Correlation and regression

Skills sheet · SL AA 4.4 · 4.10

KEY FACTS

$-1 \leq r \leq 1$: sign and strength of a linear link D1
$y = ax + b$ passes through $(\bar{x}, \bar{y})$ D2
y from x: line of y on x; x from y: line of x on y P1
Reliable: $|r| \approx 1$ and interpolation P1

Correlation is not causation W1.
Never rearrange y on x W2.
r is not the gradient.

WHAT UNLOCKS

- "describe" $\rightarrow$ strong / moderate / weak, positive / negative.
- an estimate $\rightarrow$ the right line, then "interpolation" or "extrapolation".
- a missing mean $\rightarrow$ substitute the other mean in the line.
- "does x cause y?" $\rightarrow$ a third variable.

WORKED EXAMPLE ①

$r = 0.15$. Describe the correlation; is a prediction from the line reliable? no GDC

very weak positive correlation A1
not reliable: r close to 0 R1

Method
Judging a prediction
① Read $|r|$: close to 1?
② Is the value inside the data?
③ Conclude with the reason.

WORKED EXAMPLE ②

$y = 2x + 3$, the data run from $x = 1$ to $x = 9$. Estimate y at $x = 5$. no GDC

$2 \times 5 + 3 = 13$ A1
interpolation, so reasonable R1

Method
Using the line
① Check the direction: y from x.
② Substitute.
③ Say interpolation or extrapolation.
Outside the data: extrapolation.

FLUENCY

1. Describe $r = 0.93$ no GDC
2. Describe $r = -0.55$ no GDC
3. Describe $r = 0.05$ no GDC
4. $y = 4x - 1$: estimate y at $x = 3$ no GDC
5. $y = 0.5x + 2$ and $\bar{x} = 6$: find $\bar{y}$ no GDC
6. Interpret the gradient of $y = 2.5x + 10$ no GDC

REASONING

●●● **Question 1.** For x from 2 to 10 hours of sunshine, the line of y on x is $y = 3x + 5$ and $r = 0.91$. [no GDC]

(a) Estimate y when $x = 6$. [1]

(b) Comment on the reliability of this estimate. [2]

●●● **Question 2.** For another set of data, $r = -0.2$ and $y = -0.4x + 30$. [no GDC]

(a) Comment on using the line to make predictions. [2]

●●● **Question 3.** In a primary school, shoe size and reading score give $r = 0.8$. [no GDC]

(a) Does a bigger shoe size cause better reading? Explain. [2]

PROBLEMS

●●● **Question 4.** x: 1, 2, 3, 4, 5; y: 3, 5, 4, 7, 8. [with GDC]

(a) Find r and the line of y on x. [3]

●●● **Question 5.** The line of y on x is $y = ax + 6$, with $\bar{x} = 4$ and $\bar{y} = 24$. [no GDC]

(a) Find a. [2]

ANSWERS

1 strong positive

2 moderate negative

3 no (very weak) linear correlation

4 11

5 5

6 each extra unit of x: y up 2.5 on average

Q1 (a) 23 (b) reliable: interpolation and r close to 1

Q2 not reliable: the correlation is very weak

Q3 no: correlation is not causation; older children have bigger feet and read better

Q4 $r = 0.915;\ y = 1.2x + 1.8$

Q5 $24 = 4a + 6,\ a = 4.5$

Written and verified by a practising IB mathematics teacher, produced with AI assistance; every result recomputed by software, every figure computed from its function. Not affiliated with or endorsed by the International Baccalaureate Organization. © 2026 Hence