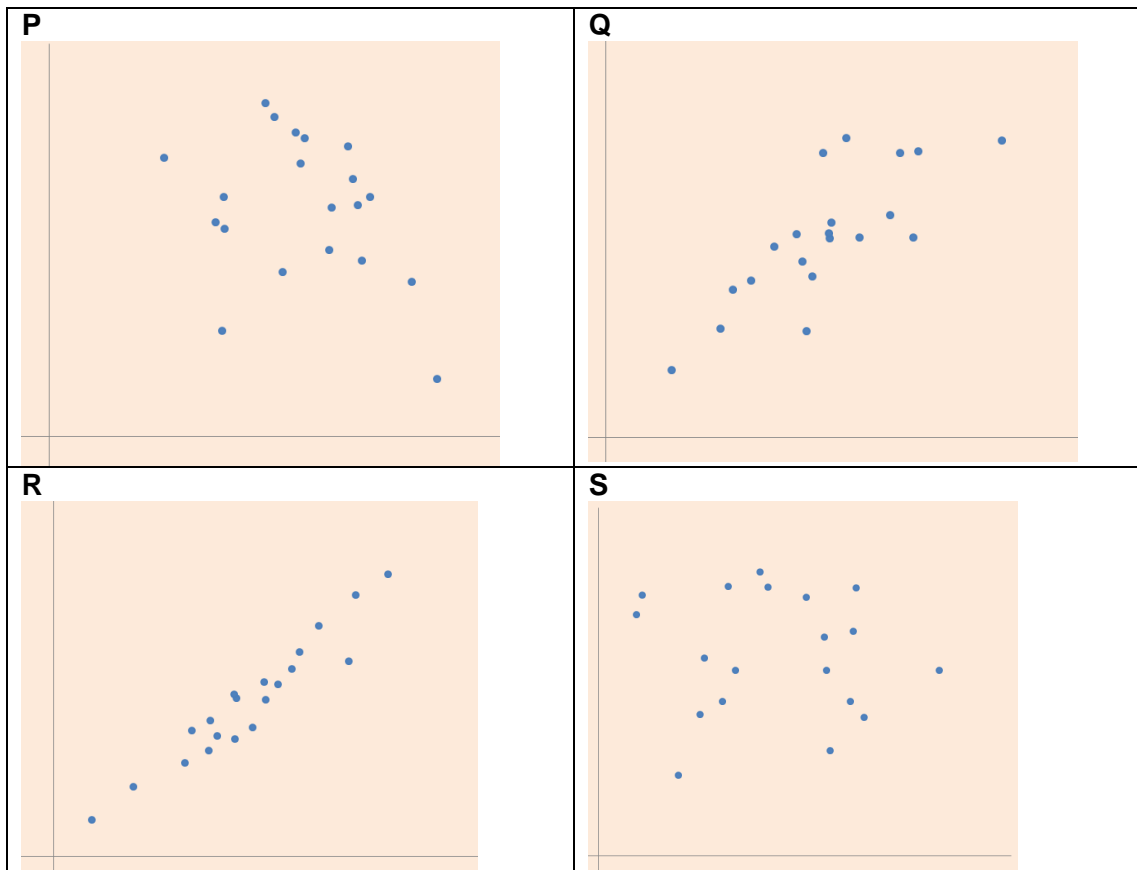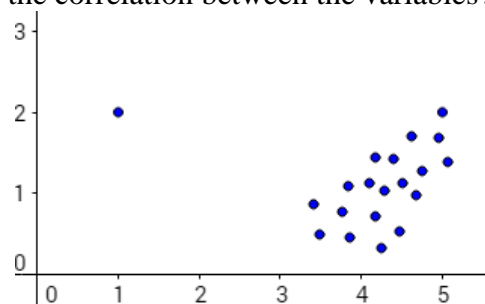## Section 3: Bivariate data

Section test

1. For the scatter diagrams shown below:
   Which scatter diagram shows strong positive correlation?
   Which scatter diagram shows moderate positive correlation?
   Which scatter diagram shows no correlation?
   Which scatter diagram shows negative correlation?

2. Look at the scatter diagram below. What effect would removing the outlier have on the correlation between the variables?


   (a) The correlation would be become stronger
   (b) The correlation would become weaker
   (c) The correlation would stay the same
   (d) It is not possible to tell

Edexcel AS Maths Data 3 section test solutions

3. The scatter diagram below is for a sample from a much larger population. Which of the statements is / are true?
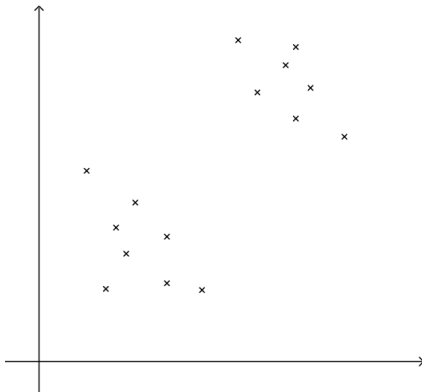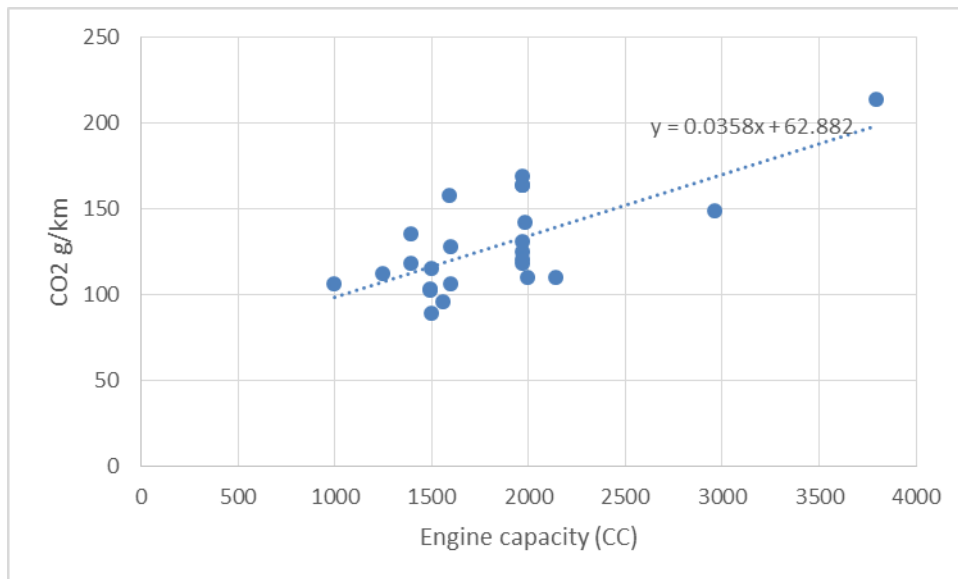

(a) There is positive correlation overall
(b) There seem to be two separate groups, each of which has positive correlation
(c) There is more than one outlier
(d) There must be two separate groups in the population

4. Which of the equations below is a possible equation for the line of best fit for the following scatter diagram?
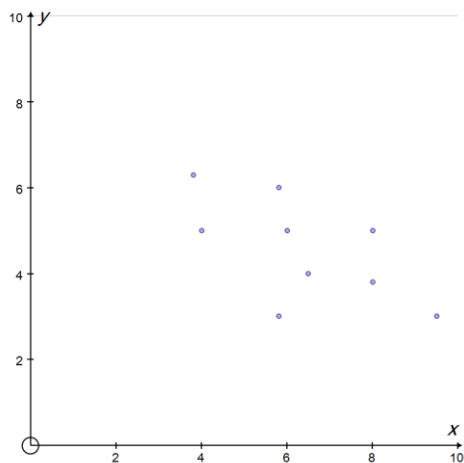
(a) $y = -0.4x + 7$
(b) $y = 0.4x - 7$
(c) $y = x$
(d) $y = 8 - x$

5. The scatter diagram shows engine capacity and CO_2 emissions per km for a sample of cars. The line of best fit has been drawn using a spreadsheet.

Edexcel AS Maths Data 3 section test solutions

Which of the following statements is true?

(a) There is positive correlation between engine capacity and CO_2 emissions

(b) Each extra cc of engine capacity increases the CO_2 emissions by nearly 63 g per km

(c) Each extra 100 cc of engine capacity increases the CO_2 emissions by around 3.6 g per km

Edexcel AS Maths Data 3 section test solutions

Solutions to section test

1. Strong positive correlation means that the data points are close to a straight line with positive gradient. This is the case in diagram R.

    Diagram Q shows moderate positive correlation; the data points are reasonably close to a straight line with positive gradient.

    Diagram S shows no correlation; the data points are not close to a straight line.

    Negative correlation means that as one variable increases, the other one tends to decrease. The only scatter diagram for which this is the case is P.

2. The correlation would become stronger, since the remaining data points are close to a straight line.

3. It is true that there is positive correlation overall; as one variable increases the other one also tends to increase

    Although there are two separate groups, if these are considered separately they each appear to have negative correlation. So this statement is not true.

    Since there are two separate groups, it does not really make sense to consider outliers.

    It is not true that there must be two separate groups in the population; we do not know.

4. The data shows negative correlation, so the line of best fit has negative gradient. The line $y = -0.4x + 7$ passes through $(0, 7)$ and $(10, 3)$ and this seems reasonable (whereas the line $y = 8 - x$ passes through $(0, 8)$ and $(8, 0)$ which lies below most of the points)

5. The gradient of the line of best fit is positive so the correlation is positive.

    The gradient of the line tells us that for every increase of 1 cc of engine capacity, there is an increase of 0.0358 g/km of CO_2 emissions. Therefore, for every 100 cc of engine capacity, there is an increase of 3.58 g/km of CO_2 emissions.