

Scatter Plots & Lines of Best Fit

Simulation: edtechsims.com/line-of-best-fit

Name: _____ Class: _____ Date: _____



Scan to open

By the end of this activity, you should be able to:

- plot a scatter plot and describe the correlation it shows
- draw a line of best fit by eye and use it to make estimates
- explain how an outlier affects a line of best fit
- explain the difference between interpolation and extrapolation
- understand that correlation does not mean one variable causes the other

Part A: Exploring the Scatter Plot

On the sim: Read Learning Goals and How to use. Click on the graph to add points; click a point to remove it. Watch the red dashed line of best fit.

A1 Complete the sentences.

[3]

A line of best fit shows the general _____ of the data points.

In a positive correlation, as one variable increases, the other _____.

In a negative correlation, as one variable increases, the other _____.

A2 Work through each Try This! challenge. Record the Correlation (r) value and the description shown (e.g. Strong Positive), and describe what happens to the line.

[4]

Try This!	r value	Description	What happens to the line of best fit?
Add points that create a perfect line			
Add an outlier far from the others			
Create a negative correlation			
Make points with no clear pattern			

Table 1

A3 What does an r value close to 1, close to -1 and close to 0 tell you about the data?

[3]

Close to 1:

Close to -1 :

Close to 0:

A4 Explain why one outlier can change the line of best fit a lot when there are only a few points. [2]

.....

Part B: Draw Your Own Line of Best Fit

On the sim: Go to the Practice Activity. For each dataset, adjust the slope and y-intercept sliders until your line fits the data, click Check My Line, then click Show Best Fit Line.

B1 Record your results in Table 2. [5]

Dataset	My slope (m)	My y-intercept (b)	Feedback	Actual best fit line
1: Positive Trend				
2: Strong Positive Trend				
3: Negative Trend				
4: Gentle Positive Trend				
5: Scattered Data				

Table 2

B2 Which dataset was the hardest to fit? Explain why. [2]

.....

.....

B3 For the Negative Trend dataset, was your slope positive or negative? Explain. [1]

.....

.....

B4 In the equation $y = mx + b$, what do m and b tell you about the line? [2]

m:

b:

B5 Tick (✓) the tips for drawing a good line of best fit by eye. [3]

- ☐ Use a ruler to draw one straight line.
- ☐ Join the first point to the last point.
- ☐ Have roughly the same number of points above and below the line.
- ☐ Make the line pass through the origin every time.
- ☐ Ignore any obvious outlier.
- ☐ Join up all the points dot-to-dot.

Part C: Practice Questions

C1 The table shows the number of hours eight students studied and their test scores.

Hours studied	1	2	3	4	5	6	7	8
Test score	35	42	50	52	61	66	74	80

(a) On Fig. 1, plot the scatter plot.

[2]

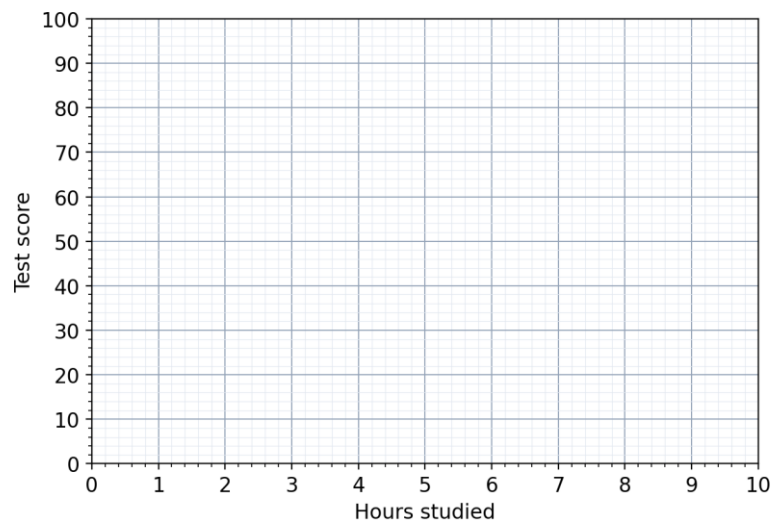


Fig. 1

(b) Describe the correlation shown.

[1]

.....

(c) Draw a line of best fit on Fig. 1.

[1]

(d) Use your line to estimate the test score of a student who studied for 5.5 hours.

[1]

score = ...

(e) Explain why it would not be reliable to use your line to estimate the score of a student who studied for 15 hours.

[2]

.....

.....

C2 Fig. 2 shows the temperature on 11 days and the number of hot drinks sold by a café.

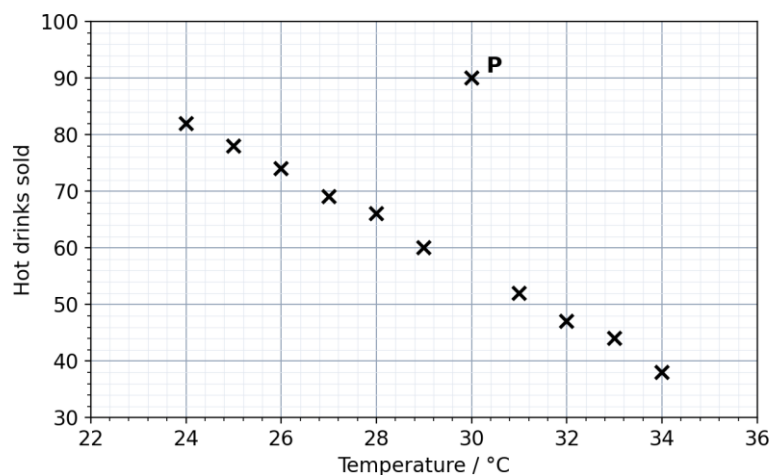


Fig. 2

(a) Describe the relationship between the temperature and the number of hot drinks sold. [2]

.....

.....

(b) Point P does not fit the pattern. What is this point called? Suggest a reason for it. [2]

.....

.....

(c) Draw a line of best fit on Fig. 2, ignoring point P. [1]

(d) Estimate the number of hot drinks sold on a day when the temperature is 27.5 °C. [1]

..... drinks

C3 A study found a positive correlation between ice cream sales and the number of people who drown at the beach. Jun says, "Eating ice cream causes people to drown." Explain why Jun is wrong. [2]

.....

.....

.....

Answer Key: Scatter Plots & Lines of Best Fit

Part A

A1 trend (pattern); increases; decreases

A2 Values from the sim. The sim labels $|r| \geq 0.9$ Very Strong, ≥ 0.7 Strong, ≥ 0.5 Moderate, ≥ 0.3 Weak, otherwise Very Weak. Perfect line: $r = 1$ (or -1), line passes through every point. Outlier: line tilts towards the outlier and r moves closer to 0. Negative: r is negative and the line slopes down. No pattern: r is close to 0 and the line is nearly flat.

A3 Close to 1: strong positive correlation. Close to -1 : strong negative correlation. Close to 0: little or no linear correlation.

A4 The line is calculated from all the points. With few points, each point has a large effect, so one point far from the others pulls the line towards it.

Part B

B1 Student values from the sim. Actual best fit lines (shown after Show Best Fit Line): Dataset 1: $y = 0.60x + 54.00$; Dataset 2: $y = 1.20x + 24.00$; Dataset 3: $y = -0.80x + 252.00$; Dataset 4: $y = 0.26x + 84.40$; Dataset 5: $y = 0.44x + 75.20$.

B2 Usually Dataset 5 (Scattered Data): the points do not lie close to a straight line, so many lines look equally good.

B3 Negative. As x increases, y decreases, so the line slopes downwards.

B4 m is the slope (gradient): how steep the line is and whether it goes up (positive) or down (negative). b is the y -intercept: where the line crosses the y -axis.

B5 Tick: use a ruler; roughly the same number of points above and below; ignore an obvious outlier.

Part C

C1(a) All eight points plotted correctly (1 mark for 6–7 correct).

C1(b) Positive correlation (strong).

C1(c) Straight ruled line with points evenly on both sides, e.g. passing close to (1, 35) and (8, 80), and through the mean point (4.5, 57.5).

C1(d) About 64 (accept 61–67, read from the student's line).

C1(e) 15 hours is far outside the range of the data (1–8 hours); this is extrapolation. The pattern may not continue, and a score cannot exceed the maximum mark.

C2(a) Negative correlation: as the temperature increases, the number of hot drinks sold decreases.

C2(b) An outlier. e.g. a special promotion, an event nearby, or a recording error.

C2(c) Straight ruled line through the remaining points, e.g. from about (24, 82) to (34, 38).

C2(d) About 67 (accept 64–70, read from the student's line).

C3 Correlation does not mean causation. Both are linked to a third factor: hot weather. More people buy ice cream and more people swim when it is hot, so both increase together.