# Linear regression

120 marks

Hints online, answers free with an account: mathswariz.uk/worksheets/a-level-statistics-linear-regression/

Date



© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

## A Do Now

Quick questions to start. Some use earlier lessons; some set a slip that costs marks.

**A1** Needed today [2 marks]

A scatter diagram of the age of a tree, $x$ years, and its height, $y$ metres, shows strong positive correlation. Interpret this in context.

**A2** Needed today [3 marks]

A straight line passes through the points $(2, 11)$ and $(6, 23)$. Find its equation in the form $y = a + bx$.

Answer .....

**A3** Needed today [2 marks]

For the line $y = 4.2 + 0.8x$, find $y$ when $x = 7.5$, and the change in $y$ when $x$ increases by $5$.

Answer .....

**A4** Common slip [2 marks]

The cost, £$c$, of printing $n$ thousand leaflets is modelled by $c = 15 + 2.4n$. Find the cost of printing $40\,000$ leaflets.

Answer .....


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**A5** Common slip [2 marks]

A regression line of $y$ on $x$ is found from data with $x$ from $10$ to $50$. Mo uses it to estimate $y$ when $x = 35$ and when $x = 80$. For each, say whether it is interpolation or extrapolation, and which estimate is reliable.


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

## B Example, then your turn

Your teacher works through each example with you. Then try the one beside it.

**EXAMPLE 1**

A café records the daily maximum temperature, $t$ °C, and the number of cold drinks sold, $d$, on $20$ days with $t$ from $12$ to $30$. The regression line of $d$ on $t$ is $d = 18.4 + 6.25t$. (a) Interpret the value $6.25$. (b) Interpret the value $18.4$, and say whether it makes sense.

**Hint** $b$ is the change in $d$ for each $1$ °C rise in $t$. $a$ is the value of $d$ when $t = 0$: is $t = 0$ inside the data?

**B1** [4 marks]

For $25$ students, the regression line of the mark in a test, $s$ %, on the hours of revision, $r$, is $s = 31.5 + 4.2r$. The hours of revision range from $2$ to $15$. (a) Interpret the value $4.2$. (b) Interpret the value $31.5$, and say whether it makes sense here.



© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**EXAMPLE 2**

A block of ice melts. Its mass, $m$ grams, $t$ minutes after it is taken out of the freezer is modelled by $m = 480 - 6.4t$, for $0 \leq t \leq 40$. Find the change in the mass of the ice in $15$ minutes.

**Hint** The change over $15$ minutes is $15$ times the gradient. Putting $t = 15$ into the line gives the mass at $15$ minutes, not the change.

**B2** For a model of car, the value, $p$ (£ thousand), of a car aged $a$ years is modelled by the regression line $p = 21.6 - 1.85a$, for $1 \leq a \leq 8$. (a) Interpret the value $-1.85$. (b) Find the change in the value of a car between age $2$ and age $6$. [4 marks]


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**EXAMPLE 3**

For $30$ fish of one species with lengths, $l$ cm, from $18$ to $45$, the regression line of mass, $w$ grams, on length is $w = -312 + 24.5l$. (a) Estimate the mass of a fish of this species that is $0.3$ m long. (b) Explain why the line should not be used for a fish $10$ cm long.

**Hint** Check the units of $l$ before you substitute. For (b), work out the value and look at the data range.

**B3** For $40$ trees with trunk diameters, $d$ cm, from $10$ to $60$, the regression line of height, $h$ metres, on diameter is $h = 1.2 + 0.35d$. (a) Estimate the height of a tree with a trunk diameter of $0.4$ m. (b) Explain why the line should not be used for a tree with a trunk diameter of $150$ cm. [4 marks]


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**EXAMPLE 4**

For $12$ adults with heights, $h$ cm, from $150$ to $190$, the regression line of mass, $w$ kg, on height is $w = -94 + 0.95h$. (a) Estimate the mass of an adult who is $170$ cm tall. (b) Nina wants to estimate the height of an adult with mass $60$ kg. She rearranges the line. Explain why she should not.

**Hint** The line of $w$ on $h$ is chosen to make the errors in $w$ as small as possible. Which variable must be the response variable for Nina's estimate?

**B4** [4 marks]

For $20$ students who studied for $5$ to $30$ hours, the regression line of the test mark, $m$, on the hours of study, $s$, is $m = 12 + 1.6s$. (a) Estimate the mark of a student who studied for $20$ hours. (b) Jess rearranges the line to estimate the hours of study of a student who scored $60$. Explain why this is not appropriate.


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**EXAMPLE 5**

Ola measures the height, $y$ cm, of a plant every two weeks. Use your calculator to find the regression line of $y$ on $x$, where $x$ is the number of weeks, and estimate the height after $7$ weeks.

| $x$ (weeks) | $y$ (cm) |
|---|---|
| 2 | 5.1 |
| 4 | 8.9 |
| 6 | 13.2 |
| 8 | 16.8 |
| 10 | 21.1 |
| 12 | 24.9 |

**Hint** Use the statistics mode for two variables, with $x$ in the first list. Give $a$ and $b$ to $3$ significant figures.

**B5** A pool records the noon temperature, $x$ °C, and the number of visitors, $y$, on $6$ days. (a) Use your calculator to find the regression line of $y$ on $x$. (b) Interpret the gradient. (c) Estimate the number of visitors on a day when the noon temperature is $22$ °C. [4 marks]

| $x$ (°C) | $y$ |
|---|---|
| 15 | 62 |
| 20 | 70 |
| 25 | 81 |
| 30 | 88 |
| 35 | 99 |
| 40 | 106 |

Answer .....


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**C1** A ball is dropped from a bridge. Its height above the river, $h$ m, is recorded at times $t$ seconds from $0.5$ to $2.5$, and the regression line is $h = 9.6 - 1.28t$. Asked to "interpret the gradient", Ria writes: "$-1.28$: there is negative correlation." Explain why this scores no marks, and write a full interpretation. [4 marks]


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**C2** The area, $A$ cm², of a patch of mould $d$ days after it is first seen is modelled by $A = 2.8 + 0.45d$, for $1 \leq d \leq 14$. Sam is asked for the increase in the area over $6$ days. He writes: "$A = 2.8 + 0.45 \times 6 = 5.5$, so the area increases by $5.5$ cm²." Explain Sam's mistake and find the correct increase. [4 marks]

Answer .....


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**C3** For $15$ animals of one species, the regression line of body length, $l$ cm, on girth, $g$ cm, is $l = 4.6 + 0.82g$, with $g$ from $14$ to $26$. Find the estimated length of an animal with girth $20$ cm. Tom wants the girth of an animal that is $17.3$ cm long. He writes: "$g = \dfrac{17.3 - 4.6}{0.82} = 15.5$ cm." Explain why Tom's method is not suitable, and what he should do instead. [4 marks]


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**C4** A chemical breaks down. The mass left, $m$ grams, $t$ hours after the start, is recorded for $0 \leq t \leq 15$, and the regression line is $m = 85 - 3.6t$. Lily finds $m = 16.6$ when $t = 19$ and says: "This is positive, so the estimate is reliable." Explain why Lily is wrong. Then find $m$ when $t = 25$ and comment. [5 marks]


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**C5** Ash records the hours of typing practice in a week, $x$, and her typing speed, $y$ words per minute, for $8$ weeks. Use your calculator to find the regression line of $y$ on $x$. Interpret the gradient, and estimate her speed in a week with $10$ hours of practice. [5 marks]

| $x$ (hours) | 3 | 5 | 6 | 8 | 9 | 11 | 12 | 14 |
|---|---|---|---|---|---|---|---|---|
| $y$ (wpm) | 41 | 48 | 50 | 57 | 61 | 66 | 70 | 75 |

Answer .....


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**C6** A shop records the daily rainfall, $w$ mm, and the number of umbrellas sold, $s$, and finds the regression line $s = 42 + 8.5w$, for $0 \leq w \leq 12$. Ben writes: "For each extra umbrella sold, the rainfall increases by $8.5$ mm." Explain Ben's mistake, and interpret both $8.5$ and $42$ correctly. [4 marks]

## D Exam-style questions

Edexcel exam style

**D1** Jo uses the large data set. For $30$ days at Leuchars in 2015, she finds the regression line of the daily maximum gust, $g$ knots, on the daily mean windspeed, $w$ knots: $g = 6.1 + 1.92w$. The values of $w$ range from $4$ to $22$.

**(a)** Interpret the value $1.92$ in context. [2 marks]


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**(b)** Estimate the daily maximum gust on a day with a mean windspeed of $15$ knots. [2 marks]

Answer .....

**(c)** $1$ knot is $1.852$ km/h. Let $G$ and $W$ be the gust and the mean windspeed in km/h. Find the regression line of $G$ on $W$, giving the values to $3$ significant figures. [3 marks]

Answer .....


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**(d)** On one day the maximum gust was $60$ knots. Jo rearranges the line to estimate the mean windspeed that day. Give two reasons why her estimate is not reliable. [3 marks]

**Total for D1: 10 marks**

**D2**

A car is driven from a full tank. The fuel left, $f$ litres, is recorded at $12$ distances, $d$ km, from $0$ to $600$. The regression line of $f$ on $d$ is $f = 52.4 - 0.068d$. Tim writes:

(1) "The gradient $-0.068$ is close to $0$, so the correlation is weak."

(2) "After $50$ km, the fuel has gone down by $52.4 - 0.068 \times 50 = 49$ litres."

(3) "The tank will be empty at $d = \dfrac{52.4}{0.068} = 771$ km."

(4) "The value $52.4$ means the tank holds exactly $52.4$ litres."

Tim has made a mistake in each statement.

**(a)** Explain what is wrong with statement (1). [2 marks]


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**(b)** Correct statement (2). [2 marks]

Answer .....

**(c)** Give two reasons why statement (3) is not reliable. [3 marks]

**(d)** Explain what is wrong with statement (4). [2 marks]

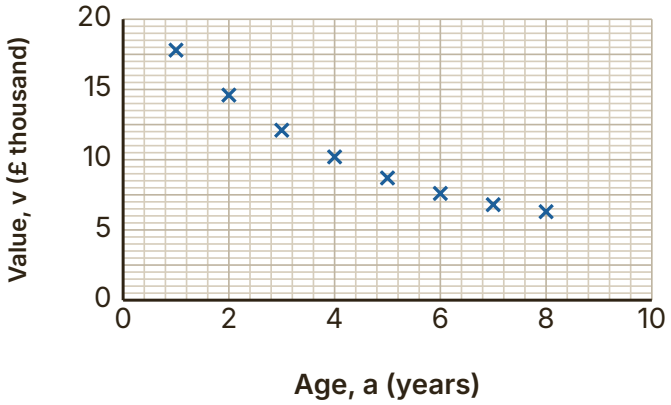
**Total for D2: 9 marks**


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**D3** On $10$ summer days, a shop records the noon temperature, $x$ °C, and the number of ice lollies sold, $y$. The temperatures range from $11$ to $26$ °C. $\sum x = 182$ and $\sum y = 1037$. The regression line of $y$ on $x$ has gradient $4.5$ and passes through the point $(\bar{x}, \bar{y})$. Find the equation of the regression line, interpret the gradient, and estimate the number of lollies sold on a day when the noon temperature is $20$ °C, saying whether your estimate is reliable. Explain whether the value of $a$ has a sensible meaning here. [8 marks]


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**D4** The scatter diagram shows the age, $a$ years, and the value, $v$ (£ thousand), of $8$ cars of the same model. For ages $1$ to $8$ the values are $17.8$, $14.6$, $12.1$, $10.2$, $8.7$, $7.6$, $6.8$ and $6.3$. The regression line of $v$ on $a$ is $v = 17.7 - 1.60a$.




**(a)** Interpret the value $-1.60$ in context. [2 marks]

**(b)** Use the line to estimate the value of a $4$-year-old car, and compare it with the value in the data. [2 marks]

Answer .....


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**(c)** Use the line to estimate the value of a $12$-year-old car, and comment on your answer. [3 marks]

**(d)** With reference to the scatter diagram, explain why a linear model is not the best model for these data. [3 marks]

**Total for D4: 10 marks**


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**D5** A shop sells bars of chocolate of different masses. For $8$ bars, the mass, $x$ grams, and the price, $y$ pence, are shown in the table. The regression line of $y$ on $x$ is $y = 11.8 + 0.958x$.

| $x$ (g) | 25 | 30 | 40 | 45 | 50 | 65 | 70 | 75 |
|---|---|---|---|---|---|---|---|---|
| $y$ (p) | 35 | 41 | 50 | 56 | 60 | 73 | 79 | 84 |

**(a)** Interpret the value $0.958$ in context. [2 marks]

**(b)** Estimate the price of a $55$ g bar. [2 marks]

Answer .....


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**(c)** Another shop sells a $60$ g bar for $90$ p. Use the line to decide whether this is a fair price. [3 marks]

**(d)** Explain why the line should not be used to estimate the price of a $500$ g bar. [2 marks]

**Total for D5: 9 marks**


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**D6** Sara measures the height, $h$ cm, and the arm span, $s$ cm, of $7$ students.

| $h$ (cm) | 152 | 158 | 161 | 165 | 170 | 174 | 180 |
|---|---|---|---|---|---|---|---|
| $s$ (cm) | 150 | 157 | 163 | 166 | 171 | 176 | 183 |

**(a)** Sara wants to estimate arm span from height. Use your calculator to find the regression line she should use. [3 marks]

Answer .....

**(b)** Estimate the arm span of a student who is $165$ cm tall. [2 marks]

Answer .....


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**(c)** Sara says that a person's arm span is equal to their height. [3 marks]
Use your line to estimate the arm spans of students of heights $155$ cm and $175$ cm, and hence comment on her claim.

Answer .....

**Total for D6: 8 marks**


© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

No route is given. The hints are at the end of the sheet. Use one at a time.

**E1** Data on two variables, $x$ and $y$, are coded with $u = \dfrac{x - 50}{10}$ and $v = \dfrac{y - 200}{5}$. The regression line of $v$ on $u$ is $v = 1.2 + 0.4u$.

**(a)** Find the regression line of $y$ on $x$ in the form $y = a + bx$. [5 marks]

Answer .....

**(b)** Estimate $y$ when $x = 72$, and the change in $y$ when $x$ increases by $15$. [2 marks]

Answer .....



© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

**(c)** The mean of $x$ is $63$. Find the mean of $y$. [2 marks]

Answer .....

**Total for E1: 9 marks**

## Extension hints

Cover these with paper. Uncover one at a time.

**E1**

**Hint 1** Replace $u$ and $v$ in the coded line by their expressions in $x$ and $y$.

**Hint 2** Multiply both sides by $5$ and add $200$ to make $y$ the subject.

**Hint 3** A regression line passes through the point of means, and coding changes the means in the same way as each value.

## How did it go?

Colour one circle on each line: not yet, nearly, yes.

| | |
|---|---|
| I can interpret the gradient and the intercept of a regression line in context, with units. | ○ ○ ○ |
| I can work out the change in y for a change in x, and estimate y from x. | ○ ○ ○ |
| I can say whether an estimate is interpolation or extrapolation, and whether it is reliable. | ○ ○ ○ |
| I can explain why the line of y on x should not be used to estimate x. | ○ ○ ○ |



© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.