Regression line *[18 marks]*

The following table shows the Mathematics test scores (x) and the Science test scores (y) for a group of eight students.

Mathematics scores (x)	64	68	72	75	80	82	85	86
Science scores (y)	67	72	77	76	84	83	89	91

The regression line of y on x for this data can be written in the form $y = ax + b$.

1a. Find the value of a and the value of b. *[2 marks]*

Markscheme

$1.01206\ldots,\ 2.45230\ldots$

$a = 1.01,\ b = 2.45\ (1.01x + 2.45)$ ***A1A1***

[2 marks]

1b. Write down the value of the Pearson's product-moment correlation coefficient, r. *[1 mark]*

Markscheme

$0.981464\ldots$

$r = 0.981$ ***A1***

Note: A common error is to enter the data incorrectly into the GDC, and obtain the answers $a = 1.01700\ldots$, $b = 2.09814\ldots$ and $r = 0.980888\ldots$ Some candidates may write the 3 sf answers, ie. $a = 1.02$, $b = 2.10$ and $r = 0.981$ or 2 sf answers, ie. $a = 1.0$, $b = 2.1$ and $r = 0.98$. In these cases award ***A0A0*** for part (a) and ***A0*** for part (b). Even though some values round to an accepted answer, they come from incorrect working.

[1 mark]

1c. Use the equation of your regression line to predict the Science test score *[2 marks]* for a student who has a score of 78 on the Mathematics test. Express your answer to the nearest integer.

Markscheme

correct substitution of 78 into **their** regression equation ***(M1)***

$81.3930\ldots$ 81.23 from 3 sf answer

81 ***A1***

[2 marks]

In Lucy's music academy, eight students took their piano diploma examination and achieved scores out of 150. For her records, Lucy decided to record the average number of hours per week each student reported practising in the weeks prior to their examination. These results are summarized in the table below.

Average weekly practice time (h)	28	13	45	33	17	29	39	36
Diploma score (D)	115	82	120	116	79	101	110	121

2a. Find Pearson's product-moment correlation coefficient, r, for these data. *[2 marks]*

Markscheme

use of GDC to give **(M1)**

$r = 0.883529\ldots$

$r = 0.884$ **A1**

Note: Award the ***(M1)*** for any correct value of r, a, b or $r^2 = 0.780624\ldots$ seen in part (a) or part (b).

[2 marks]

2b. The relationship between the variables can be modelled by the regression equation $D = ah + b$. Write down the value of a and the value of b. *[1 mark]*

Markscheme

$a = 1.36609\ldots$, $b = 64.5171\ldots$

$a = 1.37$, $b = 64.5$ **A1**

[1 mark]

2c. One of these eight students was disappointed with her result and wished she had practised more. Based on the given data, determine how her score could have been expected to alter had she practised an extra five hours per week. *[2 marks]*

Markscheme

attempt to find their difference ***(M1)***

$5 \times 1.36609\ldots$ OR
$1.36609\ldots(h+5)+64.5171\ldots-(1.36609\ldots h+64.5171\ldots)$

$6.83045\ldots$

$= 6.83 \ \ (6.85 \text{ from } 1.37)$

the student could have expected her score to increase by 7 marks.
A1

Note: Accept an increase of $6,\ 6.83$ or 6.85.

[2 marks]

2d. Lucy asserts that the number of hours a student practises has a direct effect on their final diploma result. Comment on the validity of Lucy's assertion. *[1 mark]*

Markscheme

Lucy is incorrect in suggesting there is a causal relationship.

This might be true, but the data can only indicate a correlation.
R1

Note: Accept 'Lucy is incorrect as correlation does not imply causation' or equivalent.

[1 mark]

2e. Lucy suspected that each student had not been practising as much as they reported. In order to compensate for this, Lucy deducted a fixed number of hours per week from each of the students' recorded hours. *[1 mark]*

State how, if at all, the value of r would be affected.

Markscheme

no effect **A1**

[1 mark]

Lucy sells hot chocolate drinks at her snack bar and has noticed that she sells more hot chocolates on cooler days. On six different days, she records the maximum daily temperature, T, measured in degrees centigrade, and the number of hot chocolates sold, H. The results are shown in the following table.

Maximum temperature (T)	14	8	4	18	13	11
Number of hot chocolates (H)	79	143	191	58	84	105

The relationship between H and T can be modelled by the regression line with equation $H = aT + b$.

3a. Find the value of a and of b. *[3 marks]*

Markscheme

valid approach ***(M1)***

eg correct value for a or b (or for r or $r^2 = 0.962839$ seen in (ii))

$a = -9.84636,\ b = 221.592$

$a = -9.85,\ b = 222$ ***A1A1 N3***

[3 marks]

3b. Write down the correlation coefficient. *[1 mark]*

Markscheme

-0.981244

$r = -0.981$ ***A1 N1***

[1 mark]

3c. Using the regression equation, estimate the number of hot chocolates that Lucy will sell on a day when the maximum temperature is $12\,^{\circ}\mathrm{C}$. *[2 marks]*

Markscheme

correct substitution into their equation ***(A1)***

eg $-9.85 \times 12 + 222$

103.435 (103.8 from 3 sf)

103 (hot chocolates) ***A1 N2***

[2 marks]