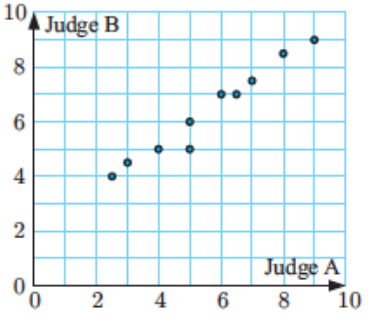
## EXERCISE 11A

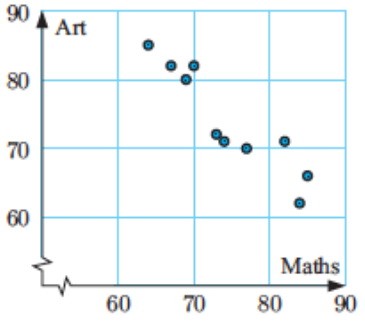
1 a weak positive correlation, linear, no outliers
b strong negative correlation, linear, one outlier
c no correlation
d strong negative correlation, not linear, one outlier
e moderate positive correlation, linear, no outliers
f weak positive correlation, not linear, no outliers

2 a




b There appears to be **strong, positive, linear** correlation between Judge A's scores and Judge B's scores. This means that as Judge A's scores increase, Judge B's scores **increase**.
c No, the scores are related to the quality of the ice-skaters' performances.

3 a




b There is a strong, negative, linear correlation between Mathematics and Art marks.

4 a **D** b **A** c **B** d **C**

5 a There is a moderate, positive, linear correlation between *hours of study* and *marks obtained*.
b The test is out of 50 marks, so the outlier ($> 50$) appears to be an error. It should be discarded.

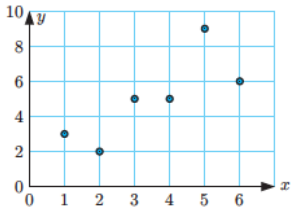
6 a Not causal, dependent on genetics and/or age.
b Not causal, dependent on the size of the fire. c Causal
d Causal e Not causal, dependent on population of town.

## EXERCISE 11B.1

1 weak positive correlation

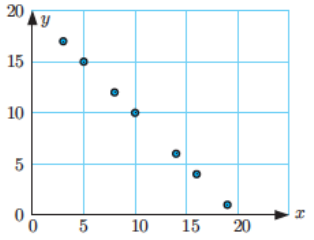
2 a B b A c D d C e E

3 a i



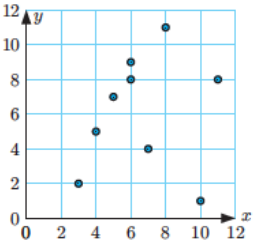

ii $r \approx 0.786$ iii moderate positive correlation

b i



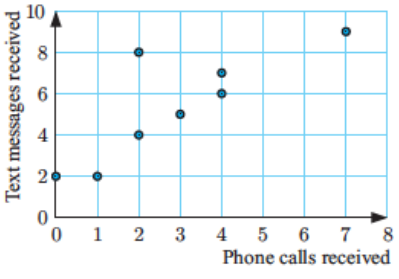

ii $r = -1$ iii perfect negative correlatio

c i




ii $r \approx 0.146$ iii weak positive correlation

4 a

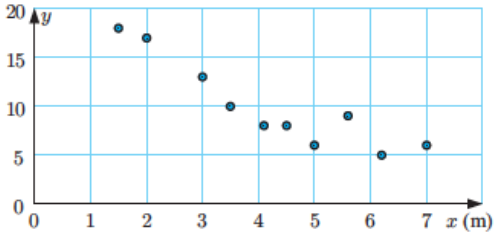



b $r \approx 0.816$ c moderate positive correlation

5 a $r \approx 0.917$

b strong positive correlation.
In general, the greater the young athlete's age, the further they can throw a discus.

6 a




b negative c $r \approx -0.911$

d strong negative correlation e decreases f yes

## EXERCISE 11B.2

1 **a** $\overline{x} = 6,\ \overline{y} = 6$

**b**

| | $x$ | $y$ | $x-\overline{x}$ | $y-\overline{y}$ | $(x-\overline{x})(y-\overline{y})$ | $(x-\overline{x})^2$ | $(y-\overline{y})^2$ |
|---|---|---|---|---|---|---|---|
| | 2 | 10 | −4 | 4 | −16 | 16 | 16 |
| | 4 | 7 | −2 | 1 | −2 | 4 | 1 |
| | 7 | 5 | 1 | −1 | −1 | 1 | 1 |
| | 11 | 2 | 5 | −4 | −20 | 25 | 16 |
| Totals: | 24 | 24 | 0 | 0 | −39 | 46 | 34 |

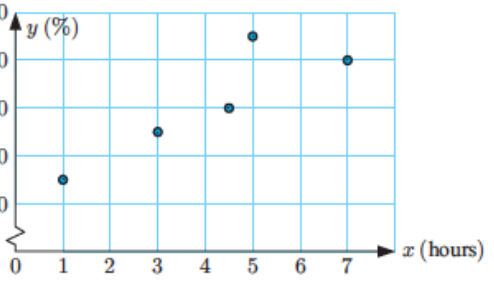
**c** $r \approx -0.986$

2 **a** $r = 1$, the data is perfectly positively linearly correlated.

**b** $r = -1$, the data is perfectly negatively linearly correlated.

**c** $r = 0$, there is no correlation.

3 **a**




**b** $r \approx 0.881$ **c** strong positive correlation

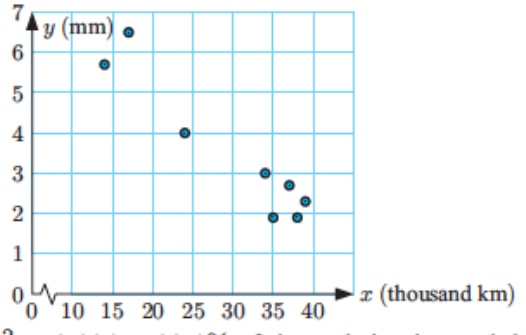
## EXERCISE 11B.3

1 57.8% of the variation in the number of visitors can be explained by the variation in maximum temperture.

2 $r^2 \approx 0.224$. 22.4% of the variation in money lost can be explained by the variation in time spent gambling.

3 $r^2 \approx 0.133$. 13.3% of the variation in heart rate can be explained by the variation in age.

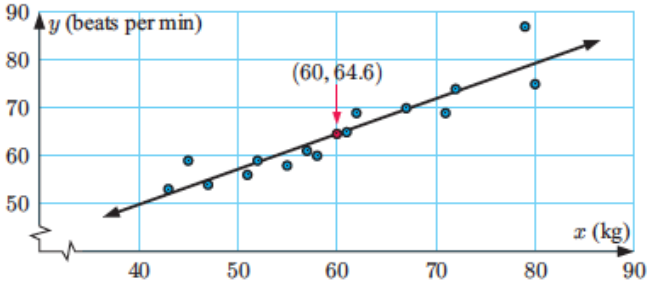
4 **a**




**b** $r^2 \approx 0.904$. 90.4% of the variation in tread depth can be explained by the variation in number of kilometres travelled.

## EXERCISE 11C

**1** **a, e**



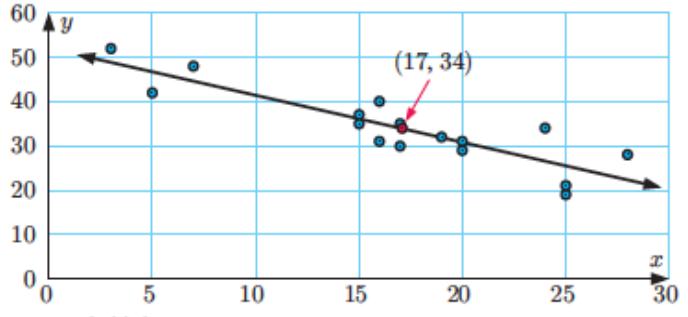

**b** $r \approx 0.929$

**c** There is a strong positive correlation between *weight* and *pulse rate*.

**d** (60, 64.6)

**f** 68 beats per minute. This is an interpolation, so the estimate is reliable.

**2** **a, d**

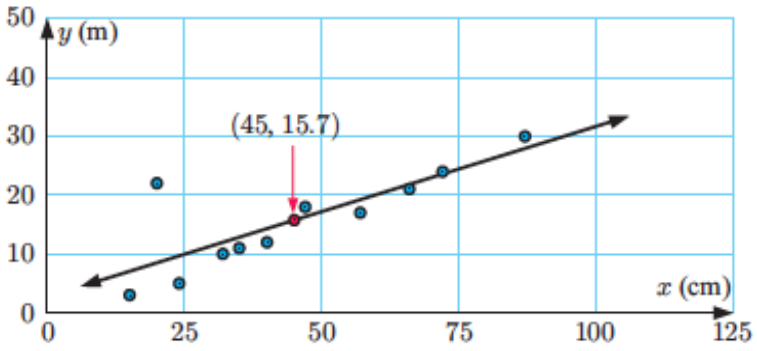



**b** $r \approx -0.878$

**c** There is a strong negative correlation between *number of speed cameras* and *number of car accidents*.

**e** At $y \approx 52$. This means that we would expect a city with no speed cameras to have approximately 52 car accidents.

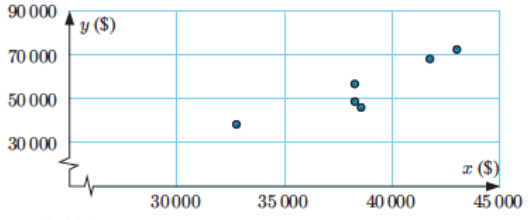
**3** **a, e**




**b** (20, 22) **c** moderate height, and thinner. **d** (45, 15.7)
**f** $\approx$ 37.4 m. This is an extrapolation, so the prediction may not be reliable.

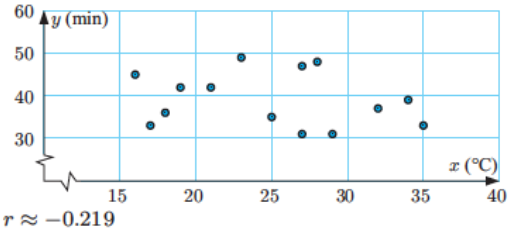
## EXERCISE 11D

**1** **a**




**b** $r \approx 0.921$
**c** There is a strong, positive, linear association between the starting salaries for Bachelor degrees and the starting salaries for PhDs.
**d** $y \approx 3.44x - 78300$
**e** **i** \$59 300
**ii** This is an interpolation, so the prediction is likely to be reliable.

**2** **a**




**b** $r \approx -0.219$
**c** There is a weak negative correlation between *temperature* and *time*.
**d** No, as there is almost no correlation.

3 a $r \approx -0.924$

b There is a strong, negative, linear correlation between the *petrol price* and the *number of customers*.

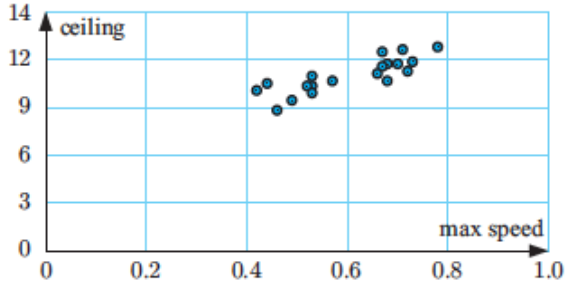
c $y \approx -4.27x + 489$

d gradient $\approx -4.27$, for every 1 cent per litre increase in the price of petrol, a service station will lose 4.27 customers.

e $-5.10$ customers

f It is impossible to have a negative number of customers. This extrapolation is not valid.

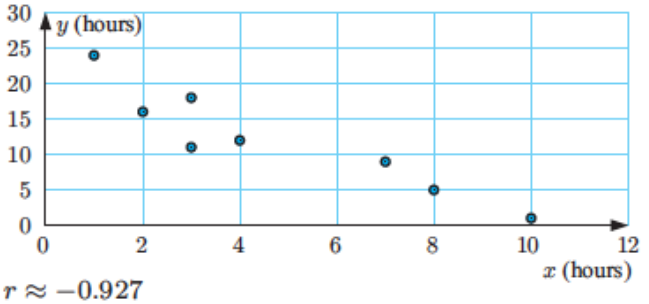
4 a




b $r \approx 0.840$ c moderate positive linear correlation

d $y \approx 8.12x + 6.09$ e 11.0 km

5 a




b $r \approx -0.927$

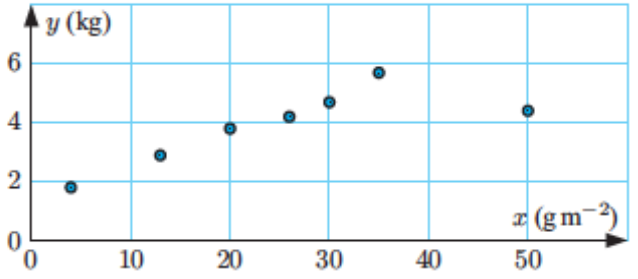
c There is a strong negative linear correlation between time exercising and time watching television.

d $y \approx -2.13x + 22.1$

e gradient $\approx -2.13$, for every hour the children spend exercising, they watch 2.13 hours less television.
$y$-intercept $\approx 22.1$ We would expect a child who does not exercise at all to watch 22.1 hours of television per week.

f 11.5 hours each week

**6** **a**




(50, 4.4) is the outlier.

**b** **i** $r \approx 0.798$ **ii** $r \approx 0.993$

**c** **i** $y \approx 0.0672x + 2.22$ **ii** $y \approx 0.119x + 1.32$

**d** The one which excludes the outlier.

**e** Too much fertiliser often kills the plants, or makes them sick.

## EXERCISE 11E.1

**1** **a**

| | *Likes chicken* | *Dislikes chicken* | *sum* |
|---|---|---|---|
| *Likes fish* | 45 | 15 | 60 |
| *Dislikes fish* | 30 | 10 | 40 |
| *sum* | 75 | 25 | 100 |

**b**

| | *Drove to work* | *Cycled to work* | *Public transport* | *sum* |
|---|---|---|---|---|
| *Male* | 25.3 | 7.7 | 11 | 44 |
| *Female* | 20.7 | 6.3 | 9 | 36 |
| *sum* | 46 | 14 | 20 | 80 |

**c**

| | *Junior school* | *Middle school* | *High school* | *sum* |
|---|---|---|---|---|
| *Plays sport* | 38.28 | 56.76 | 69.96 | 165 |
| *Does not play sport* | 19.72 | 29.24 | 36.04 | 85 |
| *sum* | 58 | 86 | 106 | 250 |

**d**

| | *Wore hat and sunscreen* | *Wore hat or sunscreen* | *Wore neither* | *sum* |
|---|---|---|---|---|
| *Sunburnt* | 10.92 | 6.16 | 3.92 | 21 |
| *Not sunburnt* | 28.08 | 15.84 | 10.08 | 54 |
| *sum* | 39 | 22 | 14 | 75 |

**2** **a**

| | *Pass Maths test* | *Fail Maths test* | *sum* |
|---|---|---|---|
| *Male* | 30 | 20 | 50 |
| *Female* | 30 | 20 | 50 |
| *sum* | 60 | 40 | 100 |

**b** In a sample of 100 students, we would expect 30 to be male and pass the Maths test.

**c**

| $f_o$ | $f_e$ | $f_o - f_e$ | $(f_o - f_e)^2$ | $\frac{(f_o - f_e)^2}{f_e}$ |
|---|---|---|---|---|
| 24 | 30 | −6 | 36 | 1.2 |
| 26 | 20 | 6 | 36 | 1.8 |
| 36 | 30 | 6 | 36 | 1.2 |
| 14 | 20 | −6 | 36 | 1.8 |
| | | | *Total* | 6 |

$\chi^2_{calc} = 6$

**3** **a** **i**

| | *Likes football* | *Dislikes football* | *sum* |
|---|---|---|---|
| *Male* | 14.56 | 11.44 | 26 |
| *Female* | 13.44 | 10.56 | 24 |
| *sum* | 28 | 22 | 50 |

**ii** $\chi^2_{calc} \approx 13.5$

**b** **i**

| | *Full-time job* | *Part-time job* | *Unemployed* | *sum* |
|---|---|---|---|---|
| *Left handed* | 16 | 14 | 10 | 40 |
| *Right handed* | 144 | 126 | 90 | 360 |
| *sum* | 160 | 140 | 100 | 400 |

**ii** $\chi^2_{calc} \approx 1.05$

**c** **i**

| | Age | | | |
|---|---|---|---|---|
| | 18 - 29 | 30 - 39 | 40+ | *sum* |
| *Married* | 13.99 | 15.67 | 17.35 | 47 |
| *Single* | 11.01 | 12.33 | 13.65 | 37 |
| *sum* | 25 | 28 | 31 | 84 |

**ii** $\chi^2_{calc} \approx 4.35$

**d** **i**

| | | Visits Art Gallery | | | |
|---|---|---|---|---|---|
| | | *Often* | *Rarely* | *Never* | *sum* |
| Visits Museum | *Often* | 37.60 | 37.95 | 35.45 | 111 |
| | *Rarely* | 36.58 | 36.93 | 34.49 | 108 |
| | *Never* | 30.82 | 31.12 | 29.06 | 91 |
| | *sum* | 105 | 106 | 99 | 310 |

**ii** $\chi^2_{calc} \approx 25.1$

## EXERCISE 11E.2

**1** $\chi^2_{calc} \approx 6.61$, $\text{df} = 2$, $p \approx 0.0368$

As $\chi^2_{calc} > 5.99$, we reject $H_0$, and conclude that the variables *weight* and *suffering diabetes* are not independent.

**2** **a** 4.61

**b** $\chi^2_{calc} \approx 8.58$, $\text{df} = 2$, $p \approx 0.0137$

As $\chi^2_{calc} > 4.61$, we reject $H_0$. So at a 10% level, we conclude that *age* and the *party to vote for* are not independent.

**3** $\chi^2_{calc} \approx 2.56$, $\text{df} = 3$, $p \approx 0.456$

As $\chi^2_{calc} < 11.34$, we do not reject $H_0$. So at a 1% level, *gender* and *favourite season* are independent.

**4** **a** $\chi^2_{calc} \approx 23.6$, $\text{df} = 3$, $p \approx 0.000\,029\,9$

As $\chi^2_{calc} > 7.81$, we reject $H_0$. So at a 5% level, *reason for travelling* and *rating* are not independent.

**b** Guests travelling for a holiday are more likely to give a higher rating.

**5** $\chi^2_{calc} \approx 18.4$, $\text{df} = 6$, $p \approx 0.005\,28$

As $\chi^2_{calc} > 12.59$, we reject $H_0$. So at a 5% level, *hair colour* and *eye colour* are not independent.

**6** $\chi^2_{calc} \approx 7.94$, $\text{df} = 6$, $p \approx 0.242$

As $\chi^2_{calc} < 10.64$, we do not reject $H_0$. So at a 10% level, *position* and *injury type* are independent.

## EXERCISE 11E.3

**1 a**

| | | *Own a pet?* | | |
|---|---|---|---|---|
| | | *Yes* | *No* | *sum* |
| *Age* | 0 - 19 | 4.02 | 3.98 | 8 |
| | 20 - 29 | 27.1 | 26.9 | 54 |
| | 30 - 49 | 50.2 | 49.8 | 100 |
| | 50+ | 36.7 | 36.3 | 73 |
| | *sum* | 118 | 117 | 235 |

**b** Yes, 4.02 and 3.98.

**c** Combine the 0 - 19 and 20 - 29 rows.

**2 a** $\chi^2_{calc} \approx 16.9$, $df = 6$, $p = 0.009\,59$

As $\chi^2_{calc} > 16.81$, we reject $H_0$. So at a 1% level, we conclude that *intelligence level* and *cigarette smoking* are not independent.

**b**

| | *Intelligence level* | | | | |
|---|---|---|---|---|---|
| | *low* | *average* | *high* | *very high* | *sum* |
| *Non smoker* | 262 | 383 | 114 | 4.69 | 763 |
| *Medium level smoker* | 133 | 194 | 57.7 | 2.38 | 387 |
| *Heavy smoker* | 107 | 157 | 46.6 | 1.93 | 313 |
| *sum* | 502 | 734 | 218 | 9 | 1463 |

**c** Combine the *high* and *very high* data columns.

**d** $\chi^2_{calc} \approx 13.2$, $df = 4$, $p = 0.0104$

As $\chi^2_{calc} < 13.28$, we do not reject $H_0$. So at a 1% level, *intelligence level* and *cigarette smoking* are independent. This is a different conclusion from the one in **a**.

## EXERCISE 11E.4

**1 a**

| | | *Result* | | |
|---|---|---|---|---|
| | | *Heads* | *Tails* | *sum* |
| *Guess* | *Heads* | 49.4 | 54.6 | 104 |
| | *Tails* | 45.6 | 50.4 | 96 |
| | *sum* | 95 | 105 | 200 |

**b** $\chi^2_{calc} \approx 1.35$

**c** As $\chi^2_{calc} < 3.84$, we do not reject $H_0$. So at a 5% level, Horace's *guess* and *result* are independent.

**d** According to this test, Horace's claim is not valid.

**2 a**

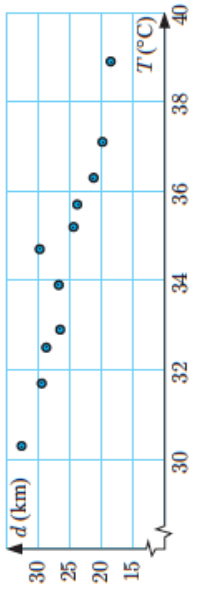
| | | *Result* | | |
|---|---|---|---|---|
| | | *Pass* | *Fail* | *sum* |
| *Country* | *France* | 63.8 | 21.2 | 85 |
| | *Germany* | 168 | 55.8 | 224 |
| | *sum* | 232 | 77 | 309 |

**b** 2.71 **c** $\chi^2_{calc} \approx 4.62$

**d** As $\chi^2_{calc} > 2.71$, we reject $H_0$. So at a 10% level, *motorbike test result* and *country* are not independent.

## REVIEW SET 11A

**1** **a**



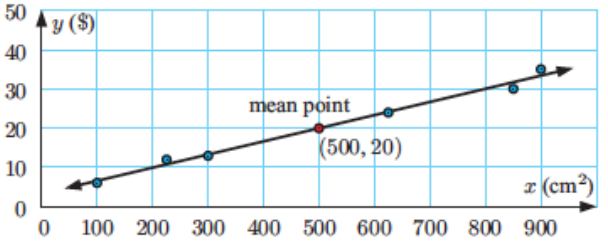

**b** $r \approx -0.928$, a strong negative linear relationship exists between the variables.

**c** $d = -1.64T + 82.3$ **d** 50.0°C

**2** $\chi^2_{calc} \approx 7.37$, df $= 2$, $p \approx 0.0251$

As $\chi^2_{calc} > 5.99$ we reject $H_0$. So, at a 5% level, *wearing a seat belt* and *severity of injury* are not independent.

**3** **a, e**



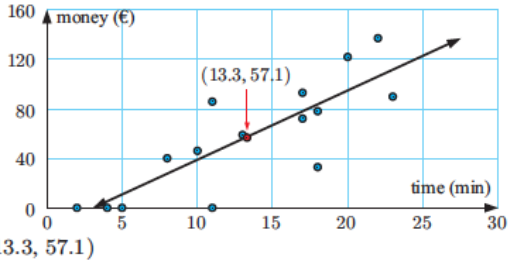

**b** $r \approx 0.994$

**c** There is a very strong positive correlation between *area* and *price*.

**d** $y \approx 0.0335x + 3.27$

**f** $43.42, this is an extrapolation, so it may be unreliable.

**4** **a, c**

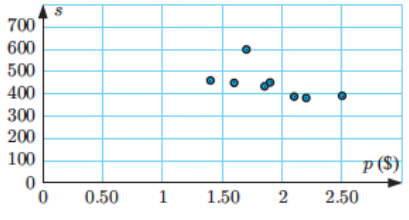



**b** (13.3, 57.1)

**d** There is a moderate positive linear correlation between *time in the store* and *money spent*.

**e** €66.80. This is an interpolation, so the estimate is reliable.

**5** **a**




**b** Yes, the point (1.7, 597) is an outlier. It should not be deleted as there is no evidence that it is a mistake.

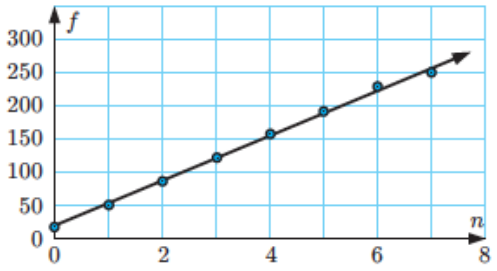
**c** $s \approx -116p + 665$

**d** No, the prediction would not be accurate, as that much extrapolation is unreliable.

**6** **a** number of waterings, $n$ **b** $f \approx 34.0n + 19.3$

**c** Yes, plants need water to grow, so it is expected that an increase in watering will result in an increase in flowers.

**d**




**e** **i** 104 ($n = 2.5$), 359 ($n = 10$)

**ii** $n = 10$ is unreliable as it is outside the poles and over watering could be a problem. $n = 2.5$ is reliable.

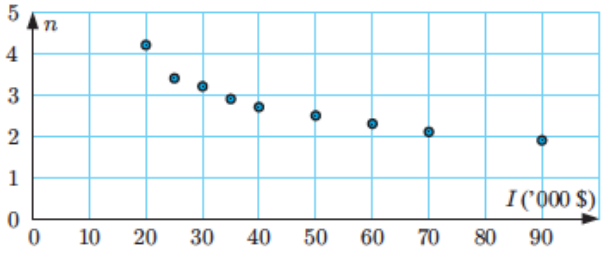
**7** $\chi^2_{calc} \approx 13.0$, df $= 6$, $p \approx 0.0433$

**a** As $\chi^2_{calc} > 12.59$, we reject $H_0$. So, at a 5% level, $P$ and $Q$ are not independent.

**b** As $\chi^2_{calc} < 16.81$, we do not reject $H_0$. So, at a 1% level, $P$ and $Q$ are independent.

## REVIEW SET 11B

**1** **a**




**b** $r \approx -0.908$ **c** $n \approx -0.0284I + 4.12$

**d** **i** 2.84 children **ii** 0.144 children

**e** The first is interpolation, so the estimate is reliable. The second is extrapolation, so the estimate may not be reliable.

**2** **a** **i** Negative correlation. As prices increase, the number of tickets sold is likely to decrease.

**ii** Causal. Less people will be able to afford tickets as the prices increase.

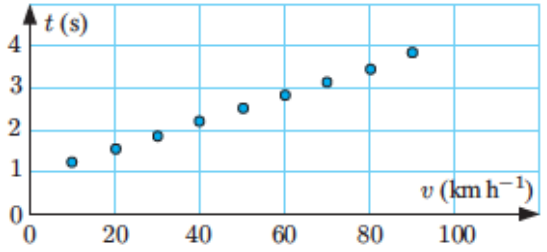
**b** **i** Positive correlation. As icecream sales increase, number of drownings is likely to increase.

**ii** Not causal. Both these variables are dependent on the number of people at the beach.

3 $\chi^2_{calc} \approx 42.1$, $\text{df} = 2$, $p \approx 7.37 \times 10^{-10}$
As $\chi^2_{calc} > 4.61$, we reject $H_0$. So at a 10% level, *age of driver* and *increasing the speed limit* are not independent.

4 $\chi^2_{calc} \approx 25.6$, $\text{df} = 9$, $p \approx 0.002\,41$
As $\chi^2_{calc} > 21.67$, we reject $H_0$. So at a 1% level, *intelligence level* and *business success* are not independent.

5 a

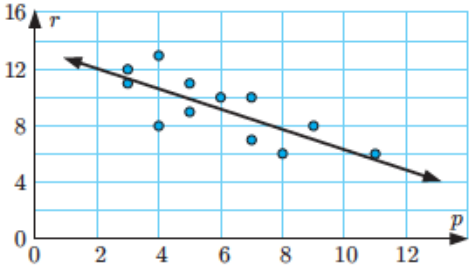



b $t \approx 0.0322v + 0.906$

c i 2.68 seconds ii 4.44 seconds

d The driver's reaction time.

6 a




b $r \approx -0.706p + 13.5$ dozen maidens

c $r \approx -0.763$. There is a moderate negative relationship. This supports Superman's suspicions.

d 9.25 dozen (111 maidens)

e This would predict that Silent Predator would abduct a negative number of maidens, which is unrealistic.

f $r$-int $\approx 13.5$, $p$-int $\approx 19.1$ These represent how many dozen maidens we would expect one villain to abduct if the other villain did not abduct any.

g Silent Predator

≈