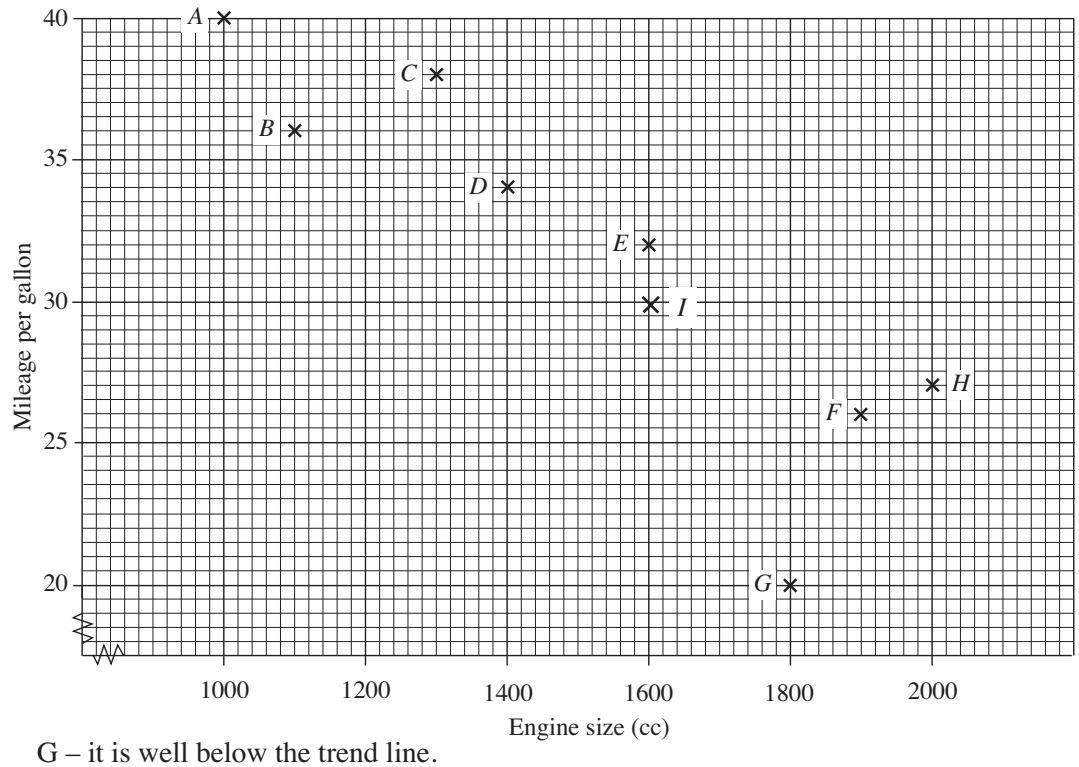
Correlation and Regression Answers

1 Correlation

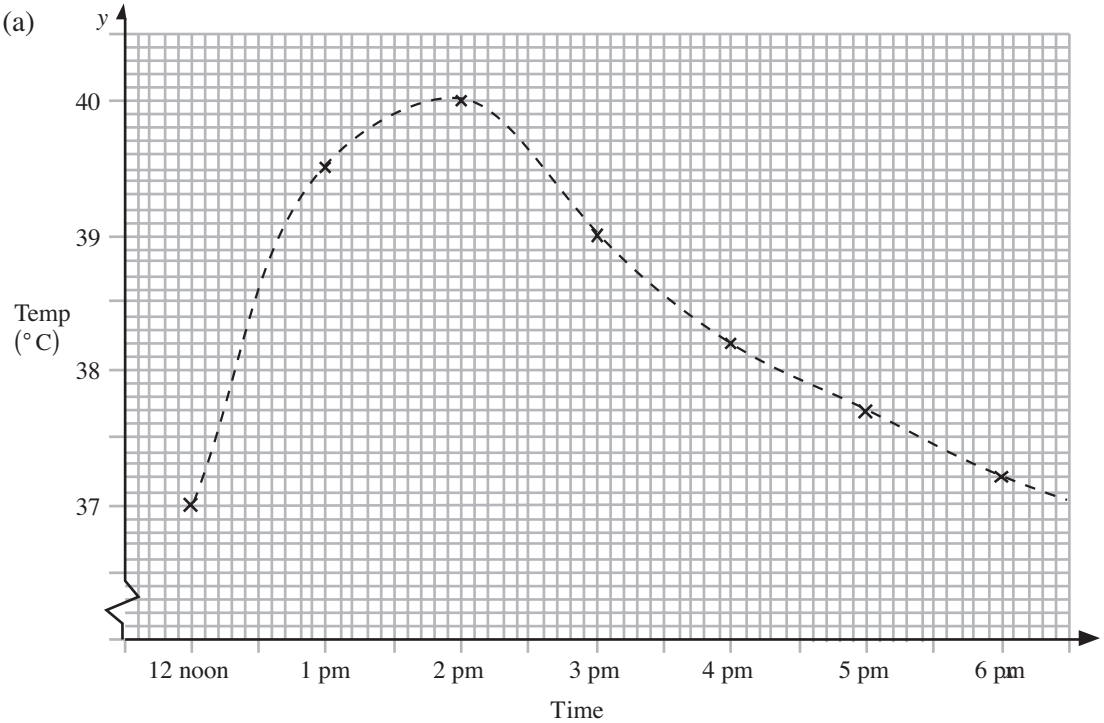
1. (a) E : 1600 cc ; F : 26 miles per gallon ; H : 2000 cc, 27 miles per gallon

(b)



(c) G – it is well below the trend line.

2. (a)

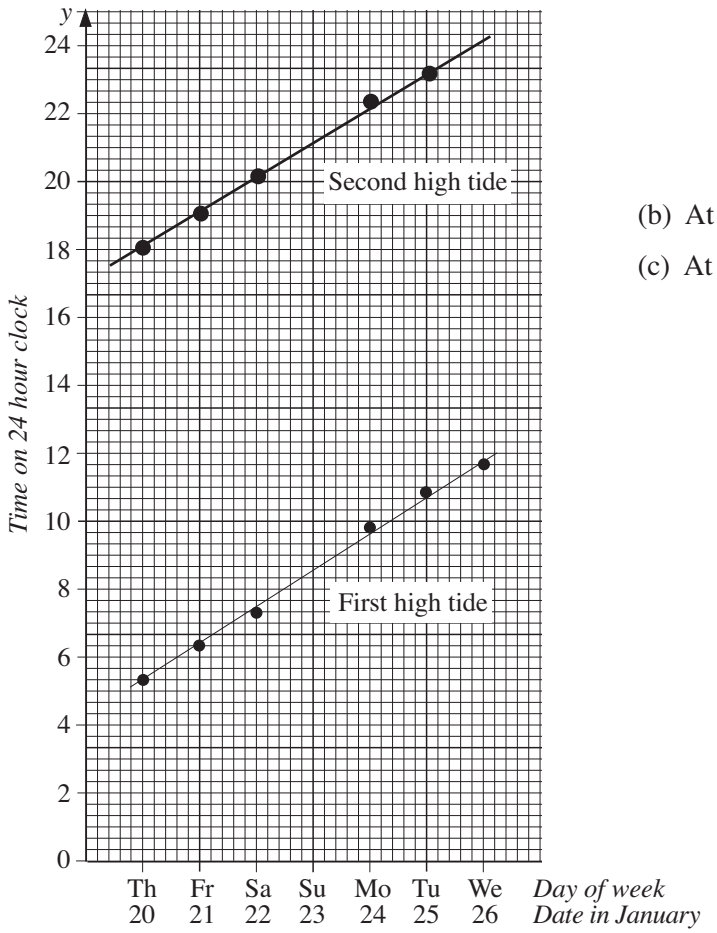


(b) (i) At about 12.45 pm (ii) 38.5 °C (c) At about 6.30 pm

Correlation and Regression Answers

1

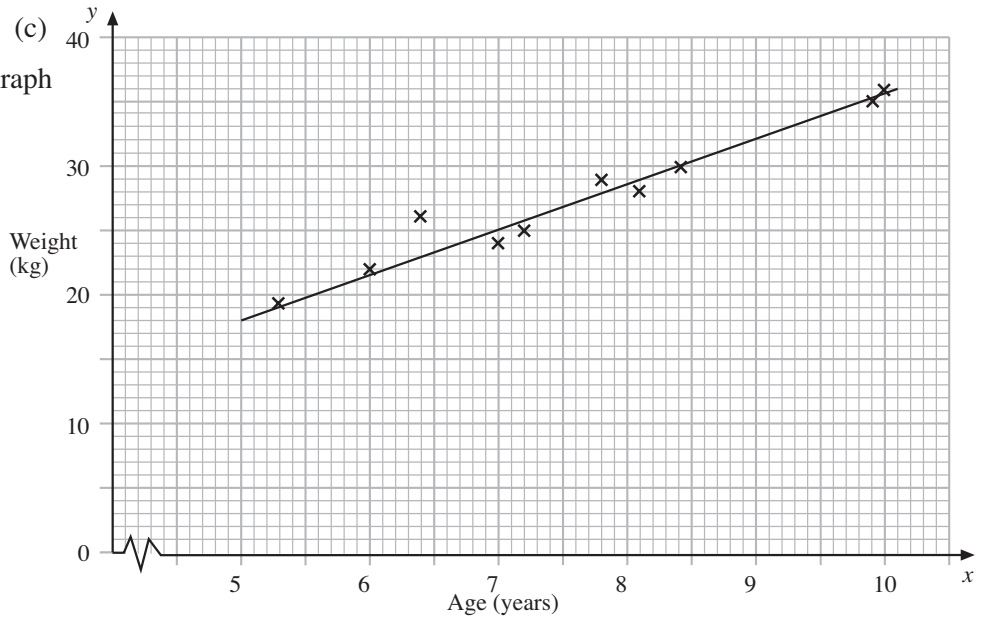
3. (a)



(b) At about 08.30 and 21.10

(c) At about 23.17 and 23.47

4. (a) and (c) graph



(b) 27.5 kg (d) $y = 3.4x + 1.6$ (e) 28.8 kg (f) Small sample size

5. (a) 28 (b) 2 (c) 80, 11

(d) No, the data shows no correlation.

Correlation and Regression Answers

2 Spearman's Rank Coefficient of Correlation

1. (a)

Typist	*A*	*B*	*C*	*D*	*E*	*F*
Rank time	3	5	2	4	1	6
Rank errors	4	2.5	5	2.5	6	1
d	1	2.5	3	1.5	5	5
d^2	1	6.25	9	2.25	25	25

(b) $\sum d^2 = 68.5$ and $r = -0.957$ (c) Quite fast and not too many errors.

2. (a) 0.006 (b) Small but positive correlation, giving some support to hypothesis.
(c) Beer does not improve with age.

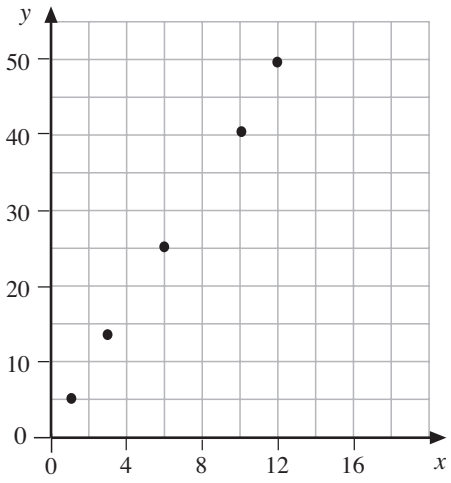
3.

Club	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*
Position in league	1	3	6	2	7	8	5	4
Average attendance	34	12	18	32	15	25	27	19
Rank of attendance	1	8	6	2	7	4	3	5
Difference in ranks (*d*)	0	5	0	0	0	4	2	1
d^2	0	25	0	0	0	16	4	1

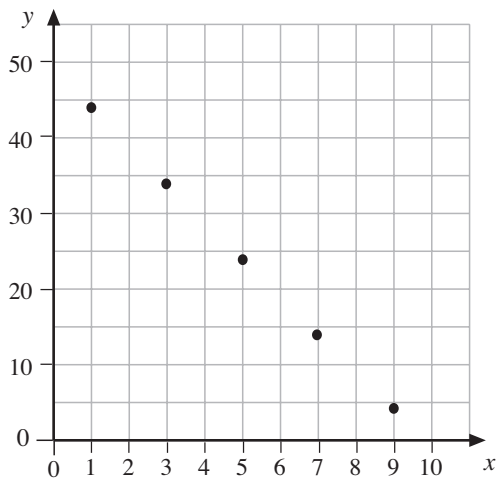
(b) 0.45 (c) Some positive correlation (c) Very positive correlation

3 Product Moment Correlation Coefficient

1. (a) (i)



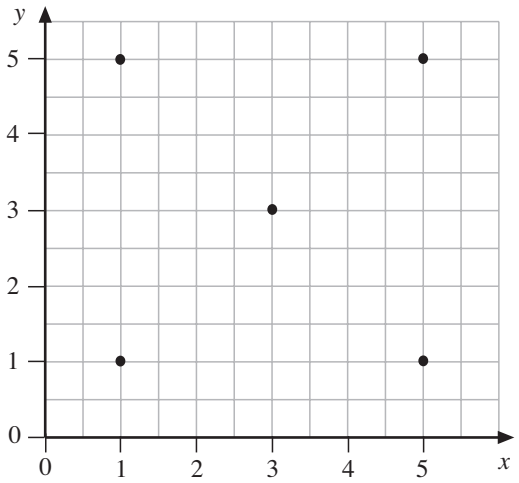
(ii)



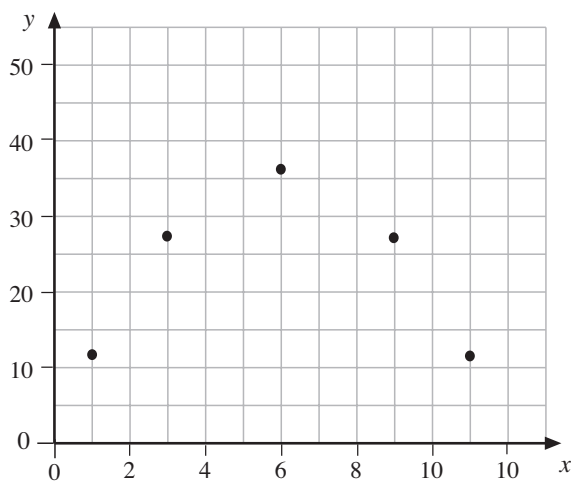
Correlation and Regression Answers

3

(iii)



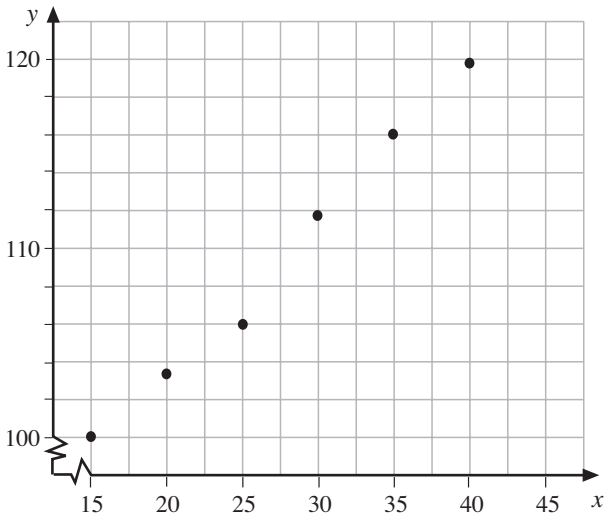
(iv)



(b) (i) $r = 1$ (ii) $r = -1$ (iii) $r = 0$ (iv) $r = 0$

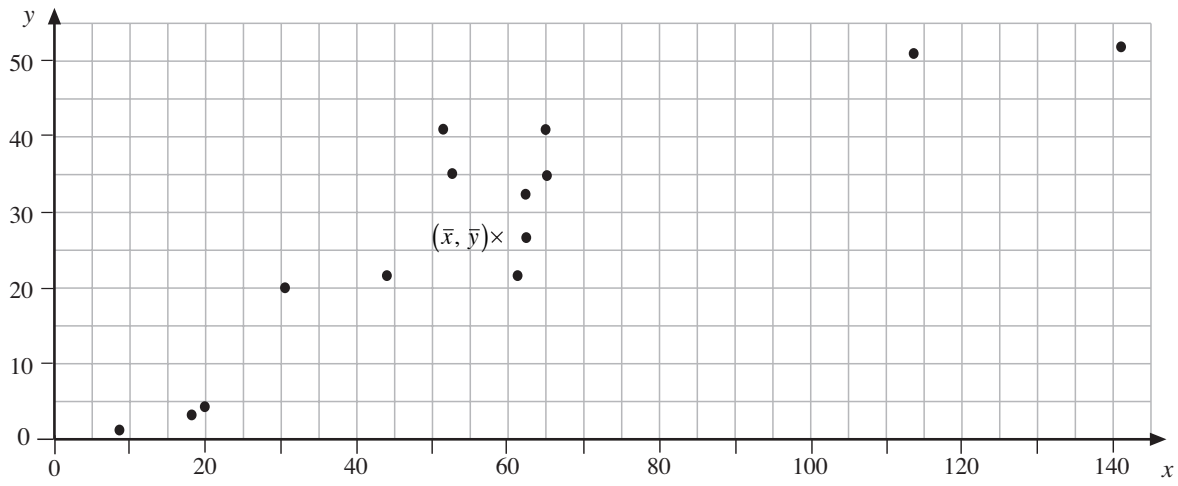
2. (a) –0.676 (b) –0.676 (c) 1

3. (a)



(b) $r = 0.995$; discard (25, 106.1) , $r = 1$

4. (a)



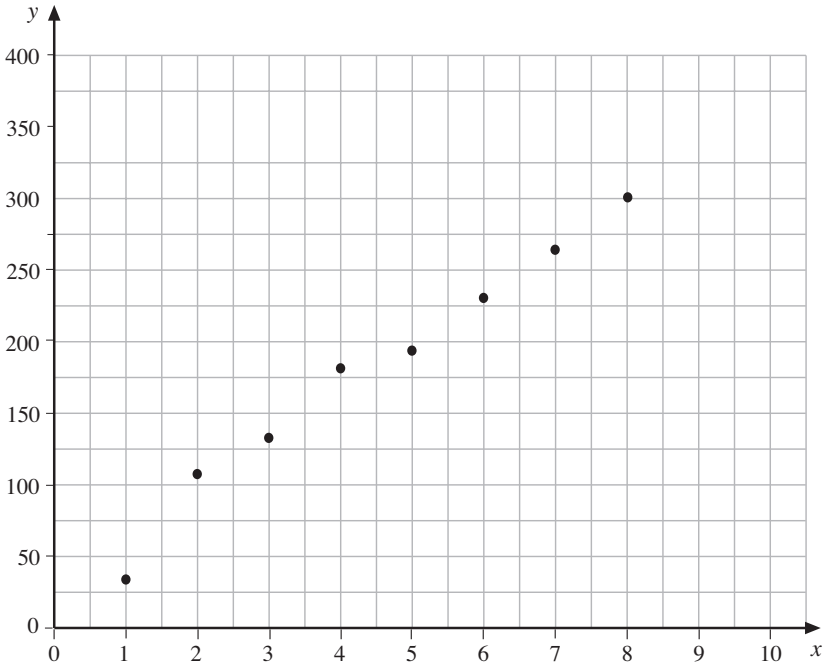
Correlation and Regression Answers

3

4. (b) $\bar{x} = 56.9,\ \bar{y} = 27.7$

(c) $s_x = 34.404, s_y = 15.966,\ r = 0.880$

5. (a)

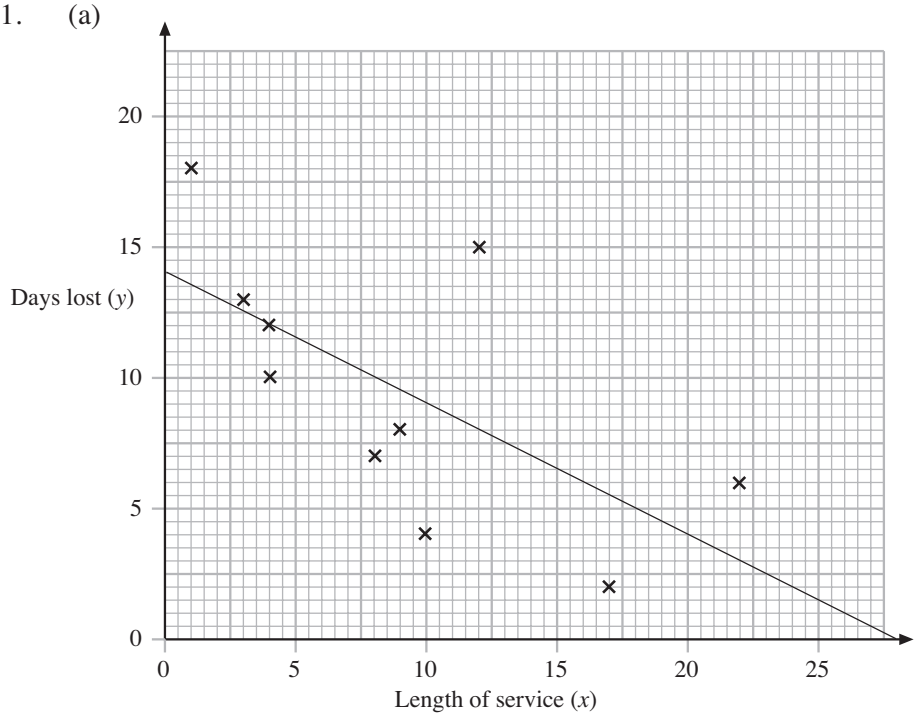


$r = 0.989$

(b) The linear growth from 1 to 8 does not continue as there is almost no growth from 13 to 15.

4 Regression Lines

1. (a)



Correlation and Regression Answers

4

1. (b) (i) $y = -0.5x + 14$

(ii) If $x = 0$, the person has not been employed so no lost days!

(c) 4

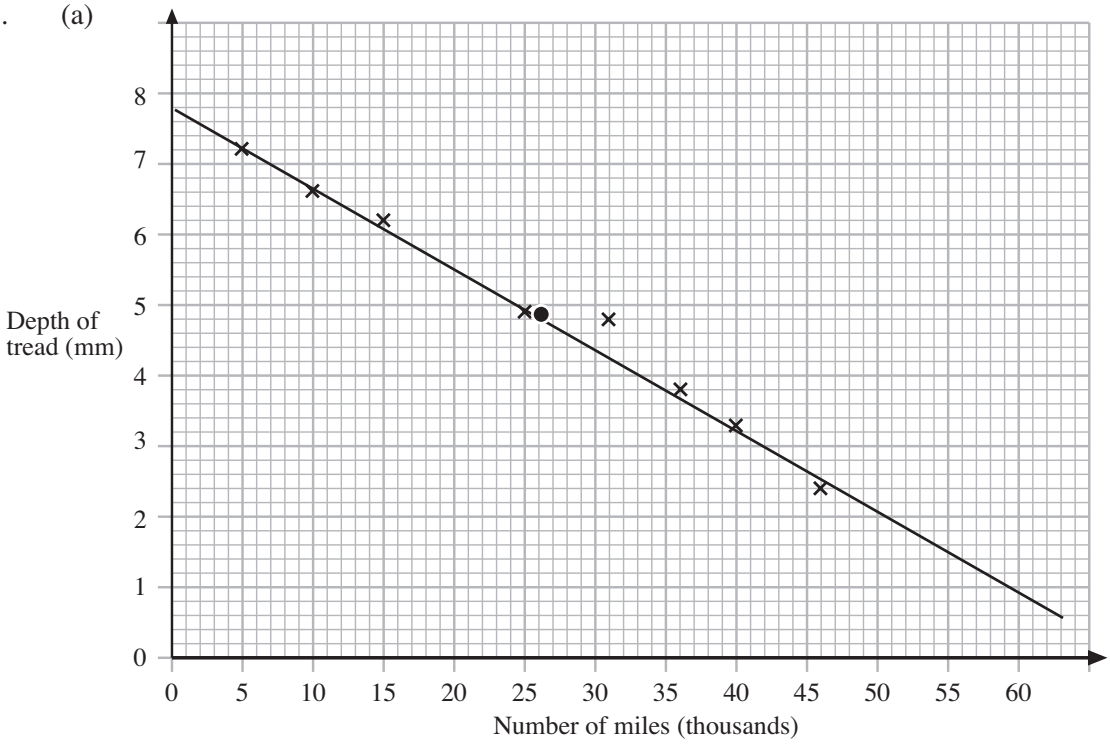
(d) E.g. The regression line is based only on employees who have worked for up to 22 years.
It would give a negative answer or 30 years of service.
The sample size, particularly for longer-serving employees, is very small..

2. (a) (i) 13 (ii) It corresponds to significantly lower takings.

(b) $y = 0.5x + 12$

(c) The store may be reaching its capacity in terms of shopfloor space and/or checkout desks; there might not be sufficient car parking space for potential customers.

3. (a)



(b) (i) 4.9 mm (ii) see diagram

(c) About 5.6 mm (d) 54 000 miles

(e) Answer to part (c) is more reliable as it is within the range of data given whereas the answer to part (d) is outside the data range.

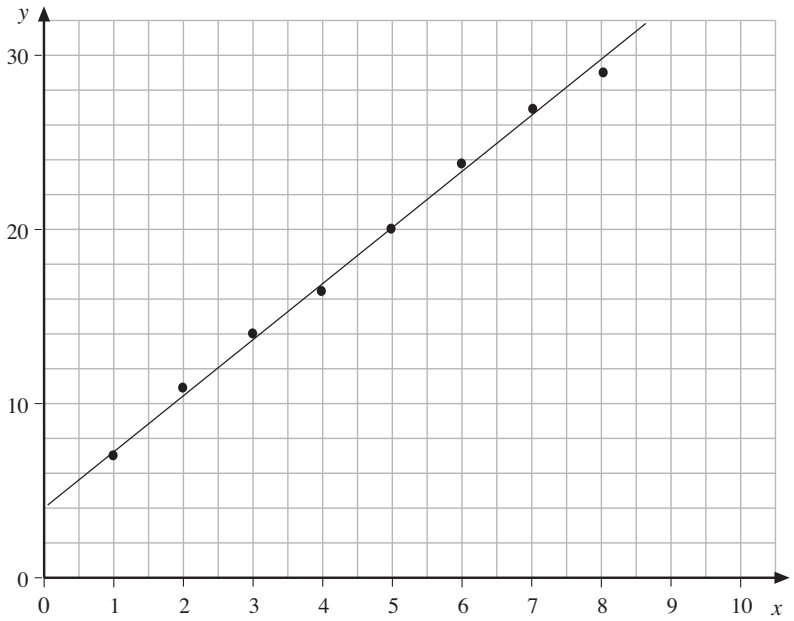
(f) Yes, as the tyres wear with increasing miles travelled.

Correlation and Regression Answers

5 Line of Regression Equation

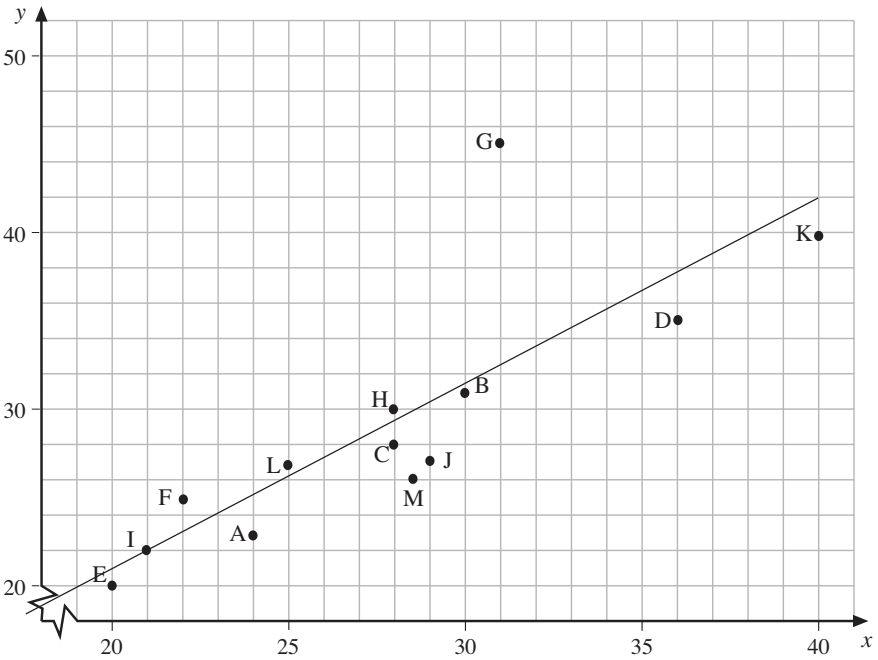
1. $y = 7.47x - 7.714$ (a) 478
   (b) 739 (c) 7464
   Last value if outside the range of data.

2. (a) (i) Diagram (ii) $y = 3.179x + 4.221$



   (b) Increase: 3.179 kg

3. (a)



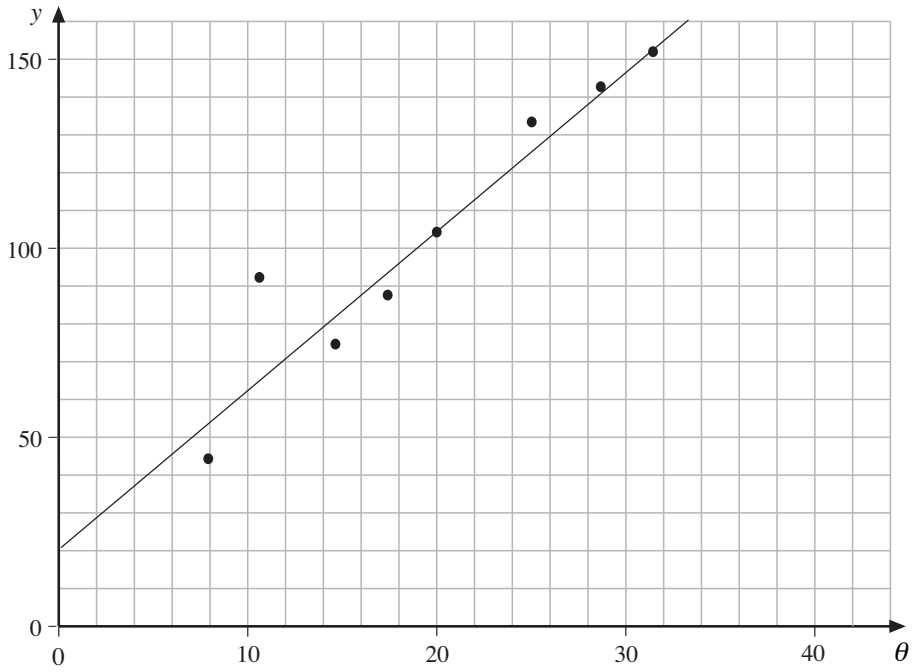
   (b) $y = 1.03x + 0.53$

   (c) All data close to line of regression except for G, which is an outlier.

Correlation and Regression Answers

5

4. (a)



(b) $y = 4.124\theta + 23.334$ (c) (10.5, 91)

(d) (i) 72.8 cm

(ii) 188.3 cm; the first answer is more likely to be a good estimate as it is inside the range of θ data (i.e. interpolation), whereas the second answer is based on a θ value outside the range of experimental data (i.e. extrapolation).

5. $H = 1579T - 11.028$

$H = 1.3$