

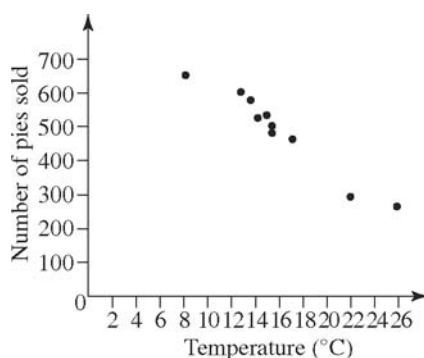
Chapter 4 — Bivariate data

Exercise 4A — Scatterplots

- 1 a Acidity is the independent variable (x -axis).
Height of corn is the dependent variable (y -axis).
- b Price is the independent variable (x -axis).
Amount purchased is the dependent variable (y -axis).
- c Time is the independent variable (x -axis).
Amount of drug is the dependent variable (y -axis).
- d Age of machine is the independent variable (x -axis).
Running cost is the dependent variable (y -axis).
- e Circumference of biceps is the independent variable (x -axis).
Performance on a lift test is the dependent variable (y -axis).

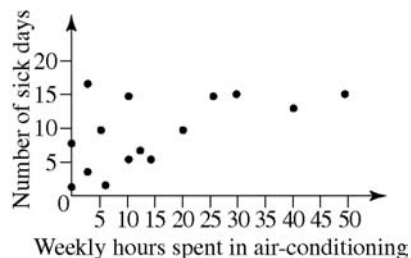
2 A & c, B & b, C & g, D & e, E & d, F & f, G & a.

3 a



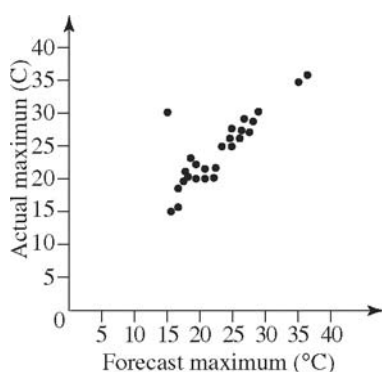
- b Strong negative correlation. The greater the temperature, the fewer pies sold.
- c Other factors could affect sale of pies, i.e. cost; the number of people at the match; the number of pies he has.

4 a



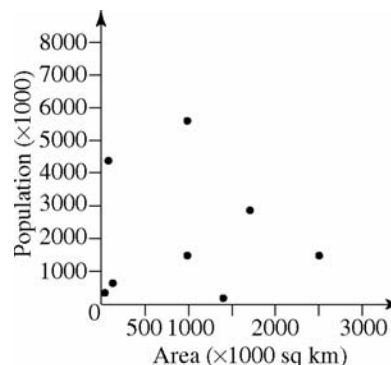
- b Moderate positive correlation.
There is evidence to show that the greater the number of hours spent in air-conditioned buildings, the greater the number of sick days.
- c It may be incorrect to say that air conditioning is the cause of poor health as there could be many other factors such as poor diet, lack of exercise, viruses etc.

5 a



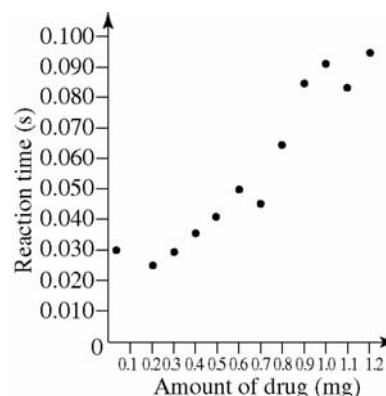
- b Strong positive correlation.
The greater the forecast maximum, the greater the actual maximum.
- c The evidence here suggests that weather forecasters tend to 'get it right', that is, their forecasts are close to the actual maximums.

6 a



- b No correlation.
There is no relationship between the variables.

7 a



- b Strong positive correlation.
The greater the amount of drug, the greater the reaction time.
 - c The age or body size of the people given the drug.
- 8 The amount of growth hormone is the independent variable (x -axis). The size is the dependent variable (y -axis).
Answer is **D**.
- 9 The greater the level of the independent variable, the smaller the level of the dependent variable.
∴ Strong negative correlation.
Answer is **E**.
- 10 There is evidence to show that the greater the level of the independent variable, the greater the level of the dependent variable.
∴ Moderate positive correlation.
Answer is **B**.
- 11 There is evidence to show that the greater the level of pollution, the lower the fish population.
Answer is **B**.

Exercise 4B — The correlation coefficient

- 1 a Strong positive correlation.
 b Weak positive correlation.
 c Strong positive correlation.
 d Moderate negative correlation.
 e Weak positive correlation.
 f No correlation.
 g Strong positive correlation.
 h Moderate negative correlation.
 i Weak negative correlation.
 j Strong negative correlation.

- 2 a i $a = 5, b = 0, c = 5, d = 0$

$$q = \frac{(a+c)-(b+d)}{a+b+c+d}$$

$$= \frac{(5+5)-(0+0)}{5+0+5+0}$$

$$= 1.0$$

- ii Strong positive correlation.

- b i $a = 4, b = 1, c = 4, d = 1$

$$q = \frac{(a+c)-(b+d)}{a+b+c+d}$$

$$= \frac{(4+4)-(1+1)}{4+1+4+1}$$

$$= 0.6$$

- ii Moderate positive correlation.

- c i $a = 0, b = 4, c = 1, d = 3$

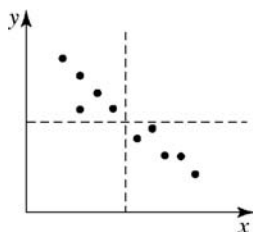
$$q = \frac{(a+c)-(b+d)}{a+b+c+d}$$

$$= \frac{(0+1)-(4+3)}{0+4+1+3}$$

$$= -0.75$$

- ii Strong negative correlation.

- 3 a The median is the $\frac{10+1}{2} = 5.5$ th score.



- i $a = 0, b = 5, c = 0, d = 5$

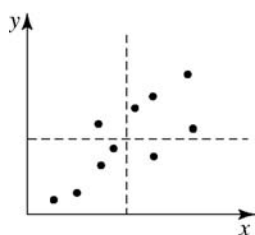
$$q = \frac{(a+c)-(b+d)}{a+b+c+d}$$

$$= \frac{(0+0)-(5+5)}{0+5+0+5}$$

$$= -1.0$$

- ii Strong negative correlation.

- b The median is the $\frac{10+1}{2} = 5.5$ th score.



- i $a = 4, b = 1, c = 4, d = 1$

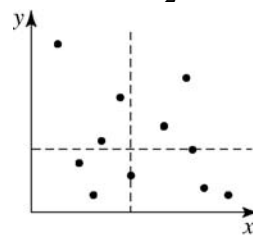
$$q = \frac{(a+c)-(b+d)}{a+b+c+d}$$

$$= \frac{(4+4)-(1+1)}{4+1+4+1}$$

$$= 0.6$$

- ii Moderate positive correlation.

- c The median is the $\frac{11+1}{2} = 6$ th score.



- i $a = 2, b = 3, c = 2, d = 2$

$$q = \frac{(a+c)-(b+d)}{a+b+c+d}$$

$$= \frac{(2+2)-(3+2)}{2+3+2+2}$$

$$= -0.11$$

- ii No correlation.

- 4 $a = 5, b = 1, c = 5, d = 1$

$$q = \frac{(a+c)-(b+d)}{a+b+c+d}$$

$$= \frac{(5+5)-(1+1)}{5+1+5+1}$$

$$= 0.67$$

Answer is **B**.

- 5 $a = 1, b = 3, c = 1, d = 3$

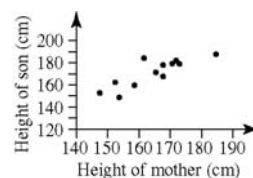
$$q = \frac{(a+c)-(b+d)}{a+b+c+d}$$

$$= \frac{(1+1)-(3+3)}{1+3+1+3}$$

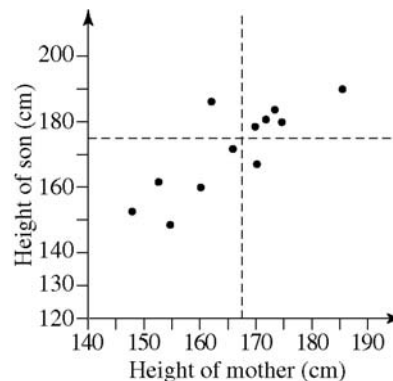
$$= -0.5$$

Answer is **C**.

- 6 a



- b



The median is the $\frac{12+1}{2} = 6.5$ th score.

- $a = 5, b = 1, c = 5, d = 1$

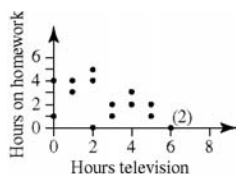
$$q = \frac{(a+c)-(b+d)}{a+b+c+d}$$

$$= \frac{(5+5)-(1+1)}{5+1+5+1}$$

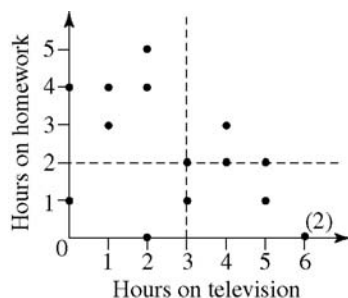
$$= 0.67$$

- c There is a moderate positive correlation.
There is evidence to show that the greater the height of the mother, the greater the height of the son.

7 a



b



The median is the $\frac{15+1}{2} = 8$ th score

$$a = 1, b = 5, c = 2, d = 3$$

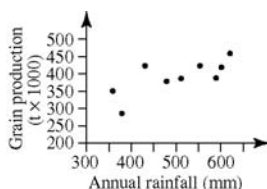
$$q = \frac{(a+c) - (b+d)}{a+b+c+d}$$

$$= \frac{(1+2) - (5+3)}{1+5+2+3}$$

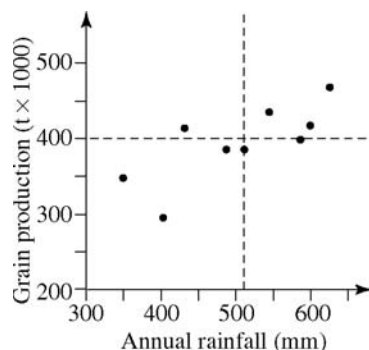
$$= -0.45$$

- c Weak negative correlation.
There is little evidence to show that the greater the amount of television watched, the less the amount of homework done.
- d Many factors affect the amount of homework done apart from the time spent watching television, e.g. organisation, motivation, recreational activities etc.

8 a



b



The median is the $\frac{9+1}{2} = 5$ th score.

$$a = 3, b = 1, c = 3, d = 0$$

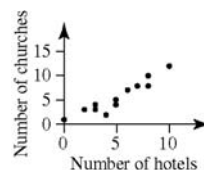
$$q = \frac{(a+c) - (b+d)}{a+b+c+d}$$

$$= \frac{(3+3) - (1+0)}{3+1+3+0}$$

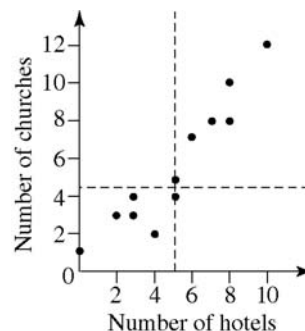
$$= 0.71$$

- c Moderate positive correlation.
There is evidence to show that the greater the amount of rainfall, the greater the grain production.
- d Other factors affect grain production e.g. the time of rainfall, amount of sunshine etc.

9 a



b



The median is the $\frac{12+1}{2} = 6.5$ th score.

$$a = 5, b = 0, c = 5, d = 0$$

$$q = \frac{(a+c) - (b+d)}{a+b+c+d}$$

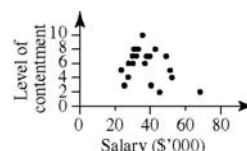
$$= \frac{(5+5) - (0+0)}{5+0+5+0}$$

$$= 1.0$$

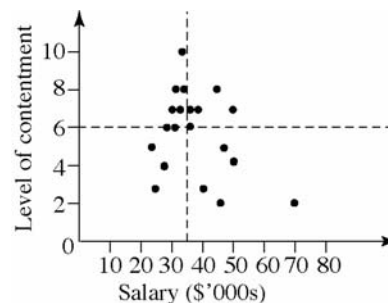
Strong positive correlation.

- c The greater the number of hotels, the greater the number of churches.
- d Other factors affect the number of churches in a town apart from the number of hotels, e.g. number of different religions present in the town.

10 a



b



The median is the $\frac{20+1}{2} = 10.5$ th score.

$$a = 4, b = 5, c = 3, d = 5$$

$$q = \frac{(a+c) - (b+d)}{a+b+c+d}$$

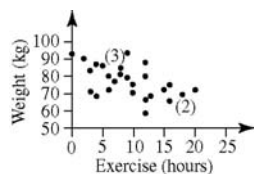
$$= \frac{(4+3) - (5+5)}{4+5+3+5}$$

$$= -0.18$$

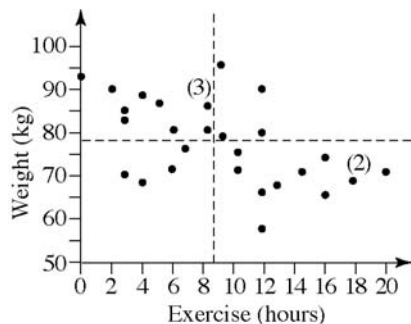
No correlation.

- c There is no relationship between salary and level of contentment.
- d Many factors contribute to happiness and salary is not necessarily one of them. Therefore the data confirm that 'Money cannot buy you happiness'.

11 a



b



The median is the $\frac{30+1}{2} = 15.5$ th score.

$$a = 4, b = 11, c = 4, d = 11$$

$$q = \frac{(a+c) - (b+d)}{a+b+c+d}$$

$$= \frac{(4+4) - (11+11)}{4+11+4+11}$$

$$= -0.47$$

$\therefore$ Weak negative correlation.

- c There is a little evidence to show that the greater the amount of exercise, the lower the body weight.
- d Many other factors affect the body weight apart from amount of exercise, e.g. diet, height, build etc.

- 12 $q = 0.82 \Rightarrow$ strong positive correlation
That is, the greater the level of incentive, the greater the work rate.

Answer is **B**.

- 13 Correlation factors lie between -1 and 1

$$q = -1.34 \Rightarrow \text{error in calculations}$$

Answer is **E**.

- 14 a $a = 6, b = 0, c = 6, d = 0$

$$q = \frac{(a+c) - (b+d)}{a+b+c+d}$$

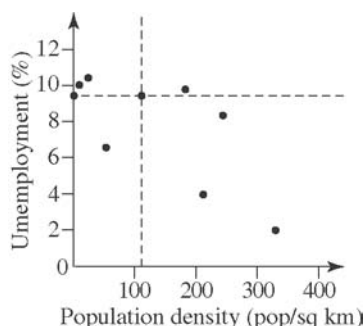
$$= \frac{(6+6) - (0+0)}{6+0+6+0}$$

$$= 1.0$$

- b No, the data is not truly linear. There is not a direct linear relationship between the two variables being measured. Other factors affect the dependent variable apart from the independent variable.

- 15 Does overcrowding cause unemployment?

Using scatter plots to find q correlation coefficient.
Interpretation of q correlation factor.



The median is the $\frac{n+1}{4} = 5$ th score.

$$a = 1, b = 2, c = 1, d = 3$$

$$q = \frac{(1+1) - (2+3)}{1+2+1+3}$$

$$= -0.43$$

$\therefore$ Weak negative correlation.

There is a little evidence to suggest that the greater the overcrowding, the lower the unemployment. Therefore the hypothesis that overcrowding causes unemployment is incorrect.

Exercise 4C — Linear modelling

- 1 a (x_1, y_1) (x_2, y_2)
(2, 1) (6, 3)

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{3-1}{6-2}$$

$$= 0.5$$

$$y - y_1 = m(x - x_1)$$

$$y - 1 = 0.5(x - 2)$$

$$y - 1 = 0.5x - 1$$

$$y = 0.5x - 1 + 1$$

$$y = 0.5x$$

- b (x_1, y_1) (x_2, y_2)
(5, 4) (7, 0)

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{0-4}{7-5}$$

$$= -2$$

$$y - y_1 = m(x - x_1)$$

$$y - 4 = -2(x - 5)$$

$$= -2x + 10$$

$$y = -2x + 10 + 4$$

$$y = -2x + 14$$

- c (x_1, y_1) (x_2, y_2)
(5, 8) (10, 18)

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{18-8}{10-5}$$

$$m = 2$$

$$y - y_1 = m(x - x_1)$$

$$y - 8 = 2(x - 5)$$

$$= 2x - 10$$

$$y = 2x - 10 + 8$$

$$y = 2x - 2$$

- d (x_1, y_1) (x_2, y_2)
(12, 17) (6, 27)

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{27-17}{6-12}$$

$$= -1.67$$

$$y - y_1 = m(x - x_1)$$

$$y - 17 = -1.67(x - 12)$$

$$= -1.67x + 20$$

$$y = -1.67x + 20 + 17$$

$$y = -1.67x + 37$$

$$\mathbf{e} \quad \begin{matrix} (x_1, y_1) & (x_2, y_2) \\ (10, 32) & (25, 67) \end{matrix}$$

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \\ &= \frac{67 - 32}{25 - 10} \\ &= 2.33 \\ y - y_1 &= m(x - x_1) \\ y - 32 &= 2.33(x - 10) \\ &= 2.33x - 23.3 \\ y &= 2.33x - 23.3 + 32 \\ y &= 2.33x + 8.7 \end{aligned}$$

$$\mathbf{f} \quad \begin{matrix} (x_1, y_1) & (x_2, y_2) \\ (14, 52) & (23, 18) \end{matrix}$$

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \\ &= \frac{18 - 52}{23 - 14} \\ &= -3.78 \\ y - y_1 &= m(x - x_1) \\ y - 52 &= -3.78(x - 14) \\ &= -3.78x + 52.92 \\ y &= -3.78x + 52.92 + 52 \\ y &= -3.78x + 104.92 \end{aligned}$$

$$\mathbf{g} \quad \begin{matrix} (x_1, y_1) & (x_2, y_2) \\ (5.2, 12.9) & (7.4, 8.6) \end{matrix}$$

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \\ &= \frac{8.6 - 12.9}{7.4 - 5.2} \\ &= -1.95 \\ y - y_1 &= m(x - x_1) \\ y - 12.9 &= -1.95(x - 5.2) \\ &= -1.95x + 10.14 \\ y &= -1.95x + 10.14 + 12.9 \\ y &= -1.95x + 23.04 \end{aligned}$$

$$\mathbf{h} \quad \begin{matrix} (x_1, y_1) & (x_2, y_2) \\ (31.5, 82.3) & (101.8, 32.6) \end{matrix}$$

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \\ &= \frac{32.6 - 82.3}{101.8 - 31.5} \\ &= -0.71 \\ y - y_1 &= m(x - x_1) \\ y - 82.3 &= -0.71(x - 31.5) \\ &= -0.71x + 22.37 \\ y &= -0.71x + 22.37 + 82.3 \\ y &= -0.71x + 104.67 \end{aligned}$$

$$\mathbf{i} \quad \begin{matrix} (x_1, y_1) & (x_2, y_2) \\ (6, 8) & (-2, 4) \end{matrix}$$

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \\ &= \frac{4 - 8}{-2 - 6} \\ &= 0.5 \\ y - y_1 &= m(x - x_1) \\ y - 8 &= 0.5(x - 6) \\ &= 0.5x - 3 \\ y &= 0.5x - 3 + 8 \\ y &= 0.5x + 5 \end{aligned}$$

$$\mathbf{j} \quad \begin{matrix} (x_1, y_1) & (x_2, y_2) \\ (5.2, -7.4) & (23.7, -36.8) \end{matrix}$$

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \\ &= \frac{-36.8 - 7.4}{23.7 - 5.2} \\ &= -1.59 \\ y - y_1 &= m(x - x_1) \\ y - 7.4 &= -1.59(x - 5.2) \\ y + 7.4 &= -1.59x + 8.27 \\ y &= -1.59x + 8.27 - 7.4 \\ y &= -1.59x + 0.87 \end{aligned}$$

$$\mathbf{2} \quad \begin{matrix} (x_1, y_1) & (x_2, y_2) \\ (3, 8) & (12, 26) \end{matrix}$$

$$\begin{aligned} m &= \frac{26 - 8}{12 - 3} \\ &= 2 \\ y - y_1 &= m(x - x_1) \\ y - 8 &= 2(x - 3) \\ &= 2x - 6 \\ y &= 2x - 6 + 8 \\ y &= 2x + 2 \end{aligned}$$

Answer is **C**.

3 Gradient is positive $\therefore m > 0$.
y-intercept is positive $\therefore c > 0$
 $\therefore$ answer is **C**.

4 Gradient is positive $\therefore m > 0$
y-intercept is negative $\therefore c < 0$
 $\therefore$ answer is **D**.

5 Answers to this question may vary slightly.

$$\mathbf{a} \quad \begin{matrix} (x_1, y_1) & (x_2, y_2) \\ (4, 32) & (10, 62) \end{matrix}$$

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \\ &= \frac{62 - 32}{10 - 4} \\ &= 5 \\ y - y_1 &= m(x - x_1) \\ y - 32 &= 5(x - 4) \\ &= 5x - 20 \\ y &= 5x - 20 + 32 \\ y &= 5x + 12 \end{aligned}$$

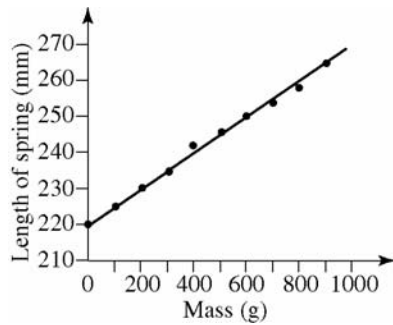
$$\mathbf{b} \quad \begin{matrix} (x_1, y_1) & (x_2, y_2) \\ (20, 59) & (100, 27) \end{matrix}$$

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \\ &= \frac{27 - 59}{100 - 20} \\ &= -0.4 \\ y - y_1 &= m(x - x_1) \\ y - 59 &= -0.4(x - 20) \\ &= -0.4x + 8 \\ y &= -0.4x + 8 + 59 \\ y &= -0.4x + 67 \end{aligned}$$

$$\mathbf{c} \quad \begin{matrix} (x_1, y_1) & (x_2, y_2) \\ (20, 1500) & (35, 3000) \end{matrix}$$

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \\ &= \frac{3000 - 1500}{35 - 20} \\ &= 100 \\ y - y_1 &= m(x - x_1) \\ y - 1500 &= 100(x - 20) \\ &= 100x - 2000 \\ y &= 100x - 2000 + 1500 \\ y &= 100x - 500 \end{aligned}$$

6 a



b (x_1, y_1) (x_2, y_2)
 $(0, 220)$ $(800, 260)$

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$m = \frac{260 - 220}{800 - 0}$$

$$= 0.05$$

$$y - y_1 = m(x - x_1)$$

$$y - 220 = 0.05(x - 0)$$

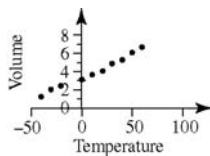
$$= 0.05x$$

$$y = 0.05x + 220$$

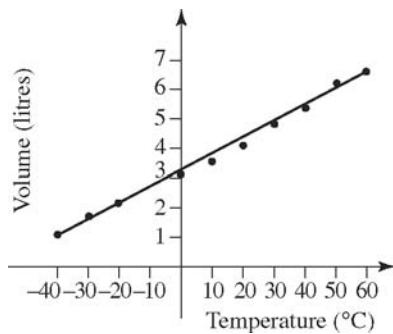
That is, $L = 0.05m + 220$

- c The y -intercept is the natural length of the spring. The gradient represents the extra length of the spring for each extra gram added.

7 a



b



(x_1, y_1) (x_2, y_2)
 $(-30, 1.8)$ $(50, 5.8)$

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{5.8 - 1.8}{50 - (-30)}$$

$$= 0.05$$

$$y - y_1 = m(x - x_1)$$

$$y - 1.8 = 0.05(x - (-30))$$

$$= 0.05x + 1.5$$

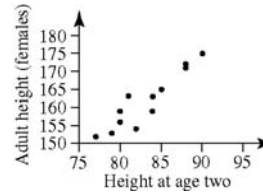
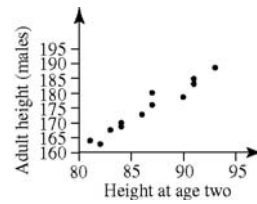
$$y = 0.05x + 1.5 + 1.8$$

$$y = 0.05x + 3.3$$

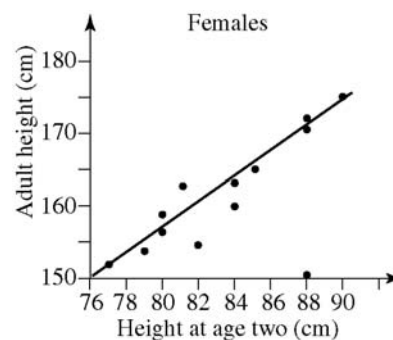
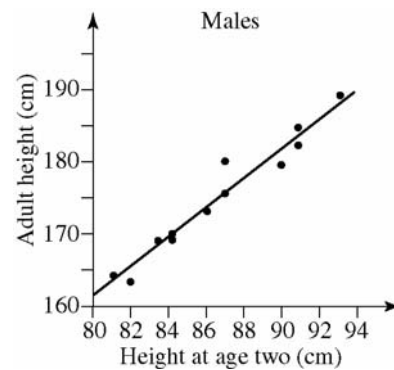
That is, $V = 0.05T + 3.3$

- c The y -intercept gives the volume of gas at 0°C . The gradient gives the increase in volume for each additional degree in temperature.
- 8 Gradient of 0.96 indicates that for each metre increase in length, the wingspan increases by 0.96 metres. Therefore, the answer is **B**.

9 a



b



Males:

(x_1, y_1) (x_2, y_2)
 $(84, 170)$ $(94, 191)$

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{191 - 170}{94 - 84}$$

$$= 2.1$$

$$y - y_1 = m(x - x_1)$$

$$y - 170 = 2.1(x - 84)$$

$$= 2.1x - 176.4$$

$$y = 2.1x - 176.4 + 170$$

$$y = 2.1x - 6.4$$

That is, $A = 2.1b - 6.4$

Females:

(x_1, y_1) (x_2, y_2)
 $(80, 157)$ $(90, 175)$

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{175 - 157}{90 - 80}$$

$$= 1.8$$

$$y - y_1 = m(x - x_1)$$

$$y - 157 = 1.8(x - 80)$$

$$\begin{aligned}
 &= 1.8x - 144 \\
 y &= 1.8x - 144 + 157 \\
 y &= 1.8x + 13
 \end{aligned}$$

That is, $A = 1.8b + 13$

10 a

x	10	20	30	40	50	60	70	80
y	24	35	43	57	66	75	89	94

lower set

upper set

The x -coordinates are in increasing order; there are 8 data points: 4 to each set.

$$\bar{x}_L = \frac{10 + 20 + 30 + 40}{4}$$

$$= 25$$

$$\bar{y}_L = \frac{24 + 35 + 43 + 57}{4}$$

$$= 39.75$$

$$(\bar{x}_L, \bar{y}_L) = (25, 39.75)$$

$$\bar{x}_U = \frac{50 + 60 + 70 + 80}{4}$$

$$= 65$$

$$\bar{y}_U = \frac{66 + 75 + 89 + 94}{4}$$

$$= 81$$

$$(\bar{x}_U, \bar{y}_U) = (65, 81)$$

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{81 - 39.75}{65 - 25}$$

$$= \frac{41.25}{40}$$

$$= 1.03125$$

$$y - y_1 = m(x - x_1)$$

$$y - 39.75 = 1.03125(x - 25)$$

$$y = 1.03125x + 13.96875$$

b First rearrange the data points so that x -coordinates are in increasing order.

x	5	10	15	20	25	30	35	40
y	90	84	79	70	65	61	40	45

lower set

upper set

There are 8 data points: 4 to each set.

$$\bar{x}_L = \frac{5 + 10 + 15 + 20}{4}$$

$$= 12.5$$

$$\bar{y}_L = \frac{90 + 84 + 79 + 70}{4}$$

$$= 80.75$$

$$(\bar{x}_L, \bar{y}_L) = (12.5, 80.75)$$

$$\bar{x}_U = \frac{25 + 30 + 35 + 40}{4}$$

$$= 32.5$$

$$\bar{y}_U = \frac{65 + 61 + 40 + 45}{4}$$

$$= 52.75$$

$$(\bar{x}_U, \bar{y}_U) = (32.5, 52.75)$$

$$m = \frac{52.75 - 80.75}{32.5 - 12.5}$$

$$= \frac{-28}{20}$$

$$= -1.4$$

$$y - 80.75 = -1.4(x - 12.5)$$

$$y = -1.4x + 98.25$$

c

x	30	32	34	36	38	40	42	44	46
y	40	53	61	68	82	90	104	111	119

lower half

upper half

The x -coordinates are in increasing order; there are 9 data points (odd number), so the middle point (38, 82) goes to both sets.

$$\bar{x}_L = \frac{30 + 32 + 34 + 36 + 38}{5}$$

$$= 34$$

$$\bar{y}_L = \frac{40 + 53 + 61 + 68 + 82}{5}$$

$$= 60.8$$

$$(\bar{x}_L, \bar{y}_L) = (34, 60.8)$$

$$\bar{x}_U = \frac{38 + 40 + 42 + 44 + 46}{5}$$

$$= 42$$

$$\bar{y}_U = \frac{82 + 90 + 104 + 111 + 119}{5}$$

$$= 101.2$$

$$(\bar{x}_U, \bar{y}_U) = (42, 101.2)$$

$$m = \frac{101.2 - 60.8}{42 - 34}$$

$$= \frac{40.4}{8}$$

$$= 5.05$$

$$y - 60.8 = 5.05(x - 34)$$

$$y = 5.05x - 110.9$$

d Rearrange the data points so that x -coordinates are in increasing order.

x	76	77	78	79	80	81	82	83	84
y	21	25	28	30	37	43	44	47	52

lower half

upper half

There are 9 data points, so middle point (80, 37) goes to both sets.

$$\bar{x}_L = \frac{76 + 77 + 78 + 79 + 80}{5}$$

$$= 78$$

$$\bar{y}_L = \frac{21 + 25 + 28 + 30 + 37}{5}$$

$$= 28.2$$

$$(\bar{x}_L, \bar{y}_L) = (78, 28.2)$$

$$\bar{x}_U = \frac{80 + 81 + 82 + 83 + 84}{5}$$

$$= 82$$

$$\bar{y}_U = \frac{37 + 43 + 44 + 47 + 52}{5}$$

$$= 44.6$$

$$(\bar{x}_U, \bar{y}_U) = (82, 44.6)$$

$$m = \frac{44.6 - 28.2}{82 - 78}$$

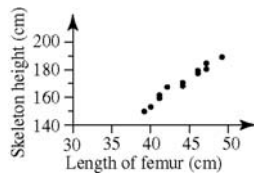
$$= \frac{16.4}{4}$$

$$= 4.1$$

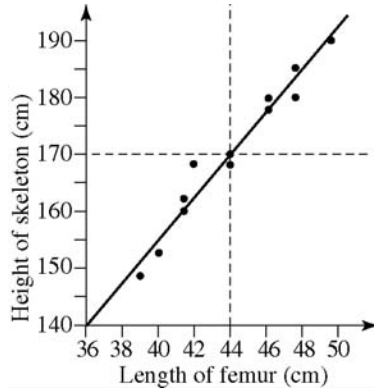
$$y - 28.2 = 4.1(x - 78)$$

$$y = 4.1x - 291.6$$

11 a



b



The median is the $\frac{12+1}{2} = 6.5$ th score

$$a = 5, b = 0, c = 5, d = 0$$

$$q = \frac{(a+c) - (b+d)}{a+b+c+d}$$

$$= \frac{(5+5) - (0+0)}{5+0+5+0}$$

$$= 1.0$$

There is a strong positive correlation.

The greater the length of the femur, the greater the height of the person.

c

<i>L</i>	39	40	41	41	42	44	44	46	46	47	47	49
<i>H</i>	150	154	160	162	168	170	171	178	180	181	185	190
	lower half						upper half					

There are 12 data points: 6 to each set.

$$\bar{x}_L = \frac{39+40+41+41+42+44}{6} = 41\frac{1}{6}$$

$$\bar{y}_L = \frac{150+154+160+162+168+170}{6} = 160\frac{2}{3}$$

$$(\bar{x}_L, \bar{y}_L) = \left(41\frac{1}{6}, 160\frac{2}{3}\right)$$

$$\bar{x}_U = \frac{44+46+46+47+47+49}{6} = 46.5$$

$$\bar{y}_U = \frac{171+178+180+181+185+190}{6} = 180\frac{5}{6}$$

$$(\bar{x}_U, \bar{y}_U) = \left(46.5, 180\frac{5}{6}\right)$$

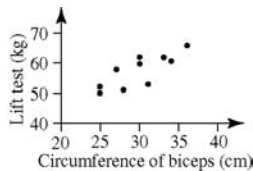
$$m = \frac{180\frac{5}{6} - 160\frac{2}{3}}{46.5 - 41\frac{1}{6}} = 3.78125$$

$$y - 160\frac{2}{3} = 3.78125 \left(x - 41\frac{1}{6}\right)$$

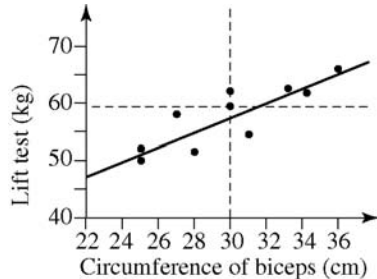
$$y = 3.78125x + 5.00521$$

So, $H = 3.78L + 5.01$ (rounded to 2 decimal places).

12 a



b



The median is the $\frac{10+1}{2} = 5.5$ th score.

$$a = 3, b = 0, c = 4, d = 1$$

$$q = \frac{(a+c)-(b+d)}{a+b+c+d}$$

$$= \frac{(3+4)-(0+1)}{3+0+4+1}$$

$$= 0.75$$

There is a strong positive correlation.

The greater the circumference of the biceps, the greater the performance on the lift test.

c

<i>b</i>	25	25	27	28	30	30	31	33	34	36
<i>s</i>	50	52	58	51	60	62	53	62	61	66

lower set

upper set

There are 10 data points: 5 to each set.

$$\bar{x}_L = \frac{25+25+27+28+30}{5} = 27$$

$$\bar{y}_L = \frac{50+52+58+51+60}{5} = 54.2$$

$$(\bar{x}_L, \bar{y}_L) = (27, 54.2)$$

$$\bar{x}_U = \frac{30+31+33+34+36}{5} = 32.8$$

$$\bar{y}_U = \frac{62+53+62+61+66}{5} = 60.8$$

$$(\bar{x}_U, \bar{y}_U) = (32.8, 60.8)$$

$$m = \frac{60.8-54.2}{32.8-27} = \frac{6.6}{5.8} = 1.138 \text{ (to 3 d. p.)}$$

$$y - 54.2 = 1.138(x - 27)$$

$$y = 1.138x + 23.474$$

$$\text{So, } S = 1.138b + 23.474$$

13 b is unsuitable since it is curved.

c is unsuitable since there is no correlation.

$\therefore$ answer is D.

Exercise 4D — Making predictions

1 a $F = 1.2d + 3$

$$= 1.2(12) + 3$$

$$= 14.4 + 3$$

$$= 17.4$$

The fare would be \$17.40.

b $F = 1.2(4.5) + 3$

$$= 5.4 + 3$$

$$= 8.4$$

The fare would be \$8.40.

c $F = 1.2d + 3$

$$26 = 1.2d + 3$$

$$23 = 1.2d$$

$$19.17 = d$$

The distance is 19.17 km.

d $F = 1.2d + 3$

$$13.20 = 1.2d + 3$$

$$10.20 = 1.2d$$

$$8.5 = d$$

The distance is 8.5 km.

2 a $H = 6.1f - 5$

$$H = 6.1(27) - 5$$

$$= 164.7 - 5$$

$$= 159.7$$

The height is 159.7 cm.

b $H = 6.1f - 5$

$$= 6.1(30) - 5$$

$$= 183 - 5$$

$$= 178$$

The height is 178 cm.

c $H = 6.1f - 5$

$$185 = 6.1f - 5$$

$$190 = 6.1f$$

$$31.1 = f$$

The footprint is 31.1 cm.

d $H = 6.1f - 5$

$$152 = 6.1f - 5$$

$$157 = 6.1f$$

$$25.7 = f$$

The footprint is 25.7 cm.

3 a $V = -500t + 2500$

$$= -500(4) + 2500$$

$$= -2000 + 2500$$

$$= 500$$

The value is \$500.

b 2 yrs 9 mths $= 2\frac{9}{12}$

$$= 2.75 \text{ yrs.}$$

$$V = -500t + 2500$$

$$= -500(2.75) + 2500$$

$$= -1375 + 2500$$

$$= 1125$$

The value is \$1125.

c When equipment is new, $t = 0$

$$V = -500t + 2500$$

$$= -500(0) + 2500$$

$$= 2500$$

The value is \$2500.

d $V = -500t + 2500$

$$1000 = -500t + 2500$$

$$-1500 = -500t$$

$$3 = t$$

It will be 3 years old.

e When equipment is worthless,

$$V = 0$$

$$V = -500t + 2500$$

$$0 = -500t + 2500$$

$$-2500 = -500t$$

$$5 = t$$

It will be worthless after 5 years.

4 a $N = -23t + 870$

$$= -23(5) + 870$$

$$= -115 + 870$$

$$= 755$$

So 755 pies will be sold.

$$\begin{aligned} \text{b } N &= -23t + 870 \\ &= -23(25) + 870 \\ &= -575 + 870 \\ &= 295 \end{aligned}$$

So 295 pies will be sold.

$$\begin{aligned} \text{c } N &= -23t + 870 \\ 400 &= -23t + 870 \\ -470 &= -23t \\ 20.4 &= t \end{aligned}$$

The temperature is likely to be 20.4°C .

$$\text{d Number of pies sold} \Rightarrow N = 0$$

$$\begin{aligned} N &= -23t + 870 \\ 0 &= -23t + 870 \\ -870 &= -23t \\ 37.8 &= t \end{aligned}$$

The temperature is likely to be 37.8°C .

$$\begin{aligned} \text{5 a } C &= 40h + 35 \\ &= 40(2) + 35 \\ &= 80 + 35 \\ &= 115 \end{aligned}$$

The cost is \$115.

$$\text{b } 1 \text{ hr. } 15 \text{ mins} = 1\frac{15}{60} = 1.25 \text{ hours.}$$

$$\begin{aligned} C &= 40h + 35 \\ &= 40(1.25) + 35 \\ &= 50 + 35 \\ &= 85 \end{aligned}$$

The cost is \$85.

$$\begin{aligned} \text{c } C &= 40h + 35 \\ 175 &= 40h + 35 \\ 140 &= 40h \\ 3.5 &= h \end{aligned}$$

At least 3.5 hours or 3 hours and 30 minutes needs to be spent on the job.

$$\begin{aligned} \text{d } C &= 40h + 35 \\ 215 &= 40h + 35 \\ 180 &= 40h \\ 4.5 &= h \end{aligned}$$

At least 4.5 hours or 4 hours and 30 minutes need to be spent on the job.

$$\begin{aligned} \text{6 } Y &= -6.2X + 20 \\ 4.5 &= -6.2X + 20 \\ -15.5 &= -6.2X \\ 2.5 &= X \end{aligned}$$

$\therefore$ answer is **D**.

7 a From the graph,

$$\begin{aligned} \text{i } x &= 12 &\Rightarrow y &= 700 \\ \text{ii } x &= 6 &\Rightarrow y &= 450 \\ \text{iii } x &= 23 &\Rightarrow y &= 1170 \\ \text{iv } y &= 600 &\Rightarrow x &= 9.5 \\ \text{v } y &= 1000 &\Rightarrow x &= 19 \\ \text{vi } y &= 320 &\Rightarrow x &= 3 \end{aligned}$$

$$\text{b } \begin{matrix} (x_1, y_1) & (x_2, y_2) \\ (6, 450) & (23, 1170) \end{matrix}$$

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \\ &= \frac{1170 - 450}{23 - 6} \\ &= 42.353 \\ &= 42 \end{aligned}$$

$$\begin{aligned} y - y_1 &= m(x - x_1) \\ y - 450 &= 42(x - 6) \\ &= 42x - 252 \\ y &= 42x - 252 + 450 \\ y &= 42x + 198 \end{aligned}$$

$$\begin{aligned} \text{c i } y &= 42(12) + 198 \\ &= 504 + 198 \\ &= 702 \end{aligned}$$

$$\begin{aligned} \text{ii } y &= 42(6) + 198 \\ &= 252 + 198 \\ &= 450 \end{aligned}$$

$$\begin{aligned} \text{iii } y &= 42(23) + 198 \\ &= 966 + 198 \\ &= 1164 \end{aligned}$$

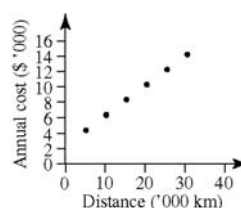
$$\begin{aligned} \text{iv } 600 &= 42x + 198 \\ 402 &= 42x \\ x &= 9.6 \end{aligned}$$

$$\begin{aligned} \text{v } 1000 &= 42x + 198 \\ 802 &= 42x \\ x &= 19.1 \end{aligned}$$

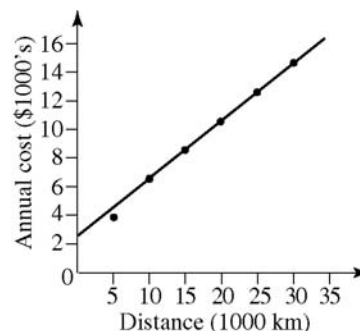
$$\begin{aligned} \text{vi } 320 &= 42x + 198 \\ 122 &= 42x \\ x &= 2.9 \end{aligned}$$

8 Average number of children $= -0.48 + 0.146 \times 60 = 8.28$.
Therefore, the answer is **A**.

9 a



b



$$\text{c } \begin{matrix} (x_1, y_1) & (x_2, y_2) \\ (10\,000, 6400) & (30\,000, 14\,400) \end{matrix}$$

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \\ &= \frac{14400 - 6400}{30000 - 10000} \\ &= 0.4 \\ y - y_1 &= m(x - x_1) \\ y - 6400 &= 0.4(x - 10\,000) \\ &= 0.4x - 4000 \\ y &= 0.4x - 4000 + 6400 \\ y &= 0.4x + 2400 \end{aligned}$$

That is, $C = 0.4d + 2400$

d i y-intercept is \$2400.

ii It is the cost of running a car which is not driven at all.

e i Gradient is 0.4.

ii It is the additional cost of running a car per kilometre travelled.

$$\begin{aligned} \text{f } C &= 0.4d + 2400 \\ &= 0.4(15\,000) + 2400 \\ &= 6000 + 2400 \\ &= 8400 \end{aligned}$$

The annual cost is \$8400.

$$\begin{aligned} \text{g } C &= 0.4d + 2400 \\ &= 0.4(1000) + 2400 \\ &= 400 + 2400 \\ &= 2800 \end{aligned}$$

The annual cost is \$2800.

$$\begin{aligned} \text{h } C &= 0.4d + 2400 \\ 8000 &= 0.4d + 2400 \\ 5600 &= 0.4d \\ 14\,000 &= d \end{aligned}$$

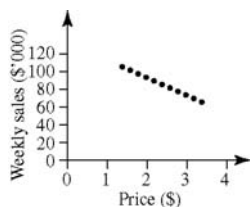
The distance driven is 14 000 km.

$$\begin{aligned} \text{i } C &= 0.4d + 2400 \\ 16\,000 &= 0.4d + 2400 \\ 13\,600 &= 0.4d \\ 34\,000 &= d \end{aligned}$$

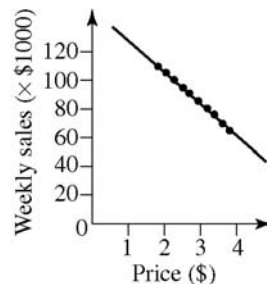
The distance driven is 34 000 km.

- j The parts **f** and **h** are reliable because this is a clear linear relationship between cost (C) and distance (d) and these results are interpolated.

10 a



b



$$\begin{array}{cc} (x_1, y_1) & (x_2, y_2) \\ (1, 115) & (3, 75) \end{array}$$

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \\ &= \frac{75 - 115}{3 - 1} \\ &= -20 \end{aligned}$$

$$\begin{aligned} y - y_1 &= m(x - x_1) \\ y - 115 &= -20(x - 1) \\ &= -20x + 20 \\ y &= -20x + 20 + 115 \\ y &= -20x + 135 \end{aligned}$$

That is, $N = -20p + 135$ where N is the number of cans sold in thousands.

$$\begin{aligned} \text{d } N &= -20p + 135 \\ &= -20(3.10) + 135 \\ &= -62 + 135 \\ &= 73 \end{aligned}$$

So 73 000 cans are sold.

$$\begin{aligned} \text{e } N &= -20p + 135 \\ &= -20(4.6) + 135 \\ &= -92 + 135 \\ &= 43 \end{aligned}$$

So 43 000 cans are sold.

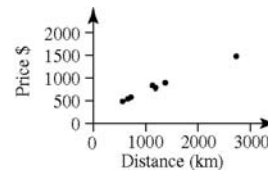
$$\begin{aligned} \text{f } N &= -20p + 135 \\ 80 &= -20p + 135 \end{aligned}$$

$$\begin{aligned} -55 &= -20p \\ 2.75 &= p \\ \text{Sell at } \$2.75 \text{ per can.} \end{aligned}$$

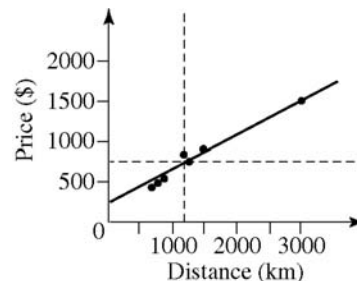
$$\begin{aligned} \text{g } 100\,000 \text{ cans} &\Rightarrow N = 100 \\ N &= -20p + 135 \\ 100 &= -20p + 135 \\ -35 &= -20p \\ 1.75 &= p \\ \text{Sell at } \$1.75 \text{ per can.} \end{aligned}$$

- h The parts **d**, **f** and **g** are all reliable since the price and number sold closely fit a linear relationship and these results are interpolated.

11 a



b and c



The median is the $\frac{7+1}{2} = 4$ th score.

$$a = 2, b = 0, c = 3, d = 0$$

$$\begin{aligned} q &= \frac{(a+c) - (b+d)}{a+b+c+d} \\ &= \frac{(2+3) - (0+0)}{2+0+3+0} \end{aligned}$$

$$q = 1.0$$

There is a strong positive correlation.

The greater the distance, the greater the air fare.

- d First arrange the points so that the x-coordinates are in increasing order.

x	559	669	713	1144	1172	1370	2728
y	520	570	580	820	790	890	1490
	lower half				upper half		

There are 7 data points, so place the middle point (1144, 820) in both sets.

$$\bar{x}_L = \frac{559 + 669 + 713 + 1144}{4} = 771.25$$

$$\bar{y}_L = \frac{520 + 570 + 580 + 820}{4} = 622.5$$

$$(\bar{x}_L, \bar{y}_L) = (771.25, 622.5)$$

$$\bar{x}_U = \frac{1144 + 1172 + 1370 + 2728}{4}$$

$$= 1603.5$$

$$\bar{y}_U = \frac{820 + 790 + 890 + 1490}{4}$$

$$= 997.5$$

$$(\bar{x}_U, \bar{y}_U) = (1603.5, 997.5)$$

$$m = \frac{997.5 - 622.5}{1603.5 - 771.25}$$

$$= \frac{375}{832.25}$$

$$\approx 0.45$$

$$y - 622.5 = 0.45(x - 771.25)$$

$$y = 0.45x + 275.44 \text{ (rounded to 2 decimal places)}$$

$$\text{So, } A = 0.45d + 275.44$$

e $d = 671 \text{ km}$

$$A = 0.45 \times 671 + 275.44$$

$$= \$577.39$$

f $d = 2125 \text{ km}$

$$A = 0.45 \times 2125 + 275.44$$

$$= \$1231.69$$

g $d = 1024$

$$A = 0.45 \times 1024 + 275.44$$

$$= \$736.24$$

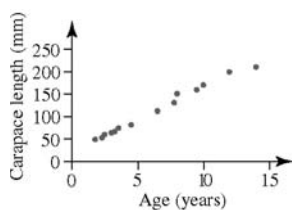
h $d = 3295$

$$A = 0.45 \times 3295 + 275.44$$

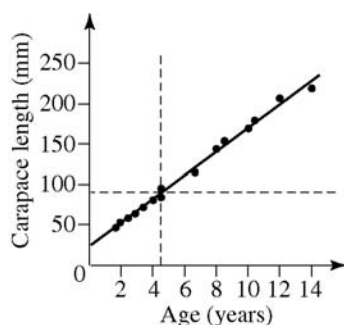
$$= \$1758.19$$

i Answers to parts **e**, **f** and **g** are reliable, because they are obtained using interpolation.

12 a



b and c



The median is the $\frac{15+2}{2} = 8\text{th}$ score.

$$a = 7, b = 0, c = 6, d = 0$$

$$q = \frac{(a+c) - (b+d)}{a+b+c+d}$$

$$= \frac{(7+6) - (0+0)}{7+0+6+0}$$

$$= 1.0$$

There is a strong positive correlation.

The greater the age, the greater the carapace length.

d The points are not perfectly linear as other factors influence carapace length such as breeding etc.

e From the graphics calculator,

$$y = 14.35x + 21.72$$

$$\text{That is, } L = 14.35a + 21.72$$

f $L = 14.35 \times 5 + 21.72$

$$= 93.47$$

It is 93.47 mm long.

g $L = 14.35 \times 16 + 21.72$

$$= 251.32 \text{ mm}$$

It is 251.32 mm long.

h $65 = 14.35a + 21.72$

$$43.28 = 14.35a$$

$$a = 3.02$$

They reach sexual maturity at approximately 3 years.

i $a = 3 + 3 = 6 \text{ years}$

$$L = 14.35 \times 6 + 21.72$$

$$= 107.82$$

The male crayfish should be 107.82 mm or more (approximately 108 mm) before they can be caught.

j Parts **f** and **h** are reliable as they are interpolated and the data closely fits a straight line.

13 The result is a prediction beyond the bounds of the original data.

$\therefore$ The result is unreliable as it is extrapolated.

Answer is **C**.

14 There is not a strong correlation between the points.

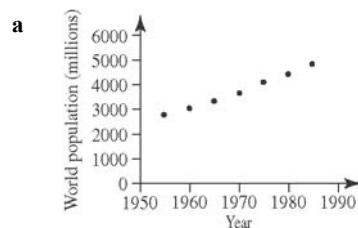
$\therefore$ The result is unreliable as the data does not closely fit a straight line.

Answer is **D**.

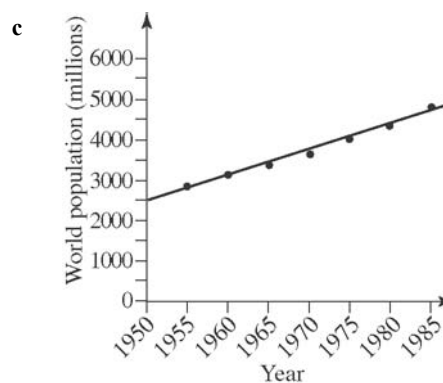
15 Extrapolation is unreliable as it assumes the relationship holds true for untested values

Answer is **C**.

16 World population



b Yes. The points form a straight line



$$(x_1, y_1)$$

$$(1959, 3000)$$

$$(x_2, y_2)$$

$$(1980, 4500)$$

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{4500 - 3000}{1980 - 1959}$$

$$= 71.4$$

$$y - y_1 = m(x - x_1)$$

$$y - 3000 = 71.4(x - 1959)$$

$$= 71.4x - 139\,873$$

$$y = 71.4x - 139\,873 + 3000$$

$$y = 71.4x - 136\,873$$

That is $W = 71.4y - 136\,873$
(millions of people)

$$\begin{aligned} \text{d } W &= 71.4y - 136\,873 \\ &= 71.4(1950) - 136\,873 \\ &= 139\,230 - 136\,873 \\ &= 2357 \end{aligned}$$

The population was 2357 million.

$$\begin{aligned} \text{e } W &= 71.4(1982) - 136\,873 \\ &= 141\,515 - 136\,873 \\ &= 4642 \end{aligned}$$

The population was 4642 million.

$$\begin{aligned} \text{f } W &= 71.4(1997) - 136\,873 \\ &= 142\,586 - 136\,873 \\ &= 5713 \end{aligned}$$

The population was 5713 million.

g Answer to part 5 is the only reliable one as the others are extrapolated.

h Lines of best fit are only approximations and can vary. Also, the relationship is not exactly linear.

Chapter review

Multiple choice

- 1 Amounts of fertiliser is the independent variable (x -axis). Total weight of potatoes is the dependent variable (y -axis).

Answer is **D**.

- 2 The points follow a line sloping downwards.

Answer is **D**.

- 3 The points follow a line sloping downwards and show some correlation

$\therefore$ The graph shows moderate negative correlation.

Answer is **D**.

- 4 Since $0.5 < q < 0.75$, moderate positive correlation exists. So there is evidence to show that the greater the number of pedestrian crossings, the greater the number of pedestrian accidents.

Answer is **C**.

- 5 Since $q < -0.75$, strong negative correlation exists. So the greater the number of weeks spent on the assembly line, the slower the assembly of components.

Answer is **B**.

- 6 $a = 0$, $b = 5$, $c = 1$, $d = 5$

$$\begin{aligned} q &= \frac{(a+c)-(b+d)}{a+b+c+d} \\ &= \frac{(0+1)-(5+5)}{0+5+1+5} \\ &= -0.82 \end{aligned}$$

Answer is **B**.

- 7 Gradient is positive $\therefore m > 0$

y -intercept is negative $\therefore c < 0$

Answer is **C**.

- 8 (x_1, y_1) (x_2, y_2)
 $(100, 40)$ $(150, 8)$

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \\ &= \frac{8 - 40}{150 - 100} \\ &= -0.64 \end{aligned}$$

$$y - y_1 = m(x - x_1)$$

$$y - 40 = -0.64(x - 100)$$

$$y - 40 = -0.64x + 64$$

$$y = -0.64x + 64 + 40$$

$$y = -0.64x + 104$$

Answer is **E**.

- 9 The y -intercept is 300 and it represents the weekly cost when no items are produced.

Answer is **D**.

- 10 Gradient of 0.87 indicates that for each 1 mm increase in diameter (independent variable), there is 0.87 mm increase in length (Dependent variable). Therefore the answer is **D**.

- 11 The y -intercept is about 210, so only options A or B are acceptable. Since the gradient of the line is negative, the correct option is **A**.

- 12 If the numbers of data points is odd, then the middle point goes to both sets. In this case, there are 11 points, so the first 5 go to the lower set, last 5 — to the upper set and the 6th point (the middle one) goes to both sets. The answer is **D**.

- 13 Interpolation involves predictions within the range of a given set of data. Answer is **B**.

- 14 These are too few data points for the result to be reliable. Answer is **A**.

- 15 Reading horizontally from $y = 10$ then down to x -axis shows that $x = 8$. Answer is **C**.

$$\begin{aligned} 16 \quad V &= -200a + 4200 \\ \text{when } a = 0, \quad V &= -200(0) + 4200 \\ &= 4200 \end{aligned}$$

Half initial value is 2100

$$2100 = -200a + 4200$$

$$-2100 = -200a$$

$$10.5 = a$$

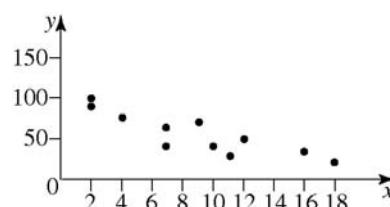
That is, half the initial value is reached after 10.5 years.

Answer is **C**.

- 17 Study time = 35; number of errors = $8.8 - 0.120 \times 35 = 4.6$. Therefore, the answer is **B**.

Short answer

1



There appears to be moderate negative correlation.

- 2 a $a = 4$, $b = 1$, $c = 4$, $d = 1$

$$\begin{aligned} q &= \frac{(a+c)-(b+d)}{a+b+c+d} \\ &= \frac{(4+4)-(1+1)}{4+1+4+1} \\ &= 0.6 \end{aligned}$$

- b $a = 3$, $b = 1$, $c = 2$, $d = 1$

$$\begin{aligned} q &= \frac{(a+c)-(b+d)}{a+b+c+d} \\ &= \frac{(3+2)-(1+1)}{3+1+2+1} \\ &= 0.43 \end{aligned}$$

- c $a = 0$, $b = 3$, $c = 1$, $d = 4$

$$\begin{aligned} q &= \frac{(a+c)-(b+d)}{a+b+c+d} \\ &= \frac{(0+1)-(3+4)}{0+3+1+4} \\ &= -0.75 \end{aligned}$$

- 3 $q = 0.9$ (> 0.75)

$\Rightarrow$ Strong positive correlation.

The greater the thickness, the greater the strength.

4 $q = -0.65$ ($-0.75 < q \leq -0.5$)

$\Rightarrow$ Moderate negative correlation.

There is evidence to show that the greater the age of the person, the lower the age of their car.

5 a (x_1, y_1) (x_2, y_2)
(3, 9) (7, 14)

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{14 - 9}{7 - 3}$$

$$= 1.25$$

$$y - y_1 = m(x - x_1)$$

$$y - 9 = 1.25(x - 3)$$

$$= 1.25x - 3.75$$

$$y = 1.25x - 3.75 + 9$$

$$y = 1.25x + 5.25$$

b (x_1, y_1) (x_2, y_2)
(8, 13) (12, 2)

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{2 - 13}{12 - 8}$$

$$= -2.75$$

$$y - y_1 = m(x - x_1)$$

$$y - 13 = -2.75(x - 8)$$

$$= -2.75x + 22$$

$$y = -2.75x + 22 + 13$$

$$y = -2.75x + 35$$

c (x_1, y_1) (x_2, y_2)
(12.5, 258) (14.8, 307)

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{307 - 258}{14.8 - 12.5}$$

$$= 21.30$$

$$y - y_1 = m(x - x_1)$$

$$y - 258 = 21.3(x - 12.5)$$

$$= 21.3x - 266.25$$

$$y = 21.3x - 266.25 + 258$$

$$y = 21.3x - 8.25$$

6 a (x_1, y_1) (x_2, y_2)
(20, 100) (60, 300)

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{300 - 100}{60 - 20}$$

$$= 5$$

$$y - y_1 = m(x - x_1)$$

$$y - 100 = 5(x - 20)$$

$$y - 100 = 5x - 100$$

$$y = 5x - 100 + 100$$

$$y = 5x$$

b (x_1, y_1) (x_2, y_2)
(0, 300) (100, -550)

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{-550 - 300}{100 - 0}$$

$$= -8.5$$

$$y - y_1 = m(x - x_1)$$

$$y - 300 = -8.5(x - 0)$$

$$= -8.5x$$

$$y = -8.5x + 300$$

7 a The **dependent** variable is the value of the car.

b Select two points on the line, say (3, 14000) and (8, 8000)

$$m = \frac{8000 - 14000}{8 - 3} = -\frac{6000}{5} = -1200$$

$$y - 8000 = -1200(x - 8)$$

$$y - 8000 = -1200x + 9600$$

$$y = 8000 - 1200x + 9600$$

$$y = 17600 - 1200x$$

$$\text{value} = 17\,600 - 1200 \times \text{age}$$

Note: answers may vary slightly, depending on accuracy of reading the coordinates of the points from the scatterplot. If calculator is used to find least squares regression line from the table, the equation (with coefficients rounded to the nearest hundred is: $\text{value} = 17\,500 - 1200 \times \text{age}$.

8 a First rearrange the data points so that x -coordinates are in increasing order.

x	20	30	40	50	60	70	80	90
y	53	49	46	41	39	33	31	26
	lower half				upper half			

There are 8 data points, so 4 points to each set.

$$\bar{x}_L = \frac{20 + 30 + 40 + 50}{4}$$

$$= 35$$

$$\bar{y}_L = \frac{53 + 49 + 46 + 41}{4}$$

$$= 47.25$$

$$(\bar{x}_L, \bar{y}_L) = (35, 47.25)$$

$$\bar{x}_U = \frac{60 + 70 + 80 + 90}{4}$$

$$= 75$$

$$\bar{y}_U = \frac{39 + 33 + 31 + 26}{4}$$

$$= 32.25$$

$$(\bar{x}_U, \bar{y}_U) = (75, 32.25)$$

$$m = \frac{47.25 - 32.25}{35 - 75}$$

$$= \frac{15}{-40}$$

$$= -0.375$$

$$y - 47.25 = -0.375(x - 35)$$

$$y = -0.375x + 60.375$$

b

x	55	56	57	58	59	60	61	62	63
y	12	24	35	48	62	70	85	99	109
	lower half					upper half			

There are 9 points, so the 5th point (59, 62) goes to both sets.

$$\bar{x}_L = \frac{55 + 56 + 57 + 58 + 59}{5}$$

$$= 57$$

$$\bar{y}_L = \frac{12 + 24 + 35 + 48 + 62}{5}$$

$$= 36.2$$

$$(\bar{x}_L, \bar{y}_L) = (57, 36.2)$$

$$\bar{x}_U = \frac{59 + 60 + 61 + 62 + 63}{5}$$

$$= 61$$

$$\bar{y}_U = \frac{62 + 70 + 85 + 99 + 109}{5}$$

$$= 85$$

$$(\bar{x}_U, \bar{y}_U) = (61, 85)$$

$$m = \frac{85 - 36.2}{61 - 57}$$

$$= \frac{48.8}{4}$$

$$= 12.2$$

$$y - 36.2 = 12.2(x - 57)$$

$$y = 12.2x - 659.2$$

9 a $C = 3.6d + 250$

$$430 = 3.6d + 250$$

$$180 = 3.6d$$

$$50 = d$$

That is, $d = 50$

b $C = 3.6d + 250$

$$= 3.6(40) + 250$$

$$= 144 + 250$$

$$= 394$$

So $C = 394$

10 Using the graph, extend the line of best fit.

a $x = 10$, gives $y = 200$

b $y = 900$, gives $x = 48$.

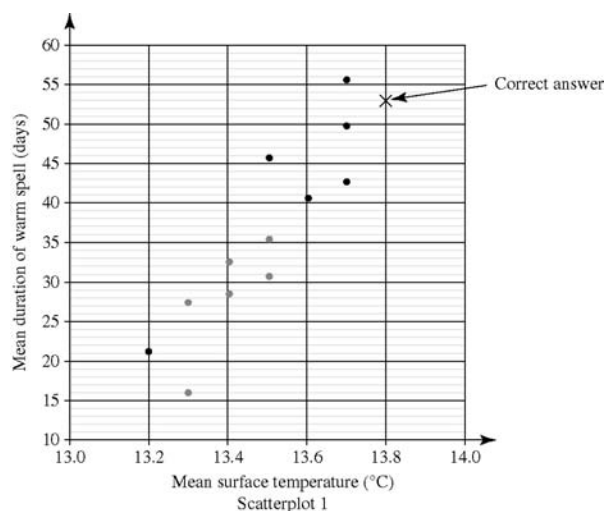
11 a The independent variable here is the Year.

i The value of the gradient tells us the increase of the average salary per year. Since the data is rescaled so that the salaries are measured in thousands of dollars, the average increase is $10.99 \times 1000 = \$10\,990$

ii 2005 is equivalent to year 15. Average salary = $20.9 + 10.99 \times 15 = 185.75$. Since the salaries are measured in thousands of dollars, in 2005 the average salary is expected to be \$185 750

Extended response

1 a



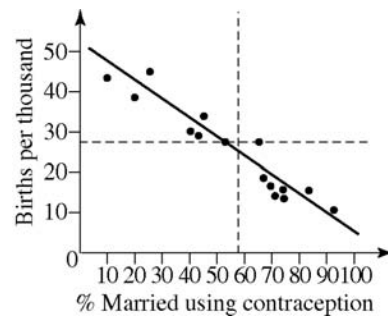
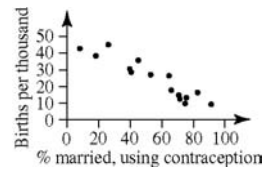
b As appears from the graph, the lowest mean surface temperature recorded was in 1964.

c i mean surface temperature = $-12.361 + 0.013 \times 2010$
 $= 13.769$

Therefore, correct to two decimal places the mean surface temperature for 2010 was 13.77°C .

ii The gradient of the trend line suggests that Australia's mean surface temperature will rise by 0.013 degrees each year.

2 a



$$a = 0, b = 6, c = 0, d = 7$$

$$q = \frac{(a + c) - (b + d)}{a + b + c + d}$$

$$= \frac{(0 + 0) - (6 + 7)}{0 + 6 + 0 + 7}$$

$$= -1.0$$

There is a strong negative correlation.

The greater the percentage of married women using contraception, the lower the birth rate.

d Other factors affect the number of births such as women's fertility, etc.

e $(x_1, y_1) \quad (x_2, y_2)$
 $(25, 38) \quad (65, 20)$

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{20 - 38}{65 - 25}$$

$$= -0.45$$

$$y - y_1 = m(x - x_1)$$

$$y - 38 = -0.45(x - 25)$$

$$= -0.45x + 11.25$$

$$y = -0.45x + 11.25 + 38$$

$$y = -0.45x + 49.25$$

That is, $B \approx -0.45w + 49$

f $B = -0.45w + 49$

$$= -0.45(25) + 49$$

$$= -11.25 + 49$$

$$= 37.75$$

The birth rate is 38 per thousand.

g $B = -0.45(38) + 49$

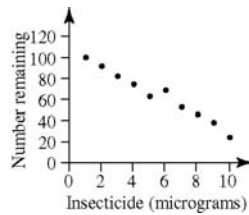
$$= -17.1 + 49$$

$$= 31.9$$

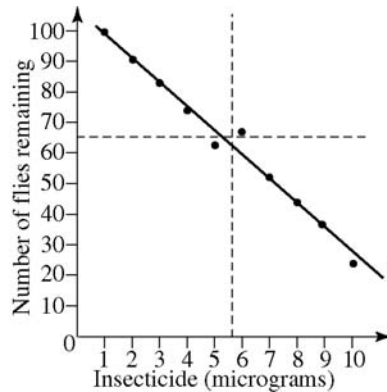
The birth rate is 32 per thousand.

h Both answers are reliable as they are interpolated and the relationship is quite close to linear.

3 a



b and c



$$a = 1, b = 4, c = 1, d = 4$$

$$q = \frac{(a+c) - (b+d)}{a+b+c+d}$$

$$= \frac{(1+1) - (4+4)}{1+4+1+4}$$

$$= -0.6$$

There is a moderate negative correlation. There is evidence to show that the greater the amount of insecticide, the lower the number of flies remaining after 2 hours.

d The points are not perfectly linear as the survival of flies depends on the location of the flies to the insecticide as well as the amount of insecticide.

e (x_1, y_1) (x_2, y_2)
 $(2, 92)$ $(8, 45)$

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$= \frac{45 - 92}{8 - 2}$$

$$= -7.8$$

$$y - y_1 = m(x - x_1)$$

$$y - 92 = -7.8(x - 2)$$

$$= -7.8x + 15.6$$

$$y = -7.8x + 15.6 + 92$$

$$y = -7.8x + 107.6$$

That is, $S = -7.8I + 108$

f The y-intercept is the number of survivors if no insecticide is used.

The gradient gives the decrease in the number of survivors per microgram of insecticide used.

g $S = -7.8I + 108$
 $= -7.8(8.5) + 108$
 $= -66.3 + 108$
 $= 41.7$

So there are about 42 survivors.

h $S = -7.8I + 108$
 $= -7.8(12.5) + 108$
 $= -97.5 + 108$
 $= 10.5$

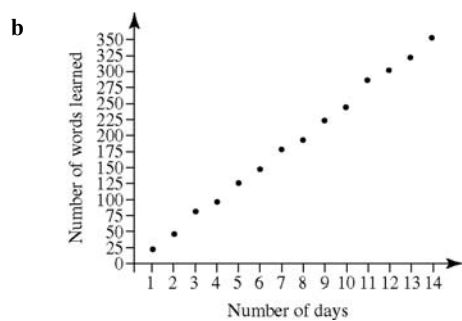
So there are about 10 survivors

i $0 = -7.8I + 108$
 $-108 = -7.8I$
 $13.85 = I$

For no survivors, approximately 13.85 micrograms of insecticide needs to be used.

j Part g is reliable as it is not extrapolated, i.e. it is interpolated.

- 4 a** Number of days from the beginning of the course is the independent variable. It is reasonable to expect the total number of words learned to depend on the number of days of studying. So the size of the vocabulary is the dependent variable.



- c The scatterplot suggests a strong positive linear correlation.

d

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14
v	24	50	78	97	127	149	175	196	223	248	280	300	321	350

lower set

upper set

There are 14 data points, so 7 goes to the lower set and 7 to the upper set.

$$\bar{x}_L = \frac{1+2+3+4+5+6+7}{7}$$

$$= 4$$

$$\bar{y}_L = \frac{24+50+78+97+127+149+175}{7}$$

$$= 100$$

$$(\bar{x}_L, \bar{y}_L) = (4, 100)$$

$$\bar{x}_U = \frac{8+9+10+11+12+13+14}{7}$$

= 11

$$\bar{y}_U = \frac{196 + 223 + 248 + 280 + 300 + 321 + 350}{7}$$

$$= 274$$

$$(\bar{x}_U, \bar{y}_U) = (11, 274)$$

$$m = \frac{274 - 100}{11 - 4}$$

$$= \frac{174}{7}$$

$$= 24\frac{6}{7}$$

$$y - 100 = 24 \frac{6}{7} (x - 4)$$

$$y = 24\frac{6}{7}x + \frac{4}{7}$$

$$\therefore V = 24 \frac{6}{7} n + \frac{4}{7}$$

- e The gradient is $24 \frac{6}{7}$ (approximately 25). So in this course, on average, Rachel learned about 25 new words per day.

f $n = 21$

$$V = 24 \frac{6}{7} \times 21 + \frac{4}{7}$$

$$= 522\frac{4}{7}$$

So by the end of the third week, she will learn 522 new words.

g $V = 500$

$$500 = 24\frac{6}{7}n + \frac{4}{7}$$

$$499\frac{3}{7} = 24\frac{6}{7}n$$

$$n = 499 \frac{3}{7} \div 24 \frac{6}{7}$$

$$n = 20.092$$

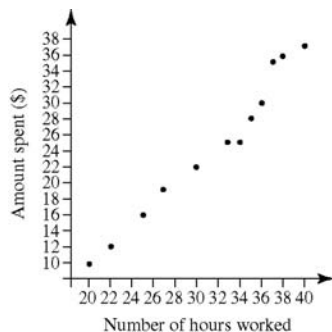
$$n \approx 20$$

It will take Rachel 20 days to learn 500 words.

- h** The answer is not reliable because $V = 500$ is outside the given data set and hence the prediction is obtained using extrapolation.

- 5 a** It is reasonable to expect the amount of money spent on take-away food to depend on the time spent at work. So, time spent working: independent; amount of money spent on take-away food: dependent.

b



- c** Strong positive correlation.

- d** Rearrange the data so that the x -values (that is, time at work) are in increasing order.

Time at work	20	22	25	27	30	33	34	35	36	37	38	40
Amount spent	10	12	16	19	22	25	25	28	30	35	36	37
	lower half						upper half					

There are 12 data points, so 6 goes to each set.

$$\bar{x}_L = \frac{20 + 22 + 25 + 27 + 30 + 33}{6}$$

$$= 26 \frac{1}{6}$$

$$\bar{y}_L = \frac{10 + 12 + 16 + 19 + 22 + 25}{6}$$

$$= 17 \frac{1}{3}$$

$$(\bar{x}_L, \bar{y}_L) = \left(26 \frac{1}{6}, 17 \frac{1}{3}\right)$$

$$\bar{x}_U = \frac{34 + 35 + 36 + 37 + 38 + 40}{6}$$

$$= 36 \frac{2}{3}$$

$$\bar{y}_U = \frac{25 + 28 + 30 + 35 + 36 + 37}{6}$$

$$= 31 \frac{5}{6}$$

$$(\bar{x}_U, \bar{y}_U) = \left(36 \frac{2}{3}, 31 \frac{5}{6}\right)$$

$$m = \frac{31 \frac{5}{6} - 17 \frac{1}{3}}{36 \frac{2}{3} - 26 \frac{1}{6}}$$

$$= \frac{14.5}{10.5}$$

$$= \frac{29}{21}$$

$$y - 17 \frac{1}{3} = \frac{29}{21} \left(x - 26 \frac{1}{6}\right)$$

$$y = \frac{29}{21}x - 18 \frac{101}{126}$$

(rounded to 2 decimal places, then $y = 1.38x - 18.80$)

So, $A = 1.38n - 18.80$

- e** With every extra hour of work, the amount of money spent on junk and take-away food per week increased by \$1.38.

- f** $n = 29$

$$A = 1.38 \times 29 - 18.80$$

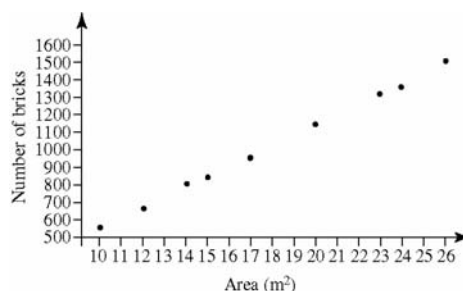
$$= 21.22$$

$$\approx \$21.20$$

- g** Interpolation, since $n = 29$ is within the given range of points.

- h** The prediction for part **f** is reliable, because interpolation has been used in obtaining it.

6 a



- b** Strong positive linear relationship (points nearly form a straight line).

- c** The horizontal and vertical line will both pass through the fifth point, which results in:

$$a = 4, b = 0, c = 4, d = 0$$

$$q = \frac{(a + c) - (b + d)}{a + b + c + d}$$

$$= \frac{(4 + 4) - (0 + 0)}{4 + 0 + 4 + 0}$$

$$= \frac{8}{8}$$

$$= 1$$

The value of $q = 1$ suggests perfectly linear relationship. This closely matches our observation in **b**.

- d** Arrange the data so that the x -values (that is, the area) are in order of increasing size.

Area	10	12	14	15	17	20	23	24	26
No. of bricks	572	686	800	863	965	1150	1323	1357	1502
	lower set					upper set			

There are 9 data points, so the 5th point (17, 965) goes to both sets.

$$\bar{x}_L = \frac{10 + 12 + 14 + 15 + 17}{5}$$

$$= 13.6$$

$$\bar{y}_L = \frac{572 + 686 + 800 + 863 + 965}{5}$$

$$= 777.2$$

$$(\bar{x}_L, \bar{y}_L) = (13.6, 777.2)$$

$$\bar{x}_U = \frac{17 + 20 + 23 + 24 + 26}{5}$$

$$= 22$$

$$\bar{y}_U = \frac{965 + 1150 + 1323 + 1357 + 1520}{5}$$

$$= 1259.4$$

$$(\bar{x}_U, \bar{y}_U) = (22, 1259.4)$$

$$m = \frac{1259.4 - 777.2}{22 - 13.6}$$

$$= \frac{482.2}{8.4}$$

$$\approx 57.40$$

$$y - 777.2 = 57.40(x - 13.6)$$

$$y = 57.40x - 3.44$$

$$\text{So, } N = 57.40 \times a - 3.44$$

- e Gradient of 57.40 means that about 57.4 bricks are needed to construct 1 m^2 of brick wall.

- f $a = 18$

$$N = 57.40 \times 18 - 3.44$$

$$= 1029.76.$$

So, Michael needs about 1030 bricks to complete the job. So, 1000 bricks is not enough.

- g Since 18 m^2 is within the given range of data, this is an example of interpolation.

- h Since we used interpolation, the result is reliable.

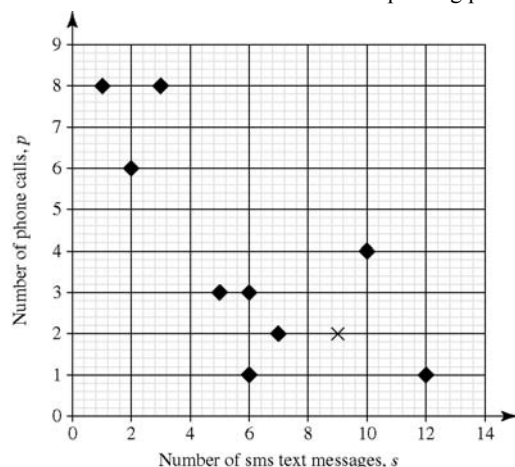
Exam practice 1

Multiple choice

- 1 E reading from the table less than 20 minutes, $4 + 6 = 10$ females
- 2 D time intervals in increasing order imply categorical and ordinal data
- 3 A this question involves expressing the data in percentage form. Particular attention should be given to the wording of the question, such as all students and of the males. 25 students use their mobile phones 20 minutes or less, not 25%
- 4 B $20 + 0.25 \times 15 = 23.75$. Students should ensure that the currency is consistent, such as all in dollars.
- 5 D similar to question 4, $C = 45 + 0.28n$
- 6 C the key point is the y intercept of 50 and a positive gradient

Extended response

- 1 a i Entering data into appropriate calculator. Using statistical programs, the following values are
Mean = 6.1 and standard deviation = 3.5
- ii 50%
This question involves recognition that 3 and 9 are the Q1 and Q3 and the percentage required is represented by the IQR.
- b Students need to be accurate when plotting points.



- c When determine (x_L, y_L) the x values need to be in ascending order. Rearranging the x values:

x	1	2	3	5	6	6	7	9	10	12
y	8	6	8	3	1	3	2	2	4	1

Divide the set of coordinates into two groups (lower and upper)

Find the average x and y values for each half:

(1 mark method: attempt to divide into two groups and find average)

$$x_L = \frac{1 + 2 + 3 + 5 + 6}{5}$$

$$= 3.4$$

$$y_L = \frac{8 + 6 + 8 + 3 + 1}{5}$$

$$= 5.2$$

Hence, (x_L, y_L) is (3.4, 5.2)

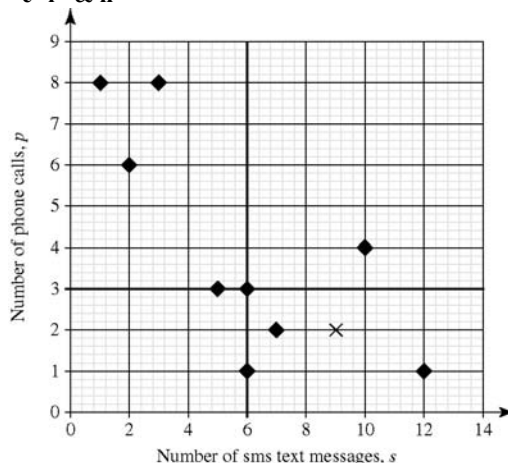
- d The general linear equation is:

$$y = mx + c$$

Substitute the points (8.8, 2.6) and (4.3, 5.2) into this equation and solve for x and y .

$$y = -0.48x + 6.83$$

- e i & ii



- iii

$$\begin{aligned} q &= \frac{(a+c)-(b+d)}{a+b+c+d} \\ &= \frac{(1+0)-(3+2)}{1+0+3+2} \\ &= \frac{-5}{7} \end{aligned}$$

- iv The q value is moderate and negative which indicates that students generally use their mobile phones for either text messaging or phone calls but not both, therefore, EasyDial should not spend money on their new campaign.