Pearson BTEC Levels 4 and 5 Higher Nationals in Engineering (RQF)

**Unit 2: Engineering Maths (core)**

# Unit Workbook 1

in a series of 3 for this unit

Learning Outcomes 1 and 2

# Mathematical Methods

# &

# Statistical Techniques

**Relevance: Assignment 1**

# Contents

# Section 1: Mathematical Methods

# INTRODUCTION

Identify the relevance of mathematical methods to a variety of conceptualised engineering examples.

**Mathematical concepts:**

**Dimensional analysis.**

**Arithmetic and geometric progressions.**

**Functions:**

**Exponential, logarithmic, circular and hyperbolic functions**

# GUIDANCE

This document is prepared to break the unit material down into bite size chunks. You will see the learning outcomes above treated in their own sections. Therein you will encounter the following structures;

**Purpose** Explains *why* you need to study the current section of material. Quite often learners are put off by material which does not initially seem to be relevant to a topic or profession. Once you understand the importance of new learning or theory you will embrace the concepts more readily.

**Theory** Conveys new material to you in a straightforward fashion. To support the treatments in this section you are strongly advised to follow the given hyperlinks, which may be useful documents or applications on the web.

**Example** The examples/worked examples are presented in a knowledge-building order. Make sure you follow them all through. If you are feeling confident then you might like to treat an example as a question, in which case cover it up and have a go yourself. Many of the examples given resemble assignment questions which will come your way, so follow them through diligently.

**Question** Questions should not be avoided if you are determined to learn. Please do take the time to tackle each of the given questions, in the order in which they are presented. The order is important, as further knowledge and confidence is built upon previous knowledge and confidence. As an Online Learner it is important that the answers to questions are immediately available to you.. Contact your Unit Tutor if you need help.

**Challenge** You can really cement your new knowledge by undertaking the challenges. A challenge could be to download software and perform an exercise. An alternative challenge might involve a practical activity or other form of research.

**Video** Videos on the web can be very useful supplements to your distance learning efforts. Wherever an online video(s) will help you then it will be hyperlinked at the appropriate point.

# 1 Mathematical Concepts

## 1.1 Dimensional analysis

Very often in engineering we encounter numbers with associated physical units, for example; 3 volts (3 V), 5 kilograms per metre (5 kg/m or 5 kg.m$^{-1}$), 12 cubic metres (12 m$^3$) etc. Dimensional Analysis is a neat way to analyse those physical units, enabling us to check the validity of an equation (an equation has just one unknown and an 'equals' sign) or a formula (a formula has two or more unknowns and an 'equals' sign). We can even use dimensional analysis to generate an equation or formula.

Dimensional analysis uses the basic dimensions of **Length (L)**, **Mass (M)** and **Time (T)** to formulate the dimensions of other quantities. For example, the area of a football field can be expressed in square metres, and square metres is 'a length times a length', of course. Let's take a look at some notation for this simple analysis…

$$[Area] = [length \times length] = L^2$$

The use of the square brackets [ ] here denotes 'dimensions of'.

Another example could be 'density'. Let's look at that one…

$$[Density] = \left[\frac{mass}{volume}\right] = \frac{M}{L^3} = ML^{-3}$$

Yet another could be velocity (speed)…

$$[Velocity] = \left[\frac{metres}{seconds}\right] = \left[\frac{length}{time}\right] = \frac{L}{T} = LT^{-1}$$

Let's take a look at some rules which we need to be aware of when undertaking dimensional analysis…

**Rule 1:**

Constants (i.e. numbers) must be ignored. For example, if we see 3 m as a length then the dimensions are just L. The 3 is ignored.

**Rule 2:**

Angles must be ignored. An angle cannot be represented with our basic dimensions of length, mass or time.

**Rule 3:**

If physical quantities are to be added or subtracted, they MUST have the same dimensions. We cannot add apples to oranges, nor can we add voltage to current. However, we can add voltage to voltage.

At this point it is worth noting the dimensions of some common quantities encountered in engineering…

| | | |
|---|---|---|
| **[mass]** | **M** | **M** |
| **[length]** | **L** | **L** |
| **[time]** | **T** | **T** |
| **[area]** | **[length × length]** | **$L^2$** |
| **[volume]** | **[area × length]** | **$L^3$** |
| **[density]** | **[mass/volume]** | **$ML^{-3}$** |
| **[velocity]** | **[length/time]** | **$LT^{-1}$** |
| **[acceleration]** | **[velocity/time]** | **$LT^{-2}$** |
| **[force]** | **[mass × acceleration]** | **$MLT^{-2}$** |
| **[moment of force, torque]** | **$MLT^{-2} \times L$** | **$ML^2T^{-2}$** |
| **[impulse]** | **[force × time]** | **$MLT^{-1}$** |
| **[momentum]** | **[mass × velocity]** | **$MLT^{-1}$** |
| **[work]** | **[force × distance]** | **$ML^2T^{-2}$** |
| **[kinetic energy]** | **[mass × (velocity)$^2$]** | **$ML^2T^{-2}$** |
| **[power]** | **[work/time]** | **$ML^2T^{-3}$** |
| **[electrical charge]** | **[length x (force)$^{0.5}$]** | **$L^{1.5}M^{0.5}T^{-1}$** |
| **[current]** | **[electrical charge/time]** | **$L^{1.5}M^{0.5}T^{-2}$** |
| **[electric potential energy]** | **[charge$^2$/distance]** | **$L^2MT^{-2}$** |
| **[voltage]** | **[electric potential energy/charge]** | **$L^{0.5}M^{0.5}T^{-1}$** |
| **[capacitance]** | **[charge/voltage]** | **L** |
| **[inductance]** | **[voltage x time/current]** | **$L^{-1}T^2$** |
| **[electrical resistance]** | **[voltage/current]** | **$L^{-1}T$** |

**Worked Example 1**

Ohm's Law relates resistance (R), voltage (V) and current (I) as;

$$I = \frac{V}{R}$$

Use the dimensions of V and R to determine the dimensions of $I$.

**ANSWER**

$$[V] = L^{0.5}M^{0.5}T^{-1}$$

$$[R] = L^{-1}T$$

$$\therefore \quad [I] = \left[\frac{V}{R}\right] = \frac{L^{0.5}M^{0.5}T^{-1}}{L^{-1}T} = L^{0.5--1}M^{0.5}T^{-1-1}$$

$$\therefore \quad [\boldsymbol{I}] = \boldsymbol{L^{1.5}M^{0.5}T^{-2}}$$

**Worked Example 2**

A guitar string has mass (m), length ($l$) and tension (F, i.e. a force). It is proposed that a formula for the period of vibration (t) of the string might be;

$$t = 2\pi\sqrt{\frac{F}{ml}}$$

Use dimensional analysis to determine whether this formula might be correct.

**ANSWER**

The first thing to note here is that $2\pi$ is just a number, so, according to Rule 1 it should be ignored. We can now use a 'proportionality' (∝) symbol instead of an equals sign to represent the proposition…

$$t \propto \sqrt{\frac{F}{ml}}$$

$$\left[\sqrt{\frac{F}{ml}}\right] = \left(\frac{MLT^{-2}}{ML}\right)^{0.5} = (T^{-2})^{0.5} = T^{-1}$$

This proposed formula has dimensions of $T^{-1}$ which are clearly not the dimensions of time (the period of vibration of the guitar string, which has dimensions of T). Therefore, we have shown, using dimensional analysis, that the proposed answer is invalid.

**Worked Example 3**

Two positively charged particles, $Q_1$ and $Q_2$, are separated by a distance r. Each particle has a mass m. We are unsure whether Q, r and m are all required in an equation to represent the force (F) between the particles. Use dimensional analysis to develop a formula which represents the force between the particles.

**ANSWER**

We commence with an overview of the quantities which are possibly involved…

**F** = force **$MLT^{-2}$**

**$Q_1$** = charge **$L^{1.5}M^{0.5}T^{-1}$**

**$Q_2$** = charge **$L^{1.5}M^{0.5}T^{-1}$**

**r** = distance **L**

**$m_1$** = mass **M**

**$m_2$** = mass **M**

We may now suggest a general expression regarding the force, F…

$$F = function\{Q, r, m\}$$

Which simply means that force is possibly a function of (or governed by) charge, distance and mass.

To make progress towards a solution, we prefer to represent the problem as follows…

$$F \propto Q^A r^B m^C$$

and then we may say…

$$[F] = [Q^A r^B m^C]$$

***Our job will be to find values for those unknowns A, B and C***.

We now replace those quantities with their dimensions…

$$MLT^{-2} = (L^{1.5}M^{0.5}T^{-1})^A L^B M^C$$

Now let's bring together those dimensions and powers…

$$MLT^{-2} = L^{1.5A+B}M^{0.5A+C}T^{-A}$$

Now we look at each dimension in turn and equate those powers, left to right…

From L: 1 = 1.5A + B [1]

From M: 1 = 0.5A + C [2]

From T: -2 = -A [3]

From [3]: **A = 2** [4]

Sub [4] into [1]: 1 = 1.5(2) + B so **B = -2**

Sub [4] into [2]: 1 = 0.5(2) + C so **C = 0**

Almost there. Earlier we had…

$$F \propto Q^A r^B m^C$$

Since we know that C = 0 then the 'm' term disappears (anything to the power zero is 1). We also know values for A and B…

$$F \propto Q^A r^B m^C$$

$$F \propto Q^A r^B m^0$$

$$F \propto Q^A r^B$$

$$F \propto Q^2 r^{-2}$$

$$\therefore \quad F \propto \frac{Q^2}{r^2}$$

An important thing to note here; we suspected that the mass of the particles may have had an influence on the force between them. Our dimensional analysis determined that this was not the case – ***the mass was irrelevant***. Also, the top of our solution contains a charge squared term, and we know that we had two charges, $Q_1$ and $Q_2$, so we can confidently write…

$$F \propto \frac{Q_1 Q_2}{r^2}$$

This result looks remarkably like Coulomb's Law.

## 1.2 Arithmetic and geometric progressions

**Purpose**

You will learn about the notation used for sequences, arithmetic and geometric progressions, the limit of a sequence, sigma notation, the sum of a series, arithmetic and geometric series, Pascal's triangle and the binomial theorem. Much of this material has uses in the areas of Digital Electronics, Probability and Statistics.

**Terminology**

**Progression**: A sequence of numbers, separated by commas, following a definite pattern

**Series**: A sequence of numbers, separated by + or - signs, following a definite pattern

**Convergent**: A series which has a limited sum when an infinite number of terms are added

**Divergent**: A series which tends to infinity when an infinite number of terms are added

### Sequence Notation and Progressions

You should be aware that many printed and electronic texts tend to blur the distinction between a sequence, progression and series. However, whenever we use the word 'sum' we are talking about a series, since there are addition signs between terms.

You can think of a progression as a sequence of terms which are formed in a regular pattern.

1, 5, 9, 13, 17,... is a progression. In fact, this is an arithmetic progression, since each term is formed by adding a common difference (4) to the previous term.

2, 6, 18, 54, 162,... is a progression. In fact, this is a geometric progression, since each term is formed by multiplying the previous term by a common factor (3).

3, 6, 7, 11, 19,... is NOT a progression, since there is no regular pattern to the numbers.

A progression may contain either an infinite or finite number of terms.

### The Limit of a Sequence

For an arithmetic sequence, where the common difference is non-zero, we can clearly see that adding more and more terms will make the sum of the terms (the series) tend to get very large indeed. In fact, if we add an infinite number of terms to such a series then it will tend to $\pm$ infinity (symbol $\infty$) and is said to be *divergent*.

For a geometric sequence we could find the series tending to a particular value, or possibly tending to infinity. The limiting value will depend upon the common multiplying factor. Such series are said to be *convergent* or *divergent* respectively.



edexcel Approved Centre

## Sigma Notation

Let's suppose we wish to sum all the integer terms between 1 and 10. The long way to do this would be to write…

$$S_{10} = 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10$$

and we would work that out to be 55, right? A shorthand way to express this problem is to say…

$$S_{10} = 1 + 2 + 3 + \cdots + 8 + 9 + 10$$

That saved a bit of time and ink. There is an even better way to express the problem, using Sigma Notation. Here we use the Greek capital letter Sigma ($\sum$) to denote the sum of terms…

$$\sum_{i=1}^{10} i = 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10$$

We have simply used the letter $i$ as a convenient counter, nothing more. We can then let $i$ be any number we want between our limits, indicated below and above Sigma as 1 and 10 respectively.

If you were ever asked write down the sum of the first 1000 integers you would not need to write them all out. You would just use Sigma notation, as follows…

$$\sum_{i=1}^{1000} i$$

How neat is that?

For a *geometric* progression we can adopt a similar approach. Suppose we had a geometric progression such as…

$$2, 4, 8, 16, 32, 64, 128, 256$$

It is clear here that each new term is constructed by multiplying the previous term by 2. That multiplier is known as the 'common ratio' and is given the symbol '$r$'. We always have a first term, of course, and we give this the symbol '$a$'. We could introduce $i$ as a counter again and write the Sigma notation for our series as…

$$\sum_{k=0}^{7} ar^k$$

Let's see whether this works…

$$\sum_{i=0}^{7} ar^k = 2(2)^0 + 2(2)^1 + 2(2)^2 + 2(2)^3 + 2(2)^4 + 2(2)^5 + 2(2)^6 + 2(2)^7$$

You can easily go about evaluating the answer by adding up all eight terms but that would be slow and error-prone.

**Video**

These videos will explain Progressions and Sigma Notation in a useful way.

The next section is very useful because it shows you handy formulae to solve such tedious problems in a quick and efficient manner.

## Progressions

Let's take a closer look at arithmetic progressions. Consider the infinite progression 1, 3, 5, 7, 9,... and decide what the sum of the first five numbers will be. No doubt you will very quickly say 25. What if you were asked to find the sum of the first five thousand terms? You certainly would not adopt the summing technique just employed. The formula which we need to use to calculate the sum of n terms in an arithmetic progression is...

$$\boldsymbol{S_n = \frac{n}{2}(2a + (n-1)d)}$$

*where $S_n$ is the sum of the $n$ terms, $a$ is the first term and $d$ is the common difference*

**Worked Example 4**

Let's see if we can find the sum of those first five thousand terms, mentioned above, using our new formula…

$$\boldsymbol{S_{5000} = \frac{5000}{2}(2(1) + (5000-1)2) = 25,000,000}$$

For an arithmetic progression we may also find the $n^{th}$ term by using the formula…

$$\boldsymbol{n^{th} term = a + (n-1)d}$$

**Worked Example 5**

**Find the $99^{th}$ term in the sequence 1, 3, 5, 7, 9,...**

$$\boldsymbol{n^{th} term = a + (n-1)d}$$

$$\boldsymbol{99^{th} term = 1 + (99-1)2 = 197}$$

That was far simpler than writing out all the terms and noting the $99^{th}$ term.

We can also present formulae to quickly ascertain the sum and individual terms for *geometric* series…

$$\boldsymbol{S_n = \frac{a(1-r^n)}{1-r}}$$

$$\boldsymbol{n^{th} term = ar^{n-1}}$$

**Worked Example 6**

**Given the infinite geometric progression 4, -12, 36, -108, 324, -972… find the sum of the first 10 terms and the value of the 8th term.**

The first thing to note is that the first term is 4, so a = 4. What's our common ratio (r) here? Well, you can easily find that by dividing the 2nd term by the 1st term…

$$r = \frac{-12}{4} = -3$$

We can now use our new formulae to find the solutions directly…

$$\sum_{k=1}^{10} ar^k = S_{10} = \frac{4(1-(-3)^{10})}{1-(-3)} = -59048$$

$$8^{th} term = ar^{n-1} = 4(-3)^{8-1} = -8748$$

The table below is a handy summary of formulae for calculations involving series.

| Series Type | $n^{th} term$ | Sum of $n$ terms ($S_n$) |
|---|---|---|
| **Arithmetic Series** | $n^{th} term = a + (n-1)d$ | $S_n = \frac{n}{2}(2a + (n-1)d)$ |
| **Geometric Series** | $n^{th} term = ar^{n-1}$ | $S_n = \frac{a(1-r^n)}{1-r}$ |

## Pascal's Triangle

Let's now take a look at Pascal's Triangle.

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1
\end{array}
$$

Snooker and Pool come to mind here. Let's do some Maths first. The first six rows of Pascal's triangle are shown above. You may construct the triangle by starting at the top (row 0) and placing a 1 there. To complete the next row down you just move to the required position and add up the numbers you see at 10 o' clock and 2 o' clock. It's as simple as that.

What use is this triangle? Well, the rows featured in Pascal's triangle form the coefficients in *binomial expansions*. Let us consider some binomial expansions...

$(x+y)^2 = x^2 + 2xy + y^2$ Coefficients feature row 2 of Pascal's triangle

$(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$ Coefficients feature row 3 of Pascal's triangle

$(x+y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4$ Coefficients feature row 4 of Pascal's triangle

## The Binomial Theorem

If we take $n$ as a positive integer then the Binomial Theorem states...

$$(a+x)^n = a^n + na^{n-1}x + \frac{n(n-1)}{2!}a^{n-2}x^2 + \frac{n(n-1)(n-2)}{3!}a^{n-3}x^3 + \cdots + x^n$$

We shall be making full use of the binomial series when we look at 'power series' in the next section.

**Video**

These videos will boost your knowledge on the Binomial Theorem.

**Questions**

Q1 For the infinite progression 2, 4, 6, 8, 10,... find the 22nd term and the sum of the first 1000 terms.

Q2 For the infinite progression -10, -20, -30, -40,... find the 31st term and the sum of the first 20 terms.

Q3 For the infinite progression 1, 2, 4, 8,... find the 20th term and the sum of the first 16 terms.

Q4 For the infinite progression -3, -6, -12, -24,... find the 22nd term and the sum of the first 9 terms.

Check your answers with this handy online calculator.

**ANSWERS**

| | |
|---|---|
| Q1 | 44, 1001000 |
| Q2 | -310, -2100 |
| Q3 | 524288, 65535 |
| Q4 | -6291456, -1533 |

**Challenge**

**Build the top ten rows of Pascal's triangle**

# 2 Functions

**Theory**

There are six main types of Algebraic Function…

**Linear Function** *of type* $\boldsymbol{y = mx + c}$

**Quadratic Function** *of type* $\boldsymbol{f(x) = ax^2 + bc + c}$ where $a \neq 0$

**Cubic Function** *of type* $\boldsymbol{f(x) = ax^3 + bx^2 + cx + d}$ where $a \neq 0$

**Polynomial Function** *of type* $\boldsymbol{f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0}$

**Rational Function** *of type* $\boldsymbol{f(x) = a(x)/b(x)}$ where $a(x)$ and $b(x)$ are Polynomial Functions

**Radical Function** *of type* $\boldsymbol{f(x) = \sqrt[n]{g(x)}}$

Each of these types of algebraic function are commonly used throughout all disciplines of engineering.

## 2.1 Exponential versus Logarithmic Functions

The logarithmic function is the inverse of the exponential function. You will remember from school work that $1000 = 10^3$ and if we take the logarithm to base 10 of BOTH sides we get…

$$\boldsymbol{log_{10}(1000) = log_{10}(10^3)}$$

You will also know that the logarithm to base 10 of 1000 really means ***raise 10 to what number to give 1000?*** You know that the answer is 3, of course. Let's review what we've just said…

$$\boldsymbol{log_{10}(1000) = log_{10}(10^3) = 3}$$

So the 'log' part has obliterated the '10' part, leaving just that 3. We can therefore confirm that the logarithmic function is the inverse of the exponential function.

**Theory**

## 2.2 Reduction of Exponential Laws to Linear Form

Assume you had an equation such as…

$$\boldsymbol{P = Ae^{kh}} \qquad \text{(where } A \text{ and } k \text{ are constants)}$$



edexcel Approved Centre

This is clearly in exponential form (because of the $e^{kh}$ part). To reduce this to linear form we need to take the logarithm of both sides. The natural logarithm is needed, matching the natural constant 'e', so we shall use $log_e$ in our notation…

$$\boldsymbol{log_e(P) = log_e(Ae^{kh})}$$

Employing the laws of logarithms learned at level 3…

$$\boldsymbol{log_e(P) = log_e(A) + log_e(e^{kh})}$$

$$\therefore \boldsymbol{log_e(P) = kh + log_e(A)}$$

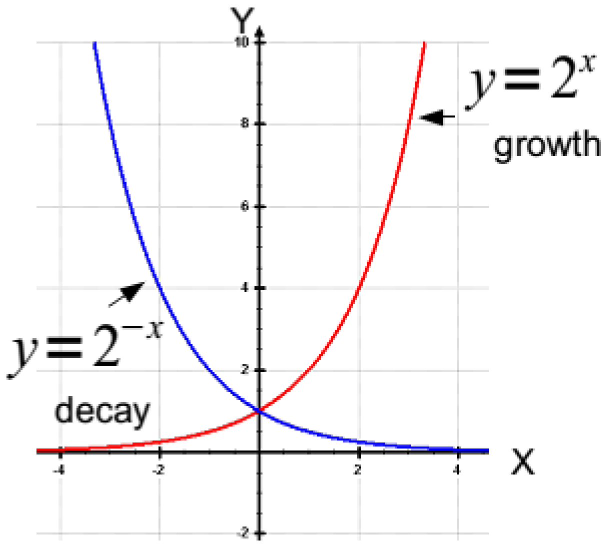
This result looks remarkably like our standard linear form $y = mx + c$ . Here $h$ is the variable rather than $x$. Thus, we have managed to reduce the exponential form of an equation to its linear form. Such manipulations will be very useful when you are solving equations or transposing formulae.

Video

Check these videos if you need to revise logarithms.

## 2.3 Solution of Exponential and Logarithmic Equations

Below is a plot of the curves $f(x) = 2^x$ and $f(x) = 2^{-x}$. You will notice that they exhibit growth and decay respectively.




We may choose any base we like and manipulate exponents when multiplied...

$$\boldsymbol{(4^x)(4^{x+1}) = 4^{(x)+(x+1)} = 4^{2x+1}}$$

Choosing another arbitrary base we may see the division rule in action...

$$\boldsymbol{\frac{6^{2x+4}}{6^{x+3}} = 6^{(2x+4)-(x+3)} = 6^{x+1}}$$

Engineers very often meet the exponential function $e^x$ and its variants. The symbol '$e$' was chosen by an 18th century mathematician named Leonhard Euler, and is known as 'Euler's number'. The number '$e$' is related to many situations in the natural universe, including…

- initial growth/decay of biological organisms
- population growth/decay
- radioactive decay
- a charging/discharging capacitor
- an energising/de-energising inductor
- financial profit/loss

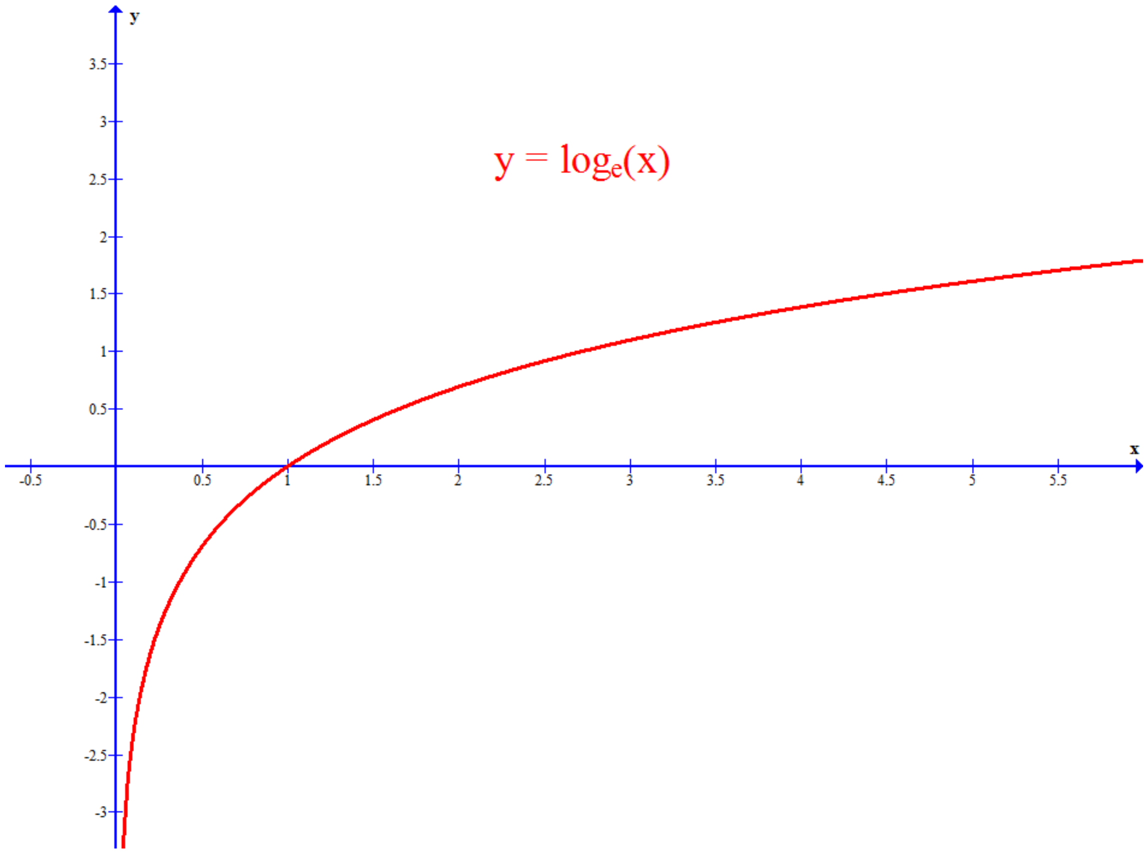
Regarding the last point on this list, consider that you have £1 to invest with a bank. The bank generously allows you to accumulate interest as many times as you like throughout the course of your deposit. Assume that the annual interest rate quoted to you was a generous 100%. You decide that you would like to know which interest period will yield the most money. You can then use the compound interest formula to calculate the value of your investment, in each case, after a full year…

| Interest period | calculation | account balance (£) |
|---|---|---|
| annual | $(1+\frac{1}{1})^{1}$ | 2.000000 |
| quarterly | $(1+\frac{1}{4})^{4}$ | 2.441406 |
| monthly | $(1+\frac{1}{12})^{12}$ | 2.613035 |
| weekly | $(1+\frac{1}{52})^{52}$ | 2.692597 |
| daily | $(1+\frac{1}{365})^{365}$ | 2.714567 |
| each minute | $(1+\frac{1}{262800})^{262800}$ | 2.718277 |
| each second | $(1+\frac{1}{15678000})^{15678000}$ | 2.718282 |

As you can see, it is more profitable to have interest calculated each second. You will also notice that the amount is reaching a limit. This limit is in fact '$e$', which, to 9 decimal places, is 2.718281828. Therefore, there will be no tangible increased profit from calculations every millisecond or microsecond. Now you know why banks like to pay interest annually, rather than more frequently!

So then, what Euler actually discovered was a universal constant of growth and decay. This constant can be applied to the analysis of any *natural* growth or decay scenario. Humans generally have 10 fingers (if you count thumbs as fingers). When you were at school you learned about $log_{10}$ of numbers. The base of 10 was chosen as an obvious human reference number. As an engineer you will almost always discard the base 10 and use the universal constant '$e$' as the regular base. You might see this function displayed on your calculator buttons as $l_n$, $log_n$ or $log_e$ but they are all the same thing.

Let's take a look at the function $y = log_e(x)$…




You will notice that $log_e(1) = 0$ since the graph crosses $y = 0$ when $x = 1$. Think of this as $e^0 = 1$, remembering that anything to the power 0 is 1.

A rule of logarithms learned at level 3 is…

$$\boldsymbol{log_e(x^n) = n\, log_e(x)}$$

This rule will be used very often to transpose formulae which you encounter in engineering problems. Let's take a look at one such problem…

**Worked Example 7**

**The instantaneous voltage ($v_c$) across a capacitor ($C$), having an initial peak voltage ($V_p$), is discharged through a resistor ($R$). The formula for the voltage across the capacitor is given by…**

$$\boldsymbol{v_c = V_p\, e^{(-t/RC)}}$$

**Rearrange this formula to find a formula for $C$.**

Our first thought in this problem is that the $V_p$ term may be moved across diagonally…

$$\frac{v_c}{V_p} = e^{(-t/RC)}$$

We know we are trying to isolate that $C$ on the right hand side, but it is trapped inside the exponential '$e$'. Remembering that logs are the inverses of exponents, we should be able to take logs of *both* sides and thereby release $C$. Let's try that…

$$log_e\left(\frac{v_c}{V_p}\right) = log_e\left(e^{(-t/RC)}\right)$$

$$log_e\left(\frac{v_c}{V_p}\right) = \frac{-t}{RC}$$

Using our transposition skills we may then write…

$$C = \frac{-t}{R\ log_e\left(\frac{v_c}{V_p}\right)}$$

That was quite useful. Let's look at another example.

**Worked Example 8**

**A coil of inductance $L$ Henries is energised by a current which also flows through a series-connected resistor ($R$). The resultant voltage across this RL combination is $V_s$. The voltage across the coil ($v_L$) is given by…**

$$v_L = V_s\, e^{(-Rt/L)}$$

**Rearrange this formula to find a formula for $L$.**

Our first thought in this problem is that the $V_s$ term may be moved across diagonally…

$$\frac{v_L}{V_s} = e^{(-Rt/L)}$$

Again remembering that logs are the inverses of exponents we should be able to take logs of *both* sides and thereby release $L$. Let's try that…

$$log_e\left(\frac{v_L}{V_s}\right) = log_e\left(e^{(-Rt/L)}\right)$$

$$\therefore\ log_e\left(\frac{v_L}{V_s}\right) = \frac{-Rt}{L}$$

Transposing this gives…

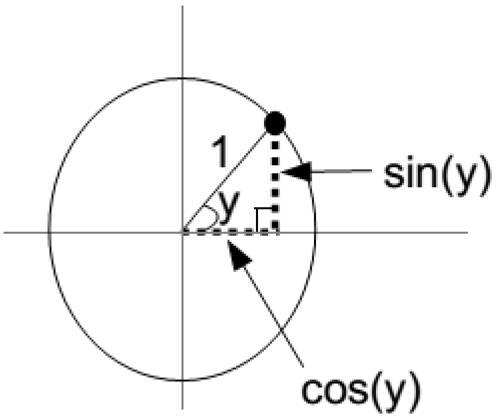
$$L = \frac{-Rt}{log_e\left(\frac{v_L}{V_s}\right)}$$

Video

Check these videos to enhance your skills on transposition of formulae.

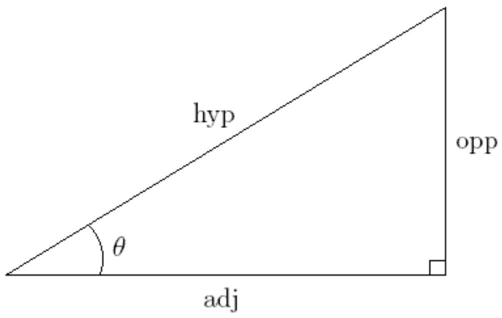
## 2.4 Trigonometric and Hyperbolic Identities

So far you have been used to thinking of trigonometry as the analysis of right-angled triangles. The graph below uses a unit circle to define the sine and cosine functions. You will notice that the triangle is contained within the circle.

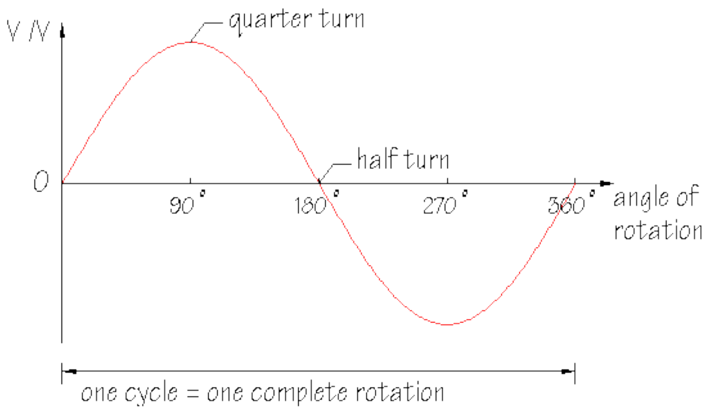



If you imagined this circle to be an electrical generator, spinning anti-clockwise with the dot starting at 3 o' clock, then measurement of the height of the dot versus time would trace out a very familiar sine wave. The mains electricity supply has a sinusoidal nature and this is produced by a circular generator at the power station. For this reason we can say that ***trigonometric functions such as sin, cos and tan are CIRCULAR functions***.

A quick reminder about the common circular functions…



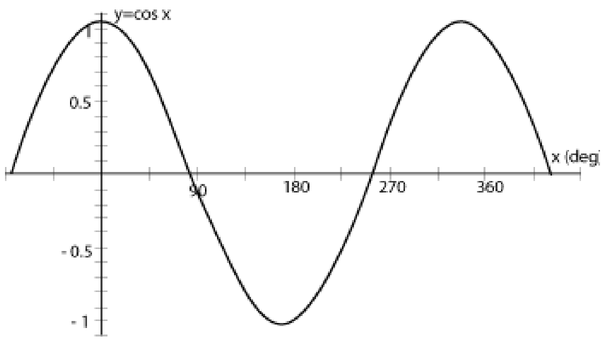

$$\sin(\theta) = \frac{opp}{hyp}; \quad \cos(\theta) = \frac{adj}{hyp}; \quad \tan(\theta) = \frac{opp}{adj}$$




If we say $\sin(90^o) = 1$ then what this means, in relation to the sine wave above, is that the sine function reaches an amplitude of 1 at $90^o$. This is reading the function from the bottom up and then left. What if we were to read the function from the left first of all and then across and to the bottom? This could be expressed mathematically as…

$$\boldsymbol{sin^{-1}(1) = 90}$$

By the same token we could look at a cosine wave…




Here we can say…

$$\boldsymbol{cos(0) = 1}$$

and we can inverse this as…

$$\boldsymbol{cos^{-1}(1) = 0}$$

These inverse trigonometric functions are very useful to us because they allow the release of variables which are 'trapped' inside circular functions. When $sin^{-1}$ directly meets $sin$ then the two disappear (handy). The same goes for $cos^{-1}$ directly meeting $cos$.

**Worked Example 9**

**An instantaneous signal voltage $(v_s)$ is described by the equation…**

$$\boldsymbol{v_s = 10\ sin(2\pi ft - \pi/8)}$$

**Make $f$ the subject of this formula.**

We commence our solution by tidying up a bit and then observing that our target variable $f$ is 'trapped' inside the sine function. Taking $sin^{-1}$ of both sides should free up our $f$ variable…

$$\boldsymbol{\frac{v_s}{10} = sin(2\pi ft - \pi/8)}$$

$$\therefore\ \boldsymbol{sin^{-1}\left(\frac{v_s}{10}\right) = sin^{-1}\left(sin(2\pi ft - \pi/8)\right)}$$

$$\therefore\ \boldsymbol{sin^{-1}\left(\frac{v_s}{10}\right) = 2\pi ft - \pi/8}$$

$$\therefore\ \boldsymbol{sin^{-1}\left(\frac{v_s}{10}\right) + \frac{\pi}{8} = 2\pi ft}$$

$$\therefore\ \boldsymbol{f = \frac{sin^{-1}\left(\frac{v_s}{10}\right) + \frac{\pi}{8}}{2\pi t}}$$

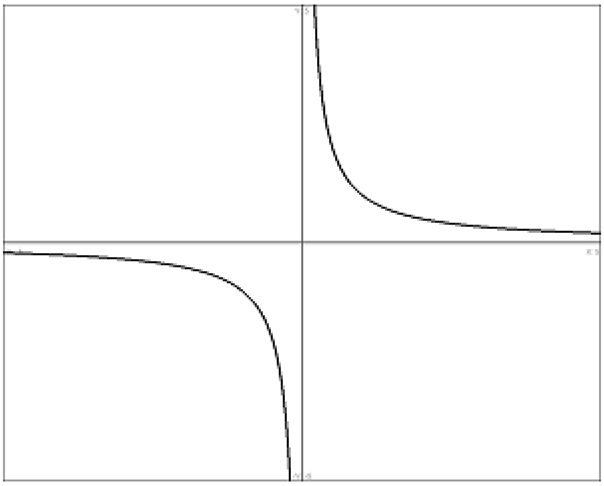
**NOTE: Here the angle involves π. Whenever π is involved in the angle you MUST have your calculator in radians mode (RAD) otherwise you will be in degrees mode and find the wrong answer.**

There are many facets of science, engineering and nature where simple trigonometry is an inadequate tool for analysis. Let's move on to look at HYPERBOLIC functions. Hyperbolae arise in many situations, such as...

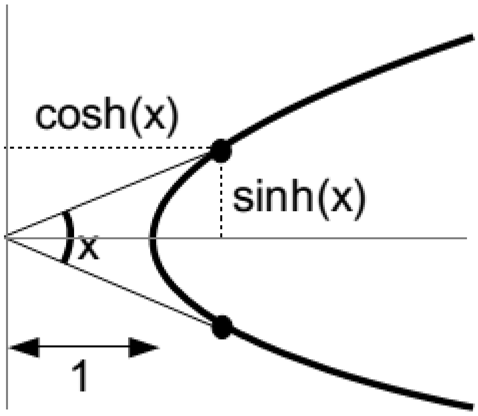
- The path of a spacecraft in open orbit around a planet and exceeding the planet's escape velocity
- The trace of a sundial shadow

- The shape of a hanging cable, chain or rope
- The scattering trajectory of a subatomic particle acted upon by a repulsive force
- The current at any point along a transmission line exhibiting leakage
- The shape of an object produced by cutting a vertical section through a cone

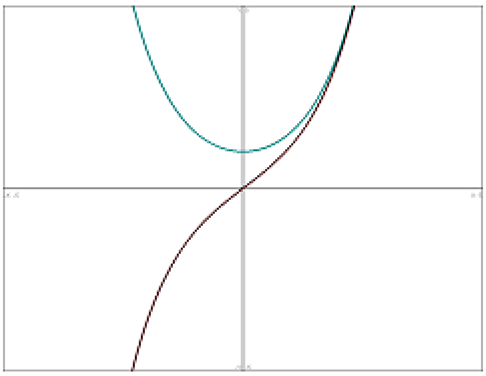
A simple hyperbola is produced by plotting the graph of $y = 1/x$...



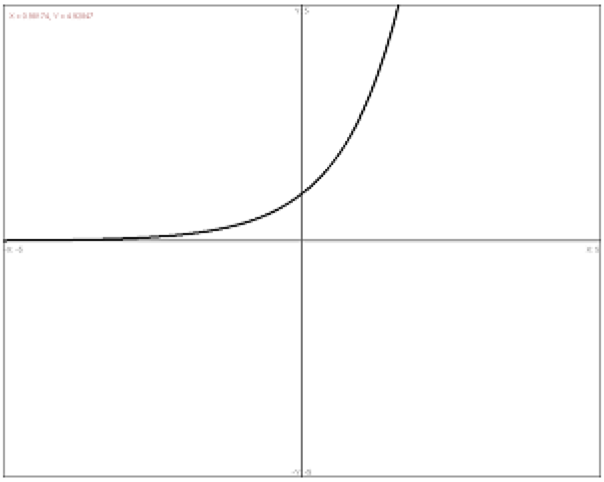
The graphic below helps us to analyse the hyperbola. From this we may define the hyperbolic sine and hyperbolic cosine functions (pronounced 'shine' and 'cosh' respectively). Notice that the apex of the hyperbola is one unit away from the vertical axis (much akin to our unit circle).




Let's take a look at the combined graphs of $\sinh(x)$ and $\cosh(x)$...



The black curve represents $\sinh(x)$ and the curve resembling a hanging cable represents $\cosh(x)$. An apparently strange thing happens if we add these two graphs together...



We end up with a graph of $e^x$, as above. Let's write down mathematically what we have just noticed graphically...

$$e^x = cosh(x) + sinh(x)$$

Now, rather than add the $\cosh(x)$ and the $\sinh(x)$ we can also subtract them. This will give us a graph of $e^{-x}$. Now we may also write...

$$e^{-x} = cosh(x) - sinh(x)$$

If we add these last two results we get...

$$cosh(x) = \frac{e^x + e^{-x}}{2}$$

and subtracting the second from the first gives...

$$sinh(x) = \frac{e^x - e^{-x}}{2}$$

We may also try $cosh^2(x) - sinh^2(x)$ and realise that this comes out to be 1...

$$cosh^2(x) - sinh^2(x) = 1$$

We can also divide $\sinh(x)$ by $\cosh(x)$ ...

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)}$$

These last four results are highlighted in blue because they are extremely important when you need to take your maths knowledge through to level 5 and beyond. Those topics will involve Complex Numbers, covered in Unit 35: Further Analytical Methods for Engineers.

Video

These videos will further boost your understanding of hyperbolic functions.

The table below lists the trigonometric and hyperbolic identities which you might meet most often.

| Trigonometric Identities | | Hyperbolic Identities |
|---|---|---|
| $cosec(x) = 1/\sin(x)$ | | $cosech(x) = 1/\sinh(x)$ |
| $\sec(x) = 1/\cos(x)$ | | $\text{sech}(x) = 1/\cosh(x)$ |
| $\cot(x) = 1/\tan(x)$ | | $\coth(x) = 1/\tanh(x)$ |
| $\tan(x) = \sin(x)/\cos(x)$ | | $\tanh(x) = \sinh(x)/\cosh(x)$ |
| $sin^2(x) + cos^2(x) = 1$ | | $cosh^2(x) - sinh^2(x) = 1$ |
| $\sin(-x) = -\sin(x)$ | | $\sinh(-x) = -\sinh(x)$ |
| $\cos(-x) = \cos(x)$ | | $\cosh(-x) = \cosh(x)$ |
| $\tan(-x) = -\tan(x)$ | | $\tanh(-x) = -\tanh(x)$ |

## 2.5 Solution of Equations containing Hyperbolic Functions

**Worked Example 10**

**Solve the equation $\sinh(x) = 4$, correct to 3 significant figures.**

The first step to finding $x$ here is to remember our definition for $\sinh(x)$ …

$$\boldsymbol{sinh(x) = \frac{e^x - e^{-x}}{2}}$$

Therefore, our question is really asking us to find $x$ given that…

$$\boldsymbol{\frac{e^x - e^{-x}}{2} = 4}$$

Multiply both sides by 2…

$$\boldsymbol{e^x - e^{-x} = 8}$$

$$\boldsymbol{\therefore\ e^x - e^{-x} - 8 = 0}$$

What do we do now? Well, if we multiply both sides by $e^x$ then a quadratic equation will appear. Fortunately we know a formula for those from our school days. Let's multiply first by $e^x$ and develop things towards that wanted quadratic form…

$$\boldsymbol{e^x(e^x - e^{-x} - 8) = e^x \times 0 = 0}$$

$$\boldsymbol{\therefore\ (e^x)^2 - e^x \cdot e^{-x} - 8e^x = 0}$$

$$\boldsymbol{\therefore\ (e^x)^2 - e^0 - 8e^x = 0}$$

$$\boldsymbol{\therefore\ (e^x)^2 - 1 - 8e^x = 0}$$



edexcel Approved Centre

$$\therefore\ (e^x)^2 - 8e^x - 1 = 0$$

We can't see a straightforward way to find factors here so we must use the quadratic formula which we loved so much back in our school days. Here's a quick reminder for you…

Given $ax^2 + bx + c = 0$ we may solve for $x$ by using the formula $x = \frac{-b\pm\sqrt{b^2-4ac}}{2a}$.

In our problem we have:

$a = 1, b = -8, c = -1$

Let's plug in our values to attempt a solution…

$$e^x = \frac{-(-8) \pm \sqrt{(-8)^2 - 4(1)(-1)}}{2(1)} = \frac{8 \pm \sqrt{68}}{2} = \frac{8 \pm 8.246}{2} = 8.123 \quad or \quad -0.123$$

Taking logs of both sides will give us…

$$x = log_e(8.123) \;\; or \;\; x = log_e(-0.123)$$

That last solution is no good because we cannot take the log of a negative number (not a real one anyway). So, the first one is what we're after…

$$x = log_e(8.123)$$

Trying that on your calculator gives $x = 2.10$, correct to 3 significant figures.

The handy inverse trigonometric functions looked at earlier also have their counterparts in the hyperbolic realm. These are known as $sinh^{-1}$, $cosh^{-1}$ and $tanh^{-1}$. Let's look at an example…

**Worked Example 11**

**Find $x$ if $\sinh(x) = 0.521$**

We simply apply the inverse, $sinh^{-1}$, to both sides…

$$\sinh^{-1}(\sinh(x)) = \sinh^{-1}(0.521)$$

$$\therefore\ x = sinh^{-1}(0.521) = 0.5$$

**Questions**

Q5 Rearrange $v_c = V_p\, e^{(-t/RC)}$ to make t the subject.

Q6 Rearrange $v_L = V_s\, e^{(-Rt/L)}$ to make R the subject.

Q7 Rearrange $v_s = 12\, sin(2\pi f t - \pi/16)$ to make t the subject

Q8 Solve the equation $\sinh(x) = 2$, correct to 3 decimal places.

Q9 Find $x$ if $\sinh(x) = 0.637$

**ANSWERS**

| | |
|---|---|
| Q5 | $t = -RC log_e\left(\frac{v_c}{V_p}\right)$ |
| Q6 | $R = -\frac{L}{t} log_e\left(\frac{v_L}{V_s}\right)$ |
| Q7 | $t = \frac{sin^{-1}\left(\frac{v_s}{12}\right) + {}^{\pi}/_{16}}{2\pi f}$ |
| Q8 | $1.444$ |
| Q9 | 0.6 |



edexcel Approved Centre

# Section 2: Statistical Techniques

# INTRODUCTION

Investigate applications of statistical techniques to interpret, organise and present data, by using appropriate computer software packages.

**Summary of data:**

**Mean and standard deviation of grouped data.**

**Pearson's correlation coefficient.**

**Linear regression.**

**Probability theory:**

**Binomial and normal distribution.**

# GUIDANCE

This document is prepared to break the unit material down into bite size chunks. You will see the learning outcomes above treated in their own sections. Therein you will encounter the following structures;

**Purpose**

Explains *why* you need to study the current section of material. Quite often learners are put off by material which does not initially seem to be relevant to a topic or profession. Once you understand the importance of new learning or theory you will embrace the concepts more readily.

**Theory**

Conveys new material to you in a straightforward fashion. To support the treatments in this section you are strongly advised to follow the given hyperlinks, which may be useful documents or applications on the web.

**Example**

The examples/worked examples are presented in a knowledge-building order. Make sure you follow them all through. If you are feeling confident then you might like to treat an example as a question, in which case cover it up and have a go yourself. Many of the examples given resemble assignment questions which will come your way, so follow them through diligently.

**Question**

Questions should not be avoided if you are determined to learn. Please do take the time to tackle each of the given questions, in the order in which they are presented. The order is important, as further knowledge and confidence is built upon previous knowledge and confidence. As an Online Learner it is important that the answers to questions are immediately available to you.. Contact your Unit Tutor if you need help.

**Challenge**

You can really cement your new knowledge by undertaking the challenges. A challenge could be to download software and perform an exercise. An alternative challenge might involve a practical activity or other form of research.

**Video**

Videos on the web can be very useful supplements to your distance learning efforts. Wherever an online video(s) will help you then it will be hyperlinked at the appropriate point.

# 2.1 Tabular and Graphical Form

## 2.1.1 Data Collection Methods

Data can be collected in a number of ways, including …

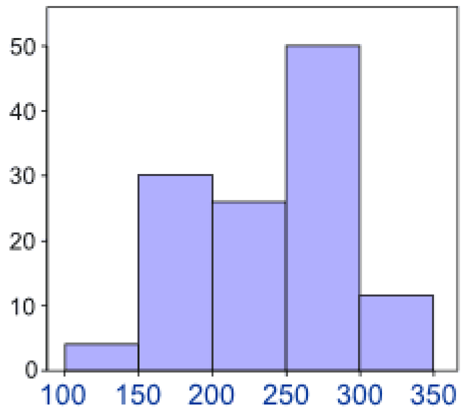
- Using test instruments such as multimeters, frequency counters, strain gauges etc.
- By observation – i.e. counting events, performance analysis
- Software simulation
- Wireless telemetry
- Artificial Intelligence
- Mobile phone applications

*Continuous* data can have any value, with potentially an unlimited number of decimal places in the data. An example is the voltage reading on an analogue voltmeter. *Discrete* data is obtained by some form of counting, such as recording the number of products on a production line.

There a several ways of presenting acquired data. These are discussed below.

## 2.1.2 Histograms

A histogram is a representation of data which is grouped into ranges. An example histogram is shown below…



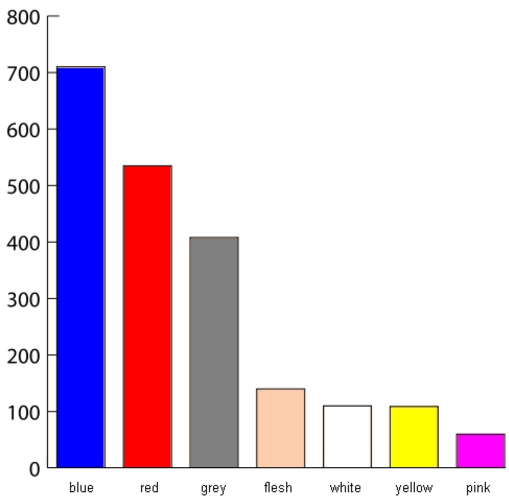

Here it can be seen, for example, that there were 30 occurrences of recorded data between 150 and 200 units.

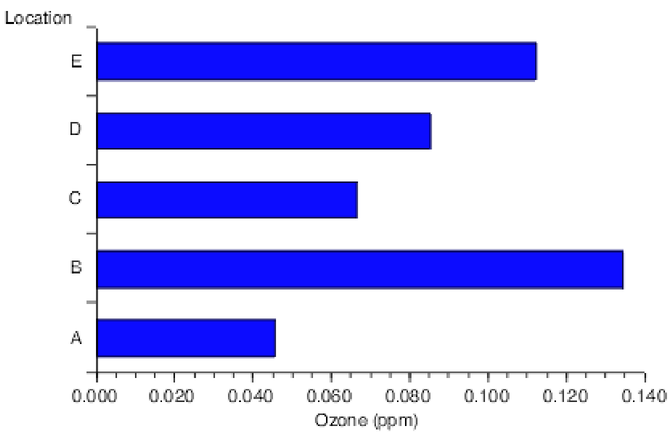
## 2.1.3 Bar Charts

A bar chart can represent data in a similar way to a histogram. There are two ***differences between a bar chart and a histogram***…

- A bar chart has spaces between the rectangles. A histogram does not.
- A bar chart can represent data as either horizontal or vertical rectangles. Histograms are vertical.

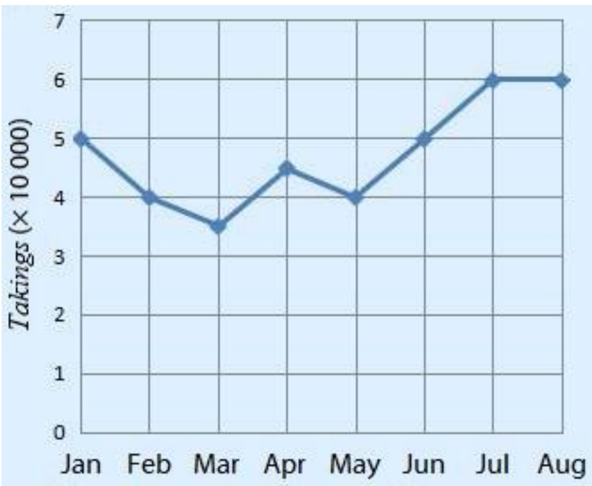
Examples of bar charts are shown below…



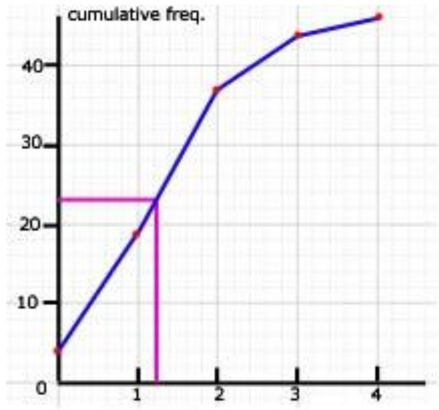

## 2.1.4 Line Diagrams

These represent discrete transitions to new data, marked by straight lines to specific data points, as per the example below…



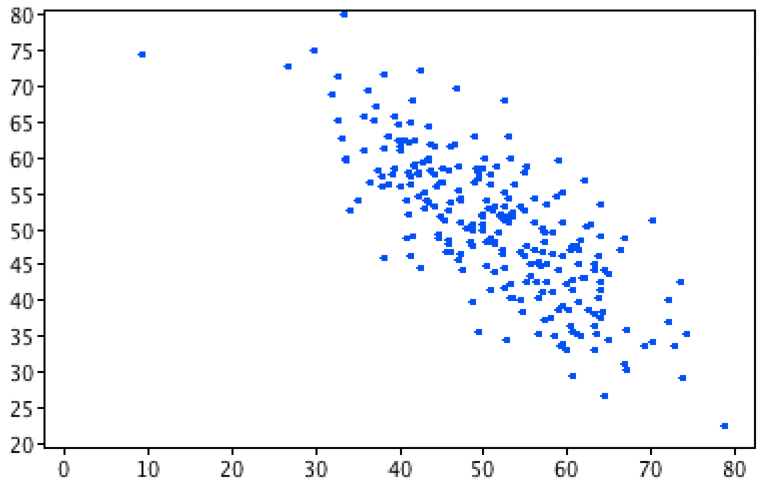
## 2.1.5 Cumulative Frequency Distribution Diagrams

These can be thought of as a running total in ascending order of the units on the horizontal axis. An example is shown below…




## 2.1.6 Scatter Plots

These are plotted with dots at matching vertices. Usually a line of 'best fit' may be drawn through the points. This line can be constructed using an equation or by estimation. A typical scatter plot is shown below…




## 2.1.7 Tally Charts

These are used to record observed data in groups. A vertical line is used to record an event. Usually a fifth event is recorded by inserting a diagonal line through the previous four vertical lines, as per the example below…

| *Mark* | *Tally* | *Frequency* |
|---|---|---|
| 2 | \| | 1 |
| 3 | \|\| | 2 |
| 4 | \|\|\| | 3 |
| 5 | \|\|\| | 3 |
| 6 | ~~\|\|\|\|~~ | 5 |
| 7 | ~~\|\|\|\|~~ \| | 6 |
| 8 | ~~\|\|\|\|~~ \|\| | 7 |
| 9 | \|\| | 2 |
| 10 | \| | 1 |

**Worked Example 1**

**The age of ten random students was recorded. The data is as follows…**

| **Sample** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Age (Years)** | **19** | **18** | **20** | **19** | **18** | **19** | **19** | **23** | **19** | **19** |

a) **Produce a Tally Chart to show the frequency of the students' ages.**
b) **Produce a Cumulative Frequency Distribution Curve for this data set.**

(a)

We see from the data that the youngest observed student is 18 years old and the oldest is 23 years old. These numbers are the bounds of our data in the tally chart…

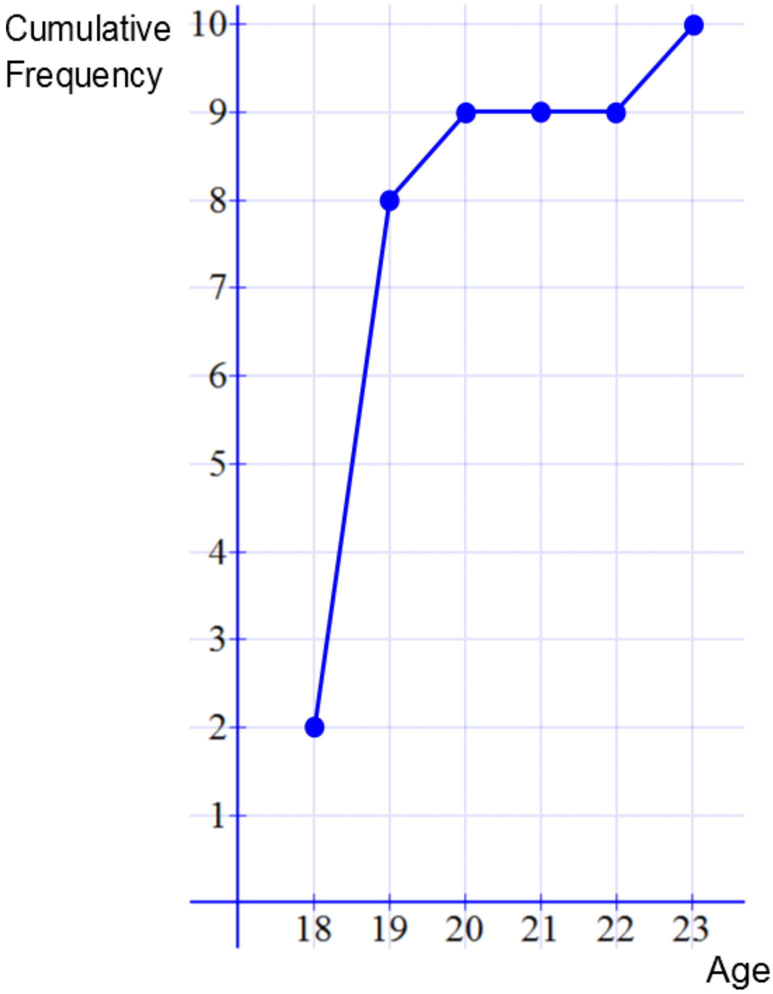
| **Age** | **Tally** | **Frequency** |
|---|---|---|
| **18** | \|\| | 2 |
| **19** | ~~\|\|\|\|~~ \| | 6 |
| **20** | \| | 1 |
| **21** | | 0 |
| **22** | | 0 |
| **23** | \| | 1 |



edexcel Approved Centre

(b)

We may start our axes anywhere we like. For this plot we choose the vertical and horizontal axes as shown below. Notice that the 'curve' never assumes a negative slope at any point since it is cumulative (a running total).




**Question**

Q2.1 **The resistance of 10 random samples from a $22\Omega$ resistor production line was recorded. The data is as follows…**

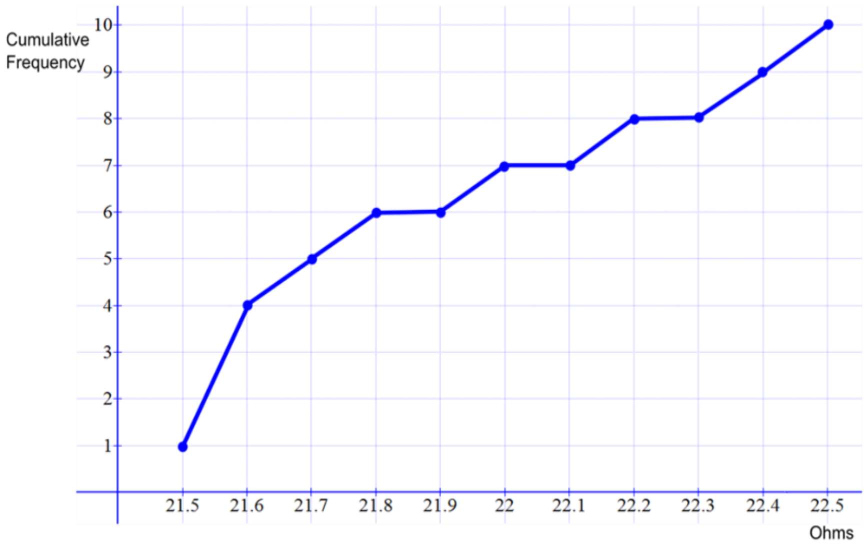
| **Sample** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Resistance ($\Omega$)** | **21.5** | **22.5** | **21.6** | **21.6** | **21.6** | **22.4** | **22.0** | **22.2** | **21.7** | **21.8** |

**a) Produce a Tally Chart to show the frequency of the resistances.**

**b) Produce a Cumulative Frequency Distribution Curve for this data set.**

**ANSWER**

| Resistance (Ω) | Tally | Frequency |
|---|---|---|
| 21.5 | \| | 1 |
| 21.6 | \|\|\| | 3 |
| 21.7 | \| | 1 |
| 21.8 | \| | 1 |
| 21.9 | | |
| 22.0 | \| | 1 |
| 22.1 | | |
| 22.2 | \| | 1 |
| 22.3 | | |
| 22.4 | \| | 1 |
| 22.5 | \| | 1 |




# 2.2 Central Tendency and Dispersion

## 2.2.1 The Concept of Central Tendency and Variance Measurement

A set of data may be represented by a single value. You will have most commonly used the term 'average' for a representative size a set of data. The correct mathematical term for this is the '*arithmetic mean*' or just '*mean'*.

On some occasions it might be more useful to represent the *most common* term in a set of data. This is termed the '*mode'* of the data. Another way is to order the data in ascending order and use the middle value, the *median*, as a representation of the data. A representative method of finding the central tendency of a data set is chosen to best reflect the nature of the data.

The idea of the *variance* of a set of data aims to quantify how much the data varies from the central point.

### 2.2.2 Mean, Median and Mode

Let us again consider the data from worked example 1 (reproduced below) to find the mean, median and mode…

| **Sample** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Age (Years)** | **19** | **18** | **20** | **19** | **18** | **19** | **19** | **23** | **19** | **19** |

The ***mean*** (average) is simply calculated by summing all the data and dividing by the number of data samples. Expressed mathematically, the mean is…

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{10} x_i = \frac{19+18+20+19+18+19+19+23+19+19}{10} = 19.3$$

Notice that we use a bar over the $x$ to represent its mean.

The ***median*** is calculated by ordering the data in ascending order and using the middle value, the median, as a representation of the data…

18, 18, 19, 19, ***19, 19***, 19, 19, 20, 23

Since we do not have an odd number of samples here then we cannot directly select the central value and call that the median. In situations like this we simply select the two central values (bold) of the ascending list and average them to find the median…

$$median = \frac{19+19}{2} = 19$$

To find the ***mode*** of the data we simply look for the most common data item. Here, this is clearly 19. Sometimes data will have two or more most common values, in which case the data is termed *bimodal* or *multimodal*.

### 2.2.3 Standard Deviation and Variance

To ascertain the amount of variation in a full given population, or sample of a population, we use the terms *standard deviation* and *variance*. Let's define these terms…

$$\boldsymbol{standard\ deviation\ for\ a\ full\ population, \sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}}$$

$$\boldsymbol{standard\ deviation\ sample\ of\ population, \sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}}$$

$$\boldsymbol{variance = \sigma^2}$$

The Greek lower case letter sigma ($\sigma$) is used for standard deviation and n is the number of samples. You will notice that the variance is simply the square of the standard deviation. Let's see these in use.

**Worked Example 2**

**Find the standard deviation and variance for the following population data set…**

| **Sample** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Age (Years)** | **19** | **18** | **20** | **19** | **18** | **19** | **19** | **23** | **19** | **19** |

We shall firstly find the standard deviation and then square that answer to get the variance. To find the standard deviation we normally construct a table, as follows…

| $\boldsymbol{x_i}$ | $\boldsymbol{x_i - \bar{x}}$ | $\boldsymbol{(x_i - \bar{x})^2}$ |
|---|---|---|
| 19 | 19.3-19=0.3 | 0.09 |
| 18 | 19.3-18=1.3 | 1.69 |
| 20 | 19.3-20=-0.7 | 0.49 |
| 19 | 19.3-19=0.3 | 0.09 |
| 18 | 19.3-18=1.3 | 1.69 |
| 19 | 19.3-19=0.3 | 0.09 |
| 19 | 19.3-19=0.3 | 0.09 |
| 23 | 19.3-23=-3.7 | 13.69 |
| 19 | 19.3-19=0.3 | 0.09 |
| 19 | 19.3-19=0.3 | 0.09 |
| $\bar{x} = \frac{1}{n}\sum_{i=1}^{10} x_i = 19.3$ | | $\sum(x_i - \bar{x})^2 = 18.1$ |

$$\therefore\ \boldsymbol{standard\ deviation, \sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} = \sqrt{\frac{18.1}{10}} = \sqrt{1.81} = 1.345}$$

$$\therefore \quad \boldsymbol{variance = \sigma^2 = 1.81}$$

**Question**

**Q2.2 (a) Calculate the mean, median and mode for the data set below.**

**(b) Find the standard deviation and variance for the population data set.**

| **Sample** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Salary (£1000's)** | **22** | **24** | **21** | **26** | **32** | **26** | **26** | **54** | **12** | **33** |

**ANSWER**

Q2.2

(a) Mean = £27,600, Median = £26,000, Mode = £26,000
(b) Standard Deviation = 10.41, Variance = 108.44

Check your answers with this handy online calculator.

## 2.2.4 Interquartile Range

Dispersion is sometimes measured in quartile values. Consider the data set…

1, 2, 2, 4, 6, 8, 9, 10, 10, 11, 12, 12, 13, 14, 17

This may be broken up into four quarters. To do this we need to make three splits. The points of these splits are known as the quartile values $Q_1, Q_2$ and $Q_3$. The value of $Q_2$ is the median and the other quartiles are determined by finding the medians to the left and right of $Q_2$. The data set below has these three quartile values highlighted in red…

1, 2, 2, **4**, 6, 8, 9, **10**, 10, 11, 12, **12**, 13, 14, 17

So, $Q_1 = 4, Q_2 = 10$ and $Q_3 = 12$. There is a range associated with a pair of these quartiles, known as the *interquartile range*. In this example the first interquartile range is (10 -4) = 6 and the second interquartile range is (12 – 10) = 2.

## 2.2.5 Application to Engineering Production

Much of the material covered so far in this workbook is used very frequently in quality control departments of manufacturing and assembly plants.

Very often in industry there will be grouped samples of data taken and these samples could be taken at random times or regular times. Each company will have its own in-house method of quality control and also data presentation.

Note that the formula for standard deviation of an incomplete sample is different than that for a full population. This is explained in the handy online calculator.

# 2.3 Regression and Linear Correlation

## 2.3.1 Regression Lines

When engineers are performing experiments or gathering data in some way it is sometimes useful to be able to plot the results and find the equation of a '*line of best fit*' between the points. This is done with *linear regression*. The standard way to express the equation of a straight line is by using the formula…

$$y = mx + c$$

Where $m$ is the slope of the line and $c$ is the intersection of the line with the $y$ axis. For linear regression analysis we do not use $m$ or $c$, instead we use $a_1$ and $a_0$ respectively…

$$y = a_1 x + a_0$$

Our task is to then find the values of $a_1$ and $a_0$ so as to produce an equation for our line of best fit. We use the following two equations to produce simultaneous equations which we then solve for $a_1$ and $a_0$…

$$\sum y_i = a_0 N + a_1 \sum x_i$$

$$\sum x_i y_i = a_0 \sum x_i + a_1 \sum {x_i}^2$$

…where N is the number of co-ordinates (points). Let's take a look at an example…

**Worked Example 3**

**A circuit component is encased in epoxy resin and therefore its type and value are unknown. In an effort to resolve this mystery an engineer places a number of voltages across the component and measures the consequent current flowing for each voltage. The results of her experiment are as follows…**

| **Applied Voltage [V]** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Measured Current [A]** | **8** | **16** | **17** | **23** | **25** | **34** | **35** | **39** | **41** | **47** |

a) **Determine an equation for the regression line of measured current against applied voltage.**
b) **Hence, determine what the applied voltage would have been for a current flow of 55A.**

(a)

To solve this problem we take a look at the standard simultaneous equations and construct a table which will facilitate determination of the terms required…

| $x_i$ | $y_i$ | $x_i y_i$ | ${x_i}^2$ |
|---|---|---|---|
| 1 | 8 | 8 | 1 |
| 2 | 16 | 32 | 4 |
| 3 | 17 | 51 | 9 |
| 4 | 23 | 92 | 16 |
| 5 | 25 | 125 | 25 |
| 6 | 34 | 204 | 36 |
| 7 | 35 | 245 | 49 |
| 8 | 39 | 312 | 64 |
| 9 | 41 | 369 | 81 |
| 10 | 47 | 470 | 100 |
| $\sum x_i = 55$ | $\sum y_i = 285$ | $\sum x_i y_i = 1908$ | $\sum {x_i}^2 = 385$ |

The number of co-ordinates N is 10. We may now use the summative data in our table to resolve the simultaneous equations…

$$\sum y_i = a_0 N + a_1 \sum x_i$$

$$\therefore \quad 285 = 10a_0 + 55a_1$$

$$\sum x_i y_i = a_0 \sum x_i + a_1 \sum {x_i}^2$$

$$\therefore \quad 1908 = 55a_0 + 385a_1$$

So we have…

$$285 = 10a_0 + 55a_1$$

$$1908 = 55a_0 + 385a_1$$

There are several ways to solve simultaneous equations such as these. The simplest way here is to multiply the first equation by $-55/10$ and then add to the second equation. That will eliminate $a_0$ and allow us to find $a_1$…

$$\frac{-55}{10}285 = \frac{-55}{10}10a_0 - \frac{55}{10}55a_1$$

$$1908 = 55a_0 + 385a_1$$

Adding these two equations gives…

$$-1567.5 + 1908 = (-302.5 + 385)a_1$$

$$\therefore \quad 340.5 = 82.5a_1$$

$$\therefore \quad a_1 = \frac{340.5}{82.5} = 4.1$$

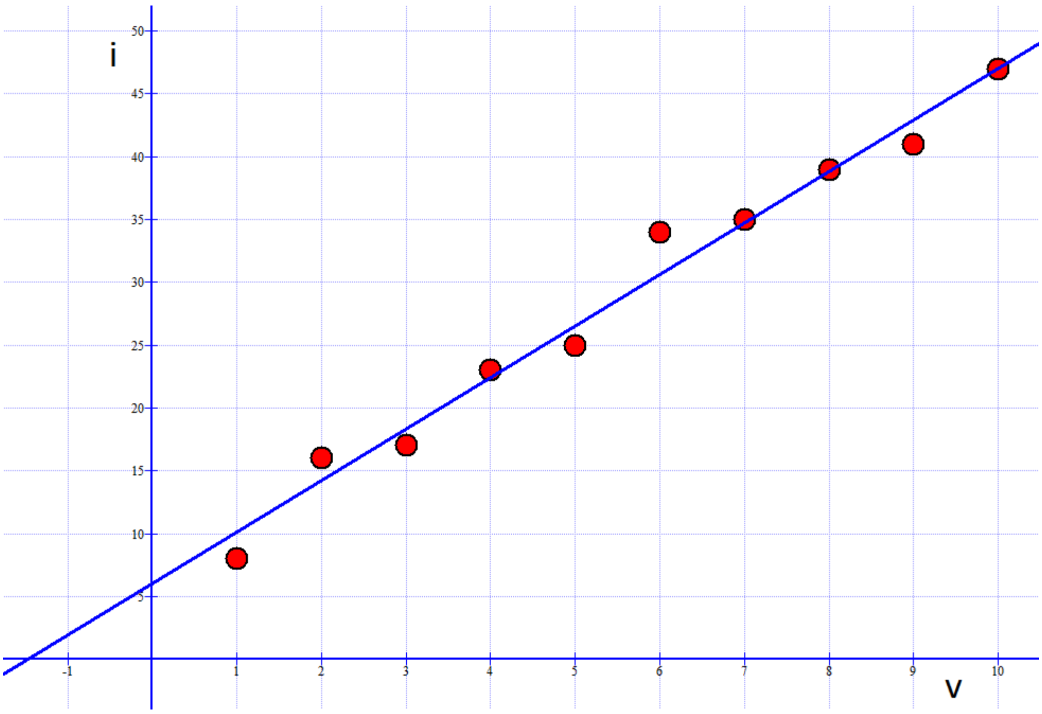
We now use the first equation and transpose to find for $a_0$…

$$285 = 10a_0 + 55a_1 \quad \therefore \quad a_0 = \frac{285 - 55a_1}{10} = \frac{285 - 55(4.1)}{10} = 6.0$$

In these calculations we have worked to 1 decimal place. Since we now know the values of $a_0$ and $a_1$ we may write the equation of the regression line…

$$i = 4.1v + 6$$

The data points and calculated regression line are shown below…




Check your answers with this handy online calculator.

(b)

Given that the current is 55A we may transpose the formula for $i$ as follows…

$$v = \frac{i-6}{4.1} = \frac{55-6}{4.1} = 12 \ volts \ \ [to\ 1\ decimal\ place]$$

## 2.3.2 Linear Correlation Coefficients

Correlation quantifies the association between two variables. We shall be looking at linear (straight line) functions and examining the amount of correlation between the data points. There are four commonly found types of correlation…

**Perfect Linear Correlation** All data points lie perfectly on a straight line and they therefore correlate with the straight line.

**Positive Linear Correlation** A straight line with a positive gradient may be drawn through the data points as a line of best fit.

**Negative Linear Correlation** A straight line with a negative gradient may be drawn through the data points as a line of best fit.

**No Correlation** No line may be drawn through the data points as a line of best fit. The data is totally random and without association.

The *product-moment formula* is commonly used to determine the linear correlation. This is expressed in terms of a coefficient, known as the linear correlation coefficient (r), and is defined by…

$$\boldsymbol{r} = \frac{\sum \boldsymbol{x_i y_i}}{\sqrt{(\sum \boldsymbol{x_i}^2)(\sum \boldsymbol{y_i}^2)}}$$

The value for $r$ may lie anywhere between $\pm 1$. Let's use this formula to work out the linear correlation coefficient for the data analysed in Worked Example 3…

**Worked Example 4**

**Calculate the linear correlation coefficient for the data below.**

| **Applied Voltage [V]** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Measured Current [A]** | **8** | **16** | **17** | **23** | **25** | **34** | **35** | **39** | **41** | **47** |

It is useful to use the table constructed to find the regression line. We just need to add one column to this table to find $\sum y_i{}^2$…

| $\boldsymbol{x_i}$ | $\boldsymbol{y_i}$ | $\boldsymbol{x_i y_i}$ | $\boldsymbol{x_i}^2$ | $\boldsymbol{y_i}^2$ |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| 1 | 8 | 8 | 1 | 64 |
| 2 | 16 | 32 | 4 | 256 |
| 3 | 17 | 51 | 9 | 289 |
| 4 | 23 | 92 | 16 | 529 |
| 5 | 25 | 125 | 25 | 625 |
| 6 | 34 | 204 | 36 | 1156 |
| 7 | 35 | 245 | 49 | 1225 |
| 8 | 39 | 312 | 64 | 1521 |
| 9 | 41 | 369 | 81 | 1681 |
| 10 | 47 | 470 | 100 | 2209 |
| $\sum x_i = 55$ | $\sum y_i = 285$ | $\sum x_i y_i = 1908$ | $\sum {x_i}^2 = 385$ | $\sum {y_i}^2 = 9555$ |

The three needed columns are highlighted in blue. We just plug in our data into the standard formula for the correlation coefficient as follows…

$$\boldsymbol{r} = \frac{\sum \boldsymbol{x_i y_i}}{\sqrt{(\sum \boldsymbol{x_i}^2)(\sum \boldsymbol{y_i}^2)}} = \frac{\mathbf{1908}}{\sqrt{(\mathbf{385})(\mathbf{9555})}} = \frac{\mathbf{1908}}{\sqrt{\mathbf{3678675}}} = \frac{\mathbf{1908}}{\mathbf{1918}} = \mathbf{0.99}$$

A figure of 0.995 for $r$ indicates very strong positive linear correlation between the data points and a straight line.

**Question**

Q2.3 **In an experiment to determine the relationship between power (P) and current (I) in a circuit the following ten measurements were taken;**

| **Current (I) [A]** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Power (P) [W]** | **19** | **32** | **38** | **53** | **58** | **71** | **79** | **94** | **97** | **112** |

(a) **Assuming a linear relationship exists between P and I, determine the equation of the regression line for P against I.**
(b) **Determine I when P is 120 Watts.**
(c) **Calculate the correlation coefficient for the given data set.**

**ANSWERS**

Q2.3

(a) $P = 10.08i + 9.86$ [W]
(b) $I = 10.93$ [A]
(c) $r = 0.99$

# 2.4 Probability

## 2.4.1 Interpretation of Probability

We can think of probability in two ways. It could represent the tendency of an event to occur or it could be a belief in the likelihood of an event occurring. We may categorise these two interpretations in two ways;

***Frequency Probability*** An event occurs in a physical system (such as a roulette wheel) in a persistent way after a very long number of trials.

***Subjective Probability*** Also known as Bayesian Probability (more on this later). These are degrees of belief based on all of the available evidence.

## 2.4.2 Probabilistic Models

A probabilistic model always contains a random variable(s). Each time your model is run it can give different results, even with the same starting conditions. The spin of a roulette wheel is a good example where random variables are at play. Consider the force placed on the thrown ball and the force applied to the wheel. These are always random, even if undertaken by a machine. The outcome of a single bet on the wheel cannot be known.

## 2.4.3 Empirical Variability

This is based upon observation or experience, rather than scientific theory. The results of such observations or experiences are very difficult to interpret.

## 2.4.4 Events and Sets

We can think of an *event* as the outcome of an experiment with an assigned probability. Consider the probability of selecting a Jack, $P(Jack)$, from a set of 52 playing cards. The probability of drawing a Jack is $4/52$ which is 0.077. A probability of 1.0 is the highest achievable (certainty) and 0 is the lowest (impossible). The Jack drawn belongs to a *Set* of Jacks. We can observe other sets, such as Hearts, Blacks, Reds, Odds, and Evens etc. To determine the probability of drawing the Jack of Spades then we have a *singular probability* of $1/52 = 0.019$.

## 2.4.5 Mutually Exclusive Events

These cannot happen together. For example, you cannot get both heads and tails with one single toss of a coin.

### 2.4.6 Independent Events

An example of two events which are independent might be to draw a Jack from a card deck, replace the card in the deck and then draw a Queen. The action of replacing the Jack in the deck has not influenced the chances of drawing the Queen.

An example of two events which are not independent might be to draw a Jack from a card deck, remove the card from the deck and then draw a King. The action of removing the Jack from the deck has influenced the chances of drawing the King.

We can therefore say that two events, A and B, are independent when event A does not affect the probability of event B occurring.

### 2.4.7 Conditional Probability

This is the probability of an event occurring after another event has taken place. For example, if a relay has been exposed to extreme heat for a long period then it is far more likely to fail than a newly manufactured relay.

Given events A and B we may express the probability of A happening given that B has already happened as…

$$P(A|B)$$

### 2.4.8 Sample Space and Probability

This is the set of all possible outcomes of an experiment. For example, the probability of a coin toss being a tail is 0.5 and the probability of a coin toss being a head is also 0.5. Since there are no other possible outcomes then the sample space is 0.5 + 0.5 = 1.

### 2.4.9 Addition Law

Given events A and B are mutually exclusive, the probability that A or B will occur is the sum of the probability of each event. For example, the probability of throwing a 2 or a 6 on a die is $1/6 + 1/6 = 1/3$.

### 2.4.10 Product Law

This is the probability of two *independent* events occurring. For example, tossing two dies simultaneously will give the probability of obtaining a pair of 6's as $1/6 \times 1/6 = 1/36$.

### 2.4.11 Bayes' Theorem

This relates current probability to prior probability and relies on conditional probability. Bayes' Theorem may be expressed mathematically as…

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

One application of Bayes' Theorem is in email spam filtering. Let event A be the probability that an email is spam. Let event B be a test for certain spam-like words. We may then write…

$$P(spam|words) = \frac{P(words|spam)\,P(spam)}{P(words)}$$

**Worked Example 5**

**A certain circuit board is installed on petrol station forecourts across the world, and is subject to a wide variation of temperatures. Let P(A) = 0.002 be the probability of an overheating component on the board and P(B) = 0.001 be the probability of a failed component on the board. If the probability of a failed component given an overheated component is $\boldsymbol{P(B|A) = 0.15}$ use Bayes' Theorem to calculate the probability that an overheated component has caused a circuit board failure.**

Let's gather the data…

$$P(A) = 0.002; \quad P(B) = 0.001; \quad P(B|A) = 0.15$$

Therefore we can say…

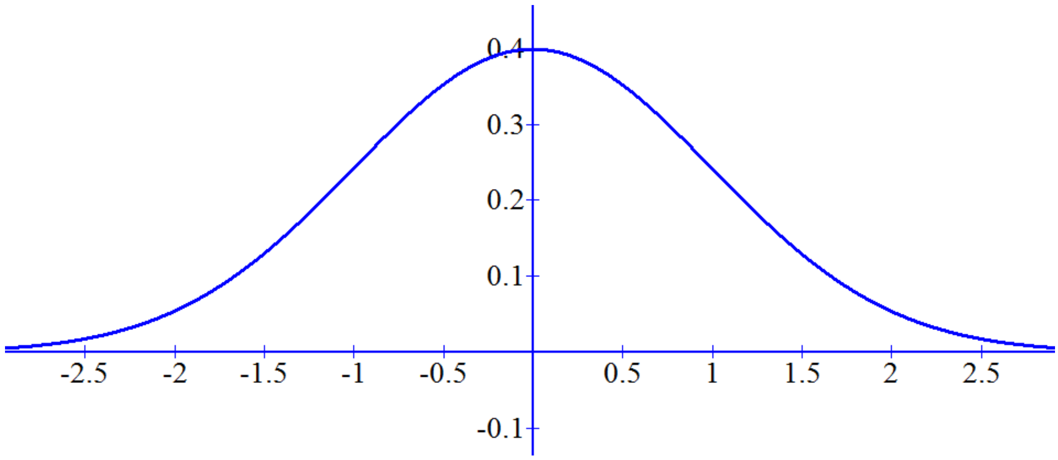
$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} = \frac{(0.15)(0.002)}{0.001} = 0.3$$

It isn't just extremes of temperature which cause components to fail on circuit boards. Other causal factors are humidity, vibration, ageing, poor manufacture, solar activity, human misuse and various other reasons.

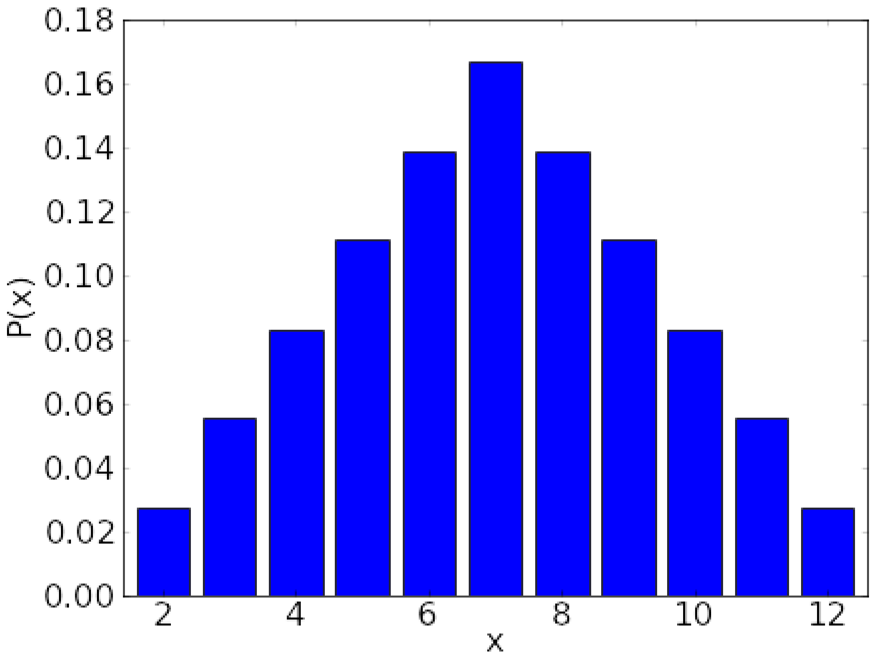
# 2.5 Probability Distributions

## 2.5.1 Discrete and Continuous Distributions

If a variable can assume any value between two limits then it is termed a *continuous variable*. Otherwise, it is known as a *discrete variable*. A *probability distribution* may therefore be classified as either continuous or discrete. A typical continuous probability distribution might look like…




A typical discrete probability distribution might look like…

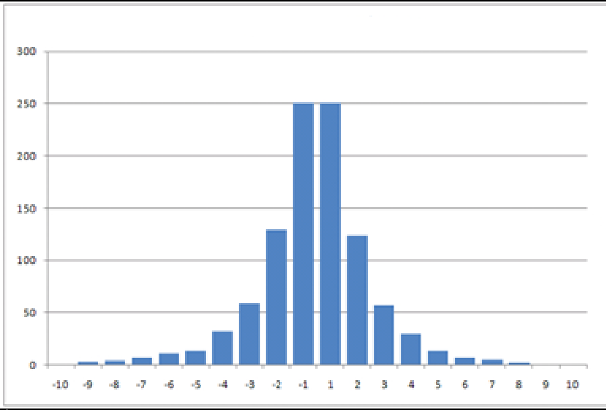
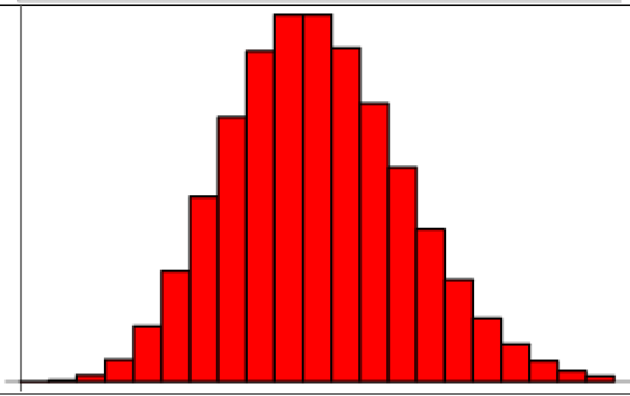
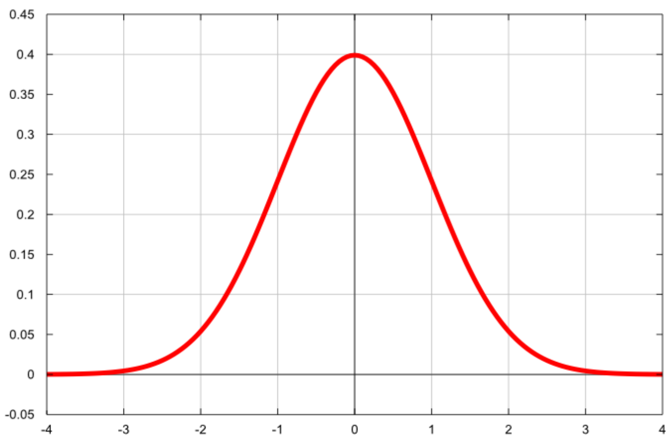





edexcel Approved Centre

## 2.5.2 Introduction to the Binomial, Poisson and Normal Distributions

The table below gives typical curves for each of these distributions…

| Distribution | Curve |
|---|---|
| **Binomial**<br>*Discrete and symmetrical* |  |
| **Poisson**<br>*Discrete and skewed* |  |
| **Normal**<br>*Continuous and symmetrical* |  |

## 2.5.3 Use of the Normal Distribution to estimate Confidence Intervals

The *Normal Distribution* (sometimes known as a *Gaussian* distribution) is a very commonly encountered continuous probability distribution curve which symmetrically falls to zero either side of a central value, normally the mean. The term 'Normal' implies 'typical'.

A Normal distribution can come about by analysing many truly random situations, such as the height of people, length of manufactured components, measurement errors, exam results and blood pressure. It is important that sampling is undertaken randomly and without bias. For example, if you wished to find the distribution of heights for the male population then your data would not reflect a 'normal' distribution if you took your samples solely from the local basketball club!

The equation for a Normal distribution, known as a *probability density function* is given by…

$$\boldsymbol{P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\bar{x})^2/2\sigma^2}}$$

where;

$$P(x) \text{ is the probability distribution of the variable } x$$

$$\sigma \text{ is the standard deviation of the data}$$

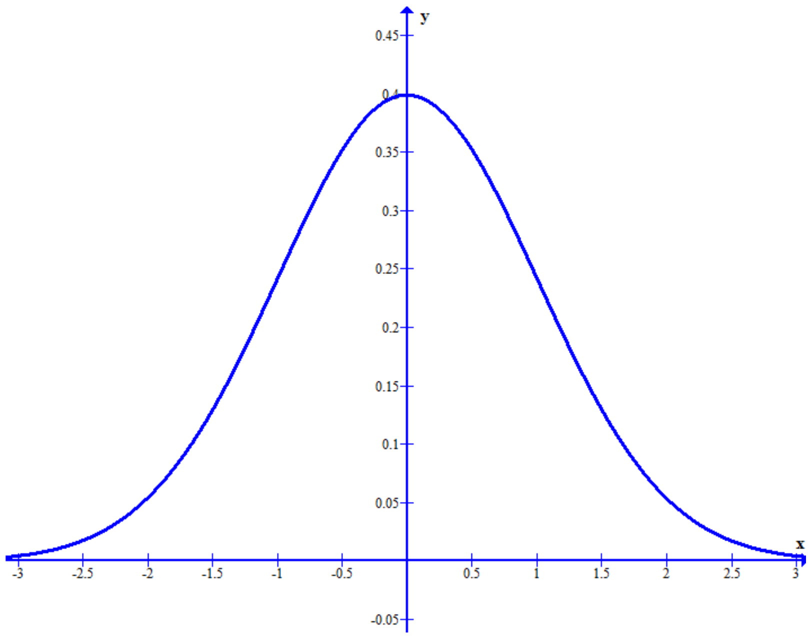
$$\bar{x} \text{ is the mean of the data}$$

$$\sigma^2 \text{ is the variance of the data}$$

For a first look at a Normal distribution we shall let the mean value $\bar{x} = 0$ and the standard deviation $\sigma = 1$. This will then give us the *standard normal distribution* of…

$$\boldsymbol{P(x) = \frac{1}{\sqrt{2\pi}} e^{-(x)^2/2}}$$

We can use the Graph application to plot this standard normal distribution…






edexcel Approved Centre

To produce this curve in Graph you need to enter (or copy/paste) the function in the following format…

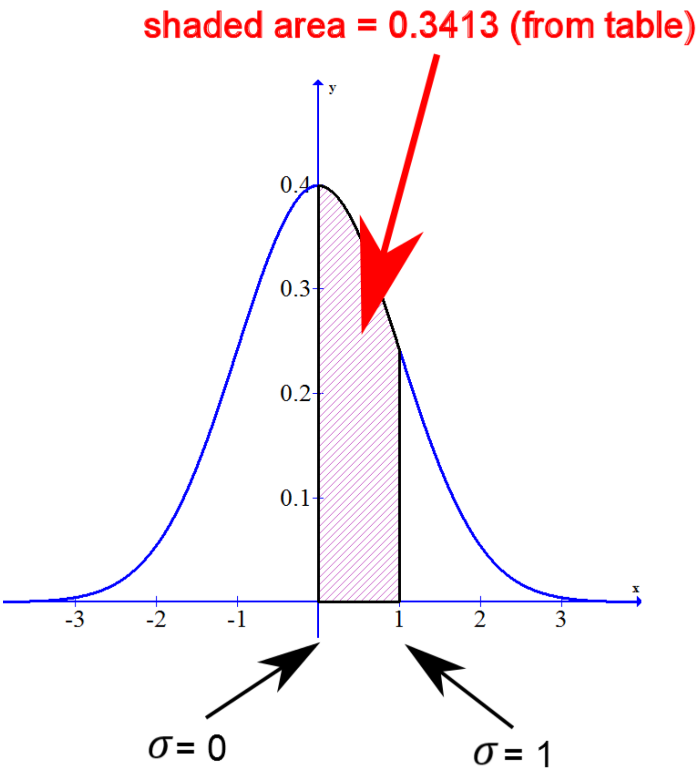
**(1/(sqrt(2*pi)))*(e^(-(x^2)/2))**

This produces the typical 'Bell Curve' we associate with Normal distributions. The central value is usually taken as the mean but sometimes the median and mode are used for the measure of central tendency.

Of course, not all mean values are zero, so we usually shift the central value to our mean value rather than zero. We also scale the horizontal axis into standard deviations and furthermore scale the curve so that the total area beneath it is equal to 1.

What we see on our standard normal distribution curve is that the horizontal axis is divided into standard deviations. The curve approaches zero around $\mp 3$ standard deviations. If we were to calculate the area under this curve we would find that it was 1.

To calculate the area under the standard normal distribution curve is much more difficult than you might imagine. In fact, numerical methods tend to be used to find such a definite integral. Let's look at the definite integral between 0 and 1 standard deviations, shown shaded below…




The Graph software produces this shading and calculation of the shaded area quite readily (click on the icon named '*Calculate the definite integral over a given interval*'). Here you can see that probability is represented by area. So, the probability of an event somewhere between 0 and 1 standard deviations is 0.3413. Because of symmetry on the graph we can therefore infer that there is a probability of $2 \times 0.3413 = 0.6826$ that an event will lie within $\mp 1$ standard deviations. The definite integral needed to determine the area under the curve is…

$$P(x) = \int_{0}^{1} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\bar{x})^2/2\sigma^2} dx$$

***You don't want to be working that integral out for all of the $x$ pairs which interest you***! Fortunately this task has been done for you by others and the full set of results are presented in the table below.

The probability for the interval between 0 and 1 standard deviation is circled in the table. For $\mp 2$ standard deviations we see from the table that the probability is $2 \times 0.4772 = 0.9544$. For $\mp 3$ standard deviations the probability is given as $2 \times 0.4987 = 0.9974$. The table is useful because it saves you having to work out all of the integrals.

| $z = \frac{x-\bar{x}}{\sigma}$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.0000 | 0.0040 | 0.0080 | 0.0120 | 0.0159 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0678 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.0832 | 0.0871 | 0.0910 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1388 | 0.1406 | 0.1443 | 0.1480 | 0.1517 |
| 0.4 | 0.1554 | 0.1591 | 0.1628 | 0.1664 | 0.1700 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| 0.5 | 0.1915 | 0.1950 | 0.1985 | 0.2019 | 0.2054 | 0.2086 | 0.2123 | 0.2157 | 0.2190 | 0.2224 |
| 0.6 | 0.2257 | 0.2291 | 0.2324 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2517 | 0.2549 |
| 0.7 | 0.2580 | 0.2611 | 0.2642 | 0.2673 | 0.2704 | 0.2734 | 0.2760 | 0.2794 | 0.2823 | 0.2852 |
| 0.8 | 0.2881 | 0.2910 | 0.2939 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| 0.9 | 0.3159 | 0.3186 | 0.3212 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.3340 | 0.3365 | 0.3389 |
| 1.0 | 0.3413 | 0.3438 | 0.3451 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| 1.1 | 0.3643 | 0.3665 | 0.3686 | 0.3708 | 0.3729 | 0.3749 | 0.3770 | 0.3790 | 0.3810 | 0.3830 |
| 1.2 | 0.3849 | 0.3869 | 0.3888 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.3980 | 0.3997 | 0.4015 |
| 1.3 | 0.4032 | 0.4049 | 0.4066 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| 1.4 | 0.4192 | 0.4207 | 0.4222 | 0.4236 | 0.4251 | 0.4265 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| 1.5 | 0.4332 | 0.4345 | 0.4357 | 0.4370 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4430 | 0.4441 |
| 1.6 | 0.4452 | 0.4463 | 0.4474 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |
| 1.7 | 0.4554 | 0.4564 | 0.4573 | 0.4582 | 0.4591 | 0.4599 | 0.4608 | 0.4616 | 0.4625 | 0.4633 |
| 1.8 | 0.4641 | 0.4649 | 0.4656 | 0.4664 | 0.4671 | 0.4678 | 0.4686 | 0.4693 | 0.4699 | 0.4706 |
| 1.9 | 0.4713 | 0.4719 | 0.4726 | 0.4732 | 0.4738 | 0.4744 | 0.4750 | 0.4756 | 0.4762 | 0.4767 |
| 2.0 | 0.4772 | 0.4778 | 0.4783 | 0.4785 | 0.4793 | 0.4798 | 0.4803 | 0.4808 | 0.4812 | 0.4817 |
| 2.1 | 0.4821 | 0.4826 | 0.4830 | 0.4834 | 0.4838 | 0.4842 | 0.4846 | 0.4850 | 0.4854 | 0.4857 |
| 2.2 | 0.4861 | 0.4864 | 0.4868 | 0.4871 | 0.4875 | 0.4878 | 0.4881 | 0.4884 | 0.4887 | 0.4890 |
| 2.3 | 0.4893 | 0.4896 | 0.4898 | 0.4901 | 0.4904 | 0.4906 | 0.4909 | 0.4911 | 0.4913 | 0.4916 |
| 2.4 | 0.4918 | 0.4920 | 0.4922 | 0.4925 | 0.4927 | 0.4929 | 0.4931 | 0.4932 | 0.4934 | 0.4936 |
| 2.5 | 0.4938 | 0.4940 | 0.4941 | 0.4943 | 0.4945 | 0.4946 | 0.4948 | 0.4949 | 0.4951 | 0.4952 |
| 2.6 | 0.4953 | 0.4955 | 0.4956 | 0.4957 | 0.4959 | 0.4960 | 0.4961 | 0.4962 | 0.4963 | 0.4964 |
| 2.7 | 0.4965 | 0.4966 | 0.4967 | 0.4968 | 0.4969 | 0.4970 | 0.4971 | 0.4972 | 0.4973 | 0.4974 |
| 2.8 | 0.4974 | 0.4975 | 0.4976 | 0.4977 | 0.4977 | 0.4978 | 0.4979 | 0.4980 | 0.4980 | 0.4981 |
| 2.9 | 0.4981 | 0.4982 | 0.4982 | 0.4983 | 0.4984 | 0.4984 | 0.4985 | 0.4985 | 0.4986 | 0.4986 |
| 3.0 | 0.4987 | 0.4987 | 0.4987 | 0.4988 | 0.4988 | 0.4989 | 0.4989 | 0.4989 | 0.4990 | 0.4990 |
| 3.1 | 0.4990 | 0.4991 | 0.4991 | 0.4991 | 0.4992 | 0.4992 | 0.4992 | 0.4992 | 0.4993 | 0.4993 |
| 3.2 | 0.4993 | 0.4993 | 0.4994 | 0.4994 | 0.4994 | 0.4994 | 0.4994 | 0.4995 | 0.4995 | 0.4995 |
| 3.3 | 0.4995 | 0.4995 | 0.4995 | 0.4996 | 0.4996 | 0.4996 | 0.4996 | 0.4996 | 0.4996 | 0.4997 |
| 3.4 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4998 |
| 3.5 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 |
| 3.6 | 0.4998 | 0.4998 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 |
| 3.7 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 |
| 3.8 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 |
| 3.9 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 0.5000 | 05000 | 0.5000 | 0.5000 | 0.5000 |

Let's have a look at a useful worked example connected with the Normal distribution…

**Worked Example 6**

**A production line manufactures $1\Omega$ resistors. Four samples of these resistors were taken each day over five separate days. The resistance values were measured and recorded as follows;**

| Sample 1 [Ω] | Sample 2 [Ω] | Sample 3 [Ω] | Sample 4 [Ω] | Sample 5 [Ω] |
|---|---|---|---|---|
| 0.99 | 1.00 | 1.05 | 1.01 | 0.98 |
| 1.03 | 0.96 | 1.03 | 0.99 | 1.05 |
| 0.96 | 0.96 | 1.04 | 1.02 | 0.96 |
| 0.95 | 0.99 | 1.03 | 1.00 | 1.02 |

a) **Determine the mean of the means.**
b) **Determine the standard deviation.**
c) **Determine both the 95% and 99.8% confidence limits.**

(a)

The mean value of sample 1 is $(0.99 + 1.03 + 0.96 + 0.95)/4 = 0.98$

The mean value of sample 2 is $(1.00 + 0.96 + 0.96 + 0.99)/4 = 0.98$

The mean value of sample 3 is $(1.05 + 1.03 + 1.04 + 1.03)/4 = 1.04$

The mean value of sample 4 is $(1.01 + 0.99 + 1.02 + 1.00)/4 = 1.01$

The mean value of sample 5 is $(0.98 + 1.05 + 0.96 + 1.02)/4 = 1.00$

We can take the mean of the means as $(0.98 + 0.98 + 1.04 + 1.01 + 1.00)/5 = 1.00$

*[Note – all working here has been to 2 decimal places]*

(b)

The standard deviation can be derived from the five mean values and the mean of means…

$$\boldsymbol{standard\ deviation, \sigma} = \sqrt{\frac{\sum(\boldsymbol{x_i} - \boldsymbol{\bar{x}})^2}{\boldsymbol{n}}} = \sqrt{\frac{(\boldsymbol{-0.02})^2 + (\boldsymbol{-0.02})^2 + (\boldsymbol{0.04})^2 + (\boldsymbol{0.01})^2 + \boldsymbol{0}^2}{\boldsymbol{5}}}$$

$$= \sqrt{\frac{0.0004 + 0.0004 + 0.0016 + 0.0001}{5}} = \sqrt{\frac{0.0025}{5}} = 0.022$$

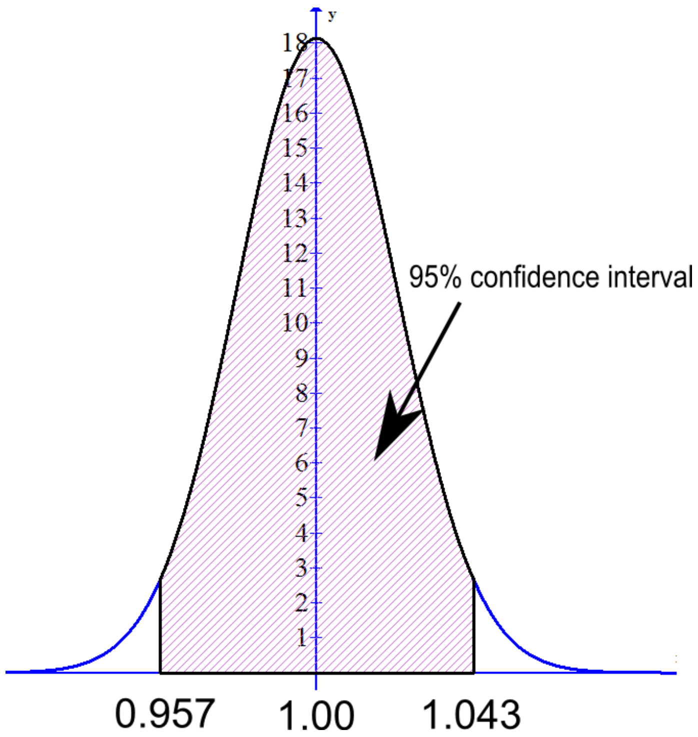
(c)

A 95% confidence interval has a probability of 0.95, of course. What we mean here is that we have 95% confidence that events/data will lie within the interval in question. What we need to do is to determine what this interval actually is.

Considering the symmetry of the Normal curve then our value for the right half area of the confidence region is 0.95/2 = 0.475. We then consult the table to determine which z value corresponds to a confidence interval of 0.475. We see that this z value is 1.96. We can then write…

$$z = \frac{x - \bar{x}}{\sigma} \quad \therefore \quad x = z\sigma + \bar{x} = 1.96(0.022) + 1.00 = 1.043$$

So we can move 0.043Ω either side of the mean to mark the 95% confidence interval.





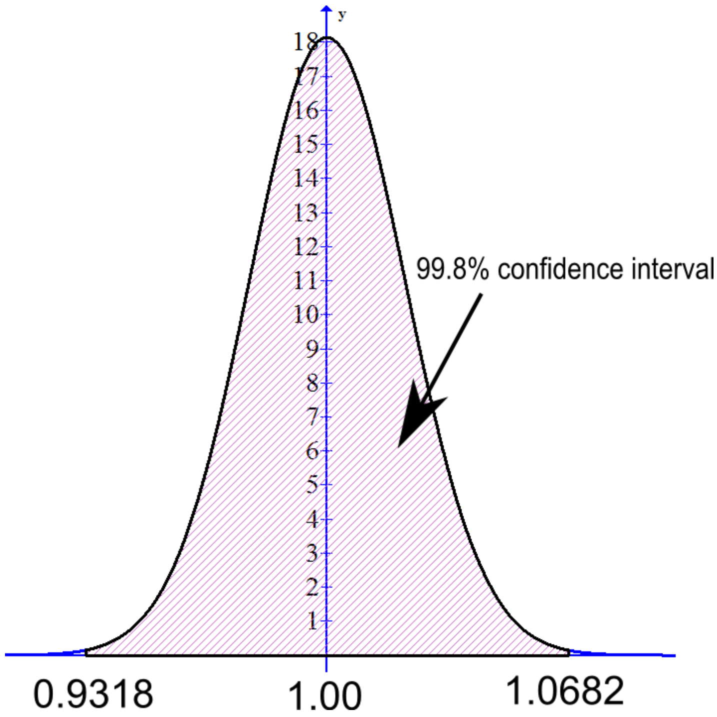

edexcel Approved Centre

We proceed in a similar manner to determine the 99.8% confidence interval.

Considering the symmetry of the Normal curve then our value for the right half area of the confidence region is 0.998/2 = 0.499. We then consult the table to determine which z value corresponds to a confidence interval of 0.499. We see that this z value is 3.1. We can then write…

$$z = \frac{x - \bar{x}}{\sigma} \quad \therefore \quad x = z\sigma + \bar{x} = 3.1(0.022) + 1.00 = 1.0682$$

So we can move 0.0682Ω either side of the mean to mark the 99.8% confidence interval. This is shown below.




**Question**

Q2.4 Using the data given in Worked Example 6 calculate the 68.26% confidence interval.

**ANSWER**

| |
|---|
| Q2.4 0.978 Ohms up to 1.022 Ohms |

**Video**

More insight into the Normal distribution with this video.

## 2.5.4 Probability Calculations for Binomially Distributed Variables

If we call p the probability that an event will happen, and q the probability that it won't happen, we are dealing with a BINOMIAL DISTRIBUTION (Bi meaning two – i.e. two probabilities only).

A good example of using a binomial distribution is the tossing of a coin. There are only two outcomes, a head on top or a tail on top. So, p could represent the probability of getting a head and q the probability of not getting a head (i.e. getting a tail). The probability of getting either a head OR tail is certain, of course, so this is allocated a probability of 1 (i.e. certainty is 1 – no chance at all is represented by 0). When unbiased coin tosses are undertaken, we can say that p = 0.5 and q = 0.5, so p and q add to 1.

When there is more than one toss of a coin, i.e. 'n' tosses, the binomial expansion is as follows…

$$(q+p)^n = q^n + nq^{n-1}p + \frac{n(n-1)}{2!}q^{n-2}p^2 + \frac{n(n-1)(n-2)}{3!}q^{n-3}p^3 + \cdots etc.$$

**Worked Example 7**

A machine produces metal bolts. In a tray of these bolts, 92% are within the allowable diameter tolerance value. The remainder exceed the tolerance.

Seven bolts are drawn at random from the tray. Determine the probabilities that;

a) Two of the seven bolts exceed the diameter.

b) More than two of the seven bolts exceed the diameter.

**ANSWER**

Let's allocate some letters to probabilities…

Given: p = probability a bolt is defective = 0.08 (i.e. 1 – 0.92 = 0.08)

Given: q = probability a bolt is within tolerance = 0.92

$$(q+p)^n = q^n + nq^{n-1}p + \boldsymbol{\frac{n(n-1)}{2!}q^{n-2}p^2} + \frac{n(n-1)(n-2)}{3!}q^{n-3}p^3 + \cdots etc.$$

(a)

The part of the binomial expansion in red, above, will give us the probability that two of the seven bolts, p(2), are defective…

$$p(2) = \frac{n(n-1)}{2!}q^{n-2}p^2$$

Remember that 2! means '2 factorial', which means 1 x 2 = 2. If we would have needed, say 4!, that would be equal to 1 x 2 x 3 x 4 which is 24. Let's put in the figures then…

$$p(2) = \frac{7(7-1)}{2}0.92^{7-2}0.08^2$$

$$\therefore \quad p(2) = (21)(0.659)(0.0064)$$

$$\therefore \quad \boldsymbol{p(2) = 0.0886}$$

(b)

Let p(>2) be the probability that we select more than 2 defective bolts.

$$\therefore \quad p(>2) = 1 - \{p(0) + p(1) + p(2)\}$$

$$\therefore \quad p(>2) = 1 - \{0.92^7 + (7)(0.92^{7-1})(0.08) + 0.0886\}$$

$$\therefore \quad \boldsymbol{p(>2) = 0.014}$$

i.e. there is a 1.4% chance of selecting more than two faulty bolts.

## 2.5.5 Probability Calculations for Normally Distributed Variables

**Worked Example 8**

The mean resistance of 500 carbon resistors is 82 Ω and the standard deviation is 6 Ω. If the resistances are normally distributed, determine the number of resistors likely to have values between 81 Ω and 83 Ω.

**ANSWER**



edexcel Approved Centre

$$\bar{x} = 82\,\Omega \quad ; \quad \sigma = 6\,\Omega$$

We now find the z-values for the given limits…

$$z_{81} = \frac{81 - 82}{6} = -0.17 \;\; standard\ deviations$$

From the z-table this corresponds to an area of 0.0678 of the total area under the function (which is 1).

$$z_{83} = \frac{83 - 82}{6} = +0.17 \;\; standard\ deviations$$

From the z-table this corresponds to an area of 0.0678 of the total area under the function (which is 1).

The total area between the limits of 81 and 83 is therefore twice 0.0678 …

Since we have 500 resistors in the sample then the number of resistors we expect to have resistance in the range 81 to 83 Ohms is…

$$2 \times 0.0678 \;\times 500 = 67.8$$

We round to the nearest integer and state that **68** resistors are expected to be within the given range.

## 2.5.6 Statistical Hypothesis Testing

We can think of a hypothesis as an educated guess about something. Some examples might be…

- The effects of a new drug
- The efficiency of a new antenna
- The benefits of a new lubricant

The most important thing about a hypothesis is that **it must be must be testable**, either by way of measurements, or simply observation.

A hypothesis should start with a statement, and this statement should be in the form of 'if' and 'then' wording. Some examples…

- **If** this new drug is given to cows **then** they will produce more milk
- **If** this new antenna design is adopted **then** urban areas will receive a stronger signal
- **If** this new lubricant is used **then** engine friction will be reduced

We use a NULL HYPOTHESIS to state what is usually accepted. Some examples of a null hypothesis might be…

- Eating too much saturated fat increases the risk of a heart attack
- Oxygen is required to support life anywhere in the universe
- All DNA has a double helix shape

We allocate $H_0$ to represent the null hypothesis. We also allocate $H_1$ to represent the **ALTERNATIVE HYPOTHESIS**.

Let's see how hypothesis testing works with an example…

**Worked Example 9**

A design engineer is testing the effects of an experimental engine lubricant for electric cars. He adds a standard sample amount of lubricant to 100 cars of the same model, and records the number of miles per full charge (mpfc) for each car after being driven around a test track at a constant speed, until the battery fully discharges. He knows that such testing undertaken without the new lubricant produces a mean mpfc figure of 250. Collecting results with the new lubricant, he notices that the mean mpfc figure is 254.2 with a sample standard deviation of 12 mpfc.

By interpreting the results of the testing, show whether you agree, or not, that the new lubricant has influenced the number of miles per full charge for the cars.

**ANSWER**

There are two hypotheses…

$H_0$ : new lubricant has no effect on mpfc, therefore the mean mpfc, $\bar{x} = 250$

$H_1$ : new lubricant does have an effect on mpfc, therefore the mean mpfc, $\bar{x} \neq 250$

Let's assume that the null hypothesis $H_0$ is true:

The sample standard deviation ($\sigma_x$) is given as 12 for 100 sample cars (n = 100), but we must convert this to the population standard deviation, usually done as follows…

$$\sigma = \frac{\sigma_x}{\sqrt{n}} = \frac{12}{\sqrt{100}} = \frac{12}{10} = 1.2$$

Since we know that…

$$z = \frac{x - \bar{x}}{\sigma}$$

$$\therefore \quad z = \frac{254.2 - 250}{1.2} = \frac{4.2}{1.2} = 3.5$$

So, z is 3.5 standard deviations away from the normalised centre. This z-value corresponds to an area under our Normal curve of (using the table)…



edexcel Approved Centre

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 3.1 | 0.4990 | 0.4991 | 0.4991 | 0.4991 | 0.4992 | 0.4992 | 0.4992 | 0.4992 | 0.4993 | 0.4993 |
| 3.2 | 0.4993 | 0.4993 | 0.4994 | 0.4994 | 0.4994 | 0.4994 | 0.4994 | 0.4995 | 0.4995 | 0.4995 |
| 3.3 | 0.4995 | 0.4995 | 0.4995 | 0.4996 | 0.4996 | 0.4996 | 0.4996 | 0.4996 | 0.4996 | 0.4997 |
| 3.4 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4997 | 0.4998 |
| 3.5 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 | 0.4998 |
| 3.6 | 0.4998 | 0.4998 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 |
| 3.7 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 |
| 3.8 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 | 0.4999 |

The ringed figure of 0.4998 only represents the area to the right of centre for the Normal curve, so we must double this to find the total area under the curve…

$$Total\ area\ under\ curve\ for\ this\ z\ value = 2 \times 0.4998 = 0.9996$$

Since the total area under the standardised Normal curve is 1, the area occupied by our z-value is…

$$\frac{0.9996}{1} \times 100\% = 99.96\%$$

We interpret this figure as meaning that there is a (100 – 99.96)% = 0.04% chance of the null hypothesis being true. Therefore, we suggest that the alternative hypothesis is likely to be true and that the new lubricant does have an influence on the number of miles per full charge for the electric cars.