Correlation

85 marks

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

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]

For a set of daily maximum gusts, $Q_1 = 18$ knots and $Q_3 = 30$ knots. An outlier is a value more than $1.5\times$ the interquartile range above Q_3 or below Q_1. Is a gust of 50 knots an outlier? Show your working.

Answer .....

A2 Needed today [3 marks]

For ten values, $\sum x = 84$ and $\sum x^2 = 762$. Find the mean and the standard deviation.

Answer .....

A3 Needed today [2 marks]

In the large data set, the daily total rainfall for some days is recorded as "tr". State what "tr" means, and the value you use for it in a calculation.

A4 Common slip [2 marks]

Sam says: "Negative correlation means that as one variable decreases, the other decreases." Correct Sam's statement.

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

A5 Common slip [2 marks]

A farmer chooses how much fertiliser to put on each of 8 plots and then measures the yield of each plot. State which variable is the explanatory variable, and which axis it goes on in a scatter diagram.

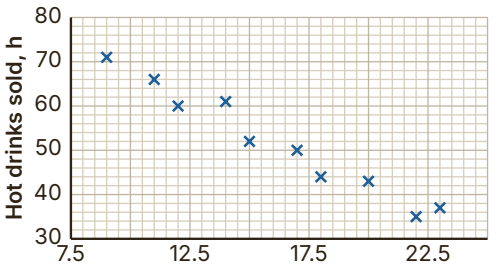
© 2026 Mathswariz. All rights reserved. Free for personal and classroom use. Not to be sold or republished.

PART 1 Describing and interpreting correlation

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

EXAMPLE 1 · I DO

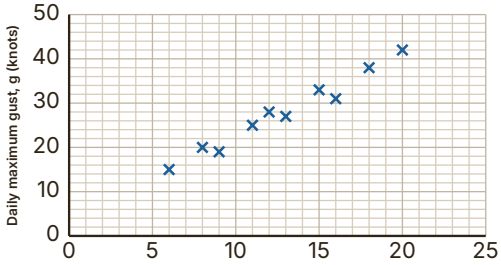
A café records the daily mean temperature, t °C, and the number of hot drinks sold, h, on 10 days. The scatter diagram shows the results. (a) State which variable is the explanatory variable. (b) Describe the correlation. (c) Interpret the correlation in context.



Hint "Describe" asks for the type and strength only. "Interpret" asks for a sentence about the two variables in the context of the question.

YOUR TURN · WE DO

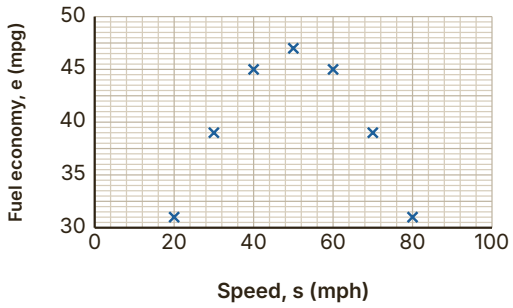
B1 The scatter diagram shows the daily mean windspeed, w knots, and the daily maximum gust, g knots, at a weather station on 10 days. (a) State, with a reason, which variable is the explanatory variable. (b) Describe the correlation. (c) Interpret the correlation in context. [4 marks]



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

EXAMPLE 2 · I DO

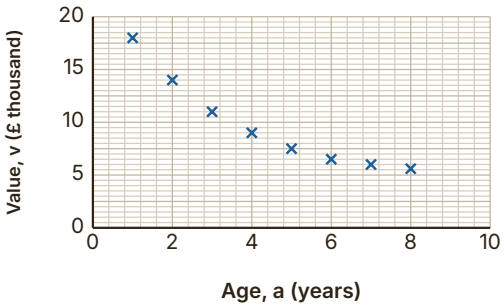
The scatter diagram shows the speed, s mph, at which a car is driven and its fuel economy, e miles per gallon, on 7 test runs. Ella says: "There is no correlation, so the fuel economy does not depend on the speed." Describe the correlation, and comment on Ella's statement.



Hint Correlation measures how close the points are to a straight line. Look for a pattern that is not a straight line.

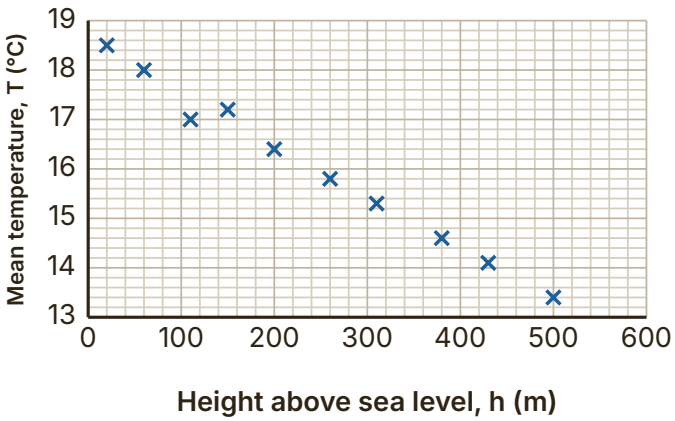
YOUR TURN · WE DO

B2 The scatter diagram shows the age, a years, and the value, v (£ thousand), of 8 cars of the same model. (a) Describe the correlation. (b) Lee says: "The value of the car falls by the same amount each year." Comment on Lee's statement, referring to the scatter diagram. [4 marks]



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

C1 The scatter diagram shows the height above sea level, h metres, and the mean daily temperature, T °C, of 10 towns on one day. The question asks: "Interpret the correlation." Ben writes: "Negative correlation." Explain why Ben's answer would not score the mark. Then describe the correlation, and give a full interpretation. [4 marks]



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

C2 Tom investigates whether hours of revision affect test marks. He plots the test mark on the horizontal axis and the hours of revision on the vertical axis, and finds strong positive correlation. (a) Explain Tom's mistake. (b) Tom says: "This shows that revising longer causes higher marks." Give one reason why the scatter diagram alone does not show this. [4 marks]

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

PART 2 Correlation and cause

EXAMPLE 3 · I DO

For 12 weeks one summer, a beach shop records the number of ice creams it sells, and a first-aid post records the number of people it treats for sunburn. The scatter diagram shows strong positive correlation. The shop manager says: "Selling ice creams causes sunburn." Give a reason why the manager's conclusion is not valid, and suggest a variable that could explain the correlation.

Hint Ask whether one variable can really change the other. If not, look for a third variable that makes both go up together.

YOUR TURN · WE DO

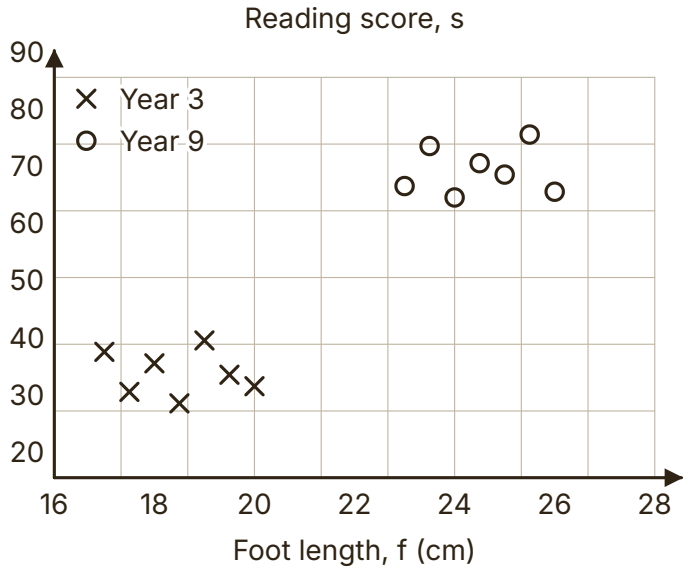
B3 For 15 towns, Ruth finds strong positive correlation between the number of libraries and the number of road accidents in a year. She says: "Building libraries leads to more road accidents." (a) Explain why Ruth's conclusion is not valid. (b) Suggest a third variable that could explain the correlation, and say how it affects both. [4 marks]

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

C3 For 30 countries, there is strong positive correlation between the number of mobile phones per 100 people and the life expectancy. Priya says: "This proves that giving people mobile phones makes them live longer." Give two things that are wrong with Priya's statement, and suggest a variable that could explain the correlation. [4 marks]

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

C4 Hal measures the foot length, f cm, and the reading score, s, of 7 pupils in Year 3 and 7 pupils in Year 9. The scatter diagram shows the results. Hal says: "There is strong positive correlation, so pupils with bigger feet read better." (a) Describe the correlation for all 14 pupils, and within each year group. (b) Explain what causes the correlation for all 14 pupils, and comment on Hal's statement. [5 marks]



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

PART 3 Outliers and the large data set

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

EXAMPLE 4 · I DO

For 9 students, Jo records the hours of sleep, x, the night before a test, and a reaction time, y milliseconds. The results are shown in the table. For the reaction times, $Q_1 = 300$ and $Q_3 = 333$. An outlier is a value more than $1.5 \times (Q_3 - Q_1)$ above Q_3 or below Q_1. (a) Show that 950 is an outlier. (b) Jo thinks a digit was typed wrongly. Explain whether she should remove this pair, and describe the correlation of the other 8 pairs.

x (hours)	y (ms)
5	340
5.5	333
6	325
6.5	318
7	950
7.5	307
8	300
8.5	296
9	291

Hint Work out the two fences. For (b): an error in recording is a reason to remove a value; a genuine but extreme value is not.

YOUR TURN · WE DO

B4 Leah records the daily total rainfall, r mm, at a weather station on 12 days. The mean is 2.4 mm and the standard deviation is 3.1 mm. An outlier is a value more than 2 standard deviations from the mean. (a) Show that a rainfall of 9.8 mm is an outlier. (b) The rainfall was recorded correctly on a day with a thunderstorm. Leah plots the rainfall against the daily total sunshine. Explain whether she should leave out this day. [4 marks]

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

EXAMPLE 5 · I DO

Kai uses the large data set. He takes the daily mean temperature and the daily mean pressure for the first 14 days of May 2015 at Hurn and draws a scatter diagram. It shows weak positive correlation. Kai says: "At every weather station, days with higher pressure have higher temperatures." (a) Name the type of sampling Kai has used. (b) Give two reasons why Kai's conclusion may not be valid.

Hint Think about how the days were chosen, how many there are, and how much of the large data set they come from.

YOUR TURN · WE DO

B5 Mia wants to use the large data set to investigate whether daily mean cloud cover and daily total sunshine are correlated at Perth. (a) Explain why she cannot do this. (b) Mia uses Heathrow instead. She takes the first 10 days of June 2015 and finds strong negative correlation. Give two reasons why her result may not describe the whole of the Heathrow data for 2015. [4 marks]

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

C5 Alex uses the large data set to draw a scatter diagram of daily total rainfall, r mm, against daily total sunshine, s hours, for 20 days at Leuchars. (a) Three of the rainfall values are "tr" and one is "n/a". Explain how Alex should deal with each. (b) Alex wants to repeat the study for Perth using daily mean cloud cover instead of rainfall. Explain why he cannot. [4 marks]

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

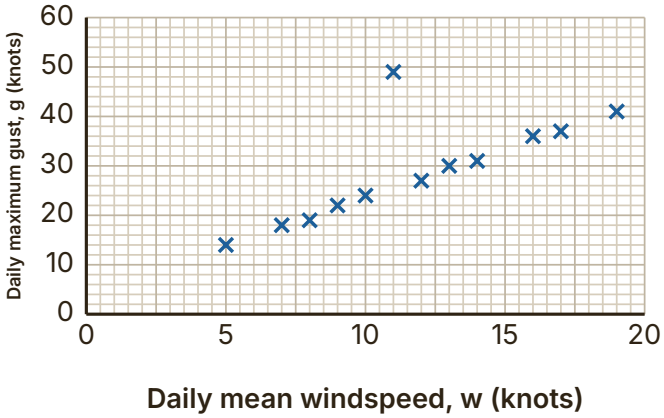
C6 For 10 students, y is the mark in a test. $\sum y = 480$ and $\sum y^2 = 24\,456$. An outlier is a value more than 2 standard deviations from the mean. Zoe looks at a scatter diagram of y against the hours of revision and says: "The mark of 70 is the highest and far from the others, so it is an outlier." Explain the mistake in Zoe's method, and use the rule to decide whether 70 is an outlier. [5 marks]

Answer .....

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

D1 Jo records the daily mean windspeed, w knots, and the daily maximum gust, g knots, at a UK weather station on 12 days. The scatter diagram shows her results. For the 12 gusts, $Q_1 = 20.5$ and $Q_3 = 36.5$.

(a) Describe and interpret the correlation between w and g. [2 marks]



(b) An outlier is a value more than $1.5 \times (Q_3 - Q_1)$ above Q_3 or below Q_1. Determine whether the gust of 49 knots is an outlier. [3 marks]

Answer .....

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

(c) Explain why the point $(11, 49)$ still stands out on the scatter diagram, and give a reason why it may be a correct reading. [3 marks]

(d) Jo says that a higher mean windspeed causes a higher maximum gust. Comment on whether a causal relationship is likely here. [2 marks]

Total for D1: 10 marks

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

D2

For 15 students in her class, Amy records the hours of television watched in a week, x, and the mark in a test, y. For the hours, the mean is 12 and the standard deviation is 4.5. Amy writes:
(1) "The scatter diagram shows negative correlation: as the hours of television go down, the marks go down."
(2) "This proves that watching television causes lower marks."
(3) "The student who watched 20 hours is an outlier, because 20 is far from the rest, so I removed this student."
(4) "So for all Year 12 students, more television means lower marks."

Amy has made a mistake in each statement.

(a) Correct statement (1). [2 marks]

(b) Explain what is wrong with statement (2), and suggest another explanation for the correlation. [3 marks]

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

(c) An outlier is a value more than 2 standard deviations from the mean. Test statement (3). [2 marks]

Answer .....

(d) Give one reason why statement (4) may not be valid. [2 marks]

Total for D2: 9 marks

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

E Extension

Stretch

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

E1 The times, in minutes, that 7 students take to solve a puzzle are $4, 6, 7, 9, 10, 12$ and k, where $k > 0$. An outlier is a value more than 2 standard deviations from the mean, where the mean and the standard deviation are worked out from all 7 times, including k. Find, in exact form, the set of values of k for which k is an outlier. Hence find the least whole number value of k for which k is an outlier. [9 marks]

Answer .....

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

Extension hints

Cover these with paper. Uncover one at a time.

E1

Hint 1 Write the mean and the variance of the 7 times in terms of k. The other six times have sum 48 and sum of squares 426.

Hint 2 For a large k, the condition is $k - \bar{x} > 2\sigma$. Square both sides (both are positive) to remove the square root.

Hint 3 Multiply through by 49 and simplify. You should reach a quadratic inequality in k with $k^2 - 16k$ in it.

How did it go?

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

I can describe the correlation on a scatter diagram, and interpret it in context.	○ ○ ○
I can say which variable is the explanatory variable and which is the response variable.	○ ○ ○
I can explain why correlation does not show that one variable causes the other.	○ ○ ○
I can test a value for an outlier with a given rule and decide whether to remove it.	○ ○ ○

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