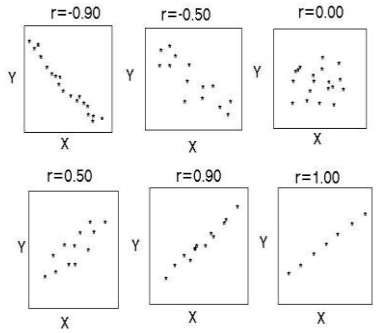
**Week** 3/4
**Term** 1
**2024**

HAWKER COLLEGE
Engage | Inspire | Achieve

**MA3**
Bivariate data analysis

## Goals



**This fortnight we are going to:**

- calculate and interpret the correlation coefficient (p and *r*) to quantify the strength of a linear association
- use the coefficient of determination to assess the strength of a linear association in terms of the explained variation
- recognise that an observation association between two variables does not necessarily mean that there is a causal relationship between them
- model a linear relationship by fitting a least-squares line to the data
- use a residual plot to assess the appropriateness of fitting a linear model to the data
- interpret the intercept and slope of the fitted line
- use the equation of a fitted line to make predications
- distinguish between interpolation and extrapolation when using the fitted line to make predictions, recognising the potential dangers of extrapolation
- write up the results of the above analysis in a systematic and concise manner

## Theoretical Components

**Resources:**
*PDF file:* Week 3/4 Notes & Exercises

Correlation vs causation
https://www.youtube.com/watch?v=VMUQSMFGBDo
Coefficient of determination
https://www.youtube.com/watch?v=qQMAjsOihYc
Line of best fit
https://www.youtube.com/watch?v=DmGLQkUm-4g
Interpolation and extrapolation
https://www.youtube.com/watch?v=bEANDlJkqcU

**Order**

1. Read through the notes and examples
2. Work through the exercises
3. Complete the investigation at the end of the booklet.
4. Complete the reflection at the end of the booklet
5. Come and see your teacher and make sure you are up to date.

## Practical Components

Work through the exercises and show the completed tasks to your teacher.

Be sure to ask for help as you need for the successful completion of all tasks.

**Remember to regularly check Google Classroom for messages.**

**Knowledge Checklist**

- q-correlation coefficient
- Pearson's correlation coefficient
- Correlation and causation
- Coefficient of determination
- Line of best fit
- Least square regression
- Interpretation, interpolation and extrapolation

## Investigation

Complete the task at the end of the booklet and submit your work for checking. 😊

## Quiz/forum/ other

**Remember to check** hawkermaths.com **for each week's learning brief.** Make sure you have joined Google Classroom. If you have not, see your teacher.

# MATHEMATICAL APPLICATIONS 3

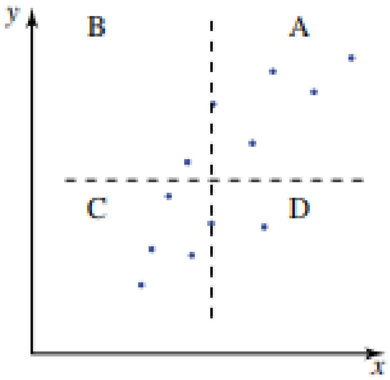
## WEEK 3/4 NOTES & EXERCISES

## Q-CORRELATION COEFFICIENT

Last week, we looked at scatter plots and the idea of correlation. It is useful to have a quantitative measurement of the strength of the relationship shown between a pair of variables. This measurement is known as a correlation coefficient.

There are several different methods used for obtaining a correlation coefficient. One relatively simple method is called the q-correlation coefficient. It can be found by using the steps detailed below.

**Step 1** - Draw a vertical line on the scatter plot of the data to indicate the position of the median of the x values. There should be an equal number of points on each side of this line. Then draw a horizontal line to indicate the position of the median of the y values. Note that these lines may or may not pass through one or more of the data points depending on whether there is an even or odd number of points.



**Step 2** - The scatter plot is now divided into four quadrants that are labelled A, B, C and D as in the diagram above. The order of the labelling is important.

**Step 3** - Count the number of data points that lie in each of the quadrants. Points which lie on either of the median lines are omitted.

**Step 4** - The correlation coefficient is calculated using the formula:

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

where:

- a is the number of data points in A
- b is the number of data points in B
- c is the number of data points in C
- d is the number of data points in D.

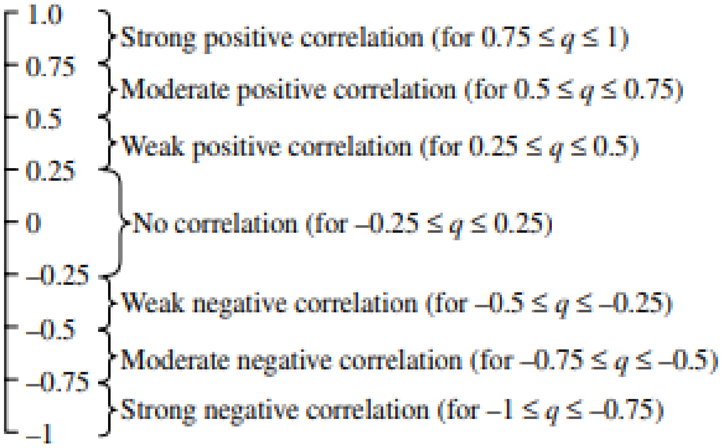
For these data: a = 5, b = 1, c = 5, d = 1

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

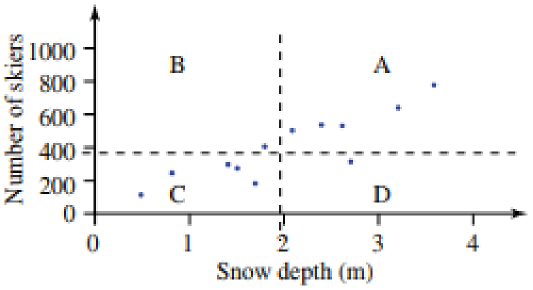
$$q = \frac{(5+5)-(1+1)}{5+1+5+1}$$

$$q = 0.67 \text{ (to 2 decimal places)}$$

The correlation coefficient may be interpreted by using the scale below.


1.0
0.75
0.5
0.25
0
–0.25
–0.5
–0.75
–1
Strong positive correlation (for 0.75 ≤ q ≤ 1)
Moderate positive correlation (for 0.5 ≤ q ≤ 0.75)
Weak positive correlation (for 0.25 ≤ q ≤ 0.5)
No correlation (for –0.25 ≤ q ≤ 0.25)
Weak negative correlation (for –0.5 ≤ q ≤ –0.25)
Moderate negative correlation (for –0.75 ≤ q ≤ –0.5)
Strong negative correlation (for –1 ≤ q ≤ –0.75)

The data used in this calculation is from the data about the ski resort on the first page of the Week 1/2 Notes.



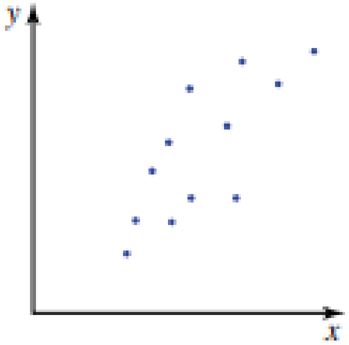
A q-correlation coefficient of 0.67 indicates a moderate positive correlation. In this case, we would conclude: There is evidence to show that the greater the depth of snow, the greater the number of skiers. Notice that the scale for correlation coefficients runs from -1 to 1. Every correlation coefficient must lie within this range. Any answers outside the range would indicate an error in workings.

**Example**

For the graph below:

a) calculate the q-correlation coefficient

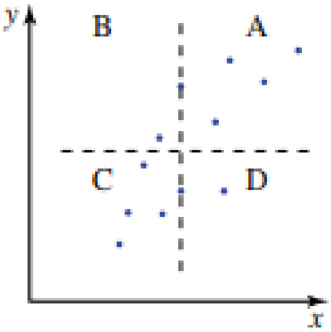
b) state what type of correlation is involved.



**Solution**

a. First find the position of the median lines and label the quadrants. Second count the number of data points in each quadrant, ignoring those that lie on a line. Lastly use the formula to calculate the q-correlation coefficient.

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



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

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

$$q = 0.6$$

b. There is a moderate positive correlation.

## EXERCISE 1

**1.** What type of correlation would be represented by scatter plots which have the following correlation coefficients?

a. 0.4

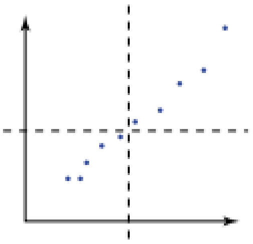
b. 0.8

c. 0.7

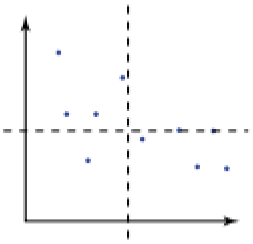
**2.** For each of the following graphs:
i) calculate the q-correlation coefficient
ii) state what type of correlation is involved.

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

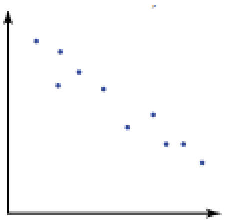
a.



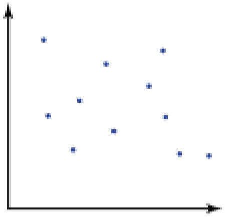
b.



c.



d.



## PEARSON'S CORRELATION COEFFICIENT (*R*)

There are limitations to the q-correlation coefficient. The q-correlation coefficient is useful because it is easy to apply to a scatter plot, but it is not the most sophisticated or reliable way of finding a measure of linear association. A much more rigorous coefficient *r*, which is found by the formula:

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)$$

where:

- $n$ is the number of pairs of data in the set
- $s_x$ is the standard deviation of the x values
- $s_y$ is the standard deviation of the y values
- $\bar{x}$ is the mean of the x values
- $\bar{y}$ is the mean of the y values

This can be complicated and tedious to calculate. Fortunately, we can use online calculators to do it for us. There are two important limitations on the use of $r$. The calculation of $r$ is applicable to sets of bivariate data which are known to be linear in form and which do not have outliers.

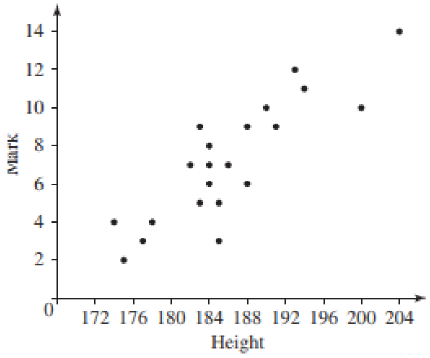
**Example**

The heights (in centimetres) of 21 football players were recorded against the number of marks they took in a game of football. The data are shown in the following table. Draw a scatter plot to show the data.

| Height (cm) | Number of marks taken |
|---|---|
| 184 | 6 |
| 194 | 11 |
| 185 | 3 |
| 175 | 2 |
| 186 | 7 |
| 183 | 5 |
| 174 | 4 |
| 200 | 10 |
| 188 | 9 |
| 184 | 7 |
| 188 | 6 |

| Height (cm) | Number of marks taken |
|---|---|
| 182 | 7 |
| 185 | 5 |
| 183 | 9 |
| 191 | 9 |
| 177 | 3 |
| 184 | 8 |
| 178 | 4 |
| 190 | 10 |
| 193 | 12 |
| 204 | 14 |
| | |

**Solution**



Height is the independent variable, so it is plotted on the x axis.

Using the calculator on https://www.socscistatistics.com/tests/pearson/default2.aspx, we get a value of *r* to be 0.86. This indicates there is a strong positive linear association between the height of a player and the number of marks they take in a game. That is, the taller the player, the more marks we might expect them to take.

## Correlation and causation

Just because variables are highly correlated, we cannot infer a causal relationship. Two variables *X* and *Y* may be highly correlated because:

1. *X* causes *Y*
2. *Y* causes *X*
3. Some third influence causes both *X* and *Y*

We cannot decide which of these applies simply based on a correlation coefficient.

**Solution (for the above example)**

From the example above about football player’s height and number of marks taken, we concluded *r* = 0.86. While we can say that there is a strong association between the height of a football player and the number of marks he takes, we cannot state that the height of a football player causes him to take a lot of marks. Being tall might assist in taking marks, but there will also be many other factors which come into play; for example, skill level, accuracy of passes from team mates, abilities of the opposing team, and so on.

A famous example of correlation and causation concerns the high positive correlation between the number of human births and the stork population of European towns.

Storks generally nested in chimneys, so the larger towns with more chimneys had a larger stork population. Since larger towns also have larger populations of people, these towns also produced more babies. The third variable, the size of the town, was responsible for the high correlation, not the theory storks bring babies. We must be careful attributing causality is a logical problem not a statistical one.

### The coefficient of determination ($r^2$)

The coefficient of determination is given by $r^2$. It is found simply by squaring $r$. its value also ranges from 0 to 1. The coefficient of determination tells us the proportion of variation in one variable that can be explained by the variation in the other variable.

### Solution (for the above example)

From the example above about football player's height and number of marks taken, $r = 0.86$ and $r^2 = 0.74$. Thus 74% of the variation in the number of marks can be explained by the variation in the height of the football players.

## EXERCISE 2

**1.** The yearly salary (x$1000) and the number of votes polled in the Brownlow medal count are given below for 10 footballers.

| **Yearly salary (x$1000)** | 180 | 200 | 160 | 250 | 190 | 210 | 170 | 150 | 140 | 180 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Number of votes** | 24 | 15 | 33 | 10 | 16 | 23 | 14 | 21 | 31 | 28 |

a. Construct a scatter plot for the data.

b. Calculate *r* using the link above and comment on the correlation between salary and number of votes.

c. Calculate the coefficient of determination for the data and interpret its value.

Thus ______ % of the variation in the number of _____________ can be explained by the variation in the _________________________ of the football players.

**2.** A set of data, obtained from 40 smokers, gives the number of cigarettes smoked per day and the number of visits per year to the doctor. The Pearson's correlation coefficient for these data was found to be 0.87. Calculate the coefficient of determination for the data and interpret its value.

**3.** The data below shows the number of people in 9 households against weekly grocery costs.

| **Number of people in households** | 2 | 5 | 6 | 3 | 4 | 5 | 2 | 6 | 3 |
|---|---|---|---|---|---|---|---|---|---|
| **Weekly grocery costs ($)** | 60 | 180 | 210 | 120 | 150 | 160 | 65 | 200 | 90 |

a. Construct a scatter plot for the data.

b. Calculate *r*.

c. Calculate the coefficient of determination and interpret its value.

**4.** A set of data was obtained from a large group of women with children under 5 years of age. They were asked the number of hours they worked per week and the amount of money they spent on childcare. The results were recorded, and the value of Pearson's correlation coefficient was found to be 0.92. Classify each of the following statements as True or False.

a. There is a positive correlation between the number of working hours and the amount of money spent on childcare.

b. The correlation between the number of working hours and the amount of money spent on childcare can be classified as strong.

c. As the number of working hours increases, the amount spent on childcare increases as well.

d. The increase in the number of hours worked causes the increase in the amount of money spent on childcare.

e. The coefficient of determination is about 0.85.

f. The number of working hours is the major factor in predicting the amount of money spent on childcare.

g. About 85% of the variation in the number of hours worked can be explained by the variation in the amount of money spent on childcare.

h. Apart from number of hours worked, there could be other factors affecting the amount of money spent on childcare.

**5.** The data below shoes the annual advertising budgets (x$1000) and the yearly profit increases (%) of 8 companies. Calculate the coefficient of determination and interpret the results.

| **Annual advertising budget (x$1000)** | 11 | 14 | 15 | 17 | 20 | 25 | 25 | 27 |
|---|---|---|---|---|---|---|---|---|
| **Yearly profit increase (%)** | 2.2 | 2.2 | 3.2 | 4.6 | 5.7 | 6.9 | 7.9 | 9.3 |

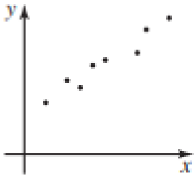
## FITTING STRAIGHT LINES TO BIVARIATE DATA

Fitting a line of 'best fit' to bivariate data (scatter plots) enables us to analyse data and possibly make predictions. Regression analysis is concerned with finding these straight lines.

In our previous work, when we displayed bivariate data as a scatterplot, the independent variable was placed on the horizontal axis and the dependent variable was placed on the vertical axis. When the relationship between two variables (x and y) is described in equation form, such as $y = mx + c$, the subject, y, is the dependent variable and x is the independent variable.

### Fitting a straight line by eye

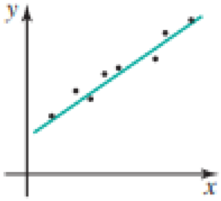
Consider the set of bivariate data points below. In this case, the x-values could be heights of married women, why y-values could be the heights of their partners. We wish to determine a linear relationship between these two random variables.



Of course, there is no single straight line which would go through all the points, so we can only estimate such a line. Furthermore, the more closely the points appear to be on or near a straight line, the more confident we are that such a linear relationship may exist and the more accurate our fitted line should be.

Consider the estimate, drawn 'by eye' in the figure below. It is clear most of the points are on or very close to this straight line. This line was easily drawn since the points are very much part of an apparent linear relationship. However, note that some points are below the line and some are above it. Furthermore, if *x* is the height of married women and *y* is the height of their partners, it seems that their partners are generally taller than them.

Regression analysis is concerned with finding these straight lines using various methods so the number of points above and below the lines are 'balanced'.

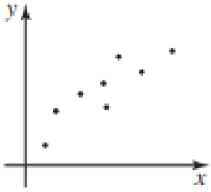


### Method of fitting lines by eye

There should be an equal number of points above and below the line. For example, if there are 12 points in the data set, 6 should be above the line and 6 below it. This may appear logical or even obvious but fitting by eye involves a considerable margin of error.

**Example**

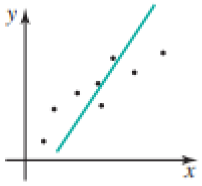
Fit a straight line to the data in the figure using the equal-number-of-points method.



**Solution**

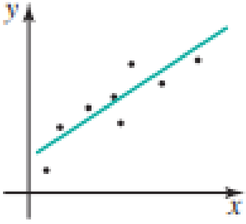
The number of points (n) is 8. We need to fit a line where 4 points are above and below the line. Using a clear plastic ruler, try to fit the best line.

Attempt 1:



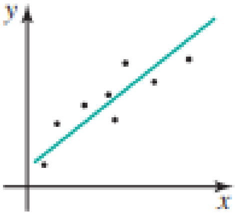
The first attempt has 3 points below the line where there should be 4. We need to make refinements.

Attempt 2:



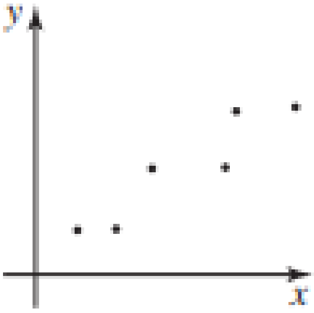
The second attempt is an improvement, but the line is too close to the points above the line. We can improve on the position of the line until there is a better ‘balance’ between the upper and lower points is achieved.

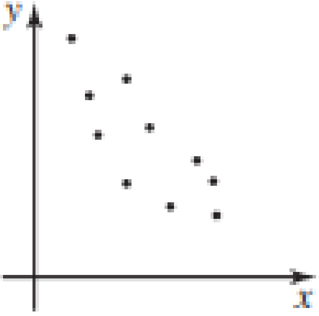
Final attempt:



## EXERCISE 3

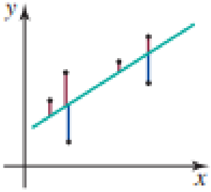
Fit a straight line to the data in the scatter plots using the equal-number-of-points method.





## FITTING A STRAIGHT LINE – LEAST-SQUARES REGRESSION

This is used when the data shows a linear relationship and there are no obvious outliers (or these are discarded). To understand the underlying theory behind least-squares, consider the regression line shown below.



We want to minimise the total of the vertical lines, or ‘errors’ in some way, such as balancing the errors above and below the line. This is reasonable, but for mathematical reasons it is preferable to minimise the sum of the squares of each of these errors. This is the essential mathematics of least-squares regression.

The formula for a straight line when using the *x* and *y* axis is:

$$\boldsymbol{y = mx + c}$$

where:

- m = gradient of the line $\left(\frac{rise}{run}\right)$
- c = y intercept (i.e. the value of *y* when *x* = 0

Fortunately, we can use an online calculator to find the equation of the line. This calculator is found at

https://www.socscistatistics.com/tests/regression/default.aspx

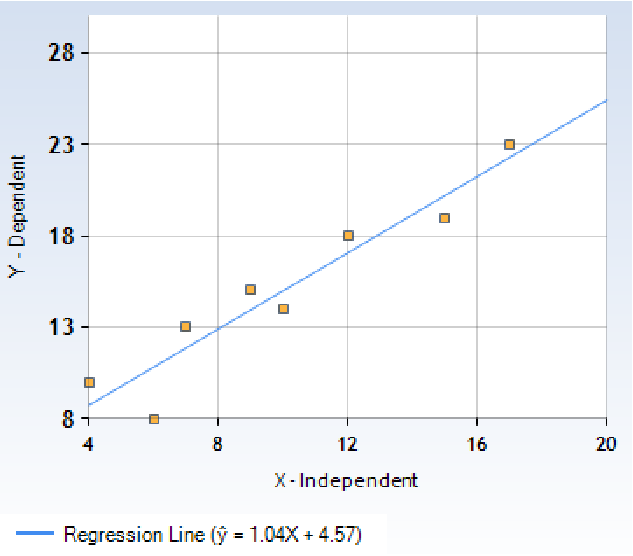
**Example**

Use the online calculator, find the line of best fit.

| x | y |
|---|---|
| 4 | 10 |
| 6 | 8 |
| 7 | 13 |
| 9 | 15 |
| 10 | 14 |
| 12 | 18 |
| 15 | 19 |
| 17 | 23 |

**Solution**
The calculator gives the equation $y = 1.04x + 4.57$ to be the line of best fit.

The calculator also shows the scatter plot and line of best fit. Ideally, the graph would start at (0,0), then we could see the y-intercept is 4.57.



## EXERCISE 4

Find the equation of the linear regression line for the following data set using the least squares method utilising the online calculator:

**1.**

| ***x*** | 4 | 6 | 7 | 9 | 10 | 12 | 15 | 17 |
|---|---|---|---|---|---|---|---|---|
| ***y*** | 10 | 8 | 13 | 15 | 14 | 18 | 19 | 23 |

**2.**

| ***x*** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| ***y*** | 35 | 28 | 22 | 16 | 19 | 14 | 9 | 7 | 2 |

## INTERPRETATION, INTERPOLATION AND EXTRAPOLATION

Once you have a linear regression line, the slope and intercept can give important information about the data set. The slope (m) indicates the rate at which the data are increasing or decreasing. The y-intercept indicates the approximate value of the data when *x* = 0.

The line can be used to make predictions such as finding a value for *y* for a given value of *x* – where *x* is not in the given data.

The line is always of the form: $y = (\text{gradient}) \times x + (\text{y} - \text{intercept})$

$$\boldsymbol{y = mx + c}$$

This line can be used to 'predict' data values for a given value of *x*. Of course, these are only approximations since the regression line itself is only an estimate of the 'true' relationship between the bivariate data. However, they can still be used, in some cases, to provide additional information about the data set such as making predictions.

### Interpolation

Interpolation is the use of the regression line to predict values in between two values already in the data set. If the data set is highly linear (*r* is near +1 or -1) then we can be confident that our interpolated value is quite accurate. If the data are not highly linear (*r* is near 0) then our confidence is duly reduced.

### Extrapolation

Extrapolation is the use of the regression line to predict values smaller than the smallest value in the data set or larger than the largest value.

Two problems may arise in attempting to extrapolate from a data set. Firstly, it may not be reasonable to extrapolate too far away from the given data values. For example, suppose there is a weather data set for 5 days. Even if it is highly linear (*r* is near +1 or -1), a regression line used to predict the same data 15 days in the future is highly risky. Weather has a habit of randomly fluctuating and patterns rarely stay stable for very long.

Secondly, the data may be highly linear in a narrow band of the given data set. For example, there may be data on stopping distances for a train at speeds of between 30 and 60 km/h. Even if they are highly linear in this range, it is unlikely that things are similar at very low spends (0-15 km/h) or high speeds (over 100 km/h).

Generally, we should feel more confident about the accuracy of a prediction derived from interpolation than on derived from extrapolation. Of course, it still depends upon the correlation coefficient (*r*). The closer to linearity the data are, the more confident our predictions in all cases.

**Example**
Using the data below, find the height of the child at age 8 and 15.

| **Age (years)** | 1 | 3 | 5 | 7 | 9 | 11 |
|---|---|---|---|---|---|---|
| **Height (cm)** | 60 | 76 | 115 | 126 | 145 | 148 |

**Solution**
The calculator gives the equation $y = 9.4x + 55.3$ to be the line of best fit.

The gradient is 9.4 thus the average growth is 9.4 cm per year. *c* = 55.3 thus at birth the height of the baby is 55.3 cm.

At age 8, *x* = 8, interpolation:
$y = mx + c$
$y = 9.4(8) + 55.3$
$y = 130.5$ cm

At age 15, *x* = 15, extrapolation:
$y = mx + c$
$y = 9.4(15) + 55.3$
$y = 196.3$ cm

The extrapolation prediction is clearly unreliable as it assumes a constant rate of growth throughout life. Thus, even though *r* = 0.96, in this case it is only useful in the range the data provides.

## EXERCISE 5

**1.** A drug company wishes to test the effectiveness of a drug to increase red blood cell counts in people who have a low count. The following data was collected.

| **Day of experiment** | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|
| **Red blood cell count** | 210 | 240 | 230 | 260 | 260 | 290 |

a. What are the independent and dependent variables in this case?

b. Find the relationship in the form of $\boldsymbol{y = mx + c}$ using the online calculator.

c. What is the rate at which the red blood cell count is changing?

d. What was the blood cell count at the beginning of the experiment (i.e. on day 0)?

**2.** A wildlife exhibition is held over 6 weekends and features still and live displays. The number of live animals that are being exhibited varies each weekend. The number of animals participating, together with the number of visitors to the exhibition each weekend, is shown below.

| **Number of animals** | 6 | 4 | 8 | 5 | 7 | 6 |
|---|---|---|---|---|---|---|
| **Number of visitors** | 311 | 220 | 413 | 280 | 379 | 334 |

a. Find the rate of increase of visitors as the number of live animals is increased.

b. Find the predicted number of visitors if there are no live animals.

**3.** An electrical goods warehouse produces the following data showing the selling price of electrical goods to retailers and the volume of those sales.

| **Selling price ($)** | 60 | 80 | 100 | 120 | 140 | 160 | 200 | 220 | 240 | 260 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Sales volume (x1000)** | 400 | 300 | 275 | 250 | 210 | 190 | 150 | 100 | 50 | 0 |

Perform a least-squares regression analysis and discuss the meaning of the gradient and y-intercept.

**4.** The following table represents the costs for shipping a consignment of shoes from Melbourne factories. The cost is given in terms of distance from Melbourne. There are two factories that can be used. The data is summarised below.

| **Distance from Melbourne (km)** | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|---|
| **Factory 1 cost ($)** | 70 | 70 | 90 | 100 | 110 | 120 | 150 | 180 |
| **Factory 2 cost ($)** | 70 | 75 | 80 | 100 | 100 | 115 | 125 | 135 |

a. Find the least-squares regression equation for each factory.

Factory 1

Factory 2

b. Which factory is likely to have the lowest shipping cost to a shop in Melbourne? (use numbers)

c. Which factory is likely to have the lowest shipping cost to Mytown, 115 kilometres from Melbourne? (use maths)

**5.** A factory produces calculators. The least-squares regression line for cost of production (C) as a function of numbers of calculators (n) produced is given by:

$$C = 7.76n + 660$$

This function is **only definitely accurate** when producing between 100 and 1000 calculators. Outside of that range, there is uncertainty in the function.

a. Find the cost to produce 200 calculators.

b. How many calculators can be produced for a cost of $2000?

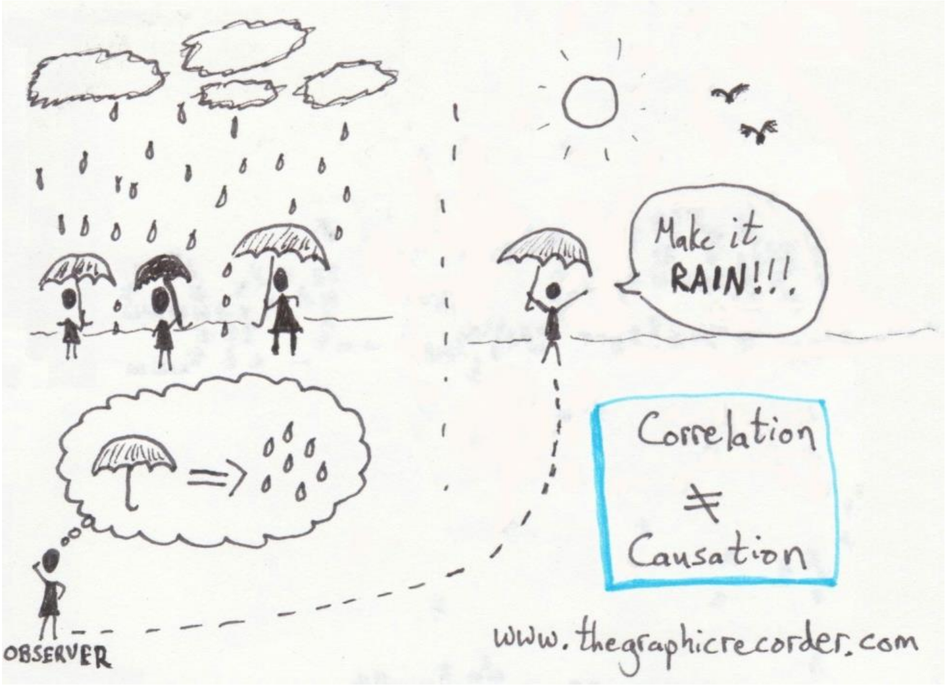
c. Find the cost to produce 10 000 calculators.

d. What are the ‘fixed’ costs for this production, before a calculator has even been made?

e. Which question/s from a. to c. is an interpolation?

## AST STYLE QUESTION

Describe/explain the following cartoon in relation to causation and correlation.



## INVESTIGATION WEEK 3/4

Complete the table below.

The first two columns give the values for age (x, in years) and systolic blood pressure (y, in mmHg) for 15 women.

| X | Y | $X^2$ | $Y^2$ | XY |
|---|---|---|---|---|
| 42 | 130 | 1764 | 16900 | 5460 |
| 46 | 115 | | | |
| 42 | 148 | | | |
| 71 | 100 | | | |
| 80 | 156 | | | |
| 74 | 162 | | | |
| 70 | 151 | | | |
| 80 | 156 | | | |
| 85 | 162 | | | |
| 72 | 158 | | | |
| 64 | 155 | | | |
| 81 | 160 | | | |
| 41 | 125 | | | |
| 61 | 150 | | | |
| 75 | 165 | | | |
| Total of X<br>$\sum = 984$ | Total of Y<br>$\sum = 2193$ | Total of $X^2$ | Total of $Y^2$ | Total of XY |

Now use the values from the table and the formulae given to estimate the parameters of the **linear regression** for the x and y. i.e use the sums and the formulae given to work out 'a', 'b', r and $r^2$. Show all working.

Check your answers on
https://www.socscistatistics.com/tests/regression/default.aspx

Write out the linear regression rule for the age (x, in years) and systolic blood pressure (y, in mmHg) for 15 women.

Use your rule to estimate the systolic blood pressure for a 59-year-old woman.

**Formulae:**

$$a = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2}$$

$$b = \bar{Y} - a\bar{X}$$

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{n}\right]\left[\sum Y^2 - \frac{(\sum Y)^2}{n}\right]}}$$

## MARKING RUBRIC

| CRITERIA | EXPECTATIONS | POSS | MULT | GIVEN | TOTAL |
|---|---|---|---|---|---|
| | | | | | |
| **Practical** | Student completes practical work of the brief to an acceptable standard set by the teacher. | **2** | **3** | | **/6** |
| **Investigation** | Student completes the investigation of the brief to an acceptable standard set by the teacher. | **2** | **2** | | **/4** |
| | | | | | |
| **Reasoning and communications** | Student responses are accurate and appropriate in presentation of mathematical ideas in different contexts, with clear and logical working out shown. | **4** | **-** | | **/4** |
| **Concepts and techniques** | Student submitted work selects and applies appropriate mathematical modelling and problem solving techniques to solve practical problems, and demonstrates proficiency in the use of mathematical facts, techniques and formulae. | **4** | **-** | | **/4** |
| | **Submission Guidelines** | | | | |
| **Timeliness** | Student submits the exercises and investigation by the set deadline. See scoring guidelines for specific details. | **2** | **-** | | **/2** |
| | | **FINAL** | | | **/20** |

**Student Reflection**:
How did you go with this week's work?

What was interesting?

What did you find easy?

What do you need to work on?