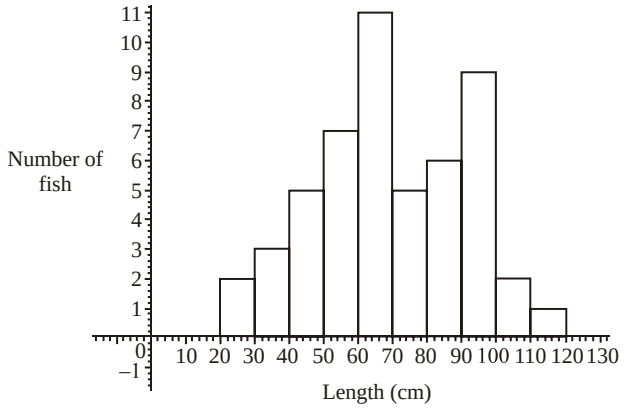
SL 2-Variable Statistics Practice from Math Studies

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.

(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?

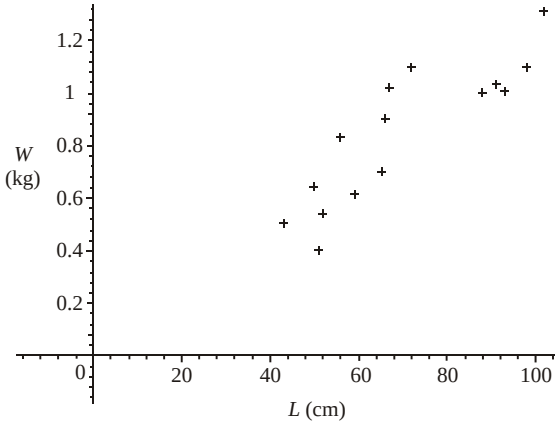
(3)

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)

2. 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)

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)

(b) (i) Write down the mean value of the load ($\bar{x}$).

(1)

(ii) Write down the standard deviation of the load.

(1)

(iii) Write down the mean value of the length ($\bar{y}$).

(1)

(iv) Write down the standard deviation of the length.

(1)

(c) Plot the mean point ($\bar{x}$, $\bar{y}$) on the scatter diagram. Name it L.

(1)

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

(d) (i) Write down the correlation coefficient, r, for these readings.

(1)

(ii) Comment on this result.

(2)

(e) Find the equation of the regression line of y on x.

(2)

(f) Draw the line of regression on the scatter diagram.

(2)

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

(1)

(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.

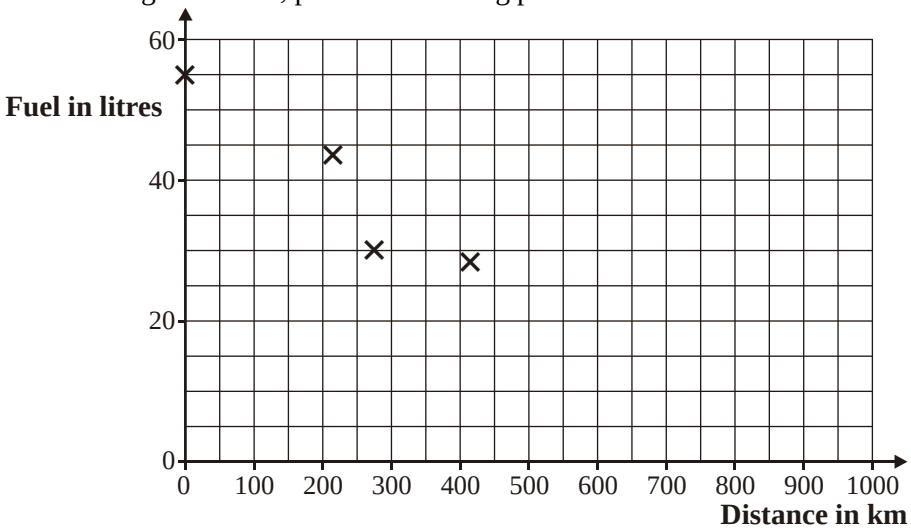
(1)

(Total 18 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. 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) (x)	**Shoe size (y)**
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 $y = mx + c$.

(3)

(b) Use your equation in part (a) to predict the shoe size of a student who is 162 cm in height.

(2)

(c) Write down the correlation coefficient.

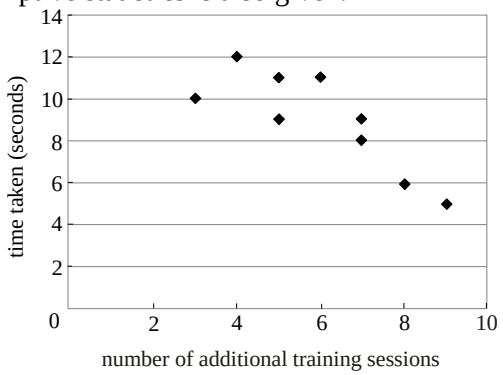
(1)

(d) Describe the correlation between height and shoe size.

(2)

(Total 8 marks)

6. A number of employees at a factory were given x additional training sessions each. They were then timed on how long (y seconds) it took them to complete a task. The results are shown in the scatter diagram below. A list of descriptive statistics is also given.



$n = 9$,
sum of x values: $\Sigma x = 54$,
sum of y values: $\Sigma y = 81$,
mean of x values: $\bar{x} = 6$,
mean of y values: $\bar{y} = 9$,
standard deviation of x: $s_x = 1.94$,
standard deviation of y: $s_y = 2.35$,
covariance: $s_{xy} = -3.77$.

(a) Determine the product-moment correlation coefficient (r) for this data.

(2)

(b) What is the nature of the relationship between the amount of additional training and the time taken to complete the task?

(2)

(c) Calculate $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ given that the covariance $s_{xy} = -3.77$.

(1)

(d) (i) Determine the equation of the linear regression line for y on x.
(ii) Find the expected time to complete the task for an employee who only attended three additional training sessions.

(4)

(Total 9 marks)

7. 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.

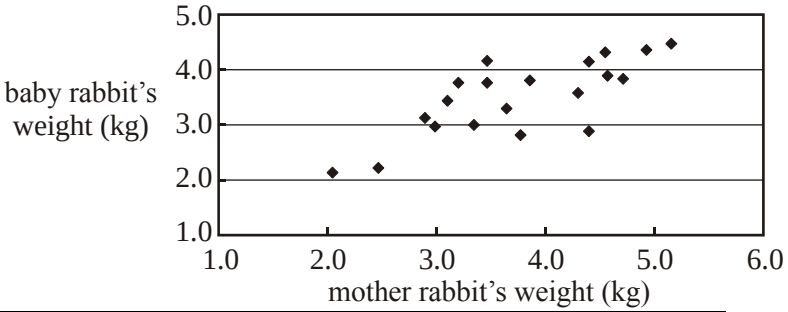
	1	2	3	4	5	6	7	8
Length of arm, x cm	79	74	72	70	77	73	64	69
Breaststroke, y seconds	135.1	135.7	139.3	141.0	132.8	137.0	152.9	144.0

(a) Calculate the mean and standard deviation of x and y. **(4)**
(b) Given that sxy = –24.82, calculate the correlation coefficient, r. **(2)**
(c) Comment on your value for r. **(2)**
(d) Calculate the equation of the regression line of y on x. **(3)**
(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)

(Total 12 marks)

8. A study was carried out to investigate possible links between the weights of baby rabbits and their mothers. A sample of 20 pairs of mother rabbits (x) and baby rabbits (y) was chosen at random and their weights noted. This information was plotted on a scatter diagram and various statistical calculations were made. These appear below.



mean of x	mean of y	s_x	s_y	s_{xy}	sum of x	sum of y
3.78	3.46	0.850	0.689	0.442	75.6	69.2

(a) Show that the product-moment correlation coefficient r for this data is 0.755.

(2)

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

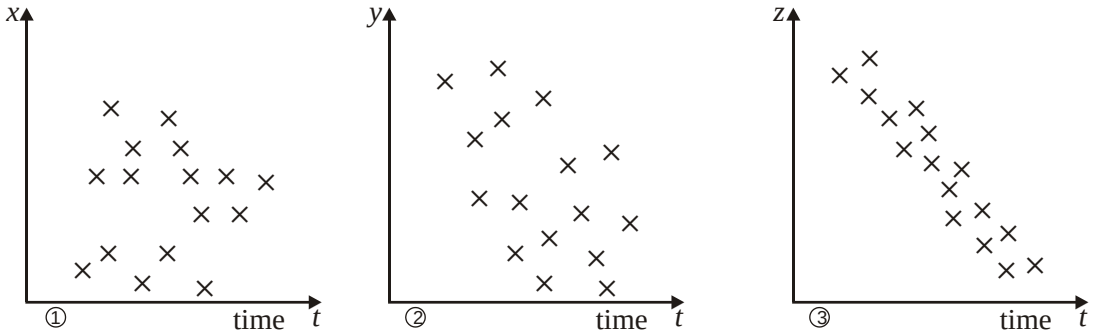
(3)

(ii) Use your equation for the regression line to estimate the weight of a rabbit given that its mother weighs 3.71 kg.

(2)

(Total 7 marks)

9. 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)

(ii) shows strong linear correlation.

(1)

(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, –0.85, –0.20, 0.04, 1.60

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

(3)

(ii) For the two remaining numbers, state why you reject them for this experiment.

(2)

(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;

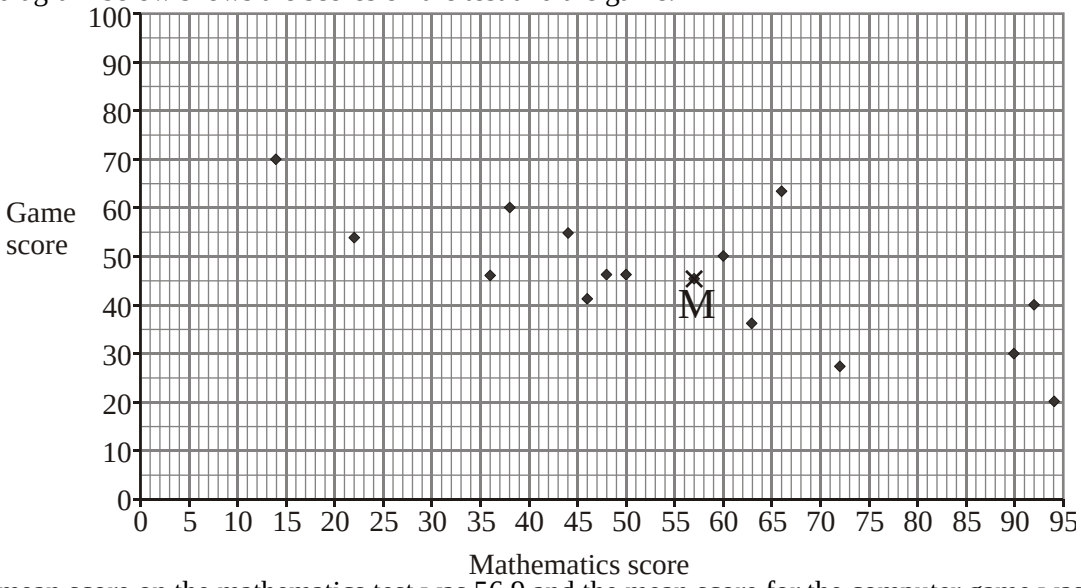
(2)

(ii) the equation of the regression line of w on t in the form $w = at + b$.

(5)

(Total 14 marks)

10. 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 mean score on the mathematics test was 56.9 and the mean score for the computer game was 45.9. The point M has coordinates (56.9, 45.9).

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

A straight line of best fit passes through the point (0, 69).

(b) On the diagram draw this straight line of best fit.

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.

(Total 8 marks)

SL 2-Variable Statistics Practice- MarkScheme

1. *Unit penalty (UP) is applicable where indicated.*

(a) Total = 2 + 3 + 5 + 7 + 11 + 5 + 6 + 9 + 2 + 1 (M1)

Note: *(M1) is for a sum of frequencies.*

= 51 (A1)(G2) 2

(b) (i) modal interval is 60 – 70 (A1)

Note: *Award (A0) for 65*

(ii) median is length of fish no. 26, (M1)

also 60 – 70 (A1)(G2)

Note: *Can award (A1)(ft) or (G2)(ft) for 65 if (A0) was awarded for 65 in part (i).*

UP (iii) mean is $\frac{2\times25+3\times35+5\times45+7\times55+\ldots}{51}$ (M1)

UP = 69.5 cm (3s.f.) (A1)(ft)(G1) 5

Note: *(M1) is for a sum of (frequencies multiplied by mid-point values) divided by candidate's answer from part (a). Accept mid-points 25.5, 35.5 etc or 24.5, 34.5 etc, leading to answers 70.0 or 69.0 (3s.f.) respectively. Answers of 69.0, 69.5 or 70.0 (3s.f.) with no working can be awarded (G1).*

UP (c) (i) standard deviation is 21.8 cm (G1)

Note: *For any other answer without working, award (G0). If working is present then (G0)**(AP)** is possible.*

(ii) $69.5 + 3 \times 21.8 = 134.9 > 120$ (M1)

no fish (A1)(ft)(G1) 3

Note: *For 'no fish' without working, award (G1) regardless of answer to (c)(i). Follow through from (c)(i) only if method is shown.*

(d) 5 fish are less than 40 cm in length, (M1)

Award (M1) for any of $\frac{5}{51}, \frac{46}{51}$, *0.098 or 9.8%, 0.902, 90.2% or 5.1 seen.*

hence no fine. (A1)(ft) 2

Note: *There is no G mark here and (M0)(A1) is never allowed. The follow-through is from answer in part (a).*

(e) (i) and (iii) are correct. (A1)(A1) 2

[14]

2. (a) $a = -0.134$, $b = 20.9$ (A1)

$y = 20.9 - 0.134x$ (A1) (C2)

(b) 17 objects (A1)(ft) (C1)

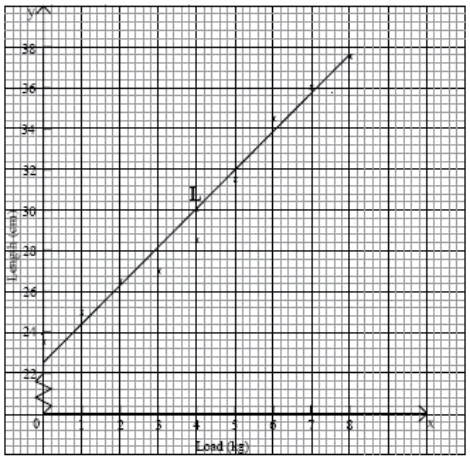
Note: *Accept only 17*

(c) $r = -0.756$ (A1) (C1)

(d) negative and moderately strong (A1)(ft)(A1)(ft) (C2)

[6]

3. (a)



(A4)

Note: *Award (A1) for correct scales and labels, (A3) for correct points, (A2) for 7 or 8 correct, (A1) for 5 or 6 correct.*

(b) (i) 4 (G1)
(ii) 2.58 (G1)
(iii) 30 (G1)
(iv) 4.78 (G1)

Note: *If wrong version of s.d. used in (ii), can (ft) in (iv) (5.07).*

(c) L correctly plotted on graph and named (A1)(ft)

(d) (i) $r = 0.986$ (0.987) (G1)
(ii) (very) strong positive correlation (R1)(ft)(R1)(ft)

(e) $y = 1.83x + 22.7$ ($y = 1.825x + 22.7$) (G1)(G1)

Award (G1) for y= 1.83x (1.825x), (G1) for 22.7

(f) Line drawn on graph. (A1)(A1)(ft)

Note: *Award (A1) for passing through the mean point, (A1) for y intercept between 22 and 23.*

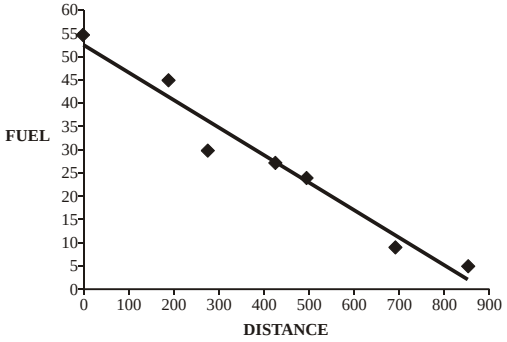
(g) (i) 32.6 cm (A1)(ft)

Note: *Allow margin of error of 0.2 from value on candidate's diagram.*

(ii) Not possible to find an answer as the value lies too far outside the given set of data. (R1)

[18]

4. (a)



For all 3 points correct (A2) (C2)

Note: *If only 2 points correct award (A1).*

(b) For straight line **with –ve gradient** for passing through the mean (A1)(A1)
For straight line intercept on *y*-axis between 50 and 55 (A1) (C3)

(c) 32 (read answer from candidate's line) (A1) (C1)

[6]

5. (a) $y = 0.070x - 3.22$ (G3) 3

Notes: *Award (G1) for correct m value, (G1) for 3.22, (G1) for negative sign. Accept 0.07x.*

(b) $y = 0.070 \times 162 - 3.22$ (M1)

$= 8.12$

Therefore shoe size 8 or 9 (8.12). (A1)

OR

$y = 8$ or 9 (G2) 2

(c) $r = 0.681$ (A1) 1

(d) Moderately strong, positive correlation. (A1)(A1) 2 **[8]**

6. (a) $r = -\frac{3.77}{1.94 \times 2.35} = -0.827$ (M1)(A1)

or (G2) 2

(b) moderate/strong (allow approximately linear) (A1)

negative (A1) 2

Note: *Comments such as: number of sessions increases as time decreases can be awarded (A1)(A0) but "inversely proportional" receives no marks.*

(c) $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = -3.77 \times 9 = -33.9$ (A1) 1

(d) (i) $(y - 9) = -\frac{3.77}{1.94^2}(x - 6)$ (M1)

$y = -x + 15$ (A1)

or (G2)

(ii) $-3 + 15 = 12$ seconds (M1)(A1)

or (G2) 4 **[9]**

7. (a) mean of $x = 72.25$ (A1)

sd of $x = 4.41$ (A1)

mean of $y = 139.7$ (140) (A1)

sd of $y = 5.99$ (A1) 4

(b) $r = -0.940$ (G2)

OR

$r = \frac{-24.82}{(4.41 \times 5.99)}$

$= -0.9396$ (= – 0.94) (M1)(A1) 2

(c) strong, negative correlation (A2) 2

Note: *Award (A1) for negative, (A1) for strong.*

(d) $y = 232 - 1.28x$ (G3)

OR

$(y - 139.7) = -\frac{24.82}{4.41^2}(x - 72.25)$

$y = -1.28x + 232$ (M1)(A1)(A1)**)** 3

(e) $y = 232 - 1.28 \times 75 = 136$ seconds (A1) 1

[12]

8. (a) $r = \frac{0.442}{0.850 \times 0.689} = 0.755$ (M1)(M1)(AG) 2

(b) (i) $y - \bar{y} = \frac{S_{xy}}{(S_x)^2}(x - \bar{x})$; $y - 3.46 = \frac{0.442}{(0.85)^2}(x - 3.78)$ (M1)

$y = 0.612x + 1.15$ (A1)(A1)

(ii) weight of rabbit = $0.612 \times 3.71 + 1.15 = 3.42$ kg (M1)(A1) 5

[7]

9. (a) (i) ① (A1)

(ii) ③ (A1) 2

(b) (i) ① 0.04 (A1)

② –0.20 (A1)

③ –0.85 (A1)

(ii) 1.60 A product–moment correlation coefficient cannot be greater than 1. (R1)

0.90 There is no diagram with a strong positive correlation. (R1) 5

(c) (i) Product-moment correlation = $\frac{55.00}{6.08 \times 10.50}$ (A1)

$= 0.8615$

$= 0.862$ (A1)

(ii) $\bar{t} = \frac{124}{20} = 6.2$ $\bar{w} = \frac{250}{20} = 12.5$ (both correct) (A1)

$w - \bar{w} = \frac{S_{tw}}{S_t^2}(t - \bar{t})$

$(w - 12.5) = \frac{55.00}{(6.08)^2}(t - 6.2)$ (M1)(A1)

$w - 12.5 = 1.4878(t - 6.2)$

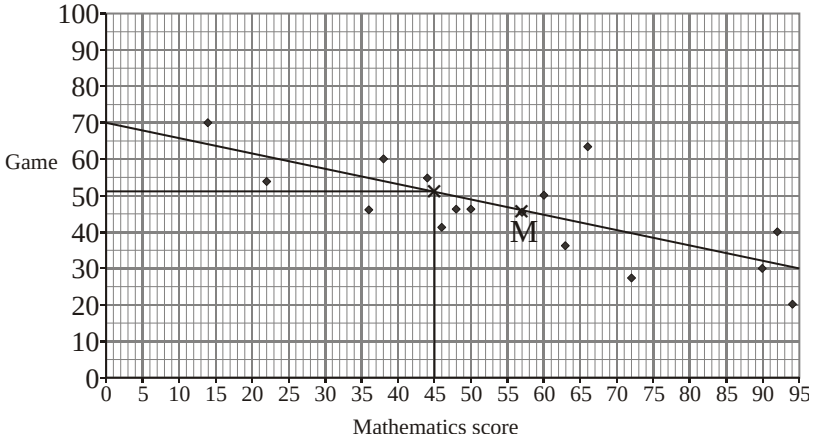
$w = 1.49t + 3.28$ (A1)(A1) 7

Note: ***ft*** *from candidate's mean values*

[14]

10. (a) The scores are negatively correlated (A2) (C2)

(b)



Line must be drawn straight. (A1)

It must pass through (0, 69). (A1)

It must pass through the mean point M = (56.9, 45.9). (A1) (C3)

(c) For method shown using intersecting coordinate lines or a mark in the right place: (M1)

51 is closest. Allow 50 or 52 (**ft** from candidate's graph). (A2) (C3)

Note: *Award only (A1) for 49, 53 or for any non-integer between 49 and 52. Assume that 50.7 is result of linear regression using two required points. Award (G2).*

[8]