

Please check the examination details below before entering your candidate information

Candidate surname		Other names	
Centre Number		Candidate Number	
<b>Pearson Edexcel</b> <b>Level 1/Level 2 GCSE (9–1)</b>		<div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid black; width: 20px; height: 20px;"></div> <div style="border: 1px solid black; width: 20px; height: 20px;"></div> <div style="border: 1px solid black; width: 20px; height: 20px;"></div> <div style="border: 1px solid black; width: 20px; height: 20px;"></div> <div style="border: 1px solid black; width: 20px; height: 20px;"></div> </div> <div style="display: flex; justify-content: space-around;"> <div style="border: 1px solid black; width: 20px; height: 20px;"></div> <div style="border: 1px solid black; width: 20px; height: 20px;"></div> <div style="border: 1px solid black; width: 20px; height: 20px;"></div> <div style="border: 1px solid black; width: 20px; height: 20px;"></div> </div>	
<h1>Thursday 13 June 2019</h1>			
Afternoon (Time: 1 hour 30 minutes)		Paper Reference <b>1ST0/1H</b>	
<h2>Statistics</h2> <h3>Paper 1</h3> <h3>Higher Tier</h3>			
<b>You must have:</b> Ruler graduated in centimetres and millimetres, protractor, pair of compasses, pen, HB pencil, eraser, scientific 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.*
- Scientific calculators may be used.
- You must **show all your working out** with **your answer clearly identified** at the **end of your solution**.



### Information

- The total mark for this paper is 80
- The marks for **each** question are shown in brackets  
– *use this as a guide as to how much time to spend on each question.*

### Advice

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

Turn over ►

R60681A

©2019 Pearson Education Ltd.

6/6/1/1/



Pearson

### Higher Tier Formulae

**You must not write on this page.**

**Anything you write on this page will gain NO credit.**

$$\text{Skew} = \frac{3(\text{mean} - \text{median})}{\text{standard deviation}}$$

$$\text{Standard deviation} = \sqrt{\frac{1}{n} \sum (x - \bar{x})^2}$$

*An alternative formula for standard deviation is*

$$\text{standard deviation} = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

Spearman's rank correlation coefficient

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

$$\text{Rates of change (e.g. Crude birth rate} = \frac{\text{number of births} \times 1000}{\text{total population}} \text{ )}$$

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Answer ALL questions.

Write your answers in the spaces provided.

You must write down all the stages in your working.

- 1 The diagram represents a children's playground that has been divided into 20 squares of equal area.

In the playground there are some children and some play equipment only.

The number of children in each square at 11 am one Saturday is shown below.

Number of children

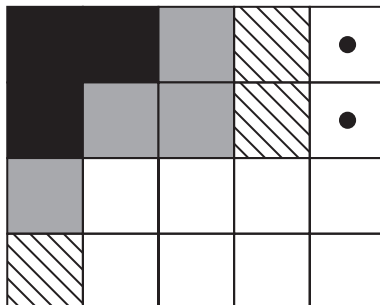
11	10	7	5	0
9	7	6	3	1
8	4	3	1	1
5	4	2	0	1

Key:

9

means 9 children  
in this square.

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



Key:

Number of children



9 or more

6–8

3–5

0–2

(2)

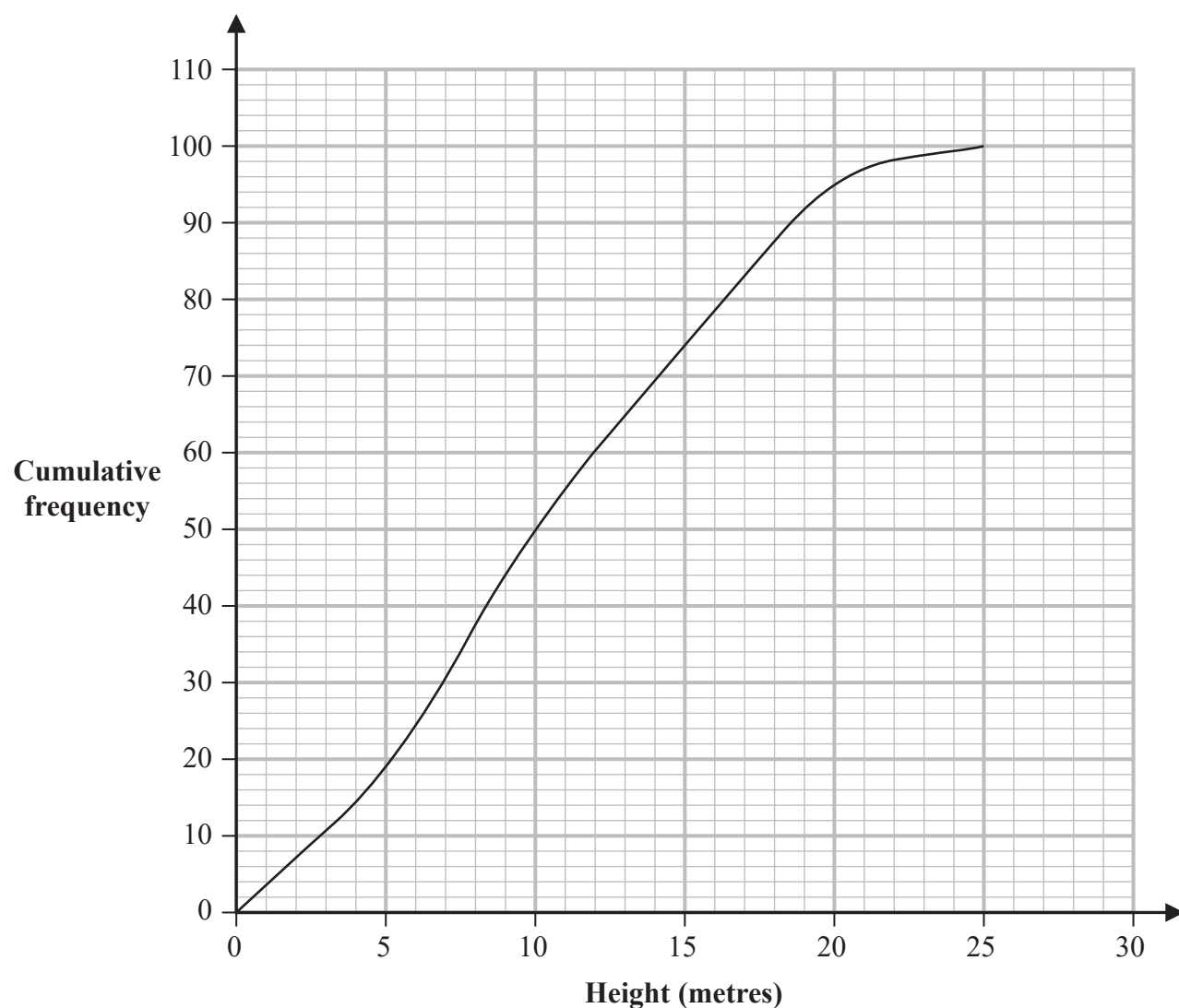
Grace concludes that there is likely to be more play equipment in that part of the playground represented by the squares in the top left hand corner of the choropleth map than elsewhere in the playground.

- (b) Assess the validity of Grace's conclusion with reference to the choropleth map.

(1)

(Total for Question 1 is 3 marks)

- 2 The cumulative frequency diagram gives information about the heights, in metres, of a sample of 100 oak trees in Camden, London.



(Source: *opendata.camden.gov.uk*)

- (a) Using the cumulative frequency diagram, complete the table below for the heights of these 100 trees.

Lower quartile	Median	Upper quartile

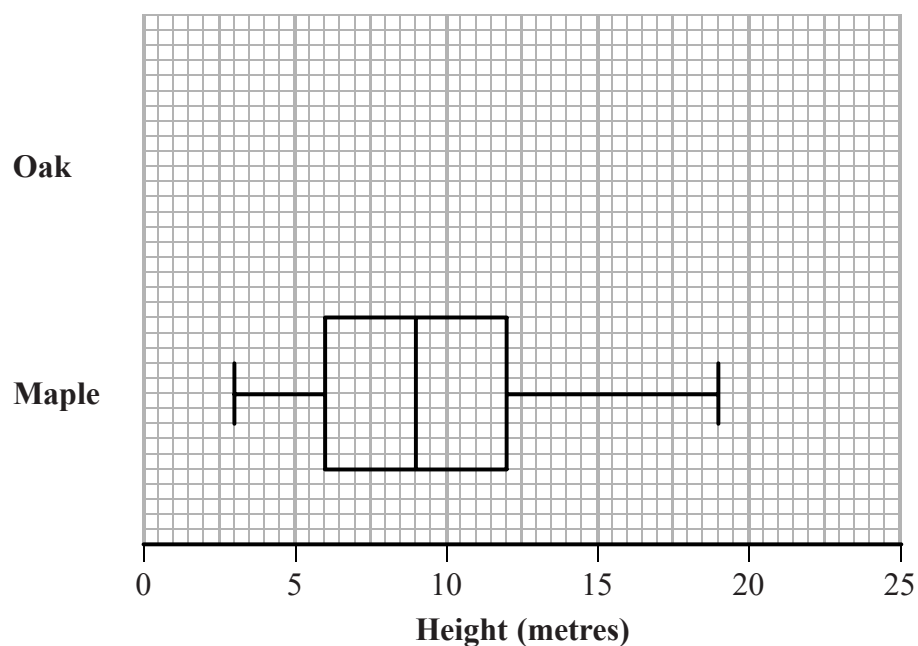
(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

The box plot shows information about the heights, in metres, of a sample of maple trees in Camden, London.



For the sample of oak trees  
the least height is 2.0 m  
the greatest height is 22.3 m

(b) On the grid above, draw a box plot for the heights of the sample of oak trees.

(2)

(c) Compare the two distributions of heights.

Give **three** comparisons and interpret one of these comparisons.

(4)

(Total for Question 2 is 8 marks)

- 3 The table gives information about the numbers of students from different types of schools who applied to Cambridge University in 2016

Type of school	Applications in 2016		
	Gender		Total
	Male	Female	
Maintained	3674	2899	6573
Independent	1510	1268	2778
Other and Overseas	300	312	612
Total	5484	4479	9963

(Source: [www.cam.ac.uk](http://www.cam.ac.uk))

Richard is going to take a sample of 200 of these students stratified by gender.

- (a) Work out how many female students there should be in this sample.

.....  
(2)

- (b) Describe a situation when it would **not** be appropriate to take a sample stratified by gender.

.....  
(1)

Richard could have used a different category for his stratified sample.

- (c) What is this different category?

.....  
(1)

A student is to be chosen at random from the 9963 students.

$F$  is the event that the student chosen is female.

$I$  is the event that the student chosen is from an independent school.

$M$  is the event that the student chosen is from a maintained school.

(d) Explain why the event  $F$  and the event  $I$  are **not** mutually exclusive.

(1)

(e) Find  $P(I \text{ or } M)$ .

(2)

(Total for Question 3 is 7 marks)

4 Diana is a journalist working for a local newspaper.

She is writing a newspaper article about how house prices in the local area have changed.

Diana has house price data for 1996 and for 2016

She plans to include in her article the median house price for 1996 and the median house price for 2016

Mika thinks that Diana should also include in her article the interquartile range of house prices for 1996 and the interquartile range of house prices for 2016

(a) Give one reason why including the interquartile ranges in the article would be appropriate.

(1)

(b) Give one reason why including the interquartile ranges in the article would **not** be appropriate.

(1)

(Total for Question 4 is 2 marks)



- 5 Chris is carrying out an experiment to see if left handed people have a better memory than right handed people. He has two groups of people with the same number of people in each group.

Group A contains left handed people only and group B contains right handed people only.

Each person is given 15 objects to memorise and then they are tested to see how many objects they remember.

The variables for the experiment are shown in the table.

- (a) For each variable, put a tick (✓) in the correct column of the table to show the type of variable.

Variable	Type of variable		
	Explanatory	Response	Extraneous
Age			
Left/Right handedness			
Gender			
Number of objects remembered			

(2)

Chris is not sure whether to work out the mean or the median number of objects each group remembered.

- (b) Describe a situation for which it would be more appropriate to work out the median than the mean.

(1)

Chris worked out these summary statistics for the number of objects people in group A remembered.

Lower quartile = 5

Median = 7

Upper quartile = 8

Sanjit is a member of group A.

He remembered 14 objects.

- (c) Determine whether or not the number of objects Sanjit remembered is an outlier.

(2)

(Total for Question 5 is 5 marks)

- 6 The table gives the crude birth rates and crude death rates, per 1000 people, for two Caribbean islands in the year 2015

Caribbean island	Crude birth rate	Crude death rate
Barbados	11.995	10.661
Saint Lucia	12.239	7.472

(Source: [www.data.worldbank.org](http://www.data.worldbank.org))

Jamil concludes that the crude birth rate and the crude death rate for Saint Lucia show that the population of Saint Lucia increased in 2015

- (a) Explain how the data can be used to support Jamil's conclusion.

(1)

- (b) Give a reason why the data might **not** support Jamil's conclusion.

(1)

In 2015, the total population of Barbados was 284 217

- (c) Using the formula below, work out the number of births in Barbados in 2015

$$\text{crude birth rate} = \frac{\text{number of births} \times 1000}{\text{total population}}$$

(2)

The standard population of Barbados and of Saint Lucia for three different age groups in 2015 is shown in the table.

Age Group	Standard population	
	Barbados	Saint Lucia
60 to 69	103	64
70 to 79	61	39
80 and over	34	22

(d) Give an interpretation of the number 64 in the table.

(1)

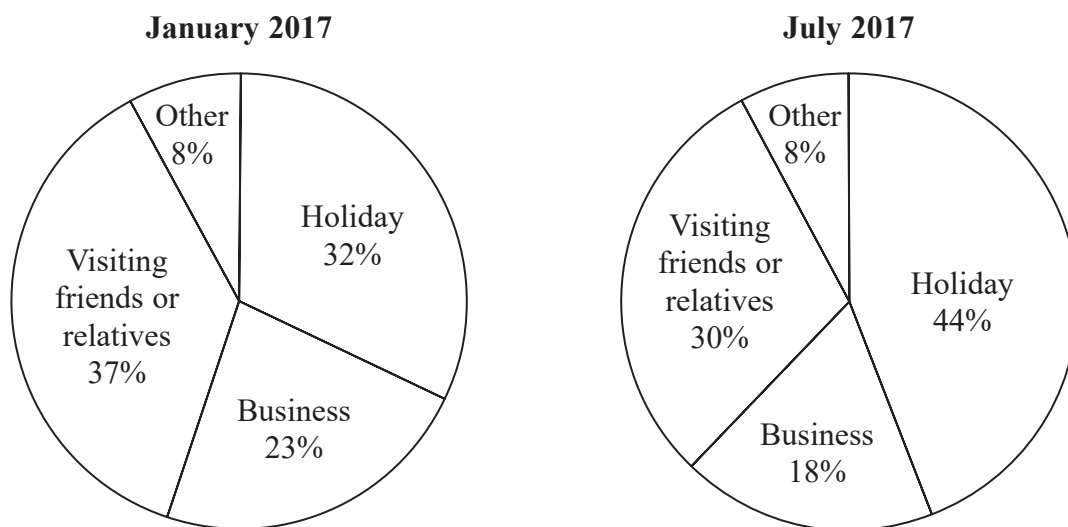
(e) Compare the crude death rate for Barbados with the crude death rate for Saint Lucia.

By referring to the information in the Standard population table, suggest a reason for your comparison.

(2)

(Total for Question 6 is 7 marks)

- 7 The pie charts show information about the numbers of visitors to the UK for the given reasons in January 2017 and in July 2017



(Source: [www.ons.gov.uk](http://www.ons.gov.uk))

- (a) The pie charts do **not** show that there were more visitors to the UK on business in January 2017 than in July 2017  
Explain why.

(1)

The number of visitors to the UK in January 2017 in order to visit friends or relatives is 1 080 733

- (b) Work out the number of visitors to the UK in January 2017 to have a holiday.

(2)

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

Month	Total number of visitors (thousands)
January 2017	2931
July 2017	4020

(c) Explain, giving reasons, how she should do this.

You must refer to the information in the table in your explanation.

(5)

**(Total for Question 7 is 8 marks)**

- 8 The table shows some of the expected percentile weights (kg) for babies up to 6 months old, based on data collected from the World Health Organisation.

Weight (kg) for Boys (B) and Girls (G)										
Age (months)	Percentile									
	9th		25th		50th		75th		91st	
	B	G	B	G	B	G	B	G	B	G
1	3.7	3.45	4.05	3.8	4.45	4.2	4.85	4.6	5.4	5
2	4.6	4.2	5.1	4.7	5.45	5.1	5.9	5.6	6.4	6.1
3	5.45	4.9	5.9	5.4	6.2	5.85	6.9	6.4	7.35	6.9
4	6.1	5.45	6.5	5.9	7	6.4	7.6	7.1	8.15	7.6
5	6.5	5.9	7	6.4	7.5	6.9	8.1	7.5	8.7	8.25
6	6.8	6.2	7.4	6.7	7.9	7.5	8.6	7.9	9.2	8.6

(Source: [www.rcpch.ac.uk](http://www.rcpch.ac.uk))

Antonia says,

“An estimate of the expected 30th percentile for a 3 month old boy is 5.8 kg”

- (a) Explain why this is **not** a good estimate.

(1)

- (b) Compare the expected 9th to the 91st interpercentile range for 5 month old girls with the expected 9th to the 91st interpercentile range for 5 month old boys.  
You must show your working.

(3)

- (c) Give an interpretation of your comparison in part (b).

(1)

(Total for Question 8 is 5 marks)

- 9 According to an internet site, an estimate of the number of reindeer in a region of Ontario is 5000

(Source: [www.ontario.ca](http://www.ontario.ca))

Giovani wants to verify this estimate.

He goes to the region of Ontario, captures a sample of 250 reindeer, attaches a tag to each reindeer and then releases the 250 reindeer back into the same region of Ontario.

Three days later, Giovani returns to the same region of Ontario and catches a sample of 98 reindeer.

He finds that 5 of these reindeer are tagged.

Giovani concludes that this information can be used to verify the estimate of 5000

Discuss the appropriateness of Giovani's method and of his conclusion.

As part of your discussion you should show your calculations and state any assumptions made.

(Total for Question 9 is 5 marks)

- 10 The table shows information about the amount of time that each member of a group of 46 teenagers spent on social media during one day.

Time spent on social media ( $t$ minutes)	Frequency
$0 \leq t < 50$	1
$50 \leq t < 100$	4
$100 \leq t < 150$	8
$150 \leq t < 200$	17
$200 \leq t < 300$	16

- (a) (i) Use linear interpolation to find an estimate of the median time spent on social media by the 46 teenagers.

..... minutes  
(3)

The average person will spend approximately 116 minutes on social media each day.

(Source: [www.socialmediatoday.com](http://www.socialmediatoday.com))

- (ii) Compare the amount of time spent on social media by the 46 teenagers with the amount of time spent on social media by the average person.

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



The table below gives the mean, the standard deviation and the median for the times spent on social media during one day by a sample of sixty year olds.

Mean (minutes)	Standard deviation (minutes)	Median (minutes)
125	25	130

(b) Calculate the skew of the times spent on social media by the sample of sixty year olds.

(2)

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

(2)

(Total for Question 10 is 8 marks)

- 11 Some students at a school walk home and some students go home by bus.

The times taken by the students at the school to walk home have a mean of 25 minutes and a standard deviation of 6 minutes.

The times taken to walk home can be modelled by a normal distribution.

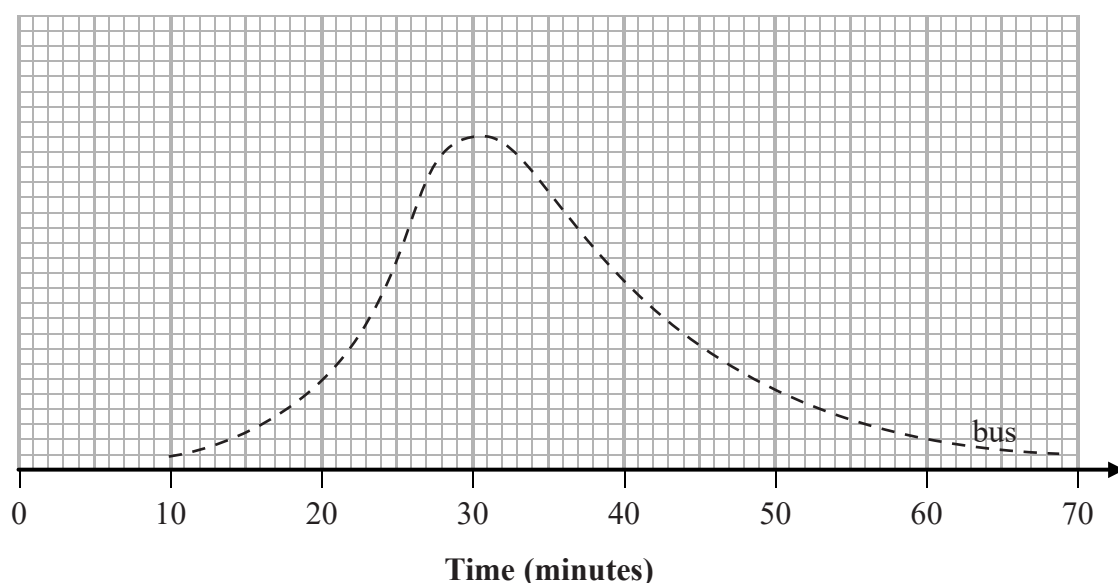
- (a) Shanaya says,

“More than 80% of the students who walk home take between 19 and 37 minutes”

Use statistical calculations to assess Shanaya’s conclusion.

(5)

The diagram below shows a sketch of the distribution of the times taken by students to go home by bus.



- (b) Explain why it is not appropriate to use a normal distribution to model the times taken by students to go home by bus.

(1)

- (c) On the same grid, sketch a diagram showing the distribution of the times taken by students to walk home.

(2)

$A$  is the event that a student walks to school, where  $P(A) = 0.6$

$B$  is the event that a student is driven to school, where  $P(B) = 0.3$

Some students are driven some of the way to school and then they walk the rest of the way to school, where  $P(A \text{ and } B) = 0.15$

A student is picked at random from the students at the school.

- (d) Work out  $P(A \text{ or } B)$ .

(2)

(Total for Question 11 is 10 marks)

- 12 Mr Singh collected data about the number of students in each of three classes in his school and the mean mark of each class in a science test.

The table gives some information about his data.

Class	Number of students in the class	Class mean mark
A	28	63
B	32	72
C	$n$	55

Mr Singh plans to use one of the following two methods to work out the mean mark of **all** the students in the three classes.

**Method 1** Work out the mean of 63, 72 and 55

**Method 2** Given that Mr Singh knows the value of  $n$ , work out the weighted mean mark for the three classes.

- (a) For each of these two methods, assess whether or not the method is an appropriate way to work out the mean mark of all the students in the three classes.

Method 1

.....

.....

Method 2

.....

.....

(2)

The weighted mean mark for the three classes is 64.1 correct to one decimal place.

- (b) Calculate the value of  $n$ .

(2)

The three classes have also taken a mathematics test.

Mr Singh thinks that there is a relationship between the science marks and the mathematics marks.

He draws a scatter diagram for each of the three classes.

For each diagram, he used the science mark,  $x$ , for each student in the class as the explanatory variable and the corresponding mathematics mark,  $m$ , as the response variable.

Mr Singh then used statistical software to find the equation of the regression line for the data in each scatter diagram.

Here are the equations.

Class	Equation of regression line
A	$m = 1.4x - 1$
B	$m = 1.2x + 5$
C	$m = -1.3x + 4$

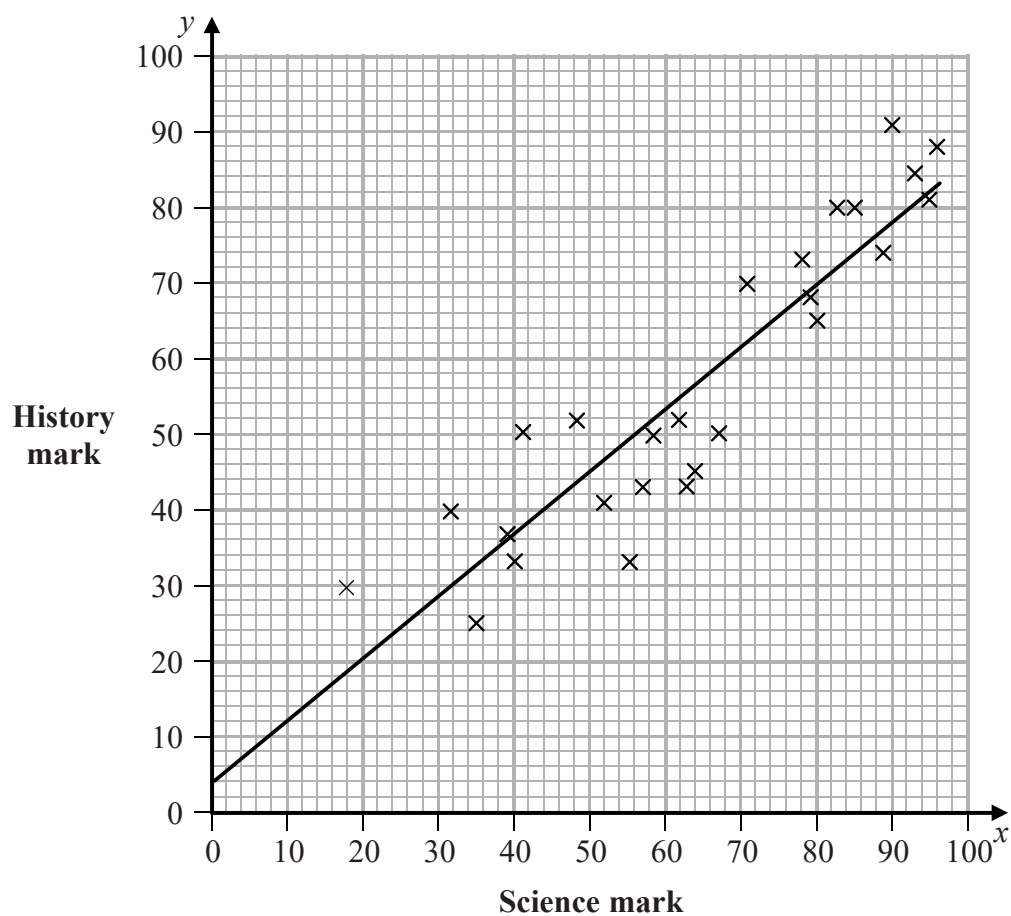
Mr Singh thinks that he has made a mistake with one of the equations.

- (c) Compare the equations of the three regression lines and explain which one is most likely to be incorrect.  
Explain your answer in context.

(2)

Mr Singh decides to compare the science marks,  $x$  and the history marks,  $y$ , of class A.

He uses statistical software to draw this scatter diagram and the regression line.



(d) Find an equation of the regression line of this scatter graph in the form  $y = a + bx$

(3)

(Total for Question 12 is 9 marks)

13 A company wants to investigate the number of sick days its employees have off work.

The company uses a questionnaire.

Here is one of the questions on the questionnaire.

Roll a fair dice.

If you get 1, 2, 3 or 4 tick box A.

If you get 5 or 6 answer this question.

Have you ever taken a sick day off work when you weren't really sick?

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

A ☐

B ☐

(a) Assess the appropriateness of the method the company uses.

(2)

All the company's employees completed the question.

615 ticked box A.

102 ticked box B.

(b) Show that an estimate of the number of employees who ticked box A because they answered yes to the question is 137

(1)

(Total for Question 13 is 3 marks)

TOTAL FOR PAPER IS 80 MARKS

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

DO NOT WRITE IN THIS AREA

**BLANK PAGE**