**1** The following table gives the average number of children for different family incomes.

| Income ($I$ thousand $) | 20 | 25 | 30 | 35 | 40 | 50 | 60 | 70 | 90 |
|---|---|---|---|---|---|---|---|---|---|
| Number of children, $n$ | 4.2 | 3.4 | 3.2 | 2.9 | 2.7 | 2.5 | 2.3 | 2.1 | 1.9 |

   **a** Construct an appropriate graph to display the data.

   **b** Find $r$.

   **c** Find the equation of the line of best fit.

   **d** Estimate the average number of children for a family income of:

      **i** $45 000             **ii** $140 000

   **e** Comment on the reliability of your estimates.

**2** For the following pairs of variables, discuss:

      **i** whether the correlation between the variables is likely to be positive or negative

      **ii** whether a causal relationship exists between the variables.

   **a** *price of tickets* and *number of tickets sold*

   **b** *ice cream sales* and *number of drownings*.

**3** The table shows the responses to a survey about whether the city speed limit should be increased.

Test at a 10% level whether there is any association between the *age of a driver* and *increasing the speed limit*.

| | Age of driver | | |
|---|---|---|---|
| | 18 to 30 | 31 to 54 | 55+ |
| Increase | 234 | 169 | 134 |
| No increase | 156 | 191 | 233 |

**4** The following table shows the results from a major investigation considering the two factors *intelligence level* and *business success*.

| | | Intelligence level | | | |
|---|---|---|---|---|---|
| | | Low | Average | High | Very high |
| Business success | No success | 35 | 30 | 41 | 25 |
| | Low success | 28 | 41 | 26 | 29 |
| | Success | 35 | 24 | 41 | 56 |
| | High success | 52 | 38 | 63 | 72 |

At a 1% level with $df = 9$, the critical value is 21.67. Test at a 1% level whether there is a link between *intelligence level* and *business success*.

**5** Safety authorities advise drivers to travel three seconds behind the car in front of them. This provides the driver with a greater chance of avoiding a collision if the car in front has to brake quickly or is itself involved in an accident.

A test was carried out to find out how long it would take a driver to bring a car to rest from the time a red light was flashed. It involved one driver in the same car under the same test conditions.

| Speed ($v$ km h$^{-1}$) | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
|---|---|---|---|---|---|---|---|---|---|
| Stopping time ($t$ s) | 1.23 | 1.54 | 1.88 | 2.20 | 2.52 | 2.83 | 3.15 | 3.45 | 3.83 |

  **a** Produce a scatter diagram of the data.

  **b** Find the linear model which best fits the data.

  **c** Hence estimate the stopping time for a speed of:    **i** 55 km h$^{-1}$    **ii** 110 km h$^{-1}$

  **d** Interpret the vertical intercept of the model.

**6** Two supervillains, Silent Predator and the Furry Reaper, terrorise Metropolis by abducting fair maidens (most of whom happen to be journalists). The superhero Superman believes that they are collaborating, alternatively abducting fair maidens so as not to compete with each other for ransom money. He plots their abduction rate below, in dozens of maidens.

| Silent Predator ($p$) | 4 | 6 | 5 | 9 | 3 | 5 | 8 | 11 | 3 | 7 | 7 | 4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Furry Reaper ($r$) | 13 | 10 | 11 | 8 | 11 | 9 | 6 | 6 | 12 | 7 | 10 | 8 |

  **a** Plot the data on a scatter diagram with Silent Predator on the horizontal axis.

  **b** Find the least squares regression line.

  **c** Calculate $r$, and hence describe the strength of Silent Predator and Furry Reaper's relationship. Is there any evidence to support Superman's suspicions?

  **d** Estimate the number of the Furry Reaper's abductions when the Silent Predator's were 6 dozen.

  **e** Why is the model inappropriate when the Furry Reaper abducts more than 20 dozen maidens?

  **f** Calculate the $p$- and $r$-intercepts of the regression line. What do these values represent?

  **g** If Superman is faced with a choice of capturing one supervillain but not the other, which should he choose?

# ANSWER

**1 a**
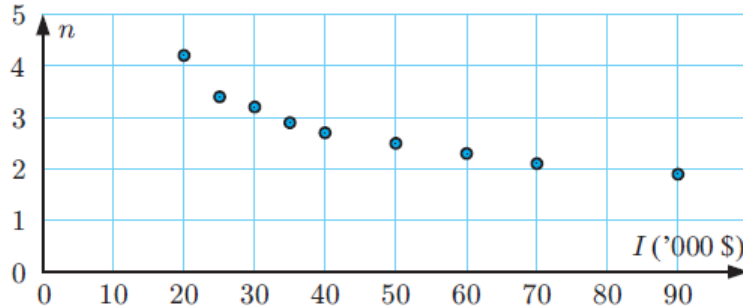


**b** $r \approx -0.908$    **c** $n \approx -0.0284I + 4.12$

**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.
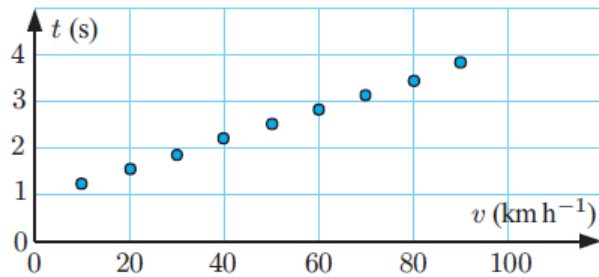
**3** $\chi^2_{calc} \approx 42.1$,   df $= 2$,   $p \approx 7.37 \times 10^{-10}$

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** $\chi^2_{calc} \approx 25.6$,   df $= 9$,   $p \approx 0.00241$

As $\chi^2_{calc} > 21.67$, we reject $H_0$. So at a 1% level, *intelligence level* and *business success* are not independent.
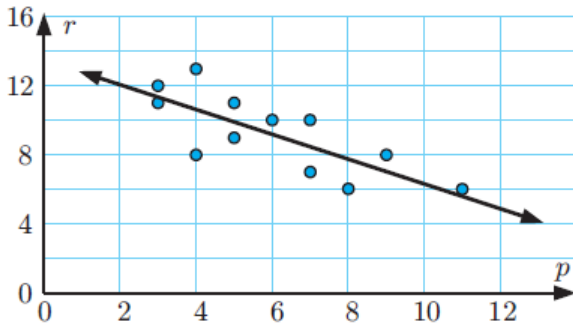
**5 a**



**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