2026

# GENERAL MATHEMATICS

## Level 3

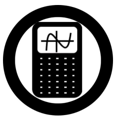
# BIVARIATE DATA

# General Mathematics: Level 3

## GM3 – BIVARIATE DATA

By Jess Bertram

With sincere thanks to John Short and Rick Smith.

| ICON: | MEANING: |
|---|---|
|  | Worked example |
|  | Complete with your teacher |
|  | Try it yourself |
|  | CAS Calculator can be used |
|  | Tips / shortcuts |



For ordering, feedback and requests please visit:

BertramEducation.com

**GENERAL MATHEMATICS - LEVEL 3**

**GM3 BIVARIATE DATA**

**BY JESS BERTRAM**

# INDEX PAGE

Part 1

# ASSOCIATION BETWEEN VARIABLES

*Have you ever wondered how sports teams decide which players to recruit, how companies know what ads work best, or how scientists prove that medicines are effective? The answer is statistics.*

Statistics is the science of collecting, analysing, interpreting, and presenting data. Understanding data allows us to make predictions, test ideas, and draw conclusions about real-world situations.

In this workbook you will focus on bivariate data analysis (relationships between two variables) and time series analysis (patterns over time).

## THE STATISTICAL INVESTIGATION PROCESS



The statistical investigation process is a systematic approach to answering questions using data. It typically follows the below steps:

**The Statistical Investigation Process**

1. Identify a problem and pose a statistical question
2. Collect or obtain data
3. Organise and analyse the data
4. Interpret the results
5. Communicate the findings

**The statistical investigation process is the foundation of this course.**

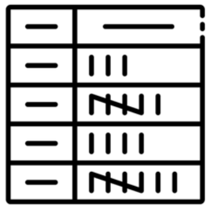
In this book you will learn how to run a full statistical investigation. You will begin with simple, guided investigations and gradually add new skills. As you progress through the book, you will learn how to collect and analyse data and draw meaningful conclusions. By the end, you will be ready to carry out a complete investigation and write a report on your findings.

This **example investigation** shows you the 'big picture' of what the investigation process looks like in action.

**Scenario:** A school canteen wants to know if increasing advertising for fruit smoothies will lead to more sales. They ask a group of TCE general maths students to help them investigate.

| | | |
|---|---|---|
| **Step 1** | Identify the problem and pose a statistical question | "Does the number of advertising posters affect smoothie sales?" |
| **Step 2** | Collect or obtain data | Students record the number of posters displayed and the number of fruit smoothies sold over the course of 10 weeks. They vary how many posters are displayed each week. |
| **Step 3** | Organise and analyse the data | Data is collected in a two-way frequency table and then plotted as a scatterplot. They use the data to create a range of graphs showing the results. Students use their statistics skills to find a linear relationship between the number of posters and the number of sales. They calculated the correlation coefficient to quantify the strength of the relationship. |
| **Step 4** | Interpret the results | Students analyse the data to look for patterns, associations and causation. They posed questions like: what happens to smoothie sales as the number of posters increase? Could things other than the number of posters be affecting sales?<br>A strong positive relationship was identified - more posters generally lead to more smoothies sold. |
| **Step 5** | Communicate the findings | The students prepared a report of their findings and advised the canteen that increasing advertising may increase sales. |

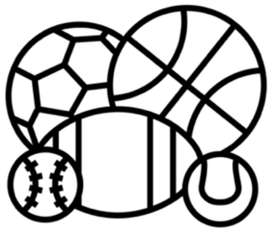
## TWO-WAY FREQUENCY TABLES



A two-way frequency table is used to organise data about two variables at the same time.
They show how often combinations of the two variables occur.

Two-way frequency tables are useful for:

- Organising raw data for analysis
- Comparing categories across variables
- Calculating percentages to identify patterns and associations



**Worked Example.** The school board wants to know how engaged TCE students are in sport. They survey 50 TCE students from their school and ask them the following questions:

Do you play a sport? (Yes / No)

Are you in Year 11 or Year 12?

From the survey they create a two-way frequency table.



| | **Year 11** | **Year 12** | **Total** |
|---|---|---|---|
| Play sport | 11 | 15 | **26** |
| Don't play sport | 8 | 16 | **24** |
| **Total** | **19** | **31** | **50** |

26 students play sport

The **rows** (horizontal →) represent whether or not students play a sport.

The **columns** (vertical ↓) represent the year level of the students.

Percentages can be used in two-way frequency tables.

30% of the students were in year 12 AND played sport.

| | **Year 11** | **Year 12** | **Total** |
|---|---|---|---|
| Play sport | 22% | 30% | **52%** |
| Don't play sport | 16% | 32% | **48%** |
| **Total** | **38%** | **62%** | **100%** |

48% of all students don't play sport.

For the table above, percentages have been calculated out of the total number of students surveyed (50).

The number of year 11 students who play sport (11) is divided by the total number of students (50). $(11 \div 50 = 0.22 \times 100 = 22\%)$

From analysing the overall percentages (on the previous page) we can see that:

- Overall, 52% of the TCE students play sport
- More year 12 students (62%) play sport than year 11 students (38%)
- Year 12 students are over-represented in the sample (62% of respondents)

We could also use percentages to divide the data by year level.

| | **Year 11** | **Year 12** |
|---|---|---|
| Play sport | 58% | 48% |
| Don't play sport | 42% | 52% |
| **Total** | **100%** | **100%** |

These percentages were calculated from the column totals. The number of year 11 students who play sports (11) was divided by the total number of year 11 students (19). $(11 \div 19 = 0.58 \times 100 = 58\%)$

From analysing the year level specific percentages, we can see:



- Among Year 11 students, most play sport (58%) but many do not.
- Among Year 12 students the group is more evenly split
- Suggests that Year 11's are *more likely* to play sport than Year 12 students.

**With your teacher:**
A community group wants to know how people prefer to spend their weekends. They ask community members whether they prefer outdoor activities (like hiking or mountain biking) or indoor activities (like movies or gaming). The responses are grouped by age: 'Under 20' and '20 & Over'

Complete the two-way frequency tables below by adding the totals to the first table, then converting to percentages in the second table.

| | **Under 20** | **20 & over** | **Total** |
|---|---|---|---|
| Outdoor | 15 | 22 | |
| Indoor | 18 | 45 | |
| **Total** | | | |

| | **Under 20** | **20 & over** | **Total** |
|---|---|---|---|
| Outdoor | % | % | **%** |
| Indoor | % | % | **%** |
| **Total** | **%** | **%** | **%** |

**With your teacher.**

A survey was conducted with people about their preferred mode of transport to school or work. They could choose between public transport or private car. The responses are also grouped by gender: Male, female, and other.

a. Complete the table below

| | **Male** | **Female** | **Other** | **TOTAL** |
|---|---|---|---|---|
| Public | 5 | 10 | 2 | |
| Private | 10 | 25 | 3 | |
| **TOTAL** | | | | |

b. Turn the table from (a) into percentages out of the total (55)

| | **Male** | **Female** | **Other** | **TOTAL** |
|---|---|---|---|---|
| Public | % | % | % | **%** |
| Private | % | % | % | **%** |
| **TOTAL** | **%** | **%** | **%** | **%** |



c. Turn the table from (a) into percentages per gender.

| | **Male** | **Female** | **Other** |
|---|---|---|---|
| Public | % | % | % |
| Private | % | % | % |
| **TOTAL** | **%** | **%** | **%** |

d. Write two observations about your findings from this data.

**Try it yourself (TWO-WAY FREQUENCY TABLES):** **Answers page 75**

1. A group of 50 students were asked about their preferred music genre.
   a. Fill in the missing totals
   b. Add overall percentages for each cell (out of 50)
   c. What percentage of students prefer rock music?

| | Year 9 | Year 10 | Total |
|---|---|---|---|
| Pop | 12 | 18 | |
| Rock | 8 | 12 | |
| **Total** | | | 50 |

2. Data was collected from 48 students:
   a. Complete the table with totals.
   b. Calculate the percentage of students who travel by bus and are in year 8.
   c. Make one observation about the data.

| | Year 7 | Year 8 | Total |
|---|---|---|---|
| Walk | 6 | 4 | |
| Bus | 8 | 10 | |
| Car | 8 | 12 | |
| **Total** | | | 48 |



3. You surveyed 40 students in year 9 and 10 to find out how they got to school this morning. The results are shown in the below frequency table.
   a. Complete the two-way frequency table by adding row totals, column totals, and grand total.
   b. Calculate the column percentages for each cell (to 2 decimal places)
   c. Make two observations about the data.

| | Walk to school | Bus to school | Car to school |
|---|---|---|---|
| Year 9 | 4 | 6 | 5 |
| Year 10 | 3 | 8 | 14 |

4. Create a two-way frequency table for the results of this survey: Rosemary asks 150 people who their favourite member of One Direction is. Of the 50 people aged 30+ surveyed, 10% said Zayn, 50% said Harry, 20% said Liam, 10% said Louis and 10% said Niall. Of the 100 people under 30 years old, 15 said Zayn, 25 said Harry, 10 said Louis, 45 said Niall and 5 said Liam.

5. Determine the missing numbers in the following tables:

| (a) | Group 1 | Group 2 | **Total** |
|---|---|---|---|
| For | 12 | 8 | |
| Against | 15 | | |
| **Total** | | | **39** |

| (b) | Group 1 | Group 2 | **Total** |
|---|---|---|---|
| For | | 20 | **28** |
| Against | | | |
| **Total** | **24** | **26** | |

| (c) | Group 1 | Group 2 | **Total** |
|---|---|---|---|
| For | 10 | | **40** |
| Against | | 20 | |
| **Total** | **50** | | |

| (d) | Group 1 | Group 2 | **Total** |
|---|---|---|---|
| For | 23 | | **38** |
| Against | | | |
| **Total** | | **100** | **145** |

6. Georgia has been accused of murdering another husband. The town of Wellsbury organised a survey to see who in the town thought she was guilty. 1,000 town people were surveyed, and it was discovered that 103 of the 550 men, and 350 of the women, believe she is guilty.
   a. Create a two-way frequency table to represent the situation.
   b. Compare the percentage of men who believe Georgia is guilty with the percentage of women who believe she's guilt. Are they fairly even?

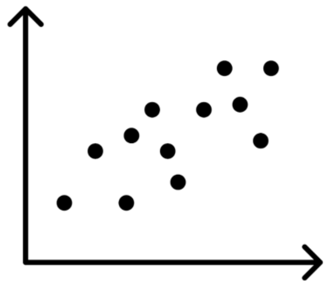


7. A company is trialing a new energy drink that claims to improve concentration levels during study sessions. 400 university students volunteered to participate in the trial. 250 of the students were given the new energy drink and the remaining 150 were given a placebo drink. After a two-hour study session, students were asked if they felt their concentration had improved. Of those who received the new energy drink, 210 reported improved concentration. Of those who received the placebo drink, 60 reported improved concentration.
   a. Present the data in a two-way frequency table.
   b. Summarise your findings in words.

8. Create a two-way table: A survey was conducted to investigate the opinions of employees and managers on the effectiveness of remote work compared to working in the office. A total of 520 people responded to the survey. Of the 150 respondents who believed that remote work was more effective, 90 were employees. 200 respondents thought that office work was more effective, with 120 of these being employees. The remaining respondents believed that both work environments were equally effective. Out of these, 110 were employees.

## SCATTER PLOTS

A scatter plot is a graph that shows the relationship between two numerical variables. Each point on the graph represents one observation, with one variable on the x-axis (horizontal) and the other on the y-axis (vertical).



Scatter plots allow us to:

- See if there is a pattern or trend between two variables (AKA association)
- Identify the strength of a relationship
- Detect unusual values (outliers)
- Decide whether one variable may help predict the other
- Visualise the data to see if a relationship exists

**An association in statistics is when two variables are related in some way. It doesn't mean one causes the other — just that as one changes, the other tends to change in a particular pattern.**

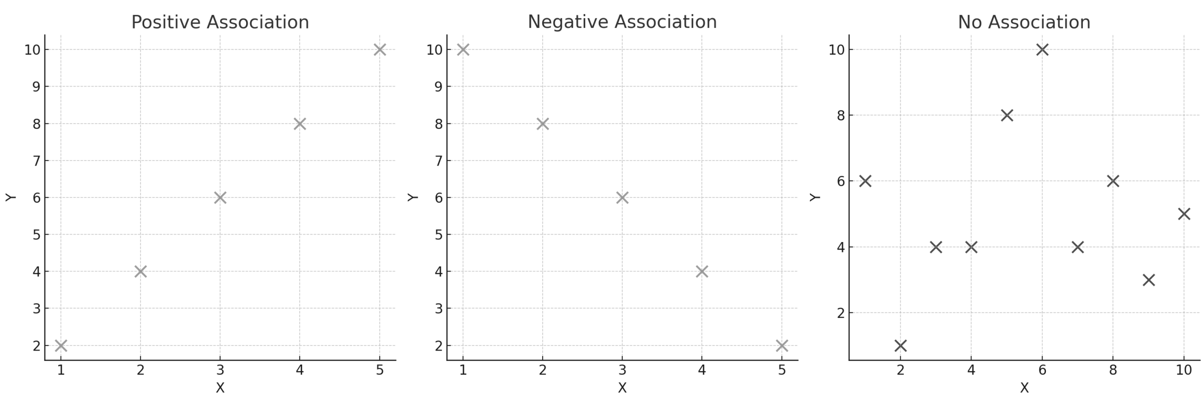


There are 3 main things you may notice from a scatter plot.

1. Direction of association
2. Strength of association
3. Form

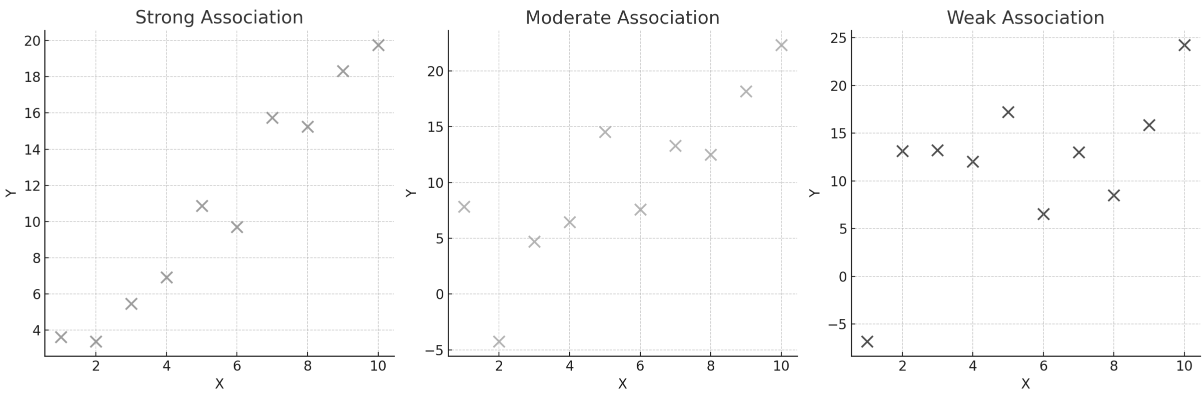
**1. Direction of association**:

- **Positive association:** As one variable increases, the other increases
- **Negative association:** As one variable increases, the other decreases
- **No association:** The points are scattered randomly with no clear trend.



**2. Strength of association (also called correlation):**

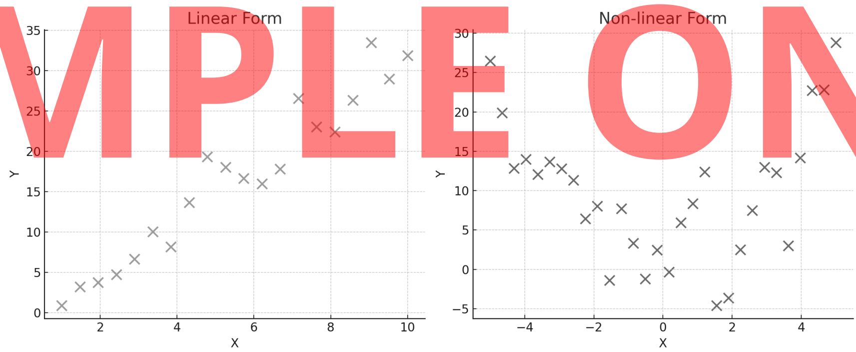
- **Strong:** Points lie close to an imaginary trend line.
- **Moderate:** Points are spread out but still show some trend.
- **Weak:** No visible trend.




**3. Form**:

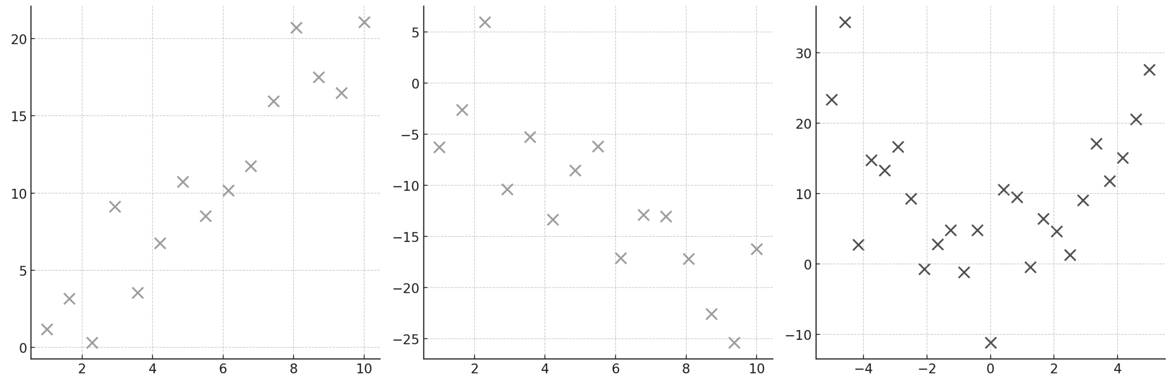
- **Linear**: Data follows an approximate straight-line pattern.
- **Non-linear**: Data curves (e.g., quadratic or exponential shapes).





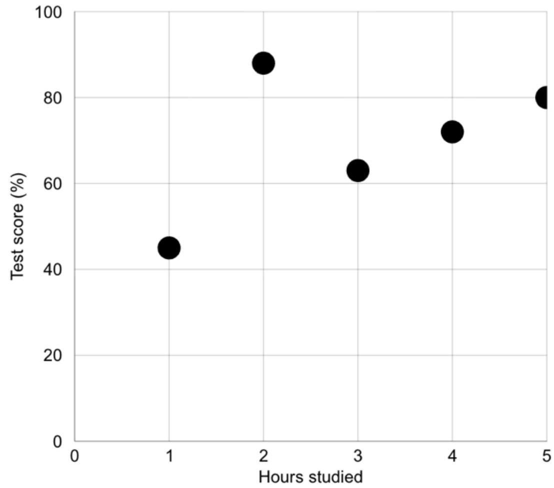

**With your teacher:**

Determine the direction of association, strength of association and form of the following scatter plots.



**Worked example:** Ms Bertram wants to determine if students who study for longer than their peers have higher test scores. She asks students to time themselves studying for a test, then records this data with their test results in a table. She uses the table to plot the results on a scatter plot.

| Hours studied | Test result (%) |
|---|---|
| 1 | 45 |
| 2 | 88 |
| 3 | 63 |
| 4 | 72 |
| 5 | 80 |




She makes the following observations from the scatterplot:

- There is a strong positive linear association.
- There is one obvious outlier.
- More study time is generally linked with higher test scores.

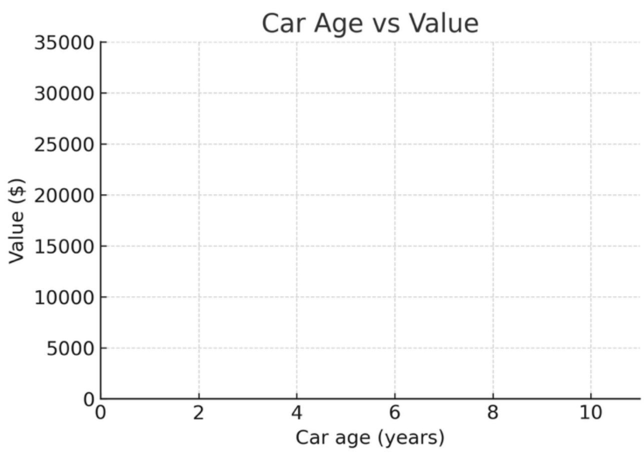


**With your teacher:**

*Alex's Awesome Automobiles* buys and sells used cars. Alex suspects there is an association between the age of the cars and their value. He records the age and value of 5 car sales over the course of 2 weeks.

a. Use the data to prepare a scatter plot.

| Car age (years) | Value ($) |
|---|---|
| 1 | 30,000 |
| 3 | 25,000 |
| 5 | 18,000 |
| 7 | 12,000 |
| 10 | 6,000 |




e. Comment on the direction of association, strength of association, and form.

**With your teacher:**

A group of students recorded the number of hours they spent on social media each day and the number of hours they slept that night. They are curious if the time on social media impacts how much sleep they get.

Their results are as follows:

| Social media (hours) | Sleep (hours) |
|---|---|
| 1 | 9 |
| 2 | 8 |
| 3 | 7 |
| 4 | 7 |
| 5 | 6 |

Your tasks:

a. Determine which column is the independent variable (X axis) and which is the dependent variable (Y axis).



b. Prepare a scatterplot from scratch using the graph paper above.

c. Describe the association

d. Write a few sentences interpreting the results in the context of the data.

## Try it yourself (SCATTERPLOTS) **Ans pg75**

1. Underline the dependent variable in each of the following situations:
   a. The number of hours spent practicing piano affects the quality of performance.
   b. The temperature of the oven influences the time it takes to bake a cake.
   c. A doctor observes patients' heart rates after different levels of exercise.
   d. The level of fertiliser given to crops impacts the height of the plants.
   e. The distance a car travels depends on the amount of petrol in the tank.

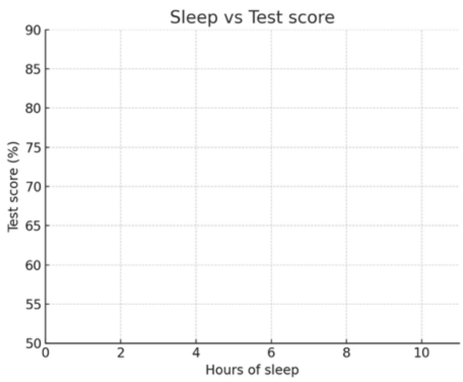
2. Draw a scatter plot of each of the below sets of data *(graph paper below)*

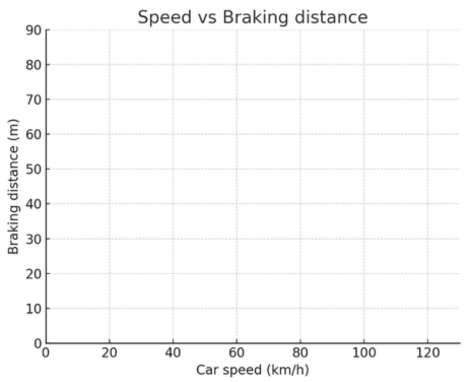
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Hours of sleep | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Test score (%) | 55 | 60 | 65 | 72 | 78 | 81 | 84 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Car speed (km/h) | 20 | 40 | 60 | 80 | 100 | 120 |
| Braking distance (m) | 5 | 10 | 20 | 35 | 55 | 80 |



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Number of practice sessions | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Goal scored | 2 | 4 | 7 | 9 | 12 | 13 | 14 |




**Try it yourself (SCATTERPLOTS)** **Ans pg75**

*Ans pg75*

3. A shop owner records the outside temperature (°C) and the number of ice creams sold at the beach.

Temp (°C): 15, 20, 25, 30, 35

Ice creams sold: 20, 35, 50, 70, 90

a. Label the axis

b. Plot the scatter plot.

c. Identify the direction and strength of the association.

d. Write one sentence summarising the findings.

4. Draw a scatterplot for each of the following sets of data: *(graph paper below)*



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Distance run (km) | 1 | 2 | 3 | 5 | 7 | 10 | 12 |
| Calories burned | 80 | 150 | 210 | 360 | 500 | 700 | 850 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Number of advertisements | 0 | 1 | 2 | 3 | 4 | 5 |
| Products sold | 20 | 25 | 40 | 50 | 65 | 80 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Number of work hours | 0 | 5 | 10 | 15 | 20 | 25 |
| Weekly income ($) | 0 | 75 | 150 | 220 | 300 | 370 |

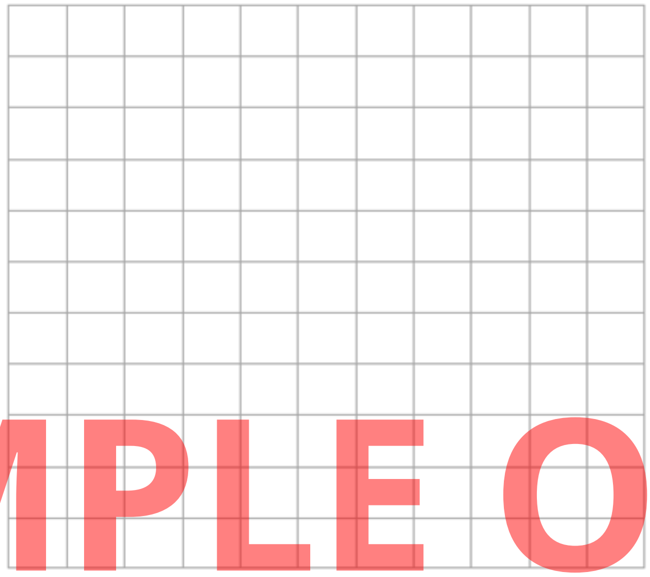
**Try it yourself (SCATTERPLOTS)** **Ans pg75**

5. The number of letters in a student's first name is compared with their test score.

Letters in name: 3, 4, 5, 6, 7

Test score (%): 62, 48, 75, 53, 70

a. Plot a scatterplot for the data
b. Comment on the direction, form, and strength of association.
c. Write a sentence interpreting the results.

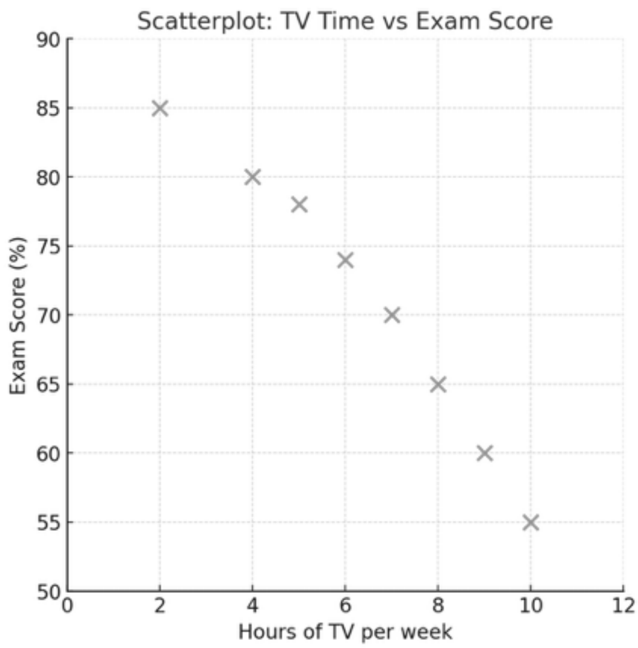




6. Anna is studying statistics and wants to draw an example of a scatterplot with a positive linear association. Which of the following two options would be more likely to have this kind of association?

- Height vs shoe size
- Shoe size vs test score

7. Describe the association shown on the Scatterplot. What does this suggest about the relationship between watching TV and exam performance?



## ASSOCIATION VS CAUSATION

> **Association** means that two variables are related in some way.
> When one changes, the other tends to change too.
>
> **Causation** means that one variable directly causes the other to change.
>
> **Two variables being associated does not mean that one causes the other.**
>
> **ASSOCIATION ≠ CAUSATION**

There are several possible reasons why two variables may appear related without a cause-and-effect link.

1. **Coincidence**
   - Sometimes data lines up by chance.
   - Example: Global temperatures and the number of pirates in the world both decreased over the past 200 years - it doesn't mean pirates control the climate!

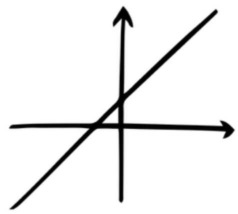


2. **Confounding**
   - A third variable could explain the relationship.
   - Example: Ice cream sales and drowning deaths are positively associated. The confounding variable is temperature - in summer, both ice cream sales and swimming increase.
3. **Reverse causation**
   - Sometimes we assume the wrong direction.
   - Example: Towns with more firefighters also have more fire damage. It's not that firefighters cause more fires – towns with larger number of fires employ more firefighters.

**Worked example.**
A researcher studies children and finds that those with larger shoe sizes tend to have better reading ability. Is this a causal relationship?

No – the larger shoe sizes are not causing an improvement in reading. There is a confounding variable: age. The older children got, the better their reading skills were, and the bigger their feet got too.
Older children generally have bigger feet *and* better reading skills.

# Part 2 LINEAR RELATIONSHIPS



Statistics doesn't just stop at finding out *whether* two variables are associated. Once we've identified an association, the next step is to see if we can model the relationship with a mathematical equation.

In this part of the workbook, we focus on linear relationships. These are situations where the data can be modelled by a straight line (linear regression).

**Why model a relationship using an equation?**

By modelling a relationship with a straight line, we can:

- Describe the strength and direction of the relationship more precisely.
- Write an equation that links the two variables.
- Use the line to make predictions about new values.
- Interpret the gradient (slope) and intercept in the context of a real problem.
- Understand when predictions are safe (interpolation) and when they may be unreliable (extrapolation).



**Example.**

Mr Smith teaches General Maths. His students are stressed and he wants to guide them on how much study they should be doing leading up to tests. He collects data on hours studied and test scores. A scatterplot showed a strong positive association.

- In Part 1, we would have said: "More hours studied tends to be associated with higher test scores. The correlation coefficient was r = 0.92, which suggests a strong positive association."
- In Part 2, we go further: we determine a line of best fit: $y = 5x + 40$

  Then we interpret:
  - The intercept (40) suggests that if no hours are studied, the predicted test score is 40%.
  - The gradient (5) means each extra hour of study is associated with an increase of 5% in the test score.

## STRAIGHT LINE EQUATION INTRODUCTION

A straight line has the general equation:

$$y = mx + c$$

*(can also be written as* $y = ax + b$*)*

where:

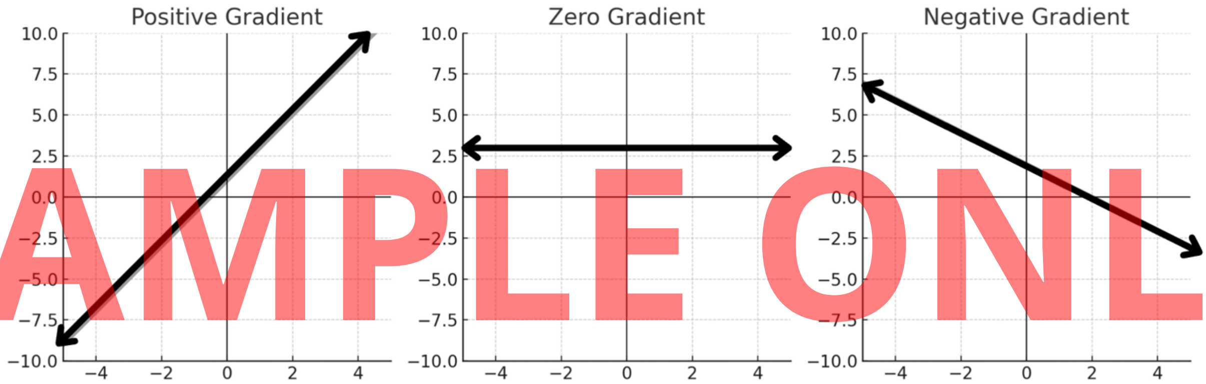
- **m** is the **gradient (slope)** of the line
- **c** is the **y-intercept** (the point where the line crosses the y-axis)

### Gradient (slope)

The gradient tells us how steep the line is.

The gradient can be positive, zero, or negative.

The larger the number, the steeper the graph.






We can find the **gradient of a line** from two points.

$$m = \frac{rise}{run} \quad OR \quad m = \frac{y_2 - y_1}{x_2 - x_1} \quad OR \quad m = \frac{\Delta y}{\Delta x}$$

There are lots of different ways to write it, but essentially gradient is calculated by the difference in Y values divided by the difference in X values.

**Worked example:** Find the gradient of the line between each pair of points.

(2,3) and (6,11)

$$m = \frac{11 - 3}{6 - 2}$$

$$m = \frac{8}{4}$$

$$m = 2$$

(−4,5) and (2, −1)

$$m = \frac{-1 - 5}{2 - -4}$$

$$m = \frac{-6}{6}$$

$$m = -1$$

(0,−2) and (3,−2)

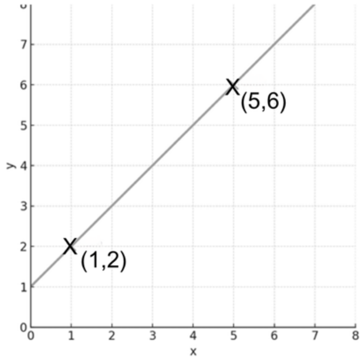
$$m = \frac{-2 - -2}{3 - 0}$$

$$m = \frac{0}{3}$$

$$m = 0$$

**With your teacher:** Find the gradient between the following pairs of points.

(1, 2) & (5, 6)   (1, 5) & (4, -1)   (1, 2) & (5, 3)




### The Y intercept

Once we know the gradient of a line, we can find the y-intercept by substituting one of the points into the straight-line equation (y = mx + c). The steps are:

1. Find the gradient $(m)$ using $m = \frac{y_2 - y_1}{x_2 - x_1}$
2. Use the gradient and one of your points in the formula $y - y_1 = m(x - x_1)$



**Worked Example 1.**

Find the equation of the line through the points (2, 3) and (6, 11).

$$m = \frac{11 - 3}{6 - 2} = \frac{8}{4} = 2$$ *← Find the gradient first*

$$y - 11 = 2(x - 6)$$

$$y - 11 = 2x - 12$$

$$y = 2x - 1$$ ***← Final equation***

**Worked Example 2.**

Find the equation of the line through (-4, 5) and (2, -1).

$$m = \frac{-1 - 5}{2 - -4} = \frac{-6}{6} = -1$$ *← Find the gradient first*

$$y - 5 = -1(x - -4)$$

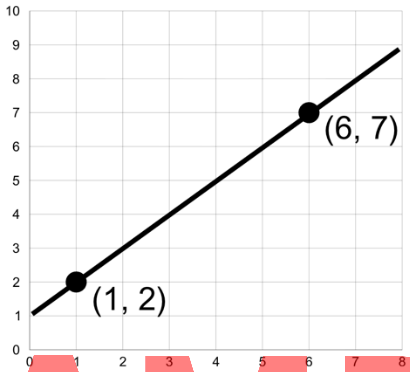
$$y - 5 = -1x - 4$$

$$y = -x + 1$$ ***← Final equation***

**With your teacher:**

1. Find the equation of the line that passes through the points (1,4) and (5,8)

2. Find the equation of the line below.






**Finding the gradient and Y intercept from two points using Casio Classpad**

1. On the main menu, tap the Statistics icon.
2. Enter your data (minimum of 2 points)
   - Type the x-values into List 1
   - Type the y-values into List 2
3. Tap Calc (top toolbar).
   - Choose regression – linear regression
   - Select List 1 as the X-list and List 2 as the Y-list.
   - Press ok
4. The calculator will display a variety of information.

   **a = gradient**

   **b = y intercept**

   r = correlation coefficient

   $r^2$ = coefficient of determination

**Try it yourself (STRAIGHT LINE INTRODUCTION):** **Ans pg77**

1. State the Y intercept (c) and gradient (m) in each equation.

a) $y = 3x + 4$ b) $y = 2x + 5$ c) $y = 6x + 3$ d) $y = 9x + 2$
e) $y = 3x - 5$ f) $y = 7x - 10$ g) $y = 10x - 6$ h) $y = 5x - 4$
i) $y = -2x + 5$ j) $y = -3x + 10$ k) $y = -4x - 10$ l) $y = -5x - 5$
m) $y = 5 + 2x$ n) $3x - 0.5 = y$ o) $y = \frac{x}{2} + \frac{3}{2}$ p) $k = 2r - 5.3$

2. Write the equation of a line given the gradient (m) and Y intercept (c).

a) m = 1, c = 0 b) m = 2, c = 3 c) m = -1, c = 4 d) m = 5, c = -2
e) m = 0, c = 6 f) m = -2, c = -3 g) m = 3, c = 1 h) m = 0.5, c = -1
i) m = -4, c = 2 j) m = 7, c = -5 k) m = -0.5, c = 3 l) m = 10, c = 0
m) m = 2/3, c = -4 n) m = -3/2, c = 5 o) m = 8, c = -10 p) m = -6, c = -7

3. Find the gradient of the straight line passing through the following points.

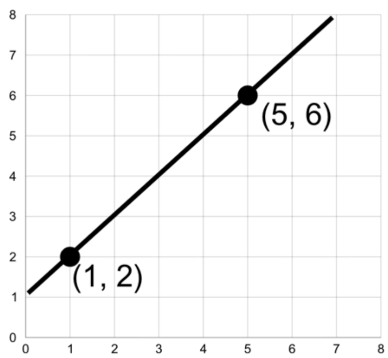
a) (0,0) and (4,4) b) (1,2) and (3,6) c) (2,5) and (4,2) d) (-1,3) and (2,3)
e) (0,5) and (5,0) f) (-2,-2) and (2,2) g) (1,1) and (4,7) h) (2,8) and (6,2)
i) (3,5) and (3,10) j) (0,-4) and (6,2) k) (-3,7) and (5,-1) l) (2,3) and (5,3)
m) (1,2) (3.5,6.5) n) (0,0) and (7,10) o) (-4,-2) and (4,6) p) (2,7) and (6,3)

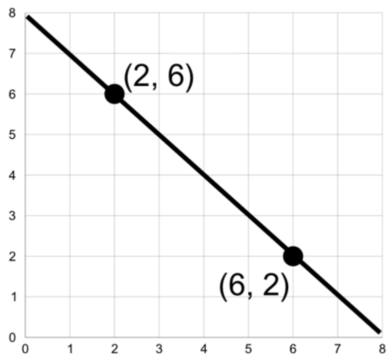


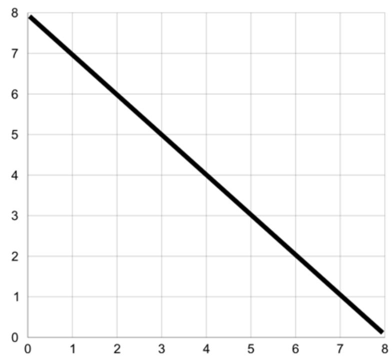
4. Find the equation of the line passing through the points.

a) (0,0) and (2,2) b) (1,1) and (3,5) c) (0,2) and (4,6) d) (2,3) and (6,3)
e) (-2,1) and (2,5) f) (0,-1) and (4,3) g) (-3,-2) and (3,4) h) (1,4) and (5,-2)
i) (-2,6) and (2,0) j) (0,5) and (5,0) k) (1.5,2), (3.5,6) l) (-1,3.4), (2,7.4)
m) $(0,\frac{3}{2})$ and $(4,\frac{11}{2})$ n) $(\frac{2}{3},1)$ and $(\frac{8}{3},5)$ o) (-4.5,-2), (3.5,5) p) (-2,7), (6,-1)

5. Find the equation of the line in the following graphs:

a)



b)



c)


**CONSTRUCTING GRAPHS – SUBSTITUTION METHOD**

There are multiple ways to draw a graph from a given equation. In this workbook we will focus on the substitution method. You may use a different method if you prefer.

**Steps to graph a line – substitution method**

1. Choose at least two values for (x)
2. Substitute the values for (x) into the equation to find (y)
3. Plot each pair (x,y)
4. Draw a straight line through the points

**Worked example.** Graph the equation $y = 2x + 1$

**Step 1** – Choose 2 values for (x)

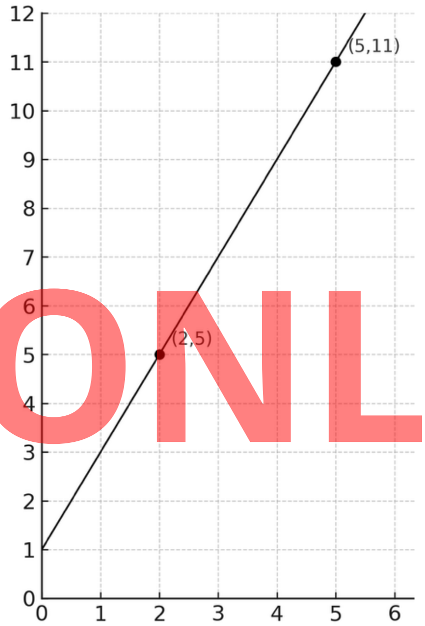
X = 2 and X = 5

**Step 2** – Substitute each X into the equation.

| (X = 2) | (X = 5) |
|---|---|
| Y = 2(2) + 1 | Y = 2(5) + 1 |
| Y = 5 | Y = 11 |
| Point @ (2,5) | Point @ (5,11) |

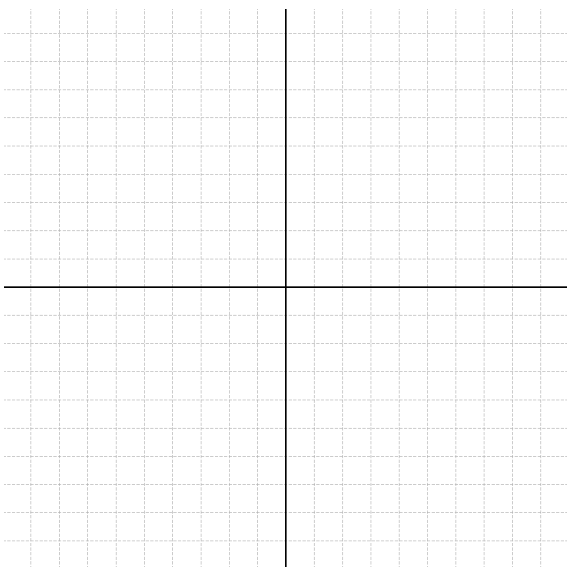
**Step 3** – Plot each pair

**Step 4** – Draw a straight line through the points






**With your teacher.** Graph the equation $y = -3x + 10$



As well as graphing the equation of a line, we can also use substitution to find the value of y for a specific value of x. This is very useful when we want to make predictions in real-life contexts.

**Worked Example**

Katy runs a small business. She has found that her profit (y, in dollars) can be modelled by the equation y = 15x – 50, where x is the number of items she sells.

She wants to know what her profit will be if she sells 20 items.

Substitute x = 20:

y = 15(20) - 50

y = 300 - 50

y = 250

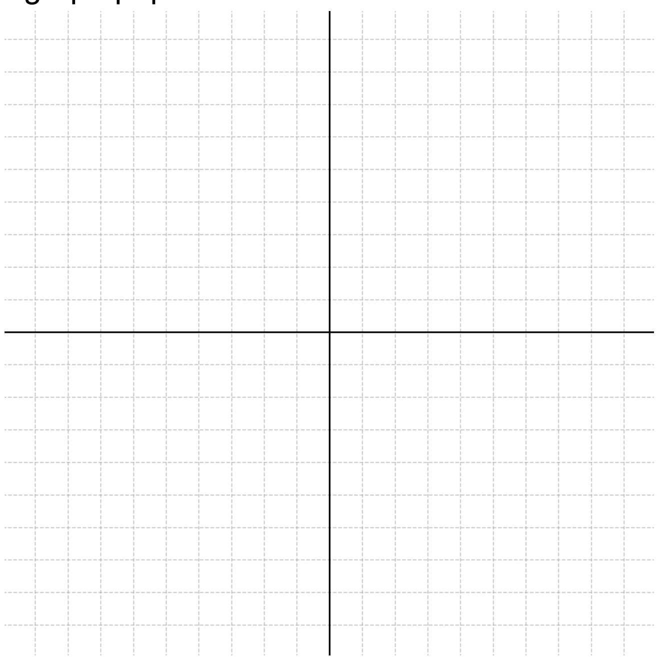
If Anna sells 20 items, she can expect to make a profit of $250.

**With your teacher.** Using the example above, find Katy's profit when she sells 10 items, 50 items, and 100 items.



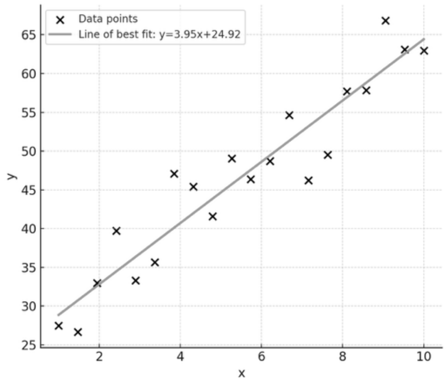
**Try it yourself (SUBSTITUTION):** **Ans pg77**

1. Graph the following equations on the same graph paper.
   a. $y = 3x - 4$
   b. $y = 5x + 1$
   c. $y = -2x + 4$
   d. $y = 0.5x + 0.5$
   e. $y = -x + 2$
2. For each of the above equations, find the value of Y when:
   X = 5
   X = 10
   X = 100



## LINE OF BEST FIT

When we have a scatterplot showing an association between two variables, we can summarise the relationship using a **line of best fit.** This line is not drawn through all the points - it is placed so that it best represents the overall trend of the data.




Fitting a line can be done by hand or using your CAS calculator.

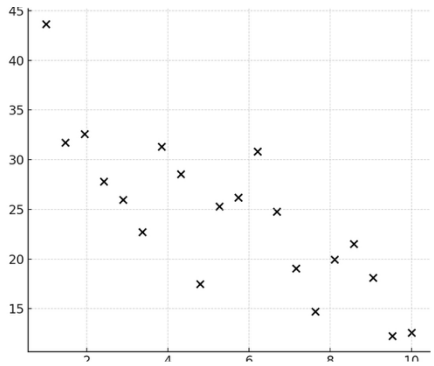
**Line of best fit - by hand (without technology)**

- Plot the scatterplot.
- Use a ruler to draw a straight line that seems to follow the trend of the data.
- Try to "balance" the points so there are roughly as many above the line as below.
- Choose two points on the line you've just drawn, and use them to calculate the equation in the form y = mx + c (like we did in the last section).

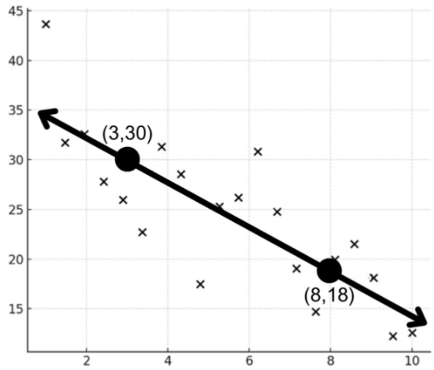


**Worked example.**

**Step 1:** Plot the scatterplot



**Step 2:** Fit the line




**Step 3:** Use two points [(3,30) and (8,18)] to calculate the equation of the line.

$$m = \frac{18-30}{8-3} = \frac{-12}{5} = -2.4$$ ← *Find the gradient first*

$$y - 30 = -2.4(x - 3)$$

$$y - 30 = -2.4x + 7.2$$

$$y = -2.4x + 37.2$$ ← ***Equation of the line***

A calculator or spreadsheet can calculate the line of best fit more precisely.

This is called the least-squares line because it minimises the total squared vertical distances between the data points and the line.

**Line of best fit – Casio Classpad (least-squares)**

1. On the main menu, tap the Statistics icon.
2. Enter your data (all your raw data)
   - Type the x-values into List 1
   - Type the y-values into List 2
3. Tap Calc (top toolbar).
   - Choose regression – linear regression
   - Select List 1 as the X-list and List 2 as the Y-list.
   - Press ok
4. The calculator will display a variety of information.

   **a = gradient**

   **b = y intercept**

   r = correlation coefficient

   $r^2$ = coefficient of determination ←

   Equation of the line $\rightarrow y = ax + b$

*$r^2$ measures how closely the plotted data points lie to the fitted trend line. It can be used to measure the reliability of predictions made.*



**Worked example.** Find the equation of the line of best fit for the following data.

| X | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Y | 12 | 14 | 13 | 16 | 18 | 19 | 20 | 25 | 24 | 30 | 25 | 27 | 33 |

Following the steps in the box above – the screen should show the following:

$a$ = 1.6428571 *← Gradient*

$b$ = 9.7307692 *← Y intercept*

$r$ = 0.958826 *← There is a strong, positive association between the data.*

$r^2$ = 0.9193472 *← 91.93% of the variation in Y-values can be explained by X.*

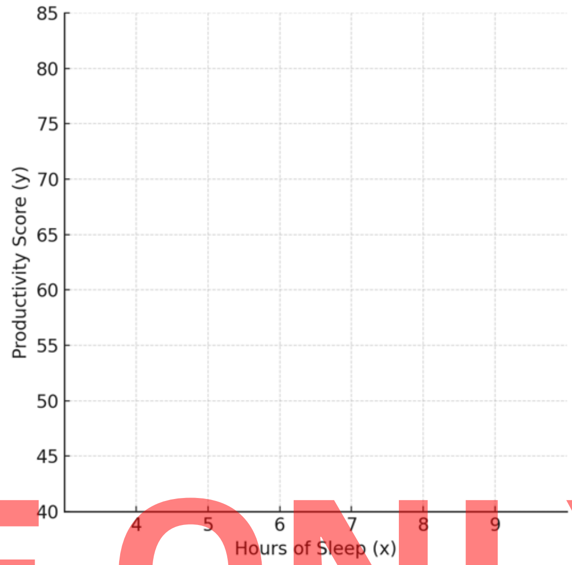
**Final equation:** $\mathbf{y = 1.64x + 9.73}$

Once you've written down the information, hit 'ok' and the calculator will show you the scatterplot with line of best fit drawn on it.

**With your teacher.** For the following data you are going to find the equation of the line of best fit. First by hand, then using your calculator.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| x (Hours of Sleep) | 4 | 5 | 5 | 6 | 6 | 7 | 7 | 8 | 8 | 9 |
| y (Productivity Score) | 45 | 50 | 54 | 56 | 61 | 65 | 70 | 72 | 76 | 79 |

**<u>By hand.</u>**





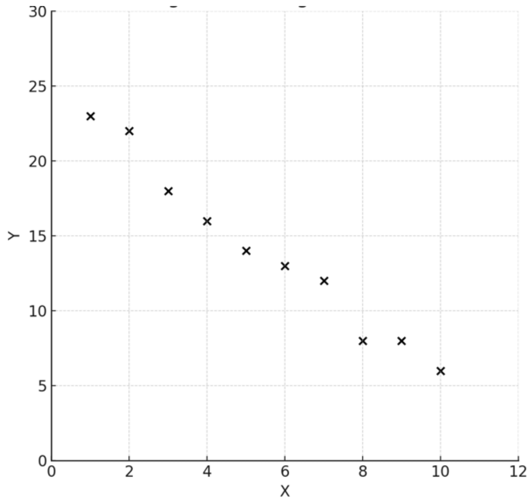

**<u>Using calculator.</u>**

**Interpret, in context, the values of a, b, r and $r^2$ found by your calculator.**

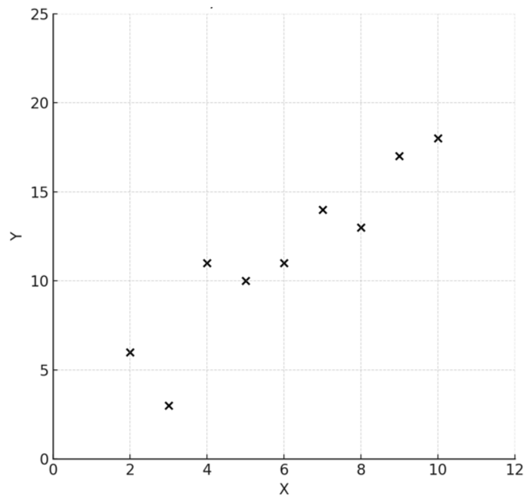
**Try it yourself (LINE OF BEST FIT):** *Ans pg77*

1. For each scatterplot below, mark the line of best fit then find the equation.

a.




b




2. The table shows the number of hours a fridge is left open and the energy used.

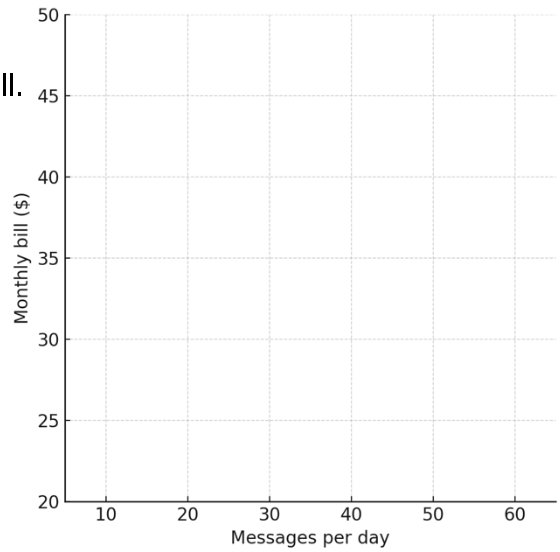
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Hours fridge open | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Energy used (kWh) | 2 | 3 | 5 | 7 | 8 | 10 | 12 |




a. Plot a scatterplot of the data.
b. Draw a line of best fit by eye.
c. Use two points on the line to calculate the equation.
d. From your graph, roughly how much energy do you think would be used if the fridge was open for 2.5 hours?

3. A study records the number of text messages students send per day and their monthly phone bill.

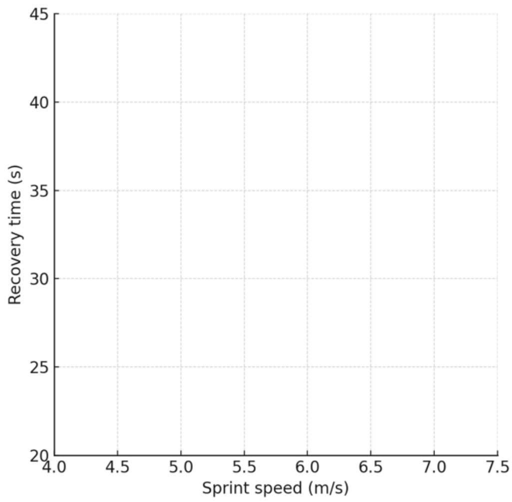
| | | | | | | |
|---|---|---|---|---|---|---|
| Messages per day | 10 | 20 | 30 | 40 | 50 | 60 |
| Monthly bill ($) | 25 | 28 | 32 | 38 | 41 | 46 |




a. Plot the data.
b. Mark a line of best fit.
c. Find the equation of the line of best fit.

4. A coach records players' sprint speed and recovery time.

| Sprint speed (m/s) | 4.5 | 5.0 | 5.5 | 6.0 | 6.5 | 7.0 |
|---|---|---|---|---|---|---|
| Recovery time (s) | 42 | 38 | 35 | 31 | 29 | 26 |




   a. Draw a scatterplot.
   b. What type of correlation exists?
   c. Use two points to find the equation of the line of best fit.

5. The table shows the weight of apples purchased and their cost. Use your calculator to find the equation of the line, then copy a rough sketch of the graph.

| Weight (kg) | 0 | 2 | 4 | 6 | 8 | 10 |
|---|---|---|---|---|---|---|
| Cost ($) | 0 | 6 | 12 | 18 | 24 | 30 |



6. A car rental company records the distance travelled and the total rental cost. Use your calculator to find the equation of the line of best fit.

| Distance (km) | 0 | 100 | 200 | 300 | 400 | 500 |
|---|---|---|---|---|---|---|
| Cost ($) | 50 | 80 | 110 | 140 | 170 | 200 |

7. The number of cups of coffee consumed per day is compared with hours of sleep per night. The data is collected over two nights.

| Coffee (cups) | 0 | 1 |
|---|---|---|
| Sleep (hours) | 8 | 7.5 |

   a. Plot the scatterplot.
   b. Draw a line of best fit.
   c. Find the equation using two points.
   d. Do you this this line of best fit is an accurate representation of the association between cups of coffee and sleep? Why/why not?

8. The data shows the distance travelled by an electric scooter and battery percentage remaining. Use your calculator to find the line of best fit.

| Distance (km) | 0 | 5 | 10 | 15 | 20 | 25 |
|---|---|---|---|---|---|---|
| Battery (%) | 100 | 92 | 85 | 77 | 70 | 62 |

## INTERPRETING AND PREDICTING.

### Interpreting the gradient and y intercept

The equation of the line of best fit has two important parts: the gradient and the Y intercept.

**Gradient / slope, (m)**

- Tells us how much the response variable (y) changes for each unit increase in the explanatory variable (x).
- Example: If the line of best fit for *hours studied* and *test score* is $y = 8x + 40$, the gradient (8) means *each extra hour of study is associated with an 8% higher test score*.

**Y Intercept (c)**

- Tells us the predicted value of y when ($x = 0$).
- Example: In the equation $y = 8x + 40$, the intercept (40) means *a student who studies 0 hours is predicted to score 40%*.
- Sometimes the Y intercept makes sense, sometimes it doesn't. If the intercept gives an unrealistic value (e.g., negative test score), we simply note that it has no meaningful interpretation.



**Worked Example.**

A group of students record their study hours and test scores.

The CAS calculator gives the line of best fit: $y = 7.9x + 42.3$, with a correlation coefficient (r) of 0.95 and a coefficient of determination ($r^2$) of 0.90.

They interpret the data as follows:

- **Gradient (7.9):** For every extra hour of study, test scores increase by 7.9%.
- **Y Intercept (42.3):** A student who does no study is predicted to score 42%.
- **Correlation coefficient ($r = 0.95$):** There is a strong positive linear association between study hours and test scores.
- **Coefficient of determination ($r^2 = 0.90$):** 90% of the variation in test scores is explained by hours studied. The remaining 10% of the variation is due to other factors (natural ability, quality of sleep, distractions, etc.).
- **Context:** The line fits the data strongly ($r = 0.95$), so we can confidently use it to describe and predict scores from study time.

**Making predictions using a line of best fit.**

Once we have a regression line, we can use it to predict values of x and y. These predictions can be helpful — but their reliability depends on where the value of x sits.

**Interpolation.**

- Interpolation is making predictions for values inside the range of your data.
- It is usually reliable, because the model has data on both sides of the value you are predicting.

**Extrapolation.**

- Extrapolation is making predictions for values outside the range of your data.
- It is less reliable, because you are assuming the same trend continues even though you have no data to support it.
- The further outside the range you go, the less reliable the prediction becomes.

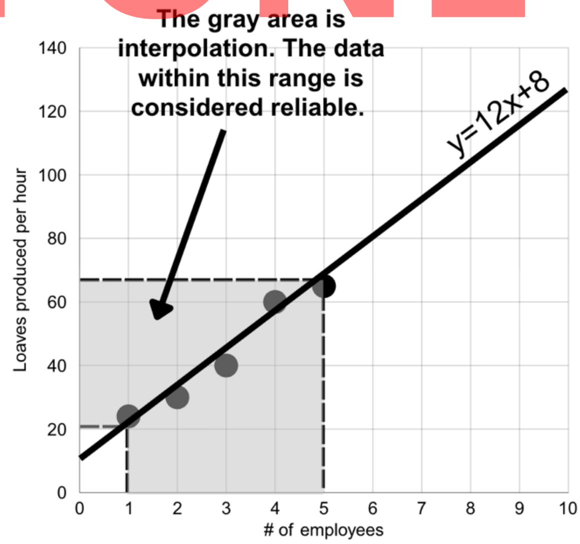


**Worked example.**

A small bakery measures how many loaves of bread it produces per hour depending on the number of employees working. From data collected with 1 to 5 employees, the bakery fits the linear model y = 12x + 8 where;

x = number of employees

y = loaves produced per hour




Predict the output for **4 employees**:

y = 12(4) + 8 = 56 loaves per hour.

This is **within** the collected data range (1–5 employees) so it is interpolation.

It is considered a reliable prediction.

Predict the output for **1000 employees**:

y = 12(1000) + 8 = 12,008 loaves per hour

This is far outside the data range and is considered extrapolation.

This prediction is **not reliable.** This makes sense in context because:

- A small bakery could not comfortably fit 1,000 workers.
- They would be crowded, unable to move, and productivity would drop.

**With your teacher.**

| Temperature (°C) | 15 | 17 | 19 | 21 | 23 | 25 | 27 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|
| Ice creams sold | 72 | 74 | 83 | 89 | 98 | 107 | 116 | 122 | 128 |

Ally owns an ice cream parlour and wants to predict sales throughout summer. She collects the above data over the course of 9 days and finds the line of best fit is $y = 3.89x + 9.74$. Help Ally interpret the results by answering the below questions.

1. What is the Y intercept and what does this represent?

2. What is the gradient and what does it represent?

3. What does it mean if (r) is 0.996 and ($r^2$) is 0.991?



4. Ally wants to predict the number of ice-creams sold if the temperature was 24 degrees.
    a. Is this considered interpolation or extrapolation?
    b. How many ice-creams would be sold?
    c. Is this a reliable estimate?

5. Ally wants to predict the number of ice-creams sold if the temperature was 100 degrees.
    a. Is this considered interpolation or extrapolation?
    b. How many ice-creams would be sold?
    c. Is this a reliable estimate?

**Try it yourself (INTERPRETING AND PREDICTING):** **Ans pg78**

1. The line of best fit for a dataset is given as y = 4x + 12
    a. What is the gradient and what does it mean?
    b. What is the y-intercept and what does it mean?

2. Another dataset gives y = -2.5x + 50 for the line of best fit. Find the value for Y for each of the below values of X.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| X | 0 | 100 | 22 | 90 | 35 | 42 | 1,500 |

3. A line of best fit is found to be y = 4x + 3. Find the value of X for each of the below Y values.

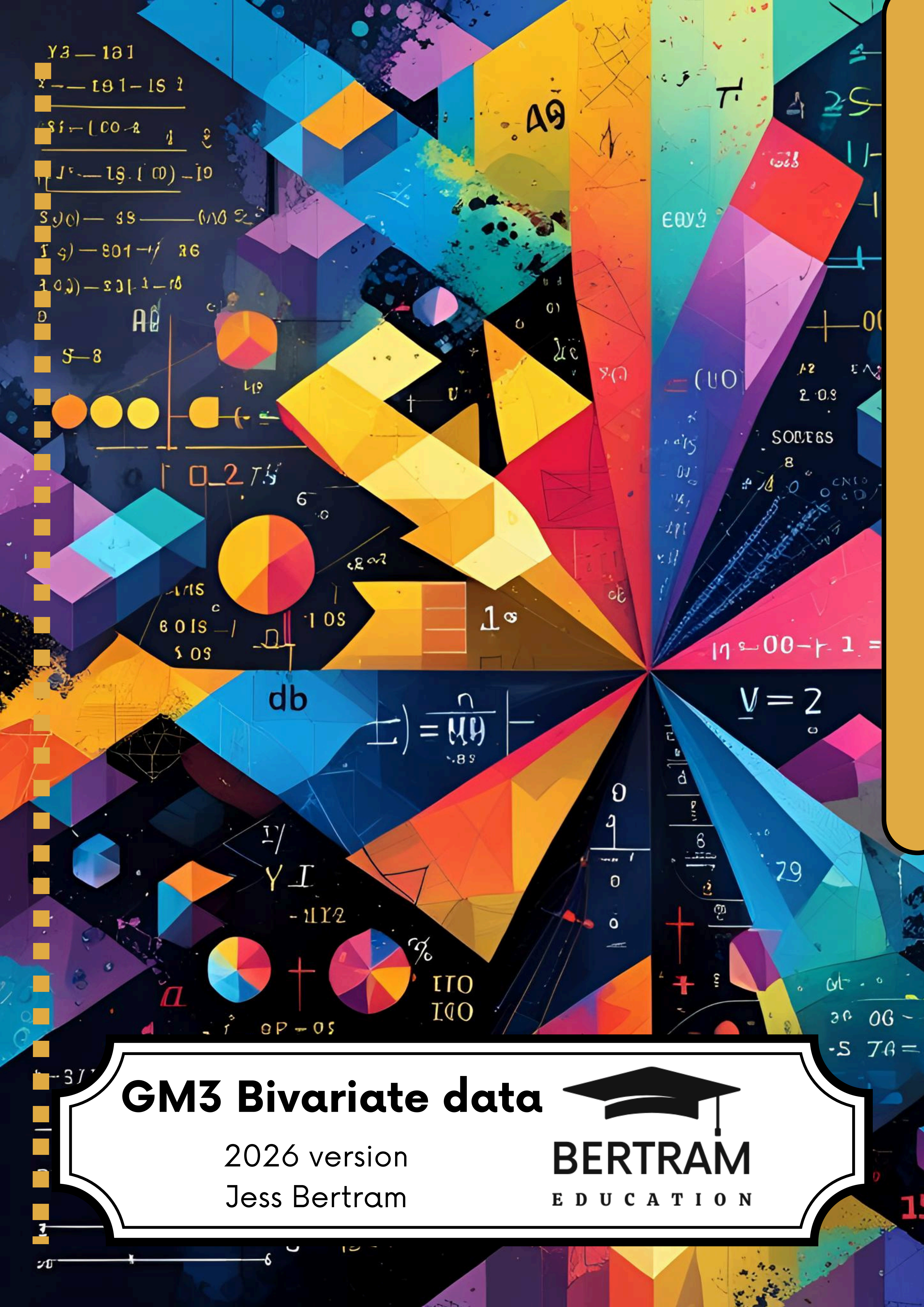
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Y | 7 | 0 | 203 | 83 | 11 | 43 | 19 |

4. A company collects data on the number of advertisements shown per day (x) and the number of sales (y). The line of best fit is found to be: y = 15x + 120
    a. Predict the number of sales if 5 advertisements are shown.
    b. How many sales would you predict if no advertisements are shown?
    c. Explain in context what the Y intercept means.
    d. Explain in context what the gradient means.



5. A line of best fit for temperature (x) and ice cream sales (y) is: y = 3.8x + 10
    a. Predict the sales when the temperature is 22°C. Is this interpolation or extrapolation?
    b. Predict the sales when the temperature is 60°C. Is this interpolation or extrapolation?
    c. Which prediction is more reliable and why?

6. A regression equation is found for hours of sleep (x) and concentration scores (y): y = 7.2x + 45, with ( $r^2$ = 0.88 ).
    a. What does the value of ( $r^2$ ) mean in this context?
    b. What percentage of the variation is *not* explained by the model? What might explain the rest?
    c. Predict the concentration score for someone after they slept 20 hours. Do you think this is realistic?



# GM3 Bivariate data

2026 version

Jess Bertram

BERTRAM EDUCATION