

INTERNATIONAL BACCALAUREATE
Mathematics: applications and interpretation

MAI

EXERCISES [MAI 4.4]
LINEAR REGRESSION
Compiled by Christos Nikolaidis

A. Paper 1 questions (SHORT)

1. [Maximum mark: 7]

Consider the following data

x	1	2	3	4
y	2	3	7	8

- (a) Find the correlation coefficient r . [1]
- (b) Describe the relation between x and y . [2]
- (c) Find the equation $y = ax + b$ of the regression line for y on x . [2]
- (d) Describe what the coefficient a represents. [1]
- (e) Describe what the constant b represents. [1]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

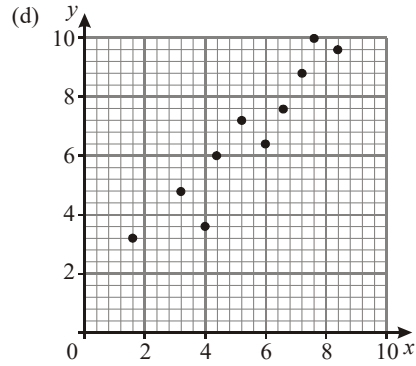
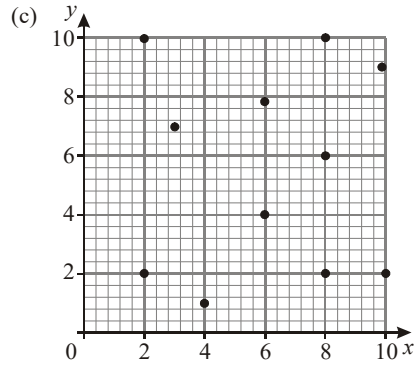
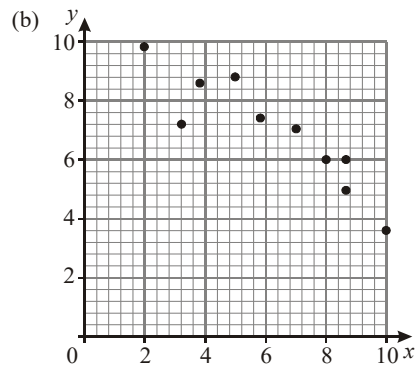
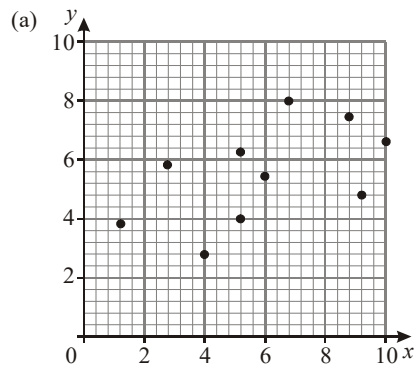
.....

4. [Maximum mark: 4]

Statements I, II, III, IV and V represent descriptions of the correlation between two variables.

- I High positive linear correlation
- II Low positive linear correlation
- III No correlation
- IV Low negative linear correlation
- V High negative linear correlation

Which statement **best** represents the relationship between the two variables shown in each of the scatter diagrams below.



.....

.....

.....

.....

.....

.....

.....

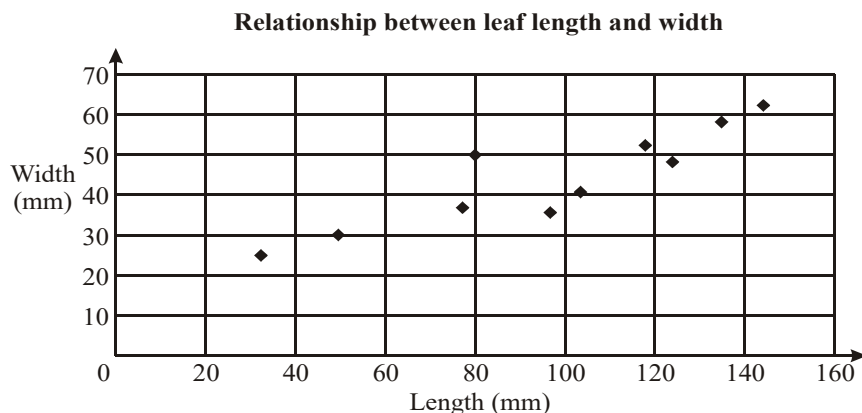
.....

.....

.....

6. [Maximum mark: 4]

The length and width of 10 leaves are shown on the scatter diagram below.



- (a) Plot the point $M(97, 43)$ which represents the mean length and the mean width. [1]
- (b) Draw a suitable line of best fit. [2]
- (c) Write a sentence describing the relationship between leaf length and leaf width for this sample. [1]

.....

.....

7. [Maximum mark: 5]

Ten students were asked for their average grade at the end of their last year of high school and their average grade at the end of their last year at university. The results were put into a table as follows:

Student	High School grade, x	University grade, y
1	90	3.2
2	75	2.6
3	80	3.0
4	70	1.6
5	95	3.8
6	85	3.1
7	90	3.8
8	70	2.8
9	95	3.0
10	85	3.5
Total	835	30.4

- (a) Find the correlation coefficient r . [1]
- (b) Describe the correlation between the high school and the university grades. [2]
- (c) Find the equation of the regression line for y on x . [2]

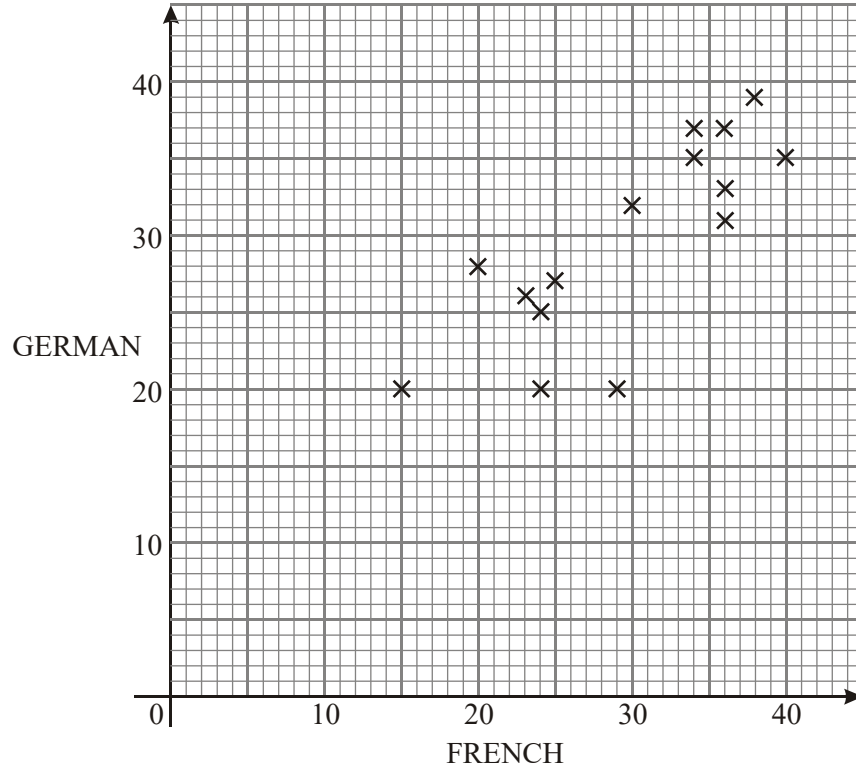
.....

.....

.....

8. [Maximum mark: 5]

The diagram below shows the marks scored by pupils in a French test and a German test. The mean score on the French test is 29 marks and on the German test is 31 marks.



- (a) Describe the relationship between the scores. [1]
- (b) On the graph mark the point M which represents the mean of the distribution. [1]
- (c) Draw a suitable line of best fit. [2]
- (d) Idris scored 32 marks on the French test. Use your graph to estimate the mark Idris scored on the German test. [1]

.....

.....

.....

.....

.....

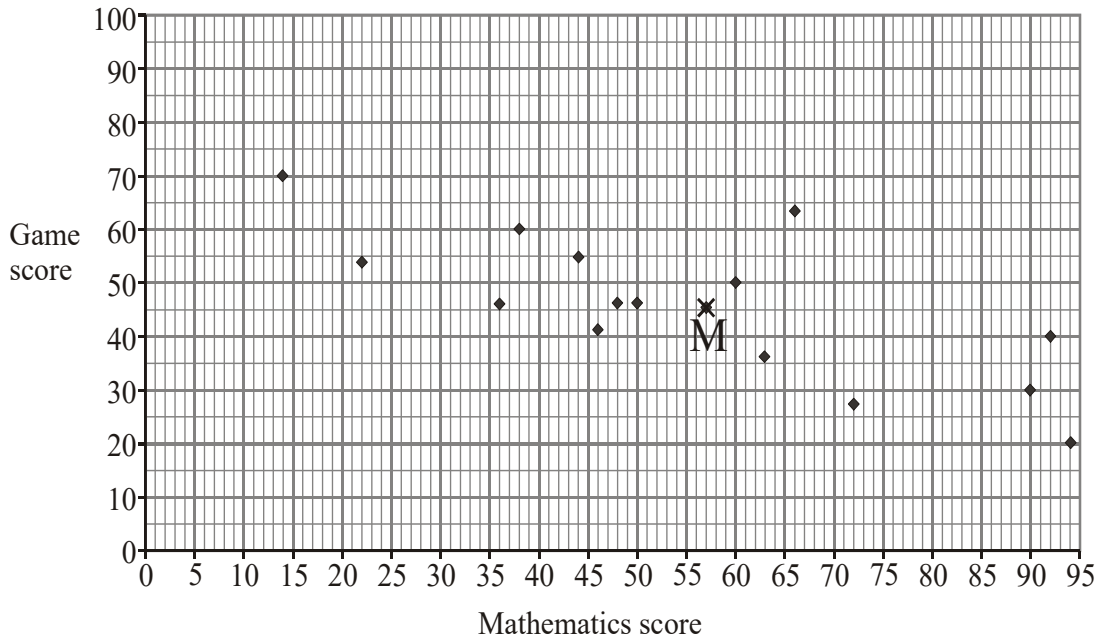
.....

.....

.....

9. [Maximum mark: 5]

A group of 15 students was given a test on mathematics. The students then played a computer game. The diagram below shows the scores on the test and the game.



The point M corresponding to the means has coordinates (56.9, 45.9).

(a) Describe the relationship between the two sets of scores. [1]

(b) On the diagram draw the straight line of best fit given that it passes through the point (0, 69). [2]

Jane took the tests late and scored 45 at mathematics.

(c) Using your graph or otherwise, estimate the score Jane expects on the computer game, giving your answer to the nearest whole number. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

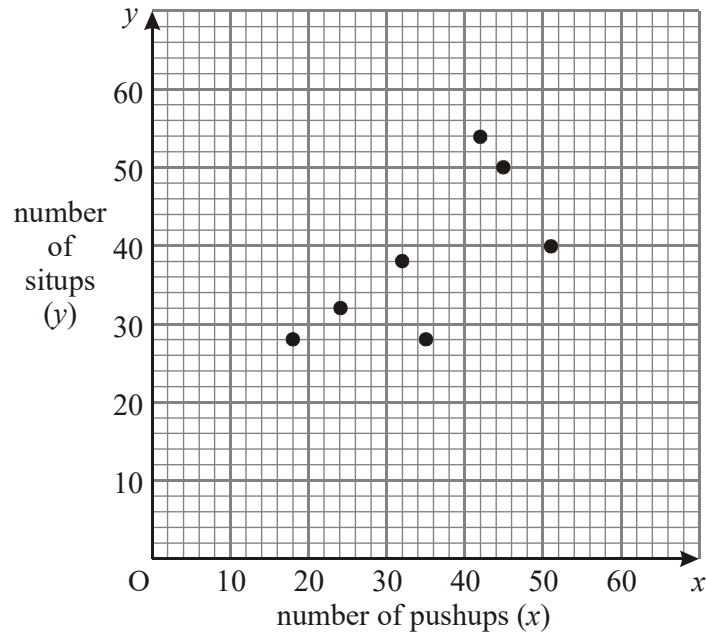
.....

10. [Maximum mark: 6]

Eight students in Mr. O'Neil's Physical Education class did pushups and situps. Their results are shown in the following table.

Student	1	2	3	4	5	6	7	8
number of pushups (x)	24	18	32	51	35	42	45	25
number of situps (y)	32	28	38	40	30	52	48	52

The graph below shows the results for the first seven students.



- (a) Plot the results for the eighth student on the graph. [1]
- (b) Find \bar{x} and \bar{y} , and draw a line of best fit on the graph. [4]
- (c) A student can do 60 pushups. How many situps can the student be expected to do? [1]

.....

.....

.....

.....

.....

.....

.....

.....

11. [Maximum mark: 4]

Ten students were given two tests, one on Mathematics and one on English.

The table shows the results of the tests for each of the ten students.

Student	A	B	C	D	E	F	G	H	I	J
Mathematics (x)	8.6	13.4	12.8	9.3	1.3	9.4	13.1	4.9	13.5	9.6
English (y)	33	51	30	48	12	23	46	18	36	50

(a) Find correct to two decimal places, the correlation coefficient (r). [2]

(b) Use your result from part (a) to comment on the statement:
'Those who do well in Mathematics also do well in English.' [2]

.....

.....

.....

.....

12. [Maximum mark: 6]

The following table contains data for two related variables x and y , and the corresponding ranks for x .

(x)	(y)	Rank x		Rank y	
		original	modified	original	modified
10	100	1	1		
20	110	2	2		
30	130	3	3		
40	120	4	5.5		
40	130	5	5.5		
40	140	6	5.5		
40	150	7	5.5		
50	140	8	8		

(a) Find the **Pearson** correlation coefficient r . [2]

(b) Explain how the modified rank 5.5 for the four equal values of x was derived. [1]

(c) Complete the table above; hence find the **Spearman** rank correlation coefficient. [3]

.....

.....

.....

.....

B. Paper 2 questions (LONG)

13. [Maximum mark: 12]

Consider the following data

x	1	2	3	5	6	7
y	11	13	12	15	17	16

- (a) Find the **Pearson** correlation coefficient r . [1]
- (b) Describe the relation between x and y . [2]
- (c) Find the equation $y = ax + b$ of the regression line for y on x . [2]
- (d) Find the equation $x = cy + d$ of the regression line for x on y . [2]
- (e) Find the **Spearman** rank correlation coefficient r_s [3]
- (f) Describe the difference between the two correlation coefficients. [2]

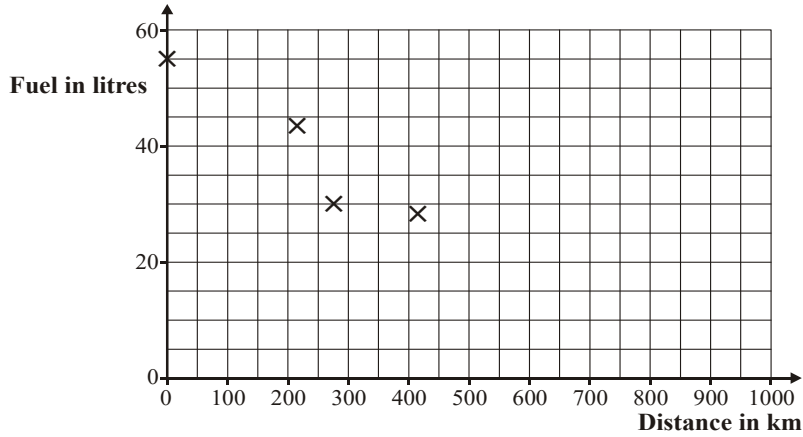
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....
.....

14. [Maximum mark: 12]

The following table gives the amount of fuel in a car's fuel tank, and the number of kilometres travelled after filling the tank.

Distance travelled (km)	0	220	276	500	680	850
Amount of fuel in tank (litres)	55	43	30	24	10	6

(a) On the scatter diagram below, plot the remaining points. [2]



The mean distance travelled is 421 km (\bar{x}), and the mean amount of fuel in the tank is 28 litres (\bar{y}). This point is plotted on the scatter diagram.

- (b) Sketch the line of best fit. [3]
- (c) A car travelled 350km. Use your line above to estimate the amount of fuel left in the tank. [1]
- (d) Find the Pearson correlation coefficient r . [2]
- (e) Find the Spearman rank correlation coefficient r_s . [2]
- (f) Describe the difference between the two correlation coefficients. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

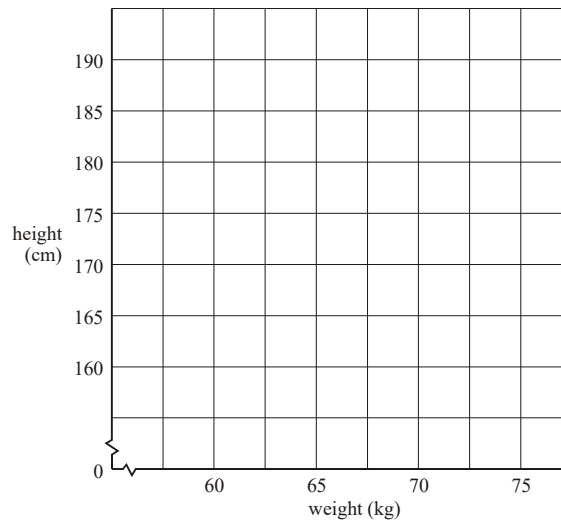
.....

15. [Maximum mark: 12]

The following table gives the heights and weights of five sixteen-year-old boys.

Name	Height	Weight
Blake	182 cm	73 kg
Jorge	173 cm	68 kg
Chin	162 cm	60 kg
Ravi	178 cm	66 kg
Derek	190 cm	75 kg

- (a) Find (i) the mean height; (ii) the mean weight. [2]
 (b) Plot the above data on the grid below and draw the line of best fit.



- (c) Find the Pearson correlation coefficient r . [4]
 (d) Find the Spearman rank correlation coefficient r_s . [2]
 (e) Explain why r_s is not 1. [3]

.....

16. [Maximum mark: 10]

It is decided to take a random sample of 10 students to see if there is any linear relationship between height and shoe size. The results are given in the table below.

Height (cm) (<i>x</i>)	Shoe size (<i>y</i>)
175	8
160	9
180	8
155	7
178	10
159	8
166	9
185	11
189	10
173	9

- (a) Write down the equation of the regression line of shoe size (*y*) on height (*x*), giving your answer in the form $y = mx + c$. [2]
- (b) State an interpretation for the coefficient *m* of the regression line in (a). [2]
- (c) A student is 162 cm in height
 - (i) Use your equation in part (a) to predict the shoe size of the student.
 - (ii) Is this an interpolation or extrapolation? Explain. [3]
- (d) Write down the correlation coefficient. [1]
- (e) Describe the correlation between height and shoe size. [2]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

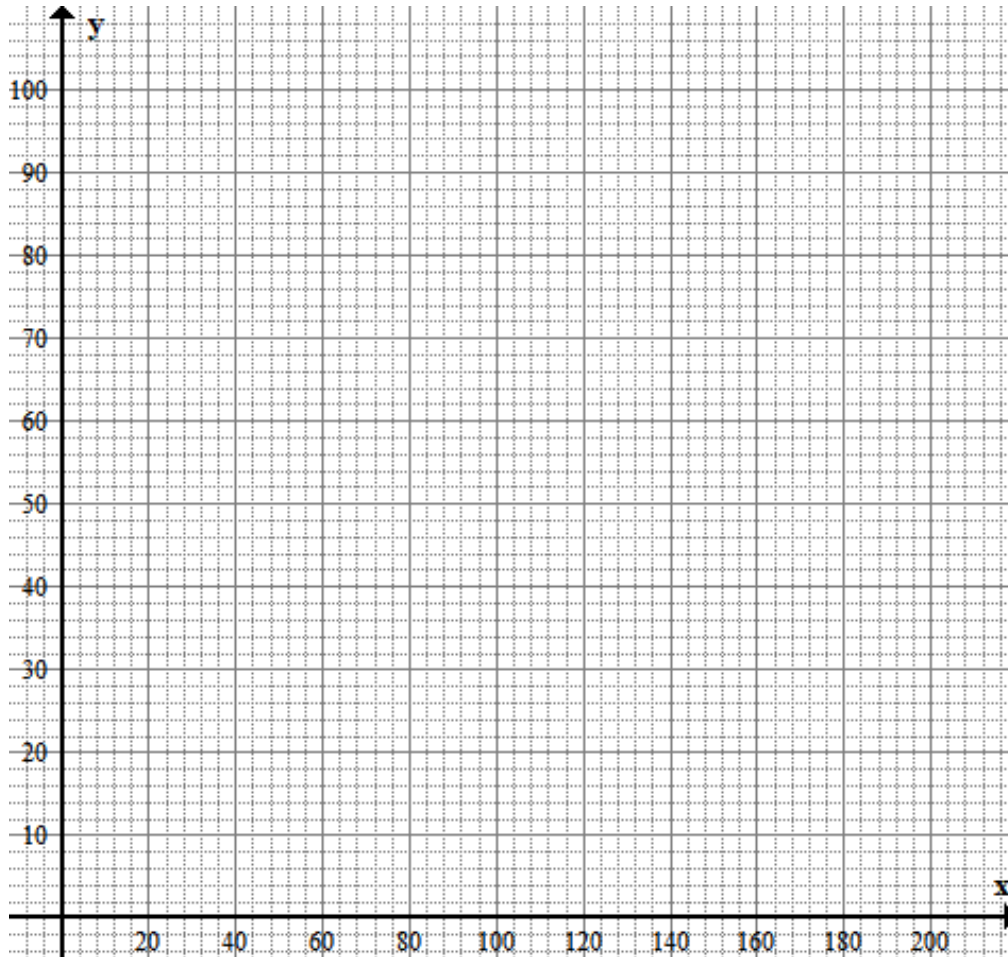
17. [Maximum mark: 12]

The heights and weights of 10 students selected at random are shown below.

Student	1	2	3	4	5	6	7	8	9	10
Height x cm	155	161	173	150	182	165	170	185	175	145
Weight y kg	50	75	80	46	81	79	64	92	74	108

(a) Plot this information on a scatter graph.

[4]



(b) Calculate the mean height and the mean weight

[2]

(c) (i) Find the equation of the line of best fit.

(ii) Draw the line of best fit on your graph.

[3]

(d) Use your line to estimate

(i) the weight of a student of height 190 cm;

(ii) the height of a student of weight 72 kg.

[2]

(e) It is decided to remove the data for student number 10 from all calculations.

Explain **briefly** what effect this will have on the line of best fit.

[1]

18. [Maximum mark: 12]

The *Type Fast* secretarial training agency has a new computer software spreadsheet package. The agency investigates the number of hours it takes people of varying ages to reach a level of proficiency using this package. Fifteen individuals are tested and the results are summarised in the table below.

Age (<i>x</i>)	32	40	21	45	24	19	17	21	27	54	33	37	23	45	18
Time (in h) (<i>y</i>)	10	12	8	15	7	8	6	9	11	16	10	13	9	17	5

- (a) (i) Find the correlation coefficient r for this data.
 (ii) What does the value of the correlation coefficient suggest about the relationship between the two variables? [2]
- (b) Write down the equation of the regression line for y on x in the form $y = ax + b$. [2]
- (c) Use your equation for the regression line to predict
 (i) the time that it would take a 30 year old person to reach proficiency, giving your answer correct to the nearest hour;
 (ii) the age of a person who would take 8 hours to reach proficiency, giving your answer correct to the nearest year. [4]
- (d) Find an estimation for the age of the person in question (c)(ii) by using the regression line of x on y . [4]

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

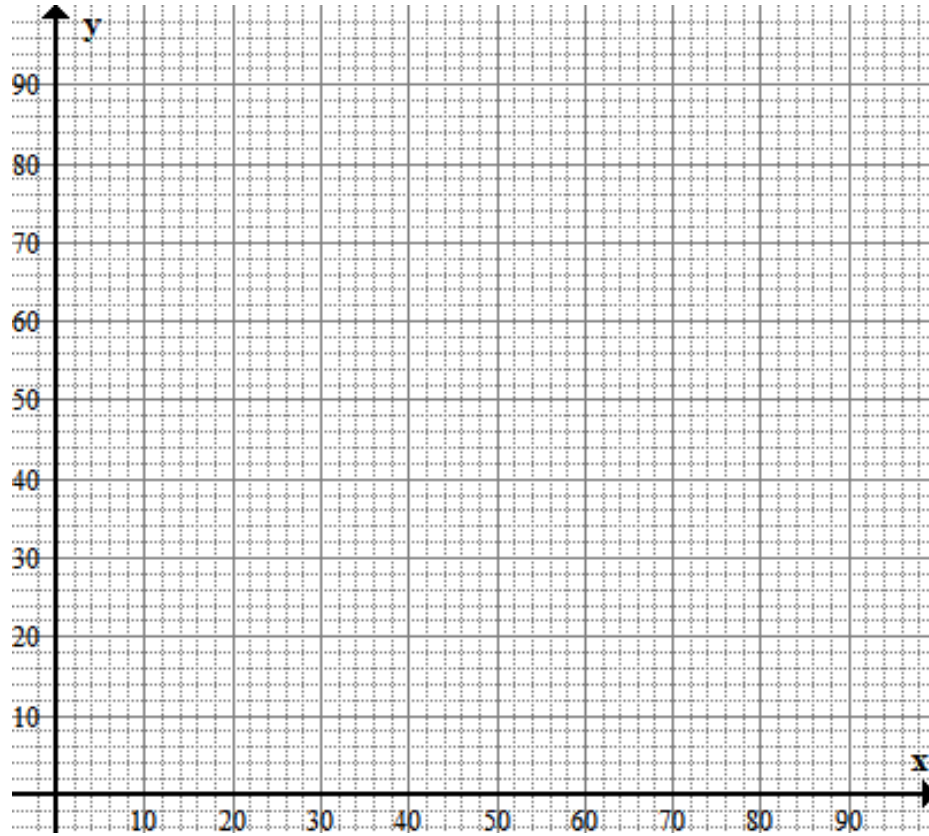
20. [Maximum mark: 12]

The following are the results of a survey of the scores of 10 people on both a mathematics (x) and a science (y) aptitude test:

Student	Mathematics (x)	Science (y)
1	90	85
2	38	60
3	58	78
4	85	70
5	73	65
6	82	71
7	56	80
8	73	90
9	95	96
10	80	85

(a) Plot this information on a scatter graph.

[4]



(b) Find and plot the point $M(\bar{x}, \bar{y})$ on the graph.

[2]

(c) Find the equation of the regression line of y on x in the form $y = ax + b$.

[2]

(d) Graph this line on the above graph.

[2]

(e) Given that a student receives an 88 on the mathematics test, what would you expect this student's science score to be? Show how you arrived at your result.

[2]

21. [Maximum mark: 11]

The following are the results of a survey of the scores of 10 people on both a mathematics (x) and a science (y) aptitude test:

Student	Mathematics (x)	Science (y)
1	90	85
2	38	60
3	58	78
4	90	70
5	73	65
6	90	71
7	56	80
8	73	90
9	95	96
10	80	85

- (a) Find the equation of the regression line of y on x . [2]
- (b) Find the equation of the regression line of x on y . [2]
- (c) Find the **Pearson** correlation coefficient r . [2]

The table below shows the data for x in increasing order and the corresponding ranks.

Mathematics (x)	Science (y)	Rank x	Rank y
38	60	1	
56		2	
58		3	
73		4	
73		5	
80		6	
90		7	
90		8	
90		9	
95		10	

- (d) (i) Complete the data for y and the corresponding ranks in the table above.
- (ii) **Hence**, find the **Spearman** rank correlation coefficient r_s . [5]

.....

.....

.....

.....

.....

.....

.....