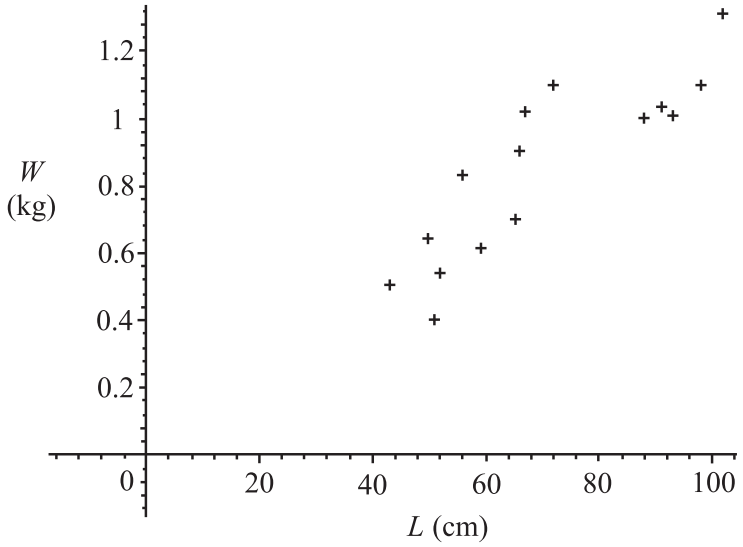
Bivariate stats 3

1)

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 marks]*

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 $W = 63.5L + 16.5$.

2)

Marty asked some of his classmates to rate their level of stress out of 10, with 10 being very high. He also asked them to measure the number of minutes it took them to get from home to school. A random selection of his results is listed below.

Travel time (x)	13	24	22	18	36	16	14	20	6	12
Stress rating (y)	3	7	5	4	8	8	4	8	2	6

(i) Write down the value of the (linear) coefficient of correlation for this information. *[1 mark]*

(ii) Explain what a positive value for the coefficient of correlation indicates. *[1 mark]*

(iii) Write down the linear regression equation of y on x in the form $y = ax + b$. *[2 marks]*

(iv) Use your equation in part (iii) to determine the stress rating for a student who takes three quarters of an hour to travel to school. *[2 marks]*

(v) Can your answer in part (iv) be considered reliable? Give a reason for your answer. *[2 marks]*

3) In an experiment a vertical spring was fixed at its upper end. It was stretched by hanging different weights on its lower end. The length of the spring was then measured. The following readings were obtained.

Load (kg) x	0	1	2	3	4	5	6	7	8
Length (cm) y	23.5	25	26.5	27	28.5	31.5	34.5	36	37.5

(a) Plot these pairs of values on a scatter diagram taking 1 cm to represent 1 kg on the horizontal axis and 1 cm to represent 2 cm on the vertical axis. *[4 marks]*

(b) (i) Write down the mean value of the load ($\overline{x}$). *[1 mark]*

(ii) Write down the standard deviation of the load. *[1 mark]*

(iii) Write down the mean value of the length ($\overline{y}$). *[1 mark]*

(iv) Write down the standard deviation of the length. *[1 mark]*

(c) Plot the mean point ($\overline{x}$, $\overline{y}$) on the scatter diagram. Name it L. *[1 mark]*

It is given that the covariance S_{xy} is 12.17.

(d) (i) Write down the correlation coefficient, ***r***, for these readings. *[1 mark]*

(ii) Comment on this result. *[2 marks]*

(e) Find the equation of the regression line of y on x. *[2 marks]*

(f) Draw the line of regression on the scatter diagram. *[2 marks]*

(g) (i) Using your diagram or otherwise, estimate the length of the spring when a load of 5.4 kg is applied. *[1 mark]*

(ii) Malcolm uses the equation to claim that a weight of 30 kg would result in a length of 62.8 cm. Comment on his claim. *[1 mark]*

4) It is thought that the breaststroke time for 200 m depends on the length of the arm of the swimmer.
Eight students swim 200 m breaststroke. Their times (y) in seconds and arm lengths (x) in cm are shown in the table below.

	1	2	3	4	5	6	7	8
Length of arm, x cm	79	74	72	70	77	73	64	69
Breaststroke, y seconds	135.1	135.7	139.3	141.0	132.8	137.0	152.9	144.0

(a) Calculate the mean and standard deviation of x and y. *[4 marks]*

(b) Given that $s_{xy} = -24.82$, calculate the correlation coefficient, r. *[2 marks]*

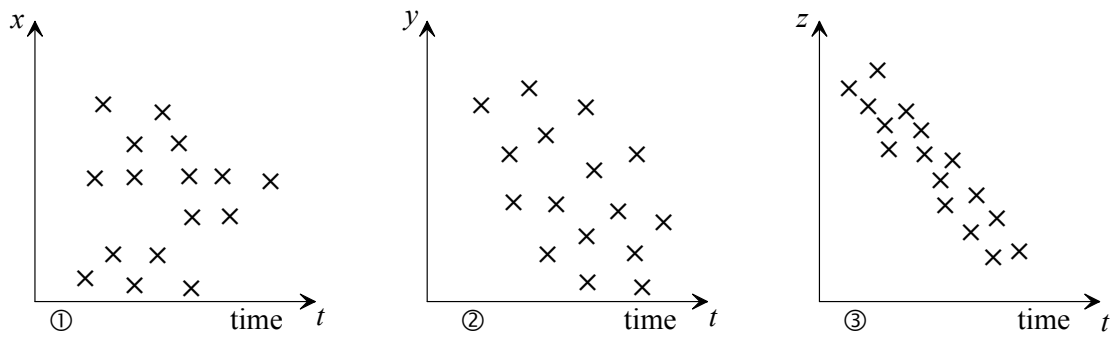
(c) Comment on your value for r. *[2 marks]*

(d) Calculate the equation of the regression line of y on x. *[3 marks]*

(e) Using your regression line, estimate how many seconds it will take a student with an arm length of 75 cm to swim the 200 m breaststroke. *[1 mark]*

4)

5) (iii) The sketches below represent scatter diagrams for the way in which variables x, y and z change over time, t, in a given chemical experiment. They are labelled ①, ② and ③.



(a) State which of the diagrams indicate that the pair of variables

(i) is not correlated. *[1 mark]*

(ii) shows strong linear correlation. *[1 mark]*

(b) A student is given a piece of paper with five numbers written on it. She is told that three of these numbers are the product moment correlation coefficients for the three pairs of variables shown above. The five numbers are

$$0.9, \quad -0.85, \quad -0.20, \quad 0.04, \quad 1.60$$

(i) For each sketch above state which of these five numbers is the most appropriate value for the correlation coefficient. *[3 marks]*

(ii) For the two remaining numbers, state why you reject them for this experiment. *[2 marks]*

(c) Another variable, w, over time, t, gave the following information

$$\sum t = 124 \quad \sum w = 250 \quad s_t = 6.08 \quad s_w = 10.50 \quad s_{tw} = 55.00$$

for 20 data points.

Calculate

(i) the product moment correlation coefficient for this data. *[2 marks]*

(ii) the equation of the regression line of w on t in the form $w = at + b$. *[5 marks]*