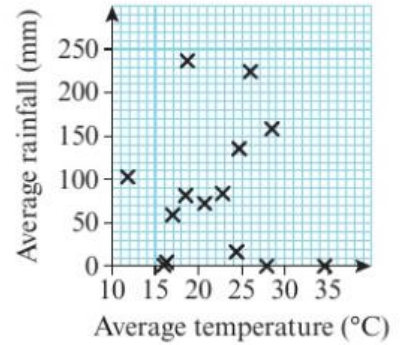# Correlation

1 The average temperature and rainfall were collected for a
number of cities around the world.
The scatter diagram shows this information.

   a Describe the correlation between average temperature
and average rainfall.

   b Comment on the claim that hotter cities have less rainfall.

2

A survey of British towns recorded the number of serious road accidents in a week (x) in each
town, together with the number of fast food restaurants (y). The data showed a strong positive
correlation. Katie states that this shows that building more fast food restaurants in her town will
cause more serious road accidents. Explain whether the data supports Katie's statement.

3 Suggest a value of r for each of these scatter diagrams:
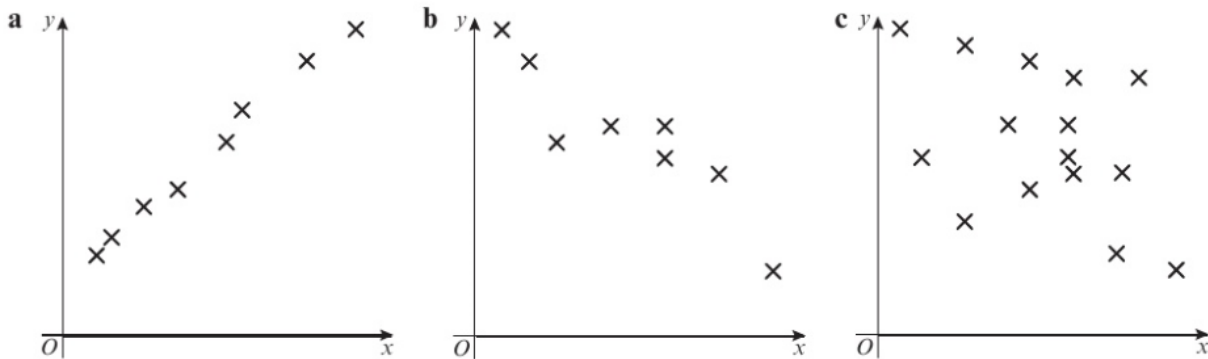
4 The following table shows 10 observations from a bivariate data set.

v	50	70	60	82	45	35	110	70	35	30
m	140	200	180	210	120	100	200	180	120	60

   a State what is measured by the product moment correlation coefficient.

   b Use your calculator to find the value of the product moment correlation coefficient between
v and m.

5

In a training scheme for young people, the average time taken for each age group to reach a
certain level of proficiency was measured. The table below shows the data.

Age, x (years)	16	17	18	19	20	21	22	23	24	25
Average time, y (hours)	12	11	10	9	11	8	9	7	6	8

   a Use your calculator to find the value of the product moment correlation coefficient for these
data.

   b Use your answer to part **a** to describe the correlation between the age and average time taken
based on this sample.

Answers to correlation

1

a There is no correlation.

b The scatter graph does not support the statement that hotter cities have less rainfall.

2

The data shows that the number of serious road accidents in a week strongly correlates with the number of fast food restaurants. However, it does not show whether the relationship is causal. Both variables could correlate with a third variable, e.g. the number of roads coming into a town.

3

a $r = 0.9$ is a good approximation, since the points lie roughly, but not exactly, on a straight line. Remember that the value of r tells you how 'close' the data is to having a perfect positive or negative linear relationship.

b Clearly r is negative, and the data is not as close to being linear as in part **a**. $r = -0.7$ is therefore a good approximation.

c The data seems to have some negative correlation, but is rather 'random'. Because so many points would lie far away from a line of best fit, $r = -0.3$ is a good approximation.

4

a The product moment correlation coefficient gives the type (positive or negative) and strength of linear correlation between v and m.

b By inputting the (ordered) data into your calculator, $r = 0.870$ (to 3 s.f.).

5

a $r = -0.854$ (to 3 s.f.)

b There is a negative correlation. The relatively older young people took less time to reach the required level.