Lecture-01-Introduction

Quantitative Biochemistry

- My name is Dr Philip Fowler and my undergraduate was in Physics, my PhD in Chemistry and I spent ten years in the Department of Biochemistry before moving up to the John Radcliffe hospital in 2016 to establish my own research group investigating antibiotic resistance.
- I took this course over in 2020 from Professor Elspeth Garman who gave this course and developed it over 25 years. About ten years ago I was a class tutor on this course.
- Please feel free to email me with any suggestions (or any other thoughts). My email is philip.fowler@ndm.ox.ac.uk.
- This term we have 12 lectures to introduce you to the mathematics you need for biochemistry. Next term someone else will introduce the statistics you need.
- You will have one examples class on Wed every week from weeks 2-8.
- The problem sheet for each class must be uploaded to Canvas by 4.30pm on the Monday immediately before your class.
- The next problem sheet will be made available at the same time from Canvas.
- I am conscious that some of you are less-experienced in maths than others, or maybe just haven't done it for a few years. Don't worry! The course is fast-paced but we've taught many students in your shoes over the years.
- At the end of the course in week 2 next term you will be required to complete a Quiz on Canvas. I'll talk more about this later, but it's sole purpose is so we can work out who might need extra help so we can set that up. To that end we only give high-level grades and you will have a generous time limit to work through the questions.

Why do Biochemists need to know Mathematics?

1. Plot and analyse experimental data
2. When will my bacteria die?
3. How fast will my reaction go? How can I make it go faster (or slower)?
4. How radioactive is my ^{32}P?
5. Is there a simple relationship between the variables in the data I've collected from my experiment?

> **Thought:**
>
> 🤔 For us, mathematics is a *tool* that helps us understand Biochemistry.

Who said this?

> During the three years which I spent at Cambridge my time was wasted.... I attempted mathematics and even went during the summer of [XXXX] with a private tutor to Barmouth but I got on very slowly. The work was repugnant to me, chiefly from my not being able to see any meaning in the early stages of algebra. This impatience was very foolish and in after years I have deeply regretted that I did not proceed far enough at least to understand something of the great leading principles of mathematics, for [people] thus endowed seem to have an extra sense.

Course principles

Maths is both a skill and knowledge.

To become competent at mathematics it is not enough to listen and understand the concepts. You must also *practice* doing mathematics. Much like other skills (training for a sport, playing an instrument), if you don't keep practicing you get rusty and need to regain your competence through more practice.

You need to warm-up your brain

Ever tried reading a textbook and then swapping to maths? Doesn't work! Just like other skills, you need to warm-up your brain. Hence we will have at the start of each lecture a small number of warm-up questions - try and answer them quickly.

In this course we will build layer upon layer

We start off with the bottom layer of bricks; graphs, some algebra, logarithms, indices and trigonometry. Then we start building the next layer; calculus (differentiation) and integration. Finally we place the top layer of bricks: differential equations.

Each layer *relies* on the one below: you need algebra and graphs to understand calculus which you need to solve differential equations which you will need in Biochemistry.

Hence this course is like surfing a wave.

You have to listen to ALL the lectures, do ALL the questions, attend ALL your examples classes and talk to your examples class tutor, otherwise the wave will crash over you and pass you by.

If you miss one lecture you might not be able to understand the next concept: the wave will crash over you and pass you by.

You might have more experience at mathematics than your friends; don't make the mistake of thinking you can wait until we hit material you don't know! The wave is going fast and it *will* pass you by. Start paddling now!

Mathematicians are lazy

There will be occasions where you will wonder "why is writen that way?". e.g. doesn't $\sin(x)$ more clearly show what is the function and what is the independent variable? Yes, but after a while, mathematicians get tired of writing brackets and simply put $\sin x$.

This course is to help you learn biochemistry.

Hence we won't do formal proofs; if we examine why something is true it is to help you understand it. It also means we are only covering the concepts you need for the rest of your course. So if you are thinking "why aren't we doing X?". The answer is probably because it isn't needed (or isn't needed very often) in the Biochemistry course. Finally, we sometimes need to introduce a concept a little earlier than ideally we'd like so you've heard about it before you encounter it elsewhere in your course, so if you are thinking "why talk about this now - wouldn't it make more sense later?" it is probably because another lecturer is about to use that concept.

In the words of Sydney Brenner

Math is the art of the perfect, physics is the art of the optimal, and biology is the art of the satisfactory

Lecture notes

This course was developed by Professor Elspeth Garman, a long-standing and distinguished member of the department, over 25+ years. I've converted her OHPs to `jupyter` notebooks so they can be recorded electronically. The lecture notes are automatically created from the notebooks using `nbconvert`. They are all stored on a GitHub private repository - it is hoped to make this public in the future.

> **For you:**
>
> Please email me any mistakes you find: there will be some since this is the second year I've given these lectures!

If you are interested, the graphs are all plotted using `Matplotlib` in `Python`, making occasional use of `numpy`.

Key mathematical concepts

In the remainder of this lecture and the start of the next we will quickly trot through some elementary concepts that hopefully will be familiar to most of you.

### Graphs and graphical representation

We nearly always are concerned with 2-dimensional graphs

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
/var/folders/5v/2144g0bs2jq2w110dhq2g_pm0000gn/T/ipykernel_2048/3482300946.py in
 ↪<module>
----> 1 x = numpy.linspace(-1, 3, 5000)
      2
      3 # plt.xkcd()
      4 page,graph=plt.subplots(nrows=1,ncols=1,figsize=(5,5))
      5

NameError: name 'numpy' is not defined
```

y is the DEPENDENT VARIABLE

x is the INDEPENDENT VARIABLE

**What information should you put on a graph?**
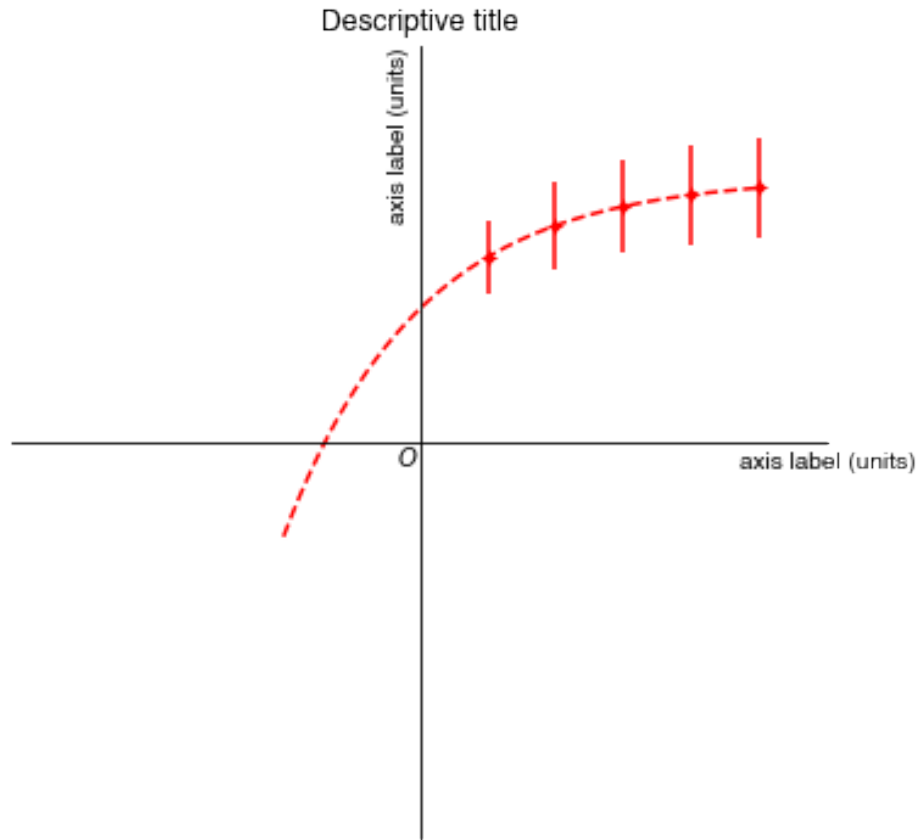
The graph should be **self-contained** i.e. understandable on its own - imagine a piece of graph paper falling out of your lab book. Can you pick it up and understand what it is showing?

The TITLE should state the relationship between the independent and dependent variables. e.g. 'The relationship between enzyme activity and external pH'.

The AXES should be LABELLED with UNITS and span an appropriate RANGE. You may need to extend an axis if later you need to read where a fitted line crosses the y-axis. Does the ORIGIN $(0, 0)$ need to be shown? Is it misleading not to?

All the DATA POINTS should be clearly plotted with ERROR BARS, if appropriate. Also if a CURVE or LINE is fitted to the data, this needs to be clearly shown. Humans are rubbish at drawing anything other than straight lines so usually we TRANSFORM the data to ensure it should lie on a STRAIGHT LINE.

Equations of a straight line
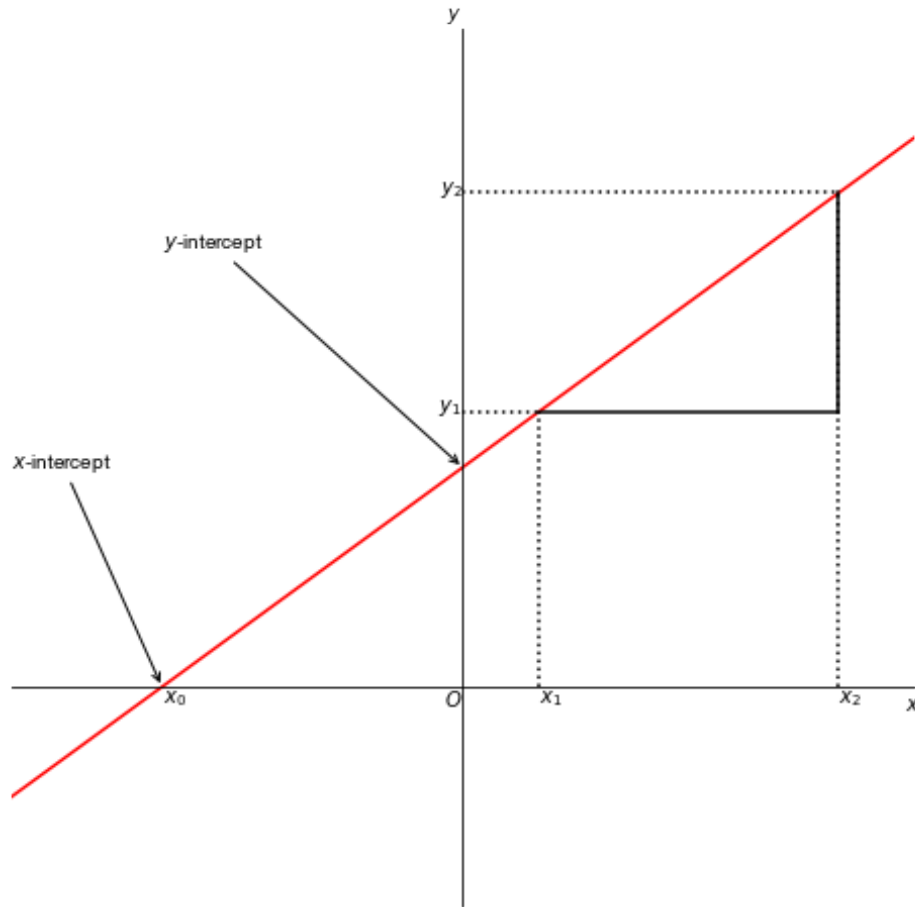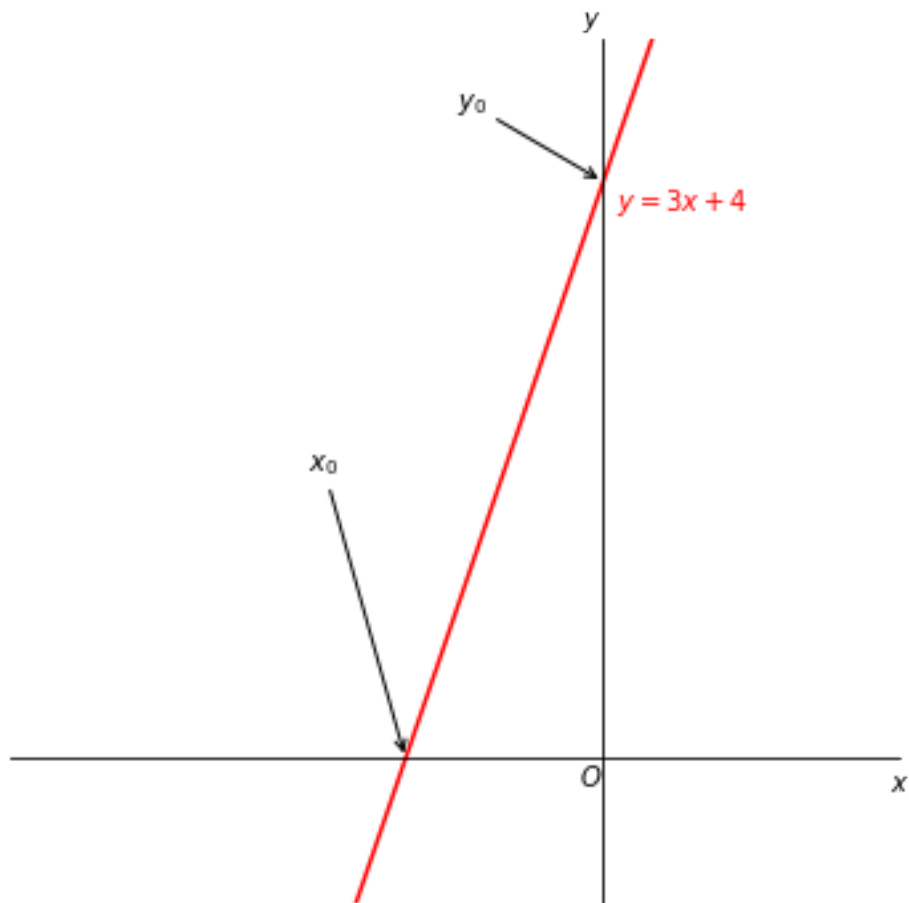

$$y = mx + c$$

where

$$m = \text{constant} \qquad = \text{GRADIENT OF LINE (+ve or -ve)}$$
$$c = \text{constant} \qquad = y\text{-INTERCEPT}$$
$$\frac{-c}{m} = \text{constant} \qquad = x\text{-INTERCEPT}$$

Example

y

y_0

$y = 3x + 4$

x_0

O

x

What are the values of x_0 and y_0 and what is the slope/gradient of the line?

$$y = 3x + 4 \qquad\qquad\qquad \text{substitute } x = 0$$
$$y_0 = 4$$

$$y = 3x + 4 \qquad\qquad\qquad \text{substitute } y = 0$$
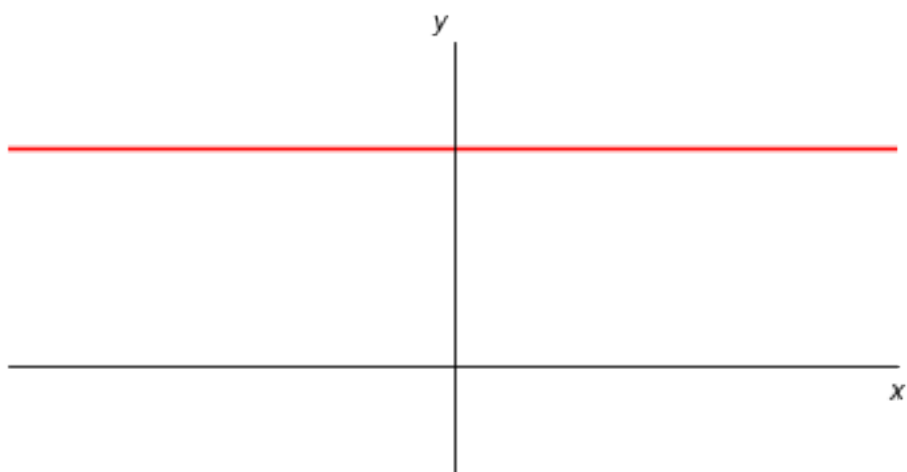$$0 = 3x_0 + 4$$
$$x_0 = \frac{-4}{3}$$

The slope is 3 (i.e. $m = 3$).
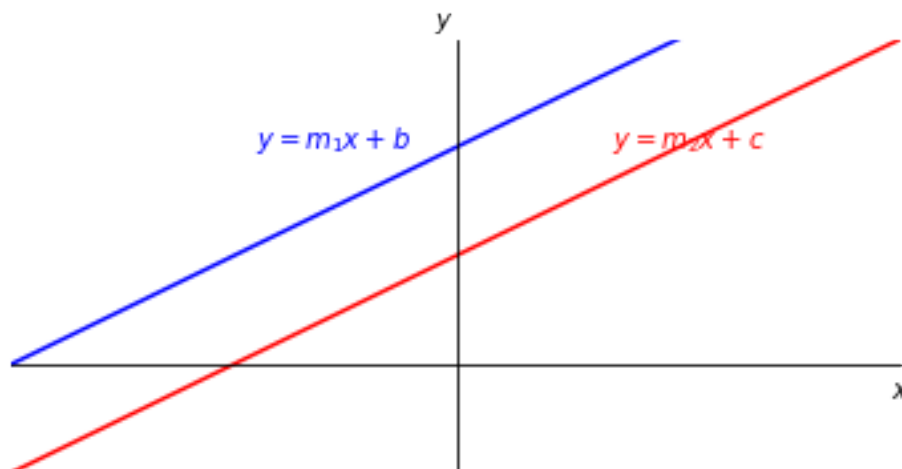
Special cases

1. When $c = 0$ then $y = mx + 0$
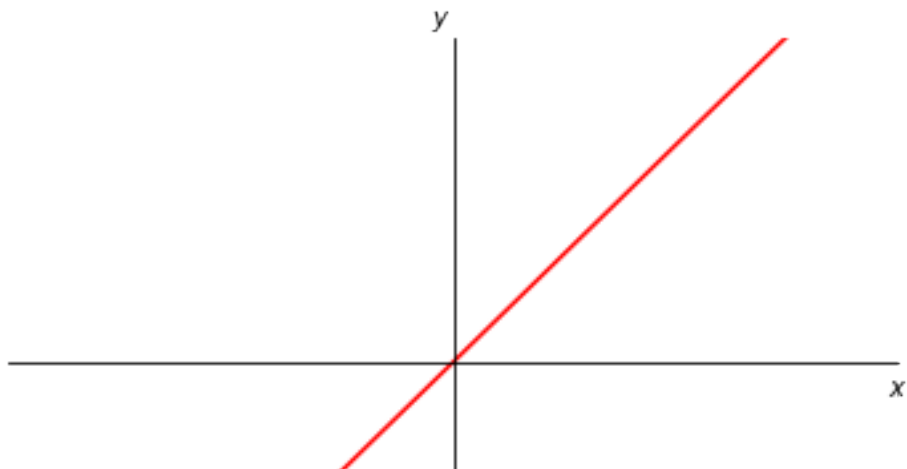
2. When $m = 0$ then $y = c$

3. Two lines are PARALLEL if and only if their slopes are the same i.e. in the graph below iff* $m_1 = m_2$.

* - shorthand for "if and only if"

Polynomials

A polynomial is a function that is a sum of one or more powers of x. Note that since $x^0 = 1$ they can include numbers.

Examples

1. $y = 2x^2$
2. $y = 5x^2 + 3x^3 + 1$
3. $y = 10x^{100} + 2x^4 + 3$

We may generalise the form of a POLYNOMIAL EQUATION

$$y = a_n x^n + a_{n-1} a^{n-1} + a_{n-2} x^{n-2} + \cdots + a_1 x^1 + a_0 x^0$$

where $a_n, a_{n-1}, a_{n-2} \ldots a_1, a_0$ are CONSTANTS. So in Example 3 above $a_{100} = 10, a_4 = 2, a_0 = 3$.
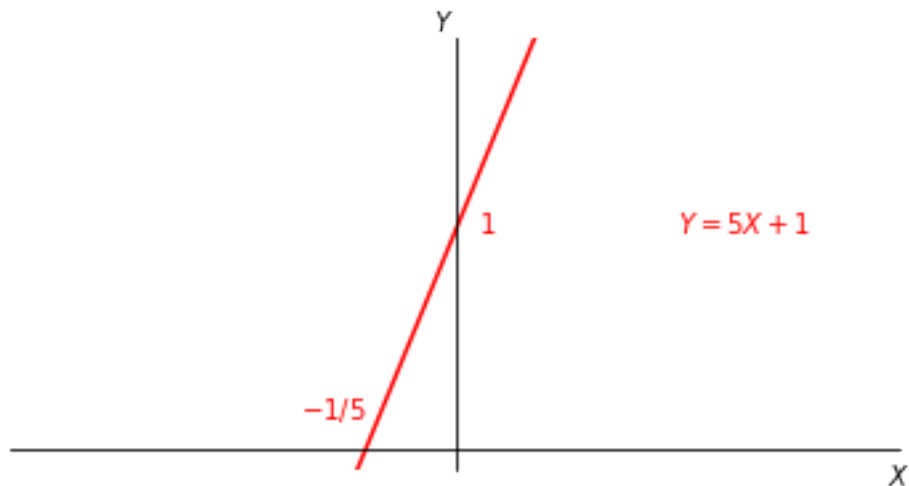
The ORDER of a polynomial is the largest power of x. Hence the examples have order 2, 3 and 100, respectively.

A polynomial of order N has $N - 1$ curves/turning points. ($N > 0$)

Can draw graphs of polynomials e.g. $y = 5x^2 + 1$ which is SECOND ORDER and so has one TURNING POINT.

$$y = 5x^2 + 1$$
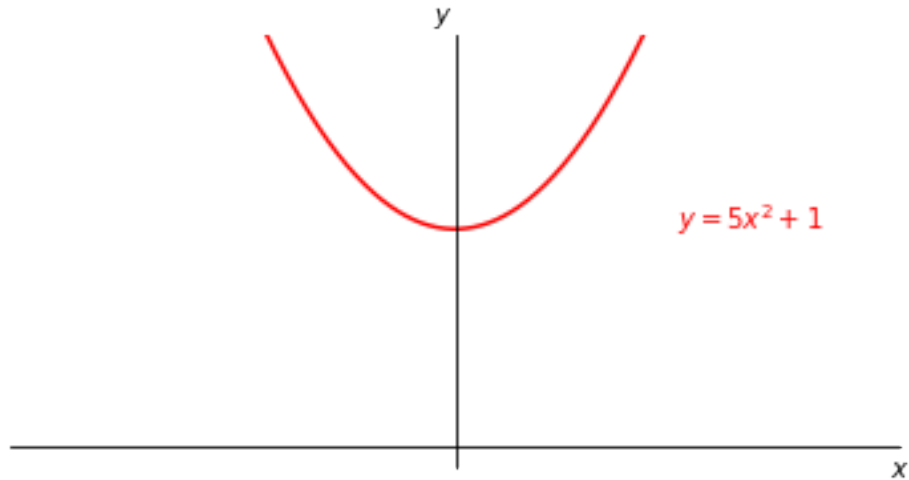
But we can TRANSFORM this curve by setting $Y = y$ and $X = x^2$, then $Y = 5X + 1$

$$Y = 5X + 1$$

Fitting experimental data to a mathematical function

This suggests a method by which we could check whether our data were described by a given functional form. Suppose we think our data have the general form

$$y = mx^2 + c$$

then we could plot a graph of y versus x^2 and see if it is a straight line with gradient m.

Many problems in Biosciences reduce down to (i) identifying the theory that describes your data best and then (ii) plotting the right graph so it comes out as a straight line and then (iii) finding the gradient and y-intercept so you can (iv) calculate the relevant constants.

One such example is that of enzyme-catalysed reactions.

Enzyme catalysis   The inverse of the rate at which a given enzyme can catalyse a reaction is dependent on the inverse of the substrate concentration plus an offset (this is the Michaelis-Menten equation and will be covered in detail elsewhere in the course).

The first thing we have to do is 'translate' the words into a mathematical function

$$\frac{1}{V} = \frac{m}{S} + c$$
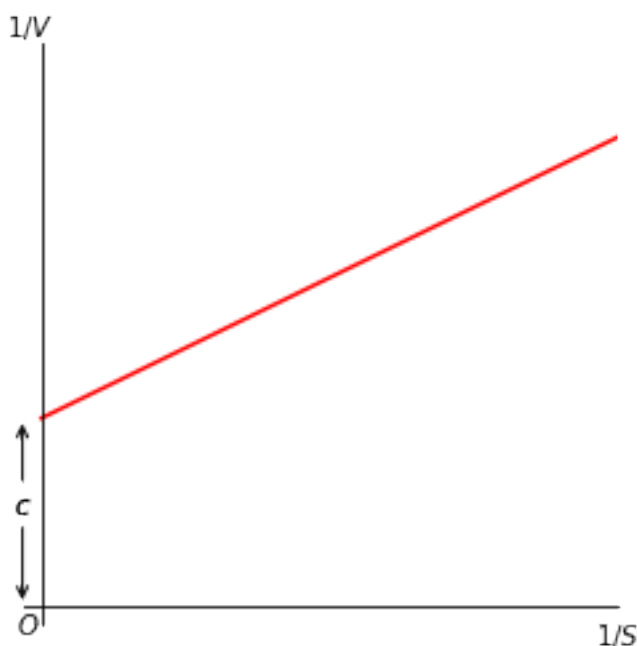
Curve sketching

Often useful to be able to assess important details of the curve WITHOUT having to draw it accurately (i.e. using graph paper).

(Aside: if you are asked to PLOT a curve, it means accurately using graph paper. SKETCH means what we are doing here; the curve can be done quickly by hand - we are interested in labelling details like where it crosses the axes, the location of any TURNING POINTS).
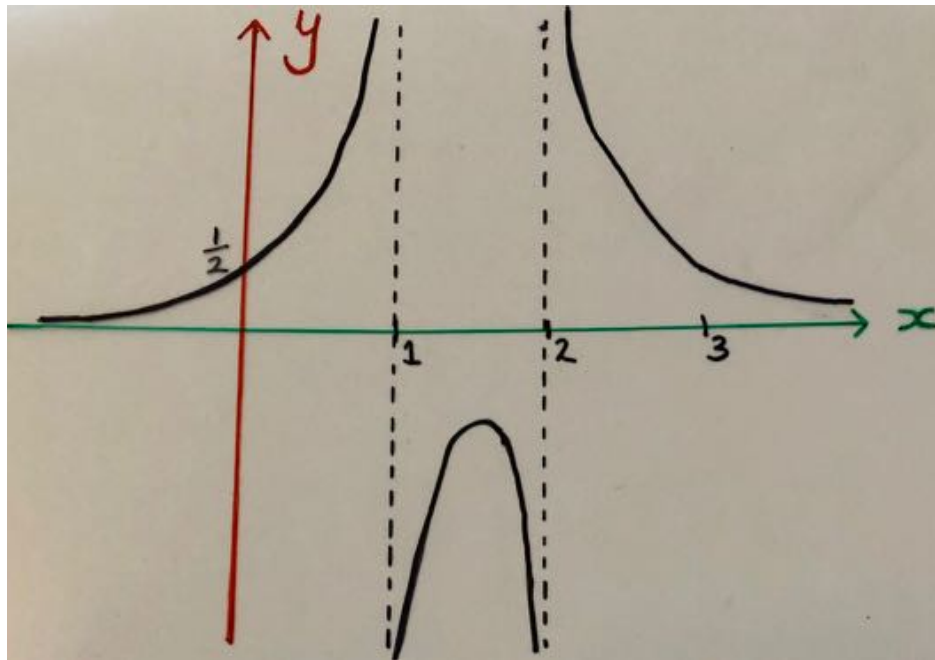
Example   SKETCH the curve

$$y = \frac{1}{(x-1)(x-2)}$$

A good approach is to work out the value of y at key points

x	y
0	1/2
very large and positive	very small and positive
very large and negative	very small and positive
1	$-\infty$
1.5	-4
2	$+\infty$
3	1/2

Now that we have this table, sketching the function is a bit like "joining the dots".

Always check how your curve behaves at $x = \pm 0, \pm\infty$.

The dotted lines are ASYMPTOTES. Note how the function never quite reaches e.g. $x = 1$. Also how the sketch never quite reaches the x-axis (since you'd have to be at $\pm\infty$ for that to be true).
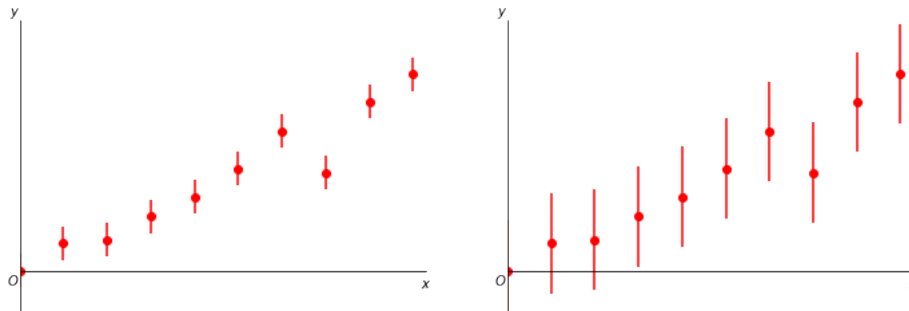
Real data

Real data collected from experiments is messy.

- It can contain OUTLIERS.
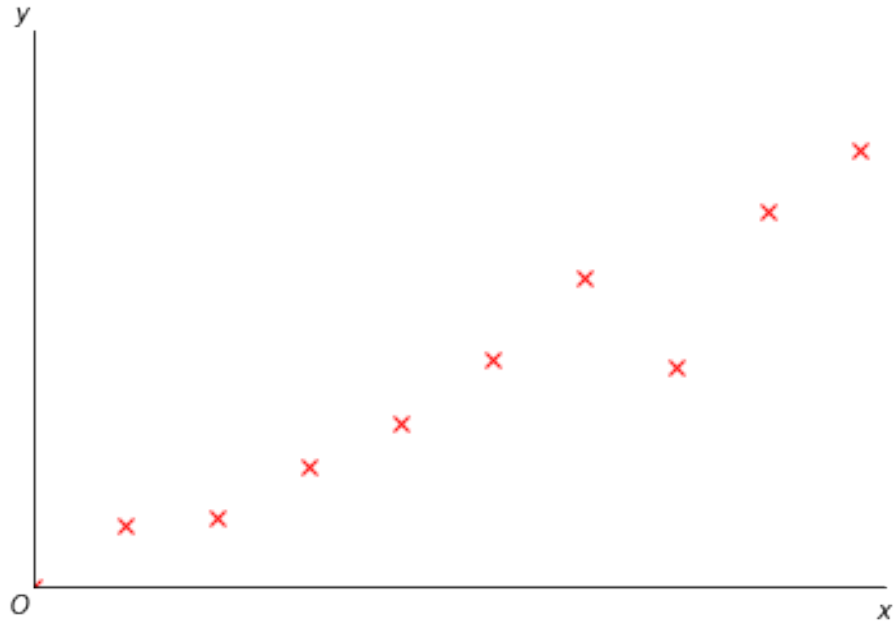
- you might have to space the data points out so that sparse areas are populated
- plotting the ERROR BARs is often key to interpretation

For example, the data below are identical but the error bars are 3x larger on the right. Is there still an outlier?

To work out if a data point is an outlier or to reduce the size of the errors the answer is to collect more data!

- Overfitting: you can always find a function that fits your data, but that doesn't mean it is correct!
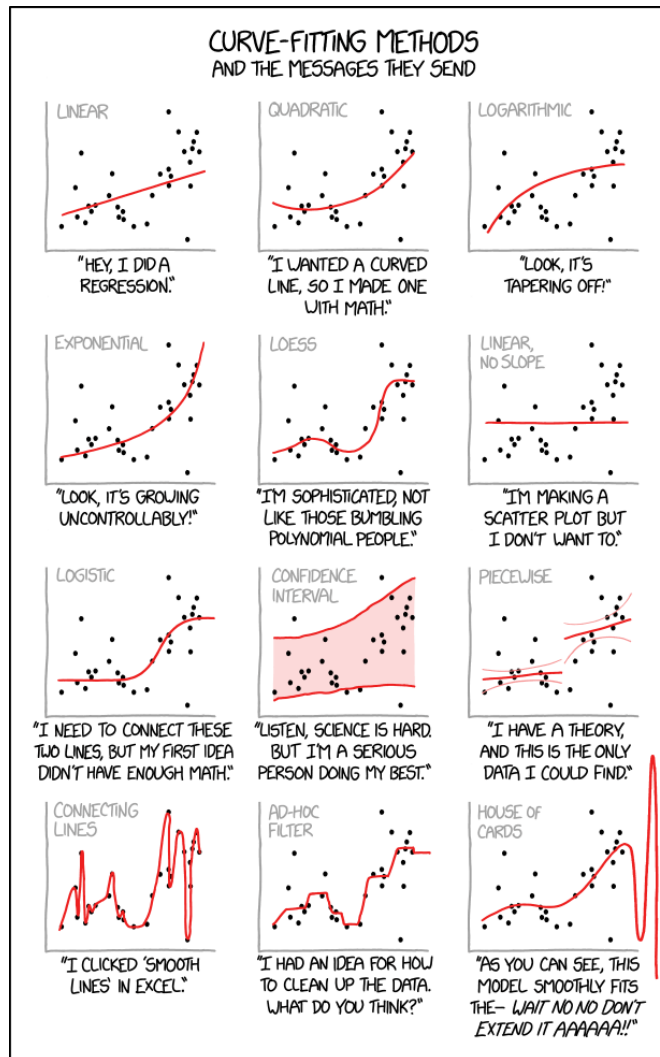
- be careful when extrapolating
- make sure you use the appropriate scales for the axes

Common misconception:

When finding the slope (gradient) of real data use the FITTED LINE not two points on the graph

Units

Length

1 m = 10 dm = 10^2 cm = 10^3 mm = $10^6 \mu$m = 10^9 nm = 10^{10} Angstroms.

In Biochemistry we commonly swap between these units

10^{-1} dm = 1 cm = 10 mm

Clearly it would be silly to use of these units in everyday life...

50 m = 5×10^{11} Angstroms = 5×10^4 mm

Area

$$
\begin{aligned}
1\text{m}^2 = 10^2\text{cm} \times 10^2\text{cm} \qquad\qquad &= 10^4\text{cm}^2 \\
= 10^3\text{mm} \times 10^3\text{mm} \qquad\qquad &= 10^6\text{mm}^2 \\
= 10^6\mu\text{m} \times 10^6\mu\text{m} \qquad\qquad &= 10^{12}\mu\text{m}^2 \\
= 10^9\text{nm} \times 10^9\text{nm} \qquad\qquad &= 10^{18}\text{nm}^2
\end{aligned}
$$

For example 50 mm^2 = $50 \times 10^6 \mu$m^2 = 50×10^{12} nm^2.

(Aside: in nuclear physics they need *really* small areas so have defined a 1 barn = 10^{-28} m^2, 1 outhouse = 10^{-34} m^2, 1 shed = 10^{-52} m^2)

Volume

$$
\begin{aligned}
1 \text{ dm}^3 = 10^3 \text{ ml} \qquad\qquad &= 10^{-3} \text{ m}^3 \\
10^{-3}\text{dm}^3 = 1 \text{ ml} \qquad\qquad &= 1 \text{ cm}^3 = 1 \text{ cc} \\
1\,\text{l} = 1000 \text{ ml} \qquad\qquad &= 1000 \text{ cc}
\end{aligned}
$$

$$
\begin{aligned}
1\text{cm}^3 = 10^{-2}\text{m} \times 10^{-2}\text{m} \times 10^{-2}\text{m} \qquad\qquad &= 10^{-6}\text{m}^3 \\
= 10^{-3}\text{dm}^3
\end{aligned}
$$

1 dm^3 = 10^3 ml (estimate of volume of water in jug)

For example 5 ml = 50×10^{-3} dm^3 = 50×10^{-9} m^3

Density

$$
\begin{aligned}
1\text{g/ml} = 1\text{g. ml}^{-1} \\
= 1 \times 10^3 \text{ g} .10^3 \text{ ml}^{-1} \\
= 1\text{kg. dm}^{-3}
\end{aligned}
$$

For proteins, density is usually quoted in mg ml^{-1}.

Concentration: Molarity

$$1 \text{ M} = 10^3 \text{ mM} \qquad\qquad = 10^6 \,\mu\text{M} = 10^9 \text{ nM}$$

hence

$$1 \text{ nM} = 10^{-6} \text{ mM} \qquad\qquad = 10^{-3} \,\mu\text{M} = 10^{-9} \text{ M}$$

Moles, molarity and Daltons

Moles (amount)

Units of mol.

$$\text{moles} = \frac{\text{weight}}{\text{molecular weight}}$$

The units have to be consistent, so if the molecular weight is measured in g/mol then the weight has to be in g.

Molarity (concentration)

Units of M

$$\text{molarity} = \frac{\text{moles of solute}}{\text{volume of solution}}$$

Again, the units have to be consistent so if the volume is measured in l, and molarity is M, which is mol/l (or equivalently mol/dm^3), then moles of solute must be measured in mol.

Daltons (mass)

1 Dalton is one twelfth the mass of a carbon-12 atom. The exact nuclide must be specified since carbon has multiple isotopes.

$$1\text{Dalton} = \frac{1}{12} \text{ mass } {}^{12}\text{C} = 1.66054 \times 10^{-27} \text{ kg}$$

Protein masses are usually quoted in thousands of Daltons (kDa).

Ratios and dilutions

Since moles of solute = molarity $\times$ volume we can often get the answer we need using ratios. i.e.

$$M_1 \times V_1 = M_2 \times V_2$$

Example 1

75 ml of a 1.4 M solution of NaCl is 'made up' to 130 ml. What is the molarity of the new solution?

$$M_1 \times V_1 = M_2 \times V_2$$
$$75 \times 1.4 = M_2 \times 130$$
$$M_2 = \frac{75 \times 1.4}{130}$$
$$=$$

Example 2

How many moles of ATP are there in a 25 μl of a 3 nM solution?

$$\text{moles of solution} = \text{molarity (M)} \times \text{volume of solution (l)}$$
$$= 3 \times 10^{-3} \times 25 \times 10^{-6}$$
$$= 75 \times 10^{-9} \text{ moles}$$
$$= 75 \text{ nmoles}$$

What is its weight? (The molecular mass of ATP is 507.18 g/mol)

$$\text{moles} = \frac{\text{weight}}{\text{molecular weight}}$$

hence

$$\text{molecular weight} = 75 \times 10^{-9} \times 507$$
$$= 3.8 \times 10^4 \times 10^{-9}$$
$$= 3.8 \times 10^{-5}$$
$$= 38 \times 10^{-6}$$
$$= 38 \mu g$$

Useful geometric formulae

Circumference of a circle

$$c = 2\pi r$$

where r is the radius, π is the irrational constant with value 3.141259...and c is the circumference of the circle.

Area of a circle

$$A_{circ} = \pi r^2$$

Volume of a sphere

$$V_{sph} = \frac{4}{3}\pi r^3$$

For example, estimate the volume of a human cell if its length is 8 μm.

Assume the cell can be approximated by a sphere but beware that we've been given the diameter, hence radius = 4 μm.

$$V_{cell} = \frac{4}{3} \times \pi \times 4^3$$
$$\approx 4 \times 64$$
$$\approx 256 \mu m^3$$

What is the volume in m^3? Well 1 μm = 10^{-6} m. Hence 1 μm^3 = $(10^{-6})^3$m^3 = 10^{-18} m^3.

$$V_{cell} = 2.56 \times 10^2 \times 10^{-18}$$
$$= 2.56 \times 10^{-16} m^3$$

What about in mm? 1 mm = 0.001 m = 10^{-3} m.

$$V_{cell} = 2.56 \times 10^2 \times (10^{-3})^3$$
$$= 2.56 \times 10^{-7} \text{ mm}^3$$

Volume of a cylinder

$$V_{cyl} = \pi r^2 l$$

where r is the radius and l is the length/height of the cylinder.

Dimensional analysis

There are many units that physical quantities can be measured and expressed with. You will be meeting even more shortly!

All units can be expressed in terms of 7 fundamental base quantities. Many quantities only need three: * Mass [M] * Length [L] * Time [T]

Note that the square brackets indicate that this is a dimension, so [M] means "has dimension of mass" and is nothing to do with molarity here.

For example,

$$F = ma$$

m has dimensions [M] and a has dimensions of $[L][T]^{-2}$. Hence force has its 'own' unit of the Newton, it has dimensions of $[M][L][T]^{-2}$. We've therefore shown that force is not fundamental since it can be expressed in terms of mass, length, and time.

Likewise, since

$$\text{pressure} = \frac{\text{force}}{\text{area}}$$
$$= \frac{[M][L][T]^{-2}}{[L]^2}$$
$$= [M][L]^{-1}[T]^{-2}$$

Hence pressure has units $kg\ m^{-1}\ s^{-2}$, otherwise known as the Pascal (Pa) in the SI system. Other units exist such as Torr.

Why is this useful? The key insight is that the dimensions in any expression we write down MUST balance i.e. LHS = RHS - we've implicitly just used this relationship. You can use this to "sense check" any formula you are using since if they don't yield the right dimensions for the quantity in question then something is missing or wrong.

The other four fundamental dimensions are less common and are * Temperature $[\theta]$ * Electric charge [Q] * Amount of a substance [N] * Luminosity [J]

Significant figures

The number of significant figures (s.f.) is how many digits we quote for a quantity.

For example,

$$6 = 1 \text{ s.f.}$$
$$6.0 = 2 \text{ s.f.}$$
$$6.05 = 3 \text{ s.f.}$$
$$3.141592654 = 10 \text{ s.f}$$
$$0.1 = 1 \text{ s.f.}$$
$$0.01 = 1 \text{ s.f.}$$
$$0.0112 = 3 \text{ s.f.}$$
$$1.12 \times 10^{-2} = 3 \text{ s.f.}$$

The accuracy of any device used to make a measurement limits the number of significant figures you can quote a value to.

Hence an accuracy of 1 in 10,000 implies 4 s.f. e.g. 73.24, whilst 1 in 1,000 implies 3 s.f. e.g. 4.52.

Any final result cannot be any more accurate than the least accurate quantity used in its calculation. Hence if we multiplied the above two numbers together, we should only quote the answer to 3.s.f.

$$73.24 \times 4.52 = 331$$

Common misconception:
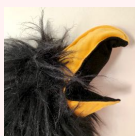
Do NOT quote spurious accuracy in your calculations.

Use of calculators

Calculators, like maths, are a tool. They help us, but if we press the wrong buttons they give us the wrong answer. It is always a good idea, therefore, to check your answer "by eye". Does it "look right"? As you practice more and more maths, you may get to the stage where it "feels" wrong but you are not sure why! Trust that feeling.

Similarly, we can sometimes simplify a calculation by doing some simple maths to reduce the number of buttons we have to press and therefore reduce the likelihood we will make a mistake.

A good example is calculations involving many large and small powers of ten: this is VERY common in Biochemistry!

What is the energy of 600 nm photons?

$$
\begin{aligned}
E &= \frac{hc}{\lambda} \\
&= \frac{6.63 \times 10^{-34} \times 3 \times 10^{8}}{600 \times 10^{-9}} \\
&= \frac{6.63 \times 3}{600} \times 10^{-34+8+9} \qquad \leftarrow \text{check that the powers of ten add up to -19} \\
&= \frac{6.63 \times 3}{6} \times 10^{-19} \times 10^{2} \\
&= 3.32 \times 10^{-17} \text{ J}
\end{aligned}
$$

In the example above, you can limit your use of the calculator by grouping the powers of ten together and then doing them yourself (this also makes it faster). And when you've got an answer you can think "does this make sense? is the power of ten in the right range?". So if you got 3.32×10^{17} J you would think at this stage NO!
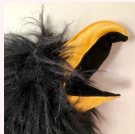
Errors

Devices used to measure quantities all have associated errors. There are two types of error you will encounter

Absolute Errors

For example, length measured using a 30cm ruler. The smallest length we can measure is 1 mm i.e. 0.1 cm. The absolute error is half this, therefore 0.05 cm. Hence if we were to measure the length of something using a 30 cm ruler, properly we should quote it as 20 ± 0.05 cm.

Relative Errors

$$\text{relative error} = \frac{\text{absolute error}}{\text{the measurement}}$$

Hence using our example above

$$\text{relative error} = \frac{0.05}{20}$$
$$= \pm 0.25\%$$

In this case, the relative error will decrease as the magnitude of the measurement increases. You will encounter errors all through your course and, as we saw in graphs, they are VERY important in interpreting data.

One is always conservative e.g. we always try to calculate the maximum error, so that in practice the real errors are probably less than these.