IB Questionbank Mathematical Studies 3rd edition

Bivariate data

179 min
172 marks

1. The heat output in thermal units from burning 1 kg of wood changes according to the wood's percentage moisture content. The moisture content and heat output of 10 blocks of the same type of wood each weighing 1 kg were measured. These are shown in the table.

Moisture content % (x)	8	15	22	30	34	45	50	60	74	82
Heat output (y)	80	77	74	69	68	61	61	55	50	45

(a) Draw a scatter diagram to show the above data. Use a scale of 2 cm to represent 10 % on the x-axis and a scale of 2 cm to represent 10 thermal units on the y-axis.

(4)

(b) Write down

(i) the mean percentage moisture content, $\bar{x}$;

(ii) the mean heat output, $\bar{y}$.

(2)

(c) Plot the point $(\bar{x}, \bar{y})$ on your scatter diagram and label this point M.

(2)

(d) Write down the product-moment correlation coefficient, r.

(2)

The equation of the regression line y on x is $y = -0.470x + 83.7$.

(e) Draw the regression line y on x on your scatter diagram.

(2)

(f) Estimate the heat output in thermal units of a 1 kg block of wood that has 25 % moisture content.

(2)

(g) State, with a reason, whether it is appropriate to use the regression line y on x to estimate the heat output in part (f).

(2)

(Total 16 marks)

2. At the end of the year, only seven of the female Science students sat examinations in Science and French.

The marks for these seven students are shown in the following table.

Science (*S*)	23	51	56	62	12	73	72
French (*F*)	65	45	45	40	70	36	30

(a) Using a scale of 2 cm to represent 10 marks for each axis, draw a labelled scatter diagram for this data.

(4)

(b) Use your graphic display calculator to find

(i) $\overline{S}$, the mean of S;

(ii) $\overline{F}$, the mean of F.

(2)

(c) Plot the point M($\overline{S}, \overline{F}$) on your scatter diagram.

(1)

(d) Use your graphic display calculator to find the equation of the regression line of F on S.

(2)

(e) Draw the regression line on your scatter diagram.

(2)

Carletta's mark on the Science examination was 44. She did not sit the French examination.

(f) Estimate Carletta's mark for the French examination.

(2)

Monique's mark on the Science examination was 85. She did not sit the French examination. Her French teacher wants to use the regression line to estimate Monique's mark.

(g) State whether the mark obtained from the regression line for Monique's French examination is reliable. Justify your answer.

(2)

(Total 15 marks)

3. In an environmental study of plant diversity around a lake, a biologist collected data about the number of different plant species (y) that were growing at different distances (x) in metres from the lake shore.

Distance (x)	2	5	8	10	13	17	23	35	40
Plant species (y)	35	34	30	29	24	19	15	13	8

(a) Draw a scatter diagram to show the data. Use a scale of 2 cm to represent 10 metres on the x-axis and 2 cm to represent 10 plant species on the y-axis.

(4)

(b) Using your scatter diagram, describe the correlation between the number of different plant species and the distance from the lake shore.

(1)

(c) Use your graphic display calculator to write down

(i) $\bar{x}$, the mean of the distances from the lake shore;

(ii) $\bar{y}$, the mean number of plant species.

(2)

(d) Plot the point $(\bar{x}, \bar{y})$ on your scatter diagram. **Label this point M.**

(2)

(e) Write down the equation of the regression line y on x for the above data.

(2)

(f) Draw the regression line y on x on your scatter diagram.

(2)

(g) Estimate the number of plant species growing 30 metres from the lake shore.

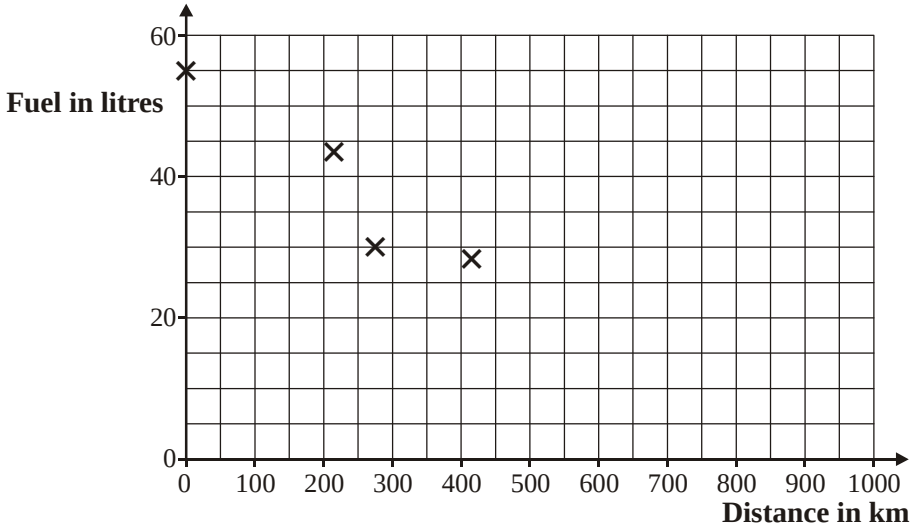
(2)

(Total 15 marks)

4. 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.



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.

A car travelled 350 km.

(c) Use your line of best fit to estimate the amount of fuel left in the tank.

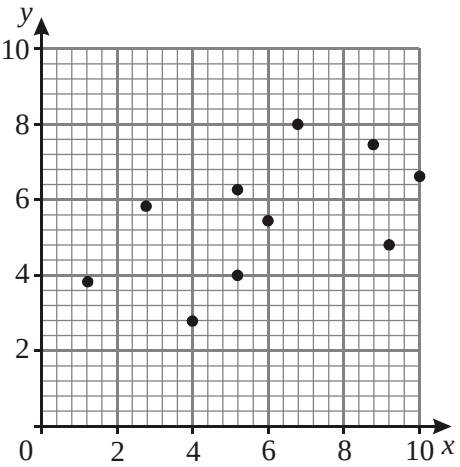
(Total 6 marks)

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

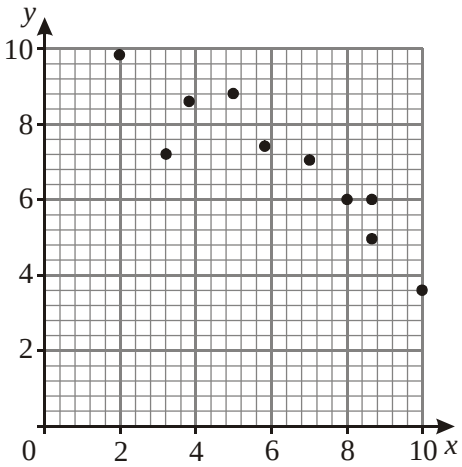
I High positive linear correlation
II Low positive linear correlation
III No correlation
IV Low negative linear correlation
V High negative linear correlation

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

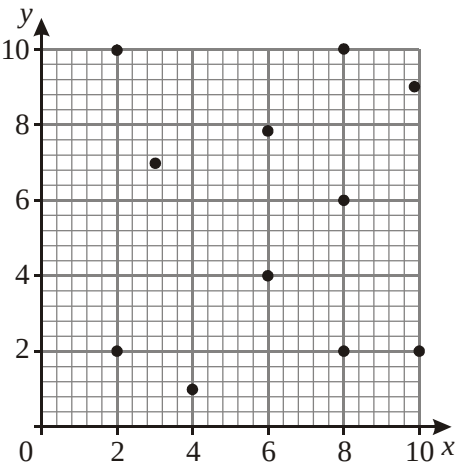
(a)



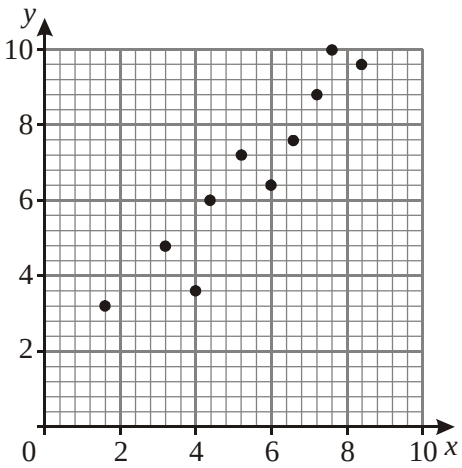
(b)



(c)



(d)



Answers:

(a) ……

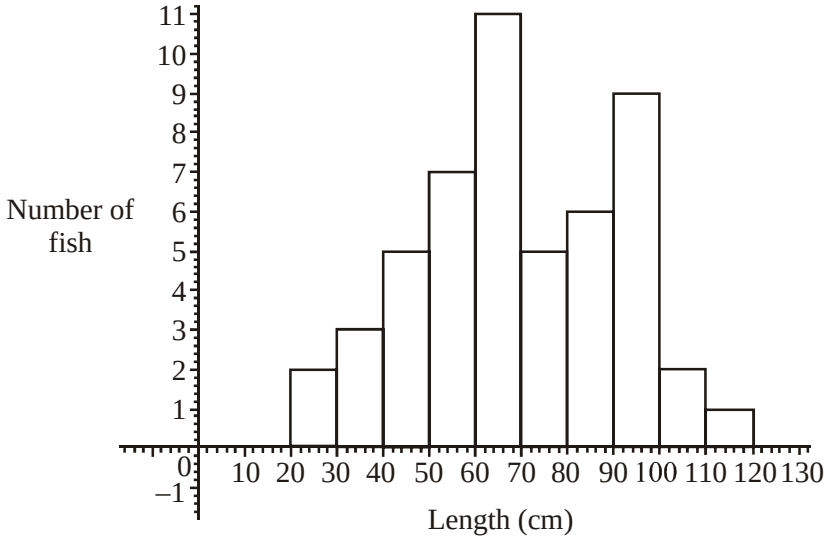
(b) ……

(c) ……

(d) ……

(Total 4 marks)

6. The figure below shows the lengths in centimetres of fish found in the net of a small trawler.



(a) Find the total number of fish in the net.

(2)

(b) Find (i) the modal length interval;

(ii) the interval containing the median length;

(iii) an estimate of the mean length.

(5)

(c) (i) Write down an estimate for the standard deviation of the lengths.

(ii) How many fish (if any) have length **greater than** three standard deviations **above** the mean?

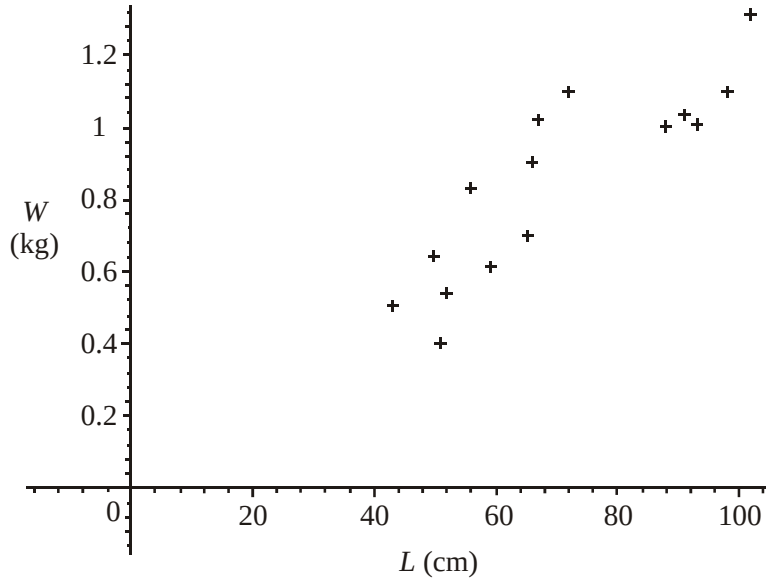
(3)

The fishing company must pay a fine if more than 10% of the catch have lengths less than 40cm.

(d) Do a calculation to decide whether the company is fined.

(2)

A sample of 15 of the fish was weighed. The weight, W was plotted against length, L as shown below.



(e) Exactly **two** of the following statements about the plot could be correct. Identify the two correct statements.

(2)

Note: You do **not** need to enter data in a GDC **or** to calculate r exactly.

(i) The value of r, the correlation coefficient, is approximately 0.871.

(ii) There is an exact linear relation between W and L.

(iii) The line of regression of W on L has equation $W = 0.012L + 0.008$.

(iv) There is negative correlation between the length and weight.

(v) The value of r, the correlation coefficient, is approximately 0.998.

(vi) The line of regression of W on L has equation $W = 63.5L + 16.5$.

(Total 14 marks)

7. Alex and Kris are riding their bicycles together along a bicycle trail and note the following distance markers at the given times.

Time (t hours)	1	2	3	4	5	6	7
Distance (d km)	57	65	72	81	89	97	107

(a) Draw a scatter diagram of the data. Use 1 cm to represent 1 hour and 1 cm to represent 10 km.

(3)

(b) Write down for this set of data

(i) the mean time, $\bar{t}$;

(ii) the mean distance, $\bar{d}$.

(2)

(c) Mark and label the point $M(\bar{t}, \bar{d})$ on your scatter diagram.

(2)

(d) Draw the line of best fit on your scatter diagram.

(2)

(e) **Using your graph**, estimate the time when Alex and Kris pass the 85 km distance marker. Give your answer correct to **one decimal place**.

(2)

(f) Write down the equation of the regression line for the data given.

(2)

(g) (i) **Using your equation** calculate the distance marker passed by the cyclists at 10.3 hours.

(ii) Is this estimate of the distance reliable? Give a reason for your answer.

(4)

(Total 17 marks)

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

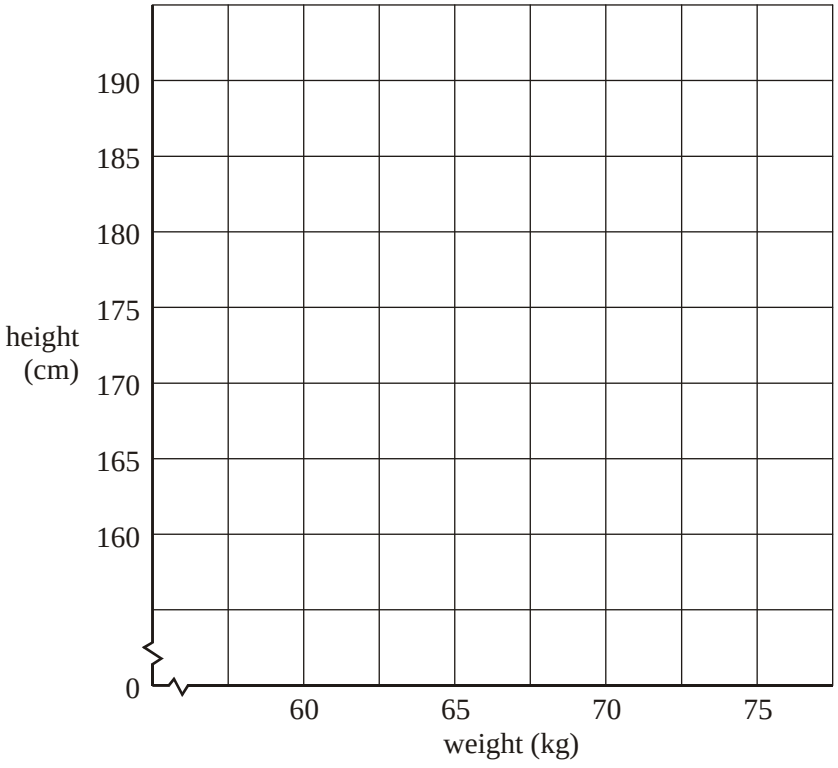
Name	Height	Weight
Blake	182 cm	73 kg
Jorge	173 cm	68 kg
Chin	162 cm	60 kg
Ravi	178 cm	66 kg
Derek	190 cm	75 kg

(a) Find

(i) the mean height;

(ii) the mean weight.

(b) Plot the above data on the grid below and draw the line of best fit.



(Total 4 marks)

9. The following table shows the cost in AUD of seven paperback books chosen at random, together with the number of pages in each book.

Book	1	2	3	4	5	6	7
Number of pages (x)	50	120	200	330	400	450	630
Cost (y AUD)	6.00	5.40	7.20	4.60	7.60	5.80	5.20

(a) Plot these pairs of values on a scatter diagram. Use a scale of 1 cm to represent 50 pages on the horizontal axis and 1 cm to represent 1 AUD on the vertical axis.

(3)

(b) Write down the linear correlation coefficient, r, for the data.

(2)

(c) Stephen wishes to buy a paperback book that has 350 pages in it. He plans to draw a line of best fit to determine the price. State whether or not this is an appropriate method in this case and justify your answer.

(2)

(Total 7 marks)

10. A shop keeper recorded daily sales s of ice cream along with the daily maximum temperature t°C. The results for one week are shown below.

t	29	31	34	23	19	20	27
s	104	92	112	48	56	72	66

(a) Write down the equation of the regression line for s on t.

(3)

(b) Use your equation to predict the ice cream sales on a day when the maximum temperature is 24 °C. Give your answer correct to the nearest whole number.

(3)

(Total 6 marks)

11. In a mountain region there appears to be a relationship between the number of trees growing in the region and the depth of snow in winter. A set of 10 areas was chosen, and in each area the number of trees was counted and the depth of snow measured.
The results are given in the table below.

Number of trees (x)	**Depth of snow in cm (y)**
45	30
75	50
66	40
27	25
44	30
28	5
60	35
35	20
73	45
47	25

(a) Use your graphic display calculator to find

(i) the mean number of trees;

(ii) the standard deviation of the number of trees;

(iii) the mean depth of snow;

(iv) the standard deviation of the depth of snow.

(4)

The covariance, $S_{xy} = 188.5$.

(b) Write down the product-moment correlation coefficient, r.

(2)

(c) Write down the equation of the regression line of y on x.

(2)

(d) If the number of trees in an area is 55, estimate the depth of snow.

(2)

(e) (i) Use the equation of the regression line to estimate the depth of snow in an area with 100 trees.

(ii) Decide whether the answer in (e)(i) is a valid estimate of the depth of snow in the area. Give a reason for your answer.

(3)
(Total 13 marks)

12. Tania wishes to see whether there is any correlation between a person's age and the number of objects on a tray which could be remembered after looking at them for a certain time.

She obtains the following table of results.

Age (x years)	15	21	36	40	44	55
Number of objects remembered (y)	17	20	15	16	17	12

(a) Use your graphic display calculator to find the equation of the regression line of y on x.

(2)

(b) Use your equation to estimate the number of objects remembered by a person aged 28 years.

(1)

(c) Use your graphic display calculator to find the correlation coefficient r.

(1)

(d) Comment on your value for r.

(2)
(Total 6 marks)

13. The number of bottles of water sold at a railway station on each day is given in the following table.

Day	0	1	2	3	4	5	6	7	8	9	10	11	12
Temperature (T°)	21	20.7	20	19	18	17.3	17	17.3	18	19	20	20.7	21
Number of bottles sold (n)	150	141	126	125	98	101	93	99	116	121	119	134	141

(a) Write down

(i) the mean temperature;

(ii) the standard deviation of the temperatures.

(2)

(b) Write down the correlation coefficient, r, for the variables n and T.

(1)

(c) Comment on your value for r.

(2)

(d) The equation of the line of regression for n on T is $n = dT - 100$.

(i) Write down the value of d.

(ii) Estimate how many bottles of water will be sold when the temperature is 19.6°.

(2)

(e) On a day when the temperature was 36° Peter calculates that 314 bottles would be sold. Give one reason why his answer might be unreliable.

(1)

(Total 8 marks)

14. A shopkeeper wanted to investigate whether or not there was a correlation between the prices of food 10 years ago in 1992, with their prices today. He chose 8 everyday items and the prices are given in the table below.

	sugar	milk	eggs	rolls	tea bags	coffee	potatoes	flour
1992 price	\$1.44	\$0.80	\$2.16	\$1.80	\$0.92	\$3.16	\$1.32	\$1.12
2002 price	\$2.20	\$1.04	\$2.64	\$3.00	\$1.32	\$2.28	\$1.92	\$1.44

(a) Calculate the mean and the standard deviation of the prices

(i) in 1992;

(ii) in 2002.

(4)

(b) (i) Given that $s_{xy} = 0.3104$, calculate the correlation coefficient.

(ii) Comment on the relationship between the prices.

(4)

(c) Find the equation of the line of the best fit in the form $y = mx + c$.

(3)

(d) What would you expect to pay now for an item costing \$2.60 in 1992?

(1)

(e) Which item would you omit to increase the correlation coefficient?

(2)

(Total 14 marks)

15. 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) Given that $s_x = 8.96$, $s_y = 0.610$ and $s_{xy} = 4.16$, find the correlation coefficient r, giving your answer to two decimal places.

(2)

(b) Describe the correlation between the high school grades and the university grades.

(2)

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

(2)

(Total 6 marks)

16. The heights and weights of 10 students selected at random are shown in the table 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. Use a scale of 1 cm to represent 20 cm on the x-axis and 1 cm to represent 10 kg on the y-axis.

(4)

(b) Calculate the mean height.

(1)

(c) Calculate the mean weight.

(1)

(d) It is given that $S_{xy} = 44.31$.

(i) By first calculating the standard deviation of the heights, correct to two decimal places, show that the gradient of the line of regression of y on x is 0.276.

(ii) Calculate the equation of the line of best fit.

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

(6)

(e) 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)

(f) 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)
(Total 15 marks)

17. The marks obtained by 8 candidates in Physics and Chemistry tests are given below.

Physics (x)	6	8	10	11	10	5	4	12
Chemistry (y)	8	11	14	13	11	7	5	15

(a) Write down the product moment correlation coefficient, r.

(1)

(b) Write down, in the form $y = mx + c$, the equation of the regression line y on x for the 8 candidates.

(2)

A ninth candidate obtained a score of 7 in the Physics test but was absent for the Chemistry test.

(c) Use your answer to (b) to estimate the score he would have obtained on the Chemistry test.

(2)

(d) Give a reason why it is valid to use this regression line to estimate the score on the Chemistry test.

(1)
(Total 6 marks)