2026

GENERAL MATHEMATICS

Level 2

DATA

General Mathematics: Level 2

GM2 - DATA

By Jess Bertram

With sincere thanks to John Short and Rick Smith

ICON:	MEANING:
	Worked example
	Complete with your teacher
	Try it yourself
	Tips / shortcuts
	Investigation

For ordering, feedback and requests please visit:

BertramEducation.com

GENERAL MATHEMATICS - LEVEL 2

GM2 - DATA

BY JESS BERTRAM

Part 1

INTRO TO DATA

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.

Throughout this workbook, you'll learn how to turn data into meaning — using graphs, tables, and statistical language to describe and compare results.

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 the advertising of 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	“Do 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.
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? 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.

TYPES OF DATA

There are two main type of data – categorical and numerical.

Categorical data	Numerical data
Categorical data consists of values that describe categories or groups. They represent qualities, labels, or names, rather than numbers. Examples: eye colour (blue, green, brown), type of car (sedan, SUV), yes/no responses.	Numerical data consists of values that are numbers and can be measured or counted. They represent quantities. Examples: height (in cm), number of pets, test scores (%).

The type of data determines what can be done with the data – how it can be collected, graphs that can be used to display information, and what techniques can be used to analyse data.

With your teacher.

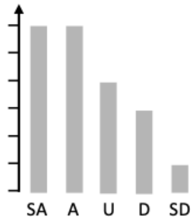
1. Classify each of the following as categorical (C) or numerical (N).
   a. Favourite ice cream flavour
   b. Number of books on a shelf
   c. Blood type
   d. Distance from home to school (km)
   e. Type of music most often listened to
   f. Number of siblings
   g. Height of students in a class (cm)

2. For each of the following survey questions, decide whether the data collected would be categorical or numerical.
   a. A survey asks students what type of pet they own (dog, cat, bird, fish, none).
   b. Students are asked which brand of phone they use (Apple, Samsung, other).
   c. A coach measures the time taken for each student to run 100 metres.
   d. A teacher records the number of books borrowed from the library.
   e. The public was surveyed about their favourite mode of transportation.

CATEGORICAL DATA

Categorical data is data with qualitative categories, not numerical.

Categorical data is often displayed with a bar chart.



The features of a bar chart are:

Categories on the X-axis.

- Each bar represents a category (e.g., favourite sport)
- Categories can be in any order (alphabetical, by size, or logical order).

Frequency (or count) on the Y-axis

- The height of each bar shows how many times that category occurs.
- Sometimes the Y-axis can show percentages instead of counts.

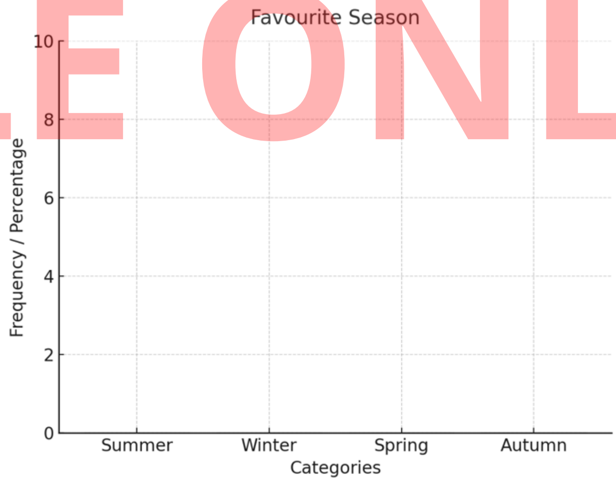
Bars are Separate. There are gaps between bars, all bars are the same width.

Title and Labels on X-axis, Y-axis, and overall.

With your teacher.

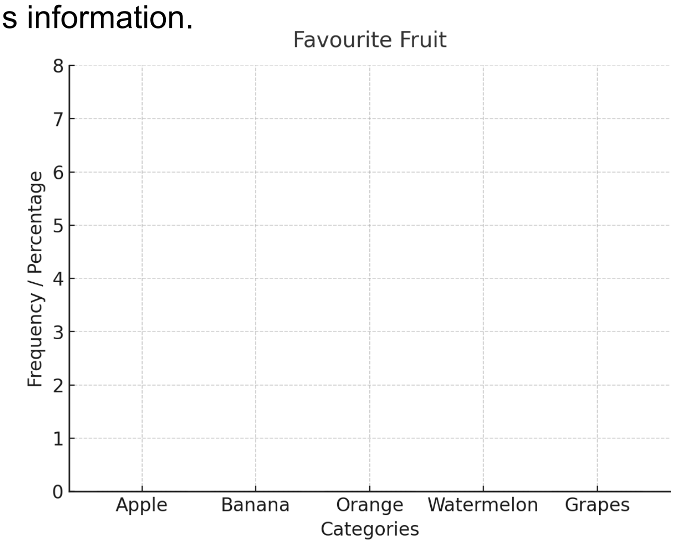
1. A survey asked students about their favourite season. Create a bar graph showing this information.

- 8 students chose summer.
- 5 students chose winter.
- 6 students chose spring.
- 3 students chose autumn.



2. Students were asked about their favourite fruit. The results are shown in the table below. Create a bar graph showing this information.

Favourite fruit	Frequency
Apple	7
Banana	4
Orange	6
Watermelon	3
Grapes	5



Categorical data can be split into two different types - ordinal and nominal.

Ordinal	Nominal
Data comes with an order. Examples: ■ Do you agree with… (disagree, agree, strongly agree, etc) ■ Olympic medals (gold, silver, bronze) ■ Grades (A, B, C, D, F)	Data has no order. Examples: ■ Mode of transport to school (car, bike, bus) ■ Type of fruit in a fruit basket (apple, banana, oranges)

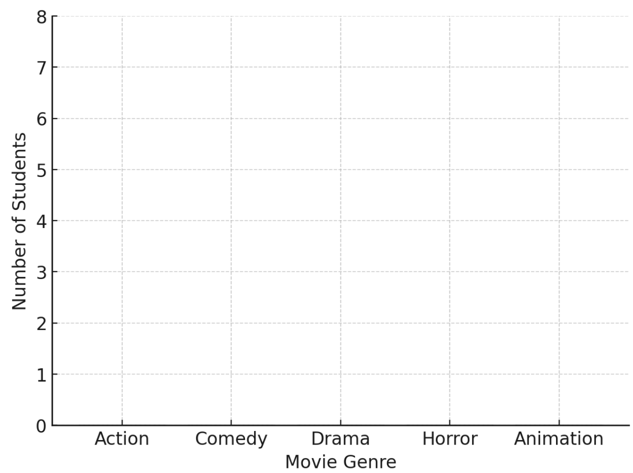
With your teacher.

1. Classify each of the following as nominal (N) or ordinal (O).
    a. Income level (high, medium or low)
    b. Country of birth (Australia, USA, New Zealand etc)
    c. T-shirt size (XS, S, M, L, XL)
    d. Favourite sport (soccer, netball, basketball, tennis)
    e. School year level (Year 7, Year 8, Year 9…)

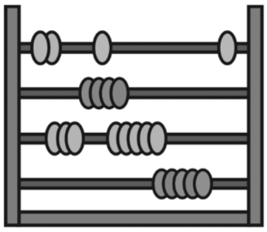
Try it yourself (CATEGORICAL DATA). *Answers page 48*

1. Classify each of the following as nominal (N) or ordinal (O).
    a. Eye colour (blue, green, brown, hazel)
    b. Brand of breakfast cereal
    c. Income level (low, medium or high)
    d. Blood group (A, B, AB, O)
2. 20 students were asked 'what type of movie do you prefer?'. 4 students chose action, 6 chose comedy, 5 chose drama, 3 chose horror and 2 chose animation.
    a. Determine whether the results are ordinal or nominal data
    b. Fill in the table below and use it to create a bar graph from the survey.

Type of movie	Frequency
Action	
Comedy	
Drama	
Horror	
Animation	



NUMERICAL DATA



Numerical data is made up of numbers that represent quantities.

This type of data is useful because it allows us to perform calculations, such as finding totals, averages, or differences.

Numerical data is often displayed in graphs and tables so that patterns and trends can be seen clearly.

Numerical data can be either **counted (discrete)** or **measured (continuous)**.

Discrete data	Continuous data
Can only take specific values, usually whole numbers that can be counted. Examples: ■ Number of siblings ■ Goals scored	Can take any value within a range. Measured, not counted. Examples: ■ Height ■ Temperature

If you can answer with a decimal, the data is usually continuous. Can you have 31.17 computers? No = discrete. Can you have a box 31.17cm wide? Yes = continuous.

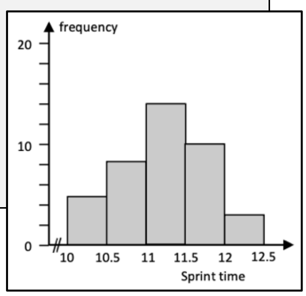
With your teacher.

Classify each of the following as continuous (C) or discrete (D).

a. The number of goals scored in a soccer match.
b. The height of a basketball player.
c. The number of rooms in a house.
d. The temperature outside at midday in degrees Celsius.
e. The time it takes to run 100 metres.
f. The number of cars in a school car park.
g. The weight of a newborn baby.

Continuous numerical data can be turned into a **histogram**, which look like a bar chart except:

- Bars are touching (no gaps)
- X axis is numerical and in ascending order.
- Can show distribution of data (we will cover in part 2)

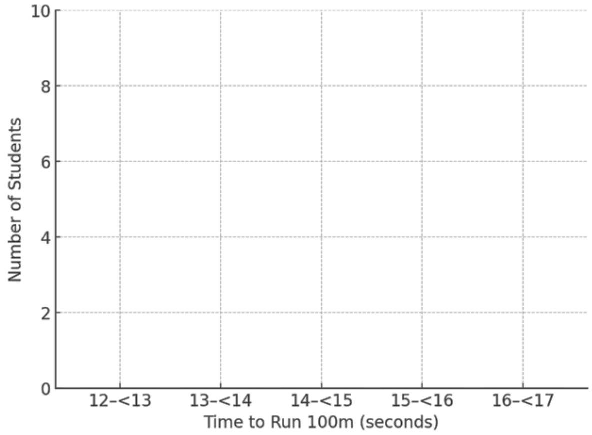


With your teacher.

The athletics team recorded how long it took each student to run 100 metres. Use the data to complete the table below, then turn the grouped data into a histogram.

Student times (in seconds): 12.1, 12.5, 12.9, 13, 13.5, 13.5, 13.6, 13.7, 13.9, 14, 14, 14.2, 14.4, 14.5, 14.7, 14.7, 14.8, 15.12, 15.23, 15.5, 15.7, 15.9, 16, 16, 16.5

Time (grouped)	Frequency
12 to <13 seconds	
13 to <14 seconds	
14 to <15 seconds	
15 to <16 seconds	
16 to <17 seconds	

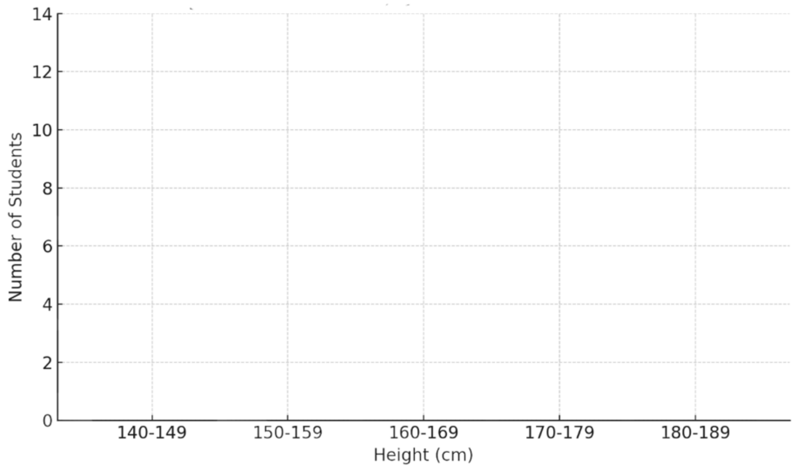


Try it yourself (NUMERICAL DATA): **Ans pg48**

1. Classify the following data as continuous (C) or discrete (D)
   a. The time spent watching TV in one evening
   b. The number of people in a movie theatre.
   c. The volume of water in a glass.
   d. The number of text messages sent in a day.
   e. The speed of a car on the highway.
   f. The shoe size of students in a class.

2. 40 year 10 students were measured (in cm). The results were; 3 students were between 140-149cm, 7 students were between 150-159cm, 12 students were between 160-169cm, 10 students were between 170-179cm, and 8 students were between 180-189cm.
   a. Classify the data as either continuous or discrete
   b. Fill in the below table with all the data, then turn into a histogram.

Height (cm)	Frequency
140-149	
150-159	
160-169	
170-179	
180-189	



Try it yourself (INTRO TO DATA MIX) **Ans pg48**

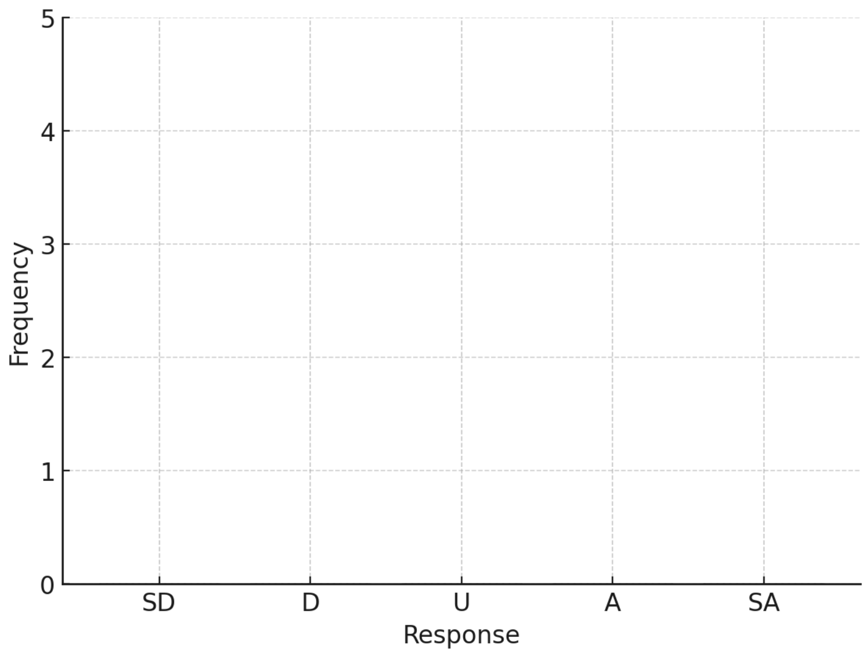
1. Classify the following as either numerical or categorical, then further classify as discrete, continuous, ordinal, or nominal.
   a. A survey asks people their marital status.
   b. A studio audience rates contestants in a talent quest on a scale of 1 to 10.
   c. A fisheries inspector measures the lengths of a fisherman's catch of fish.
   d. The heart rate of a group of athletes is measured.
   e. A graph showing the type of driver's licence held by 18-year-olds is divided into the groups: Learner's, Provisional, Full licence or No licence.
   f. The number of faults made by each tennis player in a final is counted.
   g. A survey asking people their level of agreement with the statement "*Hobart needs a cable car*" has alternative answers: Strongly agree, Agree, Unsure, Disagree, and Strongly disagree.
   h. A survey is made of preferences between different chocolate bars.

2. The data below is the result of a survey in which people were asked to show their level of agreement with the statement. *"Australia needs a new flag"* by choosing between: Strongly agree, Agree, Unsure, Disagree, and Strongly disagree.

   a. How many people were involved in the survey?
   b. What is the frequency of 'Agree'?
   c. What type of data does this survey illustrate?
   d. Turn the data into a table, then into a bar chart.

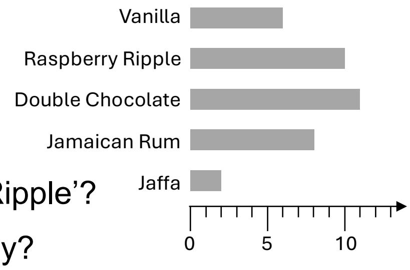
SA	D	SD	U	D
SD	D	U	A	A
SD	U	A	U	D
D	U	D	D	U
U	SA	A	U	SD

Response	Frequency
Strongly disagree	
Disagree	
Unsure	
Agree	
Strongly agree	



Try it yourself (INTRO TO DATA MIX) **Ans pg48**

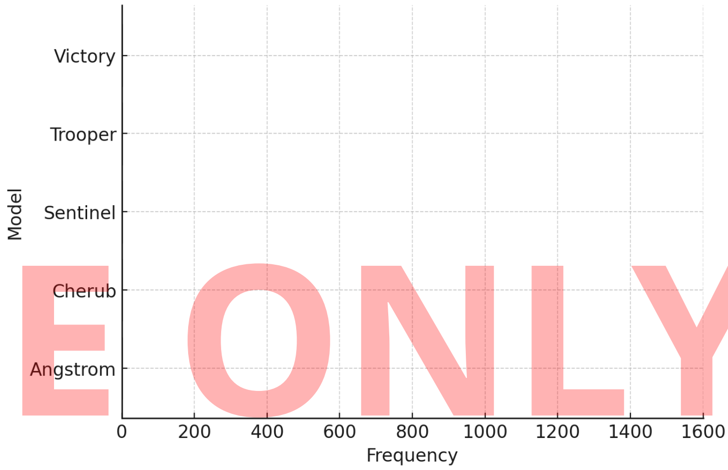
3. The horizontal bar chart shows the popularity of the different ice-cream flavours available at an ice-creamery.



   a. What was the most popular flavour?
   b. What was the least popular flavour?
   c. What was the frequency of purchase of 'Raspberry Ripple'?
   d. What type of categorical data does this survey display?
   e. Turn the bar chart into a table showing the flavours in one column and frequency in the other.

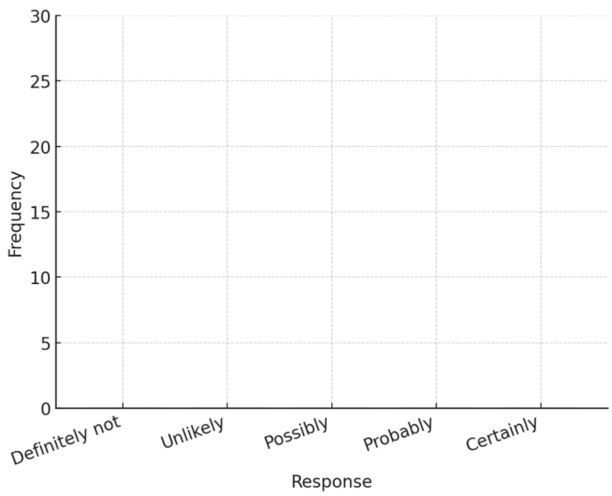
4. The frequency table shows the monthly sales of different models of a brand of car. Represent the data using a horizontal bar graph.

Model	Frequency
Angstrom	850
Cherub	1500
Sentinel	130
Trooper	650
Victory	400



5. The frequency distribution table below shows the results of a telephone poll in which people were asked: *"Would you be prepared to pay money to upgrade your household internet service?"*
   a. Represent the data using a vertical bar graph.
   b. How many people were polled?
   c. The company plans to follow-up those respondents who are at least possibly considering an upgrade. What % of the data is this?
   d. What type of categorical data does this survey display?

Response	**Frequency**
Definitely not	15
Unlikely	25
Possibly	12
Probably	6
Certainly	4



Investigation 1:

You want to know what the most popular social media app is.

Use the statistical investigation process to find out which app is the most popular.

Step 1: Pose a statistical question.

What is the most popular social media app?

Step 2: Collect data.

Find some people to survey and record (tally) their vote.

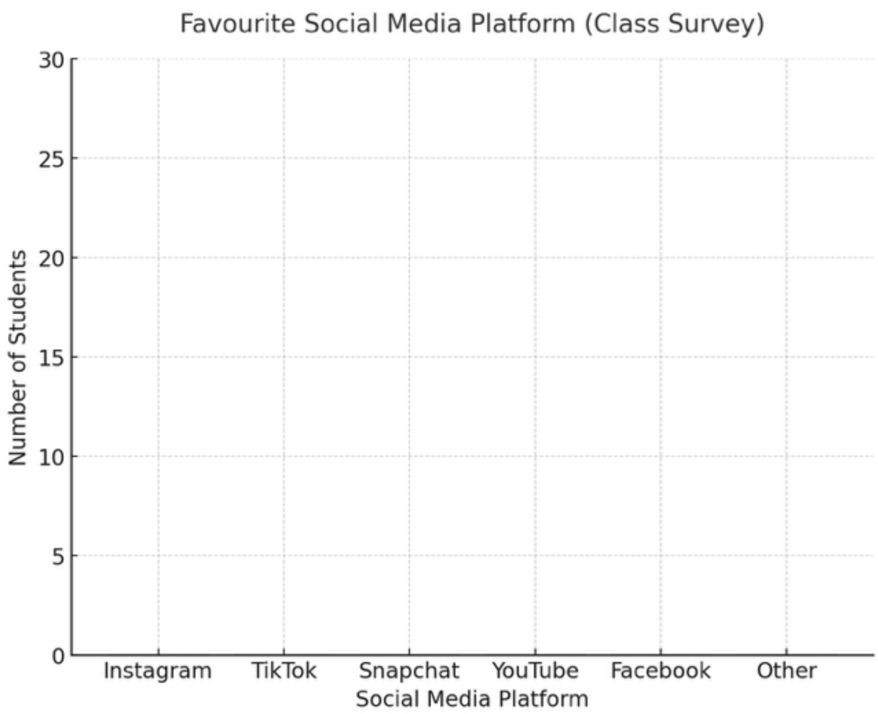
The more people you survey, the more accurate your data and analysis will be.

	Tally	Total
Instagram		
TikTok		
Snapchat		
YouTube		
Facebook		
Other		

Step 3: Organise and analyse the data.

Use your data from step 2 to create your own bar chart.

Use a ruler and colour in the different bars.



Investigation 1 (continued):

Step 4: Interpret the results. Write 3 observations about the data.

You might like to consider the following prompts for your observations:

- *Which platform was the most popular?*
- *Are the differences big or small?*

Step 5: Communicate the findings.

Refer to the original question in step 1. Write a short paragraph summarising what you found, then present your findings to a partner, small group, or class.

You might like to consider the following prompts:

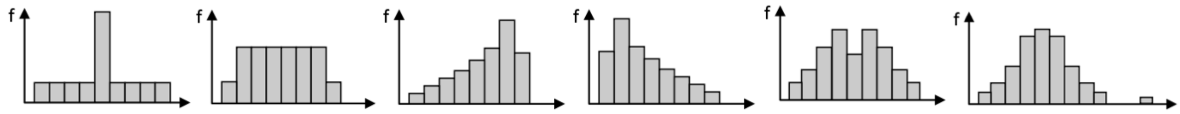
- *Did your investigation answer the question?*
- *Were there any limitations in your data (e.g. small sample size, only surveying classmates, all similar ages, surveying participants in the city vs rural)?*
- *How could this investigation be improved if you repeated it?*
- *How could an information like this be helpful to business?*

Part 2 DISTRIBUTION

DISTRIBUTION

Analysing distribution helps us make sense of large sets of numbers. Instead of looking at hundreds of raw data values, we can describe them quickly using:

- Shape - what the data looks like overall
- Spread - how varied the values are
- Outliers - what doesn't fit the pattern
- Centre - where most values sit



Understanding distribution helps us to:

1. **Summarise messy data**
    - Instead of showing every test score, we can say: *"Most students scored around 65, the data is roughly symmetric, there are a few high scores."*
2. **Compare groups**
    - Example: Compare two basketball teams' scores. One team may be more consistent, while the other has bigger highs and lows.
3. **Spot unusual results**
    - Example: A scientist can see if one measurement is an outlier that doesn't fit with the rest.
4. **Make better decisions**
    - Example: A business notes that most customers spend a small amount, but a few spend a lot. This influences how they market their products.
5. **Prepare for more advanced statistics**
    - Concepts like probability, averages, and standard deviation all rely on understanding distribution first.

Studying distribution helps us turn lists of numbers into clear stories about what the data means in real life.

In this course you will need to be able to describe distribution in terms of:

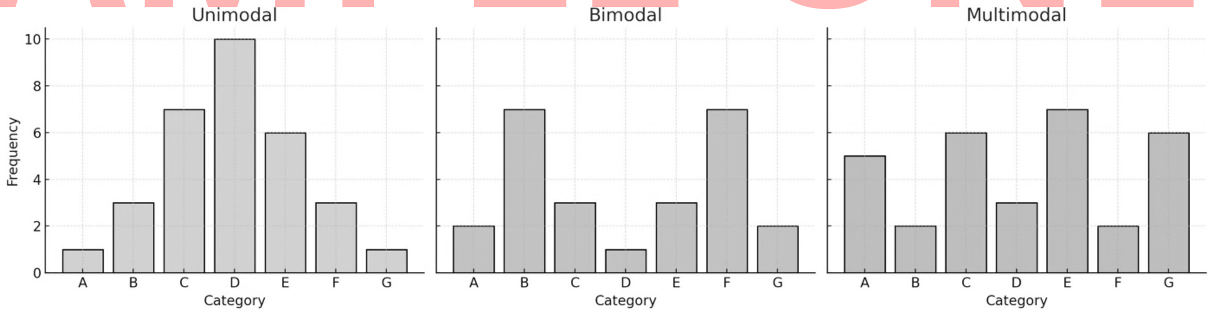
- Modality (unimodal, bimodal, multimodal)
- Shape (symmetric, positively skewed, negatively skewed)
- Spread (range)
- Outliers (extreme or unusual values)
- Location (centre - mode, mean, median) *Median is covered in part 3.*

MODALITY

When we look at a dataset, we often notice that the values "cluster" around one or more points. The number of clusters, or peaks, in a distribution is referred to as modality. Modality refers to "peaks" or "spikes" in the data.

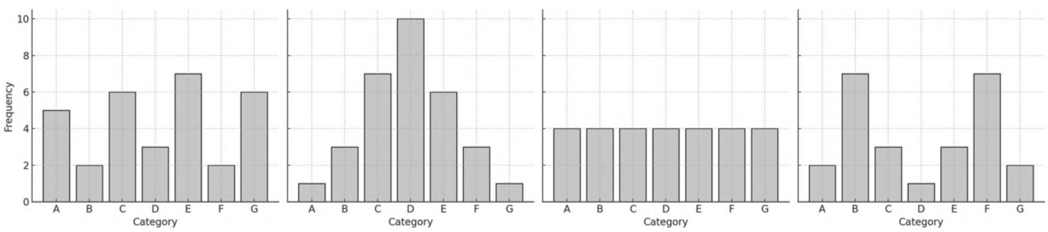
- **Unimodal** → one peak
- **Bimodal** → two peaks
- **Multimodal** → more than two peaks
- **Uniform** → no peaks

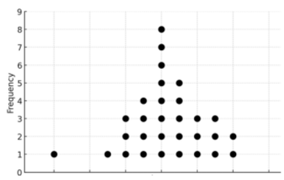
The peaks don't need to be the same height to be considered bimodal or multimodal. Modality is useful because it gives us a quick sense of the overall pattern in the data.

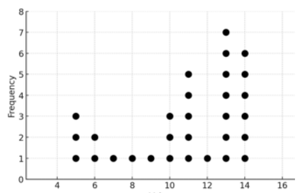


With your teacher.

Describe each of the following graphs in terms of modality.







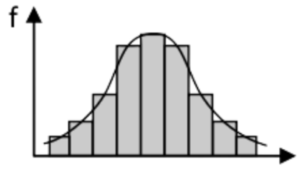
1	0, 0, 2, 2, 3, 4, 5, 6, 7, 8, 8, 9
2	0, 0, 1, 2
3	0, 1, 4, 6
4	1, 3, 5, 6, 6, 7
5	0, 3, 5, 6, 6, 6, 6, 7, 8, 9, 9, 9
6	3, 3, 3, 4, 5, 6, 7
7	2, 5, 7, 8

SHAPE

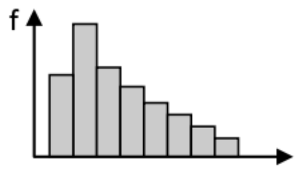
When we graph numerical data, the overall shape of the distribution tells us how the values are spread. Shape describes whether the data is balanced (symmetric) or pulled to one side (skewed).

The most important shapes to recognise are:

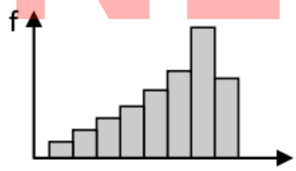
- **Symmetric**
  - The left and right sides of the graph look about the same. The peak is in the middle.
  - Example: Heights of people in a large group often form a roughly symmetric "bell" shape.



- **Positively Skewed (Right-Skewed)**
  - Most values are at the lower end, with a long tail stretching to the right.
  - Example: Hospital wait times. Most patients are seen fairly quickly, but a few cases take a long time, creating a long right tail.



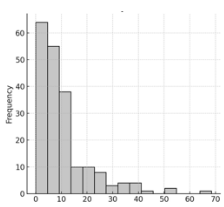
- **Negatively Skewed (Left-Skewed)**
  - Most values are at the higher end, with a long tail stretching to the left.
  - Example: Retirement age — many people retire around 65, but a few retire much earlier.

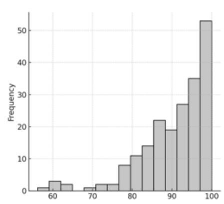


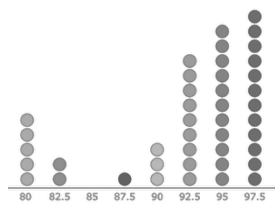
Shape helps us quickly see patterns in the data. It can reveal whether most values are typical, or if there are extreme values pulling the data to one side. This affects how we choose measures of centre (mean vs median) and spread.

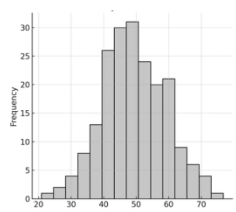
With your teacher.

Describe each of the following graphs in terms of shape.







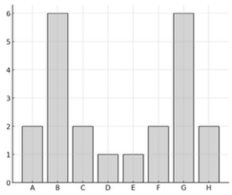
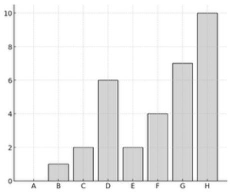
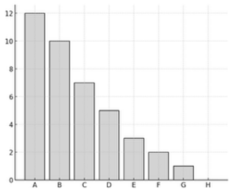
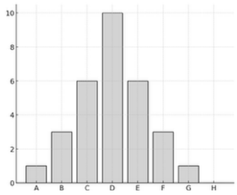


10	0, 0, 2, 2, 3, 4, 5, 6, 7, 8, 8, 9
20	0, 0, 1, 2, 5, 6, 7, 8, 8, 9
30	0, 1, 4, 6, 7, 7, 7, 7, 8
40	1, 3, 5, 6, 6, 7
50	0, 3, 5, 6
60	3, 3, 3
70	2, 5

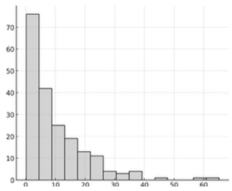
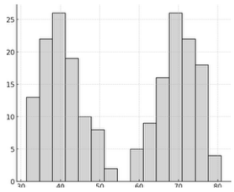
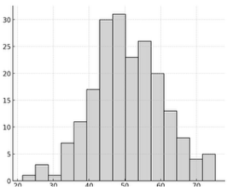
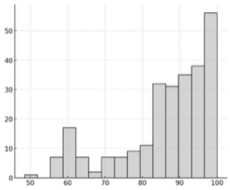
11	0, 0, 2, 2
12	0, 0, 1, 2, 5
13	0, 1, 4, 6, 7, 7
14	1, 3, 5, 6, 6, 7
15	0, 3, 5, 6
16	3, 3, 3
17	2, 5

Try it yourself (MODALITY AND SHAPE) **Ans pg48**

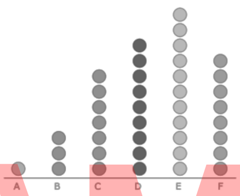
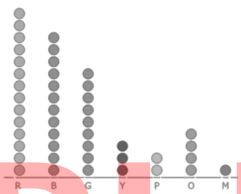
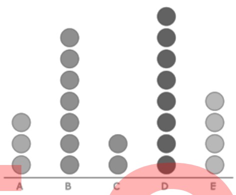
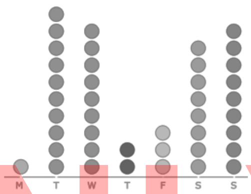
1. Describe each of the following bar charts in terms of modality and shape.

a

b

c

d


2. Describe each of the following histograms in terms of modality and shape.

a

b

c

d


3. Describe each of the following dot plots in terms of modality and shape.

a

b

c

d


4. Describe each of the following stem plots in terms of modality and shape.

a

15	0, 0, 2, 2, 4, 4, 5, 5, 6,
16	0, 0, 1, 2, 5, 6, 7, 7
17	0, 1, 4, 6, 7, 7
18	1, 3, 5, 6,
19	0, 3, 5, 6
20	3, 3, 3
21	2, 5

b

15	0, 0, 2, 3, 3, 6, 6, 7
16	0, 0, 1, 2, 5, 6, 7, 7
17	0, 1, 4, 6, 7, 7
18	1, 3, 5, 6,
19	0, 3, 5, 6, 7
20	3, 3, 3, 3, 3, 6, 6, 7
21	2, 5, 7, 7, 7, 8, 8, 8, 9

c

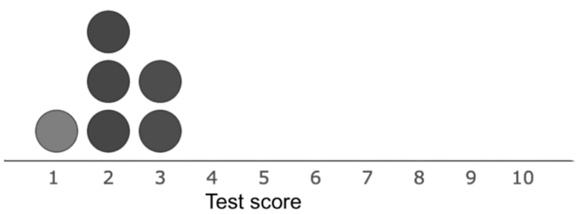
15	0, 0, 2, 3, 3, 6, 6, 7
16	0, 0, 1, 2, 5, 6, 7
17	0, 1, 4, 6, 7
18	1, 3, 5, 6, 7, 7, 8, 8, 9
19	0, 3, 5, 6
20	3, 3, 3, 3, 3, 6, 6
21	2, 5, 7, 7, 7, 8, 8, 8, 9

d

15	0, 0, 2
16	0, 0, 1, 2, 5
17	0, 1, 4, 6, 7, 9, 9
18	1, 3, 5, 6, 7, 7, 8, 8, 9
19	0, 3, 5, 6, 7, 7, 8, 9
20	3, 3, 3, 3, 3, 6
21	2, 5, 7, 7

5. A teacher records the scores for 50 students on a maths test.
   a. Create a dot plot of the data (the first 3 scores have been done for you)
   b. Describe the distribution (modality and shape)
   c. What does the distribution tell you about how students went on the test?

Score (x/10)	Frequency
1	1
2	3
3	2
4	3
5	6
6	7
7	8
8	9
9	6
10	5



SPREAD

Spread tells us how far apart the data values are. Two datasets can have the same mean (centre) but be very different in how spread out the values are.

- If the spread is small, the data values are bunched up close together.
- If the spread is large, the data values are scattered widely.

Spread is important whenever we want to know about consistency or reliability:

- Sport: Two athletes might both average 20 points per game. The one with smaller spread is more consistent.
- Science: If repeated measurements have a small spread, the experiment is reliable.
- Business: A shop with steady daily sales (small spread) is easier to plan for than one with unpredictable ups and downs (large spread).

We will go further into spread in part 3 of this workbook. For now, the only measure of spread you need to learn is **range.**

> Range is the difference between the largest and smallest value.
>
> $$Range = maximum\ value - minimum\ value$$

Worked Example.

A teacher recorded the number of texts 6 students sent during class: 5, 8, 7, 10, 6, 12

Maximum = 12

Minimum = 5

Range = 12 – 5 = 7

Interpretation: The number of texts varied by 7 messages between the least and most active students. This shows the data is moderately spread out.

With your teacher. Find the range of the below data sets.

a. 2, 4, 6, 9, 10, 15

1, 3, 6, 11, 42, 12

45, 13, 14, 23, 15

22, 22, 22, 22, 22

5, 7, 1, 0, 56, 3, 0

b.

Number of pets	Frequency
0	8
1	7
2	9
3	4

c.

Stem	Leaf
10	1, 2, 3, 4
11	2, 3
12	1, 3
13	3, 6, 8
14	3, 2, 5
15	0

Key 10/1 = 101

LOCATION (CENTRE)

In statistics, an important way to describe a dataset is to identify its centre. This tells us where most of the values are clustered. This is sometimes called the location of the distribution.

The most common measures of centre are:

- Mode → The most common value. There can be multiple modes.
- Median → the middle value when the data is ordered. (covered later in part 3)
- Mean → the average of the values.
- Standard deviation → the average distance from each score to the mean

Calculating the mean (average)

To find the mean, add all the scores together and divide by the number of scores.

This method can be summarised by the formula:

$$\bar{x} = \frac{\Sigma x}{n}$$

Where:

$\bar{x}$ = symbol for mean

Σ = Greek letter sigma, means ‘the sum of’ or to add everything together

n = Number of scores

Worked example.

A class of 11 students were asked ‘how many computers do you have at home?”

The survey revealed the following results.

0, 1, 1, 2, 6, 1, 3, 2, 2, 4, 0

$$\bar{x} = \frac{(0+1+1+2+6+1+3+2+2+4+0)}{11} = \frac{22}{11} = 2$$

The mean of the data is 2.

With your teacher.

Find the mode(s) and mean of the following data.

a. 1, 3, 4, 5, 6, 5, 4, 4, 6, 7

b. 10, 25, 20, 33, 25, 26, 14, 28, 24

c. 3, 2, 3, 4, 1, 4, 5, 3, 2, 1, 3, 4, 1, 558 (what do you notice about this?)

Calculating mean from a table

If there is a large amount of data, it may be helpful to use a frequency table to calculate the mean. There is a formula for large groups of data:

$$\bar{x} = \frac{\Sigma fx}{\Sigma f}$$

Where:

Σfx = Total of all scores, taking frequency into account

Σf = Total of all frequencies (same as ‘n’ on the last page)

Worked example.

A tuna fishing charter company keeps records of the number of fish it catches on each charter for a season. The results are shown in the table below.

Number of tuna (x)	Frequency (f)	
0	18	← There were 18 charters that caught 0 tuna
1	12	
2	6	
3	2	
4	0	← No charters caught 4 tuna
5	2	
6	5	← 5 charters caught 6 tuna each

To calculate the mean from a frequency table, add a column for frequency multiplied by score (**fx**) and calculate the totals of each column (**Σf** and **Σfx**).

Number of tuna (x)	Frequency (f)	fx	
0	18	0	← $0 \times 18 = 0$
1	12	12	
2	6	12	
3	2	6	← $3 \times 2 = 6$
4	0	0	
5	2	10	← $5 \times 2 = 10$
6	5	30	
Totals (Σ)	**Σf = 45**	**Σfx = 70**	← Find the total of each column (Σ)

$$\bar{x} = \frac{\Sigma fx}{\Sigma f} = \frac{70}{45} = 1.56$$

Interpret in context: The average number of tuna fish caught per charter was 1.56

With your teacher.

1. A survey was done on how far students travelled to school (in km) in one week:

Distance (km)	Frequency (f)	fx
2	6	
3	8	
5	10	
6	7	
7	5	
8	12	
9	6	
60	1	

a. Find the average distance travelled.

b. Which value makes the mean misleading as a description of a "typical" student?

2. A teacher surveyed a class of students about how much pocket money they receive per week. The results are as follows:

Pocket money ($)	Frequency (f)	fx
5	8	
10	10	
15	12	
20	15	
25	6	
50	3	
100	1	

a. Find the mean pocket money.

b. Do you think the mean is a good description of what most students receive? Explain.

The mean can also be calculated from a range of graphs.

Calculating the mean from a stem and leaf plot or dot plot are straightforward, as these plots list the data individually.

Calculating mean from a stem and leaf plot.

Treat each piece of data as individuals, then add all the scores together and divide by the number of scores.

Worked example.

Karen did a survey of the ages of all her relatives. She presented the ages in the below stem and leaf plot.

The Key is the most important part of a stem and leaf plot. Without it, the data makes no sense.

Ages of relatives - Key 1/5 = 15 years old

The stem shows the 'tens' and the leaves show the "ones"

Stem	Leaf	
1	5, 7, 8	← This reads 15, 17, 18
2	0, 2, 3, 5	
3	1, 4, 6	← This reads 31, 34, 36
4	0	

From the stem and leaf plot, the data values are:

15, 17, 18, 20, 22, 23, 25, 31, 34, 36, 40

$$\bar{x} = \frac{(15 + 17 + 18 + 20 + 22 + 23 + 25 + 31 + 34 + 36 + 40)}{11} = \frac{281}{11} = 25.6$$

The average age of Karen's relatives is 25.6 years old.

With your teacher.

Find the mean of both below sets of data.

Stem	Leaf
2	3, 4, 4, 5, 6
3	1, 2
4	0, 2, 4
5	1, 3, 3
6	0, 0, 5, 7, 8
7	1, 4, 5

Key 2/3 = 2.3

Stem	Leaf
10	1, 2, 3, 4
11	2, 3
12	1, 3
13	3, 6, 8
14	3, 2, 5
15	0

Key 10/1 = 101

Calculating mean from a dot plot.

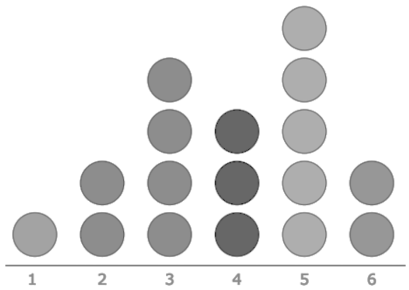
You can turn a dot plot into a table then calculate the mean.

Or. multiply each score (x) by the number of dots in that column (f) to find the frequency of x (fx). Divide the total of fx (Σfx) by the total number of dots (Σf)

Worked example.

A group of students conducted a survey about how many pets were in each household. They recorded the results on the dot plot.

Find the average number of pets per household.



To find the mean, the dot plot could be written as a table:

Score (x)	Frequency (f)	fx
1	1	1
2	2	4
3	4	12
4	3	12
5	5	25
6	2	12
Total (Σ)	**17**	**66**

$$\bar{x} = \frac{\Sigma fx}{\Sigma f}$$

$$\bar{x} = \frac{66}{17}$$

$$\boldsymbol{\bar{x} = 3.88}$$

Or put straight into the equation.

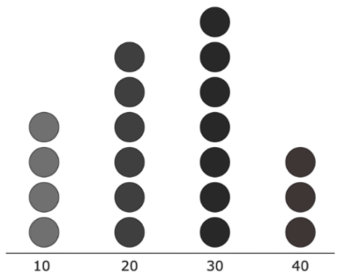
$$\bar{x} = \frac{(1 \times 1) + (2 \times 2) + (3 \times 4) + (4 \times 3) + (5 \times 5) + 6 \times 2)}{17} = \frac{66}{17} = \mathbf{3.88}$$

Interpret in context: The average number of pets per household is 3.88

With your teacher.

The dot plot shows the number of minutes a class of 20 students spent on homework last night.

Calculate the average of the data.



Interpret in context: Students spent ___ minutes on average on homework last night.

Try it yourself (RANGE, MODE AND MEAN): **Ans pg49**

1. Calculate the range, mode and mean of the following sets of data:
   a. 1, 3, 4, 5, 6, 4, 7, 5, 3, 6, 4, 5, 7
   b. 12, 15, 18, 20, 14, 16, 17, 19, 13, 15, 18, 14, 16, 20
   c. 4.3, 6.5, 5.2, 7.8, 4.9, 6.1, 5.6, 7.3, 6.8, 5.0, 6.2
   d. 217, 349, 285, 301, 278, 320, 295, 260, 330, 310, 289, 275
   e. 8, 12.5, 14, 10, 9.5, 15, 13, 11.5, 12, 14.5, 10.5, 13.5

2. Calculate the range, mode and mean from the following tables

a.

x	f
1	3
2	5
3	10
4	4
5	6
6	9
7	10

b.

x	f
4	0
5	2
6	0
7	3
8	6
9	5
10	10

c.

x	f
10.5	4
11	3
11.5	7
12	5
12.5	6
13	1
13.5	5

3. Calculate the range and mean from the following stem and leaf plots

a. Key: 1/0 = 10

1	0, 1
2	1, 2, 5
3	6, 7, 9, 9
4	1, 3, 5, 6, 7
5	0, 3, 5
6	3, 3
7	2

b. Key: 1/1 = 1.1

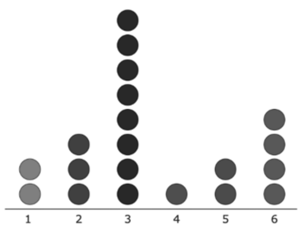
1	1, 4, 6
2	1, 3, 3, 6
3	0, 0, 4
4	7, 8
5	0
6	4, 3
7	1

c. Key: 24/3 = 24,300

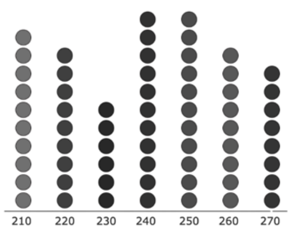
24	3
25	4, 5
26	0, 6, 9
27	8, 9
28	3
29	
30	1

4. Calculate the range, mode and mean from the following dot plots

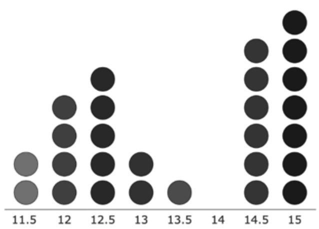
a



b



c



5. A cricketer records his batting results for 7 consecutive innings as follows: 19, 58, 62, 67, 18, 55, 34. What will the cricketer need to score in his next average if he wishes to have a batting average of 50?

6. A student has completed 5 maths tests with the following results: 72, 65, 80, 91, 77. The student wants their average after 6 tests to be at least 85. What is the minimum score they must get on the 6th test to achieve this goal?

STANDARD DEVIATION (SD)

Standard deviation is a measure of spread. It tells us, on average, roughly how far the data values are from the mean.

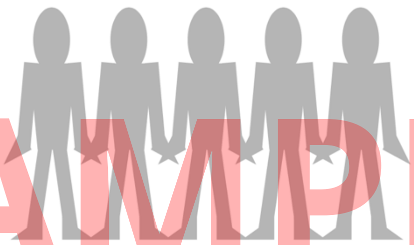
- A small SD means the data is tightly clustered around the mean (consistent).
- A large SD means the data is more spread out (less consistent).

Standard deviation formula (from a table)

$$SD = \sqrt{\frac{\Sigma f x^2}{\Sigma f} - \left(\frac{\Sigma f x}{\Sigma f}\right)^2}$$

Worked example.

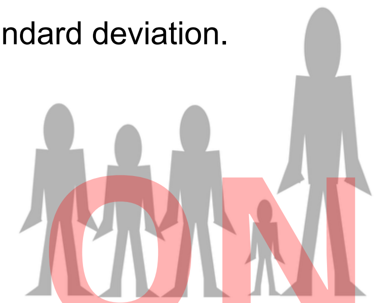
Two basketball teams consist of 5 people each. Both teams have the same average height (173cm) but look different. Calculate the standard deviation.



Team A

Smaller standard deviation.
All values (heights) are closer to the average value (height)

H = 172, 175, 174, 172, 172 (cm)



Team B

Larger standard deviation.
Individual values (heights) are further away from the average height.

H = 170, 165, 170, 140, 220 (cm)

Calculating standard deviation for team A.

Height (x)	Frequency (f)	x^2	fx	fx^2
172	1	29,584	172	29,584
175	1	30,625	175	30,625
174	1	30,276	174	30,276
172	1	29,584	172	29,584
172	1	29,584	172	29,584
Σ	**5**	**149,653**	**865**	**149,653**

$$SD = \sqrt{\frac{\Sigma f x^2}{\Sigma f} - \left(\frac{\Sigma f x}{\Sigma f}\right)^2}$$

$$SD = \sqrt{\frac{149{,}653}{5} - \left(\frac{865}{5}\right)^2}$$

$$\boldsymbol{SD = 1.26}$$

Interpretation in context: The standard deviation of 1.26 means that, on average, most player's heights are 1.26cm away from the team's average height (173cm).

With your teacher.

Calculate the standard deviation of team B.

Height (x)	Frequency (f)	x^2	fx	fx^2
170	2			
165	1			
140	1			
220	1			
Σ	**5**			

$$SD = \sqrt{\frac{\Sigma f x^2}{\Sigma f} - \left(\frac{\Sigma f x}{\Sigma f}\right)^2}$$

In practice, standard deviation is usually calculated with technology.

You can use a CAS calculator, or an online calculator like Desmos or GeoGebra.

Free online calculators for standard deviation
Desmos scientific https://www.desmos.com/scientific → Function → stdevp → stdevp(172,175,174,172,172) = 1.26
GeoGebra Scientific calculator https://www.geogebra.org/scientific → Keyboard (bottom left) → f(x) → stdevp(172,175,174,172,172) = 1.26

Limitations of standard deviation

Standard deviation is a useful measure of spread, but only as part of a bigger picture.

- SD is heavily influenced by outliers and extreme values
- SD works best with a symmetric distribution; it may be misleading when used on skewed data.
- SD only makes sense when considered alongside the mean, not on its own.

Try it yourself (STANDARD DEVIATION) **Ans pg49**

1. Calculate the standard deviations for each of the below tables.

a.

x	f	x^2	fx	fx^2
20	5			
25	10			
35	6			
40	7			
Σ				

b.

x	f	x^2	fx	fx^2
4	12			
5	14			
7	6			
8	2			
Σ				

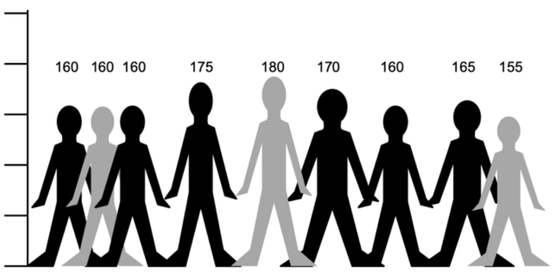
2. Use an online calculator like Desmos or GeoGebra to find the SD of this data:
   a. 12, 15, 14, 12, 11, 16, 17, 13, 18, 15, 15, 16
   b. 42.5, 47, 49.3, 50.55, 48.2, 45.7, 51.6
   c. 100, 112, 110, 101, 105, 113, 110, 110, 110, 42

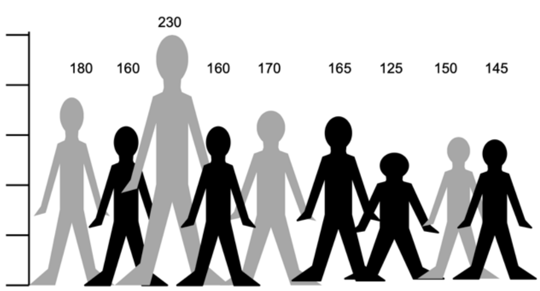
3. The table below shows the number of jellybeans in randomly chosen packets produced on an automatic packing machine. The company insists the process is consistent, with a standard deviation of less than 1 jellybean per packet.

# per pack (x)	f	x^2	fx	fx^2
20	5			
25	10			
30	4			
35	6			
40	7			
Σ				

   a. Complete the table and calculate the mean and standard deviation.
   b. Is the company correct in saying the standard deviation is less than 1?
   c. Comment on any limitations of using SD alone in this context.

4. For each of group of people pictured below (group A on the left, B on the right):
   a. Find the mean, mode, range and standard deviation for both groups
   b. If the tall person from group B was to stand in group A, how would the statistics of group A be affected?





INVESTIGATION 2 – FACTORY WAGES

A bitter dispute has broken out at a local factory.

- The CEO is running advertising claiming that employees are generously paid, with an average salary of over $40,000 per year.
- The Union, representing the factory workers, argues that this is dishonest - most workers are poorly paid, and the CEO is using misleading statistics to make the factory look good.

Both sides are determined to prove their case using the same set of salary data. It's your job to analyse the data and then sharpen your pitchforks - either as the CEO's marketing team or the Union's campaign team.

PART 1 – crunch the numbers *(part 1 can be done individually or in pairs)*

The salaries for all employees are listed below.

Employee	Salary	Frequency	x^2	fx	fx^2
Worker	$15,000	30			
Supervisor	$18,000	4			
Manager	$75,000	6			
CEO	$850,000	1			
Σ					

Calculate the mean, mode, range, and standard deviation of the salaries.

Mean	Mode	Range	SD

Is the CEO lying about the average salary?

DATA

2026 version

Jess Bertram

BERTRAM
EDUCATION