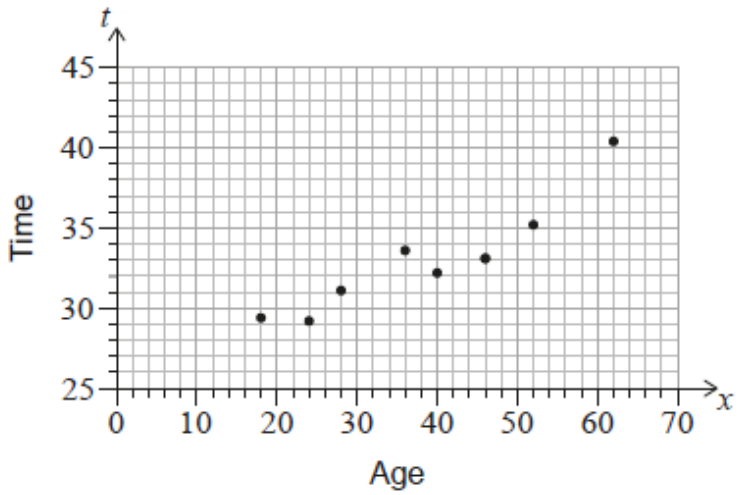
AI SL 13.09 [63 marks]

1. [Maximum mark: 6] 21N.1.SL.TZ0.1

Eduardo believes that there is a linear relationship between the age of a male runner and the time it takes them to run 5000 metres.

To test this, he recorded the age, x years, and the time, t minutes, for eight males in a single $5000\ \text{m}$ race. His results are presented in the following table and scatter diagram.

x, years	18	24	28	36	40	46	52	62
t, minutes	29.4	29.2	31.1	33.6	32.2	33.1	35.2	40.4



(a) For this data, find the value of the Pearson's product-moment correlation coefficient, r. [2]

Markscheme

$r = 0.933\ \ (0.933419\ldots)$ ***A2***

[2 marks]

Eduardo looked in a sports science text book. He found that the following information about r was appropriate for athletic performance.

Value of $\lvert r \rvert$	Description of the correlation
$0 \leq \lvert r \rvert < 0.4$	weak
$0.4 \leq \lvert r \rvert < 0.8$	moderate
$0.8 \leq \lvert r \rvert \leq 1$	strong

(b) Comment on your answer to part (a), using the information that Eduardo found. [1]

Markscheme

strong *A1*

Note: Answer may include "positive", however this is not necessary for the mark.

[1 mark]

(c) Write down the equation of the regression line of t on x, in the form $t = ax + b$. [1]

Markscheme

$t = 0.228x + 24.3 \quad (t = 0.227703\ldots x + 24.3153\ldots)$ *A1*

Note: Condone y in place of t. Answer must be an equation.

[1 mark]

(d) A 57-year-old male also ran in the 5000 m race.

Use the equation of the regression line to estimate the time he took to complete the $5000\ \text{m}$ race. [2]

Markscheme

$(t =)\ 0.227703\ldots \times 57 + 24.3153\ldots$ ***(M1)***

Note: Award ***(M1)*** for correct substitution into their regression line.

$(t =)\ 37.3\ \text{minutes}\ (37.2944)$ ***A1***

Note: Accept 37.1 and 37.4 from use of 2sf and/or 3sf values.

[2 marks]

2. [Maximum mark: 6] 20N.2.SL.TZ0.S_2

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 $H = aT + b$.

(a.i) Find the value of a and of b. [3]

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]

(a.ii) Write down the correlation coefficient. [1]

Markscheme

-0.981244

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

[1 mark]

(b) 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]

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]

3. [Maximum mark: 6] 19N.1.SL.TZ0.T_6

Galois Airways has flights from Hong Kong International Airport to different destinations. The following table shows the distance, x kilometres, between Hong Kong and the different destinations and the corresponding airfare, y, in Hong Kong dollars (HKD).

Destination	Bali, Indonesia	Sydney, Australia	Bengaluru, India	Singapore	Auckland, New Zealand	Bangkok, Thailand
Distance x, (km)	3400	7400	4000	2600	9200	1700
Airfare y, (HKD)	1550	3600	2800	1300	4000	1400

The Pearson's product–moment correlation coefficient for this data is 0.948, correct to three significant figures.

(a) Use your graphic display calculator to find the equation of the regression line y on x. [2]

Markscheme

$y = 0.384x + 629$

$y = (0.384221\ldots)\,x + (629.421\ldots)$ ***(A1)(A1) (C2)***

Note: Award ***(A1)*** for $0.384x$, ***(A1)*** for 629. If the answer is not given as an equation, award a maximum of ***(A1)(A0)***.

[2 marks]

The distance from Hong Kong to Tokyo is 2900 km.

(b) Use your regression equation to estimate the cost of a flight from Hong Kong to Tokyo with Galois Airways. [2]

Markscheme

$y = 0.384221\ldots \times 2900 + 629.421\ldots$ ***(M1)***

Note: Award ***(M1)*** for substitution into *their* regression equation.

1740 $(1744, 1743.66\ldots)$ (HKD) ***(A1)*****(ft)** ***(C2)***

Note: Follow through from part (a).

[2 marks]

(c) Explain why it is valid to use the regression equation to estimate the airfare between Hong Kong and Tokyo. [2]

Markscheme

the correlation is (very) strong ***(R1)***

2900 (km) is within the given data range (interpolation) ***(R1)*** ***(C2)***

Note: Two correct reasons are required for the awarding of ***(C2)***.

[2 marks]

4. [Maximum mark: 6] 19M.2.SL.TZ2.S_1

A group of 7 adult men wanted to see if there was a relationship between their Body Mass Index (BMI) and their waist size. Their waist sizes, in centimetres, were recorded and their BMI calculated. The following table shows the results.

Waist (x cm)	58	63	75	82	93	98	105
BMI (y)	19	20	22	23	25	24	26

The relationship between x and y can be modelled by the regression equation $y = ax + b$.

(a.i) Write down the value of a and of b. [3]

Markscheme

valid approach ***(M1)***

eg correct value for a or b (or for correct r or $r^2 = 0.955631$ seen in (ii))

0.141120, 11.1424

$a = 0.141$, $b = 11.1$ ***A1A1 N3***

[3 marks]

(a.ii) Find the correlation coefficient. [1]

Markscheme

0.977563

$r = 0.978$ ***A1 N1***

[1 mark]

(b) Use the regression equation to estimate the BMI of an adult man whose waist size is 95 cm. [2]

Markscheme

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

eg 0.141(95) + 11.1

24.5488

24.5 ***A1 N2***

[2 marks]

5. [Maximum mark: 6] 18M.2.SL.TZ2.S_1

The following table shows the mean weight, y kg , of children who are x years old.

Age (x years)	1.25	2.25	3.5	4.4	5.85
Weight (y kg)	10	13	14	17	19

The relationship between the variables is modelled by the regression line with equation $y = ax + b$.

(a.i) Find the value of a and of b. [3]

Markscheme

valid approach ***(M1)***

eg correct value for a or b (or for r seen in (ii))

$a = 1.91966$ $b = 7.97717$

$a = 1.92$, $b = 7.98$ ***A1A1 N3***

[3 marks]

(a.ii) Write down the correlation coefficient. [1]

Markscheme

0.984674

$r = 0.985$ ***A1 N1***

[1 mark]

(b) Use your equation to estimate the mean weight of a child that is 1.95 years old. [2]

Markscheme

correct substitution into their equation ***(A1)***
eg $1.92 \times 1.95 + 7.98$

11.7205

11.7 (kg) ***A1 N2***

[2 marks]

6. [Maximum mark: 17] SPM.2.SL.TZ0.3

The Malvern Aquatic Center hosted a 3 metre spring board diving event. The judges, Stan and Minsun awarded 8 competitors a score out of 10. The raw data is collated in the following table.

Competitors	A	B	C	D	E	F	G	H
Stan's score (x)	4.1	3	4.3	6	7.1	6	7.5	6
Minsun's score (y)	4.7	4.6	4.8	7.2	7.8	9	9.5	7.2

(a.i) Write down the value of the Pearson's product–moment correlation coefficient, r. [2]

Markscheme

0.909 (0.909181…) ***A2***

[2 marks]

(a.ii) Using the value of r, interpret the relationship between Stan's score and Minsun's score. [2]

Markscheme

(very) strong and positive ***A1A1***

Note: Award ***A1*** for (very) strong ***A1*** for positive.

[2 marks]

(b) Write down the equation of the regression line y on x. [2]

Markscheme

$y = 1.14x + 0.578$ $(y = 1.14033\ldots x + 0.578183\ldots)$ ***A1A1***

Note: Award ***A1*** for $1.14x$, ***A1*** for 0.578. Award a maximum of ***A1A0*** if the answer is not an equation in the form $y = mx + c$.

[2 marks]

(c.i) Use your regression equation from part (b) to estimate Minsun's score when Stan awards a perfect 10. [2]

Markscheme

1.14 × 10 + 0.578 ***M1***

12.0 (11.9814…) ***A1***

[2 marks]

(c.ii) State whether this estimate is reliable. Justify your answer. [2]

Markscheme

no the estimate is not reliable ***A1***

outside the known data range ***R1***
OR
a score greater than 10 is not possible ***R1***

Note: Do not award ***A1R0***.

[2 marks]

The Commissioner for the event would like to find the Spearman's rank correlation coefficient.

(d) **Copy** and complete the information in the following table.

Competitors	A	B	C	D	E	F	G	H
Stan's Rank		8					1	4
Minsun's Rank		8					1	4.5

[2]

Markscheme

Competitors	A	B	C	D	E	F	G	H
Stan's rank	7	8	6	4	2	4	1	4
Minsun's rank	7	8	6	4.5	3	2	1	4.5

A1A1

Note: Award ***A1*** for correct ranks for Stan. Award ***A1*** for correct ranks for Minsun.

[2 marks]

(e.i) Find the value of the Spearman's rank correlation coefficient, r_s. [2]

Markscheme

0.933 (0.932673…) ***A2***

[2 marks]

(e.ii) Comment on the result obtained for r_s. [2]

Markscheme

Stan and Minsun strongly agree on the ranking of competitors. ***A1A1***

Note: Award ***A1*** for "strongly agree", ***A1*** for reference to a rank order.

[2 marks]

(f) The Commissioner believes Minsun's score for competitor G is too high and so decreases the score from 9.5 to 9.1.

Explain why the value of the Spearman's rank correlation coefficient r_s does not change. [1]

Markscheme
decreasing the score to 9.1, does not change the rank of competitor G ***A1***
[1 mark]

7. [Maximum mark: 16] 22M.2.SL.TZ1.3

The scores of the eight highest scoring countries in the 2019 Eurovision song contest are shown in the following table.

	Eurovision score
Netherlands	498
Italy	472
Russia	370
Switzerland	364
Sweden	334
Norway	331
North Macedonia	305
Azerbaijan	302

For this data, find

(a.i) the upper quartile. [2]

Markscheme

$\frac{370+472}{2}$ ***(M1)***

Note: This ***(M1)*** can also be awarded for either a correct Q_3 or a correct Q_1 in part (a)(ii).

$\mathrm{Q}_3 = 421$ ***A1***

[2 marks]

(a.ii) the interquartile range. [2]

Markscheme

their part (a)(i) – their Q_1 (clearly stated) ***(M1)***

IQR $=(421 - 318 =)\ 103$ ***A1***

[2 marks]

(b) Determine if the Netherlands' score is an outlier for this data.
Justify your answer. [3]

Markscheme

$(\mathrm{Q}_3 + 1.5(\text{IQR}) =)\ 421 + (1.5 \times 103)$ ***(M1)***

$= 575.5$

since $498 < 575.5$ ***R1***

Netherlands is not an outlier ***A1***

Note: The ***R1*** is dependent on the ***(M1)***. Do not award ***R0A1***.

[3 marks]

Chester is investigating the relationship between the highest-scoring countries' Eurovision score and their population size to determine whether population size can reasonably be used to predict a country's score.

The populations of the countries, to the nearest million, are shown in the table.

	Population (x) (millions)	Eurovision score (y)
Netherlands	17	498
Italy	60	472
Russia	145	370
Switzerland	9	364
Sweden	10	334
Norway	5	331
North Macedonia	2	305
Azerbaijan	10	302

Chester finds that, for this data, the Pearson's product moment correlation coefficient is $r = 0.249$.

(c) State whether it would be appropriate for Chester to use the equation of a regression line for y on x to predict a country's Eurovision score. Justify your answer. [2]

Markscheme

not appropriate ("no" is sufficient) ***A1***

as r is too close to zero / too weak a correlation ***R1***

[2 marks]

Chester then decides to find the Spearman's rank correlation coefficient for this data, and creates a table of ranks.

	Population rank (to the nearest million)	Eurovision score rank
Netherlands	3	1
Italy	2	2
Russia	1	3
Switzerland	a	4
Sweden	b	5
Norway	7	6
North Macedonia	8	7
Azerbaijan	c	8

Write down the value of:

(d.i) a. [1]

Markscheme

6 **A1**

[1 mark]

(d.ii) b. [1]

Markscheme

4.5 **A1**

[1 mark]

(d.iii) c. [1]

Markscheme

4.5 **A1**

[1 mark]

(e.i) Find the value of the Spearman's rank correlation coefficient r_s. [2]

Markscheme

$r_s = 0.683 \quad (0.682646\ldots)$ **A2**

[2 marks]

(e.ii) Interpret the value obtained for r_s. [1]

Markscheme

EITHER

there is a (positive) association between the population size and the score
A1

OR

there is a (positive) linear correlation between the ranks of the population size and the ranks of the scores (when compared with the PMCC of 0.249).
A1

[1 mark]

(f) When calculating the ranks, Chester incorrectly read the Netherlands' score as 478. Explain why the value of the Spearman's rank correlation r_s does not change despite this error. [1]

Markscheme

lowering the top score by 20 does not change its rank so r_s is unchanged

R1

Note: Accept "this would not alter the rank" or "Netherlands still top rank" or similar. Condone any statement that clearly implies the ranks have not changed, for example: "The Netherlands still has the highest score."

[1 mark]