Attributions

- WACE Mathematics Methods ATAR Course Data Booklet © School Curriculum and Standards Authority WA

elucidate

Applecross, 6153, Perth, Western Australia
Tel: 0414 107 122
Email: christian.bien@elucidate.org.au

Authors

Jack Anderson – Author and Chief Innovation Officer
University of Western Australia (2019 – 2023): Bachelor of Philosophy (Honours) (Engineering Science & Business Law)

Hannah Knight – Author and Chief Product Officer
Curtin University (2020 – 2024): Bachelor of Science (Chemistry) & Bachelor of Arts (Japanese)

Jemma Schofield – Author and Head of Textbooks
Curtin University (2020 – 2023): Bachelor of Science (Data Science) & Bachelor of Commerce (Economics)

Patrick Catambay – Lead Author
Curtin University (2018 – 2023): Bachelor of Engineering (Mechanical) & Bachelor of Commerce (Finance)

Kyi Chan – Lead Author
University of Western Australia (2020 – 2023): Bachelor of Commerce (Business Law, Finance & Business Analytics)

Sarah Jade Sauzier – Lead Author
University of Western Australia (2021 – 2023): Bachelor of Science (Engineering Science)

Matthew Harms – Quality Assurance
University of Western Australia (2021 – 2023): Doctor of Medicine

Daniel Holding – Quality Assurance
University of Western Australia (2021 – 2025): Bachelor of Philosophy (Honours) & Master of Professional Engineering (Software)

Damon Clift – Co-Author
University of Western Australia (2021 – 2023): Bachelor of Philosophy (Honours) (Frontier Physics & Mathematics and Statistics)

Aaron Gibson-John – Co-Author
University of Western Australia (2022 – 2025): Bachelor of Philosophy (Honours) (Mathematics and Statistics & Finance)

Zoe Griffiths – Co-Author
Curtin University (2020 – 2023): Bachelor of Science (Actuarial Science)

Apphia Jin – Co-Author
Australian National University (2015 – 2017): Bachelor of Science (Mathematics and Statistics)

Celine Ng – Co-Author
University of Melbourne (2021- 2023): Bachelor of Biomedicine (Physiology)

Leonardo Rachetti – Co-Author
University of Western Australia (2020 – 2025): Bachelor of Science (Engineering Science)

Dylan Royce – Co-Author
University of Western Australia (2021 – 2022): Master of Professional Engineering (Electronic and Electrical Engineering)

Stephanie Schwab – Co-Author
Curtin University (2020 – 2023): Bachelor of Engineering (Electrical and Electronic Engineering)

Henry Freegard – Co-Author
University of Western Australia (2022 – 2025): Bachelor of Philosophy (Honours) (Mathematics and Statistics, Advanced French Studies)

Conan Po Dewitt – Contributing Author
University of Western Australia (2020 – 2023): Bachelor of Science (Software Engineering)

Evangelos Hajigabriel – Contributing Author
University of Western Australia (2023 – 2026): Bachelor of Science (Frontier Physics)

Zijun Jiang– Contributing Author
Curtin University (2022 – 2027): Bachelor of Engineering (Civil and Construction Engineering) (Honours) & Bachelor of Science (Mining)

Ramudhi Kottage – Contributing Author
Curtin University (2020 – 2025): Bachelor of Civil and Construction Engineering (Honours) and Bachelor of Commerce (Finance)

Amelie Ong – Contributing Author
University of Western Australia (2023 – 2029): Bachelor of Engineering (Honours) (Chemical Engineering) & Bachelor of Philosophy (Honours)

Maddie Ong – Contributing Author
University of Western Australia (2023 – 2029): Bachelor of Engineering (Honours) (Chemical Engineering) & Bachelor of Philosophy (Honours)

Charles Richmond – Contributing Author
Curtin University (2021 – 2025): Bachelor of Science (Industrial and Applied Mathematics)

Tran Thai – Contributing Author
University of Melbourne (2021 – 2023): Bachelor of Biomedicine

Aurelee Tjoeng – Contributing Author
University of Western Australia (2020 – 2022): Bachelor of Biomedical Science (Neuroscience)

Michael Wisdom – Contributing Author
University of Western Australia (2020 – 2023): Bachelor of Engineering (Electrical Engineering)

Masonic Foundation – Maths Methods Textbook Sponsor

The publication of this book is made possible through the support of Masonic Foundation. For over a century Western Australian Freemasons, through Masonic Foundation as it is known today, have been supporters of access to quality education for all.

Masonic Foundation education programmes provide opportunities for students of all levels to reach their full potential; developing primary students' public speaking abilities through "Speak Up Awards", your education through textbooks like these, and financial support for university students through Masonic Foundation Education Grants. Masonic Foundation aims to build a stronger and better society by helping others like you.

CONTENTS

Chapter 7

Logarithmic Functions

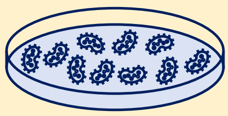
Chapter 7 – Logarithmic Functions

Introduction

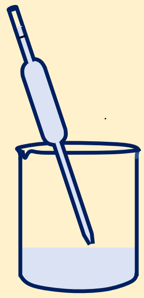
Now that we've completed Unit 3, it's time for us to start our journey into Unit 4. To start, we will be exploring our final topic of calculus which is **logarithms**. A **logarithmic function** is best understood as the inverse of the **exponential function**.

This might not make sense right now, but once you've mastered the concept of **logarithms**, you will be able to use them to determine the **magnitude of an earthquake**, the **population growth of bacteria** or the **pH** of **various chemical solutions**. The applications of logarithms are truly endless!

Measuring the scale of an Earthquake (Richter scale)



Population Growth of Bacteria in a Petri Dish



pH of a Chemical Solution (pH scale)

This first chapter will cover **five main concepts**:

1. **Fundamentals of Logarithmic Functions**
2. **The Natural Logarithm and applications of Logarithmic Functions**
3. **Differentiation and Integration of Natural Logarithmic Functions**

Let's Begin!

7.1 Introduction to Logarithmic Functions

Logarithms are the inverse operation of an **exponential**. Just as addition and subtraction are inverses of each other, **logarithms "undo" exponentiation**. They are written in the **form**: $log_a y = x$

Recall that an exponent is the number another function is raised to. For example, the **exponent** in b^y is y. Suppose we have a number, let's call it 'x', and we want to **find the exponent** to which a **base 'b' must be raised** to **obtain '** x **'**. We can express this using **logarithm notation** as:

$$x = log_b y$$

Here, 'b' is the **base** of the logarithm, 'x' is the **number** we want to **find the exponent for**, and 'y' is the exponent or power to which **'b' must be raised** to **obtain** 'x'. As we mentioned in the definition, a **logarithm** is essentially the inverse of an **exponential function,** and this is a **special relationship** that we can use for **various applications**.

The Relationship between Logarithms and Exponentials

Logarithms and **exponential functions** are **closely related**. In fact, they are like **two sides** of the **same coin**. An **exponential function** is one where the **independent variable** appears in the **exponent**, while a **logarithmic function** is one where the **independent variable** appears **inside the logarithm**.

To understand this relationship, consider the **exponential equation**:

$$y = b^x$$

Here, 'b' is the **base**, 'x' is the **exponent**, and 'y' is the **result**. To find the **inverse relationship**, we can **rewrite the equation** using **logarithm notation**:

$$x = log_b\, y$$

Notice how the **exponent** 'x' becomes the logarithm and the **result** 'y' becomes the **number we want to find** the exponent for. The **logarithm helps us "zoom-out"** from the **exponential growth** and understand the relationship from a **different perspective**. This relationship can be **summarised** as:

$$\boxed{x = log_b\, y \iff y = b^x}$$

To understand how this relationship works, let's explore some **examples** with **actual numbers** rather than just variables. For instance, if we want to find the **value** of x when b is 10 and y is 1000 in the exponential equation: $y = b^x$, we can **re-write** it in the **logarithm form** to determine the **value of** x:

$$1000 = 10^x$$

① $x = \log_{10} 1000$

② $x = 3$

① **Re-write in logarithmic form**

② **Evaluate the logarithm** to get the **value of** x

As seen above, to find the **value of** x, we need to find the **exponent** to which **10 must be raised to obtain 1000**. This is equivalent to evaluating the logarithm, which gives us: $x = 3$.

As a **second example**, let's suppose we want to find the **value of** y when b is **equal to** 2 and x is **equal to** 3 in the logarithmic form: $x = \log_b y$. To determine the **value of** y, we can **rewrite** it in the **exponential form** and then **evaluate** the exponential as follows:

$$3 = \log_2 y$$

① $y = 2^3$

② $y = 8$

① **Re-write in exponential form**

② **Evaluate the exponential** to get the **value of** y

As a third and final example, let's suppose we had the equation: $81 = 3^x$. Like the first example, we can **re-write the equation** in its **logarithmic form** to **evaluate for** x as follows:

$$81 = 3^x$$

① $x = log_3\, 81$

② $x = 3$

① **Re-write in logarithmic form**

② **Evaluate the logarithm** to get the **value of** x

Log Laws

Logarithmic laws are a **set of rules** that help us **manipulate** and **simplify** logarithmic expressions. Understanding these properties is **essential for solving** and **simplifying logarithmic expressions**. Let's start by exploring **three fundamental logarithmic laws**.

The **first key rule** is for logarithms of the form $log_b(mn)$. Our **first rule** states that when **two numbers** are **multiplied together inside a logarithm**, it is **equivalent** to **adding their logarithms**. Mathematically, this can be **expressed** as follows:

$$log_b(mn) = log_b m + log_b n$$

This law lets us **break down** a **product of two numbers inside a logarithm** into the **sum of their individual logarithms**. It is useful when dealing with **complex expressions** or **evaluating logarithms of large numbers**.

For example, $log_2(4 \times 8)$, can be **split up** into **two logarithms** and then **evaluated separately**:

① $log_2(4 \times 8) = log_2(4) + log_2(8)$

② $log_2(4 \times 8) = 2 + 3$

$\therefore log_2(4 \times 8) = 5$

① Apply our log law: $log_b(mn) = log_b m + log_b n$

② Evaluate each logarithm separately

The **second key rule** is the for logarithms of the form: $log_b\left(\frac{m}{n}\right)$. Our **second rule** states that when **two numbers** are **divided inside a logarithm**, it is **equivalent to subtracting** their logarithms. Mathematically, it can be **expressed** as follows:

$$log_b\left(\frac{m}{n}\right) = log_b m - log_b n$$

This property allows us to **simplify** the **division of numbers** inside a **logarithm** by **subtracting** their individual logarithms. For instance, $log_{10}(\frac{1000}{10})$, can be split up into **two separate logs** and then **evaluated separately** as shown below:

① $log_{10}(\frac{1000}{10}) = log_{10}(1000) - log_{10}(10)$

② $log_{10}(\frac{1000}{10}) = 3 - 1$

$\therefore log_{10}(\frac{1000}{10}) = 2$

① Apply our log law: $log_b\left(\frac{m}{n}\right) = log_b m - log_b n$

② Evaluate each logarithm separately

The **third key rule** is the for logarithms of the form: $log_b(m^n)$. This **rule states** that when a **number** is **raised to a power inside a logarithm**, it is **equivalent** to **multiplying the power by the logarithm** of the number. Mathematically, it can be expressed as follows:

$$log_b(m^n) = n log_b m$$

This property allows us to **simplify logarithms** involving exponents by **multiplying the exponent** with the **logarithm of the base**. For instance, if we had the **logarithm:** $log_5(25^2)$, we could **factor out** the **exponent 2** and then **evaluate the logarithm** as follows:

① $log_5(25^2) = 2\,log_5(25)$

② $log_5(25^2) = 2(2)$

$\therefore log_5(25^2) = 4$

① **Apply our log law:** $log_b(m^n) = n\,log_b m$

② **Evaluate** $log_5(25)$ and then **multiply by 2**

Beyond these **three main rules**, there are also **other key properties** such as: $log_b(1) = 0$ and $log_b(b) = 1$. All these **log laws** along with the **corresponding exponential laws** can be summarised in the table below:

Log laws	Index Laws
$log_b(mn) = log_b m + log_b n$	$b^m \times b^n = b^{m+n}$
$log_b\left(\frac{m}{n}\right) = log_b m - log_b n$	$\frac{b^m}{b^n} = b^{m-n}$
$log_b(m^n) = n\,log_b m$	$(b^m)^n = b^{mn}$
$log_b(b^n) = n$	$b^{log_b n} = n$
$log_b(1) = 0$	$b^0 = 1$
$log_b(b) = 1$	$a^1 = a$

Understanding and **applying these logarithmic properties** can help us **simplify** and **manipulate logarithmic expressions**. We can break down complex logarithmic equations or expressions into simpler components and **solve them step by step**, which we will **explore** in the **next section**.

Worked Example 1

Kerry is one of the star students at Mathematics College and she intends to keep it that way by **learning ahead of the class. Help Kerry use log laws** o **simplify** the **following logarithmic functions**.

(i) $log_3(9 \times 27) = log_3(9) + log_3(27)$

$log_3(9 \times 27) = 2 + 3$

$\therefore log_3(9 \times 27) = 5$

(ii) $log_2\left(\frac{2^5}{8}\right) = log_2(2^5) - log_2(8)$

$log_2\left(\frac{2^5}{8}\right) = 5log_2(2) - log_2(8)$

$log_2\left(\frac{2^5}{8}\right) = 5(1) - 3$

$\therefore log_2\left(\frac{2^5}{8}\right) = 2$

Solving Logarithmic Equations

Solving logarithmic equations involves finding the **value** of the **variable** that **satisfies** the given **logarithmic equation**. By **applying logarithmic properties** and **algebraic techniques**, we can simplify the equations and isolate the variable. To start, let's suppose we want to **solve for x** in the equation: $log_2(x) + log_2(x+2) = 3$. To **solve for x**, we can apply the log law: $log_b(mn) = log_b m + log_b n$, which will **simplify the logarithm** as follows:

$$log_2(x) + log_2(x+2) = 3$$
$$log_2[(x)(x+2)] = 3$$
$$log_2[x^2 + 2x] = 3$$

With the **logarithm simplified**, we can then **rewrite the equation** in the **exponential form** and then **re-arrange** the equation to **solve for x**:

$$2^3 = x^2 + 2x$$
$$0 = x^2 + 2x - 8$$
$$0 = (x+4)(x-2)$$
$$\therefore x = -4, x = 2$$

As a second example, let's suppose that we want to **solve for x** in the equation: $4^{2x+5} = 2^x$. To **solve** this, we can start by **taking the logarithm** with a **base 2 on both sides of the equation**:

$$4^{2x+5} = 2^x$$
$$\text{①} \quad log_2(4^{2x+5}) = log_2(2^x)$$

① **Take the log_2 of both sides**

Even though the equation doesn't look easy to solve now, we can use our **log law**: $log_b(m^n) = n log_b m$, to get the $2x+5$ and the x to the **front of both sides of the equation**:

$$log_2(4^{2x+5}) = log_2(2^x)$$
$$\text{②} \quad (2x+5)log_2(4) = x log_2(2)$$

② **Apply our log law: $log_b(m^n) = n log_b m$**

From there, we can **solve for x** by **expanding** and **re-arranging the equation** to **isolate x** as follows:

$$(2x+5)log_2(4) = x log_2(2)$$
$$\text{③} \quad 2x log_2(4) + 5 log_2(4) = x log_2(2)$$
$$5 log_2(4) = x log_2(2) - 2x log_2(4)$$
$$5 log_2(4) = x\left(log_2(2) - 2 log_2(4)\right)$$
$$\text{④} \quad x = \frac{5 log_2(4)}{log_2(2) - 2 log_2(4)} = \frac{5(2)}{1 - 2(2)} = -\frac{10}{3}$$

③ **Expand out the equation**

④ **Re-arrange the equation** to **isolate x**

As another example, let's suppose we want to **solve for x** in the equation: $log_5(3x-2) = log_5(4)$. In this case, since the **logarithms have the same base**, we can **simply equate** the **expressions inside the logarithms** as follows:

$$log_5(3x-2) = log_5(4)$$
$$3x - 2 = 4$$
$$3x = \frac{4+2}{3}$$
$$\therefore x = 2$$

As a final, **more complicated example**, let's imagine we want to **solve for x** in the equation $3^{2x} - 3^x + 6 = 0$. To **solve for x**, we can start by **rewriting the equation** as a **quadratic equation** in terms of a **new variable**, let's say 'y', where $y = 3^x$. This will give:

$$3^{2x} - 3^x + 6 = 0$$

① $$y^2 - y + 6 = 0$$

$$(y-3)(y+2) = 0$$

① Let $y = 3^x$ to rewrite the equation as a quadratic equation

Next, we can **set each factor** of the quadratic equation **equal to zero** and **solve for y**:

② $(y-3) = 0$, $y = 3$ $\qquad$ $(y+2) = 0$, $y = -2$

② Solve for y for each factor

Now, we can **substitute $y = 3^x$** into the equation to determine the **value of x**:

$$y - 3 = 0$$

③ $$3^x - 3 = 0$$

$$3^x = 3$$

$$x = log_3(3)$$

$$x = 1$$

$$y + 2 = 0$$

$$3^x + 2 = 0$$

$$3^x = -3$$

∴ No real solution (since the log of a negative number is undefined)

③ Substitute in $y = 3^x$

④ State the final answer

④ ∴ The **solution** to the equation is $x = 1$.

As shown in these examples, there are **various ways** that we can **use logarithms to solve equations**. The best way to **understand the different scenarios** is through the **problem sets**.

Worked Example 1

Seeing that Kerry is eager to get ahead with her knowledge of logarithms, Teacher Simon decides to **help Kerry** by giving her some **equations to solve using logarithms**.

(a) Help Kerry to **solve for x** in the equation: $7^{x+1} = 8$.

$$7^{x+1} = 8$$

$$log_{10}(7^{x+1}) = log_{10}(8)$$

$$(x+1)log_{10}(7) = log_{10}(8)$$

$$xlog_{10}(7) + log_{10}(7) = log_{10}(8)$$

$$\therefore x = \frac{log_{10}(8) - log_{10}(7)}{log_{10}(7)}$$

(b) Help Kerry to **solve for x** in the equation: $2^{2x} + 7(2^x) - 8 = 0$.

$$2^{2x} + 7(2^x) - 8 = 0$$

$$y^2 + 7y - 8 = 0$$

$$(y-1)(y+8) = 0$$

$$y - 1 = 0$$

$$2^x - 1 = 0$$

$$2^x = 1$$

$$\therefore x = 0$$

$$y + 8 = 0$$

$$2^x + 2 = 0$$

$$3^x = -3$$

∴ No real solution

∴ The **solution** to the equation is $x = 0$.

7.2 Natural Logarithms and Change of Base

As we have already explored, logarithms of the form $x = log_b\, y$ can have a **base** b that is essentially of **any value**. However, there are **two specific types** of logarithms we commonly use: **natural logarithms** and **logarithms with a changed base**. The difference between these two logarithms is the **type of base used**. Understanding **natural logarithms** and the concept of **changing the base of logarithms** is useful in many **real-world applications** that we will explore soon.

Natural Logarithms

The **natural logarithm** is a **logarithm** with the **base 'e'**, where **'e'** is **Euler's number**, an irrational number approximately equal to 2.71828. The **natural logarithm function** is defined as: $ln(x) = log_e(x)$

Like the other logarithms that we have explored, **natural logarithms** have a **special relationship** with **exponential functions**, particularly with the **base 'e'**. This relationship is **expressed as**:

$$x = ln(b) \iff b = e^x$$

This **relationship** can be **used to solve equations** involving exponential functions by **using the properties of natural logarithms**. Suppose we want to **solve for x** in the **equation** $3e^x = 15$. To **solve for x** we can start by **dividing both sides** of the **equation by 3** and then **re-arranging** into a **natural logarithm** as follows:

$$3e^x = 15$$
$$① \quad e^x = 5$$
$$② \quad x = ln(5)$$

① **Divide both sides** of the **equation** by **3**

② **Re-arrange** into the **natural logarithm form** to get **value of x**

As a **second example**, let's consider **solving for x** in the equation $2^{3x+1} = 6$. We can start by **taking the natural logarithm** of **both sides** which will give:

$$2^{3x+1} = 6$$
$$① \quad ln\left(2^{3x+1}\right) = ln(6)$$

① **Take the natural logarithm of both sides**

While this may look confusing now, we can use our **log law** $log_b(m^n) = n log_b m$ to get $3x + 1$ at the **front of the equation**:

$$ln\left(2^{3x+1}\right) = ln(6)$$
$$② \quad (3x+1)ln(2) = ln(6)$$

② **Apply our log law:** $log_b(m^n) = n log_b m$

From there, we can **solve for x** by **re-arranging the equation** to **isolate x** as follows:

$$(3x+1)ln(2) = ln(6)$$
$$③ \quad 3x + 1 = \frac{ln(6)}{ln(2)}$$
$$3x = \frac{ln(6)}{ln(2)} - 1$$
$$④ \quad \therefore x = \frac{ln(6)}{3ln(2)} - \frac{1}{3}$$

③ **Re-arrange the equation** to **isolate x**

④ **State the final answer**

To reinforce this process, let's consider a second example of **solving for x** in the equation: $4^{2x-3} = 10$. We can follow the **same process** above of **taking the natural logarithm** of **both sides**, applying the **log law** $log_b(m^n) = nlog_b m$, and then **rearranging** to **solve for x**.

$$4^{2x-3} = 10$$

① $ln(4^{2x-3}) = ln(10)$

② $(2x-3)ln(4) = ln(10)$

③ $2x - 3 = \frac{ln(10)}{ln(4)}$

$2x = \frac{ln(10)}{ln(4)} + 3$

④ $\therefore x = \frac{ln(10)}{2ln(4)} + \frac{3}{2}$

① **Take the natural logarithm of both sides**

② **Apply our log law:** $log_b(m^n) = nlog_b m$

③ **Re-arrange the equation** to **isolate x**

④ **State the final answer**

As a **more complicated example**, suppose we are asked to **solve** the equation: $e^{2x} + e^x - 6 = 0$. To **solve for x**, we can start by **rewriting the equation** as a **quadratic equation** in terms of **'y'**, where $y = e^x$. This will give:

$$e^{2x} + e^x - 6 = 0$$

① $y^2 + y - 6 = 0$

$(y-2)(y+3) = 0$

① **Let $y = e^x$ to rewrite the equation as a quadratic equation**

Next, we can **set each factor equal to zero** and **solve for y**:

② $(y-2) = 0$, $y = 2$ | $(y+3) = 0$, $y = -3$

② **Solve for y for each factor**

Now, we can **substitute $y = e^x$** into the equation to determine the **value of x**:

③
$y - 2 = 0$
$e^x - 2 = 0$
$e^x = 2$
$x = ln(2)$

$y + 3 = 0$
$e^x + 3 = 0$
$e^x = -3$
No real solution

③ **Substitute in $y = e^x$**

④ **State the final answer**

④ $\therefore$ The **solution** to the equation is $x = ln(2)$.

Change of Base

In logarithms, the **base determines** the **behaviour and properties** of the **logarithmic function**. However, it is often necessary to **work with logarithms** of **different bases** or to **convert logarithms from one base to another**. The **change of base formula** allows us to **convert a logarithm** with **one base** to a **logarithm with a different base**. The **change of base formula** is:

Change of Base: $log_b x = \frac{log_c x}{log_c b}$

For instance, let's evaluate $log_4 16$ using the **change of base formula** to **base 10**. Using the **change of base formula** we can write:

$$log_4 16 = \frac{log_{10} 16}{log_{10} 4}$$

Now by **using our calculators**, we can **evaluate the logarithm** as follows:

$$log_4 16 = \frac{log_{10} 16}{log_{10} 4}$$

$$log_4 16 = \frac{1.2041}{0.6021}$$

$$log_4 16 \approx 2$$

Additionally, we can use the **natural logarithm (ln)** as the **base** for the **change of base formula**. For instance, we can use it to convert $log_5 10$ to a **natural logarithm** using the **change of base formula**:

$$log_5 10 = \frac{ln(10)}{ln(5)}$$

$$log_5 10 = \frac{2.3026}{1.6094}$$

$$log_5 10 \approx 1.43$$

To **consolidate all the knowledge** you have just learnt in relation to **natural logarithms** and **change of base**, practise working through the **worked example** below.

Worked Example 1

Continuing to explore her interest in **logarithmic functions**, **help Kerry** to **determine the following**.

(a) If Kerry has the **equation:** $2^{2x-2} = 3$, help her to **determine the value of** x.

$$2^{2x-2} = 3$$

$$ln\left(2^{2x-2}\right) = ln(3)$$

$$(2x - 2)ln(2) = ln(3)$$

$$2x - 2 = \frac{ln(3)}{ln(2)}$$

$$2x = \frac{ln(3)}{ln(2)} + 2$$

$$\therefore x = \frac{ln(3)}{2ln(2)} + 1$$

(b) If Kerry has the **equation:** $e^{2x} + 2e^x - 3 = 0$, help her to **determine the value of** x.

$$e^{2x} + 2e^x - 3 = 0$$

$$y^2 + 2y - 3 = 0$$

$$(y - 1)(y + 3) = 0$$

$(y - 1) = 0$	$(y + 3) = 0$
$y = 1$	$y = -3$
Sub in $y = e^x$**:**	**Sub in** $y = e^x$**:**
$e^x - 1 = 0$	$e^x + 3 = 0$
$e^x = 1$	$e^x = -3$
$x = ln(1) = 0$	**No real solution**

$\therefore$ the **solution** to the equation is $x = 0$.

7.3 Applications of Logarithmic Functions

Logarithms are **not just** mathematical concepts; they have **practical applications** in **various real-world scenarios**. In this section and in the **problem sets**, we are going to explore some of the **most common applications** of **logarithms in real-world contexts**.

For instance, **logarithms** are often used to **measure sound intensity** using the **decibel scale**. The **decibel (dB) scale** is a **logarithmic scale** that measures the **relative loudness of sounds**. It allows us to compare and **quantify the differences** in **sound intensity levels**. The **formula** used to **calculate sound intensity** in **decibels** is:

$$dB = 10log_{10}\left(\frac{I}{I_o}\right)$$

In this formula, I is the **sound intensity being measured** and I_o is the **reference sound intensity** (typically set at the threshold of human hearing).

Applying this to a **real-world example**, let's suppose we have **two sounds**: **sound** A with an **intensity** of $I_A = 10^{-5}\ watts/m^2$ and **sound** B with an **intensity** of $I_B = 10^{-3}\ watts/m^2$, where the reference sound intensity is $I_B = 10^{-12}\ watts/m^2$. To **calculate the difference** in the **sound intensity** in **decibels**, we can use our **logarithmic formula** as follows:

For sound A:

$$dB_A = 10log_{10}\left(\frac{I}{I_o}\right)$$

$$dB_A = 10log_{10}\left(\frac{10^{-5}}{10^{-12}}\right)$$

$$dB_A = 120dB$$

For sound B:

$$dB = 10log_{10}\left(\frac{I}{I_o}\right)$$

$$dB_B = 10log_{10}\left(\frac{10^{-3}}{10^{-12}}\right)$$

$$dB_B = 140dB$$

∴ The difference in **sound intensity** between **sound** A and **sound** B is $20dB$

As a second real-world example, logarithms are also used to **calculate the pH** of a solution in chemistry. The **formula** to **calculate pH** is:

$$pH = -log_{10}[H^+]$$

In this equation, $[H^+]$ represents the **concentration of hydrogen ions** in the solution. With this knowledge in mind, let's suppose we have **two solutions**: **solution** A with a **pH of 2** and **solution** B with a **pH of 5**. If we wanted to know the **difference** in the **hydrogen ion concentration** between these two solutions, we could **re-arrange** the **logarithmic equation** above into its **exponential form** to get:

$$[H^+] = 10^{-pH}$$

We can then **substitute** the **pH values** to **compare** the **hydrogen ion concentrations**:

For solution A:

$$[H^+]_A = 10^{-pH}$$

$$[H^+]_A = 10^{-2}$$

$$[H^+]_A = 0.01\ mol\ L^{-1}$$

For solution B:

$$[H^+]_B = 10^{-pH}$$

$$[H^+]_B = 10^{-5}$$

$$[H^+]_B = 0.00001\ mol\ L^{-1}$$

Comparing the **two concentrations**, we can **calculate** the **ratio** of $[H^+]_A$ to $[H^+]_B$ to be:

$$\frac{[H^+]_A}{[H^+]_B} = \frac{0.01}{0.00001}$$

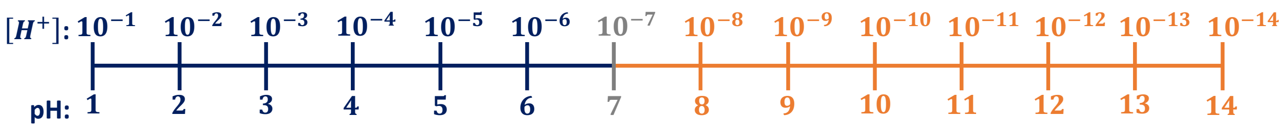
$$\frac{[H^+]_A}{[H^+]_B} = 1000$$

The **ratio** tells us that the **hydrogen ion concentration** in **solution** A is **1000 times** the **concentration** of **solution** B. This example shows us that even though the **pH** of the **solutions** is **2** and **5**, the **actual concentration difference** between the solutions is **far greater**.

Logarithmic Scales

We now know that in many real-world scenarios, we encounter data that **spans a wide range of values** such as **sound intensity** and **hydrogen ion concentration**. **Linear scales**, where **equal spacing represents equal differences**, may **not be suitable** in these cases. Logarithmic scales address this issue by using **equal spacing** to **represent multiplicative factors** or **ratios**. **Logarithmic scales** allow us to **visualize and compare data** that covers **several orders of magnitude effectively**.

For instance, in the previous section you may have noticed that the **pH scale** is **logarithmic scale**. This is because **each unit on the pH scale** represents a **tenfold difference** in the **hydrogen ion concentration**. As seen below, this **logarithmic relationship** allows us to compress a **wide range** of **concentration values** into a **more manageable scale**.



This may help you to understand why in the previous section we determined that a **decreasing the pH** from **5 to 2**, resulted in a hydrogen ion concentration being **1000 times larger** for the **pH of 2**.

Another **logarithmic scale** is the **Richter scale**. The **Richter scale** is a **logarithmic scale** that **measures** the **energy released by an earthquake**. Each **increase** of **one unit** on the **Richter scale** corresponds to a **tenfold increase** in the **amplitude of seismic waves**.

To explore this **logarithmic scale**, let's suppose an **earthquake** with a **magnitude of 5** on the **Richter scale** occurred in **Region A**, while **another earthquake** with a **magnitude of 6.5** on the Richter scale occurred in **Region B**. Let's compare the **difference in magnitude** between these **two earthquakes**:

$$Magnitude\ difference = Magnitude\ of\ Region\ B - Magnitude\ of\ Region\ A$$

$$Magnitude\ difference = 6.5 - 5.0$$

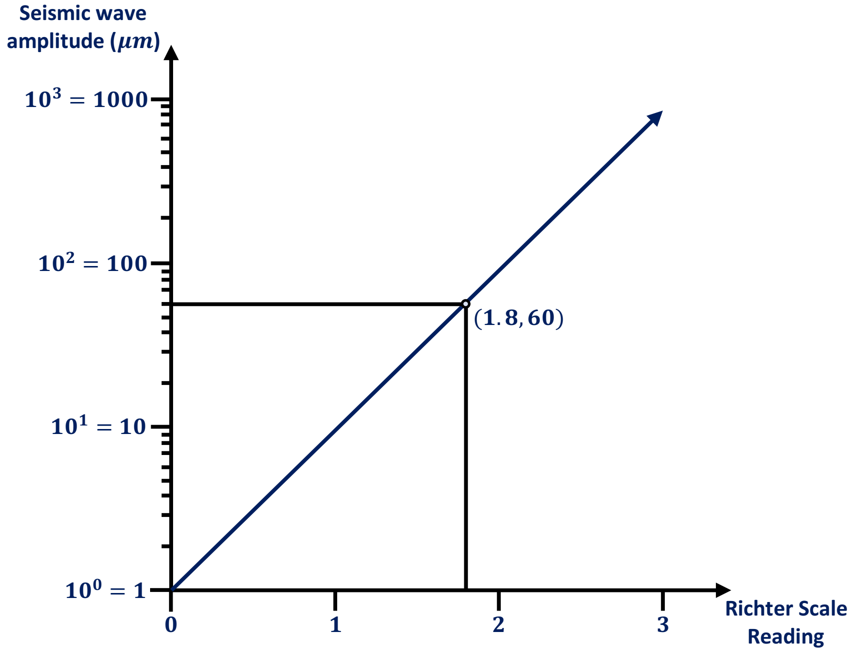
$$Magnitude\ difference = 1.5$$

As this is a **logarithmic scale**, **one unit increase** on the **Richter scale** corresponds to a **tenfold increase** in the **amplitude of seismic waves**. Therefore, the **magnitude difference** will be:

$$10^{1.5} = 31.62$$

∴ Earthquake in region B is 31.62 **times more instense** than in **region** A

Log-linear graphs are also commonly used to **represent logarithmic scales**. For instance, the **log-linear graph** for the **Richter scale** is shown below. Here you can see that a **linear relationship** is **maintained** despite it being a **logarithmic function**. This is because the **y-axis increases ten-fold** to the **x-axis**. These graphs can be used to determine key information. For instance, as shown below, for a **Richter scale reading** of 1.8 will give a **seismic wave amplitude** of $60\mu m$.



Worked Example 2

Rupert is a keen music student and knows that the **ratio of frequencies** between **two notes** can be **calculated** using the **logarithm of the frequency ratio** $log\left(\frac{f_2}{f_1}\right)$, where f_1 and f_2 are the frequencies of the two notes.

(a) If the **frequency of note 1** is $220Hz$ and the **frequency** of **note 2** is $440Hz$, help **Rupert** to **calculate** the **frequency ratio** between these two notes.

$$Frequency\ ratio = log\left(\frac{f_2}{f_1}\right)$$

$$Frequency\ ratio = log\left(\frac{440}{220}\right)$$

$$Frequency\ ratio = log(2)$$

$$Frequnecy\ ratio = 0.30103$$

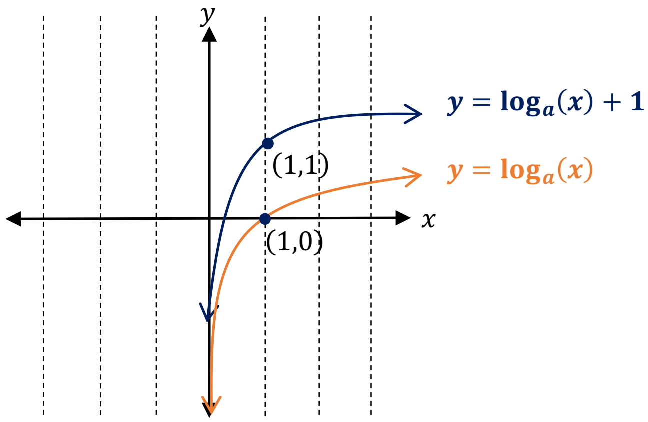
∴ **the frequency ratio between the two notes is approximately** 0.30103

7.4 Transformations of Logarithmic Curves

Conveniently, **the function transformations** which you have learned in the past **apply to logs too**. Let's define a simple log function as $f(x) = \log_a x$, with the base $a > 0$.

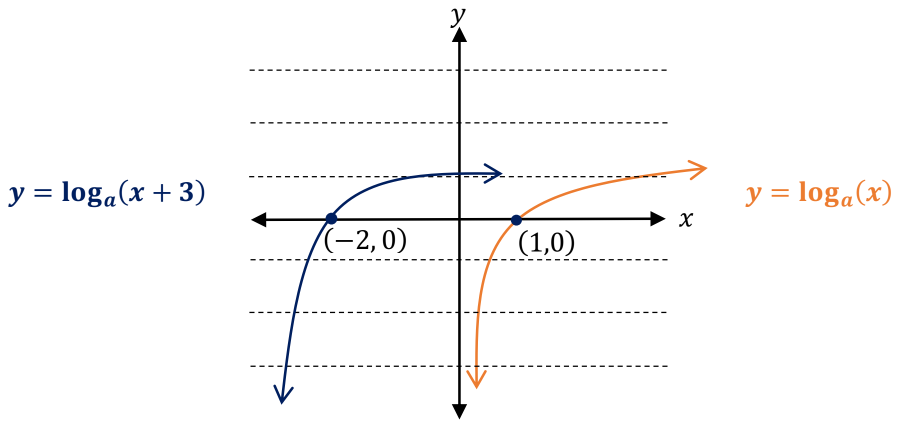
Vertical Translation: $y = f(x) + b$

If b is **positive**, the curve shifts **up**. For every point on the curve, each y value is increased by the magnitude of b. In the example below, the part of the curve which originally passed through the x-axis has moved 1 unit up. If b is **negative**, the curve shifts **down**. Every corresponding y value of the function is decreased by the magnitude of b.



Horizontal Translation: $y = f(x + c)$

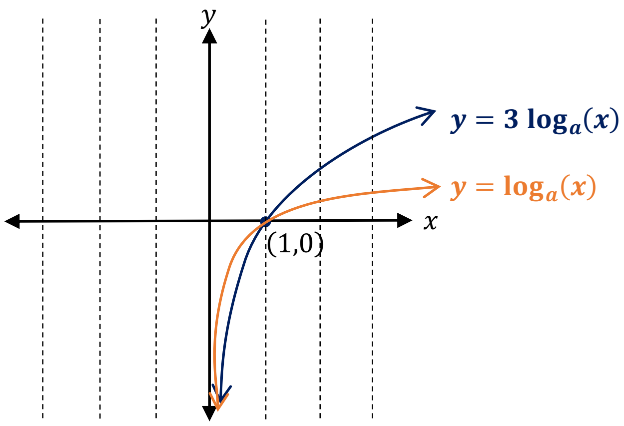
If c is **positive**, the curve shifts **left**. Since our new function is $\log_a(x + c)$, all x values of the new curve must **decrease** by the value of c to have the equivalent y values from the original function $f(x)$. In the example below, $c = 3$. If c is **negative**, the curve shifts **right**, since all x values of the new curve must **increase** by the value of c to correspond to the same y values of the original function $f(x)$.



Vertical Dilation: $y = kf(x)$

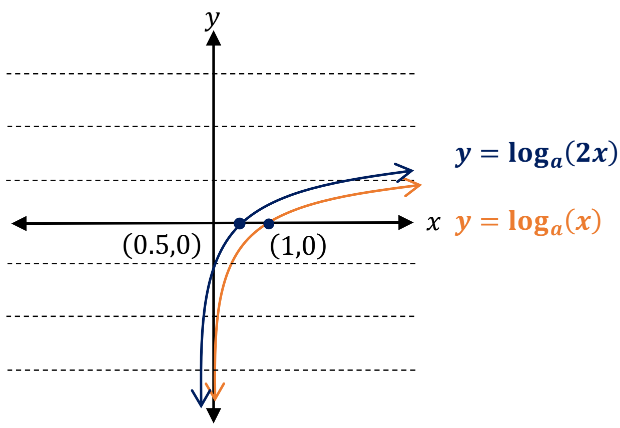
As k becomes **larger**, the curve dilates **away** from the x-axis. For every point on the curve, each y value is **scaled** by a factor **proportional** to the magnitude of k. Notice how the intercept for the example shown in the graph below did not change. As k becomes **smaller**, the curve contracts

towards the $\boldsymbol{x}$-axis. The corresponding $\boldsymbol{y}$ values of the function are again scaled proportionally to $\boldsymbol{k}$. You can also visualise this as 'compressing' the graph. For $\boldsymbol{y = kf(x)}$, the curve is stretched/compressed vertically by a factor **proportional** to $\boldsymbol{k}$.



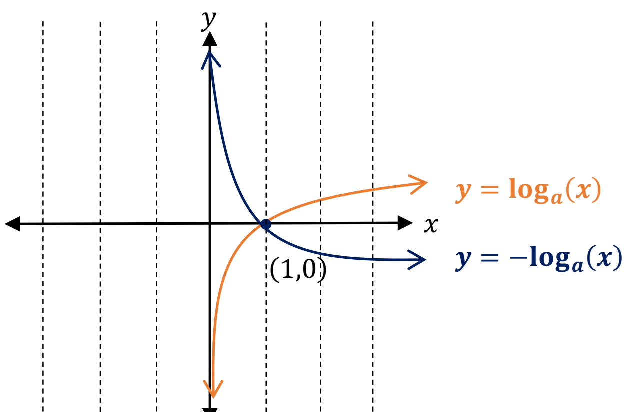
Horizontal Dilation: $\boldsymbol{y = f(kx)}$

As $\boldsymbol{k}$ becomes **larger**, the curve contracts **towards** the $\boldsymbol{y}$-axis. For every point on the curve, each $\boldsymbol{x}$ value is **scaled inversely proportional** to $\boldsymbol{k}$. Shown in the example below, is how the intercept has changed as a result. In the example, $\boldsymbol{k = 2.}$ As $\boldsymbol{k}$ becomes **smaller**, the curve dilates **away** from the $\boldsymbol{y}$-axis. You can think of this as 'stretching.' For $\boldsymbol{y = f(\mathrm{k}x)}$, the curve is stretched/compressed horizontally by a factor **inversely proportional** to $\boldsymbol{k}$.



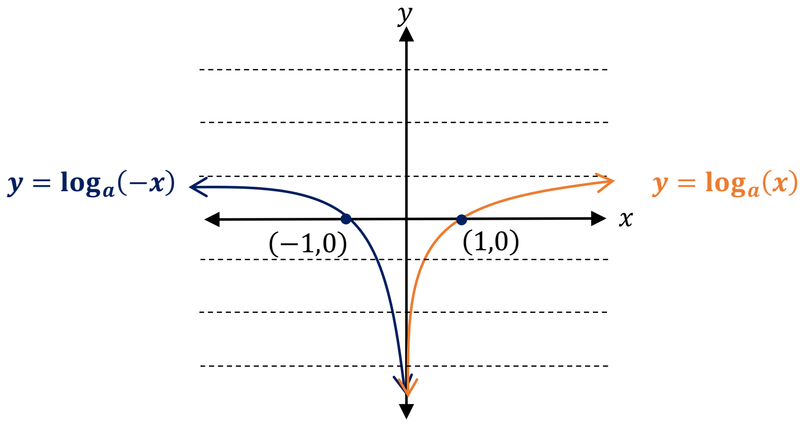
Reflection over $\boldsymbol{x}$-axis: $\boldsymbol{y = -f(x)}$

Perhaps the simplest of the transformations is changing the sign of the function by multiplying through by (-1). You can also think of it as 'mirroring' or 'reflecting' the function top-to-bottom. For every point on the curve, **the sign of each $\boldsymbol{y}$ value is inverted** while **corresponding $\boldsymbol{x}$ values are the same**.



Reflection over y-axis: $y = f(-x)$

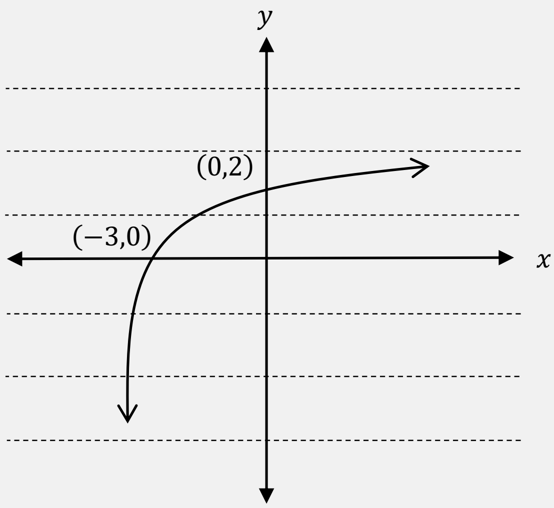
At first glance, given $y = \log_a x$, one might think that $f(-x)$ can't be used to transform the function. Indeed, you can't have the log of a negative number, but the **negative values of x become positive** as the signs cancel out. For every point on the curve, **the sign of each x value is inverted** while **the corresponding y values remain the same**.



The transformations can be applied to $f(x)$ simultaneously, with each combining to produce a result.

Worked Example 2

(a) Star student Kerry is looking to understand logarithmic transformations. Help Kerry find the equation of the following graph.



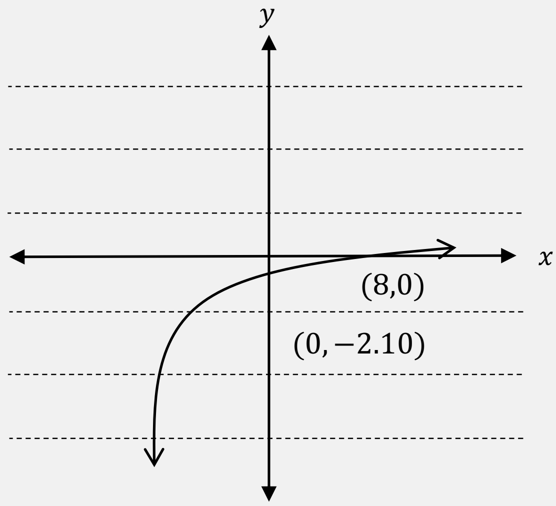
$$y = \log_a(x + b)$$
$$0 = \log_a(-3 + b)$$
$$b - 3 = 1$$
$$b = 4$$
$$2 = \log_a(4)$$
$$a^2 = 4$$
$$a = 2$$
Hence $y = \log_2(x + 4)$

Worked Example 2 continued

(b) To try the opposite process of graphing logarithmic functions, help Kerry to graph the following equation:

$$\boldsymbol{y = 3\log_{10}(x+2) - 3}$$

.



Congratulations, you have now completed the **first half of logarithms**! **Consolidate your knowledge** in the **problem sets** and once you are ready, we are going to **explore** the **differentiation** and **integration** of **logarithms**.

Problem Set 9 – Logarithms
Progressive Questions

Concept 1
Logarithms – Progressive Questions

(9 questions)

Repetitive questions: 1.11 – 1.92

The International Mathematics Championship!

Starring: Teacher Andrew, Janitor Peter, the Methods Class

Logarithmic Functions: Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9

Repetitive: 1.11-1.92

Rested and ready to tackle the second semester at Mathematics College, the students can't wait for more adventures. As they are walking to class, Peter Janitor excitedly calls the students over. The students have been selected to travel to Peru for the International Mathematics Championship! This year, the competition will be on logarithmic functions, so the students begin to study hard for the competition!

[20 marks]

1. Tyler can't remember what a **logarithmic function** is. Help Tyler evaluate all **real values** of x for all following **equations**.

(a) $\log_3 9 = x$ **(1)**

(b) $\log_6 x = 2$ **(1)**

(c) $\log_2 16 = 4x$ **(2)**

(d) $\log_3 9x = 4$ **(2)**

(e) $7\log_5 125 = 27x - 4$ **(3)**

(f) $\log_3 x = 2\log_9 81$ **(3)**

(g) $5\log_8 x^6 = 4^2 - 6$ **(4)**

(h) $\mathrm{x}^3 = 7\mathrm{x}\log_{13} 1 + \mathrm{x}\log_4 64$ **(4)**

[12 marks]

2. Janitor Peter is impressed with Tyler's skills, but before he can let the students go, he asks them all to **complete the following questions**. **Help** the students answer following questions by rewriting **(a)** to **(c)** in **exponential form** $a^m = n$ and **(d)** to **(f)** in **logarithmic form** $log_a n = m$ where a, n and m are integers.

(a) $\log_7 49 = 2$ **(1)**

(b) $6\log_6 36 - 3 = 9$ **(2)**

(c) $\log_2 1 + \frac{\log_2 32}{20} = \frac{1}{2}$ **(3)**

(d) $3^4 = 27$ **(1)**

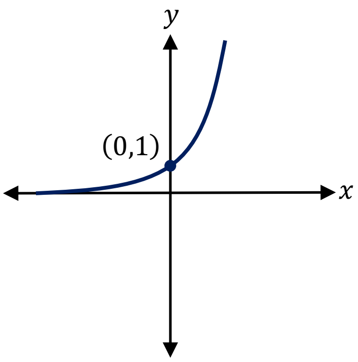
(e) $5^2 + 11 = 36$ **(2)**

(f) $\frac{4^3}{8} + 61^0 = 3^2$ **(3)**

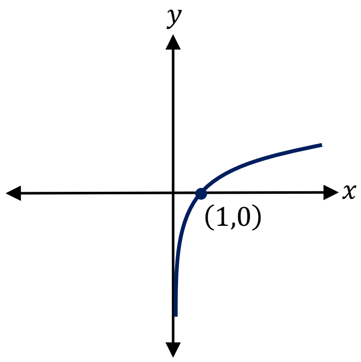
[10 marks]

3. The students impressed Janitor Peter and are all **invited to the championships**! The students board the plane. Jevon notices the team in front of him with a page of graphs. Wanting to outsmart that team, help Jevon:

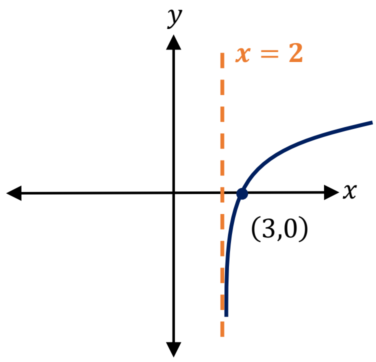
(a) **Match** the following **graphs** with the **appropriate equation.** **(3)**



Graph 1



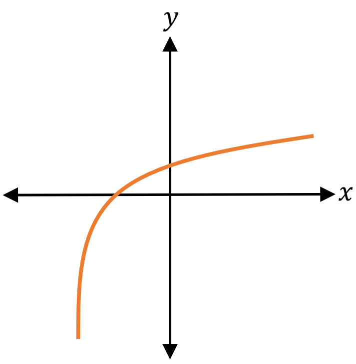
Graph 2



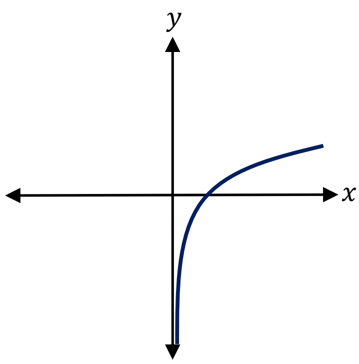
Graph 3

A. $\ln(x)$ **B.** $\log(x-2)$ **C.** e^x

(b) **Label** the **intercepts** and **asymptotes** of the following graphs. **(4)**



$\log(x+3)$



$\log(2x)$

(c) **Sketch a graph** for the function $y = \log 2x + 3$, labelling the **intercepts** and **asymptotes.** **(3)**

[14 marks]

4. The students land at the mathematics airport! For each piece of luggage they wish to collect, they must complete the following questions. Help the students **rewrite** the following equations using the **log laws** in terms of p, q, and r where $p = \log 2$, $q = \log 3$, and $r = \log 5$.

(a) $\log(6)$ **(2)**

(b) $\log\frac{3}{5}$ **(2)**

(c) $\log 25$ **(3)**

(d) $\log\frac{1}{2}$ **(2)**

(e) $\log_3 2$ **(2)**

(f) $\log_5 6$ **(3)**

[12 marks]

5. The students check-in to the **Log Hotel**. Their room numbers are expressed in terms of a logarithm. Help the students **find their room number** by **evaluating** the following expressions using the **log laws**.

(a) $\log 5 + \log 2$ **(2)**

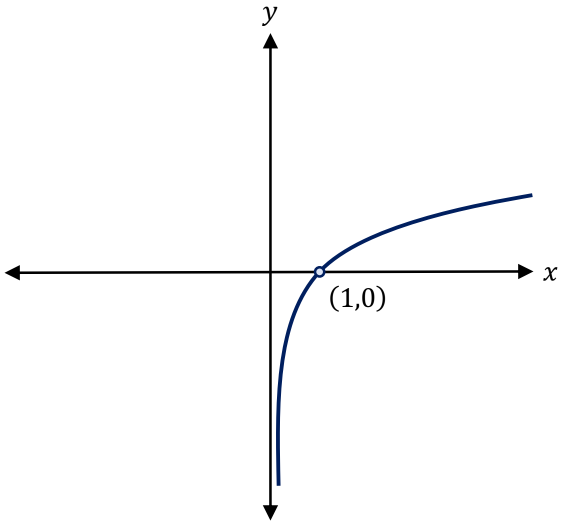
(b) $\log(\log 10)$ **(2)**

(c) $\log_2 12 - \log_2 3$ **(4)**

(d) $2\log_7 \sqrt{7} + \log_7(49^{-1})$ **(4)**

[12 marks]

6. Alexa is woken up in the middle of the night to a maths book falling from the shelf. She runs into the hall where she discovers it was a **small earthquake**! The **earthquake trend** can be **modelled** by the **curve shown below**. The following graph shows the **curve** for $\log_e(x)$:



Sketch the following equations, making sure to provide **at least one coordinate point** to describe the equation.

(a) **Vertical translation** seen in:
$\log_e(x) + 5$ **(3)**

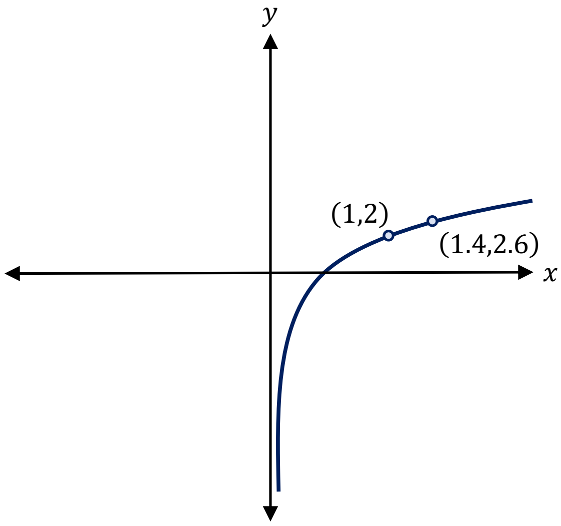
(b) **Horizontal translation** seen in:
$\log_e(x - 4)$ **(3)**

(c) **Horizontal dilation** seen in:
$\log_e(3x)$ **(3)**

(d) **Vertical dilation** seen in:
$\log_e(x^{\frac{1}{2}})$ **(3)**

[12 marks]

7. The students are told earthquakes are measured with the Richter scale with a **base** of $\mathbf{10}$. The graph of $y = \log_a(x)$ is shown below. **Using the graph**, help the students answer the following questions:



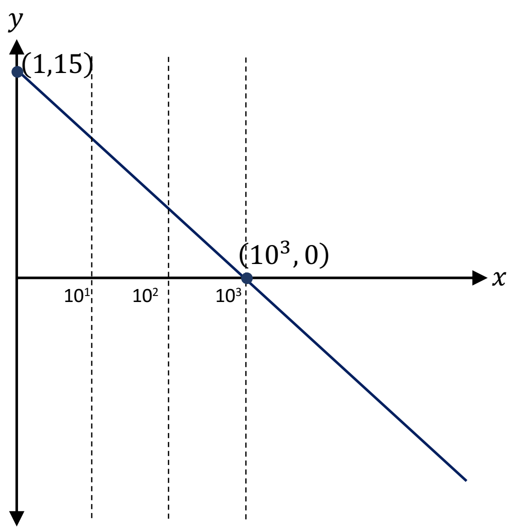
(a) Find the **value** of a, and state whether this is a **Richter scale logarithm** **(3)**

(b) Find the **value** of p such that $a^{p+0.2} - 2.6 = 0$. **(3)**

(c) **Sketch** the graph of $y = -\ln(-(x+1)) - 1$. **(4)**

[9 marks]

8. One of the **books** that **fell from the earthquake** showed a **log scale**. In this situation, for a function $y = f(x)$, instead of graphing x against y, $\log(x)$ is graphed against y. The **log scale plot** that the book contained is shown below.



(a) **Write** the function of the graph **in the form** of $y = I\log\left(\frac{x}{A}\right)$. **(4)**

(b) On the same graph, **sketch** $y = I\log\left(\frac{x}{0.1A}\right)$. **(2)**

(c) **Explain** the effect of **increasing** I and A on the log scale graph. **(3)**

[10 marks]

9. The team decide to create a new scale for measuring **how powerful earthquakes are**. They want to use the **amount of energy** it releases in **joules (E)** to calculate a **magnitude (M)** for the quake. They come up with the equation $M = 7 + Log_5(\frac{E}{E_0})$, where E_0 **is** the energy released by $100kg$ **of TNT**.

(a) The team estimates that $100kg$ **of TNT** releases 2000 **joules of energy**.

(i) State the equation for M in terms of E **(1)**

(ii) What would be the **magnitude of a quake** that released $50,000\ J$ **of energy**? **(2)**

(iii) **How much energy** would a **magnitude** 8 quake produce? **(2)**

(b) Lucky student Tom **does not know the energy released by** $100Kg$ **of TNT** but does find the team's notes containing the equation $M = 7 + Log(\frac{E}{E_0})$. He then overhears the team talking about a **magnitude** 6 and a **magnitude** 9 earthquake. Can Tom work out **how many times more energy is released by the bigger quake than the smaller one**? Support your answer with a calculation. **(5)**

The students have had an exciting day investigating all things earthquakes, but the competition is the next day. After a big day of studying and preparing the competition, the students sleep very early.

Problem Set 9 – Logarithms
Repetitive Questions

Concept 1
Logarithmic Functions: – Repetitive Questions
(14 questions)

Logarithmic Functions: Questions

[20 marks]

1.11 Solve the following equations for **all real values** of $\boldsymbol{x}$.

(a) $\log_4 64 = x$ **(1)**

(b) $\log_5 x = 2$ **(1)**

(c) $\log_6 36 = 9x$ **(2)**

(d) $\log_9 3x = 2$ **(2)**

(e) $\log_{12} 144 + 6 = 2\log_2 8 - x$ **(3)**

(f) $\log_3 9x = 14\log_7 49 - 28$ **(3)**

(g) $6\log_6 x^2 + 28 = 7^2 - 19$ **(4)**

(h) $x^3 = 53x\log_{13} 1 - 3x\log_3 27$ **(4)**

[15 marks]

1.12 Find the **value** of the following expressions **without** a calculator:

(a) $\log 0.1$ **(1)**

(b) $\log_3 27$ **(2)**

(c) $\log_4 64 - \log_4 4$ **(2)**

(d) $\log_3 \frac{1}{3} + \log 1000$ **(2)**

(e) $\log_3(\log_4 64)$ **(2)**

(f) $\log_{\frac{1}{2}} 2$ **(2)**

(g) $\log_2 \sqrt{2} - \log_2 16$ **(2)**

(h) $\log_5 125 + \log_5 \frac{1}{25}$ **(2)**

[14 marks]

1.21 Rewrite the following **logarithmic functions** as **exponential functions.**

(a) $\log_6 36 = 2$ **(1)**

(b) $7\log_3 27 - 3 = 18$ **(2)**

(c) $8\log_4 64 - 4 = 20$ **(2)**

(d) $7\log_2 32 - \log_{22} 1 = 35$ **(2)**

(e) $\log_{13} 169 = 2$ **(1)**

(f) $9 = 5\log_9 81 - 1$ **(2)**

(g) $11 = 2^2\log_{11} 121 + 3$ **(2)**

(h) $\log_{36} 1 + 3\log_9 3^6 = 9$ **(2)**

[10 marks]

1.22 Rewrite the following **exponential functions** as **logarithmic functions.**

(a) $5^3 = 125$ **(1)**

(b) $7^2 + 11 = 60$ **(2)**

(c) $\frac{2^3}{8} + 592^0 = 2$ **(2)**

(d) $4^3 = 64$ **(1)**

(e) $10^2 + 21 = 121$ **(2)**

(f) $\frac{4^3}{16} - 3 = (-1)^0$ **(2)**

[14 marks]

1.31 Rewrite the following questions using the **log laws** in terms of $\boldsymbol{p}$, $\boldsymbol{q}$, and $\boldsymbol{r}$ where $\boldsymbol{p = \log 5}$, $\boldsymbol{q = \log 7}$, and $\boldsymbol{r = \log 11}$.

(a) $\log(35)$ **(2)**

(b) $\log \frac{7}{11}$ **(2)**

(c) $\log 121$ **(2)**

(d) $\log \frac{1}{5}$ **(2)**

(e) $\log_{11} 5$ **(2)**

(f) $\log_5 35$ **(4)**

[12 marks]

1.32 Rewrite the following questions in terms of i, j, k, l, and m, as **separate terms** using the **log laws**.

(a) $\log(ijk)$ **(1)**

(b) $\log \frac{kl}{mi}$ **(2)**

(c) $\log j^0$ **(2)**

(d) $\log \frac{1}{lm}$ **(2)**

(e) $\log_i ij$ **(2)**

(f) $\log_k \frac{kl}{m}$ **(3)**

[11 marks]

1.41 Evaluate the following expressions using the **log laws.**

(a) $\log 25 + \log 4$ **(2)**

(b) $\log_4 12 - \log_4 3$ **(3)**

(c) $\log(\log_2 1024)$ **(2)**

(d) $2\log_5 \sqrt{25} + \log_5(125^{-1})$ **(4)**

[16 marks]

1.51 Evaluate each of the following using **log laws**.

(a) $\log_4 32 + \log_4 2$ **(2)**

(b) $\log_6 8 + \log_6 9 - \log_6 2$ **(3)**

(c) $\log_3 4.5 - \log_3 4 + \log_3 8$ **(2)**

(d) $\log_5 25 - \log_5 \frac{1}{125}$ **(2)**

(e) $\log(\log_2 1024)$ **(2)**

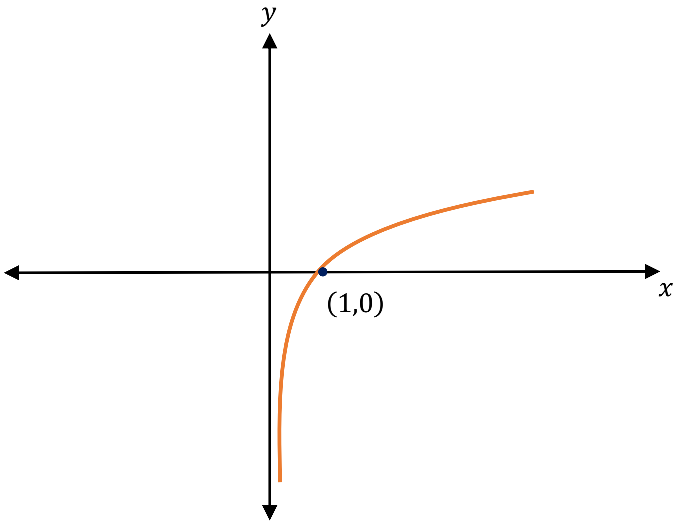
(f) $\log(\log(10) \times \log(10))$ **(2)**

(g) $\log_9 \left(\frac{1}{81}\right) - \log\left(\frac{1}{100}\right)$ **(3)**

(h) $2\log_4(16^{\frac{2}{3}}) \times 3\log_4(4^{0.25})$ **(3)**

[12 marks]

1.61 Below is the graph of $y = \log_3 x$



Sketch the following graphs, labelling all **key points** and **asymptotes**:

(a) $y = \log_3(x - 3)$ **(3)**

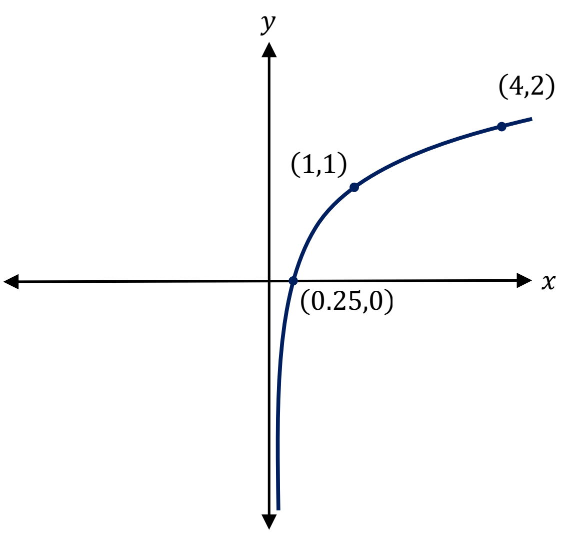
(b) $y = \log_3(x) + 2$ **(3)**

(c) $y = \log_3(x + 3) - 1$ **(3)**

(d) $y = \log_3(\sqrt[3]{x})$ **(3)**

[10 marks]

1.71 Below is the graph of $y = \log_a(x) + b$.



(a) Determine the **values** of the constants a and b. **(2)**

(b) Determine the following:

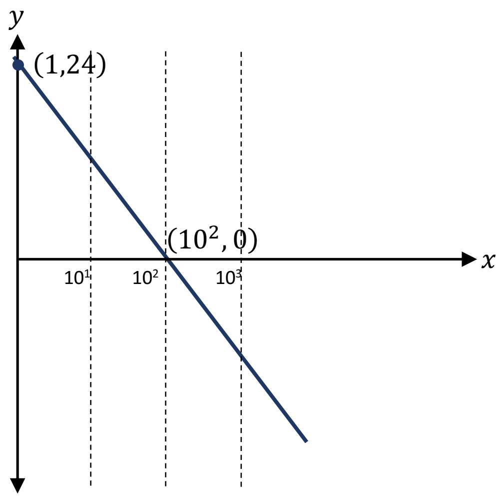
(i) The exact value of y if $x = 2$. **(3)**

(ii) The exact value of y if $x = 4\sqrt{2}$. **(3)**

(iii) The exact value of x if $y = 0.5$. **(3)**

[7 marks]

1.81 Similar to question 8 in the previous set, below is a graph of $\log(x)$. See the log scale plot below.

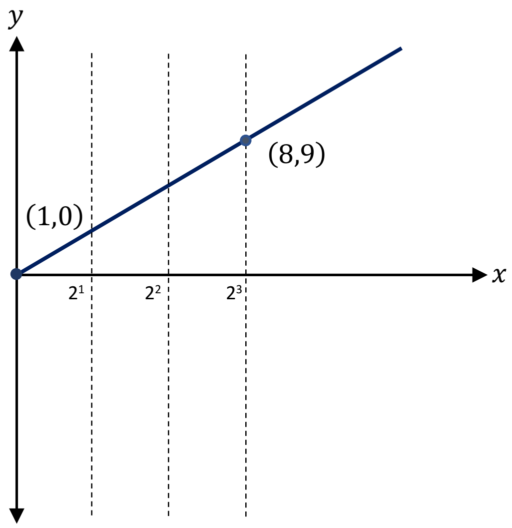


(a) **Write** the function of the graph **in the form** of $y = I\log\left(\frac{x}{A}\right)$. **(4)**

(b) **On the same graph, sketch** $y = \log\left(\frac{x}{0.5A}\right)$. **(3)**

[13 marks]

1.82 Consider the log scale shown below:



(a) Write the equation of the **graph** as $\boldsymbol{y = I\log_2\left(\frac{x}{A}\right)}$. **(4)**

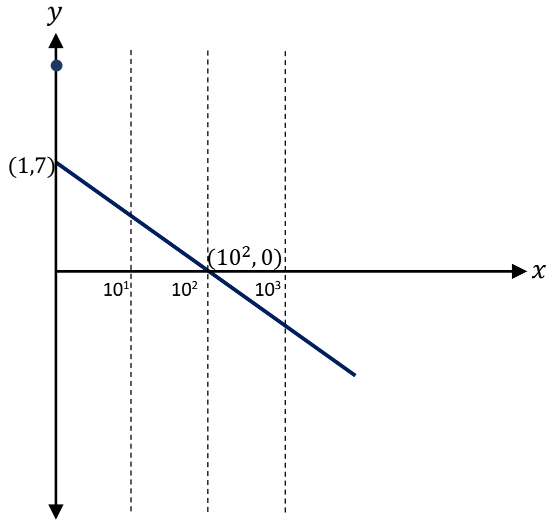
(b) **Rewrite** the equation as $\boldsymbol{y = M\log_B(x) + C}$. **(2)**

(c) Determine the value of $\boldsymbol{I}$ and $\boldsymbol{A}$ so that there will the **vertical translation** of **10** units **up**?**(4)**

(d) Using the log scale provided, **sketch** the graph of $\boldsymbol{y = \log_{10} x}$. **(3)**

[12 marks]

1.83 Consider the log scale shown below.



(a) Write the equation of the **graph** as $y = I\log\left(\frac{x}{A}\right)$ **(4)**

(b) On the same graph, sketch $y = \log\left(\frac{x}{0.7A}\right)$ **(3)**

(c) On the same graph, sketch $y = 2\log\left(\frac{x}{0.7A}\right)$ **(3)**

(d) If the value of A **increases**, what **effect** does it have on the graph? **(2)**

[8 marks]

1.91 The **loudness** of sound at a music festival featuring the band "Logger Rhythms" was found to be $L = 20\log\left(\frac{P}{2\times10^{-5}}\right)$ in decibels (dB) where P is the **pressure** of the sound wave in Pa.

(a) How much **pressure** does a sound wave make from a conversation if the **loudness** of a conversation is $60dB$? **(2)**

(b) Sounds can damage hearing when the loudness is **above** $85dB$. The bass player is delivering sounds of **pressure** $6Pa$. Will this sound cause hearing damage? **(2)**

(c) When the festival-goers returned home, they found the acoustics of their room changed the formula. If a conversation is $65dB$ given the **same pressure** found in (a) and the **loudness** of sound goes up by $10dB$ when the **pressure** is **2 times** as strong, what is the new formula in the form of $L = A\log\left(\frac{P}{P_0}\right)$? **(4)**

[6 marks]

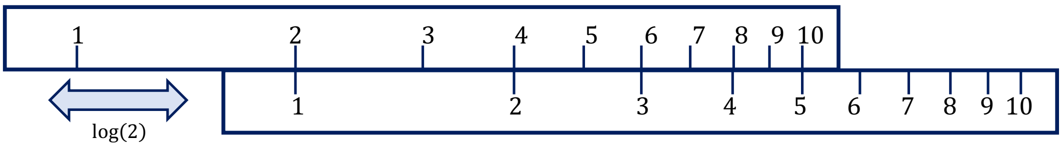
1.92 Hick's Law is a formula in psychology that estimates the average **time** (in seconds) it takes a person to decide between n choices given as $T = b\log_2(n+1)$, where b is a **constant** experimentally found.

(a) What is the value of b, given it takes someone 10 seconds to say yes **or** no? **(1)**

(b) When designing a menu, a café wants customers to take **no more** that 30 seconds to decide what to order. Previously, it took customers 20 seconds to decide between 5 options. What is the **largest** number of **options** the café should provide customers? **(4)**

[7 marks]

1.93 Before calculators were invented, there existed a tool called a "slide ruler" to help with multiplying and dividing numbers. This would consist of two rulers with markings distributed in an interesting way where the **number** x would be placed a **distance** $\log(x)$ from the marking of $\mathbf{1}$. See the slide rule below as an example:



In this situation, sliding the bottom ruler $\log(2)$ units to the **right**, causes the bottom numbers to line up with **multiples** of 2 of the upper ruler.

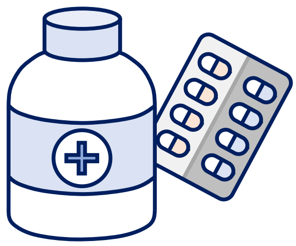
(a) What would be the distance between the **markings** 2 and 3? **(2)**

(b) Show that sliding the bottom ruler $\log(3)$ units to the **right**, causes the bottom numbers to line up with **multiples** of 3 of the upper ruler (hint, think about log **multiplication** rules). **(3)**

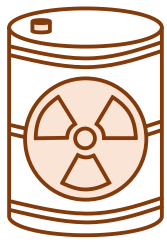
(c) **Explain** how **division** would work. **(2)**

7.5 Calculus of Natural Logarithms

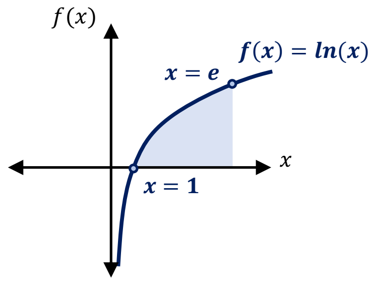
In Unit 3 we explored **exponential and trigonometric functions** and how we can **differentiate and integrate** them to obtain **useful information** for various real-world situations. By learning to **differentiate and integrate logarithmic functions** we can determine information such as the **absorption rate of medicine**, the **decay rate of a radioactive substance** or the **area under a curve**.



Rate of absorption of medicine



Rate of decay of a radioactive substance



Area under the curve of $f(x) = ln(x)$

When we explore the **differentiation** and **integration** of **logarithmic functions**, it is important to note that we are only **applying calculus** to **natural logarithms** such as $ln(x)$. In the cases that the logarithm is **not** in the natural logarithm format such as $log_{10}(x)$, we apply our **change of base formula**.

Differentiation of Logarithmic Functions

For the natural logarithms $ln(x)$, the **derivative** with respect to x is $\frac{1}{x}$, as shown in the formula below:

If $y = ln(x)$, then $\frac{dy}{dx} = \frac{1}{x}$

The best way to understand this formula is to **apply it** to some examples. For instance, let's suppose that we want to **differentiate** the functions: $y = 4ln(x)$ and $y = x^2 - 2ln(x)$. Let's go through them **step by step**.

$$y = 4ln(x)$$

① $\frac{dy}{dx} = 4\left(\frac{1}{x}\right)$

② $\frac{dy}{dx} = \frac{4}{x}$

$$y = x^2 - 2ln(x)$$

① $\frac{dy}{dx} = 2x - 2\left(\frac{1}{x}\right)$

② $\frac{dy}{dx} = 2x - \frac{2}{x}$

① **Apply** the **differentiation rule** if $y = ln(x)$ then $\frac{dy}{dx} = \frac{1}{x}$

② **State** the **final answer**

The **second rule** for the **differentiation of logarithms** is **derived** from the **chain rule**. It states that if a **function** $f(x)$ is **contained within the logarithm** (i.e. $ln(f(x))$), then the **derivative** is $f'(x)$ divided by $f(x)$. This rule can be stated as follows:

Derived from Chain Rule: If $y = ln(f(x))$ then $\frac{dy}{dx} = \frac{f'(x)}{f(x)}$

Again, the best way to understand the application of this rule is to **apply** it to some **examples**.

Suppose we have the functions $f(x) = ln(4x+2)$ **and** $y = ln(5x - x^2)$, and we want to determine the **derivative** of each function. We can apply the **shortcut rule derived from chain rule** as follows:

$$y = ln(4x+2)$$
$$① \quad \frac{dy}{dx} = \frac{f'(x)}{f(x)}$$
$$② \quad \frac{dy}{dx} = \frac{4}{4x+2}$$

$$y = ln(5x - x^2)$$
$$① \quad \frac{dy}{dx} = \frac{f'(x)}{f(x)}$$
$$② \quad \frac{dy}{dx} = \frac{5-2x}{5x - x^2}$$

① **Apply** the shortcut **rule derived** from **chain rule**

② **State** the **final answer**

To take it one step further, we can consider cases where the function contained within the **natural logarithm** is a **trigonometric function**. For instance, let's consider the functions $y = ln(sinx)$ and $y = ln(-cosx)$.

In these cases, we can again apply our **shortcut rule derived from chain rule** in **combination** with our **trigonometric differentiation rules**: $\frac{d}{dx}sin(x) = cos(x)$ and $\frac{d}{dx}cos(x) = -sin(x)$

$$y = ln(sinx)$$
$$① \quad \frac{dy}{dx} = \frac{f'(x)}{f(x)}$$
$$② \quad \frac{dy}{dx} = \frac{cos(x)}{sin(x)}$$

$$y = ln(-cosx)$$
$$① \quad \frac{dy}{dx} = \frac{f'(x)}{f(x)}$$
$$② \quad \frac{dy}{dx} = -\frac{sin(x)}{cos(x)}$$

① **Apply** the shortcut **rule derived** from **chain rule**

② **State** the **final answer**

With our new skills, we can also consider differentiating **more difficult natural logarithms** where we need to apply our understanding of **log laws**. For instance, suppose we wanted to **differentiate** $y = ln[(x+1)(x-2)]$. To differentiate this, we would need to apply our **log law**: $log_a mn = log_a m + log_a n$ to **split it up** into **two natural logarithms** and then **differentiate** each component individually.

$$y = ln[(x+1)(4x-2)]$$
$$① \quad y = ln(x+1) + ln(4x-2)$$
$$② \quad \frac{dy}{dx} = \frac{f'(x)}{f(x)}$$
$$③ \quad \frac{dy}{dx} = \frac{1}{x+1} + \frac{4}{4x-2}$$

① **Apply** the log law: $log_a mn = log_a m + log_a n$

② **Apply** the shortcut **rule derived** from **chain rule**

③ **State** the **final answer**

As a second example, if we wanted to differentiate: $y = ln\left[\frac{x^2+1}{2x-1}\right]$, we would apply our **log law**: $log_a \frac{m}{n} = log_a m - log_a n$ to **split it up** into **two natural logarithms** and then **differentiate** each component as follows:

$$y = ln\left[\frac{x^2+1}{2x-1}\right]$$
$$① \quad y = ln(x^2+1) - ln(2x-1)$$
$$② \quad \frac{dy}{dx} = \frac{f'(x)}{f(x)}$$
$$③ \quad \frac{dy}{dx} = \frac{2x}{x^2+1} - \frac{2}{2x-1}$$

① **Apply:** $log_a \frac{m}{n} = log_a m - log_a n$

② **Apply** the **shortcut rule derived** from **chain rule**

③ **State** the **final answer**

As a more complicated example involving **trigonometric functions**, let's consider the **differentiation** of $y = ln(tanx)$. To differentiate this, we first apply the property $tan(x) = \frac{sin\,(x)}{cos\,(x)}$, then split the logarithm up using the log law: $log_a \frac{m}{n} = log_a m - log_a n$, and finally **differentiate** each component separately, as shown below:

$$y = ln(tanx)$$

① $y = ln\left(\frac{sinx}{cosx}\right)$

② $y = ln(sinx) - ln(cosx)$

$$\frac{dy}{dx} = \frac{cos(x)}{sin(x)} - \frac{-sin(x)}{cos(x)}$$

③ $\frac{dy}{dx} = \frac{cos(x)}{sin(x)} + \frac{sin(x)}{cos(x)}$

① **Apply:** $tan(x) = \frac{sin\,(x)}{cos\,(x)}$

② **Apply:** $log_a \frac{m}{n} = log_a m - log_a n$

③ **Differentiate** to get **final answer**

Finally, let's consider cases when we have a logarithm that is **not in the form** of a **natural logarithm**. For $y = log_5 x$ and $y = log_{10} x$, we need to apply our **change of base formula**: $log_a b = \frac{log_c b}{log_c a}$, in order to **convert it** to a **natural logarithm form**, so we can then **differentiate** it as shown below:

$$y = log_5 x$$

① $log_a b = \frac{log_c b}{log_c a}$

$$y = \frac{ln(x)}{ln(5)}$$

$$y = ln(x) \times \frac{1}{ln(5)}$$

② $\frac{dy}{dx} = \frac{1}{x} \times \frac{1}{ln(5)}$

③ $\frac{dy}{dx} = \frac{1}{xln(5)}$

$$y = log_{10} x$$

① $log_a b = \frac{log_c b}{log_c a}$

$$y = \frac{ln(x)}{ln(10)}$$

② $y = ln(x) \times \frac{1}{ln(10)}$

③ $\frac{dy}{dx} = \frac{1}{x} \times \frac{1}{ln(10)}$

$$\frac{dy}{dx} = \frac{1}{xln(10)}$$

① **Apply change of base formula to get natural logarithm form**

② **Apply** the **shortcut rule derived** from **chain rule**

③ **Simplify** to get the **final answer**

Whilst there are **many different examples** through which logarithmic functions are differentiated, it is **important to remember** that there are **only two differentiation rules** that we are **actually applying**.

Worked Example 1

Kerry is looking to understand **logarithmic functions** and how she can **differentiate** them to use the **rate of change function** in various real world applications. Help Kerry to **determine** the **derivatives** of the following logarithmic functions.

(i) $y = ln(4x^2 - 5)$

$$\frac{dy}{dx} = \frac{f'(x)}{f(x)}$$

$$\frac{dy}{dx} = \frac{4}{4x - 5}$$

(ii) $y = ln(-sinx)$

$$\frac{dy}{dx} = \frac{f'(x)}{f(x)}$$

$$\frac{dy}{dx} = \frac{-cos(x)}{-sin(x)}$$

$$\frac{dy}{dx} = \frac{cos(x)}{sin(x)}$$

(iii) $y = ln[(2x - 8)(x^2 + 4)]$

$$y = ln(2x - 8) + ln(x^2 + 4)$$

$$\frac{dy}{dx} = \frac{f'(x)}{f(x)}$$

$$\frac{dy}{dx} = \frac{2}{2x - 8} + \frac{2x}{x^2 + 4}$$

Integration of Logarithmic Functions

At a fundamental level, for the **integration of logarithms**, we are simply **reversing** the **differentiation processes** we have just learnt. For instance, we have just learnt that the **derivative** of $ln(x) = \frac{1}{x}$. Therefore, we can work **backwards** and find that the **integral** of $\frac{1}{x}$ must be $ln(x)$ **plus a constant**.

To start, the **first integration rule** states that the **integral** of $\frac{1}{x}$ is simply $ln(x) + c$.

$$\int \frac{1}{x} dx = ln(x) + C \quad \text{for } x > 0$$

Applying this to the examples such as $f'(x) = \frac{5}{x}$ and $\frac{dy}{dx} = \frac{-3}{2x}$, the **integral** will be determined as follows:

$$f(x) = \int \frac{5}{x} dx$$

① $f(x) = 5\int \frac{1}{x} dx$

② $f(x) = 5ln(x) + C$

$$y = \int \frac{-3}{2x} dx$$

① $y = -\frac{3}{2}\int \frac{1}{x} dx$

② $y = -\frac{3}{2} ln(x) + C$

① **Apply** $\int \frac{1}{x} dx = ln(x) + C$

② **State** the **final answer**

However, what if the **function** we are trying to **integrate** isn't of the form $y = \frac{1}{x}$? This is where we will need to use our **second integration rule**.

The **second integration rule** is for exponentials of the form: $\frac{f'(x)}{f(x)}$. It states that if the **numerator** is **equal to the derivative** of the **denominator**, then the **integral** will be: $ln(f(x)) + C$. This can be summarised as:

$$\int \frac{f'(x)}{f(x)} dx = ln(f(x)) + C$$

Applying this formula to some examples, let's suppose that we had the two integrals: $y = \int \frac{4x}{2x^2+2} dx$ and $f(x) = \int \frac{10x-2}{5x^2-2x} dx$. Applying our rule above, we would **integrate** them as follows:

$$y = \int \frac{4x}{2x^2 + 2} dx$$

① $\int \frac{f'(x)}{f(x)} dx = ln(f(x)) + C$

② $y = ln(2x^2 + 2) + C$

$$f(x) = \int \frac{10x - 2}{5x^2 - 2x} dx$$

① $\int \frac{f'(x)}{f(x)} dx = ln(f(x)) + C$

② $f(x) = ln(5x^2 - 2x) + C$

① **Apply our second integration rule**

② **State** the **final answer**

In the examples above, the **numerator** was **exactly** the **derivative** of the **denominator**, so we could use the **second integration rule** straight away. However, what do we do when the **function doesn't fit** the exact format of $\frac{f'(x)}{f(x)}$? In those cases, we must **factor out** a **co-efficient** to ensure that the **numerator** is the **derivative** of the **denominator**.

For instance, let's consider $y = \int \frac{9}{3x-2}dx$. As we can see, the **numerator** is **not the exact derivative** of the **denominator**. Therefore, we must **factor** out a **co-efficient** of **3** so that the integral will be in the form $\frac{f'(x)}{f(x)}$. Once in the correct form, we will be able to integrate it as follows:

$$y = \int \frac{9}{3x-2}dx$$

① $y = 3\int \frac{3}{3x-2}dx$

② $\int \frac{f'(x)}{f(x)}dx = ln(f(x)) + C$

③ $y = 3ln(3x-2) + C$

① **Factor out the co-efficient from the numerator**

② **Apply** $\int \frac{f'(x)}{f(x)}dx = ln(f(x)) + C$

③ **State** the **final answer**

As a second example let's consider $f(x) = \int \frac{8x}{2x^2-9}dx$. Here we can see that we must **factor out** a **co-efficient** of **2** to make the **integral** in the form $\frac{f'(x)}{f(x)}$. From there, we can integrate it as follows:

$$f(x) = \int \frac{8x}{2x^2-9}dx$$

① $f(x) = 2\int \frac{4x}{2x^2-9}dx$

② $\int \frac{f'(x)}{f(x)}dx = ln(f(x)) + C$

③ $f(x) = 2ln(2x^2-9) + C$

① **Factor out the co-efficient from the numerator**

② **Apply** $\int \frac{f'(x)}{f(x)}dx = ln(f(x)) + C$

③ **State** the **final answer**

To consolidate your knowledge of these **integration rules**, practice using the **worked example below**.

Worked Example 1

Continuing to explore her **foundational understanding** of **logarithmic functions**, help Kerry to **determine** the following **integrals**.

(i) $f(x) = \int \frac{-2}{x}dx$

$f(x) = -2\int \frac{1}{x}dx$

$f(x) = -2ln(x) + C$

(ii) $y = \int \frac{-10x}{-5x^2+4}dx$

Apply: $\int \frac{f'(x)}{f(x)}dx = ln(f(x)) + C$

$y = ln(-5x^2+4) + C$

(iii) $f(x) = \int \frac{8x}{8x^2+4}dx$

$f(x) = \frac{1}{2}\int \frac{16x}{8x^2+4}dx$

Apply: $\int \frac{f'(x)}{f(x)}dx = ln(f(x)) + C$

$f(x) = \frac{1}{2}ln(8x^2+4) + C$

Now that we've learned how to **integrate indefinite integrals** containing **natural logarithms**, we can explore the **applications** of differentiation and integration. As shown in the **worked examples below**, we can determine **definite integrals of logarithms**, the **area between curves**, **rectilinear motion** and many other applications.

Worked Example 2

Pushing her knowledge of integration, Kerry wants to use **definite integrals** to determine the **area bound** between $x = 2$ and $x = 6$ for $f(x) = \frac{4x}{2x^2+7}$. Help Kerry to determine this **area**. **[CA]**

$$A = \int_2^6 \frac{4x}{2x^2+7}dx$$

$$A = \left[ln(2x^2+7)\right]_2^6$$

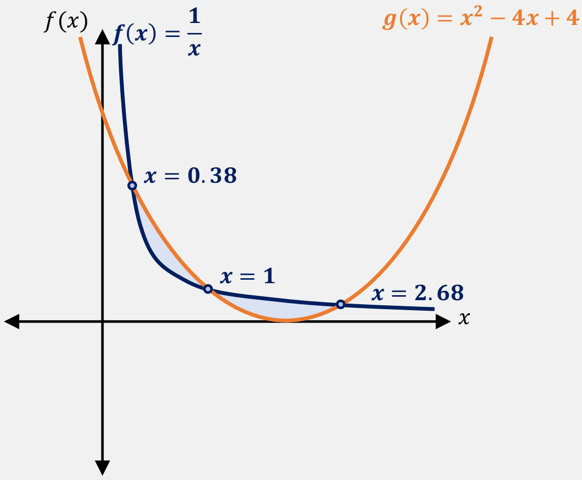
$$A = \left[ln(2(6)^2+7)\right] - \left[ln(2(2)^2+7)\right]$$

$$A = ln(79) - ln(15)$$

$$A \approx 1.66$$

Worked Example 3

Continuing to explore her interest for the **integration** of **logarithmic functions**, Kerry wants to determine the **area trapped** between the two functions: $f(x) = \frac{1}{x}$ and $g(x) = x^2 - 4x + 4$. **[CA]**



Help Kerry to **determine** the **area** trapped between $f(x)$ and $g(x)$.

$$A = \int_{-1}^{0} f(x) - g(x)dx$$

$$A = \int_{0.38}^{1} x^2 - 4x + 4 - \frac{1}{x}\,dx + \int_{1}^{2.68} \frac{1}{x} - x^2 + 4x - 4\,dx$$

$$A = \left[x^2 - 4x + 4 - \frac{1}{x}\right]_{0.38}^{1} + \left[\frac{1}{x} - x^2 + 4x - 4\right]_{1}^{2.68}$$

$$A = 0.1163 + 0.5477$$

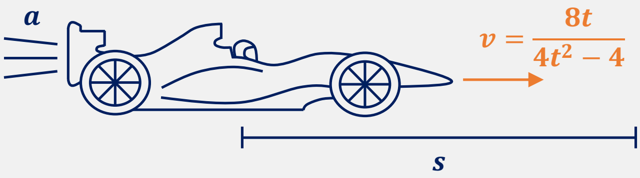
$$A = 0.664\ units^2$$

Note, **with a calculator** you could instead do:

$$A = \int_{0.38}^{2.68} \left|\frac{1}{x} - x^2 + 4x - 4\right|\,dx = 0.664\ units^2$$

Worked Example 4

Kerry has built herself a race car that has the **velocity function**: $v = \frac{8t}{4t^2-4}$ where v is in m/s and t is **time** in **seconds**. Before driving this race car, Kerry wants to do some calculations. **[CA]**



(a) Help Kerry to **determine** the **velocity** of the car after **2 seconds**.

$$v(t) = \frac{8t}{4t^2 - 4}$$
$$v(2) = \frac{8(2)}{4(2)^2 - 4}$$
$$v(2) = 1.33\ m/s$$

(b) If Kerry knows the race car has an **initial displacement** of $4m$, determine $s(t)$.

$$s(t) = \int v(t)dt$$
$$s(t) = \int \frac{8t}{4t^2 - 4}dt$$
Apply: $\int \frac{f'(x)}{f(x)}dx = ln(f(x)) + C$
$$s(t) = ln(4t^2 - 4) + 4$$

(c) Determine the **net change** in **displacement** of the **car** between the 2^{nd} **and** 8^{th} **second**.

Displacement
$$s(b) - s(a) = \int_a^b v(t)dt$$
$$s(8) - s(2) = \int_2^8 \frac{8t}{4t^2 - 4}dt$$
$$s(8) - s(2) = 3.04m$$

LOGARITHMIC FUNCTIONS TOPIC NOTES

Logarithmic Functions

Logarithms are the **inverse operation** of an **exponential**. Just as addition and subtraction are inverses of each other, **logarithms "undo" exponentiation**.

$$x = \log_b y \Longleftrightarrow y = b^x$$

Log Laws

Each **log law** and its **corresponding exponential law** is summarised below:

Log laws	Index Laws
$\log_b(mn) = \log_b m + \log_b n$	$b^m \times b^n = b^{m+n}$
$\log_b\left(\frac{m}{n}\right) = \log_b m - \log_b n$	$\frac{b^m}{b^n} = b^{m-n}$
$\log_b(m^n) = n\log_b m$	$(b^m)^n = b^{mn}$
$\log_b(b^n) = n$	$b^{\log_b n} = n$
$\log_b(1) = 0$	$b^0 = 1$
$\log_b(b) = 1$	$a^1 = a$

Natural Logarithms

The **natural logarithm** is a **logarithm** with the **base 'e'**, where **'e'** is **Euler's number**, an irrational number approximately equal to 2.71828. The **natural logarithm function** is defined as: $ln(x) = \log_e(x)$

Similar to other logarithms, **natural logarithms** have a **special relationship** with **exponential functions**. This relationship is **expressed as**:

$$x = ln(b) \Longleftrightarrow b = e^x$$

Change of Base

The **change of base formula** allows us to **convert a logarithm** with **one base** to a **logarithm with a different base**. The **change of base formula** is:

Change of Base: $\log_b x = \frac{\log_c x}{\log_c b}$

Differentiation of Logarithmic Functions

For the natural logarithms $ln(x)$, the **derivative** with respect to x is $\frac{1}{x}$, as shown in the formula below:

If $y = ln(x)$, then $\frac{dy}{dx} = \frac{1}{x}$

The **second rule** for the **differentiation of logarithms** is **derived** from the **chain rule**. It states that if a **function** $f(x)$ is **contained within the logarithm** (i.e. $ln(f(x))$), then the **derivative** is $f'(x)$ divided by $f(x)$. This rule can be stated as follows:

Derived from Chain Rule: If $y = ln(f(x))$ then $\frac{dy}{dx} = \frac{f'(x)}{f(x)}$

Integration of Logarithmic Functions

$$\int \frac{1}{x} dx = ln(x) + C \quad \text{for } x > 0$$

$$\int \frac{f'(x)}{f(x)} dx = ln(f(x)) + C$$

Calculus of Logarithmic Functions with bases other than e

When we explore the **differentiation** and **integration** of **logarithmic functions**, it is important to note that we are only **applying calculus** to **natural logarithms** such as $ln(x)$. In the cases that the logarithm is **not** in the natural logarithm format such as $log_{10}(x)$, we apply our **change of base formula**.

Problem Set 10 – Calculus of Logarithms
Progressive Questions

Concept 1
Natural Logarithms – Progressive Questions

(7 questions)

Repetitive questions: 1.11 – 1.92

The International Mathematics Championship!

Starring: Teacher Andrew, Janitor Peter, the Methods Class

Natural Logarithmic Functions: Q1, Q2, Q3, Q4, Q5, Q6

Repetitive: 2.11-2.81

[10 marks]

1. As the team approaches the venue for the competition, Tom wants to warm up on **natural logarithms** but has forgotten his calculator! Help Tom find the **value** of the following expressions **without** a calculator:

(a) $\ln\left(\frac{1}{e}\right)$ **(1)**

(b) $\ln(e^3)$ **(1)**

(c) $\ln(1) + \ln(4e) - \ln(4)$ **(2)**

(d) $\ln(e) + \ln(\sqrt{e})$ **(2)**

(e) $\log_{\frac{1}{e}}(e)$ **(2)**

(f) $\ln(\ln(e^e))$ **(2)**

[CA] [21 marks]

2. Tom finds his calculator was in his pocket all along. **Using a calculator**, solve for x in the following equations, rounding to **3 significant figures** where necessary:

(a) $e^x = 2$ **(1)**

(b) $3e^{5x} = 300$ **(2)**

(c) $10 - 4e^{x^2} = 2$ **(3)**

(d) $1000e^{-0.5x} - 500 = 0$ **(2)**

(e) $e^{2x} - 5e^x + 6 = 0$ **(4)**

(f) $e^{2x} + e^x - 6 = 0$ **(5)**

(g) $2^x = 3$ **(2)**

(h) $5^{x+1} = 3$ **(2)**

[7 marks]

3. Tom's calculator has now run out of power! Annoyed by Tom, Jevon wants to prove he doesn't need the help of a calculator. Help Jevon **simplify** these **expressions**:

(a) $e^{\ln(x)+1}$ **(1)**

(b) $\ln\left(\sqrt{\frac{x+1}{e^4}}\right)$ **(2)**

(c) $\frac{1}{\ln\left(e^{(x^2)}\right)}$ **(1)**

(d) $e^{2\ln(x)} - \sqrt{e}^{4\ln(x)}$ **(3)**

[10 marks]

4. The competition has officially began, however, another team is peeking at the answers! As quickly as you can before they are able to copy our answers, **express and simplify** x in terms of **natural logarithms**:

(a) $e^{3x} = 12$ **(1)**

(b) $3e^{\frac{x}{4}} = 72e$ **(1)**

(c) $6e^{2x} = 156e^{3x}$ **(1)**

(d) $\ln(e^{2x} + 1) = \ln(33)$ **(1)**

(a) $e^x + 16 + e^{(x+1)} = 30$ **(1)**

(b) $2^{e^{2x}} \times 4^{e^x} = 16^{e^{2x}}$ **(2)**

(c) $e^{2x} - e^x = 6$ **(2)**

(d) $8\left(\frac{e^{2x-1}}{e^{x-2}}\right) - 3 = 14$ **(2)**

[16 marks]

5. The final round before a break is a power round! Help your team **solve** the following for $\boldsymbol{x}$ in terms of $\boldsymbol{ln}$.

(a) $\mathrm{e}^{(2x-6)} = 4$ **(2)**

(b) $e^{x-2} = 3$ **(2)**

(c) $4e^{2x-1} = 600$ **(3)**

(d) $6 = 4e^{3x} - 2$ **(3)**

(e) $3e^{x+5} = 300$ **(3)**

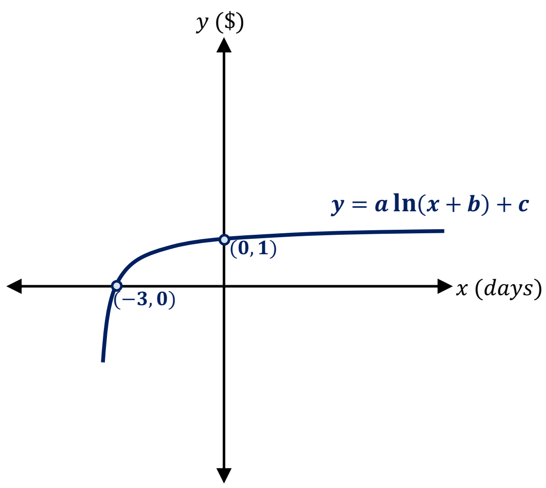
(f) $4e^{4x-2} - 12 = 0$ **(3)**

[12 marks]

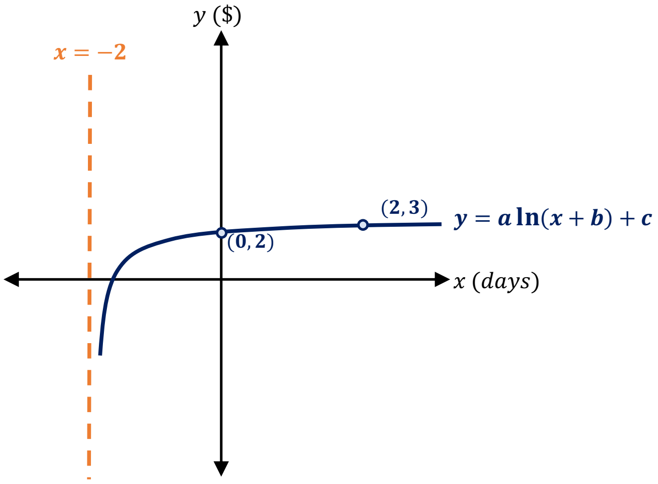
6. The team heads outside to buy some lemonade from the volunteers selling them in front of the school. Curious on why the price of lemonade is so expensive, the volunteers shows your team 3 **graphs** representing **price change trends** of the lemonade over time for each season. **Identify** the **equation** of each of these graphs by finding $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$.

Note: for clarity, the full graphs including those stretching into negative quadrants are shown.

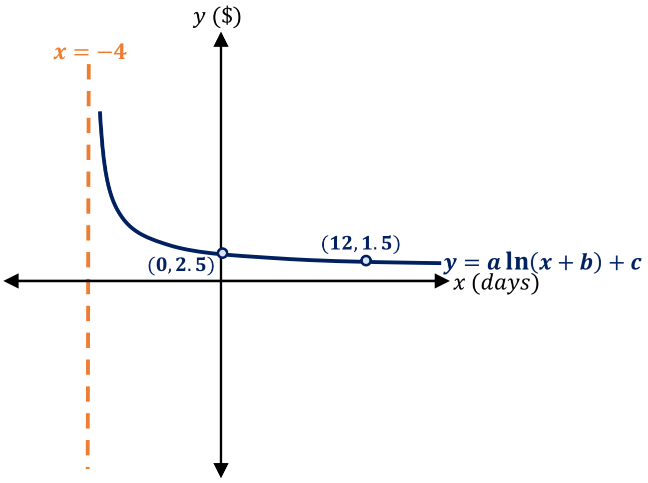
(a) Autumn: general rule $\boldsymbol{y = a\ln(x + b) + c}$ **(4)**



(b) Summer: general rule $\boldsymbol{y = a\ln(x+b) + c}$ with a **vertical asymptote** at $\boldsymbol{x = -2}$. **(4)**



(c) Winter: general rule $\boldsymbol{y = a\ln(x+b) + c}$ with a **vertical asymptote** at $\boldsymbol{x = -4}$ **(4)**



Convinced by the volunteers that their lemonade price is reasonable, the team enjoy a short break drinking lemonade. They have half an hour before they start the differentiation part of the competition and the students are excited to prove what they know!

Concept 2

Calculus of the Natural Logarithmic Function – Progressive Questions

(15 questions)

Repetitive questions: 2.1 – 2.115

The Mathematics World Championships Continued…

Starring: Teacher Andrew, Rabea, Tyler, Tom, Rupert and Wallace

---

Coming into round two of the competition, the students are required to complete their own set of questions! Nervous, but excited by the chance to show their maths skills, help the students with their question sets.

Calculus of the Logarithmic Function: Q1, Q2, Q3, Q4, Q5

Repetitive: 1.11 → 1.31 (3 questions)

[20 marks]

1. Up first is confident and patient Alexa. **Help** Alexa **differentiate** the following **logarithmic functions**!

(a) $f(x) = \ln(x)$ **(2)**

(b) $y = 2x - 3\ln(x)$ **(2)**

(c) $y = 4e^x + \ln(4x)$ **(2)**

(d) $f(x) = \ln(5x - 2)$ **(2)**

(e) $f(x) = \ln(e) + \ln(-4x)$ **(2)**

(f) $f(x) = \ln(\sin x)$ **(2)**

(g) $y = 4x^3 - \ln(2x^2 - 2)$ **(2)**

(h) $f(x) = \ln(x^3 + 2x)$ **(2)**

(i) $y = \ln(\cos x)$ **(2)**

(j) $f(x) = \frac{1}{\ln(x)}$ **(2)**

[28 marks]

2. Tom has lost track of time and needs your help! **Help** Tom **differentiate** the following **logarithmic functions**!

(a) $y = \ln(2x^2 - 5x) + x^3$ **(2)**

(b) $f(x) = \ln\left(\frac{x+1}{2x-4}\right)$ **(3)**

(c) $y = \ln(\sqrt{x})$ **(2)**

(d) $y = \ln[(x-2)(x^2+2)]$ **(3)**

(e) $f(x) = \ln\left[\frac{(x-2)^3}{x+3}\right]$ **(3)**

(f) $f(x) = \ln(\frac{\sin x}{\cos x})$ **(3)**

(g) $y = \log_e[(x^3 - 2)^5]$ **(3)**

(h) $f(x) = \ln[\sqrt{x}(x^2 - 4)]$ **(3)**

(i) $y = \ln[(x^2 - 4)(x + 2)]$ **(3)**

(j) $f(x) = x^3 - \ln\left[\frac{x(x^2+4)}{x-9}\right]$ **(3)**

[35 marks]

3. Jevon decided he wanted to challenge himself with the product and quotient rule. **Help** Jevon to find the **derivatives** of the following functions using the **product rule** and **quotient rule**.

(a) $f(x) = x^2\ln(2x)$ **(3)**

(b) $y = \frac{\ln(4x)}{3x}$ **(3)**

(c) $f(x) = \frac{1}{x}\ln(3x+2)$ **(3)**

(d) $f(x) = \frac{\ln\left(\frac{1}{x}\right)}{2e^x}$ **(3)**

(e) $f(x) = 2x^2\ln(3x+2)$ **(3)**

(f) $f(x) = \frac{\ln(5x^2-3)}{x-3}$ **(4)**

(g) $y = \ln\left[\frac{x-2}{4x^2-3x}\right]$ **(4)**

(h) $y = 4x^3\ln[x^2-3x]$ **(4)**

(i) $f(x) = \frac{e^{5x^2}}{\ln(x^2-4x)}$ **(4)**

(j) $y = (1+x)^3\ln(x)$ **(4)**

[15 marks]

4. Rabea got distracted and has forgotten how to **differentiate natural logarithms**. Help Rabea **differentiate** the following **natural logarithms** in terms of x :

(a) $y = \ln(1-2x^2)$ **(1)**

(b) $y = \ln(x^5)$ **(1)**

(c) $y = \ln(3e^{3x}+3)$ **(1)**

(d) $y = \ln(4+3x)^4$ **(1)**

(e) $y = \ln\left(\sqrt{x^4-4x^2}\right)$ **(2)**

(f) $y = \ln((x^2-1)\cdot 2x+1))$ **(1)**

(g) $f(x) = \ln 8x^4 \cdot 4x^3$ **(2)**

(h) $f(x) = \ln(\sqrt{2x^3-6x})$ **(2)**

(i) $y = \ln\frac{x^3}{7x}$ **(2)**

(j) $f(x) = \ln\left(\frac{3x^3}{3x^4}\right)$ **(2)**

[CA] [12 marks]

5. Always wanting to prove himself, Rupert chose to answer the gradient questions. **Help** Rupert to determine the **gradient** of various logarithmic functions at the **following values**.

(a) Determine the **gradient** of $\boldsymbol{f(x) = 3\ln(x)}$ at the **point** $(\mathbf{1,0})$ **(2)**

(b) Determine the **gradient** of $\boldsymbol{f(x) = 4x\ln(x)}$ at the **point** $(\boldsymbol{e, 4e})$ **(2)**

(c) Determine the **gradient** of $\boldsymbol{y = \ln(x^3)}$ at the **point** $(\boldsymbol{e, 3})$ **(2)**

(d) Determine the **gradient** of $\boldsymbol{y = \log_2(x)}$ at $(\mathbf{1,0})$ using the **change of base formula** **(3)**

(e) Determine the **gradient** of $\boldsymbol{y = \log_8(x)}$ at $(\mathbf{1,0})$ using the **change of base formula** **(3)**

Integration of Logarithmic Functions: Q6, Q7, Q8, Q9, Q10

Repetitive: 1.51 → 1.81 (4 questions)

[18 marks]

6. Now working as a team again, the students must complete this set of questions as soon as possible! **Help** the students **integrate** the **following functions**.

(a) $\int \frac{3}{x}\, dx$ **(1)**

(b) $\int -\frac{2}{x} dx$ **(1)**

(c) $\int \frac{2x}{x^2+2}\, dx$ **(1)**

(d) $\int \frac{-8}{7+2x} dx$ **(2)**

(e) $\int \frac{2(x+2)^2}{x}\, dx$ **(3)**

(f) $\int \left(1+\frac{1}{x}\right)^2 dx$ **(2)**

(g) $\int \frac{x^{-2}+1}{x} dx$ **(2)**

(h) $\int \frac{x^{-\frac{1}{2}}}{2\sqrt{x}} dx$ **(2)**

(i) $\int \frac{2x-5}{x^2-5x+6}\, dx$ **(2)**

(j) $\int \frac{2x}{(x-1)(x+1)} dx$ **(2)**

[19 marks]

7. The students were not the fastest and are placed in a knock-out round to prove themselves! **Help** the team to **integrate** the following functions.

(a) $\int \frac{\sin(x)}{\cos(x)}\, dx$ **(1)**

(b) $\int \frac{-\sin(x)}{\cos(x)} dx$ **(1)**

(c) $\int \frac{-\cos(x)}{\sin(x)}\, dx$ **(1)**

(d) $\int \tan(x) dx$ **(2)**

(e) $\int \frac{\sin(x)+\cos(x)}{\sin(x)-\cos(x)} dx$ **()**

(f) $\int \frac{\sin 2x}{\sin^2(x)} dx$ **(3)**

(g) $\int \frac{1-\sin^2(x)}{\sin(x)\cos(x)} dx$ **(3)**

(h) $\int \frac{1}{x\ln x} dx$ **(2)**

(i) $\int \frac{2}{x\ln(x^2)} dx$ **(3)**

(j) $\int \frac{4x+3}{(2x^2+3x+4)\ln(2x^2+3x+4)} dx$ **(2)**

[31 marks]

8. In the second part of the knockout round, each student must answer one question individually. **Help** the group to **solve** the following **definite integrals.**

(a) $\int_1^2 \frac{4x+1}{2x^2+x+2} dx$ **(3)**

(b) $\int_5^{10} \frac{8x+4}{x^2+x+1} dx$ **(3)**

(c) $\int_{-6}^{4} \frac{12x-8}{3x^2-4x+5} dx$ **(3)**

(d) $\int_5^7 \frac{2x^3+x}{x^4+x^2+5} dx$ **(3)**

(e) $\int_2^4 \frac{x^2+2x+1}{x^3+3x^2+3x+1} dx$ **(3)**

(f) $\int_{-1}^{1} \frac{45x^4-24x-18x^2}{2+4x^2-3x^5+2x^3} dx$ **(3)**

(g) $\int_{\pi/6}^{\pi/3} \tan x\, dx$ **(3)**

(h) $\int_3^5 \frac{1}{x\ln x} dx$ **(3)**

(i) $\int_{\sqrt{e}}^{\sqrt{e^e}} \frac{1}{x\ln(x^2)} dx$ **(3)**

(j) $\int_{\pi/4}^{\pi/3} \frac{\sin(x)\left(\cos^2\left(\frac{1}{2}x\right)-1\right)^2}{\cos(x)(1-\cos^2 x)} dx$ **(4)**

[CA] [15 marks]

9. The team has certainly proved themselves. Every question this round is worth double the number of points as the last round! This round involves **using integration** to **determine $f(x)$**. **Help** the students to do **determine $f(x)$** for each of the following scenarios.

(a) Determine $f(x)$ if $f'(x) = \frac{1}{x}$ and $f(x)$ intersects the point $(1, 1)$ **(3)**

(b) Determine $f(x)$ if $f'(x) = 6e^{x}$ and $f(x)$ intersects the point $(1, 6.38)$ **(3)**

(c) Determine $f(x)$ if $f'(x) = \frac{2x}{x^2}$ and $f(x)$ intersects the point at $(0.18, 3)$ **(3)**

(d) Determine $f(x)$ if $f'(x) = -\frac{2}{x}$ and there is a **root** in $f(x)$ at $x = 0.61$ **(3)**

(e) Determine $f(x)$ if $f'(x) = \frac{3x^2}{2x}$ and there is a **root** in $f(x)$ at $x = 7.39$ **(3)**

[13 marks]

10. It is the final round of the competition! Excitement fills the air as the last question is announced…

(a) **Integrate** the following function and give the answer **in the form of only one natural log function**. There is **no need to simplify**. Upon inspection, Rupert notices both parts of the function will **need modification** before it can be worked with. **(5)**

$$\int \frac{12x + 21}{2x^2 + 7x + 12} + \frac{7}{2x}\,dx$$

The time has come for the winner of the competition to be announced, but the Methods Team is greeted strangely with another question. Turns out, each contestant was given a number, and if their number matched the answer to this question, they are the winner.

(b) Though annoyed at the roundabout way of announcement, the students were still keen to know if they had won, so help them **evaluate** and give in the **simplest form**. **(6)**

$$\int_{-4}^{-2} \frac{2x + 5}{x^2 + 5x + 7}\,dx - \int_{2-2\sqrt{2}}^{2-\sqrt{2}} \frac{4x - 8}{x^2 - 4x + 5}\,dx$$

(c) Without a calculator, the students cannot evaluate this final answer. However, they notice they were the only one in the competition to be given a **negative number**. Did the Methods Students **win** the competition? **(2)**

Against all odds, the wonderful methods team has won the International Mathematics Championship! Janitor Peter and Teacher Andrew are so proud of their students and are excited to spend a few more days exploring Peru, but first is their celebrations for their victory!

General Applications of Logarithmic Functions: Q11, Q12, Q13, Q14, Q15

[CA] [10 marks]

11. To prepare for the party, Rabea decided to bake cookies for all the students to share! The oven is a different model and is extremely quick to preheat so Rabea doesn't want to turn it on too early. They know that the oven will have a **temperature (in ℃)** that follows the equation: $\boldsymbol{T = 24e^{0.06t}}$, where $\boldsymbol{t}$ is **time** in **seconds**.

(a) What is the **temperature** of the oven **before turning it on**? **(1)**

(b) **How hot** will the oven be **one second** after it's turned on? **(2)**

(c) The cookies bake at $\boldsymbol{200}$℃. At **what time** will it reach this **critical temperature**? **(3)**

(d) There is **36 seconds** until the brownies need to be put into the oven. Will it reach the **target temperature?** If not, how many **seconds short** will they be? **(1)**

(e) The class found a **different oven**. This oven can withstand up to $\boldsymbol{350}$℃ and will have a **temperature (in ℃)** that follows the equation: $\boldsymbol{T = 24e^{0.07t}}$. At **what time** will it reach this **critical temperature**? Is this oven a better choice? **(3)**

[CA] [12 marks]

12. Once the cookies were baked, Rabea was curious how quickly they would be eaten. He found that the **number of trays of cookies** left at any time ***t*** was given by the formula $\boldsymbol{A = A_0 e^{-0.05t}}$, where $\boldsymbol{A_0}$ is the number of trays of brownies Rabea baked and $\boldsymbol{A}$ is the number of trays of brownies left after $\boldsymbol{t}$ **minutes**.

(a) Rabea baked **10** trays of cookies. If he measured the number of trays left **12 minutes later**, **how many trays** did he find were left? **(2)**

(b) After those **initial 12 minutes**, Rabea began to get hungry. He ate **1 tray of cookies** at a time, and each tray took **two minutes** to eat. If he ate a **whole number of trays**, and considering that the cookies were still being eaten by the other students, **how many trays** does he eat and **what fraction of a tray** (as a decimal) is left once he's finished eating? *[HINT: your answer from **part a** is needed!]* **(2)**

(c) Once there was a **tenth of a tray** left, people stopped eating the cookies. **How long** did it take for this to happen? *[HINT: your answer from **part b** is needed!]* **(4)**

(d) Rabea wants to calculate the "**half-life**" of the cookies, which is the **amount of time** it takes for **half of any number of trays** to be eaten by the students (ignoring Rabea in part b). What is the half-life in **minutes**? **(4)**

[CA] [5 marks]

13. Rupert decides to inflate some balloons for the party. After he is finished, he notices that two of them have different sizes. One has a **radius of 164.5 mm** and the other has a **radius of 165 mm**. Rupert wants to inflate one of the balloons to make it the same size as the other. The **radius** of the **smaller balloon** (in **mm**) as it inflates is given by the function $\boldsymbol{r_1 = 0.5\ln(t+e) + 164}$, where $\boldsymbol{t}$ is time in **minutes**.

(a) **How long** will it take for the **smaller balloon** to be the **same size** as the **larger one**? **(2)**

(b) Rupert realises halfway through inflating the balloon that Alexa has stolen the other balloon and is inflating it as well. The **radius** of the **larger balloon** (in **mm**) while it is being inflated is given by the function $\boldsymbol{r_2 = 0.25\ln(t+e) + 164.75}$**,** where $\boldsymbol{t}$ is time in **minutes**. **After how many minutes** will they be the **same size** and **what size** will they be? **(3)**

[CA] [12 marks]

14. The students decide to bake some cakes and want to ice the tops of the cakes. Their shapes can be given by **various functions**. To determine the **amount of icing they need**, help the students **calculate** the **following areas**:

(a) Calculate the **area under the curve** $f(x) = \frac{3x}{x^2}$ between $x = 1$ and $x = 3$ **(3)**

(b) Calculate the **area under the curve** $f(x) = \frac{4x-4}{x^2}$ between $x = 2$ and $x = 3$ **(3)**

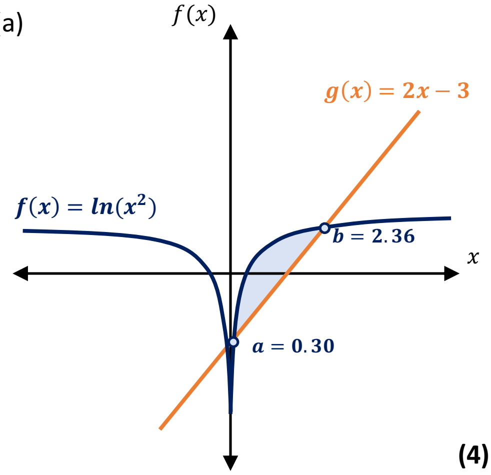
(c) Calculate the **area under the curve** $f(x) = \frac{8x^2-7x}{7x^3}$ from $x = 1$ to $x = 5$ **(3)**

(d) Calculate the **trapped between the x-axis** and $f(x) = \ln(2x)$ from $x = 1$ to $x = 3$ **(3)**

[CA] [18 marks]

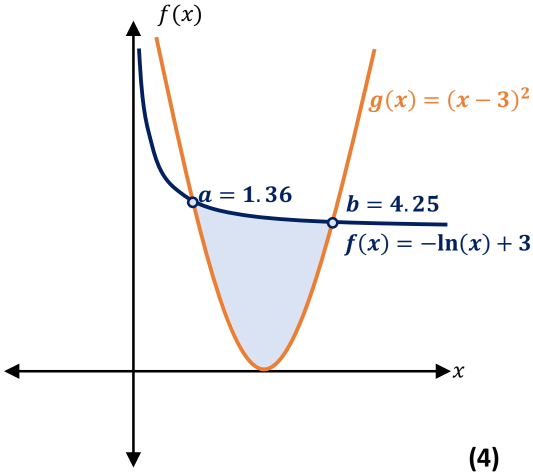
15. Some of the tops of the cakes are in the rather unusual shape of the areas created between exponential functions and other functions. Help the students to **determine** the **shaded areas below** using **definite integrals**.

(a)



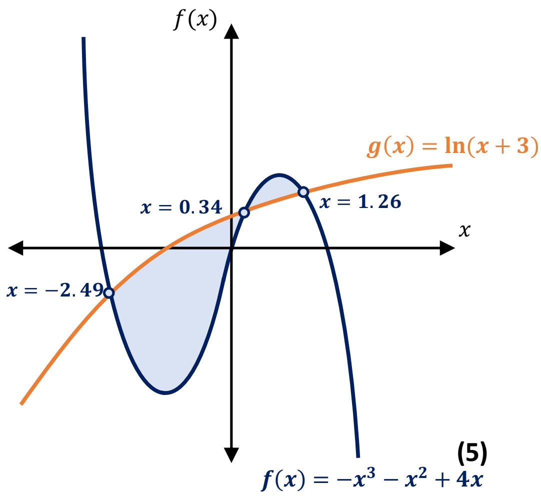
(4)

(b)



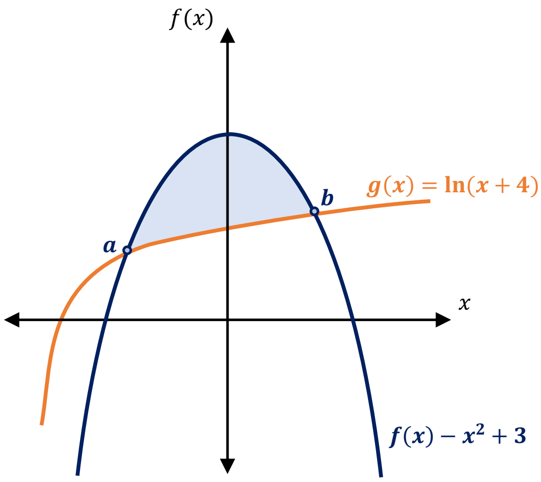
(4)

(c)



(5)

(d)



(5)

Problem Set 10 – Calculus of Logarithms
Repetitive Questions

Concept 1
Natural Logarithms – Repetitive Questions
(6 questions)

Natural Logarithms: Questions

[15 marks]

1.11 Find the **value** of the following expressions **without** a calculator.

(a) $\ln(e^2)$ **(1)**

(b) $\ln(2e) - \ln(2)$ **(2)**

(c) $\ln(e^2) + \ln(\sqrt{e})$ **(2)**

(d) $\ln\left(\frac{2}{e}\right)$ **(2)**

(e) $\ln(6e) - \ln(1)$ **(2)**

(f) $\ln\left(\frac{e^4}{e^2}\right)$ **(2)**

(g) $\ln(e^5) - \ln(e^2)$ **(2)**

(h) $\ln\left(\frac{4}{e}\right) - 5\ln(\sqrt{e})$ **(2)**

[CA][12 marks]

1.21 **Using a calculator**, solve for x in the following equations, rounding to 3 significant figures where necessary.

(a) $e^{-4x} = 12$ **(2)**

(b) $e^{6x} = 15$ **(2)**

(c) $6e^{7x} = 18$ **(2)**

(d) $2e^{-9x} = 6$ **(2)**

(e) $100e^{x^2} = 500$ **(2)**

(f) $4e^{x^2} = 16$ **(2)**

[8 marks]

1.31 Simplify the following **expressions**:

(a) $3\ln(e^{-x})$ **(2)**

(b) $4x\ln(e^{2x}) + 4\ln(e^x)$ **(2)**

(c) $\ln e^{x^2+2x+1}$ **(2)**

(d) $6\log_e(e^{-x}) + \ln(e^{4x-9})$ **(2)**

[15 marks]

1.41 **Express and simplify** x in terms of natural logarithms:

(a) $e^{6x} = 18$ **(1)**

(b) $e^{2x} = 3e$ **(2)**

(c) $4e^{3x} = 2e^{2x}$ **(2)**

(d) $5e^{x^2} = 3e^x$ **(2)**

(e) $e^{2x} - 3e^x + 15e^{2x} = 0$ **(2)**

(f) $-5e^{2x} + e^{3x} + 8e^{3x} = 0$ **(2)**

(g) $\frac{e^x}{e^{3x+2}} - e = 9$ **(2)**

(h) $\frac{2e^{3x}}{2e^{8x+4}} - e = 14$ **(2)**

[12 marks]

1.51 Solve the following for x in terms of ln.

(a) $e^{3x} = 12$ **(1)**

(b) $e^{4x+3} = 9$ **(2)**

(c) $5e^{4x} = 30$ **(2)**

(d) $16 = x^2$ **(2)**

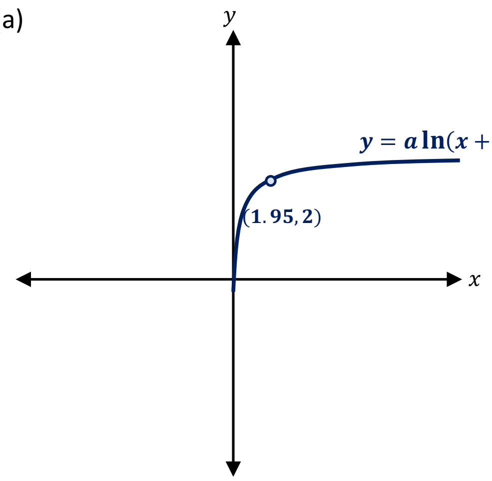
(e) $4e^{x^2} = 12$ **(2)**

(f) $9e^{x^{0.5}} - 18 = 0$ **(3)**

[CA][15 marks]

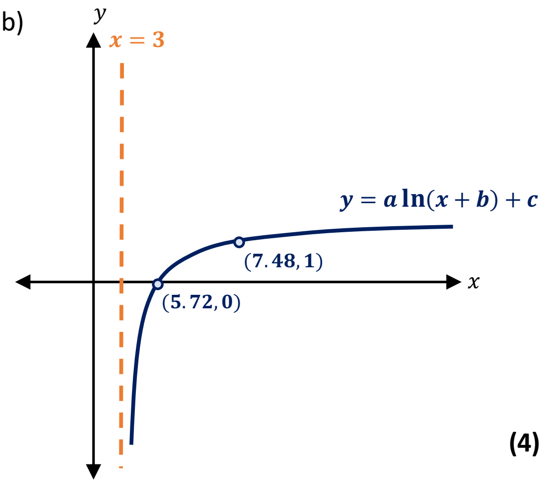
1.61 Identify the **equation** of each of these graphs by finding a, b and c. Graphs including those stretching into negative quadrants are shown.

(a)



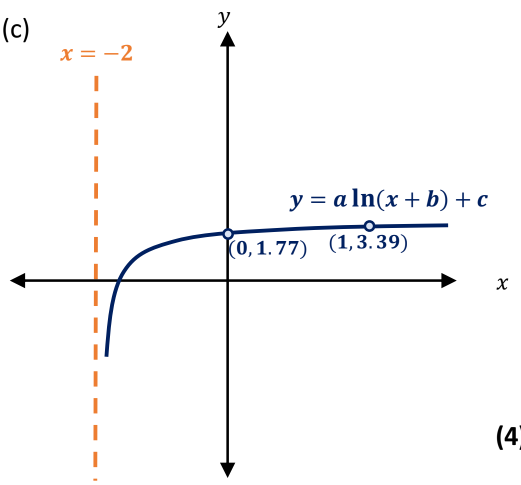
(3)

(b)



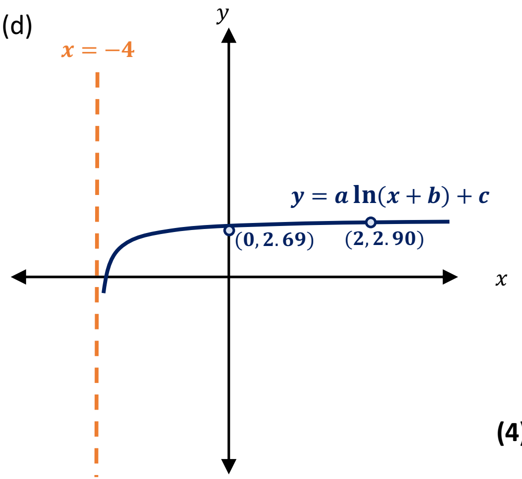
(4)

(c)



(4)

(d)



(4)

Concept 2

Calculus of Natural Logarithms – Repetitive Questions

(16 questions)

Differentiation of the Logarithmic Function:

Repetitive: 2.1 → 2.15 (16 questions)

[11 marks]

2.1 Differentiate the following **natural logarithms** in terms of x:

(a) $y = \ln(x^5) - 4$ **(1)**

(b) $y = \ln(2x^3 + 1)$ **(2)**

(c) $y = \ln(1 - x^2)$ **(2)**

(d) $y = \ln(x^3 + x)$ **(2)**

(e) $y = \ln(3x^2 - 4x)$ **(2)**

(f) $y = \ln(x^3) - 2x^5$ **(2)**

[16 marks]

2.2 Differentiate the following **logarithmic functions**.

(a) $y = \ln(x^4 - 2x) - 4x^3$ **(2)**

(b) $f(x) = \frac{2x^3}{\ln(x^2)}$ **(3)**

(c) $y = \ln(3\sqrt{x})$ **(2)**

(d) $f(x) = \frac{\ln 6x^4}{3x^2}$ **(3)**

(e) $y = \log_e[(x^2 - 3)^3]$ **(3)**

(f) $f(x) = \ln[x^2(x^4 + 4)]$ **(3)**

[18 marks]

2.3 Find the **derivatives** of the following functions using the **product rule** and **quotient rule**, **simplifying** where reasonable.

(a) $f(x) = 3x^2\ln(4x - 2)$ **(3)**

(b) $y = \ln(3x + 2)\ln(2x)$ **(3)**

(c) $f(x) = \frac{\ln(3x-3)}{6x+2x^2}$ **(3)**

(d) $f(x) = \frac{\ln(2x-3x^2)}{\ln(x^2)}$ **(3)**

(e) $y = \ln(4x^2)\ln(8x^3)$ **(3)**

(f) $y = \frac{6x^2-4x+2}{\ln(7x-4)}$ **(3)**

[13 marks]

2.4 Differentiate the following **natural logarithms** in terms of x:

(a) $y = \ln(4x - x^3)$ **(2)**

(b) $y = 6\ln\left(\frac{x^5}{x^3}\right)$ **(2)**

(c) $y = \ln(e^{2x} + 2x)$ **(2)**

(d) $y = \ln((3x^3) \cdot 4x + 3x))$ **(2)**

(e) $f(x) = \frac{\ln x^4}{x^2}$ **(3)**

(f) $f(x) = \ln(\sqrt{x} + 2)$ **(2)**

[CA] [12 marks]

2.5 Determine the **gradient** of various logarithmic functions at the **following values**.

(a) Determine the **gradient** of $\boldsymbol{f(x) = 4\ln(2x)}$ at the **point** $(2, 0)$ **(2)**

(b) Determine the **gradient** of $\boldsymbol{f(x) = \ln(x^2 + 2)}$ at the **point** $(4, 1)$ **(2)**

(c) Determine the **gradient** of $\boldsymbol{y = \log_e(3x^3)}$ at the **point** $(e, 3)$ **(2)**

(d) Determine the **gradient** of $\boldsymbol{y = \log_6(x^2)}$ at $(3, 0)$ using the **change of base formula** **(3)**

(e) Determine the **gradient** of $\boldsymbol{y = \log_4(2x^3)}$ at $(5, 0)$ using the **change of base formula** **(3)**

Integration of Logarithmic Functions:

Repetitive: 2.6 → 2.10 (7 questions)

[6 marks]

2.6 **Integrate** the following functions.

(a) $f(x) = \int \frac{4}{x}\, dx$ **(1)**

(b) $f(x) = \int -\frac{7}{x} dx$ **(1)**

(c) $f(x) = \int \frac{4x}{x^2+6}\, dx$ **(2)**

(d) $f(x) = \int \frac{16x}{4x^2+5} dx$ **(2)**

(e) $f(x) = \int \frac{6x-4}{6x^2-8x} dx$ **(2)**

(f) $f(x) = \int \frac{4x+1}{8x^2+4x} dx$ **(2)**

[12 marks]

2.7 **Integrate** the following functions.

(a) $f(x) = \int \frac{6x^2-1}{6x^3-3x} dx$ **(2)**

(b) $f(x) = \int \frac{24+18x^2}{8x+2x^3} dx$ **(2)**

(c) $f(x) = \int \frac{2e^{3x}}{e^{3x}} dx$ **(2)**

(d) $f(x) = \int \frac{4e^x-3x^2}{4e^x-x^3} dx$ **(2)**

(e) $f(x) = \int \tan(x) + 2\, dx$ **(2)**

(f) $f(x) = \int \frac{-\cos(x)+12x}{6x^2-\sin(x)} dx$ **(2)**

[14 marks]

2.8.1 Using **natural logarithms**, solve the following **definite integrals**.

(a) $\int_e^{2e} \frac{2x}{x^2} dx$ **(2)**

(b) $\int_1^2 \frac{4x+1}{2x^2+x+2} dx$ **(2)**

(c) $\int_4^7 \frac{6x+5}{3x^2+5x-3} dx$ **(2)**

(c) $\int_5^{10} \frac{8x+4}{x^2+x+1} dx$ **(2)**

(e) $\int_{-6}^{4} \frac{12x-8}{3x^2-4x+5} dx$ **(3)**

(f) $\int_2^8 \frac{24x+9}{4x^2+3x-2} dx$ **(3)**

[18 marks]

2.8.2 Using **natural logarithms**, solve the following **definite integrals**.

(a) $\int_2^4 \frac{x^2}{x^3+3x^2}dx$ **(3)**

(b) $\int_5^7 \frac{2x^3+x}{x^4+x^2+5}dx$ **(3)**

(c) $\int_{-1}^1 \frac{45x^4-24x}{2+4x^2-3x^5}dx$ **(3)**

(c) $\int_{-1}^1 \frac{6x^3-4x^2}{36x^2-16x}dx$ **(3)**

(d) $\int_1^2 \frac{4x^3-16x^2}{6x^2-16x}dx$ **(3)**

(e) $\int_3^5 \frac{1}{x\ln x}dx$ **(3)**

[CA][15 marks]

2.9 Using integration, determine $f(x)$ for each of the following scenarios.

(a) Determine $f(x)$ if $f'(x) = \frac{5}{x}$ and $f(x)$ intersects the point $(2, 5.47)$ **(3)**

(b) Determine $f(x)$ if $f'(x) = -\frac{8}{x}$ and $f(x)$ intersects the point $(1, 3)$ **(3)**

(c) Determine $f(x)$ if $f'(x) = \frac{3x^2+3}{x^3+3x}$ and $f(x)$ intersects the point at $(1, -2.6)$ **(3)**

(d) Determine $f(x)$ if $f'(x) = \frac{4x-3}{4x^2-6x}$ and $f(x)$ intersects the point at $(3, -7.56)$ **(3)**

(e) Determine $f(x)$ if $f'(x) = \frac{12+12x^2}{6x+2x^3}$ and $f(x)$ intersects the point at $(4, 7)$ **(3)**

[20 marks]

2.10 **Integrate** the following functions and give the answer **in the form of only one natural log function**. There is **no need to simplify**.

(a) $\int \frac{12x+24}{3x^2+12x} + \frac{3}{9x}dx$ **(5)**

(b) $\int \frac{42x-9}{14x^2-6x} - \frac{32x}{8x^2}dx$ **(5)**

(c) $\int \frac{4xe^{x^2}-8x}{4e^{x^2}-8x^2} + \frac{6x^2}{x^3}dx$ **(5)**

(d) $\int \frac{54x^2-84x}{6x^3-14x^2} - \frac{\ln(x)}{2}dx$ **(5)**

General Applications of Logarithmic Functions:

Repetitive: 2.11 → 2.15 (5 questions)

[CA] [7 marks]

2.11 A bullet train leaves a station with a velocity that follows the equation: $v = \ln(2t^2 + 4)$, where t is **time** in **seconds**, and **velocity** is in km/s. Note that the **one second** into the train ride, the train is accelerating at $2.67\ km\ s^{-2}$.

(a) What is the **velocity** of the train after **three seconds**? **(1)**

(b) At what **time** had the train travelled $6\ km$? **(1)**

(c) Determine the **acceleration** of the train **four seconds** after it has left the station. **(2)**

(d) Determine the spaceship's **displacement** after **3 seconds**. **(3)**

[CA] [8 marks]

2.12 An airplane is travelling at a speed that follows the equation $v(t) = \ln(2x^2) + 4x$, where t is in **minutes**.

(a) Determine the **instantaneous rate of change** in the **value** of at **6 minutes**. **(3)**

(b) Determine the **instantaneous rate of change** in the **value** of at **8 minutes**. **(2)**

(c) Determine the **average rate of change** from the time of 6 to **8 minutes**. **(3)**

[CA] [5 marks]

2.13 A tree is 1.65 **meters** tall and is growing following the function $h_1(t) = \ln(t + 5.21)$, where t is **time** in **weeks**. Another tree is 2.23 **meters** tall.

(a) **How long** will it take for the **first tree** to **match** the **height** of the **second tree**, assuming that the second tree is not growing? **(2)**

(b) The second tree is growing at a rate following the function $h_2(t) = \ln(t^2) + 3$ where t is time in **weeks**. Determine after how many **weeks** the trees will be the **same height.** State this height. **(3)**

[CA] [12 marks]

2.14 Calculate the **area** trapped between the **natural logarithmic function** and the x**-axis**.

(a) Calculate the **area under the curve** $f(x) = \frac{-6x^5}{x^6} + 6$ between $x = 1$ and $x = 4$ **(3)**

(b) Calculate the **area under the curve** $f(x) = \frac{-10x}{5x^2} + 3$ between $x = 2$ and $x = 3$ **(3)**

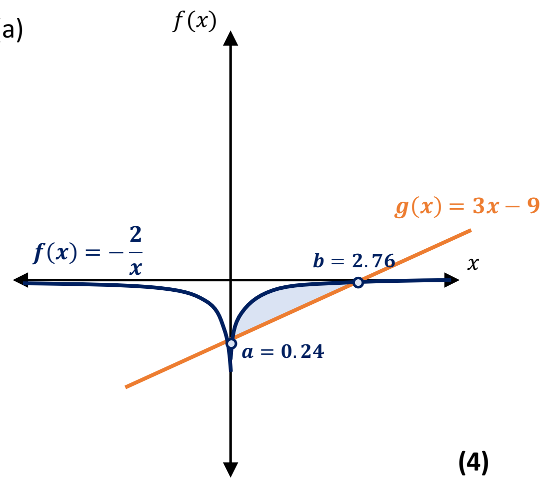
(c) Calculate the **area under the curve** $f(x) = \frac{36x}{9x^2}$ from $x = 1$ to $x = 5$ **(3)**

(d) Calculate the **trapped between the** x**-axis** and $f(x) = \frac{3x^2}{7x^3}$ from $x = 1$ to $x = 4$ **(3)**

[CA] [17 marks]

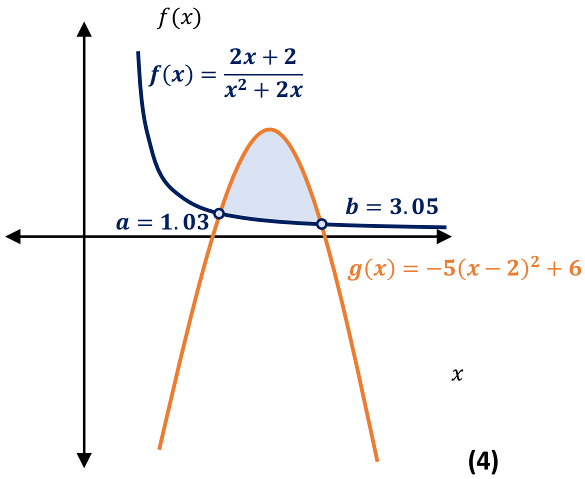
2.15 **Determine** the **shaded areas below** using **definite integrals**.

(a)



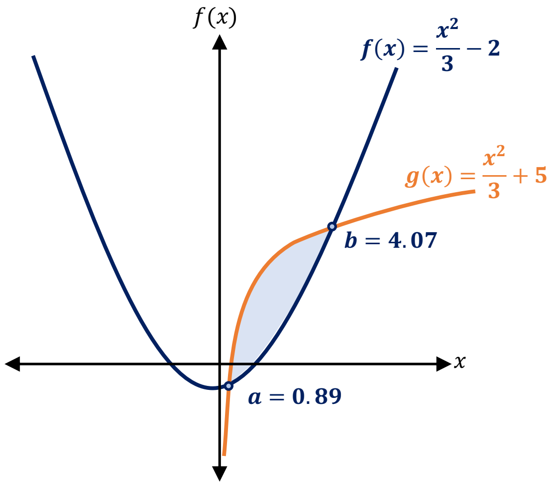
(4)

(b)



(4)

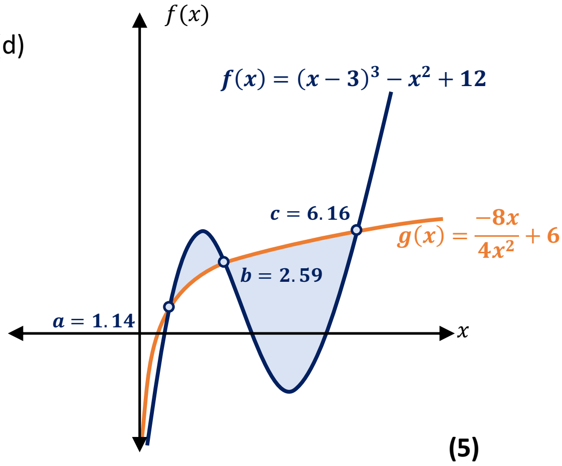
(c)



$f(x)$
$f(x) = \frac{x^2}{3} - 2$
$g(x) = \frac{x^2}{3} + 5$
$b = 4.07$
x
$a = 0.89$

(4)

(d)



$f(x)$
$f(x) = (x-3)^3 - x^2 + 12$
$c = 6.16$
$g(x) = \frac{-8x}{4x^2} + 6$
$b = 2.59$
$a = 1.14$
x

(5)

Chapter 8

Continuous Random Variables

Chapter 8 – Continuous Random Variables

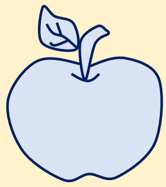
Introduction

Now we've learnt calculus and a little probability in Unit 3, it's time for us to explore more probability, starting with **Continuous Random Variables**!

This topic is quite interesting as it is often applicable to real life! **Continuous Random Variables** can be used to accurately represent the probability of certain events occurring, and you are much more likely to encounter **Continuous Random Variables** in real life than **Discrete Random Variables.**

This makes learning about them really interesting and beneficial! Here are some examples of how we use **Continuous Random Variables** in real life:

Determining the Price of Insurance



Finding the Average Weight of Fruit

Don't worry if these phrases don't make sense to you yet, we are going to take you through each concept with **explanations** and **definitions** before giving you a chance to try out the **questions**!

Let's begin!

8.1 Principles of Continuous Random Variables

A **random variable** is a variable whose value is a numerical outcome of a random phenomenon. In Unit 3, we covered **discrete random variables**. It's important that you understand what they are and how you can tell them apart from **continuous random variables**.

> **Discrete random variables** are random variables with a **countable number** of possible values.

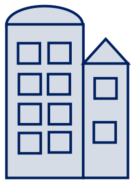
A common example of a discrete random variable is the number of heads obtained if you **flip a coin a certain number of times**. The **number of heads** can only take on **discrete**, **whole number** values such as 0, 1, 2, 3, and so on. It is not possible to get **less than one** head or **2.5** heads.

Continuous random variables can take any value in a continuous range, including fractions and infinitely long decimals.

> **Continuous random variables** are random variables with an **infinite number** of possible values.

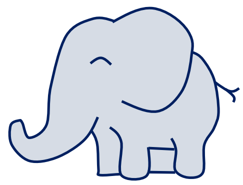
One way to differentiate **continuous random variables** and **discrete random variables**, would be while **discrete random variables** are countable items, **continuous random variables** are measured items.

An easy way to think about it is this is that we **count the outcomes** for a **DRV** and we **measure** the outcomes for a **CRV**. For instance, **height** is seen as **continuous** because it is possible to have values that lie **between whole numbers** (i.e., $x = 160, 160.1, 160.15, \ldots$). Other examples of continuous random variables also include:



$x =$ Height of a Building

$y =$ Time to Drive to School



$z =$ Mass of an Elephant

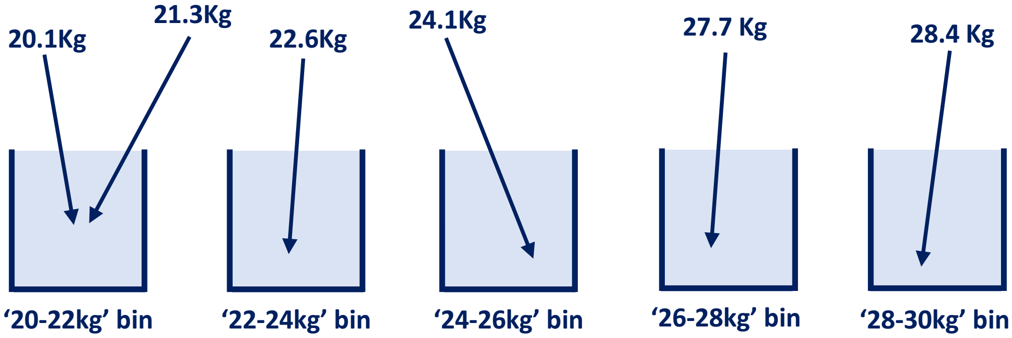
Relative Frequency using Histograms

When we collect data, we often want to **display** it in a way that makes it **easy to interpret. This allows us to observe trends over time or the relationship between variables!** For **discrete random variables**, this often takes the form of a **bar graph**, allowing easy comparison of **how common each option was**. Let's see what happens if we try to do this for **continuous random variables**!

Let's say we go on a trip to the outback to see some **friendly kangaroos**, and **weigh kangaroo** we can find. On this trip, we find 25 friendly kangaroos who weigh the following:

Skippy: 21.3Kg	**Dippy: 20.5Kg**	**Lippy: 28.6Kg**	**Kippy: 21.4Kg**	**Bippy: 21.6Kg**
Slippy: 25.3Kg	**Shippy: 22.6Kg**	**Vippy: 24.1Kg**	**Tippy: 27.7Kg**	**Sippy: 23.4Kg**
Rippy: 24.7Kg	**Blippy: 21.3Kg**	**Snippy: 23.3Kg**	**Plippy: 21.7Kg**	**Yippy: 22.3Kg**
Quippy: 22.2Kg	**Wippy: 20.9Kg**	**Grippy: 23.7Kg**	**Flippy: 25.5Kg**	**Nippy: 21.9Kg**
Mippy: 28.4Kg	**Jippy: 29.9Kg**	**Hippy: 29.4Kg**	**Xippy: 29.3Kg**	**Pippy: 20.1Kg**

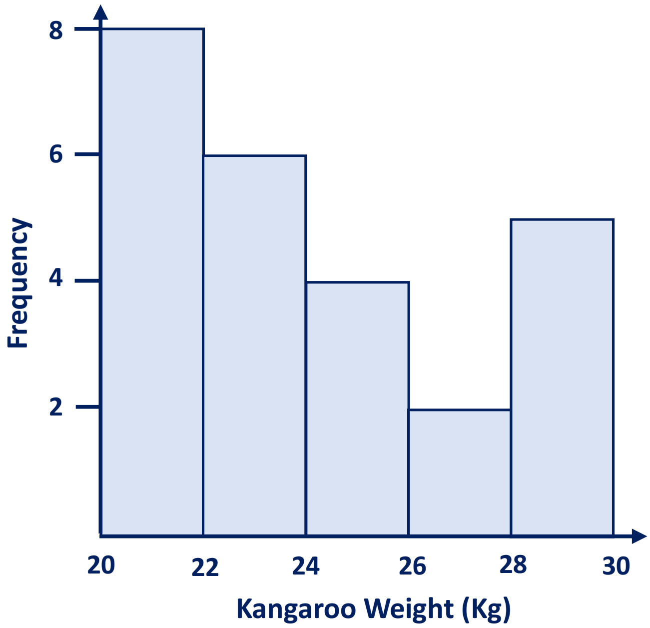
If we tried to create a **bar graph** for this, we'd find that **none of the weights appear more than once**! This means in this case a bar graph **isn't very useful at all**! Instead, we can define some '**bins**' that cover a **range of values**, and place our weights into one of those bins.



Doing this for **all our weights**, we can create a table to represent all the bins:

Bin (Kg)	Number of Kangaroos
20-22	8
22-24	6
24-26	4
26-28	2
28-30	5

We can then turn this table into a **histogram**, which is similar to a bar graph with the **height** representing the **frequency at which each weight occurs**, but the **x axis represents the bins** rather than the individual weights:



This histogram can help us estimate the **probability** of a **particular event** occurring.

Relative frequency is measured as a **ratio** between the **frequency of a certain event occurring** to the **total number of events**.

To calculate relative frequency of an event occurring, we first need to **determine** the **number of times** that this event has **occurred** in the **past**. Then we **divide** this number by the **total number of events that have happened**. A **histogram** is a useful way to help us determine these values.

Let's look at our example above to determine **relative frequency**! Using data in a **histogram**, we could estimate the **probability** of a variable being within a particular range. Using our kangaroo data, let's say we wanted to find the probability that a kangaroo in this population weighed between **22 kg and 24 kg**. We just need to divide number of kangaroos in this range, which is **6**, by the total number of kangaroos, **25**. Therefore, the probability that any given kangaroo weighs between **22 kg and 24 kg** is $\mathbf{0.24}$.

$$P(22 < x < 24)$$

① $n_{22<x<24} = \mathbf{6}$

② $n_{total} = \mathbf{25}$

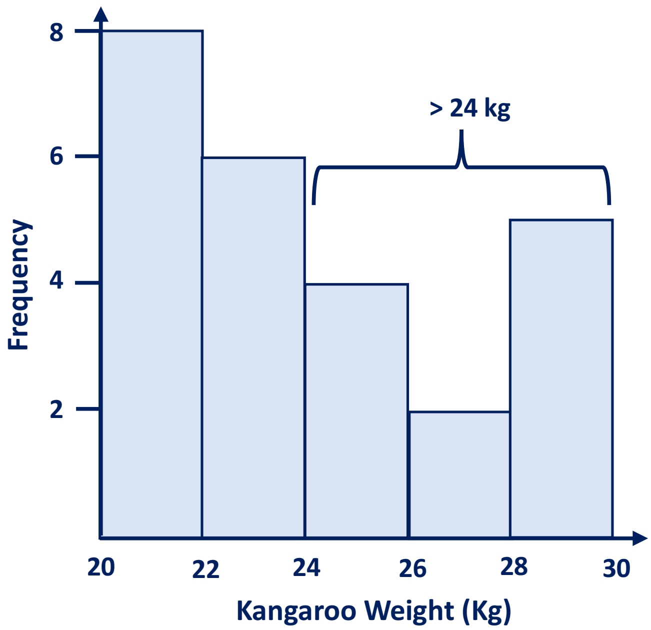
③ $p = \frac{\mathbf{6}}{\mathbf{25}} = \mathbf{0.24}$

① **Determine** number of kangaroos between **22 and 24 kg**

② **Determine total number** of kangaroos, of **any weight**

③ **Divide** number of kangaroos in **weight bracket** by **total** kangaroos

This can be extended to any range within our data. For example, to find the **probability** of finding a kangaroo that weighs **over 24 kg**, we add up the number of kangaroos in this range $(4 + 2 + 5 = 11)$, then divide it by our total of 25. This gives us a probability of $\mathbf{0.44}$.



$$P(x > 24)$$

① $n_{x>24} = \mathbf{11}$

② $n_{total} = \mathbf{25}$

③ $p = \frac{\mathbf{11}}{\mathbf{25}} = \mathbf{0.44}$

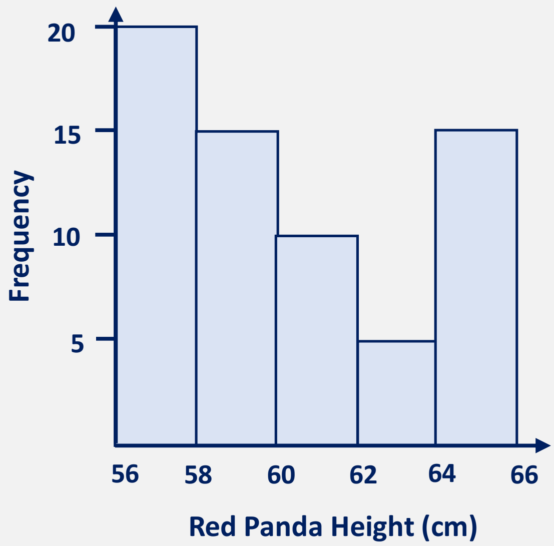
① **Determine** number of kangaroos that weigh over **24 kg**

② **Determine total number** of kangaroos, of **any weight**

③ **Divide** number of kangaroos in **weight bracket** by **total** kangaroos

Worked Example 1

Student Rupert is interested in the heights of red pandas in a colony. Rupert counted and **classified** the **heights** of the pandas and created the following **histogram**.



(a) Determine the relative frequency of the selected red panda measuring between 60 and 62 cm.

$$P(60 < x < 62)$$

$$n_{60<x<62} = \mathbf{10}$$

$$n_{total} = \mathbf{65}$$

$$p = \frac{\mathbf{10}}{\mathbf{65}} = \mathbf{0.15}$$

(b) Estimate the probability that a selected red panda will be less than 64 cm tall.

$$P(x < 64)$$

$$n_{x<64} = \mathbf{50}$$

$$n_{total} = \mathbf{65}$$

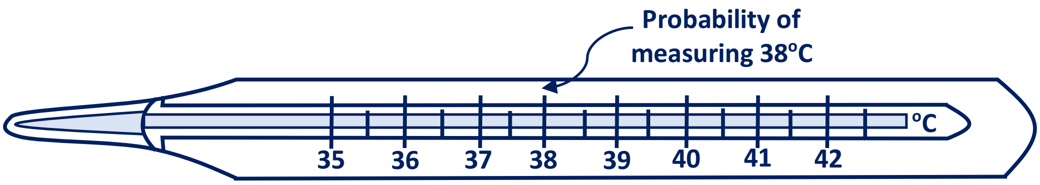
$$p = \frac{\mathbf{50}}{\mathbf{65}} = \mathbf{0.80}$$

Probability density functions

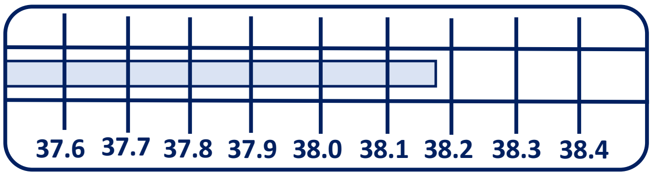
In Unit 3, we introduced **discrete random variables** and explained the **probability distribution**. Remember that the **probability distribution function** gives the probability of each outcome occurring, written as $p(x)$. Note that the probability distribution function is defined under **discrete random variables**. For example, you can determine the probability of rolling a 0, 1, 2 or 3 on a dice.

x	0	1	2	3
$P(X = x)$	0.20	0.40	0.30	0.10

Let's think about a thermometer reading. Using **discrete random variables**, we could calculate the probability of the temperature reading exactly 38 degrees.



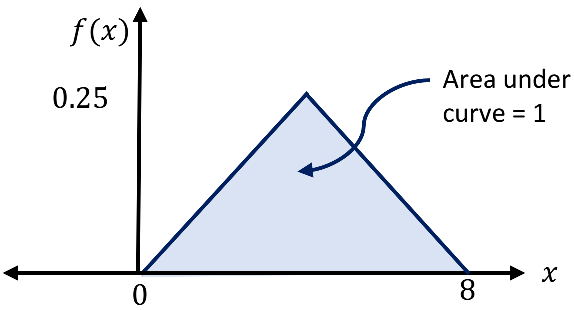
However, temperature can be a **continuous random variable**. Normally, we round temperature to a whole degree but this doesn't give us the whole picture. In the thermometer below, we would say the temperature is **38 degrees**, but in reality this is rounded from **38.82 degrees**.

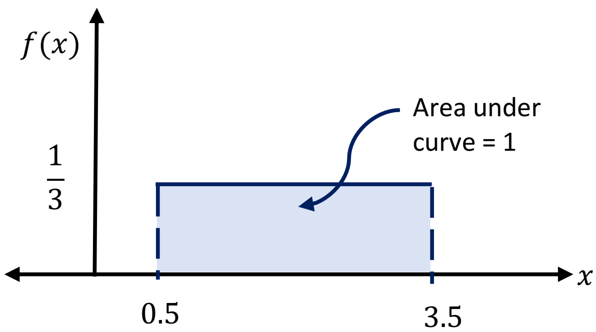


Because there are **infinite temperatures** that lie between each reading of degree, calculating the **probability** of obtaining a temperature of **exactly 38 degrees** is unhelpful. Instead, we may look at the **probability** of obtaining a value between **37.9** and **38.1°C**. When calculating the probability of a continuous random variable, we use something called a **probability density function**. Let's look at the formal definition of the **probability density function**.

> The **probability** of a **continuous random variable** taking a value **in a given interval** is equivalent to the **area under the curve** of the variable's **probability density function** in that interval.

Probability density functions can take on many different shapes. Some examples of these have been included below.





Because a **probability density function** is a just a function, we can graph it over a **specified interval**. For our thermometer example above, we would make a graph with the lower bound, 35 and upper bound 42. The **area under the probability density function** represents the **probability** of an event occurring. Within these ranges, we can specify the probability of the temperature recording within the range of 37.9 and 38.1 degrees. We would calculate the area between the interval $[37.9, 38.1]$ which represents the probability of that the event results in a number between 37.9 and 38.1.

This idea can be simplified with the following definition of probability density functions.

If $f(x)$ is the **probability density function** of a **continuous random variable** X, then the probability that X takes a value in the interval $[a, b]$ is given by $\boldsymbol{P(a \leq X \leq b) = \int_a^b f(x)dx}$.

Probability density functions $f(x)$ have two properties:

Probability Density Function properties: ① $\boldsymbol{f(x) \geq 0}$ and ② $\boldsymbol{\int_{-\infty}^{\infty} f(x)dx = 1}$

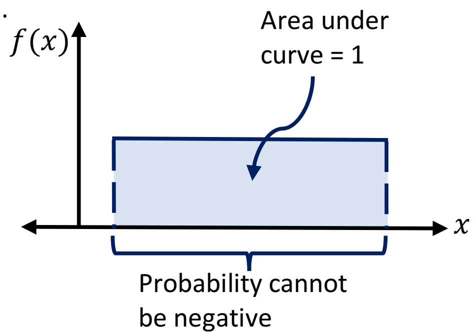
Let's have a look at an example to see what these properties mean.

① $\boldsymbol{f(x) \geq 0}$

① Probability cannot be **negative**. This means that in a probability density, **values** all must be in **positive values** on the **y-axis**.

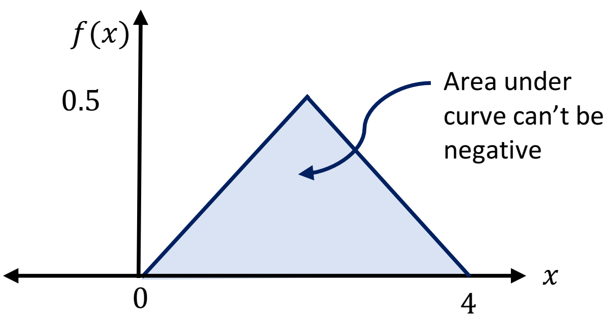
② $\boldsymbol{\int_{-\infty}^{\infty} f(x)dx = 1}$

② The total area under the curve of a **probability density function curve** must sum to one.

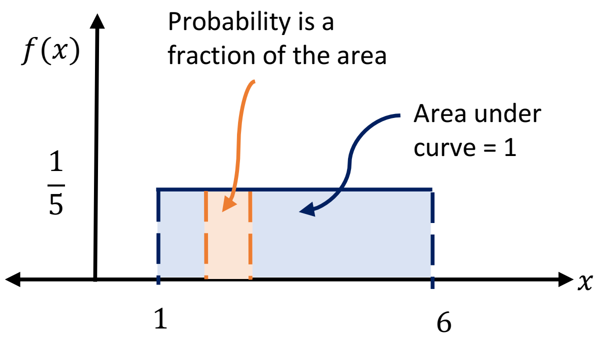


Let's unpack these properties and what they mean for calculating probability. We know that for **continuous random variables**, the values they take must lie in **an interval**. To find the probability that they lie in this interval, we need to find **the area under** the **probability density function** within the interval.

Property 1 shows that the **area under** the **probability density function** can't be negative.

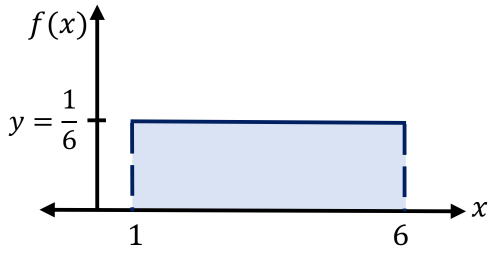


Property 2 shows us that the total **area under** the **probability density function** is 1, which suggests that the probability is represented as a fraction of the area under the **probability density function** within **the interval**.

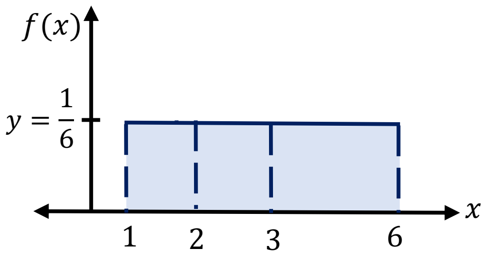


The **probability density function** can take many shapes, and the only requirements are **those two properties**. Let's look through a detailed example and see how we can use both **geometry** and **calculus** to determine the probability of a certain event occurring!

Imagine we are given a probability distribution that is modelled by the function $f(x) = \frac{1}{6}$, and is true over the bounds of 1 to 6.



We are then asked to determine the probability of this continuous random variable falling between the values of 2 to 3. Let's use our knowledge of geometry to determine this area under the curve, and therefore the probability!



① $A = \boldsymbol{l} \times \boldsymbol{w}$

② $\boldsymbol{l} = \frac{\mathbf{1}}{\mathbf{6}}, \boldsymbol{w} = (\mathbf{3} - \mathbf{2}) = \mathbf{1}$

③ $A = \frac{1}{6} \times 1 = \frac{1}{6}\ units^2$

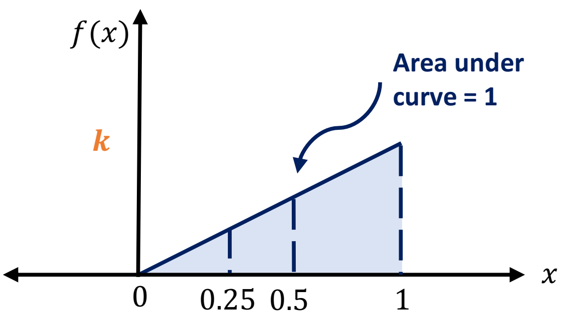
$\therefore \boldsymbol{P} = \frac{\mathbf{1}}{\mathbf{6}}$

① **Determine** general formula for the area of the shape

② Find the **width** of the **graph**

③ **Substitute** in values for **width** and **length** (determined by the specified **interval**)

Let's look at another example. Imagine we wanted to find the probability of obtaining a value of **0** to **1** degrees and also want to figure out the value for **unknown k**. From the Property 2, we know the area under the curve **must be one**. Given that, we can identify the **unknown k** from the **integrals**.

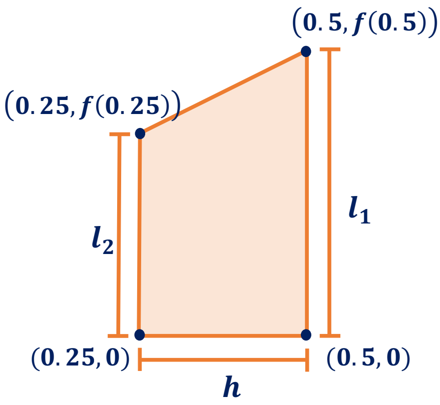


$$\mathbf{1} = \int_0^1 f(x)dx$$
$$= \int_0^1 kx\,dx$$
$$= k\left[\frac{x^2}{2}\right]_0^1$$
$$= k\left(\frac{1^2 - 0^2}{2}\right)$$
$$= \frac{k}{2}$$
$$k = 2$$

We can also use **geometry** to **determine** the value for k. We know that the area for a triangle is $\boldsymbol{A} = \frac{\boldsymbol{hb}}{\mathbf{2}}$. In this case, we know that the area under the curve is equal to one, $\boldsymbol{b} = \mathbf{1}$ and the height is defined by **unknown k**.

$$1 = \frac{1}{2} \times k$$
$$\mathbf{1} = \frac{k}{2}$$
$$k = 2$$

In some questions, you can use **geometry** or **integration**. this case, **geometry** is the easier method to use to find k. Now $P(-0.25 \leq X \leq 0.5)$ can be found. The **geometry** approach uses the equation for the area of a **trapezoid**.



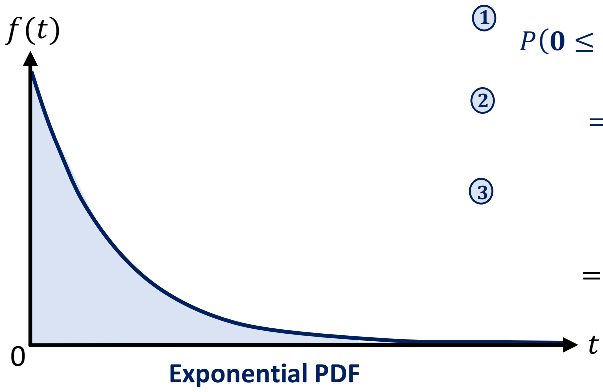
$$\begin{aligned} P(0.25 \leq X \leq 0.5) &= \frac{h(l_1 + l_2)}{2} \\ &= \frac{1}{2}(0.5 - 0.25)(f(0.25) + f(0.5)) \\ &= \frac{1}{8}(k(0.25) + k(0.5)) \\ &= \frac{1}{8}(0.5 + 1) \\ &= \frac{3}{16} \end{aligned}$$

The **integral** approach can also be calculated to find the **area under the curve.**

$$\begin{aligned} P(-70 \leq X \leq -50) &= \int_{0.25}^{0.5} f(x)dx \\ &= \int_{0.25}^{0.5} (kx)\, dx \\ &= \int_{0.25}^{0.5} (2x)\, dx \\ &= 2\left[\frac{x^2}{2}\right]_{0.25}^{0.5} \\ &= 2\left(\frac{0.5^2 - 0.25^2}{2}\right) \\ &= \frac{3}{16} \end{aligned}$$

However, sometimes the **integral** approach is the only method that can be used. Refer to the following **exponential distribution** example.

The **lifetime** of a **particle** is **exponentially distributed**. In this example, a particles lifetime is given by a **probability density function** that is the **exponential function** $f(t) = \frac{1}{9}e^{-\frac{1}{9}t}$, where t is the lifetime of the particle in **seconds**. We want to know the **probability** of it decaying within the **first 3 seconds** after it is created.



① $P(0 \leq t \leq 3) = \int_0^3 f(t)dt$

② $= \int_0^3 \frac{1}{9} e^{-\frac{1}{9}t}\, dt$

③ $= -\left[e^{-\frac{1}{9}t}\right]_0^3$

$= -\left(e^{-\frac{3}{9}} - e^0\right)$

$= 0.2835$

① **Substitute 0** and **3** as **lower** and **upper bound** respectively.

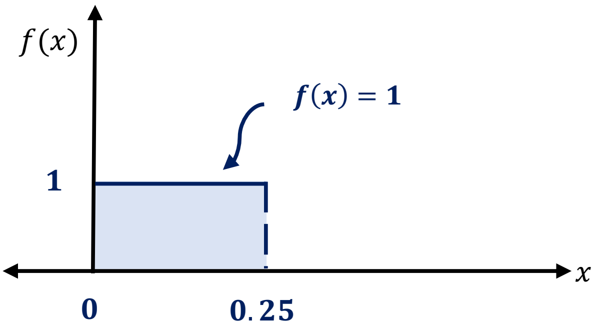
② **Substitute** in the **equation**

③ **Solve** for the **probability!**

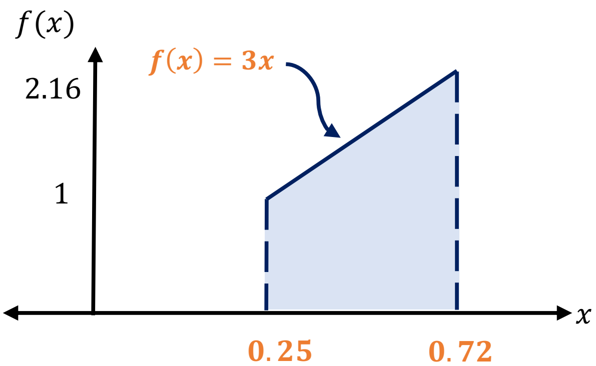
Now that we know how to calculate probabilities from the single equations and graphs given to us, we can learn how to read **probability density functions**. **Probability density functions** are often written like this:

$$f(x) = \begin{cases} \mathbf{1}, & \mathbf{0 \leq x \leq 0.25} \\ \mathbf{3x}, & \mathbf{0.25 < x \leq 0.72} \end{cases}$$

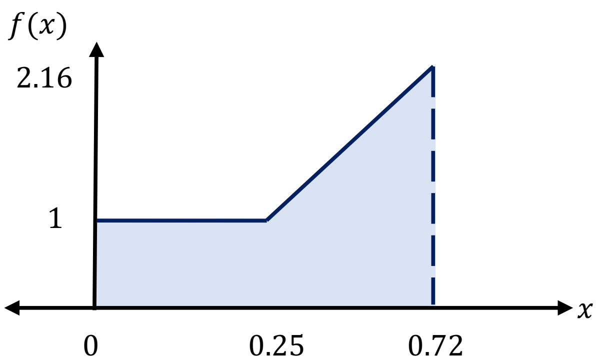
While this may look a bit confusing at first, we can read each line separately. On the top line, $\boldsymbol{f(x) = 1, 0 \leq x \leq 0.25}$, which means that the **value of the function is equal to 1** between $\boldsymbol{x = 0}$ and $\boldsymbol{x = 0.25}$. We know that this function looks like a straight, horizontal line:



On the bottom line, $\boldsymbol{f(x) = 3x, 0.25 \leq x \leq 0.72}$, which means that the **value of the function is equal to $3x$** between $\boldsymbol{x = 0.25}$ and $\boldsymbol{x = 0.72}$. We know that this function looks like a straight line with a positive slope of 3:



Putting these two functions together, we get:

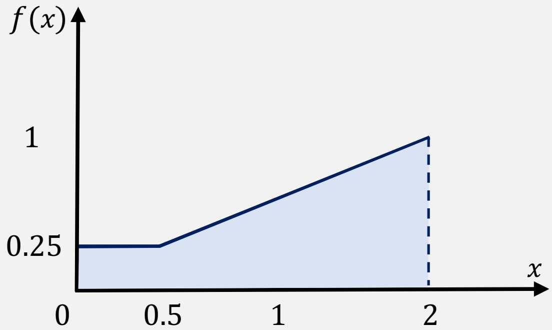


Worked Example 2

A probability density function is given below:

$$f(x) = \begin{cases} \frac{1}{4}, & 0 \leq x \leq 0.5 \\ \frac{1}{2}x, & 0.5 < x \leq 2 \end{cases}$$

What is $\boldsymbol{P(0 \leq x \leq 1)}$?



$$P(0 \leq x \leq 1) = \int_0^1 f(x)dx$$

$$= \int_0^{0.5} \frac{1}{4}dx + \int_{0.5}^{1} \frac{1}{2}xdx$$

$$= \left[\frac{1}{4}x\right]_0^{0.5} + \left[\frac{1}{4}x^2\right]_{0.5}^{1}$$

$$= \frac{1}{8} + \left(\frac{1}{4} - \frac{1}{16}\right)$$

$$= \frac{5}{16}$$

8.2 The Cumulative Distribution Function

In Unit 3, we explored the following example.

$x = 1, 2, 3, 4$
(number drawn)

x	1	2	3	4
$P(X = x)$	0.10	0.30	0.20	0.40

This example shows the probability of drawing each number between 1 and 4. The probability of getting the number 1, 2, 3, and 4 is 0.1, 0.3, 0.2, and 0.4, respectively. We can use discrete random variables to calculate any unknowns.

But what if we wanted to calculate the probability of the number drawn being **less than or equal to 3**? Only **outcome 1, 2 and 3** satisfy the condition. This means that the probability is $\mathbf{0.1 + 0.3 + 0.2 = 0.6}$. The process of calculating the probability that a variable takes on a value **less** than or **equal** to a variable can be described by a function. This function is called the **cumulative distribution function.**

The **cumulative distribution function** of **any random variable** $\boldsymbol{X}$ is defined by the function $\boldsymbol{F(x) = P(X \leq x)}$

The cumulative distribution functions can apply to both **discrete** and **continuous random variables**. When we are looking at continuous random variables, the **cumulative distribution function** (**denoted** $\boldsymbol{F(X)}$) and **probability density function** (**denoted by** $\boldsymbol{f(x)}$) have an interesting relationship.

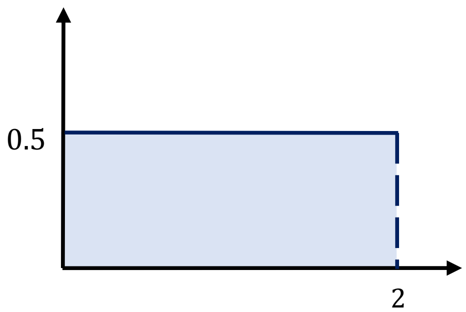
$$\boldsymbol{F(x) = \int_{-\infty}^{x} f(t)dx \Leftrightarrow f(x) = \frac{d}{dx} F(X)}$$

Note that $\boldsymbol{f(t)}$ and $\boldsymbol{f(x)}$ are the **same function**. $\boldsymbol{t}$ is used instead of $\boldsymbol{x}$ so that $\boldsymbol{x}$ can be used in the upper bound of the integral. Notice that $\boldsymbol{t}$ is replaced by $\boldsymbol{x}$ in the integral, so the choice of variable **doesn't matter**, as long as it isn't $\boldsymbol{x}$.

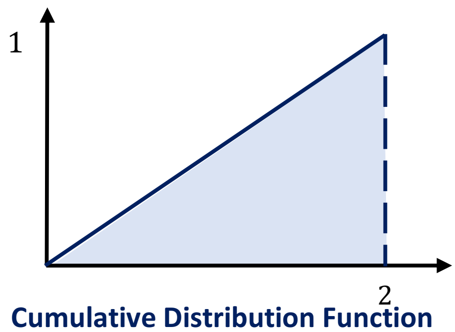
Let's look at an example. A probability density function is given by $\boldsymbol{f(x) = 0.5}$ for $\boldsymbol{0 \leq x \leq 2.}$ The relationship between the **probability density function** and **cumulative distribution function** tells us that the **cumulative distribution function** is the **integral** of the **probability density function**. Therefore, we know that over the range $\boldsymbol{0 \leq x \leq 2}$, the **cumulative distribution function** is:

$$\begin{aligned} \boldsymbol{F(x)} &= \int_{-\infty}^{x} \boldsymbol{f(t)}dx \\ &= \int_{0}^{x} (\boldsymbol{0.5})dt \\ &= [0.5t]_0^x \\ &= \boldsymbol{0.5x}. \end{aligned}$$

Going back to our graphing that we learnt in calculus, we can **graph** a **cumulative distribution function** from a **probability density function!**



Probability Density Function



Cumulative Distribution Function

Let's revisit out example from above.

$\boldsymbol{x}$	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$
$\boldsymbol{P(X = x)}$	$\mathbf{0.10}$	$\mathbf{0.30}$	$\mathbf{0.20}$	$\mathbf{0.40}$

The number drawn example is a **discrete random variable X** with a **cumulative distribution function** represented by F. Suppose we want to know the **probability** that the drawn number is **less than or equal to 4**. It equals the summation of the total probability of obtaining 1, 2, 3, and 4.

$$P(X \leq 4) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$$

As we learnt in unit 3, **mutually exclusive events** are **two events** that **cannot happen at the same time**. If one event happens, the other cannot. Let's find mutually exclusive events for our example. It's easy to find two **mutually exclusive events** in the event space, i.e. $X \leq 1$ and $1 < X \leq 4$, and the union of such two events is $X \leq 4$.

$X \leq 1$, represents the **values of X that are less than or equal to 1**, while $1 < X \leq 4$, represents the **values of X that are greater than 1 but less than or equal to 4**. These cannot happen at the same time. When we **combine** all outcome that **satisfy either of the events**, we get a **union**, $X \leq 4$. Every possible value will be less than or equal to 4.

$$P(X \leq 4) = P(X \leq 1) + P(1 < X \leq 4)$$

We rearrange the function and get the following relationship.

$$\begin{aligned} P(1 < X \leq 4) &= P(X \leq 4) - P(X \leq 1) \\ &= F(4) - F(1) \end{aligned}$$

In fact, for any random variable X with a cumulative distribution function $F(x)$, for any a, b in the event space and $a < b$, the following relationship is always true.

$$\begin{aligned} P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ &= F(b) - F(a) \end{aligned}$$

This relationship leads to some important properties of the **cumulative distribution function**.

Cumulative Distribution Function Properties
The cumulative distribution function is **non-decreasing**.
If x approaches to a negative infinity, $F(x)$ approaches to **0.**
If x approaches to a positive infinity, $F(x)$ approaches to **1.**

These three properties describe the shape of cumulative distribution functions. It always has a minimum value **0** and a maximum value **1**. If the probability distribution is **0** for all values below $x = a$, then $F(x) = 0$ when $x \leq a$. Similarly, if the probability distribution is **0** above all values of $x = b$, then $F(x) = 1$ when $x \geq b$. For any **a**, **b** in the domain and $a < b$, $F(a) < (b)$. It's worth to note that the $<$ and $\leq$ and the $>$ and $\geq$ signs are **interchangeable** for **continuous random variables.**

$$\begin{aligned} P(a < X \leq b) &= P(a < X < b) \\ &= P(a \leq X < b) \\ &= P(a \leq X \leq b) \end{aligned}$$

This is because there is still a **probability of 0** that the **continuous random variable** will take an **exact value**, so the signs are effectively **equivalent** here.

Worked Example 3

The **cumulative density function** of a continuous random variable X is

$$F(x) = e^{2x+1} - 3, 0 \leq X \leq b$$

Find the **value of b** and identify $P(0.05 < X \leq 0.1)$

$$e^{2b+1} - 3 = 1$$
$$2b + 1 = ln(4)$$
$$2b = ln(4) - 1$$
$$b = \frac{ln(4) - 1}{2}$$
$$\approx 0.193$$

$$P(0.05 < X \leq 0.1) = F(0.1) - F(0.05)$$
$$= \left[e^{2x+1} - 3\right]_{0.05}^{0.1}$$
$$= e^{2\times0.1+1} - 3 - (e^{2\times0.05+1} - 3)$$
$$= e^{1.2} - 3 - (e^{1.1} - 3)$$
$$= e^{1.2} - e^{1.1}$$
$$\approx 0.316$$

8.3 Expected Value, Variance and Standard Deviation

Continuous random variables can be defined by calculating and interpreting its parameters by the **expected value**, **variance**, and **standard deviation**. The **expected value**, **variance**, and **standard deviation** can be calculated. Even these are now applicable to continuous random variables, we can use our knowledge from unit 3 about **expected value**, **variance**, and **standard deviation.**

The **mean**, μ, also known as the **expected value**, $E(X)$, is the **average outcome** of a random variable. It is the **average value** you would **expect** if you continually repeated an event.

We can use the following formula to calculate the **mean** (μ) or **expected value** ($E(X)$) of a continuous random variable:

$$\textbf{Mean}: \quad E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

This formula tells us that the **mean** is the **integral** of the lowest possible X value in the range **(x min)** to the largest possible X value **(x max)**. This **differs** from the calculation of an expected value for a **discrete random variable** that requires the summation of all possible values. For instance, suppose we had the following function for a continuous random variable X, with the probability density function $f(x)$.

$$f(x) = \begin{cases} \frac{3(x-2)^2}{8}, & 0 \leq x \leq 2 \\ 0, & elsewhere \end{cases}$$

To determine the mean, we would find the **integral** of the lowest possible X value in the range **(x min)** to the largest possible X value (**x max**), where x is **multiplied** by the probability density function $f(x)$:

① $$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

② $$E(X) = \int_{0}^{2} x \cdot \frac{3(x-2)^2}{8} dx$$

③ $$E(X) = \frac{1}{2}$$

① **Identify** the general formula for the expected value.

② Substitute in the **upper** and **lower bound**, and the **function**.

③ **Determine** the **expected value.**

Let's look at how we can determine the **variance** and **standard deviation** of a **continuous random variable!**

> **Variance**, $Var(X)$, is also a measure of the **spread** of the **results from the mean**. It is equal to the **standard deviation squared** (i.e. $Var(X) = \sigma^2$).

> **Standard deviation** (σ) is a measure of the **spread** of **results from the mean**. A **small σ** indicates the data points are **spread close** to the **mean**, whereas a **large σ** indicates the data is **widely spread** from the mean.

As you can see above, the **variance** and **standard deviation** are closely related as they both describe the **variability** of **data**. The **variance** of a continuous random variable can be calculated by the formulas:

> **Variance:** $Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$

> **Variance:** $Var(X) = E(X^2) - \mu^2$

The **first**, **integral formula**, is almost always used when we have a continuous random variable that is defined by a function. The **second**, **expected value formula** is a simpler formula that has been derived from the first formula. Let's look at how we can use the formulas to determine variance by applying the second variance method to our previous example, assuming that our mean was $\mu = \frac{1}{2}$.

Method 1

① $$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$
$$= \int_{0}^{2} \left(x - \frac{1}{2}\right)^2 \frac{3(x-2)^2}{8} dx$$
$$= \frac{3}{20}$$

① **Substitute** values into $Var(X)$ formula

Method 2

① $$E(X^2) = \int_{0}^{2} x^2 \cdot \frac{3(x-2)^2}{8} dx$$
$$= \int_{0}^{2} x^2 \cdot \frac{3(x-2)^2}{8} dx$$
$$E(X^2) = \frac{2}{5}$$

① **Substitute** values into $E(X^2)$ formula

② $$Var(X) = E(X^2) - \mu^2$$
$$= \frac{2}{5} - \frac{1}{2}^2$$
$$= \frac{3}{20}$$

② **Substitute** $E(X^2)$ and μ^2 into the **variance formula**

To calculate the **standard deviation** of a continuous random variable, we first need to calculate the **variance**. Standard deviation can be calculated with the following formula:

Standard Deviation: $\boldsymbol{\sigma = \sqrt{Var(X)}}$

This formula tells us that **standard deviation** is the **square root** of the value for **variance**. Therefore, to find standard deviation, variance needs to be calculated first. Regardless of the method used to determine the variance, the **standard deviation** is calculated in the same way.

The final calculation we can do is for **standard deviation**. This is straightforward as it is the **square root** of **variance**:

① $\boldsymbol{Var(X) = 0.15}$

① **Determine** the **variance** using either method.

② $\boldsymbol{\sigma = \sqrt{Var(X)}}$
$\boldsymbol{\sigma = \sqrt{0.15}}$
$\boldsymbol{\sigma = 0.39}$

② **Substitute** the **variance** into the **standard deviation formula.**

The formula for expected value (mean), variance and standard deviation for continuous random variable can be summarised in the table below:

Mean (μ)	Variance ($Var(X)$)	Standard Deviation (σ)
$E(X) = \int_{-\infty}^{\infty} xf(x)dx$	$V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$ or $Var(X) = E(X^2) - \mu^2$	$\sigma = \sqrt{Var(X)}$

Linear transformations of Continuous Random Variables

From our understanding of earlier parts in this chapter, things that are considered 'continuous' include height, weight and temperature as they can be any values within a given interval.

The distribution of a continuous random variable can be **changed** by applying **addition**, **subtraction**, **multiplication** or **division** to the random variable. Formally, these are called **linear transformations**. We defined linear transformations in unit 3.

Linear transformations are **changes** to a variable characterised by **multiplying** the random variable by a **constant** (a) and/or by **adding** a **constant** (b) to the random variable.

Though there are four different linear transformations that can be applied, these changes are categorised into either a **change of scale** or a **change of origin**. Change of **scale** and change of **origin** are defined as:

A **change of scale** is the **stretching/contracting** of a random variable by **multiplying** it by a **constant** (a).

A **change of origin** is the **shifting** of each value of a random variable by **adding** a **constant** (b).

When we have a function of a continous random variable X, linear transformations, such as changing scale or origin, creates a new continous random variable Y. The **mean**, **variance**, and **standard deviation** of the continous random variable Y are different compared to X.

From our previous knowledge on discrete random variables, we know that changing the **origin** affects the **mean**, but **not** the **variance**, and changing the **scale** affects the **mean** and the **variance**, and **standard deviation**.

If the continous random variable $\boldsymbol{X}$ has a mean of $\boldsymbol{E(X)}$, variance of $\boldsymbol{\sigma^2}$, and a standard deviation of $\boldsymbol{\sigma}$, then with the application of the variables of linear transofrmation, the random variable $\boldsymbol{aX+b}$ will have a mean of $\boldsymbol{aE(X)+b}$, a variance of $\boldsymbol{a^2\sigma^2}$, and a standard deviation of $|a|\boldsymbol{\sigma}$.
This is summed by the formulas:

Distribution X	**Mean ($E(X)$)**	**Variance ($Var(X)$)**	**Standard Deviation (σ)**
Linear Transformation ($Y = aX + b$)	$\boldsymbol{E(Y) = aE(X) + b}$	$\boldsymbol{Var(Y) = a^2Var(X)}$	$\boldsymbol{\sigma(Y) = \|a\|\sigma(X)}$

Remember that a **change in origin** will result in the **change** of the **mean**, but **not the v ariance**, and a **change in the scale** will result in the **change** of all the **mean**, **variance** and **standard deviation**.

Worked Example 4

Sarah and Jason wanted to prove that linear transformation does changes the original mean and standard deviation. If their original continuous random variable X has a mean of $\boldsymbol{\mu = 16}$ and a standard deviation of $\boldsymbol{\sigma = 4}$, determine their **new mean** and the **standard deviation** of the continuous random variable Y in each of the following situations of linear transformations:

(a) $Y = 3X$ $\boldsymbol{\mu = 3(16) = 48}$ $\boldsymbol{\sigma = 3(4) = 12}$

(b) $Y = X + 3$ $\boldsymbol{\mu = 16 + 3 = 19}$ $\boldsymbol{\sigma = 4}$

(c) $Y = E(2X + 5)$ $\boldsymbol{\mu = 2(16) + 5 = 37}$ $\boldsymbol{\sigma = 2(4) = 8}$

Worked Example 5

After finding the above examples easy for Sarah to do, Jason asked Sarah if she could determine the answers from a given **function** without a given **mean**, **standard deviation** or **variance**.

$$\text{Given } f(x) = \begin{cases} 5x^4, & 0 \le x \le 1 \\ 0, & elsewhere \end{cases}$$

(a) Find the **original mean** and **original variance**.

$$\mu = E(X) = \int x\,(f(x))\,dx$$

$$\mu = \int_0^1 x(5x^4)dx = 5\int_0^1 x^5 dx$$

$$\mu = 5\left[\frac{x^{(5+1)}}{5+1}\right]_0^1 = 5\left[\frac{x^6}{6}\right]_0^1$$

$$\mu = \frac{5}{6}$$

$$\sigma^2 = Var(X) = \int (x-\mu)^2\,(f(x))dx$$

$$\sigma^2 = \int \left(x - \frac{5}{6}\right)^2 (5x^4)dx$$

$$\sigma^2 = 5\int_0^1 \left(x - \frac{5}{6}\right)^2 (x^4)dx$$

$$\sigma^2 = \frac{5}{252}$$

(b) Determine the **exact values** of $E(X+2)$.

$$E(X+2) = E(X) + 2$$

$$E(X+2) = \frac{5}{6} + 2 = \frac{17}{6}$$

(c) Determine the **exact value** of $Var(2X+1)$

$$Var(Y) = a^2 Var(X)$$

$$Var(2X+1) = 2^2\,Var(X)$$

$$Var(2X+1) = 4\left(\frac{5}{252}\right)$$

$$Var(2X+1) = \frac{5}{63}$$

Continuous Random Variables

Continuous random variables are random variables with an **infinite number** of possible values. This is contrasted to **discrete random variables**, which are random variables with a **countable number** of possible values.

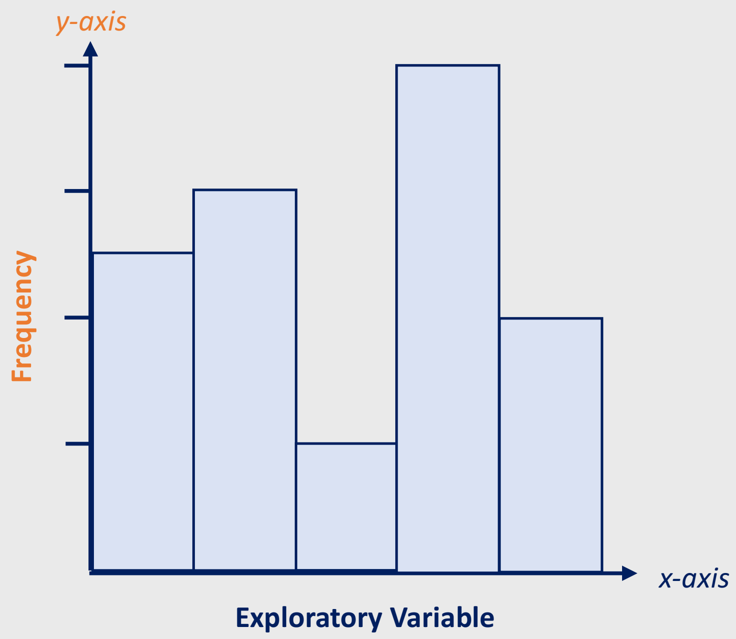
One way to differentiate **continuous random variables** and **discrete random variables**, would be while **discrete random variables** are **countable items**, **continuous random variables** are **measured items**.

Relative Frequency using Histograms

Histograms provide a way to display data that makes it **easy to interpret** and allows us to observe the **relationship between variables**. Histograms can also help us estimate the **probability** of a **particular event** occurring.

> **Relative frequency** is measured as a **ratio** between the **frequency of a certain event occurring** to the **total number of events**.

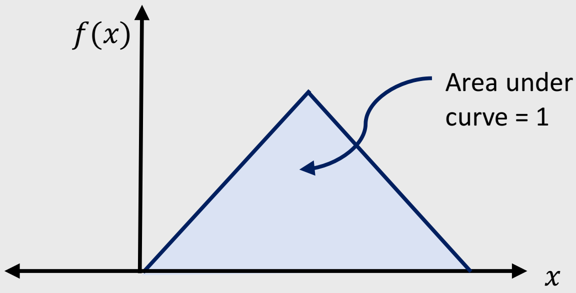
$$Relative\ Frequency = \frac{Number\ of\ times\ this\ event\ occured}{Total\ number\ of\ events}$$

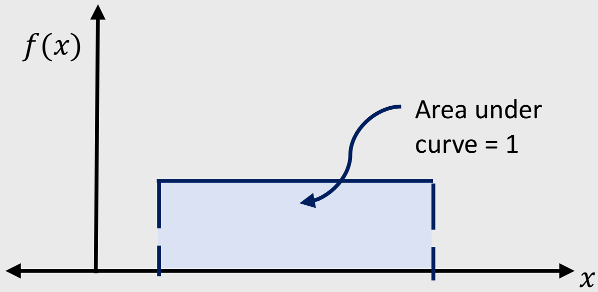


Probability density functions

The **probability distribution function**, written as $f(x)$, gives the probability of each outcome occurring. however, with **continuous random variables**, since there are an **infinite range of possibilities**, it would be more helpful if we looked at the **probability** of **obtaining a value within a range**.

The **probability** of a **continuous random variable** taking a value **in a given interval** is equivalent to the **area under the curve** of the variable's **probability density function** in that interval.





If $f(x)$ is the **probability density function** of a **continuous random variable** X, then the probability that X takes a value in the interval $[a, b]$ is given by $P(a \leq X \leq b) = \int_a^b f(x)dx.$

Probability density functions $f(x)$ have two properties:

Probability Density Function properties: ① $f(x) \geq 0$ and ② $\int_{-\infty}^{\infty} f(x)dx = 1$

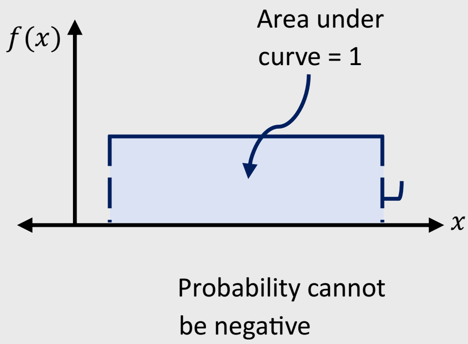
Let's have a look at an example to see what these properties mean.

① $f(x) \geq 0$

② $\int_{-\infty}^{\infty} f(x)dx = 1$

① Probability cannot be **negative**. This means that in a probability density, **values** all must be in

② The total area under the curve of a **probability density function curve** must sum to one.



To **calculate** the **probability** of obtaining a value within a range, the **area under the graph** between the 2 values must be calculated. Both **geometry** and **calculus** are helpful in helping us calculate this!

$$\textit{Area of a Rectangle}: h \times l$$

$$\textit{Area of a Triangle}: \frac{1}{2} \times h \times l$$

$$\textit{Area of a Trapezium}: \frac{h(l_1 + l_2)}{2}$$

$$P(a \leq x \leq b) = \int_a^b f(x)dx$$

The Cumulative Distribution Function

The **cumulative distribution function** of **any random variable** X is defined by the function $F(x) = P(X \leq x)$. The cumulative distribution functions can apply to both **discrete** and **continuous random**

variables. When we are looking at continuous random variables, the **cumulative distribution function** (**denoted by $F(X)$**) and **probability density function** (**denoted by $f(x)$**) have a relationship.

$$F(x) = \int_{-\infty}^{x} f(t)dx \Leftrightarrow f(x) = \frac{d}{dx} F(X)$$

For any random variable X with a cumulative distribution function $F(x)$ and for any a, b in the event space where $a < b$, the following relationship is always true.

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

This relationship leads to some important properties of the **cumulative distribution function**.

Cumulative Distribution Function Properties
The cumulative distribution function is **non-decreasing**.
If x approaches to a negative infinity, $F(x)$ approaches to **0.**
If x approaches to a positive infinity, $F(x)$ approaches to **1.**

These three properties describe the shape of cumulative distribution functions. It always has a minimum value **0** and a maximum value **1**. If the probability distribution is **0** for all values below $x = a$, then $F(x) = 0$ when $x \leq a$. Similarly, if the probability distribution is **0** above all values of $x = b$, then $F(x) = 1$ when $x \geq b$. For any **a**, **b** in the domain and $a < b$, $F(a) < (b)$.

Expected Value, Variance, and Standard Deviation

Continuous random variables can be defined by calculating and interpreting its parameters by the **expected value**, **variance**, and **standard deviation**.

Mean: $$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Variance: $$Var(X) = \sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

Variance: $$Var(X) = E(X^2) - \mu^2$$

Standard Deviation: $$\sigma = \sqrt{Var(X)}$$

Linear transformations of Continuous Random Variables

The distribution of a continuous random variables can be **changed** by applying **addition**, **subtraction**, **multiplication** or **division** to the random variable.

Linear transformations are **changes** to a variable characterised by **multiplying** the random variable by a **constant** (a) and/or by **adding** a **constant** (b) to the random variable.

Though there are four different linear transformations that can be applied, these changes are categorised into either a **change of scale** or a **change of origin**. Change of **scale** and change of **origin** are defined as:

A **change of scale** is the **stretching/contracting** of a random variable by **multiplying** it by a **constant** (a).

A **change of origin** is the **shifting** of each value of a random variable by **adding** a **constant** (b).

Distribution X	Mean ($E(X)$)	Variance ($Var(X)$)	Standard Deviation (σ)
Linear Transformation ($Y = aX + b$)	$E(Y) = aE(X) + b$	$Var(Y) = a^2Var(X)$	$\sigma(Y) = \lvert a \rvert \sigma(X)$

Remember that a **change in origin** will result in the **change** of the **mean** but **not the variance** and a **change in the scale** will result in the **change** of all the **mean**, **variance** and **standard deviation**.

Problem Set 11 – Continuous Random Variables
Progressive Questions

Concept 1
Continuous Random Variables – Progressive Questions

(8 questions)
Repetitive questions: 1.11 – 1.91

Exploring Peru

Starring: Teacher Andrew, Janitor Peter, the Maths Methods students

---

After the student's massive win at the International Mathematics Championship, the students decide to spend a few days exploring Peru before they head back home. Of course, Janitor Peter couldn't not take this opportunity to teach the students some more about maths, so he decided to make this a competition between teams in the maths class. Whoever answers all the questions right first wins!

Continuous Random Variables: Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9
Repetitive: 1.11 – 1.92

[9 marks]

1. Before the students embark on their travels, help them answer the following questions so they can form a basic understanding! Explain why the following examples are either **discrete** or **continuous random variables**:

(a) The **height** of each student in a class of 30. **(1)**

(b) The **number of heads** after flipping 8 coins. **(1)**

(c) T: the **time** it takes for a seed to germinate **(1)**

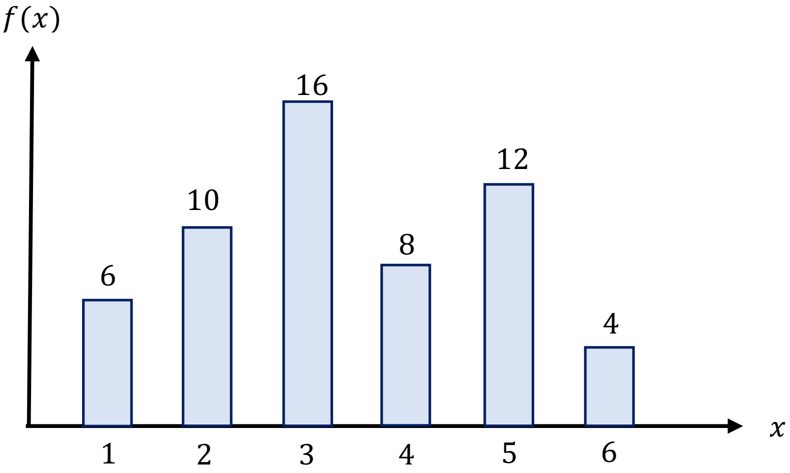
(d) **(1)**

t	$t < 7$	$7 \leq t$	$10 \leq t$	$13 \leq t$	$16 \leq t$	$19 \leq t < 22$	$t \geq 22$
Number calculators	3	7	10	13	16	11	7

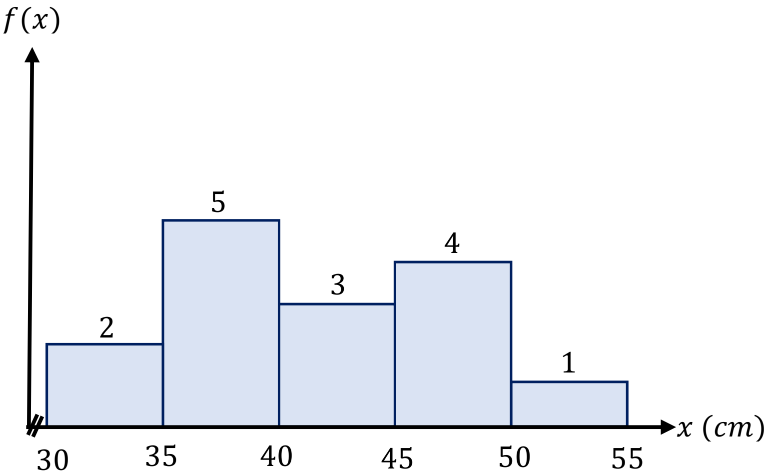
(e) **(1)**

x	0	1	2	3	4	5
Number of textbooks	4	9	1	16	2	3

(f) **(1)**



(g) **(1)**



(h) $f(x) = \begin{cases} \frac{1}{8}x, & 0 \leq x \leq 4 \\ 0, & x < 0\,,\ \ x > 4 \end{cases}$ **(1)**

(i) $f(x) = \begin{cases} \frac{1}{6}x, & x = 0, 1, 2, 3 \\ 0, & x < \ 0, x > 3 \end{cases}$ **(1)**

[12 marks]

2. The first stop on their tour is a mountainside with wild giant alpacas roaming free. Alexa and Tyler arrive first and are tasked with exploring the traits of giant alpacas.

(a) Given the **continuous random variable** H is defined as the height of the giant alpacas, find the **expected value** and **variance** of H. **(3)**

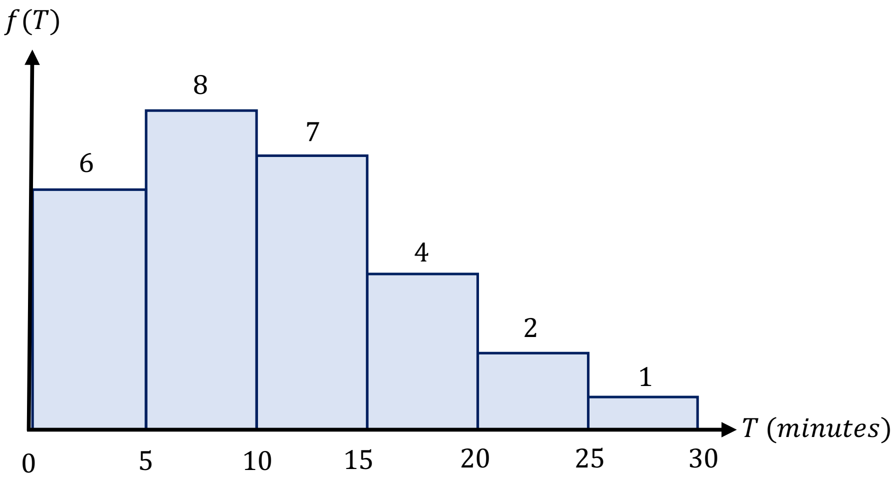
h	$150 \leq h < 160$	$160 \leq h < 170$	$170 \leq h < 180$	$180 \leq h < 190$	$190 \leq h < 200$
Number of Students	**4**	**7**	**12**	**4**	**3**

(b) The **probability mass function** for the **time** taken for an alpaca to **run 100 metres** is shown below.

(i) Find $P(X > 12)$. **(1)**

(ii) If $P(X < k) = 0.8$, find k **(3)**

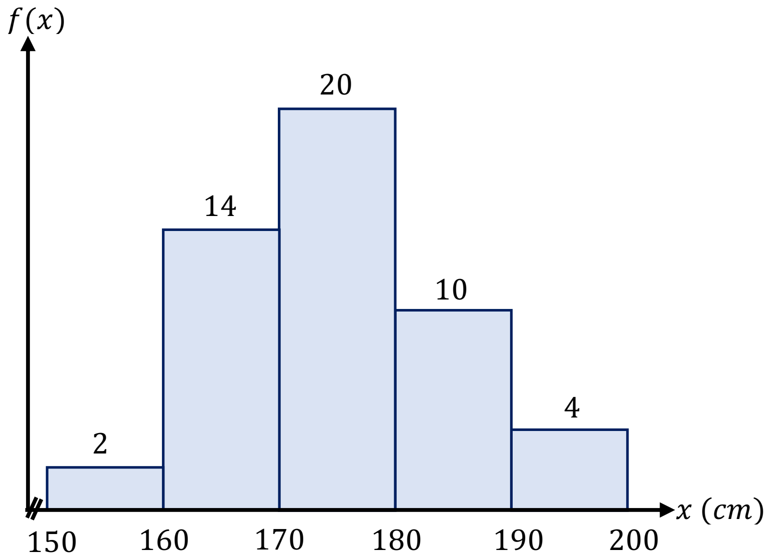
(iii) State what k represents. **(1)**



(c) The following probability mass function represents the distance an alpaca can spit. Find:

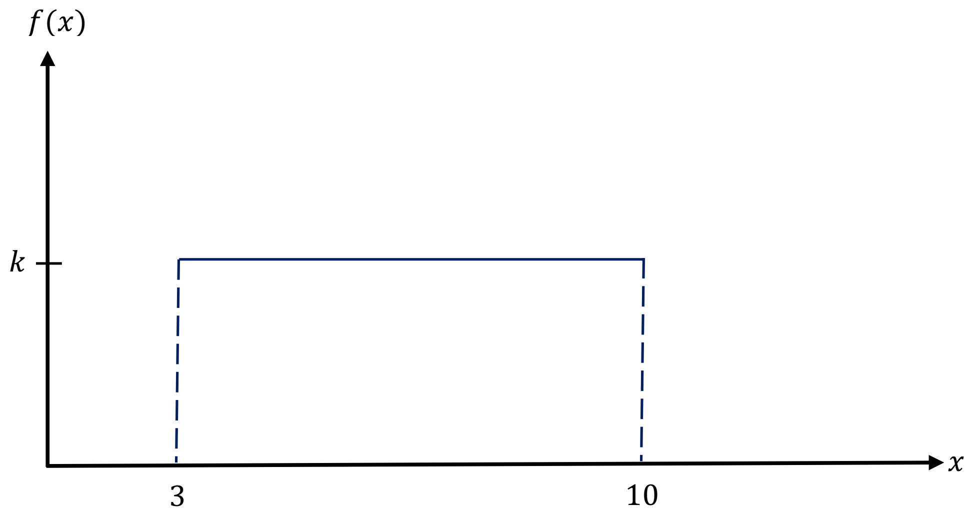
(i) the **expected value** and **standard deviation** **(3)**

(ii) $P(E(X) - \sigma < X < E(X) + \sigma)$. **(3)**



[10 marks]

3. Alexa and Tyler have been overtaken by Shauna and Kerry! To arrive at their next stop, they must wait for a bus, with the waiting times shown by the following **uniform probability distribution**. A **uniform probability distribution** of a **continuous random variable** X is shown below.



(a) Find the **value of** k **(1)**

(b) Find the **probability** that the bus takes between 5 and 7 minutes **(1)**

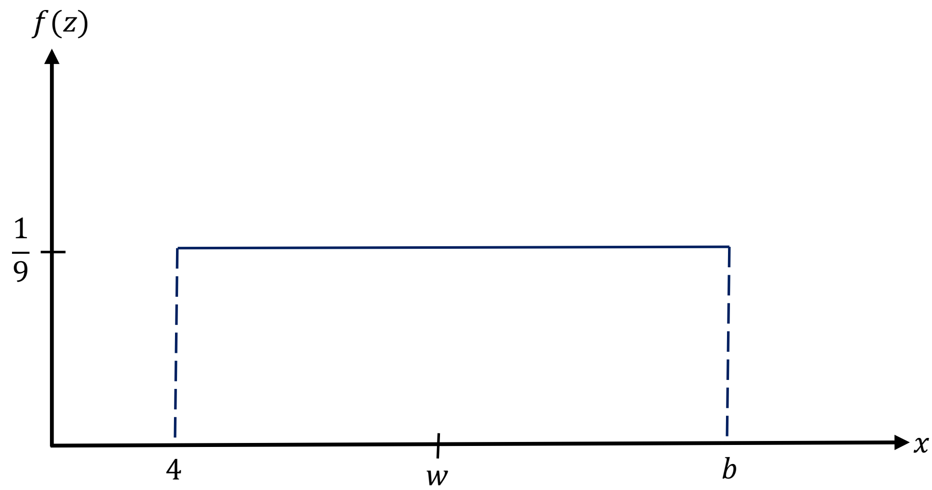
(c) Find the **probability** that the bus takes between 1 and 6 minutes **(1)**

(d) Find the **probability** that this **continuous random variable** takes a value of **exactly 4** **(1)**

(e) Find the **expected time** the students must wait for the bus **(1)**

(f) Find the **standard deviation** of the waiting time for the bus **(1)**

Team Harriet and Tom decide to shortcut the waiting time and catch a train. Another **uniform probability distribution** of a **continuous random variable** Z, exists below. This time, it shows the average wait time for the train.



(g) Find the **value** of b **(1)**

(h) Find $Var(Z)$ **(2)**

(i) If $P(w \leq Z \leq 14) = \frac{2}{3}$, Find the **value** of w **(1)**

[8 marks]

4. Bored at the bus stop, Jevon convinces his partner Rabea to practice their maths so they will be quicker later! Rabea reluctantly agrees. Help Rabea find the maximum point for the following **probability density functions**

(a) A **binomial distribution** where $n = 8$ and $p = 0.25$ **(2)**

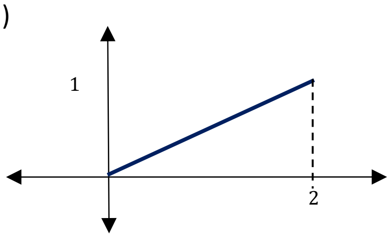
(b) A **normally distributed pdf** where 68% of values lie between 2 and 5 **(3)**

(c) A **normally distributed pdf** where 95% of values lie between 3.8 and 11.2 **(3)**

[16 marks]

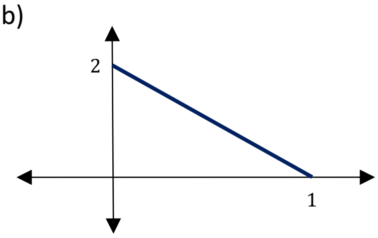
5. After a lot of travelling, the teams reach Machu Picchu! They are tasked to model the total time tourists spent at the popular destination using **cumulative distribution functions**. Find and draw the **cumulative distribution functions** for the following **probability density function** graphs

(a)



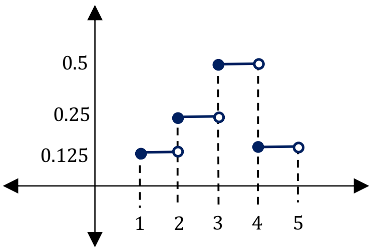
(3)

(b)



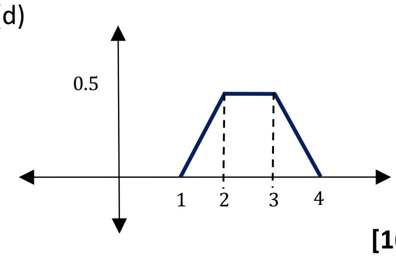
(3)

(c)



(5)

(d)

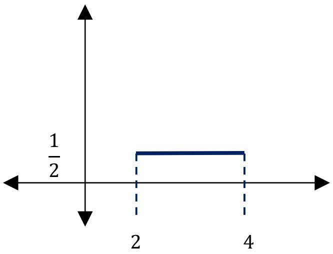


(5)

[16 marks]

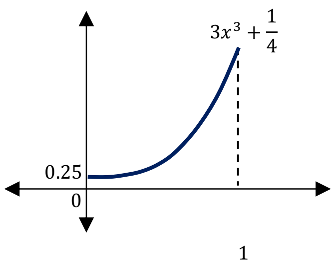
6. While waiting in line to hike to Machu Pichu, Tyler decided to make some probability density graphs. Find the **expected value**, **standard deviation**, and **variance** for the following **probability density function** graphs.

(a)



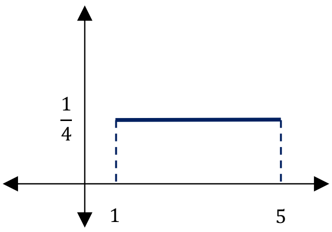
(4)

(c)



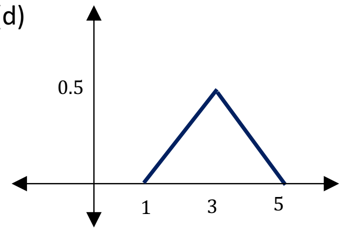
(4)

(b)



(4)

(d)



(4)

[14 marks]

7. The tour guide tells the students about different facts regarding the hike. Find the **expected value**, **standard deviation**, and **variance** from the following **probability density functions**.

(a) Where X is the time taken to hike Machu Pichu

$$f(x) = \begin{cases} \frac{1}{6}, & 3 \le x \le 9 \\ 0, & for\ \ all\ other\ values\ of\ x \end{cases}$$ **(3)**

(b)

$$f(x) = \begin{cases} 3(x-1)\text{^}2, & 0 \le x \le 1 \\ 0, & for\ \ all\ other\ values\ of\ x \end{cases}$$ **(3)**

(c)

$$f(x) = \begin{cases} \frac{7}{16}, & 3 \le x < 4 \\ \frac{1}{8}x - \frac{1}{2}, & 4 \le x \le 7 \\ 0, & for\ \ all\ other\ values\ of\ x \end{cases}$$ **(4)**

(d)

$$f(x) = \begin{cases} \frac{1}{16}x - \frac{1}{16}, & 1 \le x < 5 \\ \frac{9}{16} - \frac{1}{16}x, & 5 \le x \le 9 \\ 0, & for\ \ all\ other\ values\ of\ x \end{cases}$$ **(4)**

[10 marks]

8. The **satisfaction of the students** is defined by a continuous random variable $\boldsymbol{X}$ which is given by the following probability distribution function:

$$f(x) = \begin{cases} 4x^3 & for\ 0 \le x \le 1 \\ 0 & otherwise \end{cases}$$

(a) Find the **mean, μ_X** **(2)**

(b) Find **standard deviation, σ_X** **(3)**

(c) If $\boldsymbol{Y = X + 2}$, find the **mean** and **standard deviation** of Y **(2)**

(d) If $\boldsymbol{Y = 3X + 2}$, find the **mean** and **standard deviation** of Y **(3)**

[14 marks]

9. To close out their trip to Peru, Janitor Peter decides to test the students with one final question. Given the probability density function:

$$\begin{cases} k \cdot x(5-x), & for\ 0 \le x \le 5 \\ 0, & otherwise \end{cases}$$

(a) Find the value of $\boldsymbol{k}$ **(2)**

(b) Find the **probability** that x is **greater than 3** **(3)**

(c) Find **expected value** **(2)**

(d) Find **variance** and **standard deviation** **(4)**

(e) If $\boldsymbol{Y = 2X + 5}$ find the new **expected value** and **standard deviation** **(3)**

Problem Set 11 – Continuous Random Variables
Repetitive Questions

Concept 1
Continuous Random Variables – Repetitive Questions
(12 questions)

Continuous Random Variables:

[9 marks]

1.11 Determine if the following are discrete, or **continuous random variables**. Explain your answer.

(a) The **number** of red marbles in a bag. **(1)**

(b) The **time** it takes to run 2km. **(1)**

(c) The **length** of each horn from 20 goats. **(1)**

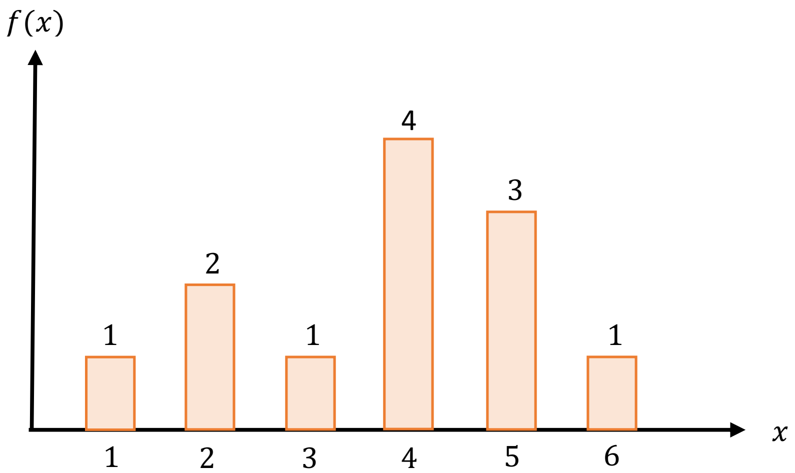
(d) **(1)**

t	0	1	2	3	4	5
Number of goals scored	5	7	2	9	3	1

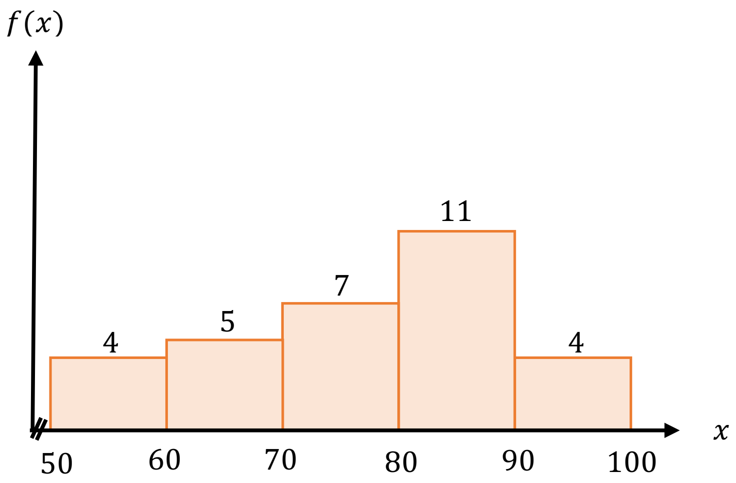
(e) **(1)**

t	$0 \le t < 50$	$50 \le t < 60$	$60 \le t < 70$	$70 \le t < 80$	$80 \le t < 90$	$90 \le t \le 100$
Number of test scores	5	10	15	10	5	1

(f) **(1)**



(g) **(1)**



(h) **(1)**

$$f(x) = \begin{cases} \frac{1}{12}x, & 0 \le x \le 6 \\ 0, & x < 0\,,\ \ x > 6 \end{cases}$$

(i) **(1)**

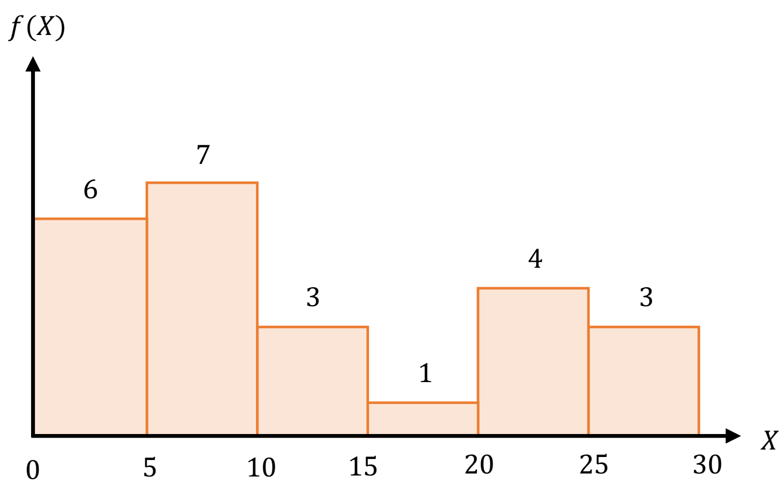
$$f(x) = \begin{cases} \frac{1}{8}x, & x = 0, 1, 2, 3,4 \\ 0, & x < \ 0, x > 4 \end{cases}$$

[12 marks]

1.21 Solve the following:

z	$0 \le z < 20$	$20 \le z < 40$	$40 \le z < 60$	$60 \le z < 80$	$80 \le z < 100$
y	11	2	8	9	4

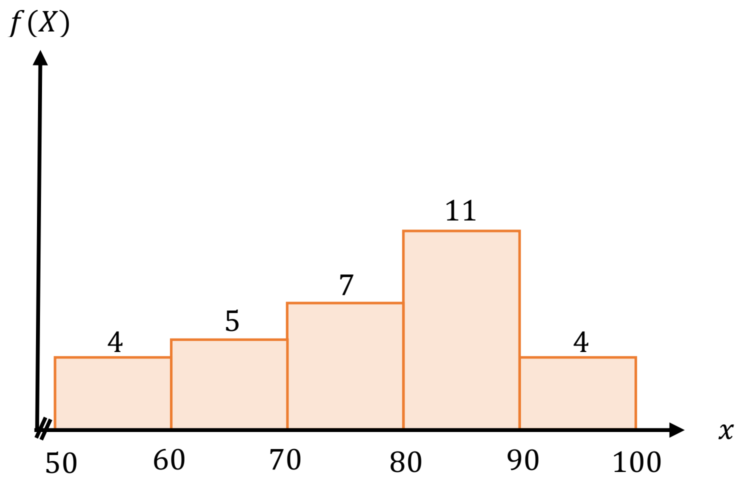
(a) Find the **expected value** and variance of z. **(1+2)**



(b)

i. Find $P(X > 14)$. **(1)**

ii. If $P(X > k) = 0.8$, find k. **(3)**



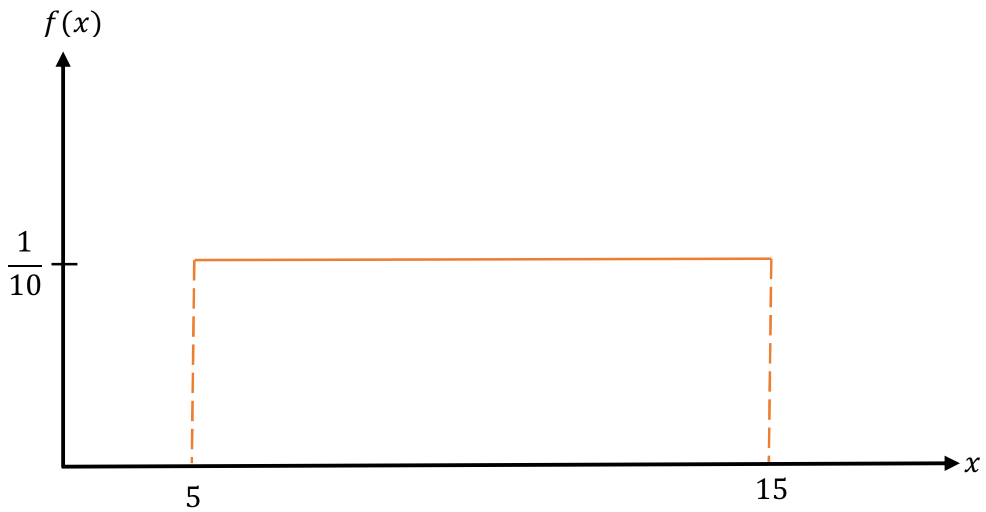
(c)

i. Find the **expected value** and **variance** of x. **(2)**

ii. Then find $P(E(X) - \sigma < X < E(X) + \sigma)$. **(3)**

[3 marks]

1.31 A **uniform probability distribution** of a **continuous random variable** X exists below



(a) Write the **probability density function** of this **distribution** **(1)**

(b) Write the **cumulative distribution function** of this **distribution** **(1)**

1.32 **[6 marks]**

(a) Find the **value** of **b**. **(1)**
(b) Find the **expected value** and **standard deviation** of **X**. **(3)**
(c) If $P(w < X < b - 2) = 0.5$, find w. **(1)**
(d) Find $P(X < w)$. **(1)**

[7 marks]

1.41 The **continuous random variable X** is defined by the **probability distribution**

$f(x) = \frac{1}{9}x^2, \ 0 \leq x \leq k$

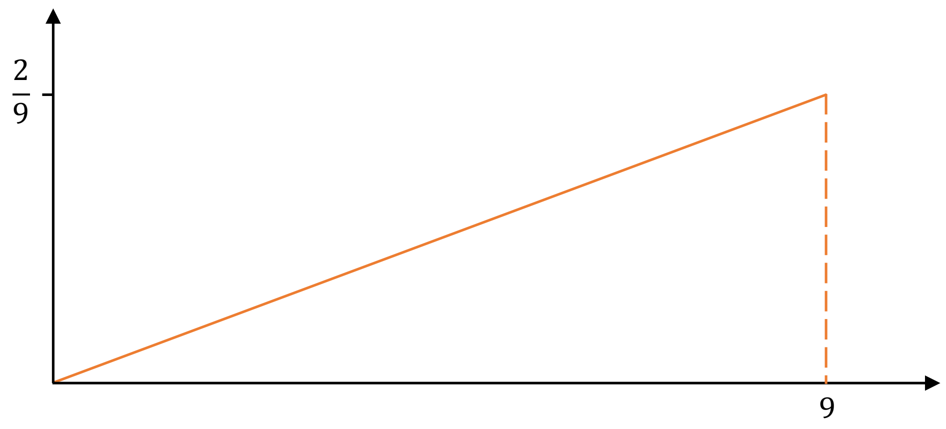
(a) Find the **value** of k. **(2)**
(b) Find $P(2 < X < 3)$ **(2)**
(c) Find $P(2.5 < X \mid X < 3)$ **(3)**

[8 marks]

1.42 Find the **probability** that the following **continuous random variables** with the shown **probability density functions** will take a value in the **specified range.**

(a) $P(3 < X < 5)$ **(1)**

(b) $P(3 \le X \le 6)$ **(2)**



(c) $P(49 < X < 81)$ for the PDF with **equation** $f(x) = \frac{1}{64}$ for $36 < x < 100$ and 0 elsewhere. **(1)**

(d) $P(0 \le X \le \frac{1}{8})$ for the PDF with **equation** $f(x) = 5 - 8x$ for $0 \le x \le \frac{1}{4}$ and 0 elsewhere. **(2)**

(e) $P(\frac{1}{16} \le X \le \frac{1}{4})$ for the PDF with **equation** $f(x) = e^{3x}$ for $0 \le x \le \frac{2}{3}\ln(2)$ and 0 elsewhere. **(2)**

[28 marks]

1.51 From the probability density function and parameters provided; find the **cumulative distribution function** for:

(a) $f(x) = \frac{1}{4}$ when $2 \le x \le 6$ and 0 elsewhere **(2)**

(b) $f(x) = \frac{1}{2}x$ when $0 \le x \le 2$ and 0 elsewhere **(2)**

(c) $f(x) = e^x$ when $0 \le x \le \ln(2)$ and 0 elsewhere **(2)**

By finding $\boldsymbol{k}$ first, find the **cumulative distribution function** for the following probability density functions for:

(d) $f(x) = k$ when $3 \le x \le 9$ and 0 elsewhere **(3)**

(e) $f(x) = kx + 2$ when $2 \le x \le 5$ and 0 elsewhere **(4)**

(f) $f(x) = k(\pi^{2x})$ when $2 \le x \le 3$ and 0 elsewhere **(5)**

Prove the following:

(g) For any continuous uniform distribution, $f(x) = k$, its **cumulative distribution function** is $F(x) = \frac{x-a}{b-a}$, where a and b are the lowest and highest values the random variable can take respectively. **(3)**

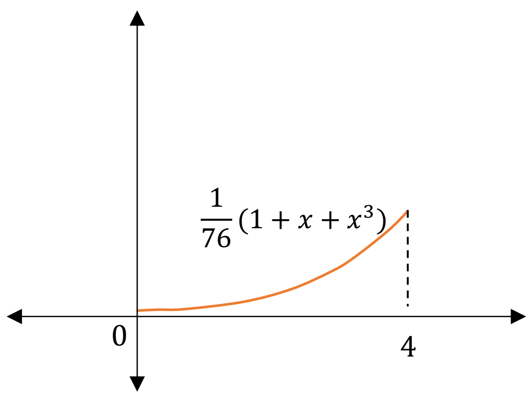
(h) For any continuous linear distribution, $f(x) = mx + c$, its **cumulative distribution function** is $F(x) = (x-a)(\frac{m(x+a)+2c}{2})$ where a is the smallest value the random variable can take. **(4)**

(i) For any **continuous exponential distribution**, $f(x) = c\left(k^{\lambda x}\right)$, its **cumulative distribution function** is $F(x) = \frac{c}{\lambda \ln(k)}\left(k^{x\lambda} - k^{a\lambda}\right)$ where a is the smallest value the random variable can take. **(3)**

[8 marks]

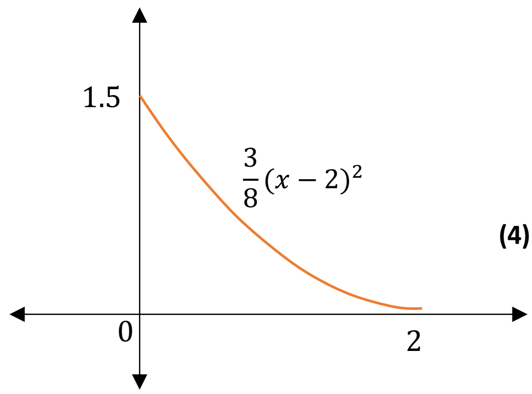
1.61 Find the **expected value**, **standard deviation**, and **variance** for the following **probability density function** graphs.

(a)



(4)

(b)



(4)

[8 marks]

1.62 Find the **expected value**, **standard deviation**, and **variance** from the following **probability density functions**.

(a)

$$f(x) = \begin{cases} \frac{2}{6}, & 3 \geq x \leq 4 \\ \frac{2}{6}, & 6 \geq x \leq 8 \\ 0, & for\ \ all\ other\ values\ of\ x \end{cases}$$ **(4)**

(b)

$$f(x) = \begin{cases} \frac{1}{15}(8 - x^2), & 0 \leq x \leq 3 \\ 0, & for\ \ all\ other\ values\ of\ x \end{cases}$$ **(4)**

[9 marks]

1.71 A continuous random variable has the following probability distribution function.

$$\begin{cases} k \cdot x(2-x), & for\ 0 \leq x \leq 1 \\ 0, & otherwise \end{cases}$$

(a) Find the value of $\boldsymbol{k}$ **(2)**

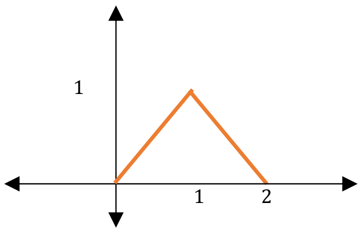
(b) Find the **mean, $\boldsymbol{\mu_X}$** **(2)**

(c) Find **standard deviation, $\boldsymbol{\sigma_X}$** **(4)**

(d) If $\boldsymbol{Y = 5X + 1}$, find the **mean** and **standard deviation** of $\boldsymbol{Y}$ **(2)**

[11 marks]

1.81 A continuous random variable has a probability density function shown below.



(a) Derive an expression for the probability density function **(1)**

(b) Find the **probability** that X is **between 0.3 and 1.8** **(2)**

(c) Find the **mean, μ_X** **(2)**

(d) Find **standard deviation, σ_X** **(3)**

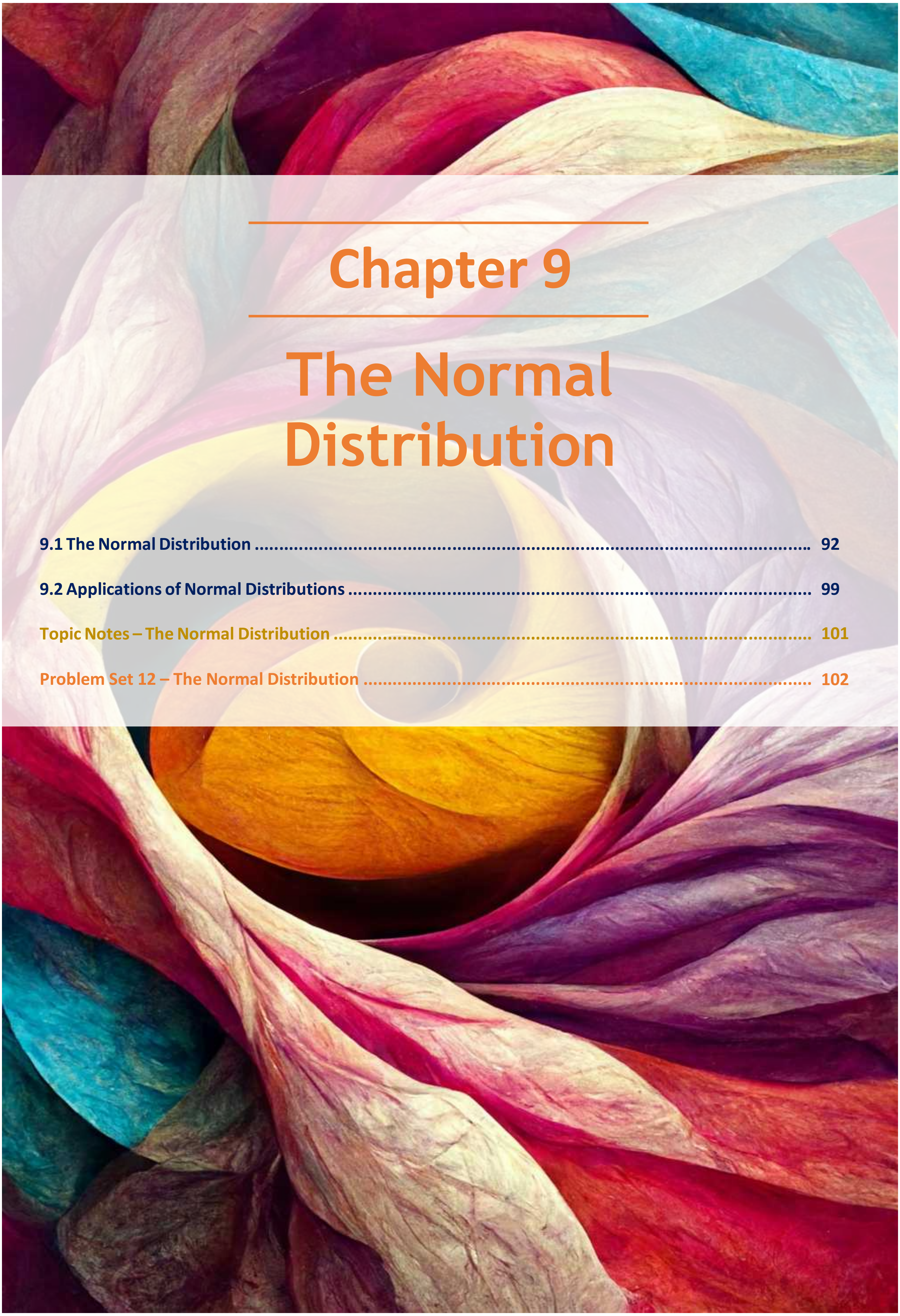
(e) If $Y = 2X + 7$, find the **mean**, **standard deviation,** and **variance** of Y **(3)**

[10 marks]

1.91 **Prove** the following for the **uniform probability distribution** $f(x) = \frac{1}{k}$ (where a and b are the **lowest** and **highest** bounds of the function).

(a) $E(X) = \frac{a+b}{2}$ **(4)**

(b) $Var(X) = \frac{(b-a)^2}{12}$ [**Hint:** $b^3 - a^3 = (b^2 + a^2 + ab)(b - a)$] **(6)**



Chapter 9

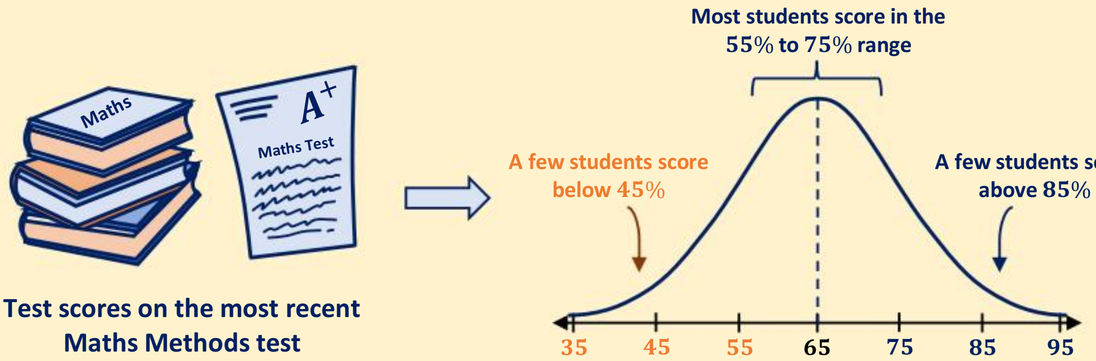
The Normal Distribution

Chapter 9 – Normal Distribution

Introduction

Our next stop on our probability journey is exploring **normal distributions**! A **normal distribution** is a **probability distribution** that is **symmetric** around its **mean**. It is often referred to as a **bell curve** because it has a **characteristic bell shape**.

A **normal distribution** might be used to describe the **distribution of the test scores** for a cohort of Year 12 students. In your Maths Methods cohort, the teachers might want to create a test where the **mean score is 65%** and a bell curve is created around this mean. So, a few students will **score exceptionally well**, a few students will **score below 45%**, but **most of the students** will score between the **55% to 75% range**.



To explore normal distributions, we have broken this chapter up into **three main concepts**:

1. **Fundamental Normal Distributions Concepts**
2. **68%, 95% and 99.7% Rule**
3. **Applications of Normal Distributions**

Let's Begin!

9.1 The Normal Distribution

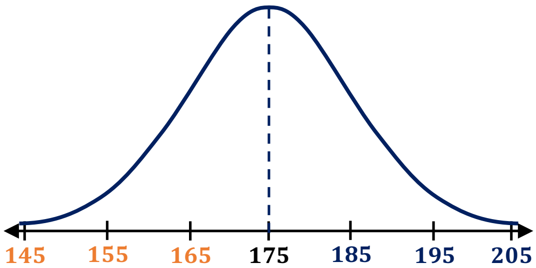
A **normal distribution** is a type of **probability distribution** that is **symmetric** around its **mean** and it often **characterised by its bell shaped curve**. We write them as $X \sim N(\mu, \sigma^2)$.

Normal distributions are always the **exact same shape** and can be found in an **endless number** of **real-life scenarios**. In the context **of Year 12 alone**, normal distributions will be found in the **heights of Year 12 students**, the distribution of **ATAR results**, the **weight** of the backpacks you carry to school, the **times students run** in the **$100m$ race**, and the **scores** on any **school test**.

To explore one of these examples further, the **heights** of all the **Year 12 students** is likely to be **approximately normal**, with a **mean** of $\mu = 175cm$ and a **standard deviation** of a $\sigma = 10cm$.

Using the **normal distribution notation**, this distribution would be written as: $X \sim N(175, 10^2)$ and the distribution would be plotted as follows:

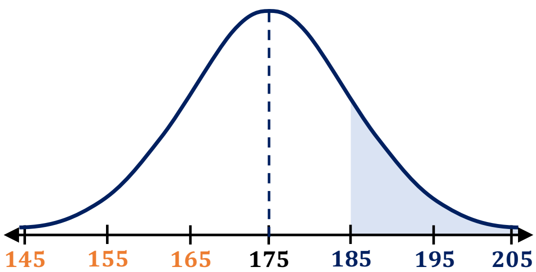
Distribution of Year 12 student heights $X \sim N(175, 10^2)$

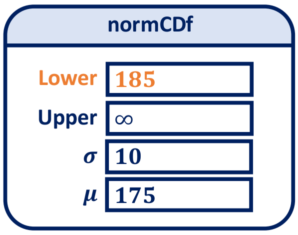


With a **normal distribution** such as the one above, we can use it to determine important information about a **population**.

For instance, if we wanted to determine the **probability** that a **Year 12 student** has a **height above $185cm$**, we could use the **'normCDf' function** on our **calculators**. This function can be found on the **main section** of your calculator, then selecting **'Interactive'**, then **'Distribution/Inv. Dist'**, then **'Continuous'** and then **'normCDf'**. By substituting in the values: $lower = 185$, $upper = \infty$, $\mu = 175$ and $\sigma = 10$, we can solve for $P(X > 185)$:

If $X \sim N(175, 10^2)$ determine $P(X > 185)$





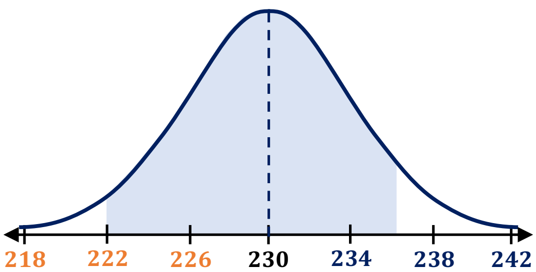
Solving on **calculator** gives:

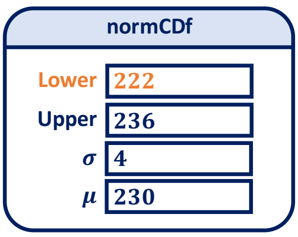
$P(X > 185) = 0.159$

$\therefore$ **15.9% chance of student being taller than $185cm$**

As a second example, suppose an **apple orchard grew apples** with a **mean weight** of **230 grams** and a **standard deviation** of **4 grams**. The weight distribution of the apples could be **approximated to be normal**, where: $X \sim N(230, 4^2)$. If the grocery stores **won't accept** apples **below a weight of $222g$** or **above a weight of $236g$**, we could **determine** the **probability** that one of our **apples won't be accepted** as follows:

If $X \sim N(230, 4^2)$ determine $P(222 < X < 236)$





Solving on **calculator** gives:

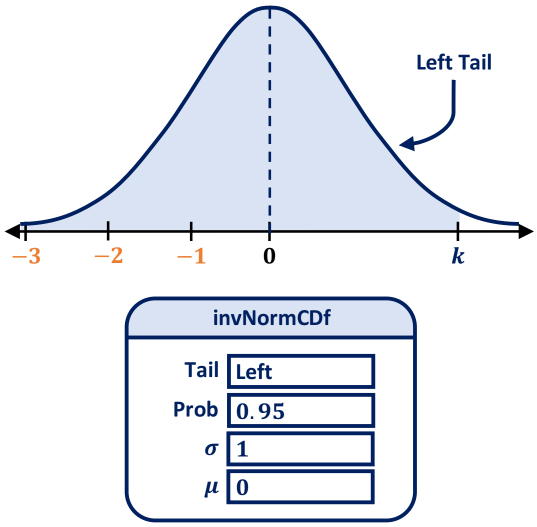
$P(222 < X < 236) = 0.910$

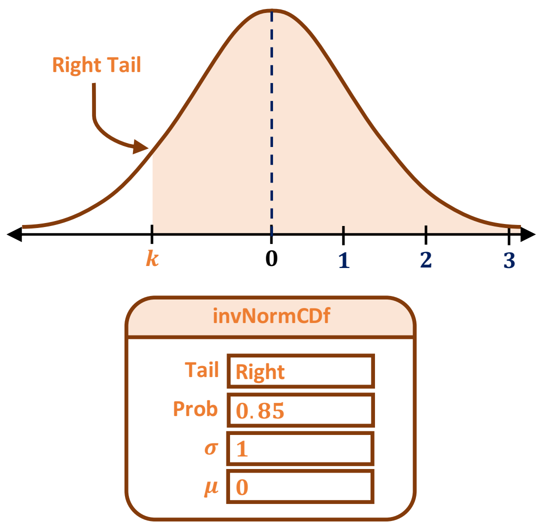
$\therefore$ **9% chance apple gets rejected**

In an alternate scenario, suppose we were **given the proportion** but wanted to know the **value of x** for this proportion. For instance, maybe we wanted to know the **mass** that **94%** of the apples in our apple orchard **were above** or the **score** in a **physics test** that **65%** of Year 12 students **scored below**.

A **second function** that we can utilise in instances like these is the **invNormCDf function**, which can again be found on the **main section** of your calculator by selecting **'Interactive'**, then **'Distribution/Inv. Dist'**, then **'Inverse'** and then **'invNormCDf'**. For this function, it is important to understand that **'tail'** means the **side of the curve that is shaded**.

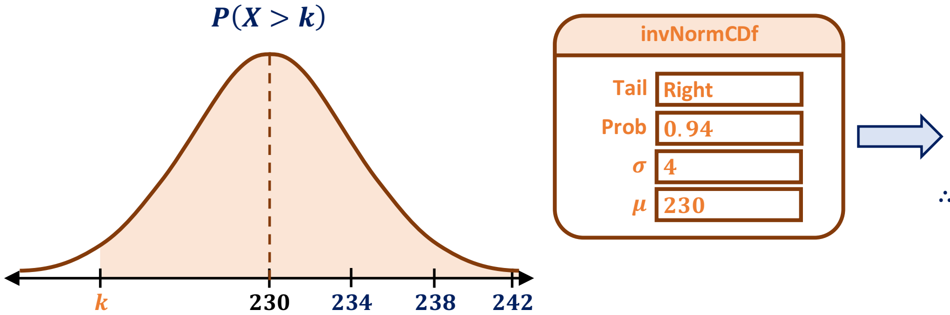
For instance, for a **probability of 95%** with the **left tail selected**, this means it is considered that **0.95** is **shaded to the left**. In contrast, **85% to the right** means that the **shaded area is to the right**. The **'tail' setting** that **you select** is **pivotal** because it will determine the **value of 'k'** that is **outputted** by the function, as seen below.





Returning to our apple orchard example where $X \sim N(230, 4^2)$. Let's suppose that we wanted to know the **mass that 94% of all apples** will **be above**. To determine this value, we would **substitute** in: $tail = right$, $prob = 0.94$, $\sigma = 4$ and $\mu = 230$, as follows:

If $X \sim N(230, 4^2)$ determine
$P(X > k)$



Solving $P(X > k) = 0.94$
on **calculator** gives:
$k = 223.8g$
$\therefore$ **94% apples are above $223.8g$**

Another common context through which we use the **invNormCDf function** is when we are trying to determine **quantiles**.

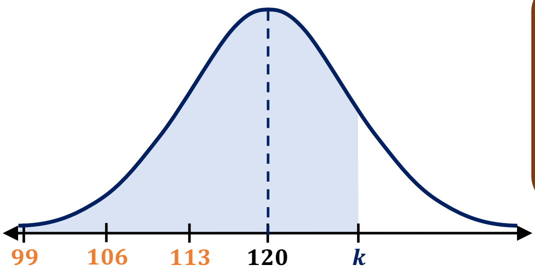
Quantiles

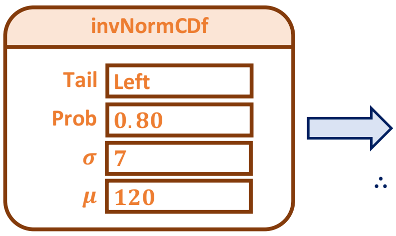
Quantiles are **values** that a set fraction of the population fall below. For instance, the $\mathbf{0.5}$ **quantile** of a **normal distribution** represents the **value** at which **50% of the data falls below**.

Quantiles represent the position of one observation under a distribution and is used in an endless number of **day-to-day situations**. Your **ATAR score** is a **quantile**, where **an 85 ATAR represents** that you **scored above 85%** of the students in Year 12. In medicine, many of your **medical records** such as **blood pressure** and **cholesterol** are **measured in percentiles**. To determine the **quantile** value of an observation, we can use the **invNormCDf function**.

For instance, suppose that a **Year 12 cohort** of **Maths Methods students** all competed in an **Australia-wide mathematics competition** where the scores were normally distributed with the **mean** of the test being $\mathbf{120}$ and the **standard deviation** being $\mathbf{7}$. If the teachers wanted to **determine the scores** required to be in the $\mathbf{80^{th}}$ **percentile** they could do this by using the **invNormCDf function** as follows:

If $X \sim N(120, 7^2)$ determine
$P(X < k) = 0.80$





Solving $P(X < k) = 0.80$
on **calculator** gives:

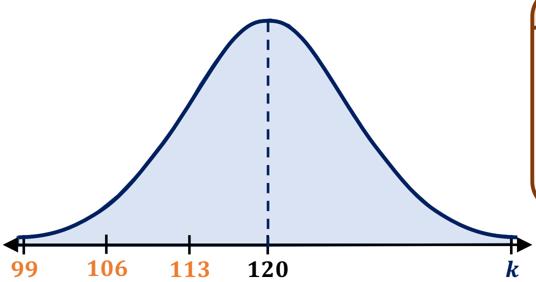
$k = 125.9$

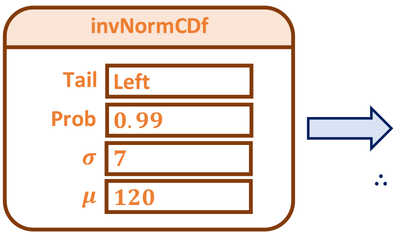
$\therefore$ **80% of students score below 125.9**

As we can see, the Year 12 Maths students will need to **achieve a score greater than 125.9** in order to score in the $\mathbf{80^{th}}$ **percentile of Australia**.

For the high achieving students, if the teachers then wanted to know the **score students require** to get in the $\mathbf{99^{th}}$ **percentile**, they can **repeat the same process** but change the **probability** to $\mathbf{0.99}$ and calculate that students must **achieve a score greater than 136.3** to score in the $\mathbf{99^{th}}$ **percentile of Australia**.

If $X \sim N(120, 7^2)$ determine
$P(X < k) = 0.99$





Solving $P(X < k) = 0.99$
on **calculator** gives:

$k = 136.9$

$\therefore$ **99% of students score below 136.3**

To consolidate all the **foundational understandings** of **normal distributions**, you can work through the **example below**.

Worked Example 1

As one of Teacher Andrew's unique hobbies, he has set up a **bean production company** that **sells canned beans**. These cans of beans are **normally distributed** with a **mean** of $220g$ and a **standard deviation** of $6g$.

(a) State the **notation** for Teacher Andrew's normal distribution.

$$X \sim N(220, 6^2)$$

(b) Determine the **probability** that a randomly selected can contains **less than $216g$** of beans.

To determine $P(X < 216)$ we use the **normCDf function** as follows:

normCDf	
Lower	0
Upper	216
σ	6
μ	220

Solving on **calculator** gives:

$P(X < 216) = 0.252$

$\therefore$ 25.2% chance of a can containing less than $216g$

(c) If cans are **rejected** when they contain **less than $210g$ of beans** or **more than $230g$ of beans**, how many cans will be approximately **accepted** in a random sample of **200 cans of beans**.

To start, det/ermine $P(210 < X < 230)$ using the **normCDf function**:

normCDf	
Lower	210
Upper	230
σ	6
μ	220

Solving on **calculator** gives:

$P(210 < X < 230) = 0.904$

$$\therefore Cans\ accepted = 200 \times 0.904$$

$$\therefore Cans\ accepted = 181\ cans$$

(d) Finally, if Teacher Andrew wants to know the **weight that 98%** of his cans of beans are **below**. Help Teacher Andrew to **determine this weight**.

invNormCDf	
Tail	Left
Prob	0.98
σ	6
μ	220

Solving $P(X < k) = 0.98$ on **calculator** gives:

$k = 232.3$

$\therefore$ 98% of cans have a weight below $232.3g$

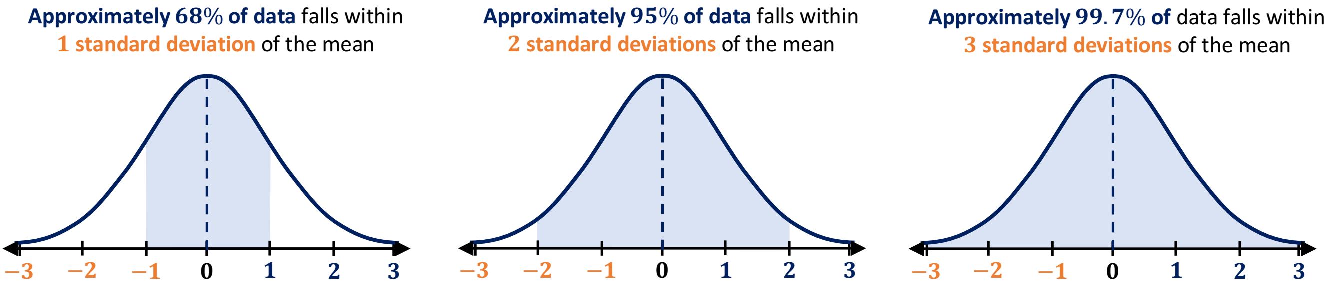
Beyond using our calculators, another way that we can **approximate** the results of **normal distributions** is through applying an **approximation rule** known as the **68%, 95%, 99.7% rule**, which we will explore in the next section.

Approximate Spread of Normal Distributions – 68%, 95%, 99.7% Rule

In a normal distribution, **approximately 68% of the data** falls within **one standard deviation** of the mean, **95% of the data** falls within **two standard deviations** of the mean, and **99.7% of the data** falls within **three standard deviations** of the mean.

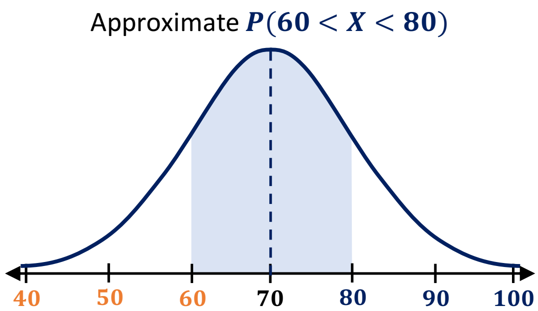
This rule is helpful because it allows us to quickly get a sense of **how spread out the data** is.

This rule is **best summarised** in the diagram below.



We can apply the **68%, 95%, 99.7% rule** to any of the normal distributions we have already explored to **approximate key values** about the distribution. For instance, if the teacher sets a maths test with a **mean score** of 70% and **standard deviation** of 10%, we can use the **68%, 95%, 99.7% rule** to **approximate key values** in relation to the Year 12 students' scores.

If we wanted to **approximate** the **proportion of students** that **scored between 60% and 80%**, we could approximate it to be 68% since this is within **one standard deviation either side of the mean**.



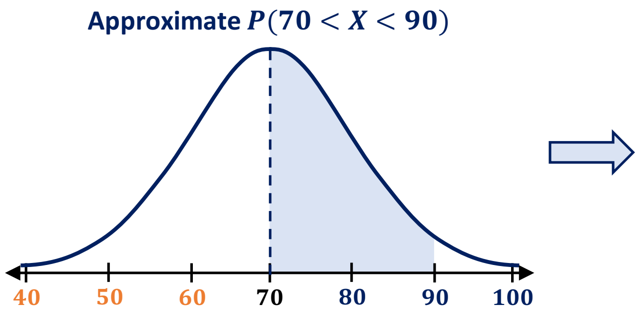
Applying the 68%, 95%, 99.7% rule:

60% and 80% are both **one standard deviation either side of the mean**

$$P(60 < X < 80) \approx 0.68$$

$\therefore$ **68% students score between 60% and 80%**

Taking it up a notch, if the teacher wanted to know the **proportion of students** that **scored between 70% and 90%**, we could **approximate** it to be 47.5%. This is because a score of 90% is within **two standard deviations of the mean**, but the **range** is **only between 70% to 90%**, so we **half the 95% rule:**



Applying the 68%, 95%, 99.7% rule:

90% is within **two standard deviations of the mean**

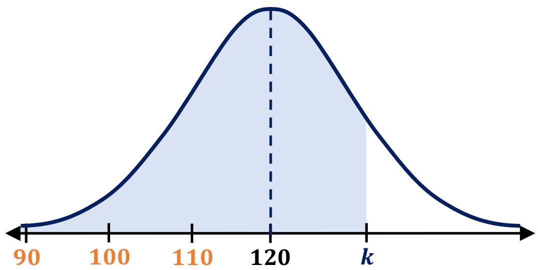
$$P(70 < X < 90) \approx 0.475$$

$\therefore$ **47.5% students score between 70% and 90%**

Similar to how we used the **invNormCDf function** to determine the **value of x** for a **given proportion**, we can also use the **$68\%, 95\%, 99.7\%$ rule**. Using a common example in healthcare, let's suppose that the **systolic blood pressure** of a population has a normal distribution with a **mean** of $120mmHg$ and a **standard deviation** of $10mmHg$.

Let's suppose we wanted to know the **blood pressure** that 84% of the population is **below**. Using the **$68\%, 95\%, 99.7\%$ rule** we can identify that 84% of the population is **one standard deviation away from the mean** (i.e. $130mmHg$). We do this by realising that the **first 50% of the population** is **below the mean value** of $120mmHg$ and the **remaining 34% represents half of the 68%**, which is **one standard deviation**.

Approximate $P(X < k) = 0.84$



Reasoning with the $68\%, 95\%, 99.7\%$ rule:

50% of the population is **below the mean**

The **remaining 34%** is **one standard deviation away from the mean**

$$\therefore k = 120 + 10$$
$$k = 130mmHg$$

Worked Example 2

Teacher Andrew is eager to continue his applications of normal distributions to various parts of his life. Another company Teacher Andrew runs is a **Kombucha production company**. The **Kombucha bottles** are **normally distributed** with a **mean of $330ml$** and a **standard deviation** of $3ml$.

(a) First, Teacher Andrew wants to know what the probability is that the next Kombucha bottle will be **between $327ml$ and $336ml$**. Help Teacher Andrew to **approximate this probability**.

Distribution is **one standard deviation below the mean** and **two standard deviations above the mean**, so:

$$P(327 < X < 336) \approx 0.34 + 0.475$$
$$\therefore P(327 < X < 336) \approx 0.815$$

(b) The Kombucha bottles are **rejected** by the grocery stores if the bottles contain **less than $323ml$**. Out of a **random sample of 80 bottles**, approximate **how many bottles** of Teacher Andrew's Kombucha will be **rejected**.

Since $323ml$ is **two standard deviations below the mean:**

$$P(X < 323) \approx \frac{0.025}{2}$$
$$P(X < 323) \approx 0.0125$$
$$\therefore Bottles\ rejected = 80 \times 0.0125$$
$$\therefore Bottles\ rejected = 1\ bottle$$

9.2 Applications of Normal Distributions

Now that we've explored the **properties of normal distributions** and how to use them to make predictions and analyse data, let's take a look at some **real-world applications**.

A common application of normal distributions is in **finance**. For instance, suppose a **stock analyst** receives **daily returns of a particular stock**, and we have historical data that tells us the **daily returns** follow a **normal distribution** with a **mean** of 1.5% and a **standard deviation** of 0.2%.

To begin with, let's say that the **stock analyst** wants to determine the probability that they get a return of **at least** 1.6% from the stock today. To do this, we use our **normCDf function**:

normCDf	
Lower	1.6
Upper	∞
σ	0.2
μ	1.5

Solving on **calculator** gives:

$$P(X > 1.6) = 0.309$$

$\therefore$ 30.9% **probability** of getting a **return of at least** 1.6%

To keep his investors happy, the **stock analyst** is **expected** to generate **at least** 1.6% **returns** on **at least four days of each week**. To determine the probability that the stock analyst meets these demands, we now consider our distribution as a **binomial distribution** where $Y \sim Bin(7, 0.309)$. Applying our knowledge of **binomial distributions**, we can calculate the **probability** the **stock analyst meets the investor demands** as follows:

① $P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}$

② $P(Y = 4) = \binom{7}{4} 0.309^4 \times 0.691^3$

$P(Y = 4) = 0.105$

③ $\therefore$ **stock analyst** has a 10.5% **probability of meeting investor demands**

① **State** the **binomial distribution function**

② **Substitute** in the **values** of n, y **and** p

③ **State** the **final probability** and a **concluding statement**

Understanding that the current stock is unlikely to provide the returns he needs, he decides to pick a new stock with **daily returns** that follow a **normal distribution** with a **mean** of 2% and a **standard deviation** of 1%. If the stock analyst gets a **negative daily return**, he could **lose his investors**. As a result, the stock analyst wants to **determine the probability** that he gets a **negative daily return** (i.e. $P(X < 0)$). To do this, we again return to the **normCDf function**:

normCDf	
Lower	$-\infty$
Upper	0
σ	1
μ	2

Solving on **calculator** gives:

$$P(X < 0) = 0.0228$$

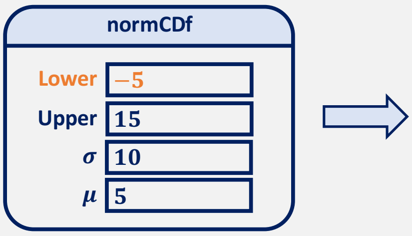
$\therefore$ 2.28% **probability** that he **loses his investors**

As we have seen throughout this chapter, the **applications for normal distributions** are **incredibly diverse**. To **consolidate your knowledge**, have a go at working through the **problem set,** particularly the **application-style questions**.

Worked Example 3

Shauna is notorious for being late to class each day. Suppose the **time Shauna arrives at school** each morning follows a **normal distribution** with a **mean** of $8:05am$ and a **standard deviation** of 10 **minutes**.

(a) What **percentage of the time** is Shauna expected to arrive at school **between** $7:55am$ and $8:15am$?

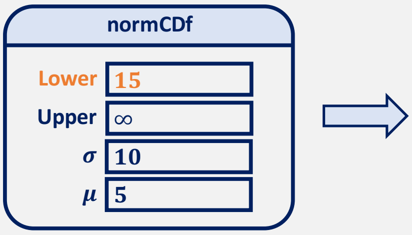


Solving on **calculator** gives:

$$P(-5 < X < 15) = 0.683$$

∴ 68.3% **probability** of **arriving between these times**

(b) If **school starts at** $8:15am$, what **percentage of the time** is Shauna expected to be **late for school**?



Solving on **calculator** gives:

$$P(X > 15) = 0.159$$

∴ 15.9% **probability** of being **late to school**

(c) Over the course of the **five days of school**, what is the **probability** that Shauna will be **late** on **two of the days**?

$$Y \sim Bin(5, 0.159)$$

$$P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}$$

$$P(Y = 2) = \binom{5}{2} 0.159^2 \times 0.841^3$$

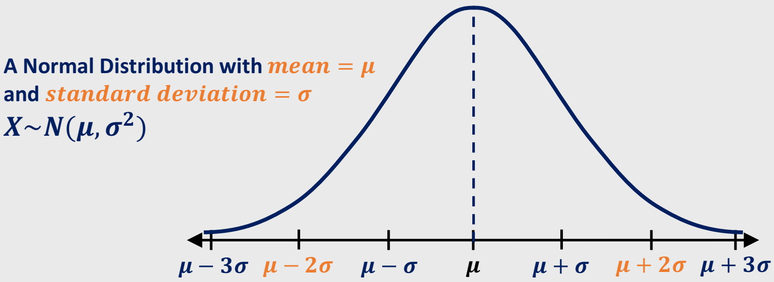
$$P(Y = 2) = 0.150$$

∴ **Shauna** has a 15% **probability of being late to school on two days of the week**

NORMAL DISTRIBUTION TOPIC NOTES

Normal Distribution

A **normal distribution** is a type of **probability distribution** that is **symmetric** around its **mean** and it often **characterised by its bell shaped curve**. We write them as $X \sim N(\mu, \sigma^2)$. The **normCDf** and **invNormCDf** functions on your calculator can help you **find different values** associated with a normal distribution.

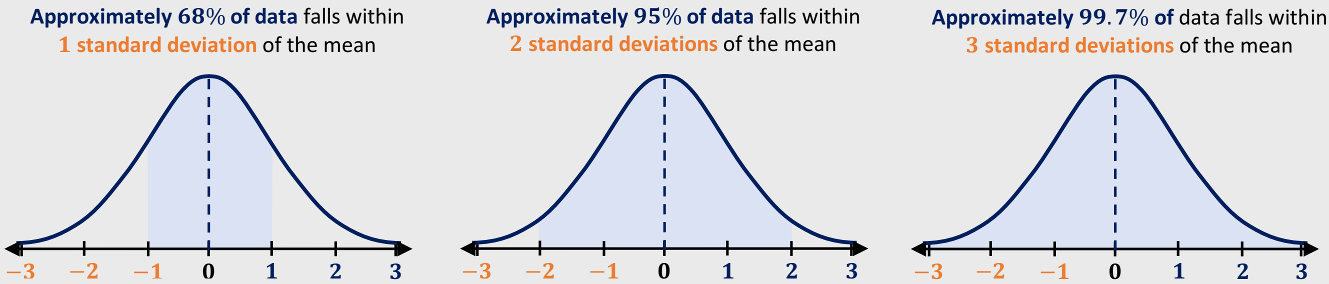


Quantiles

Quantiles are **values** in which a **specific proportion** of the distribution **falls below**. For observations under a normal distribution, we can calculate the quantile through a calculator. We need the **mean** and the **standard deviation** of the normal distribution, and the **quantile value**. The quantile value can be calculated with the **invNormCDf function**.

Approximate Spread of Normal Distribution – 68%, 95%, 99.7% Rule

The **68%, 95%, 99.7% rule** is summarised below.



We can use the **68%, 95%, 99.7% rule** to **approximate key values** within the normal distribution by determining how many **standard deviations they lie from the mean**.

Applications of Normal Distribution

Normal distributions have a wide variety of **real world-applications**, across a variety of industries. They can be used to model returns on stocks, for example, or the distributions of blood pressures, ingredient contents, or the weight or size of a manufactured product. When dealing with **application-based questions**, you will often be met with a **worded question** detailing a few **distribution parameters** or perhaps the **probability of an event occurring**. To solve these questions, we can often make use of the **normCDF function**.

Problem Set 12 – The Normal Distribution
Progressive Questions

Concept 1

The Normal Distribution – Progressive Questions

(7 questions)

Repetitive questions: 1.11 → 1.61 (7 questions)

A Fun Plane Ride Home

Starring: Teacher Simon and the Maths Methods Students

Having Hiked Machu Pichu, Teacher Simon, Janitor Peter and the are now set to return home to Mathemiatcs college to continue learning. However, on the plane flight Teacher Simon and Janitor Peter see this as a perfect opportunity to get some more mathematics teaching in.

Normal Distribution: Q1, Q2, Q3, Q4, Q5, Q6, Q7

Repetitive: 1.11 → 1.71 (8 questions)

[4 marks]

1. On the plane flight home, teacher Simon is handing out snacks and sees this as the perfect opportunity to get some practice in for normal distributions by **calculating standard scores**. Based on the snacks provided to each student, help each of them to determine the **standard scores** of their snack distributions.

(a) Rupert's orange weighs $460g$ that typically has a **mean weight** of $450g$ and a **standard deviation** of $5g$. **(1)**

(b) Wallace's mango weighs $150g$ that typically has a **mean weight** of $200g$ and a **standard deviation** of $10g$ **(1)**

(c) Rabea's packet of chips weighs $97g$ that typically has a **mean weight** of $90g$, and a **standard deviation** of $2g$ **(1)**

(d) Tyler's yoghurt weighs $95g$. This is $5g$ less than the typical **mean weight** and $86g$ **more** than the **variance.** **(1)**

[4 marks]

2. As the students finish eating their snacks, teacher Simon decides to teach them more about normal distributions. He says, "**if X is a normal variable** with a **mean of 144** and has a **standard deviation 12**, and if $P(x \geq 180) = a$, determine the following in terms of a."

Help the students **estimate** the following **probabilities** as teacher Simon reads them out.

(a) $P(X \leq 180)$ **(1)**

(b) $P(144 \leq X \leq 180)$ **(1)**

(c) $P(X \geq 108)$ **(1)**

(d) $P(108 \leq X \leq 180)$ **(1)**

[CA][4 marks]

3. After a few minutes of loud discussion, the students take in it turns to call out the correct answers. Teacher Simon is thoroughly impressed with their solid grasp of the new topic and calls Janitor Peter over to witness their progress. However, Janitor Peter decides he wants to see them in action and gives the students the **following problem**. Help them work through it:

Given that $X \sim N(\mathbf{100}, \mathbf{2}^2)$, determine:

(a) the 0.95 **quantile** **(1)**

(b) the 0.30 **quantile** **(1)**

(c) the 21^{st} **percentile** **(1)**

(d) the 78^{th} **percentile** **(1)**

[CA] [8 marks]

4. As further practice, Janitor Peter decides to hand out their test scores from some of their tests they have completed throughout the year. Help the students to **standardise these test scores** and **determine which student scored the best**.

(a) What is Rupert's standard score if he **scored** $\mathbf{74\%}$ when the **mean** was $\mathbf{89\%}$ and the **standard deviation** was 6% **(2)**

(b) What is Wallace's standard score if he **scored** $\mathbf{95\%}$ when the **mean** was $\mathbf{92\%}$ was and the **standard deviation** was 12% **(2)**

(c) What is Tyler's standard score if he **scored** $\mathbf{66\%}$ when the **mean** was $\mathbf{75\%}$ and then **standard deviation** was 8% **(2)**

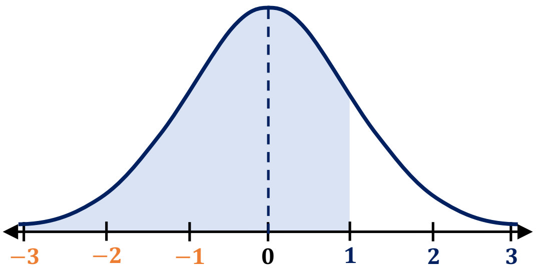
(d) Which student **scored the best** and which student **scored the worst**? **(2)**

To continue pushing the Maths methods students, Teacher Simon and Janitor Peter decide to create some more difficult questions for the plane ride home.

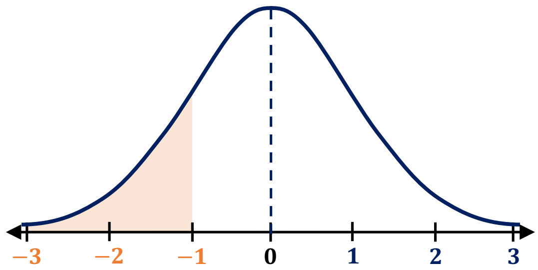
[CA] [6 marks]

5. Teacher Simon decides to set a question for the iconic duo, Rupert and Rabea. In this problem they are provided with a **series of normal distributions** for which they must **determine** the **probabilities**. Help Rupert and Rabea to do so based on the information provided:

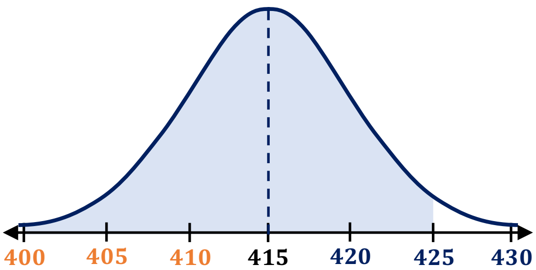
(a) If $\mu = 0$ and $\sigma = 1$ determine $P(X \leq 1)$



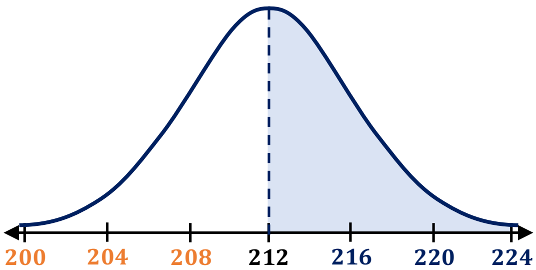
(b) If $\mu = 0$ and $\sigma = 2$ determine $P(X < -1)$



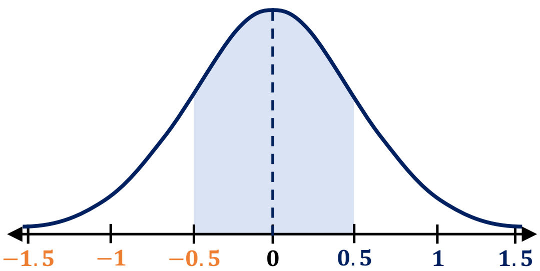
(c) If $\mu = 415$ and $\sigma = 5$ determine $P(X < 425)$



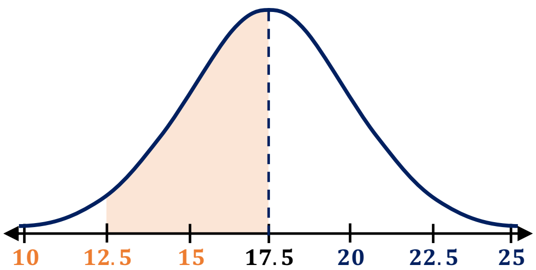
(d) If $\mu = 212$ and $\sigma = 4$ determine $P(X > 212)$



(e) If $\mu = 0$ and $\sigma = 0.5$ determine $P(-0.5 < X < 0.5)$



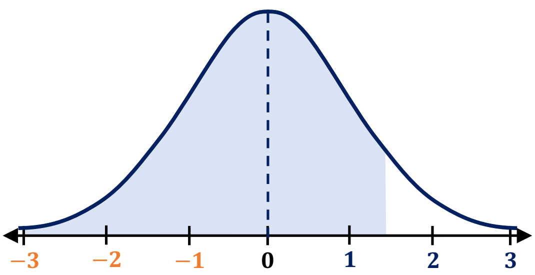
(f) If $\mu = 17.5$ and $\sigma = 2.5$ determine $P(12.5 < X < 17.5)$



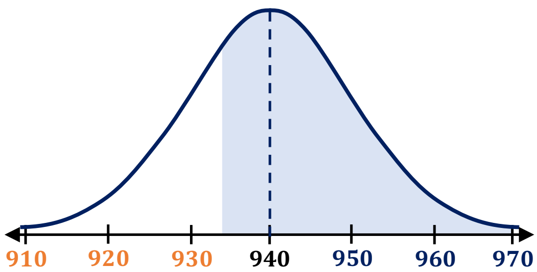
[CA] [16 marks]

6. Ramping up the difficulty of the problems, Rabea and Rupert are now faced with a much more difficult problem. Now Teacher Simon wants the pair to determine the **value of k** based on the **information provided** about a normal distribution. Help Jevon and Tom to **determine the value** of k for the following distributions:

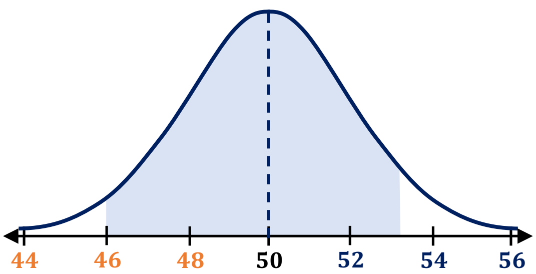
(a) If $\mu = 0$, $\sigma = 1$ and $P(X < k) = 0.935$ determine k



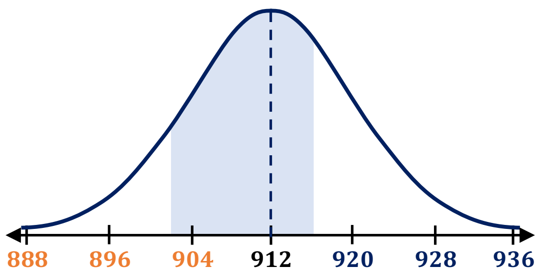
(b) If $\mu = 940$, $\sigma = 50$ and $P(X > k) = 0.7$ determine k



(c) If $\mu = 50$, $\sigma = 2$ and $P(46 < X < k) = 0.874$ determine k



(d) If $\mu = 912$, $\sigma = 8$ and $P(k < X < 916) = 0.688$ determine k



(e) Determine the **value** of k if $\mu = 5$, $\sigma = 5$ and $P(X < k) = 0.987$ **[2]**

(f) Determine the **value** of k if $\mu = 0.5$, $\sigma = 0.1$ and $P(X > k) = 0.425$ **[2]**

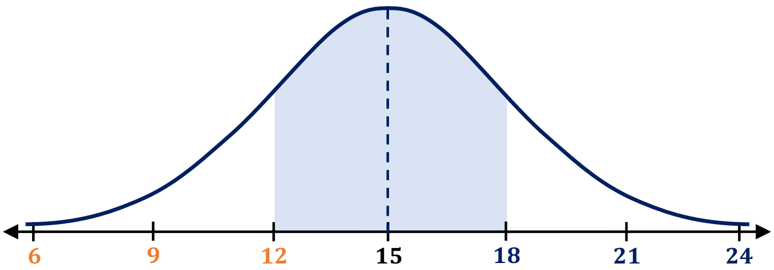
(g) Determine the **value** of k if $\mu = 420$, $\sigma = 69$ and $P(k < X < 410) = 0.210$ **[2]**

(h) Determine the **value** of k if $\mu = -2$, $\sigma = 1$ and $P(-4 < X < k) = 0.665$ **[2]**

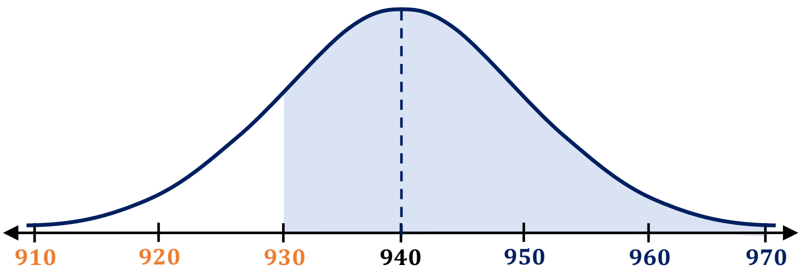
[CA] [10 marks]

7. The **probabilities** and **k values** for normal distributions can also be predicted using the **68%, 95%, 99.7% rule**. Kerry and Wallace are tasked with using this rule to predict the probabilities of the **normal distributions below**. Help Kerry and Wallace to determine these **probabilities** and **k values** based on the information provided.

(a) If $\mu = 15$ and $\sigma = 3$ determine $P(12.5 < X < 17.5)$



(b) If $\mu = 940$ and $\sigma = 10$ determine $P(930 < X < 970)$



(c) Suppose Kerry and Wallace are handed a **large can of baked beans** which has a **mean** of $300g$ and a **standard deviation** of $3g$. What is the **probability** that there can of baked beans contains **less than $294g$ of baked beans**?

Concept 2

Challenging Normal Distribution – Progressive Questions

(3 questions)

Repetitive questions: 2.11 → 2.31 (3 questions)

A Fun Plane Ride Home continued

Starring: Teacher Simon and the Maths Methods Students

Having absolutely excelled through all of the initial problems, Teacher Simon and Janitor Peter decides to give the students a much more difficult set of normal distribution questions that they must work through. Help the methods students to work through the scenarios given below.

Challenging Normal Distribution Questions: Q1, Q2, Q3

Repetitive: 2.11 → 2.31 (3 questions)

[10 marks]

1. John's local supermarket, Moles, is being accused of putting in, **on average** less than the advertised **25 potatoes per bag** when filling bags of potatoes, so John decides he wants to find out whether the supermarket is lying in their advertisements. John brings home a ridiculous **35 bags of potatoes** and then counts the number of potatoes in all of them.

(a) John found that the bags had a **mean of 26 and a standard deviation of 3**. Is it correct to use a normal distribution to model this data if **24 of the bags fell within one standard deviation of the mean?** **(2)**

(b) From this data, **was the accusation against Moles correct**? **Explain why**. **(2)**

(c) From this data, **was Moles' advertisement false**? **Explain why**. **(2)**

(d) Suppose Moles corrected their error and made the **mean potato per bag 28 and the standard deviation 1.4**. What is the probability that a bag of potatoes is **more than 2.2 standard deviations** away from the mean? **(4)**

[6 marks]

2. An Olympic-level athlete, Chelsea, has a complex training regimen and diet, and their coach knows that they need to replenish **250 mg of sodium** after every training session to prevent muscle cramps. The athlete replenishes their sodium only by taking an electrolyte powder called **'Serious Citrus'** that claims to have **250 mg** of sodium per packet, but they are still experiencing muscle cramps **20% of the time**. This is unacceptable, so you need to help the sports nutritionist with investigating the electrolyte powder's sodium content. Assume **normally distributed parameters**, $X \sim \mathcal{N}(\mu, \sigma^2)$.

(a) The machine for **'Serious Citrus'** production is set to fill the packets with an **average of 255 mg of Sodium**. What is the **standard deviation** for the amount of Sodium in packets of **'Serious Citrus'**? *Hint: You can apply the standard normal distribution* $Z \sim N(0, 1^2)$. **(2)**

(b) Using your mathematical knowledge, explain why the athlete might be experiencing the muscle cramps despite taking the **'Serious Citrus'** after training? **(2)**

(c) **Using your understanding of the properties of normal distributions**, what could the manufacturer of the electrolyte powder change about their production methods to reduce the occurrence of muscle cramps? **(2)**

[6 marks]

3. To help Chelsea, the same athlete from before, the sports nutritionist will be trying to **choose between two** new electrolyte powder suppliers: **'Apple Annihilation'** and **'Fruity Fury.'** Since there are random variations during manufacturing, assume **normally distributed parameters**, $X \sim \mathcal{N}(\mu, \sigma^2)$.

(a) Packets of **'Apple Annihilation'** reduced the occurrence of muscle cramps to **2.5%.** If each packet has **standard deviation of 6 mg of Sodium**, what is the **mean amount** of Sodium contained in the packets of **'Apple Annihilation'**? **(2)**

(b) The Sodium content for Packets of **'Fruity Fury'** have a **mean of 265 mg** and a **standard deviation of 5 mg**. What is the athlete's **probability of a muscle cramp** if the sports nutritionist supplies **'Fruity Fury'**? **(2)**

(c) Between the two new electrolyte powders, **which should the sports nutritionist choose? Justify your answer**. **(2)**

Problem Set 12 – The Normal Distribution
Repetitive Questions

Concept 1
The Normal Distribution – Repetitive Questions
(7 questions)

The Normal Distribution: Qs 1.11, 1.21, 1.22, 1.31, 1.41, 1.51, 1.61

[7 marks]

1.11 Given the following **test scores**:

(a) Determine the **standardised score** for each of these **test results**:

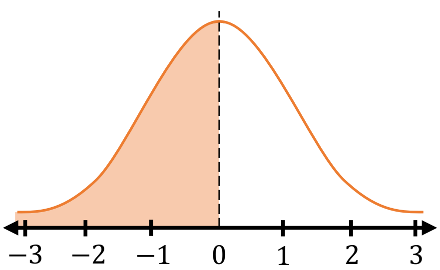
(i) **Result** of **15** in a test of **mean 12** and **standard deviation 3** **(1)**
(ii) **Result** of **60** in a test of **mean 70** and **standard deviation 9** **(1)**
(iii) **Result** of **44** in a test of **mean 38** and **standard deviation 3.5** **(1)**
(iv) **Result** of **5** in a test of **mean 8** and **standard deviation 1.2** **(1)**
(v) **Result** of **19** in a test of **mean 20** and **standard deviation 2** **(1)**

(b) Can you **explain** what it means when a **standardised score** is **positive** or **negative**? **(2)**

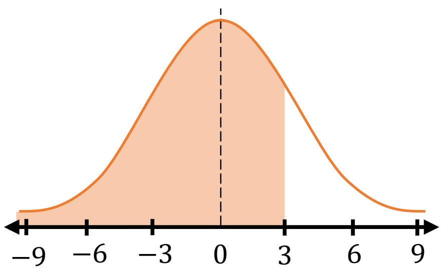
[17 marks]

1.21 Find **probability** of the **shaded area** given the **mean**, **standard deviation** and **graph**. Use the **68%**, **95%**, **99.7%** rule. Keep in mind your results should be an **approximation**.

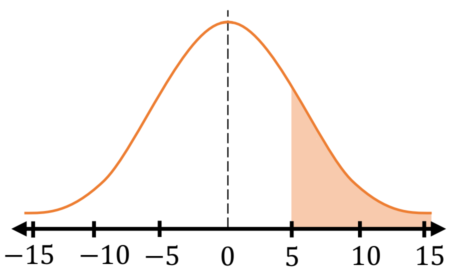
(a) Mean $(\mu) = 0$
Standard deviation $(\sigma) = 1$ **(2)**



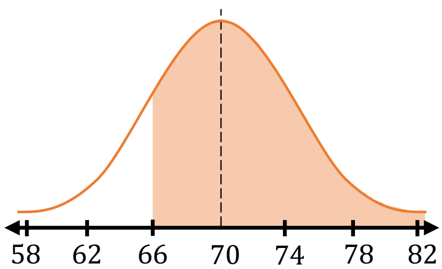
(b) Mean $(\mu) = 0$
Standard deviation $(\sigma) = 3$ **(3)**



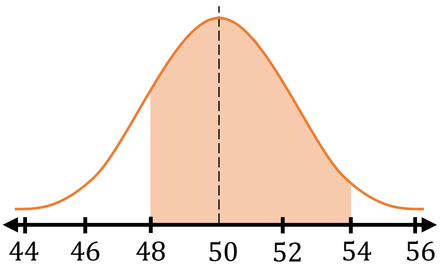
(c) Mean $(\mu) = 0$
Standard deviation $(\sigma) = 5$ **(3)**



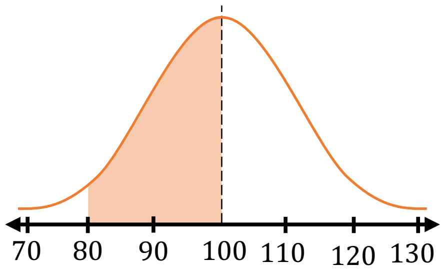
(d) Mean $(\mu) = 70$
Standard deviation $(\sigma) = 4$ **(3)**



(e) Mean $(\mu) = 50$
Standard deviation $(\sigma) = 2$ **(3)**



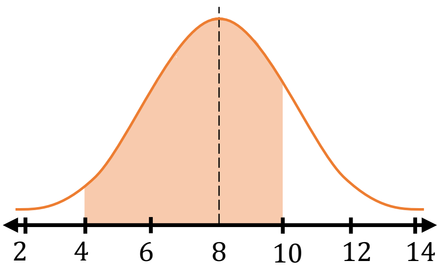
(f) Mean $(\mu) = 100$
Standard deviation $(\sigma) = 10$ **(3)**



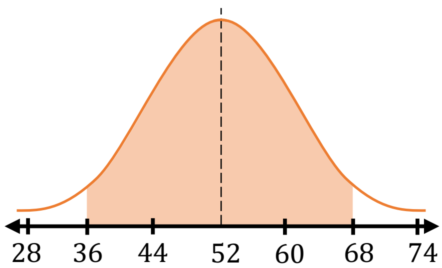
[18 marks]

1.22 Solve the following **probabilities** of the **shaded areas** by **standardising** the **scores**. Use the **68%**, **95%**, **99.7%** rule. Keep in mind your results should be an **approximation**.

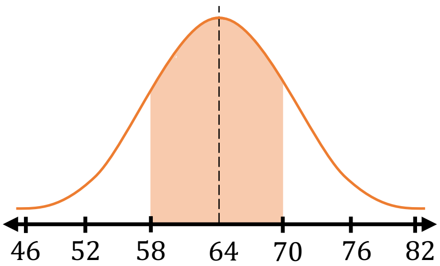
(a) Mean $(\mu) = 8$
Standard deviation $(\sigma) = 2$ **(3)**



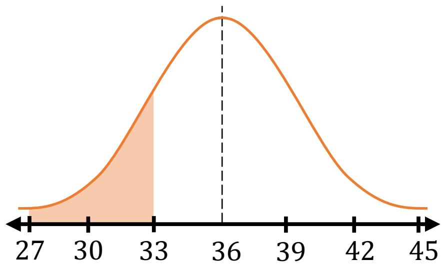
(b) Mean $(\mu) = 52$
Standard deviation $(\sigma) = 8$ **(3)**



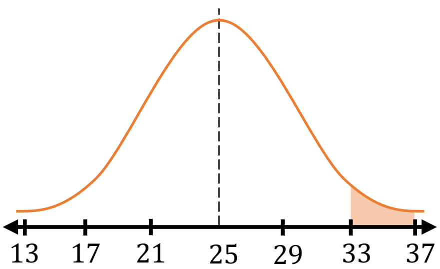
(c) Mean $(\mu) = 64$
Standard deviation $(\sigma) = 6$ **(3)**



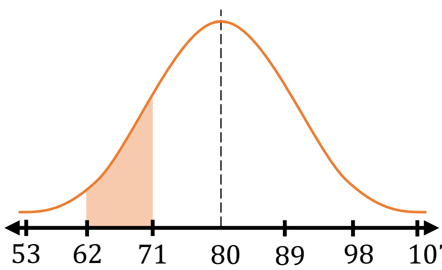
(d) Mean $(\mu) = 36$
Standard deviation $(\sigma) = 3$ **(3)**



(e) Mean $(\mu) = 25$
Standard deviation $(\sigma) = 4$ **(3)**



(f) Mean $(\mu) = 80$
Standard deviation $(\sigma) = 9$ **(3)**



[9 marks]

1.31 Using the **68% 95% 99.7% rule** for normal distributions and **without using a calculator**, determine the following:

(a) The **0.5** quantile for X, if $X \sim N(8, 4)$ **(1)**

(b) The **0.84** quantile for X, if $X \sim N(-3, 4^2)$ **(2)**

(c) The **0.975** quantile for X, if $X \sim N(5,\ 1.5^2)$ **(2)**

(d) The **0.16** quantile for X, if $X \sim N(25, 9)$ **(2)**

(e) The **0.025** quantile for X, if $X \sim N(100, 15^2)$ **(2)**

[12 marks]

1.41 Using the **68% 95% 99.7% rule** for normal distributions and **without using a calculator**, **determine the value of k** in the following:

(a) $P(k < x < 12) = 0.68,\ if\ X \sim N(10, 4)$ **(2)**

(b) $P(k < x) = 0.16\ ,\ if\ X \sim N(0, 1)$ **(2)**

(c) $P(4 < x < k) = 0.815,\ if\ X \sim N(8, 2^2)$ **(2)**

(d) $P(x < k) = 0.975,\ if\ X \sim N(15, 16)$ **(2)**

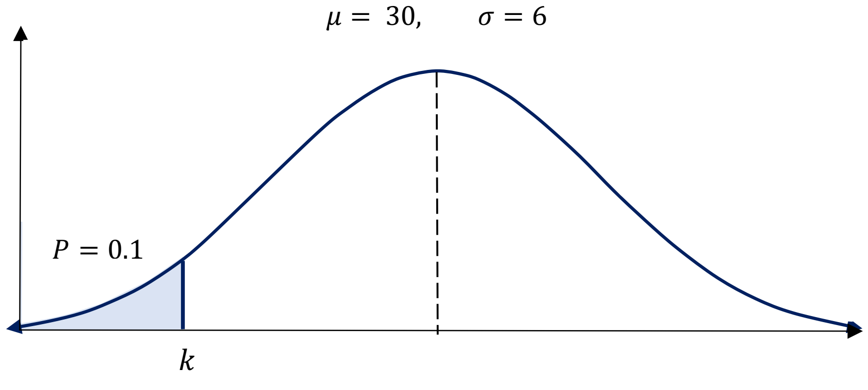
(e) $P(k < x < 29) = 0.9735,\ if\ X \sim N(20, 3^2)$ **(2)**

(f) $P(-k < x < k) = 0.95,\ if\ X \sim N(0, 1^2)$ **(2)**

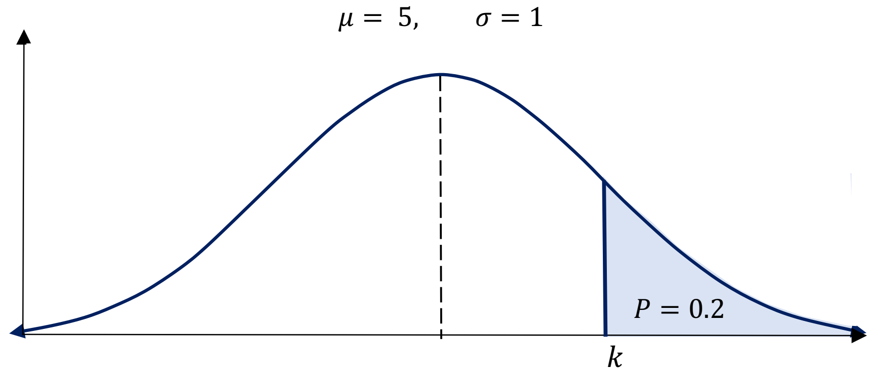
[8 marks]

1.51 For the following **normal distributions**, the probability of the value falling within a shaded area is shown. Using a calculator, **determine the value of k and/or n** in the following situations:

(a) **(1)**

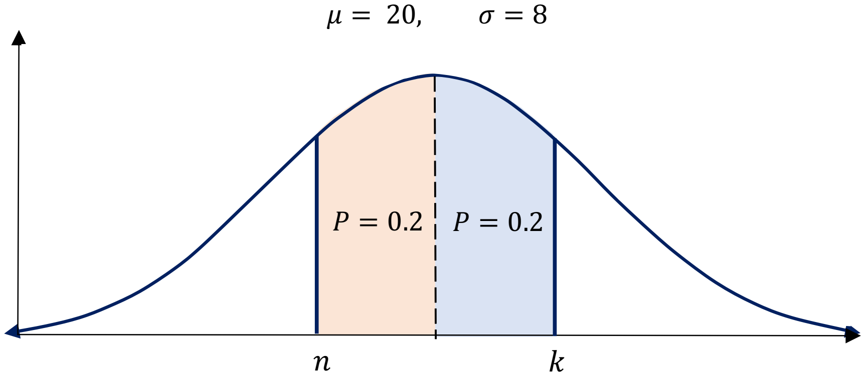


(b) **(1)**



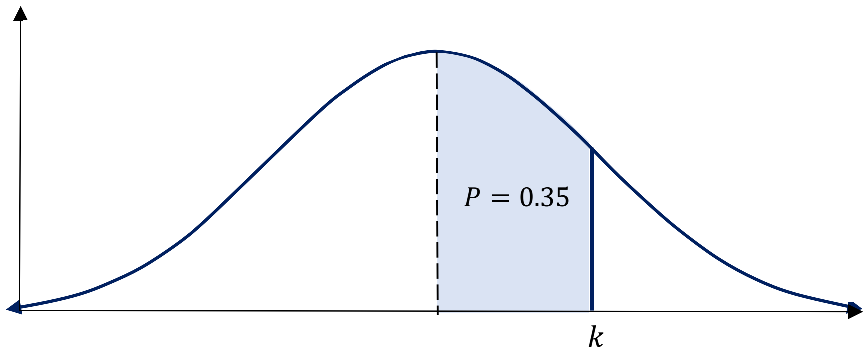
(c) **(1)**

$\mu = 20, \quad \sigma = 8$



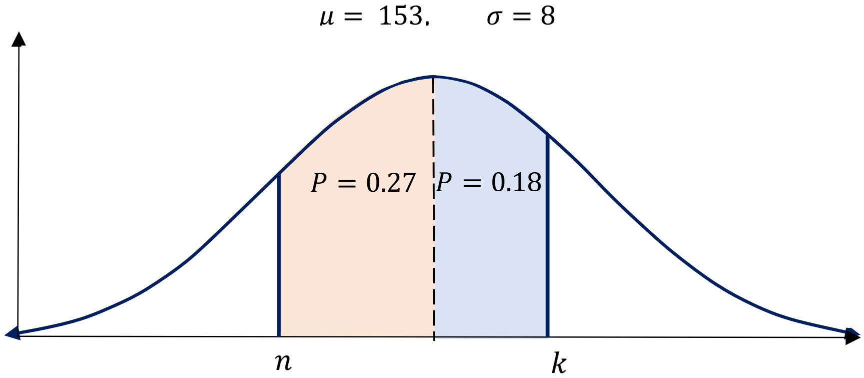
(d) **(1)**

$\mu = 17, \quad \sigma = 2.3$



(e) **(2)**

$\mu = 153, \quad \sigma = 8$



[16 marks]

1.61 State whether the following are likely to be normally distributed. **Provide a brief justification of your answer.**

(a) The **heights** measured for students in a school. **(2)**
(b) **IQ scores** of the general population. **(2)**
(c) The distribution of **WACE exam results**. **(2)**
(d) The **heights** of the players on the court in a basketball game. **(2)**
(e) The **frequency distribution** of digits in π **(2)**
(f) **100 students** at an elite sporting school completed a 400m sprint with an **average** time of 60 seconds. 60 students took longer than one minute to complete the 400m sprint. **(3)**
(g) The **swim times** were measured for 8 students at a rural high school who swam in a 50m freestyle race. The **average** time was 30 seconds. 4 students finished in **less** than 30s. **(3)**

Concept 2

Challenging Normal Distribution – Repetitive Questions

(3 questions)

Challenging Normal Distribution Questions: Qs 2.11, 2.21, 2.31

[9 marks]

2.11 A small beverage company produces canned soft drinks and is reviewing their quality control process due to customer complaints. The volume of liquid in each can is labelled as **375 mL**. Out of **four hundred and eighty thousand** cans sold in the last month, **twelve thousand** cans contained **less** than the labelled amount. It is assumed that the volume of liquid in each can is **normally distributed**, with $X \sim \mathcal{N}(\mu, \sigma^2)$.

(a) What **percentage** of the cans contained **less** than the labelled amount of soft drink? **(1)**

(b) The machine on the production line is programmed to fill the cans with an average of 376.96mL. What is the **standard deviation** for the volume of soft drink in the cans? *Hint: You can apply the standard normal distribution* $Z \sim N(0, 1^2)$. **(2)**

(c) To please their customers the company wishes to ensure that the number of cans containing less than 375 mL is reduced. **How can the quality control engineer achieve this**? **(2)**

(d) To reduce the number of cans containing less than 375 mL **from 2.5% to 0.5%,** the company has **increased the average** amount of soft drink in each can, while **the standard deviation remains the same**. What is the **mean amount** of soft drink in each can? **(2)**

(e) To reduce the number of cans containing less than 375 mL **from 2.5% to 1%**, the company has **decreased the standard deviation** while **maintaining the original average** amount of soft drink in each can. What must the **new standard deviation** of soft drink be to achieve this? Round to 4 decimal places. **(2)**

[6 marks]

2.21 A fish monger sells gutted sardines at $\$22$ for a kilogram box. The **weight** of a single sardine is normally distributed with an **average of $120g$**. During a single day, the fish monger expects to earn revenues of $\$176$ from sardines alone. The **number of sardine box sales** is also normally distributed with **standard deviation of 2**. The monger also knows that **less than one fortieth of the time** he sells a fish that is less than **51% the weight of an average sardine**.

(a) Define the distribution of sardine weight **(2)**

(b) Define the distribution of sardine box sales **(2)**

(c) What is the probability that on any given day the monger makes a sale containing a sardine of **weight less than $80g$** *and* earns **revenues of less than $\$132$.** Assume that **sardine weight** and **total sales** are independent of each other. **(2)**

[CA][6 marks]

2.31 A researcher is trying to study the link between sugar content in locally made chocolate bars and its effect on children's energy levels. To do this, she buys three different bars from a local supermarket: **'Choc Delight'**, **'Almond Crunch'** and **'Chunky Caramel'**. Due to manufacturing variations, it can be assumed that the sugar content in each bar is **normally distributed**, with $X \sim \mathcal{N}(\mu, \sigma^2)$.

(a) The researcher finds that for the children who consumed the **'Choc Delight'** bar, the probability that their energy levels would be **raised** was $\mathbf{12.2\%}$. If each bar has a **standard deviation** of $\mathbf{3g}$ of sugar and claimed to have $\mathbf{25g}$ of sugar per bar, what is the **mean amount of sugar** in the **'Choc Delight'** bars? **(2)**

(b) The **'Almond Crunch'** bars have a **mean** of $\mathbf{27g}$ of sugar, despite each bar claiming to have $\mathbf{25g}$ of sugar, and a **standard deviation** of $\mathbf{2g}$. What is the **probability** that the energy levels of children who consumed these bars would be **raised**? **(2)**

(c) Through a chemical experiment, the researcher finds that $\mathbf{2.2\%}$ of the **'Chunky Caramel'** bars had less sugar than the advertised amount of $\mathbf{20g}$. To reduce the amount from $\mathbf{2.2\%}$ to $\mathbf{1\%}$, the researcher believes that the chocolate manufacturer should **increase the average amount of sugar** in each **'Chunky Caramel'** bar. After she contacts the manufacturer, they decide to **increase the average amount** of sugar. If the **standard deviation is** $\mathbf{2g}$, what is the *new* **mean amount of sugar** in each **'Chunky Caramel'** bar? **(2)**

Chapter 10

Sampling and Proportions

Chapter 10 –Sampling and Proportions

Introduction

Welcome to our **final topic** in Year 12 Maths Methods! For the conclusion of our long journey together, we will be learning about **estimation and proportions**. **Estimation and proportions** at a Year 12 level entails learning about the techniques we use to **correctly select** and **analyse samples**. Taking **smaller samples** of a **bigger population** is the name of the game, and with **correct techniques**, we can gain important insights into everyday aspects of life in a timely and cost-effective way.

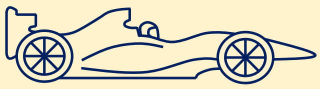
To explore this topic, we have broken this chapter up into **two main concepts**:

1. **Random Sampling**
2. **Sample Proportions**

In essence, this whole topic is about taking careful short cuts. We start with a question we want to know the answer to, like **how many Australians really think cereal is a soup**. Sampling is the art of giving a **reasonably accurate answer** to that question, **without having to ask every person** in the country! But "reasonably accurate" means different things to different people, so we want to do better than that. We want to be able to **quantify exactly how accurate our guess is**, or even calculate how we need to plan our survey so that even the most passionate soup advocate can never doubt us.

Some examples of how sampling can be used are:

Gathering Health Data about the Australian Population



Evaluating the Safety of a Sports Car Model

Determining the Satisfaction of Skiers on a Ski Slope

Let's Begin!

10.1 Random Sampling

A **random sample** is a **subset** of a **statistical population** in which each member of the subset has an **equal probability** of being chosen. This sample is used to **make inferences** about the **whole population**.

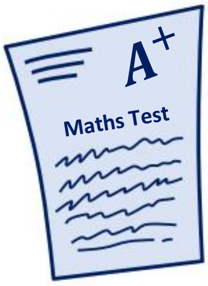
Examples of random samples include surveying the **number of patients** that have **contracted a specific virus** in the **public hospitals** of Western Australia or the **number of cows** in Australia that are **above a specific weight**. By taking a sample of a few hospitals or farms in an area, we can use the results from those populations to **make inferences** about the **whole population**.

When creating a **random sample,** it is essential that the population of the sample is **large enough** to be a fair reflection of the entire population. For instance, a **random sample** of **ten students is not a sufficiently sized sample** to make inferences about all the students in Australia.

A **common rule** is that the **population** of the **random sample** should be **more than 30** (e.g. 30 cows, 30 students, 30 cars), otherwise there is **significant room** for **doubt and error**.

Whilst $\mathbf{30}$ is the **minimum** for a **random sample** to be sufficient, you should aim to have as **large of a sample as practical** to **reduce uncertainty**. It is also worth noting that when a sample is **larger than 30**, certain **properties** of the **random sample** can be **approximated** as being **normally distributed** (which we will explore in section 10.2).

Sampling Methods to Ensure Randomness



Suppose you wanted to survey the **interest of Australian students towards mathematics**, but you **only surveyed students from one school**. The results produced from the survey will be **highly biased**. That **one school** may have **incredible maths teachers** or **uninspiring maths teachers**, which will **heavily influence** the student's answers on the survey.

When **creating a random sample**, the goal is to always have a **population** that is **truly selected at random**, such that every member of the population has an **equal chance** of **being selected**.

To achieve this, there are **various sampling methods** that are commonly used which each have their own benefits and **shortfalls**. These methods include **systematic random sampling**, **stratified sampling**, **convenience sampling** and **self-selection sampling**.

Method	Description	Example
Systematic random sampling	**Systematic sampling** selects **every n^{th}** member of the entire population. This is considered one of the **most reliable** and **least biased ways** to sample, but it is **difficult** and **expensive**.	A statistician is surveying the **attitudes** of the residents of Perth towards **cats and dogs**, so they look through a **phonebook** and decide to **select every 1000^{th} person** for sampling.
Stratified Sampling	**Stratified sampling** divides a **population** into **subpopulations** based on a **specific characteristic**, like **age** or **location**, and selects members of each group in **proportion** to the **size of the group**. This is also **reliable**, but **difficult** and **expensive**.	A university wants to know **student satisfaction**, so based on the number of students in each major, they survey **12% from English**, **15% from mathematics**, **20% from computer science** and so on.
Convenience Sampling	**Convenience sampling** selects members of a population because they are **convenient** in some way. These **samples** are **easy** and **inexpensive**, but **heavily affected** by **bias**.	A company wants to know **how many people enjoy using their products**, so a representative stands outside of a shopping centre and surveys the **first 50 people** who walk past.
Volunteer or self-selection sampling	**Volunteer sampling** involves asking for **volunteers** where people decide whether they want to be part of the **sample**. These samples are **heavily affected** by **bias**.	**Residents of a suburb** are **sent a survey** asking them whether they **approve** of a new **building proposal**. Most respondents **lived close to the site** and therefore had strong views about it.

A final sampling method worth exploring is **capture-recapture**, which is commonly used to **estimate the populations** of a **specific species** of animals. **Capture-recapture method** is where an **initial sample** is **captured** and **marked**. The marked individuals are then assumed to **evenly mixed** with the rest of the population after some time passes. It is also assumed that **no individuals leave the area under investigation.** Then another **sample** is captured **independently**, and it is noted how many of the **second sample** are **marked**. The **overall size** of the population can then be **estimated** using the **following ratio**:

$$\frac{Total\ Marked}{Total\ Population} = \frac{Number\ of\ Marked\ Recaptured}{Total\ Recaptured\ on\ Second\ Occasion}$$

The best way to understand this ratio is through an example. Suppose a researcher wants to **estimate** the **size of a kangaroo population** on a **specific wildlife reserve**. To do this, they could use the **capture-recapture method** as follows. On the **first day**, **110 kangaroos** are **marked** and released. Therefore, the **ratio** for the **left-hand side** of the equation is:

$$\frac{Total\ Marked}{Total\ Population} = \frac{110}{Total\ Population}$$

After a week, once the marked kangaroos have **evenly mixed** with the rest of the population, and assuming **none have left the reserve**, the researchers return and **capture 200 kangaroos** of which **17 are marked**. Therefore, the **ratio** for the **right-hand side** of the equation will be:

$$\frac{Number\ of\ Marked\ Recaptured}{Total\ Recaptured\ on\ second\ occasion} = \frac{17}{200}$$

If we **combine both ratios together**, we can then **solve** for the **'Total Population'** of **kangaroos** at this nature reserve:

① $$\frac{110}{Total\ population} = \frac{17}{200}$$

① **Substitute** the **known values** into the **ratio formula**

② $$Total\ population = 110 \times \frac{200}{17}$$

② **Re-arrange** the equation to **solve** for **'Total Population'**

③ $$\therefore Total\ population = 1294\ kangaroos$$

③ **Calculate the total kangaroo population**

While knowing the **different types** of sampling methods is valuable, it is also **very important** to be able to **determine** the **biases** in a sampling method and **conclude whether a sample is truly random**.

Biases in Sampling

When creating a 'random' sample, it is important to understand that there are various **causes for biases** which can **skew the results**.

A sampling method will have **bias** if it **systematically favours one outcome over others**. A source of bias can be intentional or unintentional and it can lead to **inaccurate conclusions** about a population.

For instance, let's suppose Big Bang Burgers, a **burger chain** that **mostly caters to teenagers**, wants to **gauge customer satisfaction** with its restaurants. They decide to interview their **first 50 customers** on a **Tuesday morning** at **one** of their stores in Perth. As we have learned before, this is a type of **convenience sampling**. From this method of sampling, we can **quickly identify** some **sources of bias**:

Bias Source	Description
1	**Only one store was surveyed** – customers at **other burger store locations** may hold **different opinions** that **aren't reflected** in the data collected.
2	The **selection scheme is flawed** - the **first 50 people** is **not** a **random sample** of **all customers**.
3	The sample **only** surveys customers who **dine on a Tuesday morning**. This **excludes** people who **work during the day** and **school students**, who are their main customer base.

Determining the **biases in a sampling method** will **vary in every situation**, however, there are **two general questions** that you should ask yourself when determining whether there is **bias present**.

General Questions to Determine Bias

① How does the **sampling method** influence the **type/groups of people** that will be **responding** to the **survey**?

② How will the **sampling method** create **restrictions for people to respond** to the survey? Consider **factors** such as the **time and location of the survey**, and whether the **survey is voluntary** or **compulsory**.

To **practise determining bias in sampling methods**, try working through the examples below.

Worked Example 1

Tyler is eager to get ahead of his classmates and decides to explore the **concept of random sampling outside of his class time**. To do so, he creates a **television advertisement** that asks people to **ring a number** to give their **opinion** about whether **housing prices are too high** in Australia.

(a) What **method of sampling** is Tyler conducting and what **sources of biases** could be present?

Tyler is conducting a form of **voluntary/self-selection sampling**. There are **significant biases present**, including that only **people with strong opinions** about housing prices are likely to **call-up**, responses are **more likely** to come from **adults and elderly people who watch television**, and the responses are **likely to come** from **people living in urban areas** with better connection to the TV network.

(b) To get a more 'random' sample, Tyler decides to **ask 80 of friends** from **Mathematics College** about their **perspectives on housing prices**. What **method of sampling** is Tyler conducting and will there be any **sources of bias present**?

Tyler is conducting **convenience sampling** because it is **convenient to interview his friends**. There are **significant biases present**, including **only interviewing high school students** rather than the diverse Australian population, and all the students **likely live in the same suburb** so they will hold the same perspectives about housing prices. In future, Tyler should aim to **collect data** from **different age levels, genders and socioeconomic backgrounds**.

10.2 Sample Proportions

A **sample proportion**, represented as $\hat{p}$, is the **proportion of individuals** in a sample that **share a specific trait**, such as the **proportion of Australians** that have **travelled outside the country**.

The **sample proportion**, $\hat{p}$, is **calculated** using the formula below, where X is the **number of people with the specific trait** and n is the **number of people selected**:

Sample Proportion: $\hat{p} = \frac{X}{n}$

It is important to understand that the **value** of the **sample proportion** will **vary** in **each new sample** that you create. For instance, if you wanted to determine the **proportion of Australians** with an **IQ above 100**, it might be $\hat{p} = 52\%$ for **one sample**, but $\hat{p} = 58\%$ for **another sample**. In other words, $\hat{p}$ is itself a **random variable**, with its **own distribution**.

As a **general rule of thumb**, when large samples are created with **more than 30 people** (i.e. $n > 30$), certain properties of a sample **tend towards** being **normally distributed**. When the properties of a sample tend towards being **normally distributed**, we can determine the **mean** and **standard deviation** for that **specific sample size** using the **formulas** we use for **all normal distributions**. The formulas for the **mean** and **standard deviation** of $\hat{p}$ are:

Mean: $\mu = p$

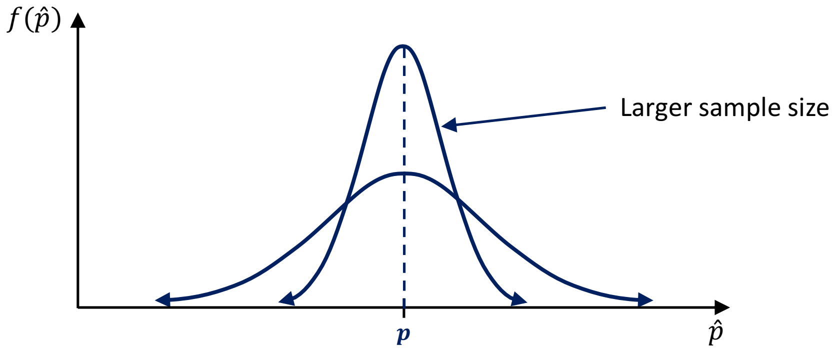
Standard Deviation: $\sigma = \sqrt{\frac{p(1-p)}{n}}$

Before we have a go using these equations, lets quickly unpack the meaning of a **population proportion** (p).

The population proportion (p) is the **true proportion** of individuals with the **specific trait in the population**. It is a type of **population parameter**, which are numbers used to **summarise** a **characteristic** of the **population**.

With these formulas, it is important to note that when the **sample size is larger than 30**, we can assume that $p \approx \hat{p}$, **which means** we can use $\hat{p}$ to **calculate** the **mean** and **standard deviation**. p itself is almost always unknown and it is a fixed value.

Here are some examples of what different **distributions** for $\hat{p}$ could be. **Larger** sample sizes **reduce** the standard deviation of $\hat{p}$ since $\hat{p}$ becomes a **better estimate for p** as the sample size increases.



Continuing with our example of the **proportion of Australians** with an **IQ above 100**, let's suppose that from **conducting an IQ test** on **389 people**, it was found that **220 people** had an **IQ of more than 100**. We would start by **calculating** the **sample proportion ($\hat{p}$)** as follows:

$$\hat{p} = \frac{X}{n}$$
$$\hat{p} = \frac{220}{389}$$
$$\hat{p} = 0.132$$

With the value of $\hat{p}$ determined, we can then substitute this into the **standard deviation formula** since we can assume that $p \approx \hat{p}$. This gives us the **standard deviation** for samples with a **size of $n = 389$**:

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$
$$\sigma = \sqrt{\frac{0.132(1-0.132)}{389}}$$
$$\sigma = 0.0172$$

Worked Example 1

Continuing his exploration of random sampling, Tyler decides to determine the **proportion of students** at Mathematics College that **regularly catch the bus to school**. To do so, he conducts a survey of **464 students** and finds that **only 102 students** regularly catch the bus to school.

(a) Determine the **sample proportion** of students who **regularly catch the bus to school**.

$$\hat{p} = \frac{102}{464}$$
$$\hat{p} = 0.2198$$

(b) Estimate the **standard deviation** of $\hat{p}$, for **samples of size 464 people**.

$$\sigma = \sqrt{\frac{0.2198(1-0.2198)}{464}}$$
$$\sigma = 0.0192$$

(c) If the Australian census concluded that **44% of Australian students catch the bus to school**, **comment on the difference** in your **result determined in (a)**.

The **sample proportion** of 21.98% is more than **11 standard deviations below** the **population proportion**. This suggests that the sample of Mathematics College students is **highly unique** and is **not an accurate representation** of **Australian students** in relation to catching the bus to school.

10.3 Confidence Intervals

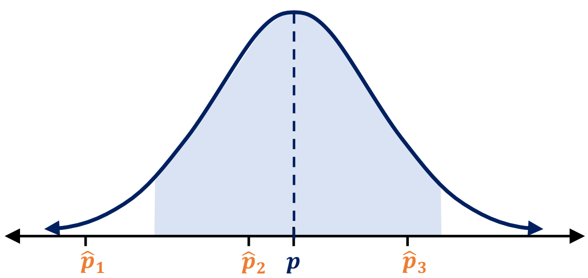
A **confidence interval** is an interval that has a **confidence level** (given as a percentage) which indicates how **certain** we can be that a **population parameter** falls within the interval.

For instance, for the **proportion of the population** that **doesn't vote** in **local government elections**, a **90% confidence interval** of $\mathbf{0.42 \leq p \leq 0.51}$, means we can **expect with 90% confidence** that between **42% and 51%** of Australians **don't vote** in local government elections.

The **formula** for **calculating** the **confidence interval** can be seen below, where z is the **z-score** based on the **confidence interval selected**.

Interval Estimate: $$\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

To understand this formula, think back to the probability distribution of $\hat{p}$ and how we would find an estimate for the **size of an interval** centred at p given it has a certain **probability of containing $\hat{p}$**. Notice if this interval contains a $\hat{p}$, an interval of the **same size** centred instead at that $\hat{p}$ will also **contain p**. This is useful because a **random confidence interval** centred at a **random $\hat{p}$** is expected to **contain p** with a **probability** equal to the **confidence level** of the confidence interval.

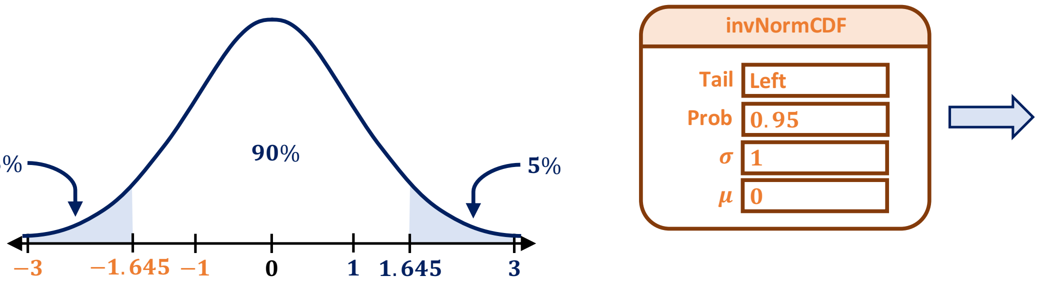


Realise that p is a fixed value, so any given interval around any **certain $\hat{p}$** will either contain p or it won't. This is not a random scenario since both $\hat{p}$ and p have **definite values** here. This means that the level of confidence represents an estimate for the probability that a **random** confidence interval will contain p, **not a specific one**. This is important to remember because you will be calculating specific ones.

At the first glance, this formula can appear to be **incredibly complex**. However, once we **break it down** it **becomes easier to use**. When it comes to **calculating confidence intervals** in **Year 12**, there are **three confidence intervals** we are particularly interested in. These are **confidence intervals** of 90%, 95% and 99%. Each of these confidence intervals has a **z-score** that **needs to be calculated** through using **normal distributions** and our **calculator**.

To calculate the **z-scores** of these **confidence intervals**, we need to determine **how many standard deviations** the confidence interval will be away from the **mean**. To do this, we use the **'invNormCDF' function** on our **calculator** to **solve** for $P(Z < z) = 0.95$, which gives $z = 1.645$.

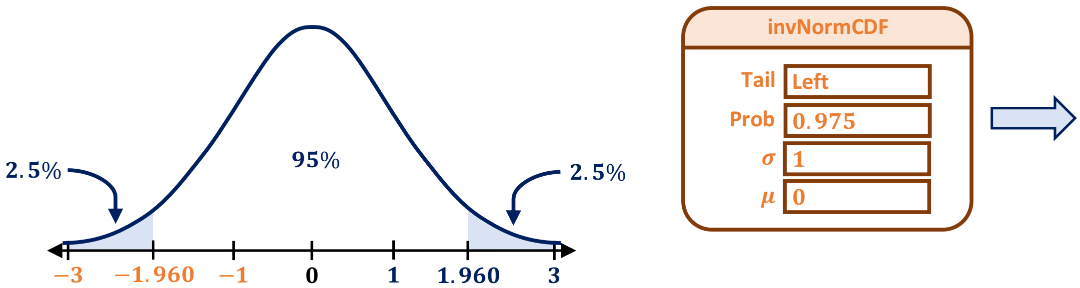
For a **90% confidence interval: $z = 1.645$**



$P(Z < z) = 0.95$
Solving for z using **calculator** gives:
$z = 1.645$

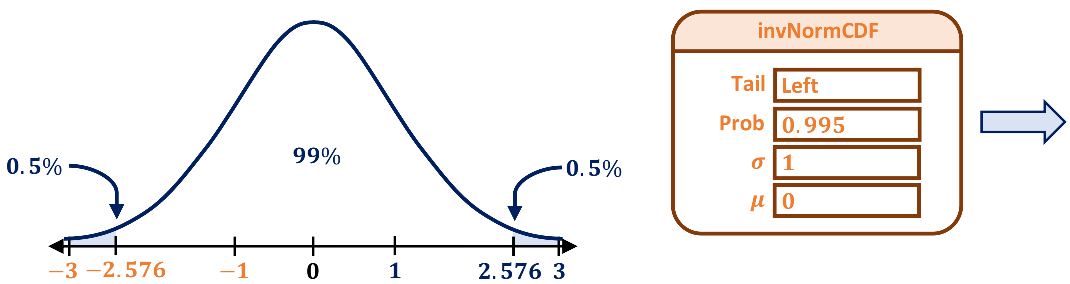
Applying this same process to **confidence intervals** of 95% and 99%, we get **z-scores** of $z = 1.960$ and $z = 2.576$ respectively.

For a **95% confidence interval: $z = 1.960$**



$P(Z < z) = 0.975$
Solving for z using **calculator** gives:
$z = 1.960$

For a **99% confidence interval: $z = 2.576$**



$P(Z < z) = 0.995$
Solving for z using **calculator** gives:
$z = 2.576$

Whilst the derivations are important, the **actual values** of z are the **most important part** and they can be **summarised** in the table on the next page. These **z values** do not change for different samples.

Confidence Interval	Value of z
90%	$z = 1.645$
95%	$z = 1.960$
99%	$z = 2.576$

With the **z-scores determined**, we can now explore some examples to understand how to determine the **confidence interval**. Suppose we **conducted a survey** that concluded that **186 people** are **left-handed** of a **random sample** of **1000 people**. If we wanted to determine the **95% confidence interval**, we would start by determining the **sample proportion $\hat{p}$** and the **standard deviation σ**.

Sample Proportion:

$$\hat{p} = \frac{186}{1000}$$

$$\hat{p} = 0.186$$

Standard Deviation:

$$\sigma = \sqrt{\frac{0.186(1 - 0.186)}{1000}}$$

$$\hat{p} = 0.0123$$

We can use this information along with knowing that $z = 1.960$ for a **95% confidence interval** to **determine** the following **confidence interval**:

① $$\hat{p} - z\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \leq p \leq \hat{p} + z\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

② $$0.186 - 1.960(0.0123) \leq p \leq 0.186 + 1.960(0.0123)$$

③ $$0.162 \leq p \leq 0.210$$

① **State** the **confidence interval formula**

② **Substitute** in the **sample proportion**, **standard deviation** and **z-score**

③ **State** the **confidence interval**

This result means we **expect** with **95% confidence** that **between 16.2% and 21%** of Australians will be **left-handed**.

One final question we can ask ourselves is: if we were to **calculate** the **90% confidence interval** or the **99% confidence interval**, would the **size** of the **confidence interval** get **smaller** or **larger**?

To answer these questions, let's **calculate** the **90%** and **99% confidence intervals** for our previous random sample, using $z = 1.645$ for the **90% confidence interval** and $z = 2.576$ for the **99% confidence interval**.

90% Confidence Interval

$$0.186 - 1.645(0.0123) \leq p \leq 0.186 + 1.645(0.0123)$$

$$0.1658 \leq p \leq 0.2062$$

99% Confidence Interval

$$0.186 - 2.576(0.0123) \leq p \leq 0.186 + 2.576(0.0123)$$

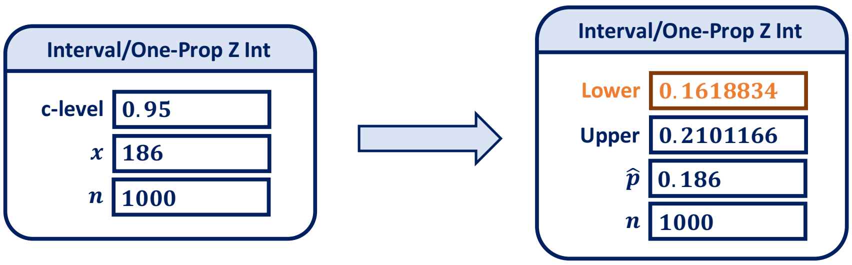
$$0.1543 \leq p \leq 0.2176$$

As we can see, the **90% confidence interval** is smaller than the **95% confidence interval** and the **99% confidence interval** is **larger** than the **95% confidence interval**. This trend can be **best understood** through the following reasoning:

> A **higher level of confidence** means that we **want to be more certain** that the **true population parameter falls within the interval**, so we need to **widen the interval** to account for the **increased certainty**.

In Year 12, **confidence intervals** are more commonly **calculated** using your **calculators**. This can be done by accessing the **statistics component** of your calculator, then **selecting 'Calc'**, then **'Interval'** and then **'One-Prop Z Int'**. You can then **input** the **confidence level**, the **sample size n** and the **number of individuals** with the **specific trait x**.

For instance, for our previous example of the **95% confidence interval** for **left-handers**, we can use the inputs $c - level = 0.95$, $x = 186$ and $n = 1000$ to get our **confidence interval**:



$\therefore$ **the 95% confidence interval is $0.162 \leq p \leq 0.210$**

Either method is **equally valid**, however, in an exam situation, the **calculator** is always **faster**.

Worked Example 1

Walking around Mathematics College, Tyler is eager to determine the **proportion of Year 12 students** that have **some form of colour blindness**. Of the **212 students** Tyler surveyed, he found that **28** of the students had **some form of colour blindness**. **[CA]**

(a) Help Tyler to determine the **95% confidence interval** for the **sample proportion**.

$$\hat{p} = \frac{28}{212}$$
$$\hat{p} = 0.132$$

Confidence Interval:

$$\hat{p} - z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.132 - 1.960\sqrt{\frac{0.132(1-0.132)}{212}} \leq p \leq 0.132 + 1.960\sqrt{\frac{0.132(1-0.132)}{212}}$$

$$0.0864 \leq p \leq 0.178$$

(b) Despite having successfully calculated the confidence interval, Tyler is **struggling** to **interpret what this result means**. Can you **interpret this result** for Tyler?

This means we expect with 90% **confidence** that between 8.64% and 17.8% of Australian students will have **some form of colour blindness**.

(c) Would you expect a **95% confidence interval** to be **larger** or **smaller** than the interval in part (a)?

We would expect the 95% **confidence interval** to be **larger** than the interval in part (a).

Margin of Error

From the **part (a)** of the previous worked example, we determined that the 95% **confidence interval** was $0.0864 \leq p \leq 0.178$. Another way that we could express this interval is: 0.132 ± 0.0456. The **second value** of 0.0456 is known as the **margin of error**.

The **margin of error** is a **measure** of the **accuracy** of a **confidence interval**. It represents the **amount of error** that we expect to have in our estimate of the population parameter.

The **margin of error** can simply be calculated using the **relevant component** of the **confidence interval formula**:

Margin of Error: $E = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

For instance, suppose Tyler surveyed that from 400 **cars** that passed along the freeway, 58 of them were **electric models of cars**. If Tyler wanted to determine the **margin of error** for a 95% **confidence interval**, he would calculate the **margin of error** as follows:

$$\hat{p} = \frac{58}{400}$$

① $\hat{p} = 0.145$ — ① **Calculate** the **sample proportion**

② $E = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ — ② **State** the **margin of error formula**

③ $E = 1.960 \times \sqrt{\frac{0.145(1-0.145)}{400}}$ — ③ **Substitute** in $z = 1.960$, $\hat{p} = 0.145$ and $n = 400$

④ $E = 0.0345$ — ④ **Calculate** the **margin of error**

After seeing that the **margin of error is larger** than he'd hoped for, Tyler decides to keep surveying cars until he gets a **maximum margin of error** of 0.025. To achieve this, Tyler will need to know **how many cars (n)** he will need to survey. To determine the **value of n**, we **substitute** in the values $E = 0.025$, $z = 2.576$ and $\hat{p} = 0.145$ and then **use our calculators** to **solve** for n:

① $E = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	① **State** the **margin of error**
② $0.025 = 1.960 \times \sqrt{\frac{0.145(1-0.145)}{n}}$	② **Substitute** in $E = 0.025$, $z = 1.960$ and
$n = 762.02$	③ **Use calculator** to **solve** for n and **round up**
③ $\therefore n = 763$	

As we can see, Tyler would need to survey **at least 763 cars** to have a **maximum margin of error** of 0.025. It is also important to note that you should **always round up. Rounding down** to **762 cars** would result in a **margin of error slightly larger** than 0.025.

Like confidence intervals, a **lower confidence interval** will result in a **smaller margin of error,** whereas a **higher confidence interval** will lead to a **larger margin of error**. As shown below, if we assume that $n = 762$ and **calculate** the **margin of error** for a **90% confidence interval**, we get $E = 0.021$ and if we **calculate** the **margin of error** for a **99% confidence interval**, we get $E = 0.0329$.

For 90% confidence interval:

$$E = 1.645 \times \sqrt{\frac{0.145(1-0.145)}{762}}$$

$$E = 0.021$$

For 99% confidence interval:

$$E = 2.576 \times \sqrt{\frac{0.145(1-0.145)}{762}}$$

$$E = 0.0329$$

To solidify our **final concept** of **margin of error**, try working through the example below.

Worked Example 2

Continuing his exploration of sample proportions, Tyler is eager to determine the **margin of error** for **two random samples** he has gathered in relation to the **number of languages spoken by students**.

(a) Of the **1132 students** Tyler surveyed at Mathematics College, **640** of the students were **fluent in another language**. Using this sample, **calculate** the **margin of error** for **a 90% confidence interval**.

$$\hat{p} = \frac{640}{1132}$$

$$\hat{p} = 0.5654$$

$$E = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$E = 1.645 \times \sqrt{\frac{0.5654(1-0.5654)}{1132}}$$

$$E = 0.0242$$

(b) Tyler decides to collect a new sample proportion from the students. If Tyler wants a **maximum margin of error** of 0.01 for a **95% confidence interval**, determine the **size of the sample required**.

$$0.01 = 1.960 \times \sqrt{\frac{0.5654(1-0.5654)}{n}}$$ } **Solve for n using calculator**

$$n = 9439.69$$

$$\therefore n = 9440$$

Congratulations! You have officially **completed all the content** for Year 12 Maths Methods! Take some time to celebrate, reflect on how far you have come, and finish out the year strong by trying your very best to **consolidate all the content** you have learned through the **problem sets**.

SAMPLE PROPORTIONS TOPIC NOTES

Random sampling

A **random sample** is a **subset** of a **statistical population** in which each member of the subset has an **equal probability** of being chosen. This sample is used to **make inferences** about the **whole population**.

When creating a **random sample** it is essential that the population of the sample is **large enough** so it fairly reflects the entire population. A **common rule** is that the **population** of the **random sample** should be **30 or more**, otherwise there is **significant room** for **doubt and error**.

Sampling methods to ensure randomness

For a **random sample**, the goal is to have a **population** that is **truly selected at random**, such that every member of the actual population has an **equal chance** of **being selected**. To achieve this, there are **various sampling methods** that are commonly used.

Random sampling methods				
Systematic random sampling	Stratified sampling	Convenience sampling	Volunteer/self-selection sampling	Capture-recapture sampling

Biases in Sampling

Biases from 'random' sampling can greatly **skew results**.

A sampling method will have **bias** if it **systematically favours an outcome over others**. A source of bias can be intentional or unintentional and it can lead to **inaccurate conclusions** about a population.

There are two general **questions** that can help us to **identify bias**:

① How does the **sampling method** influence the **type/groups of people** that will be **responding** to the **survey**?

② How will the **sampling method** create **restrictions for people to respond** to the survey? Consider **factors** such as **time and location of the survey**, and whether the survey is **voluntary** or **compulsory**.

Sample proportions

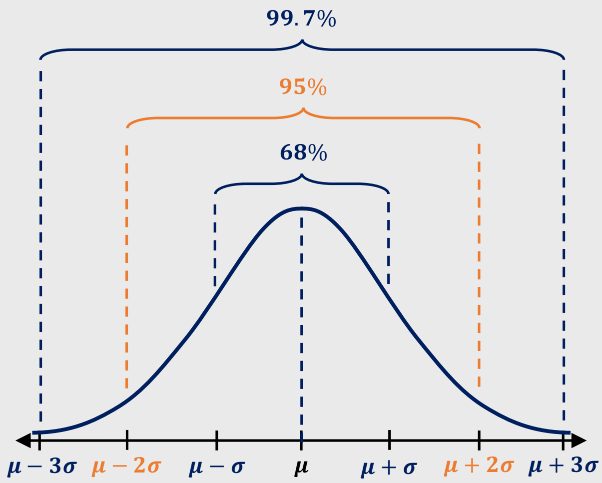
Sample Proportion: $\hat{p} = \frac{X}{n}$	**Mean:** $E(X) = p$	**Standard Deviation:** $\sigma = \sqrt{\frac{p(1-p)}{n}}$

Confidence intervals

A confidence interval is an interval in which we are $x\%$ **certain** that the **true value of p** lies within. For example, if we use a 90% confidence interval to estimate p, out of 10 intervals calculated, we would expect nine of them to contain the true value p.

In a **90% confidence interval**:	In a **95% confidence interval**:	In a **99% confidence interval**:
$P(Z < k) = 0.95$	$P(Z < k) = 0.975$	$P(Z < k) = 0.995$
$k = 1.645$	$k = 1.960$	$k = 2.576$

Important to remember is the **68%, 95%, 99.7% rule** (otherwise known as the **three-sigma rule of thumb**) in which 68%, 95% and 99.7% of the values of a normal distribution lie within **one**, **two** and **three standard deviations of the mean**, respectively.



Margin of error

The **margin of error** is a **measure** of the **accuracy** of a **confidence interval**. It represents the **amount of error** that we expect to have in our estimate of the population parameter.

The **margin of error** can simply be calculated using the **relevant component** of the **confidence interval formula**:

Margin of Error: $E = z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

A **lower confidence interval** will result in a **smaller margin of error**, whereas a **higher confidence interval** will lead to a **larger margin of error**.

Problem Set 13 – Sample Proportions
Progressive Questions

Concept 1
Sample Proportions – Progressive Questions

(6 questions)

Repetitive questions: 1.11 – 1.61

Quest for Leadership

Starring: Janitor Peter, Rupert, Shauna, Fraser, Harriet, Isla and Tom

---

After an exciting journey to Peru, it's time for the final few weeks of school at Mathematics College. Every year, final year students from Mathematics College apply to be representatives on the International Mathematics Board. This board is open to anyone of any age who has completed Year 12 Mathematics. However, this year, the students decide who better to apply than Janitor Peter himself? The students begin their quest to ensure Janitor Peter wins the election!

Sample Proportions: Q1, Q2, Q3, Q4, Q5, Q6, Q7

Repetitive: 1.11-1.61

[5 marks]

1. The students decide to interview some of the local mathematicians to see who they are likely to vote for. To make sure their samples are accurate, help the students answer some **questions about sampling**!

(a) To gauge the distribution of ethnicities residing in Australia, the government decides to send out a census **to all citizens**. Does this represent a **population** or a **sample**? **(1)**

(b) Fraser wants to find out if **high school students** believe energy drinks enhance their **sprinting performance**. To find out, he interviews **his school's 30-person athletics team**. Does this represent a **population** or a **sample**? **(1)**

(c) Fraser also wants to find out what the favourite energy drink of **his school's athletics team** is, so in the same interview he also asks **each person** what their favourite energy drink is. Does this represent a **population** or a **sample**? **(1)**

(d) **One of the year 12 methods classes** was interviewed about their favourite canteen menu item to figure out what sort of food to provide for **a year 12** end of year party. Who is the **population**? **(1)**

(e) What is the main difference between a **sample statistic** and **population parameter**? **(1)**

[6 marks]

2. Rupert forgot to do his sampling, so did it all at the very last minute. There were some issues with his sampling. Help the methods team **identify a source of bias** in each method, **explain why it introduces bias,** and **suggest a way** to **correct** it!

(a) Rupert, interested in the amount of work musicians around the world put in to prepare for their performances, interviews **Perth Symphony Orchestra members** performing on an event night to ask how many hours of practice they do each day. **(2)**

(b) To obtain opinions about the equality of men and women's income, Rupert **surveys** people from **morning until noon** on a **weekday** passing by a local supermarket. **(2)**

(c) In order to find out the **average height** of koalas in Western Australia, Rupert randomly selects 20 of the 50 koalas kept in zoos in WA. **(2)**

[CA] [6 marks]

3. Shauna needs to **estimate** the **number of red pandas** on a reserve. She finds and sneakily places tags on the legs of **25 sleeping red pandas**. After a couple of weeks, carefully making sure that she doesn't count any twice, she randomly spots **40 red pandas** and notes that **10** of them had tags on their legs.

(a) **Estimate** the **number of red pandas** on the reserve. **(2)**

(b) What did you **assume** to do this estimate? **(2)**

(c) Shauna, while removing the tags from all of the red pandas, finds that because it was dark when she tagged them, **10** of the **25** pandas that she tagged were actually small panda shaped logs. Calculate an **estimate** for the **number of red pandas** on the reserve based on this new knowledge. **(2)**

[8 marks]

4. Since Rupert's sampling from question 2 wasn't very good, the students are each assigned a group to survey. **Identify** the **type of sampling** each student performed and whether each method produces a **random sample**.

(a) Shauna, interested to find out whether people in her school like liquorice or not, **interviews** the **first 10 people** she sees that morning. **(2)**

(b) Harriet, curious about others' opinions of the school's cafeteria, **divides** those at the school according to year level and asks **5 students from each year** to rate the cafeteria's food. **(2)**

(c) Isla, intrigued that some students in her class might prefer to write in running writing, takes out a class list and selects **every 3rd name** on the list to ask about their writing style. **(2)**

(d) Fraser, interested about fellow classmates' favourite colours, draws **paper slips** out of a hat containing all the names of those in the class to ask. **(2)**

[CA] [7 marks]

5. Tom decided to make an online poll for students about their favourite representative candidate from a list of four people. Each person could respond with **Janitor Peter**, **Summer**, **Johan** or **Maria**. Out of a total of **40 votes**, **12** voted for **Peter**, **6** voted for **Summer**, **14** voted for **Johan** and **8** voted for **Maria**.

(a) Based on this data, what is our **estimate** of the **probability** of a student voting for peter? **(1)**

(b) If we were to take a bigger sample of **150 students**, how many would we expect to vote for each candidate? **(4)**

(c) Based on the above data, what is the **chance** that a student voted for Maria, given **they did not vote for Peter**. **(2)**

[CA] [9 marks]

6. Unsure of the results of his previous poll, Tom decides to conduct another survey to find out which candidates students prefer to vote for. Out of a total of **40 votes**, 16 voted for **Peter**, 5 voted for **Summer**, 10 voted for **Johan** and 9 voted for **Maria**. Use this data to **estimate** the following:

(a) What is the **chance** a randomly selected student will vote for **Peter**? **(1)**

(b) If a random student is asked whether they would vote for **Peter**, what is the **expected distribution** of their response? **(2)**

(c) What is the **expected distribution** for the **number of people** voting for **Peter** out of a sample of **40**? **(2)**

(d) Tom is getting tired of the long calculations needed for the above distribution, so he wants to find another way to model the **proportion** of people in a sample that will vote for Peter. He proposes that for a **sample size of 40**, he can use a **normal distribution** instead. With reference to the **central limit theorem**, explain whether or not this is reasonable. **(2)**

(e) Would it be reasonable to use a normal distribution if his sample size was **15** instead? **(2)**

[CA] [11 marks]

7. Tom and Isla, taking a break from conducting surveys, are playing with **two fair six sided dice**. The two dice are rolled with the numbers on top then added together. Tom rolls the two dice **50 times**, producing a **total less than seven 23 times**.

(a) Determine the **population proportion** p of rolling a **total less than 7**. **(1)**

(b) Determine the **sample proportion** $\hat{p}$ of rolling a **total less than 7**. **(1)**

(c) Calculate the **mean** and **standard deviation** of the **sample proportion** $\hat{p}$. **(2)**

(d) State the approximate distribution of the sample proportion $\hat{p}$, and **give a reason** why this is an appropriate approximation. **(2)**

(e) Calculate the **probability** of obtaining a value of $\hat{p}$ that is greater than **0.6** from a random sample. **(2)**

(f) Isla takes her turn rolling the two more dice that she thinks might not be fair (i.e. might be more likely to land on certain numbers than others). She produces a **total less than 7** in **56 out of 100 rolls**.

  a. Now determine Isla's **sample proportion** $\hat{p}$ for rolling a **total less than 7**. **(1)**

  b. State the approximate distribution of the sample proportions $\hat{p}$ of Isla's dice. **(2)**

Concept 2

Sample Proportions – Progressive Questions

(9 questions)

Repetitive questions: 2.11 – 2.121

One Last Victory

Starring: Janitor Peter, Rupert, Tom, Fraser and Maria

Now that the students are confident Janitor Peter will win the election from their sampling, it's time to advertise for him!

Confidence Intervals and Margins of Error: Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q8, Q9

Repetitive: 2.11-2.121

[CA] [15 marks]

1. Rupert wants to improve his sampling skills to more effectively advertise to the right people. Help him calculate the **point estimates for p**, and describe the **expected distributions** of similar points estimates in the given scenarios, including their **means** and **standard deviations**.

(a) Imagine a spinner divided into **four quadrants**: red, yellow, green, and blue. After spinning the spinner **120** times, it landed on the **red quadrant 35 times**. **(3)**

(b) A **six-sided die** is rolled **79 times** and the **number 5** lands face-up **10 times**. **(3)**

(c) A survey of **150 children** found that **89** preferred eating apples compared to bananas. **(3)**

(d) A number generator can generate the numbers **1 to 10**. Out of **60 tries**, the **number 7** is generated **8 times**. **(3)**

(e) In a Year 12 Methods class, **27** out of the **48** students said that Sample Proportions was their favourite Methods topic. **(3)**

[CA] [10 marks]

2. Tom, still focused on his surveys, wants to know how accurate his estimates are. Help him learn how to calculate **margins of error** by finding them for the relevant **confidence intervals** in the following scenarios:

(a) What is the **margin of error** for a **90%** confidence interval estimating the proportion of people who drive black cars? A survey found that out of **200** people, **70** of them drive black cars. **(2)**

(b) Calculate the **margin of error** for a **99%** confidence interval estimating the proportion of people who like broccoli. In a survey of **60** people, it was found that **17** like broccoli. **(2)**

(c) Determine the margin of error for a **95%** confidence interval estimating the proportion of people who drink apple juice. A survey of **140** people revealed that **60** of them drink apple juice. **(2)**

(d) Find the **margin of error** for an **87%** confidence interval estimating the proportion of people who catch the bus to school. From a survey of **200** people, it was observed that **70** of them catch the bus to school. **(2)**

(e) Calculate the **margin of error** for a **93%** confidence interval estimating the proportion of people who participate in the athletics carnival. Based on a survey of **180** people, it was discovered that **20** of them did not participate in the carnival. **(2)**

[CA] [15 marks]

3. Rupert, after creating surveys to find more information about the people they are advertising to, wants to know how accurate his estimates are. For the following scenarios, construct the relevant **confidence interval**.

(a) What is the **90%** confidence interval estimating the proportion of people who support longer lunch breaks? A survey revealed that out of **180** people, **160** of them support longer lunch breaks. **(3)**

(b) Calculate the **99%** confidence interval estimating the proportion of people who think canteen prices are expensive. Based on a survey of **200** people, it was found that **90** of them think canteen prices are expensive. **(3)**

(c) Determine the **95%** confidence interval estimating the proportion of people who hate sardines. In a survey of **70** people, it was discovered that **21** of them hate sardines. **(3)**

(d) Find the **92%** confidence interval estimating the proportion of people who get dropped at school in the morning. From a survey of **64** people, it was observed that **16** of them get dropped at school. **(3)**

(e) Calculate the **85%** confidence interval estimating the proportion of people who participate in the swimming carnival. A survey of **180** people revealed that **50** of them did not participate in the carnival. **(3)**

[CA] [12 marks]

4. Maria needs to conduct surveys for Janitor Peter but a specific level of accuracy is required. **Calculate** the **sample size** needed to achieve the desired **margin of error** in the following scenarios:

(a) In a previous survey, **160** out of **180** students supported longer lunch breaks. Maria wants to conduct another survey to determine if the proportion has changed. What sample size does she need to ensure that the width of her confidence interval is no more than **0.02** for a **99% confidence interval**? **(3)**

(b) Based on a previous survey, **250** out of **300** students believed that canteen prices are too expensive. Maria plans to conduct another survey to investigate any changes in the proportion of people who believe this. What sample size does she need to ensure that the width of her confidence interval is no more than **0.01** for a **90% confidence interval**? **(3)**

(c) From a previous survey, **16** out of **64** people get dropped off at school. Maria aims to determine if that proportion has changed. What sample size does she need to ensure that the margin of error is no more than **0.015** for a **95% confidence interval**? **(3)**

(d) A previous survey found that out of **180** students, **50** of them did not attend their swimming carnival. Maria intends to determine whether this proportion has changed or not. What sample size does she need to ensure that the margin of error is no more than **0.04** for a **68.27% confidence interval**? **(3)**

[4 marks]

5. Fraser, known for zoning out in class, seeks help from Janitor Peter to catch up on his studies. He knows that there was a certain value for $\boldsymbol{p}$ that gave the **widest confidence interval** for **any confidence level**. Show how Janitor Peter might use a **calculus method** to determine this value of p that **maximises** the margin of error.

[CA] [6 marks]

6. The students want to investigate Janitor Peter's competition in the election, so they calculated the **confidence intervals** based on data from surveys of a few other participants on the leader board. The students want to do a thorough investigation and so they decide to check how **uncertain** the results of the survey are. Given the following **confidence intervals** the students calculated, find the **margin of error**:
   (a) $[0.7453, 0.9542]$ **(2)**
   (b) $[0.2672, 0.9810]$ **(2)**
   (c) $[0.3593, 0.9165]$ **(2)**

[CA] [7 marks]

7. The students then ran a survey to investigate whether Janitor Peter's competitors had good public speaking skills. The students sent out surveys asking whether or not certain competitors were good public speakers. They then calculated **confidence intervals** based on the results, however they wanted to see **how many students** answered each survey. Given the **90% confidence intervals**, find the **sample size**.
   (a) 0.256 ± 0.05 **(3)**
   (b) $[0.369, 0.409]$ **(4)**

[CA] [12 marks]

8. Based on a recent report, Janitor Peter is leading the election! The students decide that they should investigate the **proportion** of Peter's votes that are coming from **high schoolers**. They **randomly select** a sample of $\mathbf{45}$ **high schoolers** and find that $\mathbf{27}$ of these students **voted for Peter**.
   (a) Determine the **sample proportion** $\hat{\boldsymbol{p}}$ that voted for Peter. **(1)**
   (b) What is the **expected distribution** for $\hat{\boldsymbol{p}}$? **(2)**
   (c) Based on this estimate, what is the **probability** of obtaining a $\hat{\boldsymbol{p}}$ **value less than** $\mathbf{0.46}$? **(2)**
   (d) Calculate the **z score** for this scenario to construct a $\mathbf{90\%}$ **confidence interval**. **(1)**
   (e) Using your answer from **part (d)**, calculate the $\mathbf{90\%}$ **confidence interval** for this sample. **(3)**
   (f) Justify whether the students' results are **valid** or not. (**Hint:** consider whether the sample is **unusual** and if so, give **reasons** as to why it is or isn't unusual). **(3)**

[CA] [14 marks]

9. Finally, the results are in, and Janitor Peter won the election! During the celebration of his victory, Rupert playfully challenges Peter to a statistical inference competition. For the following scenarios, help him by identifying the **distribution** of the variables and state their **parameters** or their **estimates**. Justify whether the variables can be modelled by the **normal distribution**, and if so, state the **approximate normal distribution** and its **parameters**.

(a) The number of heads when a **fair coin** is tossed 12 times. **(3)**

(b) The number of votes for dogs in a classroom vote on whether cats or dogs are better pets, if 16 students out of a total of 29 students **voted for dogs**. **(3)**

(c) The number of people who preferred to travel by bus, if in a survey on public transport, 105 participants were asked whether they preferred to travel by car or by bus and 32 participants reported that they **preferred to travel by bus**. **(4)**

(d) In a school election to vote for the new School Captain out of Candidates Alice and Bob, 127 students out of the total 300 **voted for Alice**. **(4)**

Problem Set 13 – Sample Proportions
Repetitive Questions

Concept 1
Sample Proportions – Repetitive Questions
(5 questions)

Sample Proportions: Qs 1.11, 1.21, 1.31, 1.51, 1.61

[7 marks]

1.11 A Year 12 Methods class is learning about sample proportions. Help them answer these questions.

(a) Joey interviews a Year 11 class at his school to find out which canteen food is most loved there. Is this class a **population** or a **sample**? **(2)**

(b) Joey also asks the same class who their favourite teacher is to find out which teacher the year 11s at her school like best. What is the **population** for this sample? **(1)**

(c) In a recent **census** conducted by the Western Australian government, it was found that 4.5% of people identified as Aboriginal or Torres Strait Islander. Would this be classified as a **sample statistic** or a **population parameter**? Remember to justify your answer. **(2)**

(d) Mr Smith, to determine **ratio of boys to girls** joining the school Athletics Day carnival, picked the **first 30 students** to arrive and found out that it was 3:2. Would this be classified as a **sample statistic** or a **population parameter**? Remember to justify your answer. **(2)**

[6 marks]

1.21 Angela was absent for the class in which they discussed sampling, so she is stuck on the homework about sampling bias. Help her **determine any biases** in her sampling and **suggest a way to correct it!** Any reasonable answers are accepted.

(a) Her classroom teacher, Ms Bose, put up a poster **in the classroom** inviting sign ups to participate in an experiment to determine who can type the fastest **in her class**. **(2)**

(b) To investigate the number of people who go to his **local swimming pool** every week, Benjamin surveys **athletes** coming from the **nearby secondary school.** **(2)**

(c) Angela and Rupert meet up to interview those walking out of a train station from **morning until noon** on a **weekday** to ask what they think about a café in the station. **(2)**

[CA] [6 marks]

1.31 Steve, who is curious about the **number of turtles** in a large aquarium, randomly selects **20 turtles** and gently marks their shells. After a week, he captures and then checks the shells of **30 turtles** at random and notes that **15** of them had marked shells.

(a) Use this to **estimate** the **number of turtles** in the aquarium. **(2)**

(b) What did you **assume** to make this estimate? **(2)**

(c) Steve realises that he accidentally marked some pieces of wood instead of turtles because he forgot his glasses on the day that he made the marks. After some searching, he finds that he marked **5 pieces of wood** in total. Assuming that he found all of the wood, calculate the new estimate for the number of turtles in the aquarium. **(2)**

[3 marks]

1.41 Identify the **type of sampling** used in the following situations.

(a) Shauna, interested to find out whether people in her school like chocolate or not, **interviews** every **4th person** that she sees during the day. **(1)**

(b) Harriet, curious about others' opinions of the school's cafeteria, **divides** the school **according to year level** and asks **20 students in individual grades** to rate the cafeteria's food. **(1)**

(c) Fraser, interested about fellow classmates' favourite foods, creates **surveys** to leave in the classroom for people to complete. **(1)**

[CA] [10 marks]

1.51 Joey has bought a **spinner** with **four colours** on it: red, green, blue and yellow. The box says it has a **50%** chance to land on blue, **20%** chance to land on green, **18%** chance to land on red and **12%** chance to land on yellow. Joey wants to test this and so spins it **500** times:

(a) **How many times** would joey expect it to land on **each** of the different colours? **(2)**

(b) A generated sample is shown below. What are the **point estimates** of the **probabilities** based on this sample? **(2)**

	Blue	Green	Red	Yellow
Frequency	247	101	97	55

(c) Joey is particularly interested in the **percentage of spins** that **land on green**. What is the **expected distribution** for the number of spins that **land on green**? **(2)**

(d) Joey is getting tired of the calculations needed for the above distribution, so he wants to find another way to model the number of spins that will land on green. He proposes that for a **sample size of 500**, he can use a **normal distribution** instead. With reference to the **central limit theorem**, explain whether or not this is reasonable. **(2)**

(e) Would it be reasonable to use a normal distribution if his **sample size was 15** instead? **(2)**

[CA] [11 marks]

1.61 Tommy and Isla are playing with **two fair six sided dice**. The dice are rolled with the numbers on top then added together. Tommy rolls the dice **200 times**, producing a **total more than seven 78 times**.

(a) Determine the **population proportion** p of rolling a **total more than 7**. **(1)**

(b) Determine the **sample proportion** $\hat{p}$ of rolling a **total more than 7**. **(1)**

(c) Calculate the **mean** and **standard deviation** of the **sample proportions** $\hat{p}$. **(2)**

(d) State the approximate distribution of the sample proportions $\hat{p}$**, and give a reason why this is an appropriate approximation.** **(2)**

(e) Calculate the probability of obtaining a $\hat{p}$ **value greater than 0.5** from a random sample. **(2)**

(f) Isla takes her turn rolling two more dice that she thinks might not be fair (i.e. might be more likely to land on certain numbers than others). She produces a **total more than 7** in **56 out of 100 rolls**.

   a. Now determine Isla's **sample proportion** $\hat{p}$ for rolling a **total less than 7**. **(1)**

   b. State the **approximate distribution** of the sample proportions $\hat{p}$ of Isla's dice **(2)**

Concept 2

Sample Proportions – Repetitive Questions

(13 questions)

Confidence Intervals and Margins of Error: Qs 2.11, 2.21, 2.31, 2.41, 2.51 2.61, 2.71, 2.72, 2.81, 2.91, 2.101, 2.111, 2.121

[CA] [5 marks]

2.11 A student asks 100 **students** from their school what their **favourite colour** is.

(a) Calculate the **sample proportion** of students with the favourite colour orange if 22 **students** said that that was their favourite colour. **(1)**

(b) Determine the **standard deviation** for the **sample proportion** from **part (a)**. **(1)**

(c) The same student now asks another 100 **students** from a different school the same question. If the **sample proportion from part (a)** is used to estimate the **proportion of all students** with the favourite colour orange, **is it likely** that this new school will have 35 **students** with that favourite colour? (**Hint:** calculate how many **standard deviations** this **probability** is from the **mean**). **(3)**

[CA] [4 marks]

2.21 Determine the **point estimate** for the **population proportion** given the **provided data**.

(a) From a survey of 100 **people**, 13 said they were **left-handed**. **(1)**

(b) From a survey of 1500 **people**, 978 possessed a **driver's licence**. **(1)**

(c) From a survey of 600 people, 257 spoke a **language other than English**. What would the **point estimate** be for those **who only spoke English**? **(1)**

(d) From a survey of 1440 people, it was determined that 1037 had **brown eyes**. What is the **point estimate** for people **who do not have brown eyes**? **(1)**

[CA] [12 marks]

2.31 Determine the **confidence interval** for the following scenarios.

(a) In a survey of 100 **people**, it was found that 50% agreed that pineapple should not belong on pizza. Find the 99% **confidence interval** for this **population proportion**. **(3)**

(b) In a survey of 250 **people**, 32% were satisfied with the online service of an internet provider. Find the 95% **confidence interval** for this **population proportion**. **(3)**

(c) In a survey of 500 **people**, 401 said that they played team sports in their spare time. Find the 90% **confidence interval** for this **population proportion**. **(3)**

(d) A survey was undertaken where 2000 **people** were asked about a local government proposal to change the speed limit around residential roads, with 1420 **people agreeing**. Find the 90% **confidence interval** for the proportion of those **who disagreed**. **(3)**

[CA] [6 marks]

2.41 Find the **confidence interval** for each of the following scenarios, being sure to indicate the **point estimate**:

(a) In a sample of **120 Year 12 students**, **75** indicated that they work part-time. Use this information to find a **90% confidence interval** for the proportion of students who work part-time. **(2)**

(b) A manufacturing company proclaimed that a sample of **260** out of **320 computers** were able to run the latest operating system. Use this information to find a **95% confidence interval** for the proportion of computers which can run the latest operating system. **(2)**

(c) In a pizza shop on a certain day, it was reported that out of **3800 orders**, **1500** of the orders were a pineapple pizza. Use this information to find a **99% confidence interval** for the proportion of pineapple pizza orders. **(2)**

[CA] [4 marks]

2.51 Consider a **random variable $\hat{p}$** that represents the proportion of right-handed people in a sample. $\hat{p} \sim Bin(50, p)$ and a point estimate for p is $\frac{9}{10}$.

(a) Calculate the **90% confidence interval** based on this point estimate. **(2)**

(b) Would it be correct to say that our **90% confidence interval** must be **narrower** than the **95% interval**? **Explain why or why not**. **(2)**

[CA] [5 marks]

2.61 An experiment that aims to find the time it takes for an unrefrigerated bottle of milk to spoil found that after 2 hours, **40** out of the **50 bottles of milk** they had were considered spoilt. Based on the data collected:

(a) Calculate the **margin of error** for the **95% and 99% confidence intervals**. **(3)**

(b) Based on your values in (a), calculate both **confidence intervals**. **(2)**

[CA] [5 marks]

2.71 Johnny owns a factory manufacturing cars, and he wants to test the safety of the vehicles his factory makes, so he takes **70 factory-new cars** and examined them. It turns out **3 of his cars out of 70** had broken airbags!

(a) Johnny's factory advisor tells him that he is **95% confident** that this result is the true proportion of cars with broken airbags. Based on this assumption, what is the **margin of error to 4 decimal places** for this sample of cars? **(3)**

(b) Using the information given and **backing it up with the calculated margin of error**, justify **whether or not** Johnny should improve the safety of the vehicles he produces. **(2)**

[CA] [5 marks]

2.72 A company producing laptops claims that **98% of laptops sold** will not have problems with overheating. A reviewing body, wanting to test their claims, took a sample of **300 laptops** and found that **11 out of the 300** displayed problems with overheating.

(a) Construct a **97% confidence interval** for the sample. **(3)**

(b) **Using the calculated confidence interval**, comment on **whether or not** the company's claim is true. **(2)**

[CA] [2 marks]

2.81 Evan is a quality inspector for electronic devices. To test whether a brand of phone charger was faulty, he took a sample of chargers and constructed a **90% confidence interval** for the sample. If the true proportion of faulty chargers is **not within his confidence interval**, does this mean that his confidence interval was wrong?

[CA] [8 marks]

2.91 Calculate a **confidence interval** with the specified **level of confidence** for the **given sample data**.

(a) A **98% confidence interval** for a sample with sample size $n = 42$ and sample proportion $\hat{p} = 0.23$ **(2)**

(b) A **96% confidence interval** for a sample with sample size $n = 112$ and sample proportion $\hat{p} = 0.04$ **(2)**

(c) A **93% confidence interval** for a sample with sample size $n = 63$ and sample proportion $\hat{p} = 0.56$ **(2)**

(d) An **89% confidence interval** for a sample with sample size $n = 79$ and sample proportion $\hat{p} = 0.17$ **(2)**

[CA] [6 marks]

2.101 Sam calculates the confidence interval for results of maths tests for three different classes. Given the following **confidence intervals** for each class, find the **margin of error**:

(a) Class A: $[0.1245, 0.3261]$ **(2)**

(b) Class B: $[0.2450, 0.8755]$ **(2)**

(c) Class C: $[0.5543, 0.9538]$ **(2)**

[CA] [8 marks]

2.111 Rina found some results for an old survey which asked the students at two different schools whether they took public transport to school. However, the results only showed a different **confidence interval** for each school. Rina wanted to calculate **how many students** from each school answered that survey. Given the following **confidence intervals**, find the **sample size**:

(a) School A's **95% confidence interval** for the portion of students taking public transport was $[0.578 \pm 0.04]$. Find the **sample size**. **(4)**

(b) School B's **99 % confidence interval** for the portion of students taking public transport was $[0.03 \pm 0.01]$. Find the **sample size**. **(4)**

[CA] [5 marks]

2.121 Ben runs a lemonade stand and he wants to expand his menu. One day, he decides to survey the customers and asks if he should add pink lemonade to the menu. Out of the **200 customers** Ben received that day, **105 of them said yes** to pink lemonade. Given this information, answer the following questions:

(a) **Determine** the **sample proportion**, $\hat{p}$. **(1)**

(b) Ben wants to calculate a **95% confidence interval**. What is the **Z-score**? **(1)**

(c) For the **95% confidence interval**, what is the **margin of error?** **(1)**

(d) What does the **95% confidence interval** represent. Should Ben add pink lemonade to the menu? **Explain.** **(2)**