Regression line - revision questions (preDP1) [25 marks]

1. [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]

2. [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]

3. [Maximum mark: 7] 23N.1.SL.TZ1.1

Billy is a keen walker who keeps a record of his performance. The following table shows the time, in minutes, it takes him to walk one kilometre up hills with different gradients. The gradient of each hill is constant.

Gradient G (%)	0	4	10	15	20
Time T (min.)	6.85	8.42	11.20	14.49	17.88

(a.i) Find the equation of the regression line of T on G.

[2]

Markscheme

$T = 0.552G + 6.36 \ \ (= 0.552139\ldots G + 6.35703\ldots)$ ***A1A1***

Note: Award ***A1*** for **correct** values of a and b, ***A1*** for an equation using these **correct** values.

[2 marks]

(a.ii) Describe the correlation between T and G with reference to the value of r, the Pearson's product-moment correlation coefficient. [2]

Markscheme

$(r =)\ 0.994\ (= 0.993910\ldots)$ ***A1***

there is a (very) strong positive linear correlation ***R1***

Note: If r is missing award ***A0R0***.

[2 marks]

On Sunday, Billy intends to walk up a hill with a gradient of $13\,\%$.

(b) Estimate the time it will take Billy to walk one kilometre up the hill. [2]

Markscheme

attempt to substitute 13 into their regression equation ***(M1)***

$T = 0.552139\ldots(13) + 6.35703\ldots$

$13.5\,(\text{mins})\ \ (= 13.5348\ldots)$ ***A1***

[2 marks]

This morning, Billy walked one kilometre up a hill, and it took 22 minutes.

(c) Explain why it would be inappropriate to use the equation found in part (a) to estimate the gradient of this hill. [1]

Markscheme

EITHER

using the T on G regression line cannot (always) reliably make a prediction for G ***R1***

OR

equation is for Time on Gradient not Gradient on Time. ***R1***

OR

this estimate is an extrapolation ***R1***

OR

there is no reason to assume this new hill has constant gradient ***R1***

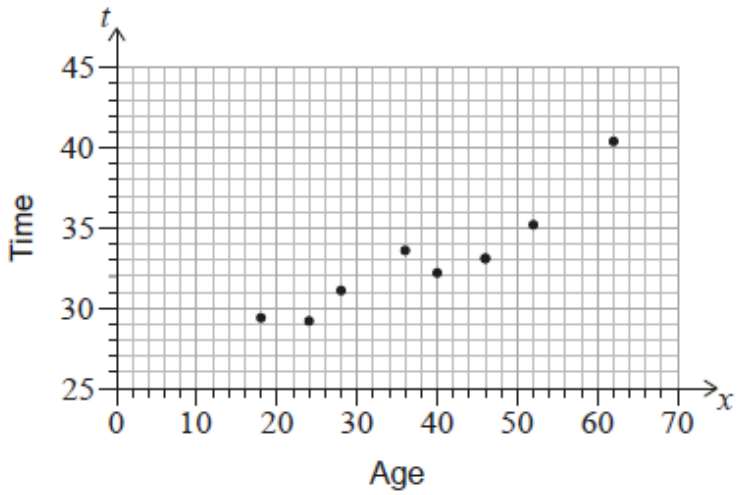
[1 mark]

4. [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 $5000\,\text{m}$ 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]