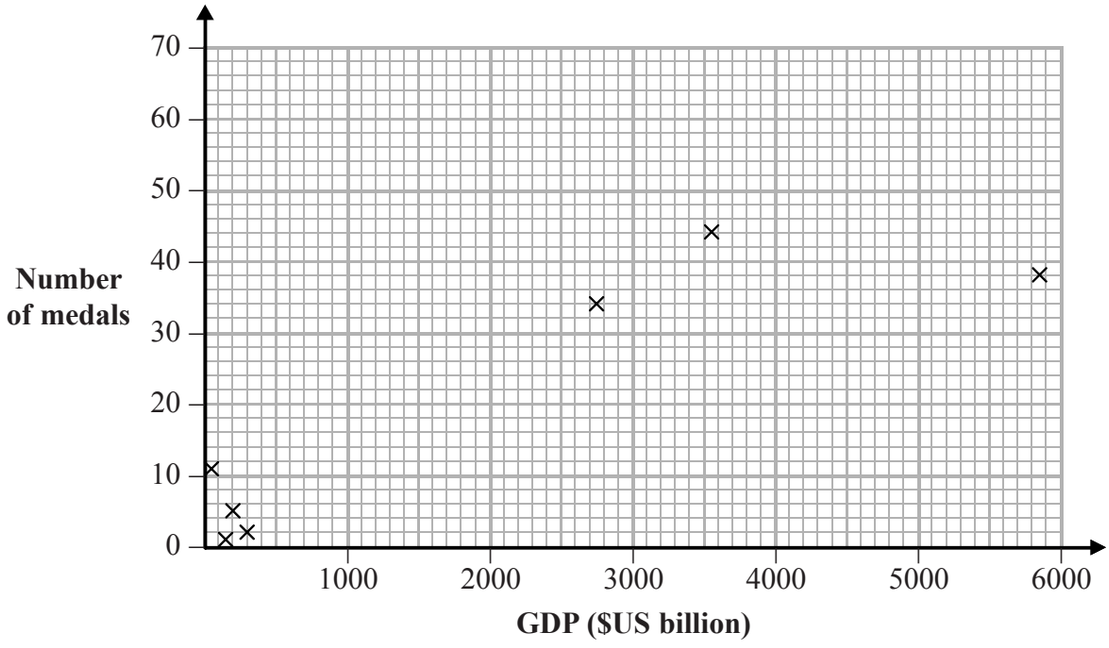
Seb takes a random sample of ten countries that won at least one medal in the London Olympics. He uses GDP (in billions of US dollars) as a measure of wealth.

The table shows information about these ten countries.

Country	GDP ($US billions)	Number of medals	GDP rank	Number of medals rank		
France	2750	34	3			
Germany	3550	44	2			
Greece	300	2	7			
Ireland	200	5	8			
Japan	5850	38	1			
Kenya	50	11	10			
Kuwait	150	1	9			
Mexico	1150	7	5			
Poland	500	10	6			
UK	2450	65	4			

(*Data source: adapted from theguardian.com*)

Some of this information is shown on the scatter diagram.



(b) Complete the scatter diagram by plotting the points for Mexico, Poland and the UK.

(2)

The GDPs of the countries have been ranked in the table.

(c) (i) Complete the column for the Number of medals rank.

(ii) Calculate Spearman's rank correlation coefficient for these data.
You may use the blank columns in the table to help with your calculations.

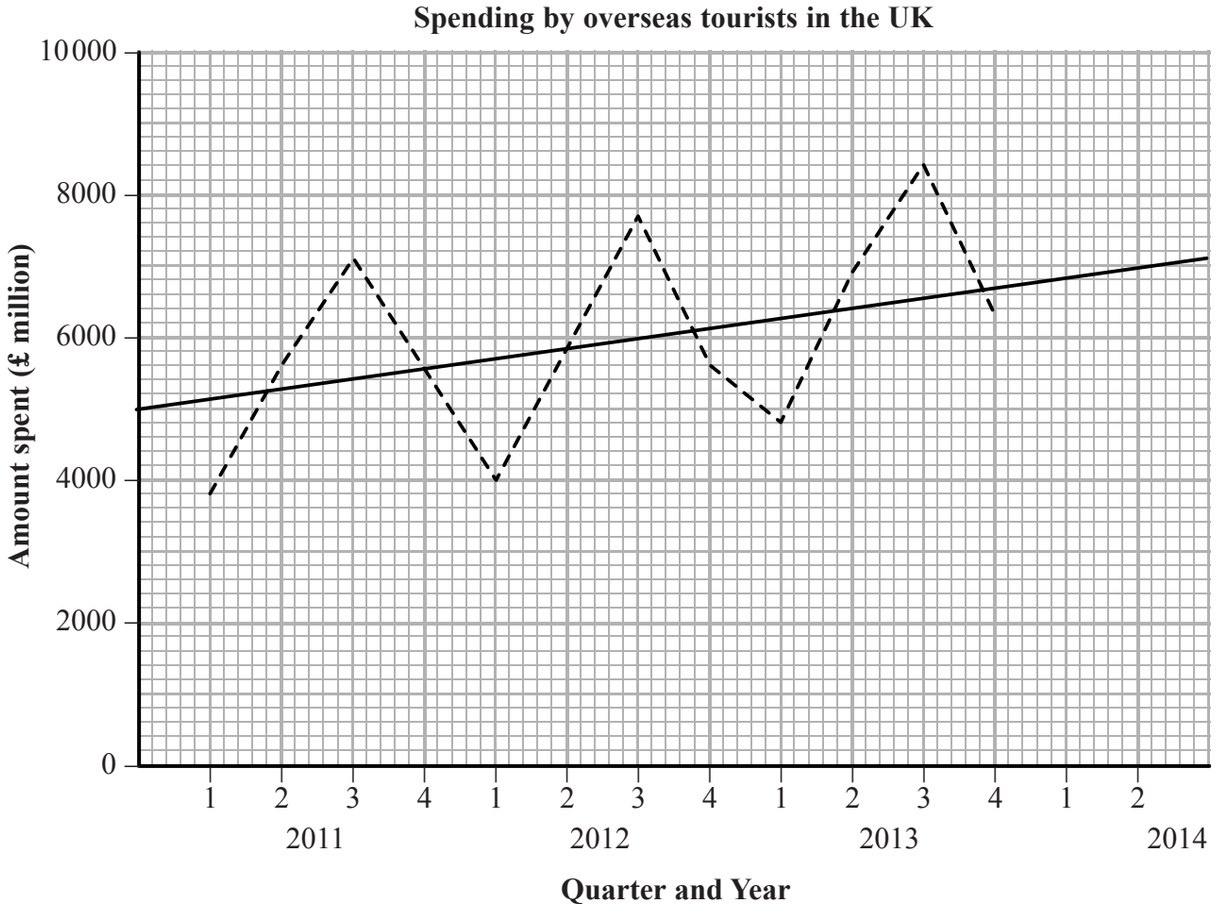
(4)

*(d) What conclusion can you reach about your hypothesis in part (a)?
You must give a reason for your answer.

(2)

(Total for Question 11 is 9 marks)

12 The time series graph gives information about the total amount of money spent by overseas tourists in the UK for each quarter for the years 2011 to 2013



(*Data source: Office for National Statistics*)

A trend line has been drawn on the graph.

The trend line is based on 4-point moving averages.

(a) Explain why **4-point** moving averages were chosen.

.....

.....

(1)

(b) Discuss any seasonal variation shown by the graph.

Do **not** do any calculations.

.....

.....

.....

.....

(2)

*(c) (i) Work out the gradient of the trend line.

.....

(ii) Interpret your answer.

.....

.....

(3)

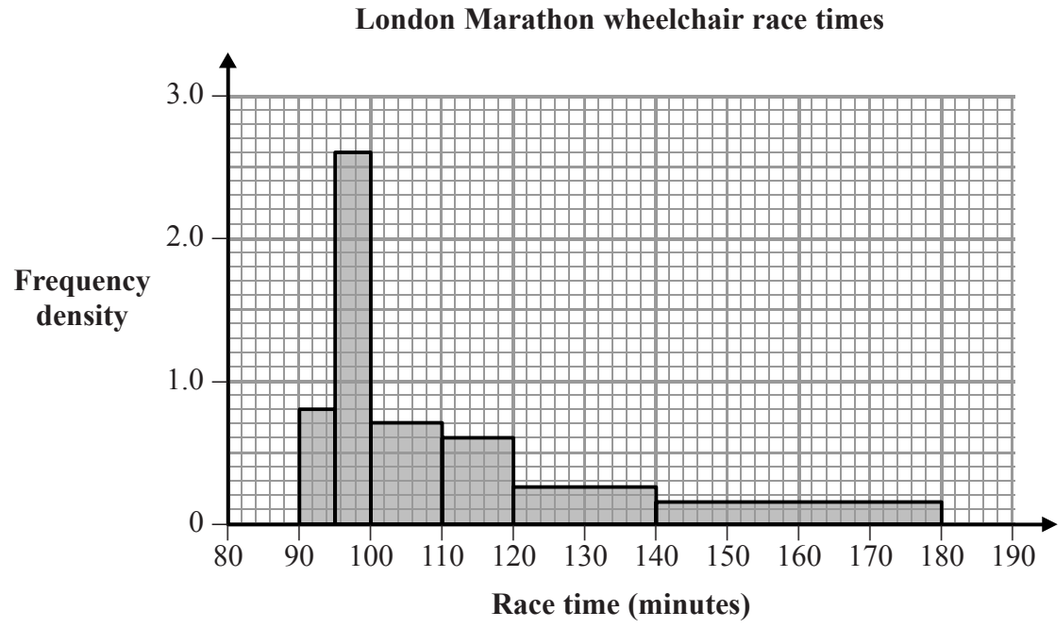
(d) Calculate an estimate for the total amount of money spent by overseas tourists for Quarter 1 in 2014
You must show your working.

£ ..... million

(3)

(Total for Question 12 is 9 marks)

13 The histogram shows information about the race times for the 41 wheelchair competitors in the 2014 London Marathon.



(*Data source: virginmoneylondonmarathon.com*)

(a) Describe the shape of the distribution.

.....

(1)

4 of the competitors had a race time in the class interval 90 minutes to 95 minutes.

(b) Show why the frequency density for this class interval is 0.8

(1)

(c) Calculate the number of competitors with a race time between 95 minutes and 110 minutes.

.....

(2)

(d) Use the histogram to work out an estimate for how long it took the first 21 competitors to finish the race.

..... minutes

(2)

David wants to predict the proportion of wheelchair competitors who will finish the New York Marathon in less than two hours.

(e) Explain whether or not it is sensible to use the London Marathon race times for the prediction.

.....

.....

.....

.....

.....

(2)

(Total for Question 13 is 8 marks)

14 Kirstin took tests in Maths, in Physics and in French.

The table shows information about the marks of all students who took the tests.

	Maths	**Physics**	**French**
Mean	53	69	48
Standard deviation	8	10	6

Kirstin scored 63 marks in Maths.

(a) Show that Kirstin's standardised score in Maths is 1.25

(1)

Kirstin scored 78 marks in Physics.

(b) Work out whether Kirstin did better in Maths or in Physics.
You must explain your answer.

.....

.....

(3)

Kirstin's standardised score in French was –0.5

(c) Work out Kirstin's mark in French.

.....

(2)

(Total for Question 14 is 6 marks)

15 Eve measures the heights, in metres, of 12 fully grown sunflowers.

Here are her results.

2.30 3.71 2.61 2.61 2.45 3.10 2.87 2.09 2.86 2.96 2.65 3.28

(a) (i) Find the mean.

..... m

(ii) Show that the standard deviation is 0.4 m, correct to 1 significant figure.

(4)

(b) Find the percentage of these sunflowers with heights within two standard deviations of the mean.

..... %

(2)

Eve thinks that the heights of the sunflowers could be modelled by a normal distribution.

(c) Discuss whether or not your answer to part (b) supports what Eve thinks.

.....

.....

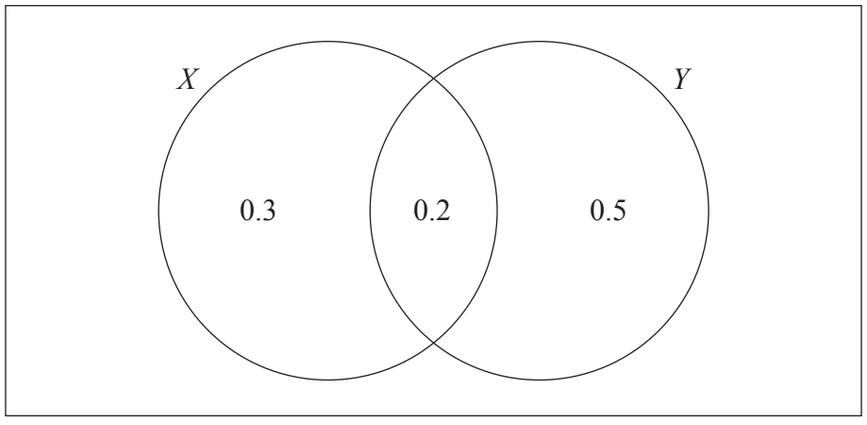
.....

.....

(2)

(Total for Question 15 is 8 marks)

16 The Venn diagram shows probabilities relating to two events, X and Y.



(a) Explain whether or not X and Y are exhaustive events.

.....

.....

(1)

(b) (i) Work out $P(X) \times P(Y)$.

.....

(ii) Explain why the events X and Y are **not** independent.

.....

.....

(3)

Two different events, A and B, are such that

$$P(A) = 0.6 \qquad P(B) = 0.5 \qquad P(A \cap B) = 0.25$$

(c) Find $P(A \cup B)$.

.....

(2)

(Total for Question 16 is 6 marks)

TOTAL FOR PAPER IS 100 MARKS

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA