Tool 3: Mathematics

IB Biology - Revision Notes

Applying General Mathematics in Biology

Applying General Mathematics in Biology

- Biology often requires the use of calculations, which can include
  - **Decimals**
    - Most biological calculations use decimals, e.g. calculating the size of a bacterial cell
  - **Fractions**
    - Most scientific calculators will initially give answers as fractions
      - Make sure you know where the S⇔D button is so that you convert the fraction into a decimal
  - **Percentages**
    - There are many percentage calculations, including percentage change and percentage difference
  - **Ratios**
    - The most common ratio requiring understanding is that of surface area to volume ratio
  - **Proportions**
    - Proportionality can be used to understand quantity and scale and is important in biology in topics such as cell biology when creating biological drawings of cells and tissues from a microscope image or micrograph
  - **Frequencies**
    - This is most commonly used in understanding change in allele frequency
  - **Densities**
    - We often look at and examine population density in ecology or stomatal density in plant biology
  - **Approximations**
    - This is used to obtain an approximate value for example when using the magnification formula
  - **Reciprocals**
    - We frequently used reciprocals (1/n) when dealing with concentration versus rate graphs, using 1/T where T is time

Measures of central tendency

- Measures of central tendency involve calculations of **mean, median and mode** which you should be able to apply to a range of scenarios and contexts
  - **Mean**
    - The mean is an **average of a group of numbers** calculated by totaling all values and dividing by the number of values
    - Mean is used to summarise a dataset with a single number which represents the data's typical value

- **Median**
  - This is the **middle number** which can be found by ordering all values and picking out the one in the middle
  - It helps us to understand that 50% of values have are smaller or equal to the median and 50% of values are higher or equal to the median
- **Mode**
  - This is the **most frequent value** in a dataset
  - It can be useful to understand the most common value in categorical data when the mean and median can't be used

Measures of dispersion

- Measures of dispersion involve applying calculations of **standard deviation (SD), standard error (SE) and interquartile range (IQR)** to a range of contexts
- These ideas are also considered here with reference to the use of error bars on graph
  - **Standard Deviation**
    - The mean is a more informative statistic when it is provided alongside standard deviation
    - Standard deviation **measures the spread of data around the mean** value
      - It is very useful when comparing consistency between different data sets
    - The mean must be calculated before working out the standard deviation
  - **Standard Error**
    - Standard error of the mean **measures how far the mean of the data is likely to be from the true mean**
    - It measures the accuracy with which a sample represents a population
    - The SE is always smaller than the SD
  - **Interquartile Range**
    - This is another method of analysing dispersion of data
    - It is the **difference between the 75th and 25th percentiles of the data**
      - Quartiles are the values that divide the whole series into four equal parts

Scientific notation

- **Scientific notation** is also known as **standard form**
- It is a system of writing and working with very large or very small numbers
- Numbers in scientific notation are written as:

$$\mathbf{a \times 10^n}$$

- They follow these rules:
  - **a** is a number above 1 and below 10
  - For large numbers, **n** is an integer that is greater than 0
    - i.e It shows how many times **a** is multiplied by 10
  - For small numbers, **n** is an integer that is less than 0
    - i.e It shows how many times **a** is divided by 10

- n < 0 for small numbers i.e how many times a is divided by 10

Approximation and estimation

- Approximation and estimation are both methods used to **obtain values that are close to the true or accurate values**
  - While they share some similarities, they have distinct characteristics and are used in different contexts

Approximation

- Approximation involves finding a value that is close to the actual value of a quantity
  - It may not necessarily be very precise or accurate
- It is often used when an exact calculation is challenging or time-consuming and a reasonably close value is sufficient

Estimation

- Estimation involves making an educated guess or assessment based on available information or data
- It is used when the true value of a quantity is unknown or cannot be directly measured
  - For example biologists estimate dates of the first living cells and the last universal common ancestor or the method of estimating times by use of the "molecular clock"

Scales of magnification

- Magnification is an important skill used widely in biology and frequently assessed in examinations
- For more information and worked examples see our revision note on microscope skills

Rates of change

- The rate of change tells us **how something changes over time**
  - For example oxygen consumption in germinating seeds over a period of days
- To determine rates of change from tabulated data, you can use the average rate of change or gradient, if the data has been plotted as a graph
- The average rate of change between two points on a graph or in a table is:

$$\text{Rate of change} = \frac{\text{Change in the dependent variable}}{\text{Change in the independent variable}}$$

Proportionality and correlations

- There are a number of terms that are commonly applied to trends, particularly in graphs
  - **Direct and inverse proportionality**
    - Direct proportionality applies to a trend that has a clearly **linear** relationship which means the relationship can be described as "**when one variable increases, the other increases**" or "if x doubles, then y doubles"

        - Inverse proportionality means that the relationship can be described as "**when one variable increases, the other decreases**" or "if x doubles, then y halves"
    - **Positive and negative correlations**
        - Positive correlations show when the gradient of the graph is positive / slopes or curves upwards and describes a relationship where as x increases, y also increases
        - Negative correlations is when the gradient of the graph is negative / slopes or curves downwards; this describes a relationship where as x increases, y decreases

Percentage change and percentage difference

- Percentage change and percentage difference are commonly used to express **the relative change between two values**
    - They are useful for comparing experimental results, determining reaction yields and analysing other chemical data

Percentage change

- Percentage change is used to express the relative change between an initial value and a final value
- It is calculated using the following formula:

$$\text{Percentage Change} = \frac{\text{Final value} - \text{Intial value}}{\text{Initial value}} \times 100$$

Percentage difference

- Percentage difference is used to compare two values to determine how much they differ from each other as a percentage
- It is calculated using the following formula

$$\text{Percentage Difference} = \frac{(\text{Value 1} - \text{Value 2})}{\text{Average value 1 and value 2}} \times 100$$

Continuous and discrete data

- **Discrete** data is quantitative
    - It consists of separate, distinct and countable values
    - For example:
        - Number of an organism in a sample
- **Continuous** data is also quantitative
    - It is based on measurements and can include decimal numbers or fractions
    - This allows for an infinite number of values
    - For example:
        - The temperature of an enzyme reaction as time progresses
        - The volume of oxygen gas produced during a photosynthesis reaction

Statistical tests

- Statistical tests can be used to analyse a range of different data sets
- The type of test used will depend on a number of factors such as
  - The size of the sample
  - They type of data, i.e. is it discrete or continuous
  - The nature of the question being investigated

Simpson's reciprocal index

- The Simpson's reciprocal index can be used to **measure the relative biodiversity of a given community**
- It accounts for both the number of species present (richness) and the number of individuals per species (evenness)
- A **higher index value is indicative of a greater degree of biodiversity** within the community

The Lincoln index.

- This calculation allows an **estimate of population sizes of individual animal species**
- You can read more about the Lincoln Index here

Chi-squared test

- A chi-square test is a statistical test that is used **to compare observed and expected results**
- Our revision notes here cover this in detail

The *t*-test

- The **t-test** can be used to **compare the means of two sets of data** and determine whether they are **significantly** different or not
- The sets of data must follow a rough **normal distribution**, be **continuous** and the **standard deviations** should be approximately equal

Exam Tip

You will be provided with the formulae for these statistical tests in the exam, your job is to apply them to a range of contexts and data.

Using Units, Symbols & Numerical Values in Biology

Using Appropriate Units

- The **International System of Units** (SI) is also called the metric system
  - This is the international standard for measurement
- There are several SI base units that are used in science

SI Base Units Table

Quantity	SI base unit	Symbol
length	metre	m
mass	kilogram	kg
time	second	s
temperature	Kelvin	K
amount of substance	mole	mol
current	Ampere	A
luminous intensity	candela	cd

- Measurements of physical quantities can require very large and very small values, for example:
  - The diameter of an atom is about 10^{-10} m or 0.0000000001 m
  - One mole of a substance contains 6.02×10^{23} or 602 000 000 000 000 000 000 000 particles
- Powers of ten are numbers that can be achieved by multiplying 10 times itself
- These come under two categories of units:
  - **Multiples** e.g. 10^2, 10^3
  - **Sub-multiples** e.g. 10^{-1}, 10^{-2}
- Each power of ten is defined by a prefix, the most common ones used in biology are listed in the table below

Table of common prefixes in biology

Prefix	Abbreviation	Power of ten
kilo-	k	10^3
centi-	c	10^{-2}
milli-	m	10^{-3}
micro-	µ	10^{-6}
nano-	n	10^{-9}

- It essential that the correct scientific measurements are used when discussing biological experiments
- Ensure that the **correct symbols** are used in conjunction with the unit of measurement
  - E.g. m^3 for cubic metres

Units of Measurement Table

Measurement	Base unit	Symbol	Units used
Length	Metre	m	1000 m = 1 km 0.01 m = 1 cm 0.001 m = 1 mm 0.000001 m = 1 µm
Volume	Cubic metre	m^3	10^9 m^3 = 1 km^3 0.000001 m^3 = 1 cm^3 10^{-9} m^3 = 1 mm^3 10^{-18} m^3 = 1 μm^3
Volume	Cubic decimetre	dm^3	0.001 dm^3 = 1 cm^3
Area	Square metre	m^2	10 000 m^2 = 1 ha 0.0001 m^2 = 1 cm^2
Mass	Kilogram	kg	1000 kg = 1 tonne 0.001 kg = 1 g

			0.000001 kg = 1 mg 10^{-9} kg = 1 μg
Time	Second	s	60 s = 1 min 60 min = 1 hour
Pressure	pascal	Pa	1000 Pa = 1 kPa
Energy	joule	J	1000 J = 1 kJ
Temperature	degree Celcius	°C	
Amount of substance	mole	mol	0.001 mol = 1 millimole

- cm^3 is the same as millilitre (ml)
- dm^3 is the same as litre (l)

Exam Tip

Be careful when using the word "amount" in your answers. "Amount" has a very specific meaning in science - "mole". Instead refer to the mass, volume or concentration of a substance!

Significant figures

- Significant figures must be used when dealing with quantitative data
- Significant figures are the digits in a number that are **reliable and absolutely necessary** to indicate the quantity of that number
- There are some important **rules** to remember for significant figures
  - All non-zero digits are significant
  - Zeros between non-zero digits are significant
    - 4107 (4.s.f.)
    - 29.009 (5.s.f)
  - Zeros that come before all non-zero digits are not significant
    - 0.00079 (2.s.f.)
    - 0.48 (2.s.f.)
  - Zeros after non-zero digits within a number without decimals are not significant
    - 57,000 (2.s.f)
    - 640 (2.s.f)
  - Zeros after non-zero digits within a number with decimals are significant
    - 689.0023 (7.s.f)

- When rounding to a certain number of significant figures:
  - Identify the significant figures within the number using the rules above
  - Count from the first significant figure to the specified number
  - Use the next number as the 'rounder decider'
  - If the decider is 5 or greater, increase the previous value by 1

Worked example

Write 1.0478 to 3 significant figures.

Answer:

Step 1: Identify the significant figures

They are all significant figures

Step 2: Count to the specified number (3rd s.f.)

1.0**4**78

Step 3: Round up or down

1.05

Exam Tip

An exam question may sometimes specify how many significant figures the answer should be, make sure you keep an eye out for this!

Processing Uncertainties in Biology

Processing Uncertainties in Biology

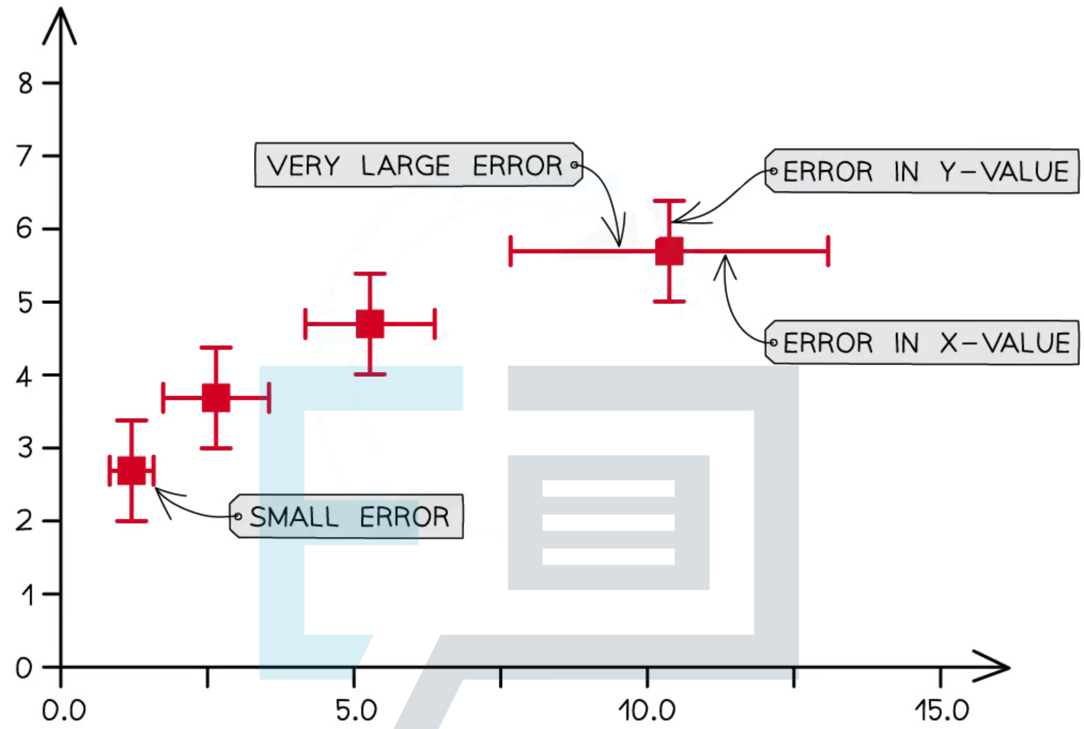
What is uncertainty?

- Uncertainty is **quantitative** indication of the quality of numerical results
  - It is the **difference between the actual measurement**, resulting from the **equipment** or techniques used to collect data, **and the true value**
  - It is a **range of values** around a measurement within which the true value is expected to lie
- Uncertainties are **not** the same as errors
  - Errors arise from equipment or practical techniques that cause a reading to be different from the true value
- Uncertainties in measurements are **recorded as a range (±)** to an appropriate level of precision, e.g.
  - If a balance that measures mass shows scale graduations of 10 g, then mass is measured to the nearest 10 g (this is known as the margin of error)
    - The true value could be 5 g higher or lower than the measured value, so the uncertainty would be ±5 g
  - If a pipette shows scale graduations every 0.1 cm^3, then volume is measured to the nearest 0.1 cm^3
    - The true value could be 0.05 cm^3 more or less than this, so the uncertainty would be ±0.05 cm^3

Error bars

- The uncertainty in a measurement can be shown on a graph as an **error bar**
  - This bar is drawn above and below the point (or from side to side) and shows the **uncertainty** in that measurement
  - Usually, error bars will be in the vertical direction, for y-values, but can also be plotted horizontally, for x-values
- **Range**, **degree of precision**, **standard error** and **standard deviation**; can be expressed on a graph using error bars
  - Range = the difference between the lowest and highest value
  - Degree of precision = how close a set of data points are to each other
  - Standard error = an estimate of the reliability of the mean
  - Standard deviation = the spread of data around the mean
- Note that it is important that you know **what is represented** by error bars on a graph, e.g. whether they represent standard deviation or standard error; in an exam this information would be provided in the question
  - Error bars that represent standard deviation can be used to assess whether or not two data sets are **significantly different** to each other

- Overlapping error bars indicate that two sets of data are not significantly different
- Error bars are used in the specification when measuring osmotic concentration



Error bars on a graph can be used to show uncertainty

Level of precision

- Measurements and processed uncertainties must be expressed to an appropriate level of precision
  - E.g. number of **decimal places**
- This may depend on the sensitivity of the apparatus used to collect data; the level of precision used to express the data should not exceed the level of precision at which the data is initially measured
- Values in a raw data set should all be expressed to the same level of precision

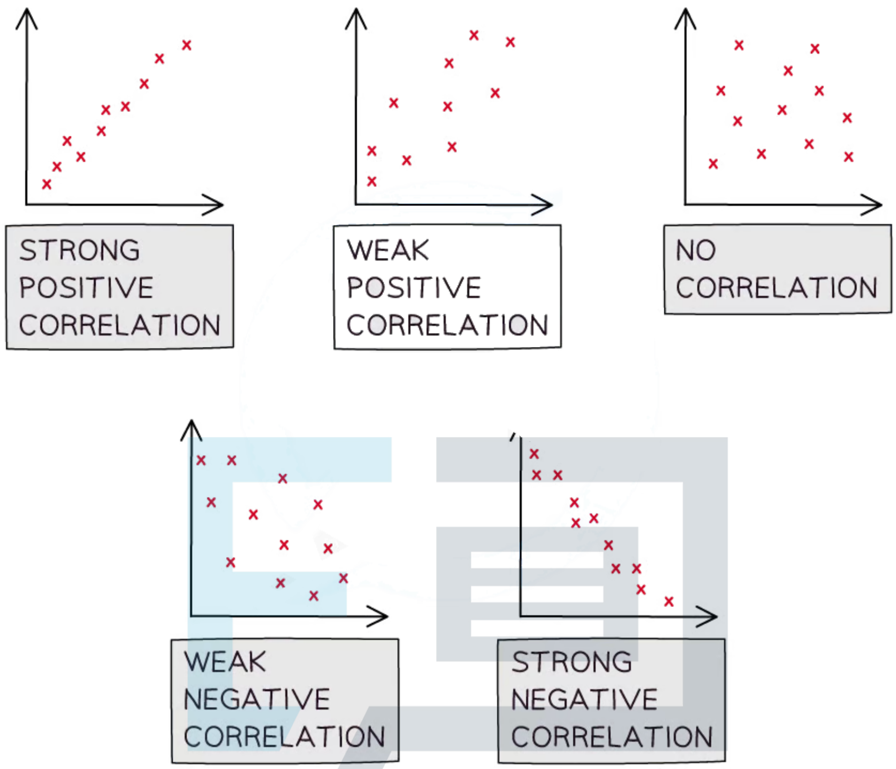
The coefficient of determination, R^2

- The **coefficient of determination** is a **measure of fit** that can be applied to lines and curves on graphs
- The coefficient of determination is written as R^2
- It is used to evaluate the fit of a trend line / curve with its data set:
  - $R^2 = 0$

        - The dependent variable cannot be predicted from the independent variable.
        - R^2 is usually greater than or equal to zero
    - **R^2 between 0 and 1**
        - The dependent variable can be predicted from the independent variable, although the degree of success depends on the value of R^2
        - The closer to 1, the better the fit of the trend line / curve
    - **R^2 = 1**
        - The dependent variable can be predicted from the independent variable
        - The trend line / curve is a perfect fit
        - **Note:** This does not guarantee that the trend line / curve is a good model for the relationship between the dependent and independent variables
- Coefficient of determination is used in the specification when comparing the speed of nerve impulse transmission

Correlation

- Correlation is an **association**, or relationship, between variables
    - Note that there is a clear distinction between correlation and causation: correlation **does not necessarily indicate a causal relationship**
    - Causation occurs when one variable has an influence or is influenced by another
- Correlation can be **positive** or **negative**
    - Positive correlation: as variable A increases, variable B increases
    - Negative correlation: as variable A increases, variable B decreases
- The **correlation coefficient (r)** can be calculated to determine **whether a linear relationship exists** between variables and **how strong** that relationship is
    - Perfect correlation occurs when all of the data points lie on a straight line; this will give a correlation coefficient of 1 or -1
        - 1 = a perfect positive correlation
        - -1 = a perfect negative correlation
    - A less-than perfect correlation will give a correlation coefficient between 1 and 0, or between 0 and -1
        - The closer to 1, or -1, the coefficient is, the stronger the correlation
    - If there is no correlation between variables the correlation coefficient will be 0
- Correlation coefficients are used in the specification when evaluating data on coronary heart disease



A strong correlation will have a correlation coefficient close to 1, a weak correlation will have a correlation coefficient close to 0, while a lack of any correlation will give a correlation coefficient of 0

Statistical tests

- Statistical tests are used to assess whether or not a data set **supports a particular hypothesis**. e.g.
  - A **null hypothesis** will state that there is **no** significant difference, or association, between two variables
  - An **alternative hypothesis** will state that there **is** a significant difference, or association, between two variables
- Statistical analysis allows researchers to **accept** or **reject** the null hypothesis
- If a statistical test shows that there is no significant difference, or association, between variables, then it is said that any visible difference is **due to chance** alone
- **Different statistical tests** are used for **different types of data set**, e.g.
  - A t-test determines whether the means of two data sets differ significantly
  - A correlation test determines the presence and strength of a correlation
  - A chi-squared test determines whether the difference between observed and expected values is significant
- You should be able to **select** and **apply** the correct statistical test
- The chi-squared test is used in the specification as follows:
  - To test for difference between observed and expected outcomes of a genetic cross
  - To test for association between species

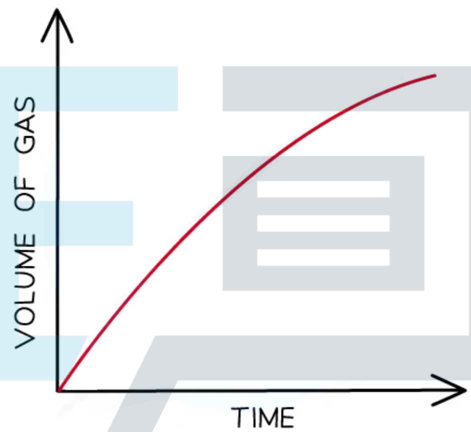
Graphing in Biology

Graphing in Biology

Sketch graphs

- **Sketch graphs** are a way to represent qualitative trends where the variables shown are often **proportional** or **inversely proportional**

A simple sketch graph



A sketch graph of the relationship between time and volume of gas given off, these two variables show a proportional relationship trend

General guidance on drawing graphs

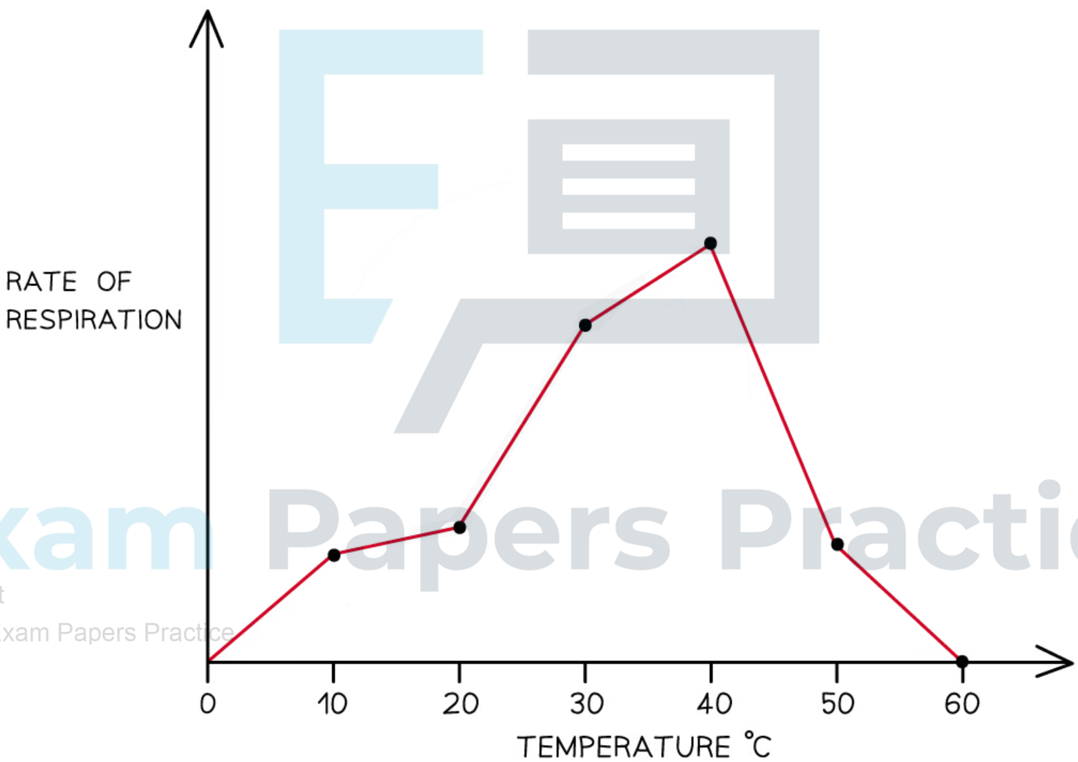
- The types of graphs that students are expected to be able to draw include:
  - Bar charts
  - Histograms
  - Scatter graphs
  - Line / curve graphs
  - Logarithmic graphs
  - Pie charts
  - Box-and-whisker plots

Tips for plotting data

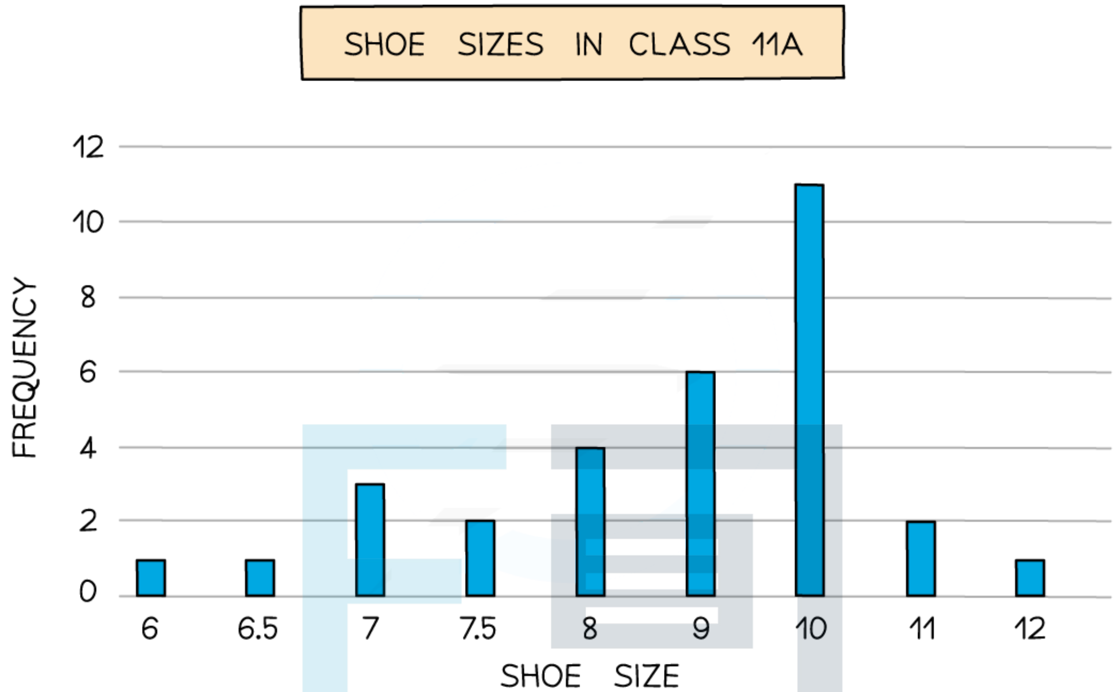
- Whatever type of graph you use, remember the following:
  - The data should be plotted with the **independent** variable on the **x-axis** and the **dependent** variable on the **y-axis**
  - Plot data points **accurately**

- Use **appropriate** linear **scales** on axes
- Choose scales that enable **all** data points to be plotted **within** the **graph area**
- **Label axes**, with **units** included
- Make graphs that **fill the space** the exam paper gives you
- Draw a **line of best fit**. This may be **straight** or **curved** depending on the **trend** shown by the data. If the line of best fit is a curve make sure it is drawn **smoothly**. A line of best-fit should have a **balance of data points** above and below the line
- In some cases, the line or curve of best fit should be drawn through the **origin** (but only if the data and trend allow it)

Continuous data represented in a line graph



Discontinuous data represented in a bar chart



The line graph has been used to display continuous data over time while the bar chart has been used to display grouped data

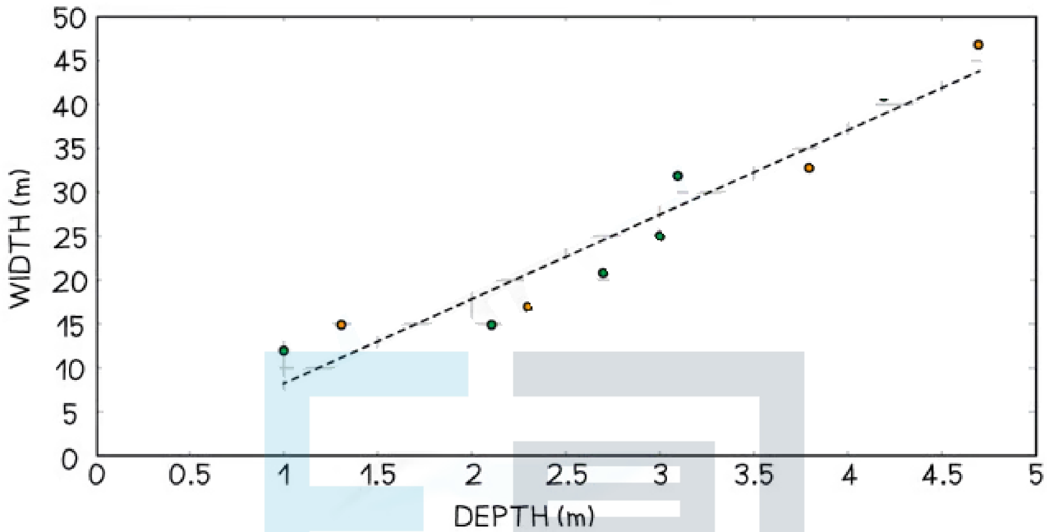
- **Remember:** The independent variable is the one you control or manipulate and the dependent variable is the one that changes as a result of your manipulation
- Always draw data points in pencil as it makes it easier to make corrections and adjustments

Best fit lines

- Students often confuse the term **lines of best fit** with straight lines
- **Lines of best fit** can be straight lines or curves and:
  - They show the trend of the data
    - It does not have to go through all the points, but shows the general trend
  - They must go through the majority of the points
  - Where the data is scattered the points should be evenly distributed on either side of the **best fit line**

Graph to show use of a best fit line

RELATIONSHIP BETWEEN WIDTH AND DEPTH



SCATTERGRAPH TO SHOW THE RELATIONSHIP BETWEEN WIDTH AND DEPTH ON A RIVER LONG PROFILE

Other features of graphs

Using a tangent to find the initial rate of a reaction

- **For linear graphs (i.e. graphs with a straight-line), the gradient is the same throughout**

  - This makes it easy to calculate the rate of change (rate of change = change ÷ time)
- However, many **enzyme rate experiments** produce non-linear graphs (i.e. graphs with a **curved line**), meaning they have an **ever-changing gradient**
  - They are shaped this way because the **reaction rate** is **changing over time**
- In these cases, a **tangent** can be used to find the **reaction rate** at any **one point** on the graph:
  - A tangent is a **straight line** that is drawn so it just **touches** the curve at a **single point**
  - The **slope** of this tangent **matches** the slope of the **curve** at just that point
  - You then simply find the **gradient** of the straight line (tangent) you have drawn
- The **initial** rate of reaction is the rate of reaction at the **start** of the reaction (i.e. where **time = 0**)