## IB MATHEMATICS SL

## Topic: Bivariate Data

NAME: $\qquad$ DATE: $\qquad$

1. 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.
(b) Find (i) the modal length interval;
(ii) the interval containing the median length;
(iii) an estimate of the mean length.
(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?

The fishing company must pay a fine if more than $10 \%$ of the catch have lengths less than 40 cm .
(d) Do a calculation to decide whether the company is fined.

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.

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 $W=0.012 L+0.008$.
(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 $W=63.5 L+16.5$.
(Total 14 marks)
2. The length and width of 10 leaves are shown on the scatter diagram below.

## Relationship between leaf length and width


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

## Working:

Answer:
(c) $\qquad$
3. 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 marks scored in the two tests.
(b) On the graph mark the point M which represents the mean of the distribution.
(c) Draw a suitable line of best fit.
(d) Idris scored 32 marks on the French test. Use your graph to estimate the mark Idris scored on the German test.

## Working:

Answers:
(a)
(d) $\qquad$
4. 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.


Working:

Answers:
(a) (i)
(ii)
$\qquad$
$\qquad$
5. 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 <br> $(x)$ | 32 | 40 | 21 | 45 | 24 | 19 | 17 | 21 | 27 | 54 | 33 | 37 | 23 | 45 | 18 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Time <br> (in hours) <br> $(y)$ | 10 | 12 | 8 | 15 | 7 | 8 | 6 | 9 | 11 | 16 | $t$ | 13 | 9 | 17 | 5 |

(a) (i) Given that $S_{y}=3.5$ and $S_{x y}=36.7$, calculate the product-moment correlation coefficient $r$ for this data.
(ii) What does the value of the correlation coefficient suggest about the relationship between the two variables?
(b) Given that the mean time taken was 10.6 hours, write the equation of the regression line for $y$ on $x$ in the form $y=a x+b$.
(c) Use your equation for the regression line to predict
(i) the time that it would take a 33 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.
(Total 12 marks)
6. 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 <br> $x \mathrm{~cm}$ | 155 | 161 | 173 | 150 | 182 | 165 | 170 | 185 | 175 | 145 |
| Weight <br> $y \mathrm{~kg}$ | 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.
(b) Calculate the mean height.
(c) Calculate the mean weight.
(d) It is given that $S_{x y}=44.31$.
(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.
(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 .
(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.
7. The scatter diagram below shows the relationship between the number of vehicles per thousand of population and the number of people killed in road accidents over an eight year period in Calmville.

Relationship between number of vehicles and people
killed in road accidents in Calmville


Let $x$ be the number of vehicles per thousand and $y$ be the number of people killed. The following information is known.

$$
\bar{x}=270, \quad \bar{y}=650 \quad s_{x}=22.3 \quad s_{y}=96.2, \quad s_{x y}=2077.75
$$

(a) (i) Calculate the product-moment correlation coefficient $(r)$.
(ii) Explain clearly the statistical relationship between the variables $x$ and $y$
(b) Write the equation of the regression line of $y$ on $x$, expressing it in the form $y=m x+c$ (where $m$ and $c$ are given correct to 3 significant figures).
(c) Use your equation in part (b) to answer the following questions.
(i) There were 250 vehicles per 1000 of population. Find the number of people killed.
(ii) Explain why it is not a good idea to use the regression line to estimate the number of people killed when the number of vehicles is 150 per thousand.

