13.1 Overview

Numerous videos and interactivities are embedded just where you need them, at the point of learning, in your learnON title at www.jacplus.com.au. They will help you to learn the content and concepts covered in this topic.

13.1.1 Why learn this?

Observations of two or more variables are often recorded, for example, the heights and weights of individuals. Studying the data allows us to investigate whether there is any relationship between the variables, how strong the relationship is, and whether one variable can be effectively predicted from information about another variable. Statistics can be applied to medical research, sport, agriculture, sustainability, weather forecasting and fashion trends, to name but a few fields. The capacity to analyse data and draw conclusions is an essential skill in a world where information is readily available and often manipulated.

13.1.2 What do you know?

1. THINK List what you know about interpreting bivariate data. Use a thinking tool such as a concept map to show your list.
2. PAIR Share what you know with a partner and then with a small group.
3. SHARE As a class, create a thinking tool such as a large concept map that shows your class’s knowledge of interpreting bivariate data.
13.2 Bivariate data

13.2.1 Bivariate data

- **Bivariate data** are data with two variables.
- The list of bivariate data can be considered as numerical pairs of the type: 
  \((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\)
- Bivariate data are usually represented graphically on scatterplots.

13.2.2 Scatterplots

- A **scatterplot** is a graph that shows whether there is a relationship between two variables.
- Each data value on a scatterplot is shown by a point on a Cartesian plane.

13.2.3 Dependent and independent variables

- One variable is generally the **dependent variable**, and the other the **independent variable**.
- The dependent variable, as the name suggests, is the one whose value depends on the other variable. The independent variable takes on values that do not depend on the value of the other variable.
- When data are expressed in the form of a table, generally the independent variable is written in the first row or the first column.
- The independent variable is placed on the x-axis and the dependent variable on the y-axis.

**WORKED EXAMPLE 1**

The table shows the total revenue from selling tickets for a number of different chamber music concerts. Represent these data on a scatterplot.

<table>
<thead>
<tr>
<th>Number of tickets sold</th>
<th>400</th>
<th>200</th>
<th>450</th>
<th>350</th>
<th>250</th>
<th>300</th>
<th>500</th>
<th>400</th>
<th>350</th>
<th>250</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total revenue ($)</td>
<td>8000</td>
<td>3600</td>
<td>8500</td>
<td>7700</td>
<td>5800</td>
<td>6000</td>
<td>11000</td>
<td>7500</td>
<td>6600</td>
<td>5600</td>
</tr>
</tbody>
</table>

**THINK**

1 Determine which is the dependent variable and which is the independent variable.

2 Rule up a set of axes. Label the title of the graph. Label the horizontal axis ‘Number of tickets sold’ and the vertical axis ‘Total revenue ($)’.

3 Use an appropriate scale on the horizontal and vertical axes.

4 Plot the points on the scatterplot.

**WRITE/DRAW**

The total revenue depends on the number of tickets being sold, so the number of tickets is the independent variable and the total revenue is the dependent variable.
13.2.4 Correlation

- **Correlation** is a descriptive measure of the relationship between two variables.
- Correlation describes the **strength**, the **direction** and the **form** of the relationship between the two variables.
- The closer the points are to a straight line form, the stronger the correlation between the two variables. The strength is described as **weak, moderate or strong**.
- If the points on a scatterplot have a generally positive slope, the relationship has a positive direction. If the slope is negative, the direction is negative.

![Graphs showing different types of correlation](image)

**WORKED EXAMPLE 2**

State the type of correlation between the variables \( x \) and \( y \), shown on the scatterplot.

**THINK**
Carefully analyse the scatterplot and comment on its form, direction and strength.

**WRITE**
The points on the scatterplot are close together and constantly increasing therefore the relationship is linear.

- The path is directed from the bottom left corner to the top right corner and the value of \( y \) increases as \( x \) increases. Therefore the correlation is positive.
- The points are close together so the correlation can be classified as strong.
- There is a strong, positive, linear relationship between \( x \) and \( y \).

13.2.5 Correlation and causation

- Even a strong correlation does not necessarily mean that the increase or decrease in the level of one variable *causes* an increase or decrease in the level of the other. It is best to avoid statements such as: ‘An increase in rainfall *causes* an increase in the wheat growth.’
- The following guidelines should be closely followed in order to draw a conclusion about the relationship between the two variables based on the scatterplot.
  - If the correlation between \( x \) and \( y \) is weak, we can conclude that *there is little evidence to show* that the larger \( x \) is, the larger (positive correlation) or smaller (negative correlation) \( y \) is.
  - If the correlation between \( x \) and \( y \) is moderate, we can conclude that *there is evidence to show* that the larger \( x \) is, the larger (positive correlation) or smaller (negative correlation) \( y \) is.
  - If the correlation between \( x \) and \( y \) is strong, we can conclude that the larger \( x \) is, the larger (positive correlation) or smaller (negative correlation) \( y \) is.
WORKED EXAMPLE 3

Mary sells business shirts in a department store. She always records the number of different styles of shirt sold during the day. The table below shows her sales over one week.

<table>
<thead>
<tr>
<th>Price ($)</th>
<th>14</th>
<th>18</th>
<th>20</th>
<th>21</th>
<th>24</th>
<th>25</th>
<th>28</th>
<th>30</th>
<th>32</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of shirts sold</td>
<td>21</td>
<td>22</td>
<td>18</td>
<td>19</td>
<td>17</td>
<td>17</td>
<td>15</td>
<td>16</td>
<td>14</td>
<td>11</td>
</tr>
</tbody>
</table>

a Construct a scatterplot of the data.
b State the type of correlation between the two variables and, hence, draw a corresponding conclusion.

THINK

a Draw the scatterplot showing ‘Price ($)’ (independent variable) on the horizontal axis and ‘Number of shirts sold’ (dependent variable) on the vertical axis.

b 1 Carefully analyse the scatterplot and comment on its form, direction and strength.

2 Draw a conclusion corresponding to the analysis of the scatterplot.

WRITE/DRAW

The points on the plot form a path that resembles a straight, narrow band, directed from the top left corner to the bottom right corner. The points are close to forming a straight line. There is a strong, negative, linear correlation between the two variables.

b The price of the shirt appears to affect the number sold; that is, the more expensive the shirt the fewer sold.

learnON RESOURCES — ONLINE ONLY

- Complete this digital doc: SkillSHEET: Substitution into a linear rule (doc-5405)
- Complete this digital doc: SkillSHEET: Solving linear equations that arise when finding x- and y-intercepts (doc-5406)
- Complete this digital doc: SkillSHEET: Transposing linear equations to standard form (doc-5407)
- Complete this digital doc: SkillSHEET: Measuring the rise and the run (doc-5408)
- Complete this digital doc: SkillSHEET: Finding the gradient given two points (doc-5409)
- Complete this digital doc: SkillSHEET: Graphing linear equations using the x- and y-intercept method (doc-5410)
- Complete this digital doc: SkillSHEET: Determining independent and dependent variables (doc-5411)
- Complete this digital doc: SkillSHEET: Determining the type of correlation (doc-5413)
Exercise 13.2 Bivariate data

Individual pathways

■ PRACTISE
Questions: 1a–d, 2–6, 8, 12

■ CONSOLIDATE
Questions: 1d–g, 2–5, 7, 9, 12

■ MASTER
Questions: 1e–h, 2–4, 7–13

Individual pathway interactivity: int-4626

To answer questions online and to receive immediate feedback and sample responses for every question, go to your learnON title at www.jacplus.com.au. Note: Question numbers may vary slightly.

Fluency

1. For each of the following pairs, decide which of the variables is independent and which is dependent.
   a. Number of hours spent studying for a Mathematics test and the score on that test
   b. Daily amount of rainfall (in mm) and daily attendance at the Botanical Gardens
   c. Number of hours per week spent in a gym and the annual number of visits to the doctor
   d. The amount of computer memory taken by an essay and the length of the essay (in words)
   e. The cost of care in a childcare centre and attendance at the childcare centre
   f. The cost of the property (real estate) and the age of the property
   g. The entry requirements for a certain tertiary course and the number of applications for that course
   h. The heart rate of a runner and the running speed

2. The following table shows the cost of a wedding reception at 10 different venues. Represent the data on a scatterplot.

<table>
<thead>
<tr>
<th>No of guests</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
<th>110</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total cost (× $1000)</td>
<td>1.5</td>
<td>1.8</td>
<td>2.4</td>
<td>2.3</td>
<td>2.9</td>
<td>4</td>
<td>4.3</td>
<td>4.5</td>
<td>4.6</td>
<td>4.6</td>
</tr>
</tbody>
</table>

3. State the type of relationship between $x$ and $y$ for each of the following scatterplots.

a. 

b. 

c. 

TOPIC 13 Bivariate data 553
4. Eugene is selling leather bags at the local market. During the day he keeps records of his sales. The table below shows the number of bags sold over one weekend and their corresponding prices (to the nearest dollar).

<table>
<thead>
<tr>
<th>Price ($) of a bag</th>
<th>30</th>
<th>35</th>
<th>40</th>
<th>45</th>
<th>50</th>
<th>55</th>
<th>60</th>
<th>65</th>
<th>70</th>
<th>75</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of bags sold</td>
<td>10</td>
<td>12</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

a. Construct a scatterplot of the data.
b. State the type of correlation between the two variables and, hence, draw a corresponding conclusion.

Understanding
5. The table below shows the number of bedrooms and the price of each of 30 houses.

<table>
<thead>
<tr>
<th>Number of bedrooms</th>
<th>Price (× $1000)</th>
<th>Number of bedrooms</th>
<th>Price (× $1000)</th>
<th>Number of bedrooms</th>
<th>Price (× $1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>180</td>
<td>3</td>
<td>279</td>
<td>3</td>
<td>243</td>
</tr>
<tr>
<td>2</td>
<td>160</td>
<td>2</td>
<td>195</td>
<td>3</td>
<td>198</td>
</tr>
<tr>
<td>3</td>
<td>240</td>
<td>6</td>
<td>408</td>
<td>3</td>
<td>237</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
<td>4</td>
<td>362</td>
<td>2</td>
<td>226</td>
</tr>
<tr>
<td>2</td>
<td>155</td>
<td>2</td>
<td>205</td>
<td>4</td>
<td>359</td>
</tr>
</tbody>
</table>
### Bivariate data

<table>
<thead>
<tr>
<th>Number of bedrooms</th>
<th>Price (× $1000)</th>
<th>Number of bedrooms</th>
<th>Price (× $1000)</th>
<th>Number of bedrooms</th>
<th>Price (× $1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>306</td>
<td>7</td>
<td>420</td>
<td>4</td>
<td>316</td>
</tr>
<tr>
<td>3</td>
<td>297</td>
<td>5</td>
<td>369</td>
<td>2</td>
<td>200</td>
</tr>
<tr>
<td>5</td>
<td>383</td>
<td>1</td>
<td>195</td>
<td>2</td>
<td>158</td>
</tr>
<tr>
<td>2</td>
<td>212</td>
<td>3</td>
<td>265</td>
<td>1</td>
<td>149</td>
</tr>
<tr>
<td>4</td>
<td>349</td>
<td>2</td>
<td>174</td>
<td>3</td>
<td>286</td>
</tr>
</tbody>
</table>

a. Construct a scatterplot of the data.
b. State the type of correlation between the number of bedrooms and the price of the house and, hence, draw a corresponding conclusion.
c. Suggest other factors that could contribute to the price of the house.

6. The table below shows the number of questions solved by each student on a test, and the corresponding total score on that test.

<table>
<thead>
<tr>
<th>Number of questions</th>
<th>2</th>
<th>4</th>
<th>7</th>
<th>10</th>
<th>5</th>
<th>2</th>
<th>6</th>
<th>3</th>
<th>9</th>
<th>4</th>
<th>8</th>
<th>3</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total score (%)</td>
<td>22</td>
<td>39</td>
<td>69</td>
<td>100</td>
<td>56</td>
<td>18</td>
<td>60</td>
<td>36</td>
<td>87</td>
<td>45</td>
<td>84</td>
<td>32</td>
<td>63</td>
</tr>
</tbody>
</table>

a. Construct a scatterplot of the data.
b. What type of correlation does the scatterplot suggest?
c. Give a possible explanation as to why the scatterplot is not perfectly linear.

7. A sample of 25 drivers who had obtained a full licence within the last month was asked to recall the approximate number of driving lessons they had taken (to the nearest 5), and the number of accidents they had while being on P plates. The results are summarised in the table that follows.

<table>
<thead>
<tr>
<th>Number of lessons</th>
<th>Number of accidents</th>
<th>Number of lessons</th>
<th>Number of accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>25</td>
<td>3</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>35</td>
<td>0</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>35</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>2</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>10</td>
<td>4</td>
</tr>
</tbody>
</table>

a. Represent these data on a scatterplot.
b. Specify the relationship suggested by the scatterplot.
c. Suggest some reasons why this scatterplot is not perfectly linear.
Reasoning

8. Each point on the scatterplot below shows the time (in weeks) spent by a person on a healthy diet and the corresponding mass lost (in kg).

![Image of scatterplot]

Study the scatterplot and state whether each of the following statements is true or false.

a. The number of weeks that the person stays on a diet is the independent variable.

b. The y-coordinates of the points represent the time spent by a person on a diet.

c. There is evidence to suggest that the longer the person stays on a diet, the greater the loss in mass.

d. The time spent on a diet is the only factor that contributes to the loss in mass.

e. The correlation between the number of weeks on a diet and the number of kilograms lost is positive.

9. The scatterplot that best represents the relationship between the amount of water consumed daily by a certain household for a number of days in summer and the daily temperature is:

A. Temperature (°C) Water usage (L)
B. Temperature (°C) Water usage (L)
C. Temperature (°C) Water usage (L)
D. Temperature (°C) Water usage (L)

10. The scatterplot at right shows the number of sides and the sum of interior angles for a number of polygons.

Which of the following statements is not true?

A. The correlation between the number of sides and the angle sum of the polygon is perfectly linear.

B. The increase in the number of sides causes the increase in the size of the angle sum.

C. The number of sides depends on the sum of the angles.

D. The correlation between the two variables is positive.
11. **MC** After studying a scatterplot, it was concluded that there was evidence that the greater the level of one variable, the smaller the level of the other variable. The scatterplot must have shown a:
   A. strong, positive correlation
   B. strong, negative correlation
   C. moderate, positive correlation
   D. moderate, negative correlation

**Problem solving**

12. The table below gives the number of kicks and handballs obtained by the top 8 players in an AFL game.

<table>
<thead>
<tr>
<th>Player</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of kicks</td>
<td>20</td>
<td>27</td>
<td>21</td>
<td>19</td>
<td>17</td>
<td>18</td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td>Number of handballs</td>
<td>11</td>
<td>3</td>
<td>11</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>9</td>
<td>7</td>
</tr>
</tbody>
</table>

a. Represent this information on a scatterplot by using the $x$-axis as the number of kicks and the $y$-axis as the number of handballs.
b. Does the scatterplot support the claim that the more kicks a player obtains, the more handballs he gives?

13. The scatterplot shown gives the marks obtained by students in two mathematics tests. Mardi’s score in the tests is represented by M. Which point represents each of the following students?
   a. Mandy, who got the highest mark in both tests
   b. William, who got the top mark in test 1 but not in test 2
   c. Charlotte, who did better on test 1 than Mardi but not as well in test 2
   d. Dario, who did not do as well as Charlotte in both tests
   e. Edward, who got the same mark as Mardi in test 2 but did not do so well in test 1
   f. Cindy, who got the same mark as Mardi for test 1 but did better than her for test 2
   g. Georgina, who was the lowest in test 1
   h. Harrison, who had the greatest discrepancy between his two marks

**Reflection**

How could you determine whether the change in one variable causes the change in another variable?
13.3 Lines of best fit

13.3.1 Lines of best fit

- If the points on a scatterplot appear to lie fairly closely distributed in a linear pattern, a straight line can be drawn through the data. The line can then be used to make predictions about the data.
- A line of best fit is a line on a scatterplot that is positioned so that it is as close as possible to all the data points.
- A line of best fit is used to generalise the relationship between two variables.

13.3.2 Lines of best fit by eye

- A line of best fit can be drawn on a scatterplot by eye. This means that a line is positioned so that there is an equal number of points above and below the line.
- Once a line of best fit has been placed on the scatterplot, an equation for this line can be established, using the coordinates of any two points on the line. These two points do not necessarily have to be actual data points, but if any data points do lie on the line, these are chosen as their values are known accurately.
- The equation for the line passing through the two selected points can then be calculated.
  
  The equation through the two points \((x_1, y_1)\) and \((x_2, y_2)\) is given by:
  
  \[ y = mx + c \text{ where } m = \frac{y_2 - y_1}{x_2 - x_1}. \]

WORKED EXAMPLE 4

The data in the table show the cost of using the internet at a number of different internet cafes based on hours used per month.

<table>
<thead>
<tr>
<th>Hours used per month</th>
<th>10</th>
<th>12</th>
<th>20</th>
<th>18</th>
<th>10</th>
<th>13</th>
<th>15</th>
<th>17</th>
<th>14</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total monthly cost ($)</td>
<td>15</td>
<td>18</td>
<td>30</td>
<td>32</td>
<td>18</td>
<td>20</td>
<td>22</td>
<td>23</td>
<td>22</td>
<td>18</td>
</tr>
</tbody>
</table>

a Construct a scatterplot of the data.
b Draw in the line of best fit by eye.
c Find the equation of the line of best fit in terms of the variables \(n\) (number of hours) and \(C\) (monthly cost).

THINK

a Draw the scatterplot placing the independent variable (hours used per month) on the horizontal axis and the dependent variable (total monthly cost) on the vertical axis.
  Label the axes.

WRITE/DRAW

a
13.3.3 Making predictions

- The line of best fit can be used to predict the value of one variable from that of another. Because of the subjective nature of the line, it should be noted that predictions are not accurate values, but rough estimates. Although this is the case, predictions using this method are considered valuable when no other methods are available.
- If the equation of the line of best fit is known, or can be derived, predictions can be made by substituting known values into the equation of the line of best fit.
WORKED EXAMPLE 5

Use the given scatterplot and line of best fit to predict:

a  the value of \( y \) when \( x = 10 \)

b  the value of \( x \) when \( y = 10 \).

THINK

a  1 Locate 10 on the \( x \)-axis and draw a vertical line until it meets with the line of best fit. From that point, draw a horizontal line to the \( y \)-axis. Read the value of \( y \) indicated by the horizontal line.

b  1 Locate 10 on the \( y \)-axis and draw a horizontal line until it meets with the line of best fit. From that point draw a vertical line to the \( x \)-axis. Read the value of \( x \) indicated by the vertical line.

WRITE/DRAW

When \( x = 10 \), \( y \) is predicted to be 35.

When \( y = 10 \), \( x \) is predicted to be 27.

WORKED EXAMPLE 6

The table below shows the number of boxes of tissues purchased by hayfever sufferers and the number of days affected by hay fever during the blooming season in spring.

<table>
<thead>
<tr>
<th>Number of days affected by hay fever ((d))</th>
<th>3</th>
<th>12</th>
<th>14</th>
<th>7</th>
<th>9</th>
<th>5</th>
<th>6</th>
<th>4</th>
<th>10</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of boxes of tissues purchased ((T))</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

a  Construct a scatterplot of the data and draw a line of best fit.

b  Determine the equation of the line of best fit.

c  Interpret the meaning of the gradient.
d Use the equation of the line of best fit to predict the number of boxes of tissues purchased by people suffering from hay fever over a period of:

i 11 days
ii 15 days.

THINK

a 1 Draw the scatterplot showing the independent variable (number of days affected by hay fever) on the horizontal axis and the dependent variable (total number of boxes of tissues purchased) on the vertical axis.

2 Position the line of best fit on the scatterplot so there is approximately an equal number of data points on either side of the line.

b 1 Select two points on the line that are not too close to each other.

2 Calculate the gradient of the line.

3 Write the rule for the equation of a straight line.

4 Substitute the known values into the equation, say, 3, 1, into the equation to calculate C.

5 Replace x with d (number of days with hay fever) and y with T (total number of boxes of tissues used) as required.
13.3.4 Interpolation and extrapolation

- **Interpolation** is the term used for predicting a value of a variable from within the range of the given data.
- **Extrapolation** occurs when the value of the variable being predicted is outside the range of the given data.

In Worked example 6, the number of days ranged from 3 days to 14 days. Making a prediction for 11 days is an example of interpolation, whereas making a prediction for 15 days is an example of extrapolation.

Predictions involving interpolation are considered to be quite reliable. Those involving extrapolation should be treated with caution, as they rely on the trend of the line remaining unchanged beyond the range of the data.

13.3.5 Reliability of predictions

- When predictions of any type are made, it is useful to know whether they are reliable or not.
- If the line of best fit is used to make predictions, they can be considered to be reliable if each of the following is observed.
  - The number of data values is large.
  - The scatterplot indicates reasonably strong correlation between the variables.
  - The predictions are made using interpolation.
13.3.6 Least squares regression

- Least squares regression involves an exact mathematical approach to fitting a line of best fit to bivariate data that show a strong linear correlation.
- Consider the regression line shown at right. The vertical lines give an indication of how well the line best 'fits' the data. The line of best fit is placed so that these 'error' lines are minimised, by balancing the errors above and below the line.
- Least-squares regression takes these error lines, forms squares, and minimises the sum of the areas of the squares.
- The actual calculation of the equation of a least squares regression line is complicated; however, a calculator can generate the equation with ease.

13.3.7 Correlation coefficient

- Once a relationship between two variables has been established, it is helpful to develop a quantitative value to measure the strength of the relationship. One method is to calculate a correlation coefficient \( r \). This is easily done using a calculator, but a manual method is shown below.
- The formula for the correlation coefficient \( r \) is:

\[
r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}
\]

where \( x \) and \( y \) are the two sets of scores
\( \bar{x} \) and \( \bar{y} \) are the means of those scores
the symbol \( \sum \) representing the sum of the expressions indicated.

- The correlation coefficient is a value in the range \(-1\) to \(+1\). The value of \(-1\) indicates a perfect negative relationship between the two variables, while the value of \(+1\) indicates a perfect positive relationship. For values within this range, a variety of descriptors are used.

WORKED EXAMPLE 7

The percentages from two tests (English and Maths) for a group of 5 students are as shown.

\[\text{Student} \quad \begin{array}{|c|c|c|} \hline \\[-0.5ex] \text{English} & \text{Maths} \\ \hline \\[-0.5ex] 1 & 95 & 85 \\ 2 & 85 & 95 \\ 3 & 80 & 70 \\ 4 & 70 & 65 \\ 5 & 60 & 70 \\ \hline \end{array}\]

\[\text{Strong positive linear association} \quad \text{Moderate positive linear association} \quad \text{Weak positive linear association} \quad \text{No linear association} \quad \text{Weak negative linear association} \quad \text{Moderate negative linear association} \quad \text{Strong negative linear association}\]
THINK

a  1  Draw up a table to calculate all the necessary data: 
\( \bar{x}, \bar{y}, (x - \bar{x}), (y - \bar{y}), 
(x - \bar{x})^2, (y - \bar{y})^2 \)

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>85</td>
<td>17</td>
<td>289</td>
<td>8</td>
<td>64</td>
</tr>
<tr>
<td>85</td>
<td>95</td>
<td>7</td>
<td>49</td>
<td>18</td>
<td>324</td>
</tr>
<tr>
<td>80</td>
<td>70</td>
<td>2</td>
<td>4</td>
<td>-7</td>
<td>49</td>
</tr>
<tr>
<td>70</td>
<td>65</td>
<td>-8</td>
<td>64</td>
<td>-12</td>
<td>144</td>
</tr>
<tr>
<td>60</td>
<td>70</td>
<td>-18</td>
<td>324</td>
<td>-7</td>
<td>49</td>
</tr>
<tr>
<td>Σ</td>
<td>390</td>
<td>385</td>
<td>730</td>
<td>630</td>
<td>470</td>
</tr>
<tr>
<td>Mean</td>
<td>78</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2  Substitute into the formula to calculate the correlation coefficient \( r \).

\[
\begin{align*}
    r &= \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}} \\
    &= \frac{470}{\sqrt{730 \times 630}} \\
    &= 0.69
\end{align*}
\]

b  Describe the relationship.

b  A correlation coefficient of 0.69 indicates that the relationship between English and Maths marks for this group of students is only moderate. This seems to indicate that students who are good at English are not necessarily good at Maths, and vice versa.

RESOURCES — ONLINE ONLY

- Try out this interactivity: Applying lines of best fit (int-2798)
- Try out this interactivity: Extrapolation (int-1154)
- Complete this digital doc: WorkSHEET: Lines of best fit (doc-14599)

Exercise 13.3 Lines of best fit

Individual pathways

- **PRACTISE**
  - Questions: 1–7, 11

- **CONSOLIDATE**
  - Questions: 1–8, 10, 11

- **MASTER**
  - Questions: 1–12

To answer questions online and to receive immediate feedback and sample responses for every question, go to your learnON title at www.jacplus.com.au. Note: Question numbers may vary slightly.
Fluency

1. The data in the table below show the distances travelled by 10 cars and the amount of petrol used for their journeys (to the nearest litre).
   a. Construct a scatterplot of the data.
   b. Draw in the line of best fit.
   c. Determine the equation of the line of best fit in terms of the variables $d$ (distance travelled) and $P$ (petrol used).

<table>
<thead>
<tr>
<th>Distance travelled (km), $d$</th>
<th>52</th>
<th>36</th>
<th>83</th>
<th>12</th>
<th>44</th>
<th>67</th>
<th>74</th>
<th>23</th>
<th>56</th>
<th>95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Petrol used (L), $P$</td>
<td>7</td>
<td>5</td>
<td>9</td>
<td>2</td>
<td>7</td>
<td>9</td>
<td>12</td>
<td>3</td>
<td>8</td>
<td>14</td>
</tr>
</tbody>
</table>

2. A random sample of ten Year 10 students who have part-time jobs was selected. Each student was asked to state the average number of hours they work per week and their average weekly earnings (to the nearest dollar). The results are summarised in the table below.

<table>
<thead>
<tr>
<th>Hours worked, $h$</th>
<th>4</th>
<th>8</th>
<th>15</th>
<th>18</th>
<th>10</th>
<th>5</th>
<th>12</th>
<th>16</th>
<th>14</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly earnings ($), $E$</td>
<td>23</td>
<td>47</td>
<td>93</td>
<td>122</td>
<td>56</td>
<td>33</td>
<td>74</td>
<td>110</td>
<td>78</td>
<td>35</td>
</tr>
</tbody>
</table>

   a. Construct a scatterplot of the data using technology.
   b. Draw in the line of best fit using technology.
   c. Write the equation of the line of best fit, in terms of variables $h$ (hours worked) and $E$ (weekly earnings).
   d. Interpret the meaning of the gradient.

Understanding

3. Use the given scatterplot and line of best fit to predict:

   a. the value of $y$ when $x = 45$
   b. the value of $x$ when $y = 15$

4. Analyse the following graph.

   a. Use the line of best fit to predict the value of $y$ when the value of $x$ is:
      i. 7  
      ii. 22  
      iii. 36.  
   b. Use the line of best fit to predict the value of $x$ when the value of $y$ is:
      i. 120  
      ii. 260  
      iii. 480.  
   c. Determine the equation of the line of best fit, if it is known that it passes through the points $(5, 490)$ and $(40, 80)$.
   d. Use the equation of the line to verify the values obtained from the graph in parts a and b.
5. **WE6** The following table shows the average weekly expenditure on food for households of various sizes.

<table>
<thead>
<tr>
<th>Number of people in a household</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>7</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost of food ($ per week)</td>
<td>70</td>
<td>100</td>
<td>150</td>
<td>165</td>
<td>150</td>
<td>140</td>
<td>120</td>
<td>155</td>
</tr>
<tr>
<td>Number of people in a household</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Cost of food ($ per week)</td>
<td>90</td>
<td>160</td>
<td>160</td>
<td>160</td>
<td>125</td>
<td>75</td>
<td>135</td>
<td></td>
</tr>
</tbody>
</table>

a. Construct a scatterplot of the data and draw in the line of best fit.

b. Determine the equation of the line of best fit. Write it in terms of variables \( n \) (for the number of people in a household) and \( C \) (weekly cost of food).

c. Interpret the meaning of the gradient.

d. Use the equation of the line of best fit to predict the weekly food expenditure for a family of:

i. 8
ii. 9
iii. 10.

6. The following table shows the gestation time and the birth mass of 10 babies.

<table>
<thead>
<tr>
<th>Gestation time (weeks)</th>
<th>31</th>
<th>32</th>
<th>33</th>
<th>34</th>
<th>35</th>
<th>36</th>
<th>37</th>
<th>38</th>
<th>39</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth mass (kg)</td>
<td>1.080</td>
<td>1.470</td>
<td>1.820</td>
<td>2.060</td>
<td>2.230</td>
<td>2.540</td>
<td>2.750</td>
<td>3.110</td>
<td>3.080</td>
<td>3.370</td>
</tr>
</tbody>
</table>

Use technology to answer the following questions.

a. Construct a scatterplot of the data. What type of correlation does the scatterplot suggest?

b. Draw in the line of best fit and determine its equation. Write it in terms of the variables \( t \) (gestation time) and \( M \) (birth mass).

c. What does the value of the gradient represent?

d. Although full term of gestation is considered to be 40 weeks, some pregnancies last longer. Use the equation obtained in part b to predict the birth mass of babies born after 41 and 42 weeks of gestation.

e. Many babies are born prematurely. Using the equation obtained in part b, predict the birth mass of a baby whose gestation time was 30 weeks.

f. If the birth mass of the baby was 2.390 kg, what was his or her gestation time (to the nearest week)?

7. **WE7** The number of hours spent studying, and the marks obtained by a group of students on a test are shown in this table.

<table>
<thead>
<tr>
<th>Hours spent studying</th>
<th>45</th>
<th>30</th>
<th>90</th>
<th>60</th>
<th>105</th>
<th>65</th>
<th>90</th>
<th>80</th>
<th>55</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marks obtained</td>
<td>40</td>
<td>35</td>
<td>75</td>
<td>65</td>
<td>90</td>
<td>50</td>
<td>90</td>
<td>80</td>
<td>45</td>
<td>65</td>
</tr>
</tbody>
</table>
8. As a part of her project, Rachel is growing a crystal. Every day she measures the crystal’s mass using special laboratory scales and records it. The table below shows the results of her experiment.

<table>
<thead>
<tr>
<th>Day number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass (g)</td>
<td>2.5</td>
<td>3.7</td>
<td>4.2</td>
<td>5.0</td>
<td>6.1</td>
<td>8.4</td>
<td>9.9</td>
<td>11.2</td>
<td>11.6</td>
<td>12.8</td>
<td>16.1</td>
<td>17.3</td>
</tr>
</tbody>
</table>

Measurements on days 6, 7, 13 and 14 are missing, since these were 2 consecutive weekends and, hence, Rachel did not have a chance to measure her crystal, which is kept in the school laboratory.

a. Construct the scatterplot of the data and draw in the line of best fit.
b. Determine the equation of the line of best fit. Write the equation, using variables $d$ (day of the experiment) and $M$ (mass of the crystal).
c. Interpret the meaning of the gradient.
d. For her report, Rachel would like to fill in the missing measurements (that is, the mass of the crystal on days 6, 7, 13 and 14). Use the equation of the line of best fit to help Rachel find these measurements. Is this an example of interpolation or extrapolation? Explain your answer.
e. Rachel needed to continue her experiment for 2 more days, but she fell ill and had to miss school. Help Rachel to predict the mass of the crystal on those two days (that is, days 17 and 18), using the equation of the line of best fit. Are these predictions reliable? Explain your answer.

9. MC Consider the figure shown below.

![Graph](image)

The line of best fit on the scatterplot is used to predict the values of $y$ when $x = 15, x = 40$ and $x = 60$.

a. Interpolation would be used to predict the value of $y$ when the value of $x$ is:
   A. 15 and 40  
   B. 15 and 60  
   C. 15 only  
   D. 40 only  
   E. 60 only

b. The prediction of the $y$-value(s) can be considered reliable when:
   A. $x = 15$ and $x = 40$  
   B. $x = 15, x = 40$ and $x = 60$  
   C. $x = 40$  
   D. $x = 40$ and $x = 60$  
   E. $x = 60$

10. MC The scatterplot below is used to predict the value of $y$ when $x = 300$. This prediction is:

   ![Graph](image)

   A. reliable, because it is obtained using interpolation  
   B. not reliable, because it is obtained using extrapolation  
   C. not reliable, because only $x$-values can be predicted with confidence
**Problem solving**

11. For his birthday, Ari was given a small white rabbit. To monitor the rabbit’s growth, Ari decided to measure it once a week. The table below shows the length of the rabbit for various weeks.

<table>
<thead>
<tr>
<th>Week number, $n$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>13</th>
<th>14</th>
<th>17</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length (cm), $l$</td>
<td>20</td>
<td>21</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>30</td>
<td>32</td>
<td>35</td>
<td>36</td>
<td>37</td>
<td>39</td>
</tr>
</tbody>
</table>

a. Construct a scatterplot of the data.
b. Draw a line of best fit and determine its equation.
c. As can be seen from the table, Ari did not measure his rabbit on weeks 5, 7, 9, 11, 12, 15, 16, 18 and 19. Use the equation of the line of best fit to predict the length of the rabbit for those weeks.
d. Were the predictions made in part c an example of interpolation or extrapolation? Explain.
e. Predict the length of the rabbit in the next three weeks (that is, weeks 21–23), using the line of best fit from part c.
f. Are the predictions that have been made in part e reliable? Explain.
g. Check your results using technology.

12. Laurie is training for the long jump, hoping to make the Australian Olympic team. His best jump each year is shown in the table below.

<table>
<thead>
<tr>
<th>Age ($a$)</th>
<th>Best jump ($B$) (metres)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>4.31</td>
</tr>
<tr>
<td>9</td>
<td>4.85</td>
</tr>
<tr>
<td>10</td>
<td>5.29</td>
</tr>
<tr>
<td>11</td>
<td>5.74</td>
</tr>
<tr>
<td>12</td>
<td>6.05</td>
</tr>
<tr>
<td>13</td>
<td>6.21</td>
</tr>
<tr>
<td>14</td>
<td>—</td>
</tr>
<tr>
<td>15</td>
<td>6.88</td>
</tr>
<tr>
<td>16</td>
<td>7.24</td>
</tr>
<tr>
<td>17</td>
<td>7.35</td>
</tr>
<tr>
<td>18</td>
<td>7.57</td>
</tr>
</tbody>
</table>

a. Plot the points generated by the table on a scatterplot.
b. Join the points generated with straight line segments.
c. Draw a line of best fit and determine its equation.
d. The next Olympic Games will occur when Laurie is 20 years old. Use the equation of the line of best fit to estimate Laurie’s best jump that year and whether it will pass the qualifying mark of 8.1 metres.
e. Is a line of best fit a good way to predict future improvement in this situation? What problems are there with using a line of best fit?
f. Olympic Games will also be held when Laurie is 24 years old and 28 years old. Using extrapolation, what length would you predict Laurie could jump at these two ages? Is this realistic?
g. When Laurie was 14, he twisted a knee in training and did not compete for the whole season. In that year, a national junior championship was held. The winner of that championship jumped 6.5 metres. Use your line of best fit to predict whether Laurie would have won that championship.
Reflection
Why is extrapolation considered to be not reliable?

CHALLENGE 13.1
Sam has a mean score of 88 per cent for his first nine tests of the semester. In order to receive an A* his score must be 90 per cent or higher. There is one test remaining for the semester. Is it possible for him to receive an A*? Why or why not?

13.4 Time series

13.4.1 Time series

- Time series data are any data that have time as the independent variable.
- The data are graphed and the graphs are used to determine if a trend is present in the data.
- Identifying a trend can help when making predictions about the future.

13.4.2 Types of trends

- A general upward or downward trend is a graph that overall goes up or down as illustrated in the graph shown below.

- A seasonal pattern displays fluctuations that repeat at the same time each week, month or quarter and usually last less than one year. The graph below illustrates that the peak selling time for houses is in the spring.
• A **cyclical pattern** displays fluctuations that repeat but will usually take longer than a year to repeat. An example of this is shown in the graph below, which depicts software products sold.

![Graph showing cyclical pattern](image)

- **Random patterns** do not show any regular fluctuation. They are usually caused by unpredictable events such as the economic recession illustrated in the graph below.

![Graph showing random pattern](image)

- Trends can work in combinations; for example, you can have a seasonal pattern with an upward trend.

![Graph showing combined trend](image)

**WORKED EXAMPLE 8**

Classify the trend suggested by the time series graph at right as being linear or non-linear, and upward, downward or no trend.

**THINK**
Carefully analyse the given graph and comment on whether the graph resembles a straight line or not and whether the values of y increase or decrease over time.

**WRITE**
The time series graph does not resemble a straight line and overall the level of the variable, y, decreases over time. The time series graph suggests a non-linear downward trend.
13.4.3 Trend lines

- A trend line is a type of line of best fit. Trend lines indicate the general trend of the data.
- Trend lines are useful in forecasting, or making predictions about the future, by extrapolation. Extrapolation can have limited reliability, as predictions are based on the assumption that the current trend will continue into the future.

WORKED EXAMPLE 9

The data below show the average daily mass of a person (to the nearest 100 g), recorded over a period of 4 weeks.
63.6, 63.8, 63.5, 63.7, 63.2, 63.0, 62.8, 63.3, 63.1, 62.7, 62.6, 62.5, 62.9, 63.0, 63.1, 62.9, 62.6, 62.8, 63.0, 62.6, 62.5, 62.1, 61.8, 62.2, 62.0, 61.7, 61.5, 61.2

a Plot these masses as a time series graph.
b Comment on the trend.

THINK

a 1 Draw the points on a scatterplot with day on the horizontal axis and mass on the vertical axis.

2 Join the points with straight line segments to create a time series plot.

b Carefully analyse the given graph and comment on whether the graph resembles a straight line or not and whether the values of y (in this case, mass) increase or decrease over time.

WRITE/DRAW

a The graph resembles a straight line that slopes downwards from left to right (that is, mass decreases with increase in time). Although a person’s mass fluctuates daily, the time series graph suggests a downward trend. That is, overall, the person’s mass has decreased over the 28-day period.
WORKED EXAMPLE 10

The graph at right shows the average cost of renting a one-bedroom flat, as recorded over a 10-year period.

a If appropriate, draw in a line of best fit and comment on the type of the trend.
b Assuming that the current trend will continue, use the line of best fit to predict the cost of rent in 5 years’ time.

THINK

a 1 Analyse the graph and observe what occurs over a period of time. Draw a line of best fit.

2 Comment on the type of trend observed.

b 1 Extend the line of best fit drawn in part a. The last entry corresponds to the 10th year and we need to predict the cost of rent in 5 years’ time; that is, in the 15th year.

2 Locate the 15th year on the time axis and draw a vertical line until it meets with the line of best fit. From the trend line (line of best fit) draw a horizontal line to the cost axis.

3 Read the cost from the vertical axis.

WRITE/DRAW

The graph illustrates that the cost of rent increases steadily over the years. The time series graph indicates an upward linear trend.

b

Cost of rent = $260

Assuming that the cost of rent will continue to increase at the present rate, in 5 years we can expect the cost of rent to reach $260 per week.
13.4.4 Using spreadsheets to determine the correlation

It is possible to draw scatterplots and time series graphs using a variety of digital technologies, including spreadsheets.

**WORKED EXAMPLE 11**

Data were recorded about the number of families who moved from Sydney to Newcastle over the past 10 years.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number moved</td>
<td>97</td>
<td>118</td>
<td>125</td>
<td>106</td>
<td>144</td>
<td>155</td>
<td>162</td>
<td>140</td>
<td>158</td>
<td>170</td>
</tr>
</tbody>
</table>

a Use technology to construct a time series graph, with a line of best fit, that represents the data.
b Describe the trend.
c Measure the correlation.
d Comment on the results.

**THINK**

a Enter the data into a spreadsheet. Select the type of graph — a scatter-plot with the points joined and a trend line fitted.

b Describe the trend.

c To measure the correlation, place the cursor in a cell — see the purple cell. From the menu ribbon, select ‘More formulas’, then ‘Statistics’, then ‘CORREL’. Complete the values for the relevant cells and press Enter. The correlation value will be shown.

d Interpret the results.

**WRITE**

a There appears to be an upward trend over the 10 years.

c The correlation is 0.8761.

d Over the last 10 years, an increasing number of families have decided to make the move from Sydney to Newcastle. The correlation is strong and positive (0.8761), making it possible to predict that this trend is likely to continue.

**RESOURCES — ONLINE ONLY**

- Watch this eLesson: Fluctuations and cycles (eles-0181)
- Complete this digital doc: WorkSHEET: Time series (doc-14600)
Exercise 13.4 Time series

Individual pathways

- **PRACTISE**
  Questions: 1–3, 5–7, 9

- **CONSOLIDATE**
  Questions: 1–7, 9, 10

- **MASTER**
  Questions: 1–11

To answer questions online and to receive immediate feedback and sample responses for every question, go to your learnON title at www.jacplus.com.au. Note: Question numbers may vary slightly.

**Fluency**

1. **WE8** Classify the trend suggested by each time series graph as being linear or non-linear, and upward, downward or stationary in the mean (no trend).

   - **a.** Data
   - **b.** Data
   - **c.** Data
   - **d.** Data
   - **e.** Data
   - **f.** Data
   - **g.** Data
   - **h.** Data

2. **WE9** The data below show the average daily temperatures recorded in June.
   17.6, 17.4, 18.0, 17.2, 17.5, 16.9, 16.3, 17.1, 16.9, 16.2, 16.0, 16.6, 16.1, 15.4, 15.1 15.5, 16.0, 16.0, 15.4, 15.2, 15.0, 15.5, 15.1, 14.8, 15.3, 14.9, 14.6, 14.4, 15.0, 14.2
   - **a.** Plot these temperatures as a time series graph.
   - **b.** Comment on the trend.
Understanding

3. The data below show the quarterly sales (in thousands of dollars) recorded by the owner of a sheepskin product store over a period of 4 years.

<table>
<thead>
<tr>
<th>Quarter</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>57</td>
<td>59</td>
<td>50</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>102</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>125</td>
<td>127</td>
<td>120</td>
<td>124</td>
</tr>
<tr>
<td>4</td>
<td>74</td>
<td>70</td>
<td>72</td>
<td>73</td>
</tr>
</tbody>
</table>

a. Plot the time series.
b. The time series plot displays seasonal fluctuations of period 4 (since there are four quarters). Explain in your own words what this means. Also write one or two possible reasons for the occurrence of these fluctuations.
c. Overall, does the time series plot indicate upward, downward or no trend?

4. The table below shows the total monthly revenue (in thousands of dollars) obtained by the owners of a large reception hall. The revenue comes from rent and catering for various functions over a period of 3 years.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>60</td>
<td>65</td>
<td>40</td>
<td>45</td>
<td>40</td>
<td>50</td>
<td>45</td>
<td>50</td>
<td>55</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>2008</td>
<td>70</td>
<td>65</td>
<td>60</td>
<td>65</td>
<td>60</td>
<td>65</td>
<td>70</td>
<td>75</td>
<td>80</td>
<td>85</td>
<td>85</td>
</tr>
<tr>
<td>2009</td>
<td>80</td>
<td>70</td>
<td>65</td>
<td>70</td>
<td>65</td>
<td>70</td>
<td>75</td>
<td>80</td>
<td>85</td>
<td>90</td>
<td>100</td>
</tr>
</tbody>
</table>

a. Construct a time series plot for these data.
b. Describe the graph (peaks and troughs, long-term trend, any other patterns).
c. Try to give possible reasons for monthly fluctuations.
d. Does the graph show seasonal fluctuations of period 12? Are there any patterns that repeat from year to year?

5. The owner of a motel and caravan park in a small town keeps records of the total number of rooms and total number of camp sites occupied per month. The time series plots based on his records are shown below.

a. Describe each graph, discussing general trend, peaks and troughs and so on. Explain particular features of the graphs and give possible reasons.
b. Compare the two graphs and write a short paragraph commenting on any similarities and differences between them.
6. **WE10** The graph below shows enrolments in the Health and Nutrition course at a local college over a 10-year period.

![Graph showing enrolments](image)

a. If appropriate, draw in a line of best fit and comment on the type of the trend.
b. Assuming that the trend will continue, use the line of best fit to predict the enrolment for the course in 5 years’ time; that is, in the 15th year.

**Reasoning**

7. **WE11** In June a new childcare centre was opened. The number of children attending full time (according to the enrolment at the beginning of each month) during the first year of operation is shown in the table below.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>12</td>
<td>10</td>
<td>11</td>
<td>13</td>
<td>12</td>
<td>14</td>
</tr>
</tbody>
</table>

a. Plot this time series using a digital technology. *(Hint: Let June = 1, July = 2 etc.)*
b. Is the child care business going well? Justify your answer.
c. Draw a line of best fit and find its equation, using coordinates of any two points on the line.
d. Use your equation of the line of best fit to predict the enrolment in the centre during the second year of operation at the beginning of:
   i. August
   ii. January.

What assumptions have you made?

8. The graph below shows the monthly sales of a certain book since its publication. Explain in your own words why linear trend forecasting of the future sales of this book is not appropriate.

![Graph showing monthly sales](image)

9. a. Choose an object or subject that is of interest to you and that can be observed and measured during one day. For example, you might decide to measure your own pulse rate.
b. Prepare a table where you will record your results every hour within the school day. For example, for
the pulse rate the table might look like this.

<table>
<thead>
<tr>
<th>Time</th>
<th>8 am</th>
<th>9 am</th>
<th>10 am</th>
<th>11 am</th>
<th>12 pm</th>
<th>1 pm</th>
<th>2 pm</th>
<th>3 pm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pulse rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

c. Take your measurements at the regular time intervals you have decided on and record them in the

d. Plot the time series obtained as a result of your experiment.

e. Describe the graph and comment on the trend.

f. If appropriate, draw in a line of best fit and predict the values (that is, your pulse rate) for the
next 2–3 hours.

g. Take the actual measurements during the hours you have made predictions for. Compare your
predictions with the actual measurements. Were your predictions accurate? Why or why not?

Problem solving

10. Running a small business is difficult when your business is affected by seasonal patterns. It can make it
difficult to do a budget when you have a fluctuating income. For example, if you owned a cafe at
Mt Buller you would have a lot of business during the ski season but very little business over summer. If
you wish to remain open over the summer, you need to find a way to remove the impact of the seasons from
your sales data to see if it is viable. Investigate a way of removing the effects of the seasons from the data.

11. The table below gives the quarterly sales figures for a second-hand car dealer over a three-year period.

<table>
<thead>
<tr>
<th>Year</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>75</td>
<td>65</td>
<td>92</td>
<td>99</td>
</tr>
<tr>
<td>2013</td>
<td>91</td>
<td>79</td>
<td>115</td>
<td>114</td>
</tr>
<tr>
<td>2014</td>
<td>93</td>
<td>85</td>
<td>136</td>
<td>118</td>
</tr>
</tbody>
</table>

a. Represent this data on a time series plot.

b. Briefly describe how the car sales have altered over the time period.

c. Does it appear that the car dealer can sell more cars in a particular period each year?

Reflection

Why are predictions in the future appropriate for time series even though they involve extrapolation?

CHALLENGE 13.2

Mr MacDonald recorded the test marks for his Year 10 class of 25 students. He calculated the average mark to be 72. Sandra’s mark of 86 was incorrectly marked as 36. What was the correct average mark for the test?
13.5 Review

13.5.1 Review questions

Fluency

1. As preparation for a Mathematics test, a group of 20 students was given a revision sheet containing 60 questions. The table below shows the number of questions from the revision sheet successfully completed by each student and the mark, out of 100, of that student on the test.

<table>
<thead>
<tr>
<th>Number of questions</th>
<th>9</th>
<th>12</th>
<th>37</th>
<th>60</th>
<th>55</th>
<th>40</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>48</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test result</td>
<td>18</td>
<td>21</td>
<td>52</td>
<td>95</td>
<td>100</td>
<td>67</td>
<td>15</td>
<td>50</td>
<td>97</td>
<td>85</td>
<td>89</td>
</tr>
<tr>
<td>Number of questions</td>
<td>50</td>
<td>48</td>
<td>35</td>
<td>29</td>
<td>19</td>
<td>44</td>
<td>20</td>
<td>16</td>
<td>58</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Test result</td>
<td>97</td>
<td>85</td>
<td>62</td>
<td>54</td>
<td>30</td>
<td>70</td>
<td>82</td>
<td>37</td>
<td>28</td>
<td>99</td>
<td>80</td>
</tr>
</tbody>
</table>

a. State which of the variables is dependent and which is independent.
b. Construct a scatterplot of the data.
c. State the type of correlation between the two variables suggested by the scatterplot and draw a corresponding conclusion.
d. Suggest why the relationship is not perfectly linear.

2. a. Use the line of best fit shown on the graph at right to predict the value of \( y \), when the value of \( x \) is:
   i. 10
   ii. 35.

b. Use the line of best fit to predict the value of \( x \), when the value of \( y \) is:
   i. 15
   ii. 30.

c. Find the equation of a line of best fit if it is known that it passes through the points \((5, 5)\) and \((20, 27)\).
d. Use the equation of the line to algebraically verify the values obtained from the graph in parts a and b.

3. The graph shows the number of occupants of a large nursing home over the last 14 years.
a. Comment on the type of trend displayed.
b. Explain why it is appropriate to draw in a line of best fit.
c. Draw a line of best fit and use it to predict the number of occupants in the nursing home in 3 years’ time.
d. What assumption has been made when predicting figures for part c?
4. The table below shows the advertised sale price (’000s dollars) and the land size (m²) for ten vacant blocks of land.

<table>
<thead>
<tr>
<th>Land size (m²)</th>
<th>Sale price (×$1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>632</td>
<td>36</td>
</tr>
<tr>
<td>1560</td>
<td>58</td>
</tr>
<tr>
<td>800</td>
<td>40</td>
</tr>
<tr>
<td>1190</td>
<td>44</td>
</tr>
<tr>
<td>770</td>
<td>41</td>
</tr>
<tr>
<td>1250</td>
<td>52</td>
</tr>
<tr>
<td>1090</td>
<td>43</td>
</tr>
<tr>
<td>1780</td>
<td>75</td>
</tr>
<tr>
<td>1740</td>
<td>72</td>
</tr>
<tr>
<td>920</td>
<td>43</td>
</tr>
</tbody>
</table>

a. Construct a scatterplot and determine the equation of the line of best fit.
b. What does the gradient represent?
c. Using the line of best fit, predict the approximate sale price, to the nearest thousand dollars for a block of land with an area of 1600 m².
d. Using the line of best fit, predict the approximate land size, to the nearest 10 square metres, you could purchase with $500000.

5. The table below shows, for fifteen students, the amount of pocket money they receive and spend at the school canteen in an average week.

<table>
<thead>
<tr>
<th>Pocket money ($)</th>
<th>Canteen spending ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>16</td>
</tr>
<tr>
<td>40</td>
<td>17</td>
</tr>
<tr>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>25</td>
<td>14</td>
</tr>
<tr>
<td>40</td>
<td>16</td>
</tr>
<tr>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>30</td>
<td>16</td>
</tr>
<tr>
<td>30</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pocket money ($)</th>
<th>Canteen spending ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>50</td>
<td>19</td>
</tr>
<tr>
<td>20</td>
<td>14</td>
</tr>
<tr>
<td>35</td>
<td>17</td>
</tr>
<tr>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>13</td>
</tr>
</tbody>
</table>

a. Construct a scatterplot and determine the equation of the line of best fit.
b. What does the gradient represent?
c. Using your line of best fit, predict the amount of money spent at the canteen for a student receiving $45 pocket money a week.
d. Using your line of best fit, predict the amount of money spent at the canteen by a student who receives $100 pocket money each week? Does this seem reasonable? Explain.

6. The table below shows, for 10 ballet students, the number of hours a week spent training and the number of pirouettes in a row they can complete.

<table>
<thead>
<tr>
<th>Training (h)</th>
<th>11</th>
<th>11</th>
<th>2</th>
<th>8</th>
<th>4</th>
<th>16</th>
<th>11</th>
<th>16</th>
<th>5</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of pirouettes</td>
<td>15</td>
<td>13</td>
<td>3</td>
<td>12</td>
<td>7</td>
<td>17</td>
<td>13</td>
<td>16</td>
<td>8</td>
<td>5</td>
</tr>
</tbody>
</table>

a. Construct a scatterplot and determine the equation of the line of best fit.
b. What does the gradient represent?
c. Using your line of best fit, predict the number of pirouettes that could be complete if a student undertakes 14 hours of training.

d. Professional ballet dancers may undertake up to 30 hours of training a week. Using your line best fit, predict the number of pirouettes they should be able to do in a row. Comment on your findings.

7. a. Use the data given below to draw a scatterplot and determine the equation of the line of best fit.

<table>
<thead>
<tr>
<th>Age in years</th>
<th>7</th>
<th>11</th>
<th>8</th>
<th>16</th>
<th>9</th>
<th>8</th>
<th>14</th>
<th>19</th>
<th>17</th>
<th>10</th>
<th>20</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours of television watched in a week</td>
<td>20</td>
<td>19</td>
<td>25</td>
<td>55</td>
<td>46</td>
<td>50</td>
<td>53</td>
<td>67</td>
<td>59</td>
<td>25</td>
<td>70</td>
<td>58</td>
</tr>
</tbody>
</table>

b. Use the line of best fit to predict the value of the number of hours of television watched by a person aged 15.

c. Determine the age when the number of hours of television watched is 60.

Problem solving

8. Describe the trends present in the following time series data that shows the mean monthly daily hours of sunshine in Melbourne from January to December.

<table>
<thead>
<tr>
<th>Month</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily hours of sunshine</td>
<td>8.7</td>
<td>8.0</td>
<td>7.5</td>
<td>6.4</td>
<td>4.8</td>
<td>4.0</td>
<td>4.5</td>
<td>5.5</td>
<td>6.3</td>
<td>7.3</td>
<td>7.5</td>
<td>8.3</td>
</tr>
</tbody>
</table>

9. The existence of the following situations is often considered an obstacle to making estimates from data.

a. Outlier

b. Extrapolation

c. Small range of data

d. Small number of data points

Explain why each of these situations is considered an obstacle to making estimates of data and how each might be overcome.

10. The table below shows the heights of 10 students and the distances along the ground between their feet as they attempt to do the splits.

<table>
<thead>
<tr>
<th>Height (cm)</th>
<th>Distance stretched (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>134.5</td>
<td>150</td>
</tr>
<tr>
<td>156</td>
<td>160</td>
</tr>
<tr>
<td>133.5</td>
<td>147</td>
</tr>
<tr>
<td>145</td>
<td>160</td>
</tr>
<tr>
<td>160</td>
<td>162</td>
</tr>
<tr>
<td>135</td>
<td>149</td>
</tr>
<tr>
<td>163</td>
<td>163</td>
</tr>
<tr>
<td>138</td>
<td>149</td>
</tr>
<tr>
<td>152</td>
<td>158</td>
</tr>
<tr>
<td>159</td>
<td>160</td>
</tr>
</tbody>
</table>

Using the data, estimate the distance a person 1.8 m tall can achieve when attempting the splits. Write a detailed analysis of your result. Include:

- an explanation of the method(s) used
- any plots or formula generated
- comments on validity of the estimate
- any ways the validity of the estimate could be improved.
Language

It is important to learn and be able to use correct mathematical language in order to communicate effectively. Create a summary of the topic using the key terms below. You can present your summary in writing or using a concept map, a poster or technology.

- **bivariate**
- **continuous**
- **denominator**
- **dependent variable**
- **direction**
- **discrete**
- **form**
- **independent**
- **independent variable**
- **line of best fit**
- **moderate**
- **predictions**
- **relationship**
- **scatterplot**
- **strength**
- **strong**
- **time series**
- **trend**
- **univariate**
- **variable**
- **weak**

Investigation | Rich Task

Collecting, recording and analysing data over time

A time series is a sequence of measurements taken at regular intervals (daily, weekly, monthly and so on) over a certain period of time. Time series are best represented using time-series plots, which are line graphs with the time plotted on the horizontal axis. Examples of time series include daily temperature, monthly unemployment rates and daily share prices.

When data are recorded on a regular basis, the value of the variable may go up and down in what seems to be an erratic pattern. These are called fluctuations. However, over a long period of time,
the time series usually suggests a certain trend. These trends can be classified as being linear or non-linear, and upward, downward or stationary (no trend).

**Time series are often used for forecasting, that is, making predictions about the future. The predictions made with the help of time series are always based on the assumption that the observed trend will continue in the future.**

1. Choose a subject that is of interest to you and that can be observed and measured during one day or over the period of a week or more. (Suitable subjects are shown in the list below.)
2. Prepare a table for recording your results. Select appropriate regular time intervals. An example is shown below.

<table>
<thead>
<tr>
<th>Time</th>
<th>8 am</th>
<th>9 am</th>
<th>10 am</th>
<th>11 am</th>
<th>12 pm</th>
<th>1 pm</th>
<th>2 pm</th>
<th>3 pm</th>
<th>4 pm</th>
<th>5 pm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pulse rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Take your measurements at the selected time intervals and record them in the table.
4. Use your data to plot the time series. You can use software such as Excel or draw the scatterplot by hand.
5. Describe the graph and comment on its trend.
6. If appropriate, draw a line of best fit and predict the next few data values.
7. Take the actual measurements during the hours you have made predictions for. Compare the predictions with the actual measurements. Were your predictions good? Give reasons.

Here are some suitable subjects for data observation and recording:
- minimum and maximum temperatures each day for 2 weeks
  (use the TV news or newspaper as resources)
- the value of a stock on the share market (e.g. Telstra, Wesfarmers and Rio Tinto)
- your pulse over 12 hours (ask your teacher how to do this or check on the internet)
- the value of sales each day at the school canteen
- the number of students absent each day
- the position of a song in the Top 40 over a number of weeks
- petrol prices each day for 2 weeks
- other measurements (check with your teacher)
- world population statistics over time.
### Answers

#### Topic 13 Bivariate data

#### Exercise 13.2 Bivariate data

1. **Independent** | **Dependent**
   - a. Number of hours | Test results
   - b. Rainfall | Attendance
   - c. Hours in gym | Visits to the doctor
   - d. Lengths of essay | Memory taken
   - e. Cost of care | Attendance
   - f. Age of property | Cost of property
   - g. Number of applicants | Cut-off OP score
   - h. Running speed | Heart rate

2. 

3. **a.** Perfectly linear, positive  
   **b.** No correlation  
   **c.** Non-linear, negative, moderate  
   **d.** Strong, positive, linear  
   **e.** No correlation  
   **f.** Non-linear, positive, strong  
   **g.** Strong, negative, negative  
   **h.** Non-linear, moderate, negative  
   **i.** Weak, negative, linear  
   **j.** Non-linear, moderate, positive  
   **k.** Positive, moderate, linear  
   **l.** Non-linear, strong, negative  
   **m.** Strong, negative, linear  
   **n.** Weak, positive, linear  
   **o.** Non-linear, moderate, positive

4. **a.**

   **b.** Negative, linear, moderate. The price of the bag appeared to affect the numbers sold; that is, the more expensive the bag, the fewer sold.

5. **a.**
b. Moderate positive linear correlation. There is evidence to show that the larger the number of bedrooms, the higher the price of the house.
c. Various answers; location, age, number of people interested in the house, and so on.

6. a.

\[ \text{Moderate positive linear correlation. There is evidence to show that the larger the number of bedrooms, the higher the price of the house.} \]

\[ \text{Various answers; location, age, number of people interested in the house, and so on.} \]

b. Strong, positive, linear correlation
c. Various answers — some students are of different ability levels and they may have attempted the questions but had incorrect answers.

7. a.

\[ \text{Weak, negative, linear relation} \]

b. Weak, negative, linear relation
c. Various answers, such as some drivers are better than others, live in lower traffic areas, traffic conditions etc.

8. a. T
d. F
e. T

9. B

10. C

11. D

12. a.

b. This scatterplot does not support the claim.

c. Using \((23, 3)\) and \((56, 8)\), the equation is \(P = \frac{5}{33}d - \frac{16}{33}\).
2. a, b.

\[ E = 6.75h - 7 \]

c. Using \((8, 47)\) and \((12, 74)\), the equation is \(E = 6.75h - 7\).

d. On average, students were paid $6.75 per hour.

3. a. 38  b. 18

4. a. i. 460  ii. 290  iii. 130
   b. i. 39  ii. 24  iii. 6
   c. \(y = -11.71x + 548.57\)
   d. \(y\)-values:
      i. 466.60  ii. 290.95  iii. 127.01
      \(x\)-values:
      i. 36.60  ii. 24.64  iii. 5.86

5. a.

\[ C = 18.75n + 56.25 \]

b. Using \((1, 75)\) and \((5, 150)\), the equation is \(C = 18.75n + 56.25\).

c. On average, weekly cost of food increases by $18.75 for every extra person.

d. i. $206.25  ii. $225.00  iii. $243.75

6. a.

Positive, strong, linear correlation

b. Using \((32, 1.470)\) and \((35, 2.230)\), \(M = 0.25t - 6.5\).

c. With every week of gestation the mass of the baby increases by approximately 250 g.

d. 3.75 kg; 4 kg  e. Approximately 1 kg  f. Between 35 and 36 weeks
7. a. \( r = 0.9 \)
   b. There is a strong positive relationship between the number of hours spent studying and the marks obtained. This seems to indicate that greater dedication to studying will produce better results.

8. a. 
   \[
   \begin{array}{c|cccccccccc}
   \end{array}
   \]
   b. Using \((2, 3.7)\) and \((10, 11.2)\), \( M = 0.973d + 1.285 \).
   c. Each day Rachel’s crystal gains 0.973 g in mass.
   d. 7.123 g; 8.096 g; 13.934 g; 14.907 g; interpolation (within the given range of 1–16)
   e. 17.826 g; 18.799 g; predictions are not reliable, since they were obtained using extrapolation.

9. a. D
   b. C

10. E

11. a. 
   \[
   \begin{array}{c|cccccccccc}
   \text{Week} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 \\
   \end{array}
   \]
   b. Using \((2, 21)\) and \((17, 37)\), \( L = 1.07n + 18.9 \).
   c. 24.25 cm; 26.39 cm; 28.53 cm; 30.67 cm; 31.74 cm; 34.95 cm; 36.02 cm; 38.16 cm; 39.23 cm
   d. Interpolation (within the given range of 1–20)
   e. 41.37 cm; 42.44 cm; 43.51 cm
   f. Not reliable, because extrapolation has been used.

12. a. 
   \[
   \begin{array}{c|cccccccccc}
   \text{Age} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 \\
   \text{Best jump (metres)} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 & 22 \\
   \end{array}
   \]
   b. 
   \[
   \begin{array}{c|cccccccccc}
   \text{Age} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 \\
   \text{Best jump (metres)} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & 19 & 20 & 21 & 22 \\
   \end{array}
   \]
c. Yes. Using points (9, 4.85) and (16, 7.24), \( B = 0.34a + 1.8 \); estimated best jump = 8.6 m.

e. No, trends work well over the short term but long term are affected by other variables.

f. 24 years old: 9.97 m; 28 years old: 11.33 m. It is unrealistic to expect his jumping distance to increase indefinitely.

g. No, trends work well over the short term but long term are affected by other variables.

**Challenge 13.1**

No. He would have to get 108%, which would be impossible on a test.

**Exercise 13.4 Time series**

1. a. Linear, downward  b. Non-linear, upward  c. Non-linear, stationary in the mean  
   d. Linear, upward  e. Non-linear, downward  f. Non-linear, stationary in the mean  
   g. Non-linear, stationary in the mean  h. Linear, upward

2. a. May temperature  

b. Linear downward trend

3. a.  

b. Sheepskin products more popular in the third quarter (presumably winter) — discount sales, increase in sales, and so on.  
   c. No trend

4. a.  

b. Sheepskin products more popular in the third quarter (presumably winter) — discount sales, increase in sales, and so on.  
   c. No trend
b. General upward trend with peaks around December and troughs around April.
c. Peaks around Christmas where people have lots of parties, troughs around April where weather gets colder and people less inclined to go out.
d. Yes. Peaks in December, troughs in April.
5. a. Peaks around Christmas holidays and a minor peak at Easter. No camping in colder months.
   b. Check with your teacher.
6. a.  
   ![Graph showing enrolment trend over the years]
   
   Upward linear
   
   b. In 15th year the expected amount = 122
7. a.  
   ![Graph showing number of children per month]
   
   b. Yes, the graph shows an upward trend.
   
   c. \( y = \frac{4}{7}x + \frac{45}{7} \)
   
   d. i. 15
   
   ii. 18 (The assumption made was that business will continue on a linear upward trend.)
8. The trend is non-linear, therefore unable to forecast future sales.
9. Answers will vary.
10. Check with your teacher. Some methods could involve ‘smoothing’ and ‘seasonal adjustments’.
11. a.  
   ![Graph showing cars sold per quarter]
   
   b. Secondhand car sales per quarter have shown a general upward trend but with some major fluctuations.
   
   c. More cars are sold in the third and fourth quarters compared to the first and second quarters.

**Challenge 13.2**
74
13.5 Review

1. a. Number of questions — independent; test result — dependent

b. [Graph showing scatter plot]

c. Strong, positive, linear correlation; the larger the number of completed revision questions, the higher the mark on the test.
d. Different abilities of the students

2. a. i. 12.5
   ii. 49
b. i. 12
   ii. 22.5
c. \( y = \frac{22}{15}x - \frac{2}{3} \)
d. i. 12.33
   ii. 49
and
   i. 11.82
   ii. 22.05

3. a. Linear downwards
   b. The trend is linear.
   c. About 65 occupants
   d. Assumes that the current trend will continue.

4. a. \( P = 31.82a + 13070.4 \), where \( P \) is the sale price and \( a \) is the land area.
b. The price of land is approximately $31.82 per square metre.
c. $64,000
   d. 1160 m²

5. a. \( C = 0.15p + 11.09 \), where \( C \) is the money spent at the canteen and \( p \) represents the pocket money received.
b. Students spend 15 cents at the canteen per dollar received for pocket money.
c. $18
   d. $26. This involves extrapolation, which is considered unreliable. It does not seem reasonable that, if a student receives more money, they will eat more or have to purchase more than any other student.

6. a. \( P = 0.91t + 2.95 \), where \( P \) is the number of pirouettes and \( t \) is the number of hours of training.
b. Ballet students can do approximately 0.91 pirouettes for each hour of training.
c. Approximately 15 pirouettes
   d. Approximately 30 pirouettes. This estimate is based on extrapolation, which is considered unreliable. To model this data linearly as the number of hours of training becomes large is unrealistic.

7. a. Hours of TV watched = 3.31 × age + 3.05
   b. Approximately 53 hours
   c. Approximately 17 years old

8. Overall the data appears to be following a seasonal trend, with peaks at either end of the year and a trough in the middle.

9. a. Outliers can unfairly skew data and as such dramatically alter the line of best fit. Identify and remove any outliers from the data before determining the line of best fit.
b. Extrapolation involves making estimates outside the data range and this is considered unreliable. When extrapolation is required, consider the data and the likelihood that the data would remain linear if extended. When giving results, make comment on the validity of the estimation.
c. A small range may not give a fair indication if a data set shows a strong linear correlation. Try to increase the range of the data set by taking more measurements or undertaking more research.

d. A small number of data points may not be able to establish with confidence the existence of a strong linear correlation. Try to increase the number of data points by taking more measurements or undertaking more research.

10. About 170 cm; data was first plotted as a scatter plot. (145, 160) was identified as an outlier and removed from the data set. A line of best fit was then fitted to the remaining data and its equation determined as $d = 0.5h + 80$, where $d$ is the distance stretched and $h$ is the height. Substitution was used to obtain the estimate.

The estimation requires extrapolation and cannot be considered reliable. The presence of the outlier may indicate variation in flexibility rather than a strong linear correlation between the data. Estimate is based on a small set of data. More data should be collected in order to determine the suitability of least squares regression.

Investigation — Rich task
Answers will vary; teacher to check.